Εθνικό Μετσόβιο Πολυτεχνείο

Τμήμα Εφαρμοσμένων μαθηματικών και φυσικών επιστημών

ΜΠΣ Εφαρμοσμένες μαθηματικές επιστήμες

Θέμα

Συνελίξεις κατανομών με βαριά ουρά

ΛΟΣΙΔΗΣ ΣΩΤΗΡΙΟΣ

ΔΙΠΛΩΜΑΤΙΚΗ ΔΙΑΤΡΙΒΗ

Υποβλήθηκε στο

Τμήμα εφαρμοσμένων Μαθηματικών και Φυσικών επιστημών

Τριμελής επιτροπή

Λουλάκης Μιχαήλ

Καρώνη Χρυσηίς

Πολίτης Κωνσταντίνος

Αθήνα

Απρίλιος 2014

Στη Σοφία της ζωής μου

Περιεχόμενα

Γραφικές παραστάσεις

Σχήμα 1γραφική απεικόνιση συνάρτησης πυκνότητας πιθανότητας εκθετικής κατανομής με $\mu = 10$

Σχήμα 2 :Γραφική παράσταση συνάρτησης πυκνότητας πιθανότητας της κανονικής κατανομής για διαφορετικές διασπορές σ

Σχήμα 3:*Γραφική απεικόνιση κατανομής Γάμμα με διαφορετικές τιμές παραμέτρου κλίμακας*

Σχήμα4: (αριστερά) συνάρτηση κατανομής με παράμετρο σχήματος μικρότερη της μονάδας και (δεξιά) συνάρτηση κατανομής με παράμετρο σχήματος μικρότερη της μονάδας

Σχήμα5: (αριστερά) συνάρτηση κατανομής με παράμετρο σχήματος ίση με 5 και (δεξιά) συνάρτηση κατανομής με παράμετρο σχήματος ίση με 2

Σχήμα6: συνάρτηση κατανομής Wiebull με παράμετρο με διαφορετικά επίπεδα παραμέτρου κλίμακας

Σχήμα7: *γραφική απεικόνιση κατανομών Gamma με διαφορετικές τιμές της παραμέτρου α*

Σχήμα 8*: συνάρτησης πυκνότητας της κατανομής Γάμμα για διάφορες τιμές της παραμέτρου σχήματος α.*

Σχήμα 9 : *Inverse Gaussian*

Κώδικας R

Για τις γραφικές απεικονίσεις του τυχαίου περιπάτου στη σελίδα 75 χρησιμοποιήθηκαν οι κώδικες

```
# Generate k random walks across time {0, 1, ... , T}
T <- 100
k <- 250
initial.value <- 10
GetRandomWalk <- function() {
  # Add a standard normal at each step
  initial.value + c(0, cumsum(rnorm(T)))
}
# Matrix of random walks
values <- replicate(k, GetRandomWalk())
# Create an empty plot
dev.new(height=8, width=12)
plot(0:T, rep(NA, T + 1), main=sprintf("%s Random Walks", k),
     xlab="time", ylab="value",
     ylim=10 + 4.5 * c(-1, 1) * sqrt(T))
mtext(sprintf("%s%s} with initial value of %s",
              "Across time {0, 1, ... , ", T, initial.value))
for (i in 1:k) {
  lines(0:T, values[ , i], lwd=0.25)
}
for (sign in c(-1, 1)) {
  curve(initial.value + sign * 1.96 * sqrt(x), from=0, to=T,
        n=2*T, col="darkred", lty=2, lwd=1.5, add=TRUE)
}
legend("topright", "1.96 * sqrt(t)",
       bty="n", lwd=1.5, lty=2, col="darkred")
savePlot("random_walks.png")
```

```
c <- 1
d <- 2

a <- -2
b <- 3.5

ll <- pnorm(a, c, d)
ul <- pnorm(b, c, d)

x <- qnorm( runif(3000, ll, ul), c, d )
hist(x)
range(x)
mean(x)
sd(x)
plot(x, type='l')
```

Κεφάλαιο 1

Εισαγωγή βασικών εννοιών

1.1 Εισαγωγή

Η ανάγκη μαθηματικής περιγραφής και μοντελοποίησης συστημάτων τα οποία εξελίσσονται χρονικά κατά τρόπο που περιέχει την τυχαιότητα (randomness , stochasticity) και όχι κατά τρόπο ντετερμινιστικό (προσδιοριστικό) οδήγησε στην ανάπτυξη της θεωρίας των στοχαστικών διαδικασιών.

1.2 Στοχαστική διαδικασία

Στοχαστική διαδικασία ή στοχαστική ανέλιξη όπως πολλές φορές συναντάται στη βιβλιογραφία είναι μια οικογένεια τυχαίων μεταβλητών $\{X(t): t \in T\}$ πάνω σε ένα κοινό **χώρο πιθανότητας** $(\Omega, F, \mathbb{P})$

1.3 Χώρος πιθανότητας

Ονομάζεται μια τριάδα$(\Omega, F, \mathbb{P})$ όπου

- $\Omega \neq \emptyset$ δηλαδή το Ω είναι ένα μη κενό υποσύνολο
- $\boldsymbol{F}$ είναι σ-άλγεβρα υποσυνόλων του Ω
- $\mathbb{P}: \boldsymbol{F} \to [\mathbf{0}, \mathbf{1}]$ μια σύνολο συνάρτηση ώστε :
    i. $P(A) \geq 0 \; για \; κάθε \; A \in F$
    ii. $P(\Omega) = 1$
    iii. Για κάθε ακολουθία $\{E_n\}_{n \in N}$ στοιχείων της $\boldsymbol{F}$ με $E_n \cap E_m = \emptyset$ για κάθε $m \neq n \in N$ ισχύει $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$
    Η ιδιότητα $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$ είναι γνωστή στη θεωρία μέτρου ως αριθμήσιμη προσθετικότητα (countable additivity)
    Το σύνολο Ω ονομάζεται **δειγματοχώρος** και τα υποσύνολα του Ω που ανήκουν στην $\boldsymbol{F}$ ονομάζονται **ενδεχόμενα.**

Σημειώνεται πως αν δεν ισχύει η δεύτερη από τις τρείς ιδιότητες του παραπάνω ορισμού τότε μιλάμε για **χώρο μέτρου** και όχι χώρο πιθανότητας.

1.4 Χώρος μέτρου

Ένας χώρος μέτρου (*measure space)* είναι μια τριάδα $(\boldsymbol{\Omega}, \boldsymbol{F}, \boldsymbol{\mu})$ όπου

- $\Omega \neq \emptyset$ δηλαδή το Ω είναι ένα μη κενό υποσύνολο
- $\boldsymbol{F}$ είναι σ-άλγεβρα υποσυνόλων του Ω

- $\boldsymbol{\mu}: \boldsymbol{F} \rightarrow [0, \infty]$ μια σύνολοσυνάρτηση ώστε :

  I. $\mu(\emptyset) = 0$
  II. Για κάθε ακολουθία $\{E_n\}_{n \in N}$ στοιχείων της $\boldsymbol{F}$ με $E_n \cap E_m = \emptyset$ για κάθε $m \neq n \in N$ ισχύει $\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$

1.5 Σύνολο δεικτών

Το σύνολο T λέγεται σύνολο δεικτών της διαδικασίας ή παραμετρικός χώρος .Αν το σύνολο T είναι αριθμήσιμο τότε η στοχαστική διαδικασία $\{X(t): t \in T\}$ λέγεται διαδικασία διακριτής παραμέτρου ή διαδικασία διακριτού χρόνου. Αν το σύνολο T είναι μη αριθμήσιμο τότε η διαδικασία $\{X(t): t \in T\}$ λέγεται διαδικασία συνεχούς παραμέτρου ή διαδικασία συνεχούς χρόνου.

Ο παραμετρικός χώρος δεν είναι κατ ανάγκη μονοδιάστατος. Για παράδειγμα η τυχαία μεταβλητή $X(t)$ μπορεί να είναι διδιάστατη και να αφορά την θερμοκρασία και την υγρασία ή μπορεί ακόμη να είναι και πολυδιάστατη παραδείγματος χάρη οι γεωγραφικές συντεταγμένες ενός σημείου σε συνδυασμό με τον χρόνο.

1.6 Χώρος καταστάσεων

Το σύνολο των δυνατών τιμών των τυχαίων μεταβλητών $X(t)$ $t \in T$ ονομάζεται χώρος καταστάσεων και συνήθως συμβολίζεται με S.Σημειώνουμε σε αυτό το σημείο πως όπως ο παραμετρικός χώρος έτσι και ο χώρος καταστάσεων δεν είναι κατ΄ ανάγκη μονοδιάστατος .

Ανάλογα με το αν ο χώρος καταστάσεων και ο παραμετρικός χώρος είναι συνεχής ή διακριτός διακρίνουμε τις παρακάτω κατηγορίες στοχαστικών διαδικασιών:

- Συνεχής με συνεχή παραμετρικό χώρο.
- Συνεχής με διακριτό παραμετρικό χώρο.
- Διακριτή με συνεχή παραμετρικό χώρο.
- Διακριτή με διακριτό παραμετρικό χώρο.

1.7 Παραδείγματα

Παραδείγματα των παραπάνω κατηγοριών είναι:

Παράδειγμα 1ο Έστω $X(t)$ η τιμή της μετοχής μιας εισηγμένης στο χρηματιστήριο εταιρείας τη χρονική στιγμή t. Τότε η στοχαστική διαδικασία $\{X(t): t \in T\}$ είναι ***μια συνεχής στοχαστική διαδικασία με συνεχή παραμετρικό χώρο.***

Παράδειγμα 2ο Έστω $X(t)$ η τιμή της μετοχής μιας εισηγμένης στο χρηματιστήριο εταιρείας στο τέλος της τελευταίας εργάσιμης ημέρας κάθε εβδομάδας του μήνα Ιανουαρίου, $t \in \{1,2,3,4\}$ μιας και ο μήνας έχει τέσσερις εβδομάδες. Τότε η στοχαστική διαδικασία $\{X(t): t \in T\}$ είναι ***μια συνεχής στοχαστική διαδικασία με διακριτό παραμετρικό χώρο.***

Παράδειγμα **3º** Ο αριθμός των μετοχών με ανοδική κίνηση σε συγκεκριμένη ημέρα αποτελεί ένα παράδειγμα ***διακριτής στοχαστικής ανέλιξης*** $\{\boldsymbol{X}(\boldsymbol{t}): \boldsymbol{t} \in \boldsymbol{T}\}$ ***με συνεχή παραμετρικό χώρο.***

Παράδειγμα 4º Έστω η τυχαία μεταβλητή X(t) παριστάνει το μέσο όρο της τιμής της μετοχής κατά τη διάρκεια μιας ημέρας και εμείς θέλουμε να μελετήσουμε σε μια περίοδο 100 ημερών σε πόσες από αυτές τις μέρες ο μέσος όρος της μετοχής ήταν πάνω από ένα συγκεκριμένο κατώφλι. Σε αυτή την περίπτωση τόσο ο χρόνος ο οποίος μετράται σε ημέρες είναι διακριτή μεταβλητή όσο και ο αριθμός των μετοχών θα είναι διακριτή μεταβλητή επίσης ,πρόκειται λοιπόν για μια περίπτωση ***διακριτής στοχαστικής ανέλιξης σε διακριτό παραμετρικό χώρο.***

Τα περισσότερα παραδείγματα ανελίξεων με ενδιαφέρον στη θεωρία κινδύνου γενικά αλλά και στη θεωρία χρεοκοπίας ποιο συγκεκριμένα αφορούν στοχαστικές ανελίξεις σε συνεχή χρόνο με διακριτές τιμές. Το απλούστερο παράδειγμα μιας τέτοιας ανέλιξης είναι η ανέλιξη *Poisson.*

1.8 Μαρκοβιανή στοχαστική ανέλιξη

Μια στοχαστική ανέλιξη $\{X(t): t \in T\}$ λέγεται Μαρκοβιανή εάν για κάθε $n < m$ $και\, για\, κάθε$ $t_0 < t_1 < t_2 \ldots \ldots . < t_n < \cdots < t_m$ η δεσμευμένη κατανομή των

$X(t_{n+1}), X(t_{n+2}), \ldots \ldots, X(t_m)$ δεδομένων των $X(t_0), X(t_1), \ldots, X(t_n)$ εξαρτάται μόνο απο την $X(t_n)$.

Δηλαδή μια Μαρκοβιανή στοχαστική ανέλιξη είναι μια στοχαστική ανέλιξη στην οποία η γνώση του παρόντος προσδιορίζει από πιθανοθεωρητική άποψη το μέλλον ενώ το παρελθόν δεν παίζει κανένα ρόλο. Δηλαδή από όλο το παρελθόν μιας Μαρκοβιανή στοχαστικής ανέλιξης μόνο η πρόσφατη κατάσταση καθορίζει το μέλλον.

1.9 Στοχαστική διαδικασία ομοιογενών προσαυξήσεων

Μια στοχαστική διαδικασία $\{X(t): t \in T\}$ καλείται στοχαστική διαδικασία ομοιογενών προσαυξήσεων αν για κάθε $t_1 < t_2 \, με \, t_1 > 0, t_2 > 0$ η κατανομή της τυχαίας μεταβλητής $X(t_2) - X(t_1)$ εξαρτάται μόνο από τη διαφορά $t_2 - t_1$**[1].**

1.10 Διαδικασία *Poisson*

Η απλούστερη ομοιογενής Μαρκοβιανή στοχαστική διαδικασία σε διακριτό χώρο και σε συνεχή παραμετρικό χώρο είναι η ομογενής διαδικασία *Poisson.*

Για τη διαδικασία Poisson δώσουμε δύο ισοδύναμους ορισμούς

Η στοχαστική διαδικασία $\{X(t) : t \geq 0\}$ καλείται ομοιογενής διαδικασία Poisson εάν:

Ορισμός 1ος :

- $X(0) = 0$ δηλαδή τη στιγμή έναρξης των εργασιών δεν έχει συμβεί καμία ζημιά.
- Η $\{X(t): t \geq 0\}$ έχει ανεξάρτητες προσαυξήσεις .

- Ο αριθμός των ενδεχομένων σε κάθε διάστημα μήκους t ακολουθεί την κατανομή Poisson με μέση τιμή λt **[2]**

Εναλλακτικά

Ορισμός 2ος

- $X(0) = 0$ δηλαδή τη στιγμή έναρξης των εργασιών δεν έχει συμβεί καμία ζημιά καθώς επίσης η τυχαία μεταβλητή $X(t)$ παίρνει μη αρνητικές ακέραιες τιμές ,δηλαδή η τυχαία μεταβλητή παίρνει τιμές στο $\{0,1,2,3, \dots\}$
- Η $\{X(t): t \geq 0\}$ έχει ομογενείς και ανεξάρτητες προσαυξήσεις
- Η πιθανότητα δηλαδή να έχουμε παραπάνω από ένα ζημιογόνο συμβάν είναι πολύ μικρή και τείνει στο μηδέν .

$$P(X(t) \geq 2) = o(t) = P(X(t) > 1)$$

Σημειώνουμε πως η συνάρτηση $o(t)$ έχει την ιδιότητα

$$\lim_{t \to 0} \frac{o(t)}{t} = 0$$

δηλαδή ο αριθμητής φτάνει γρήγορα στο μηδέν σε σχέση με τον παρονομαστή.

Η πιθανότητα ατυχήματος σε ένα μικρό χρονικό διάστημα δεν εξαρτάται από την αρχή του χρονικού διαστήματος αλλά από το μήκος λ του διαστήματος.

$$P(X(t) = 1) = \lambda t + o(t)$$

1.11 Στοχαστική διαδικασία ανεξάρτητων προσαυξήσεων

Μια στοχαστική ανέλιξη $\{X(t): t \geq 0\}$ λέγεται ανεξάρτητων προσαυξήσεων εάν για κάθε $n\ και\ κάθε\ t_0 < t_1 < t_2 < \cdots < t_n$ οι διαφορές

$$Y(t_j) = X(t_j) - X(t_j) \quad j = 1,2,3, \dots$$

Είναι ανεξάρτητες τυχαίες μεταβλητές

Κεφάλαιο 2

Συνελίξεις

2.1 Ορισμός

Έστω ότι έχουμε δύο συναρτήσεις $f\ και\ g$ η συνέλιξη τους $f * g$ ορίζεται ως εξής

$$(f * g)(z) = \int_{-\infty}^{+\infty} f(z - x)g(x)dx$$

Εφόσον στην παραπάνω σχέση τα όρια του ολοκληρώματος εκτείνονται $στο\ \mathbb{R}$ ο υπολογισμός της συνάρτησης $(f * g)(z)$ μπορεί να είναι δύσκολος σε αρκετές περιπτώσεις .

Προς αυτή την κατεύθυνση αναφέρουμε ότι τα όρια στην ολοκλήρωση στην παραπάνω σχέση είναι ουσιαστικά πεπερασμένα όταν οι μεταβλητές X και Y παίρνουν μόνο μη αρνητικές τιμές. Ποιό συγκεκριμένα αν οι X,Y είναι συνεχείς με τιμές στο $[0, \infty)$ τότε οι αντίστοιχες πυκνότητες $f_X(x) και\ f_Y(y)$ μηδενίζονται στον αρνητικό ημιάξονα και έτσι η αρχική μας σχέση μπορεί πλέον να γραφεί

$$(f * g)(z) = \int_0^z f(z - x)f(x)dx \qquad z \geq 0$$

2.2 Κατανομή αθροίσματος δύο ανεξάρτητων τυχαίων μεταβλητών

Υποθέτουμε ότι $X\ και\ Y$ είναι ανεξάρτητες τυχαίες μεταβλητές με συναρτήσεις πυκνότητας πιθανότητας $f_x(x)\ και\ f_Y(y)$ αντίστοιχα. Τότε η τυχαία μεταβλητή $Z = X + Y$ έχει συνάρτηση πυκνότητας πιθανότητας

$$f_Z = f_X * f_Y$$

Απόδειξη

(Συνεχής τυχαίες μεταβλητές)

Γνωρίζουμε ότι η συνάρτηση κατανομής μιας τυχαίας μεταβλητής δίνεται από τη σχέση

$$F(x) = P(Z \leq x) = \int_{-\infty}^{x} f_Z\,(z)dz$$

Παραγωγίζοντας την παραπάνω σχέση μπορούμε να πάρουμε τη συνάρτηση πυκνότητας πιθανότητας. Έτσι έχουμε τα παρακάτω

$$f_z(z) = \frac{\partial F(z)}{\partial z} = \lim_{\Delta Z \to 0} \frac{F(z + \Delta z) - F(z)}{\Delta z} = \lim_{\Delta Z \to 0} \frac{P(z \leq Z \leq x + \Delta z)}{\Delta z}$$

Έτσι για μικρό Δx έχουμε κατα προσέγγιση ότι

$$\mathbb{P}(z \leq Z \leq z + \Delta z) \cong f_z(z)dz$$

Όμως

$$Z = X + Y$$

Και η παραπάνω σχέση θα μας δώσει

$$\mathbb{P}(z \leq X + Y \leq z + \Delta z) \cong f(z)dz$$

Όμως

$$\mathbb{P}(z \leq X + Y \leq z + \Delta z) = \int_{-\infty}^{+\infty} \mathbb{P}(z - x \leq Y \leq z - x + \Delta z) f_x(x)dx$$

Όμως

$$\mathbb{P}(z - x \leq Y \leq z - x + \Delta z) = \int_{z-x}^{z-x+\Delta z} f_Y(y)dy = f_Y(z - x)dz$$

Και τελικά

$$\int_{-\infty}^{+\infty} \int_{z-x}^{z-x+\Delta z} f_x(x) f_Y(y)\, dydx = dz \int_{-\infty}^{+\infty} f_x(x) f_Y(z - x)dx$$

Δηλαδή καταλήξαμε στη σχέση

$$f_z(z)dz = dz \int_{-\infty}^{+\infty} f_x(x) f_Y(z - x)dx$$

από την οποία προκύπτει πως

$$f_z(z) = \int_{-\infty}^{+\infty} f_x(x) f_Y(z - x)dx$$

(Διακριτές τυχαίες μεταβλητές)

Στην περίπτωση όπου οι τυχαίες μεταβλητές X, Y, Z είναι διακριτές θα αποδείξουμε τη σχέση

$$f_Z = f_X * f_Y$$

Δουλεύοντας με την από κοινού συνάρτηση πυκνότητας πιθανότητας των τυχαίων μεταβλητών

$$f_{Y,Z}(y, z) = \mathbb{P}(Y = y, Z = z) = \mathbb{P}(Y = y, X + Y = z) = \mathbb{P}(Y = y, X = z - y)$$

Όμως οι τυχαίες μεταβλητές X, Y είναι ανεξάρτητες και έτσι

$$f_{Y,Z}(y,z) = \mathbb{P}(Y = y)\mathbb{P}(X = z - y)$$

Τώρα έχοντας υπολογίσει την από κοινού συνάρτηση πυκνότητας πιθανότητας μπορούμε να υπολογίσουμε την περιθώρια συνάρτηση πιθανότητας

$$f_Z(z) = \sum_{y\in\mathbb{R}} \mathbb{P}(Y = y)\mathbb{P}(X = z - y) = \sum_{y\in\mathbb{R}} f_x(z - y)\, f_Y(y)$$

Η συνέλιξη περιγράφει την συνάρτηση πυκνότητας πιθανότητας αντίστοιχα ενός αθροίσματος ανεξάρτητων τυχαίων μεταβλητών. Τα χρονικά διαστήματα που μεσολαβούν μεταξύ διαδοχικών ανανεώσεων είναι ανεξάρτητα μεταξύ τους έτσι η συνάρτηση κατανομής της n-στής ανανέωσης μπορεί να περιγραφεί με τη βοήθεια μιας συνέλιξης .

2.3 Συνελίξεις γνωστών κατανομών

1.Εκθετική κατανομή

Η απλούστερη κατανομή χρόνου ζωής είναι η εκθετική κατανομή.

Εάν δηλαδή μια συνεχής τυχαία μεταβλητή X ακολουθεί την εκθετική κατανομή έχει συνάρτηση πυκνότητας :

$$f(x) = \begin{cases} \mu e^{-\mu x} & \gamma\iota\alpha\ x \geq 0 \\ 0 & \gamma\iota\alpha\ x < 0 \end{cases}$$

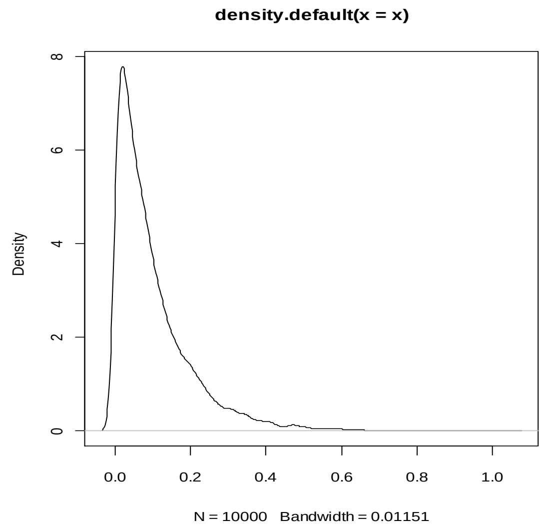
Όπου μ > 0 με μέση τιμή $E(X) = \frac{1}{\mu}$ και διακύμανση $Var(X) = \frac{1}{\mu^2}$

και συνάρτηση κατανομής

$$F(x) = \begin{cases} 1 - e^{-\mu x} & \gamma\iota\alpha\ x \geq 0 \\ 0 & \gamma\iota\alpha\ x < 0 \end{cases}$$

Με τη βοήθεια του στατιστικού προγράμματος R

Έχουμε μια γραφική απεικόνιση μιας εκθετικής κατανομής απο ένα τυχαίο δείγμα 10000 παρατηρήσεων με παράμετρο $\mu = 10$



Σχήμα 1γραφική απεικόνιση συνάρτησης πυκνότητας πιθανότητας εκθετικής κατανομής με $\mu = 10$

Έστω X, Y δύο ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν την εκθετική κατανομή με παράμετρο μ. Στην συγκεκριμένη περίπτωση $f = g$ λόγω ισονομίας.

$$f_Z(z) = (f * g)(z) = \int_{-\infty}^{+\infty} f(z-x)g(x)dx = \int_{0}^{+\infty} \mu e^{-\mu x}\, \mu e^{-\mu(z-x)} dx$$

$$= \int_0^z \mu^2 e^{-\mu z}\, dx = z\mu^2 e^{-\mu z}$$

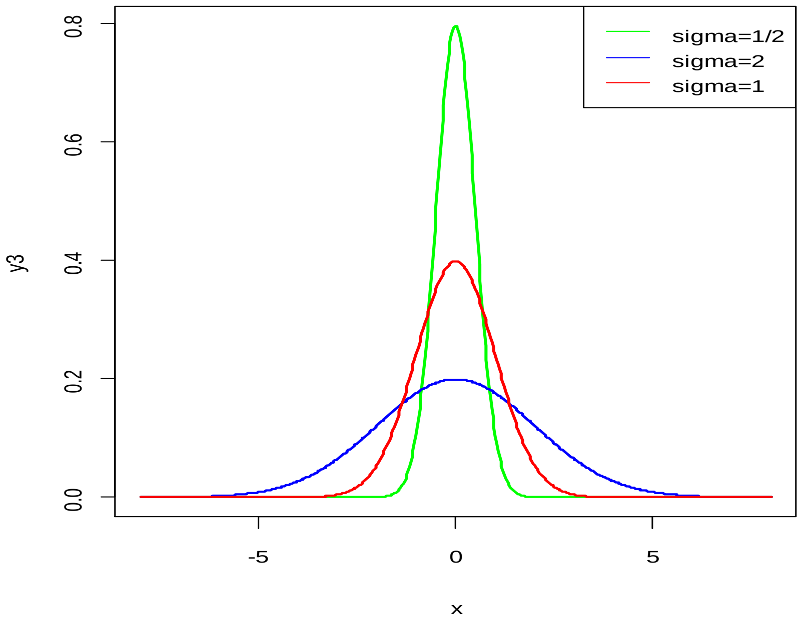
Πρόκειται δηλαδή για μια συνάρτηση πυκνότητας πιθανότητας Gamma(2,μ).

2.Κανονική κατανομή

Η κανονική κατανομή (συχνά στη βιβλιογραφία αναφέρεται και ως κατανομή του Gauss)

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\{-\frac{(t-\mu)^2}{2\sigma^2}\} \qquad t\epsilon(-\infty, +\infty)$$

με μ παράμετρο θέσης να παίρνει οποιαδήποτε τιμή στο ℝ και σ παράμετρο κλίμακας με $\sigma\, \epsilon(0, +\infty)$



Σχήμα 2 :Γραφική παράσταση συνάρτησης πυκνότητας πιθανότητας της κανονικής κατανομής για διαφορετικές διασπορές σ

Η συνάρτηση πυκνότητας πιθανότητας του αθροίσματος δύο τυχαίων μεταβλητών $Z = X + Y$ όπου X, Y $ακολουθούν$ $N(0,1)$ μπορεί να υπολογιστεί ως εξής

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x)dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-x)^2}{2}} dx =$$

$$= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} e^{-\frac{(z-x)^2}{2}} dx = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} e^{-\frac{z^2-2zx+x^2}{2}} dx =$$

$$= \frac{1}{2\pi} e^{-\frac{z^2}{4}} \int_{-\infty}^{+\infty} e^{-\left(x-z/2\right)^2} dx = \frac{1}{2\pi} e^{-\frac{z^2}{4}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(\frac{2x-z}{\sqrt{2}})^2} dx =$$

$$= \frac{1}{2\pi} e^{-\frac{z^2}{4}} \sqrt{\pi} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}\left(\frac{2x-z}{\sqrt{2}}\right)^2} \frac{1}{\sqrt{\pi}} dx$$

Όμως το ολοκλήρωμα της παραπάνω σχέσης ισούται με τη μονάδα μιας και το εσωτερικό του ολοκληρώματος είναι μια συνάρτηση πυκνότητας πιθανότητας της κανονικής κατανομής με με παράμετρο θέσης μηδέν και παράμετρο κλίμακας σ ίση με $\sqrt{2}$.

Έτσι καταλήγουμε στο συμπέρασμα ότι

$$f_Z(z) = \frac{1}{2\pi} e^{-\frac{z^2}{4}} \sqrt{\pi} = \frac{1}{\sqrt{4\pi}} e^{-\frac{z^2}{4}}$$

Πρόκειται για τη συνάρτηση πυκνότητας πιθανότητας μιας κανονικής κατανομής με μέση τιμή μηδέν και διακύμανση ίση με δύο.

3.Κατανομή Cauchy

Μια τυχαία μετβλητή X λέμε οτι ακολουθεί την κατανομή Cauchy και συμβολίζουμε $X \sim C(\mu, \sigma)$ $\mu \in \mathbb{R}$ $\sigma > 0$ όταν έχει συνάρτηση πυκνότητας πιθανότητας

$$f(x) = \frac{1}{\sigma\pi} \frac{1}{1 + \frac{(x-\mu)^2}{\sigma^2}} \qquad -\infty \leq x \leq +\infty$$

Στην ειδική περίπτωση όπου μ=0 και σ=1 η κατανομή λέγεται τυπική κανονική Cauchy

Και με τη βοήθεια του ***Mathematica*** αφού εισάγουμε την συνάρτηση πυκνότητας της τυπικής κανονικής Cauchy

$$f[\mathrm{x\_}] = (1/(\mathrm{Pi})) * (1/(1 + (x)^2)$$

μπορούμε να λύσουμε το ολοκλήρωμα

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) dx$$

με την εντολή

$$\mathrm{Integrate}[f[z-x] * f[x], \{x, -\mathrm{Infinity}, \mathrm{Infinity}\}, \mathrm{Assumptions} \rightarrow \{-1 < \mathrm{Im}[z] < 1\}]$$

$$\frac{1}{\pi^2(8+2z^2)}(4\pi + i\text{Log}[1-iz] + i\text{Log}[1+iz] - i\left(\begin{array}{cc}\text{Conjugate}[\text{Log}[1-iz]] & \text{Re}[z]==0\&\&\text{Im}[z]<0 \\ \text{Log}[1-iz] & \text{True}\end{array}\right) - i\left(\begin{array}{cc}\text{Conjugate}[\text{Log}[1+iz]] & \text{Re}[z]==0\&\&\text{Im}[z]>0 \\ \text{Log}[1+iz] & \text{True}\end{array}\right.$$

Και κάνοντας κάποιες πράξεις στο πραγματικό μέρος του παραπάνω μιγαδικού έχουμε

$$f_Z(z) = \frac{2}{\pi(4+z^2)}$$

4.Κατανομή Gamma

Έστω η τυχαία μεταβλητή $X \sim G(\alpha_1, \beta)$ και η τυχαία μεταβλητή $Y \sim G(\alpha_2, \beta)$ με αντίστοιχες συναρτήσεις πυκνότητας πιθανότητας

$$f_X(x) = \frac{e^{-x\beta} x^{\alpha_1 - 1} \beta^{\alpha_1}}{\Gamma(\alpha_1)} \quad \kappa\alpha\iota\; f_y(x) = \frac{e^{-x\beta}\, x^{\alpha_2 - 1} \beta^{\alpha_2}}{\Gamma(\alpha_2)}$$

Για $x \geq 0 \;\; \geq \kappa\alpha\iota\; y \geq 0$.Αντικαθιστώντας αυτές τις συναρτήσεις στη σχέση

$$f_Z(z) = \int_0^z f_x\,(x) f_Y(z-x)dx = \int_0^z f_x(x)\, f_Y(z-y)dx$$

$$f_Z(z) = \int_0^z \frac{e^{-x\beta}\, \chi^{\alpha_1 - 1}\beta^{\alpha_1}}{\Gamma(\alpha_1)} \frac{e^{-(z-y)\beta}\,(z-x)^{\alpha_2 - 1}\beta^{\alpha_2}}{\Gamma(\alpha_2)} dx =$$

$$f_Z(z) = \frac{\beta^{\alpha_1 + \alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} e^{-\beta(x+y)} \int_0^z x^{\alpha_1 - 1}\,(z-x)^{\alpha_2 - 1} dx =$$

$$= \frac{\beta^{\alpha_1 + \alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)} e^{-\beta(x+y)} \int_0^1 t^{\alpha_1 - 1}\,(1-t)^{\alpha_2 - 1} dx$$

Όμως

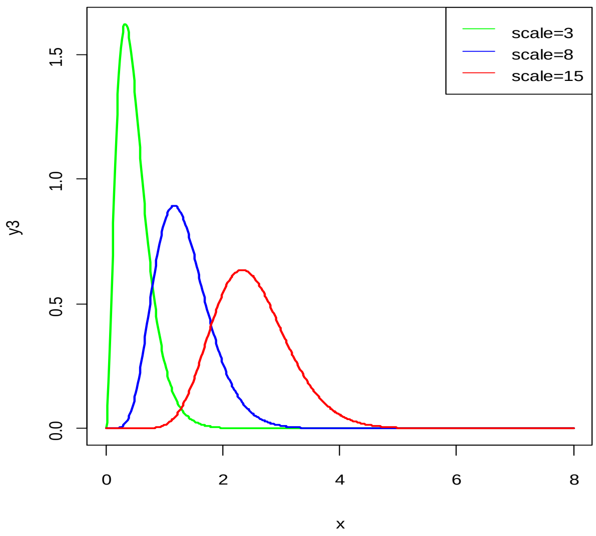
$$\int_0^1 t^{\alpha_1 - 1}(1-t)^{\alpha_2 - 1} dx = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)}$$

Με τη βοήθεια των παραπάνω βρίσκουμε πως η πυκνότητα του αθροίσματος $Z = X + Y$ δίνεται απο τη σχέση

$$f_Z(z) = \frac{\beta^{\alpha_1 + \alpha_2}}{\Gamma(\alpha_1 + \alpha_2)} z^{\alpha_1 + \alpha_2 - 1} e^{-\beta z} \qquad Z \geq 0$$

Δηλαδή αν η τυχαία μεταβλητή $X \sim Gamma(\alpha_1, \beta)$ και η $Y \sim Gamma(\alpha_2, \beta)$ τότε το άθροισμα $X + Y \sim Gamma(\alpha_1 + \alpha_2, \beta)$.

Με τη βοήθεια του στατιστικού πακέτου R σ το σχήμα που ακολουθεί έχουμε της πυκνότητες κατανομών Gamma με διαφορετικές παραμέτρους κλίμακας .



Σχήμα 3: *Γραφική απεικόνιση κατανομής Γάμμα με διαφορετικές τιμές παραμέτρου κλίμακας*

Κατανομή Weibull

Η κατανομή Weibull (από τον Σουηδό W.Weibull) είναι η πλέον διαδεδομένη κατανομή στην ανάλυση αξιοπιστίας. Εμπειρικά αποτελεί ένα ικανοποιητικό μοντέλο για πολλά φαινόμενα.

Η συνάρτηση πυκνότητας πιθανότητας ορίζεται ως εξής

$$f(t) = at^{a-1} \exp\{-t^{a}\} \quad t \geq 0$$

Όπου $\alpha > 0$ είναι η παράμετρος κλίμακας και $n > 0$ είναι η παράμετρος σχήματος.

Η Εκθετική κατανομή αποτελεί ειδική περίπτωση της Wiebull όταν $n = 1$

Η συνάρτηση κατανομής της Wiebull είναι η εξής

$$F(t) = 1 - \exp\{-(t)^{n}\} \quad t \geq 0$$

Θα προσπαθήσουμε να υπολογίσουμε τη συνέλιξη δύο τυχαίων και ανεξάρτητων (ισόνομων) μεταβλητών $X\ και\ Y$.

$$f_Z(z) = \int_0^z f_x\,(x) f_Y(z-x)dx = \int_0^z f_x(x)\, f_Y(z-y)dy$$

$$= \int_0^z at^{a-1} \exp\{-t^a\} a(z-t)^{a-1} \exp\{-(z-t)^a\}dt \quad =$$

$$a^2 \int_0^z t^{a-1}\,(z-t)^{a-1} e^{-[(z-t)^a + t^a]} dt =$$

Στην παραπάνω σχέση θέτουμε όπου $\frac{t}{z}$=x και έχουμε

$$= a^2 z^{a-1} z^{a-1} z \int_0^1 x^{a-1}\,(1-x)^{a-1} e^{-z^a[(1-x)^a + x^a]} dx$$

Έχουμε υποθέσει την κατανομή της τυχαίας μεταβλητής $t/_Z$ υπο τη συνθήκη ότι το άθροισμα $X + Y = Z$

Αυτή η υπο συνθήκη κατανομή έχει πυκνότητα ίση με

$$g_{a,d}(x) = cx^{a-1}(1-x)^{a-1} e^{-z^a[(1-x)^a + x^a]}$$

2.4 Συνελίξεις διακριτών τυχαίων μεταβλητών

Πρόταση

Έστω X ,Y δύο ανεξάρτητες τυχαίες μεταβλητές και Z μια τρίτη μεταβλητή που ορίζεται απο τη σχέση $Z = X + Y$.

Αν οι X,Y είναι διακριτές τυχαίες μεταβλητές με συναρτήσεις πιθανότητας $(x)\ \kappa\alpha\iota\ f_Y(y)$ τότε η συνάρτηση πιθανότητας της Z δίνεται απο τη σχέση

$$f_Z(z) = \sum_{x \in \mathbb{R}_X} f_X\,(x) f_Y(z-x) = \sum_{y \in \mathbb{R}_Y} f_Y\,(y) f_Y(z-y)$$

Απόδειξη

Αν οι X,Y είναι διακριτές τότε για την απο κοινού συνάρτηση πιθανότητας των μεταβλητών Y και $Z = X + Y$ έχουμε

$$f_{Y,Z}(y,z) = \mathrm{P}(Y = y, Z = z) = \mathrm{P}(Y = y, X + Y = z) = \mathrm{P}(Y = y, X = z - y)$$

Εφόσον οι X,Y είναι ανεξάρτητες πάιρνουμε ότι

$$f_{Y,Z}(y,z) = f_Y(y) f_Y(z-x)$$

Προσθέτωντας τις πιθανότητες $f_{Y,Z}(y,z)$ για όλες τις τιμές $y \in \mathrm{R}_Y$ και δεδομένου ότι το άθροισμα ισούται με $f_Z(z)$= $\mathrm{P}(Z = z)$ προκύπτει η δεύτερη σχέση.

Η πρώτη σχέση προκύπτει κάνοντας την αντικατάσταση $z - x = y$ και παρατηρώντας ότι κάθε $X \in \mathrm{R}_X$ μπορεί να γραφεί στη μορφή $z - y$ όπου z, y είναι στο πεδίο τιμών των Z *και* Y αντίστοιχα.

2.5 Συνέλιξη κατανομής

Η συνέλιξη για δύο κατανομές F και G συμβολίζεται με $F * G$ και για κάθε Borel σύνολο δίνεται απο τη σχέση

$$(F * G)(B) = \int_{-\infty}^{+\infty} F(B - y)d(G(y) = \int_{-\infty}^{+\infty} G(B - y)d(F(y)$$

Όπου

$$B - y = \{x - y : x \in B\}$$

Έστω $(\boldsymbol{\Omega}, \boldsymbol{F}, \mathbf{P})$ένας χώρος πιθανότητας και X, Y είναι δύο ανεξάρτητες τυχαίες μεταβλητές με αντίστοιχές συναρτήσεις κατανομής F *και* G τότε

$$(F * G)(B) = \mathrm{P}\{X + Y \in B\}$$

Πιο συγκεκριμένα

- Αν η συνάρτηση κατανομής G είναι συνεχής με πυκνότητα g τότε

$$(F * G)(x) = \int_{-\infty}^{+\infty} F(x - y)d(G(y) = \int_{-\infty}^{+\infty} F(x - y)g(y)dy$$

- Στη γενική περίπτωση όπου η G είναι μια μεικτή κατανομή με συνάρτηση πυθανότητας g_1 για το διακριτό τμήμα και g_2 για το συνεχές τμήμα τότε

$$(F * G)(B) = \int_{-\infty}^{+\infty} F(x - y)g_2(y)dy + \sum_{y \in R_y} F(x - y)g_1(y)$$

Πρόταση

Έστω X, Y δύο ανεξάρτητες τυχαίες μεταβλητές με συναρτήσεις κατανομής F *και* G αντίστοιχα. Τότε η συνάρτηση κατανομής του αθροίσματος

$$Z = X + Y$$

είναι η συνέλιξη $F * G$

Απόδειξη

(Διακριτή περίπτωση)

$$\mathrm{P}(Z \leq x) = \ \textstyle\sum_{y \in R_y} \ \mathrm{P}(\, Z \leq x/Y = y)\ \mathrm{P}(Y = y) =$$

$$\textstyle\sum_{y \in R_y} \ .\mathrm{P}\ (X \leq x - y/Y = y)\ \mathrm{P}(Y = y)$$

Όμως οι τυχαίες μεταβλητές X, Y είναι ανεξάρτητες και έτσι η παραπάνω σχέση μας δίνει

$$\sum_{y \in R_y} F(x - y)\ \mathrm{P}(Y = y)$$

(Συνεχής περίπτωση)

$$P(Z \leq x) = \int_{-\infty}^{+\infty} P(\, Z \leq x/Y = y) g(y) dy =$$

$$\int_{-\infty}^{+\infty} P(\, X \leq x - y/Y = y) g(y) dy =$$

$$\int_{-\infty}^{+\infty} F(x - y) g(y) dy$$

2.6 Συνέλιξη τυχαίου αθροίσματος ισόνομων και ανεξάρτητων τυχαίων μεταβλητών

Έστω μια τυχαία μεταβλητή

$$S = X_1 + X_2 + \cdots + X_N$$

Όπου $X_i \quad i = 1,2,3, \ldots N$ τυχαίες μεταβλητές με διακριτή ή συνεχή κατανομή και N διακριτή τυχαιά μεταβλητή που μετράει τον αριθμό των φορών που εμφανίζεται κάποιο συμβάν.

Έστω το ενδεχόμενο $\{S > x\}$ δηλαδή το ενδεχόμενο το άθροισμα

$$\sum_{i=1}^{N} X_i > x$$

Όμως το S εξαρτάται απο τις τυχαίες μεταβλητές $X_i \quad i = 1,2,3, \ldots N$ και την τυχαία μεταβλητή N.

Με τη βοήθεια της θεωρίας συνόλων μπορούμε να γράψουμε το ενδεχεχόμενο

$$\{S > x\} = \bigcup_{n=0}^{\infty} \left\{ \{S > x\} \bigcap \{N = n\} \right\}$$

Και η αντίστοιχη πιθανότητα το ενδεχομένου είναι

$$P\{S > x\} = P(\bigcup_{n=0}^{\infty} \left\{ \{S > x\} \bigcap \{N = n\}) \right\}$$

$$P\{S > x\} = \sum_{n=0}^{\infty} P(\{S > x\} \bigcap \{N = n\})$$

$$P\{S > x\} = \sum_{n=0}^{\infty} P(S > x/N = n)\, P(N = n)$$

$$P\{S > x\} = \sum_{n=0}^{\infty} P(X_1 + X_2 + \cdots + X_N > x/N = n)\, P(N = n)$$

$$P\{S > x\} = \sum_{n=0}^{\infty} P(X_1 + X_2 + \cdots + X_n > x/N = n)\, P(N = n)$$

$$P\{S > x\} = \sum_{n=0}^{\infty} P(X_1 + X_2 + \cdots + X_n > x)\, P(N = n)$$

Ορίζουμε με

$$\overline{F^{*n}}(x) = P(X_1 + X_2 + \cdots + X_n > x)$$

την ουρά της νιοστής συνέλιξης και έτσι

$$P\{S > x\} = \sum_{n=0}^{\infty} \overline{F^{*n}}(x)\, P(N = n)$$

Αντίστοιχα η πιθανότητα του ενδεχομένου

$$\{S \leq x\} = \bigcup_{n=0}^{\infty} \left\{ \{S \leq x\} \bigcap \{N = n\} \right\}$$

$$P\{S \leq x\} = P(\bigcup_{n=0}^{\infty} \left\{ \{S \leq x\} \bigcap \{N = n\}) \right\}$$

$$P\{S \le x\} = \sum_{n=0}^{\infty} P(\{S \le x\} \bigcap \{N = n\})$$

$$P\{S \le x\} = \sum_{n=0}^{\infty} P(S \le x/N = n)\, P(n = n)$$

$$P\{S \le x\} = \sum_{n=0}^{\infty} P(X_1 + X_2 + \cdots + X_N \le x/N = n)\, P(N = n)$$

$$P\{S \le x\} = \sum_{n=0}^{\infty} P(X_1 + X_2 + \cdots + X_n \le x/N = n)\, P(n = n)$$

$$P\{S \le x\} = \sum_{n=0}^{\infty} P(X_1 + X_2 + \cdots + X_n \le x)\, P(N = n)$$

2.7 Εφαρμογή

Έστω μια τυχαία μεταβλητή

$$S = X_1 + X_2 + \cdots + X_N$$

Όπου $X_i \quad i = 1,2,3, \ldots N$ ανεξάρτητες και ισόνομες τυχαίες μεταβλητές που ακολουθούν την εκθετική κατανομή με παράμετρο λ και N διακριτή τυχαία μεταβλητή που μετράει τον αριθμό των φορών που εμφανίζεται κάποιο συμβάν.

Έστω το ενδεχόμενο $\{S > x\}$ δηλαδή το ενδεχόμενο το άθροισμα

$$\sum_{i=1}^{N} X_i > x$$

Σκοπός μας να φτιάξουμε έναν κλειστό τύπο για την ουρά της τυχαίας μεταβλητής S

$$P\{S \ge x\} = \sum_{n=0}^{\infty} P(X_1 + X_2 + \cdots + X_n > x)\, P(N = n)$$

Το άθροισμα n τυχαίων ανεξάρτητων και ισόνομων τυχαίων μεταβλητών που ακολουθούν την εκθετική κατανομή με παράμετρο λ ή ισοδύναμα της κατανομή Γάμμα με παραμέτρους 1 και λ ακολουθεί την κατανομή Γάμμα με παραμέτρους n, λ

$$\overline{F^{*n}}(x) = 1 - F^{*n}(x) = 1 - \int_{-\infty}^{x} f^{*n}(y)dy = 1 - \int_{0}^{x} f^{*n}(y)dy = \int_{x}^{\infty} f^{*n}(y)dy$$

με

$$f^{*n}(y) = \frac{\lambda^n y^{n-1} e^{-\lambda x}}{\Gamma(n)} \qquad y \geq 0$$

Έτσι

$$\overline{F^{*n}}(x) = \int_x^{\infty} f^{*n}\,(y) dy = \int_x^{\infty} \frac{\lambda^n y^{n-1} e^{-\lambda y}}{\Gamma(n)} dy = \int_{u=\lambda x}^{\infty} \frac{u^{n-1} e^{-u}}{\Gamma(n)} du = \frac{\Gamma(n, \lambda x)}{\Gamma(n)}$$

Όπου

$$\frac{\Gamma(n, \lambda x)}{\Gamma(n)}$$

Είναι η άνω μη πλήρης συνάρτηση Γάμμα (*upper incomplete)*

Εφαρμόζουμε παραγοντική ολοκλήρωση και έχουμε

$$\bar{F}^{*n}(x) = \int_x^{\infty} \frac{\lambda^n y^{n-1} e^{-\lambda x}}{\Gamma(n)} dy = -\frac{\lambda^n}{\Gamma(n)} \left\{ \left[y^{n-1} \frac{e^{-\lambda y}}{\lambda} \right]_x^{\infty} - (n-1) \int_x^{\infty} y^{n-2} \frac{e^{-\lambda y}}{\lambda} dy \right\}$$

$$\int_x^{\infty} \frac{\lambda^n y^{n-1} e^{-\lambda x}}{\Gamma(n)} dy = \frac{\lambda^n}{\Gamma(n)} \left\{ \left[x^{n-1} \frac{e^{-\lambda x}}{\lambda} \right] - (n-1) \overline{F *^{(n-1)}}\,(x) \right\}$$

$$= \frac{\lambda^n}{\Gamma(n)} x^{n-1} \frac{e^{-\lambda x}}{\lambda} + \int_x^{\infty} \frac{\lambda^{n-1} y^{n-2} e^{-\lambda y}}{(n-2)!} d\,y =$$

$$= \frac{\lambda^n}{\Gamma(n)} x^{n-1} \frac{e^{-\lambda x}}{\lambda} + 1 - F^{*(n-1)}(x)$$

Τελικά προκύπτει πως

$$1 - F^{*n}(x) = \frac{\lambda^n}{\Gamma(n)} x^{n-1} \frac{e^{-\lambda x}}{\lambda} + 1 - F^{*(n-1)}(x)$$

Από την οποία λαμβάνουμε τον αναδρομικό τύπο

$$F^{*n}(x) = F^{*(n-1)}(x) - \frac{\lambda^{n-1}}{(n-1)!} x^{n-1} e^{-\lambda x}$$

$$F^{*1}(x) = F^{*0}(x) - \frac{(\lambda x)^0}{0!} e^{-\lambda x}$$

$$F^{*2}(x) = F^{*1}(x) - \frac{(\lambda x)^1}{1!} e^{-\lambda x}$$

$$F^{*3}(x) = F^{*2}(x) - \frac{(\lambda x)^2}{2!} e^{-\lambda x}$$

.

.

.

.

$$F^{*(n-1)}(x) = F^{*(n-2)}(x) - \frac{(\lambda x)^{n-2}}{(n-2)!} e^{-\lambda x}$$

$$F^{*n}(x) = F^{*(n-1)}(x) - \frac{(\lambda x)^{n-1}}{(n-1)!} e^{-\lambda x}$$

Προσθέτοντας κατά μέλη τις παραπάνω σχέσεις προκύπτει

$$F^{*n}(x) = 1 - e^{-\lambda x} \sum_{i=0}^{n-1} \frac{(\lambda x)^i}{i!} \quad n \geq 1$$

Κεφάλαιο 3

Τυχαίος Περίπατος

3.1 Εισαγωγή

Ο υπολογισμός ποσοτήτων, όπως οι πιθανότητες πραγματοποίησης ορισμένων γεγονότων σε καθορισμένο τμήμα του χώρου ή/και του χρόνου, σε πολλές περιπτώσεις είναι είτε δύσκολος και χρονοβόρος με μια **ντετερμινιστική** μέθοδο, δηλαδή με χρήση καθορισμένων εξισώσεων που περιγράφουν το εξεταζόμενο φαινόμενο, οι οποίες ενδεχομένως δεν υπάρχουν ή είναι αδύνατο να εξαχθούν. Στις περιπτώσεις αυτές εξετάζεται η δυνατότητα εφαρμογής μιας **μη ντετερμινιστικής** ή **στοχαστικής** (stochastic) μεθόδου.

3.2 Τυχαίος περίπατος

Ο όρος τυχαίος περίπατος χρησιμοποιήθηκε για πρώτη φορά από τον Pearson το 1905. Το μοντέλο του τυχαίου περιπάτου εφαρμόστηκε αρχικά στην φυσική για να περιγράψει τις κινήσεις των σωματιδίων (Brownian motion). Η πρώτη οικονομική εφαρμογή του μοντέλου έγινε από τον Cowles (1933), ο οποίος παρουσίασε μια εμπειρική μελέτη πάνω στις επιλογές μετοχών ενός μεγάλου αριθμού επενδυτών και κατέληξε στο συμπέρασμα πως δεν υπάρχουν σοβαρές ενδείξεις που να δείχνουν ότι κάποιος μπορεί να κερδίσει στην αγορά. Όταν κατέστη δυνατή η χρήση των ηλεκτρονικών υπολογιστών για την στατιστική ανάλυση χρονολογικών σειρών, αναπτύχθηκαν περισσότερες και πιο σύνθετες οικονομικές εφαρμογές του μοντέλου. Ο Kendall (1953) εξέτασε τις ιστορικές τιμές 22 μετοχών και εμπορευμάτων, από την Μεγάλη Βρετανία και διαπίστωσε πως παρουσιάζουν σχεδόν μηδενική γραμμική συσχέτιση. Τέτοιου είδους έρευνες πραγματοποίησαν στην συνέχεια και μια σειρά επιστημόνων όπως ο Working, ο Fama, ο Roberts, ο Alexander κ.α (Dimson and Massoud, 1998)

3.3 Ένα απλό παράδειγμα τυχαίου περιπάτου

Υποθέτουμε ότι είναι καλοκαίρι και βρισκόμαστε με μια παρέα φίλων σε διακοπές, το καθημερινό πρόγραμμα που έχουμε αποφασίσει να ακολουθήσουμε είναι ιδιαίτερα χαλαρό και περιλαμβάνει πολύ ξεκούραση και διασκέδαση.

Κάποιο από τα βράδια των διακοπών σας καταναλώνεται μεγάλη ποσότητα αλκοόλ και επειδή η οδική σας συμπεριφορά δεν το επιτρέπει αποφασίζεται να γυρίσετε στο δωμάτιο σας με τα πόδια.

Θέλοντας να απλουστεύσουμε λίγο την παραπάνω κατάσταση ας υποθέσουμε ότι όλα τα νυχτερινά καταστήματα διασκέδασης βρίσκονται στον ίδιο δρόμο σε μια ευθεία από το ξενοδοχείο στο οποίο διαμένουμε. Σκοπός λοιπόν εκείνη τη νύχτα είναι να μπορέσουμε να μεταφερθούμε με ασφάλεια από το σημείο Α που θα είναι η ακριβής τοποθεσία του νυχτερινού κέντρου διασκέδασης στο σημείο Β που θα είναι το δωμάτιο μας. Για την δική μας ευκολία και χωρίς αυτό να επηρεάζει τη δομή του προβλήματος αλλά και τη λύση του θεωρούμε πως η αρχική μας κατάσταση (το σημείο Α δηλαδή) είναι ίσο με το μηδέν και ότι το δωμάτιο μας βρίσκεται n βήματα μακριά μας (μετράμε την απόσταση με βήματα μιας και το μεταφορικό μας μέσο δεν το χρησιμοποιούμε λόγο της ποσότητας αλκοόλ που έχουμε χρησιμοποιήσει).

Επειδή βρίσκεστε σε κατάσταση μέθης και έχετε έλλειψη προσανατολισμού κινείστε με ακανόνιστα τυχαία βήματα άλλοτε προς τα αριστερά του νυχτερινού κέντρου και άλλοτε δεξιά βρίσκεστε πάντα όμως στο δρόμο που συνδέει το κέντρο διασκέδασης με το ξενοδοχείο που μένετε. Η πιο απλή περίπτωση που μπορούμε να φανταστούμε είναι πως μπορείτε να πάτε δεξιά ή αριστερά του δρόμου με την ίδια ακριβώς πιθανότητα μιας και τα βήματα σας είναι τυχαία και ακανόνιστα. Η πιθανότητα να κινηθείτε δεξιά λοιπόν είναι ίση με την πιθανότητα να κινηθείτε αριστερά και ίση με 0,5 μιας και δεν έχουμε λάβει στο παράδειγμα μας υπόψη πως μπορούμε να μείνουμε ακίνητοι στη μέση του δρόμου!

3.4 Απλός τυχαίος περίπατος

Την παραπάνω ιστορία θα προσπαθήσουμε να την περιγράψουμε με την βοήθεια των πιθανοτήτων και των στοχαστικών διαδικασιών .

Ας θεωρήσουμε αρχικά πώς ο πρωταγωνιστής μας κινείται με σταθερό μήκος βήματος l πάνω στον δρόμο με πιθανότητα p να κινηθεί δεξιά και με πιθανότητα $q = 1 - p$ να κινηθεί αριστερά.

Έστω επίσης X_n η θέση του ανθρώπου ύστερα από n βήματα.

Η θέση του μεθυσμένου μας πρωταγωνιστή μετά από n βήματα μπορεί να περιγραφεί με την βοήθεια μιας στοχαστικής διαδικασίας $\{X_n : n = 1,2,3, \dots\}$ ανεξάρτητων προσαυξήσεων με την τυχαία μεταβλητή

$$X_n = X_0 + Z_1 + Z_2 + \cdots Z_n \qquad n = 1,2,3, \dots$$

όπου X_0 είναι η θέση εκκίνησης και στο παράδειγμα μας έχουμε ορίσει ως $X_0 = 0$ και Z_i ανεξάρτητες διακριτές τυχαίες μεταβλητές που περιγράφουν τα βήματα του τυχαίου περιπάτου οι οποίες έχουν κατανομή

$$Z_i = \begin{cases} +1 & \mu\varepsilon\ \pi\iota\theta\alpha\nu\acute{o}\tau\eta\tau\alpha \quad p \\ -1 & \mu\varepsilon\ \pi\iota\theta\alpha\nu\acute{o}\tau\eta\tau\alpha \quad q \end{cases}$$

Δηλαδή η τυχαία μεταβλητή Z_i περιγράφει τα ακανόνιστα και τυχαία βήματα του ζαλισμένου από την υπερβολική κατανάλωση πρωταγωνιστή της ιστορίας μας.

Θέση της τυχαίας μεταβλητής X_n

Η θέση που θα βρεθεί ο άνθρωπος του παραδείγματος μας μετά από n το πλήθος βήματα εξαρτάται αρχικά αν αυτά τα n βήματα είναι άρτιος αριθμός ή περιττός. Πιο συγκεκριμένα :

- Αν $n = 2\kappa$ δηλαδή άρτιος αριθμός βημάτων τότε ο άνθρωπος θα βρεθεί σε άρτια θέση. Αναφέρουμε ένα απλό παράδειγμα για την παραπάνω περίπτωση έστω ότι ο μεθυσμένος έχει διανύσει δέκα το πλήθος ακανόνιστα βήματα λόγο έλλειψης προσανατολισμού τότε η θέση του που περιγράφεται από την τυχαία μεταβλητή X_n θα είναι μια από τις -10,-8,-6,-4,-2,0,2,4,6,8,10.

Γενικεύοντας το παραπάνω απλό παράδειγμα μπορούμε να πούμε πως η τυχαία μεταβλητή X_n *παίρνει τιμές στο σύνολο* $\{-2\kappa, -2\kappa+2, \ldots, 0,2,4, \ldots, 2\kappa-2, 2\kappa\}$

- Αν $n = 2\kappa + 1$ δηλαδή περιττός αριθμός βημάτων τότε ο άνθρωπος θα βρεθεί σε περιττή θέση. Αναφέρουμε ένα απλό παράδειγμα για την παραπάνω περίπτωση έστω ότι ο μεθυσμένος έχει διανύσει έντεκα το πλήθος ακανόνιστα βήματα λόγω έλλειψης προσανατολισμού τότε η θέση του που περιγράφεται από την τυχαία μεταβλητή X_n θα είναι μια από τις -11,-9,-7,-5,-3,-1,1,3,5,7,9.

Γενικεύοντας το παραπάνω απλό παράδειγμα μπορούμε να πούμε πως η τυχαία μεταβλητή X_n *παίρνει τιμές στο σύνολο* $\{-2\kappa-1, -2\kappa+1, \ldots, 0,2,4, \ldots, 2\kappa-1, 2\kappa+1\}$

Σύμφωνα λοιπόν με τα παραπάνω η θέση του πρωταγωνιστή μας θα μπορούσε να είναι μια από τις παραπάνω, το επόμενο που θα μπορεί να αναρωτηθούμε είναι με τι πιθανότητα μπορεί να βρεθεί σε καθεμία από τις παραπάνω θέσεις

Υπολογισμός της πιθανότητας η τυχαία μεταβλητή X_n να πάρει μια τιμή μεταξύ ενός τυχαίου διαστήματος

Είναι λογικό να θέλουμε να υπολογίσουμε με τι πιθανότητα μπορεί να βρεθεί η στοχαστική μας ανέλιξη σε κάποια από τις πιθανές θέσεις που μπορεί πάρει.

Ζητάμε λοιπόν να υπολογίσουμε την πιθανότητα

$$P(-r < X_n \leq m)$$

Αν ο αριθμός των βημάτων του τυχαίου περιπάτου είναι αρκετά μεγάλος ώστε να μπορέσουμε να εφαρμόσουμε κεντρικό οριακό θεώρημα σύμφωνα με το οποίο μπορούμε να προσεγγίσουμε την ζητούμενη πιθανότητα της διακριτής τυχαίας μεταβλητής με τη βοήθεια της τυποποιημένης κατανομής $N(0,1)$.

3.5 Μέση τιμή της τυχαίας μεταβλητής Z_i

Όπως έχουμε αναφέρει η τυχαία μεταβλητή Z_i έχει κατανομή

$$Z_i = \begin{cases} +1 & \mu\varepsilon\ \pi\iota\theta\alpha\nu\acute{o}\tau\eta\tau\alpha\ \ p \\ -1 & \mu\varepsilon\ \pi\iota\theta\alpha\nu\acute{o}\tau\eta\tau\alpha\ \ q \end{cases}$$

Η μέση τιμή της Z_i δίνεται από τη σχέση

$$E(Z_i) = +1P(Z_i = 1) - 1P(Z_i = -1)$$

Όμως

$$P(Z_i = 1) = p \quad \kappa\alpha\iota\ P(Z_i = -1) = q$$

Έτσι

$$E(Z_i) = p - q = \mu = 2p - 1$$

3.6 Ισχυρός Νόμος των μεγάλων αριθμών και μέση τιμή

Ο Ισχυρός Νόμος των Μεγάλων Αριθμών (*The Strong Law of Large Numbers)* είναι ένα από τα πιο γνωστά αποτελέσματα της θεωρίας πιθανοτήτων. Σύμφωνα με το θεώρημα κάτω από κατάλληλες υποθέσεις ο δειγματικός μέσος μιας ακολουθίας ανεξάρτητων τυχαίων μεταβλητών που ακολουθούν μια κοινή κατανομή συγκλίνει σχεδόν βεβαίως προς τον θεωρητικό μέσο της κατανομής (τη μέση τιμή της κατανομής)

Πιο αναλυτικά θα αναφέρουμε παρακάτω δύο θεωρήματα και ένα πόρισμα που θα μας βοηθήσουν στην ανάλυση μας

Θεώρημα 1 (Kolmogorov) Έστω $\{X_n : n = 1,2,3, \ldots\}$ ακολουθία ανεξάρτητων τυχαίων μεταβλητών για τις οποίες οι μέσες τιμές $\mu_n = E(X_n) και\ οι\ διασπορές\ \sigma_n^2 = V(X_n)$ είναι πεπερασμένες και ικανοποιούν την συνθήκη

$$\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} < \infty$$

Τότε

$$\frac{S_n - E(S_n)}{n} \to^{\sigma\beta} 0$$

Άμεσο συμπέρασμα του θεωρήματος είναι το ακόλουθο πόρισμα.

Πόρισμα

Έστω $\{X_n : n = 1,2,3, \ldots\}$ ακολουθία ανεξάρτητων τυχαίων μεταβλητών με κοινή πεπερασμένη μέση τιμή $\mu\ = E(X_n) και\ πεπερασμένη\ διασπορα\ \sigma_n^2 = V(X_n)$.

Τότε

$$\frac{S_n}{n} \to^{\sigma\beta} \mu$$

Θεώρημα 2 (Kolmogorov) Έστω $\{X_n : n = 1,2,3, \ldots\}$ ακολουθία ανεξάρτητων και ισόνομων τυχαίων μεταβλητών. Αν η κοινή μέση τιμή είναι πεπερασμένη και ίση με με μ τότε

$$\frac{S_n}{n} \to^{\sigma\beta} \mu\ για\ n \uparrow \infty$$

Το θεώρημα αυτό προυποθέτει επιπλέον ισόνομες τυχαίες μεταβλητές όμως σε ότι αφορά τις ροπές αρκείται στο πεπερασμένο της μέσης τιμής .

Σύμφωνα λοιπόν με τον Ισχυρό Νόμο των Μεγάλων Αριθμών

$$\frac{X_n}{n} \to^{\sigma\beta} E(Z_i) = p - q = \mu\ για\ n \uparrow \infty$$

Και συνεπώς η ακολουθία $\{X_n : n = 1,2,3, \dots\}$ αποκλίνει θετικά δηλαδή $\mathbb{P}(limX_n = \infty) = 1$ όταν $p > q \Rightarrow$

$$p - q > 0 \rightarrow \mu > 0$$

Και αποκλίνει αρνητικά για $p < q$

3.7 Διακύμανση της τυχαίας μεταβλητής Z_i

Όπως έχουμε αναφέρει η τυχαία μεταβλητή Z_i έχει κατανομή

$$Z_i = \begin{cases} +1 & \mu\varepsilon\ \pi\iota\theta\alpha\nu\acute{o}\tau\eta\tau\alpha \quad p \\ -1 & \mu\varepsilon\ \pi\iota\theta\alpha\nu\acute{o}\tau\eta\tau\alpha \quad q \end{cases}$$

Κάνοντας χρήση του τύπου της διακύμανσης έχουμε

$$Var(Z_i) = E\left({Z_i}^2\right) - E(Z_i)^2$$

$$E\left({Z_i}^2\right) = p + q = 1$$

$$Var(Z_i) = E\left({Z_i}^2\right) - E(Z_i)^2 = p + q - (p - q)^2 = 4pq$$

3.8 Κεντρικό οριακό θεώρημα

Έστω $\{X_n : n = 1,2,3, \dots\}$ ακολουθία ανεξάρτητων και ισόνομων τυχαίων μεταβλητών με πεπερασμένη μέση τιμή $\mu\ = E(X_n) \kappa\alpha\iota\ \delta\iota\alpha\sigma\pi o\rho\acute{\alpha}\ \sigma^2 = V(X_n)$.Αν

$$S_n = \sum_{n=1}^{\infty} X_n \qquad n = 1,2,3 \dots$$

Τότε

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \rightarrow^{D} N(0,1)$$

Τώρα έχουμε όλα τα απαραίτητα εργαλεία ώστε να μπορέσουμε να εφαρμόσουμε το Κεντρικό Οριακό Θεώρημα στην

$$P(-r < X_n \leq m)$$

$$= P\left(\frac{-r - n(p-q)}{\sqrt{4npq}} < \frac{X_n - (p-q)}{\sqrt{4npq}} \leq \frac{m - n(p-q)}{\sqrt{4npq}}\right)$$

Μπορούμε να ορίσουμε την συνάρτηση κατανομής της τυποποιημένης κανονικής κατανομής $N(0,1)$ ως εξής

$$\Phi(x) = \int_{-\infty}^{x} \varphi(u)du$$

Με τη βοήθεια της συνάρτησης κατανομής η (2.6) μπορεί να γραφεί ως εξής

$$P\left(\frac{-r - n(p-q)}{\sqrt{4npq}} < \frac{X_n - n(p-q)}{\sqrt{4npq}} \leq \frac{m - n(p-q)}{\sqrt{4npq}}\right) =$$

$$= \Phi\left(\frac{m - n(p-q)}{\sqrt{4npq}}\right) - \Phi\left(\frac{-r - n(p-q)}{\sqrt{4npq}}\right) \qquad \gamma\iota\alpha \quad n \uparrow \infty$$

Από την παραπάνω σχέση παρατηρούμε τα εξής

- Αν $p - q > 0 \; \gamma\iota\alpha \; n \uparrow \infty$ τότε $\tau\alpha \; \Phi\left(\frac{m-n(p-q)}{\sqrt{4npq}}\right), \Phi\left(\frac{-r-n(p-q)}{\sqrt{4npq}}\right)$ τείνουν στο μηδέν
- Αν $p - q < 0 \; \gamma\iota\alpha \; n \uparrow \infty$ τότε $\tau\alpha \; \Phi\left(\frac{m-n(p-q)}{\sqrt{4npq}}\right), \Phi\left(\frac{-r-n(p-q)}{\sqrt{4npq}}\right)$ τείνουν στη μονάδα
- Αν $p - q = 0 \; \gamma\iota\alpha \; n \uparrow \infty$ τότε $\tau\alpha \; \Phi\left(\frac{m-n(p-q)}{\sqrt{4npq}}\right), \Phi\left(\frac{-r-n(p-q)}{\sqrt{4npq}}\right)$ τείνουν στο 0,5
- Τα παραπάνω αποτελέσματα θα ήταν περισσότερο ικανοποιητικά αν κάναμε χρήση και της διόρθωσης της συνέχειας.

Συμπέρασμα

Σύμφωνα λοιπόν με όσα έχουμε αναφέρει μέχρι τώρα η πιθανότητα να ταλαιπωρηθεί ο πρωταγωνιστής του παραδείγματος μας και να μην φτάσει κάποια στιγμή στο δωμάτιο του είναι μηδέν!

Υπολογισμός της πιθανότητας η τυχαία μεταβλητή X_n να πάρει μια τιμή m

Είναι λογικό να θέλουμε να υπολογίσουμε με τι πιθανότητα μπορεί να βρεθεί η στοχαστική μας ανέλιξη σε κάποια από τις πιθανές θέσεις που μπορεί πάρει.

Ζητάμε λοιπόν να υπολογίσουμε την πιθανότητα του ενδεχομένου $\{X_n = m\}$

$$P(X_n = m) = P(X_n \geq m) - P(X_n \geq m + 1) =$$

$$\{1- P(X_n \leq m)\}-\{1- P(X_n \leq m+1)\} =$$
$$1-\Phi\left(\frac{m-n(p-q)}{\sqrt{4npq}}\right)-\{1-\Phi\left(\frac{m+1-n(p-q)}{\sqrt{4npq}}\right)$$
$$=\Phi\left(\frac{m+1-n(p-q)}{\sqrt{4npq}}\right)-\Phi\left(\frac{m-n(p-q)}{\sqrt{4npq}}\right)$$

Μπορούμε όμως και χωρίς να κάνουμε χρήση του κεντρικού οριακού θεωρήματος να υπολογίσουμε την πιθανότητα η στοχαστική ανέλιξη των ανεξάρτητων προσαυξήσεων μας να βρεθεί στο σημείο m.Ορίσαμε στην αρχή του κεφαλαίου τη στοχαστική διαδικασία ανεξάρτητων προσαυξήσεων $\{X_n : n = 1,2,3, \ldots\}$ ανεξάρτητων προσαυξήσεων με την τυχαία μεταβλητή X_n να ορίζεται ως εξής

$$X_n = X_0 + Z_1 + Z_2 + \cdots Z_n \qquad n = 1,2,3, \ldots$$

Όπου X_0 είναι η θέση εκκίνησης(ορίζουμε ως αρχική θέση εκκίνησης $X_0 = 0$ και Z_i ανεξάρτητες διακριτές τυχαίες μεταβλητές που περιγράφουν τα βήματα του τυχαίου περιπάτου οι οποίες έχουν κατανομή

$$Z_i = \begin{cases} +1 & \mu\varepsilon\ \pi\iota\theta\alpha\nu\acute{o}\tau\eta\tau\alpha \quad p \\ -1 & \mu\varepsilon\ \pi\iota\theta\alpha\nu\acute{o}\tau\eta\tau\alpha \quad q \end{cases}$$

δηλαδή η τυχαία μεταβλητή Z_i περιγράφει τα ακανόνιστα και τυχαία βήματα του τυχαίου περιπάτου μας. Η κατανομή της τυχαίας μεταβλητής Z_i είναι διακριτή αλλά δεν ταιριάζει με καμία διακριτή κατανομή από τις ήδη γνωστές.

Μπορούμε να εισάγουμε στην ανάλυση μας μια καινούργια μεταβλητή την

$$Y_i = \tfrac{1}{2}(Z_i + 1) \qquad i = 1,2,3, \ldots$$

Η παραπάνω τυχαία μεταβλητή είναι και αυτή διακριτή μεταβλητή και παίρνει μόνο δύο τιμές

$$Y_i = \frac{1}{2}(Z_i + 1) = \begin{cases} 1 & \mu\varepsilon\ \pi\iota\theta\alpha\nu\acute{o}\tau\eta\tau\alpha \quad p \\ 0 & \mu\varepsilon\ \pi\iota\theta\alpha\nu\acute{o}\tau\eta\tau\alpha \quad q \end{cases}$$

Η $Y_i \sim B(1,p)$ και το άθροισμα $\sum_{i=1}^{n} Y_i \sim B(n,p)$.Σύμφωνα λοιπόν με όλα τα παραπάνω τα ενδεχόμενα $\{X_n = m\}\kappa\alpha\iota\ \{\sum_{i=1}^{n} Y_i = \frac{1}{2}(m+n)\}$ είναι ισοπίθανα
Έτσι χωρίς την υπόθεση του μεγάλου αριθμού βημάτων του τυχαίου περιπάτου μπορούμε να υπολογίσουμε την πιθανότητα

$$P(\{X_n = m\}) = P\left(\left\{\sum_{i=1}^{n} Y_i = \frac{1}{2}(m+n)\right\}\right) =$$

$$\binom{n}{\frac{1}{2}(m+n)} p^{\frac{1}{2}(m+n)} q^{n-\frac{1}{2}(m+n)} \,, \quad m=-n, \ldots \ldots n \;.$$

Η παραπάνω σχέση στην ειδική περίπτωση που $p = q = 1/2$ δηλαδή στην περίπτωση που ο τυχαίος περίπατος είναι συμμετρικός μπορεί να γραφεί

$$P\left(\left\{\sum_{i=1}^{n} Y_i = \frac{1}{2}(m+n)\right\}\right) =$$

$$= \binom{n}{\frac{1}{2}(m+n)} \; 2^{-n} \qquad .$$

Όπως είναι κατανοητό ο διωνυμικός συντελεστής παίρνει την τιμή μηδέν εκτός αν το $\frac{1}{2}(m+n)$ είναι ένας ακέραιος αριθμός μεταξύ του 0 $και\ n$.
Για να βρεθεί ο άνθρωπος του παραδείγματος μας στο σημείο μηδέν θα πρέπει το άθροισμα των βημάτων του τυχαίου περιπάτου να ισούται με μηδέν. Θα πρέπει δηλαδή

$$S_n = \sum_{i=1}^{n} Y_i = 0$$

Και η αντίστοιχη πιθανότητα θα είναι

$$P\left(\left\{\sum_{i=1}^{n} Y_i = 0\right\}\right) = \binom{n}{\frac{1}{2}n} \; 2^{-n}$$

Και για άρτιο αριθμό βημάτων δηλαδή για $n = 2v$ η παραπάνω σχέση θα μας δώσει

$$P(\{S_n = 0\}) = \binom{2v}{v} \; 2^{-2v}$$

Με τη βοήθεια της σχέσης που ο *Stirling[6]* προτείνει

$$n! \sim \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n}$$

Η παραπάνω σχέση δίνει το εξής αποτέλεσμα

$$P(\{S_n = 0\}) = \binom{2v}{v} \; 2^{-2v} = \; \frac{1}{\sqrt{\pi v}} = u_n$$

Ο παραπάνω τύπος μας δίνει την δυνατότητα να υπολογίσουμε την πιθανότητα να βρεθεί ο τυχαίος περίπατος στην κατάσταση $X_0 = 0$.

Ιδιαίτερο ενδιαφέρον έχει η πιθανότητα να βρεθεί ο τυχαίος περίπατος για πρώτη φορά στο σημείο $X_0 = 0$.

Για να βρεθεί ο τυχαίος περίπατος $\{X_n : n = 1,2,3,\dots\}$ για πρώτη φορά από την αρχή της πορείας του στην κατάσταση $X_0 = 0$ θα πρέπει σύμφωνα με τη λογική να ισχύει το παρακάτω ενδεχόμενο

$$\{S_1 \neq 0, S_2 \neq 0, \dots S_{2n-1} \neq 0, S_{2n} = 0\}$$

Η πιθανότητα του παραπάνω ενδεχομένου μπορεί να γραφτεί ως εξής

$$P(\{S_1 \neq 0, S_2 \neq 0, \dots S_{2n-1} \neq 0, S_{2n} = 0\}) = f_{2n}$$

Αν σκεφτούμε λογικά θα πρέπει να υπάρχει κάποια σχέση μεταξύ των πιθανοτήτων να βρεθεί ο τυχαίος περίπατος στην κατάσταση $X_0 = 0$ ύστερα από άρτιο αριθμό βημάτων κάποια χρονική στιγμή στο μέλλον και της πιθανότητας να βρεθεί στην αρχική κατάσταση για πρώτη φορά.

$$u_{2n} = f_2 u_{2n-2} + f_4 u_{2n-4} + \dots + f_{2n} u_0 = \sum_{i=2j}^{n} f_i \, u_{2n-i} \qquad i = 1,2,3 \dots$$

Δηλαδή το ενδεχόμενο να βρεθούμε στην αρχική μας κατάσταση είναι ίδιο με το ενδεχόμενο να βρεθούμε μόνο μια φορά στην αρχική μας κατάσταση μετά από $2n$ βήματα ή να βρεθούμε για πρώτη φορά μετά από 2 βήματα και να επανέλθουμε ξανά κάποια στιγμή στα επόμενα $2n - 2$ βήματα, ή να βρεθούμε για πρώτη φορά στην αρχική μας κατάσταση μετά από 4 βήματα και να επανέλθουμε ξανά κάποια στιγμή στα επόμενα $2n - 4$ βήματα κοκ.

Η πιθανότητα

$$P(\{S_1 \neq 0, S_2 \neq 0, \dots S_{2n-1} \neq 0, S_{2n} = 0\}) = f_{2n}$$

μπορεί να υπολογιστεί με τη βοήθεια ενός βασικού λήμματος

Η πιθανότητα του ενδεχομένου στα πρώτα $2n$ βήματα να μην βρεθούμε στην αρχική μας κατάσταση ισούται με την πιθανότητα του ενδεχομένου την χρονική στιγμή $2n$ να βρεθούμε στην αρχική μας κατάσταση. Δηλαδή

$$P(\{S_1 \neq 0, S_2 \neq 0, \dots S_{2n-1} \neq 0, S_{2n} \neq 0\}) = P(S_{2n} = 0) = u_{2n}$$

Η πιθανότητα λοιπόν

$$f_{2n} = P(\{S_1 \neq 0, S_2 \neq 0, \dots S_{2n-2} \neq 0, S_{2n} = 0\}) =$$

$$P(\{S_1 \neq 0, S_2 \neq 0, \dots S_{2n-2} \neq 0\}) - P(\{S_1 \neq 0, S_2 \neq 0, \dots S_{2n-2} \neq 0, S_{2n} \neq 0\}) =$$

$$= u_{2n-2} - u_{2n}$$

3.9 Το θεώρημα της αντανάκλασης(reflection principle)

Πριν αναφέρουμε το θεώρημα της αντανάκλασης είναι σημαντικό να κάνουμε μερικές διευκρινήσεις.

Αρχικά μέχρι στιγμής θεωρούμε ότι έχουμε έναν απλό τυχαίο περίπατο με μέγεθος βήματος ίσο με τη μονάδα. Δηλαδή κάθε βήμα Z_i παίρνει τις τιμές ένα και μείον ένα με πιθανότητα $p \; και \; q$ αντίστοιχα.

Η θέση του σωματιδίου ύστερα από n βήματα δίνεται από τη σχέση

$$X_n = X_0 + \sum_{i=1}^{n} Z_i$$

Η παραπάνω σχέση θα μπορούσε να γραφτεί με τη βοήθεια μερικών αθροισμάτων ως εξής

$$X_n = X_0 + \sum_{i=1}^{\kappa} Z_i + \sum_{\kappa+1}^{n} Z_i$$

μπορούμε να συμβολίσουμε με

$$S_k = \sum_{i=1}^{\kappa} Z_i$$

το μερικό άθροισμα των k πρώτων όρων της ακολουθίας και με

$$S_{k+1} = \sum_{i=1}^{\kappa+1} Z_i$$

Το μερικό άθροισμα των $k+1$ πρώτων όρων της ακολουθίας. Έτσι η διαφορά που προκύπτει αν αφαιρέσουμε τις δύο αυτές ποσότητες είναι ίση με

$$S_{k+1} - S_k = \sum_{i=1}^{\kappa+1} Z_i - \sum_{i=1}^{\kappa} Z_i = Z_k$$

Όπου

$$Z_k = \begin{cases} +1 \; με \; πιθανότητα \; p \\ -1 \; με \; πιθανότητα \; q \end{cases}$$

Και η μέση τιμή της διαφοράς των μερικών αθροισμάτων ισούται με τη μέση τιμή της τυχαίας θέσης που θα βρεθεί ο τυχαίος μας περίπατος τη χρονική στιγμή k δηλαδή

$$E\left(\sum_{i=1}^{\kappa+1} Z_i - \sum_{i=1}^{\kappa} Z_i\right) = E(Z_k) = p - q$$

Με S_n ορίζουμε το άθροισμα των n όρων του τυχαίου μας περιπάτου δηλαδή το S_n θα ταυτίζεται με το X_n

Έστω λοιπόν ότι τις πρώτες n ακέραιες χρονικές στιγμές P ήταν οι κινήσεις προς τα δεξιά με πιθανότητα p και $Q = n - P$ οι κινήσεις προς τα αριστερά με πιθανότητα q.

Μπορούμε ακόμη να ορίσουμε

$$S_0 = 0$$

Στην προσπάθεια μας να μελετήσουμε τον τυχαίο περίπατο θα εισάγουμε στην ανάλυση μας έννοιες από την γεωμετρία.

Αρχικά θεωρούμε ένα δισδιάστατο χώρο με οριζόντιο άξονα τον t και κάθετο άξονα τον x.

Η ακολουθία $\{Z_1, Z_2, \ldots, Z_n\}$ μπορεί να παρασταθεί γραφικά σαν μια πολυγωνική γραμμή στην οποία η $k - o\sigma\tau$ή πλευρά έχει κλίση Z_k και η $k - o\sigma\tau$ή κορφή έχει τεταγμένη S_k.Τέτοιες γραμμές στην ανάλυση μας θα τις ονομάζουμε μονοπάτια *(paths)*

Ορισμός

Έστω ακέραιοι x και $n > 0$.Ένα μονοπάτι $(S_1, S_2, .., S_n)$ από την αρχή των αξόνων δηλαδή από το σημείο $O(0,0)$ μέχρι το σημείο (n, x) είναι μια πολυγωνική γραμμή της οποίας οι κορυφές έχουν τετμημένη $0,1,2, \ldots, n$ και τεταγμένη $S_0, S_1, S_2, .., S_n$ $\alpha\nu\tau\acute{\iota}\sigma\tau o\iota\chi\alpha$

και ικανοποιούν τις σχέσεις

$$S_{k+1} - S_k = \sum_{i=1}^{\kappa+1} Z_i - \sum_{i=1}^{\kappa} Z_i = Z_k$$

και

$$S_0 = 0 \; \kappa\alpha\iota \; S_n = x = P - Q$$

Τώρα που ορίσαμε το μονοπάτι μπορούμε να ορίσουμε το μέγεθος του το οποίο θα είναι ίσο με n. Λογικό αν σκεφτούμε πως ο χρόνος είναι διακριτός και κάθε ακέραια χρονική στιγμή μπορούμε να ορίσουμε τη θέση του τυχαίου μας περιπάτου.

Το μέγεθος λοιπόν n θα είναι ίσο με

$$n = P + Q$$

Γενικότερα μπορούμε να πούμε ότι υπάρχουν 2^n διαφορετικά μονοπάτια μεγέθους n και αυτό διότι κάθε χρονική στιγμή το σωματίδιο μπορεί να κινηθεί είτε δεξιά είτε αριστερά

$$Z_k = \begin{cases} +1 \; \mu\varepsilon \; \pi\iota\theta\alpha\nu\acute{o}\tau\eta\tau\alpha \; p \\ -1 \; \mu\varepsilon \; \pi\iota\theta\alpha\nu\acute{o}\tau\eta\tau\alpha \; q \end{cases}$$

Δηλαδή υπάρχουν δύο δυνατά ενδεχόμενα ,η θέση όμως του τυχαίου περιπάτου τη χρονική στιγμή n δίνεται από τη σχέση

$$X_n = X_0 + \sum_{i=1}^{n} Z_i$$

Και αν κάνουμε την παραδοχή πως $X_0 = 0$

$$X_n = \sum_{i=1}^{n} Z_i$$

Λόγω ανεξαρτησίας των Z_i κάθε θέση του μονοπατιού $(S_1, S_2, .., S_n)$ μπορεί να πάρει δύο τιμές και χρησιμοποιώντας βασικές αρχές της συνδυαστικής όπως τον πολλαπλασιαστικό κανόνα οι πιθανοί συνδυασμοί όλων των δυνατών μονοπατιών είναι ίσοι με

$$\underbrace{2 * 2 * 2 * \ldots . * 2}_{n\ \varphi o\rho \acute{\varepsilon} \varsigma} = 2^n$$

Πιο συγκεκριμένα ένα μονοπάτι από την αρχή των αξόνων έως το σημείο (n, x) υπάρχει μόνο αν πληρούνται οι παραπάνω προϋποθέσεις σε αυτή την περίπτωση τα P διαφορετικά σημείο που η αντίστοιχη Z_i είναι ίση με τη μονάδα μπορούν να επιλεγούν ως εξής

$$N_{n,x} = \binom{P+Q}{P} = \frac{(P+Q)!}{P!\,Q!} = \binom{P+Q}{Q}$$

Αν επιλέξουμε λοιπόν τα P σημεία την αρχή αλλά και το τέλος του μονοπατιού τότε ο παραπάνω συνδυασμός δίνει όλες τα δυνατά μονοπάτια από το σημείο $O(0,0)$ μέχρι το σημείο (n, x).Θεωρούμε επίσης πως όποια $n\ και\ x$ δεν ικανοποιούν τις σχέσεις

$$n = P + Q$$

και

$$x = P - Q$$

τότε θα έχουν

$$N_{n,x} = 0$$

Με τη βοήθεια της σχέσης που ο *Stirling* προτείνει

$$n! \sim \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n}$$

η σχέση

$$N_{n,x} = \binom{P+Q}{P} = \frac{(P+Q)!}{P!\,Q!}$$

μπορεί να γραφτεί

$$N_{n,x} \cong \frac{\sqrt{2\pi}(P+Q)^{P+Q+\frac{1}{2}}e^{-(P+Q)}}{\sqrt{2\pi}P^{P+\frac{1}{2}}e^{-P}\sqrt{2\pi}Q^{Q+\frac{1}{2}}e^{-Q}}$$

και με πράξεις η παραπάνω σχέση μπορεί να μας δώσει

$$N_{n,x} \cong \frac{(P+Q)^{P+Q+\frac{1}{2}}}{\sqrt{2\pi}P^{P+\frac{1}{2}}Q^{Q+\frac{1}{2}}} = \left(\frac{P+Q}{P}\right)^{P+\frac{1}{2}}\left(\frac{P+Q}{Q}\right)^{Q}\frac{1}{\sqrt{2\pi Q}}$$

Αρχή αντανάκλασης

Το πλήθος των μονοπατιών από το σημείο Α στο σημείο Β που ακουμπάει η διασχίζει τον άξονα x είναι ακριβώς ίσο με το πλήθος των μονοπατιών από το σημείοA'στο Β

Απόδειξη

Θεωρείστε ένα μονοπάτι $(S_a = a, S_{a+1}, \dots S_b = \beta)$ από το σημείο Α στο σημείο Β το οποίο έχει περισσότερες από μια κορυφές στον άξονα t.Έστω Τα να είναι η τετμημένη της πρώτης τέτοιας κορυφής, αυτό θα σημαίνει πως $S_a > 0, S_{a+1} > 0, \dots S_{T-1} > 0, S_T = 0$.

Τότε το μονοπάτι $(-S_a, -S_{a+1}, \dots, -S_{T-1}, S_T = 0, S_{T+1}, S_{T+2\dots,}S_b)$ είναι ένα άλλο μονοπάτι που οδηγεί από το σημείο A'στο Β και έχει στο σημείο$T(t,0)$ την πρώτη κορυφή στον άξονα t.Το τμήμα $A'T$ είναι η αντανάκλαση του τμήματος ΑΤ και υπάρχει μια ένα προς ένα αντιστοιχία μεταξύ όλων των μονοπατιών από το A'στο Β και από το Α στο Β.

3.10 Απλός τυχαίος περίπατος με απορροφητικά φράγματα

Σε αυτή την περίπτωση ο τυχαίος περίπατος θα σταματήσει μόλις περάσει σε μία από τις δύο ακραίες καταστάσεις $-r, m$. Η πιθανότητα να περάσει κάποια στιγμή από μια από τις δύο ακραίες αυτές καταστάσεις και να βρεθεί εκτός του διαστήματος $(-r, m)$ είναι ίση με τη μονάδα μιας και έχουμε ήδη δείξει ότι

$$P(-r < X_n \le m) = \Phi\left(\frac{m - n(p-q)}{\sqrt{4npq}}\right) - \Phi\left(\frac{-r - n(p-q)}{\sqrt{4npq}}\right) = 0 \quad \text{για} \quad n \uparrow \infty$$

Υποθέτοντας πως τα απορροφητικά φράγματα $-r, m$ δεν είναι πολύ κοντά στο μηδέν έτσι ώστε ο τυχαίος περίπατος να έχει διάρκεια.

Ιδιαίτερο ενδιαφέρον έχει ο υπολογισμός της πιθανότητας να περάσει ο τυχαίος περίπατος από το ένα απορροφητικό φράγμα ή από το άλλο απορροφητικό φράγμα.

Θα προσπαθήσουμε να υπολογίσουμε την πιθανότητα ο τυχαίος περίπατος να περάσει από το απορροφητικό φράγμα m.

Ορίζουμε την πιθανότητα

$$\beta_i = P(X_n = m/X_0 = i)$$

Με λόγια η παραπάνω πιθανότητα είναι η πιθανότητα του ενδεχομένου ο τυχαίος περίπατος να περάσει το απορροφητικό φράγμα m δεδομένου ότι την χρονική στιγμή μηδέν (στην αρχή του πειράματος ή του παιχνιδιού) βρισκόμασταν στην κατάσταση i.

Κάνοντας ανάλυση ως το πρώτο βήμα μπορούμε να πούμε τα εξής

Μπορεί ο παίκτης στην πρώτη του απόπειρα να κερδίσει ή μπορεί και να χάσει έτσι

$$\beta_i = P(X_n = m, X_1 = i+1/X_0 = i) + P(X_n = m, X_1 = i-1/X_0 = i)$$

Όμως

$$P(X_n = m, X_1 = i+1/X_0 = i) = P(X_1 = i+1/X_0 = i)P(X_n = m/X_1 = i+1)$$

Κάνοντας πάλι χρήση της ιδιότητας της αναλλοίωτης στο χρόνο κατανομής,δηλαδή

$$P(X_n = m/X_1 = i+1) = P(X_{n-1} = m/X_0 = i+1) = \beta_{i+1}$$

Αντίστοιχα με τον ίδιο ακριβώς τρόπο

$$P(X_n = m, X_1 = i-1/X_0 = i) = P(X_1 = i-1/X_0 = i)P(X_n = m/X_1 = i-1)$$

Η παραπάνω σχέση γράφεται έτσι λόγο της μαρκοβιαννής ιδιότητας δηλαδή με απλά λόγια πως μόνο το πιο πρόσφατο παρελθόν επηρεάζει την επόμενη μέρα .

Θα κάνουμε επίσης χρήση της ιδιότητας της αναλλοίωτης στο χρόνο κατανομής δηλαδή

$$P(X_n = m/X_1 = i-1) = P(X_{n-1} = m/X_0 = i-1) = \beta_{i-1}$$

Τέλος θεωρούμε τις πιθανότητες μετάβασης ενός βήματος

$$P(X_1 = i+1/X_0 = i) = p$$

και

$$P(X_1 = i-1/X_0 = i) = q$$

Έτσι η πιθανότητα απορρόφησης στο σημείο m θα είναι ίση με

$$\beta_i = p\beta_{i+1} + q\beta_{i-1}$$

Η παραπάνω εξίσωση είναι μια εξίσωση διαφορών δευτέρας τάξης

Στην απλούστερη περίπτωση όπου $i=1$ η παραπάνω εξίσωση διαφορών δεύτερης τάξης γράφεται

$$\beta_1 = p\beta_{1+1} + q\beta_{1-1} = p\beta_2 + q\beta_0$$

Με πλευρικές ή οριακές συνθήκες

$$\beta_{-r} = 0 \; \kappa\alpha\iota \; \beta_m = 1$$

Θα πρέπει να βρούμε την χαρακτηριστική εξίσωση της παραπάνω γραμμικής εξίσωσης διαφορών .

Θέτουμε

$$\beta_{i+1} = z^{i+1}, \beta_i = z^i, \beta_{i-1} = z^{i-1}$$

Και η εξίσωση

$$\beta_i = p\beta_{i+1} + q\beta_{i-1}$$

Γράφεται στην μορφή

$$pz^{i+1} + qz^{i-1} - z^i = 0$$

Βγάζουμε κοινό παράγοντα την ποσότητα z^{i-1} και έχουμε

$$z^{i-1}(pz^2 - z + q) = 0$$

Θα πρέπει λοιπόν

$$pz^2 - z + q = 0$$

Η διακρίνουσα της παραπάνω εξίσωσης είναι ίση με

$$\Delta = 1^2 - 4pq = (p+q)^2 - 4pq = (p-q)^2$$

1η περίπτωση η διακρίνουσα θετική δηλαδή $p \neq q$

$$Z_{1,2} = \frac{1 \pm \sqrt{(p-q)^2}}{2p} = \begin{cases} \dfrac{1+(p-q)}{2p} \\ \dfrac{1-(p-q)}{2p} \end{cases}$$

Και τελικά προκύπτει πως

$$Z_{1,2} = \begin{cases} 1 \\ \\ \dfrac{q}{p} \end{cases}$$

Η γενική λύση της εξίσωσης διαφορών δεύτερης τάξης θα είναι της μορφής

$$\beta_i = c_1 Z_1^i + c_2 Z_2^i = c_1 + c_2 \left(\frac{q}{p}\right)^i$$

Υπολογισμός σταθερών της παραπάνω λύσης

Θα χρησιμοποιήσουμε της οριακές ή πλευρικές συνθήκες για τον καθορισμό των σταθερών

Η πρώτη οριακή συνθήκη είναι η εξής

$$\beta_m = 1$$

$$\beta_m = c_1 Z_1^m + c_2 Z_2^m = c_1 + c_2 \left(\frac{q}{p}\right)^m = 1$$

Η δεύτερη οριακή συνθήκη είναι η

$$\beta_{-r} = 0$$

$$\beta_{-r} = c_1 Z_1^{-r} + c_2 Z_2^{-r} = c_1 + c_2 \left(\frac{q}{p}\right)^{-r} = 0$$

Έτσι έχουμε να λύσουμε το σύστημα δύο εξισώσεων

της

$$c_1 + c_2 \left(\frac{q}{p}\right)^m = 1$$

και της

$$c_1 + c_2 \left(\frac{q}{p}\right)^{-r} = 0$$

Αφαιρώντας κατά μέλη προκύπτει ότι

$$c_2 = \left[\left(\frac{q}{p}\right)^m - \left(\frac{q}{p}\right)^{-r}\right] = 1$$

Άρα

$$c_2 = \frac{1}{\left(\frac{q}{p}\right)^m - \left(\frac{q}{p}\right)^{-r}}$$

Και αντικαθιστώντας στην πρώτη από τις δύο εξισώσεις του συστήματος έχουμε

$$c_1 = 1 - \frac{\left(\frac{q}{p}\right)^m}{\left(\frac{q}{p}\right)^m - \left(\frac{q}{p}\right)^{-r}} = - \frac{\left(\frac{q}{p}\right)^{-r}}{\left(\frac{q}{p}\right)^m - \left(\frac{q}{p}\right)^{-r}}$$

Η λύση λοιπόν της διαφορικής εξίσωσης είναι

$$\beta_i = - \frac{\left(\frac{q}{p}\right)^{-r}}{\left(\frac{q}{p}\right)^m - \left(\frac{q}{p}\right)^{-r}} + \frac{\left(\frac{q}{p}\right)^i}{\left(\frac{q}{p}\right)^m - \left(\frac{q}{p}\right)^{-r}} = \frac{\left(\frac{q}{p}\right)^i - \left(\frac{q}{p}\right)^{-r}}{\left(\frac{q}{p}\right)^m - \left(\frac{q}{p}\right)^{-r}}$$

2η περίπτωση $p = q$

Σε αυτή την περίπτωση η διακρίνουσα

$$\Delta = 1^2 - 4pq = (p+q)^2 - 4pq = (p-q)^2 = 0$$

Μπορούμε να πάρουμε το όριο της

$$\beta_i = \frac{\left(\frac{q}{p}\right)^i - \left(\frac{q}{p}\right)^{-r}}{\left(\frac{q}{p}\right)^m - \left(\frac{q}{p}\right)^{-r}}$$

$$\lim_{\frac{q}{p} \to 1} \frac{\left(\frac{q}{p}\right)^i - \left(\frac{q}{p}\right)^{-r}}{\left(\frac{q}{p}\right)^m - \left(\frac{q}{p}\right)^{-r}} = \frac{0}{0}$$

Μπορούμε κάνοντας *De l hospital* δηλαδή να παραγωγίσουμε αριθμητή και παρονομαστή στην παραπάνω σχέση να έχουμε

$$\lim_{\frac{q}{p} \to 1} \frac{i\left(\frac{q}{p}\right)^{i-1} + r\left(\frac{q}{p}\right)^{-r-1}}{m\left(\frac{q}{p}\right)^{m-1} + r\left(\frac{q}{p}\right)^{-r-1}} = \frac{i+r}{m+r}$$

Καταλήξαμε λοιπόν στο συμπέρασμα πως η πιθανότητα απορρόφησης στο σημείο m ισούται με

$$\beta_i = \begin{cases} \dfrac{\left(\frac{q}{p}\right)^i - \left(\frac{q}{p}\right)^{-r}}{\left(\frac{q}{p}\right)^m - \left(\frac{q}{p}\right)^{-r}} & \alpha\nu\ p \neq q \\[2ex] \dfrac{i+r}{m+r} & \alpha\nu\ p = q \end{cases}$$

3.11 Πιθανότητα απορρόφησης και μέση τιμή βήματος τυχαίου περιπάτου

Δείξαμε παραπάνω πως στην περίπτωση που η διακρίνουσα είναι θετική

$$\Delta = 1^2 - 4pq = (p+q)^2 - 4pq = (p-q)^2$$

Τότε

$$Z_{1,2} = \frac{1 \pm \sqrt{(p-q)^2}}{2p} = \begin{cases} \frac{1+(p-q)}{2p} \\ \frac{1-(p-q)}{2p} \end{cases}$$

Έχουμε ήδη δείξει πως η μέση τιμή του βήματος ενός τυχαίου περιπάτου

$$Z_i = \begin{cases} +1 & \mu\varepsilon\ \pi\iota\theta\alpha\nu\acute{o}\tau\eta\tau\alpha \quad p \\ -1 & \mu\varepsilon\ \pi\iota\theta\alpha\nu\acute{o}\tau\eta\tau\alpha \quad q \end{cases}$$

υπολογίζεται ως εξής

$$E(Z_i) = +1P(Z_i = 1) - 1P(Z_i = -1)$$

Όμως

$$P(Z_i = 1) = p \quad \kappa\alpha\iota\ P(Z_i = -1) = q$$

Έτσι

$$E(Z_i) = p - q = \mu$$

Και η διακύμανση

$$Var(Z_i) = E\left({Z_i}^2\right) - E(Z_i)^2$$

$$E\left({Z_i}^2\right) = p + q = 1$$

$$Var(Z_i) = E\left({Z_i}^2\right) - E(Z_i)^2 = p + q - (p-q)^2 = 4pq$$

$$Z_{1,2} = \frac{1 \pm \sqrt{1-4pq}}{p+1-q} = \begin{cases} \frac{1+\sqrt{1-var(Z_i)}}{1+E(Z_i)} \\ \frac{1-\sqrt{1-var(Z_i)}}{1+E(Z_i)} \end{cases}$$

Εναλλακτικά θα μπορούσαμε να υπολογίσουμε το $Z_{1,2}$ ως εξής

$$Z_{1,2} = \frac{1 \pm \sqrt{(p-q)^2}}{2p} = \begin{cases} 1 \\ \dfrac{1-E(Z_i)}{1+E(Z_i)} \end{cases}$$

Έτσι η γενική λύση της εξίσωσης διαφορών της δεύτερης τάξης είναι η εξής

$$\beta_i = c_1 Z_1^i + c_2 Z_2^i = c_1 + c_2 \left(\frac{1-E(Z_i)}{1+E(Z_i)}\right)^i$$

Υπολογισμός σταθερών της παραπάνω λύσης

Θα χρησιμοποιήσουμε της οριακές ή πλευρικές συνθήκες για τον καθορισμό των σταθερών

Η πρώτη οριακή συνθήκη είναι η εξής

$$\beta_m = 1$$

$$\beta_m = c_1 Z_1^m + c_2 Z_2^m = c_1 + c_2 \left(\frac{1-E(Z_i)}{1+E(Z_i)}\right)^m = 1$$

Η δεύτερη οριακή συνθήκη είναι η

$$\beta_{-r} = 0$$

$$\beta_{-r} = c_1 Z_1^{-r} + c_2 Z_2^{-r} = c_1 + c_2 \left(\frac{1-E(Z_i)}{1+E(Z_i)}\right)^{-r} = 0$$

Έτσι έχουμε να λύσουμε το σύστημα δύο εξισώσεων

της

$$c_1 + c_2 \left(\frac{1-E(Z_i)}{1+E(Z_i)}\right)^m = 1$$

και της

$$c_1 + c_2 \left(\frac{1-E(Z_i)}{1+E(Z_i)}\right)^{-r} = 0$$

Αφαιρώντας κατά μέλη προκύπτει ότι

$$c_2 = \left[\left(\frac{1-E(Z_i)}{1+E(Z_i)}\right)^m - \left(\frac{1-E(Z_i)}{1+E(Z_i)}\right)^{-r}\right] = 1$$

Άρα

$$c_2 = \frac{1}{\left(\frac{1-E(Z_i)}{1+E(Z_i)}\right)^m - \left(\frac{1-E(Z_i)}{1+E(Z_i)}\right)^{-r}}$$

Και αντικαθιστώντας στην πρώτη από τις δύο εξισώσεις του συστήματος έχουμε

$$c_1 = 1 - \frac{\left(\frac{1-E(Z_i)}{1+E(Z_i)}\right)^m}{\left(\frac{1-E(Z_i)}{1+E(Z_i)}\right)^m - \left(\frac{1-E(Z_i)}{1+E(Z_i)}\right)^{-r}} = -\frac{\left(\frac{1-E(Z_i)}{1+E(Z_i)}\right)^{-r}}{\left(\frac{1-E(Z_i)}{1+E(Z_i)}\right)^m - \left(\frac{1-E(Z_i)}{1+E(Z_i)}\right)^{-r}}$$

Η λύση λοιπόν της διαφορικής εξίσωσης είναι

$$\beta_i = -\frac{\left(\frac{1-E(Z_i)}{1+E(Z_i)}\right)^{-r}}{\left(\frac{1-E(Z_i)}{1+E(Z_i)}\right)^m - \left(\frac{1-E(Z_i)}{1+E(Z_i)}\right)^{-r}} + \frac{\left(\frac{1-E(Z_i)}{1+E(Z_i)}\right)^{i}}{\left(\frac{1-E(Z_i)}{1+E(Z_i)}\right)^m - \left(\frac{1-E(Z_i)}{1+E(Z_i)}\right)^{-r}}$$

Τελικά προκύπτει πως

$$\beta_i = \frac{\left(\frac{1-E(Z_i)}{1+E(Z_i)}\right)^{i} - \left(\frac{1-E(Z_i)}{1+E(Z_i)}\right)^{-r}}{\left(\frac{1-E(Z_i)}{1+E(Z_i)}\right)^m - \left(\frac{1-E(Z_i)}{1+E(Z_i)}\right)^{-r}} \quad \gamma\iota\alpha\ E(Z_i) \neq 0$$

Στην περίπτωση που $\boldsymbol{p} = \boldsymbol{q}$ παίρνουμε το όριο της παραπάνω σχέσης το οποίο οδηγεί σε απροσδιόριστη μορφή και κάνουμε χρήση κανόνα *De l Hospital*

$$\lim_{E(Z_i)\to 0} \frac{i\left(\frac{1-E(Z_i)}{1+E(Z_i)}\right)^{i-1} + r\left(\frac{1-E(Z_i)}{1+E(Z_i)}\right)^{-r-1}}{m\left(\frac{1-E(Z_i)}{1+E(Z_i)}\right)^{m-1} + r\left(\frac{1-E(Z_i)}{1+E(Z_i)}\right)^{-r-1}} = \frac{i+r}{m+r}$$

$$\beta_i = \begin{cases} \dfrac{\left(\frac{1-E(Z_i)}{1+E(Z_i)}\right)^{i} - \left(\frac{1-E(Z_i)}{1+E(Z_i)}\right)^{-r}}{\left(\frac{1-E(Z_i)}{1+E(Z_i)}\right)^m - \left(\frac{1-E(Z_i)}{1+E(Z_i)}\right)^{-r}} & \alpha\nu\ p \neq q \\[2ex] \dfrac{i+r}{m+r} & \alpha\nu\ p = q \end{cases}$$

3.12 Πιθανότητα απορρόφησης εκφρασμένη με τη βοήθεια της χαρακτηριστικής συνάρτησης

Έχουμε ήδη δείξει πως η πιθανότητα απορρόφησης στο σημείο m δίνεται από τη σχέση

$$\beta_i = \begin{cases} \dfrac{\left(\dfrac{1-E(Z_i)}{1+E(Z_i)}\right)^i - \left(\dfrac{1-E(Z_i)}{1+E(Z_i)}\right)^{-r}}{\left(\dfrac{1-E(Z_i)}{1+E(Z_i)}\right)^m - \left(\dfrac{1-E(Z_i)}{1+E(Z_i)}\right)^{-r}} & \alpha\nu\ p \neq q \\ \\ \dfrac{i+r}{m+r} & \alpha\nu\ p = q \end{cases}$$

Όμως

$$E(Z_i^r) = \frac{1}{i^r}\Phi_Z^{(r)}(0)$$

Και σύμφωνα με την τελευταία σχέση θα έχουμε

$$E(Z_i) = \frac{1}{i}\Phi_Z^{(1)}(0)$$

$$\beta_i = \begin{cases} \dfrac{\left(\dfrac{1-\frac{1}{i}\Phi_Z^{(1)}(0)}{1+\frac{1}{i}\Phi_Z^{(1)}(0)}\right)^i - \left(\dfrac{1-\frac{1}{i}\Phi_Z^{(1)}(0)}{1+\frac{1}{i}\Phi_Z^{(1)}(0)}\right)^{-r}}{\left(\dfrac{1-\frac{1}{i}\Phi_Z^{(1)}(0)}{1+\frac{1}{i}\Phi_Z^{(1)}(0)}\right)^m - \left(\dfrac{1-\frac{1}{i}\Phi_Z^{(1)}(0)}{1+\frac{1}{i}\Phi_Z^{(1)}(0)}\right)^{-r}} & \alpha\nu\ p \neq q \\ \\ \dfrac{i+r}{m+r} & \alpha\nu\ p = q \end{cases}$$

Μεγάλο ενδιαφέρον παρουσιάζει η χρονική στιγμή στην οποία θα περάσει η στοχαστική ανέλιξη $\{X_n : n = 1,2,3,\dots\}$ από κάποιο από τα απορροφητικά φράγματα.

Μεγάλο ενδιαφέρον έχει και ο υπολογισμός της πιθανότητας με την οποία ο τυχαίος περίπατος θα περάσει από την κατάσταση $-r$ ή από την κατάσταση m.

3.13 Χρόνος απορρόφησης

Μέχρι στιγμής θεωρούμε τον χρόνο διακριτή μεταβλητή και ταυτίζεται κάθε φόρα με τον αριθμό των βημάτων που κάνει ο τυχαίος περίπατος.

Μπορούμε να φανταστούμε λοιπόν τον χρόνο ως το άθροισμα των ενδιάμεσων χρόνων ή χρόνων αναμονής. Για να ορίσουμε όμως τον ενδιάμεσο χρόνο θα πρέπει πρώτα να εισάγουμε στην ανάλυση μας τον χρόνο άφιξης ενός γεγονότος

$$Y_i = min\{t : N(t) = i\}$$

Με λόγια η παραπάνω σχέση μας λέει πως Y_i είναι ο ελάχιστος χρόνος που απαιτείται για την επέλευση του i γεγονότος και καλείται χρόνος άφιξης του i γεγονότος .

3.14 Ενδιάμεσοι χρόνοι ή χρόνοι αναμονής :

Με βάση την ακολουθία των χρόνων άφιξης μπορούμε να ορίσουμε την ακολουθία $\{T_\kappa : k = 1,2, \dots\dots\}$ ως εξής :

$$T_1 = Y_1$$

$$T_2 = Y_2 - Y_1$$

$$\dots\dots\dots\dots$$

$$T_k = Y_k - Y_{k-1}$$

Οι μεταβλητές T_k ονομάζονται ενδιάμεσοι χρόνοι ή χρόνοι αναμονής[7].Αξίζει να σημειωθεί πως ενώ η $N(t)$ είναι μια διακριτή τυχαία μεταβλητή, οι Y_i και T_i είναι συνεχείς τυχαίες μεταβλητές ή διακριτές τυχαίες μεταβλητές ανάλογα με τη μονάδα μέτρησης που χρησιμοποιούμε για να περιγράψουμε μια χρονική στιγμή.

Τώρα είμαστε σε θέση να περιγράψουμε με μαθηματικό τρόπο τον χρόνο απορρόφηση μιας στοχαστικής διαδικασίας ως εξής

$$T = T_1 + T_2 + \cdots + T_k \; \mu\varepsilon \; k \geq \min\{-r, m\}$$

Ο χρόνος απορρόφησης είναι ο χρόνος που απαιτείται ώστε η ανέλιξη μας να βρεθεί σε κάποια από τις απορροφητικές καταστάσεις

$$T = \inf\{n : X_n = -r \text{ ή } X_n = m\}$$

3.15 Καθορισμός της κατανομής του χρόνου απορρόφησης

Θα προσπαθήσουμε να υπολογίσουμε την πιθανότητα η διακριτή τυχαία μεταβλητή T που περιγράφει τον χρόνο απορρόφησης του τυχαίου περιπάτου σε ένα από τα δύο απορροφητικά φράγματα να πάρει μια ακέραια τιμή k δοθέντος ότι η αρχική κατάσταση του τυχαίου περιπάτου είναι $X_0 = i$.

Στην γλώσσα της στατιστικής λοιπόν αναζητούμε την δεσμευμένη πιθανότητα

$$P_k^{(i)} = P(T = k / X_0 = i)$$

Αναλύοντας ως προς το πρώτο βήμα του τυχαίου περιπάτου έχουμε

$$P_k^{(i)} = P(T = k, X_1 = i + 1 / X_0 = i) + P(T = k, X_1 = i + 1 / X_0 = i)$$

$$P_k^{(i)} = pP(T = k / X_1 = i + 1) + qP(T = k / X_1 = i + 1)$$

με

$$P(X_1 = i + 1 / X_0 = i) = p$$

Και

$$P(X_1 = i-1/X_0 = i) = q$$

Και επειδή η κατανομή είναι αναλλοίωτη στον χρόνο έχουμε

$$P(T = k/X_1 = i+1) = P(T = k-1/X_0 = i+1) = P_{k-1}^{(i+1)}$$

Και αντίστοιχα

$$P(T = k/X_1 = i-1) = P(T = k-1/X_0 = i-1) = P_{k-1}^{(i-1)}$$

Έτσι η πιθανότητα

$$P_k^{(i)} = pP(T = k/X_1 = i+1) + qP(T = k/X_1 = i+1)$$

Είναι ίση με

$$P_k^{(i)} = pP_{k-1}^{(i+1)} + qP_{k-1}^{(i-1)}$$

Στην περίπτωση που ο τυχαίος περίπατος ξεκινάει από την κατάσταση $X_0 = -\alpha$ ή b

Τότε

$$P_k^{(-a)} = P_k^{(b)} = \begin{cases} 1 \ \alpha\nu \ k = 0 \\ 0 \ \alpha\nu \ k \neq 0 \end{cases}$$

3.16 Υπολογισμός πιθανογεννήτριας του χρόνου απορρόφησης

Μπορούμε να ορίσουμε την πιθανογεννήτρια ως εξής

$$\pi_i(s) = E(S^T/X_0 = i) = \sum_{k=0}^{\infty} s^k \, P(T = k/X_0 = i)$$

$$\pi_i(s) = \sum_{k=0}^{\infty} s^k \left(pP_{k-1}^{(i+1)} + qP_{k-1}^{(i-1)} \right) =$$

$$= \sum_{k=0}^{\infty} s^k \, pP_{k-1}^{(i+1)} + \sum_{k=0}^{\infty} s^k qP_{k-1}^{(i-1)} =$$

$$= s \sum_{k=1}^{\infty} s^{k-1} \, pP_{k-1}^{(i+1)} + s \sum_{k=1}^{\infty} s^{k-1} qP_{k-1}^{(i-1)}$$

Και τελικά καταλήγουμε σε μια εξίσωση διαφορών δεύτερης τάξης ως προς i

$$\pi_i(s) = sp\pi_{i+1}(s) + sq\pi_{i-1}(s)$$

Με πλευρικές ή οριακές συνθήκες

$$\pi_{-\alpha}(s) = \sum_{k=0}^{\infty} s^k \, P(T = k/X_0 = -a)$$

$$\pi_{-\alpha}(s) = \mathrm{s}^0 P(T = 0/X_0 = -a) + \sum_{k=1}^{\infty} s^k \, P(T = k/X_0 = -a) =$$

Και τελικά προκύπτει πως

$$\pi_{-\alpha}(s) = 1$$

Διότι

$$P_k^{(-a)} = P(T = k/X_0 = -a) = \begin{cases} 1 \; \alpha\nu \; k = 0 \\ 0 \; \alpha\nu \; k \neq 0 \end{cases}$$

Και έτσι το άθροισμα

$$\sum_{k=1}^{\infty} s^k \, P(T = k/X_0 = -a) = 0$$

Αντίστοιχα

$$\pi_b(s) = 1$$

Γενική λύση εξίσωσης διαφορών δεύτερης τάξης

Στην ουσία έχουμε να διαχειριστούμε μια ομογενή εξίσωση διαφορών δεύτερης τάξης

$$sp\pi_{i+1}(s) - \pi_i(s) + sq\pi_{i-1}(s) = 0$$

Με πλευρικές συνθήκες

$$\pi_{-\alpha}(s) = \pi_b(s) = 1$$

Θέτουμε

$$\pi_{i-1}(s) = \lambda^{i-1}, \pi_i(s) = \lambda^i, \pi_{i+1}(s) = \lambda^{i+1}$$

Και η διαφοροεξίσωση γράφεται

$$sp\lambda^{i+1} - \lambda^i + sq\lambda^{i-1} = 0$$

$$\lambda^{i-1}(sp\lambda^2 - \lambda + sq) = 0$$

Ισοδύναμα

$$sp\lambda^2 - \lambda + sq = 0$$

Το παραπάνω τριώνυμο έχει διακρίνουσα ίση με

$$\Delta = 1 - 4pqs^2 = 1 - s^2 var(Z_i)$$

Σε περίπτωση που η διακρίνουσα είναι θετική δηλαδή

$$1 - 4pqs^2 > 0 \Rightarrow 1 > 4pqs^2 \Rightarrow \frac{1}{4pq} > s^2 \Rightarrow \frac{1}{2\sqrt{pq}} > |s|$$

Υπάρχουν δύο ρίζες πραγματικές και άνισες

$$\lambda_{1,2} = \frac{1 \pm \sqrt{1 - 4pqs^2}}{2sp} = \frac{1 \pm \sqrt{1 - var(Z_i)s^2}}{(1 + E(Z_i))s}$$

Και η γενική λύση είναι

$$\pi_i(s) = c_1\lambda_1^i + c_2\lambda_2^i$$

Ειδικότερα

$$\pi_i(s) = c_1\left(\frac{1 + \sqrt{1 - 4pqs^2}}{2sp}\right)^i + c_2\left(\frac{1 - \sqrt{1 - 4pqs^2}}{2sp}\right)^i$$

Εναλλακτικά θα μπορούσαμε την παραπάνω γενική λύση να την εκφράσουμε ως εξής

$$\pi_i(s) = c_1\left(\frac{1 + \sqrt{1 - var(Z_i)s^2}}{(1 + E(Z_i))s}\right)^i + c_2\left(\frac{1 - \sqrt{1 - var(Z_i)s^2}}{(1 + E(Z_i))s}\right)^i$$

Υπολογισμός σταθερών $\boldsymbol{c_1, c_2}$

Υπολογισμός σταθερών της παραπάνω λύσης

Θα χρησιμοποιήσουμε της οριακές ή πλευρικές συνθήκες για τον καθορισμό των σταθερών

Η πρώτη οριακή συνθήκη είναι η εξής

$$\pi_{-\alpha}(s) = 1$$

$$\pi_{-\alpha}(s) = c_1\left(\frac{1 + \sqrt{1 - 4pqs^2}}{2sp}\right)^{-\alpha} + c_2\left(\frac{1 - \sqrt{1 - 4pqs^2}}{2sp}\right)^{-\alpha} = 1$$

Η δεύτερη οριακή συνθήκη είναι η

$$\pi_b(s) = 1$$

$$\pi_b(s) = c_1\left(\frac{1+\sqrt{1-4pqs^2}}{2sp}\right)^b + c_2\left(\frac{1-\sqrt{1-4pqs^2}}{2sp}\right)^b = 1$$

Έτσι έχουμε να λύσουμε το σύστημα δύο εξισώσεων

της

$$c_1\left(\frac{1+\sqrt{1-4pqs^2}}{2sp}\right)^{-\alpha} + c_2\left(\frac{1-\sqrt{1-4pqs^2}}{2sp}\right)^{-\alpha} = 1$$

και της

$$c_1\left(\frac{1+\sqrt{1-4pqs^2}}{2sp}\right)^b + c_2\left(\frac{1-\sqrt{1-4pqs^2}}{2sp}\right)^b = 1$$

Το παραπάνω σύστημα μπορεί να λυθεί με τη βοήθεια των οριζουσών

$$D = \begin{vmatrix} \left(\frac{1+\sqrt{1-4pqs^2}}{2sp}\right)^{-\alpha} & \left(\frac{1-\sqrt{1-4pqs^2}}{2sp}\right)^{-\alpha} \\ \left(\frac{1+\sqrt{1-4pqs^2}}{2sp}\right)^b & \left(\frac{1-\sqrt{1-4pqs^2}}{2sp}\right)^b \end{vmatrix}$$

$$D=\left(\frac{1+\sqrt{1-4pqs^2}}{2sp}\right)^{-\alpha}\left(\frac{1-\sqrt{1-4pqs^2}}{2sp}\right)^b - \left(\frac{1-\sqrt{1-4pqs^2}}{2sp}\right)^{-\alpha}\left(\frac{1+\sqrt{1-4pqs^2}}{2sp}\right)^b$$

$$D_1 = \begin{vmatrix} 1 & \left(\frac{1-\sqrt{1-4pqs^2}}{2sp}\right)^{-\alpha} \\ 1 & \left(\frac{1-\sqrt{1-4pqs^2}}{2sp}\right)^b \end{vmatrix}$$

$$D_1 = \left(\frac{1-\sqrt{1-4pqs^2}}{2sp}\right)^b - \left(\frac{1-\sqrt{1-4pqs^2}}{2sp}\right)^{-\alpha}$$

$$D_2 = \begin{vmatrix} \left(\frac{1+\sqrt{1-4pqs^2}}{2sp}\right)^{-\alpha} & 1 \\ \left(\frac{1+\sqrt{1-4pqs^2}}{2sp}\right)^{b} & 1 \end{vmatrix}$$

$$D_2 = \left(\frac{1+\sqrt{1-4pqs^2}}{2sp}\right)^{-a} - \left(\frac{1+\sqrt{1-4pqs^2}}{2sp}\right)^{b}$$

Η λύση λοιπόν του συστήματος είναι η

$$c_1 = \frac{\left(\frac{1-\sqrt{1-4pqs^2}}{2sp}\right)^{b} - \left(\frac{1-\sqrt{1-4pqs^2}}{2sp}\right)^{-\alpha}}{\left(\frac{1+\sqrt{1-4pqs^2}}{2sp}\right)^{-\alpha}\left(\frac{1-\sqrt{1-4pqs^2}}{2sp}\right)^{b} - \left(\frac{1-\sqrt{1-4pqs^2}}{2sp}\right)^{-\alpha}\left(\frac{1+\sqrt{1-4pqs^2}}{2sp}\right)^{b}}$$

Και

$$c_2 = \frac{\left(\frac{1+\sqrt{1-4pqs^2}}{2sp}\right)^{-a} - \left(\frac{1+\sqrt{1-4pqs^2}}{2sp}\right)^{b}}{\left(\frac{1+\sqrt{1-4pqs^2}}{2sp}\right)^{-\alpha}\left(\frac{1-\sqrt{1-4pqs^2}}{2sp}\right)^{b} - \left(\frac{1-\sqrt{1-4pqs^2}}{2sp}\right)^{-\alpha}\left(\frac{1+\sqrt{1-4pqs^2}}{2sp}\right)^{b}}$$

3.17 Υπολογισμός ροπογεννήτριας του χρόνου απορρόφησης

Μπορούμε να ορίσουμε την πιθανογεννήτρια ως εξής

$$M_i(s) = E(e^{sT}/X_0 = i) = \sum_{k=0}^{\infty} e^{sk}\, P(T = k/X_0 = i)$$

$$M_i(s) = \sum_{k=0}^{\infty} e^{sk}\left(pP_{k-1}^{(i+1)} + qP_{k-1}^{(i-1)}\right) =$$

$$= \sum_{k=0}^{\infty} e^{sk}\, pP_{k-1}^{(i+1)} + \sum_{k=0}^{\infty} e^{sk} qP_{k-1}^{(i-1)} =$$

$$= pe^{s} \sum_{k=1}^{\infty} s^{k-1}\, P_{k-1}^{(i+1)} + qe^{s} \sum_{k=1}^{\infty} s^{k-1} P_{k-1}^{(i-1)}$$

Και τελικά καταλήγουμε σε μια εξίσωση διαφορών δεύτερης τάξης ως προς i

$$M_i(s) = pe^{s} M_{i+1}(s) + qe^{s} M_{i-1}(s)$$

Με πλευρικές ή οριακές συνθήκες

$$M_{-\alpha}(s) = \sum_{k=0}^{\infty} e^{sk}\, P(T = k/X_0 = -a)$$

$$M_{-\alpha}(s) = \mathrm{s}^{0} P(T = 0/X_0 = -a) + \sum_{k=1}^{\infty} s^{k}\, P(T = k/X_0 = -a) =$$

Και τελικά προκύπτει πως

$$M_{-\alpha}(s) = 1$$

Διότι

$$P_k^{(-a)} = P(T = k/X_0 = -a) = \begin{cases} 1 \ \alpha\nu \ k = 0 \\ 0 \ \alpha\nu \ k \neq 0 \end{cases}$$

Και έτσι το άθροισμα

$$\sum_{k=1}^{\infty} e^{sk}\, P(T = k/X_0 = -a) = 0$$

Αντίστοιχα

$$M_b(s) = 1$$

Γενική λύση εξίσωσης διαφορών δεύτερης τάξης

Στην ουσία έχουμε να διαχειριστούμε μια ομογενή εξίσωση διαφορών δεύτερης τάξης

$$pe^{s} M_{i+1}(s) - M_i(s) + qe^{s} M_{i-1}(s) = 0$$

Με πλευρικές συνθήκες

$$M_{-\alpha}(s) = M_b(s) = 1$$

Θέτουμε

$$M_{i-1}(s) = \lambda^{i-1}, M_i(s) = \lambda^i, M_{i+1}(s) = \lambda^{i+1}$$

Και η διαφοροεξίσωση γράφεται

$$e^s p\lambda^{i+1} - \lambda^i + e^s q\lambda^{i-1} = 0$$

$$\lambda^{i-1}(e^s p\lambda^2 - \lambda + e^s q) = 0$$

Ισοδύναμα

$$e^s p\lambda^2 - \lambda + e^s q = 0$$

Το παραπάνω τριώνυμο έχει διακρίνουσα ίση με

$$\Delta = 1 - 4pq(e^s)^2 = 1 - (e^s)^2 var(Z_i)$$

Σε περίπτωση που η διακρίνουσα είναι θετική δηλαδή

$$1 - 4pq(e^s)^2 > 0 \Rightarrow 1 > 4pq(e^s)^2 \Rightarrow \frac{1}{4pq} > (e^s)^2 \Rightarrow \log\left(\frac{1}{4pq}\right) > 2s$$

δηλαδή

$$\frac{-\log 4 - \log(pq)}{2} > s$$

Υπάρχουν δύο ρίζες πραγματικές και άνισες

$$\lambda_{1,2} = \frac{1 \pm \sqrt{1 - 4pq(e^s)^2}}{2e^s p} = \frac{1 \pm \sqrt{1 - var(Z_i)(e^s)^2}}{(1 + E(Z_i))e^s}$$

Και η γενική λύση είναι

$$M_i(s) = c_1\lambda_1^i + c_2\lambda_2^i$$

Ειδικότερα

$$M_i(s) = c_1\left(\frac{1 + \sqrt{1 - 4pq(e^s)^2}}{2e^s p}\right)^i + c_2\left(\frac{1 - \sqrt{1 - 4pq(e^s)^2}}{2e^s p}\right)^i$$

Εναλλακτικά θα μπορούσαμε την παραπάνω γενική λύση να την εκφράσουμε ως εξής

$$M_i(s) = c_1\left(\frac{1+\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^i + c_2\left(\frac{1-\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^i$$

Υπολογισμός σταθερών c_1, c_2

Υπολογισμός σταθερών της παραπάνω λύσης

Θα χρησιμοποιήσουμε της οριακές ή πλευρικές συνθήκες για τον καθορισμό των σταθερών

Η πρώτη οριακή συνθήκη είναι η εξής

$$M_{-\alpha}(s) = 1$$

$$M_{-\alpha}(s) = c_1\left(\frac{1+\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^{-\alpha} + c_2\left(\frac{1-\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^{-\alpha} = 1$$

Η δεύτερη οριακή συνθήκη είναι η

$$M_b(s) = 1$$

$$M_b(s) = c_1\left(\frac{1+\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^b + c_2\left(\frac{1-\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^b = 1$$

Έτσι έχουμε να λύσουμε το σύστημα δύο εξισώσεων

της

$$c_1\left(\frac{1+\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^{-\alpha} + c_2\left(\frac{1-\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^{-\alpha} = 1$$

και της

$$c_1\left(\frac{1+\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^b + c_2\left(\frac{1-\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^b = 1$$

Το παραπάνω σύστημα μπορεί να λυθεί με τη βοήθεια των οριζουσών

$$D=\begin{vmatrix}\left(\frac{1+\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^{-\alpha} & \left(\frac{1-\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^{-\alpha}\\ \left(\frac{1+\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^{b} & \left(\frac{1-\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^{b}\end{vmatrix}$$

$$D=\left(\frac{1+\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^{-\alpha}\left(\frac{1-\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^{b} -\left(\frac{1-\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^{-\alpha}\left(\frac{1+\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^{b}$$

$$D_1=\begin{vmatrix}1 & \left(\frac{1-\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^{-\alpha}\\ 1 & \left(\frac{1-\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^{b}\end{vmatrix}$$

$$D_1=\left(\frac{1-\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^{b}-\left(\frac{1-\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^{-\alpha}$$

$$D_2=\begin{vmatrix}\left(\frac{1+\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^{-\alpha} & 1\\ \left(\frac{1+\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^{b} & 1\end{vmatrix}$$

$$D_2=\left(\frac{1+\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^{-\alpha}-\left(\frac{1+\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^{b}$$

Η λύση λοιπόν του συστήματος είναι η

$$c_1=\frac{\left(\frac{1-\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^{b}-\left(\frac{1-\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^{-\alpha}}{\left(\frac{1+\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^{-\alpha}\left(\frac{1-\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^{b}-\left(\frac{1-\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^{-\alpha}\left(\frac{1+\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^{b}}$$

Και

c_2

$$= \frac{\left(\frac{1+\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^{-\alpha} - \left(\frac{1+\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^{b}}{\left(\frac{1+\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^{-\alpha}\left(\frac{1-\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^{b} - \left(\frac{1-\sqrt{1-var(Z_i)(e^s)^2}}{(1+E(Z_i))e^s}\right)^{-\alpha}\left(\frac{1+\sqrt{1-var(Z_i}{(1+E(Z_i))e}\right.}$$

3.18 Χαρακτηριστική συνάρτηση χρόνου απορρόφησης

Οι χαρακτηριστικές συναρτήσεις είναι λίγο πιο πολύπλοκες από τις ροπογεννήτριες (και τις πιθανογεννήτριες) διότι στον ορισμό τους εμπλέκονται μιγαδικοί αριθμοί. Έχουν όμως δύο σημαντικά πλεονεκτήματα σε σύγκριση με τις ροπογεννήτριες. Το πρώτο είναι το γεγονός ότι, σε αντίθεση με τις ροπογεννήτριες οι οποίες απαιτούν την ύπαρξη της μέσης τιμής $E(e^{tX})$ για τιμές του t σε κάποιο διάστημα της μορφής $(-\delta, \delta)$, $\delta > 0$, οι χαρακτηριστικές συναρτήσεις υπάρχουν πάντοτε δηλαδή για όλες τις κατανομές και για όλες τις τιμές του $t \in R$.

Το δεύτερο είναι το γεγονός ότι η συνάρτηση κατανομής αλλά και η συνάρτηση πιθανότητας (ή πυκνότητας) μιας τυχαίας μεταβλητής προκύπτουν από την χαρακτηριστική της συνάρτηση μέσω τύπων αντιστροφής.

Ορίζουμε με $\Phi_l(u)$ τη χαρακτηριστική συνάρτηση του χρόνου απορρόφησης σε ένα από τα δύο απορροφητικά φράγματα δοθέντος ότι τη χρονική στιγμή μηδέν ο τυχαίος περίπατος βρισκόταν στην κατάσταση $X_0 = l$.

Πιο συγκεκριμένα

$$\Phi_l(u) = E(e^{iTu}/X_0 = l) = \sum_{k=0}^{\infty} e^{iku}\, P(T = k/X_0 = l)$$

Δηλαδή

$$\Phi_l(u) = \sum_{k=0}^{\infty} e^{iku}\, P_k^{(l)}$$

Αναλύοντας την παραπάνω σχέση κάνοντας χρήση της

$$P_k^{(l)} = pP_{k-1}^{(l+1)} + qP_{k-1}^{(l-1)}$$

$$\Phi_l(u) = \sum_{k=0}^{\infty} e^{iku}\left(pP_{k-1}^{(l+1)} + qP_{k-1}^{(l-1)}\right) =$$

$$= p\sum_{k=0}^{\infty} e^{iku}\, P_{k-1}^{(l+1)} + q\sum_{k=0}^{\infty} e^{iku}\, P_{k-1}^{(l-1)} =$$

$$= \; pe^{iu}\sum_{k=0}^{\infty} e^{iu(k-1)}\, P_{k-1}^{(l+1)} + qe^{iu}\sum_{k=0}^{\infty} e^{iu(k-1)}\, P_{k-1}^{(l-1)}$$

Τελικά προκύπτει η αναδρομική σχέση

$$\Phi_l(u) = pe^{iu}\Phi_{l+1}(u) + qe^{iu}\Phi_{l-1}(u)$$

Με οριακές συνθήκες

$$\Phi_{-\alpha}(u) = \sum_{k=0}^{\infty} e^{iku}\, P(T = k/X_0 = -\alpha) =$$

$$e^0 P(T = 0/X_0 = -\alpha) + \sum_{k=1}^{\infty} e^{iku}\, P(T = k/X_0 = -\alpha)$$

Αν λάβουμε υπόψη ότι στην περίπτωση που ο τυχαίος περίπατος ξεκινάει από την κατάσταση $X_0 = -\alpha$ ή b

Τότε

$$P_k^{(-a)} = P_k^{(b)} = \begin{cases} 1 \; \alpha\nu \; k = 0 \\ 0 \; \alpha\nu \; k \neq 0 \end{cases}$$

Τότε

$$\Phi_{-\alpha}(u) = 1$$

Διότι το άθροισμα

$$\sum_{k=1}^{\infty} e^{iku}\, P(T = k/X_0 = -\alpha) = \sum_{k=1}^{\infty} e^{iku}\, 0 = 0$$

Αντίστοιχα εργαζόμενοι

$$\Phi_b(u) = 1$$

Γενική λύση εξίσωσης διαφορών δεύτερης τάξης

Στην ουσία έχουμε να διαχειριστούμε μια ομογενή εξίσωση διαφορών δεύτερης τάξης

$$pe^{iu}\Phi_{l+1}(u) - \Phi_l(u) + qe^{iu}\Phi_{l-1}(u) = 0$$

Με πλευρικές συνθήκες

$$\Phi_{-\alpha}(s) = \Phi_b(s) = 1$$

Θέτουμε

$$\Phi_{i-1}(s) = \lambda^{i-1}, \Phi_i(s) = \lambda^i, \Phi_{i+1}(s) = \lambda^{i+1}$$

Και η διαφοροεξίσωση γράφεται

$$pe^{iu}\lambda^{i+1} - \lambda^i + qe^{iu}\lambda^{i-1} = 0$$

$$\lambda^{i-1}\left(pe^{iu}\lambda^2 - \lambda + qe^{iu}\right) = 0$$

Ισοδύναμα

$$e^{iu}\lambda^2 - \lambda + qe^{iu} = 0$$

Το παραπάνω τριώνυμο έχει διακρίνουσα ίση με

$$\Delta = 1 - 4pqe^{2iu} = 1 - e^{2iu}var(Z_i)$$

Σε περίπτωση που η διακρίνουσα είναι θετική δηλαδή

$$1 - 4pqe^{2iu} > 0 \Rightarrow 1 > 4pqe^{2iu} \Rightarrow \frac{1}{4pq} > e^{2iu} \Rightarrow \log\left(\frac{1}{4pq}\right) > 2iu$$

δηλαδή

$$\frac{-\log 4 - \log(pq)}{2i} > u$$

Υπάρχουν δύο ρίζες πραγματικές και άνισες

$$\lambda_{1,2} = \frac{1 \pm \sqrt{1 - 4pqe^{2iu}}}{2e^{2iu}p} = \frac{1 \pm \sqrt{1 - var(Z_i)e^{2iu}}}{(1 + E(Z_i))e^{2iu}}$$

Και η γενική λύση είναι

$$\Phi_l(s) = c_1\lambda_1^l + c_2\lambda_2^l$$

Ειδικότερα

$$\Phi_l(s) = c_1\left(\frac{1 + \sqrt{1 - 4pqe^{2iu}}}{2e^{2iu}p}\right)^l + c_2\left(\frac{1 - \sqrt{1 - 4pqe^{2iu}}}{2e^{2iu}p}\right)^l$$

Εναλλακτικά θα μπορούσαμε την παραπάνω γενική λύση να την εκφράσουμε ως εξής

$$\Phi_l(s) = c_1 \left(\frac{1 + \sqrt{1 - var(Z_i)e^{2iu}}}{(1 + E(Z_i))e^{2iu}} \right)^l + c_2 \left(\frac{1 - \sqrt{1 - var(Z_i)e^{2iu}}}{(1 + E(Z_i))e^{2iu}} \right)^l$$

Υπολογισμός σταθερών $\boldsymbol{c_1, c_2}$

Υπολογισμός σταθερών της παραπάνω λύσης

Θα χρησιμοποιήσουμε της οριακές ή πλευρικές συνθήκες για τον καθορισμό των σταθερών

Η πρώτη οριακή συνθήκη είναι η εξής

$$\Phi_{-\alpha}(s) = 1$$

$$\Phi_{-\alpha}(s) = c_1 \left(\frac{1 + \sqrt{1 - 4pqe^{2iu}}}{2e^{2iu}p} \right)^{-\alpha} + c_2 \left(\frac{1 - \sqrt{1 - var(Z_i)e^{2iu}}}{(1 + E(Z_i))e^{2iu}} \right)^{-\alpha} = 1$$

Η δεύτερη οριακή συνθήκη είναι η

$$\Phi_b(s) = 1$$

$$\Phi_b(s) = c_1 \left(\frac{1 + \sqrt{1 - 4pqe^{2iu}}}{2e^{2iu}p} \right)^{b} + c_2 \left(\frac{1 - \sqrt{1 - var(Z_i)e^{2iu}}}{(1 + E(Z_i))e^{2iu}} \right)^{b} = 1$$

Έτσι έχουμε να λύσουμε το σύστημα δύο εξισώσεων

της

$$\Phi_{-\alpha}(s) = c_1 \left(\frac{1 + \sqrt{1 - 4pqe^{2iu}}}{2e^{2iu}p} \right)^{-\alpha} + c_2 \left(\frac{1 - \sqrt{1 - var(Z_i)e^{2iu}}}{(1 + E(Z_i))e^{2iu}} \right)^{-\alpha} = 1$$

και της

$$\Phi_b(s) = 1$$

$$\Phi_b(s) = c_1\left(\frac{1+\sqrt{1-4pqe^{2iu}}}{2e^{2iu}p}\right)^b + c_2\left(\frac{1-\sqrt{1-var(Z_i)e^{2iu}}}{(1+E(Z_i))e^{2iu}}\right)^b = 1$$

Το παραπάνω σύστημα μπορεί να λυθεί με τη βοήθεια των οριζουσών

$$D = \begin{vmatrix} \left(\frac{1+\sqrt{1-4pqe^{2iu}}}{2e^{2iu}p}\right)^{-\alpha} & \left(\frac{1-\sqrt{1-var(Z_i)e^{2iu}}}{(1+E(Z_i))e^{2iu}}\right)^{-\alpha} \\ \left(\frac{1+\sqrt{1-4pqe^{2iu}}}{2e^{2iu}p}\right)^b & \left(\frac{1-\sqrt{1-var(Z_i)e^{2iu}}}{(1+E(Z_i))e^{2iu}}\right)^b \end{vmatrix}$$

$$D = \left(\frac{1+\sqrt{1-4pqe^{2iu}}}{2e^{2iu}p}\right)^{-\alpha}\left(\frac{1-\sqrt{1-var(Z_i)e^{2iu}}}{(1+E(Z_i))e^{2iu}}\right)^b - \left(\frac{1-\sqrt{1-var(Z_i)e^{2iu}}}{(1+E(Z_i))e^{2iu}}\right)^{-\alpha}\left(\frac{1+\sqrt{1-4pqe^{2iu}}}{2e^{2iu}p}\right)^b$$

$$D_1 = \begin{vmatrix} 1 & \left(\frac{1-\sqrt{1-var(Z_i)e^{2iu}}}{(1+E(Z_i))e^{2iu}}\right)^{-\alpha} \\ 1 & \left(\frac{1-\sqrt{1-var(Z_i)e^{2iu}}}{(1+E(Z_i))e^{2iu}}\right)^b \end{vmatrix}$$

$$D_1 = \left(\frac{1-\sqrt{1-var(Z_i)e^{2iu}}}{(1+E(Z_i))e^{2iu}}\right)^b - \left(\frac{1-\sqrt{1-var(Z_i)e^{2iu}}}{(1+E(Z_i))e^{2iu}}\right)^{-\alpha}$$

$$D_2 = \begin{vmatrix} \left(\frac{1+\sqrt{1-4pqe^{2iu}}}{2e^{2iu}p}\right)^{-\alpha} & 1 \\ \left(\frac{1+\sqrt{1-4pqe^{2iu}}}{2e^{2iu}p}\right)^b & 1 \end{vmatrix}$$

$$D_2 = \left(\frac{1+\sqrt{1-4pqe^{2iu}}}{2e^{2iu}p}\right)^{-\alpha} - \left(\frac{1+\sqrt{1-4pqe^{2iu}}}{2e^{2iu}p}\right)^{b}$$

Η λύση λοιπόν του συστήματος είναι η

$$c_1 = \frac{\left(\frac{1-\sqrt{1-var(Z_i)e^{2iu}}}{(1+E(Z_i))e^{2iu}}\right)^{b} - \left(\frac{1-\sqrt{1-var(Z_i)e^{2iu}}}{(1+E(Z_i))e^{2iu}}\right)^{-\alpha}}{\left(\frac{1+\sqrt{1-4pqe^{2iu}}}{2e^{2iu}p}\right)^{-\alpha}\left(\frac{1-\sqrt{1-var(Z_i)e^{2iu}}}{(1+E(Z_i))e^{2iu}}\right)^{b} - \left(\frac{1-\sqrt{1-var(Z_i)e^{2iu}}}{(1+E(Z_i))e^{2iu}}\right)^{-\alpha}\left(\frac{1+\sqrt{1-4pqe^{2iu}}}{2e^{2iu}p}\right)^{b}}$$

Και

$$c_2 = \frac{\left(\frac{1+\sqrt{1-4pqe^{2iu}}}{2e^{2iu}p}\right)^{-\alpha} - \left(\frac{1+\sqrt{1-4pqe^{2iu}}}{2e^{2iu}p}\right)^{b}}{\left(\frac{1+\sqrt{1-4pqe^{2iu}}}{2e^{2iu}p}\right)^{-\alpha}\left(\frac{1-\sqrt{1-var(Z_i)e^{2iu}}}{(1+E(Z_i))e^{2iu}}\right)^{b} - \left(\frac{1-\sqrt{1-var(Z_i)e^{2iu}}}{(1+E(Z_i))e^{2iu}}\right)^{-\alpha}\left(\frac{1+\sqrt{1-4pqe^{2iu}}}{2e^{2iu}p}\right)^{b}}$$

3.19 Ταυτότητα του Wald

Μέχρι στιγμής στην ανάλυση μας έχουμε υποθέσει πως τα βήματα του τυχαίου περιπάτου είναι σταθερού μήκους ίσου με τη μονάδα.

Γενικεύουμε την ανάλυση μας και ορίζουμε ως τυχαίο περίπατο ως μια στοχαστική ανέλιξη $\{X_n \ , n = 0,1,2,\dots\}$ ισόνομων και ανεξάρτητων προσαυξήσεων $Z_i, i = 0,1,2..$ με πεπερασμένη μέση τιμή $E(Z_i) = \mu \ \gamma\iota\alpha \ \ i = 1,2,3,\dots$

Ορίζουμε λοιπόν την ταυτότητα του Wald ως εξής

$$E\{g(s)^{-T}\exp{(sX_T)}\} = 1$$

Όπου

$$g(s) = E(e^{sZ})$$

Η ροπογεννήτρια του βήματος του τυχαίου περιπάτου

Αν ο τυχαίος περίπατος δεν είναι συμμετρικός και ισορροπημένος τότε το βήμα του μπορεί να οριστεί ως εξής

$$Z_i = \begin{cases} +c \; με \; πιθανότητα \; p \\ -l \; με \; πιθανότητα \; q \end{cases}$$

$$E(e^{sZ}) = pe^{sc} - qe^{-sl}$$

Η μπορούμε να το ορίσουμε και ακόμη γενικότερα χωρίς αναγκαστικά να είναι σταθερά τα βήματα της μετατόπισης αλλά να εκφράζονται με τη βοήθεια συναρτήσεων

$$Z_i = \begin{cases} g(x) \; με \; πιθανότητα \; p \\ h(x) \; με \; πιθανότητα \; q \end{cases}$$

Απόδειξη της ταυτότητας του Wald

$$E\{g(s)^{-T}\exp(sX_T)\} = E_T(E(g(s)^{-T}e^{sX_T}/T) =$$

$$= \sum_{n=0}^{\infty} E(g(s)^{-T}e^{sX_T}/T = n)P(T = n) =$$

$$\sum_{n=0}^{\infty} E(\,g(s)^{-n}e^{sX_n})\,P(T = n)$$

Όμως

$$X_n = \sum_{i=1}^{n} Z_i$$

Και

$$g_{X_n}(s) = E(e^{sX_n}) = E\left(e^{s\sum_{i=1}^{n} Z_i}\right) = g(s)^n$$

Και

$$E\{g(s)^{-T}\exp(sX_T)\} = \sum_{n=0}^{\infty} E(\,g(s)^{-n}e^{sX_n})\,P(T = n)$$

$$= \sum_{n=0}^{\infty} E(\,g(s)^{-n}e^{sX_n})\,P(T = n)$$

$$= \sum_{n=0}^{\infty} E(\,g(s)^{-n}g(s)^{n})\,P(T = n)$$

$$= \sum_{n=0}^{\infty} P(T = n) = 1$$

3.20 Ταυτότητα του Little

Μπορούμε να γράψουμε την ταυτότητα του *Wald* ως εξής

$$E\{g(s)^{-T}\exp\,(sX_T)\} = E(e^0)$$

$$E\left(e^{-T\log(g(s))+sX_T}\right) = E(e^0)$$

Από την παραπάνω σχέση συμπεραίνουμε πως

$$-T\log\big(g(s)\big) + sX_T = 0$$

Αναδιατάσσοντας τους όρους

$$T\log\big(g(s)\big) = sX_T$$

Παραγωγίζοντας ως προς s και τα δύο μέλη της παραπάνω ισότητας

$$T\frac{d}{ds}\log\big(g(s)\big) = X_T\frac{d}{dx}s$$

$$T\frac{1}{g(s)}\frac{d}{ds}(g(s)) = X_T$$

Για $s = 0$

$$T\frac{d}{ds}(g(0)) = X_T$$

Όμως

$$\frac{d}{ds}\big(g(s)\big) = \frac{d}{ds}\big(E(e^{sZ})\big) = E(Ze^{sZ})$$

Και για $s = 0$

$$\frac{d}{ds}\big(g(0)\big) = E(Z)$$

Τελικά η σχέση

$$T\frac{d}{ds}(g(0)) = X_T$$

Μπορεί να γραφεί σύμφωνα με τα παραπάνω

$$TE(Z) = X_T$$

Και τελικά

$$E(X_T) = E(T)E(Z)$$

Η παραπάνω σχέση για να έχει αξία προϋποθέτει πως ο τυχαίος περίπατος δεν είναι ισορροπημένος δηλαδή η πιθανότητα να κινηθεί προς τη μια κατεύθυνση δεν είναι η ίδια με την οποία κινείται προς την άλλη κατεύθυνση.

Από την παραπάνω σχέση λοιπόν και εφόσον $E(Z) \neq 0$ ο μέσος χρόνος απορρόφησης από κάποιο απορροφητικό φράγμα είναι

$$E(T) = \frac{E(X_T)}{E(Z)}$$

Όμως

$$E(T) = \sum_{n=0}^{\infty} n\,P(N = n) = \sum_{n=1}^{\infty} P(N \geq n) = \sum_{n=1}^{\infty} P(N > n - 1) = \sum_{n=1}^{\infty} \bar{F}(n - 1)$$

$$E(X_T) = -aP(X_T = -a) + b\big(1 - P(X_T = -a)\big) = b - (a + b)P(X_T = -a))$$

$$E(Z) = pc - ql$$

Έτσι

$$\sum_{n=1}^{\infty} P(N > n - 1) = \frac{b - (a + b)P(X_T = -a))}{pc - ql}$$

Στην περίπτωση που $E(Z) = 0$ δεν μπορούμε να χρησιμοποιήσουμε τον παραπάνω τύπο.

Σε αυτή την περίπτωση μπορούμε να παραγωγίσουμε δύο φορές την ταυτότητα του $Wald$ και να βάλουμε όπου $s = 0$

Έτσι θα προκύψει

$$E(X_T^2) = \sigma^2 E(T)$$

Καταλήξαμε λοιπόν στο συμπέρασμα πως

$$E(T) = \begin{cases} \dfrac{E(X_T^2)}{var(Z)} & \alpha\nu \;\; E(Z) = 0 \\ \dfrac{E(X_T)}{E(Z)} & \alpha\nu \; E(Z) \neq 0 \end{cases}$$

Όμως

$$E(X_T) = -aP(X_T = -a) + b\big(1 - P(X_T = -a)\big) = b - (a + b)P(X_T = -a))$$

και

$$E(X_T^2) = \alpha^2 P(X_T = -a) + b^2\big(1 - P(X_T = -a)\big) = b^2 + (\alpha^2 - b^2)P(X_T = -a)$$

Και

$$E(Z) = pc - ql$$

Επίσης

$$E(Z^2) = pc^2 + ql^2$$

έτσι

$$var(Z) = E(Z^2) - (E(Z))^2$$

$$var(Z) = pc^2 + ql^2 - (pc - ql)^2 = pq(c + l)^2$$

και η παραπάνω σχέση μπορεί πλέον να γραφτεί ως εξής

$$E(T) = \begin{cases} \dfrac{b^2 + (\alpha^2 - b^2)P(X_T = -a)}{pq(c + l)^2} & \alpha\nu \;\; E(Z) = 0 \\ \dfrac{b - (a + b)P(X_T = -a))}{pc - ql} & \alpha\nu \; E(Z) \neq 0 \end{cases}$$

3.21 Πιθανότητες απορρόφησης και ταυτότητα Wald

Όπως έχουμε ήδη αναφέρει η ταυτότητα του *Wald* περιέχει την ροπογεννήτρια συνάρτηση του βήματος του τυχαίου περιπάτου.

Η ροπογεννήτρια συνάρτηση δίνεται από τη σχέση

$$g(s) = E(e^{sZ})$$

Πρόκειται για μια κυρτή συνάρτηση αφού η δεύτερη παράγωγος είναι πάντα θετική

$$\frac{\partial^2}{\partial s^2} g(s) = E(Z^2 e^{sZ}) > 0$$

Αξίζει να σημειωθεί πως η ροπογεννήτρια είναι ίση με το μηδέν αν η τυχαία μεταβλητή που εκφράζει το βήμα του τυχαίου περιπάτου παίρνει την τιμή μηδέν με πιθανότητα τη μονάδα.

Δεσμεύοντας ως προς την θετική ή την αρνητική τιμή που μπορεί να πάρει η τυχαία μεταβλητή Z έχουμε

$$g(s) = E(e^{sZ}) =$$

$$E(E(e^{sZ}/Z) = P(Z > 0)E(e^{sZ}/Z > 0) + P(Z < 0)E(e^{sZ}/Z < 0)$$

$$\lim_{s\to+\infty} g(s) = P(Z > 0) \lim_{s\to\infty} E(e^{sZ}/Z > 0) + P(Z < 0) \lim_{s\to\infty} E(e^{sZ}/Z < 0) = \infty$$

$$\lim_{s\to-\infty} g(s) = P(Z > 0) \lim_{s\to-\infty} E(e^{sZ}/Z > 0) + P(Z < 0) \lim_{s\to-\infty} E(e^{sZ}/Z < 0) = \infty$$

Αρα υπάρχει για $E(Z) \neq 0$ μη μηδενική λύση της $E(e^{sZ}) = 1$

Έστω λοιπόν ότι $s_0 \neq 0$ τέτοιο ώστε $E(e^{s_0 Z}) = 1$

Μπορούμε λοιπόν να χρησιμοποιήσουμε την ταυτότητα του *Wald* και να βάλουμε όπου $s = s_0 \neq 0$

$$E\{g(s_0)^{-T} \exp\,(s_0 X_T)\} = 1$$

$$E(\exp(s_0 X_T)) = 1$$

Κάνοντας δέσμευση ως προς την θέση του τυχαίου περιπάτουX_T τη στιγμή της απορρόφησης $T = min\{n \in N : X_n \notin (-a, b)\}$ έχουμε

$$E(\exp(s_0 X_T)) \cong E(\exp(-s_0 \alpha))P(X_T = -a) + E(\exp(s_0 b))(1 - P(X_T = -a))$$

Δηλαδή έχουμε φτιάξει τη σχέση

$$E(\exp(-s_0 \alpha))P(X_T = -a) + E(\exp(s_0 b))(1 - P(X_T = -a)) = 1$$

$$P(X_T = -a)(E \exp(-s_0 a) - E(\exp s_0 b)) = 1 - E(\exp s_0 b)$$

Τελικά προκύπτει πως η πιθανότητα απορρόφησης στο σημείο $-a$ είναι ίση με

$$P(X_T = -a) = \frac{1 - E(\exp s_0 b)}{E \exp(-s_0 a) - E(\exp s_0 b)}$$

Και επειδή ο τυχαίος περίπατος θα απορροφηθεί σίγουρα (δηλαδή με πιθανότητα τη μονάδα) σε κάποιο από τα απορροφητικά φράγματα

$$P(X_T = -a) + P(X_T = b) = 1$$

Και

$$P(X_T = b) = 1 - \frac{1 - E(\exp s_0 b)}{E \exp(-s_0 a) - E(\exp s_0 b)}$$

Παραπάνω εργαστήκαμε με την προϋπόθεση ότι $E(Z) \neq 0$

Αν $Z = 0$ τότε $E(Z) = 0$ και $E(e^{s_0 Z}) = E(e^{s_0 0}) = 1 = E(e^{0Z}) = 1$

Δηλαδή όταν το $Z \to 0$ προκαλείται το ίδιο αποτέλεσμα με αυτό όταν το $s_0 \to 0$

Αρκεί λοιπόν να πάρουμε το όριο

$$\lim_{s_0 \to 0} P(X_T = -a) = \lim_{s_0 \to 0} \frac{1 - E(\exp s_0 b)}{E \exp(-s_0 a) - E(\exp s_0 b)} = \frac{-b}{-a-b} = \frac{b}{a+b}$$

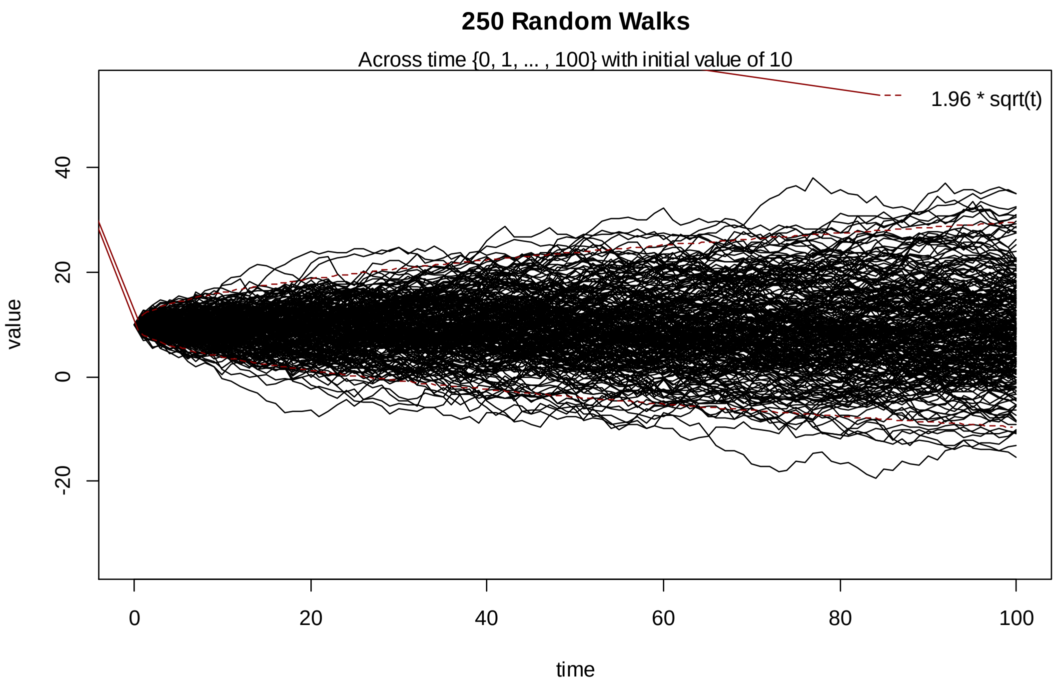
και

$$\lim_{s_0 \to 0} P(X_T = b) = \lim_{s_0 \to 0} \left(1 - \frac{1 - E(\exp s_0 b)}{E \exp(-s_0 a) - E(\exp s_0 b)}\right) = \frac{a}{a+b}$$

3.24 Γραφική απεικόνιση τυχαίου περιπάτου με τη χρήση του στατιστικού προγράμματος R

```
# Generate k random walks across time {0, 1, ... , T}
T <- 100
k <- 250
initial.value <- 10
GetRandomWalk <- function() {
  # Add a standard normal at each step
  initial.value + c(0, cumsum(rnorm(T)))
}
# Matrix of random walks
values <- replicate(k, GetRandomWalk())
# Create an empty plot
dev.new(height=8, width=12)
plot(0:T, rep(NA, T + 1), main=sprintf("%s Random Walks", k),
     xlab="time", ylab="value",
     ylim=10 + 4.5 * c(-1, 1) * sqrt(T))
mtext(sprintf("%s%s} with initial value of %s",
              "Across time {0, 1, ... , ", T, initial.value))
for (i in 1:k) {
  lines(0:T, values[ , i], lwd=0.25)
}
for (sign in c(-1, 1)) {
  curve(initial.value + sign * 1.96 * sqrt(x), from=0, to=T,
        n=2*T, col="darkred", lty=2, lwd=1.5, add=TRUE)
}
legend("topright", "1.96 * sqrt(t)",
       bty="n", lwd=1.5, lty=2, col="darkred")
savePlot("random_walks.png")
```

Κάνοντα
ς χρήση
του παραπάνω κώδικα μπορούμε να κατασκευάσουμε το γράφημα



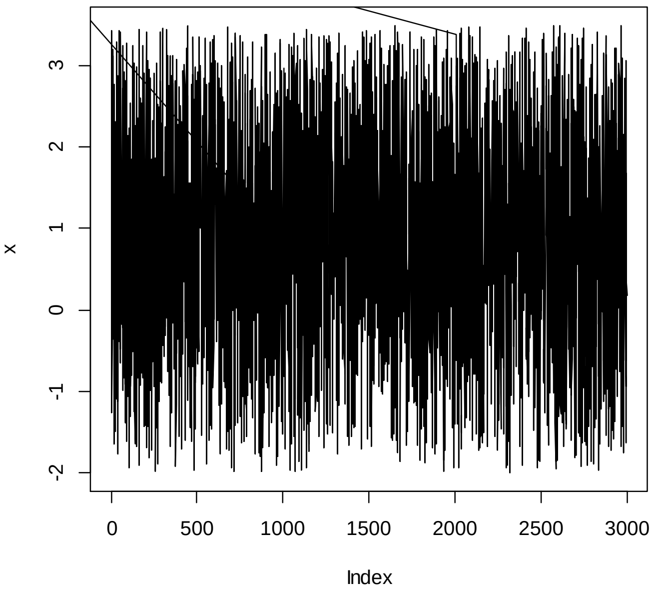
Αντίστοιχα χρησιμοποιώντας τον κώδικα

```
c <- 1
d <- 2

a <- -2
b <- 3.5

ll <- pnorm(a, c, d)
ul <- pnorm(b, c, d)

x <- qnorm( runif(3000, ll, ul), c, d )
hist(x)
range(x)
mean(x)
sd(x)
plot(x, type='l')
```



Κεφάλαιο 4
Θεωρία χρεοκοπίας

4.1 Εισαγωγή

Η θεωρία χρεοκοπίας στο συλλογικό πρότυπο της θεωρίας κινδύνου είναι ένας κλάδος της αναλογιστικής επιστήμης που μελετά την φερεγγυότητα του ασφαλιστή βασιζόμενη σε μαθηματικά μοντέλα του πλεονάσματος.

Η θεωρητική θεμελίωση της θεωρίας χρεοκοπίας αλλά και του συλλογικού προτύπου γενικότερα που είναι γνωστή στη διεθνή βιβλιογραφία ως classical compound Poisson risk model εισήχθη το 1903 από τον Σουηδό αναλογιστή Fillip Lundberg.Το κλασικό μοντέλο αργότερα επεκτάθηκε για να χαλαρώσει υποθέσεις σχετικά με τον ενδιάμεσο χρόνο έλευσης των απαιτήσεων, την κατανομή του ύψους της απαίτησης κτλ.

Βασικός στόχος του κλασικού μοντέλου αλλά και των μετέπειτα επεκτάσεων του είναι ο υπολογισμός της πιθανότητας χρεοκοπίας.

Η μελέτη των συνολικών αποζημιώσεων ενός ασφαλιστή έχει σημαντικό ρόλο στη θεωρία της χρεοκοπίας . Εδώ οι αποζημιώσεις αυτές δεν εξετάζονται σε σταθερό χρόνο, αλλά δυναμικά καθώς εξελίσσονται με την πάροδο του χρόνου και με τη βοήθεια του συλλογικού προτύπου μακράς χρονικής περιόδου. Στο πρότυπο αυτό, κεντρικό πρόβλημα είναι το πρόβλημα της χρεοκοπίας, στο οποίο εξετάζεται η μεταβολή του πλεονάσματος που προκύπτει από τα έσοδα (ασφάλιστρα) μείον τα έξοδα (αποζημιώσεις) για έναν ασφαλιστή, τόσο σε διακριτό όσο και σε συνεχή χρόνο. Βασικά εργαλεία για τη θεωρία της χρεοκοπίας αποτελούν οι σύνθετες κατανομές πιθανότητας (ιδιαίτερα η σύνθετη γεωμετρική και η σύνθετη Poisson) και η θεωρία στοχαστικών ανελίξεων

4.2 Οι στοχαστικές διαδικασίες στη θεωρία χρεοκοπίας

Στη θεωρία χρεοκοπίας εξετάζουμε τις συνολικές αποζημιώσεις από ένα χαρτοφυλάκιο ασφαλισμένων μιας ασφαλιστικής εταιρείας όπως αυτές εξελίσσονται στο χρόνο. Γι αυτό το λόγο χρησιμοποιούμε τη στοχαστική διαδικασία (στοχαστική ανέλιξη) $\{S(t) : t \geq 0\}$ τέτοια ώστε η τυχαία μεταβλητή $S(t)$ ορίζεται ως εξής :

$$S(t) = \begin{cases} \displaystyle\sum_{i=1}^{N(t)} X_i \;\; για \; N(t) \geq 1 \\ 0 \quad για \;\; N(t) = 0 \end{cases}$$

Όπως είναι απολύτως λογικό το ύψος των συνολικών αποζημιώσεων θα εξαρτάται από τον αριθμό των απαιτήσεων προς την εταιρεία. Ο αριθμός των απαιτήσεων δεν είναι και δεν θα μπορούσε να είναι εκ των προτέρων γνωστός γι αυτό και θεωρείται τυχαία μεταβλητή.

Στη θεωρία χρεοκοπίας δεν μελετάμε μεγέθη θεωρώντας το χρόνο σταθερό ,αλλά μας ενδιαφέρει ο τρόπος που τα μεγέθη διαμορφώνονται στο χρόνο. Έτσι για τον αριθμό των απαιτήσεων προς την εταιρεία θα χρησιμοποιήσουμε μια άλλη στοχαστική διαδικασία την $\{N(t) : t \geq 0\}$ η οποία είναι μια στοχαστική ανέλιξη τέτοια ώστε η τυχαία μεταβλητή $N(t)$ παριστάνει τον αριθμό των ενδεχομένων τα οποία συμβαίνουν το χρονικό διάστημα (0,t).

Η $N(t)$ είναι μια απαριθμήτρια στοχαστική ανέλιξη, δηλαδή ανέλιξη σταθερή κατά διαστήματα με άλματα ύψους ένα σε ορισμένα σημεία που είναι οι τυχαίες στιγμές επέλευσης ενός ζημιογόνου γεγονότος (ζημιά).

4.3 Πλεόνασμα (surplus)

Η ασφαλιστική εταιρεία είναι μια επιχείρηση η οποία έχει σκοπό και το κέρδος .Μας ενδιαφέρουν δηλαδή ως μέγεθος που μεταβάλλεται στο χρόνο τα έσοδα της εταιρείας .Έσοδα τα οποία προέρχονται από την είσπραξη των συνολικών ασφαλίστρων που εισπράττει η εταιρεία.

Η ασφαλιστική εταιρεία θα πρέπει να έχει κεφάλαια επαρκή για την αποτελεσματική άσκηση των δραστηριοτήτων της αλλά να έχει και πρόσθετα κεφάλαια (αποθεματικά) για την αντιμετώπιση απρόοπτων εξελίξεων που θα μπορούσαν να θέσουν σε κίνδυνο την ομαλή λειτουργία της επιχείρησης.

Στην ασφαλιστική εταιρεία τα αποθεματικά αυτά καλούνται γενικά πλεόνασμα (surplus)[8]

Ποιό συγκεκριμένα, πλεόνασμα είναι η θετική διαφορά ανάμεσα στο ενεργητικό της εταιρείας και στην καλύτερη δυνατή αναλογιστική εκτίμηση των συνολικών υποχρεώσεων της.

Η τιμή του πλεονάσματος τη χρονική στιγμή t $(με\ t \geq 0)$ γράφεται $U(t)$. Η $U(0)$ είναι η τιμή του πλεονάσματος τη χρονική στιγμή t=0 (στιγμή έναρξης των εργασιών) και καλείται αρχικό αποθεματικό το οποίο συμβολίζεται με u και αποτελεί βασική παράμετρο της θεωρίας χρεοκοπίας.

Η στοχαστική διαδικασία του πλεονάσματος $\{U(t) : t \ \geq 0\}$ ορίζεται για κάθε $t > 0$

$$U(t) = u + P(t) - S(t)$$

όπου:

$S(t)$: Η στοχαστική ανέλιξη των συνολικών αποζημιώσεων στο χρονικό διάστημα (0,t]

u: αρχικό αποθεματικό

$P(t)$: σύνολο ασφαλίστρων που εισπράττονται το χρονικό διάστημα (0,t]

4.4 Η κλασική θεωρία του κινδύνου

Η θεωρία αυτή θεωρεί τη συνάρτηση $P(t)$ αιτιοκρατική (deterministic) .Υποθέτει ακόμη ειδικότερα πως :

$$P(t) = ct\ για\ t \geq 0$$

λύνοντας ως προς c παίρνουμε c = $\frac{P(t)}{t}$ όπου c μια θετική σταθερά.

Δηλαδή ο ρυθμός είσπραξης ασφαλίστρων παραμένει σταθερός διαχρονικά, με άλλα λόγια το ασφάλιστρο που πληρώνεται από τον ασφαλισμένο στη μονάδα του χρόνου (ή το έσοδο που εισπράττει η ασφαλιστική εταιρεία στη μονάδα του χρόνου) δεν μεταβάλλεται. Το c καλείται στη διεθνή βιβλιογραφία premium rate (ένταση ασφαλίστρου). Σύμφωνα λοιπόν με τα παραπάνω η σχέση (1) μπορεί να γραφεί ως εξής :

$$U(t) = u + ct - \sum_{i=1}^{N(t)} X_i \ , t \geq 0$$

όπου $\{N(t) : t \geq 0\}$ στοχαστική διαδικασία τέτοια ώστε η τυχαία μεταβλητή $N(t)$ παριστάνει τον αριθμό των ενδεχομένων τα οποία συμβαίνουν το χρονικό διάστημα (0,t). Η στοχαστική διαδικασία $\{N(t) : t \geq 0\}$ καλείται ομογενής διαδικασία Poisson εάν:

Ορισμός 1ος

- $N(0) = 0$ δηλαδή τη στιγμή έναρξης των εργασιών δεν έχει συμβεί καμία ζημιά.
- Η $\{N(t): t \geq 0\}$ έχει ανεξάρτητες προσαυξήσεις .
- Ο αριθμός των ενδεχομένων σε κάθε διάστημα μήκους t ακολουθεί την κατανομή Poisson με μέση τιμή λt[9].

Εναλλακτικά

Ορισμός 2ος :

- $N(0) = 0$ δηλαδή τη στιγμή έναρξης των εργασιών δεν έχει συμβεί καμία ζημιά.
- Η $\{N(t): t \geq 0\}$ έχει ομογενείς και ανεξάρτητες προσαυξήσεις
- Η πιθανότητα δηλαδή να έχουμε παραπάνω από ένα ζημιογόνο συμβάν είναι πολύ μικρή και τείνει στο μηδέν[10]

$$P(N(t) \geq 2) = o(t) = P(N(t) > 1)$$

Σημειώνουμε πως η συνάρτηση $o(t)$ έχει την ιδιότητα

$$\lim_{t \to 0} \frac{o(t)}{t} = 0$$

δηλαδή ο αριθμητής φτάνει γρήγορα στο μηδέν σε σχέση με τον παρονομαστή.

Η πιθανότητα ατυχήματος σε ένα μικρό χρονικό διάστημα δεν εξαρτάται από την αρχή του χρονικού διαστήματος αλλά από το μήκος λ του διαστήματος.

$$P(N(t) = 1) = \lambda t + o(t)$$

4.5 Χρόνοι άφιξης

Μια ανέλιξη Poisson, όπως και κάθε απαριθμήτρια ανέλιξη, ορίζει μια ακολουθία από τυχαίες μεταβλητές $Y_1, Y_2 \ldots\ldots, Y_n$, που λέγεται ακολουθία χρόνων άφιξης κάποιου γεγονότος με τον ακόλουθο τρόπο :

$$Y_i = min\{t: N(t) = i\}$$

Όπως φαίνεται από τη σχέση (2) η τιμή $U(t)$ του πλεονάσματος υφίσταται πτώσεις μεγέθους X_i κατά τις χρονικές στιγμές Y_i

όπου

$$Y_i = min\{t: N(t) = i\}$$

Με λόγια η παραπάνω σχέση μας λέει πως Y_i είναι ο ελάχιστος χρόνος που απαιτείται για την επέλευση του i ζημιογόνου γεγονότος και καλείται χρόνος άφιξης της i ζημιάς .

4.6 Ενδιάμεσοι χρόνοι ή χρόνοι αναμονής

Με βάση την ακολουθία των χρόνων άφιξης μπορούμε να ορίσουμε την ακολουθία $\{T_\kappa: k = 1,2, \dots\}$ ως εξής :

$$T_1 = Y_1$$

$$T_2 = Y_2 - Y_1$$

$$\dots\dots\dots\dots$$

$$T_k = Y_k - Y_{k-1}$$

Οι μεταβλητές T_k ονομάζονται ενδιάμεσοι χρόνοι ή χρόνοι αναμονής.Αξίζει να σημειωθεί πως ενώ η $N(t)$ είναι μια διακριτή τυχαία μεταβλητή, οι Y_i και T_i είναι συνεχείς τυχαίες μεταβλητές.

4.7 Ορισμός χρεοκοπίας

Ένα μεγάλο και σημαντικό τμήμα της έρευνας στο χώρο της θεωρίας των κινδύνων είναι αφιερωμένο στη μελέτη της χρεοκοπίας. Εξίσου αξιοσημείωτη με την έκταση της σχετικής έρευνας είναι η ποικιλία των μεθόδων που έχουν χρησιμοποιηθεί (μεταξύ των οποίων ροπογεννήτριες, μετασχηματισμοί Laplace, ανανεωτικές και άλλες ολοκληροδιαφορικές εξισώσεις.)

Ιδιαίτερη σημασία έχει το ενδεχόμενο κάποια στιγμή T η αποζημίωση X_t που συνδέεται με τη ζημιά που συμβαίνει τη χρονική στιγμή t=T σε συνδυασμό με την τιμή $u(T-)$ πριν από το γεγονός να οδηγήσει σε τιμή του πλεονάσματος αρνητική (τότε λέμε ότι έχει επέλθει χρεοκοπία).Με άλλα λόγια χρεοκοπία σημαίνει πως κάποια μελλοντική στιγμή T έχουμε για πρώτη φορά $U(T) < 0$.

Ο όρος χρεοκοπία δεν χρησιμοποιείται κυριολεκτικά αλλά είναι περισσότερο ένας τεχνικός όρος, που χρησιμοποιείται σαν μέτρο της φερεγγυότητας του χαρτοφυλακίου.

Στην πράξη, η διαδικασία πλεονάσματος δεν είναι ο μοναδικός πόρος μιας ασφαλιστικής εταιρίας, η καταβολή μιας αποζημίωσης δεν είναι στιγμιαίο γεγονός (απαιτεί κάποιο χρονικό διάστημα, που συνεπάγεται εισροή ασφαλίστρου πέρα από το εισπραγμένο μέχρι τη στιγμή Τ), η εταιρία μπορεί να αυξήσει το μετοχικό της κεφάλαιο ή να συνάψει δάνειο.

Ανάμεσα στα σχετικά αντικείμενα μελέτης της θεωρίας κινδύνου (και της θεωρίας χρεοκοπίας) κεντρικό ρόλο παίζει η πιθανότητα χρεοκοπίας ως συνάρτηση του αρχικού αποθεματικού.

4.8 Χρεοκοπία από ένα μοναδικό γεγονός

Ένα πρόβλημα που μπορεί να δώσει μια πρώτη αίσθηση πως λειτουργεί το ενδεχόμενο χρεοκοπίας είναι ο υπολογισμός της πιθανότητας χρεοκοπίας στα πλαίσια μιας διαδικασίας πλεονάσματος για την κάλυψη ενός μοναδικού ζημιογόνου γεγονότος.

Για τον υπολογισμό της πιθανότητας χρεοκοπίας εξαιτίας ενός μοναδικού γεγονότος, απαιτείται η κατανομή του χρόνου επέλευσης του γεγονότος και η κατανομή του ύψους της αποζημίωσης. Η διαδικασία πλεονάσματος είναι

$$U(t) = u + P(t) - S(t).$$

Όπου

$$S(t) = 0 \; για \; t < T \; και \; S(t) = X \; για \; t \geq T.$$

Χρεοκοπία θα έχουμε εάν κατά τη χρονική στιγμή, έστω t, επέλευσης του γεγονότος, το υπάρχον αποθεματικό, $u + ct$ είναι μικρότερο από το ύψος της αποζημίωσης X.

4.9 Χρόνος χρεοκοπίας

Για την πιθανότητα χρεοκοπίας με άπειρο χρονικό ορίζοντα έχουμε δώσει τη σχέση

$$\psi(u) = P\,(T < \infty \mid u(0) = u)$$

Με αντίστοιχο ορισμό της πιθανότητας μη χρεοκοπίας :

$$\delta(u) = 1 - \psi(u) = \; P\,(T = \infty | u(0) = u)$$

$με \; T := \; inf\,\{t : t > 0 \; και \; U\,(t) < 0\}.$

Η τυχαία μεταβλητή T καλείται χρόνος χρεοκοπίας . Αναλυτικότερα ο χρόνος χρεοκοπίας μπορεί να οριστεί ως εξής :

$$T = \begin{cases} \inf\{t : U(t) < 0\} \\ \propto \; αν \; U(t) > 0 \; \forall t \end{cases}$$

Παρατηρούμε πως το ενδεχόμενο το πλεόνασμα μας να είναι θετικό για κάθε χρονική στιγμή είναι ισοπίθανο με το ενδεχόμενο ο χρόνος χρεοκοπίας να είναι άπειρος. Η πιθανότητα λοιπόν ο χρόνος χρεοκοπίας να είναι άπειρος υπάρχει και παίρνει θετική τιμή.

Αξίζει να σημειωθεί δε πως :

$$P(T = \infty) = P(U(t) > 0 \; \forall \; t) = \delta(u) = 1 - \psi(u)$$

Από πρακτική σκοπιά, πιο ρεαλιστική είναι η μελέτη του ενδεχομένου χρεοκοπίας σε πεπερασμένο χρονικό διάστημα $(0, t)$.

4.10 Πιθανότητα χρεοκοπίας και τυχαίος περίπατος

Σε αυτή την ενότητα θα χρησιμοποιήσουμε την έννοια του τυχαίου περιπάτου για να υπολογίσουμε την πιθανότητα της χρεοκοπίας μιας ασφαλιστικής εταιρίας. Με τη λέξη χρεοκοπία δεν εννοούμε την χρεοκοπία της επιχείρησης αλλά το ενδεχόμενο το αποθεματικό της ,το πλεόνασμα της να γίνει αρνητικό κάποια χρονική στιγμή.

Μια ασφαλιστική εταιρεία έχει έσοδα τα οποία προσδιορίζονται από την τιμολογιακή πολιτική της εταιρείας (αυτό το κομμάτι ανήκει στους αναλογιστές της εταιρείας) Στην ανάλυση μας θα θεωρήσουμε πως τα ασφάλιστρα καθορίζονται με ακρίβεια από τον ασφαλιστή και έτσι δεν υπάρχει αβεβαιότητα ως προς την εξέλιξη τους στο χρόνο.

Από την άλλη η ασφαλιστική εταιρεία καταβάλει αποζημιώσεις. Σκοπός της ασφαλιστικής εταιρείας είναι να αναλαμβάνει μέρος ή και το σύνολο του κινδύνου του ασφαλισμένου και στην περίπτωση ζημιογόνου γεγονότος να καταβάλει αποζημίωση. Σε αντίθεση με τα έσοδα που θεωρούνται σταθερά στη μονάδα του χρόνου η ασφαλιστική εταιρεία δεν είναι σε θέση να γνωρίζει το ύψος των αποζημιώσεων που θα πρέπει να καταβάλει σε ένα χρονικό διάστημα μήκους t.

Σε αυτό το σημείο πρέπει να αναφέρουμε πως το συνολικό ποσό αποζημιώσεων που θα καταβάλει η ασφαλιστική εταιρεία εξαρτάται από δύο διαφορετικά στοιχεία. Εξαρτάται από τον αριθμό των ζημιογόνων γεγονότων αλλά και από το ύψος της ζημίας κάθε ζημιογόνου γεγονότος.

Όπως έχουμε ορίσει παραπάνω T_i είναι οι ενδιάμεσοι χρόνοι ή χρόνοι αναμονής και Y_i οι χρόνοι άφιξης των ζημιογόνων γεγονότων.

Αν και ο χρόνος είναι συνεχής εμάς μας ενδιαφέρουν συγκεκριμένες χρονικές στιγμές και μάλιστα οι χρονικές στιγμές άφιξης μιας απαίτησης για αποζημίωση.

Θεωρούμε τη χρονική στιγμή $T = n$ πως το πλεόνασμα της εταιρείας δίνεται από τη σχέση

$$U_n = u + c\sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} X_i = u - \sum_{i=1}^{n}(X_i - cY_i)$$

Η πιθανότητα χρεοκοπίας όπως την έχουμε ορίσει μέχρι στιγμής είναι η πιθανότητα το πλεόνασμα της επιχείρησης να γίνει κάποια στιγμή στο μέλλον αρνητικό δηλαδή

$$P(U_n < 0, \text{για κάποιο } n = 1,2, \dots)$$

Η παραπάνω σχέση θα μπορούσε να γραφτεί ως εξής

$$P(u - \sum_{i=1}^{n}(X_i - cY_i) < 0, \text{για κάποιο } n = 1,2, \ldots)$$

$$= P(\sum_{i=1}^{n}(X_i - cY_i) > u\,, \text{για κάποιο } n = 1,2, \ldots)$$

Μπορούμε λοιπόν να θεωρήσουμε την ποσότητα

$$W_n = \sum_{i=1}^{n}(X_i - cY_i) = \sum_{i=1}^{n} W_i$$

Ως ένα τυχαίο περίπατο με βήμα

$$W_i = X_i - cY_i$$

Το βήμα του τυχαίου περιπάτου μας εκφράζει τα καθαρά έσοδα της εταιρείας μετά την επέλευση της i αποζημίωσης.

Στην ουσία πρόκειται για έναν τυχαίο περίπατο με απορροφητικό φράγμα το u_0 προς τα πάνω και κανένα όριο προς τα κάτω (μπορεί να θεωρηθεί ως κάτω φράγμα το $-\infty$)

Αν η μέση τιμή του βήματος του τυχαίου περιπάτου είναι θετική δηλαδή αν

$$E(W_i) = E(X_i - cY_i) = E(X_i) - cE(Y_i) > 0$$

Δηλαδή αν το ύψος των αποζημιώσεων είναι μεγαλύτερο από το ύψος των ασφαλίστρων που εισπράττει η ασφαλιστική

$$E(X_i) > cE(Y_i)$$

Τότε η επιχείρηση θα αντιμετωπίσει σίγουρα πρόβλημα ρευστότητας δηλαδή με πιθανότητα τη μονάδα σε κάποια μελλοντική χρονική στιγμή το πλεόνασμα θα γίνει αρνητικό.

Αν η μέση τιμή του βήματος του τυχαίου περιπάτου είναι αρνητική δηλαδή αν

$$E(W_i) = E(X_i - cY_i) = E(X_i) - cE(Y_i) < 0$$

Δηλαδή αν το ύψος των αποζημιώσεων είναι μικρότερο από το ύψος των ασφαλίστρων που εισπράττει η ασφαλιστική

$$E(X_i) < cE(Y_i)$$

Τότε η επιχείρηση θα αντιμετωπίσει πρόβλημα ρευστότητας με πιθανότητα p δηλαδή με πιθανότητα p σε κάποια μελλοντική χρονική στιγμή το πλεόνασμα θα γίνει αρνητικό.

Η πιθανότητα p αντιστοιχεί στην πιθανότητα απορρόφησης στην κατάσταση u_0 και ταυτίζεται με την πιθανότητα χρεοκοπίας.

Μπορούμε λοιπόν να χρησιμοποιήσουμε την ταυτότητα του *Wald* και να βάλουμε όπου $s = s_0 \neq 0$

$$E\{g(s_0)^{-T} \exp\,(s_0 X_T)\} = 1$$

$$E(\exp(s_0 X_T)) = 1$$

Κάνοντας δέσμευση ως προς την θέση του τυχαίου περιπάτουX_T τη στιγμή της απορρόφησης $T = min\{n \in N : X_n \notin (-a, u_0\}$ έχουμε

$$E(\exp(s_0 X_T)) \cong E(\exp(-s_0 \infty))P(X_T = -\infty) + E(\exp(s_0 u))(1 - P(X_T = -\infty))$$

Δηλαδή έχουμε φτιάξει τη σχέση

$$E(\exp(-s_0 \infty))P(X_T = -\infty) + E(\exp(s_0 u))(1 - P(X_T = -\infty)) = 1$$

$$P(X_T = -\infty)(E\exp(-s_0 \infty) - E(\exp s_0 u)) = 1 - E(\exp s_0\, u)$$

Τελικά προκύπτει πως η πιθανότητα απορρόφησης στο σημείο $-a$ είναι ίση με

$$P(X_T = -\infty) = \frac{E(\exp s_0 u) - 1}{E(\exp s_0 u)} = \frac{e^{s_0 u} - 1}{e^{s_0 u}}$$

Και επειδή ο τυχαίος περίπατος θα απορροφηθεί σίγουρα (δηλαδή με πιθανότητα τη μονάδα) σε κάποιο από τα απορροφητικά φράγματα

$$P(X_T = -\infty) + P(X_T = u_0) = 1$$

Και

$$P(X_T = u\,) = 1 + \frac{1 - E(\exp s_0\, u)}{E(\exp s_0 u)} = \frac{1}{E(\exp s_0 u)} = e^{-s_0 u}$$

Με $s_0 \neq 0$ την μη αρνητική ρίζα της εξίσωσης

$$g(s_0) = 1$$

$$E(e^{s_0 W}) = 1$$

Όμως

$$W_i = X_i - cY_i$$

Και

$$E(e^{s_0 W}) = E\left(e^{s_0 (X_i - cY_i)}\right) = 1$$

Οι τυχαίες μεταβλητές X_i και Y_i είναι ανεξάρτητες τυχαίες μεταβλητές δηλαδή

$$cov(X_i\,, Y_i) = 0$$

Και έτσι η σχέση

$$E\left(e^{s_0(X_i-cY_i)}\right)=E(e^{s_0X_i})E(e^{-s_0cY_i})=g_{X_i}(s_0)g_{Y_i}(-s_0c)=1$$

4.11 Εφαρμογή

Δείξαμε παραπάνω πως το μπορούμε να εισάγουμε στην θεωρία χρεοκοπίας τον τυχαίο περίπατο και να υπολογίσουμε την πιθανότητα χρεοκοπίας.

Θεωρήσαμε σαν βήμα του τυχαίου περιπάτου την ποσότητα

$$W_i=X_i-cY_i$$

Θα υποθέσουμε πως η τυχαία μεταβλητή X ακολουθεί μείξη εκθετικών κατανομών και έχει συνάρτηση πυκνότητας πιθανότητας

$$f_x(x)=\sum_{i=1}^{k}\beta_i\,d_i e^{-d_ix}$$

Όπου

$$\sum_{i=1}^{k}\beta_i=1\;\;d_i>0, i=1,2,\dots k\;και\;x\geq 0$$

Θα υποθέσουμε επίσης πως η τυχαία μεταβλητή Y ακολουθεί εκθετική κατανομή με παράμετρο α. Έχει δηλαδή συνάρτηση πυκνότητας πιθανότητας

$$f_Y(y)=ae^{-ay}\quad a>0\;y\geq 0$$

Η συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής W δίνεται από την σχέση

$$f_W(w)=\int_{-\infty}^{+\infty}f_x(x)\,f_Y\left(\frac{x-z}{c}\right)dx=\int_{z}^{+\infty}f_x(x)\,f_Y\left(\frac{x-z}{c}\right)dx$$

Με αντικατάσταση η παραπάνω σχέση μας δίνει

$$f_W(w)=\int_{z}^{+\infty}ae^{-a\left(\frac{x-z}{c}\right)}\left(\sum_{i=1}^{k}\beta_i\,d_ie^{-d_ix}\right)dx=$$

$$=ae^{\frac{az}{c}}\sum_{i=1}^{k}\int_{z}^{+\infty}\beta_i\,d_ie^{-\left(\frac{a}{c}+d_i\right)x}dx=ae^{\frac{az}{c}}\sum_{i=1}^{k}\beta_i\,d_i\int_{z}^{+\infty}e^{-\left(\frac{a}{c}+d_i\right)x}dx$$

Το ολοκλήρωμα της παραπάνω σχέσης λύνεται ως εξής

$$\int_{z}^{+\infty} e^{-\left(\frac{a}{c}+d_i\right)x}dx = \left[-\frac{e^{-\left(\frac{a}{c}+d_i\right)x}}{\frac{a}{c}+d_i}\right]_z^{+\infty} = \frac{e^{-d_i z}e^{\frac{-az}{c}}}{\frac{a}{c}+d_i}$$

Το παραπάνω ολοκλήρωμα θα μπορούσε να λυθεί και με τη βοήθεια του Mathematica ως εξής

In[14]:= Integrate[E^(-(a/c+d) x),{x,z,\[Infinity]},Assumptions->{Re[a/c+d]>0}]

Out[14]= (c E^(-(a/c+d) z))/(a+c d)

Η συνάρτηση πυκνότητας πιθανότητας της τυχαίας μεταβλητής w μπορεί να γραφτεί

$$f_W(w) = ae^{\frac{az}{c}}\sum_{i=1}^{k}\beta_i\, d_i \frac{e^{-d_i z}e^{\frac{-az}{c}}}{\frac{a}{c}+d_i} = \alpha\sum_{i=1}^{k}\beta_i\, d_i \frac{e^{-d_i z}}{\frac{a}{c}+d_i} = \alpha\sum_{i=1}^{k}\frac{\beta_i d_i}{\frac{a}{c}+d_i}e^{-d_i z} \qquad z \geq 0$$

Παρατήρηση

Στην ιδιαίτερη περίπτωση όπου

$$d_1 = d_2 = \cdots = d_k = d$$

Η συνάρτηση πυκνότητας πιθανότητας της w γράφεται

$$f_W(w) = \frac{\alpha d}{\frac{a}{c}+d}e^{-dz}\sum_{i=1}^{k}\beta_i = \frac{\alpha d}{\frac{a}{c}+d}e^{-dz} = \frac{ac}{a+cd}de^{-dz}$$

Παρατηρούμε δηλαδή πως

$$\frac{a+cd}{ac}f_W(w) = \ de^{-dz}$$

Δηλαδή με λόγια η συνάρτηση πυκνότητας πιθανότητας $f_W(w)$ πολλαπλασιασμένη με μία θετική ποσότητα ταυτίζεται με τη συνάρτηση πυκνότητας πιθανότητας της εκθετικής κατανομής με παράμετρο d.

Αν η διαφορά

$$W_i = X_i - cY_i$$

Είναι αρνητική τότε

$$f_W(w) = \int_{-\infty}^{+\infty} f_x(z+cy)\, f_Y(y)dy =$$

$$= \int_{-z/c}^{+\infty} ae^{-ay}\left(\sum_{i=1}^{k}\beta_i\, d_i e^{-d_i(z+cy)}\right)dy = a\sum_{i=1}^{k}\beta_i d_i\, e^{-d_i z}\int_{-z/c}^{+\infty} e^{-(a+d_i c)y}dy$$

Το ολοκλήρωμα

$$\int_{-z/c}^{+\infty} e^{-(a+d_i c)y} dy = \left[-\frac{e^{-(a+d_i c)y}}{a+d_i c}\right]_{-z/c}^{+\infty} = \frac{e^{zd_i} e^{\frac{az}{c}}}{a+d_i c}$$

Τελικά

$$f_W(w) = a\sum_{i=1}^{k} \beta_i d_i\, e^{-d_i z} \frac{e^{zd_i} e^{\frac{az}{c}}}{a+d_i c} = \alpha e^{\frac{az}{c}} \sum_{i=1}^{k} \frac{\beta_i d_i}{a+d_i c} \qquad z<0$$

Στην περίπτωση όπου

$$d_1 = d_2 = \cdots = d_k = d$$

$$f_W(w) = \frac{d}{\alpha+dc}\alpha e^{\frac{az}{c}}$$

Συνοψίζοντας

$$f_W(w) = \begin{cases} \alpha \displaystyle\sum_{i=1}^{k} \frac{\beta_i d_i}{\frac{a}{c}+d_i} e^{-d_i z} & z \geq 0 \\ \alpha e^{\frac{az}{c}} \displaystyle\sum_{i=1}^{k} \frac{\beta_i d_i}{a+d_i c} & z<0 \end{cases}$$

Και στην ειδική περίπτωση όπου

$$d_1 = d_2 = \cdots = d_k = d$$

$$f_W(w) = \begin{cases} \dfrac{ac}{a+cd} d e^{-dz} & z \geq 0 \\ \dfrac{d}{\alpha+dc} \alpha e^{\frac{az}{c}} & z<0 \end{cases}$$

4.12 Ροπογεννήτρια της τυχαίας μεταβλητής W

Σε αυτή την ενότητα θα υπολογίσουμε την ροπογεννήτρια του βήματος του τυχαίου περιπάτου ως εξής

$$M_w(t) = E\left(e^{t(X_i - cY_i)}\right) = E(e^{tX_i})E(e^{-tcY_i}) = M_{X_i}(t)M_{Y_i}(-tc)$$

Η παραπάνω σχέση ισχύει λόγο ανεξαρτησίας των τυχαίων μεταβλητών $X \; και \; Y$

Θα πρέπει λοιπόν να υπολογίσουμε

$$M_X(t) = \int_0^{+\infty} e^{tx} f_x(x)dx =$$

$$= \int_0^{+\infty} e^{tx} \left(\sum_{i=1}^{k} \beta_i \, d_i e^{-d_i x} \right) dx = \sum_{i=1}^{k} \beta_i d_i \int_0^{+\infty} e^{-(d_i - t)x} \, dx$$

Και ύστερα από στοιχειώδεις πράξεις

$$M_X(t) = \sum_{i=1}^{k} \frac{\beta_i d_i}{d_i - t}$$

Θα πρέπει επίσης να υπολογίσουμε την $M_{Y_i}(-tc)$

$$M_{Y_i}(-tc) = \int_0^{+\infty} e^{-tcy} f_Y(y)dy = \int_0^{+\infty} e^{-tcy} a e^{-ay} \, dy =$$

$$= a \int_0^{+\infty} e^{-(tc+a)y} \, dy = \frac{a}{tc + a}$$

Έτσι καταλήξαμε στο συμπέρασμα πως

$$M_w(t) = \left(\sum_{i=1}^{k} \frac{\beta_i d_i}{d_i - t} \right) \frac{a}{tc + a}$$

Από την παραπάνω σχέση μπορούμε να υπολογίσουμε την κ-οστή ροπή περί την αρχή ως εξής

$$\frac{dM_w(0)^k}{dt^k} = E(W^k)$$

Σύμφωνα λοιπόν με την παραπάνω σχέση

$$\frac{d}{dt}\left(\left(\sum_{i=1}^{k}\frac{\beta_i d_i}{d_i - t}\right)\frac{a}{tc + a}\right) =$$

$$= \frac{d}{dt}\left(\sum_{i=1}^{k}\frac{\beta_i d_i}{d_i - t}\right)\frac{a}{tc + a} + \left(\sum_{i=1}^{k}\frac{\beta_i d_i}{d_i - t}\right)\frac{d}{dt}\left(\frac{a}{tc + a}\right) =$$

$$= \sum_{i=1}^{k}\frac{\beta_i d_i}{(d_i - t)^2}\frac{a}{tc + a} - \frac{ac}{(tc + a)^2}\left(\sum_{i=1}^{k}\frac{\beta_i d_i}{d_i - t}\right)$$

Η μέση τιμή προκύπτει αν βάλουμε στην παραπάνω σχέση όπου $t = 0$

Έτσι

$$E(W) = \sum_{i=1}^{k}\frac{\beta_i}{d_i} - \frac{c}{a}\sum_{i=1}^{k}\beta_i$$

Όμως

$$\sum_{i=1}^{k}\beta_i = 1$$

Και

$$E(W) = \sum_{i=1}^{k}\frac{\beta_i}{d_i} - \frac{c}{a}$$

Στην περίπτωση που $d_1 = 3, \beta_1 = 0{,}5 = \beta_2, d_2 = 2, a = 4, c = 5$ και έχουμε μείξη δύο εκθετικών τότε η ροπογεννήτρια δίνεται από τη σχέση

$$M_w(t) = \left(\frac{1{,}5}{3 - t} + \frac{1}{2 - t}\right)\frac{4}{5t + 4}$$

Με τη βοήθεια του Mathematica η λύση της εξίσωσης

$$M_w(t) = 1$$

Είναι ίση με

```
In[6]:= Solve[ropogennitria[t_]==1,t]
```

Out[6]= {{t->InverseFunction[Pattern,1,2][0,_]},{t->InverseFunction[Pattern,1,2][1.459687576256715131351178232538,_]},{t->InverseFunction[Pattern,1,2][2.740312423743284868648821767462,_]}}

Και η πιθανότητα χρεοκοπίας είναι ίση με

$$P(X_T = u) = 1 + \frac{1 - E(\exp s_0\, u)}{E(\exp s_0 u)} = \frac{1}{E(\exp s_0 u)} = e^{-s_0 u}$$

Με s_0 τη μη μηδενική λύση της

$$M_w(t) = 1$$

Κεφάλαιο 5

Κατανομές με βαριά ουρά

Είναι γνωστό από τον νόμο των μεγάλων αριθμών πως όταν το μέγεθος του δείγματος τείνει στο άπειρο (είναι πολύ μεγάλο) τότε ο δειγματικός μέσος τείνει κατά πιθανότητα στην θεωρητική μέση τιμή της κατανομής (εκ των προτέρων μέση τιμή της κατανομής).Στη γλώσσα της στατιστικής αυτό σημαίνει πως αν X_n μια ακολουθία τότε

$$\overline{X_n} \xrightarrow{p} \mu \quad \acute{o}\tau\alpha\nu\; n \to \infty$$

όπου

$$\mu = \int_{-\infty}^{+\infty} x f(x) dx$$

Η στοιχειώδης λογική λέει πως όταν το μέγεθος του δείγματος τείνει στο άπειρο τότε δειγματικός μέσος θα τείνει στο μηδέν, δηλαδή

$$\frac{\sum_{i=1}^{n} X_i}{n} \to 0 \quad \acute{o}\tau\alpha\nu\; n \to \infty$$

Και αντίστοιχα η πιθανότητα

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| < \delta\right) \to 1 \quad \acute{o}\tau\alpha\nu\; n \to \infty$$

Ιδιαίτερο ενδιαφέρον παρουσιάζει η ταχύτητα με την οποία η μέση τιμή πηγαίνει στο μηδέν, δηλαδή ιδιαίτερο ενδιαφέρον έχει η ταχύτητα σύγκλισης.

Αναφέρουμε ως παράδειγμα την κανονική κατανομή η οποία ορίζεται σε ολόκληρο το $\mathbb{R}$ και είναι συμμετρική κατανομή.

Όπως αναφέρουμε στο παράρτημα της εν λόγω εργασίας η κανονική κατανομή συμβολίζεται $N(\mu, \sigma^2)$ όπου μ παράμετρος θέσης η οποία παίρνει οποιαδήποτε τιμή στο $\mathbb{R}$ και σ παράμετρο κλίμακας με $\sigma \,\epsilon (0, +\infty)$

$$\varphi(t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \quad x\epsilon(-\infty, +\infty)$$

Αρχικά θα προσπαθήσουμε να υπολογίσουμε την πιθανότητα

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| < \delta\right) = \mathbb{P}\left(-\delta < \frac{1}{n}\sum_{i=1}^{n} X_i < \delta\right) = \int_{-\delta}^{\delta} \sqrt{\frac{n}{2\pi\sigma^2}} e^{-\frac{nx^2}{2\sigma^2}} dx =$$

$$= \int_{\frac{-\delta\sqrt{n}}{\sigma}}^{\frac{\delta\sqrt{n}}{\sigma}} \sqrt{\frac{1}{2\pi}} e^{-\frac{z^2}{2}} dz$$

Δηλαδή καταλήξαμε στο εξής αποτέλεσμα

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| < \delta\right) = \int_{\frac{-\delta\sqrt{n}}{\sigma}}^{\frac{\delta\sqrt{n}}{\sigma}} \sqrt{\frac{1}{2\pi}} e^{-\frac{z^2}{2}} dz = 2\int_{0}^{\frac{\delta\sqrt{n}}{\sigma}} \sqrt{\frac{1}{2\pi}} e^{-\frac{z^2}{2}} dz$$

και

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| < \delta\right) = 1$$

Για να βρούμε την ταχύτητα σύγκλισης αρκεί να παρατηρήσουμε πως

$$\int_{\frac{-\delta\sqrt{n}}{\sigma}}^{\frac{\delta\sqrt{n}}{\sigma}} \sqrt{\frac{1}{2\pi}} e^{-\frac{z^2}{2}} dz = 2\int_{0}^{\frac{\delta\sqrt{n}}{\sigma}} \sqrt{\frac{1}{2\pi}} e^{-\frac{z^2}{2}} dz$$

Όμως κάνοντας παραγοντική ολοκλήρωση

$$2\int_{0}^{\frac{\delta\sqrt{n}}{\sigma}} \sqrt{\frac{1}{2\pi}} e^{-\frac{z^2}{2}} dz = \frac{2}{\sqrt{2\pi}}\int_{0}^{\frac{\delta\sqrt{n}}{\sigma}} (z)' \, e^{-\frac{z^2}{2}} dz =$$

$$= \frac{2}{\sqrt{2\pi}}\left[z e^{-\frac{z^2}{2}}\right]_{0}^{\frac{\delta\sqrt{n}}{\sigma}} + \frac{2}{\sqrt{2\pi}}\int_{0}^{\frac{\delta\sqrt{n}}{\sigma}} z^2 \, e^{-\frac{z^2}{2}} dz =$$

$$= \sqrt{\frac{2}{\pi}} \frac{\delta\sqrt{n}}{\sigma} e^{-\frac{n\delta^2}{2\sigma^2}} + \sqrt{\frac{2}{\pi}}\int_{0}^{\frac{\delta\sqrt{n}}{\sigma}} z^2 \, e^{-\frac{z^2}{2}} dz \leq \sqrt{\frac{2}{\pi}} \frac{\delta\sqrt{n}}{\sigma} e^{-\frac{n\delta^2}{2\sigma^2}} + \sqrt{\frac{2}{\pi}}\int_{0}^{+\infty} z^2 \, e^{-\frac{z^2}{2}} dz$$

Όμως το ολοκλήρωμα

$$\int_{0}^{+\infty} z^2 \, e^{-\frac{z^2}{2}} dz = \Gamma(3)2^3 = 32$$

Έτσι

$$\sqrt{\frac{2}{\pi}} \frac{\delta\sqrt{n}}{\sigma} e^{-\frac{n\delta^2}{2\sigma^2}} \leq \int_{\frac{-\delta\sqrt{n}}{\sigma}}^{\frac{\delta\sqrt{n}}{\sigma}} \sqrt{\frac{2}{\pi}} e^{-\frac{z^2}{2}} dz \leq \sqrt{\frac{2}{\pi}} \frac{\delta\sqrt{n}}{\sigma} e^{-\frac{n\delta^2}{2\sigma^2}} + \sqrt{\frac{2}{\pi}} 32$$

Τότε θα υπάρχει μια συνάρτηση $\gamma(z)$ τέτοια ώστε $\gamma(z)e^{-\frac{n\delta^2}{2\sigma^2}} \geq \sqrt{\frac{2}{\pi}}32 \;\; για\ κάθε\ z$

Η παραπάνω ανισότητα γράφεται

$$\sqrt{\frac{2}{\pi}}\frac{\delta\sqrt{n}}{\sigma}e^{-\frac{n\delta^2}{2\sigma^2}} \leq \int_{\frac{-\delta\sqrt{n}}{\sigma}}^{\frac{\delta\sqrt{n}}{\sigma}}\sqrt{\frac{2}{\pi}}e^{-\frac{z^2}{2}}dz \leq \sqrt{\frac{2}{\pi}}\frac{\delta\sqrt{n}}{\sigma}e^{-\frac{n\delta^2}{2\sigma^2}} + \gamma(z)e^{-\frac{n\delta^2}{2\sigma^2}}$$

$$\sqrt{\frac{2}{\pi}}\frac{\delta\sqrt{n}}{\sigma}e^{-\frac{n\delta^2}{2\sigma^2}} \leq \int_{\frac{-\delta\sqrt{n}}{\sigma}}^{\frac{\delta\sqrt{n}}{\sigma}}\sqrt{\frac{2}{\pi}}e^{-\frac{z^2}{2}}dz \leq \left(\sqrt{\frac{2}{\pi}}\frac{\delta\sqrt{n}}{\sigma} + \gamma(z)\right)e^{-\frac{n\delta^2}{2\sigma^2}}$$

Και βάζοντας τον λογάριθμο και στα δύο μέλη θα έχουμε

$$\log\sqrt{\frac{2}{\pi}}\frac{\delta\sqrt{n}}{\sigma} - \frac{n\delta^2}{2\sigma^2} \leq \log\int_{\frac{-\delta\sqrt{n}}{\sigma}}^{\frac{\delta\sqrt{n}}{\sigma}}\sqrt{\frac{2}{\pi}}e^{-\frac{z^2}{2}}dz \leq \log\left(\sqrt{\frac{2}{\pi}}\frac{\delta\sqrt{n}}{\sigma} + \gamma(z)\right) - \frac{n\delta^2}{2\sigma^2}$$

Διαιρούμε και με το n και έχουμε

$$\frac{1}{n}\log\sqrt{\frac{2}{\pi}}\frac{\delta\sqrt{n}}{\sigma} - \frac{\delta^2}{2\sigma^2} \leq \frac{1}{n}\log\int_{\frac{-\delta\sqrt{n}}{\sigma}}^{\frac{\delta\sqrt{n}}{\sigma}}\sqrt{\frac{2}{\pi}}e^{-\frac{z^2}{2}}dz \leq \frac{1}{n}\log\left(\sqrt{\frac{2}{\pi}}\frac{\delta\sqrt{n}}{\sigma} + \gamma(z)\right) - \frac{\delta^2}{2\sigma^2}$$

Και το όριο

$$\lim_{n\to\infty}\frac{1}{n}\log\int_{\frac{-\delta\sqrt{n}}{\sigma}}^{\frac{\delta\sqrt{n}}{\sigma}}\sqrt{\frac{2}{\pi}}e^{-\frac{z^2}{2}}dz = -\frac{\delta^2}{2\sigma^2}$$

Δηλαδή καταλήξαμε στο συμπέρασμα πως

$$\frac{1}{n}\log\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i\right| < \delta\right) \to -\frac{\delta^2}{2\sigma^2} \quad όταν\ n \to \infty$$

Η θεωρία των μεγάλων αποκλίσεων βρήκε αρχικά εφαρμογή στη πληροφορική, στην στατιστική, στις τηλεπικοινωνίες, και αργότερα στον αναλογισμό και στις ασφαλίσεις.

Κάθε ασφαλιστικός φορέας προσπαθεί να καλυφθεί από την εμφάνιση ενός όχι πιθανού ενδεχομένου το οποίο θα έχει αρνητικές συνέπειες για την βιωσιμότητα του φορέα. Για αυτό το λόγο θα πρέπει η εταιρεία να αντασφαλιστεί, μοιράζοντας μέρος του κινδύνου που αναλαμβάνει σε άλλους φορείς.

Στον τομέα της αντασφάλισης υπάρχει πάντοτε κάποια πιθανότητα να συμβεί μια πολύ μεγάλη ζημία, γι αυτό και στην ανάλυση μας θα χρησιμοποιήσουμε κατανομές με βαριά ουρά (σε αυτή την κατηγορία ανήκουν οι κατανομές με μακριά ουρά καθώς και οι υποεκθετικές κατανομές)

Πριν προχωρήσουμε στο θεώρημα του Crammer καλό είναι να αναφέρουμε κάποια εργαλεία που θα χρησιμοποιήσουμε. Γι αυτό το λόγο θα αναφερθούμε στα ***φράγματα του Chernoff***

φράγματα του Chernoff

Έστω X τυχαία μεταβλητή με ροπογεννήτρια συνάρτηση $M_X(t)$,τότε για κάθε $c \geq 0$ ισχύει

$$\mathbb{P}(X \geq c) \leq e^{-ct} M_X(t) \quad \text{για } t \geq 0$$

και

$$\mathbb{P}(X \leq c) \leq e^{-ct} M_X(t) \quad \text{για } t < 0$$

Παρατήρηση (πρόταση)

Αν πάρουμε την σχέση

$$\mathbb{P}(X \geq c) \leq e^{-ct} M_X(t) \quad \text{για } t \geq 0$$

και ολοκληρώσουμε και τα δύο μέλη ως προς c με $c \geq 0$ τότε η παραπάνω σχέση θα μας δώσει

$$\int_0^{+\infty} \mathbb{P}(X \geq c) dc \leq \int_0^{+\infty} e^{-ct} M_X(t) dc$$

Όμως

$$\int_0^{+\infty} e^{-ct} M_X(t) dc = M_X(t) \int_0^{+\infty} e^{-ct} dc = \frac{M_X(t)}{t}$$

Έτσι καταλήγουμε στο συμπέρασμα πως

$$\int_0^{+\infty} \mathbb{P}(X \geq c) dc \leq \frac{M_X(t)}{t}$$

Και αν αντικαταστήσουμε το c $\mu\varepsilon$ x έχουμε

$$\int_0^{+\infty} \mathbb{P}(X \geq x) dx \leq \frac{M_X(t)}{t}$$

Δηλαδή

$$\mathbb{E}(X) \leq \frac{M_X(t)}{t}$$

Το θεώρημα του Crammer

Το παράδειγμα που παραθέτουμε αφορά την κατανομή ανεξάρτητων και ισόνομων τυχαίων μεταβλητών όχι κατ ανάγκη από την κανονική κατανομή όπως στο προηγούμενο παράδειγμα.

Θα υποθέσουμε πως η κατανομή από την οποία προέρχονται έχει συνάρτηση κατανομής $F(x)$ και συνάρτηση πυκνότητας πιθανότητας $f(x)$.

Θα υποθέσουμε επίσης ότι η ροπογεννήτρια συνάρτηση της παραπάνω κατανομής είναι πεπερασμένη δηλαδή

$$M_X(t) = \int_{-\infty}^{+\infty} e^{tx}\, dF(x) = \int_{-\infty}^{+\infty} e^{tx} f(x)dx < \infty \; \gamma\iota\alpha\ \kappa\acute{\alpha}\theta\varepsilon\ t \in (-\delta, \delta)$$

Από το πρώτο παράδειγμα μας απασχόλησε η πιθανότητα

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| < \delta\right)$$

Χρησιμοποιώντας το φράγμα του ***Chernoff*** και θεωρώντας $t \geq 0$ μπορούμε να υπολογίσουμε ένα άνω φράγμα για την πιθανότητα

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| \geq \delta\right) = \mathbb{P}\left(\left|\sum_{i=1}^{n} X_i\right| \geq n\delta\right)$$

Έτσι

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} X_i\right| \geq n\delta\right) \leq e^{-n\delta t} M_{|\sum_{i=1}^{n} X_i|}(t)$$

Δηλαδή

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} X_i\right| \geq n\delta\right) \leq e^{-n\delta t} M_{|\sum_{i=1}^{n} X_i|}(t) \leq e^{-n\delta t}\, \mathbb{E}(e^{t\sum_{i=1}^{n} X_i}) \leq e^{-n\delta t} M_X(t)^n$$

Λογαριθμώντας τις παραπάνω σχέσεις και διαιρώντας με το n έχουμε

$$\frac{1}{n}\log\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| < \delta\right) \leq -(\delta t - \log M_X(t))\; \gamma\iota\alpha\ \kappa\acute{\alpha}\theta\varepsilon\ t \geq 0$$

άρα

$$\frac{1}{n}\log\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i\right| < \delta\right) \leq -Sup\;(\delta t - \log M_X(t))$$

Μπορούμε να ορίσουμε με $I(t)$

$$I(t) = -Sup\;(\delta t - \log M_X(t))$$

Λήμμα[5]

Για κάθε $\delta > 0\; και\; x \geq \bar{x}$ έχουμε

$$liminf\,\frac{1}{n}log\mathbb{P}\left(x \leq \frac{1}{n}\sum_{i=1}^{n}X_i \leq x + \delta\right) \geq -I(x)$$

Παρατήρηση 1

$$\mathbb{P}\left(\frac{1}{n}\left|\sum_{i=1}^{n}X_i\right| \geq \delta\right) \cong e^{-n\delta t}e^{\mathrm{nln}\,M_X(t)}$$

Αν οι τυχαίες μεταβλητές X_i παίρνουν μη αρνητικές τιμές τότε

$$\mathbb{P}(\bar{X} \geq \delta) \cong e^{-n\delta t}e^{\mathrm{nln}\,M_X(t)}$$

Μπορούμε να ολοκληρώσουμε και τα δύο μέλη της παραπάνω ανισότητας και έτσι έχουμε

$$\int_0^{+\infty}\mathbb{P}(\bar{X} \geq \delta)d\delta \cong e^{\mathrm{nln}\,M_X(t)}\int_0^{+\infty}e^{-n\delta t}d\delta$$

Όμως

$$\int_0^{+\infty}e^{-n\delta t}d\delta = \frac{1}{nt}$$

Άρα

$$\int_0^{+\infty}\mathbb{P}(\bar{X} \geq \delta)d\delta \cong \frac{e^{\mathrm{nln}\,M_X(t)}}{nt}$$

Ισοδύναμα η παραπάνω σχέση γράφεται

$$\int_0^{+\infty}\mathbb{P}(\bar{X} \geq \delta)d\delta \cong \frac{M_X(t)^n}{nt}$$

Έχουμε όμως δείξει πως

$$\mathbb{E}(X) \leq \frac{M_X(t)}{t}$$

$$\int_0^{+\infty}\mathbb{P}(\bar{X} \geq \delta)d\delta \geq \mathbb{E}(X)\,\frac{M_X(t)^{n-1}}{n}$$

Δηλαδή

$$\mathbb{E}(\bar{X}) \geq \frac{\mathbb{E}(X)}{n}$$

Και

$$\mathbb{E}(\bar{X}) \to 0 \ \text{ όταν } n \to \infty$$

Παρατήρηση 2

$$\sum_{\kappa=0}^{\infty} \frac{\lambda^{\kappa}}{\kappa!} E(X^{\kappa})$$

$$\int_{0}^{+\infty} \mathbb{P}(\bar{X} \geq \delta) d\delta \cong \frac{M_X(t)^n}{nt} \cong \frac{1}{nt} \left(\sum_{\kappa=0}^{\infty} \frac{t^{\kappa}}{\kappa!} E(X^{\kappa}) \right)^n$$

$$\cong \frac{1}{t} \frac{\left(\sum_{\kappa=0}^{\infty} \frac{t^{\kappa}}{\kappa!} E(X^{\kappa}) \right)^n}{n}$$

Δηλαδή

$$\mathbb{E}(\bar{X}) \cong \frac{1}{t} \frac{\left(\sum_{\kappa=0}^{\infty} \frac{t^{\kappa}}{\kappa!} E(X^{\kappa}) \right)^n}{n}$$

Και ορίζουμε μια συνάρτηση

$$h(n) = \frac{\left(\sum_{\kappa=0}^{\infty} \frac{t^{\kappa}}{\kappa!} E(X^{\kappa}) \right)^n}{n}$$

Έτσι

$$\mathbb{E}(\bar{X}) \cong \frac{1}{t} h(n)$$

Και επειδή

$$\int_{-\infty}^{+\infty} e^{tx} f(x) dx < \infty \text{ για κάθε } t \in (-\delta, \delta)$$

$$\sum_{\kappa=0}^{\infty} \frac{t^{\kappa}}{\kappa!} E(X^{\kappa}) = c < \infty$$

Η μορφή της $h(n)$ θα είναι

$$h(n) = \frac{c^n}{n}$$

Παρατήρηση 3

Έχουμε δείξει πως

$$\int_0^{+\infty} \mathbb{P}(\bar{X} \geq \delta)d\delta \cong \frac{M_X(t)^n}{nt} \cong \frac{1}{nt}\left(\sum_{\kappa=0}^{\infty} \frac{t^\kappa}{\kappa!} E(X^\kappa)\right)^n$$

Την παραπάνω σχέση μπορούμε να την γράψουμε ως εξής

$$\frac{t\int_0^{+\infty} \mathbb{P}(\bar{X} \geq \delta)d\delta}{\left(\left(\sum_{\kappa=0}^{\infty} \frac{t^\kappa}{\kappa!} E(X^\kappa)\right)^n\right)} \cong \frac{1}{n}$$

$$\frac{nt\mathbb{E}(\bar{X})}{\left(\left(\sum_{\kappa=0}^{\infty} \frac{t^\kappa}{\kappa!} E(X^\kappa)\right)^n\right)} \cong 1$$

Δηλαδή

$$\left(\sum_{\kappa=0}^{\infty} \frac{t^\kappa}{\kappa!} E(X^\kappa)\right)^n \cong nt\mathbb{E}(\bar{X})$$

$$1 + t\mathbb{E}(X) + \sum_{\kappa=2}^{\infty} \frac{t^\kappa}{\kappa!} E(X^\kappa) \cong \left(nt\mathbb{E}(\bar{X})\right)^{\frac{1}{n}} \cong \left(nt\mathbb{E}(X)\right)^{\frac{1}{n}}$$

Γνωρίζουμε επίσης ότι ισχύει

$$1 + t\mathbb{E}(X) + \sum_{\kappa=2}^{\infty} \frac{t^\kappa}{\kappa!} E(X^\kappa) \geq 1 + t\mathbb{E}(X) + \sum_{\kappa=2}^{\infty} \frac{t^\kappa}{\kappa!} E(X)^k$$

Δηλαδή

$$\sum_{\kappa=0}^{\infty} \frac{t^\kappa}{\kappa!} E(X^\kappa) \geq \sum_{\kappa=0}^{\infty} \frac{\left(tE(X)\right)^\kappa}{\kappa!}$$

Δηλαδή

$$\left(\sum_{\kappa=0}^{\infty} \frac{t^\kappa}{\kappa!} E(X^\kappa)\right)^n \geq \left(\sum_{\kappa=0}^{\infty} \frac{\left(tE(X)\right)^\kappa}{\kappa!}\right)^n$$

Καταλήξαμε λοιπόν στο συμπέρασμα

$$\left(\sum_{\kappa=0}^{\infty} \frac{t^\kappa}{\kappa!} E(X^\kappa)\right)^n \geq e^{ntE(X)}$$

Λογαριθμούμε και ταδύο μέλη της παραπάνω σχέσης και έχουμε

$$\log\left(\sum_{\kappa=0}^{\infty}\frac{t^{\kappa}}{\kappa!}E(X^{\kappa})\right) \geq tE(X)$$

Κατασκευάσαμε ένα άνω φράγμα την μέση τιμή

$$\frac{\log\left(\sum_{\kappa=0}^{\infty}\frac{t^{\kappa}}{\kappa!}E(X^{\kappa})\right)}{t} \geq E(X)$$

Τα παραπάνω ισχύουν όταν το ολοκλήρωμα

$$\int_{-\infty}^{+\infty} e^{tx} f(x)dx < \infty \; για \; κάθε \; t \geq 0$$

Στην περίπτωση όπου

$$\int_{-\infty}^{+\infty} e^{tx} f(x)dx = \infty \; για \; κάθε \; t \geq 0$$

λέμε ότι η κατανομή F είναι κατανομή με βαριά ουρά (θυμίζουμε πως αν μια κατανομή είναι υποεκθετική τότε είναι και κατανομή με βαριά ουρά).

Για τις κατανομές που χαρακτηρίζονται ως υποεκθετικές ισχύουν τα εξής

$$\frac{\mathbb{P}(S_n > x)}{n\mathbb{P}(X > x)} \to 1 \; όταν \; x \to \infty$$

Από την παραπάνω σχέση μπορούμε να πάρουμε

$$\left|\frac{\mathbb{P}(S_n > x)}{n\mathbb{P}(X > x)} - 1\right| \to 0 \;\; όταν \; x \to \infty$$

Όμως για κάθε $n \in N$ υπάρχει X_n τέτοιο ώστε

$$\left|\frac{\mathbb{P}(S_n > x)}{n\mathbb{P}(X > x)} - 1\right| \leq \frac{1}{n}$$

Η παραπάνω ανισότητα γράφεται

$$-\frac{1}{n} \leq \frac{\mathbb{P}(S_n > x)}{n\mathbb{P}(X > x)} - 1 \leq \frac{1}{n}$$

Παίρνοντας το όριο για $n \to \infty$ προκύπτει

$$\frac{\mathbb{P}(S_n > x)}{n\mathbb{P}(X > x)} - 1 \to 0 \;\; για \; n \to \infty$$

Στην περίπτωση όμως που το μέγεθος του δείγματος είναι αρκετά μεγάλο μπορούμε να χρησιμοποιήσουμε την κανονική κατανομή και ποιο συγκεκριμένα αν $x = n\mu + c\sqrt{\log n}\,\sigma$

έχουμε

$$\left|\frac{\mathbb{P}(S_n > x)}{n\bar{\Phi}(c\sqrt{\log n})} - 1\right| \to 0 \;\; \gamma\iota\alpha \; n \to \infty$$

5.1 Δεξιά ουρά κατανομής

Δεξιά ουρά (***survival function***) μιας κατανομής που έχει συνάρτηση κατανομής $F(x) =$ P$(X > x)$ συμβολίζεται με $\overline{F}(x)$ ή $S(x)$ και είναι ίση με την πιθανότητα του ενδεχομένου η τυχαία μεταβλητή X να πάρει τιμές μεγαλύτερες του x. Στην ουσία πρόκειται για την πιθανότητα μια τυχαία μεταβλητή X να πάρει μεγαλύτερη τιμή απο μια σταθερά x, πρόκειται δηλαδή για την πιθανότητα του ενδεχομένου $\{X > x\}$.

Ορίζουμε με $\bar{F}$ την δεξιά ουρά μιας κατανομής F στο R ως εξής

$$\bar{F}(x) = F(x, \infty) = \mathrm{P}(X > x)$$

για κάθε x.

Έτσι

$$\bar{F}(x) = \begin{cases} \int_x^\infty f(x)dx \;(\; \sigma\upsilon\nu\varepsilon\chi\acute{\eta}\varsigma \; \pi\varepsilon\rho\acute{\iota}\pi\tau\omega\sigma\eta \;) \\ \sum_{\kappa=\kappa+1}^{\infty} P\,(X = k)\;(\delta\iota\alpha\kappa\rho\iota\tau\acute{\eta}\;\pi\varepsilon\rho\iota\pi\tau\omega\sigma\eta) \end{cases}$$

Γιατί όμως μας ενδιαφέρει η πιθανότητα $P(X \geq x)$; Η απάντηση δεν είναι δύσκολη ένα σκεφτούμε πως στον κλάδο τις ασφάλισης και της αντασφάλισης η τυχαία μεταβλητή X εκφράζει το ύψος της ζημιάς (άρα και της απαίτησης του ασφαλισμένου στην ασφαλιστική εταιρία).Η τυχαία μεταβλητή X είναι μια μη αρνητική τυχαία μεταβλητή (P $(X \geq 0) = 1$) και έχει μεγάλο ενδιαφέρον για ασφαλιστή και αντασφαλιστή να ξέρουν, να μπορούν να υπολογίσουν δηλαδή ποια είναι η πιθανότητα να προκληθεί μια πολύ μεγάλη ζημιά.

Στην ανάλυση μας θα χρησιμοποιήσουμε όρους όπως βαριά ουρά και λεπτή ουρά. Καλό θα είναι λοιπόν να εξηγήσουμε στον αναγνώστη τι εννοούμε όταν αναφέρουμε τέτοιους όρους .

5.2 Βαριά ουρά (*heavy tailed distributions)*

Μια κατανομή με βαριά ουρά (heavy tailed distribution) είναι μια κατανομή της οποίας η ουρά (δηλαδή η συνάρτηση επιβίωσης) συγκλίνει πολύ αργά στο μηδέν όταν το x τείνει στο άπειρο.

Τυχαίες μεταβλητές που ακολουθούν κατανομές με βαριές δεξιές ουρές δηλαδή τμ των οποίων η $P(X \geq x)$ όταν $x \rightharpoonup \infty$ φθίνει πολύ αργά στο μηδέν (με άλλα λόγια υπάρχει το ενδεχόμενο να συμβεί μια πολύ μεγάλη ζημιά) δεν έχουν ροπογεννήτρια συνάρτηση χωρίς αυτό να σημαίνει πως δεν έχουν ροπές.Κατανομές με βαρία ουρά ικανοποιούν τη σχέση $\bar{F}(x) \geq 0 \quad για\ κάθε\ x.$

Μια κατανομή F ορισμένη στο R θα λέμε οτι έχει δεξιά βαριά ουρά αν

$$\int_{-\infty}^{+\infty} e^{\lambda x} dF(x) = \infty\ για\ κάθε\ \lambda > 0$$

- Αν η συνάρτηση κατανομής F είναι συνεχής με πυκνότητα f τότε

$$\int_{-\infty}^{+\infty} e^{\lambda x} dF(x) = \int_{-\infty}^{+\infty} e^{\lambda x} f(x) dx$$

- Στη γενική περίπτωση όπου η Fείναι μια μεικτή κατανομή με συνάρτηση πυθανότητας f_1 για το διακριτό τμήμα και f_2 για το συνεχές τμήμα τότε

$$\int_{-\infty}^{+\infty} e^{\lambda x} dF(x) = \int_{-\infty}^{+\infty} e^{\lambda x} f_2(x) dx + \sum_{x \in R_x} e^{\lambda x} f_1(x)$$

Ένα βασικό χαρακτηριστικό μιας κατανομής F με βαριά ουρά είναι ότι έχει αυτό που στη διεθνή βιβλιογραφία αναφέρεται ως *right unbounded support* δηλαδή ισχύει

$$\overline{F}(x) > 0\ για\ κάθε\ x$$

Ένας εναλλακτικός ορισμός για την ιδιότητα της βαρίας ουράς μιας κατανομής είναι αυτός που έδωσαν οι *Asmussen και Soren στο βιβλίο Applied probability and queues*

$$\lim_{x \to \infty} sup e^{\lambda x} \overline{F}(x) = \infty\ για\ όλα\ τα\ \lambda > 0$$

5.3 Παραδείγματα κατανομών με βαρία ουρά

- *Lognormal*
- *Pareto*
- *Weibull με παράμετρο σχήματος μικρότερη της μονάδας*
- *Pareto type distributions*

5.4 Αριθμητικά παράδειγμα κατανομής με βαριά ουρά

$\boldsymbol{Pareto(a = 2, b = 1)}$

Στον παρακάτω πίνακα βλέπουμε πως συμπεριφέρεται μια κατανομή $Pareto$ με παραμέτρους $\alpha = 2, b = 1\ και\ για\ \lambda = 2\ e = 2{,}71$

Έχουμε κάνει επίσης την παραδοχή πως η ποσότητα

$$\overline{F}(x) = \left(\frac{b}{b+x}\right)^{a} \cong x^{-a} \, για \, x \to \infty$$

X	$e^{\lambda x}$	x^{-a}	$e^{\lambda x}x^{-a}$
1	7,3341	1	7,3441
10	456442229	0,01	4564422
100	3,925167e+86	1e-04	3,925167e+82
300	6,047481e+259	1,111111e-05	6,719424e+254
350	1,198129e+303	8,163265e-06	9,780646e+297
355	2,559744e+307	7,934934e-06	2,03114e+302
356	Inf	7,890418e-06	Inf

Παρατηρούμε πως για τιμές του x πάνω απο 355 παρατηρήσεις ο όρος $e^{\lambda x}$ απειρίζεται.

Weibull (α=0.5,c=2)

X	$e^{\lambda x}$	$e^{(x/c)^{-a}}$	$e^{\lambda x}e^{(x/c)^{-a}}$
1	7,3341	0,4941337	3,628967
10	456442229	0,1076097	49117611
100	3,925167e+86	0,0008678501	3,406457e+83
300	6,047481e+259	4,980011e-06	3,011652e+254
350	1,198129e+303	1,872204e-06	2,243142e+279
355	2,559744e+307	1,704469e-06	4,363004e+301
356	Inf	1,672904e-06	Inf

Wiebull(α=2,c=2)

X	$e^{\lambda x}$	$e^{(x/c)^{-a}}$	$e^{\lambda x}e^{(x/c)^{-a}}$
1	7,3341	0,7793951	5,716162
10	456442229	1,498883e-11	0,006841535
100	3,925167e+86	0	0
300	6,047481e+259	0	0
350	1,198129e+303	0	0
355	2,559744e+307	0	0
356	Inf	0	NaN

Απο τους δύο παραπάνω πίνακες μπορούμε να βγάλουμε το εξής συμπέρασμα αν η παράμετρος σχήματος $\alpha > 1$ τότε η κατανομή Wiebull συμπεριφέρεται σαν μια κατανομή με ελαφριά ουρά μιας και η ποσότητα

$$\lim_{x\to\infty} supe^{\lambda x}\,\overline{F}(x) < \infty \; \gamma\iota\alpha \; \acute{o}\lambda\alpha \; \tau\alpha \; \lambda > 0$$

Ενώ αν αν η παράμετρος σχήματος $\alpha < 1$ τότε η κατανομή Wiebull συμπεριφέρεται σαν μια κατανομή με βαριά ουρά μιας και η ποσότητα

$$\lim_{x\to\infty} supe^{\lambda x}\,\overline{F}(x) = \infty \; \gamma\iota\alpha \; \acute{o}\lambda\alpha \; \tau\alpha \; \lambda > 0$$

5.5 κατανομή τύπου Pareto

Μια κατανομή F θα λέμε ότι είναι τύπου Pareto με παράμετρο $\alpha > 0$ αν

$$\overline{F}(x) \sim L(x)x^{-a} \quad x \to \infty$$

Με τη συνάρτηση $L(x)$ να είναι μια θετική συνάρτηση L:R+→ $(0,\infty)$ που ικανοποιεί την σχέση

$$\frac{L(xy)}{L(x)} \to 1 \qquad y > 0 \qquad x \to \infty$$

Παραδείγματα κατανομών τύπου Pareto είναι

- Η κατανομή $Par(a)$ με συνάρτηση πυκνότητας πιθανότητας

$$f(x) = a(1+x)^{-(a+1)}$$

- Η κατανομή $loggamma$ με συνάρτηση πυκνότητας πιθανότητας

$$f(x) = \frac{\lambda^{\alpha}}{\Gamma(\alpha)}(logx)^{a-1}x^{-\lambda-1} \qquad x > 1$$

- Κατανομή $Burr$ με δεξιά ουρά

$$\overline{F}(x) = \left(\frac{b}{b+x^r}\right)^a \qquad x \geq 0$$

5.6 Βαθμίδα αποτυχίας (ή συνάρτηση διακυνδύνευσης)

Έχοντας ως στόχο την ποσοτικοποίηση και την ανάλυση του μεγέθους της δεξιάς ουράς (βαριάς ή λεπτής) εισάγουμε τον όρο failure rate ή wizard rate .Συμβολίζουμε τον βαθμό αποτυχίας :

$$r(x) = \frac{f(x)}{\overline{F}(x)} = -\frac{d}{d(x)}\ln \overline{F}\,(x) \qquad \gamma\iota\alpha \; x \geq 0$$

Γνωρίζουμε ότι η παράγωγος μιας συνάρτησης $F(x)$ ορίζεται ως εξής :

$$\frac{dF(x)}{dx} = f(x) = \lim_{h\to 0} \frac{F(x+h) - F(x)}{h}$$

Έτσι λοιπόν μια εναλλακτική σχέση που μας δίνει τη βαθμίδα αποτυχίας είναι η εξής :

$$r(x) = \frac{f(x)}{\bar{F}(x)} = \lim_{h\to 0} \frac{1}{h}\left\{1 - \frac{\bar{F}(x+h)}{\bar{F}(x)}\right\}$$

5.7 Βαθμίδα αποτυχίας στην διακριτή περίπτωση:

Στη διακριτή περίπτωση ο βαθμός αποτυχίας γνωστός και ως ένταση θνησιμότητας (force of mortality) μπορεί να οριστεί ως εξής :

$$h_u = P(N = n|N \geq n) = \frac{P(N = n, N \geq n)}{P(N \geq n)} = \frac{p_n}{p_n + a_n} \quad n = 0,1,2, \dots ..$$

Πρόταση

Η κατανομή F έχει βαριά δεξιά ουρά όταν το όριο της βαθμίδα αποτυχιας τείνει στο μηδέν όταν το x τείνει στο άπειρο δηλαδή όταν $\lim_{x\to\infty} r(x) = 0$

5.8 Σωρευτική συνάρτηση διακυνδύνευσης

Η σωρευτική συνάρτηση διακινδύνευσης ορίζεται ως

$$R(x) = \int_{-\infty}^{x} r(u)du = \int_{-\infty}^{x} \frac{f(u)}{\bar{F}(u)} du =$$

$$R(x) = \int_{0}^{x} r(u)du = \int_{0}^{x} \frac{f(u)}{\bar{F}(u)} du = -\ln \bar{F}(t)$$

5.9 Κατανομές με μακριά ουρά (*long tailed distributions)*

Μια κατανομή F ορισμένη στο R θα λέμε οτι έχει μακριά ουρά αν $\bar{F}(x) > 0$ *(right unbounded support) και για κάθε σταθερό* $y > 0$ *ισχύει*

$$\frac{\bar{F}(x+y)}{\bar{F}(x)} \to 1 \quad όταν\ x \to \infty$$

Η παραπάνω σχέση απαιτεί έναν βαθμό ομαλότητας της $\bar{F}(x)$ που δεν εμφανίζεται σε κάθε κατανομή με βαριά ουρά.

Μια κατανομή με μακριά ουρά είναι και κατανομή με βαριά ουρά, χώρις να ισχύει πάντα και το αντίθετο δηλαδή μια κατανομή με βαριά ουρά δεν είναι βέβαιο πως θα είναι και κατανομή με μακριά ουρά.

Εναλλάκτικα θα μπορούσαμε κάνοντας χρήση της συνάρτησης σωρευτικής διακινδύνευσης να δώσουμε τον ισοδύναμο ορισμό

Μια κατανομή F ορισμένη στο R θα λέμε οτι έχει μακριά ουρά αν $F(x) > 0$ *(right unbounded support) και για κάθε σταθερό* $y > 0$ *ισχύει*

$$R(x) - R(x+y) \ \to 0 \quad όταν\ x \to \infty$$

Απόδειξη

$$\lim_{x\to\infty} \frac{\bar{F}(x+y)}{\bar{F}(x)} = 1$$

Παίρνωντας τον λογάριθμο της παραπάνω σχέσης έχουμε

$$\lim_{x\to\infty} \ln\left(\frac{\bar{F}(x+y)}{\bar{F}(x)}\right) = \ln 1 = 0$$

$$\lim_{x\to\infty}\{\ln \bar{F}(x+y) - \ln \bar{F}\,(x)\} = 0$$

Η παραπάνω σχέση μπορεί να γραφεί με την βοήθεια της συνάρτησης σωρευτικής διακινδύνευσης ως εξής

$$\lim_{x\to\infty}(R(x) - R(x-y)) = 0\ για\ y \geq 0$$

5.10 Θεώρημα ομοιόμορφης σύγκλισης και βαριά ουρά

Σύμφωνα με το θεώρημα ομοιόμορφης σύγκλισης (*uniform convergence theorem)* για κάθε $\alpha > 0$ έχουμε

$$sup_{|y|<a}|\bar{F}(x+y) - \bar{F}(x)| = o\big(\bar{F}(x)\big) \quad x \to \infty$$

Από την παραπάνω σχέση παρατηρούμε τα εξής

$$\lim_{x\to\infty} \frac{\bar{F}(x+y) - \bar{F}(x)}{\bar{F}(x)} = \lim_{x\to\infty} \frac{\bar{F}(x+y)}{\bar{F}(x)} - 1 = 0$$

Δηλαδή

$$\lim_{x\to\infty} \frac{\bar{F}(x+y)}{\bar{F}(x)} = 1$$

Σύμφωνα λοιπόν με όλα τα παραπάνω αν μια συνάρτηση κατανομής ικανοποιεί το θεώρημα ομοιόμορφης σύγκλισης τότε αυτή η κατανομή θα έχει βαριά δεξιά ουρά.

Μπορούμε να επεκτείνουμε το θεώρημα ομοιόμορφης σύγκλισης αν στη θέση του $\alpha > 0$ βάλουμε μια αυστηρά θετική μη φθίνουσα συνάρτηση h *τέτοια ώστε* $h(x) \to \infty \;\; x \to \infty$

Έτσι το θεώρημα παίρνει τη μορφή

$$sup_{|y|<h(x)}|\bar{F}(x+y) - \bar{F}(x)| = o\big(\bar{F}(x)\big) \quad x \to \infty$$

Και θα λέμε ότι η συνάρτηση $\bar{F}(x)$ είναι *h-insensitive*

Καταλήξαμε λοιπόν στο συμπέρασμα πως κάθε κατανομή με μακριά ουρά είναι μια *h-insensitive* κατανομή.

Αν αντικαταστήσουμε στη σχέση

$$\lim_{x\to\infty} \frac{\bar{F}(x+y)}{\bar{F}(x)}$$

Όπου $y = h(x)$ με $h(x) \to \infty \;\; x \to \infty$ θα έχουμε

$$\lim_{x\to\infty} \frac{\bar{F}(x+h(x))}{\bar{F}(x)}$$

Η κατανομή $\bar{F}(x)$ είναι κατανομή με μακριά ουρά άρα και κατανομή με βαριά ουρά.Έστω λοιπόν οτι πρόκειται για μια κατανομή Pareto type με δεξιά ουρά να δίνεται απο τη σχέση

$$\bar{F}(x) \sim L(x) x^{-a} \qquad x \to \infty$$

η συνάρτηση $h(x) = kx\,, x \geq 0.$

Έτσι η παραπάνω σχέση θα μας δώσει

$$\lim_{x\to\infty} \frac{\bar{F}(x+h(x))}{\bar{F}(x)} = \frac{L(x+h(x))(x+h(x))^{-a}}{L(x)x^{-a}} = \lim_{x\to\infty}\left\{\frac{L\big(x+h(x)\big)}{L(x)} \frac{\big(x+h(x)\big)^{-a}}{x^{-a}}\right\} =$$

$$\lim_{x\to\infty} \frac{L\big(x+h(x)\big)}{L(x)} \lim_{x\to\infty}\left(\frac{x+h(x)}{x}\right)^{-a} =$$

$$= \lim_{\substack{x\to\infty \\ k\to 0}} \left(1 + \frac{x(1+k)}{x}\right)^{-a} = 1$$

5.11 Intermediate regularly varying distribution

Σύμφωνα με το παραπάνω θεώρημα η y μπορεί να ισούται με την $h(x)$ Αν λοιπόν για μια συνάρτηση $h(x)$ με $h(x) = o(x)$ ισχύει η σχέση

$$\lim_{x\to\infty} \frac{\overline{F}(x + h(x))}{\overline{F}(x)} = 1$$

Τότε η κατανομή F ονομάζεται ***Intermediate regularly varying distribution***

Πρόταση

Αν $\lim_{x\to\infty} r(x) = 0$ τότε η κατανομή F είναι κατανομή με βαριά ουρά.

Απόδειξη

$$r(x) = \frac{f(x)}{\overline{F}(x)} = -\frac{d}{d(x)} \ln \overline{F}\,(x) \qquad \gamma\iota\alpha\ x \geq 0$$

Γνωρίζουμε ότι η παράγωγος μιας συνάρτησης $F(x)$ ορίζεται ως εξής :

$$\frac{dF(x)}{dx} = f(x) = \lim_{h\to 0} \frac{F(x+h) - F(x)}{h}$$

Έτσι λοιπόν μια εναλλακτική σχέση που μας δίνει τη βαθμίδα αποτυχίας είναι η εξής :

$$r(x) = \frac{f(x)}{\overline{F}(x)} = \lim_{h\to 0} \frac{1}{h} \left\{ 1 - \frac{\overline{F}(x+h)}{\overline{F}(x)} \right\}$$

Απο την παραπάνω σχέση καταλαβαίνουμε πως το $\lim_{x\to\infty} r(x) = 0$ είναι γεγονός όταν και μόνο όταν

$$\lim_{x\to\infty} \left\{ 1 - \frac{\overline{F}(x+h)}{\overline{F}(x)} \right\} = 0$$

ισοδύναμα

$$\lim_{x\to\infty} \frac{\overline{F}(x+h)}{\overline{F}(x)} = 1$$

Πρόταση

Αν $\lim sup_{x\to\infty} \frac{R(x)}{x} = 0$ τότε η F είναι κατανομή με βαριά ουρά.

Απόδειξη

$$r(x) = \frac{f(x)}{\bar{F}(x)} = -\frac{d}{d(x)} \ln \bar{F}\,(x) \qquad \text{για } x \geq 0$$

Και

$$R(x) = \int_{-\infty}^{x} r(u)du = \int_{-\infty}^{x} \frac{f(u)}{\bar{F}(u)} du =$$

$$R(x) = \int_{0}^{x} r(u)du = -\int_{0}^{x} \frac{f(u)}{\bar{F}(u)} du = -\ln(\bar{F}(x))$$

Τότε υπάρχει x' τέτοιο ώστε για κάθε $x \geq x'$

$$-\ln\big(\bar{F}(x)\big) \leq \varepsilon x$$

$$\ln\big(\bar{F}(x)\big) \geq -\varepsilon x$$

Και εισάγοντας μια σταθερά c μπορούμε να πούμε πως ισχύει η σχέση

$$\bar{F}(x) \geq ce^{-\varepsilon x} \text{ για κάθε } x \geq 0$$

Μια κατανομή λέμε ότι έχει βαριά ουρά όταν ισχύει

$$\int_{0}^{+\infty} e^{\lambda x} \bar{F}(x)dx \geq \int_{0}^{+\infty} e^{\lambda x}\, ce^{-\varepsilon x} dx$$

Όμως το ολοκλήρωμα

$$\int_{0}^{+\infty} e^{\lambda x}\, ce^{-\varepsilon x} dx = c\int_{0}^{+\infty} e^{x(\lambda-\varepsilon)}\, dx =$$

$$\frac{c}{s-\varepsilon}\left[e^{x(s-\varepsilon)}\right]_{0}^{+\infty} = \infty \qquad \mu\varepsilon\ s > \varepsilon$$

Και επειδή το ε είναι αυθαίρετο μπορούμε να πούμε ότι η παραπάνω σχέση ισχύει για $s > 0$

Έτσι καταλήξαμε στο συμπέρασμα πως αν $\lim sup_{x\to\infty} \frac{R(x)}{x} = 0$ τότε η κατανομή F είναι μια κατανομή με βαριά ουρά.

5.12 Βαριά ουρά και ροπές κ τάξης

Έχουμε ήδη αναφέρει πως μια κατανομή θα λέμε ότι έχει βαριά ουρά αν ισχύει η σχέση

$$\int_{-\infty}^{+\infty} e^{\lambda x} dF(x) = \infty \; \gamma\iota\alpha \; \kappa\acute{\alpha}\theta\varepsilon \; \lambda > 0$$

Κάνοντας χρήση δυναμοσειρών και μάλιστα με τη βοήθεια της σειράς *Taylor* η παραπάνω σχέση μπορει να γραφεί ως εξής

$$\int_{-\infty}^{+\infty} e^{\lambda x} dF(x) = \int_{-\infty}^{+\infty} \sum_{\nu=0}^{\infty} \frac{(\lambda x)^{\nu}}{\nu!} =$$

$$\int_{-\infty}^{+\infty} f(x)dx + \lambda \int_{-\infty}^{+\infty} xf(x)dx + \frac{\lambda^2}{2!}\int_{-\infty}^{+\infty} x^2 f(x)dx + \cdots =$$

$$1 + \lambda E(X) + \frac{\lambda^2}{2!}E(X^2) + \cdots . + \frac{\lambda^{\kappa}}{\kappa!}E(X^{\kappa}) + \cdots =$$

$$= \sum_{\kappa=0}^{\infty} \frac{\lambda^{\kappa}}{\kappa!} E(X^{\kappa})$$

Έτσι προκύπτει πως μια κατανομή θα λέμε ότι έχει δεξιά βαριά ουρά αν και μόνο αν

$$\sum_{\kappa=0}^{\infty} \frac{\lambda^{\kappa}}{\kappa!} E(X^{\kappa}) = \infty \qquad \gamma\iota\alpha \; \kappa\acute{\alpha}\theta\varepsilon \; \lambda > 0$$

Αρκεί λοιπόν να υπάρξει έστω και μια ροπή η οποία να απειρίζεται και τότε η κατανομή F θα λέμε ότι έχει βαριά ουρά.

Έστω ότι η κατανομή F είναι ορισμένη στον θετικό ημιάξονα ℝ+ και δεν είναι όλες οι ροπές της πεπερασμένες δηλαδή

$$\int_{0}^{+\infty} x^{\kappa} f(x)dx = \infty \qquad \gamma\iota\alpha \; \kappa\acute{\alpha}\pi o\iota o \; \kappa$$

σε αυτή την περίπτωση μπορούμε να βρούμε ένα τέτοιο $\kappa \geq 1$ τέτοιο ώστε η κ ροπή να απειρίζεται και η $(\kappa - 1)$ ροπή να είναι πεπερασμένη.

Αξίζει να σημειωθεί πως αν η κ ροπή απειρίζεται τότε και κάθε ροπή τάξης μεγαλύτερης της κ θα απειρίζεται.

Σύμφωνα λοιπόν με τα παραπάνω έχουμε

$$\int_0^{+\infty} x^{\kappa} f(x)dx = \int_0^{+\infty} e^{\kappa \ln x} f(x)dx = \int_0^{+\infty} xe^{(\kappa-1)\ln x} f(x)dx = \infty$$

και

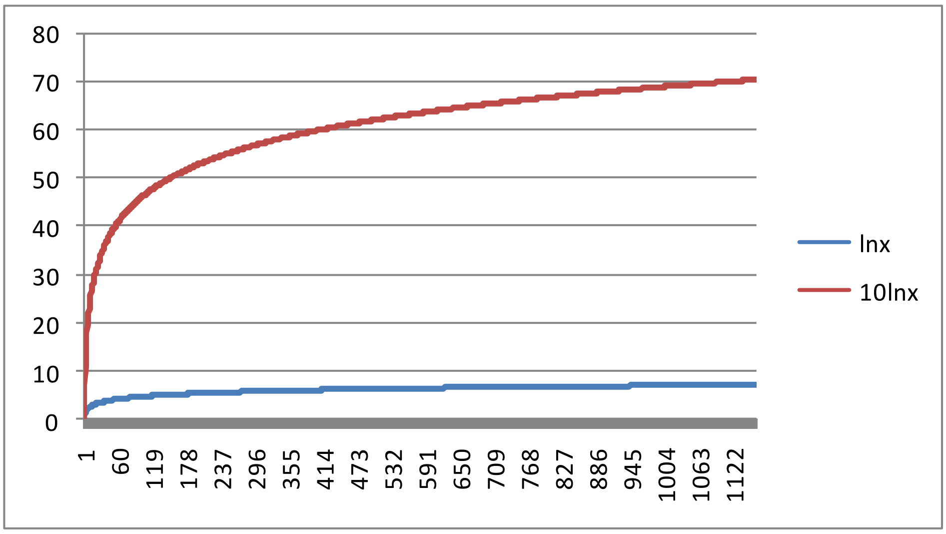
$$\int_0^{+\infty} x^{\kappa-1} f(x)dx = \int_0^{+\infty} e^{(\kappa-1)\ln x} f(x)\, dx < \infty$$

Θεώρημα

Έστω $\xi \geq 0$ μια μη αρνητική τυχαία μεταβλητή που ακολουθεί μια κατανομή με βαριά δεξιά ουρά.Έστω επίσης μια αύξουσα συνάρτηση $g(x)$ τέτοια ώστε $g(x) \to \infty$ $όταν\ x \to \infty$.

Τότε υπάρχει μια μονότονη κοίλη συνάρτηση h: R+ → R+ τέτοια ώστε
$h(x) = o(x)$ $όταν\ x \to \infty$, $\mathrm{E}e^{h(\xi)} < \infty$ και $\mathrm{E}e^{h(\xi)+g(\xi)} = \infty$

Το παραπάνω θεώρημα εγγυάται την ύπαρξη μιας κοίλης συνάρτησης h για κάθε g την οποία μπορούμε να την κατασκευάσουμε όσο λεπτή εμείς θέλουμε.



Στο παραπάνω γράφημα έχουμε πάρει ως $g(x) = \ln x$ και ως $h(x) = 10 \ln x$ δηλαδή την συνάρτηση ως $h(x) = (\kappa - 1) \ln x$ για $\kappa = 11$.

Εύκολα παρατηρούμε πως για την $g(x) = \ln x$ ισχύει $g(x) \to \infty$ $όταν\ x \to \infty$.

και για την $h(x) = (\kappa - 1) \ln x$ ισχύει

$$\lim_{x\to\infty} \frac{h(x)}{x} \to 0$$

Το παραπάνω θεώρημα όχι απλώς εγγυάται την ύπαρξη μιας μονότονης συνάρτησης h για κάθε g αλλά μπορούμε να επιλέξουμε μια συνάρτηση όσο λεπτή θέλουμε εμείς να είναι.

5.13 Παραδείγματα κατανομών με βαριά ουρά

Ακολουθούν παραδείγματα κατανομών με βαριά ουρά και γραφικές παραστάσεις για καλύτερη κατανόηση του θεωρήματος

Κατανομή Wiebull$(\boldsymbol{c},\boldsymbol{\gamma})$

Η κατανομή Wiebull (από τον Σουηδό W.Wiebull) είναι η πλέον διαδεδομένη κατανομή στην ανάλυση αξιοπιστίας. Εμπειρικά αποτελεί ένα ικανοποιητικό μοντέλο για πολλά φαινόμενα.

Η συνάρτηση πυκνότητας πιθανότητας ορίζεται ως εξής

$$f(x) = na^{-n}t^{n-1}\exp\left\{-\left(\frac{t}{a}\right)^n\right\} \quad t > 0$$

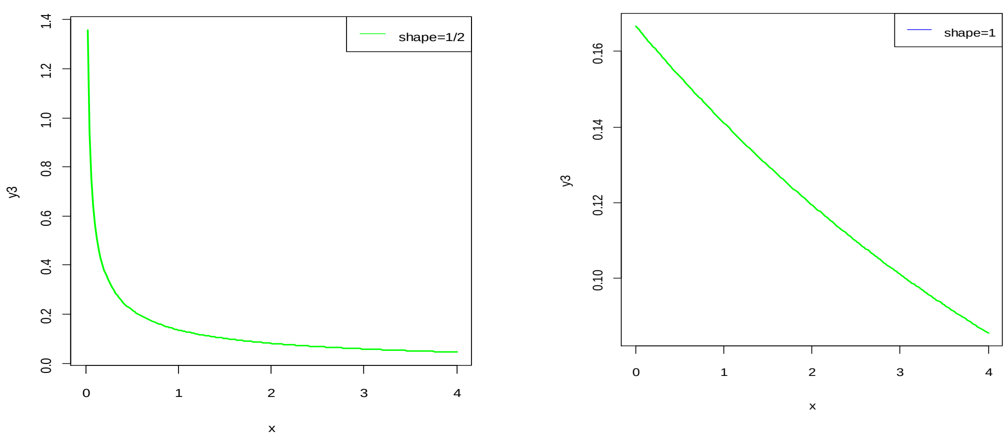
Όπου $\alpha > 0$ είναι η παράμετρος κλίμακας και $n > 0$ είναι η παράμετρος σχήματος.

Η Εκθετική κατανομή αποτελεί ειδική περίπτωση της Wiebull όταν $n = 1$

Η συνάρτηση κατανομής της Wiebull είναι η εξής

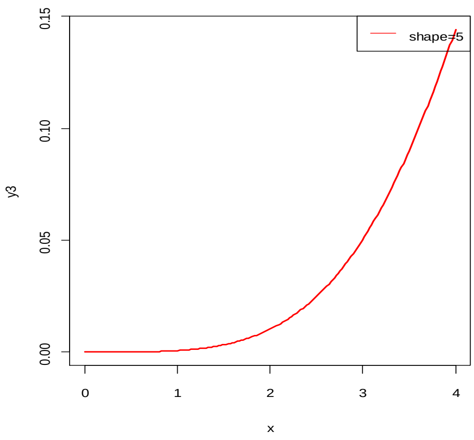
$$F(x) = 1 - \exp\left\{-\left(\frac{t}{a}\right)^n\right\}$$

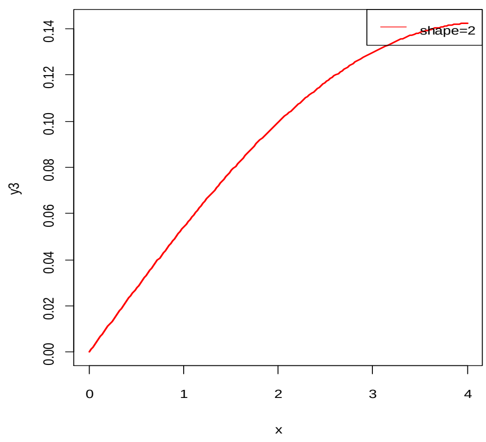
Παρατηρούμε στο διπλανό διάγραμμα πώς αλλάζει η μορφή της συνάρτησης κατανομής όταν η παράμετρος σχήματος είναι μικρότερη της μονάδας (πρώτο γράφημα) όταν η παράμετρος σχήματος είναι ίση με τη μονάδα (δεύτερο γράφημα)



Σχήμα4: (αριστερά) συνάρτηση κατανομής με παράμετρο σχήματος μικρότερη της μονάδας και (δεξιά) συνάρτηση κατανομής με παράμετρο σχήματος μικρότερη της μονάδας

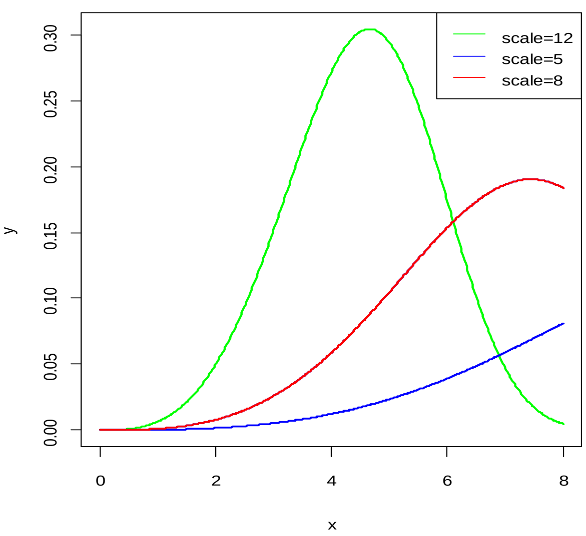
Αντίστοιχα στα παρακάτω γραφήματα βλέπουμε πως συμπεριφέρεται η κατανομή όταν επιλέγουμε παράμετρο σχήματος μεγαλύτερη της μονάδας και συγκεκριμένα παράμετρο σχήματος πέντε (αριστερό γράφημα) και δύο (για το δεξιό γράφημα)





Σχήμα5: (αριστερά) συνάρτηση κατανομής με παράμετρο σχήματος ίση με 5 και (δεξιά) συνάρτηση κατανομής με παράμετρο σχήματος ίση με 2

Μεγάλο ενδιαφέρον παρουσιάζει και η συμπεριφορά της κατανομής όταν αλλάζει η παράμετρος κλίμακας διατηρώντας σταθερή την παράμετρο σχήματος.



Σχήμα6: συνάρτηση κατανομής Wiebull με παράμετρο με διαφορετικά επίπεδα παραμέτρου κλίμακας

Ροπές r τάξης για την κατανομή Wiebull

$$E(T^r) = \int_0^\infty t^r f(t)dt = \int_0^\infty a^r u^{r/n} e^{-u} du \quad όπου\, u = (\frac{t}{a})^n$$

$$= a^r \int_0^\infty u^{r/n} e^{-u} du = a^r \Gamma\left(1 + \frac{r}{n}\right)$$

Όπου $\Gamma(a) = \int_0^{+\infty} t^{a-1} e^{-t}\, dt \quad a > 0\; η\; συνάρτηση\; Γάμμα.$

Από την παραπάνω σχέση μπορούμε να υπολογίσουμε τις ροπές πρώτης και δεύτερης τάξης ως εξής

$$E(T) = \alpha\Gamma\left(1 + \frac{1}{n}\right)$$

$$E(T^2) = a^2\Gamma\left(1 + \frac{2}{n}\right)$$

Μπορούμε επίσης να υπολογίσουμε την διακύμανση

$$Var(T) = E(T^2) - \{E(T)\}^2 =$$

$$= a^2\left[\Gamma\left(1 + \frac{2}{n}\right) - \Gamma\left(1 + \frac{1}{n}\right)^2\right]$$

Η παραπάνω κατανομή έχει βαρύτερη ουρά για παράμετρο κλίμακας $\alpha > 1$.

Δηλαδή για $\alpha > 1$ ισχύει η σχέση

$$\sum_{\kappa=0}^{\infty} \frac{\lambda^\kappa}{\kappa!} E(X^\kappa) = \sum_{\kappa=0}^{\infty} \frac{\lambda^\kappa}{\kappa!} a^\kappa \Gamma\left(1 + \frac{\kappa}{n}\right) = \infty$$

Για να ισχύει η παραπάνω σχέση θα πρέπει να υπάρχει κάποιο $\kappa > 0$ απο το οποίο όλες οι ροπές κ τάξης και άνω να απειρίζονται.

Για διευκόλυνση των πράξεων αλλάζουμε την παραμετροποίηση της κατανομής Wiebull χώρις αυτό να αλλοιώνει τα αποτελέσματα μας

Έτσι ως παράμετρο κλίμακας ορίζεται ως $c = a^{-n}$ και η παράμετρος σχήματος $n = \gamma$.

$$E(X^k) = \int_0^{\infty} x^k f(t)dx = \int_0^{+\infty} e^{\ln x}\, e^{(k-1)\ln x} c x^{\gamma-1} e^{-cx^{\gamma}} dx$$
$$= \int_0^{+\infty} e^{\ln x}\, e^{(k-1)\ln x} c e^{(\gamma-1)\ln x}\, e^{-cx^{\gamma}}\, dx$$

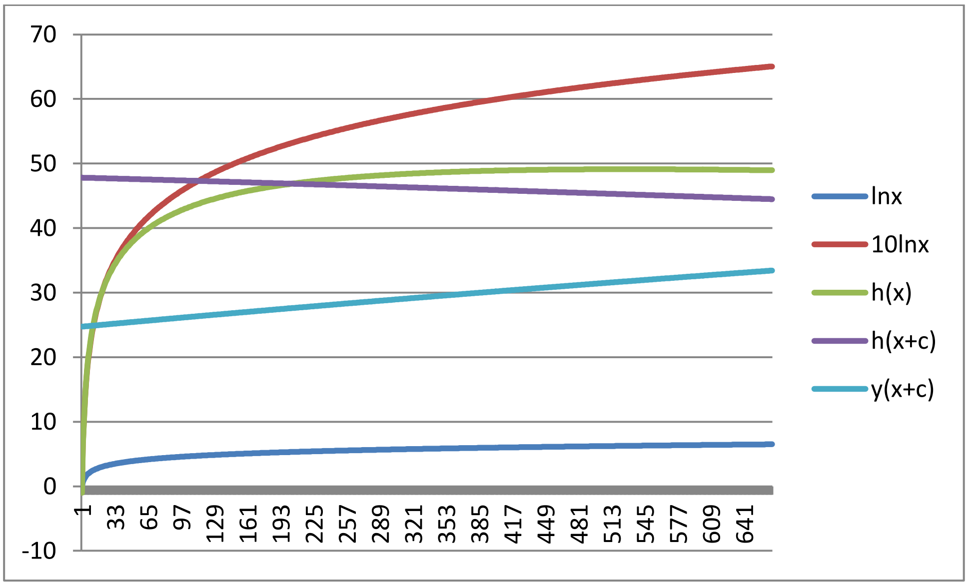
$$= \int_0^{+\infty} c e^{\ln x}\, e^{(k+\gamma-2)\ln x - cx^{\gamma}} dx = c\int_0^{+\infty} e^{\ln x}\, e^{(k+\gamma-2)\ln x - cx^{\gamma}} dx$$

Για τη διευκόλυνση των πράξεων μας στο εξής θα περιγράψουμε μια κατανομή $Wiebull(1, n)$

Έτσι

$$E(X^k) = \int_0^{+\infty} e^{\ln x}\, e^{(k+\gamma-2)\ln x - x^{\gamma}} dx$$

Παρατηρούμε ότι στην παραπάνω σχέση εμφανίζεται η συνάρτηση $g(x) = \ln x$ για την οποία ισχύει $g(x) \to \infty$ $όταν\ x \to \infty$.Θα μπορούσε η συνάρτηση $h(x) = (k + \gamma - 2)\ln x - x^{\gamma}$ να είναι η μονότονη κοίλη συνάρτηση που περιγράφει το θεώρημα.



Στο παραπάνω διάγραμμα φαίνεται γραφικά η συμπεριφορά των συναρτήσεων που έχουμε ήδη αναφέρει. Έχουμε υποθέσει για την γραφική ανάλυση ότι $\kappa = 11\ \kappa\alpha\iota\ \alpha = 0{,}5$

Στο διάγραμμα περιλαμβάνουμε και την συνάρτηση $Y(x) = (x + c)^a + \ln(x + c)$ για αρκετά μεγάλο c.

Παρατηρούμε αυτό ακριβώς που αναφέραμε ως παρατήρηση στο θεώρημα οτι αν υπάρχει η συνάρτηση $g(x)$ τέτοια ώστε $g(x) \to \infty$ $όταν\ x \to \infty$ τότε υπάρχει μια μονότονη κοίλη συνάρτηση h: R+ → R+ τέτοια ώστε $h(x) = o(x)\ όταν\ x \to \infty$, $\mathrm{E}e^{h(\xi)} < \infty$ και $\mathrm{E}e^{h(\xi)+g(\xi)} = \infty$ και αυτή δεν είναι μοναδική.

Ιδιαίτερο ενδιαφέρον παρουσιάζει το παραπάνω θεώρημα διότι δεν ισχύει για κατανομές που δεν έχουν βαρία ουρά.

Μια κατανομή έχει ελαφριά ουρά αν και ισχύει η σχέση

$$\mathrm{E}(e^{\lambda x}) < \infty \text{ για κάποια } \lambda > 0$$

δηλαδή αν ισχύει

$$\int_{-\infty}^{+\infty} e^{\lambda x} dF(x) < \infty\ για\ κάποια\ \lambda > 0$$

Το θεώρημα χρησιμοποιεί μια μονότονη κοίλη συνάρτηση h: R+ → R+ τέτοια ώστε $h(x) = o(x)\ όταν\ x \to \infty$, αυτό σημαίνει πως

$$\frac{h(x)}{x} \to \infty\,,\ \ x \to \infty$$

Με απλά λόγια η παραπάνω σχέση ισχύει όταν

$$h(x) < x$$

και θα ισχύει προφανώς η σχέση

$$h(x) < \lambda x$$

$$\int_{-\infty}^{+\infty} e^{h(x)} f(x)dx < \int_{-\infty}^{+\infty} e^{\lambda x} f(x)dx < \infty\ για\ κάποιο\ \lambda > 0$$

Σύμφωνα με το παραπάνω θεώρημα θα πρέπει επίσης να υσχύει η σχέση

$$\mathrm{E}e^{h(\xi)+g(\xi)} = \infty$$

Όμως

$$\int_{-\infty}^{+\infty} e^{h(x)+g(x)} f(x)dx < \int_{-\infty}^{+\infty} e^{\lambda x+x} f(x)dx = \int_{-\infty}^{+\infty} e^{x} e^{\lambda x} f(x)dx = \int_{-\infty}^{+\infty} e^{\ln x} e^{\lambda x} f(x)dx$$
$$= \int_{-\infty}^{+\infty} e^{(\lambda+1)\chi} f(x)dx < \infty$$

Εφόσον η κατανομή έχει ελαφριά ουρά.Παρατηρούμε λοιπόν πως η ύπαρξη μιας κοίλης μονότονης συνάρτησης h: R+ → R+ τέτοια ώστε $h(x) = o(x)\ όταν\ x \to \infty$, $\mathrm{E}e^{h(\xi)} < \infty$ και $\mathrm{E}e^{h(\xi)+g(\xi)} = \infty$ δεν ισχύει στην περίπτωση που η κατανομή μας έχει ελαφριά ουρά.

Δημιουργείται λοιπόν η σκέψη αν μπορούμε να γενικεύσουμε λίγο ακόμη το παραπάνω θεώρημα.

Έστω ξ μια τυχαία μεταβλητή που ακολουθεί μια κατανομή με βαριά δεξιά ουρά.Έστω επίσης μια αύξουσα συνάρτηση $g(x)$ τέτοια ώστε $g(x) \to \infty \;\; όταν\; x \to \infty$.

Τότε υπάρχει μια μονότονη συνάρτηση h: R → R τέτοια ώστε $h(x) = o(x)$ $όταν\; x \to \infty$, $\mathrm{E}e^{h(\xi)} < \infty$ και $\mathrm{E}e^{h(\xi)+g(\xi)} = \infty$

Το θεώρημα χρησιμοποιεί μια μονότονη συνάρτηση h: R → R τέτοια ώστε $h(x) = o(x)\; όταν\; x \to \infty$, αυτό σημαίνει πως

$$\frac{h(x)}{x} \to \infty\,, \;\; x \to \infty$$

Με απλά λόγια η παραπάνω σχέση ισχύει όταν

$$|h(x)| < |x|$$

και θα ισχύει προφανώς η σχέση

$$h(x) < \lambda x$$

$$\int_0^{+\infty} e^{h(x)} f(x)dx < \int_0^{+\infty} e^{\lambda x} f(x)dx < \infty\; για\; κάποιο\; \lambda > 0$$

Σύμφωνα με το παραπάνω θεώρημα θα πρέπει επίσης να υσχύει η σχέση

$$\mathrm{E}e^{h(\xi)+g(\xi)} = \infty$$

Όμως

$$\int_0^{+\infty} e^{h(x)+g(x)} f(x)dx < \int_0^{+\infty} e^{\lambda x+x} f(x)dx = \int_0^{+\infty} e^{x} e^{\lambda x} f(x)dx =$$
$$= \int_0^{+\infty} e^{(\lambda+1)\chi} f(x)dx = \infty$$

ή θα πρέπει να ισχύει η σχέση

$$h(x) > -\lambda x \Rightarrow$$

$$\int_{-\infty}^{0} e^{h(x)} f(x)dx > \int_{-\infty}^{0} e^{-\lambda x} f(x)dx =$$

$$-\int_{0}^{+\infty} e^{h(x)} f(x)dx > -\int_{0}^{+\infty} e^{-\lambda x} f(x)dx \Rightarrow$$

$$\int_{0}^{+\infty} e^{h(x)} f(x)dx < \int_{0}^{+\infty} e^{-\lambda x} f(x)dx$$

Θα προσπαθήσουμε να βρούμε ένα άνω φράγμα για το ολοκλήρωμα

$$\int_{0}^{+\infty} e^{-\lambda x} f(x)dx$$

Μια συνάρτηση $f(x)$ είναι συνάρτηση πυκνότητας αν παίρνει τιμές μεταξύ του μηδενός και του ένα,δηλαδή

$$0 \le f(x) \le 1 \ \Rightarrow$$

$$0 \le \int_{0}^{+\infty} e^{-\lambda x} f(x)dx \le \int_{0}^{+\infty} e^{-\lambda x} dx = \frac{1}{\lambda} \qquad \lambda > 0$$

Συμπέρασμα το ολοκλήρωμα

$$\int_{0}^{+\infty} e^{h(x)} f(x)dx \begin{cases} \le \dfrac{1}{\lambda} \\ < \displaystyle\int_{0}^{+\infty} e^{\lambda x} f(x)dx \end{cases}$$

Σε οποιαδήποτε και απο τις περιπτπωσεις και αν είμαστε όμως ισχύει

Όμως

$$\int_{0}^{+\infty} e^{h(x)+g(x)} f(x)dx < \int_{0}^{+\infty} e^{\lambda x+x} f(x)dx = \int_{0}^{+\infty} e^{x} e^{\lambda x} f(x)dx =$$

$$= \int_{0}^{+\infty} e^{(\lambda+1)x} f(x)dx = \infty$$

5.14 Κάτω φράγματα για ουρές συνελίξεων

Σε αυτή την ενότητα θα μας απασχολήσει ο υπολογισμός της πιθανότητας

$$\mathrm{P}(X_1 + X_2 + \cdots + X_n > x) \ .$$

Στην ουσία πρόκειται για την νιοστή συνέλιξη

$$\overline{F_1 * F_2 * \ldots * F_n}(x)$$

Η πραγματοποίηση πλήθους διαφορετικών ενδεχομένων μπορεί να οδηγήσει το άθροισμα n τυχαίων μεταβλητών να υπερβαίνει ένα σταθερό αριθμό x.Ένα απο τα πολλά ενδεχόμενα είναι μία μόνο απο τις τυχαίες μεταβλητές $X_i \ i = 1,2,3,\ldots n$ να πάρει τιμή μεγαλύτερη απο το x και οι υπόλοιπες $n-1$ τυχαίες μεταβλητές να πάρουν τιμές

$$\overline{F_1 * F_2 * \ldots * F_n}(x) \geq \sum_{\kappa=1}^{n} P(\{X_k > x\} \cap \{X_i \in [0,x] \gamma\iota\alpha\ \kappa\acute{\alpha}\theta\varepsilon\ i \neq \kappa\})$$

Οι τυχαίες μεταβλητές $X_1, X_2, \ldots, X_n$ είναι ανεξάρτητες μεταξύ τους και έτσι η παραπάνω σχέση μπορεί να γραφεί ως εξής

$$\overline{F_1 * F_2 * \ldots * F_n}(x) \geq \sum_{\kappa=1}^{n} P(\{X_k > x\} \prod_{i \neq k} P(\{X_i \leq x\}) \Rightarrow$$

$$\overline{F_1 * F_2 * \ldots * F_n}(x) \geq \sum_{\kappa=1}^{n} \bar{F}_k\,(x) \prod_{i \neq k} F_i\,(x)$$

Όπου

$F_i(x)$ είναι η συνάρτηση κατανομής της τυχαίας μεταβλητής $X_i \ i = 1,2,3,\ldots n$ έτσι μπορούμε να πούμε πως

$$\lim_{x\to\infty} F_i(x) = F(\mathrm{R+}) = 1+$$

και

$$\overline{F_1 * F_2 * \ldots * F_n}(x) \sim \sum_{\kappa=1}^{n} \bar{F}_k\,(x) \Rightarrow$$

$$\lim_{x\to\infty} \frac{\overline{F_1 * F_2 * \ldots * F_n}(x)}{\sum_{\kappa=1}^{n} \bar{F}_k\,(x)} = 1$$

Μια δεύτερη προσέγγιση

$$\overline{F_1 * F_2 * \ldots * F_n}(x) \geq P(\{\max\{X_1, X_2, \ldots, X_n\} > x\}) \Rightarrow$$

$$\overline{F_1 * F_2 * \ldots * F_n}(x) \geq 1 - P(\{X_1 \leq x, X_2 \leq x, \ldots, X_n \leq x) \Rightarrow$$

$$\overline{F_1 * F_2 * \dots * F_n}(x) \geq 1 - P(\{X_1 \leq x)P(X_2 \leq x), \dots P(X_n \leq x)$$

Διότι $X_1, X_2, \dots, X_n$ είναι ανεξάρτητες

$$\overline{F_1 * F_2 * \dots * F_n}(x) \geq 1 - F^n(x) = \big(1 - F(x)\big)(1 + F(x) + F^2(x) + \cdots + F^{n-1}(x)) \Rightarrow$$

$$\overline{F_1 * F_2 * \dots * F_n}(x) \geq \bar{F}(x)(1 + \sum_{i=1}^{n-1} F^i(x))$$

Όμως

$$\lim_{x\to\infty} \sum_{i=1}^{n-1} F^i(x) = \sum_{i=1}^{n-1} \lim_{x\to\infty} F^i(x) = n - 1$$

Και έτσι η παραπάνω σχέση μας δίνει για $x \to \infty$

$$\overline{F_1 * F_2 * \dots * F_n}(x) \geq n\bar{F}(x)$$

$$\frac{\overline{F_1 * F_2 * \dots * F_n}(x)}{\bar{F}(x)} \geq n$$

Μια τρίτη προσέγγιση

Το ενδεχόμενο

$$\{X_1 + X_2 + \cdots + X_n > x\} \supseteq \{X_1 > x, X_2 > x, \dots X_n > x\}$$

Και η πιθανότητα

$$P(\{X_1 + X_2 + \cdots + X_n > x\}) \geq P(\{X_1 > x, X_2 > x, \dots X_n > x\}) \Rightarrow$$

$$\overline{F_1 * F_2 * \dots * F_n}(x)) \geq P(\{X_1 > x\})P(\{X_2 > x\}) \dots P(\{X_n > x\}) \Rightarrow$$

$$\overline{F_1 * F_2 * \dots * F_n}(x)) \geq \prod_{i=1}^{n} \bar{F}\,(x)$$

5.15 Υπολογισμός αθροίσματος τριών τυχαίων μεταβλητών

Έστω τρείς τυχαίες μεταβλητές ορισμένες στον θετικό ημιάξονα.Θα προσπαθήσουμε να υπολογίσουμε την συνέλιξη

$$\overline{F_1 * F_2 * F_3}(x) = \mathrm{P}(X_1 + X_2 + X_3 > x)$$

Τα πιθανά ενδεχόμενα είναι τα εξής

$$\big\{\{X_1 > x, X_2 < x, X_3 < x\} \cup \{X_2 > x, X_1 < x, X_3 < x\} \cup \{X_3 > x, X_1 < x, X_2 < x\}\big\} = A$$

Η πιθανότητα του ενδεχομένου A είναι ίση με

$$P(A) = \sum_{\kappa=1}^{3} \bar{F}_k(x) \prod_{\kappa\neq j} F_j\,(x)$$

$$\big\{\{X_1 > x, X_2 > x, X_3 < x\} \cup \{X_2 > x, X_1 < x, X_3 > x\} \cup \{X_3 > x, X_1 > x, X_2 < x\}\big\} = B$$

Η πιθανότητα του ενδεχομένου B είναι ίση με

$$P(B) = \sum_{\kappa=1}^{3} F_k(x) \prod_{\kappa\neq j} \bar{F}_j(x)$$

$$\{X_1 > x, X_2 > x, X_3 > x\} = \Gamma$$

Η πιθανότητα του ενδεχομένου Γ είναι ίση με

$$P(\Gamma) = \prod_{j=1}^{3} \bar{F}_j\,(x)$$

$$\big\{\{X_1 < x, X_2 < x, X_3 < x\} \cap \{X_1 + X_2 + X_3 > x\}\big\} = \Delta$$

Η πιθανότητα του ενδεχομένου Δ είναι ίση με

$$P(\Delta) = \sum_{\substack{0\leq\,\kappa<\chi \\ 0<\lambda<\chi}} P(X_1 = \kappa, X_2 = \lambda, x - \kappa - \lambda < X_3 < x) = \alpha$$

$$P(\Delta) = \sum_{\substack{0\leq\,\kappa<\chi \\ 0<\lambda<\chi}} f_1(\kappa) f_2(\lambda)\{F_3(x) - F_3(x - \kappa - \lambda)\}$$

Έτσι ο ακριβής υπολογισμός της

$$\overline{F_1 * F_2 * F_3}(x) = \mathrm{P}(X_1 + X_2 + X_3 > x) =$$

$$= \sum_{\kappa=1}^{3} \bar{F}_k(x) \prod_{\kappa\neq j} F_j\,(x) + \sum_{\kappa=1}^{3} F_k(x) \prod_{\kappa\neq j} \bar{F}_j(x) + \prod_{j=1}^{3} \bar{F}_j\,(x) + \alpha$$

Και όταν το $x \to \infty$ η παραπάνω σχέση μας δίνει

$$\overline{F_1 * F_2 * F_3}(x) \sim \sum_{\kappa=1}^{3} \bar{F}_k(x)$$

Κεφάλαιο 6

Υποεκθετικές κατανομές

6.1 εισαγωγή

Οι υποεκθετικές κατανομές είναι μια ειδική κατηγορία κατανομών με δεξιά βαριά ουρά. Το όνομα τους προκύπτει από το γεγονός ότι η ουρά μιας τέτοιας κατανομής συγκλίνει στο μηδέν πιο αργά από οποιαδήποτε εκθετική κατανομή. Είναι λοιπόν λογικό σύμφωνα με όσα περιγράφουμε παραπάνω να ισχύει

$$\lim_{x \to \infty} sup e^{\lambda x} \overline{F}(x) = \infty \; για \; όλα \; τα \; \lambda > 0$$

Μιας και η ουρά $\overline{F}(x)$ συγκλίνει στο μηδέν πιο αργά απο την ουρά μιας οποιαδήποτε εκθετικής κατανομής .

Αυτή η διαπίστωση προκύπτει από το γεγονός ότι η παραπάνω σχέση μπορεί να γραφτεί ως εξής

$$\lim_{x \to \infty} sup e^{\lambda x} \overline{F}(x) = \lim_{x \to \infty} sup \frac{\overline{F}(x)}{e^{-\lambda x}} = \infty$$

Αναλόγως προκύπτει πως

$$\lim_{x \to \infty} sup \frac{e^{-\lambda x}}{\overline{F}(x)} = 0$$

Αυτό πρακτικά σημαίνει πως αν πάρουμε ένα δείγμα τιμών από μια υποεκθετική κατανομή, έχουμε πολύ μεγάλες τιμές στο δείγμα με μη αμελητέα πιθανότητα.

Για το λόγο αυτό οι κατανομές αυτές χρησιμοποιούνται στην αναλογιστική επιστήμη ως μοντέλα για την κατανομή των αποζημιώσεων σε ειδικές κατηγορίες ασφαλίσεων όπου οι αποζημιώσεις παίρνουν πολύ μεγάλες τιμές με όχι πολύ μικρή πιθανότητα.

Τέτοιες κατηγορίες ασφαλίσεων είναι οι ασφαλίσεις πυρός ,οι ασφαλίσεις από φυσικές καταστροφές κτλ.

Όπως έχουμε ήδη αποδείξει για οποιαδήποτε κατανομή F ορισμένη στον θετικό ημιάξονα $\mathbb{R}^+$ ισχύει

$$\overline{F * F}(x) = P(X_1 + X_2 > x) \geq P(X_1 > x \text{ ή } X_2 > x) \sim 2P(X_1 > x) \quad x \to \infty$$

$$\lim_{x \to \infty} inf \frac{\overline{F * F}(x)}{P(X_1 > x)} \geq 2$$

Αν ισχύει ταυτόχρονα

$$\lim_{x \to \infty} inf \frac{\overline{F * F}(x)}{P(X_1 > x)} \leq 2$$

Τότε συμπεραίνουμε πως

$$\lim_{x\to\infty} inf \frac{\overline{F*F}(x)}{P(X_1 > x\,)} = 2$$

Και τότε λέμε πως η κατανομή F ορισμένη στον θετικό ημιάξονα $\mathbb{R}^+$ είναι κατανομή με βαριά ουρά.

Γενικότερα αν μια κατανομή είναι κατανομή με βαριά ουρά και το όριο του λόγου

$$\frac{\overline{F*F}(x)}{\bar{F}(x)} \to c \qquad x \to \infty$$

Με $c \in (0, +\infty]$ τότε υποχρεωτικά το $c = 2$.

6.2 Ορισμός υποεκθετικής

Ορισμός(1)

Μια κατανομή F ορισμένη στον θετικό ημιάξονα $\mathbb{R}^+$ με την ιδιότητα $\bar{F}(x) > 0$ για όλα τα x θα λέμε ότι ανήκει στην οικογένεια των υποεκθετικών κατανομών και θα γράφουμε $F \in S$ αν

$$\overline{F*F}(x) \sim 2\bar{F}(x) \quad για\ x \to \infty$$

Η παραπάνω σχέση θα μπορούσε να γραφεί διαφορετικά ως εξής

$$P(X_1 + X_2 > x) \sim 2P(X_1 > x\,) \quad x \to \infty$$

Αν τώρα στην ανάλυση μας εισάγουμε την έννοια της δεσμευμένης πιθανότητας τότε παρατηρούμε το εξής

$$P(X_1 > x / X_1 + X_2 > x) = \frac{P(X_1 > x, X_1 + X_2 > x)}{P(X_1 + X_2 > x)}$$

Όμως έχουμε εξ αρχής τονίσει πως οι τυχαίες μεταβλητές X_1, X_2 ανήκουν στο $\mathbb{R}^+$ και έτσι η πιθανότητα

$$P(X_1 > x, X_1 + X_2 > x) = P(X_1 > x\,)$$

Έτσι

$$P(X_1 > x / X_1 + X_2 > x) \ = \frac{P(X_1 > x)}{P(X_1 + X_2 > x)}$$

Σύμφωνα λοιπόν με όσα αναφέραμε παραπάνω μπορούμε να γράψουμε την σχέση που δηλώνει υποεκθετικότητα

$$P(X_1 + X_2 > x) \sim 2P(X_1 > x\,) \quad x \to \infty$$

Ως εξής

$$\frac{P(X_1 + X_2 > x)}{P(X_1 > x\,)} = \frac{1}{P(X_1 > x / X_1 + X_2 > x)} \sim 2 \quad x \to \infty$$

Από την άλλη πλευρά η πιθανότητα

$$P(X_1 + X_2 > x) \geq P(max\{X_1, X_2\} > x)$$

και

$$P(max\{X_1, X_2\} > x) = 1 - P(max\{X_1, X_2\} \leq x)$$

Όμως αν το μέγιστο από τις τυχαίες μεταβλητές X_1, X_2 είναι μικρότερο ή ίσο με μια θετική σταθερά x τότε και το X_1και το X_2 θα είναι μικρότερα ή ίσα του x.

Σύμφωνα λοιπόν με τα παραπάνω η πιθανότητα

$$P(X_1 + X_2 > x) \geq 1 - P(max\{X_1, X_2\} \leq x) \geq 1 - P(X_1 \leq x, X_2 \leq x)$$

Λόγο ανεξαρτησίας η από κοινού συνάρτηση κατανομής ισούται

$$P(X_1 \leq x, X_2 \leq x) = P(X_1 \leq x)P(X_2 \leq x)$$

Και λόγο ισονομίας

$$P(X_1 \leq x, X_2 \leq x) = P(X_1 \leq x)^2 = F^2(x)$$

Έτσι η πιθανότητα

$$P(X_1 + X_2 > x) \geq 1 - P(X_1 \leq x, X_2 \leq x) = 1 - F^2(x) =$$

$$= \big(1 - F(x)\big)\big(1 + F(x)\big) = \bar{F}(x)(1 + F(x))$$

Όταν όμως το $x \to \infty$

$$P(X_1 + X_2 > x) \sim 2\bar{F}(x)$$

Καταλήξαμε λοιπόν στο συμπέρασμα πως αν η κατανομή που μελετάμε ανήκει στην οικογένεια των υποεκθετικών κατανομών τότε

$$P(X_1 + X_2 > x) \sim P(max\{X_1, X_2\} > x) \sim 2\bar{F}(x) \quad x \to \infty$$

Μια δεύτερη εξίσου σημαντική διαπίστωση που ισχύει όταν η κατανομή ανήκει στην οικογένεια των υποεκθετικών κατανομών και εφόσον η ανάλυση μας γίνεται στο πλαίσιο του θετικού ημιάξονα δηλαδή στον $\mathbb{R}^+$ είναι πως η πιθανότητα

$$P(X_1 + X_2 > x, max\{X_1, X_2\} \leq x) \sim P(max\{X_1, X_2\} > x, max\{X_1, X_2\} \leq x) = 0$$

Διότι δεν μπορεί ταυτόχρονα να ισχύουν τα παραπάνω δύο ενδεχόμενα.

Αντίστοιχα μιας και ισχύει η ιδιότητα πως η ουρά της κατανομής είναι πάντα θετική για κάθε τιμή της τυχαίας μεταβλητής μας η παραπάνω σχέση γράφεται

$$\frac{P(X_1 + X_2 > x, max\{X_1, X_2\} \leq x)}{\bar{F}(x)} = 0 \quad x \to \infty$$

ή ισοδύναμα

$$P(X_1 + X_2 > x, max\{X_1, X_2\} \leq x) = o\big(\bar{F}(x)\big) \quad x \to \infty$$

Θα είχε ενδιαφέρον να δούμε λίγο αναλυτικότερα την παραπάνω δεσμευμένη πιθανότητα .

Από την μια πλευρά

$$P(X_1 + X_2 > x, max(X_1, X_2) \leq x) = P(max(X_1, X_2) \leq x / X_1 + X_2 > x) P(X_1 + X_2 > x)$$

$$= P(max(X_1, X_2) \leq x / X_1 + X_2 > x) \bar{F}^{*2}(x)$$

Από την άλλη πλευρά

$$P(X_1 + X_2 > x, max(X_1, X_2) \leq x) =$$

$$= P(X_1 + X_2 > x / max(X_1, X_2) \leq x)\, P(\,max(X_1, X_2) \leq x) =$$

$$= P\,(X_1 + X_2 > x / max(X_1, X_2) \leq x)\, P(\,X_1 \leq x) P(X_1 \leq x)^2$$

$$= F(x)^2 P((X_1 + X_2 > x / max(X_1, X_2) \leq x)$$

Θα πρέπει οι παραπάνω σχέσεις να είναι ίσες αφού τα πρώτα τους μέλη είναι ίσα θα πρέπει και τα δεύτερα μέλη τους να είναι ίσα.

$$P(max(X_1, X_2) \leq x / X_1 + X_2 > x) \bar{F}^{*2}(x) = F(x)^2 P((X_1 + X_2 > x / max(X_1, X_2) \leq x)$$

Η παραπάνω σχέση μπορεί ισοδύναμα να γραφτεί ως εξής

$$\frac{P((X_1 + X_2 > x / max(X_1, X_2) \leq x)}{P(max(X_1, X_2) \leq x / X_1 + X_2 > x)} = \frac{\bar{F}^{*2}(x)}{F(x)^2}$$

Όμως

$$P(max(X_1, X_2) \leq x / X_1 + X_2 > x) =$$

$$P(X_1 \leq x, X_2 \leq x / X_1 + X_2 > x) = P(X_1 \leq x / X_1 + X_2 > x)^2$$

καταλήγουμε λοιπόν στο συμπέρασμα πως

$$\frac{P((X_1 + X_2 > x / max(X_1, X_2) \leq x)}{P(max(X_1, X_2) \leq x / X_1 + X_2 > x)} = \frac{P((X_1 + X_2 > x / max(X_1, X_2) \leq x)}{P(X_1 \leq x / X_1 + X_2 > x)^2} = \frac{\bar{F}^{*2}(x)}{F(x)^2}$$

Έτσι στην περίπτωση που η κατανομή είναι υποεκθετική η συνέλιξη της ουράς ισούται με

$$\bar{F}^{*2}(x) = F(x)^2 \frac{P((X_1 + X_2 > x / max(X_1, X_2) \leq x)}{P(X_1 \leq x / X_1 + X_2 > x)^2}$$

Ορισμός 2

Μια κατανομή F ορισμένη στον R+ Θα λέμε ότι είναι υποεκθετική εάν

$$\lim_{x \to \infty} \frac{\overline{F^{*2}}(x)}{\bar{F}(x)} = 2$$

Ξέρουμε πως για να ισχύει η ισότητα στο παραπάνω όριο θα πρέπει να ισχύουν ταυτόχρονα οι σχέσεις

$$\lim_{x \to \infty} \frac{\overline{F^{*2}}(x)}{\bar{F}(x)} \geq 2 \; και \; \lim_{x \to \infty} \frac{\overline{F^{*2}}(x)}{\bar{F}(x)} \leq 2$$

$$\overline{F^{*2}}(x) \geq P(\{\max\{X_1, X_2\} > x\})$$

$$\overline{F^{*2}}(x) \geq 1 - P(\{\max\{X_1, X_2\} \leq x\}) \Rightarrow$$

$$\overline{F^{*2}}(x) \geq 1 - P(\{X_1 \leq x) P(X_2 \leq x) \Rightarrow$$

$$\overline{F^{*2}}(x) \geq 1 - F^2(x) = \big(1 - F(x)\big)\big(1 + F(x)\big) \Rightarrow$$

$$\frac{\overline{F^{*2}}(x)}{1 - F(x)} \geq 1 + F(x) \Rightarrow$$

$$\lim_{x \to \infty} \frac{\overline{F^{*2}}(x)}{1 - F(x)} \geq \lim_{x \to \infty} (1 + F(x))$$

Όμως

$$\lim_{x \to \infty} (F(x)) = 1$$

Και έτσι

$$\lim_{x\to\infty}\frac{\overline{F^{*2}}(x)}{1-F(x)} \geq 2$$

Θα πρέπει επίσης να ισχύει η σχέση

$$\lim_{x\to\infty}\frac{\overline{F^{*2}}(x)}{\bar{F}(x)} \leq 2$$

Όρισμος 2

Θα μπορούσαμε λοιπόν να πούμε πως υποεκθετική είναι μια κατανομή F για την οποία ισχύει η σχέση

$$\lim_{x\to\infty}\frac{\overline{F^{*2}}(x)}{\bar{F}(x)} \leq 2$$

Απο τον ορισμό της υποεκθετικής κατανομής παρατηρούμε ότι ο παρονομαστής της

$$\frac{\overline{F^{*2}}(x)}{\bar{F}(x)}$$

Είναι διάφορος του μηδενός ,και επειδή $\bar{F}(x)$ είναι η δεξιά ουρά κατανομής η $\bar{F}(x) > 0$ για κάθε $x \in$ R+. Δηλαδή η κατανομή φθίνει πολύ αργά στο μηδέν όταν το $x \to \infty$.

Έτσι η σχέση

$$\lim_{x\to\infty} sup e^{\lambda x}\,\bar{F}(x) = \infty \; για \; όλα \; τα \; \lambda > 0$$

Ισχύει και η κατανομή είναι μια κατανομή με βαριά ουρά.Η κλάση λοιπόν των υποεκθετικών κατανομών είναι υποσύνολο της κλάσης των κατανομών με βαριά ουρά.

Απο την άλλη υπάρχουν κατανομές με την ιδιότητα $\bar{F}(x) > 0$ αλλά δεν είναι υποεκθετικές κατανομές.

6.3 Παράδειγμα κατανομής που δεν ανήκει στην οικογένεια των υποεκθετικών κατανομών

Ένα παράδειγμα γνωστής κατανομής που δεν είναι υποεκθετική είναι η εκθετική κατανομή exp (λ) η οποία έχει συνάρτηση πυκνότητας πιθανότητας

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & για \; x \geq 0 \\ 0 & για \; x < 0 \end{cases}$$

Όπου $\lambda > 0$ με μέση τιμή $E(X) = \frac{1}{\lambda}$ και διακύμανση $Var(X) = \frac{1}{\lambda^2}$

και συνάρτηση κατανομής

$$F(x) = \begin{cases} 1 - e^{-\lambda x} \; για \; x \geq 0 \\ 0 \qquad\quad\; για \; x < 0 \end{cases}$$

Για την εκθετική κατανομή η πιθανότητα

$$P(\{X_1 + X_2 > x\}) = \overline{F^{*2}}(x) = 1 - F^{*2}(x)$$

Όμως

$$F^{*2}(x) = \int_{-\infty}^{+\infty} F(x-y)dF(y) = \int_0^x F(x-y)dF(y)$$

$$\int_0^x \left\{1 - e^{-\lambda(x-y)}\right\} \lambda e^{-\lambda y} dy = \lambda \int_0^x e^{-\lambda y} dy - \int_0^x \lambda e^{-\lambda y} \, e^{-\lambda(x-y)} dy$$

$$F^{*2}(x) = -e^{-\lambda x} + 1 - \lambda x e^{-\lambda x}$$

Έτσι η συνέλιξη

$$\overline{F^{*2}}(x) = 1 - F^{*2}(x) = 1 - \left\{-e^{-\lambda x} + 1 - \lambda x e^{-\lambda x}\right\}$$

$$\overline{F^{*2}}(x) = e^{-\lambda x}(1 + \lambda x)$$

Και το όριο

$$\lim_{x \to \infty} \frac{\overline{F^{*2}}(x)}{\bar{F}(x)} = \lim_{x \to \infty} \frac{e^{-\lambda x}(1 + \lambda x)}{e^{-\lambda x}} = \lim_{x \to \infty} (1 + \lambda x) = \infty$$

6.4 Παράδειγμα κατανομής που ανήκει στην οικογένεια των υποεκθετικών

Έστω X_1, X_2 ανεξάρτητες και ισόνομες (όμοια κατανεμημένες) τυχαίες μεταβλητές που ακολουθούν την κατανομή pareto με δεξιά ουρά κατανομής

$$\overline{F}(x) = \left(\frac{b}{b+x}\right)^a \cong x^{-a} \; για \; x \to \infty$$

Το ενδεχόμενο $\{X_1 + X_2 > (1+\varepsilon)x\} \supseteq \{X_1 + X_2 > x\}$ για $\varepsilon \in (0,1)$

Έτσι

$$\{X_1 > (1-\varepsilon)x\} \cup \{X_2 > (1-\varepsilon)x\} \cup \{X_1 > \varepsilon x, X_2 > \varepsilon x\}$$

Και οι αντίστοιχες πιθανότητες των παραπάνω ενδεχομένων δίνουν

$$\mathrm{P}\{X_1 + X_2 > x\} \leq \mathrm{P}(X_1 > (1-\varepsilon)x) + \mathrm{P}(X_2 > (1-\varepsilon)x) + \mathrm{P}(X_1 > \varepsilon x, X_2 > \varepsilon x)$$

$$\mathrm{P}\{X_1 + X_2 > x\} \leq 2\ \mathrm{P}(X > (1-\varepsilon)x) + \mathrm{P}(X > \varepsilon x)^2$$

$$\overline{F^{*2}}(x) \leq 2\{(1-\varepsilon)x\}^{-a} + (\varepsilon x)^{-2a}$$

$$\frac{\overline{F^{*2}}(x)}{\bar{F}(x)} \leq \frac{2\{(1-\varepsilon)x\}^{-a} + (\varepsilon x)^{-2a}}{x^{-a}}$$

$$\frac{\overline{F^{*2}}(x)}{\bar{F}(x)} \leq 2\{(1-\varepsilon)\}^{-a} + x^{-a}\varepsilon^{-2\alpha}$$

Και το όριο της παραπάνω σχέσης μας δίνει

$$\lim_{x\to\infty} \frac{\overline{F^{*2}}(x)}{\bar{F}(x)} \leq \lim_{x\to\infty} \{2\{(1-\varepsilon)\}^{-a} + x^{-a}\varepsilon^{-2\alpha}\}$$

$$\lim_{x\to\infty} \frac{\overline{F^{*2}}(x)}{\bar{F}(x)} \leq 2\{(1-\varepsilon)\}^{-a}$$

Και επειδή το ε είναι αυθαίρετο ισχύει

$$\lim_{x\to\infty} \frac{\overline{F^{*2}}(x)}{\bar{F}(x)} \leq 2$$

Αποδείξαμε δηλαδή ότι η κατανομή pareto ανήκει στην οικογένεια των υποεκθετικών κατανομών.

Θα μπορούσαμε να γενικεύσουμε το παραπάνω συμπέρασμα για μια οικογενεια κατανομών και όχι απλώς για μια συγκεκριμένη κατανομή,αυτό είναι ένα φυσιολογικό ερώτημα που γεννιέται.

6.5 Κατανομή τύπου Pareto και σχέση με υποεκθετικές κατανομές

Μια κατα*νομή F* θα λέμε ότι είναι τύπου Pareto με παράμετρο $\alpha > 0$ αν

$$\overline{F}(x) \sim L(x)x^{-a} \quad x \to \infty$$

Με τη συνάρτηση $L(x)$ να είναι μια θετική συνάρτηση L:R+→ $(0, \infty)$ που ικανοποιεί την σχέση

$$\frac{L(xy)}{L(x)} \to 1 \qquad y > 0 \qquad x \to \infty$$

Το ενδεχόμενο $\{X_1 + X_2 > (1+\varepsilon)x\} \supseteq \{X_1 + X_2 > x\}$ για $\varepsilon \in (0,1)$

Έτσι

$$\{X_1 > (1-\varepsilon)x\} \cup \{X_2 > (1-\varepsilon)x\} \cup \{X_1 > \varepsilon x, X_2 > \varepsilon x\}$$

Και οι αντίστοιχες πιθανότητες των παραπάνω ενδεχομένων δίνουν

$$\mathrm{P}\{X_1 + X_2 > x\} \le \mathrm{P}(X_1 > (1-\varepsilon)x) + \mathrm{P}(X_2 > (1-\varepsilon)x) + \mathrm{P}(X_1 > \varepsilon x, X_2 > \varepsilon x)$$

$$\mathrm{P}\{X_1 + X_2 > x\} \le 2\ \mathrm{P}(X > (1-\varepsilon)x) + \mathrm{P}(X > \varepsilon x)^2$$

$$\overline{F^{*2}}(x) \le 2L((1-\varepsilon)x)\{(1-\varepsilon)x\}^{-a} + L(\varepsilon x)(\varepsilon x))^{-2a}$$

$$\frac{\overline{F^{*2}}(x)}{\bar{F}(x)} \le \frac{2L((1-\varepsilon)x)\{(1-\varepsilon)x\}^{-a} + L(\varepsilon x)(\varepsilon x))^{-2a}}{L(x)x^{-a}}$$

$$\frac{\overline{F^{*2}}(x)}{\bar{F}(x)} \le \frac{2L((1-\varepsilon)x)\{(1-\varepsilon)x\}^{-a}}{L(x)x^{-a}} + \frac{L(\varepsilon x)(\varepsilon x))^{-2a}}{L(x)x^{-a}}$$

Όμως εξ ορισμού της κατανομής τύπου pareto ισχύει

$$\frac{L((1-\varepsilon)x)}{L(x)} \to 1 \qquad \frac{L(\varepsilon x)}{L(x)} \to 1 \qquad x \to \infty$$

$$\lim_{x\to\infty} \frac{\overline{F^{*2}}(x)}{\bar{F}(x)} \le \lim_{x\to\infty} \{2\{(1-\varepsilon)\}^{-a} + x^{-a}\varepsilon^{-2\alpha}\}$$

$$\lim_{x\to\infty} \frac{\overline{F^{*2}}(x)}{\bar{F}(x)} \le 2\{(1-\varepsilon)\}^{-a}$$

Και επειδή το ε είναι αυθαίρετο ισχύει

$$\lim_{x\to\infty} \frac{\overline{F^{*2}}(x)}{\bar{F}(x)} \le 2$$

Θα μπορούσαμε να γράψουμε την σχέση

$$\bar{F}(x) \sim L(x)x^{-a} \quad x \to \infty$$

ως εξής

$$\overline{F}(x) \sim L(x) e^{-a \ln x} \quad x \to \infty$$

και λύνοντας ως προς $L(x)$

$$L(x) = \frac{\overline{F}(x)}{e^{-a \ln x}} \quad x \to \infty$$

Αντίστοιχα

$$L(xy) = \frac{\overline{F}(xy)}{e^{-a \ln xy}} \quad x \to \infty$$

όμως

$$\frac{L(xy)}{L(x)} \to 1 \qquad y > 0 \qquad x \to \infty$$

Άρα

$$\frac{\frac{\overline{F}(xy)}{e^{-a \ln xy}}}{\frac{\overline{F}(x)}{e^{-a \ln x}}} \to 1 \qquad y > 0 \qquad x \to \infty$$

Και τελικά προκύπτει

$$\frac{\overline{F}(xy)}{\overline{F}(x)} \sim e^{-a \ln y} \; y > 0$$

Δηλαδή όταν μια κατανομή ανήκει στην οικογένεια κατανομών τύπου Pareto τότε ικανοποιεί τη σχέση

$$\frac{\overline{F}(xy)}{\overline{F}(x)} \sim e^{-a \ln y} \; y > 0$$

Θεώρημα

Αν μια κατανομή είναι κατανομή τύπου Pareto τότε ανήκει στην κλάση των υποεκθετικών κατανομών

Λήμμα

Μια συνάρτηση $\bar{F}$ (και κατα αντιστοιχία F) θα λέμε ότι ανήκει στην κλάση των κατανομών με μακριά ουρά εάν και μόνο εάν η συνάρτηση $\bar{G}(x) = \bar{F}(\log x)$ συγκλίνει πολύ άργα στο μηδέν όταν το $x \to \infty$. Η συνάρτηση $\bar{G}(x)$ είναι ορισμένη για θετικά x.

Τα παραπάνω με τη βοήθεια των μαθηματικών αποτυπώνονται ως εξής

$$\frac{\bar{G}(ax)}{\bar{G}(x)} \to 1 \quad x \to \infty$$

Ισοδύναμη σχέση της παραπάνω που προκύπτει με αντικατάσταση απο τη σχέση

$$\bar{G}(x) = \bar{F}(\log x)$$

είναι η

$$\frac{\bar{F}(\log \alpha x)}{\bar{F}(\log x)} = \frac{\bar{F}(\log x + \log \alpha)}{\bar{F}(\log x)} \to 1 \quad \alpha > 0$$

Παρατηρούμε πως αν παραπάνω αντικαταστήσουμε όπου $x = e^x \; και \; a = e^y$ πρόκύπτει η γνωστή ιδιότητα των κατανομών με μακριά ουρά

$$\frac{\bar{F}(x+y)}{\bar{F}(x)} \to 1 \quad x \to \infty \;\; y \geq 0$$

Σύμφωνα λοιπόν με τα παραπάνω για $\alpha > 0$ έχουμε

$$x^a \bar{F}(\log x) \to \infty \quad x \to \infty$$

Αντικαθιστώντας όπου $x = e^x$ έχουμε

$$\frac{1}{e^{xa}\bar{F}(x)} \to 0 \quad x \to \infty$$

$$\lim_{x \to \infty} \frac{e^{-xa}}{\bar{F}(x)} = 0$$

Στην παραπάνω σχέση οφείλει και το όνομα της η οικογένεια των υποεκθετικών κατανομών και αυτό διότι αν παρατηρήσουμε ο αριθμήτης της παραπάνω σχέσης είναι η δεξιά ουρά μιας εκθετικής κατανομής με παράμετρο $\alpha > 0$ και ο παρονομαστής η δεξιά ουρά μιας κατανομής για την οποία ισχύει $\bar{F}(x) > 0$ για κάθε x.Η παραπάνω σχέση μας περιγράφει πως οποιαδήποτε κατανομή F με την παραπάνω ιδιότητα συγκλίνει πιο αργά στο μηδέν όταν το $x \to \infty$ σε σχέση με την εκθετική κατανομή, εξ ου λοιπόν και ο όρος υποεκθετική κατανομή.

6.6 Παραδείγματα κατανομών τύπου Pareto

- Η κατανομή $Par(a)$ με συνάρτηση πυκνότητας πιθανότητας

$$f(x) = a(1+x)^{-(a+1)}$$

- Η κατανομή $loggamma$ με συνάρτηση πυκνότητας πιθανότητας

$$\frac{\lambda^{\alpha}}{\Gamma(\alpha)}(logx)^{a-1}x^{-\lambda-1} \qquad x > 1$$

- Κατανομή $Burr$ με δεξιά ουρά

$$\overline{F}(x) = \left(\frac{b}{b+x^r}\right)^a \qquad x \geq 0$$

Δεν είναι πάντα εύκολο να υπολογίσουμε την συνέλιξη $\overline{F^{*2}}(x)$ διότι το ολοκλήρωμα

$$(F*G)(x) = \int_0^{+\infty} F(x-y)d(G(y) = \int_0^{+\infty} F(x-y)g(y)dy$$

Μπορεί να είναι πολύ δύσκολο να λυθεί. Χαρακτηριστική περίπτωση είναι η κατανομή $Burr(a,b,r)$ με συνάρτηση κατανομής

$$F(x) = 1 - \left(\frac{b}{b+x^r}\right)^a \qquad x \geq 0$$

Η συνέλιξη δεύτερης τάξης δίνεται απο τη σχέση

$$F^{*2}(x) = \int_{-\infty}^{+\infty} F(x-y)dF(y) = \int_0^x F(x-y)dF(y)$$

$$= \int_0^{+\infty} \{1 - b^a(b+(x-y)^r)^{-a}\} rab^a(b+y^r)^{-a-1}dy =$$

$$= \int_0^{+\infty} rab^a(b+y^r)^{-a-1}dy - \int_0^{+\infty} b^a(b+(x-y)^r)^{-a} rab^a(b+y^r)^{-a-1}dy$$

Τα παραπάνω ολοκληρώματα είναι δύσκολο να υπολογιστούν χωρίς τη βοήθεια κάποιου υπολογιστικού πακέτου.

Όμως τα πράγματα απλουστεύονται με τη βοήθεια του θεωρήματος ,δηλαδή αν αποδείξουμε οτι η κατανομή $Burr$ είναι κατανομή τύπου Pareto τότε θα είναι και υποεκθετική κατανομή.

Η δεξιά ουρά της κατανομής δίνεται απο τη σχέση

$$\overline{F}(x) = \left(\frac{b}{b+x^r}\right)^a = \left(1 + \frac{X^r}{b}\right)^{-a} \cong b^{-a}X^{-ar}$$

Όμως αν θέσουμε $ar = q$ και $L(x) = b^{-a}$

Με

$$\frac{L(xy)}{L(x)} = \frac{b^{-a}}{b^{-a}} = 1$$

Παρατηρούμε ότι πρόκειται για κατανομή τύπου Pareto και ανήκει στην οικογένεια των υποεκθετικών κατανομών.

6.7 Σχέση κατανομών με μακριά ουρά και υποεκθετικών κατνομών

Σε αυτή την ενότητα θα αποδείξουμε πως αν μια κατανομή ανήκει στην κλάση των υποεκθετικών κατανομών τότε θα είναι και μια κατανομή με μακριά ουρά.

Έχουμε ήδη αναφέρει πως μια κατανομή F ορισμένη στο R θα λέμε οτι έχει μακριά ουρά αν F(x) $>$ 0 (right unbounded support) και για κάθε σταθερό y $>$ 0 ισχύει

$$\frac{\bar{F}(x+y)}{\bar{F}(x)} \to 1 \quad \text{όταν } x \to \infty$$

Η παραπάνω σχέση σημαίνει πως

$$\lim_{x\to\infty} \frac{\bar{F}(x+y)}{\bar{F}(x)} = 1$$

Αν λοιπόν η κατανομή έχει μακριά ουρά είναι λογικό μα ισχύει ταυτόχρονα η σχέση

$$\lim_{x\to\infty} \frac{\bar{F}(x-y)}{\bar{F}(x)} = 1$$

Στην ουσία τα παραπάνω δύο όρια δείχνουν πως η συμπεριφορά της κατανομής F δεν αλλάζει μετά απο κάποια μεγάλη τιμή που μπορεί να πάρει το x.Πρακτικά η πιθανότητα η τυχαία μεταβλητή X να πάρει μεγαλύτερη τιμή ή μικρότερη τιμή απο το x είναι σταθερή.

Η παραπάνω πρόταση μπορεί να αποτυπωθεί μαθηματικά ως εξής

$$\int_{x}^{+\infty} f(z)dz = \int_{x-y}^{+\infty} f(z)dz = \int_{x+y}^{+\infty} f(z)dz \qquad x \to \infty$$

Γνωρίζουμε οτι

$$\overline{F^{*2}}(x) \geq P(\{\max\{X_1, X_2\} > x\})$$

$$\overline{F^{*2}}(x) \geq 1 - P(\{\max\{X_1, X_2\} \leq x\}) \Rightarrow$$

$$\overline{F^{*2}}(x) \geq 1 - P(\{X_1 \leq x)P(X_2 \leq x) \Rightarrow$$

$$\overline{F^{*2}}(x) \geq 1 - F^2(x) = \big(1 - F(x)\big)\big(1 + F(x)\big) \Rightarrow$$

$$\frac{\overline{F^{*2}}(x)}{1 - F(x)} \geq 1 + F(x) \Rightarrow$$

$$\lim_{x\to\infty} \frac{\overline{F^{*2}}(x)}{1 - F(x)} \geq \lim_{x\to\infty}(1 + F(x))$$

Όμως

$$\lim_{x\to\infty}(F(x)) = 1$$

Και έτσι

$$\lim_{x\to\infty} \frac{\overline{F^{*2}}(x)}{1 - F(x)} \geq 2$$

Στην περίπτωση που η παραπάνω σχέση ισχύει με την ισότητα τότε η κατανομή F ανήκει στην κλάση των υποεκθετικών κατανομών.

Θα αποδείξουμε ότι στην περίπτωση όπου η κατανομή F ανήκει στην οικογένεια των υποεκθετικών κατανομών τότε

$$\frac{\overline{F^{*2}}(x)}{\bar{F}(x)} = 1 + \int_0^x \frac{\bar{F}(x-y)dF(y)}{\bar{F}(\chi)}$$

Τα πάντα ξεκινάνε με την συνέλιξη

$$\overline{F^{*2}}(x) = \int_0^{+\infty} \bar{F}(x-y)dF(y) = \int_0^x \bar{F}(x-y)dF(y) + \int_x^{+\infty} \bar{F}(x-y)dF(y)$$

Όμως το ολοκλήρωμα

$$\int_x^{+\infty} \bar{F}(x-y)dF(y) = \int_x^{+\infty} (1 - F(x-y))dF(y) = \int_x^{+\infty} dF(y) + \int_x^{+\infty} F(x-y))dF(y)$$

Σύμφωνα όμως με τους περιορισμούς μας θα πρέπει

$$F(0) = \int_{-\infty}^0 f(x)dx = 0$$

Θα πρέπει η τυχαία μεταβλητή $y \in (0, +\infty)\ \kappa\alpha\iota\ x - y\ \in (0, +\infty)$ ή ισοδύναμα $y < x$ έτσι σύμφωνα με τα παταπάνω το ολοκλήρωμα

$$\int_{x}^{+\infty} F(x-y))dF(y) = 0$$

$$\int_{x}^{+\infty} \bar{F}(x-y)dF(y) = [F(y)]_{x}^{+\infty} = \lim_{x\to\infty} F(x) - F(x) = 1 - F(x) = \bar{F}(x)$$

Και

$$\overline{F^{*2}}(x) = \int_{0}^{x} \bar{F}(x-y)dF(y) + \bar{F}(x)$$

Και τελικά προκύπτει πως

$$\frac{\overline{F^{*2}}(x)}{\bar{F}(x)} = 1 + \int_{0}^{x} \frac{\bar{F}(x-y)dF(y)}{\bar{F}(\chi)}$$

Ας προσπαθήσουμε να δούμε με μια διαφορετική προσέγγυση την συνέλιξη των δύο τυχαίων μεταβλητών X, Y.

$$\overline{F^{*2}}(x) = \mathrm{P}\{X_1 + X_2 > x\} = \mathrm{P}(X + Y > x; Y > x) + \mathrm{P}(X + Y > x; Y < x) =$$

$$= \mathrm{P}\{X > x \cap Y > x\} + \mathrm{P}\{X < x \cap Y > x\} + \mathrm{P}(X + Y > x; Y < x) =$$

$$\mathrm{P}(X > x)\ \mathrm{P}(Y > x) + \mathrm{P}\ (X < x)\ \mathrm{P}(Y > x) + \mathrm{P}(X + Y > x; Y < x) =$$

$$= \bar{F}(x) + \int_{0}^{x} \bar{F}\ (x-y)dF(y)$$

Καταλήξαμε λοιπόν στο συμπέρασμα πως ισχύει η σχέση

$$\overline{F^{*2}}(x) = \bar{F}(x) + \int_{0}^{x} \bar{F}\ (x-y)dF(y)$$

Και

$$\frac{\overline{F^{*2}}(x)}{\bar{F}(x)} = 1 + \int_{0}^{x} \frac{\bar{F}(x-y)dF(y)}{\bar{F}(\chi)}$$

Βέβαια για να γίνει αυτό υποθέτουμε πως η δεξιά ουρά της κατανομής είναι διάφορη του μηδενός για κάθε x άρα και το όριο

$$\lim_{x\to\infty} sup e^{\lambda x}\, \bar{F}(x) = \infty\ για\ όλα\ τα\ \lambda > 0$$

Με λίγα λόγια έχουμε υποθέσει πως η κατανομή έχει βαριά ουρά.

Στην περίπτωση όπου η κατανομή είναι υποεκθετική ισχύει

$$\lim_{x\to\infty}\frac{\overline{F^{*2}}(x)}{\bar{F}(x)}=2=1+\lim_{x\to\infty}\int_0^x\frac{\bar{F}(x-y)dF(y)}{\bar{F}(x)}$$

Όρισμός υποεκθετικής κατανομής 3

Μια κατανομή F θα λέμε οτι ανήκει στην κλάση των υποεκθετικών κατανομών αν και μόνο αν

$$\lim_{x\to\infty}\int_0^x\frac{\bar{F}(x-y)dF(y)}{\bar{F}(\chi)}=1$$

Αν ορίσουμε την ουρά της κατανομής ως εξής

$$G(x)=\begin{cases}F(x) & x<0\\ 1-F(x) & x>0\end{cases}$$

Μπορούμε να γράψουμε την παραπάνω σχέση ως εξής

$$\lim_{x\to\infty}\int_0^{+\infty}\frac{G(x-y)dF(y)}{G(x)}=1$$

Μια κατανομή F θα λέμε οτι ανήκει στην κλάση των υποεκθετικών κατανομών αν και μόνο αν

$$\lim_{x\to\infty}\int_0^{+\infty}\frac{G(x-y)dF(y)}{G(x)}=1$$

Δηλαδή στην περίπτωση της υποεκθετικής κατανομής θα ισχύει

$$\lim_{x\to\infty}\int_0^x\frac{\bar{F}(x-y)dF(y)}{\bar{F}(\chi)}=1$$

Για κάποιο $x^{'}\leq x$ το ολοκλήρωμα

$$\int_0^x\frac{\bar{F}(x-y)dF(y)}{\bar{F}(\chi)}=\int_0^{x^{'}}\frac{\bar{F}(x-y)dF(y)}{\bar{F}(\chi)}+\int_{x^{'}}^{+\infty}\frac{\bar{F}(x-y)dF(y)}{\bar{F}(\chi)}$$

Όμως το ολοκλήρωμα

$$\int_0^{x'} \frac{\bar{F}(x-y)dF(y)}{\bar{F}(\chi)} = \left[\frac{\bar{F}(x-y)F(y)}{\bar{F}(\chi)}\right]_0^{x'} - \int_0^{x'} \frac{F(y)d\bar{F}(x-y)}{\bar{F}(\chi)} =$$

$$\frac{\bar{F}(x-x')F(x')}{\bar{F}(\chi)} - \frac{\bar{F}(x-0)F(0)}{\bar{F}(\chi)} - \int_0^{x'} \frac{F(y)d\bar{F}(x-y)}{\bar{F}(\chi)}$$

Ξέρουμε όμως ότι

$$F(x) = \int_{-\infty}^{x} f(x)dx = \int_0^{x} f(x)dx$$

Και σύμφωνα με την παραπάνω σχέση

$$F(0) = 0$$

Και το ολοκλήρωμα θα ισούται

$$\int_0^{x'} \frac{\bar{F}(x-y)dF(y)}{\bar{F}(x)} = \frac{\bar{F}(x-x')F(x')}{\bar{F}(x)} - \int_0^{x'} \frac{F(y)d\bar{F}(x-y)}{\bar{F}(x)}$$

Ας δουλέψουμε λίγο με το ολοκλήρωμα

$$\int_0^{x'} \frac{F(y)d\bar{F}(x-y)}{\bar{F}(x)} = -\int_0^{x'} \frac{F(y)dF(x-y)}{\bar{F}(x)}$$

Και έτσι

$$\int_0^{x'} \frac{\bar{F}(x-y)dF(y)}{\bar{F}(x)} = \frac{\bar{F}(x-x')F(x')}{\bar{F}(x)} + \int_0^{x'} \frac{F(y)dF(x-y)}{\bar{F}(x)}$$

Αν θέσουμε $z = x - y \Rightarrow y = x - z$ $\mu\varepsilon$ $z \in (x - x', x)$ το ολοκλήρωμα

$$\int_0^{x'} \frac{F(y)dF(x-y)}{\bar{F}(x)} = \int_{x-x'}^{x'} \frac{F(x-z)dF(z)}{\bar{F}(x)}$$

Και τελικά το ολοκλήρωμα

$$\int_0^{x'} \frac{\bar{F}(x-y)dF(y)}{\bar{F}(x)} = \frac{\bar{F}(x-x')F(x')}{\bar{F}(x)} + \int_{x-x'}^{x'} \frac{F(x-z)dF(z)}{\bar{F}(x)} \geq F(x')$$

Αντίστοιχα το ολοκλήρωμα

$$\int_{x'}^{x} \frac{\bar{F}(x-y)dF(y)}{\bar{F}(\chi)} = \left[\frac{\bar{F}(x-y)F(y)}{\bar{F}(x)}\right]_{x'}^{x} - \int_{x'}^{x} \frac{F(y)d\bar{F}(x-y)}{\bar{F}(\chi)} =$$

$$\frac{\bar{F}(x-x)F(x)}{\bar{F}(x)}-\frac{\bar{F}(x-x')F(x')}{\bar{F}(\chi)}+\int_{x'}^{x}\frac{F(y)dF(x-y)}{\bar{F}(\chi)}\geq\frac{\bar{F}(x-x')F(x)}{\bar{F}(x)}-\frac{\bar{F}(x-x')F(x')}{\bar{F}(\chi)}$$

Έτσι προκύπτει ότι

$$\int_{x'}^{x}\frac{\bar{F}(x-y)dF(y)}{\bar{F}(\chi)}\geq\frac{\bar{F}(x-x')}{\bar{F}(x)}(F(x)-F(x'))$$

Και τελικά το ολοκλήρωμα

$$\int_{0}^{x}\frac{\bar{F}(x-y)dF(y)}{\bar{F}(x)}\geq F(x')+\frac{\bar{F}(x-x')}{\bar{F}(x)}(F(x)-F(x'))$$

Με αναδιάταξη των όρων της παραπάνω σχέσης έχουμε

$$\left\{\int_{0}^{x}\frac{\bar{F}(x-y)dF(y)}{\bar{F}(x)}-F(x')\right\}\{F(x)-F(x')\}^{-1}\geq\frac{\bar{F}(x-x')}{\bar{F}(x)}$$

Και το όριο

$$\lim_{x\to\infty}\left[\left\{\int_{0}^{x}\frac{\bar{F}(x-y)dF(y)}{\bar{F}(x)}-F(x')\right\}\{F(x)-F(x')\}^{-1}\right]\geq\lim_{x\to\infty}\frac{\bar{F}(x-x')}{\bar{F}(x)}$$

$$\left\{\lim_{x\to\infty}\left\{\int_{0}^{x}\frac{\bar{F}(x-y)dF(y)}{\bar{F}(x)}\right\}-F(x')\right\}\left\{\lim_{x\to\infty}F(x)-F(x')\right\}^{-1}$$

Στην περίπτωση της υποεκθετικής κατανομής θα ισχύει

$$\lim_{x\to\infty}\int_{0}^{x}\frac{\bar{F}(x-y)dF(y)}{\bar{F}(\chi)}=1\ \kappa\alpha\iota\ \lim_{x\to\infty}F(x)=1$$

Δηλαδή το όριο

$$\lim_{x\to\infty}\frac{\bar{F}(x-x')}{\bar{F}(x)}\leq 1$$

Θεώρημα

Αν η κατανομή Fανήκει στην οικογένεια των υποεκθετικών κατανομών τότε θα ανήκει και στην οικογένεια κατανομών με βαριά ουρά.

Απόδειξη

Έχουμε ήδη αποδείξει πως όταν το όριο

$$\lim_{x\to\infty}\frac{H(x)}{x}=0$$

Τότε η κατανομή είναι μια κατανομή με βαριά ουρά.Με $H(x)$έχουμε ορίσει την σωρευτική συνάρτηση διακυνδίνευσης και μάλιστα έχουμε δείξει πως

$$H(x)=-\ln\bar{F}\,(x)$$

Έχουμε επίσης αποδείξει πως αν η F ανήκει στηνοικογένεια υποεκθετικών κατανομών τότε ισχύει πως

$$\lim_{x\to\infty}\frac{\bar{F}(x-y)}{\bar{F}(x)}=1$$

Παίρνωντας τον λογάριθμο της παραπάνω σχέσης έχουμε

$$\lim_{x\to\infty}\ln\left(\frac{\bar{F}(x-y)}{\bar{F}(x)}\right)=\ln 1=0$$

$$\lim_{x\to\infty}\{\ln\bar{F}(x-y)-\ln\bar{F}\,(x)\}=0$$

Η παραπάνω σχέση μπορεί να γραφεί με την βοήθεια της συνάρτησης σωρευτικής διακυνδίνευσης ως εξής

$$\lim_{x\to\infty}(H(x)-H(x-y))=0\;\gamma\iota\alpha\; y\geq 0$$

Τότε για αυθαίρετο $\varepsilon>0$ υπάρχει $X_0>0$ τετοιο ώστε για όλα τα $x\geq x_0$ έχουμε

$$H(x)-H(x-1)<\varepsilon$$

Μπορούμε να φτιάξουμε την αναδρομική σχέση

$$H(x)\leq H(x-1)+\varepsilon\leq H(x-2)+2\varepsilon\leq\cdots H(x-n)+n\varepsilon$$

Όπου ορίζουμε ως $n=x-x_0$ και $x_0\leq x-n<x_0+1$

Έτσι η $H(x)$ φράσεται απο το πάνω μέρος της απο την ποσότητα $H(x-n)+n\varepsilon$ και

$$H(x)\leq\sup_{x_0\leq x'\leq x_0+1}H(x')+(x-x_0)\varepsilon\;\gamma\iota\alpha\; x\geq x_0$$

και επειδή το ε είναι αυθαίρετο

$$\lim_{x\to\infty}\frac{H(x)}{x}=0$$

6.8 Συμπέρασμα

Αν η κατανομή F ανήκει στην οικογένεια των υποεκθετικών κατανομών τότε θα ανήκει και στην οικογένεια των κατανομών με μακριά ουρά,και στην οικογένεια κατανομών με βαριά ουρά.

Αν λοιπόν συμβολίσουμε με B την κλάση κατανομών με βαριά ουρά L την οικογένεια των κατανομών με μακριά ουρά και S την οικογένεια των υποεκθετικών κατανομών τότε σύμφωνα με τα παραπάνω είναι προφανές πως

$$B \supseteq L \supseteq S$$

6.9 Συνελίξεις κατανομών που χαρακτηρίζονται από μακριά ουρά

Αρχικά μια συνέλιξη δεν είναι τίποτε παραπάνω από μια πιθανότητα και έτσι θα την προσεγγίσουμε.

$$\overline{F * G}(x) = P(X + \eta > x)$$

Μπορούμε να κάνουμε ανάλυση της παραπάνω πιθανότητας ως προς την τιμή ή το διάστημα τιμών τις οποίες μπορεί να πάρει η τυχαία μεταβλητή X.

Έστω λοιπόν μια θετική σταθερά h .Τότε η τυχαία μεταβλητή X μπορεί να πάρει τιμές μεγαλύτερες από την σταθερά αυτή ή και μικρότερες ή ίσες με τη σταθερά αυτή.Πιο συγκεκριμένα

$$\overline{F * G}(x) = P(X + \eta > x, X > h) + P(X + \eta > x, X \leq h)$$

$$P(X + \eta > x/X > h)P(X > h) + P(X + \eta > x/X \leq h)P(X \leq h)$$

Αν επιπροσθέτως

$$h \leq \frac{x}{2}$$

$$\overline{F * G}(x) = P(X + \eta > x, X \leq h) + P(X + \eta > x, \eta \leq h) + P(X + \eta > x, X > h, \eta > h)$$

Η παραπάνω σχέση θα μπορούσε να γραφτεί με τη βοήθεια ολοκληρωμάτων αλλά πριν την περιγράψουμε με ολοκληρώματα καλό θα ήταν να αναφέρουμε τα εξής

$$P(X + \eta > x, X \leq h) = \int_{-\infty}^{h} \bar{G}\,(x - y)dF(y)$$

Λογικά αυτή η πιθανότητα θα είναι ίση με την πιθανότητα

$$P(X+\eta > x, \eta > x-h) = \int_{x-h}^{+\infty} \bar{F}\,(x-y)dG(y)$$

Τα παραπάνω θεωρούμε πως είναι ίσα διότι αν στην πρώτη σχέση η τυχαία μεταβλητή

$$X \leq h \leq \frac{x}{2}$$

Παραδείγματος χάρη $X = \frac{x}{4}$ τότε για να πραγματοποιηθεί το ενδεχόμενο $\{X+\eta > x, X \leq h\}$ θα πρέπει

$$\eta > x - h = x - \frac{x}{4}$$

Δηλαδή τα ενδεχόμενα $\{X+\eta > x, X \leq h\}$ και $\{X+\eta > x, \eta > x-h\}$ είναι ισοπίθανα και αυτό έχει ως αποτέλεσμα

$$\int_{-\infty}^{h} \bar{G}\,(x-y)dF(y) = \int_{x-h}^{+\infty} \bar{F}\,(x-y)dG(y)$$

Αρκεί λοιπόν να υπολογίσουμε το ένα από τα δύο ολοκληρώματα.

Θα υπολογίσουμε λοιπόν το ολοκλήρωμα

$$\int_{-\infty}^{h(x)} \bar{G}\,(x-y)dF(y) \leq \bar{G}(x-h(x))$$

Η παραπάνω παρατήρηση προκύπτει με στοιχειώδεις πράξεις ως εξής

$$\int_{-\infty}^{h(x)} \bar{G}\,(x-y)dF(y) = [\bar{G}(x-y)F(y)]_{-\infty}^{h(x)} - \int_{-\infty}^{h(x)} F(y)d\bar{G}(x-y)$$

$$\leq [\bar{G}(x-y)F(y)]_{-\infty}^{h(x)} \leq \bar{G}(x-h(x))$$

Από την άλλη πλευρά

$$\int_{-\infty}^{h(x)} \bar{G}\,(x-y)dF(y) \geq \int_{-h(x)}^{h(x)} \bar{G}\,(x-y)dF(y)$$

Όμως

$$\int_{-h(x)}^{h(x)} \bar{G}\,(x-y)dF(y) =$$

$$= \bar{G}\big(x - h(x)\big)F\big(h(x)\big) - \bar{G}\big(x + h(x)\big)F\big(-h(x)\big) - \int_{-h(x)}^{h(x)} F(y)d\bar{G}(x - y)$$

Αν όμως η συνάρτηση G είναι συνάρτηση με μακριά ουρά (ή ακόμη γενικότερα *h-insensitive)* τότε

$$\bar{G}\big(x - h(x)\big) \sim \bar{G}(x) \sim \bar{G}\big(x + h(x)\big) \quad x \to \infty$$

Έτσι η παραπάνω σχέση μας δίνει

$$\int_{-\infty}^{h(x)} \bar{G}\,(x - y)dF(y) \geq \bar{G}\big(x + h(x)\big)F(-h(x), h(x))$$

$$\geq \bar{G}\big(x + h(x)\big)$$

Προκύπτει στην ουσία η διπλή ανισότητα

$$\bar{G}\big(x + h(x)\big) \leq \int_{-\infty}^{h(x)} \bar{G}\,(x - y)dF(y) \leq \bar{G}(x - h(x))$$

και όταν $x \to \infty$

$$\int_{-\infty}^{h(x)} \bar{G}\,(x - y)dF(y) \sim \bar{G}(x)$$

Θεώρημα

Για οποιαδήποτε κατανομή $F_1, F_2, \dots, F_n$ ορισμένες στο $\mathbb{R}^+$ με την ιδιότητα $\bar{F}(x) > 0$ για οποιοδήποτε x ισχύει η παρακάτω σχέση

$$\lim_{x\to\infty} inf \frac{\bar{F}^{*n}(x)}{\bar{F}_1(x) + \bar{F}_2(x) + .. + \bar{F}_n(x)} \geq 1$$

Η παραπάνω σχέση ισχύει ακόμη και σε περιπτώσεις που οι τυχαίες μεταβλητές δεν είναι ισόνομες αλλά παραμένουν ανεξάρτητες.

Έτσι αν έχουμε $F_1, F_2, \dots, F_n, G$ κατανομές ορισμένες στο $\mathbb{R}^+$με ουρές που είναι πάντα θετικές και ποτέ μηδέν τότε ισχύει η σχέση

$$\lim_{x\to\infty} inf \frac{\overline{F^{*n} * G}(x)}{\bar{F}_1(x) + \bar{F}_2(x) + .. + \bar{F}_n(x) + \bar{G}(x)} \geq 1$$

Την απόδειξη του θεωρήματος την έχουμε ήδη παρουσιάσει σε προηγούμενη ενότητα αυτού του κεφαλαίου.

Το παραπάνω θεώρημα γενικεύεται και στον $\mathbb{R}$ με την προϋπόθεση πως το άθροισμα

$$\bar{F}_1(x) + \bar{F}_2(x) + .. + \bar{F}_n(x)$$

να είναι μια συνάρτηση με μακριά ουρά.

Πιο συγκεκριμένα έχουμε να παραθέσουμε το παρακάτω θεώρημα.

Θεώρημα

Έστω οι κατανομές $F_{1,}F_{2,} \dots, F_n$ κατανομές ορισμένες στο $\mathbb{R}$ τέτοιες ώστε η συνάρτηση

$$\bar{F}_1(x) + \bar{F}_2(x) + .. + \bar{F}_n(x)$$

Να χαρακτηρίζεται από μακριά ουρά.

Τότε για οποιαδήποτε κατανομή G ισχύει η παρακάτω σχέση

$$\lim_{x\to\infty} inf \frac{\overline{F^{*n} * G}(x)}{\bar{F}_1(x) + \bar{F}_2(x) + .. + \bar{F}_n(x)} \geq 1$$

Ειδικότερα η παραπάνω σχέση ισχύει όταν όλες οι συναρτήσεις κατανομής F_i $i = 1,2,3, \dots . n$ είναι κατανομές με μακριά ουρά.

Απόδειξη

Η απόδειξη του παραπάνω θεωρήματος είναι σχετικά εύκολη αν σκεφτούμε τα εξής

Αρχικά όλες οι τυχαίες μεταβλητές μας ορίζονται στο $\mathbb{R}$ δηλαδή μπορούν να πάρουν θετικές ή αρνητικές τιμές με θετική πιθανότητα. Για την ανάλυση μας θα θεωρήσουμε πως καμία τυχαία μεταβλητή δεν μπορεί να πάρει τιμή μικρότερη από το $-\alpha$. Μπορούμε λοιπόν να φανταστούμε το ενδεχόμενο

$$\{X_1 + X_2 + \cdots + X_n + \eta > x\}$$

Αν τώρα όλες οι τυχαίες μεταβλητές εκτός από μια πάρουν την ελάχιστη τιμή που μπορούν να πάρουν που εμείς έχουμε υποθέσει πως είναι το $-\alpha$ τότε για να πραγματοποιηθεί το παραπάνω ενδεχόμενο θα πρέπει μια μεταβλητή πάντα να πάρει τιμή μεγαλύτερη από

$$x + a + a + a \dots + a = x + na$$

Δηλαδή σύμφωνα με όσα είπαμε στην προηγούμενη παράγραφο

$$\{X_1 + X_2 + \cdots + X_n + \eta > x\} \supseteq \bigcup_{k=1}^{n} \{X_k > x, a < X_j \leq x, \eta > -\alpha\}$$

Και βάζοντας στην παραπάνω σχέση ενδεχομένων το μέτρο της πιθανότητας προκύπτει η εξής ανισότητα

$$P(X_1 + X_2 + \cdots + X_n + \eta > x) \geq \sum_{k=1}^{n} P\,(X_k > x + na, a < X_j \leq x, \eta > -\alpha)$$

Όμως

$$\sum_{k=1}^{n} P\left(X_k > x + na, a < X_j \leq x, \eta > -\alpha\right) = \bar{G}(-\alpha) \sum_{k=1}^{n} P\,(X_k > x + na) \prod_{j \neq k} F(-a, x]$$

Υποθέτουμε πως $\bar{G}(-a) = 1 - \varepsilon$ και $F(-a, x] = 1 - \varepsilon$ με $\varepsilon > 0$ και μικρότερο της μονάδας

Η συνέλιξη λοιπόν

$$\overline{F^{*n} * G}(x) = (1 - \varepsilon) \sum_{k=1}^{n} P\,(X_k > x + na)(1 - \varepsilon)^{n-1}$$

$$= (1 - \varepsilon)^n \sum_{k=1}^{n} P\,(X_k > x + na) = (1 - \varepsilon)^n \sum_{k=1}^{n} \bar{F}_k\,(x + na)$$

Μέχρι στιγμής στην ανάλυση μας δεν έχουμε χρησιμοποιήσει την έννοια της μακριάς ουράς σύμφωνα με τον ορισμό της οποίας για ασυμπτωτικές καταστάσεις όπου το $x \to \infty$

$$\bar{F}(x \pm y) \sim \bar{F}(x) \quad x \to \infty$$

Σύμφωνα λοιπόν με την παραπάνω σχέση

$$\sum_{k=1}^{n} \bar{F}_k\,(x + na) = \bar{F}_1(x + na) + \bar{F}_2(x + na) + \ldots + \bar{F}_n(x + na)$$

$$\sim \bar{F}_1(x) + \bar{F}_2(x) + \ldots + \bar{F}_n(x) = \sum_{k=1}^{n} \bar{F}_k\,(x)$$

Κάνοντας λοιπόν χρήση της υπόθεσης πως κάθε μια από τις κατανομές F_i $i = 1,2,3, \ldots . n$ είναι κατανομές με μακριά ουρά μπορούμε να περιγράψουμε τη συνέλιξη ως εξής

$$\overline{F^{*n} * G}(x) = (1 - \varepsilon)^n \sum_{k=1}^{n} \bar{F}_k\,(x)$$

Το ε είναι μια αυθαίρετη μη αρνητική ποσότητα και μπορούμε να της δώσουμε την τιμή μηδέν

Έτσι καταλήγουμε στη σχέση που θέλαμε να αποδείξουμε.

Πόρισμα

Η σχέση

$$\lim_{x\to\infty} inf \frac{\overline{F^{*n} * G}(x)}{\bar{F}_1(x) + \bar{F}_2(x) + .. + \bar{F}_n(x)} \geq 1$$

Ισχύει και στην ειδική περίπτωση που έχουμε $F_i \; i = 1,2,3, \dots . n - 1$ και $G = F$

Τότε

$$\lim_{x\to\infty} inf \frac{\overline{F^{*n}}(x)}{\bar{F}_1(x) + \bar{F}_2(x) + .. + \bar{F}_n(x)} \geq 1$$

και επειδή οι $F_1, F_2, \dots, F_n$ ακολουθούν την ίδια κατανομή η παραπάνω σχέση μπορεί να γραφτεί ισοδύναμα ως εξής

$$\lim_{x\to\infty} inf \frac{\overline{F^{*n}}(x)}{\bar{F}(x)} \geq n$$

Ή ακόμη και

$$\frac{1}{n} \lim_{x\to\infty} inf \, \overline{F^{*n}}(x) \geq \bar{F}(x)$$

Πόρισμα

Έστω μια κατανομή F ορισμένη στο $\mathbb{R}$ με την ιδιότητα της μακριάς ουράς και έστω μια κατανομή G τέτοια ώστε $\bar{G}(\alpha) = 0$ για κάποιο α.Τότε ισχύει η σχέση

$$\frac{\overline{F * G}(x)}{\bar{F}(x)} \sim 1 \quad x \to \infty$$

Απόδειξη

Γνωρίζουμε ότι η συνέλιξη

$$\overline{F * G}(x) = P(X + \eta > x)$$

Το ενδεχόμενο όμως

$$\{X + \eta > x\} = \{X + \eta > x, \eta > \alpha\} \bigcup \{X + \eta > x, \eta \leq \alpha\}$$

Και η αντίστοιχη πιθανότητα ισούται με

$$P(X+\eta>x)=P\left(\{X+\eta>x,\eta>\alpha\}\bigcup\{X+\eta>x,\eta\leq\alpha\}\right)=$$

$$=\mathrm{P}(X+\eta>x,\eta>\alpha)+P(X+\eta>x,\eta\leq\alpha)=$$

$$P(X+\eta>x/\eta>\alpha)P(\eta>\alpha)+P(X+\eta>x/\eta\leq\alpha)P(\eta\leq\alpha)=$$

$$=P(X+\eta>x/\eta>\alpha)\bar{G}(a)+P(X+\eta>x/\eta\leq\alpha)G(a)$$

Όμως βασική υπόθεση του πορίσματος μας είναι πως $\bar{G}(a)=0$ έτσι η παραπάνω σχέση μας δίνει

$$P(X+\eta>x)=P(X+\eta>x/\eta\leq\alpha)G(a)=P(X+\eta>x/\eta\leq\alpha)$$

Η πιθανότητα όμως

$$P(X+\eta>x/\eta\leq\alpha)\leq P(X+a>x)=P(X>x-a)$$

Καταλήξαμε λοιπόν στο συμπέρασμα πως

$$\overline{F*G}(x)\leq\bar{F}(x-a)$$

Από την άλλη πλευρά όμως έχουμε ήδη αποδείξει πως

$$\lim_{x\to\infty}inf\frac{\overline{F^{*n}*G}(x)}{\bar{F}_1(x)+\bar{F}_2(x)+..+\bar{F}_n(x)}\geq 1$$

Για $n=1$ η παραπάνω σχέση μας δίνει

$$\frac{\overline{F*G}(x)}{\bar{F}(x)}\geq 1$$

Καταλήξαμε λοιπόν στο συμπέρασμα πως

$$\frac{\overline{F*G}(x)}{\bar{F}(x)}\sim 1\ \ x\to\infty$$

Θεώρημα

Άθροισμα μη αρνητικών τυχαίων μεταβλητών που ακολουθούν υποεκθετική κατανομή δίνει υποεκθετική κατανομή.

Απόδειξη

Έστω

$$S_i=X_{i1}+X_{i2+..}+X_{in}\qquad i=1,2,..,m$$

με

X_{ij} να ακολουθεί υποεκθετική κατανομή για $j=1,2,..n\ \kappa\alpha\iota\ i=1,2,..,m$

Τότε η πιθανότητα

$$\mathbb{P}\left(\sum_{i=1}^{m} S_i > x\right) = \mathbb{P}(X_{i1} + X_{i2} + \cdots + X_{mn} > x)$$

$$\sim nm\mathbb{P}(X_{i1} > x) \qquad x \to \infty$$

Από την παραπάνω σχέση προκύπτει

$$\frac{\mathbb{P}(\sum_{i=1}^{m} S_i > x)}{m\mathbb{P}(S_1 > x)} \sim \frac{n\mathbb{P}(X_{i1} > x)}{\mathbb{P}(S_1 > x)} \to 1 \quad x \to \infty$$

Κεφάλαιο 7

Υποεκθετικές κατανομές ορισμένες σε ολόκληρο το $\mathbb{R}$

7.1 Εισαγωγή

Μέχρι τώρα στην ανάλυση μας έχουμε υποθέσει πως οι τυχαίες μεταβλητές μπορούν να πάρουν μόνο θετικές τιμές .

Σε αυτό το κεφάλαιο θα εισάγουμε τα θεμέλια εκείνα ώστε η θεωρία των υποεκθετικών κατανομών να επεκταθεί σε ολόκληρο το $\mathbb{R}$**.**

Μέχρι στιγμής από το προηγούμενο κεφάλαιο έχουμε καταλήξει στην διαπίστωση πως όσο η ανάλυση μας κινείται στον θετικό ημιάξονα τότε ισχύει το εξής

αν η κατανομή F ανήκει στην οικογένεια των υποεκθετικών κατανομών τότε θα ανήκει και στην οικογένεια των κατανομών με μακριά ουρά,και στην οικογένεια κατανομών με βαριά ουρά.

Με το επόμενο παράδειγμα μπορούμε σχετικά εύκολα να διαπιστώσουμε πως δεν ισχύει το ίδιο και σε ολόκληρο το $\mathbb{R}$

Έστω λοιπόν μια κατανομή

$$F(x) = 1 - (x + A + 1)^{-2} e^{-(x+A)} \qquad x \geq -A$$

Η ουρά της παραπάνω κατανομής ισούται με

$$\bar{F}(x) = (x + A + 1)^{-2} e^{-(x+A)} \qquad x \geq -A$$

Η συνέλιξη δύο τυχαίων μεταβλητών που ακολουθούν κατανομή με συνάρτηση κατανομής αυτή που ήδη αναφέραμε θα ισούται με

$$\overline{F * F}(x) = P\left(X + \eta > x, X \leq \frac{x}{2}\right) + P\left(X + \eta > x, \eta \leq \frac{x}{2}\right) + P\left(X + \eta > x, X > \frac{x}{2}, \eta > \frac{x}{2}\right)$$

$$= \int_{-\infty}^{\frac{x}{2}} \bar{F}(x - y)\, dF(y) + \int_{-\infty}^{\frac{x}{2}} \bar{F}(x - y)\, dF(y) + P\left(X + \eta > x, X > \frac{x}{2}, \eta > \frac{x}{2}\right)$$

$$= 2\int_{-\infty}^{\frac{x}{2}} \bar{F}(x - y)\, dF(y) + P\left(X + \eta > x, X > \frac{x}{2}, \eta > \frac{x}{2}\right)$$

Η πιθανότητα

$$P\left(X+\eta>x, X>\frac{x}{2}, \eta>\frac{x}{2}\right)=P(X+\eta>x/X>\frac{x}{2}, \eta>\frac{x}{2}) P\left(X>\frac{x}{2}\right) P(\eta>\frac{x}{2})$$

Όμως η δεσμευμένη πιθανότητα

$$P(X+\eta>x/X>\frac{x}{2}, \eta>\frac{x}{2})=1$$

Και οι ανεξάρτητες τυχαίες μεταβλητές X, η είναι και ισόνομες ά,ρα

$$P\left(X>\frac{x}{2}\right)=P\left(\eta>\frac{x}{2}\right)=\bar{F}(x/2)$$

Έτσι με αυτές τις δύο παρατηρήσεις α από κοινού πιθανότητα

$$P\left(X+\eta>x, X>\frac{x}{2}, \eta>\frac{x}{2}\right)=\bar{F}(x/2)^2$$

Από την άλλη πλευρά η πιθανότητα

$$P\left(X+\eta>x, X>\frac{x}{2}, \eta>\frac{x}{2}\right)=\int_{\frac{x}{2}}^{+\infty} \int_{\frac{x}{2}}^{+\infty} \bar{F}(x-y)\, dF(y) dx$$

$$=\int_{\frac{x}{2}}^{+\infty} \bar{F}\left(max\left(x-y, \frac{x}{2}\right)\right) dF(y)$$

Σύμφωνα λοιπόν με τα παραπάνω έχουμε το εξής αποτέλεσμα

$$\int_{\frac{x}{2}}^{+\infty} \bar{F}\left(max\left(x-y, \frac{x}{2}\right)\right) dF(y)=\bar{F}(x/2)^2$$

Η συνέλιξη μας πλέον μπορεί να γραφτεί ως εξής

$$\overline{F*F}(x)=2\int_{-\infty}^{\frac{x}{2}} \bar{F}(x-y)\, dF(y)+\int_{\frac{x}{2}}^{+\infty} \bar{F}\left(max\left(x-y, \frac{x}{2}\right)\right) dF(y)$$

$$=2\int_{-\infty}^{\frac{x}{2}} \bar{F}(x-y)\, dF(y)+\bar{F}(x/2)^2$$

Ήρθε η ώρα να κάνουμε αντικατάσταση της ουράς που στο παραπάνω ολοκλήρωμα με την ουρά που εξ αρχής έχουμε υποθέσει πως η κατανομή μας έχει

$$\int_{-\infty}^{\frac{x}{2}} \bar{F}(x-y)\, dF(y) = 2e^{-x-A} \int_{-\boldsymbol{A}}^{\frac{x}{2}} (x-y)^{-\mathbf{2}} e^{y}\, dF(y)$$

Σύμφωνα λοιπόν με τον τελευταίο υπολογισμό η συνέλιξη μας ισούται με

$$\overline{F*F}(x) = 2e^{-x-A} \int_{-\boldsymbol{A}}^{\frac{x}{2}} (x-y)^{-\mathbf{2}} e^{y}\, dF(y) + \bar{F}(x/2)^2$$

Όταν το $x \to \infty$

$$\overline{F*F}(x) \sim 2e^{-x-A} \int_{-\boldsymbol{A}}^{\frac{x}{2}} (x-y)^{-\mathbf{2}} e^{y}\, dF(y) + \lim_{x\to\infty} \bar{F}(x/2)^2$$

$$\sim 2x^{-2} e^{-x-A} \int_{-\boldsymbol{A}}^{\frac{x}{2}} e^{y}\, dF(y) + \lim_{x\to\infty} \bar{F}(x/2)^2$$

$$\sim 2\bar{F}(x) \int_{-\boldsymbol{A}}^{\infty} e^{y}\, dF(y) + \lim_{x\to\infty} \bar{F}(x/2)^2$$

Και

$$\frac{\overline{F*F}(x)}{\bar{F}(x)} \sim 2 \int_{-\boldsymbol{A}}^{\infty} e^{y}\, dF(y) + \lim_{x\to\infty} \frac{\bar{F}(x/2)^2}{\bar{F}(x)}$$

Όμως

$$\lim_{x\to\infty} \frac{\bar{F}(x/2)^2}{\bar{F}(x)} = 0(\bar{F}(x))$$

Η παραπάνω σχέση μας λέει με απλά λόγια πως ο αριθμητής τείνει (τρέχει) πολύ πιο γρήγορα στο μηδέν σε σχέση με τον παρονομαστή

Έτσι

$$\frac{\overline{F*F}(x)}{\bar{F}(x)} \sim 2 \int_{-\boldsymbol{A}}^{\infty} e^{y}\, dF(y) + 0(\bar{F}(x))$$

Και με κατάλληλη επιλογή του A το ολοκλήρωμα

$$\int_{-\boldsymbol{A}}^{\infty} e^{y}\, dF(y) = 1$$

Καταλήξαμε λοιπόν στο συμπέρασμα πως

$$\frac{\overline{F*F}(x)}{\bar{F}(x)} \sim 2 \quad x \to \infty$$

ή ισοδύναμα

$$\overline{F*F}(x) \sim 2\bar{F}(x)\ x \to \infty$$

Αν δουλεύαμε στον θετικό ημιάξονα τότε η κατανομή F θα άνηκε στην οικογένεια των υποεκθετικών κατανομών η οποία είναι υποσύνολο της οικογένειας των κατανομών με μακριά ουρά. Δηλαδή θα ήταν και κατανομή με μακριά ουρά.

Θα είχε δηλαδή την ιδιότητα

$$\frac{\bar{F}(x \pm y)}{\bar{F}(x)} \sim 1 \qquad x \to \infty$$

Στο παράδειγμα μας όμως

$$\frac{\bar{F}(x+y)}{\bar{F}(x)} = \frac{(x+y+A+1)^{-2}e^{-(x+y+A)}}{(x+A+1)^{-2}e^{-(x+A)}} = e^{-y} \neq 1 \quad x \to \infty$$

Καταλήγουμε λοιπόν στο συμπέρασμα πως στην περίπτωση που δουλεύουμε σε ολόκληρο το $\mathbb{R}$ ο ορισμός της υποεκθετικότητας απαιτεί περισσότερες συνθήκες από την

$$\overline{F*F}(x) \sim 2\bar{F}(x)\ x \to \infty$$

Λήμμα

Έστω F μια κατανομή ορισμένη στον $\mathbb{R}$ και ξ μια τυχαία μεταβλητή με κατανομή F τότε τα ακόλουθα είναι ισοδύναμα

- Η F είναι κατανομή με μακριά ουρά και επίσης $\overline{F*F}(x) \sim 2\bar{F}(x)\ x \to \infty$
- Η κατανομή F^+ της τυχαίας μεταβλητής ξ^+ ανήκει στην οικογένεια των υποεκθετικών κατανομών.
- Η δεσμευμένη κατανομή $G(B) = P(\xi \in B/\xi \geq 0)$ ανήκει στην οικογένεια των υποεκθετικών κατανομών.

Απόδειξη

Θα προσπαθήσουμε αρχικά να πάμε από την πρώτη πρόταση στη δεύτερη.

Για να ανήκει η F^+ στην οικογένεια των υποεκθετικών κατανομών θα πρέπει

$$P\left({\xi^+}_1 + {\xi^+}_2 > x\right) \sim 2P\left({\xi^+}_1 > x\right) \qquad x \to \infty$$

Πριν προχωρήσουμε στην ανάλυση μας καλό θα ήταν να ορίσουμε την τυχαία μεταβλητή ${\xi^+}_i$

$${\xi^+}_i = \begin{cases} 0\ \alpha\nu & -A < \xi_i < 0 \\ \xi_i\ \alpha\nu & \xi_i \geq 0 \end{cases} = \xi_i \mathbb{I}\{\xi_i \geq 0\}$$

Μια πρώτη διαπίστωση είναι πως η τυχαία μεταβλητή

$$\xi^{+}{}_{i} \leq \xi_i + A$$

Έτσι το ενδεχόμενο

$$\{\xi^{+}{}_{1} + \xi^{+}{}_{2} > x\} \subseteq \{\xi_1 + A + \xi_2 + A > x\}$$

Είναι πλέον σαφές πως

$$P(\xi^{+}{}_{1} + \xi^{+}{}_{2} > x) \leq P(\xi_1 + \xi_2 > x - 2A)$$

Όμως

Η F είναι κατανομή με μακριά ουρά και επίσης $\overline{F * F}(x) \sim 2\bar{F}(x)$ $x \to \infty$

Και

$$P\big(\xi^{+}{}_{1} + \xi^{+}{}_{2} > x\big) \leq P(\xi_1 + \xi_2 > x - 2A) = \bar{F}^{*2}(x - 2A)$$

$$P\big(\xi^{+}{}_{1} + \xi^{+}{}_{2} > x\big) \leq 2\bar{F}(x - 2A)$$

Όμως λόγο της ιδιότητας της μακριάς ουράς

$$\frac{\bar{F}(x-2A)}{\bar{F}(x)} \to 1 \qquad x \to \infty$$

Δηλαδή

$$\bar{F}(x - 2A) \sim \bar{F}(x) \qquad x \to \infty$$

Σύμφωνα λοιπόν με τα παραπάνω έχουμε

$$\frac{P\big(\xi^{+}{}_{1} + \xi^{+}{}_{2} > x\big)}{\overline{F^{+}}(x)} \leq 2\frac{\bar{F}(x-2A)}{\overline{F^{+}}(x)}$$

Με

$$\overline{F^{+}}(x) = \begin{cases} \bar{F}(x) & \alpha\nu\ x \geq 0 \\ 0 & \alpha\nu\ x < 0 \end{cases}$$

Τελικά μπορούμε να γράψουμε

$$\frac{P\big(\xi^{+}{}_{1} + \xi^{+}{}_{2} > x\big)}{\overline{F^{+}}(x)} \leq 2\frac{\bar{F}(x-2A)}{\bar{F}(x)}$$

Όμως

$$\frac{\bar{F}(x-2A)}{\bar{F}(x)} \sim \frac{\bar{F}(x-2A)}{\bar{F}(x-2A)} = 1$$

Αποδείξαμε λοιπόν πως

$$\frac{P\big(\xi^{+}{}_{1} + \xi^{+}{}_{2} > x\big)}{\overline{F^{+}}(x)} \leq 2$$

Από την άλλη μεριά όμως

$$P({\xi^+}_1 + {\xi^+}_2 > x) \geq P(\max({\xi^+}_1, {\xi^+}_2) > x)$$

$$P({\xi^+}_1 + {\xi^+}_2 > x) \geq P({\xi^+}_1 > x \text{ ή }, {\xi^+}_2 > x)$$

$$P({\xi^+}_1 + {\xi^+}_2 > x) \geq 2\, P({\xi^+}_1 > x)$$

$$\frac{P({\xi^+}_1 + {\xi^+}_2 > x)}{P({\xi^+}_1 > x)} \geq 2$$

Συμπεραίνουμε λοιπόν πως η κατανομή F^+ της τυχαίας μεταβλητής ξ^+ ανήκει στην οικογένεια των υποεκθετικών κατανομών.

Θα προσπαθήσουμε τώρα να δουλέψουμε αντίστροφα και να πάμε από τη δεύτερη πρόταση στην πρώτη

Υποθέτουμε πως οι κατανομή F^+ που ακολουθεί η τυχαία μεταβλητή ξ^+ ανήκει στην οικογένεια των υποεκθετικών κατανομών .Τότε σύμφωνα με γνωστό λήμμα που αναφέρεται στο βιβλίο του Foss κάθε κατανομή που ορίζεται στον θετικό ημιάξονα $\mathbb{R}^+$ και ανήκει στην οικογένεια των υποεκθετικών κατανομών τότε θα ανήκει και στην οικογένεια κατανομών με βαριά ουρά. Γενικότερα κάθε κατανομή που χαρακτηρίζεται ως υποεκθετική είναι κατανομή με βαριά ουρά.

Η κατανομή F^+ είναι υποεκθετική άρα είναι και κατανομή με μακριά ουρά και σύμφωνα με τον τρόπο που ορίσαμε την F^+ παραπάνω και η F θα είναι κατανομή με μακριά ουρά.

Θα πρέπει επίσης να δείξουμε ότι ισχύει η σχέση

$$\overline{F * F}(x) \sim 2\bar{F}(x)\ x \to \infty$$

Από την μια μεριά έχουμε αποδείξει στο προηγούμενο κεφάλαιο πως αν $F_1, F_2, \dots, F_n$ κατανομές ορισμένες στο $\mathbb{R}$ τέτοιες ώστε η συνάρτηση

$$\bar{F}_1(x) + \bar{F}_2(x) + \ldots + \bar{F}_n(x)$$

Να χαρακτηρίζεται από μακριά ουρά.

Τότε για οποιαδήποτε κατανομή G ισχύει η παρακάτω σχέση

$$\lim_{x\to\infty} inf \frac{\overline{F^{*n} * G}(x)}{\bar{F}_1(x) + \bar{F}_2(x) + \ldots + \bar{F}_n(x)} \geq 1$$

Ειδικότερα η παραπάνω σχέση ισχύει όταν όλες οι συναρτήσεις κατανομής $F_i\ i = 1,2,3, \dots . n$ είναι κατανομές με μακριά ουρά

Για $n = 1\ και\ G = F$ το παραπάνω συμπέρασμα μπορεί να γραφτεί ως εξής

$$\lim_{x\to\infty} inf \frac{\overline{F^{*2}}(x)}{\bar{F}_1(x) + \bar{F}_2(x)} \geq 1$$

Η παραπάνω σχέση είναι ισοδύναμη με

$$\lim_{x\to\infty} inf \frac{\overline{F^{*2}}(x)}{\bar{F}(x)} \geq 2$$

Αρκεί λοιπόν να αποδείξουμε πως

$$\lim_{x\to\infty} inf \frac{\overline{F^{*2}}(x)}{\bar{F}(x)} \leq 2$$

Η τυχαία μεταβλητή

$$\xi^{+}{}_{i} = max\,(0, \xi_i) = \begin{cases} 0\ \alpha\nu & -A < \xi_i < 0 \\ \xi_i\ \alpha\nu & \xi_i \geq 0 \end{cases} = \xi_i \mathbb{I}\{\xi_i \geq 0\}$$

Παρατηρούμε πως πάντα από τον τρόπο που έχει οριστεί η τυχαία μεταβλητή

$$\xi_i \leq \xi^{+}{}_{i}$$

Δηλαδή

$$\xi_1 \leq \xi^{+}{}_{1}$$

και

$$\xi_2 \leq \xi^{+}{}_{2}$$

Προσθέτοντας κατά μέλη τις παραπάνω δύο ανισότητας έχουμε

$$\xi_1 + \xi_2 \leq \xi^{+}{}_{1} + \xi^{+}{}_{2}$$

Σύμφωνα λοιπόν με την παραπάνω ανισότητα είναι πιο πιθανό το άθροισμα $\xi^{+}{}_{1} + \xi^{+}{}_{2}$ να ξεπεράσει ένα θετικό αριθμό σε σχέση με το άθροισμα $\xi_1 + \xi_2$.

$$P(\xi_1 + \xi_2 > x) \leq P(\xi^{+}{}_{1} + \xi^{+}{}_{2} > x)$$

Όμως λόγο υποεκθετικότητας

$$P\left(\xi^{+}{}_{1} + \xi^{+}{}_{2} > x\right) \sim 2P\left(\xi^{+}{}_{1} > x\right) = 2\bar{F}^{+}(x) = 2\bar{F}(x)$$

Έτσι η αρχική ανισότητα μας

$$P(\xi_1 + \xi_2 > x) \leq P(\xi^{+}{}_{1} + \xi^{+}{}_{2} > x)$$

Γράφεται

$$P(\xi_1 + \xi_2 > x) \leq 2\bar{F}(x)$$

$$\frac{\overline{F^{*2}}(x)}{\bar{F}(x)} \leq 2 \qquad x \to \infty$$

Αποδείξαμε λοιπόν πως αν η κατανομή F^{+} της τυχαίας μεταβλητής ξ^{+} ανήκει στην οικογένεια των υποεκθετικών κατανομών τότε η F είναι κατανομή με μακριά ουρά και επίσης $\overline{F * F}(x) \sim 2\bar{F}(x)\ x \to \infty$.

Τέλος θα αποδείξουμε πως όταν η κατανομή F^{+} της τυχαίας μεταβλητής ξ^{+} ανήκει στην οικογένεια των υποεκθετικών κατανομών τότε η δεσμευμένη κατανομή $G(B) = P(\xi \in B/\xi \geq 0)$ ανήκει στην οικογένεια των υποεκθετικών κατανομών.

Η πιθανότητα

$$P\left({\xi^{+}}_{1} + {\xi^{+}}_{2} > x\right) = P(\max(0, \xi_1) + \max(0, \xi_2) > x)$$

$$= P(\xi_1 > x, \xi_2 \leq 0) + P(\xi_2 > x, \xi_1 \leq 0) + P(\xi_1 + \xi_2 > x, \xi_1 > 0, \xi_2 > 0)$$

Οι τυχαίες μεταβλητές όμως ξ_1, ξ_2 είναι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές και έτσι

$$P\left({\xi^{+}}_{1} + {\xi^{+}}_{2} > x\right) =$$

$$P(\xi_1 > x)P(\ \xi_2 \leq 0) + P(\xi_2 > x)P(\ \xi_1 \leq 0) + P(\xi_1 + \xi_2 > x, \xi_1 > 0, \xi_2 > 0)$$

Επίσης η πιθανότητα

$$P(\xi_1 + \xi_2 > x, \xi_1 > 0, \xi_2 > 0) = P\ \xi_1 + \xi_2 > x/\xi_1 > 0, \xi_2 > 0)P(\xi_1 > 0, \xi_2 > 0) =$$

$$= \overline{G^{*2}}(x)P(\xi_2 > 0)P(\ \xi_1 > 0)$$

$$= (1-p)^2\overline{G^{*2}}(x)$$

Όπου

$$p = P(\ \xi_1 \leq 0)$$

Έτσι η συνέλιξη

$$P\left({\xi^{+}}_{1} + {\xi^{+}}_{2} > x\right) = 2p\bar{F}(x) + (1-p)^2\overline{G^{*2}}(x)$$

Έχουμε ορίσει επίσης

$$G(B) = P(\xi \in B/\xi \geq 0)$$

$$\bar{G}(x) = P(\xi > x/\xi \geq 0) = \frac{P(\xi > x, \xi \geq 0)}{P(\xi \geq 0)} = \frac{\bar{F}^{+}(x)}{1-p}$$

Δηλαδή

$$(1-p)\bar{G}(x) = \bar{F}^{+}(x)$$

Όμως

$$\overline{F^{+}}(x) = \begin{cases} \bar{F}(x) & \alpha\nu\ \xi \geq 0 \\ 0 & \alpha\nu\ \xi < 0 \end{cases}$$

Έτσι η συνέλιξη

$$P\left({\xi^{+}}_{1} + {\xi^{+}}_{2} > x\right) = 2p\bar{F}(x) + (1-p)^{2}\overline{G^{*2}}(x)$$

Μπορεί να πάρει τη μορφή

$$P\left({\xi^{+}}_{1} + {\xi^{+}}_{2} > x\right) = 2p\overline{F^{+}}(x) + (1-p)^{2}\overline{G^{*2}}(x)$$

και

$$P\left({\xi^{+}}_{1} + {\xi^{+}}_{2} > x\right) = 2p\bar{F}^{+}(x) + (1-p)^{2}\overline{G^{*2}}(x)$$

Όμως

$$\overline{G^{*2}}(x) = P(\xi_1 + \xi_2 > x/\xi_1 > 0, \xi_2 > 0)$$

Και

$$P(\xi_1 + \xi_2 > x/\xi_1 > 0, \xi_2 > 0) \sim P(\max{(\xi_1, \xi_2)} > x/\xi_1 > 0, \xi_2 > 0) \sim 2P(\xi_1 > x/\xi_1 > 0)$$

Δηλαδή

$$\overline{G^{*2}}(x) \sim 2\bar{G}(x) \qquad x \to \infty$$

Τελικά καταλήγουμε στο συμπέρασμα

$$P\left({\xi^{+}}_{1} + {\xi^{+}}_{2} > x\right) = 2p\overline{F^{+}}(x) + (1-p)^{2}\overline{G^{*2}}(x)$$

$$= 2p\overline{F^{+}}(x) + 2(1-p)^{2}\bar{G}(x)$$

$$= 2p\overline{F^{+}}(x) + 2(1-p)\overline{F^{+}}(x) = 2\overline{F^{+}}(x)$$

7.3 Ορισμός υποεκθετικότητας σε ολόκληρο το $\mathbb{R}$

Το παραπάνω λήμμα μας δίνει τη δυνατότητα να δώσουμε τον ορισμό της υποεκθετικότητας σε ολόκληρο το $\mathbb{R}$. Θα λέμε ότι μια κατανομή F ορισμένη στο $\mathbb{R}$ με την ιδιότητα $\bar{F}(x) > 0$ για κάθε x ότι ανήκει στην οικογένεια των υποεκθετικών κατανομών αν η κατανομή F ανήκει στην οικογένεια κατανομών με μακριά ουρά και επίσης ισχύει

$$\overline{F*F}(x) \sim 2\bar{F}(x)\ x \to \infty$$

Ισοδύναμα μια τυχαία μεταβλητή ξ θα λέμε ότι ακολουθεί κατανομή που ανήκει στην οικογένεια των υποεκθετικών σε ολόκληρο το $\mathbb{R}$ αν η τυχαία μεταβλητή ξ^+ έχει κατανομή που ανήκει στην οικογένεια των υποεκθετικών.

7.4 Βασικά θεωρήματα

Θεώρημα

Έστω μια κατανομή F ορισμένη στον $\mathbb{R}$ η οποία ανήκει στην οικογένεια των κατανομών με μακριά ουρά, και ξ_1, ξ_2 δύο ισόνομες και ανεξάρτητες τυχαίες μεταβλητές που ακολουθούν αυτή την κατανομή. Έστω επίσης μια συνάρτηση h με την ιδιότητα $h(x) \to \infty\ όταν\ x \to \infty$

(η ύπαρξη μιας τέτοιας συνάρτησης κάνει την κατανομή $F\ h - insensitive)\ με\ h(x) \leq \frac{x}{2}$.

Αν ισχύουν όλα τα παραπάνω η κατανομή F ανήκει στην οικογένεια των υποεκθετικών κατανομών αν και μόνο αν

$$P\big(\xi_1 + \xi_2 > x, \xi_1 > h(x), \xi_2 > h(x)\big) = 0\big(\bar{F}(x)\big) \quad x \to \infty$$

Απόδειξη

Αρχικά θα υποθέσουμε πως η συνάρτηση

$$h(x) \leq \frac{\chi}{2}$$

Κάτω από αυτόν τον άτυπο περιορισμό η πιθανότητα

$$\overline{F*F}(x) = P\big(X + \eta > x, X \leq h(x)\big) + P\big(X + \eta > x, \eta \leq h(x)\big) \\ + P\big(X + \eta > x, X > h(x), \eta > h(x)\big)$$

Η παραπάνω σχέση θα μπορούσε να γραφτεί με τη βοήθεια ολοκληρωμάτων αλλά πριν την περιγράψουμε με ολοκληρώματα καλό θα ήταν να αναφέρουμε τα εξής

$$P(X + \eta > x, X \leq h(x)) = \int_{-\infty}^{h(x)} \bar{F}\ (x - y) dF(y)$$

Λογικά αυτή η πιθανότητα θα είναι ίση με

$$P(X + \eta > x, X \leq h(x)) = P\big(X + \eta > x, \eta \leq h(x)\big)$$

Μιας και οι τυχαίες μεταβλητές είναι ισόνομες και ανεξάρτητες.

Έτσι

$$\overline{F*F}(x) = 2\int_{-\infty}^{h(x)} \bar{F}\,(x-y)dF(y) + \; P\big(X+\eta > x, X > h(x), \eta > h(x)\big)$$

Έχουμε ήδη αποδείξει όμως ότι

$$\int_{-\infty}^{h(x)} \bar{F}\,(x-y)dF(y) \sim \bar{F}(x) \quad x \to \infty$$

και έτσι

$$\overline{F*F}(x) \sim 2\bar{F}(x) + P\big(X+\eta > x, X > h(x), \eta > h(x)\big)$$

$$\frac{\overline{F*F}(x)}{\bar{F}(x)} \sim 2 + \frac{P\big(X+\eta > x, X > h(x), \eta > h(x)\big)}{\bar{F}(x)} \quad x \to \infty$$

Και για να ισχύει η υποεκθετικότητα θα πρέπει

$$\frac{\overline{F*F}(x)}{\bar{F}(x)} \sim 2 \quad x \to \infty$$

Συγκρίνοντας τις τελευταίες δύο σχέσεις αντιλαμβανόμαστε πως μια κατανομή ανήκει στην οικογένεια των υποεκθετικών κατανομών όταν και μόνο όταν

$$\frac{P\big(X+\eta > x, X > h(x), \eta > h(x)\big)}{\bar{F}(x)} = 0$$

ή ισοδύναμα

$$P\big(X+\eta > x, X > h(x), \eta > h(x)\big) = 0\big(\bar{F}(x)\big) \quad x \to \infty$$

Η παραπάνω σχέση είναι λογική αν το δούμε από μια άλλη οπτική γωνία

$$P\big(X+\eta > x, X > h(x), \eta > h(x)\big) = \int_{h(x)}^{+\infty} \bar{F}\big(max(x-y, h(x))\big)\,dF(y)$$

Από την άλλη πλευρά

$$P\big(X+\eta > x, X > h(x), \eta > h(x)\big) = P\big(X > x - h(x), \eta > h(x)\big) =$$

$$= P\big(X > x - h(x)\big)P\big(\eta > h(x)\big) = \bar{F}\big(x - h(x)\big)\bar{F}\big(h(x)\big)$$

Όμως η κατανομή F είναι κατανομή με μακριά ουρά και κατά συνέπεια

$$\bar{F}\big(x-h(x)\big)\sim\bar{F}(x) \qquad x\to\infty$$

Από την άλλη μεριά

$$\bar{F}\big(h(x)\big)\to 0 \quad h(x)\to\infty$$

Έτσι καταλήγουμε στο συμπέρασμα πως

$$\int_{h(x)}^{+\infty}\bar{F}\left(max\big(x-y,h(x)\big)\right)dF(y)=\bar{F}\big(x-h(x)\big)\bar{F}\big(h(x)\big)\sim\bar{F}(x)\bar{F}\big(h(x)\big)=0(\bar{F}(x))$$

Θεώρημα

Έστω μια κατανομή F ορισμένη στον $\mathbb{R}$ η οποία ανήκει στην οικογένεια των κατανομών με μακριά ουρά τότε αν η F ανήκει στην οικογένεια των υποεκθετικών κατανομών σε ολόκληρο το $\mathbb{R}$ τότε θα υπάρχει συνάρτηση h με την ιδιότητα $h(x)\to\infty$ $όταν$ $x\to\infty$ με $h(x)\leq\frac{x}{2}$

για την οποία το ολοκλήρωμα

$$\int_{h(x)}^{x-h(x)}\bar{F}(x-y)dF(y)=0\big(\bar{F}(x)\big) \quad x\to\infty$$

Η ύπαρξη μιας τέτοιας συνάρτησης h κάνει την F *h insensitive (h* flat)

Απόδειξη

Αρκεί να γράψουμε την συνέλιξη $\overline{F*F}(x)$ ως εξής

$$\overline{F*F}(x)=\int_{-\infty}^{+\infty}\bar{F}\,(x-y)dF(y)=$$

$$=\int_{-\infty}^{h(x)}\bar{F}(x-y)dF(y)+\int_{h(x)}^{x-h(x)}\bar{F}(x-y)dF(y)+\int_{x-h(x)}^{+\infty}\bar{F}(x-y)dF(y)$$

Έχουμε ήδη αποδείξει πως σε ασυμπτωτικές συνθήκες όταν $x\to\infty$

$$\int_{-\infty}^{h(x)}\bar{F}(x-y)dF(y)\sim\int_{x-h(x)}^{+\infty}\bar{F}(x-y)dF(y)\sim\bar{F}(x)$$

Σύμφωνα λοιπόν με την παραπάνω προσέγγιση

$$\overline{F*F}(x)\sim 2\bar{F}(x)+\int_{h(x)}^{x-h(x)}\bar{F}(x-y)dF(y)$$

Υποεκθετικότητα σημαίνει αυτομάτως

$$\int_{h(x)}^{x-h(x)} \bar{F}(x-y)dF(y) = 0\big(\bar{F}(x)\big) \quad x \to \infty$$

Λήμμα

Υποθέστε ότι οι κατανομές F και G είναι ορισμένες στον $\mathbb{R}$ και είναι τέτοιες ώστε το άθροισμα των ουρών τους $\bar{F} + \bar{G}$ να είναι κατανομή με μακριά ουρά. Έστω επίσης ότι υπάρχει μια συνάρτηση h με την ιδιότητα $h(x) \to \infty$ $όταν$ $x \to \infty$.Η ύπαρξη της συνάρτησης h κάνει την $\bar{F} + \bar{G}$ *h-insensitive* .

Όταν όλα τα παραπάνω ισχύουν τότε

$$\int_{-\infty}^{h(x)} \bar{G}\,(x-y)dF(y) + \int_{-\infty}^{h(x)} \bar{F}\,(x-y)dG(y) \sim \bar{F}(x) + \bar{G}(x) \quad x \to \infty$$

Απόδειξη

Αρχικά παρατηρούμε πως

$$\int_{-\infty}^{h(x)} \bar{G}\,(x-y)dF(y) \leq \bar{G}(x-h(x))$$

Η παραπάνω παρατήρηση μπορεί να προκύψει ή με ανάλυση της πιθανότητας που κρύβεται πίσω από το ολοκλήρωμα ή με απλές πράξεις. Εμείς θα δοκιμάσουμε και τα δύο

$$\int_{-\infty}^{h(x)} \bar{G}\,(x-y)dF(y) = P(X+\eta > x, X \leq h(x))$$

Παρατηρούμε πως αν η X πάρει την μέγιστη τιμή της τότε η διαφορά $x - X$ θα πάρει την ελάχιστη τιμή της για x θετική σταθερά. Και έτσι

$$P\big(X+\eta > x, X \leq h(x)\big) \leq P\big(\eta > x - h(x), X \leq h(x)\big) \leq P(\eta > x - h(x))$$

Καταλήξαμε λοιπόν στο συμπέρασμα πως

$$\int_{-\infty}^{h(x)} \bar{G}\,(x-y)dF(y) \leq P\big(\eta > x - h(x)\big) = \bar{G}(x-h(x))$$

Η παραπάνω παρατήρηση προκύπτει με στοιχειώδεις πράξεις ως εξής

$$\int_{-\infty}^{h(x)} \bar{G}\,(x-y)dF(y) = [\bar{G}(x-y)F(y)]_{-\infty}^{h(x)} - \int_{-\infty}^{h(x)} F(y)d\bar{G}(x-y)$$

$$\leq [\bar{G}(x-y)F(y)]_{-\infty}^{h(x)} \leq \bar{G}(x-h(x))$$

Αντίστοιχα

$$\int_{-\infty}^{h(x)} \bar{F}\,(x-y)dG(y) \leq \bar{F}(x-h(x))$$

Από την άλλη πλευρά

$$\int_{-\infty}^{h(x)} \bar{G}\,(x-y)dF(y) \geq \int_{-h(x)}^{h(x)} \bar{G}\,(x-y)dF(y)$$

Όμως

$$\int_{-h(x)}^{h(x)} \bar{G}\,(x-y)dF(y) = P(X+\eta > x, -h(x) < X < h(x))$$

Στην ουσία το παραπάνω ολοκλήρωμα ισούται με την πιθανότητα του ενδεχομένου

$$\{X+\eta > x, -h(x) < X < h(x)\}$$

Δηλαδή η τυχαία μεταβλητή X μπορεί να πάρει και θετικές αλλά και αρνητικές τιμές καταλήγουμε λοιπόν στο συμπέρασμα πως

$$\{X+\eta > x, -h(x) < X < h(x)\} \supseteq$$

$$\{X+\eta > x, X \in (-h(x), 0)\} \cup \{X+\eta > x, X \in (0, h(x))\}$$

Αν δηλαδή το X είναι αρνητικό τότε ο μόνος τρόπος για να πραγματοποιηθεί το ενδεχόμενο

$$\{X+\eta > x, X \in (-h(x), 0)\}$$

Είναι η τυχαία μεταβλητή η να πάρει τιμές μεγαλύτερες από $x + h(x)$

$$\eta > x + h(x)$$

Έτσι το ενδεχόμενο

$$\{X+\eta > x, X \in (-h(x), 0)\}$$

Είναι ισοπίθανο με το ενδεχόμενο

$$\{\eta > x + h(x), -h(x) < X \leq 0\}$$

Αντίστοιχα όταν η τυχαία μεταβλητή X είναι αρνητική τότε ο μόνος τρόπος για να πραγματοποιηθεί το ενδεχόμενο

$$\{X+\eta>x, X\in(0,h(x))\}$$

Είναι η τυχαία μεταβλητή η να πάρει τιμές μεγαλύτερες από $x-h(x)$

Δηλαδή

$$\eta>x-h(x)$$

Έτσι το ενδεχόμενο

$$\{X+\eta>x, X\in(0,h(x))\}$$

Είναι ισοπίθανο με το ενδεχόμενο

$$\{\eta>x-h(x), h(x)0<X<h(x)\}$$

Σύμφωνα λοιπόν με τα παραπάνω η πιθανότητα

$$P\big(X+\eta>x, -h(x)<X<h(x)\big)\geq$$

$$P(\eta>x+h(x), -h(x)<X\leq 0)+P(\eta>x-h(x), 0<X<h(x))$$

Δηλαδή

$$P\big(X+\eta>x, -h(x)<X<h(x)\big)\geq$$

$$\bar{G}\big(x+h(x)\big)P(-h(x)<X<0)+\bar{G}\big(x-h(x)\big)P(0<X<h(x))\geq$$

$$\geq \bar{G}\big(x+h(x)\big)P(-h(x)<X<0)+\bar{G}\big(x+h(x)\big)P(0<X<h(x))$$

$$=\bar{G}\big(x+h(x)\big)(P(-h(x)<X\leq 0)+P(0<X<h(x)))$$

$$=\bar{G}\big(x+h(x)\big)\left(\int_{-h(x)}^{0}dF(x)+\int_{0}^{h(x)}dF(x)\right)=$$

$$=\bar{G}\big(x+h(x)\big)\int_{-h(x)}^{h(x)}dF(x)$$

Και όταν $x\to\infty$ τότε $h(x)\to\infty$ και το ολοκλήρωμα

$$\int_{-h(x)}^{h(x)} dF(x) = 1$$

Καταλήξαμε λοιπόν στο συμπέρασμα

$$\int_{-\infty}^{h(x)} \bar{G}\,(x-y)dF(y) \geq \int_{-h(x)}^{h(x)} \bar{G}\,(x-y)dF(y) = P(X+\eta > x, -h(x) < X < h(x))$$

$$\geq \bar{G}\big(x+h(x)\big)\int_{-h(x)}^{h(x)} dF(x) \sim \bar{G}\big(x+h(x)\big)$$

Τελικά στην ουσία προέκυψε η διπλή ανισότητα

$$\bar{G}\big(x+h(x)\big) \leq \int_{-\infty}^{h(x)} \bar{G}\,(x-y)dF(y) \leq \bar{G}(x-h(x))$$

Αντίστοιχα εργαζόμαστε για το ολοκλήρωμα

$$\int_{-\infty}^{h(x)} \bar{F}\,(x-y)dG(y)$$

Και καταλήγουμε στην διπλή ανισότητα

$$\bar{F}\big(x+h(x)\big) \leq \int_{-\infty}^{h(x)} \bar{F}\,(x-y)dG(y) \leq \bar{F}(x-h(x))$$

Σύμφωνα λοιπόν με όσα έχουμε αναφέρει παραπάνω το άθροισμα των ολοκληρωμάτων

$$\int_{-\infty}^{h(x)} \bar{G}\,(x-y)dF(y) + \int_{-\infty}^{h(x)} \bar{F}\,(x-y)dG(y)$$

Θα φράσσεται από κάτω και από πάνω ως εξής

$$Q_1(x) \leq \int_{-\infty}^{h(x)} \bar{G}\,(x-y)dF(y) + \int_{-\infty}^{h(x)} \bar{F}\,(x-y)dG(y) \leq Q_2(x)$$

Όπου

$$Q_1(x) = \bar{G}\big(x + h(x)\big) + \bar{F}\big(x + h(x)\big)$$

$$Q_2(x) = \bar{G}\big(x - h(x)\big) + \bar{F}\big(x - h(x)\big)$$

και όταν $x \to \infty$

έχοντας υποθέσει από την αρχή του λήμματος ότι η $\bar{F} + \bar{G}$ να είναι κατανομή με μακριά ουρά

$$\bar{G}\big(x + h(x)\big) + \bar{F}\big(x + h(x)\big) \sim \bar{F}(x) + \bar{G}(x)$$

Και

$$\bar{G}\big(x - h(x)\big) + \bar{F}\big(x - h(x)\big) \sim \bar{F}(x) + \bar{G}(x)$$

Καταλήγουμε λοιπόν στο συμπέρασμα πως

$$\int_{-\infty}^{h(x)} \bar{G}\,(x-y)dF(y) + \int_{-\infty}^{h(x)} \bar{F}\,(x-y)dG(y) \sim \bar{F}(x) + \bar{G}(x) \;\; x \to \infty$$

Λήμμα

Υποθέστε ότι οι κατανομές F και G είναι ορισμένες στον $\mathbb{R}$ και είναι τέτοιες ώστε το άθροισμα των ουρών τους $\bar{F} + \bar{G}$ να είναι κατανομή με μακριά ουρά. Έστω επίσης ότι υπάρχει μια συνάρτηση h με την ιδιότητα $h(x) \to \infty\ όταν\ x \to \infty$.Η ύπαρξη της συνάρτησης h κάνει την $\bar{F} + \bar{G}$ *h-insensitive* .

Όταν όλα τα παραπάνω ισχύουν τότε

$$\int_{x-h(x)}^{+\infty} \bar{G}\,(x-y)dF(y) + \int_{x-h(x)}^{+\infty} \bar{F}\,(x-y)dG(y) \sim \bar{F}(x) + \bar{G}(x) \;\; x \to \infty$$

Απόδειξη

Αρκεί να δουλέψουμε για το ένα από τα δύο ολοκληρώματα και το άλλο θα προκύψει ακριβώς με τον ίδιο τρόπο.

Θα δουλέψουμε με το πρώτο ολοκλήρωμα

$$\int_{x-h(x)}^{+\infty} \bar{G}\,(x-y)dF(y) = P\big(\xi + n > x, \xi > x - h(x)\big) =$$

$$= P\big(n > h(x), \xi > x - h(x)\big) =$$

$$= P\big(n > h(x)\big)P\big(\xi > x - h(x)\big) = \bar{G}\big(h(x)\big)\bar{F}\big(x - h(x)\big)$$

Θεωρούμε πως η συνάρτηση h είναι μια συνάρτηση με τις εξής ιδιότητες

$$h(x) \to \infty \qquad x \to \infty$$

$$h(x) < \frac{x}{2}$$

Από την δεύτερη ιδιότητα που έχουμε απαιτήσει για την απόδειξη του λήμματος παρατηρούμε το εξής

Η πιθανότητα $P\big(n > h(x)\big)$ ισούται

$$P\big(n > h(x)\big) = \int_{h(x)}^{+\infty} g(x)dx \geq \int_{\frac{x}{2}}^{+\infty} g(x)dx$$

Είναι στην ουσία ένα εμβαδό που αρχίζει από ένα σημείο $h(x) < \frac{x}{2}$

Από την άλλη μεριά η πιθανότητα $P\big(n > x - h(x)\big)$ είναι ίση με το ολοκλήρωμα

$$P\big(n > x - h(x)\big) = \int_{x-h(x)}^{+\infty} g(x)dx$$

Όμως για $x > 0$ η ποσότητα

$$x - h(x) > \frac{x}{2}$$

Πρόκειται στην ουσία για ένα εμβαδό που ξεκινάει από ένα σημείο μετά το $\frac{x}{2}$

Έτσι η πιθανότητα

$$P\big(n > x - h(x)\big) \leq P\big(n > h(x)\big)$$

Ύστερα από αυτή την διαπίστωση

$$\int_{x-h(x)}^{+\infty} \bar{G}\,(x-y)dF(y) = P\big(n > h(x)\big)P\big(\xi > x - h(x)\big) \leq P\big(n > x - h(x)\big)$$

$$\leq \bar{G}\big(x - h(x)\big)$$

Από την άλλη πλευρά θα παρατηρούμε πως το ολοκλήρωμα

$$\int_{x-h(x)}^{+\infty} \bar{G}\,(x-y)dF(y) = P\big(\xi + n > x, \xi > x - h(x)\big) =$$

$$= P\big(n > h(x), \xi > x - h(x)\big) =$$

$$P\big(n > h(x)\big)P(\xi > x - h(x)) \geq P\big(n > x + h(x)\big)P(\xi > x - h(x))$$

Όμως η πιθανότητα

$$P\big(\xi > x - h(x)\big) = \int_{x-h(x)}^{+\infty} f(z)dz$$

Εκφράζει ένα εμβαδό από το $x - h(x)$ έως το $+\infty$ και θεωρώ λογικό να υποθέσουμε πως αυτό το εμβαδό θα είναι μεγαλύτερο ή ίσο από το εμβαδό που αρχίζει από το – $(x - h(x))$ και φτάνει μέχρι το $x + h(x)$.

Δηλαδή υποθέτουμε πως

$$\int_{x-h(x)}^{+\infty} f(z)dz \geq \int_{-(x-h(x))}^{x-h(x)} f(z)dz$$

Όμως όπως έχουμε ήδη αναφέρει αυτά τα εμβαδά εκφράζουν πιθανότητες και αντίστοιχα

$$P\big(\xi > x - h(x)\big) \geq P\left(-\big(x - h(x)\big) \leq \xi \leq x - h(x)\right) = F(-\big(x - h(x)\big), x - h(x))$$

Τελικά προκύπτει πως

$$\int_{x-h(x)}^{+\infty} \bar{G}\,(x-y)dF(y) \geq P\big(n > x + h(x)\big)F(-\big(x - h(x)\big), x - h(x))$$

Και όταν $x \to \infty$

$$\int_{-(x-h(x))}^{x-h(x)} f(z)dz = 1$$

Και

$$\int_{x-h(x)}^{+\infty} \bar{G}\,(x-y)dF(y) \geq \bar{G}\big(x + h(x)\big)$$

Στην ουσία έχουμε δημιουργήσει μια διπλή ανισότητα που είναι η

$$\bar{G}\big(x + h(x)\big) \leq \int_{x-h(x)}^{+\infty} \bar{G}\,(x-y)dF(y) \leq \bar{G}\big(x - h(x)\big)$$

Δουλεύοντας αντίστοιχα για το δεύτερο ολοκλήρωμα καταλήγουμε στην διπλή ανισότητα

$$\bar{F}\big(x+h(x)\big) \le \int_{x-h(x)}^{+\infty} \bar{F}\,(x-y)dG(y) \le \bar{F}\big(x-h(x)\big)$$

Προσθέτοντας κατά μέλη τις δύο ανισότητες έχουμε

$$\sum_{i=1}^{2} \bar{F}_i\,(x+h(x)) \le \int_{x-h(x)}^{+\infty} \bar{F}\,(x-y)dG(y) + \int_{x-h(x)}^{+\infty} \bar{G}\,(x-y)dF(y) \le \sum_{i=1}^{2} \bar{F}_i\,(x-h(x))$$

Όπου

$$\sum_{i=1}^{2} \bar{F}_i\,\big(x+h(x)\big) = \bar{G}\big(x+h(x)\big) + \bar{F}\big(x+h(x)\big)$$

Και

$$\sum_{i=1}^{2} \bar{F}_i\,\big(x-h(x)\big) = \bar{G}\big(x-h(x)\big) + \bar{F}\big(x-h(x)\big)$$

Υποθέσαμε στο λήμμα ότι οι κατανομές F και G είναι ορισμένες στον $\mathbb{R}$ και είναι τέτοιες ώστε το άθροισμα των ουρών τους $\bar{F}+\bar{G}$ να είναι κατανομή με μακριά ουρά

Αυτό σημαίνει πως

$$\bar{G}\big(x+h(x)\big) + \bar{F}\big(x+h(x)\big) \sim \bar{G}(x) + \bar{F}(x) \quad x \to \infty$$

Και

$$\bar{G}\big(x-h(x)\big) + \bar{F}\big(x-h(x)\big) \sim \bar{G}(x) + \bar{F}(x) \quad x \to \infty$$

Και αποδείξαμε πως το άθροισμα

$$\int_{x-h(x)}^{+\infty} \bar{G}\,(x-y)dF(y) + \int_{x-h(x)}^{+\infty} \bar{F}\,(x-y)dG(y) \sim \bar{F}(x) + \bar{G}(x) \;\; x \to \infty$$

Στο ίδιο συμπέρασμα θα καταλήγαμε αν υποθέταμε πως τόσο η κατανομή G όσο και η κατανομή F είναι κατανομές με μακριά ουρά πιο συγκεκριμένα

Αν τώρα υποθέσουμε ότι η κατανομή G ανήκει στην οικογένεια κατανομών με μακριά ουρά τότε

$$\lim_{x\to\infty} \bar{G}\big(x+h(x)\big) = \lim_{x\to\infty} \bar{G}\big(x-h(x)\big) = \bar{G}(x)$$

Και

$$\int_{x-h(x)}^{+\infty} \bar{G}\,(x-y)dF(y)\sim\bar{G}(x) \quad x\to\infty$$

Αντίστοιχα εργαζόμενοι για το ολοκλήρωμα

$$\int_{x-h(x)}^{+\infty} \bar{F}\,(x-y)dG(y)\sim\bar{F}(x) \quad x\to\infty$$

Και αποδείξαμε πως το άθροισμα

$$\int_{x-h(x)}^{+\infty} \bar{G}\,(x-y)dF(y)+\int_{x-h(x)}^{+\infty} \bar{F}\,(x-y)dG(y)\sim\bar{F}(x)+\bar{G}(x)\ \ x\to\infty$$

Σημαντικό είναι λοιπόν να σημειώσουμε πως το άθροισμα $\bar{F}+\bar{G}$ έχει μακριά ουρά όταν ισχύουν τα εξής

- η κατανομή G όσο και η κατανομή F είναι κατανομές με μακριά ουρά πιο συγκεκριμένα
- $H\ F$ είναι κατανομή με μακριά ουρά και ισχύει

$$\bar{G}(x)=0\big(\bar{F}(x)\big) \quad x\to\infty$$

7.5 Subexponentiality and weak tail-equivalence

Ορισμός

Δύο κατανομές G και F μη φραγμένες από δεξιά δηλαδή με την ιδιότητα $\bar{G}(x)>0$, $\bar{F}(x)>0$, αντίστοιχα θα ονομάζονται weakly tail equivalence αν υπάρχουν θετικές σταθερές c_1, c_2 με $c_1>0\ και\ c_2<\infty$ τέτοιες ώστε για κάθε $x>0$

$$c_1\leq\frac{\bar{F}(x)}{\bar{G}(x)}\leq c_2$$

Η παραπάνω ανισότητα είναι ισοδύναμη με τη συνθήκη

$$0<\lim_{x\to\infty} inf\frac{\bar{F}(x)}{\bar{G}(x)}\leq\lim_{x\to\infty} sup\frac{\bar{F}(x)}{\bar{G}(x)}<\infty$$

Λήμμα

Έστω h μια αύξουσα συνάρτηση ορισμένη στον $\mathbb{R}_+$τέτοια ώστε $h(x) \to \infty$.Τότε για οποιαδήποτε κατανομή F_1, F_1, G_1, G_2 στον $\mathbb{R}$ ισχύει

$$\lim_{x\to\infty} sup \frac{P(\xi_1 + \eta_1 > x, \xi_1 > h(x), \eta_1 > h(x))}{P(\xi_2 + \eta_2 > x, \xi_2 > h(x), \eta_2 > h(x)} \leq \lim_{x\to\infty} sup \frac{\bar{F}_1(x)}{\bar{F}_2(x)} \lim_{x\to\infty} sup \frac{\bar{G}_1(x)}{\bar{G}_2(x)}$$

Όπου$\xi_{1,}\xi_2, \eta_{1,}\ \eta_2$ ανεξάρτητες και ισόνομες τυχαίες μεταβλητές που ακολουθούν τις κατανομές F_1, F_2, G_1, G_2 αντίστοιχα.

Απόδειξη

Γνωρίζουμε πως

$$P\big(\xi_1 + \eta_1 > x, \xi_1 > h(x), \eta_1 > h(x)\big) = \int_{h(x)}^{+\infty} \overline{F_1}\big(max(x-y, h(x))\big)\, dG_1(y)$$

$$\leq \sup_{z>h(x)} \frac{\overline{F_1}(z)}{\overline{F_2}(z)} \int_{h(x)}^{+\infty} \overline{F_2}\left(max\big(x-y, h(x)\big)\right) dG_1(y) =$$

Όμως το ολοκλήρωμα

$$\int_{h(x)}^{+\infty} \overline{G_1}\left(max\big(x-y, h(x)\big)\right) dF_2(y) = \int_{h(x)}^{+\infty} \overline{G_1}\left(max\big(x-y, h(x)\big)\right) dF_2(y)$$

$$\leq \sup_{z>h(x)} \frac{\overline{G_1}(z)}{\overline{G_2}(z)} \int_{h(x)}^{+\infty} \overline{G_2}\left(max\big(x-y, h(x)\big)\right) dF_2(y)$$

Έτσι

$$P\big(\xi_1 + \eta_1 > x, \xi_1 > h(x), \eta_1 > h(x)\big)$$
$$\leq \sup_{z>h(x)} \frac{\overline{F_1}(z)}{\overline{F_2}(z)} \sup_{z>h(x)} \frac{\overline{G_1}(z)}{\overline{G_2}(z)} \int_{h(x)}^{+\infty} \overline{G_2}\left(max\big(x-y, h(x)\big)\right) dF_2(y)$$

$$\leq \sup_{z>h(x)} \frac{\overline{F_1}(z)}{\overline{F_2}(z)} \sup_{z>h(x)} \frac{\overline{G_1}(z)}{\overline{G_2}(z)} P\big(\xi_2 + \eta_2 > x, \xi_2 > h(x), \eta_2 > h(x)\big)$$

Έτσι ο λόγος

$$\frac{P\big(\xi_1 + \eta_1 > x, \xi_1 > h(x), \eta_1 > h(x)\big)}{P\big(\xi_2 + \eta_2 > x, \xi_2 > h(x), \eta_2 > h(x)\big)} \leq \sup_{z>h(x)} \frac{\overline{F_1}(z)}{\overline{F_2}(z)} \sup_{z>h(x)} \frac{\overline{G_1}(z)}{\overline{G_2}(z)}$$

Και για $x \to \infty$

$$\lim_{x\to\infty} sup \frac{P(\xi_1 + \eta_1 > x, \xi_1 > h(x), \eta_1 > h(x))}{P(\xi_2 + \eta_2 > x, \xi_2 > h(x), \eta_2 > h(x)} \leq \lim_{x\to\infty} sup \frac{\bar{F}_1(x)}{\bar{F}_2(x)} \lim_{x\to\infty} sup \frac{\bar{G}_1(x)}{\bar{G}_2(x)}$$

Θεώρημα

Έστω F_1, F_2, G είναι κατανομές ορισμένες στον $\mathbb{R}$ έτσι ώστε $\overline{F_1}(x) \sim \bar{F}_2(x) \quad x \to \infty$. Αν η κατανομή , G είναι κατανομή με μακριά ουρά τότε ισχύει

$$\overline{F_1 * G}(x) \sim \overline{F_2 * G}(x) \quad x \to \infty.$$

Απόδειξη

$$\overline{F_1 * G}(x) = \int_{-\infty}^{+\infty} \bar{F}_1(x-y)dG(y) =$$

$$\int_{-\infty}^{x-h(x)} \bar{F}_1(x-y)dG(y) + \int_{x-h(x)}^{+\infty} \bar{F}_1(x-y)dG(y)$$

Και επειδή η G είναι κατανομή με μακριά ουρά το ολοκλήρωμα

$$\int_{x-h(x)}^{+\infty} \bar{F}_1(x-y)dG(y) \sim \bar{G}(x) \quad x \to \infty$$

Έτσι καταλήγουμε στη σχέση

$$\overline{F_1 * G}(x) \sim \int_{-\infty}^{x-h(x)} \bar{F}_1(x-y)dG(y) + \bar{G}(x) \; x \to \infty$$

Αντίστοιχα

$$\overline{F_2 * G}(x) = \int_{-\infty}^{+\infty} \bar{F}_2(x-y)dG(y) =$$

=

$$\int_{-\infty}^{x-h(x)} \bar{F}_2(x-y)dG(y) + \int_{x-h(x)}^{+\infty} \bar{F}_2(x-y)dG(y)$$

Έτσι καταλήγουμε στη σχέση

$$\overline{F_2 * G}(x) \sim \int_{-\infty}^{x-h(x)} \bar{F}_2(x-y)dG(y) + \bar{G}(x) \; x \to \infty$$

Αν τώρα ισχύει η ιδιότητα $\bar{F}_1(x) \sim \bar{F}_2(x) \quad x \to \infty$ τότε

$$\bar{F}_1(x-y) \sim \bar{F}_2(x-y) \qquad x \to \infty$$

Και αντίστοιχα

$$\int_{-\infty}^{x-h(x)} \bar{F}_2(x-y)dG(y) \sim \int_{-\infty}^{x-h(x)} \bar{F}_2(x-y)dG(y) \qquad x \to \infty$$

Καταλήγουμε λοιπόν στο συμπέρασμα πως

$$\int_{-\infty}^{x-h(x)} \bar{F}_1(x-y)dG(y) + \bar{G}(x) \sim \int_{-\infty}^{x-h(x)} \bar{F}_2(x-y)dG(y) + \bar{G}(x) \; x \to \infty$$

Θεώρημα

Έστω κατανομές F_1, F_2, G_1, G_2 ορισμένες στον $\mathbb{R}$ τέτοιες ώστε $\overline{F_1}(x) \sim \bar{F}_2(x)$ και $\overline{G_1}(x) \sim \bar{G}_2(x)$ όταν $x \to \infty$ Αν η συνάρτηση ώστε $\overline{F_1} + \overline{G_1}$ είναι κατανομή με μακριά ουρά τότε

$$\overline{F_1 * G_1}(x) \sim \overline{F_2 * G_2}(x) \qquad x \to \infty$$

Απόδειξη

Αρχικά το θεώρημα έχει ως βασική συνθήκη ότι

$$\overline{F_1}\big(x-h(x)\big) + \overline{G_1}\big(x+h(x)\big) \sim \overline{F_1}(x) + \overline{G_1}(x) \quad x \to \infty$$

Ας υπολογίσουμε αρχικά την συνέλιξη

$$\overline{F_1 * G_1}(x) = \int_{-\infty}^{+\infty} \overline{F_1}(x-y)dG_1(y) =$$

$$= \int_{-\infty}^{h(x)} \overline{F_1}(x-y)dG_1(y) + \int_{-\infty}^{h(x)} \overline{G_1}(x-y)dF_1(y) + \int_{h(x)}^{+\infty} \overline{F_1}(\max{(h(x),(x-y))}dG_1(y)$$

Όμως για $x \to \infty$ $\overline{F_1}(x) \sim \bar{F}_2(x)$ και $\overline{G_1}(x) \sim \bar{G}_2(x)$ και κατά συνέπεια

$$\int_{-\infty}^{h(x)} \overline{F_1}(x-y)dG_1(y) \sim \int_{-\infty}^{h(x)} \overline{F_2}(x-y)dG_2(y)$$

Επίσης

$$\int_{-\infty}^{h(x)} \overline{G_1}(x-y)dF_1(y) \sim \int_{-\infty}^{h(x)} \overline{G_2}(x-y)dF_2(y)$$

Επίσης

$$\int_{h(x)}^{+\infty} \overline{F_1}(\max{(h(x),(x-y))}dG_1(y) \sim \int_{h(x)}^{+\infty} \overline{F_2}(\max{(h(x),(x-y))}dG_2(y)$$

Προσθέτοντας κατά μέλη τις τρείς σχέσεις παραπάνω έχουμε

$$\overline{F_1 * G_1}(x) \sim \overline{F_2 * G_2}(x) \qquad x \to \infty$$

Λήμμα

Έστω $F \; και \; G$ δύο κατανομές στον $\mathbb{R}_+$ τέτοιες ώστε η πρώτη να μην είναι δεξιά φραγμένη ($\bar{F}(x) > 0$ για κάθε x και η δεύτερη είναι μη εκφυλισμένη στο μηδέν

Εάν ισχύει $\bar{G}(x) \leq c\bar{F}(x)$

Και επιπλέον

$$\lim_{x\to\infty} sup \frac{\overline{F * G}(x)}{\bar{F}(x) + \bar{G}(x)} \leq 1$$

Τότε η F είναι κατανομή με μακριά ουρά.

Απόδειξη

Θα υπολογίσουμε αρχικά τη συνέλιξη

$$\overline{F * G}(x) = P(\xi + \eta > x) = P(\xi + \eta > x, \eta \leq x) + P(\xi + \eta > x, \eta > x)$$

Η πιθανότητα

$$P(\xi + \eta > x, \eta \leq x) = \int_0^x \bar{F}\,(x-y)dG(y)$$

Και η πιθανότητα

$$P(\xi + \eta > x, \eta > x) = \int_x^{+\infty} \bar{F}(x-y)dG(y)$$

Η πιθανότητα γράφεται

$$P(\xi + \eta > x, \eta > x) = P(\xi + \eta > x/\eta > x)P(\eta > x) = \bar{G}(x)$$

Με αυτή λοιπόν την παρατήρηση έχουμε

$$\overline{F * G}(x) = \int_0^x \bar{F}\,(x-y)dG(y) + \bar{G}(x)$$

Από την άλλη μεριά

$$\overline{F * G}(x) = P(\xi + \eta > x) \geq P(\xi > x) + P(\eta > x)$$
$$= \bar{G}(x) + \bar{F}(x) + 0(\bar{G}(x) + \bar{F}(x))$$

Το λήμμα μας δίνει και τη σχέση

$$\lim_{x \to \infty} sup \frac{\overline{F * G}(x)}{\bar{F}(x) + \bar{G}(x)} \leq 1$$

Από την οποία μπορούμε να συμπεράνουμε πως

$$\overline{F * G}(x) \leq \bar{F}(x) + \bar{G}(x) + 0(\bar{F}(x) + \bar{G}(x))$$

Καταλήγουμε λοιπόν στο συμπέρασμα πως

$$\overline{F * G}(x) = \bar{F}(x) + \bar{G}(x) + 0\big(\bar{F}(x) + \bar{G}(x)\big) = \bar{F}(x) + \bar{G}(x) + 0\big(\bar{F}(x)\big)$$

Ισοδύναμα η παραπάνω σχέση γράφεται

$$\int_0^x \bar{F}\,(x - y)dG(y) + \bar{G}(x) = \bar{F}(x) + \bar{G}(x) + 0\big(\bar{F}(x)\big)$$

Δηλαδή

$$\int_0^x \bar{F}\,(x - y)dG(y) = \bar{F}(x) + 0\big(\bar{F}(x)\big)$$

Για να είναι η κατανομή F είναι κατανομή με μακριά ουρά θα πρέπει

$$\frac{\bar{F}(x - y)}{\bar{F}(x)} \to 1 \qquad x \to \infty$$

Ισοδύναμη της παραπάνω σχέσης είναι η

$$\frac{\bar{F}(x - y) - \bar{F}(x)}{\bar{F}(x)} \to 0 \quad x \to \infty$$

Δηλαδή

$$\bar{F}(x - y) - \bar{F}(x) = 0\big(\bar{F}(x)\big) \;\; x \to \infty$$

Και ολοκληρώνοντας από μηδέν έως x

$$\int_0^x (\bar{F}(x - y) - \bar{F}(x))dG(y) = 0\big(\bar{F}(x)\big)$$

Σύμφωνα με τα παραπάνω αν στη σχέση

$$\int_0^x \bar{F}\,(x-y)dG(y) = \bar{F}(x) + 0\big(\bar{F}(x)\big)$$

Αφαιρέσουμε και από τα δύο μέλη τη ποσότητα

$$\int_0^x \bar{F}(x)dG(y) = \bar{F}(x)G(x)$$

Φτιάχνουμε τη σχέση

$$\int_0^x \bar{F}\,(x-y)dG(y) - \int_0^x \bar{F}(x)dG(y) = \bar{F}(x) - \bar{F}(x)G(x) + 0\big(\bar{F}(x)\big)$$

.

$$\int_0^x (\bar{F}(x-y) - \bar{F}(x))dG(y) = \bar{F}(x)\big(1 - G(x)\big) + 0\big(\bar{F}(x)\big) = 0\big(\bar{F}(x)\big)$$

Έτσι δείξαμε πως η κατανομή Fείναι κατανομή με μακριά ουρά

Λήμμα

Έστω δύο κατανομές G και F οι οποίες είναι weakly tail equivalence ορισμένες στον $\mathbb{R}$. Υποθέστε επίσης ότι ή και οι δύο είναι κατανομές με μακριά ουρά ή και οι δύο είναι επικεντρωμένες στον θετικό ημιάξονα $\mathbb{R}_+$ και επιπλέον ισχύει

$$\lim_{x\to\infty} sup \frac{\overline{F*G}(x)}{\bar{F}(x) + \bar{G}(x)} \leq 1$$

Τότε και οι δύο κατανομές G και F ανήκουν στην οικογένεια των υποεκθετικών κατανομών

Απόδειξη

Θα πρέπει να δείξουμε πως και οι δύο κατανομές ανήκουν στην οικογένεια των υποεκθετικών κατανομών.

Όταν οι κατανομές ορίζονται στον $\mathbb{R}$ για την ύπαρξη υποεκθετικότητας απαιτούνται δύο συνθήκες

- Και οι δύο κατανομές να είναι κατανομές με μακριά ουρά
- $\overline{F*G}(x) \sim \bar{F}(x) + \bar{G}(x) \quad x \to \infty$

Ας ασχοληθούμε αρχικά με τη δεύτερη συνθήκη

$$\overline{F*G}(x) = P\big(\xi+\eta > x, \xi \le h(x)\big) + P\big(\xi+\eta > x, \eta \le h(x)\big)$$

$$+P\big(\xi+\eta > x, \xi > h(x), \eta > h(x)\big)$$

Όμως χρησιμοποιώντας τις σχέσεις που ήδη έχουμε αποδείξει ότι ισχύουν

$$P\big(\xi+\eta > x, \xi \le h(x)\big) = \int_{-\infty}^{h(x)} \bar{G}(x-y)dF(y) \sim \bar{G}(x) \quad x \to \infty$$

$$P\big(\xi+\eta > x, \eta \le h(x)\big) = \int_{-\infty}^{h(x)} \bar{F}(x-y)dG(y) \sim \bar{F}(x) \quad x \to \infty$$

$$P\big(\xi+\eta > x, \xi > h(x), \eta > h(x)\big) = \int_{h(x)}^{+\infty} \bar{G}((\max{(h(x), x-y)}dF(y) =$$

$$= \int_{h(x)}^{+\infty} \bar{F}((\max{(h(x), x-y)}GF(y)$$

Μπορούμε να γράψουμε πλέον την συνέλιξη

$$\overline{F*G}(x) = \bar{G}(x) + \bar{F}(x) + \int_{h(x)}^{+\infty} \bar{F}((\max{(h(x), x-y)}GF(y) \quad x \to \infty$$

Και έτσι η παραπάνω σχέση μπορεί να γραφτεί ως εξής

$$\frac{\overline{F*G}(x)}{\bar{G}(x)+\bar{F}(x)} = 1 + \frac{\int_{h(x)}^{+\infty} \bar{F}((\max{(h(x), x-y)}GF(y)}{\bar{G}(x)+\bar{F}(x)} \quad x \to \infty$$

Η παραπάνω σχέση σε συνδυασμό με

$$\lim_{x\to\infty} sup \frac{\overline{F*G}(x)}{\bar{F}(x)+\bar{G}(x)} \le 1$$

Μας οδηγεί στο συμπέρασμα πως

$$\frac{\int_{h(x)}^{+\infty} \bar{F}((\max{(h(x), x-y)}GF(y)}{\bar{G}(x)+\bar{F}(x)} = 0 \qquad x \to \infty$$

Δηλαδή μας οδηγεί στο συμπέρασμα πως

$$\int_{h(x)}^{+\infty} \bar{F}((\max{(h(x), x-y)}GF(y) = 0\big(\bar{G}(x)+\bar{F}(x)\big) \ x \to \infty$$

Ισοδύναμα η παραπάνω σχέση γράφεται

$$P\big(\xi+\eta > x, \xi > h(x), \eta > h(x)\big) = 0\big(\bar{G}(x)+\bar{F}(x)\big) \ x \to \infty$$

Για να είναι κάθε μια από τις συναρτήσεις G και F υποεκθετικές κατανομές θα πρέπει να ισχύει

$$\overline{F * F}(x) \sim 2\bar{F}(x) \quad x \to \infty$$

Και

$$\overline{G * F = G}(x) \sim 2\bar{G}(x) \quad x \to \infty$$

Αν στην σχέση

$$P\big(\xi + \eta > x, \xi > h(x), \eta > h(x)\big) = 0\big(\bar{G}(x) + \bar{F}(x)\big) \ x \to \infty$$

Υποθέσουμε πως οι τυχαίες μεταβλητές ακολουθούν την ίδια κατανομή F τότε η παραπάνω σχέση γράφεται

$$P\big(\xi + \eta > x, \xi > h(x), \eta > h(x)\big) = 0\big(\bar{F}(x) + \bar{F}(x)\big) = 0(\bar{F}(x)) \ x \to \infty$$

Αντίστοιχα αν οι τυχαίες μεταβλητές ακολουθούν την ίδια κατανομή G τότε η σχέση

$$P\big(\xi + \eta > x, \xi > h(x), \eta > h(x)\big) = 0\big(\bar{G}(x) + \bar{F}(x)\big) \ x \to \infty$$

Γράφεται

$$P\big(\xi + \eta > x, \xi > h(x), \eta > h(x)\big) = 0\big(\bar{G}(x) + \bar{G}(x)\big) = 0(\bar{G}(x)) \ x \to \infty$$

Αποδείξαμε λοιπόν πως οι κατανομές δύο κατανομές G και F ανήκουν στην οικογένεια των υποεκθετικών κατανομών

Θεώρημα

Έστω μια κατανομή F η οποία ανήκει στην οικογένεια των υποεκθετικών κατανομών σε όλο το $\mathbb{R}$ και έστω κατανομές $G_1, G_2, \dots, G_n$ τέτοιες ώστε για κάθε i η συνάρτηση $\bar{F} + \bar{G}_i$ είναι κατανομή με μακριά ουρά και επιπλέον $\bar{G}_i(x) = O(\bar{F}(x))$ όταν $x \to \infty$.

Τότε

$$\overline{G * G * G \dots \dots * G}(x) = \bar{G}_1(x) + \bar{G}_2(x) + \dots + \bar{G}_n(x) + o(\bar{F}(x))$$

Θεώρημα

έστω ότι ισχύουν οι συνθήκες του προηγούμενου θεωρήματος και επιπλέον η G_1 ανήκει στην οικογένεια κατανομών με μακριά ουρά και επιπλέον η G_1 χαρακτηρίζεται ως *weakly tail equivalent* της F

τότε

$\overline{G * G * G \dots \dots * G}(x)$ ανήκει στην οικογένεια των υποεκθετικών κατανομών σε όλο το $\mathbb{R}$

*7.6 Η κλάση S**

Έστω F μια κατανομή στον $\mathbb{R}$ με $\bar{F}(x) > 0$ και διακριτή θετική μέση τιμή. Θα λέμε ότι η F ανήκει στην κλάση S^* αν

$$\int_0^x \bar{F}(x-y)\bar{F}(y)\,dy \sim 2m\bar{F}(x) \qquad x \to \infty$$

Πολλές φορές για να προχωρήσουμε στην ανάλυση μας χρειαζόμαστε πιο ισχυρές δομές από αυτές που η οικογένεια των υποεκθετικών κατανομών μας παρέχει γι αυτό και στην ανάλυση μας θα εισάγουμε την κλάση S^*.

Αρχικά για κάθε κατανομή F ορισμένη στον $\mathbb{R}$ με $\bar{F}(x) > 0$ για κάθε x ισχύει η ανισότητα

$$\int_0^x \bar{F}(x-y)\,\bar{F}(y)dy = \int_0^{x/2} \bar{F}(x-y)\,\bar{F}(y)dy + \int_{x/2}^x \bar{F}(x-y)\,\bar{F}(y)dy$$

Το δεύτερο ολοκλήρωμα μπορεί να γραφτεί ως εξής

Αν θέσουμε

$$z = x - y \Rightarrow y = x - z \Rightarrow \left|\frac{dy}{dz}\right| = 1 \Rightarrow dy = dz$$

Επίσης η τυχαία μεταβλητή $z \in [0, \frac{x}{2}]$

Έτσι το ολοκλήρωμα

$$\int_{x/2}^x \bar{F}(x-y)\,\bar{F}(y)dy = \int_0^{x/2} \bar{F}(z)\,\bar{F}(x-z)dy = \int_0^{x/2} \bar{F}(x-y)\,\bar{F}(y)dy$$

Με αυτή την παρατήρηση

$$\int_0^x \bar{F}(x-y)\,\bar{F}(y)dy = \int_0^{\frac{x}{2}} \bar{F}(x-y)\,\bar{F}(y)dy + \int_0^{\frac{x}{2}} \bar{F}(x-y)\,\bar{F}(y)dy =$$

$$= 2\int_0^{\frac{x}{2}} \bar{F}(x-y)\,\bar{F}(y)dy$$

Από την άλλη πλευρά γνωρίζουμε πως αν μια τυχαία μεταβλητή παίρνει μη αρνητικές τιμές τότε η μέση της τιμή ισούται με

$$\mathbb{E}(X) = \int_0^{+\infty} xf(x)dx = -\int_0^{+\infty} x(1-\bar{F}(x))'dx =$$

$$-[x(1-\bar{F}(x))]_0^\infty + \int_0^\infty \bar{F}(x)dx$$

$$= \int_0^\infty \bar{F}(x)dx = m$$

Θα αποδείξουμε αρχικά πως ισχύει η ανισότητα

$$\int_0^x \bar{F}(x-y)\,\bar{F}(y)dy \geq 2\bar{F}(x)\int_0^{\frac{x}{2}} \bar{F}(y)\,dy$$

$$\int_0^x \bar{F}(x-y)\,\bar{F}(y)dy = 2\int_0^{\frac{x}{2}} \bar{F}(x-y)\,\bar{F}(y)dy$$

Όμως

$$\bar{F}(x-y) \geq \bar{F}(x)$$

Και το παραπάνω ολοκλήρωμα

$$\int_0^{\frac{x}{2}} \bar{F}(x-y)\,\bar{F}(y)dy \geq \int_0^{\frac{x}{2}} \bar{F}(x)\,\bar{F}(y)dy = \bar{F}(x)\int_0^{\frac{x}{2}} \bar{F}(y)\,dy$$

Έτσι καταλήξαμε

$$\int_0^x \bar{F}(x-y)\,\bar{F}(y)dy \geq 2\bar{F}(x)\int_0^{\frac{x}{2}} \bar{F}(y)\,dy$$

Και αν $\bar{F}(x) > 0$ για κάθε x

$$\frac{1}{\bar{F}(x)}\int_0^x \bar{F}(x-y)\,\bar{F}(y)dy \geq 2\int_0^{\frac{x}{2}} \bar{F}(y)\,dy$$

Και

$$\lim_{x\to\infty} inf\,\frac{1}{\bar{F}(x)}\int_0^x \bar{F}(x-y)\,\bar{F}(y)dy \geq 2\lim_{x\to\infty} inf\int_0^{\frac{x}{2}} \bar{F}(y)\,dy$$

Όμως το γενικευμένο ολοκλήρωμα δεύτερου είδους

$$\lim_{x\to\infty} inf\int_0^{\frac{x}{2}} \bar{F}(y)\,dy = \int_0^\infty \bar{F}(y)\,dy = m$$

Και καταλήξαμε στη σχέση

$$\lim_{x\to\infty} inf \frac{1}{\bar{F}(x)}\int_0^{x/2} \bar{F}(x-y)\,\bar{F}(y)dy \geq 2m$$

Αν επιπλέον η κατανομή F είναι κατανομή με βαριά ουρά (θα είναι και κατανομή με μακριά ουρά) τότε

$$\bar{F}(x-y){\sim}\bar{F}(x) \quad x\to\infty$$

και σε αυτή την περίπτωση

$$\lim_{x\to\infty} inf \frac{1}{\bar{F}(x)}\int_0^{x} \bar{F}(x-y)\,\bar{F}(y)dy = 2\lim_{x\to\infty} inf \frac{1}{\bar{F}(x)}\bar{F}(x)\int_0^{x/2} \bar{F}(y)\,dy = 2m$$

Θεώρημα

Έστω ότι η κατανομή F στον $\mathbb{R}$ ανήκει στο S^*.Τότε η F θα είναι και κατανομή με μακριά ουρά

Έχουμε ήδη αναφέρει πως μια κατανομή με μακριά ουρά είναι μια κατανομή για την οποία ισχύει

$$\frac{\bar{F}(x-y)}{\bar{F}(y)}{\sim}1 \quad y>0 \quad x\to\infty$$

Πιο συγκεκριμένα

$$\frac{\bar{F}(x-1)}{\bar{F}(y)}{\sim}1 \quad y>0 \quad x\to\infty$$

Αποδείξαμε νωρίτερα πως

$$\int_0^{x} \bar{F}(x-y)\,\bar{F}(y)dy \geq 2\int_0^{x/2} \bar{F}(x-y)\,\bar{F}(y)dy$$

Το δεξί μέλος της παραπάνω σχέσης μπορεί να γραφεί ως εξής

$$2\bar{F}(x)\int_0^{\frac{x}{2}} \bar{F}(y)\,dy = 2\left(\int_0^{1} \bar{F}(x-y)\,\bar{F}(y)dy + \int_1^{\frac{x}{2}} \bar{F}(x-y)\bar{F}(y)\,dy\right)$$

$$\geq 2\bar{F}(x)\int_0^{1} \bar{F}(y)\,dy + 2\bar{F}(x-1)\int_1^{\frac{x}{2}} \bar{F}(y)\,dy$$

Έτσι μέχρι στιγμής έχουμε δείξει πως

$$\int_0^{x} \bar{F}(x-y)\,\bar{F}(y)dy \geq 2\bar{F}(x)\int_0^{1} \bar{F}(y)\,dy + 2\bar{F}(x-1)\int_1^{\frac{x}{2}} \bar{F}(y)\,dy$$

$$\frac{1}{2}\int_0^x \bar{F}(x-y)\,\bar{F}(y)dy \geq \bar{F}(x)\int_0^1 \bar{F}(y)\,dy + \bar{F}(x-1)\int_1^{\frac{x}{2}} \bar{F}(y)\,dy$$

Αφαιρώ και από τα δύο μέλη την ποσότητα

$$\bar{F}(x)\int_1^{\frac{x}{2}} \bar{F}\,(y)dy$$

Έτσι προκύπτει η ανισότητα

$$\frac{1}{2}\int_0^x \bar{F}(x-y)\,\bar{F}(y)dy - \bar{F}(x)\int_1^{\frac{x}{2}} \bar{F}\,(y)dy$$

$$\geq \bar{F}(x)\int_0^1 \bar{F}(y)\,dy + \bar{F}(x-1)\int_1^{\frac{x}{2}} \bar{F}(y)\,dy - \bar{F}(x)\int_1^{\frac{x}{2}} \bar{F}\,(y)dy$$

Και με πράξεις

$$\frac{1}{2}\int_0^x \bar{F}(x-y)\,\bar{F}(y)dy - \bar{F}(x)\int_1^{\frac{x}{2}} \bar{F}\,(y)dy \geq$$

$$\geq \bar{F}(x)\int_0^1 \bar{F}(y)\,dy + (\bar{F}(x-1) - \bar{F}(x))\int_1^{\frac{x}{2}} \bar{F}\,(y)dy$$

Συνεχίζοντας τις πράξεις και αναδιατάσσοντας τους όρους έχουμε

$$\frac{1}{2}\int_0^x \bar{F}(x-y)\,\bar{F}(y)dy - \bar{F}(x)\left(\int_0^1 \bar{F}(y)\,dy + \int_1^{\frac{x}{2}} \bar{F}\,(y)dy\right) \geq$$

$$\geq (\bar{F}(x-1) - \bar{F}(x))\int_1^{\frac{x}{2}} \bar{F}\,(y)dy$$

Όμως

$$\int_0^1 \bar{F}(y)\,dy + \int_1^{\frac{x}{2}} \bar{F}\,(y)dy = \int_0^{x/2} \bar{F}(y)\,dy$$

Και τελικά

$$\frac{1}{2}\int_0^x \bar{F}(x-y)\,\bar{F}(y)dy - \bar{F}(x)\int_0^{x/2} \bar{F}(y)\,dy \geq (\bar{F}(x-1) - \bar{F}(x))\int_1^{\frac{x}{2}} \bar{F}\,(y)dy$$

Από την παραπάνω σχέση εύκολα καταλήγουμε διαιρώντας και τα δύο μέλη με

$$\bar{F}(x)\int_{1}^{\frac{x}{2}}\bar{F}\,(y)dy$$

$$\frac{\frac{1}{2}\int_0^x \bar{F}(x-y)\,\bar{F}(y)dy - \bar{F}(x)\int_0^{x/2}\bar{F}(y)\,dy}{\bar{F}(x)\int_1^{\frac{x}{2}}\bar{F}\,(y)dy} \geq \frac{\bar{F}(x-1)-\bar{F}(x)}{\bar{F}(x)}$$

Αν η κατανομή F στον $\mathbb{R}$ ανήκει στο S^* τότε

$$\int_0^x \bar{F}(x-y)\bar{F}(y)\,dy \sim 2m\bar{F}(x) \qquad x\to\infty$$

και ο αριθμητής

$$\frac{1}{2}\int_0^x \bar{F}(x-y)\,\bar{F}(y)dy - \bar{F}(x)\int_0^{x/2}\bar{F}(y)\,dy = \ m\bar{F}(x) - \ m\bar{F}(x) = 0$$

Και επειδή η ποσότητα

$$\frac{\bar{F}(x-1)-\bar{F}(x)}{\bar{F}(x)}$$

Δεν μπορεί να είναι αρνητική τότε

$$\frac{\bar{F}(x-1)-\bar{F}(x)}{\bar{F}(x)} \to 0 \quad x\to\infty$$

Αποδείξαμε λοιπόν πως η κατανομή F ανήκει στην οικογένεια κατανομών με μακριά ουρά.

Θεώρημα

Έστω ότι η κατανομή F στον $\mathbb{R}$ τότε τα ακόλουθα είναι ισοδύναμα

(i) F ανήκει στο S^*

(ii) Η F είναι κατανομή με μακριά ουρά και για κάθε συνάρτηση h $\mu\varepsilon$ $h(x) < \frac{x}{2}$ τέτοια ώστε $h(x)\to\infty \quad x\to\infty$ ισχύει

$$\int_{h(x)}^{x-h(x)} \bar{F}(x-y)\bar{F}\,(y)dy = o\big(\bar{F}(x)\big) \quad x \to \infty$$

Απόδειξη

Αρχικά τόσο η πρώτη περίπτωση όσο και η δεύτερη περίπτωση υποθέτουν πως η κατανομή F είναι μια κατανομή με μακριά ουρά.

Το ολοκλήρωμα

$$\int_0^x \bar{F}(x-y)\,\bar{F}(y)dy =$$

$$\int_0^{h(x)} \bar{F}(x-y)\bar{F}(y)dy + \int_{h(x)}^{x-h(x)} \bar{F}(x-y)\bar{F}(y)dy + \int_{x-h(x)}^{x} \bar{F}(x-y)\bar{F}(y)dy$$

Θα ασχοληθούμε αρχικά με το ολοκλήρωμα

$$\int_{x-h(x)}^{x} \bar{F}(x-y)\bar{F}(y)dy$$

Θέτουμε αρχικά

$$z = x - y \Rightarrow \left|\frac{dy}{dz}\right| = 1$$

Επίσης

$$x - h(x) \leq y \leq x \Rightarrow 0 \leq z \leq h(x)$$

Έτσι σύμφωνα με τα παραπάνω

$$\int_{x-h(x)}^{x} \bar{F}(x-y)\bar{F}(y)dy = \int_0^{h(x)} \bar{F}(z)\bar{F}(x-z)dz$$

Και σύμφωνα με αυτή την παρατήρηση το ολοκλήρωμα

$$\int_0^x \bar{F}(x-y)\,\bar{F}(y)dy =$$

$$\int_0^{h(x)} \bar{F}(x-y)\bar{F}(y)dy + \int_0^{h(x)} \bar{F}(z)\bar{F}(x-z)dz + \int_{h(x)}^{x-h(x)} \bar{F}(x-y)\bar{F}(y)dy$$

$$= 2\int_0^{h(x)} \bar{F}(x-y)\bar{F}(y)dy + \int_{h(x)}^{x-h(x)} \bar{F}(x-y)\bar{F}(y)dy$$

Επιπλέον

$$\int_0^{h(x)} \bar{F}(x-y)\bar{F}(y)dy \sim \bar{F}(x)\int_0^{\infty} \bar{F}(y)dy = m\bar{F}(x) \quad x \to \infty$$

Δηλαδή η αρχική μας σχέση

$$\int_0^{x} \bar{F}(x-y)\,\bar{F}(y)dy = 2m\bar{F}(x) + \int_{h(x)}^{x-h(x)} \bar{F}(x-y)\bar{F}(y)dy$$

Όμως μια κατανομή ανήκει στο S^* όταν

$$\int_0^{x} \bar{F}(x-y)\,\bar{F}(y)dy = 2m\bar{F}(x)$$

Δηλαδή θα πρέπει

$$\int_{h(x)}^{x-h(x)} \bar{F}(x-y)\bar{F}(y)dy = o(\bar{F}(x))$$

Όντως το παραπάνω ολοκλήρωμα θα μπορούσαμε να αποδείξουμε ότι ισούται με $o(\bar{F}(x))$ εργαζόμενοι ως εξής

Θα εισάγουμε στην ανάλυση μας τη σωρευτική συνάρτηση διακινδύνευσης ορίζεται ως

$$R(x) = \int_{-\infty}^{x} r(u)du = \int_{-\infty}^{x} \frac{f(u)}{S(u)}du =$$

$$R(x) = \int_0^{x} r(u)du = \int_0^{x} \frac{f(u)}{\bar{F}(u)}du = -\ln \bar{F}(x)$$

Από την τελευταία σχέση προκύπτει πως

$$e^{-R(x)} = \bar{F}(x)$$

Έτσι το ολοκλήρωμα

$$\int_{h(x)}^{x-h(x)} \bar{F}(x-y)\bar{F}(y)dy = \int_{h(x)}^{x-h(x)} e^{-R(x-y)}e^{-R(y)}dy =$$

$$= \int_{h(x)}^{x-h(x)} e^{-\left(R(x-y)+R(y)\right)}\, dy$$

Παρατηρούμε πως

$$\lim_{x\to\infty} \frac{e^{-\left(R(x-y)+R(y)\right)}}{\bar{F}(x)} = 0$$

Δηλαδή

$$e^{-\left(R(x-y)+R(y)\right)} \sim o(\bar{F}(x))$$

Ποιο συγκεκριμένα

$$\frac{1}{\bar{F}(x)} \int_{h(x)}^{x-h(x)} e^{-\left(R(x-y)+R(y)\right)}\, dy = \int_{h(x)}^{x-h(x)} \frac{e^{-\left(R(x-y)+R(y)\right)}}{e^{-R(x)}}\, dy =$$

$$= \int_{h(x)}^{x-h(x)} e^{-\left(R(x-y)+R(y)\right)+R(x)}\, dy = \int_{h(x)}^{x-h(x)} e^{R(x)-R(x-y)}\, e^{-R(y)} dy$$

$$= \int_{h(x)}^{x-h(x)} \frac{\bar{F}(x-y)}{\bar{F}(x)}\, e^{-R(y)} dy \sim \int_{h(x)}^{x-h(x)} e^{-R(y)}\, dy \quad x \to \infty$$

Και το ολοκλήρωμα

$$\int_{h(x)}^{x-h(x)} e^{-R(y)}\, dy = -\left[\frac{e^{-R(y)}}{r(y)}\right]_{h(x)}^{x-h(x)}$$

$$= \frac{e^{-R(h(x))}}{r(h(x))} - \frac{e^{-R(x-h(x))}}{r(x-h(x))} \to 0 \quad x \to \infty$$

Έτσι αποδείξαμε πως όντως το ολοκλήρωμα

$$\int_{h(x)}^{x-h(x)} \bar{F}(x-y)\bar{F}(y) dy = o(\bar{F}(x))$$

Θεώρημα

Αν F ανήκει στο S^* τότε η F ανήκει στην οικογένεια των υποεκθετικών σε όλο το $\mathbb{R}$ η F_I ανήκει στην κλάση S.

Θεώρημα

Έστω F μια κατανομή με μακριά ουρά ορισμένη στο $\mathbb{R}$ υποθέτουμε επίσης ότι υπάρχει $\gamma < 1$ και $A < \infty$ τέτοιο ώστε η συνάρτηση $R(x)$ να ικανοποιεί την παρακάτω ανισότητα

$$R(x) - R(x-y) \le \gamma R(y) + A$$

για κάθε $x > 0$, $y \in [0, \frac{x}{2}]$. Αν η συνάρτηση είναι ολοκληρώσιμη σε όλο το $\mathbb{R}^+$ τότε F ανήκει στο S^*.

Όμως

$$R(x) - R(x-y) = log\left(\frac{\bar{F}(x-y)}{\bar{F}(x)}\right)$$

και

$$R(y) = -log\bar{F}(y)$$

Έτσι η σχέση

$$R(x) - R(x-y) \le \gamma R(y) + A$$

γράφεται

$$log\left(\frac{\bar{F}(x-y)}{\bar{F}(x)}\right) \le log\bar{F}(y)^{-\gamma} + logB = \log\left(\bar{F}(y)^{-\gamma}B\right)$$

Όπου

$$logB = A$$

Σύμφωνα λοιπόν με την παραπάνω σχέση

$$\frac{\bar{F}(x-y)}{\bar{F}(x)} \le \bar{F}(y)^{-\gamma}B = \frac{e^A}{\bar{F}(y)^{\gamma}}$$

Στην περίπτωση που

$$\frac{\bar{F}(x-y)}{\bar{F}(x)} \to 1 \quad x \to \infty$$

$$\frac{e^A}{\bar{F}(y)^{\gamma}} \ge 1$$

Και έχουμε ένα άνω φράγμα για την δεξιά ουρά της κατανομής

$$\bar{F}(y) < e^{A/\gamma}$$

Κεφάλαιο 8

Μέγιστο τυχαίων περιπάτων

8.1 Κλιμακωτή δομή του μέγιστου ενός τυχαίου περιπάτου

Σε αυτή την ενότητα θα προσπαθήσουμε να προσεγγίσουμε τον μέγιστο ενός τυχαίου περιπάτου μέσου υψών αυξανόμενης κλίμακας και μέσω της ανανεωτικής θεωρίας.

Αρχικά θα υποθέσουμε πως το βήμα του τυχαίου περιπάτου έχει αρνητική μέση τιμή δηλαδή αν ο τυχαίος περίπατος περιγράφεται από τη σχέση

$$S_n = \sum_{i=1}^{n} X_i$$

Τότε

$$E(X_i) < 0$$

Αν λοιπόν η μέση τιμή του βήματος είναι αρνητική τότε με πιθανότητα τη μονάδα η θέση του τυχαίου περιπάτου θα πάει στο $-\infty$ $\gamma\iota\alpha$ $n \to \infty$.

$$\lim_{n\to\infty} P(S_n = -\infty) = 1$$

Ήρθε η ώρα να ορίσουμε το μέγιστο ενός τυχαίου περιπάτου ως εξής

Αρχικά ορίζουμε το μέγιστο ενός τυχαίου περιπάτου σε πεπερασμένο χρονικό διάστημα μήκους n

$$M_n = max\{S_i\ ,0 \le i \le n\}$$

Δηλαδή

$$M_1 = max\{S_i\ ,0 \le i \le 1\} = max\{S_0, S_1\} = S_1$$

$$M_2 = max\{S_i\ ,0 \le i \le 2\} = max\{S_1, S_2\} = max\{M_1, S_2\}$$

.

.

.

.

$$M_n = max\{S_1, S_{2,\dots,}S_n\}$$

Ορίζουμε επίσης το ολικό μέγιστο ως

$$M = sup\{S_n, n \geq 0\} = sup\{M_1, M_{2,\dots,}M_n\}$$

Σύμφωνα λοιπόν με όσα είπαμε μέχρι στιγμής και κάτω από την υπόθεση της υποεκθετικότητας της κατανομής του βήματος του τυχαίου περιπάτου μπορούμε να γράψουμε την ανισότητα

$$S_n \leq max\{S_1, S_{2,\dots,}S_n\} \leq \sum_{k=1}^{n} max\{0, X_k\}$$

Ισοδύναμα η παραπάνω σχέση γράφεται

$$S_n \leq M_n \leq \sum_{k=1}^{n} X_i^+$$

Και το ενδεχόμενο

$$P(M_n > x) = P\big(max\{S_1, S_{2,\dots,}S_n\} > x\big) \geq$$

$$= 1 - \mathrm{P}\big(max\{S_1, S_{2,\dots,}S_n\} \leq x\big) = 1 - \big(P(\{S_1, S_{2,\dots,}S_n\} \leq x)\big)$$

$$= 1 - P(S_1 \leq x, S_2 \leq x, \dots, S_n \leq x) = 1 - (P(S_1 \leq x))^n$$

Και τελικά καταλήγουμε στο συμπέρασμα πως

$$P(M_n > x) \geq 1 - (F(x))^n \sim n\bar{F}(x) \;\; x \to \infty$$

Το παραπάνω προκύπτει είτε με διαίρεση του πολυωνύμου $1 - (F(x))^n$ με το $1 - F(x)$ οπότε και θα προκύψει

$$1 - (F(x))^n = (1 - F(x)\,)(1 + F(x) + F(x)^2 + \cdots + F(x)^{n-1})$$

και για $x \to \infty$

$$\lim_{x\to\infty} F(x)^n = 1 \; n = 1,2, \dots$$

έτσι

$$1 - (F(x))^n \sim n\bar{F}(x)$$

8.2 Πρώτο βήμα σε αυξανόμενη κλίμακα

Ορίζουμε επίσης

$$n_+(1) = n_+ = min\{k \geq 1 : S_k > 0\}$$

Η παραπάνω μεταβλητή χαρακτηρίζεται ως ελλειμματική ή προβληματική (*defective)* και αυτό διότι

$$P(n_+(1) = \infty) < 1$$

Δηλαδή υπάρχει το ενδεχόμενο να μην υπάρξει ποτέ κάποια χρονική στιγμή k τέτοια ώστε

$$S_k > 0$$

Την μεταβλητή $n_+(1)$ την ονομάζουμε πρώτο βήμα σε αυξανόμενη κλίμακα στη διεθνή βιβλιογραφία γνωστό ως ***first strictly ascending ladder epoch***

Έτσι τη χρονική στιγμή $n_+(1)$ ο τυχαίος περίπατος για πρώτη φορά βρίσκεται σε θέση με θετική τιμή δηλαδή

$$S_{n_+(1)} > 0$$

Και το ονομάζουμε το πρώτο αυξανόμενης κλίμακας ύψος στη διεθνή βιβλιογραφία γνωστό ως *first strictly ascending ladder height.*

Πριν προχωρήσουμε την ανάλυση μας καλό θα ήταν να εισάγουμε τον ορισμό των χρόνων διακοπής *(stopping times)*

8.3 Χρόνος διακοπής

Ένας χρόνος διακοπής είναι ένας τυχαίος χρόνος μια ακέραια τυχαία μεταβλητή T η τιμή της οποίας εξαρτάται από τις προηγούμενες αυτής της χρονικής στιγμής καταστάσεις X_i.Δηλαδή η T είναι τυχαία μεταβλητή με τιμές στο $\mathbb{N}_0 \cup \{\infty\}$ με την ιδιότητα για κάθε $n \in \mathbb{N}_0 \cup \{\infty\}$ το ενδεχόμενο

$$\mathbb{I}\{T = n\} = \varphi(X_1, X_2, \dots, X_n)$$

Λογικό αν σκεφτούμε πως δεν είναι δυνατόν να σταματήσουμε τη διακοπή μιας διαδικασίας σε χρόνο μεταγενέστερο από τη στιγμή της διακοπής.

Αν οι τυχαίες μεταβλητές που περιγράφουν τα βήματα του τυχαίου περιπάτου X_i είναι ισόνομες και ανεξάρτητες τυχαίες μεταβλητές και αν $n_+(1)$ είναι χρόνος διακοπής και μάλιστα ισχύει $n_+(1) < \infty$ τότε οι θέσεις του τυχαίου περιπάτου μετά τη χρονική στιγμή $n_+(1)$ δεν θα εξαρτώνται από τις προηγούμενες χρονικές στιγμές όπως και οι τυχαίες μεταβλητές που περιγράφουν τα βήματα του τυχαίου περιπάτου σε κάθε χρονική στιγμή από τη μια θα είναι ανεξάρτητες και ισόνομες για χρονικές στιγμές που επόνται της $n_+(1)$ και από την άλλη θα είναι ανεξάρτητες από τις μεταβλητές $\{n_+(1), X_1, X_2, \dots, X_{n+}\}$.

Σημειώνεται πως σε περίπτωση που η τυχαία μεταβλητή τ περιγράφει χρόνο διακοπής τότε το ενδεχόμενο $\{\tau \geq n\}$ θα εξαρτάται από τις $S_0, S_1, \dots, S_{n-1}$ θέσεις του τυχαίου περιπάτου ή ισοδύναμα θα εξαρτάται από τις τυχαίες μεταβλητές βήματα του τυχαίου περιπάτου $\xi_1, \xi_1, \dots, \xi_{n-1}$.

Με λίγα λόγια και χωρίς τη χρήση μαθηματικών σχέσεων αυτό που λέμε παραπάνω είναι πως το μέλλον εξαρτάται από το παρελθόν.

Η υπό ακολουθία των αυξανόμενης κλίμακας βήματος (ή διαστήματος χρονικού) ορίζεται με τη βοήθεια της επαγωγής .

$$n_+(2) = min\{k \geq 1 : S_k > S_{n_+(1)}\}$$

Ο χρόνος $n_+(2)$ είναι ο χρόνος που αντιστοιχεί σε εκείνη τη χρονική στιγμή κατά την οποία για πρώτη φορά ο τυχαίος περίπατος ξεπερνάει την αμέσως προηγούμενη μέγιστή τιμή του $S_{n_+(1)}$. ο χρόνος $n_+(2)$ μπορεί να θεωρηθεί χρόνος διακοπής και τότε οι τυχαίες μεταβλητές που περιγράφουν τη θέση του τυχαίου περιπάτου μετά την χρονική στιγμή $n_+(2)$ θα είναι ανεξάρτητες από όλες τις προηγούμενες τυχαίες μεταβλητές και από το $n_+(2)$.Στην ουσία είναι σαν να τερματίζουμε ένα παιχνίδι και να αρχίζουμε ένα άλλο με χρονικό σημείο τερματισμού το σημείο που θα ξεπεράσει το τέλος του προηγούμενου.

$$n_+(3) = min\{k \geq 1 : S_k > S_{n_+(2)}\}$$

.

.

.

.

$$n_+(n) = min\{k \geq n_+(n-1) : S_k > S_{n_+(n-1)}\}$$

$$n_+(n+1) = min\{k \geq n_+(n) : S_k > S_{n_+(n)}\}$$

Η τυχαία μεταβλητή $S_{n_+(n)}$ καλείται νιοστό αυξανόμενο κλιμακωτό ύψος *(strictly ascending ladder height).*

Θεωρώντας και πάλι πως $n_+(n) < \infty$ δηλαδή κάποια στιγμή (η οποία χρονικά θα έπεται της $n_+(n-1)$) ο τυχαίος περίπατος με βήματα $X_{n_+(n)+k}$ $k = 1,2,3,..$θα ξεπεράσει το $S_{n_+(n)}$ και επιπλέον οι τυχαίες μεταβλητές $X_{n_+(n)+k}$ $k = 1,2,3,.$ είναι ανεξάρτητες από τις τυχαίες μεταβλητές $n_+(i)$, $i = 1,2,3, \dots n$ και τις μεταβλητές X_i $με$ $i = 1,2, \dots n_+(n)$.

Σκοπός αυτής της ενότητας είναι ο προσδιορισμός της κατανομής του ολικού μεγίστου του τυχαίου περιπάτου.

Η πρώτη πιθανότητα που μπορούμε σχετικά εύκολα να υπολογίσουμε είναι η

$$P(M = 0) = P(\sup\{S_n \leq 0\, , για\ κάθε\ n \geq 1\} = P(n_+ = \infty)$$

Λογικός ο παραπάνω ισχυρισμός αν σκεφτεί κανείς πως αν όλες οι θέσεις του τυχαίου περιπάτου είναι αρνητικές τότε δεν θα έρθει ποτέ η χρονική στιγμή κατά την οποία ο τυχαίος περίπατος θα βρεθεί να πάρει θετική τιμή. Και αφού δεν θα έρθει ποτέ αυτή η στιγμή $n_+ = \infty$.

Μέχρι στιγμής θεωρούμε τον χρόνο διακριτό και κάθε χρονική στιγμή μετράει και τον αριθμό των προσπαθειών που κάνει ο τυχαίος περίπατος για να φτάσει στις θετικές τιμές .Είναι σαν να λέμε ότι μετράει τον αριθμό των αποτυχιών μέχρι να φτάσει στην πρώτη επιτυχία. Διαισθητικά λοιπόν περιμένουμε στην ανάλυση μας να παίξει σημαντικό ρόλο η γεωμετρική κατανομή.

Μπορούμε εύκολα να εισάγουμε στην ανάλυση μας ότι μια χρονική στιγμή μπορεί να έρθει ή να μην έρθει ποτέ.

$$n_+(k) = \begin{cases} +\infty \; με \; πιθανότητα \; p \\ < \infty \; με \; πιθανότητα \; q \end{cases}$$

Για να υπάρξει η χρονική στιγμή $n_+(k)$ θα πρέπει να έχουν προϋπάρξει και όλες οι προηγούμενες χρονικές στιγμές $n_+(i)$ $i = 1,2,3, \dots . k - 1$έτσι η πιθανότητα να υπάρχει και να είναι πεπερασμένη είναι ίση με

$$P(n_+(k) < \infty) = P(n_+(1) < \infty, n_+(2) < \infty, \dots n_+(k-1) < \infty, n_+(k) < \infty)$$

$$= (1-p)^k$$

Με p την πιθανότητα το μέγιστο του τυχαίου περιπάτου να μην ξεπεράσει πότε το μηδέν

$$p = P(M = 0) = P(max\{S_n, n \geq 1\} = 0)$$

Από την παραπάνω σχέση μπορούμε να κάνουμε την εξής παρατήρηση, αφού το μέγιστο από όλες τις δυνατές τιμές που παίρνει ο τυχαίος μας περίπατος είναι ίσο με το μηδέν αυτό το ενδεχόμενο θα είναι ισοπίθανο με το ενδεχόμενο όλες οι θέσεις S_i $i = 1,2,3 \dots n$ είναι μικρότερες ή το πολύ ίσες με το μηδέν.

$$P(max\{S_n, n \geq 1\} = 0) = P(S_i \leq 0, για \; κάθε \; i = 1,2,3, \dots)$$

Σύμφωνα λοιπόν με την τελευταία σχέση και σύμφωνα με τον ορισμό της χρονικής στιγμής

$$n_+(1) = n_+ = min\{k \geq 1 : S_k > 0\}$$

$$P(M = 0) = P(min\{k \geq 1 : S_k > 0\} = \infty)$$

Και τελικά

$$P(M = 0) = P(n_+(1) = \infty) = p$$

Και η πιθανότητα

$$P(M > x) = P(max\{S_n, n \geq 1\} > x)$$

Καλό θα ήταν να αναλύσουμε την παραπάνω σχέση .Ζητάμε στην ουσία να βρούμε την πιθανότητα η μέγιστη θέση του τυχαίου περιπάτου να υπερβαίνει ένα θετικό σημείο.

Η μέγιστη θέση θα σημειωθεί σε κάποια χρονική στιγμή $n_+(k)$ και αφού θα συμβεί αυτή η χρονική στιγμή θεωρούμε το

$$n_+(k) < \infty$$

Από την άλλη από τη στιγμή που τη χρονική στιγμή $n_+(k)$ η θέση του τυχαίου περιπάτου $S_{n_+(k)}$ παίρνει την μέγιστη τιμή του δεν θα υπάρξει χρονική στιγμή μεγαλύτερη από $n_+(k)$ ώστε ο τυχαίος περίπατος να πάρει ακόμη μεγαλύτερη τιμή.

Σύμφωνα λοιπόν με όλα τα παραπάνω

$$n_+(k+1) = \infty$$

8.4 Ουρά κατανομής μεγίστου

Είμαστε λοιπόν σε θέση να υπολογίσουμε την πιθανότητα

$$P(M > x) = \sum_{k=1}^{\infty} P(S_{n_+(k)} > x, n_+(k) < \infty, n_+(k+1) = \infty)$$

$$= \sum_{k=1}^{\infty} P(S_{n_+(k)} > x, n_+(k) < \infty) P(n_+(k+1) = \infty))$$

Και με τη βοήθεια της δεσμευμένης πιθανότητας και του κανόνα του Bayes η παραπάνω σχέση γράφεται

$$P(M > x) = \sum_{k=1}^{\infty} P\big(S_{n_+(k)} > x / n_+(k) < \infty\big) P(n_+(k) < \infty) P(n_+(k+1) = \infty)$$

$$= \sum_{k=1}^{\infty} p(1-p)^k \, P\big(S_{n_+(k)} > x / n_+(k) < \infty\big)$$

Θα πρέπει να υπολογίσουμε την πιθανότητα

$$P\big(S_{n_+(k)} > x / n_+(k) < \infty\big)$$

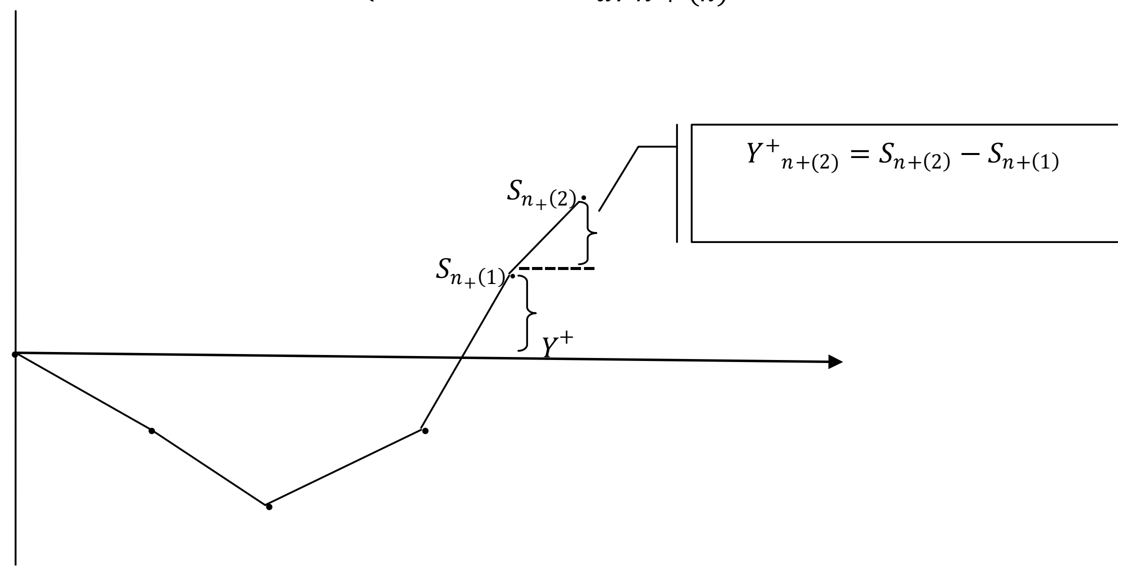
Για να υπολογίσουμε αυτή την πιθανότητα θα εισάγουμε στην ανάλυση μας μια καινούργια μεταβλητή που θα περιγράφει την υπέρβαση *(overshoot)* πάνω από ένα σημείο.

Συμβολίζουμε με Y^+ την θετική απόσταση της θέσης του τυχαίου περιπάτου που για πρώτη φορά γίνεται θετική από το μηδέν.

$$Y^+ = \begin{cases} S_{n_+(1)} & \alpha\nu \;\; n_+(1) < \infty \\ \infty & \alpha\nu \;\; n_+(1) = \infty \end{cases}$$

Ορίζουμε την **μεταβλητή υπέρβασης**

$$Y^{+}{}_{n+(k)} = \begin{cases} S_{n+(k)} - S_{n+(k-1)} & \alpha\nu \ \ n+(k) < \infty \\ \infty & \alpha\nu \ n+(k) = \infty \end{cases}$$



Με τη βοήθεια του παραπάνω διαγράμματος παρατηρούμε ότι το άθροισμα των τυχαίων μεταβλητών $Y^{+}{}_{n+(k)}$ ισούται με τη θετική μέγιστη θέση που παίρνει ο τυχαίος περίπατος.

Δηλαδή

$$M = \sum_{i=1}^{N} Y^{+}{}_{i}$$

Αν τη χρονική στιγμή k το $S_{n_+(k)} = M$

τότε

$$P\left(S_{n_+(k)} > x / n_+(k) < \infty\right) = P\left(\sum_{i=1}^{k} Y^{+}{}_{i} > x\right) = \bar{G}^{*(k)}(x)$$

Και η πιθανότητα

$$P(M > x) = \sum_{k=1}^{\infty} p(1-p)^k \, P\big(S_{n_+(k)} > x / n_+(k) < \infty\big) =$$

$$= \sum_{k=1}^{\infty} p(1-p)^k \, \bar{G}^{*(k)}(x)$$

Θεώρημα

Υποθέτουμε ότι η μέση τιμή του βήματος $\mathbb{E}\xi = -\alpha < 0$.Τότε για οποιοδήποτε $x \geq 0$

$$P(M > x) \geq \frac{\int_x^{\infty} \bar{F}(y)dy}{a + \int_x^{\infty} \bar{F}(y)dy}$$

Απόδειξη

Αρχικά θα αναφέρουμε πως οι τυχαίες μεταβλητές $\xi \; και \; M$ είναι ανεξάρτητες και θα ορίσουμε τη συνάρτηση

$$L_z(y) = \begin{cases} x & αν \; x \geq y \\ y \; αν \; x < y \leq x + y \\ x + z \; αν \; y > x + z \end{cases}$$

Η συνάρτηση $L_z(y)$ είναι άνω και κάτω φραγμένη, αυτό σημαίνει πως ο μέσος της συνάρτησης $\mathbb{E}(L_z(y))$ είναι φραγμένος και επιπλέον

$$\mathbb{E}\big(L_z(M)\big) = \mathbb{E}\big(L_z(M + \xi)\big)$$

Ισοδύναμα

$$\mathbb{E}\big(L_z(M)\big) - \mathbb{E}\big(L_z(M + \xi)\big) = 0$$

Θα μελετήσουμε τι συμβαίνει στην συνάρτηση $L_z(y)$ ασυμπτωτικά όταν $z \to \infty$.

$$\lim_{z \to \infty} L_z(y) \begin{cases} x \;\; αν \; x \geq y \\ y \; αν \; y > x \end{cases}$$

Για $\boldsymbol{y} \in [\boldsymbol{0}, \boldsymbol{x}]$

$$L(y + \xi) - L(y) = (y + \xi - x)\mathbb{I}\{\xi + y > x\} \geq (\xi + x)\mathbb{I}\{\xi > x\}$$

Και κατά συνέπεια

$$\mathbb{E}(L((M + \xi) - L(M); M \leq x) = \mathbb{E}((M + \xi - x)\mathbb{I}\{M + \xi > x\})$$

$$\geq \mathbb{E}\big((\xi - x)\mathbb{I}\{\xi > x\}\mathbb{I}\{M \leq x\}\big) \geq \mathbb{E}(\xi - x; \xi > x)\mathbb{E}(\mathbb{I}\{M \leq x\}) =$$

$$= \mathbb{E}(\xi - x; \xi > x)P(M \leq x)$$

Για $\boldsymbol{y} > x$

$$L(y+\xi)-L(y)=(y+\xi-y)\mathbb{I}\{\xi+y>x\}\geq \xi\mathbb{I}\{\xi>x\}$$

Και αντίστοιχα η μέση τιμή

$$\mathbb{E}\big(L(y+\xi)-L(y)\big)=\mathbb{E}\xi P(M>x)$$

Τότε η εξίσωση

$$\mathbb{E}\big(L(M+\xi)-L(M)\big)=$$

$$\mathbb{E}(L(M+\xi)-L(M);M\leq x)+\ \mathbb{E}(L(M+\xi)-L(M);M>x)=0$$

$$\geq \mathbb{E}(\xi-x;\xi>x)P(M\leq x)+\mathbb{E}\xi P(M>x)$$

Ισοδύναμα

$$-\mathbb{E}\xi P(M>x)\geq \mathbb{E}(\xi-x;\xi>x)P(M\leq x)+\mathbb{E}\xi P(M>x)$$

Καταλήξαμε λοιπόν στη σχέση

$$P(M>x)\geq \frac{\int_x^\infty \bar{F}(y)dy}{a+\int_x^\infty \bar{F}(y)dy}$$

Θεώρημα

Υποθέστε επιπροσθέτως με τη συνθήκη $\mathbb{E}\xi=-\alpha<0$ και η F_I η ολοκληρώσιμη ουρά κατανομής *(integrated tail distribution)* είναι υποεκθετική .Τότε

$$P(M>x)\sim a^{-1}\bar{F}_I(x)\quad x\to\infty$$

Απόδειξη

Γνωρίζουμε ήδη από το προηγούμενο θεώρημα πως

$$P(M>x)\geq \frac{\int_x^\infty \bar{F}(y)dy}{a+\int_x^\infty \bar{F}(y)dy}$$

Επίσης έχουμε ορίσει την

$$\bar{F}_I(x)=min\left\{1,\int_x^\infty \bar{F}(y)dy\right\}$$

Με το ολοκλήρωμα

$$\int_x^\infty \bar{F}(y)dy=\mathbb{E}(\xi-x;\xi>x)=$$

$$\mathbb{E}(\xi;\xi > x) - \mathbb{E}(x;\xi > x)$$

Η σχέση

$$P(M > x) \geq \frac{\int_x^\infty \bar{F}(y)dy}{a + \int_x^\infty \bar{F}(y)dy}$$

Γράφεται σύμφωνα με τα παραπάνω

$$P(M > x) \geq \frac{\bar{F}_I(x)}{\alpha + \bar{F}_I(x)}$$

και

$$\frac{P(M > x)}{\bar{F}_I(x)} \geq \frac{1}{\alpha + \bar{F}_I(x)}$$

Όταν $x \to \infty$

$$\frac{P(M > x)}{\bar{F}_I(x)} \geq \frac{1}{\alpha}$$

Αν λοιπόν καταφέρουμε να δείξουμε πως

$$\frac{P(M > x)}{\bar{F}_I(x)} \leq \frac{1}{\alpha}$$

Τότε

$$\frac{P(M > x)}{\bar{F}_I(x)} \to \frac{1}{\alpha} \qquad x \to \infty$$

Αρκεί λοιπόν να αποδείξουμε

$$\frac{P(M > x)}{\bar{F}_I(x)} \leq \frac{1}{\alpha}$$

Αρχικά θα ορίσουμε τους χρόνους ανανέωσης

$$0 = \leq \tau_1 \leq \tau_2 \leq \cdots$$

Για τη στοχαστική διαδικασία $\{S_n\}$ ως εξής

$$\tau_1 = \min\{j \geq 1: S_j > A - j(a - \varepsilon)\} =$$

$$= \min\{j \geq 1: S_j > A - j(-\mathbb{E}\xi_\iota - \varepsilon)\} =$$

$$= \min\{j \geq 1: S_j > A + j(\mathbb{E}\xi_\iota - \varepsilon)\}$$

Ορίζουμε επίσης

$$\tau_k = \infty \; \alpha\nu \;\; \tau_{k-1} = \infty$$

και

$$\tau_k = \tau_{k-1} + min\left\{j \geq 1 \quad S_{\tau_{\kappa-1+j}} - S_{\tau_{\kappa-1}} > A - j(a-\varepsilon)\right\} \quad \alpha\nu \; \tau_{k-1} < \infty$$

$$= \tau_{k-1} + min\left\{j \geq 1 \quad S_{\tau_{\kappa-1+j}} - S_{\tau_{\kappa-1}} > A + j(\mathbb{E}\xi_\iota - \varepsilon)\right\}$$

Όμως οι τυχαίες μεταβλητές

$$\tau_1 - \tau_0, \tau_2 - \tau_1, \tau_3 - \tau_2, \dots, \tau_k - \tau_{k-1}$$

είναι ανεξάρτητες .

Ανεξάρτητες είναι και οι τυχαίες μεταβλητές

$$S_{\tau_1} - S_{\tau_0}, S_{\tau_2} - S_{\tau_1}, \dots S_{\tau_k} - S_{\tau_{k-1}}$$

Ορίζουμε

$$\gamma = P(\tau_1 < \infty)$$

Όμως

$$\tau_1 = \min\{j \geq 1 : S_j > A + j(\mathbb{E}\xi_\iota - \varepsilon)\}$$

Έτσι

$$\gamma = P(\min\{j \geq 1 : S_j > A + j(\mathbb{E}\xi_\iota - \varepsilon)\} < \infty)$$

Από την παραπάνω σχέση παρατηρούμε πως το $\gamma \to \infty \; \acute{o}\tau\alpha\nu \; A \to \infty$

Τη χρονική στιγμή τ_1 ο τυχαίος περίπατος θα βρίσκεται στη θέση S_{τ_1} και η ουρά της τυχαίας μεταβλητής που περιγράφει την θέση του τυχαίου περιπάτου περιγράφεται ως εξής

Έστω πως $\tau_1 = n$ τότε το ενδεχόμενο

$$\{S_{\tau_1} > x\} = \bigcup_{n=1}^{\infty} \{S_n > x, \tau_1 = n\}$$

$$P(S_{\tau_1} > x) = P\left(\bigcup_{n=1}^{\infty} \{S_n > x, \tau_1 = n\}\right) =$$

$$= \sum_{n=1}^{\infty} P(S_n > x, \tau_1 = n)$$

Όμως

$$\{\tau_1 = n\} \subseteq \{S_{n-1} \leq A - (n-1)(a - \varepsilon)\}$$

Και έτσι η πιθανότητα

$$P\big(S_{\tau_1} > x\big) \leq P(S_n > x, S_{n-1} \leq A - (n-1)(a - \varepsilon))$$

Όμως

$$\{S_n > x\} = \{\xi_1 + \xi_2 + \cdots + \xi_n > x\}$$

ισοδύναμα

$$\{S_n > x\} = \{\xi_1 + \xi_2 + \cdots + \xi_{n-1} > x - \xi_n\}$$

και

$$\{S_{n-1} \leq A - (n-1)(a - \varepsilon)\} = \{\xi_1 + \xi_2 + \cdots + \xi_{n-1} \leq A - (n-1)(a - \varepsilon)\}$$

Αν τα παραπάνω δύο ενδεχόμενα ισχύουν ταυτόχρονα τότε

$$x - \xi_n < \xi_1 + \xi_2 + \cdots + \xi_{n-1} \leq A - (n-1)(a - \varepsilon)$$

Και από την παραπάνω διπλή ανισότητα

$$\xi_n \geq x - A + (n-1)(a - \varepsilon)$$

Έτσι η παραπάνω πιθανότητα μπορεί με αντικατάσταση να γραφτεί

$$P\big(S_{\tau_1} > x\big) \leq P(S_n > x, S_{n-1} \leq A - (n-1)(a - \varepsilon))$$

$$= P\left(\sum_{i=1}^{n} \xi_i > x, \sum_{i=1}^{n-1} \xi_i \leq A - (n-1)(a - \varepsilon)\right) =$$

$$= P(\xi_n > x - A + (n-1)(a - \varepsilon))$$

Και η πιθανότητα

$$P\big(S_{\tau_1} > x\big) \leq \sum_{n=1}^{\infty} P(\xi_n > x - A + (n-1)(a - \varepsilon)$$

Η παραπάνω σχέση θα μπορούσε να γραφτεί ως εξής

$$P\left(S_{\tau_1} > x\right) \leq \sum_{n=0}^{\infty} P(\xi_n > x - A + n(a - \varepsilon))$$

Όμως

$$P(\xi_n > x - A + n(a - \varepsilon)) = \bar{F}(x - A + n(a - \varepsilon))$$

$$P\left(S_{\tau_1} > x\right) \leq \sum_{n=0}^{\infty} \bar{F}\,(x - A + n(a - \varepsilon))$$

Παρατηρούμε επίσης πως

$$\bar{F}\big(x - A + n(a - \varepsilon)\big) \leq \bar{F}(y)$$

Για κάθε $y \leq x - A + n(a - \varepsilon)$.

Και έτσι το ολοκλήρωμα

$$\int_{x-A+(n-1)(a-\varepsilon)}^{x-A+n(a-\varepsilon)} \bar{F}\big(x - A + n(a - \varepsilon)\big)dy \leq \int_{x-A+(n-1)(a-\varepsilon)}^{x-A+n(a-\varepsilon)} \bar{F}(y)dy \leq$$

$$\leq \int_{x-A+(n-1)(a-\varepsilon)}^{\infty} \bar{F}(y)dy = \bar{F}_I\,(x - A + n(a - \varepsilon) - \alpha + \varepsilon)$$

Όμως

$$\int_{x-A+(n-1)(a-\varepsilon)}^{x-A+n(a-\varepsilon)} \bar{F}\big(x - A + n(a - \varepsilon)\big)dy =$$

$$= \bar{F}\big(x - A + n(a - \varepsilon)\big) \int_{x-A+(n-1)(a-\varepsilon)}^{x-A+n(a-\varepsilon)} dy =$$

$$= (\alpha - \varepsilon)\bar{F}\big(x - A + n(a - \varepsilon)\big)$$

Από την άλλη μεριά

Λόγο υποεκθετικότητας της κατανομής F_I

$$\bar{F_I}(x - A + n(a - \varepsilon) - \alpha + \varepsilon) = \bar{F_I}(x - A - \alpha + \varepsilon)$$

Σύμφωνα λοιπόν με τις παραπάνω σχέσεις η ανισότητα

$$\int_{x-A+(n-1)(a-\varepsilon)}^{x-A+n(a-\varepsilon)} \bar{F}\big(x - A + n(a - \varepsilon)\big)dy \leq$$

$$\leq \int_{x-A+(n-1)(a-\varepsilon)}^{\infty} \bar{F}(y)dy = \bar{F}_I\,(x - A + n(a - \varepsilon) - \alpha + \varepsilon)$$

Μπορεί να γραφτεί ως εξής

$$(\alpha-\varepsilon)\bar{F}\big(x-A+n(\alpha-\varepsilon)\big)\leq \overline{F_I}(x-A-\alpha+\varepsilon)$$

Έτσι καταλήγουμε στη σχέση

$$\bar{F}\big(x-A+n(\alpha-\varepsilon)\big)\leq \frac{\overline{F_I}(x-A-\alpha+\varepsilon)}{(\alpha-\varepsilon)}$$

$$P\big(S_{\tau_1}>x\big)\leq \frac{\overline{F_I}(x-A-\alpha+\varepsilon)}{(\alpha-\varepsilon)}$$

Τώρα θα ορίσουμε τις ισόνομες και ανεξάρτητες τυχαίες μεταβλητές $\varphi_1,\varphi_2,\dots,\varphi_n$

με δεξιά ουρά ως εξής

$$P(\varphi_1>x)=P(S_{\tau_1}>x/\tau_1<\infty)$$

Στην ουσία ορίσαμε τη δεσμευμένη πιθανότητα ως προς τον χρόνο ανανέωσης $\tau_1<\infty$ (δηλαδή να γίνει ανανέωση)

Κάνοντας χρήση του θεωρήματος του Bayes έχουμε

$$P\big(S_{\tau_1}>x/\tau_1<\infty\big)=\frac{P(S_{\tau_1}>x,\tau_1<\infty)}{P(\tau_1<\infty)}=\frac{P(S_{\tau_1}>x,\tau_1<\infty)}{\gamma}$$

Και έτσι

$$P(\varphi_1>x)\leq\frac{1}{\gamma}P\big(S_{\tau_1}>x\big)\leq\frac{1}{\gamma}\frac{1}{\alpha-\varepsilon}\overline{F_I}(x)$$

Θέτουμε

$$\bar{G}(x)=\frac{1}{\gamma}\frac{1}{\alpha-\varepsilon}\overline{F_I}(x)\quad x\to\infty$$

Ισοδύναμα η παραπάνω σχέση γράφεται

$$\frac{\bar{G}(x)}{\overline{F_I}(x)}=\frac{1}{\gamma(\alpha-\varepsilon)}\qquad x\to\infty$$

Θα προσπαθήσουμε να υπολογίσουμε την κατανομή που ακολουθεί το

$$S_n=\xi_1+\xi_2+\cdots+\xi_n$$

Με $S_0=0\ \kappa\alpha\iota\ \xi_i$ να παίρνουν τιμές στο $\mathbb{Z}_+$

Το ενδεχόμενο

$$\{S_{\tau_1} > x\} = \bigcup_{n=0}^{\infty} \{S_{\tau_1} > x, \tau_1 = n\}$$

Και η αντίστοιχη πιθανότητα

$$P(S_{\tau_1} > x) = \sum_{n=0}^{\infty} P(S_{\tau_1} > x, \tau_1 = n) =$$

$$= \sum_{n=0}^{\infty} P(S_{\tau_1} > x/\tau_1 = n)P(\tau_1 = n) = \sum_{n=0}^{\infty} P(\xi_1 + \xi_2 + \cdots + \xi_n > x)P(\tau_1 = n)$$

$$= \sum_{n=0}^{\infty} \bar{F}^{*n}(x)P(\tau_1 = n)$$

Αν επιπλέον η κατανομή F ανήκει στην οικογένεια των υποεκθετικών κατανομών τότε

$$\bar{F}^{*n}(x) = n\bar{F}(x) \quad x \to \infty$$

Και η πιθανότητα

$$P(S_{\tau_1} > x) = \sum_{n=0}^{\infty} n\bar{F}(x)\, P(\tau_1 = n) = \bar{F}(x) \sum_{n=0}^{\infty} n\, P(\tau_1 = n)$$

$$= \bar{F}(x)\mathbb{E}(T) \quad x \to \infty$$

Ισοδύναμα η παραπάνω σχέση γράφεται

$$\frac{P(S_{\tau_1} > x)}{\bar{F}(x)} \to \mathbb{E}(T) \;\; x \to \infty$$

Ο χρόνος τ_1 είναι ο χρόνος που για πρώτη φορά ο τυχαίος περίπατος θα περάσει ένα φράγμα . έτσι μπορούμε να πούμε πως η τυχαία μεταβλητή $\tau_1 \sim Ge(1-\gamma)$ και έχει συνάρτηση πιθανότητας

$$P(\tau_1 = n) = (1-\gamma)\gamma^n$$

$$P(S_{\tau_1} > x) = (1-\gamma) \sum_{n=0}^{\infty} \bar{G}^{*n}(x)\gamma^n$$

και

$$\frac{P(S_{\tau_1} > x)}{\bar{G}(x)} \to \mathbb{E}(T) = \frac{\gamma}{1-\gamma} \;\; x \to \infty$$

ή ισοδύναμα

$$P(S_{\tau_1} > x) \to \frac{\gamma}{1-\gamma}\bar{G}(x) \;\; x \to \infty$$

Έτσι

$$\frac{\gamma}{1-\gamma}\bar{G}(x) = (1-\gamma)\sum_{n=0}^{\infty}\bar{G}^{*n}(x)\gamma^n$$

$$\sum_{n=0}^{\infty}\bar{G}^{*n}(x)\gamma^n = \frac{\gamma}{(1-\gamma)^2}\bar{G}(x) \quad x \to \infty$$

Έχουμε ήδη δείξει πως

$$\bar{G}(x) = \frac{1}{\gamma}\frac{1}{\alpha-\varepsilon}\overline{F_I}(x) \quad x \to \infty$$

Και έτσι

$$\sum_{n=0}^{\infty}\bar{G}^{*n}(x)\gamma^n = \frac{\gamma}{(1-\gamma)^2}\frac{1}{\gamma}\frac{1}{\alpha-\varepsilon}\overline{F_I}(x) \quad x \to \infty$$

$$\sum_{n=0}^{\infty}\bar{G}^{*n}(x)\gamma^n \sim \frac{\gamma}{(1-\gamma)^2}\frac{1}{\gamma}\frac{1}{\alpha-\varepsilon}\overline{F_I}(x) \quad x \to \infty$$

Τέλος το ενδεχόμενο

$$\{M > x\} = \{\sup(S_n : n \geq 0) > x\} \subseteq \left\{S_{\tau_k} > x - A + a - \varepsilon\right\}$$

Με αντίστοιχη πιθανότητα

$$P(M > x) \leq P(S_{\tau_k} > x - A + a - \varepsilon) \leq \frac{\gamma}{(1-\gamma)^2}\bar{G}(x)$$

$$\leq \frac{\gamma}{(1-\gamma)^2}\frac{1}{\gamma}\frac{1}{\alpha-\varepsilon}\overline{F_I}(x)$$

$$\frac{P(M > x)}{\overline{F_I}(x)} \leq \frac{\gamma}{(1-\gamma)^2}\frac{1}{\gamma}\frac{1}{\alpha-\varepsilon}$$

Αν $\gamma \to 0 \; και \; \varepsilon \to 0$

$$\frac{P(M > x)}{\overline{F_I}(x)} \leq \frac{1}{\alpha}$$

8.5 Συνάρτηση κατανομής του μεγίστου του τυχαίου περιπάτου

Θα προσπαθήσουμε τώρα να βρούμε την πιθανότητα

$$P(M \leq x) = \sum_{k=0}^{\infty}P(\sum_{i=1}^{N}Y^+{}_i \leq x, N = k) = \sum_{k=0}^{\infty}P(Y^+{}_i \leq x/N = k)P(N = k)$$

Με

$$N = max\{k \geq 1: n_+(k) < \infty\}$$

Μπορούμε δηλαδή να ισχυριστούμε πως η τυχαία μεταβλητή N ακολουθεί την γεωμετρική κατανομή με παράμετρο p. Και η πιθανότητα

$$P(N = k) = p(1-p)^k$$

Έτσι το μέγιστο του τυχαίου περιπάτου M θα ακολουθεί σύνθετη γεωμετρική κατανομή.

Τελικά προκύπτει ότι

$$P(M \leq x) = \sum_{k=0}^{\infty} P({Y^+}_i \leq x/N = k)P(N = k) = \sum_{k=0}^{\infty} P({Y^+}_i \leq x/N = k)p(1-p)^k$$

$$= \sum_{k=0}^{\infty} G^{*k}(x)p(1-p)^k$$

8.6 renewal measures

Ορίζουμε στον θετικό ημιάξονα $\mathbb{R}_+$ το εξής μέτρο

$$H_+(B) = \mathbb{I}\{0 \in B\} + \sum_{n=1}^{\infty} P(S_1 > 0, S_2 > 0, S_3 > 0, \dots, S_n > 0, S_n \in B)$$

Ιδιαίτερο ενδιαφέρον παρουσιάζει η από κοινού πιθανότητα μέσα στο άθροισμα της παραπάνω σχέσης την οποία και θα αναλύσουμε.

$$P(S_1 > 0, S_2 > 0, S_3 > 0, \dots, S_n > 0, S_n \in B) =$$

$$P(S_n - S_{n-1} > 0, S_n - S_{n-2} > 0, \dots, S_n - S_1 > 0, S_n - S_0 > 0, S_n \in B)$$

Φαίνεται από την παραπάνω σχέση πως η θέση του τυχαίου περιπάτου τη χρονική στιγμή S_n είναι η υψηλότερη που έχει πάρει ο τυχαίος περίπατος η θέση θα ισούται με το άθροισμα των *overshoots.*

$$S_n = \sum_{i=1}^{n} {Y^+}_i$$

$$P\left(\sum_{i=1}^{n} {Y^+}_i \in B\right) = \sum_{k=1}^{\infty} P(\sum_{i=1}^{n} {Y^+}_i \in B \Big/ n = k)P(n + (n) = k)$$

$$= \sum_{k=1}^{\infty} \sum_{i=1}^{n} p(1-p)^k P({Y^+}_1 + {Y^+}_2 + \ldots + {Y^+}_k \in B)$$

Η πιθανότητα

$$P(x \in A) = \mathbb{E}(\mathbb{I}\{x \in A\}$$

Και αυτό διότι

$$\mathbb{I}\{x \in A\} = \begin{cases} 1 \ \alpha\nu \ x \in A \\ 0 \ \alpha\nu \ x \notin A \end{cases}$$

Και η μέση τιμή της παραπάνω δείκτριας συνάρτησης είναι

$$\mathbb{E}(\mathbb{I}\{x \in A\} = 1P(x \in A\ ) + 0P(x \notin A) = P(x \in A\ )$$

Σύμφωνα λοιπόν με αυτή την παρατήρηση έχουμε

$$H_+(B) = \mathbb{I}\{0 \in B\} + \sum_{n=1}^{\infty} P(\, S_1 > 0, S_2 > 0, S_3 > 0, \dots, S_n > 0, S_n \in B)$$

$$= \mathbb{I}\{0 \in B\} + \sum_{k=1}^{\infty} p(1-p)^k \sum_{i=1}^{n} \mathbb{E}(\mathbb{I}\{Y^+{}_1 + Y^+{}_2 +..+ Y^+{}_k \in B\}$$

Εάν $B \subseteq (0, +\infty)$ τότε

$$H_+((0,+\infty)) = \mathbb{I}\{0 \in (0,+\infty)\} + \sum_{n=1}^{\infty} P(\, S_1 > 0, S_2 > 0, \dots, S_n > 0, S_n \in (0,+\infty))$$

$$= \sum_{k=1}^{\infty} p(1-p)^k \sum_{i=1}^{n} \mathbb{E}(\mathbb{I}\{Y^+{}_1 + Y^+{}_2 +..+ Y^+{}_k \in (0,+\infty)\}$$

$$= \sum_{k=1}^{\infty} kp(1-p)^k$$

Το παραπάνω άθροισμα ισούται με

$$\sum_{k=1}^{\infty} kp(1-p)^k = p\sum_{k=1}^{\infty} k(1-p)^k = p\sum_{k=1}^{\infty} kq^k = p(1-p)\sum_{k=1}^{\infty} \frac{\partial}{\partial q} q^k$$

$$= p\frac{\partial}{\partial q}\left(\sum_{k=1}^{\infty} q^k\right) = p\frac{\partial}{\partial q}(q + q^2 + \cdots)$$

Το άθροισμα στην παραπάνω παρένθεση δεν είναι τίποτε άλλο από ένα άθροισμα απείρων όρων γεωμετρικής προόδου.

Έτσι

$$\sum_{k=1}^{\infty} q^k = q + q^2 + \cdots = \frac{q}{1-q}$$

Και έτσι το άθροισμα

$$\sum_{k=1}^{\infty} kp(1-p)^k = p\frac{\partial}{\partial q}\left(\frac{q}{1-q}\right) = \frac{pq}{p^2} = \frac{q}{p} = \frac{1-p}{p}$$

Και το μέτρο

$$H_+([0,+\infty)) = \mathbb{I}\{0 \in [0,+\infty)\} + \sum_{n=1}^{\infty} P(\,S_1 > 0, S_2 > 0, \dots, S_n > 0, S_n \in (0,+\infty))$$

$$= 1 + \sum_{k=1}^{\infty} p(1-p)^k \sum_{i=1}^{n} \mathbb{E}(\mathbb{I}\{Y^+{}_1 + Y^+{}_2 + .. + Y^+{}_k \in (0,+\infty)\}$$

$$= 1 + \sum_{k=1}^{\infty} kp(1-p)^k$$

$$= 1 + \frac{1-p}{p} = \frac{1}{p}$$

Αντίστοιχα με την μεταβλητή

$$n_+(1) = n_+ = min\{k \geq 1 : S_k > 0\}$$

Μπορούμε να ορίσουμε την

$$n_-(1) = n_- = min\{k \geq 1 : S_k \leq S_0 = 0\}$$

Έτσι το ενδεχόμενο

$$\{n_-(1) = k\} = \{S_1 > 0, S_2 > 0, \dots S_{k-1} > 0, , S_k \leq 0\}$$

$$= \{S_0 - S_1 < 0, S_1 - S_2 \leq 0, S_2 - S_3 \leq 0, \dots, S_k - S_{k-1} \leq 0\}$$

Ορίζουμε επίσης τη μεταβλητή

$$n_-(n) = min\left\{k \geq n_-(n-1) : S_k \leq S_{n_-(n-1)}\right\}$$

$$= \min\left\{k \geq n_-(n-1) : S_k \leq S_{n_-(n-1)} \leq S_{n_-(n-2)} \dots \leq S_1 \leq S_0 = 0\right\}$$

Πρόκειται δηλαδή για την χρονική στιγμή κατά την οποία ο τυχαίος περίπατος θα βρεθεί σε ακόμη αρνητικότερη θέση από την προηγούμενη πιο αρνητική δηλαδή για το χρονικό εκείνο σημείο κατά το οποίο ο τυχαίος περίπατος θα επιτύχει καινούργιο συνολικό ελάχιστο.

Αντίστοιχα ορίζουμε

$$n_{-}(n+1) = min\{k \geq n_{-}(n) : S_k \leq S_{n_{-}(n)}\}$$

Στην ουσία το ενδεχόμενο $\{n_{-}(n) = k\}$ θα είναι ισοδύναμο ή ισοπίθανο με το ενδεχόμενο

$$\bigcup_{n=1}^{\infty}\{n_{-}(n) = k\} = \{S_k \leq S_{k-1} \leq S_{k-2} \dots \leq S_1 \leq S_0\}$$

Δηλαδή από την παραπάνω σχέση προκύπτει πως

$$\bigcup_{n=1}^{\infty}\{n_{-}(n) = k\} = \{S_k \leq S_{k-1}, S_k \leq S_{k-2} \dots, S_k \leq S_1, S_k \leq S_0\}$$

$$\bigcup_{n=1}^{\infty}\{n_{-}(n) = k\} = \{S_k - S_{k-1} \leq 0, S_k - S_{k-2} \leq 0 \dots, S_k - S_1 \leq 0, S_k \leq 0\}$$

Όμως

$$S_k = \sum_{i=1}^{k} Y_i$$

Έτσι αντικαθιστώντας τα $S_i, i = 1,2,3, \dots k$ με την παραπάνω σχέση έχουμε

$$\bigcup_{n=1}^{\infty}\{n_{-}(n) = k\} = \left\{Y_k \leq 0, Y_k + Y_{k-1} \leq 0 \dots, \sum_{i=2}^{k} Y_i \leq 0, \sum_{i=1}^{k} Y_i \leq 0\right\}$$

Μπορούμε πλέον να υπολογίσουμε την πιθανότητα

$$P\left(\bigcup_{n=1}^{\infty}\{n_{-}(n) = k\}\right) = \mathbb{E}\left(\sum_{n=1}^{\infty} \mathbb{I}\{n_{-}(n) = k\}\right) =$$

$$= P\left(Y_k \leq 0, Y_k + Y_{k-1} \leq 0 \dots, \sum_{i=2}^{k} Y_i \leq 0, \sum_{i=1}^{k} Y_i \leq 0\right)$$

$$= P\left(Y_1 \le 0, Y_1 + Y_2 \le 0 \dots, \sum_{i=1}^{k-1} Y_i \le 0, \sum_{i=1}^{k} Y_i \le 0\right) =$$

$$= P(S_1 \le 0, S_2 \le 0, \dots, S_k \le 0)$$

Όμως αν όλες οι θέσεις του τυχαίου περιπάτου $S_1, S_2, \dots, S_k$ είναι μικρότερες ή ίσες με το μηδέν αυτό ισοδύναμα σημαίνει πως δεν έχει έρθει η στιγμή κατά την οποία για πρώτη φορά ο τυχαίος περίπατος θα βρεθεί σε θετική θέση. Σύμφωνα λοιπόν με αυτή τη διαπίστωση

$$P\left(\bigcup_{n=1}^{\infty}\{n_-(n) = k\}\right) = \mathbb{E}\left(\sum_{n=1}^{\infty} \mathbb{I}\{n_-(n) = k\}\right) =$$

$$= P(S_1 \le 0, S_2 \le 0, \dots, S_k \le 0) = P(n_+(1) \ge k + 1) = \mathbb{E}(\mathbb{I}\{n_+(1) \ge k + 1\})$$

Μπορούμε πλέον να ορίσουμε ένα μέτρο στον $\mathbb{R}_-$

$$H_0^-(B) = \sum_{k=0}^{\infty} (G^-)^{*k}(B)$$

Γνωρίζουμε πως

$$(G^-)^{*0} = \begin{cases} 1 \ \alpha\nu \ 0 \epsilon B \\ 0 \ \alpha\nu \ 0 \notin B \end{cases}$$

Μπορούμε λοιπόν να γράψουμε την αρχική μας σχέση που περιγράφει το μέτρο στον αρνητικό ημιάξονα $\mathbb{R}_-$ ως εξής

$$H_0^-(B) = (G^-)^{*0} + \sum_{k=1}^{\infty} (G^-)^{*k}(B)$$

Ιδιαίτερο ενδιαφέρον έχει το άθροισμα

$$\sum_{k=1}^{\infty} (G^-)^{*k}(B)$$

με το οποίο και θα ασχοληθούμε .

Πριν λοιπόν προχωρήσουμε στην ανάπτυξη του αθροίσματος θα εισάγουμε στην επιμέρους ανάλυση μας μια νέα μεταβλητή την

$$Y^- = \begin{cases} S_{n-(1)} & \alpha\nu \ \ n_-(1) < \infty \\ \infty & \alpha\nu \ \ n_-(1) = \infty \end{cases}$$

Γενικότερα ορίζουμε την μεταβλητή υπέρβασης

$$Y^-{}_{n-(k)} = \begin{cases} S_{n-(k)} - S_{n-(k-1)} & \alpha\nu \ \ n + (k) < \infty \\ \infty & \alpha\nu \ n + (k) = \infty \end{cases}$$

Και ονομάζεται νιοστή φθίνουσα κλίμακα ύψους (*n-th descending ladder height)*

Η συνέλιξη

$$\sum_{k=1}^{\infty}(G^{-})^{*k}\,(B)=\sum_{k=1}^{\infty}P(Y_1^{-}+Y_2^{-}+\cdots+Y_k^{-}\epsilon B)$$

Όμως

$$\sum_{i=1}^{k}Y_i^{-}=Y_1^{-}+Y_2^{-}+\cdots+Y_k^{-}=$$

$$(S_{n-(1)}-0)+(S_{n-(2)}-S_{n-(1)})+\cdots+(S_{n-(k)}-S_{n-(k-1)})$$

$$=S_{n-(k)}$$

Σύμφωνα λοιπόν με την παραπάνω ανάλυση το άθροισμα

$$\sum_{n=1}^{\infty}(G^{-})^{*n}\,(B)=\sum_{n=1}^{\infty}P(Y_1^{-}+Y_2^{-}+\cdots+Y_n^{-}\epsilon B)=\sum_{n=1}^{\infty}P(\,S_{n-(n)}\epsilon B)$$

Και το μέτρο που ορίσαμε στον αρνητικό ημιάξονα τώρα μπορεί να γραφτεί ως εξής

$$H_0^{-}(B)=(G^{-})^{*0}+\sum_{n=1}^{\infty}P(\,S_{n-(n)}\epsilon B)=\mathbb{I}\{0\epsilon B\}+\sum_{n=1}^{\infty}\mathbb{E}\mathbb{I}\{S_{n-(n)}\epsilon B\}$$

$$=\mathbb{I}\{0\epsilon B\}+\mathbb{E}\left\{\sum_{n=1}^{\infty}\mathbb{I}\{S_{n-(n)}\epsilon B\}\right\}$$

Όμως το ενδεχόμενο $\{S_{n-(n)}\epsilon B\}$ μπορούμε με τη βοήθεια τομών και ενώσεων να γραφεί ως εξής

$$\{S_{n-(n)}\epsilon B\}=\bigcup_{k=1}^{\infty}\left(\{S_{n-(n)}\epsilon B\}\bigcap\{n-(n)=k\}\right)$$

Και η δείκτρια συνάρτηση

$$\mathbb{I}\{S_{n-(n)}\epsilon B\}=\mathbb{I}\left\{\bigcup_{k=1}^{\infty}\left(\{S_{n-(n)}\epsilon B\}\bigcap\{n-(n)=k\}\right)\right\}=$$

$$=\sum_{k=1}^{\infty}\mathbb{I}\left(\{S_{n-(n)}\epsilon B\}\bigcap\{n-(n)=k\}\right)=\sum_{k=1}^{\infty}\mathbb{I}\left(\{S_{n-(n)}\epsilon B\},\{n-(n)=k\}\right)$$

Και η

$$H_0^-(B) = \mathbb{I}\{0\epsilon B\} + \mathbb{E}\left(\sum_{n=1}^{\infty}\sum_{k=1}^{\infty} \mathbb{I}\left(\{S_{n-(n)}\epsilon B\}, \{n-(n)=k\}\right)\right) =$$

$$= \mathbb{I}\{0\epsilon B\} + \mathbb{E}\left(\sum_{n=1}^{\infty}\sum_{k=1}^{\infty} \mathbb{I}\left(\{S_{n-(n)}\epsilon B\}\mathbb{I}\{n-(n)=k\}\right)\right) =$$

$$= \mathbb{I}\{0\epsilon B\} + \sum_{k=1}^{\infty} \mathbb{E}\left(\sum_{n=1}^{\infty}\left(\mathbb{I}\left(\{S_{n-(n)}\epsilon B\}, \{n-(n)=k\}\right)\right)\right) =$$

$$= \mathbb{I}\{0\epsilon B\} + P\left(Y_k \le 0, Y_k + Y_{k-1} \le 0 \dots, \sum_{i=2}^{k} Y_i \le 0, \sum_{i=1}^{k} Y_i \le 0, S_k \epsilon B\right) =$$

$$= \mathbb{I}\{0\epsilon B\} + P\left(Y_1 \le 0, Y_1 + Y_2 \le 0 \dots, \sum_{i=1}^{k-1} Y_i \le 0, \sum_{i=1}^{k} Y_i \le 0, S_k \epsilon B\right) =$$

$$= \mathbb{I}\{0\epsilon B\} + \sum_{k=1}^{\infty} P(S_1 \le 0, S_2 \le 0, \dots, S_k \le 0, S_k \epsilon B) =$$

$$= \mathbb{I}\{0\epsilon B\} + \sum_{k=1}^{\infty} P(n_+(1) \ge k+1, S_k \epsilon B) =$$

$$= \mathbb{I}\{0\epsilon B\} + \sum_{k=1}^{\infty} \mathbb{E}(\mathbb{I}\{n_+(1) \ge k+1, S_k \epsilon B\})$$

$$= \mathbb{I}\{0\epsilon B\} + \mathbb{E}\left(\sum_{k=1}^{\infty} (\mathbb{I}\{n_+(1) \ge k+1, S_k \epsilon B\})\right)$$

$$= \mathbb{I}\{0\epsilon B\} + \sum_{k=1}^{n_+(1)-1} \mathbb{I}\{S_k \epsilon B\}$$

8.7 Ταυτότητα του Wald (2η απόδειξη)

Έστω τ μια ακέραια τυχαία μεταβλητή τέτοια ώστε για κάθε n το ενδεχόμενο $\{t \ge n\}$ εξαρτάται από τις τυχαίες μεταβλητές $\xi_1, \xi_1, \dots, \xi_{n-1}$ και δεν εξαρτάται από τις τυχαίες μεταβλητές $\xi_n, \xi_{n+1}, \dots$

Τότε

$$\mathbb{E}(S_\tau) = \mathbb{E}(\tau)\mathbb{E}(\xi)$$

Όπως έχουμε ήδη αναφέρει η παραπάνω ταυτότητα είναι γνωστή ως ταυτότητα του *Little*.Οι παραπάνω μέσες τιμές $\mathbb{E}(\tau)$ και $\mathbb{E}(\xi)$ θα πρέπει να είναι πεπερασμένες.

Απόδειξη

Αρχικά θα υποθέσουμε πως η τυχαία μεταβλητή τα είναι φραγμένη από μια ακέραια θετική τιμή M

Η θέση του τυχαίου περιπάτου S_τ στη χρονική στιγμή τ μπορεί να γραφεί με τη βοήθεια του ακόλουθου τεχνάσματος

$$S_\tau = \sum_{n=0}^{M} S_\tau \, \mathbb{I}\{\tau = n\} = \sum_{k=1}^{M} \xi_\kappa \, \mathbb{I}\{\tau \geq k\}$$

Η μέση τιμή της παραπάνω ποσότητας είναι ίση με

$$\mathbb{E}(S_\tau) = \mathbb{E}\left(\sum_{k=1}^{M} \xi_\kappa \, \mathbb{I}\{\tau \geq k\}\right)$$

Επειδή το άθροισμα μας είναι πεπερασμένο δεν υπάρχει κάποιο πρόβλημα να περάσουμε την μέση τιμή μέσα και έτσι

$$\mathbb{E}(S_\tau) = \sum_{n=1}^{M} \mathbb{E}(S_\tau; \tau = n) = \sum_{n=1}^{M} \mathbb{E}\,(S_\tau)\mathbb{I}\{\tau = n\} = \sum_{n=1}^{M}\sum_{k=1}^{n} \mathbb{E}(\xi_k)\mathbb{I}\{\tau = n\}$$

$$= \sum_{k=1}^{M} \mathbb{E}\xi_k \mathbb{I}\{\tau \geq k\} = \sum_{k=1}^{M} \mathbb{E}(\xi_k; \tau \geq k)$$

Όμως όπως έχουμε ήδη αναφέρει η το ενδεχόμενο $\{\tau \geq k\}$ δεν εξαρτάται από το κ οστό βήμα του τυχαίου περιπάτου και επίσης αν το βήμα του τυχαίου περιπάτου είναι ίσο με την ποσότητα

$$\xi_i = X_i - cT_i$$

Τότε η μέση τιμή του βήματος θα ισούται με

$$\mathbb{E}\xi_\iota = \mathbb{E}\xi = \mathbb{E}X - c\mathbb{E}T = \mu$$

$$\mathbb{E}(S_\tau) = \mathbb{E}\xi_k \mathbb{E}\mathbb{I}\{\tau \geq k\} = \sum_{k=1}^{M} \mathbb{E}(\xi_k) P(\tau \geq k)$$

Και τελικά

$$\mathbb{E}(S_\tau) = \mu \sum_{k=1}^{M} P(\tau \geq k) = \mu \mathbb{E}T$$

Στην περίπτωση που η τυχαία μεταβλητή τ δεν είναι φραγμένη τότε θα πρέπει να είμαστε ιδιαίτερα προσεκτικοί με την χρήση της μέσης τιμής και τη μετάβαση της μέσα στο άθροισμα.

Χρησιμοποιώντας το θεώρημα κυριαρχούμενης σύγκλισης και με την συνθήκη πως οι $|\xi_k|$ και $\mathbb{I}\{\tau \geq k\}$ είναι ανεξάρτητες έχουμε

$$\sum_{k=1}^{\infty} \mathbb{E}|\xi_k|\mathbb{I}\{\tau \geq k\} = \mathbb{E}|\xi_k|\mathbb{E}\mathbb{I}\{\tau \geq k\} = \sum_{k=1}^{\infty} \mathbb{E}|\xi_k|(P(\tau \geq k) = \mathbb{E}|\xi_k| \sum_{k=1}^{\infty} P(\tau \geq k)$$

$$= \mathbb{E}|\xi_k|\mathbb{E}T < \infty$$

Έτσι και ενώ ισχύει το θεώρημα κυριαρχούμενης σύγκλισης έχουμε

$$\mathbb{E}(S_\tau) = \mathbb{E}\left(\sum_{k=1}^{\infty} \xi_\kappa \, \mathbb{I}\{\tau \geq k\}\right)$$

$$= \sum_{n=1}^{\infty} \mathbb{E}(S_\tau; \tau = n) = \sum_{n=1}^{\infty} \mathbb{E}\,(S_\tau)\mathbb{I}\{\tau = n\} = \sum_{n=1}^{\infty}\sum_{k=1}^{n} \mathbb{E}(\xi_k)\mathbb{I}\{\tau = n\}$$

$$= \sum_{k=1}^{\infty} \mathbb{E}\xi_k\mathbb{I}\{\tau \geq k\} = \sum_{k=1}^{\infty} \mathbb{E}(\xi_k; \tau \geq k)$$

8.8 Pollaczek-Khinchin formulae

Μέχρι τώρα έχουμε υποθέσει πως

$$U_n = u + c\sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} X_i = u - \sum_{i=1}^{n}(X_i - cY_i)$$

$$= u - \sum_{i=1}^{n} X_i - c\sum_{i=1}^{n} Y_i = u - \sum_{i=1}^{n} X_i + cT_i$$

$$= u - (\textstyle\sum_{i=1}^{n} X_i - cT_i)$$

Η πιθανότητα χρεοκοπίας θα συμβεί με πιθανότητα τη μονάδα αν τα έξοδα της ασφαλιστικής εταιρείας είναι μεγαλύτερα σε μέγεθος από τα έσοδα. Με άλλα λόγια αν το ποσοστό κέρδους του ασφαλιστή είναι μικρότερο της μονάδας τότε το πλεόνασμα της ασφαλιστικής εταιρείας κάποια στιγμή στο μέλλον θα γίνει αρνητικό.

Σύμφωνα λοιπόν με την παραπάνω σχέση αρνητικό πλεόνασμα έχουμε όταν

$$\mathbb{E}\left(\sum_{i=1}^{n} X_i - cT_i\right) = \sum_{i=1}^{n} \mathbb{E}\,(X_i) - c\mathbb{E}(T_i) > 0$$

Το περιθώριο ασφαλείας της ασφαλιστικής εταιρείας (που εκφράζει το ποσοστιαίο κέρδος του ασφαλιστή) δίνεται όπως έχουμε αναφέρει σε προηγούμενο κεφάλαιο από τη σχέση

$$\theta = \frac{c}{\lambda\mathbb{E}(X)} - 1 = \frac{c - \lambda\mathbb{E}(X)}{\lambda\mathbb{E}(X)} = \frac{1}{\rho} - 1$$

Όπου

$$\rho = \frac{\lambda\mathbb{E}(X)}{c}$$

Από την άλλη πλευρά ο λόγος

$$\frac{c\mathbb{E}(Y_i) - \mathbb{E}(X_i)}{\mathbb{E}(X_i)}$$

Μας δείχνει τι μέρος των μέσων αποζημιώσεων καλύπτουν τα καθαρά έσοδα.

Κάνοντας κάποιες πράξεις στην παραπάνω σχέση έχουμε

$$\frac{c\mathbb{E}(Y_i)}{\mathbb{E}(X_i)} - \frac{\mathbb{E}(X_i)}{\mathbb{E}(X_i)} = \frac{c\mathbb{E}(Y_i)}{\mathbb{E}(X_i)} - 1$$

Έτσι καταλήγουμε στο συμπέρασμα πως

$$\frac{1}{\rho} = \frac{c\mathbb{E}(Y_i)}{\mathbb{E}(X_i)}$$

Αν λοιπόν ο συντελεστής $\rho > 1$ τότε με πιθανότητα τη μονάδα κάποια στιγμή η ασφαλιστική θα αποκτήσει αρνητικό πλεόνασμα.

Αν όμως ο συντελεστής $\rho \leq 1$ τότε υπάρχει θετική πιθανότητα το πλεόνασμα να γίνει κάποια στιγμή αρνητικό και αυτή θα πρέπει να υπολογίσουμε.

Όπως έχουμε ήδη αναφέρει το πλεόνασμα της ασφαλιστικής εταιρείας τη χρονική στιγμή $T = n$ δίνεται από τη σχέση

$$U_n = u\, + c\sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} X_i = u\, - \sum_{i=1}^{n}(X_i - cY_i)$$

Η πιθανότητα χρεοκοπίας όπως την έχουμε ορίσει μέχρι στιγμής είναι η πιθανότητα το πλεόνασμα της επιχείρησης να γίνει κάποια στιγμή στο μέλλον αρνητικό δηλαδή

$$P(U_n < 0, για\ κάποιο\ n = 1,2, \dots)$$

Η παραπάνω σχέση θα μπορούσε να γραφτεί ως εξής

$$P(u - \sum_{i=1}^{n}(X_i - cY_i) < 0, \text{για κάποιο } n = 1,2,\dots)$$
$$= P(\sum_{i=1}^{n}(X_i - cY_i) > u\,, \text{για κάποιο } n = 1,2,\dots)$$

Μπορούμε λοιπόν να θεωρήσουμε την ποσότητα

$$W_n = \sum_{i=1}^{n}(X_i - cY_i) = \sum_{i=1}^{n} W_i$$

Ως ένα τυχαίο περίπατο με βήμα

$$W_i = X_i - cY_i$$

Το βήμα του τυχαίου περιπάτου μας εκφράζει τα καθαρά έσοδα της εταιρείας μετά την επέλευση της i αποζημίωσης.

Εναλλακτικά μπορούμε να σκεφτούμε ότι ο τυχαίος περίπατος θα περάσει το απορροφητικό φράγμα u_0 όταν το μέγιστο M του τυχαίου περιπάτου περάσει το u_0 .

$$\psi(u) = P(M > u)$$

Θυμίζουμε πως το μέγιστο ενός τυχαίου περιπάτου ορίζεται ως εξής

$$M = max\{0, S_1, S_2, \dots, S_n, \dots.\} = max\{S_n, n \geq 1\}$$

Έχουμε ήδη δείξει πως το μέγιστο ενός τυχαίου περιπάτου ισούται με

$$M = \sum_{i=1}^{N} w_i^{+}$$

Όπου N είναι μια τυχαία μεταβλητή που μετράει τον αριθμό των επαναλαμβανόμενων προσπαθειών ώστε ο τυχαίος περίπατος να απορροφηθεί στο u_0.Στην ουσία μετράει αριθμό προσπαθειών μέχρι την πρώτη επιτυχία και μπορεί να περιγραφεί με την γεωμετρική κατανομή με πιθανότητα επιτυχίας p και αποτυχίας $1 - p$

Η πιθανότητα χρεοκοπίας λοιπόν γράφεται

$$\psi(u) = P(M > u\,) = P\left(\sum_{i=1}^{N} w_i^+ > u\,\right) = \sum_{n=1}^{\infty} P(\sum_{i=1}^{N} w_i^+ > u \Big/ N = n)P(N = n)$$

$$\sum_{n=1}^{\infty} P(\sum_{i=1}^{N} w_i^+ > u \Big/ N = n)(1-p)^n p =$$

$$= p\sum_{n=1}^{\infty}(1-p)^n\, P(w_1^+ + w_2^+ +..+ w_n^+ > u_0)$$

$$= p\sum_{n=1}^{\infty}(1-p)^n\, \bar{G}^{*n}(u\,)$$

Έτσι η πιθανότητα χρεοκοπίας γράφεται ως εξής

$$\psi(u) = p\sum_{n=1}^{\infty}(1-p)^n\, \bar{G}^{*n}(u\,) = p\sum_{n=1}^{\infty}(1-p)^n \int_0^{u_0} \bar{G}\;\,(u\,-y)dG^{*(k-1)}(y)$$

8.9 Η πιθανότητα χρεοκοπίας και υποεκθετικές κατανομές

Αν η κατανομή G ανήκει στην οικογένεια των υποεκθετικών κατανομών

$$\frac{\psi(u)}{\bar{G}\;(u)} = p\sum_{n=1}^{\infty}(1-p)^n \frac{\bar{G}^{*n}(u)}{\bar{G}\;(u)} \sim p^2 \sum_{n=1}^{\infty} n(1-p)^{n-1} \qquad u \to \infty$$

Όμως

$$\sum_{n=1}^{\infty} n(1-p)^{n-1} = \frac{1}{p^2}$$

Καταλήξαμε λοιπόν στο συμπέρασμα πως

$$\frac{\psi(u)}{\bar{G}\;(u)} \sim 1 \quad u \to \infty$$

8.10 Πιθανότητα χρεοκοπίας με την πρώτη αποζημίωση και υποεκθετικές κατανομές

$$\psi(u) = \mathbb{P}(U(t) < 0) = \mathbb{P}(u + cT - \sum_{i=1}^{N(t)} X_i < 0)$$

$$\psi(u) = \int_0^{+\infty} \mathbb{P}\left(\sum_{i=1}^{N(t)} X_i > u + ct\right) f(t)dt$$

Η πιθανότητα να χρεοκοπήσει η επιχείρηση από την πρώτη αποζημίωση είναι

$$\psi(u) = \int_0^{+\infty} \mathbb{P}(X_1 > u + ct) f(t)dt$$

$$\psi(u) = \int_0^{+\infty} \bar{F}\,(u + ct) f(t)dt$$

Αν η κατανομή των αποζημιώσεων ανήκει στην υποεκθετική οικογένεια κατανομών τότε

$$\frac{\psi(u)}{\bar{F}(u)} \sim 1 \quad u \to \infty$$

Παράρτημα

Κατανομές χρόνων ζωής

Κυριότερες συνεχείς κατανομές χρόνων ζωής

Εκθετική κατανομή

Η απλούστερη κατανομή χρόνου ζωής είναι η εκθετική κατανομή.

Εάν δηλαδή μια συνεχής τυχαία μεταβλητή X ακολουθεί την εκθετική κατανομή έχει συνάρτηση πυκνότητας :

$$f(x) = \begin{cases} \mu e^{-\mu x} & \gamma\iota\alpha\ x \geq 0 \\ 0 & \gamma\iota\alpha\ x < 0 \end{cases}$$

Όπου μ > 0 με μέση τιμή $E(X) = \frac{1}{\mu}$ και διασπορά $Var(X) = \frac{1}{\mu^2}$

και συνάρτηση κατανομής

$$F(x) = \begin{cases} 1 - e^{-\mu x} & \gamma\iota\alpha\ x \geq 0 \\ 0 & \gamma\iota\alpha\ x < 0 \end{cases}$$

Κατανομή Weibull

Η κατανομή Wiebull (από τον Σουηδό W.Wiebull) είναι η πλέον διαδεδομένη κατανομή στην ανάλυση αξιοπιστίας. Εμπειρικά αποτελεί ένα ικανοποιητικό μοντέλο για πολλά φαινόμενα.

Η συνάρτηση πυκνότητας πιθανότητας ορίζεται ως εξής

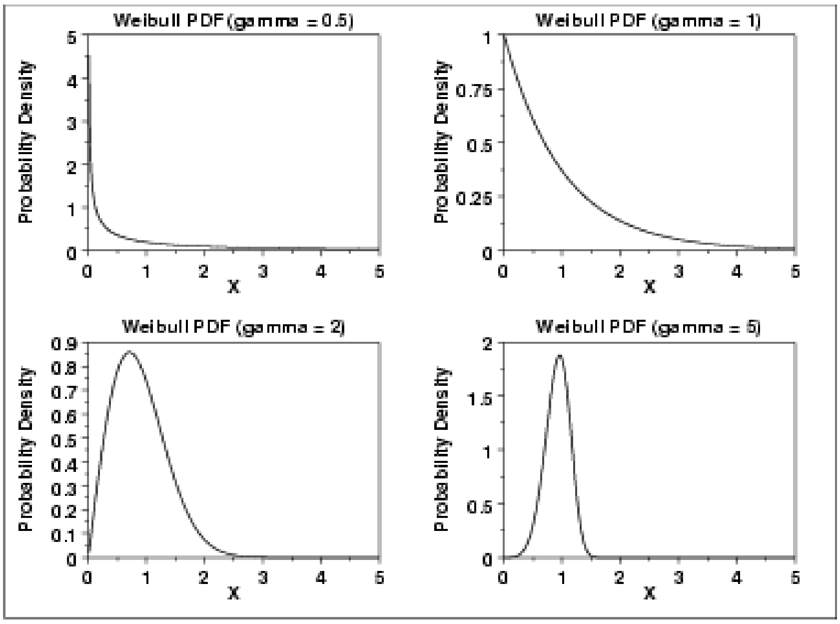
$$f(x) = n a^{-n} t^{n-1} \exp\left\{-\left(\frac{t}{a}\right)^n\right\} \quad t > 0$$

Όπου $\alpha > 0$ είναι η παράμετρος κλίμακας και $n > 0$ είναι η παράμετρος σχήματος.

Η Εκθετική κατανομή αποτελεί ειδική περίπτωση της Weibull όταν $n = 1$

Η συνάρτηση κατανομής της Wiebull είναι η εξής

$$F(x) = 1 - \exp\left\{-\left(\frac{t}{a}\right)^n\right\}$$



Σχήμα7 : *γραφική απεικόνιση κατανομών Gamma με διαφορετικές τιμές της παραμέτρου α*

Ροπές r τάξης για την κατανομή Weibull

$$E(T^r) = \int_0^\infty t^r f(t)dt = \int_0^\infty a^r u^{r/n} e^{-u} du \quad όπου\ u = (\frac{t}{a})^n$$

$$= a^r \int_0^\infty u^{r/n} e^{-u} du = a^r \Gamma\left(1 + \frac{r}{n}\right)$$

Όπου $\Gamma(a) = \int_0^{+\infty} t^{a-1} e^{-t}\, dt \quad a > 0\ η\ συνάρτηση\ Γάμμα.$

Από την παραπάνω σχέση μπορούμε να υπολογίσουμε τις ροπές πρώτης και δεύτερης τάξης ως εξής

$$E(T) = \alpha\Gamma\left(1 + \frac{1}{n}\right)$$

$$E(T^2) = a^2\Gamma\left(1 + \frac{2}{n}\right)$$

Μπορούμε επίσης να υπολογίσουμε την διακύμανση

$$Var(T) = E(T^2) - \{E(T)\}^2 =$$

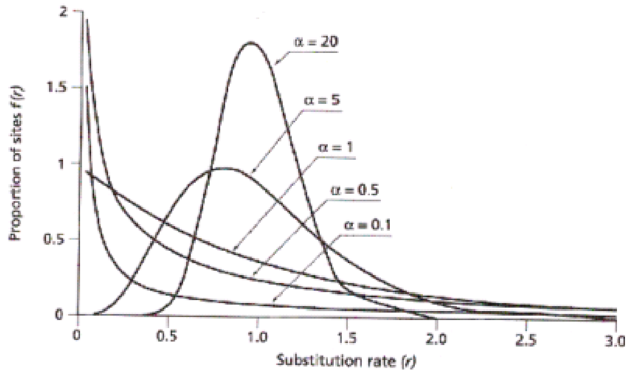
$$= a^2\left[\Gamma\left(1+\frac{2}{n}\right) - \Gamma\left(1+\frac{1}{n}\right)^2\right]$$

Κατανομή Γάμμα

Η κατανομή Γάμμα με παραμέτρους κλίμακας και σχήματος $\lambda > 0\; και\; \alpha > 0$ αντίστοιχα έχει συνάρτηση πυκνότητας πιθανότητας

$f(t) = \frac{\lambda^{\alpha} t^{a-1} e^{-\lambda t}}{\Gamma(\alpha)}$ t>0

Σχήμα 8: *συνάρτησης πυκνότητας της κατανομής Γάμμα για διάφορες τιμές της παραμέτρου σχήματος α.*



Ροπές r τάξης για την κατανομή Γάμμα

$$E(T^r) = \int_0^{\infty} t^r f(t)dt = \int_0^{\infty} \frac{\lambda^{\alpha} t^{a+r-1} e^{-\lambda t}}{\Gamma(\alpha)}\, dt =$$

$$= \frac{\Gamma(\alpha+r)}{\Gamma(\alpha)} \frac{\lambda^{\alpha}}{\lambda^{\alpha+r}} \int_0^{\infty} \frac{\lambda^{\alpha+r} t^{a+r-1} e^{-\lambda t}}{\Gamma(\alpha+r)}\, dt =$$

$$= \frac{\Gamma(\alpha+r)}{\lambda^r \Gamma(\alpha)}$$

Από την παραπάνω σχέση μπορούμε να υπολογίσουμε ροπές πρώτης και δεύτερης τάξης

$$E(T) = \frac{\Gamma(\alpha+1)}{\lambda\Gamma(\alpha)} = \frac{\alpha\Gamma(\alpha)}{\lambda\Gamma(\alpha)} = \frac{\alpha}{\lambda}$$

$$E(T^2) = \frac{\Gamma(\alpha+2)}{\lambda^2\Gamma(\alpha)} = \quad \frac{\alpha(\alpha+1)}{\lambda^2}$$

Και η διασπορά της κατανομής είναι ίση με

$$Var(T) = \ Var(T) = E(T^2) - \{E(T)\}^2 =$$

$$\frac{\alpha(\alpha+1)}{\lambda^2} - (\frac{\alpha}{\lambda})^2 = \frac{a}{\lambda^2}$$

Λογαριθμοκανονική κατανομή

Ως κατανομή χρόνου ζωής μπορούμε να θεωρήσουμε έναν μετασχηματισμό της κανονικής κατανομής τη λεγόμενη Λογαριθμοκανονική κατανομή.
Η κανονική κατανομή (συχνά στη βιβλιογραφία αναφέρεται και ως κατανομή του Gauss)

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(t-\mu)^2}{2\sigma^2}\right\} \qquad t\epsilon(-\infty, +\infty)$$

με μ παράμετρο θέσης να παίρνει οποιαδήποτε τιμή στο R και σ παράμετρο κλίμακας με $\sigma\ \epsilon(0, +\infty)$ σπάνια χρησιμοποιείται ως μοντέλο διάρκειας ζωής.
Αυτό συμβαίνει για δύο λόγους.
Ο πρώτος λόγος είναι ιδιαίτερα προφανής μιας και η τυχαία μεταβλητή περιγράφει τον χρόνο αυτός θα παίρνει θετικές τιμές και όχι αρνητικές. Το ενδεχόμενο $\{T < 0\}$ είναι ένα σημαντικό μειονέκτημα όταν η σχετική πιθανότητα του ενδεχομένου αυτού είναι μεγάλη.
Ο δεύτερος λόγος συνδέεται με τη συμμετρικότητα της κατανομής και αυτό έρχεται σε αντίθεση με τη λοξότητα που εμφανίζεται στην πράξη.
Αν λοιπόν η τυχαία μεταβλητή που περιγράφει τον χρόνο ακολουθεί την λογαριθμοκανονική κατανομή δηλαδή αν

$$T{\sim}LN(\mu, \sigma^2)$$

Τότε η τυχαία μεταβλητή

$$Y = lnT{\sim}\ N(\mu, \sigma^2)$$

Η συνάρτηση κατανομής της T υπολογίζεται ως εξής

$$F(t) = P(T \leq t) = P(Y \leq lnt) =$$

$$P\left(\frac{Y-\mu}{\sigma} \leq \frac{lnt-\mu}{\sigma}\right) = \Phi(\frac{lnt-\mu}{\sigma})=$$

$$\int_{-\infty}^{lnt} \frac{1}{\sigma\sqrt{2\pi}}\exp\{-\frac{(x-\mu)^2}{2\sigma^2}dx = \int_0^t \frac{1}{\sigma u\sqrt{2\pi}}\exp\{-\frac{(lnu-\mu)^2}{2\sigma^2}du$$

όπου $u = e^x$.

Η συνάρτηση πυκνότητας πιθανότητας υπολογίζεται ως εξής

$$f(t) = \frac{d}{dt}\Phi\left(\frac{lnt-\mu}{\sigma}\right) = \Phi'\left(\frac{lnt-\mu}{\sigma}\right)\frac{d}{dt}\left(\frac{lnt-\mu}{\sigma}\right) = \frac{1}{\sigma t\sqrt{2\pi}}\exp\left\{-\frac{(lnt-\mu)^2}{2\sigma^2}\right\} \qquad t > 0$$

Υπολογισμός ροπής r τάξης της λογαριθμοκανονικής κατανομής

Αν η τυχαία μεταβλητή $Y = lnT \sim N(\mu, \sigma^2)$ τότε $T = e^Y \sim LN(\mu, \sigma^2)$

$$E(T^r) = E(e^{Yr}) = M_Y(r) = \int_0^{\infty} t^r f(t)dt = \int_{-\infty}^{\infty} e^{yr} f(y)dy$$

Με $f(y)$ η συνάρτηση πυκνότητας πιθανότητας της κανονικής κατανομής $N(\mu, \sigma^2)$

Μπορούμε να δουλέψουμε το εσωτερικό του παραπάνω ολοκληρώματος ως εξής

$$E(e^{Yr}) = E\left(e^{\frac{\sigma r(y-\mu)}{\sigma}+r\mu}\right) = e^{r\mu}E(e^{\frac{\sigma r(y-\mu)}{\sigma}} = e^{r\mu}E(e^{\sigma rZ}) = e^{r\mu}M_Z(\sigma r)$$

Με $Z = \frac{Y-\mu}{\sigma} \sim N(0,1)$

Η ροπογεννήτρια της τυποποιημένης κανονικής κατανομής Ζ θα δίνεται από τη σχέση

$$M_Z(r) = E(e^{Zr}) = \int_{-\infty}^{\infty} e^{zr} f(z)dz = \int_{-\infty}^{\infty} e^{zr}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{z^2}{2}\right\}dz$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(z^2-2zr)}\,dz = e^{\frac{r^2}{2}}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(z-r)^2}dz = e^{\frac{r^2}{2}}$$

Το παραπάνω αποτέλεσμα προέκυψε διότι η ποσότητα $\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(z-r)^2}$ είναι η συνάρτηση πυκνότητας πιθανότητας μιας $N(r, 1)$ και ως εκ τούτου

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(z-r)^2}dz = 1$$

$$E(T^r) = E(e^{Yr}) = e^{r\mu}M_Z(\sigma r) = e^{r\mu}e^{\frac{(\sigma r)^2}{2}} = e^{r\mu+\frac{(\sigma r)^2}{2}}$$

Σύμφωνα λοιπόν με τον παραπάνω τύπο μπορούμε να υπολογίσουμε τις ροπές πρώτης και δεύτερης τάξης

$$E(T) = e^{\mu+\frac{\sigma^2}{2}}$$

$$E(T^2) = e^{2\mu+\frac{(2\sigma)^2}{2}} = e^{2\mu+2\sigma^2}$$

Και μπορούμε να υπολογίσουμε τη διακύμανση

$$Var(T) = Var(T) = E(T^2) - \{E(T)\}^2 =$$

$$Var(T) = e^{2\mu+2\sigma^2} - \left(e^{\mu+\frac{\sigma^2}{2}}\right)^2$$

Η γενικευμένη κατανομή Γάμμα

Η κατανομές που παρουσιάστηκαν στις προηγούμενες ενότητες είναι ειδικές περιπτώσεις της Γενικευμένης κατανομής Γάμμα.

Η γενικευμένη κατανομή Γάμμα έχει συνάρτηση πυκνότητας πιθανότητας

$$f(t) = \frac{n}{a\Gamma(k)}(\frac{t}{a})^{kn-1}\exp\{-\left(\frac{t}{a}\right)^n\}$$

Με $\alpha > 0$ παράμετρο κλίμακας και $n > 0, k > 0$ παράμετροι σχήματος.

Οι κατανομές των παραπάνω ενοτήτων προκύπτουν όταν

παράμετροι	κατανομή
k=1	Weibull
n=1	Gamma
k=1,n=1	Exponencial
$k \to \infty$	Logexponencial

Κατανομή Gumbel

Η κατανομή Gumbel (από τον E.Gumbel) έχει συνάρτηση πυκνότητας πιθανότητας

$$f(t) = \sigma^{-1}\exp\{{}^{(t-\mu)}/_{\sigma}\}\exp[-exp\left\{{}^{(t-\mu)}/_{\sigma}\right\} \qquad -\infty < t < \infty$$

Η αναμενόμενη τιμή και διασπορά είναι

$$E(T) = \mu - \gamma\sigma$$

$$Var(T) = (\pi^2/6)\sigma^2$$

με

$$\gamma = -\int_0^\infty e^{-\chi} lnxdx = -\Gamma'(1) = 0{,}5772 \ \ \eta\ \sigma\tau\alpha\theta\varepsilon\rho\acute{\alpha}\ Euler$$

Η κατανομή Gumbel μπορεί να χρησιμοποιηθεί ως μοντέλο διάρκειας ζωής παρότι ορίζεται στο διάστημα $(-\infty, +\infty)$,αρα επιτρέπει και αρνητικές τιμές .Ωστόσο αυτό δεν αποτελεί πρακτικό πρόβλημα αν οι τιμές των παραμέτρων είναι τέτοιες ώστε η πιθανότητα του ενδεχομένου $\{T < 0\}$ να είναι αμελητέα.

Σχέση κατανομών Gumbel και Wiebull

Αν η τυχαία μεταβλητή που περιγράφει τον χρόνο$T{\sim}Weibull$

$$\tau\acute{o}\tau\varepsilon\ \eta\ lnT{\sim}Gumbel$$

Η σχέση των παραμέτρων των δύο κατανομών

$$F(x) = P(X \leq x) = P(lnT \leq x) =$$

$$= P(T \leq e^x) = 1 - \exp\{-\left({e^x}/{a}\right)^n\} =$$

$$= 1 - \exp\{-\left(e^{nx}e^{-nlna}\right)\} = 1 - \exp\{-\exp[n(x - lna)]\} =$$

$$= 1 - \exp\{-\exp\left[\frac{x-\mu}{\sigma}\right]\}$$

Η παραπάνω συνάρτηση είναι η συνάρτηση κατανομής της κατανομής Gumbel με παραμέτρους

$$\mu = lna$$

$$\sigma = \frac{1}{n}$$

Αντίστροφη Γκαουσιανή κατανομή

Το όνομα της οφείλεται στο ότι η γεννήτρια συνάρτηση αθροιστικών της (*cumulant generating function)* είναι η αντίστροφη αυτής της κανονικής (ή του Gauss) κατανομής.

Η κατανομή ανακαλύφθηκε από τον διάσημο φυσικό Schrödinger το 1915 για να περιγράψει τον χρόνο πρώτης διέλευσης της κίνησης Brown. Η ονομασία inverse Gaussian προτάθηκε από τον Tweedie το 1945. O Wald επαναφέρει εκ νέου την κατανομή το 1947 για την οριακή μορφή δείγματος σε διαδοχικά τεστ αναλογιών πιθανότητας

Η συνάρτηση της πυκνότητας πιθανότητας της αντίστροφης Γκαουσιανής κατανομής (*Inverse Gaussian)* είναι η

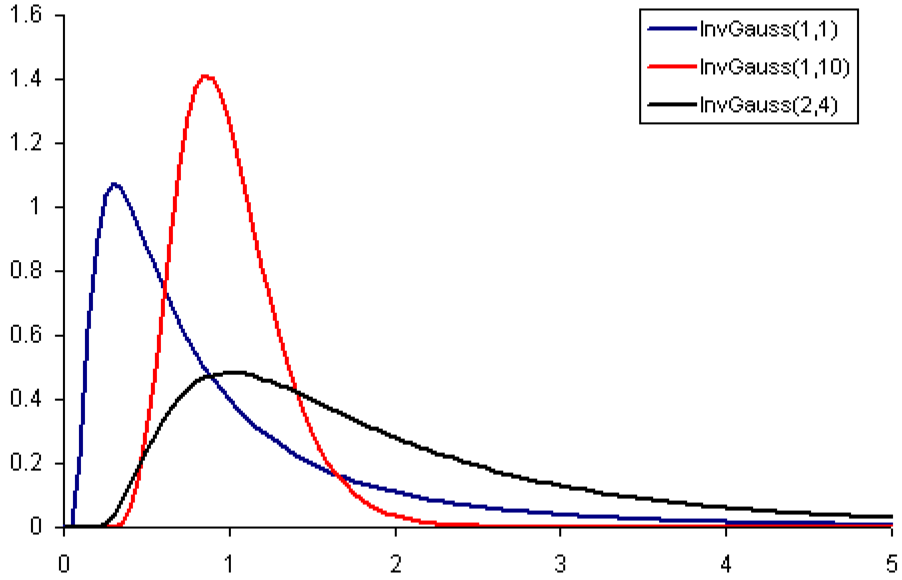
$$f(t) = \frac{\sqrt{\lambda}}{\sqrt{2\pi t^3}} \exp\left\{\frac{-\lambda(t-\mu)^2}{2\mu t^2}\right\} \qquad t > 0, \mu > 0, \lambda > 0$$

Και η συνάρτηση κατανομής γίνεται από τη σχέση

$$F(x) = \int_0^x \frac{\sqrt{\lambda}}{\sqrt{2\pi t^3}} \exp\left\{\frac{-\lambda(t-\mu)^2}{2\mu t^2}\right\} dt =$$

$$= 1 - \Phi\left[\sqrt{\lambda/_t}\left(1 - \frac{t}{\mu}\right)\right] - e^{2\lambda/\mu}\Phi\left[-\sqrt{\lambda/_t}\left(1 + \frac{t}{\mu}\right)\right]$$

Στο παρακάτω διάγραμμα φαίνεται πως μεταβάλλεται η συνάρτηση πυκνότητας της *inverse Gaussian* για διαφορετικές τιμές των παραμέτρων κλίμακας και σχήματος



Σχήμα 9 :*Inverse Gaussian*

Οι κυριότερες διακριτές κατανομές χρόνων

Γεωμετρική κατανομή

Ορισμός

Έστω ότι έχουμε μια ακολουθία ανεξάρτητων δοκιμών Bernoulli(1,p) με πιθανότητα επιτυχίας $p = 1 - \varphi$ και αποτυχίας $q = \varphi$ σταθερές για όλες τις δοκιμές. Έστω επίσης X ο αριθμός των προσπαθειών μέχρι την εμφάνιση της πρώτης επιτυχίας. Η κατανομή που ακολουθεί η τυχαία μεταβλητή X καλείται γεωμετρική κατανομή και συμβολίζεται στη διεθνή βιβλιογραφία $G(p)$.

Η συνάρτηση πυκνότητας πιθανότητας θα δίνεται από τη σχέση

$$P(T=n) = f(x) = pq^{n-1} \ , n = 1,2, \ldots.$$

Όταν η τυχαία μεταβλητή X ακολουθεί τη γεωμετρική κατανομή η μέση τιμή της και η διακύμανση της δίνονται από τις σχέσεις :

$$E(T) = \frac{1}{p} \ \kappa\alpha\iota \ V(T) = \frac{q}{p^2}$$

Η τυχαία μεταβλητή Τ είναι ο αριθμός των ανεξάρτητων δοκιμών Bernoulli μέχρις ότου εμφανιστεί για πρώτη φορά ένα γεγονός που έχει σταθερή πιθανότητα p να προκύψει σε οποιαδήποτε δοκιμή.

Κατανομή Poisson

Ένα από τα σημαντικότερα προβλήματα τα οποία αντιμετωπίζουμε κατά τη χρήση των διωνυμικών συντελεστών

$$\binom{n}{k} \qquad k = 0,1,2 \ldots n$$

Είναι οι μεγάλες τιμές που παίρνουν όταν το $n \to \infty$ και το k όχι πολύ κοντά στο μηδέν ή στο n.

Επειδή ο διωνυμικός συντελεστής αποτελεί μέρος της συνάρτησης πυκνότητας πιθανότητας της διωνυμικής κατανομής

$$f(t) = \binom{n}{t} p^t (1-p)^{n-t} \qquad t = 0,1,2 \ldots, n$$

Ο απευθείας υπολογισμός πιθανοτήτων με τον παραπάνω τύπο παρουσιάζει υπολογιστικές δυσκολίες.

Για να ξεπεράσουμε τη δυσκολία αυτή θα μπορούσαμε να χρησιμοποιήσουμε τον αναδρομικό τύπο του *Panjer*.

Σύμφωνα με τον αναδρομικό τύπο του Panjer ο οποίος συμβολίζεται $R(a, b)$ έχουμε για την τυχαία μεταβλητή Τ και για α, b σταθερές ότι

$$P(T=n) = \left(a + \frac{b}{n}\right) P(T = n-1), \qquad n = 1,2, \ldots$$

Για

$$\alpha = -\frac{p}{1-p} \; και \; b = \frac{(n+1)p}{1-p}$$

Και με αρχική συνθήκη την

$$f(0) = \binom{n}{0} p^0 (1-p)^{n-0} = (1-p)^n$$

Μπορούμε να έχουμε ένα σχετικά εύκολο τρόπο υπολογισμού των πιθανοτήτων.

Σημαντικό μειονέκτημα της παραπάνω διαδικασίας είναι πως αν επιθυμούμε να υπολογίσουμε την πιθανότητα για μια μεγάλη τιμή του Τ θα πρέπει να υπολογίσουμε αναγκαστικά όλες τις προηγούμενες πιθανότητες.

Η κατανομή Poisson μπορεί να θεωρηθεί ως μια βολική προσέγγιση της Διωνυμικής κατανομής για μεγάλο μέγεθος δείγματος n και μικρή πιθανότητα επιτυχίας p.

Η συνάρτηση πιθανότητας της κατανομής *Poisson* είναι η

$$P(T = n) = e^{-\lambda} \frac{\lambda^n}{n!} \qquad n = 0,1,2,\ldots$$

Ο *Simeon-Denis Poisson (1781-1840)* πραγματοποίησε μια συστηματική μελέτη της διωνυμικής κατανομής με στόχο την ανακάλυψη κάποιας έμμεσης μεθόδου υπολογισμού των πιθανοτήτων σε περιπτώσεις που το $n \to \infty$ τo $p \to 0$.

Σύμφωνα λοιπόν με τα παραπάνω

$$\alpha = \lim_{p \to 0}\left(-\frac{p}{1-p}\right) = 0 = \; και \; b = = \lim_{p \to 0} \frac{(n+1)p}{1-p} = np = \lambda$$

Σύμφωνα λοιπόν με τα παραπάνω όταν το $n \to \infty$ τo $p \to 0$ η μέση τιμή της Διωνυμικής κατανομής ισούται με τη μέση τιμή της κατανομής *Poisson* και θα μπορούσαμε να γράψουμε $p = \frac{\lambda}{n}$

Ενδιαφέρον έχει και ο υπολογισμός του ορίου

$$\lim_{\substack{p \to 0 \\ n \to \infty}} (1-p)^n) = \lim_{\substack{p \to 0 \\ n \to \infty}} (1 + \frac{(-\lambda)}{n})^n = e^{-\lambda}$$

Αναφορές

Αναφορές ξένες

[1] Feller W An introduction to probability theory and its applications Vol I,II, 1966

[2] Foss Sergey An introduction to heavy tailed and subexponential distributions Springer 2009

[3] Gut Allan Stopped Random Walks Limit Theorems and applications Springer 1988

[4] Rolsky Tomasz Stochastic Processes for Insurance and Finance Wiley New Tork

Αναφορές ελληνικές

[1] Θεόδωρος Π. Αρτίκης , «Μαθήματα στοχαστικών διαδικασιών Τόμοι 1,2, 3» ,εκδόσεις Α. Σταμούλης Πειραιάς 1991

[2] Κούτρας Μάρκος Εισαγωγή στις πιθανότητες θεωρία και εφαρμογές τόμοι 1,2

[3] Κουτσόπουλος Κ.Ι «Αναλογιστικά μαθηματικά μέρος 1 Θεωρία των κινδύνων», εκδόσεις Συμμετρία Αθήνα 1999

[4] *Λουλάκης Μιχαήλ Θεωρία μεγάλων αποκλίσεων θερινό σχολείο μαθηματικών ,Ηράκλειο 2005*

[5] Κωνσταντίνος Πολίτης, Πανεπιστημιακές σημειώσεις θεωρίας χρεοκοπίας 2011