Προσομοίωση με εφαρμογές στις κατανομές Πιθανοτήτων Διακριτές και Συνεχείς

Μωυσιάδης Πολυχρόνης
Πιθανοθεωρητική Προσομοίωση και Γραφήματα
ΠΜΣ Στατιστική και Μοντελοποίηση

Αναγκαιότητα και Τεχνικές

- Πιθανοθεωρητικά προβλήματα μεγάλης πολυπλοκότητας (σελ. 1)
- Έλεγχος θεωρητικής μελέτης
- Σύγκριση μεθόδων

- Καλή γνώση της Θ. Πιθανοτήτων
- Κατασκευή Αλγορίθμων και προγραμματισμός τους σε κάποια γλώσσα προγραμματισμού (Fortran, C++, Basic) ή σε κατάλληλα πακέτα (Mathematica, Matlab, Maple, Splus, κλπ)

Πρόβλημα

Ο Engel έθεσε το εξής πρόβλημα: Λέει ο Άβελ στον Κάϊν: Θέλω να ρίξουμε ένα νόμισμα, με μια πλευρά Κ και την άλλη Γ, μέχρις ότου να εμφανιστεί μια από τις ακολουθίες ΚΚΚΚ ή ΓΓΚΚ. Στην πρώτη περίπτωση θα κερδίσεις εσύ, στη δεύτερη εγώ. Το παιχνίδι είναι δίκαιο αφού κάθε ακολουθία έχει πιθανότητα 1/16 να εμφανιστεί.

Έχει δίκιο ο Άβελ; Και αν ναι ποιες οι πιθανότητες κέρδους καθενός παίκτη;

Πρόγραμμα Fortran '90 το coins.f90 $\Rightarrow$ coins.exe

που βρήκε P(ΚΚΚΚ)≈0.25, P(ΓΓΚΚ)≈0.75

Ο Αλγόριθμος

Συμβολίζω Κ=1, Γ=0

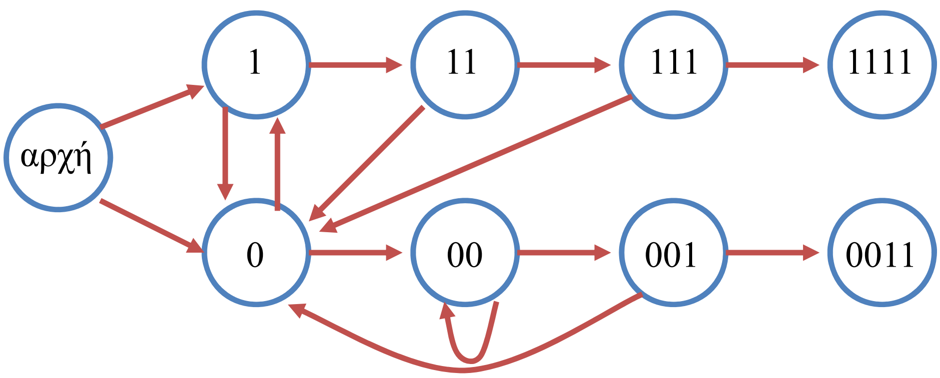
(1) Ρίχνω ένα νόμισμα 4 φορές

(2) Έστω x οι τελευταίες 4 ενδ.

(3) Αν x=1111 : κερδίζει ο Κάιν

(4) Αν x=0011 : κερδίζει ο Άβελ

(5) Αλλιώς ρίξε άλλη μία φορά και πάνε στο (2)



Οι κάρτες των ποδοσφαιριστών

Ο Don Piele καθηγητής Στατιστικής έλυσε με προσομοίωση και με το Mathematika το εξής πρόβλημα. Οι φωτογραφίες 12 σπουδαίων ποδοσφαιριστών είναι τυπωμένες σε κάρτες που στην πίσω πλευρά τους έχουν και σύντομο βιογραφικό. Κυκλοφορούν ως δώρο σε σοκολάτες που κάθε μία περιέχει μία κάρτα με αδιαφανή τρόπο. Πόσες σοκολάτες πρέπει να αγοραστούν, κατά μέσο όρο, ώστε να συμπληρωθεί η συλλογή;

Η απάντηση μπορεί να δοθεί και με άλλα προγράμματα και θεωρητικά. Κάντε τέτοιες προσπάθειες με τα προγράμματα που γνωρίζετε. Η λύση του Don Piele είναι 37 τουλάχιστον και την έδωσε και με θεωρητικό τρόπο πάλι με το Mathematika.

Ψευδο-Τυχαίοι Αριθμοί

Με seed, και με χρήση κάποιας συνάρτησης

Αποδείχθηκε ελαττωματική

$(3712)^2=13[7789]44$ παράγοντας 7789,
$(7789)^2=60[6685]21$ παράγοντας 6685,
$(6685)^2=44[6892]25$ παράγοντας 6892,
$(6892)^2=47[4996]64$ παράγοντας 4996
$(4996)^2=24[9600]16$ παράγοντας 9600
$(9600)^2=92[1600]16$ παράγοντας 1600
$(1600)^2=02[5600]00$ παράγοντας 5600
$(5600)^2=31[3600]00$ παράγοντας 3600
$(1600)^2=12[9600]00$ παράγοντας 9600

$$\boxed{x_{i+1}=(ax_i+c)[\mathrm{mod}(m)]}$$

m μεγάλο συνήθως $2^{31}-1$

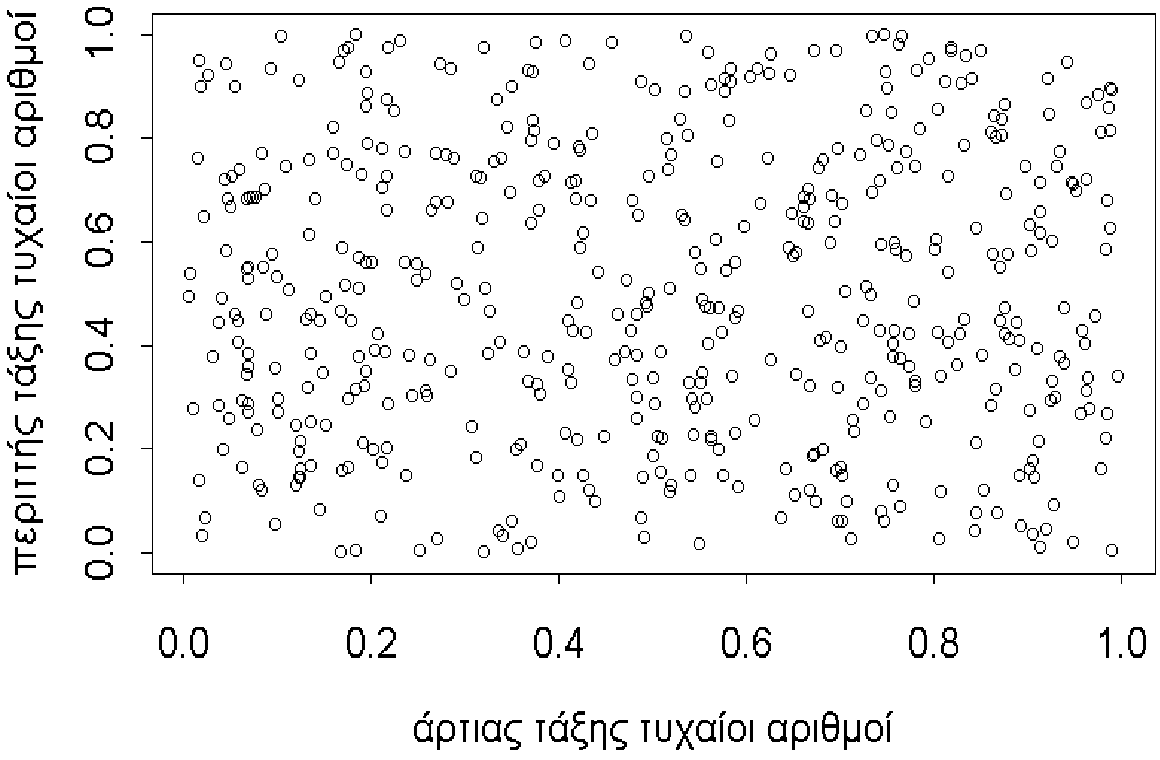
a π.χ. $7^5=16807$, c=0

c=0 multiplicative congruential generator

c≠0 mixted congruential generator

Steve Selvin,Ph.D.,Modern Applied Biostasistical Methods Using S-Plus,New York,Oxford University Press,Oxfor

1000 τ.α. με το S-Plus



Μέθοδος Monte Carlo

Ο νόμος των μεγάλων αριθμών: Αν $X_1, X_2, \ldots, X_n$ ανεξ. ισ.τ.μ.

με $\mathrm{E}X_i = m$, και $VarX_i = \sigma^2$, τότε: $\frac{X_1+X_2+\cdots+X_n}{n} \xrightarrow{P} m$

Αν $X_i \sim U(0,1)$ και $Y_i = f(Xi)$, με $m = \mathrm{E}f(Xi) = \int_0^1 f(x)dx$

και $\sigma^2 = \int_0^1 f(x-m)^2 dx$ τότε $\frac{f(X_1)+\cdots+f(X_n)}{n} \xrightarrow{P} m$

Δηλαδή: $\frac{f(X_1)+\cdots+f(X_n)}{n}$ εκτιμά το $\int_0^1 f(x)dx$ με ακρίβεια $\sqrt{n}$

Αυτό συνιστά τη μέθοδο Monte Carlo

Για εκτίμηση του $\theta = \int_0^1 f(x)dx$	Παίρνουμε k (μεγάλο) τ.αρ. x_i από την U(0,1) κατανομή και υπολογίζουμε $f(x_i)$	Τότε: $\theta \approx \frac{1}{k}\sum_{i=1}^{k} f(x_i)$

Εφαρμογές

$$\theta=\int_a^b f(x)dx$$

Θέτουμε
$t=(x-a)/(b-a)$

$$\boxed{\theta=\int_0^1 f(a+(b-a)t)(b-a)dt}$$

$$\theta=\int_0^\infty f(x)dx$$

Θέτουμε
$t=1/(x+1)$

$$\boxed{\theta=\int_0^1 f(\tfrac{1}{t}-1)/t^2\,dt}$$

$$\theta=\int_0^1\int_0^1\cdots\int_0^1 f(x_1x_2\ldots x_n)dx_1dx_2\ldots dx_n \qquad \boxed{\theta\approx\frac{1}{k}\sum_{i=1}^k f(U_1^iU_2^i\ldots U_n^i)}$$

όπου $\left\{U_1^1,U_2^1,\ldots,U_n^1\right\},\ldots,\left\{U_1^k,U_2^k,\ldots,U_n^k\right\}$ nk τυχαίοι αριθμοί του (0,1)

Η μέθοδος Monte Carlo είναι πολύ καλύτερη για τα πολλαπλά ολοκληρώματα απ’ ότι για τα απλά ολοκληρώματα αλλά

Ανάλυση Συναρτήσεων

Αν η συνάρτηση f(x) αναλύεται ως γινόμενο μιας άλλης συνάρτησης g(x) και μιας συνάρτησης πυκνότητας πιθανότητας p(x) μιας τ.μ. X, δηλαδή

$$f(x)=g(x)\cdot p(x) \quad \text{με} \quad \int p(x)dx=1$$

Τότε

$$\theta=\int f(x)dx=\int g(x)\cdot p(x)dx=E(g(X))$$

Άρα, αν $x_1, x_2, x_3, \ldots, x_m$ είναι τ.δ. από την τ.μ. X, θα είναι

$$\hat{\theta}=\frac{\sum_{i=1}^m g(x_i)}{m}$$

Είναι φανερό ότι υπάρχουν χρήσιμες και μη αναλύσεις

Εκτίμηση της Διασποράς

Αν στην εκτίμηση του $\hat{\theta} = \frac{\sum_{i=1}^{m} g(x_i)}{m}$ τα $g(X_i)$ είναι ανεξάρτ.
τότε:

$$\hat{V}(\hat{\theta}) = \frac{\sum_{i=1}^{m}\left(g(x_i) - \overline{g(x)}\right)^2}{m-1}$$

Γενικά ισχύει αν τα X_i είναι ανεξάρτητα

Αυτό δικαιολογείται από το ότι οι συνδιασπορές των $g(X_i)$ είναι 0

Εκτίμηση της Διασποράς με κατά δέσμες μέσους

Έστω $G_1, \ldots, G_b, G_{b+1}, \ldots, G_{2b}, G_{2b+1}, \ldots, G_{kb}$ είναι τ.μ. που η συνδιασπορά των G_i και G_{i+b} είναι περίπου 0. (π.χ σε χρονο-σειρές) Η διασπορά του γενικού μέσου $\overline{G}$ δίνεται από:

$$V(\overline{G}) = V\left(\frac{1}{m}\sum G_i\right) = V\left(\frac{1}{k}\sum_{j=1}^{k}\left(\frac{1}{b}\sum_{i=(j-1)b+1}^{jb} G_i\right)\right) \approx$$

$$\approx \frac{1}{k^2}\sum_{j=1}^{k} V\left(\frac{1}{b}\sum_{i=(j-1)b+1}^{jb} G_i\right) \approx \frac{1}{k} V(\overline{G}_b),$$

είναι ο μέσος μιας δέσμης μήκους b.

Μπορούμε να δεχτούμε ότι οι μέσοι των δεσμών έχουν κοινή διασπορά την

$$V(\overline{G}_b) = \frac{\sum(\overline{g}_j - \overline{g})^2}{k-1}.$$

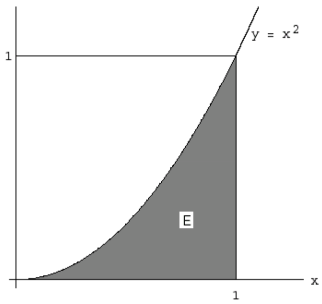
Άρα η εκτίμηση της διασποράς δίνεται

$$\hat{V}(\overline{G}) = \frac{\sum(\overline{g}_j - \overline{g})^2}{k(k-1)}.$$

Υπολογισμός εμβαδών με προσομοίωση

Να υπολογιστεί το εμβαδόν κάτω από την παραβολή $y=x^2$ από 0 έως 1

```
> set.seed(100)
> x<-runif(1000)
> xsq2 <- x^2
> mean(xsq2)
[1] 0.339692
```



$$\int_0^1 x^2\,dx = \frac{1}{3}$$

```
> set.seed(100)
> x_runif(10000)
> xsq2 <- x^2
> mean(xsq2)
[1] 0.337657
```

100000

0.333791

Άλλος τρόπος

Υποθέτουμε ότι (x,y) επιλέγεται με τρόπο ώστε x, y είναι τυχαία σημεία στο διάστημα (0,1). Αν $E=\{(x,y): y\leq x^2\}$

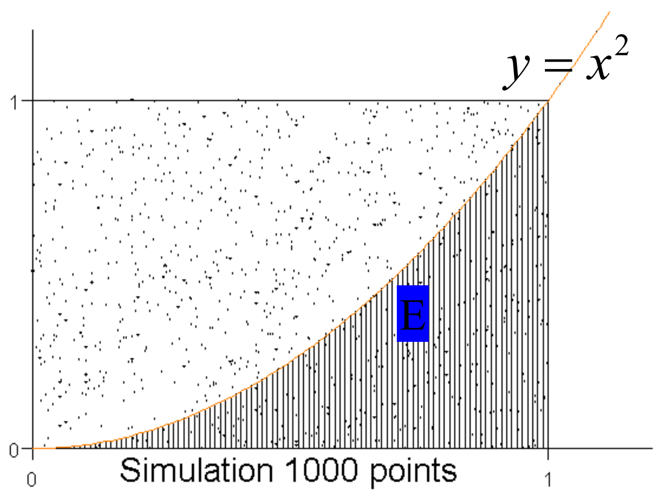
Γεωμετρική πιθανότητα $P((x,y)\in E) = P(y<x^2) = \frac{\mathrm{E}\mu\beta(\mathrm{E})}{\mathrm{E}\mu\beta(\Omega)} = \mathrm{E}\mu\beta(\mathrm{E})$

Έστω I δείκτρια με:

$$\mathrm{I}(\omega) = \begin{cases} 1, & \alpha\nu\ \omega\in \mathrm{E} \\ 0, & \alpha\lambda\lambda\iota\acute{\omega}\varsigma \end{cases}$$

τότε E(I)=P(E)= Εμβ(E)

```
> k <- 1000;
set.seed(100)
> x <- runif(k);  y <-
runif(k)
> z <- y < x^2;
sum(z)/k
[1] 0.341
k=10000  …. 0.3386
k=10000  …. 0.33402
```



Ο παλιός νόμος των παικτών (ΠΝΠ)

$p<0.5$ πιθ. 1 επιτυχίας σε 1 παρτίδα

$1-(1-p)^2$ πιθ.τουλ. 1 επιτυχίας σε 2 παρτίδες

$1-(1-p)^n$ πιθ.τουλ. 1 επιτυχίας σε n παρτίδες

παιχνίδι ευνοϊκό

$$\left(1-(1-p)^n\right) > (1-p)^n$$

$$n > \frac{\ln 2}{-\ln(1-p)}$$

p, n χαρακτηρ. παιχνιδιού

---

p_1, n_1
p_2, n_2
δύο παιχνί-δια

ΠΝΠ

$$n_2 = \frac{p_1}{p_2} \times n_1$$

$$n > \frac{\ln 2}{p + \frac{p^2}{2} + \frac{p^3}{3} + \ldots + \frac{p^n}{n} + \ldots}$$

πρώτη προσέγγιση

Παράδοξο De Mere: $p_1=1/6$, $n_1=4$, $p_2=1/36$, $n_2=25 \neq 4 \cdot 6=24$

Odds και Πιθανότητα

αναλογία υπερ-κατά

στοίχημα «r προς 1»
(r:1) ή « r to 1 odds”
ότι A κερδίζει B

P(A)+P(B)=1
P(A)=r P(B)

$$\boxed{p(A) = \frac{r}{r+1}}$$

στοίχημα «r προς s»
(r:s) ή « r to s odds »
ότι A κερδίζει B

«r/s προς 1»

$$\boxed{p(A) = \frac{r}{r+s}}$$

Αν για το A
«r προς s» odds

Τότε για το A A (δύο φορές στη σειρά)
«r^2 προς s^2+2rs» odds

Πόρισμα
E συμβαίνει σε k από n ισοπίθανες περιπτώσεις

odds για το E k:(n-k)
odds για το EE k^2:(n^2-k^2)

Διακριτές τυχαίες μεταβλητές

Υλοποίηση Πειράματος

Δίνεται ο πιθανοχώρος
$\Omega = \{\omega_1, \omega_2, \omega_3\}$,
με $P(\omega_1) = \frac{1}{2}, P(\omega_2) = \frac{1}{3}, P(\omega_3) = \frac{1}{6}$
Πως μπορεί να υλοποιηθεί με πείραμα;

Ζάρι με 3 πλευρές με το χρώμα 1 (ω_1), 2 με το χρώμα 2 (ω_2) και 1 με το χρώμα 3(ω_3)

Το ίδιο για τον πιθανοχώρο
$\Omega = \{\omega_1, \omega_2, \dots, \omega_k\}$, με $P(\omega_j)=\frac{n_j}{n}$,
όπου n ΕΚΠ(παρονομαστών)

Κατασκευάζουμε ένα πρίσμα με τομή κανονικό n-γωνο του οποίου nj πλευρές βάφονται με το χρώμα j

Συζήτηση για τους τρόπους υλοποίησης με υλικά μέσα (ζάρια, κάλπες, κλπ) ή με προγραμματισμό Η/Υ.

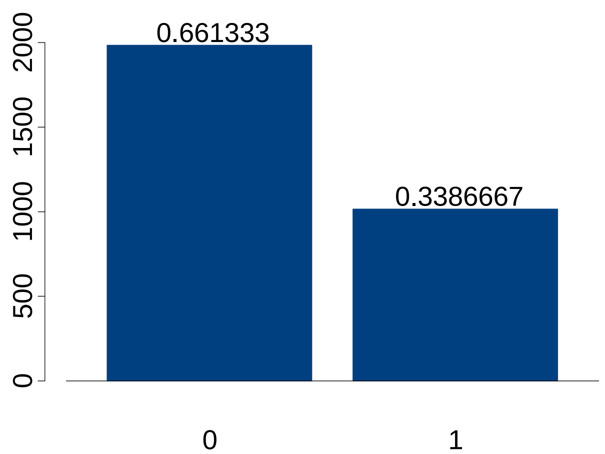
Κατανομή Bernoulli

$X = 0$ με πιθανότητα $1 - p$
$X = 1$ με πιθανότητα p

Μπορεί να θεωρηθεί ως η τ.μ. που παριστάνει την επιτυχία στην εκτέλεση ενός πειράματος. Παίρνει την τιμή 1 στην επιτυχία και την 0 στην αποτυχία

Για την προσομοίωσή της αρκεί να πάρουμε ένα τυχαίο αριθμό (στο (0,1)) και να τον συγκρίνουμε με τη δοθείσα πιθανότητα επιτυχίας p. Αν η τιμή είναι μικρότερη θέτουμε την τιμή 1, αλλιώς την τιμή 0.

Το ραβδόγραμμα των 1-δων και 0-κών σε ένα πλήθος εκτελέσεων της προσομοίωσης μπορεί να θεωρηθεί ως γράφημα της συνάρτησης πιθανότητας



Διωνυμική Κατανομή

$X \sim B(n, p)$. Μετρά το πλήθος επιτυχιών σε n δοκιμές Bernoulli

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \; x = 0, 1, \ldots, n$$

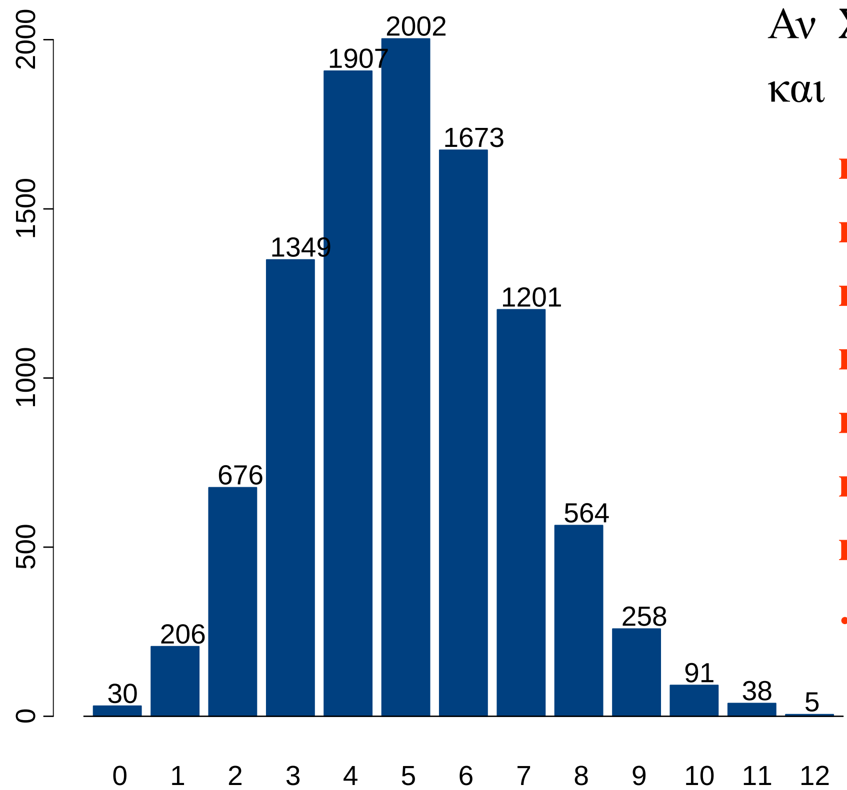
Είναι γνωστό ότι αν $X = X_1 + X_2 + \cdots + X_n$, όπου X_k ακολουθεί κατανομή Bernoulli με πιθ. επιτ. p και X_k ανεξ., τότε η X ακολουθεί τη διωνυμική $B(n, p)$.

Για την απόδειξη μπορούμε να χρησιμοποιήσουμε ροπογεννήτριες. Διαισθητικά ισχύει το παραπάνω αφού αν έχουμε επιτυχία στην k-στη δοκιμή τότε και μόνον θα είναι $X_k = 1$. Άρα το άθροισμα όλων των X_i θα ισούται με το πλήθος των επιτυχιών.

Για την προσομοίωση της $B(n, p)$ παίρνουμε n κατανομές Bernoulli με πιθανότητα επιτυχίας p και τις προσθέτουμε

Το ραβδόγραμμα των αθροισμάτων που προκύπτουν σε ένα πλήθος εκτελέσεων της προσομοίωσης μπορεί να θεωρηθεί ως γράφημα της συνάρτησης πιθανότητας

Προσομοίωση με n=10000



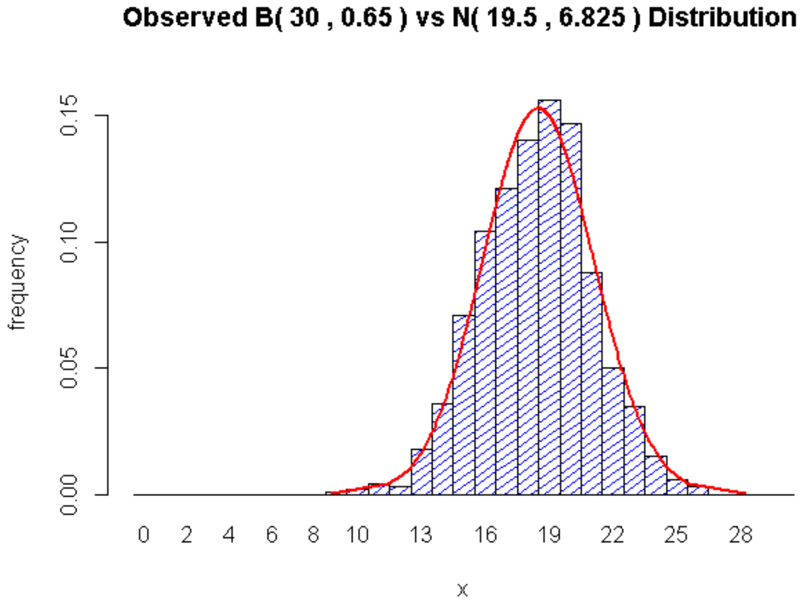
Αν $X \sim B(20, ¼)$
και n=10000

n·P(X=0)=31.7
n·P(X=1)=211.4
n·P(X=2)=669.5
n·P(X=3)=1339.0
n·P(X=4)=1896.9
n·P(X=5)=2023.3
n·P(X=6)=1686.1
………………

Διωνυμική Κατανομή

Είναι γνωστό ότι η διωνυμική κατανομή $B(n,p)$ προσεγγίζεται από την $N(np, np(1-p))$. Έτσι αν προσομοιώσουμε όπως πριν τη διωνυμική πρέπει οριακά να ταυτίζεται με την αντίστοιχη κανονική

Για να γίνει αυτό πρέπει το ραβδόγραμμα να το κάνουμε με τις σχετικές συχνότητες και να φροντίσουμε να κάνουμε τη γραφική παράσταση της κανονικής στο ίδιο σχήμα.



Προσομοίωση διακριτής τ.μ.

Έστω X τ.μ. με: $P(X = x_j) = p_j,\ j = 0,1,2,\ldots,n,\ \sum_{j=0}^{n} p_j = 1$

Αν U τ.μ. $\sim U(0,1)$ και

$$X = \begin{cases} x_0, \alpha\nu\ U < p_0 \\ x_1, \alpha\nu\ p_0 < U < p_0 + p_1 \\ \vdots \\ x_j, \alpha\nu \sum_{t=1}^{j-1} p_t < U < \sum_{t=1}^{j} p_t \\ \vdots \end{cases}$$

τότε $P(X = x_j) = p_j$

Αλγόριθμος

Βήμα 1. αν $U < p_0$,

θέσε $X = x_0$ και stop

Βήμα 2. αν $U < p_0 + p_1$,

θέσε $X = x_1$ και stop

……………………………………..

Παρατηρούμε ότι

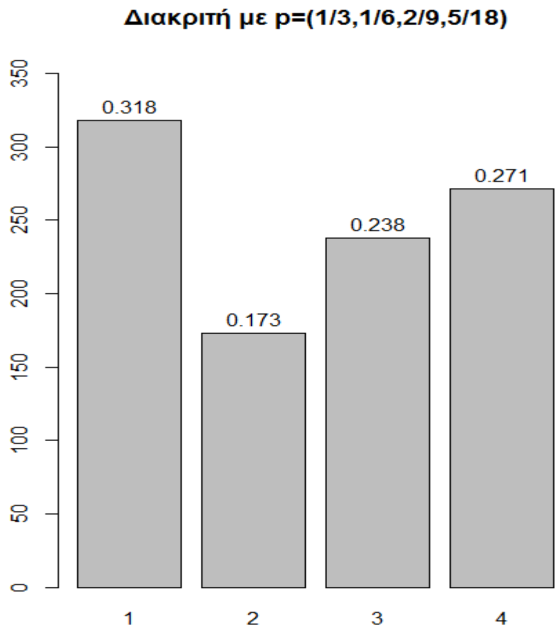
$$\sum_{t=0}^{j} p_t = F(x_j)$$

Παράδειγμα

Παράδειγμα

$p_1=1/3$, $p_2=1/6$, $p_3=2/9$, $p_4=5/18$

```
•k=1000;n=4
•p<-c(1/3,1/6,2/9,5/18)
•f<-cumsum(p);f
•s<-NULL
•for(j in 1:k){
•  z<-rep(1,n)
•  u<-runif(1)
•  for(i in 1:(n-1)) {
•    if (u<f[i]) z[i]<-0
•  }
•s[j]<-sum(z)
•}
•s
```



Απαριθμητή Ομοιόμορφη

$X = 1, 2, \dots, n$

με $P(X = j) = \frac{1}{n}$

Από προηγούμενο αν $U \sim U(0,1)$

$X = j$ αν $\frac{j-1}{n} \le U \le \frac{j}{n}$

$\Rightarrow X = j$ αν $j - 1 \le nU < j$

$\Rightarrow X = j$ αν $j = [nU + 1]$

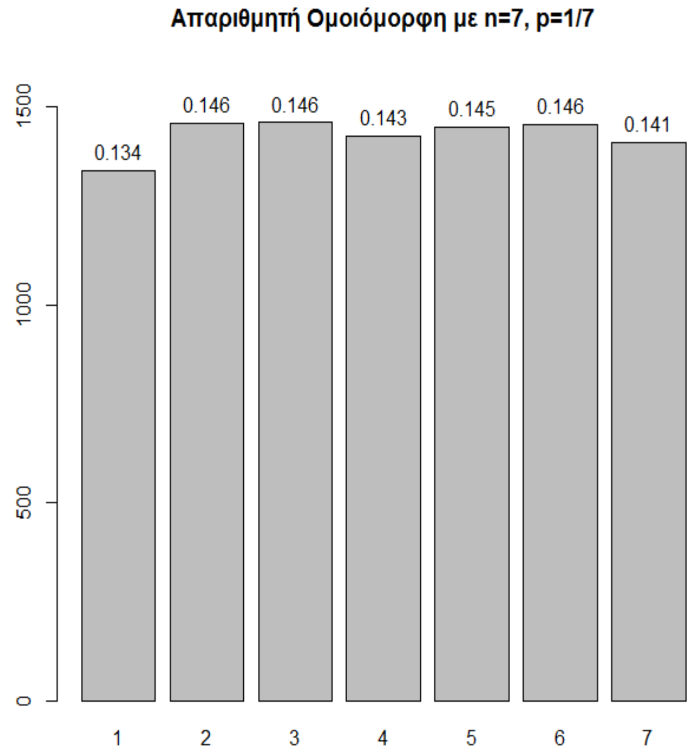
$\boxed{\Rightarrow X = [nU + 1]}$

Δίνεται ο πιθανοχώρος
$\Omega = \{\omega_1, \omega_2, \omega_3\}$,
με $P(\omega_1) = \frac{1}{2}, P(\omega_2) = \frac{1}{3}, P(\omega_3) = \frac{1}{6}$
Πως μπορεί να υλοποιηθεί με πείραμα;

$\Rightarrow X = [6U + 1]$
και $\omega_1 : X \in \{1,2,3\}$
$\omega_2 : X \in \{4,5\}$
$\omega_3 : X = 6$

Παράδειγμα

```
n<-7
k<-10000
x<-floor(n*runif(k)
   +1)
b<-barplot(
  t<-table(x),
  names=names(t),
  ylim=c(0,1600))
text(b,t+k/200,
  round(t/k,3))
title(main="Απαριθμητή
  Ομοιόμορφη με n=7,
  p=1/7")
```



Τυχαία μετάθεση ή Διάταξη

Βήμα 1. Έστω $P_1, P_2, \dots, P_n$ τυχαία μετάθεση των $1.2, \dots, n$

Βήμα 2. Θέσατε $k = n$

Βήμα 3. Επιλέξτε ένα τυχ. αρ. $U \in (0,1)$ και θέσατε $i = [kU + 1]$

Βήμα 4. Εναλλάξτε τις τιμές P_i, P_k

Βήμα 5. Θέσατε $k = k - 1$ και αν $k > 1$ πηγαίνετε στο Βήμα 3.

Βήμα 6. Η μετάθεση $P_1, P_2, \dots, P_n$ είναι η ζητούμενη.

αν $k > n - r + 1$ παίρνουμε τυχαία διάταξη r αριθμών από τους n.

αρκεί $r \leq \frac{n}{2}$

Καταγραφή Μεταθέσεων
σε αύξουσα σειρά

Βήμα 1. Θέσατε π=1 2 3 ... n και καταγράψτε την π.

Βήμα 2. Αν $\pi=\pi_1\pi_2\pi_3...\pi_n$ και ισχύει $\pi_i > \pi_{i+1}$ για όλα τα i, σταματήστε (διότι η λίστα έχει ολοκληρωθεί).

Βήμα 3. Βρέστε το μεγαλύτερο i για το οποίο $\pi_i < \pi_{i+1}$.

Βήμα 4. Βρέστε το μικρότερο π_j για το οποίο $i < j$ και $\pi_i < \pi_j$.

Βήμα 5. Εναλλάξτε αμοιβαία τα π_i και π_j.

Βήμα 6. Διατάξτε σε αύξουσα φυσική σειρά τα σύμβολα που ακολουθούν το π_j, συμβολίστε με π την μετάθεση που προκύπτει, καταγράψτε την π και πηγαίνετε στο Βήμα 2.

- Α Β Γ, Α Γ Β, Β Α Γ, Β Γ Α, Γ Α Β, Γ Β Α, ή
- 1 2 3, 1 3 2, 2 1 3, 2 3 1, 3 1 2, 3 2 1

Διάταξη σε αύξουσα φυσική σειρά

Αλγόριθμος bubble sort

Βήμα 1. Θέσατε m=n-1

Βήμα 2. Για i=1, 2, …, m: αν ισχύει $x_i > x_{i+1}$, εναλλάξτε τα $x_i > x_{i+1}$.

Βήμα 3. Ελαττώσατε το m κατά 1. Αν το m είναι τώρα 0, σταματήστε, αλλιώς πηγαίνετε στο Βήμα 2.

- Διάταξη σε αύξουσα σειρά των 3, 4, 1, 5, 2

1η φάση	**3**	**4**	1	5	2
	3	**1**	**4**	5	2
	3	1	**4**	**5**	2
	3	1	4	**2**	**5**
2η φάση	**1**	**3**	4	2	5
	1	**3**	**4**	2	5
	1	3	**2**	**4**	5
3η φάση	**1**	**3**	2	4	5
	1	**2**	**3**	4	5
4η φάση	**1**	**2**	3	4	5

Γεωμετρική Κατανομή

Y ακολουθεί γεωμετρική όταν Y=πλήθος αποτυχιών μέχρι την πρώτη επιτυχία. Επιλέγουμε τυχαίους αριθμούς του (0,1) X έως ότου να πετύχουμε για πρώτη φορά $X < p$. Δεν είναι αποδοτική μέθοδος, ιδιαίτερα για μικρά p (μέση τιμή 1/p).

Θεωρούμε X τυχαίο αριθμό στο (0,1) και ας είναι Y ο μικρότερος ακέραιος για τον οποίο $\boxed{1 - q^{Y+1} \geq X \ \ (1)}$ (υπάρχει;)

Τότε έχουμε: $P(Y = j) = P(1 - q^{j+1} \geq X > 1 - q^{j}) =$
$$= (1 - q^{j+1}) - (1 - q^{j}) = pq^{j}, \ j = 0, 1, \ldots$$

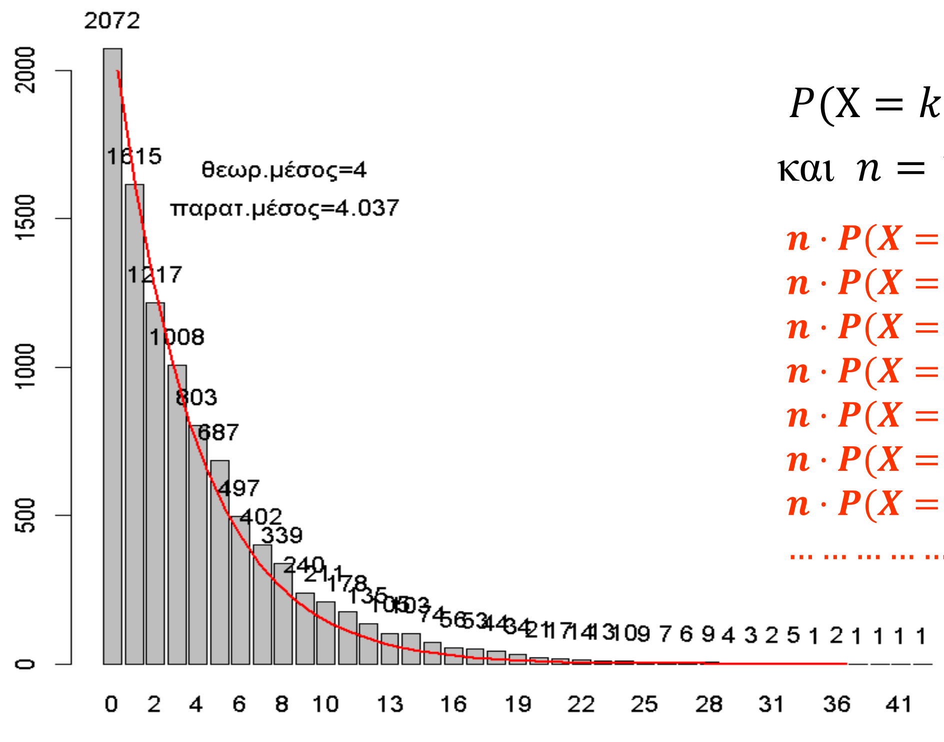
Δηλαδή η Y είναι τότε γεωμετρική

X και $1 - X$
ίδια κατανομή

Λογαριθμίζουμε την (1) και λύνουμε ως προς Y οπότε: $Y = \left[\frac{\ln(1-X)}{lnq}\right]$

$$\boxed{Y = \left[\frac{lnX}{lnq}\right]}$$

Προσομοίωση με $n = 10000, p = .2$



$P(X = k) = 0.2(0.8)^{k}$
και $n = 10000$

$\boldsymbol{n \cdot P(X = 0) = 2000.0}$
$\boldsymbol{n \cdot P(X = 1) = 1600.0}$
$\boldsymbol{n \cdot P(X = 2) = 1280.0}$
$\boldsymbol{n \cdot P(X = 3) = 1024.0}$
$\boldsymbol{n \cdot P(X = 4) = 819.2}$
$\boldsymbol{n \cdot P(X = 5) = 655.4}$
$\boldsymbol{n \cdot P(X = 6) = 524.3}$
… … … … … …

Αρνητική Διωνυμική κατανομή

Είναι γνωστό ότι αν $X = X_1 + X_2 + \cdots + X_n$, όπου X_k ακολουθεί γεωμετρική κατανομή με πιθ. επιτ. p και X_k ανεξ. τότε η X ακολουθεί την αρνητική διωνυμική, δηλ.

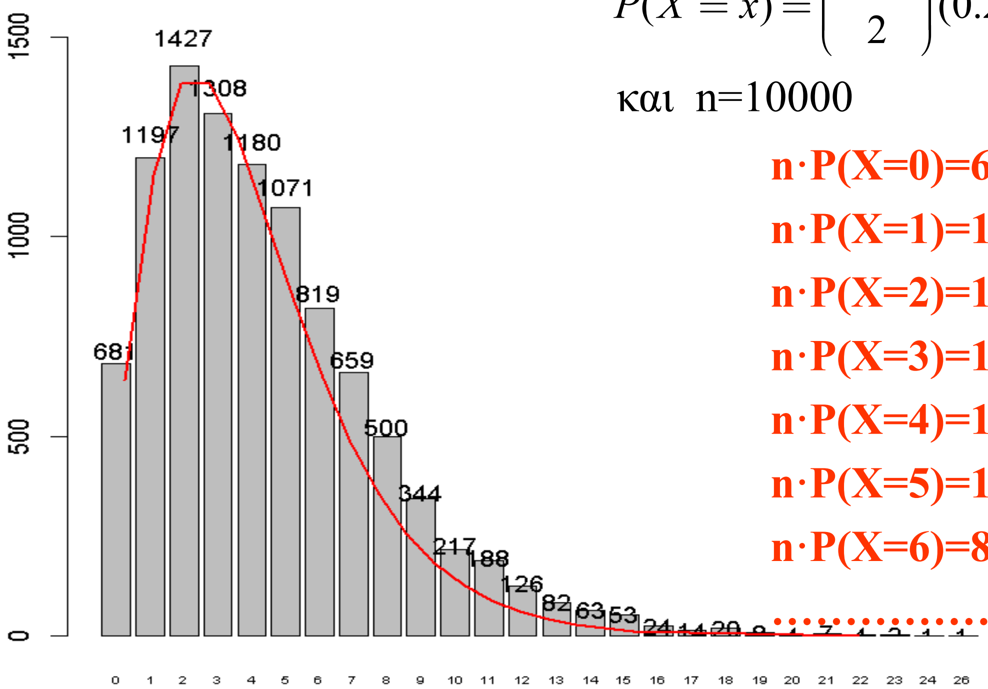
$$P(X=x)=\binom{k+x-1}{x}p^k q^x,\ \ x=0,1,\ldots$$

Για την απόδειξη μπορούμε να χρησιμοποιήσουμε ροπογεννήτριες. Διαισθητικά ισχύει το παραπάνω αφού οι αποτυχίες μέχρι την k-στη επιτυχία ισούνται με τις αποτυχίες μέχρι την 1η επιτυχία αυξημένες με τις αποτυχίες από τότε και μέχρι τη 2η επιτυχία. Και συνεχίζουμε μέχρι να συμπληρωθούν οι k επιτυχίες. Είναι προφανώς ανεξάρτητες οι ομάδες δοκιμών που απαιτούνται.

Άρα θα έχουμε: $\boxed{Y = \sum_{i=1}^{k}\left[\frac{lnX_i}{lnq}\right]}$

όπου X_i τυχαίοι αριθμοί του (0,1)

Προσομ. Αρν. Διων. Κατανομής με p= 0.4 k = 3 και N= 10000



$$P(X=x)=\binom{2+x}{2}(0.2)^3(0.8)^x$$

και n=10000

n·P(X=0)=640

n·P(X=1)=1152

n·P(X=2)=1382

n·P(X=3)=1382

n·P(X=4)=1244

n·P(X=5)=1045

n·P(X=6)=836

………………

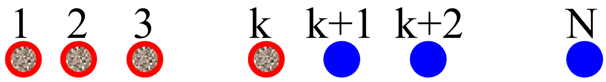
Υπεργεωμετρική Κατανομή

k κόκκινα

N-k μπλε

$$P(X=x)=\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \quad x=0,1,\ldots,n$$

Έστω:

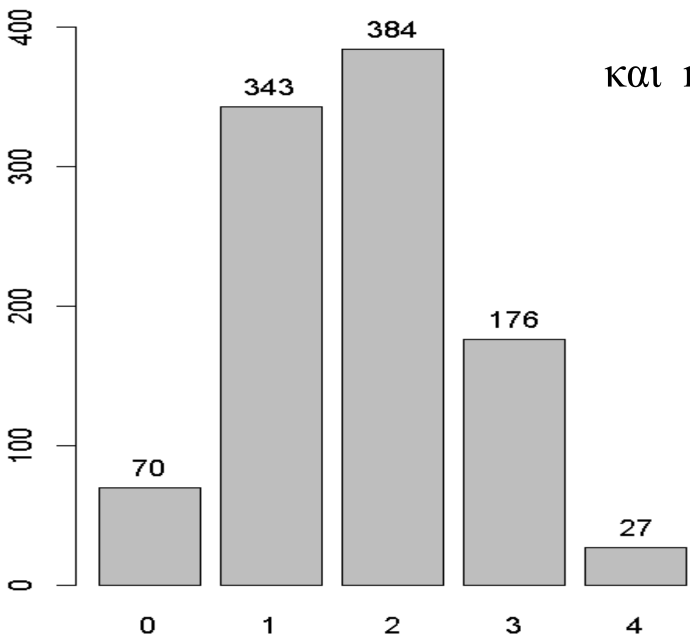


Έστω έχω επιλέξει μέχρι τώρα m σφαιρ. j από τα οποία είναι **K** (αρχικές τιμές m=0, j=0). Για το επόμενο υπάρχουν k-j (**K**) από συνολικά N-m. Το **K** έχει πιθανότητα (k-j)/(N-m). Έτσι επιλέγω X τυχαίο αριθμό στο (0,1). Αν είναι X≤(k-j)/(N-m) αυξάνω το jκαι το m, αλλιώς μόνο το m, ώσπου να γίνει n. Το τελικό j είναι τιμή της τ.μ. με υπεργεωμετρική κατανομή.

Προσομοίωση m=1000, N=20, k=5, n=7

$$P(X=x)=\frac{\binom{5}{x}\binom{15}{7-x}}{\binom{20}{7}}, x=0,1,\ldots,7$$

και n=1000



n·P(X=0)=83.0

n·P(X=1)=322.8

n·P(X=2)=387.4

n·P(X=3)=176.1

n·P(X=4)=29.3

n·P(X=5)=1.4

n·P(X=6)=0

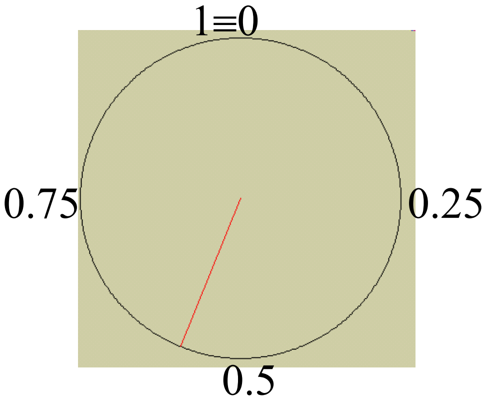
n·P(X=7)=0

Συνεχείς τυχαίες μεταβλητές

Πειράματα για συνεχείς τ.μ.

Τροχός με δείκτη περιστρεφόμενο

$\Omega = [0,1)$ (ή $[\alpha,\beta)$ για δοσμένα α,β)



Κάθε σημείο ίδια πιθανότητα (πρέπει)

Αλλά

Για κάθε

$p\neq 0$

$$\boxed{\sum_{i\in[0,1)} p = \infty \neq 1}$$

Είναι όμως αναμενόμενο ο δείκτης να σταματήσει σε σημείο που ανήκει σε διάστημα μήκους δ έχει πιθανότητα δ. **U(0,1)**

Το πείραμα είναι ισοδύναμο με την επιλογή σημείου με τυχαίο τρόπο στο διάστημα (0,1)

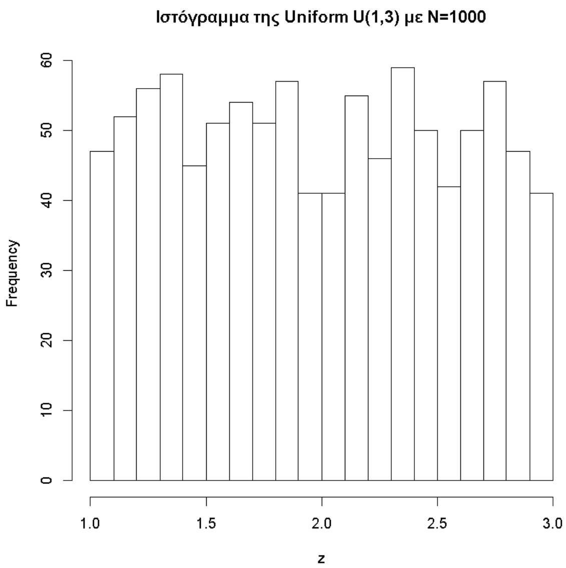
Ομοιόμορφη κατανομή U(a,b)

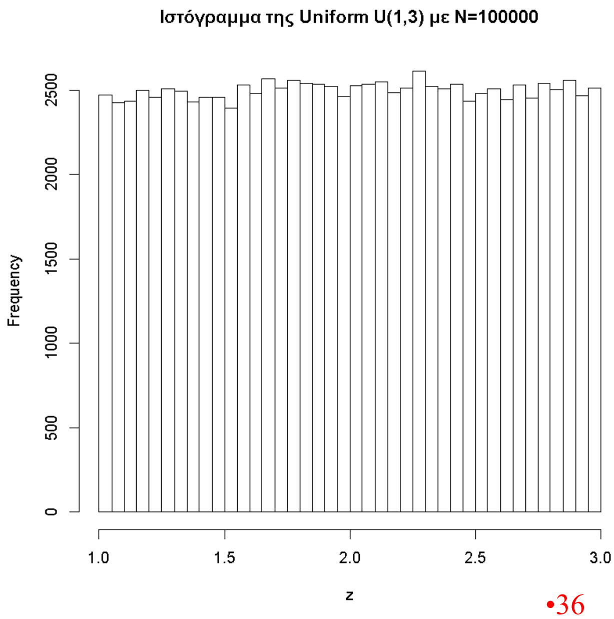
$$f(x) = \begin{cases} 1/(b-a), & a \le x \le b \\ 0, & \alpha\lambda\lambda o\acute{\upsilon} \end{cases}$$

Αν $X \sim U(0,1)$ τότε

$a + (b-a)X \sim U(a,b)$

Μεγαλύτερα n καλύτερη συμπεριφορά



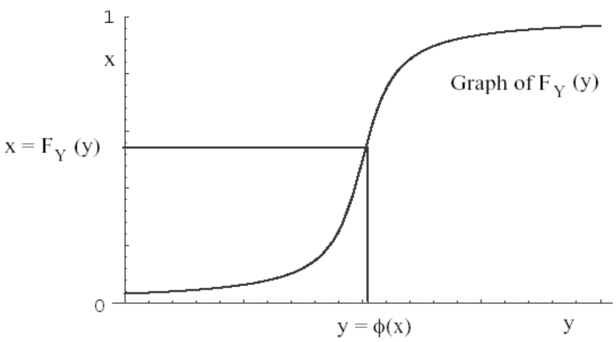


Συνεχής Κατανομή με σ.κ. F(x)

Θεώρημα: Έστω U τ.μ. με ομοιόμορφη κατανομή στο $(0, 1)$.
Για κάθε συνεχή τ.μ. με συν. κατανομής $F(x)$ αν ορίσουμε την τ.μ. Χ ως $X = F^{-1}(U)$ τότε η τ.μ. Χ θα έχει σ.κ. $F(x)$.

Απόδειξη:

$$
\begin{aligned}
F_X(x) &= P(X \le x) = \\
&= P(F^{-1}(U) \le x) = \\
&\overset{F \text{ αύξουσα}}{=} P(F(F^{-1}(U)) \le F(x)) = \\
&= P(U \le F(x)) \\
&= F(x)
\end{aligned}
$$



Αρκεί να μπορεί να βρεθεί η συνάρτηση $F^{-1}(x)$

Παράδειγμα

Αν σ.π.π.

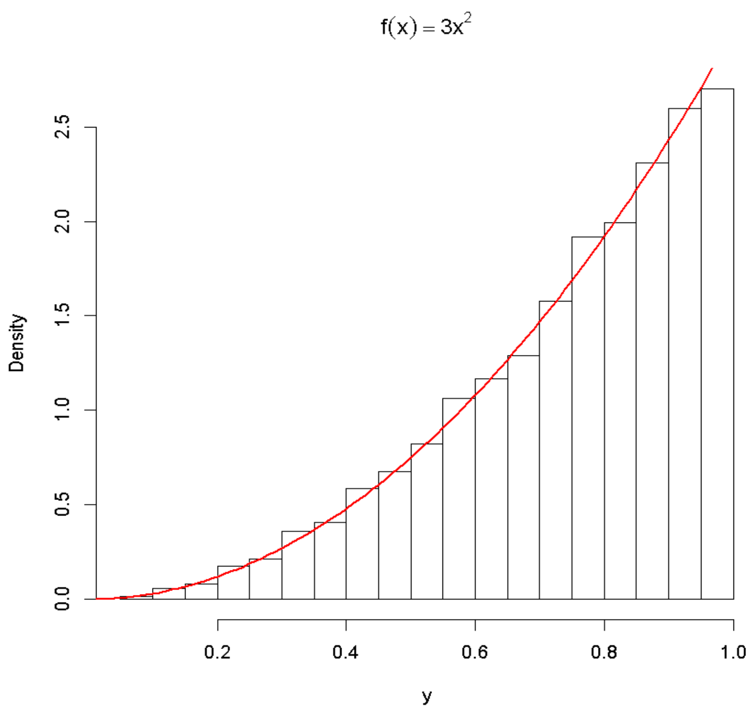
$$f(x) = \begin{cases} 3x^2, x \in (0,1) \\ 0, \ αλλού \end{cases}$$

τότε

$$F(x) = \begin{cases} 0, \ x < 1 \\ x^3, x \in (0,1) \\ 1, \ x \ge 1 \end{cases}$$

τότε

$$F^{-1}(x) = \begin{cases} 0, \ x < 1 \\ \sqrt[3]{x}, x \in (0,1) \\ 1, \ x \ge 1 \end{cases}$$



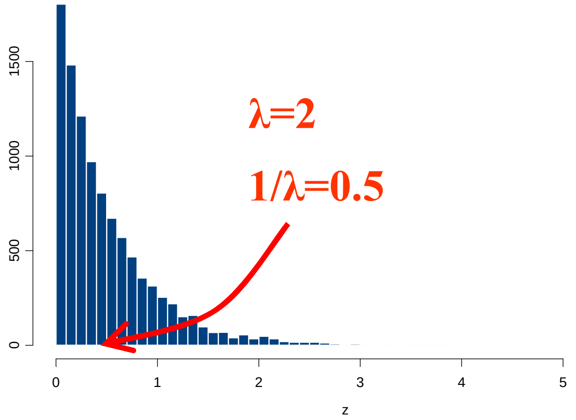
Εκθετική κατανομή παραμ. λ

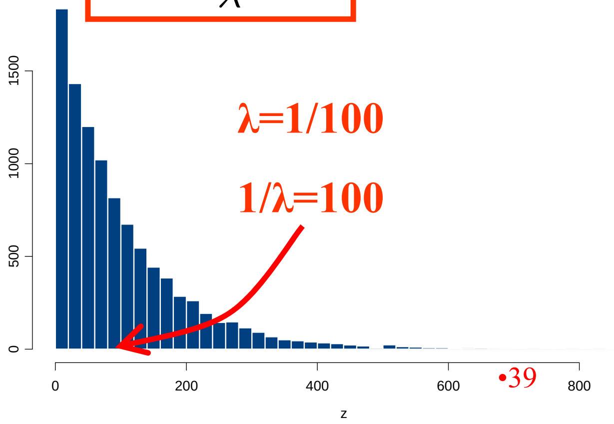
$$f(x)=\begin{cases}\lambda e^{-\lambda x} & x\ge 0\\ 0, & \alpha\lambda\lambda o\acute{\upsilon}\end{cases} \qquad F(x)=\begin{cases}1-e^{-\lambda x} & x\ge 0\\ 0, & \alpha\lambda\lambda o\acute{\upsilon}\end{cases}$$

$$y=1-e^{-\lambda x}\Rightarrow e^{-\lambda x}=1-y\Rightarrow x=-\frac{1}{\lambda}\ln(1-y) \quad \text{Άρα} \quad F^{-1}(x)=\begin{cases}-\frac{1}{\lambda}\ln(1-x) & 0\le x\le 1\\ 0, & \alpha\lambda\lambda o\acute{\upsilon}\end{cases}$$

Αν $U\sim U(0,1)$ τότε αρκεί να ορίσουμε $X=F^{-1}(U)=-\frac{1}{\lambda}\ln(1-U)$

και επειδή και η $1-U\sim U(0,1)$ αρκεί $\boxed{X=-\frac{1}{\lambda}\ln(U)}$



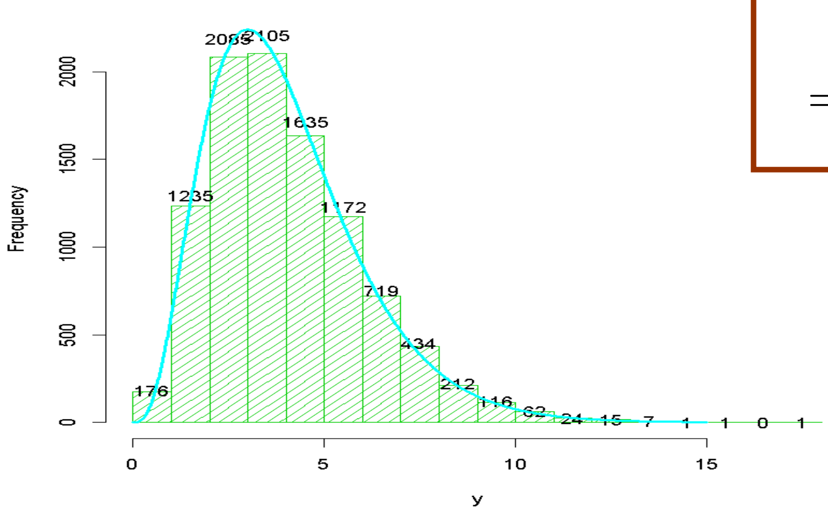


Κατανομή G(n,λ)

Εύκολα δείχνουμε ότι αν $X=X_1+X_2+\cdots+X_n$, όπου X_k ακολουθεί εκθετική κατανομή με παράμ. λ και X_k ανεξ., τότε η X ακολουθεί την κατανομή $G(n,\lambda)$, δηλ.

$$f(x)=\frac{\lambda^n}{(n-1)!}x^{n-1}e^{-\lambda x},\ \ \gamma\iota\alpha\ x\ge 0$$

$$\boxed{\begin{aligned}\mathrm{Y}&=-\frac{1}{\lambda}\sum_{i=1}^{n}\ln(U_i)=\\ &=-\frac{1}{\lambda}\ln\left(\prod_{i=1}^{n}U_i\right)\end{aligned}}$$



Κατανομή Poisson P(λ)

Γνωρίζουμε ότι η κατανομή των ενδιάμεσων χρόνων μεταξύ δύο γεγονότων που συμβαίνουν ακολουθώντας την κατανομή Poisson με παράμετρο λ, είναι εκθετική με παράμετρο λ. Άρα αν η Poisson έχει παράμετρο 1, οι ενδιάμεσοι χρόνοι ακολουθούν εκθετική με παράμετρο 1, και σύμφωνα με το προηγούμενο προσομοιώνονται από την $-\log(U_i)$, όπου U_i ομοιόμορφη στο (0,1). Αυτό σημαίνει επίσης ότι η τυχαία μεταβλητή:

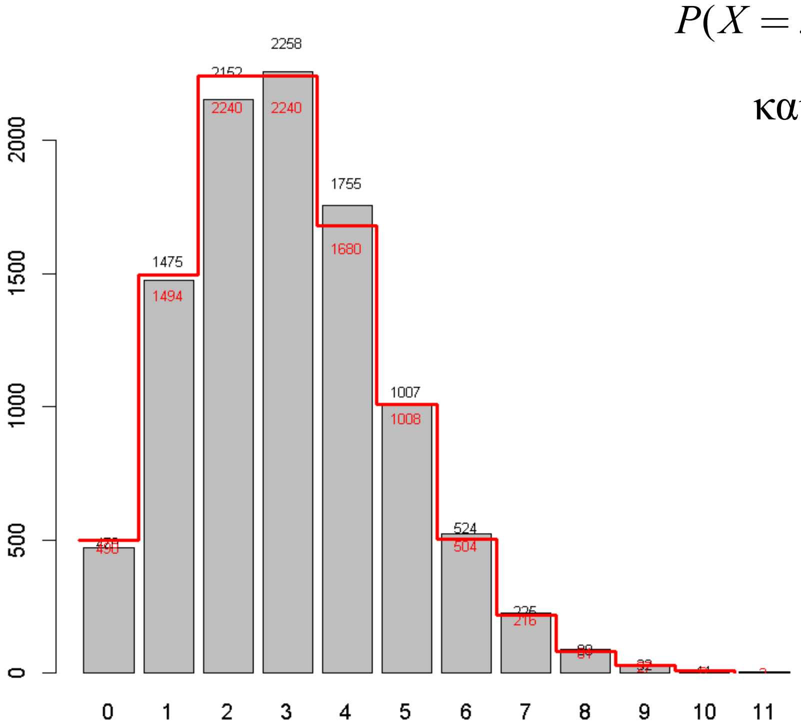
$$N = \max\left\{ n : \sum_{i=1}^{n} -\ln(U_i) < \lambda \right\}$$

θα ακολουθεί κατανομή Poisson(λ).

Πράγματι το N μπορούμε να το ερμηνεύσουμε ως το πλήθος γεγονότων από την ανέλιξη Poisson με παράμετρο 1 που συμβαίνουν στο χρόνο (0,λ).

Εδώ, χρησιμοποιήσαμε το θεώρημα ότι: Η τ.μ. X(λ) που μετρά το πλήθος των γεγονότων σε μια ανέλιξη Poisson με παράμετρο ν, ακολουθεί κατανομή Poisson με παράμετρο νλ, δηλ: $P(X(\lambda) = k) = e^{-\nu\lambda}\frac{(\nu\lambda)^k}{k!}$ και θέσαμε ν=1.

Προσομοίωση με n=10000, λ=3



$$P(X = x) = e^{-\lambda}\frac{3^x}{x!},\ \ x = 0,1,\ldots$$

και n=10000

n·P(X=0)=498

n·P(X=1)=1494

n·P(X=2)=2240

n·P(X=3)=2240

n·P(X=4)=1680

n·P(X=5)=1008

n·P(X=6)=504

n·P(X=7)=216

n·P(X=8)=81

……………….

Άλλη μέθοδος

$$p_i = P(X=i) = e^{-\lambda}\frac{\lambda^i}{i!},\ i=0,1,\ldots \qquad \boxed{\Rightarrow\ p_{i+1} = \frac{\lambda}{i+1}p_i,\ i=0,1,\ldots}$$

Βήμα 1. Επιλέξατε ένα τυχ. αρ. $U \in (0,1)$

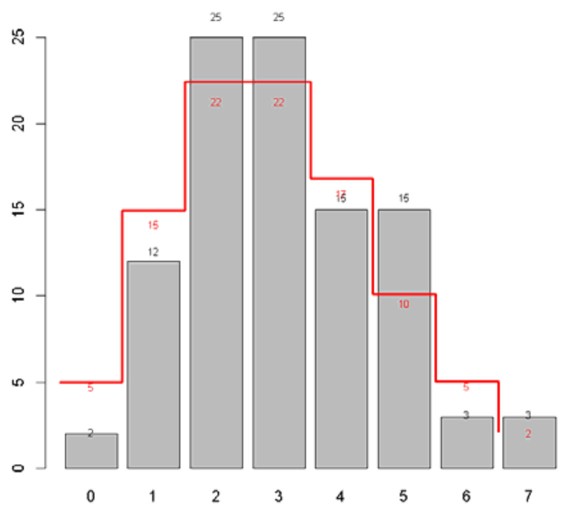
Βήμα 2. Θέσατε i=0, $p=e^{-\lambda}$, F=p

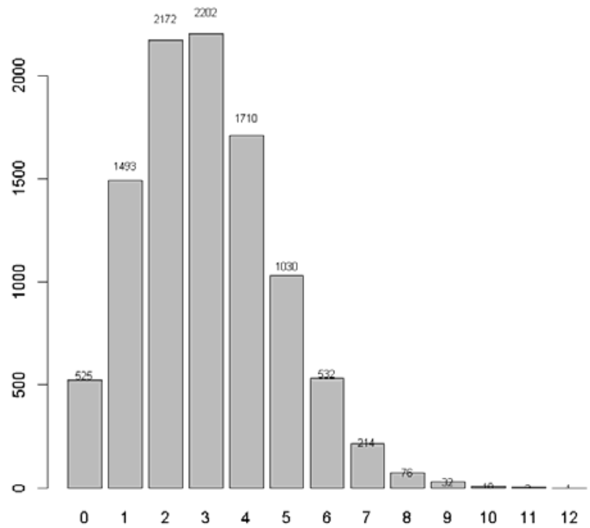
Βήμα 3. Αν U<F, θέσατε X=i και Σταματήστε

Βήμα 4. p=λp/(i+1), F=F+p, i=i+1

Βήμα 5. Πηγαίνετε στο Βήμα 3.

100 και 10000 Poisson τιμές για λ=3





Η μέθοδος απόρριψης (the rejection method)

Έστω ότι γνωρίζουμε να προσομοιώσουμε την τ.μ. Y με πυκνότητα $g(y)$. Θέλουμε να προσομοιώσουμε την τ.μ. X με πυκνότητα $f(x)$.

Υπολογίζουμε πρώτα ένα άνω φράγμα (ή μέγιστο) για το λόγο $\frac{f(y)}{g(y)}$, δηλ. έστω $\frac{f(y)}{g(y)} \leq c, \ \forall y$

Στη συνέχεια ακολουθούμε την παρακάτω μέθοδο:

Βήμα 1: Παίρνουμε μια τιμή από την τ.μ. Y και μία τιμή από την ομοιόμορφη U στο $(0,1)$

Βήμα 2: Αν $U \leq \frac{f(Y)}{c \cdot g(Y)}$ θέτουμε $X = Y$. Αλλιώς πηγαίνουμε στο Βήμα 1.

Απόδειξη

Αν X η τιμή που προέκυψε μετά N επαναλήψεις, τότε

$$
\begin{aligned}
P(X \leq x) &= P(Y_N \leq x) = \\
&= P(Y \leq x \,|\, U \leq f(Y)/c \cdot g(Y)) = \\
&= \frac{P(Y \leq x,\ U \leq f(Y)/c \cdot g(Y))}{K} = \\
&= \frac{\int P(Y \leq x,\ U \leq f(Y)/c \cdot g(Y) \,|\, Y = y) g(y)\,dy}{K} = \\
&= \frac{\int_{-\infty}^{x} P(U \leq f(y)/c \cdot g(y)) g(y)\,dy}{K} = \\
&= \frac{\int_{-\infty}^{x} (f(y)/c \cdot g(y)) g(y)\,dy}{K} = \frac{\int_{-\infty}^{x} f(y)\,dy}{cK} = \frac{F_X(x)}{cK}
\end{aligned}
$$

$$K = P(U \leq f(Y)/c \cdot g(Y))$$

Αφήνοντας το x να τείνει στο άπειρο έχουμε $K = 1/c$ και ο.ε.δ.

Παράδειγμα

Με την μέθοδο απόρριψης θα προσομοιώσουμε την κατανομή με πυκνότητα $f(x) = 20x(1-x)^3, 0 \le x \le 1$, δηλ. μιας βήτα B(2,4) .

Επειδή είμαστε στο διάστημα (0,1) θεωρούμε ως βάση την ομοιόμορφη στο (0, 1) δηλ. $g(x) = 1, 0 \le x \le 1$

Για τη σταθερά c παραγωγίζουμε την $\frac{f(x)}{g(x)} = 20x(1-x)^3$ ως προς x και βρίσκουμε $20(1-x)^2(1-4x)$ ότι δηλαδή έχουμε μέγιστο στο $\frac{1}{4}$ την τιμή $c = \frac{135}{64}$.

Άρα έχουμε $\frac{f(x)}{c \cdot g(x)} = \frac{256}{27}x(1-x)^3$ και η μέθοδος γίνεται:

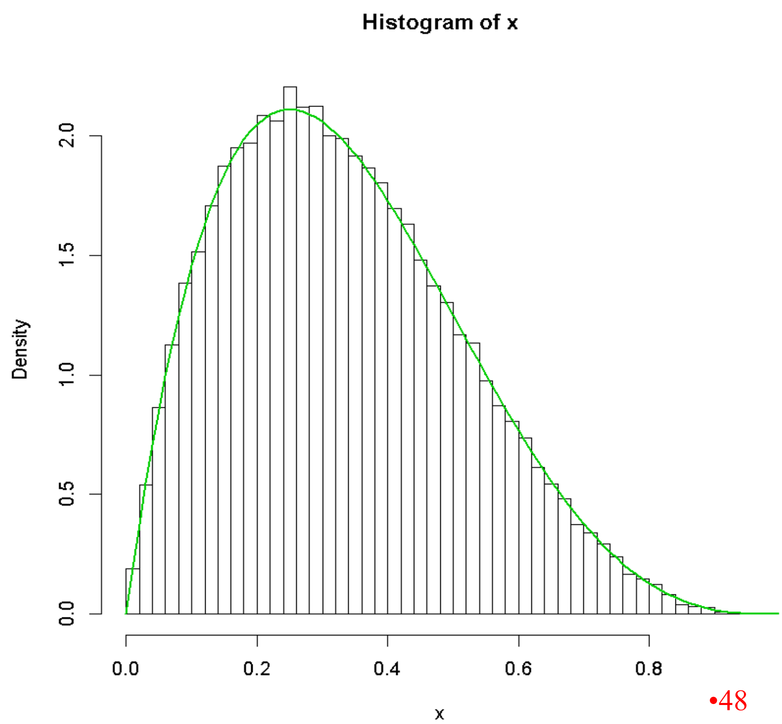
Βήμα 1: Παίρνουμε δύο τιμές U_1, U_2 από την ομοιόμορφη U στο $(0,1)$

Βήμα 2: Αν $U_2 \le \frac{256}{27}U_1(1-U_1)^3$ θέτουμε $X = U_1$ και σταματούμε. Αλλιώς πηγαίνουμε στο Βήμα 1.

Η προσομοίωση

```
x=0
i=0
for (k in 1:100000)  {
          u1<-runif(1)
          u2<-runif(1)
          b=(256/27)*u1*(1-u1)^3
          if(u2<b) {
                    i <- i+1
                    x[i] <- u1
          }
}

x
hist(x,nclass=50,probability=T)
x1=seq(0,1,.01)
y1=20*x1*(1-x1)^3
points(x1,y1,type="l",col=3,lwd=2)
```



Κανονική κατανομή (με μέθ. απόρ.)

Αν Z~N(0,1) τότε η |Z| έχει πυκνότητα $f(x)=\dfrac{2}{\sqrt{2\pi}}e^{-x^2/2}, 0<x<\infty$

Χρησιμοποιούμε ως κατανομή βάσης την εκθετική με

$$g(x)=e^{-x}, 0<x<\infty$$

Βρίσκουμε $c=\sqrt{\dfrac{2e}{\pi}}$ διότι: $\dfrac{f(x)}{g(x)}=\sqrt{\dfrac{2e}{\pi}}\, e^{-(x-1)^2/2} \le \sqrt{\dfrac{2e}{\pi}}$

Άρα $\dfrac{f(x)}{c\,g(x)}=e^{-(x-1)^2/2}$ και η μέθοδος δίνεται από:

Βήμα 1: Παίρνουμε μία τυχ. τιμή Y από την εκθετική και μία U από την ομοιόμορφη U στο (0, 1)

Βήμα 2: Αν $U \le e^{-(Y-1)^2/2}$ ή ισοδύναμα $-\ln U \ge (Y-1)^2$ θέτουμε X=Y και σταματούμε. Αλλιώς πηγαίνουμε στο Βήμα 1.

συνέχεια

Επειδή η $-lnU$ έχει εκθετική τιμή η μέθοδος γράφεται:

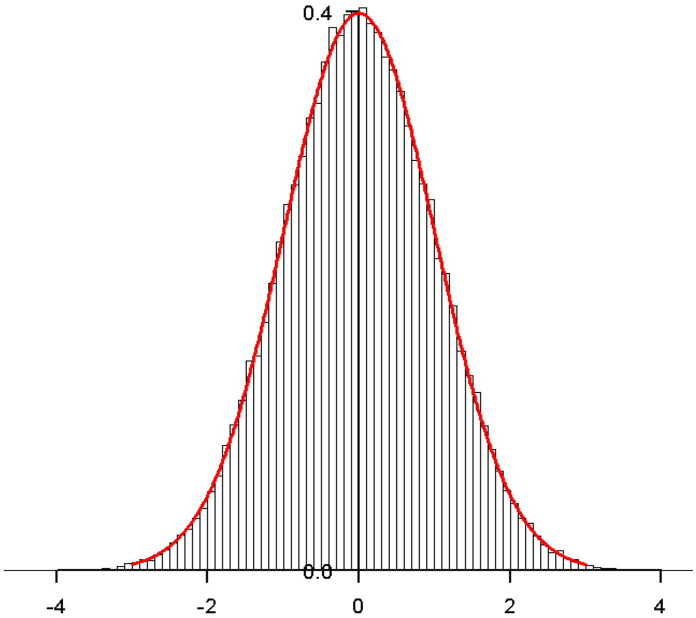
Βήμα 1: Παίρνουμε δύο τυχ. τιμές Y_1, Y_2 από την εκθετική με παρ.1
Βήμα 2: Αν $Y_2 \ge (Y_1-1)^2$ θέτουμε $X=Y_1$ και σταματούμε. Αλλιώς πηγαίνουμε στο Βήμα 1.

Με τα παραπάνω προσομοιώνουμε το |Z| . Για να πάρουμε το Z χρησιμοποιούμε μια τυχαία τιμή στο (0, 1) και αν είναι μικρότερη του $\frac{1}{2}$ κρατούμε τη θετική τιμή που βρήκαμε, αλλιώς την αρνητική της. Δηλαδή:

Βήμα 1: Παίρνουμε δύο τυχ. τιμές Y_1, Y_2 από την εκθετική με παρ.1
Βήμα 2: Αν $Y_2-(Y_1-1)^2 \ge 0$ πηγαίνουμε στο Βήμα 3 και σταματούμε. Αλλιώς πηγαίνουμε στο Βήμα 1.
Βήμα 3: Παίρνουμε μία τυχαία τιμή U στο (0,1). και θέτουμε $X=Y_1$ αν $U \le 1/2$ και $X=-Y_1$ αν U>1/2

Η προσομοίωση για N=100000

```
x=0 ; i=0
for (k in 1:100000)  {
  u1<-runif(1)
  u2<-runif(1)
  y1<--log(u1)
  y2<--log(u2)
  b=y2-(y1-1)^2/2
  if(b>0) {
    i <- i+1
    u<-runif(1)
      if (u<=1/2)
        x[i] <- y1
      else
        x[i] <- -y1
   }
u1;u2;y1;y2;b;u;x
}
x
hist(x,nclass=100,probability=T,
 axes=F,xlab=NULL,ylab=NULL,
 main=NULL)
axis(1,pos=0,line=T,lwd=2)
axis(2,pos=0,line=T,lwd=2,
   at=c(0,0.4),las=1)
x1=seq(-3,3,.01)
y1=(1/sqrt(2*pi))*exp(-x1^2/2)
points(x1,y1,type="l",col=2,lwd=3)
```



Κανονική Κατανομή με άλλο τρόπο

$$F(x)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-\frac{(t-\mu)^2}{2\sigma^2}}\,dt$$

Δεν υπάρχει αναλυτική έκφραση για το $F^{-1}(x)$.

$\Rightarrow$

Δεν εφαρμόζεται το προηγούμενο θεώρημα

Αποδεικνύεται ότι:

Αν $U, V \sim U(0,1)$,

και U, V ανεξάρτητες, τότε

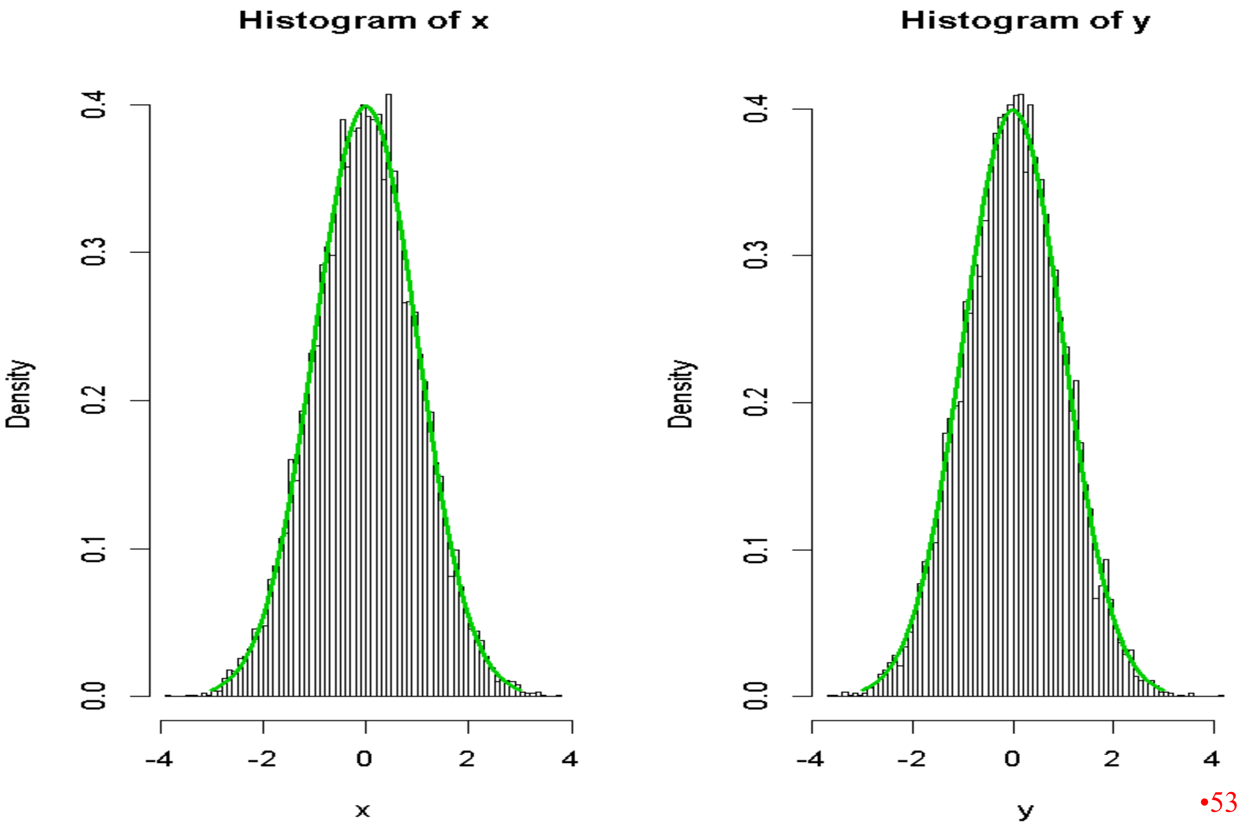
$$\mathrm{X}=\sqrt{-2\ln U}\,\sigma\upsilon\nu 2\pi V \qquad Y=\sqrt{-2\ln U}\,\eta\mu 2\pi V$$

είναι ανεξάρτητες με τυπική N(0,1) κατανομή η κάθε μία

Αν $\mathbf{Z}\sim\mathbf{N}(0,1)$ τότε ισχύει $\mathbf{X}=\boldsymbol{\mu}+\boldsymbol{\sigma}\mathbf{Z}\sim\mathbf{N}(\boldsymbol{\mu},\boldsymbol{\sigma}^2)$

και έτσι προσομοιώνεται οποιαδήποτε κανονική.

Προσομοίωση με n=10000



Με συσκευή Galton

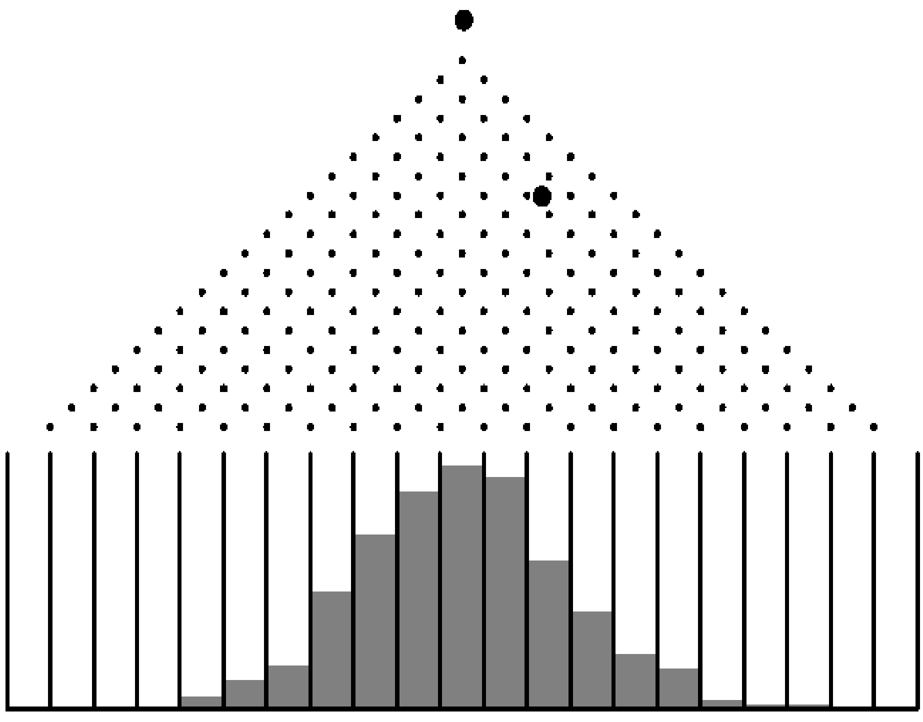


Figure 3.6: Simulation of the Galton board.

Σχέδιο του K.Pearson

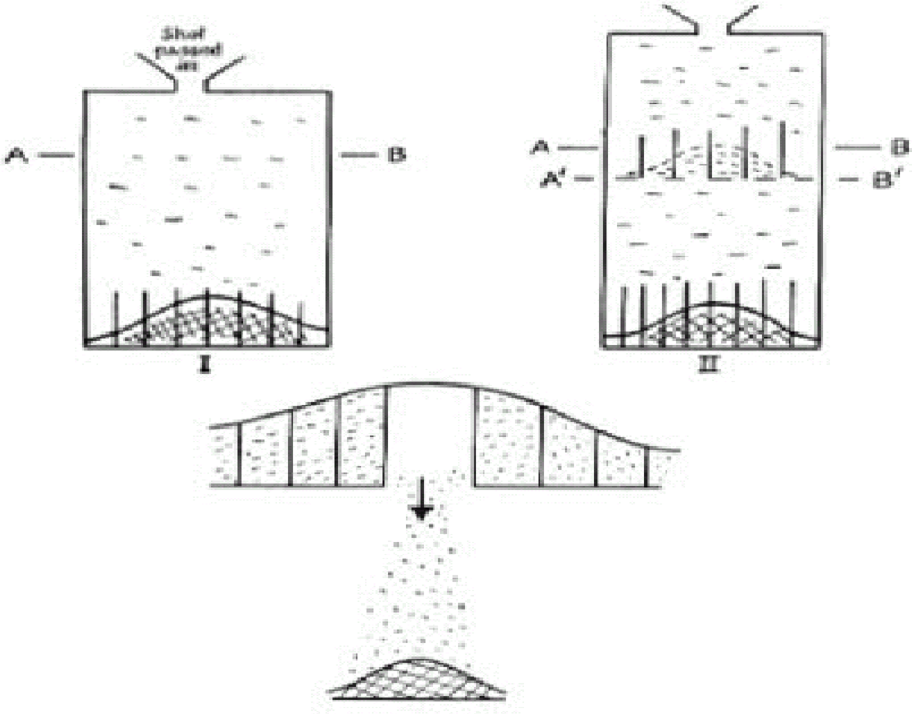
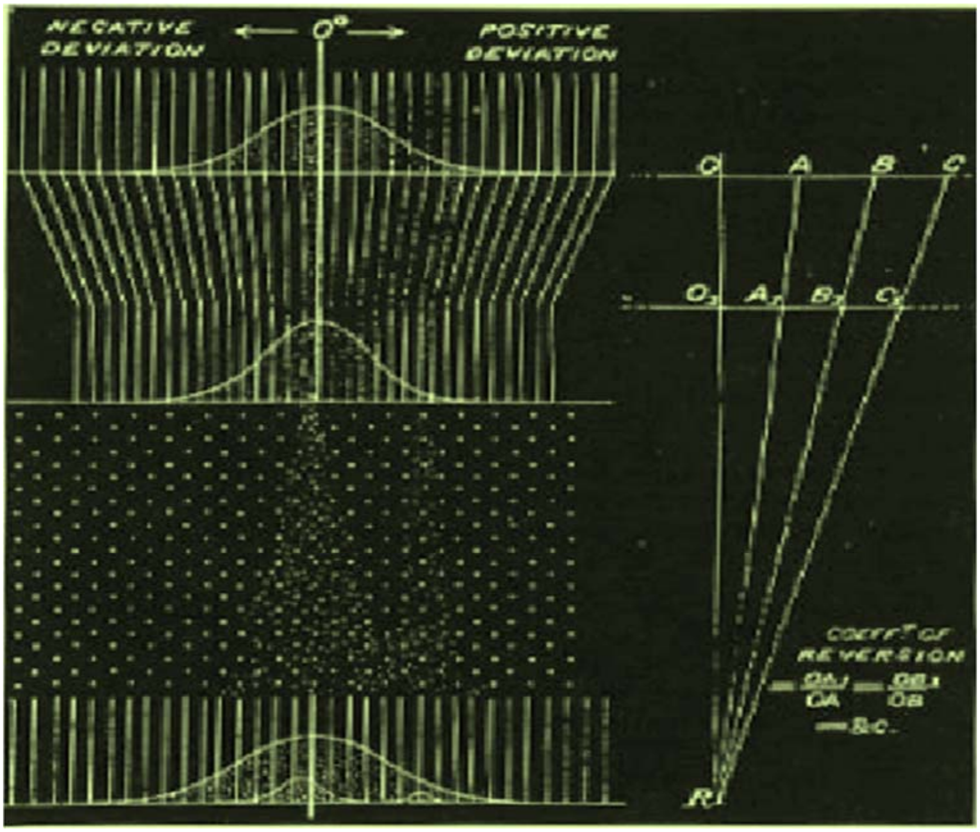
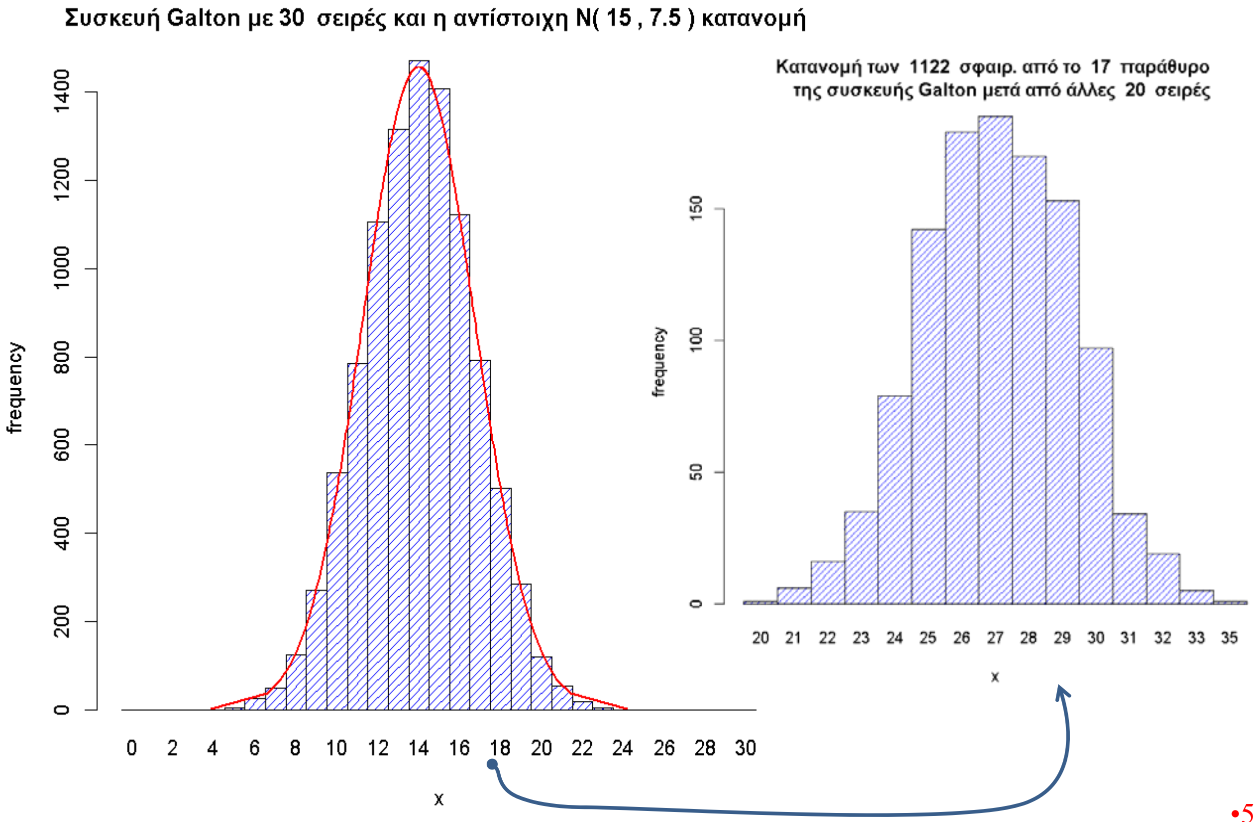


Figure 9.11: Two-stage version of the quincunx.

Reversion ή Regression to the mean



Επιβεβαίωση



Maxwell και Rayleigh κατανομές

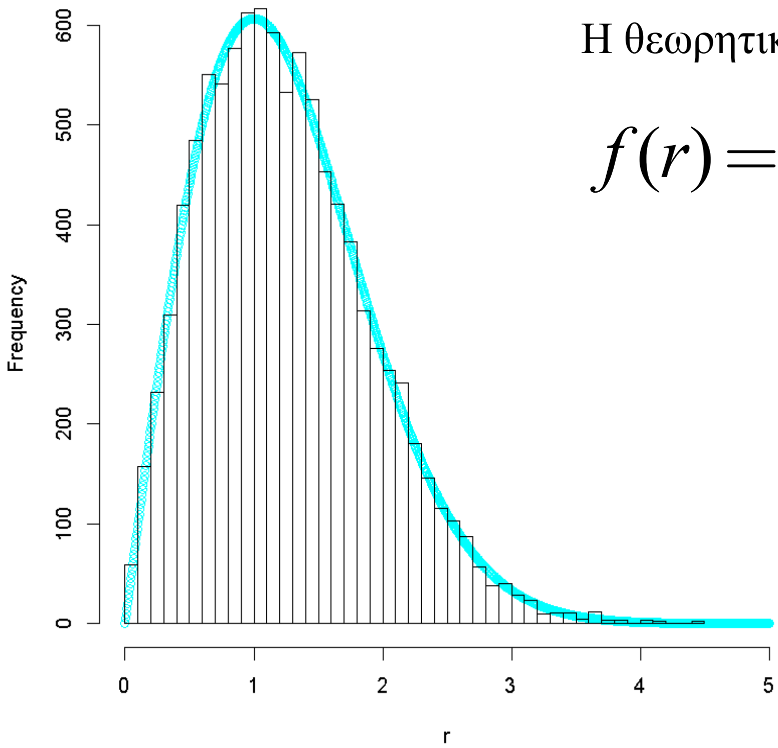
Ένα κινούμενο μόριο στον R^n έχει τις n συνιστώσες της ταχύτητας ανεξάρτητες ανά δύο και κανονικά κατανεμημένες. Ζητείται η κατανομή της ταχύτητας του μορίου.

Αν $n = 3$ τότε έχουμε κατανομή Maxwell,

αν $n = 2$ τότε έχουμε κατανομή Rayleigh

Για $n = 2$ ισοδύναμο πείραμα είναι της ρίψης βελών στο επίπεδο xOy, με την υπόθεση ότι οι συντεταγμένες του σημείου που το βέλος βρίσκει το στόχο είναι από $\mathrm{N}(0,1)$ κατανομή και εμείς ενδιαφερόμαστε για την απόσταση $r = \sqrt{x^2 + y^2}$ του σημείου από την αρχή

Προσομοίωση για n=10000



Η θεωρητική Rayleigh κατανομή

$$f(r) = re^{-r^2/2}, \;\; \gamma\iota\alpha\; r \geq 0$$

χ^2 Κατανομή

Αν $Y = (Y_1, \ldots Y_{k-1})$ έχει κατανομή πολυωνυμική με παραμέτρους n και $p_1, \ldots, p_{k-1}$, $(p_1 + \cdots + p_k = 1)$
τότε (όπως έδειξε ο K. Pearson το 1900)
το στατιστικό Q ακολουθεί $\chi_{k-1}{}^2$

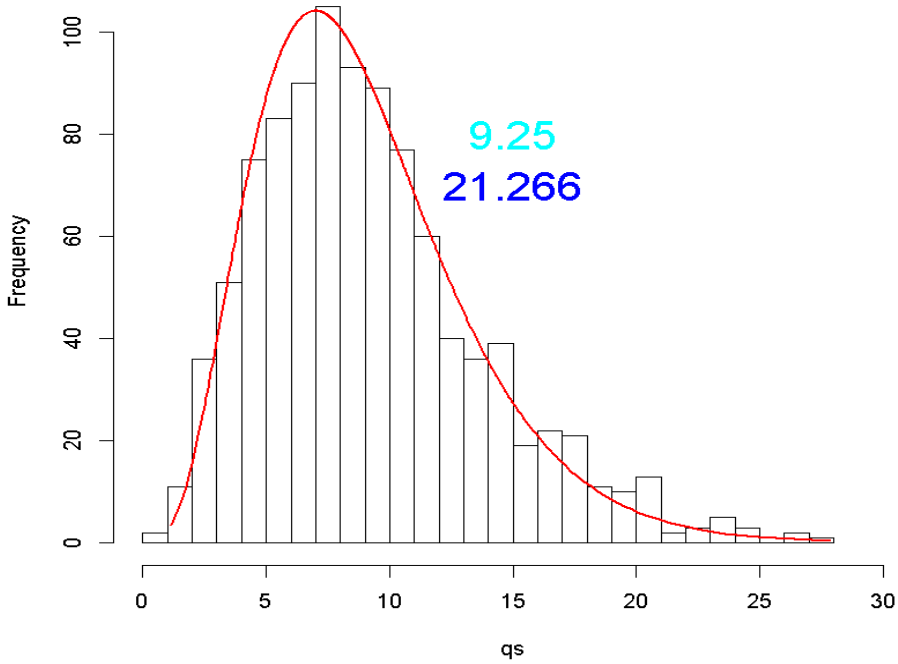
$$Q = \sum_{i=1}^{k} \frac{(Y_i - np_i)^2}{np_i}$$

Παίρνουμε 1000 τυχαίες τιμές στο (0,1) και θέτουμε Y_i το πλήθος των τιμών x που ικανοποιούν την $\frac{i-1}{10} < x < \frac{i}{10}$, $i = 1,2, \ldots, 10$. Τότε το στατιστικό Q θα έχει κατανομή $\chi_9{}^2$, με

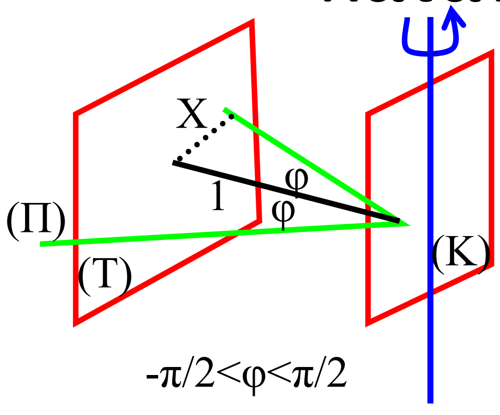
$\mathrm{E}(Q) = 9$ και $Var(Q) = 18$

$$Q = \sum_{i=1}^{10} \frac{(Y_i - 100)^2}{100}$$

Προσομοίωση της χ_9^2



Κατανομή Cauchy



Πρόβλημα Feller (1966)

Ποια η κατανομή της απόστασης Χ στο διπλανό σχήμα. Ο καθρέπτης (Κ) περιστρέφεται, ενώ ο άξονάς του απέχει 1 από τον τοίχο (Τ). Η ακτίνα από την πηγή (Π) ανακλάται κατά γωνία φ που θεωρείται ομοιόμορφα κατανεμημένη στο διάστημα (-π/2, π/2)

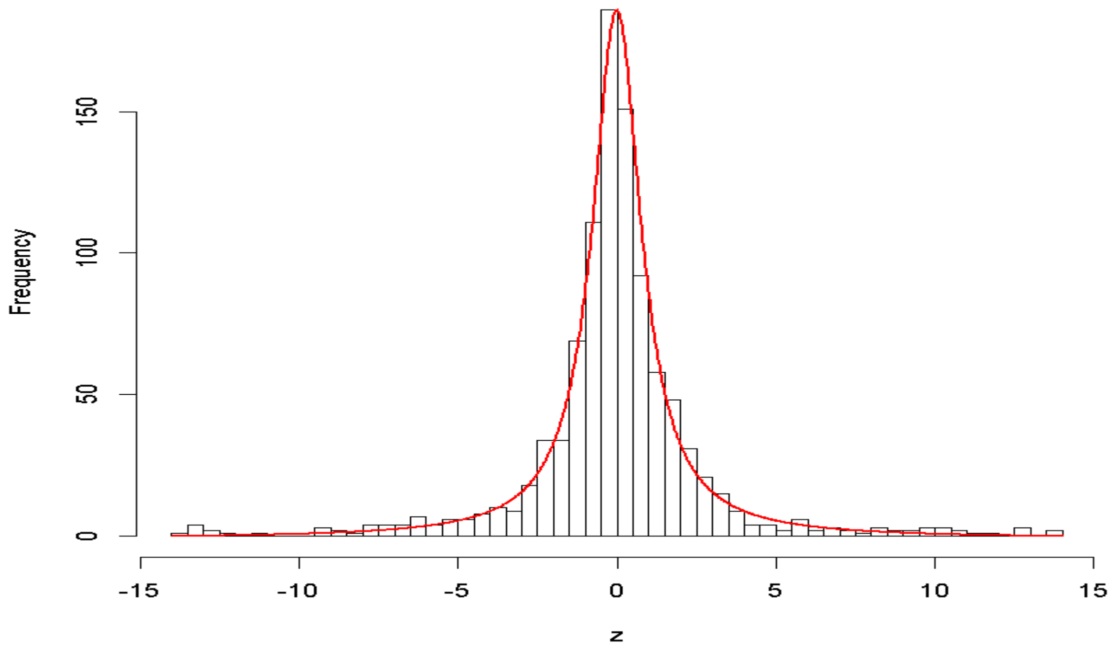
$\varepsilon\varphi\varphi = x$. Αν x σταθερή απόσταση, τότε

$P(X \geq x) = P(\varepsilon\varphi(\Phi) \geq x) = P(\Phi \geq \tau o\xi\varepsilon\varphi x) = (\pi/2 - \tau o\xi\varepsilon\varphi x)/\pi.$

Άρα $F_X(x) = (1/2) + (1/\pi)\tau o\xi\varepsilon\varphi x$, επομένως

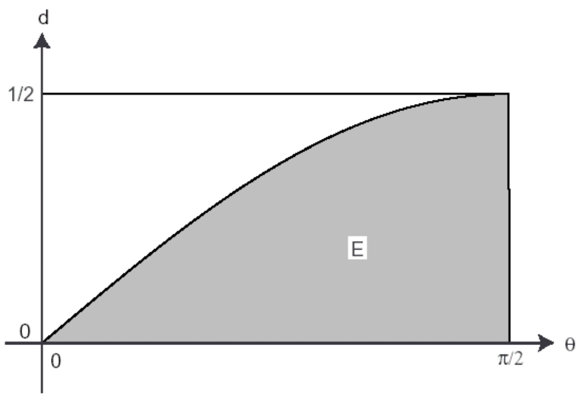
$$\boxed{f_X(x) = \frac{1}{\pi(1+x^2)} \quad \gamma\iota\alpha\; x \in \mathbb{R}}$$

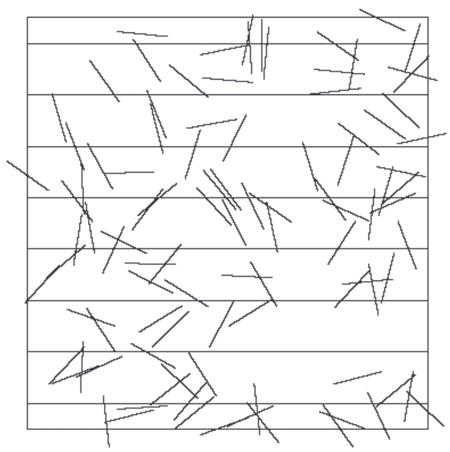
Προσομοίωση για n=1000



Η βελόνα του Buffon

Προσομοιώσαμε το πρόβλημα της βελόνας του Buffon θεωρώντας τη βελόνα να έχει μήκος 1 και τις παράλληλες να απέχουν μήκος 1. Η βελόνα καθορίζεται από τα d (απόσταση από την πλησιέστερη παράλληλο) και θ (οξεία γωνία της βελόνας με τις παράλληλες). Ισχύει $0 \le d \le 1/2$ και $0 \le \theta \le \pi/2$.





$d < ½\, \eta\mu\theta \qquad P(E) = 2/\pi$

κάθε 100 βελόνες /10000

εδώ $\pi = 3.139$