ΤΕΧΝΙΚΕΣ ΠΡΟΣΟΜΟΙΩΣΗΣ

Επικ. Καθ. Στέλιος Ζήμερας

Τμήμα Μαθηματικών

Κατεύθυνση Στατιστικής και Αναλογιστικών – Χρηματοοικονομικών Μαθηματικών

2017

Διακριτές κατανομές

επιθυμούμε να παράγουμε αριθμούς από μία κατανομή με συνάρτηση πιθανότητας

$$p_i = \Pr(X = x_i),\ i = 0,1,2,\ldots \qquad \sum_i p_i = 1$$

Ολες οι μέθοδοι βασίζονται στη δημιουργία μίας

ακολουθίας τυχαίων μεταβλητών U~U(0,1) οι οποίες μετά από κατάλληλο μετασχηματισμό δίνουν τις X.

Διακριτές κατανομές

Υποθέτουμε ότι δύο τ.μ. X και Y έχουν την ίδια κατανομή. Τότε σε κάποιο πείραμα προσομοίωσης που χρησιμοποιεί την X μπορούμε εναλλακτικά να χρησιμοποιήσουμε την Y.

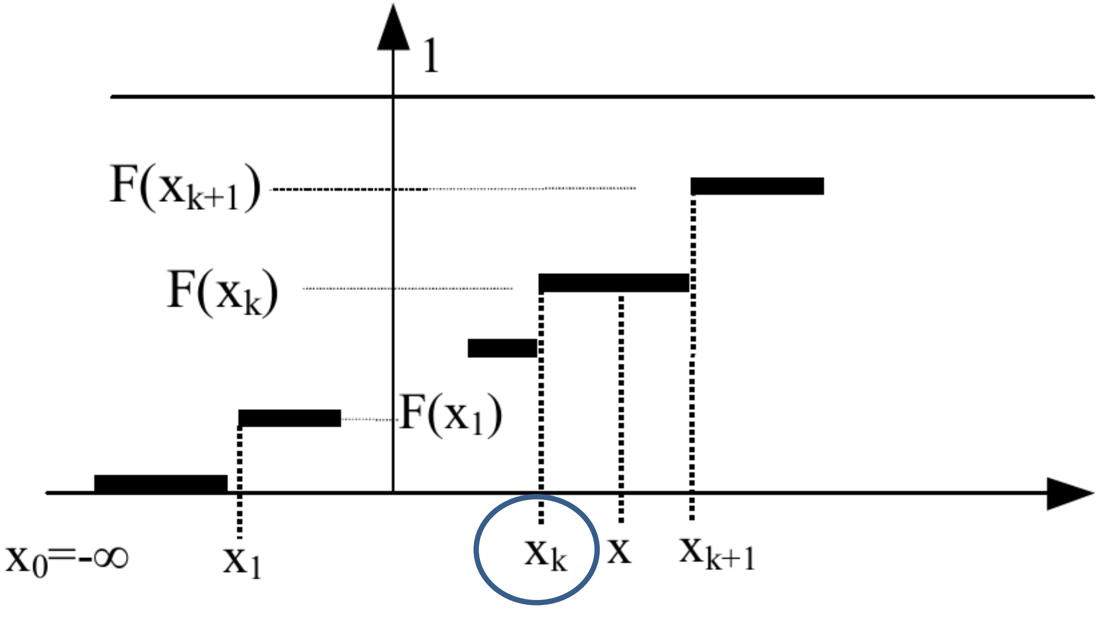
Οι X και Y ονομάζονται στοχαστικά ίσες τ.μ κατά κατανομή.

Αν η τ.μ. $X < \infty$ παίρνει τιμές από ένα αριθμήσιμο σύνολο $\{x_1, x_2, \ldots, x_k, \ldots\}$ με αύξουσα διάταξη, τότε η συνάρτηση κατανομής της είναι

$$F(x)=P(X\leq x)= P(X=x_1)+ \ldots +P(X=x_k)$$

Διακριτές κατανομές

όπου x_k είναι η τιμή της X που είναι η πιο κοντινή και ≤ από την x



Διακριτές κατανομές

- **Η μέθοδος της αντιστροφής**

Η μέθοδος αντιστροφής εφαρμόζεται όταν η F έχει αντίστροφη F^{-1}.

F(x) = U έχει πολλές λύσεις $\Rightarrow$ $\nexists$ αντίστροφη F^{-1}

$\Downarrow$

αν $U = F(x_k)$ τότε $F(x)=U$ για κάθε x: $x_k \leq x < x_{k+1}$.

Για την γέννηση τυχαίων αριθμών X έχουμε:

Διακριτές κατανομές

B1. Γέννησε τον τυχαίο αριθμό $U \sim U(0,1)$.

B2. Βρές τις δύο διαδοχικές τιμές x_k και x_{k-1} της X για τις οποίες $F(x_{k-1}) < U \leq F(x_k)$ και θέσε

$$X^* = x_k$$

Θα πρέπει η X^* να είναι στοχαστικά ίση με την X

$\mathbf{P(X^* = x_k)} = P[F(x_{k-1}) < U \leq F(x_k)] =$

$= P[\, U \leq F(x_k)\,] - P[U \leq F(x_{k-1})\,] =$

$= F(x_k) - F(x_{k-1})\ (U \sim U(0,1)) =$

$= [P(X=x_k)+P(X=x_{k-1})+\ldots] - [P(X=x_{k-1})+\ldots]$ (F(x) ορισμό)

$\mathbf{= P(X=x_k)}$ άρα X^* στοχαστικά ίση με X

Διακριτές κατανομές

- **Η μέθοδος της αντιστροφής**

Ένας τρόπος να παράγουμε τυχαίους αριθμούς από την κατανομή με σ.π. $\{p_0, p_1,\ldots\}$ χρησιμοποιώντας τυχαίους αριθμούς από την ομοιόμορφη κατανομή είναι ο ακόλουθος:

Έστω U~ U(0,1). Η νέα τυχαία μεταβλητή

$$X = \begin{cases} x_0, & 0 \leq U < p_0 \\ x_1, & p_0 \leq U < p_0 + p_1 \\ \vdots & \\ x_k, & \sum_{i=0}^{k-1} p_i \leq U < \sum_{i=0}^{k} p_i \\ \vdots & \end{cases}$$

θα έχει συνάρτηση πιθανότητας $\{p_0,p_1,\ldots\}$

Διακριτές κατανομές

Πράγματι

$$\Pr(X = x_k) = \Pr(\sum_{i=0}^{k-1} p_i \le U < \sum_{i=0}^{k} p_i) = \sum_{i=0}^{k} p_i - \sum_{i=0}^{k-1} p_i = p_k \,, \quad k=0,1,2,\ldots\,.$$

Εάν θεωρήσουμε ότι $x_0 < x_1 < x_2$ ... η αντίστοιχη συνάρτηση κατανομής στα σημεία $\{x_0, x_1, \ldots\}$ θα είναι

$$F(x_k) = \Pr(X \le x_k) = \sum_{i=0}^{k} \Pr(X = x_i) = \sum_{i=0}^{k} p_i \,, \quad k = 0,1,\ldots$$

Σε αυτή την περίπτωση η X παράγεται από την U ως εξής:

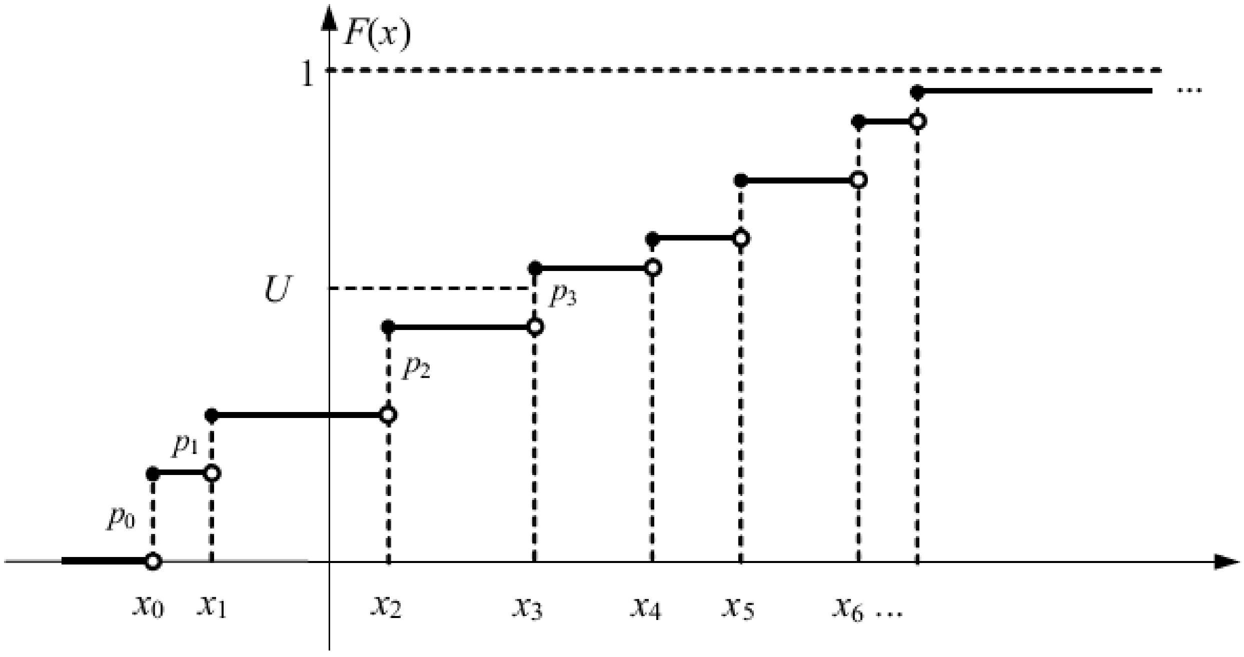
$$X = \{x_k : F(x_{k-1}) \le U < F(x_k)\} \quad \text{ή} \quad X = \{x_k : F(x_{k-1}) < U \le F(x_k)\}$$

Διακριτές κατανομές

Η γενικευμένη αντίστροφη συνάρτηση μιας σ.κ. F είναι η $F^{-1}(y) = \inf F^{-1}([y,1]) = \inf\{x : F(x) \in [y,1]\}$,

η οποία συμπίπτει με την συνήθη αντίστροφη όταν η F αντιστρέφεται.

Επαληθευτεί ότι $X = F^{-1}(U)$ αν $X = \{x_k : F(x_{k-1}) < U \le F(x_k)\}$



Διακριτές κατανομές

B1. Παράγουμε έναν τυχαίο αριθμό U~ U(0,1)

B2. Αν U< p_0 τότε θέτουμε X = x_0 και σταματάμε, αλλιώς πάμε στο B3

B3. Αν U< $p_0 + p_1$ τότε θέτουμε X = x_1 και σταματάμε, αλλιώς πάμε στο B4, κ.λ.π.

Αν N είναι το πλήθος των βημάτων αυτού του αλγορίθμου, τότε

$$E(N) = \sum_{i=1}^{\infty} i \Pr(N = i) = \sum_{i=1}^{\infty} i p_{i-1} \, .$$

Διακριτές κατανομές

Παράδειγμα

Θέλουμε να παράγουμε τυχαίους αριθμούς X από την κατανομή με συνάρτηση πιθανότητας

$$p_1=0.3,\ p_2=0.2,\ p_3=0.5$$

Έχουμε $p_1=0.3$, $p_1+p_2=0.5$, $p_1+p_2+p_3=1$

$$E(N)= 1*0.3+2*0.2+3*0.5=2.2$$

Παράγουμε τυχαίους αριθμούς με βάση την σχέση

$$X=\begin{cases} 1, & U\le 0.3 \\ 2, & 0.3<U\le 0.5 \\ 3, & 0.5<U\le 1 \end{cases}$$

Διακριτές κατανομές

Γενικά:

B1. Παραγωγή τ.α. από U~(0,1)

B2. Αν $U \leq p_1$ τότε $X=x_1$

αλλιώς αν $U \leq p_1 + p_2$ τότε $X=x_2$

....

αλλιώς αν $U \leq p_1 + p_2 + \ldots\ldots + p_i$ τότε $X=x_i$

....

αλλιώς αν $U \leq p_1 + p_2 + \ldots\ldots + p_{n-1}$ τότε $X=x_{n-1}$

αλλιώς $X=x_n$

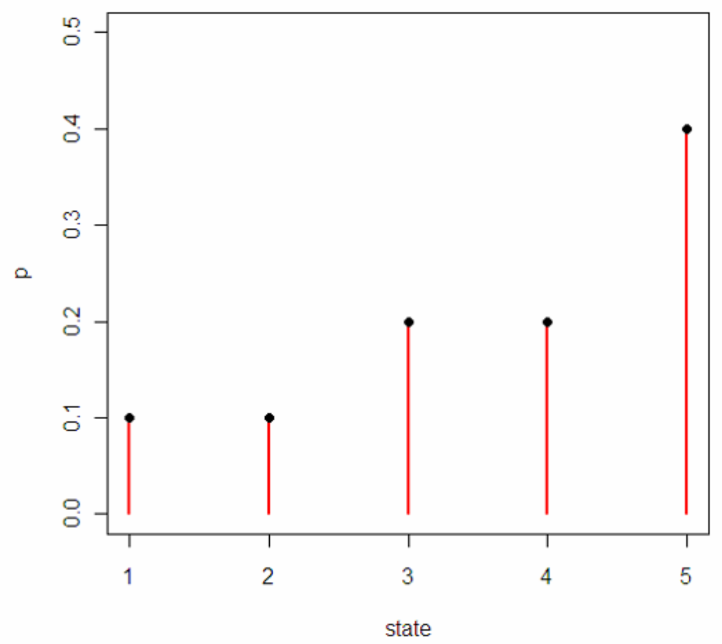
τέλος

Διακριτές κατανομές

Πράγματι

$$P\left(X=x_i\right)=P\left(p_1+p_1+\ldots+p_{i-1}<U\le p_1+p_1+\ldots+p_i\right)=F_{i-1}-F_i=p_i$$

$$\text{Αν } \mathrm{U}_i \sim U(0,1),\ 1\le \mathrm{j}\le \mathrm{N} \text{ τότε } \mathrm{x}_i \sim p\left(.\right),\ 1\le \mathrm{j}\le \mathrm{N}$$



Διακριτές κατανομές

Τυχαίοι αριθμοί

πιθανότητες

```
n = 1000; X = zeros(1,n); x = 1:3; pr = [0.3 0.2 0.5];
for i = 1:n
    u = rand;
    if u <= pr(1)
      X(i) = x(1);
      elseif u <= sum(pr(1:2))
        X(i) = x(2);
     else
        X(i) = x(3);
    end
end
```

Τα ποσοστά κάθε αριθμού

```
x1 = length(find(X==1))/n;
x2 = length(find(X==2))/n;
x3 = length(find(X==3))/n;
```

$p_1 + p_2$

$$X = \begin{cases} 1, & U \le 0.3 \\ 2, & 0.3 < U \le 0.5 \\ 3, & 0.5 < U \le 1 \end{cases}$$

Διακριτές κατανομές

Ομοιόμορφη στο{1, 2, ..., n} κατανομή(U{1,2,...,n})

Κατανομή:

$$p_i = \Pr(X = i) = \frac{1}{n},\ i = 1,2,\ldots,n$$

παράγουμε U~ U(0,1)

αν U< 1/n τότε θέτουμε X = 1 και σταματάμε,

αν U< 2/n τότε θέτουμε X = 2 και σταματάμε κ.λ.π.

$$X = \{k : \tfrac{k-1}{n} \leq U < \tfrac{k}{n}\}$$

Διακριτές κατανομές

Εναλλακτικά μπορούμε να παράγουμε από

$$\boxed{X = \lfloor nU \rfloor + 1}$$

με

$$\Pr(X = i) = \Pr(\lfloor nU \rfloor = i - 1) = \Pr(i - 1 \le nU < i) = \Pr(\frac{i-1}{n} \le U < \frac{i}{n}) = \frac{1}{n},\ x = 1,2,\ldots,n$$

B1. Παράγουμε έναν τυχαίο αριθμό U~ U(0,1)

B2. Θέτουμε $X = \lfloor nU \rfloor + 1$

Διακριτές κατανομές

N = 5; (παράγουμε 5 νούμερα)

n = 500; (500 φορές)

```
X = ceil(N*rand(1,n));
freq = zeros(1,N);
for i = 1:N
    freq(i) = length(find(X==i))/n;
end
```

Τα αποτελέσματα πρέπει να είναι κοντά στα θεωρητικά 1/N = 1/5=0.2

Διακριτές κατανομές

Εναλλακτικά

```
 N = 5; (παράγουμε 5 νούμερα)
n = 500; (500 φορές)
x = unif(N,n);
freq = zeros(1,N);
for i = 1:N
    freq(i) = length(find(x==i))/n;
end
```

```
function X = unif(N,n)
X = ceil(N*rand(1,n));
```

Διακριτές κατανομές

Ομοιόμορφη διακριτή κατανομή U~(α, β)

συνάρτηση πυκνότητας πιθανότητας

$$f(x) = \frac{1}{\beta - \alpha + 1}$$

με κατανομή:

$$p = P(X = i) = \frac{1}{\beta - \alpha + 1}, \quad x = \alpha, \alpha+1, \ldots, \beta-1, \beta.$$

Διακριτές κατανομές

B1. Παράγουμε τυχαίο δείγμα U από U~U(0, 1)

B2. Θέτουμε x = α + [U(β-α+1)].

ακέραιο μέρος

η πιθανότητα η τυχαία μεταβλητή X, που ακολουθεί την ομοιόμορφη διακριτή κατανομή, να πάρει μια τιμή στο διάστημα [α, β] είναι

$$p = P(X = i) = \frac{1}{\beta - \alpha + 1} \Rightarrow (\beta\text{-}\alpha+1)\cdot p = 1 \Rightarrow (\beta\text{-}\alpha+1) = \frac{1}{p}$$

B2. Θέτουμε x = α + [u/p]

Διακριτές κατανομές

Αποτελέσματα

α	$0 \leq u \leq p$
$\alpha+1$	$p \leq u \leq 2p$
$\alpha+2$	$2p \leq u \leq 3p$
……	………………
$\alpha+i$	$i \cdot p \leq u \leq (i+1) \cdot p$
……	………………
β	$(\beta-\alpha) \cdot p \leq u \leq (\beta-\alpha+1) \cdot p = 1$

Διακριτές κατανομές

Γεωμετρική κατανομή (Ge(p))

συνάρτηση πιθανότητας

$$p_i = \Pr(X = i) = pq^{i-1},\ i = 1,2,\ldots \qquad (q = 1 - p)$$

Η τ.μ. Χ εκφράζει το πλήθος των δοκιμών μέχρι και την εμφάνιση της πρώτης επιτυχίας σε μία ακολουθία ανεξάρτητων δοκιμών με πιθανότητα επιτυχίας p

$$F(k) = \sum_{i=1}^{k} p_i = \sum_{i=1}^{k} pq^{i-1} = p\sum_{i=0}^{k-1} q^i = p\frac{1-q^k}{1-q} = 1 - q^k$$

Διακριτές κατανομές

Για τυχαίο αριθμό U~ U(0,1) θα ισχύει

$$X=\{k\geq 1:1-q^{k-1}\leq U<1-q^{k}\}$$

$$X=k \Leftrightarrow 1-q^{k}<U<1-q^{k+1} \Leftrightarrow q^{k}>1-U\geq q^{k+1}$$

$$\Downarrow$$

$$k<\frac{\ln(1-U)}{\ln q}\leq k+1 \quad \Rightarrow \quad \frac{\ln(1-U)}{\ln q}-1\leq k<\frac{\ln(1-U)}{\ln q}$$

Διακριτές κατανομές

$$E(N) = \sum_{i=1}^{\infty} i \Pr(N = i) = \sum_{i=1}^{\infty} i p_i = \frac{1}{p}$$

Εναλλακτικά

$$\boxed{X = \left\lfloor \frac{\ln U}{\ln q} \right\rfloor + 1}$$

Διακριτές κατανομές

η τ.μ. X ακολουθεί την γεωμετρική κατανομή με παράμετρο p

$$P(X=k)=P\left(\left\lfloor\frac{\ln U}{\ln q}\right\rfloor=k-1\right)=P\left(k-1\le\frac{\ln U}{\ln q}<k\right)$$

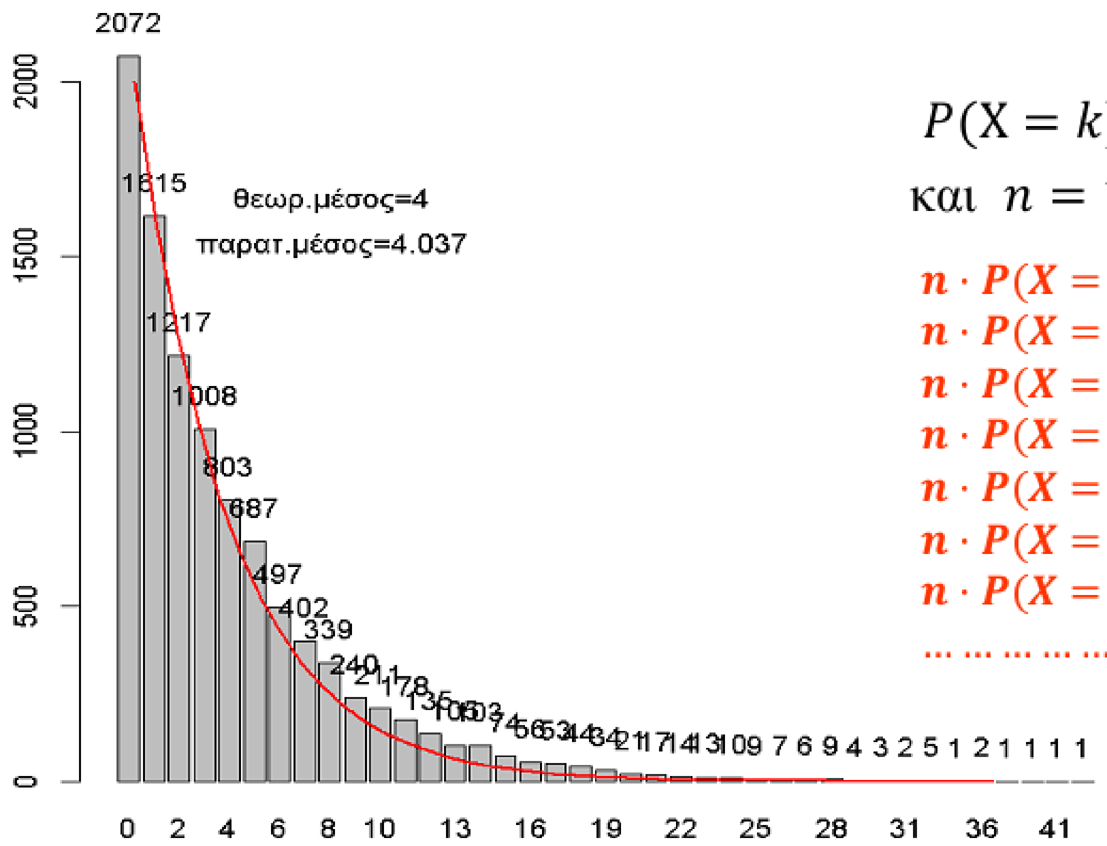
$$=P\left(k\ln q<\ln U\le(k-1)\ln q\right)=P\left(q^{k}<U\le q^{k-1}\right)=q^{k-1}-q^{k}=pq^{k-1}$$

B1. Παράγουμε τυχαίο αριθμό U~ U(0,1)

B2. Θέτουμε

$$X=\left\lfloor\frac{\ln U}{\ln q}\right\rfloor+1$$

Διακριτές κατανομές



$$P(\mathrm{X}=k)=0.2(0.8)^k$$

και $n=10000$

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

$$n=10000, p=.2$$

Διακριτές κατανομές

κατανομή Poisson (Po(λ))

συνάρτηση πιθανότητας

$$p_i = \Pr(X = i) = e^{-\lambda} \frac{\lambda^i}{i!}, \quad i = 0,1,2,\ldots$$

Η τ.μ. X εκφράζει το πλήθος των επιτυχιών σε

ένα μεγάλο αριθμό δοκιμών n → ∞ με πιθανότητα επιτυχίας p →0 σε κάθε δοκιμή ώστε np → λ.

Διακριτές κατανομές

B1. Παράγουμε έναν τυχαίο αριθμό U~ U(0,1)

B2. Αν U< $e^{-\lambda}$ τότε θέτουμε X = 0 και σταματάμε, αλλιώς πάμε στο B3

B3. Αν U< $e^{-\lambda}$+ $e^{-\lambda}\lambda/1!$ τότε θέτουμε X = 1 και σταματάμε, αλλιώς πάμε στο B4

Διακριτές κατανομές

Εναλλακτικά:

$$p_{i+1} = e^{-\lambda} \frac{\lambda^{i+1}}{(i+1)!} = \frac{\lambda}{i+1} e^{-\lambda} \frac{\lambda^{i}}{i!} = \frac{\lambda}{i+1} p_i$$

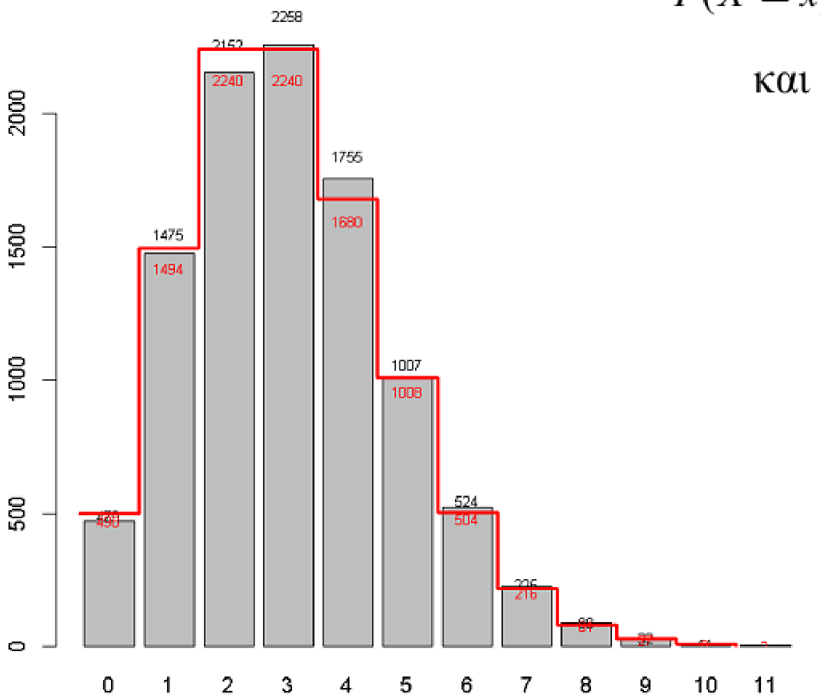
B1. Παράγουμε τυχαίο αριθμό U~ U(0,1)

B2. Θέτουμε p = $e^{-\lambda}$, F = p, i = 0

B3. Αν U< F τότε θέτουμε X = i και σταματάμε

B4. Θέτουμε p = p·λ/(i+1), F = F + p, i = i + 1 και επιστρέφουμε στο B3.

Διακριτές κατανομές



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

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

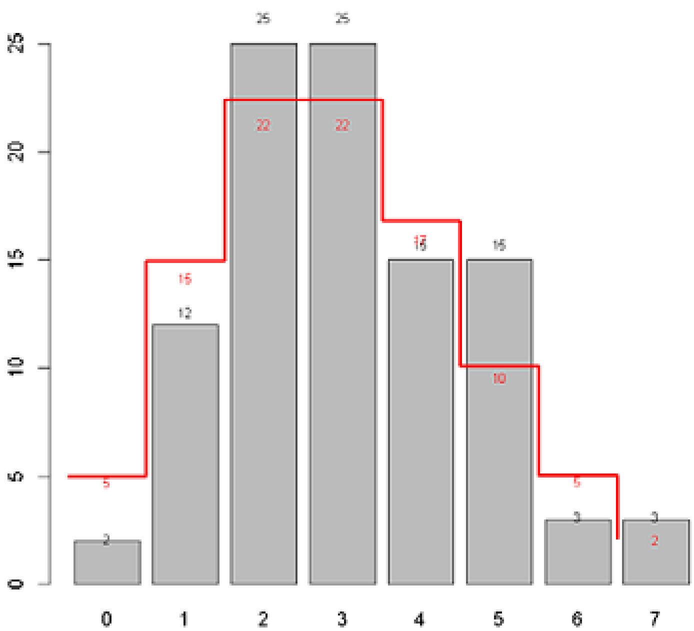
n·P(X=7)=216

n·P(X=8)=81

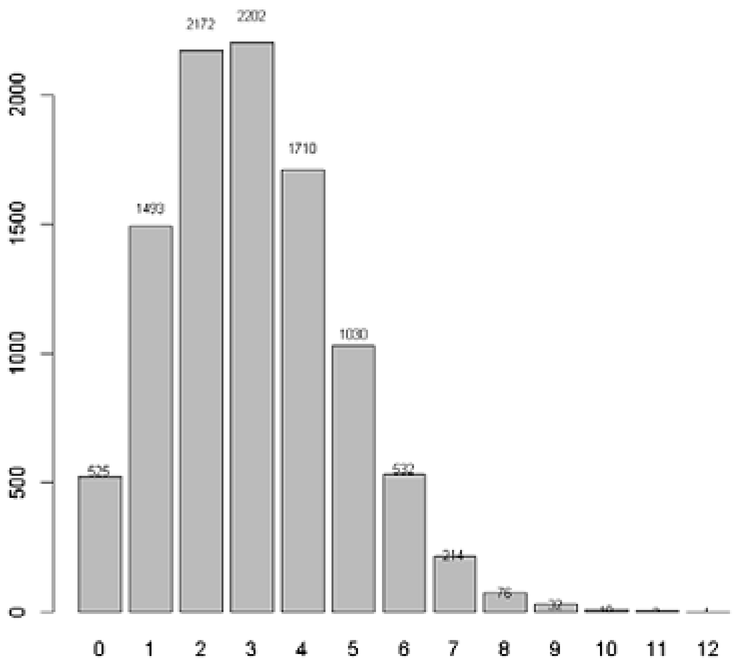
n=10000, λ=3

Διακριτές κατανομές

λ=3



100



10000

Διακριτές κατανομές

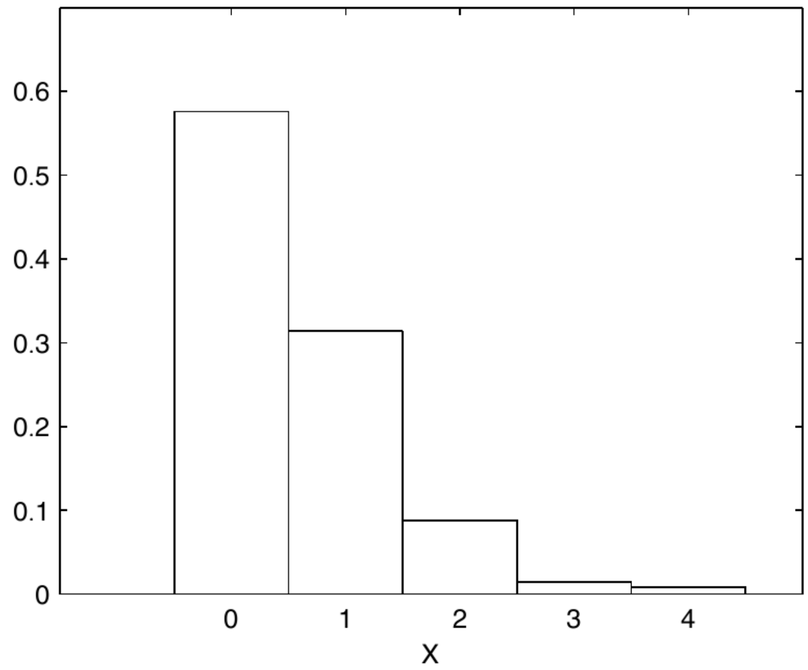
B1. Παράγουμε τυχαίο αριθμό U~ U(0,1)
B2. Θέτουμε p = $e^{-\lambda}$, F = p, i = 0
B3. Αν U< F τότε θέτουμε X = i και σταματάμε
B4. Θέτουμε p = p·λ/(i+1), F = F + p, i = i + 1
και επιστρέφουμε στο B3.

```
function x = dpois(lam,n)
x = zeros(1,n); j = 1;
while j <= n
    flag = 1; u = rand(1); i = 0; p = exp(-lam); F = p;
    while flag
        if u <= F
        x(j) = i; flag = 0; j = j+1;
       else
        p = lam*p/(i+1); i = i+1; F = F + p;
      end
   end
end
```

Διακριτές κατανομές

```
lam = 0.5;
N = 5000;
x = dpois (lam,N);
edges = 0:max(x);
f = histc(x,edges);
bar(edges,f/N,1,'w')
```

```
function x = dpois(lam,n)
```



Διακριτές κατανομές

Διωνυμική κατανομή(Bi(n,p))

συνάρτηση πιθανότητας

$$p_i = \Pr(X = i) = \binom{n}{i} p^i (1-p)^{n-i}, \quad i = 0,1,2,\ldots,n$$

Η τ.μ. X εκφράζει το πλήθος των επιτυχιών σε n ανεξάρτητες δοκιμές σε κάθε μία από τις οποίες εμφανίζεται επιτυχία με p και αποτυχία με 1–p.

$$p_{i+1} = \frac{n!}{(i+1)!(n-i-1)!} p^{i+1}(1-p)^{n-i-1} = \frac{n-i}{i+1}\frac{p}{1-p}\frac{n!}{i!(n-i)!} p^i (1-p)^{n-i} = \frac{n-i}{i+1}\frac{p}{1-p} p_i$$

αναδρομική σχέση

Διακριτές κατανομές

Μια δοκιμή Bernoulli ακολουθεί τα εξής βήματα:

B1. Γέννηση τυχαίου δείγματος u από ομοιόμορφη κατανομή U(0, 1).

B2. Εάν u < p τότε έχουμε επιτυχία, ενώ εάν u > p τότε έχουμε αποτυχία.

Δυωνυμική

Γίνεται ένας αριθμός δοκιμών Bernoulli, σύμφωνα με τη παραπάνω διαδικασία, και στη συνέχεια μετριέται ο αριθμός των επιτυχιών (x). Αυτός ο αριθμός είναι ένα δείγμα από μια δυωνυμική κατανομή B(n, p).

Διακριτές κατανομές

Αλγόριθμος

B1. Παράγουμε έναν τυχαίο αριθμό U ~ U(0,1)

B2. Θέτουμε pr = $(1-p)^n$, F = pr,i = 0.

B3. Αν U< F τότε θέτουμε X = i και σταματάμε.

B4. Θέτουμε

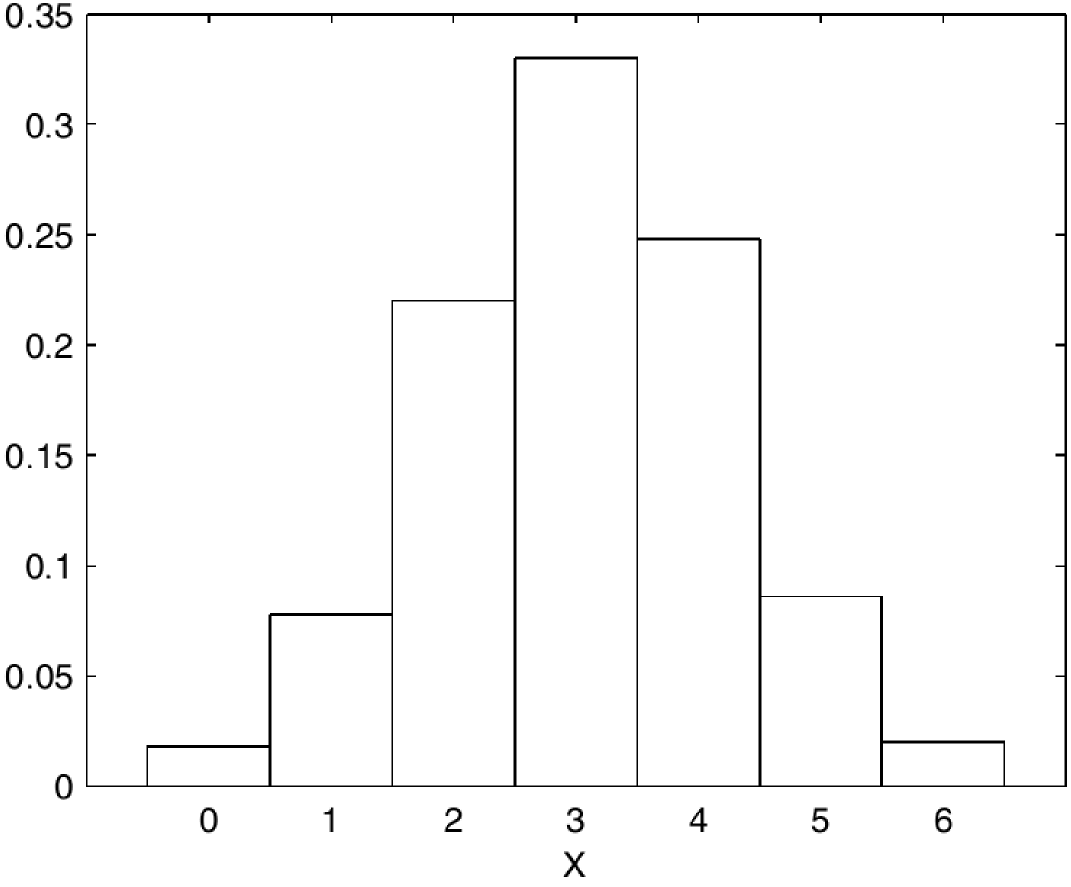
$$pr = \frac{n-1}{i+1}\frac{p}{1-p}pr, F = F + pr, i = i+1$$

και επαναλαμβάνουμε το3

Διακριτές κατανομές

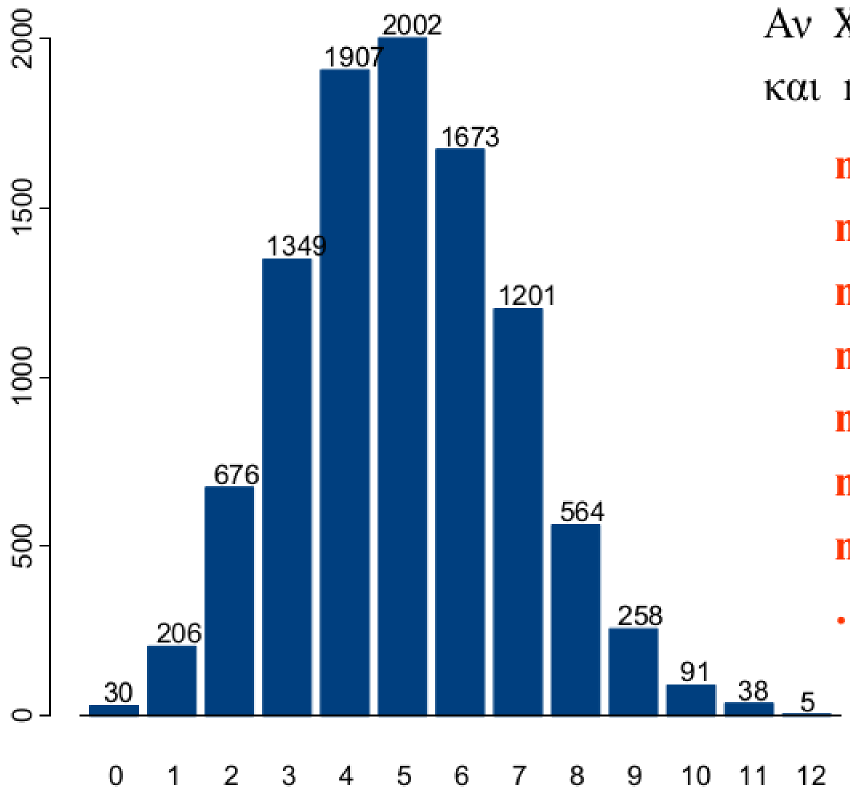
```
function X = dbin(n,p,N)
X = zeros(1,N);
U = rand(N,n);
for i = 1:N
        ind = find(U(i,:) <= p);
        X(i) = length(ind);
end
```

Διακριτές κατανομές



n=6, p=0.5

Διακριτές κατανομές



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

....