Κατανομές

Κολοβού Αθανασία
http://users.uoa.gr/~akolovou/

Bernoulli Trials-Binomial Distribution

- Bernoulli πείραμα
  - Περιγράφει ένα τυχαίο πείραμα με δυο πιθανά αποτελέσματα (επιτυχία - αποτυχία) και με πιθανότητα επιτυχίας p , και αποτυχίας 1-p

- Η τυχαία μεταβλητή Χ που εκφράζει τον αριθμό των επιτυχιών σε n ανεξάρτητα πειράματα με πιθανότητα επιτυχίας p κάθε φορά λέγεται Διωνυμική τυχαία μεταβλητή

- Δίνεται από τον τύπο $P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$ $\quad k=0,1,2\ldots,n$

- Όπου $\binom{n}{k}=\dfrac{n!}{k!(n-k)!}$ και $\mu=\mathrm{np}$, $\sigma^2=np(1-p)$

Bernoulli Trials-Binomial Distribution

- Η εντολή στο Matlab είναι η Y = binopdf(X,N,P)
- Υπολογίζει την συνάρτηση πυκνότητας πιθανότητας (Probability Density Function ή pdf)
- Ποια είναι η πιθανότητα να πάρουμε 3 φορές Κ(Κεφαλή) σε 5 ρίψεις νομίσματος ;

  Αριθμός επαναλήψεων: 5

  Ακριβής αριθμός επιτυχιών: 3

  Πιθανότητα επιτυχίας: 0.5

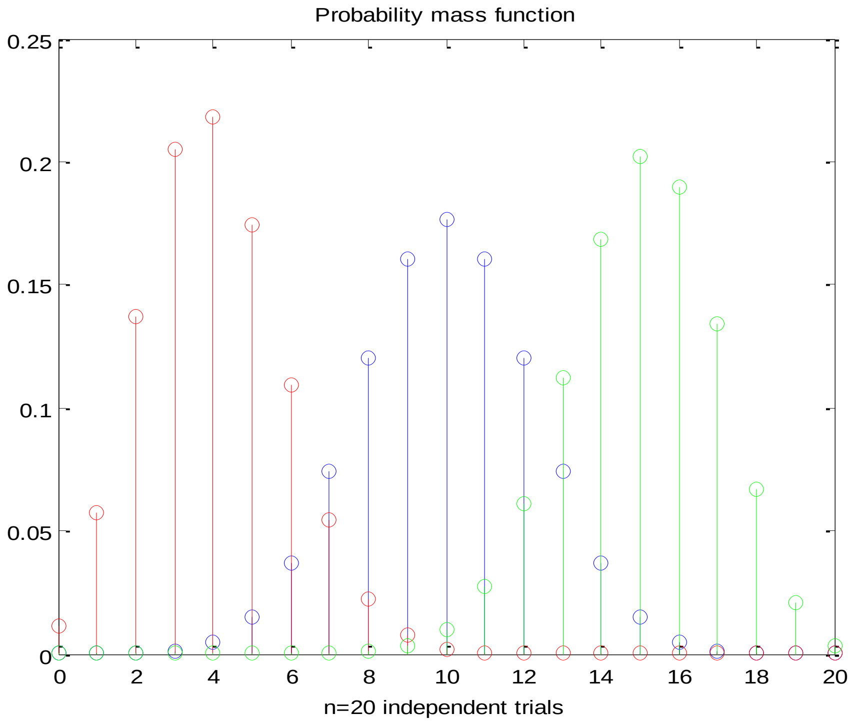
  ```
  binopdf(3,5,1/2)

  ans = 0.3125
  ```

Bernoulli Trials-Binomial Distribution

- Παράδειγμα με τυχαία μεταβλητή και διαφορετικές τιμές πιθανότητας επιτυχίας.

- Χρησιμοποιούμε τη stem για να τονίσουμε ότι η συνάρτηση πυκνότητας πιθανότητας παίρνει θετικές τιμές στο διάστημα 0,1,…n .

```
k=0:20;
y=binopdf(k,20,0.5);
h=binopdf(k,20,0.5);
t=binopdf(k,20,0.75);
stem(k,y);hold on;
stem(k,h,'r');
stem(k,t,'g');
```



Bernoulli Trials-Binomial Distribution

- Η αθροιστική συνάρτηση κατανομής (cumulative distribution function – cdf) περιγράφει την πιθανότητα

$$P(X \leq x)$$

- Και δίνεται με την εντολή Y = binocdf(X,N,P)

- Για παράδειγμα έστω ότι θέλουμε να υπολογίσουμε την πιθανότητα ότι μία διωνυμική τυχαία μεταβλητή X με πιθανότητα επιτυχία 0.2 βρίσκεται μεταξύ του 4 και του 16. Στην πραγματικότητα αυτό είναι

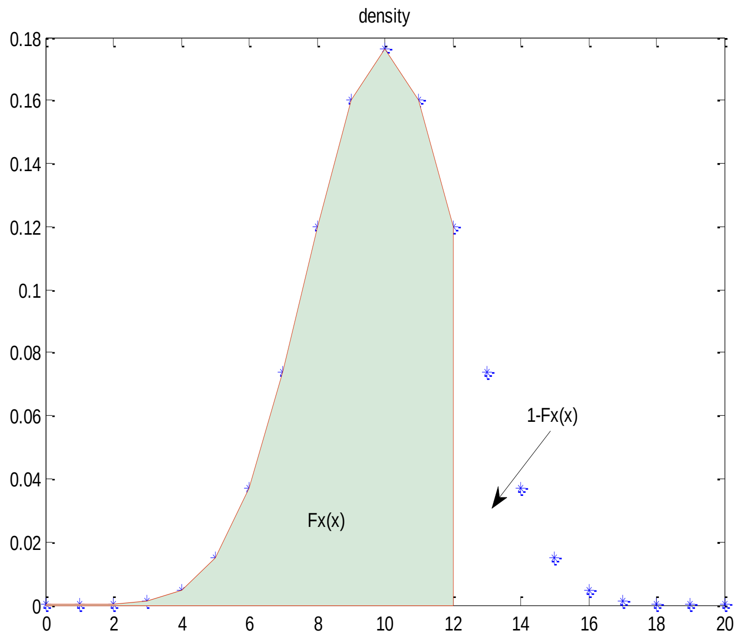
$$P(4 \leq X \leq 16) = P(X \leq 16) - P(X \leq 4)$$

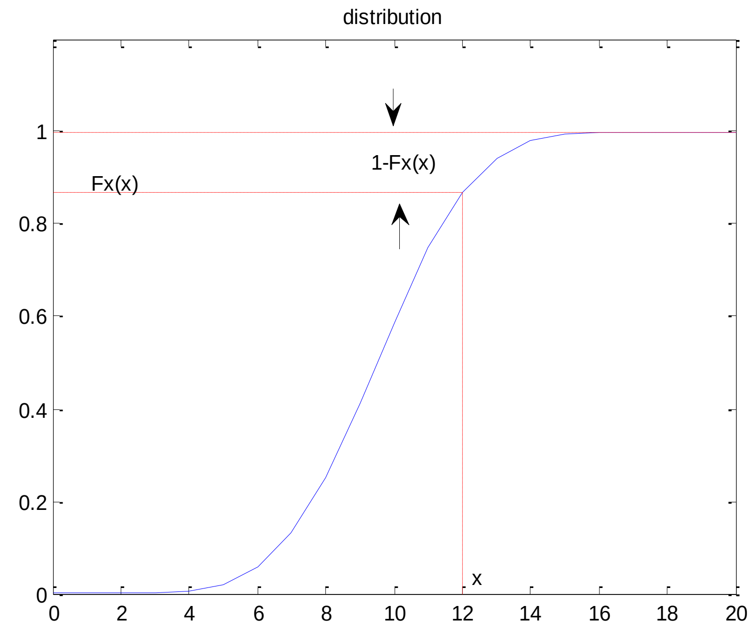
- In MATLAB γράφουμε

```
binocdf(16,10,.2)-binocdf(4,10,.2)
ans = 0.3228
```

Bernoulli Trials-Binomial Distribution

- Η αθροιστική συνάρτηση κατανομής γραφικά

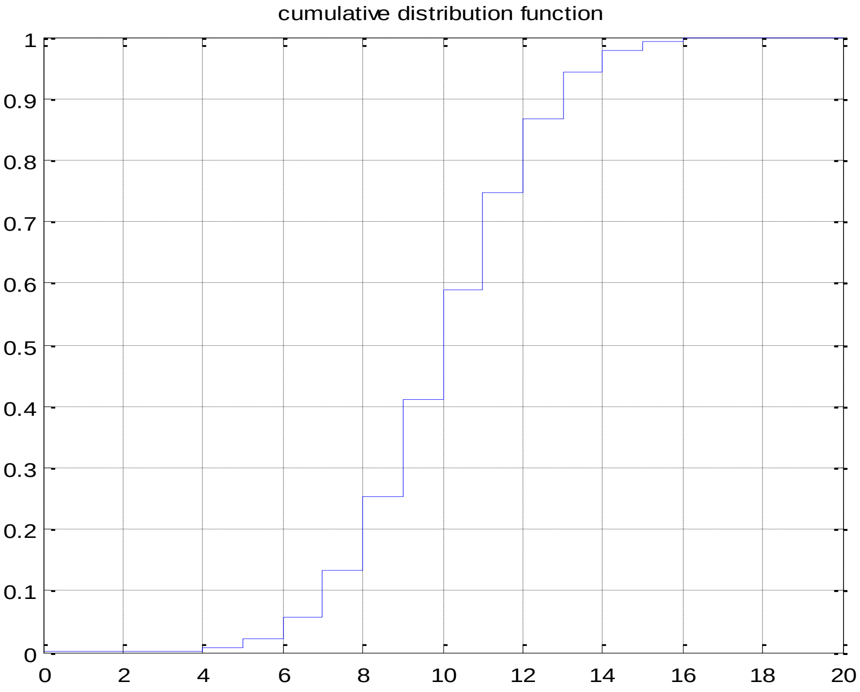




Bernoulli Trials-Binomial Distribution

- Το γράφημα της αθροιστικής συνάρτησης κατανομής φαίνεται πιο ενδιαφέρον αν χρησιμοποιήσουμε την stairs

```
k=0:20;
y=binocdf(k,20,0.5);
stairs(k,y)
grid on
```



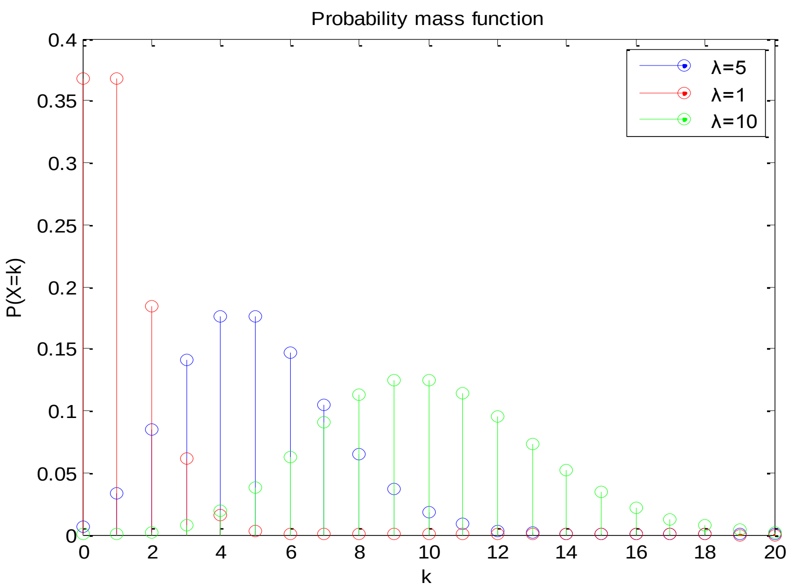
Poisson Distribution

- Παράδειγμα τυχαίων μεταβλητών
  - Πόσα hits έχει μια web σελίδα σε ορισμένο χρόνο
  - Εισερχόμενες τηλεφωνικές κλήσεις σε ένα αστυνομικό κέντρο
- Θέλουμε να υπολογίσουμε πιθανότητες για τέτοια γεγονότα
  - Ο αριθμός των κλήσεων σε ένα αστυνομικό κέντρο σε διάστημα 5 λεπτών ακολουθεί των κατανομή poisson με ρυθμό αφίξεων λ=2. Ποια είναι η πιθανότητα να έχουμε 3 τουλάχιστον κλήσεις στα επόμενα 5 λεπτά ?)
- Ρυθμός με τον οποίο καταμετρώνται τα γεγονότα είναι *λ*
- Probability density function $P(k) = \frac{l^{k}}{k!} e^{-l} \quad k = 0,1,..n$

Poisson Distribution

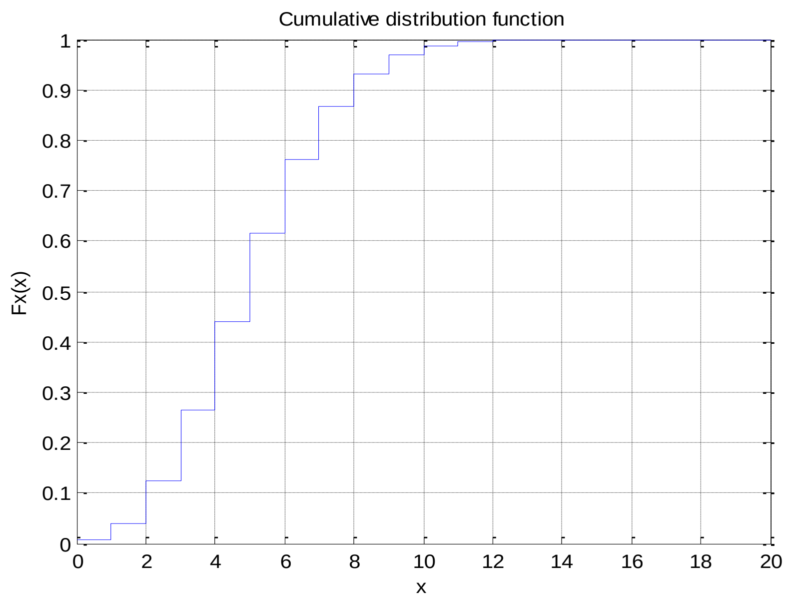
Y = poisspdf(X,lambda)

```
k=0:20;
y=poisspdf(k,5);
stem(k,y)
```



- P = poisscdf(X,lambda)

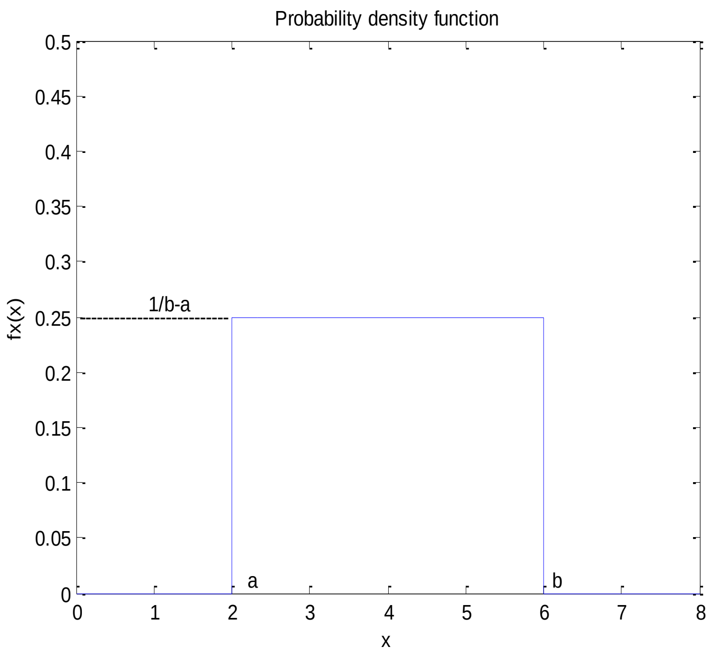
```
k=0:20;
y=poisscdf(k,5);
stairs(k,y)
```



Uniform Distribution

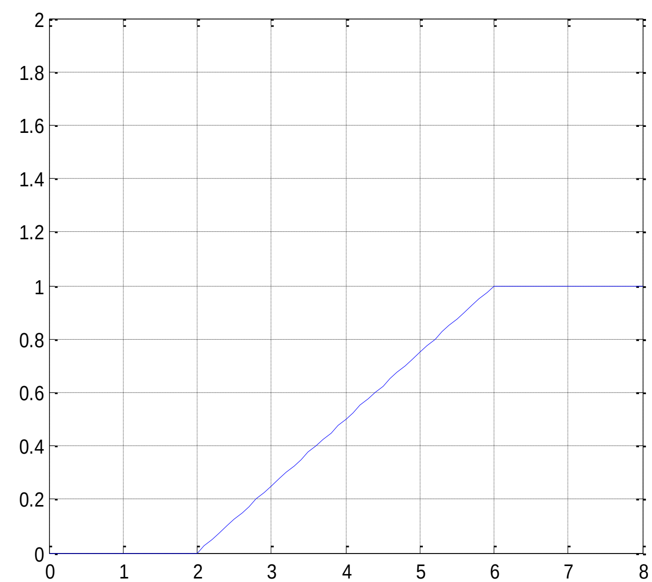
- Συνεχής τυχαία μεταβλητή που λαμβάνει τιμές μεταξύ *a και b με ίσες πιθανότητες*
- Probability density function

```
x=0:0.001:8;
y=unifpdf(x,2,6);
plot(x,y)
```



- Cumulative Distribution Function

```
x=0:0.1:8;
y=unifcdf(x,2,6);
plot(x,y)
axis([0 8 0 2])
```



Normal (Gaussian) Distribution

- Η κανονική κατανομή αναφέρεται σε συνεχείς μεταβλητές (σε αντίθεση με την poisson και την διωνυμική που είναι ασυνεχείς κατανομές)
- Αποτελεί μια συνεχή συνάρτηση πυκνότητας πιθανότητας και η μαθηματική έκφραση αυτής είναι:

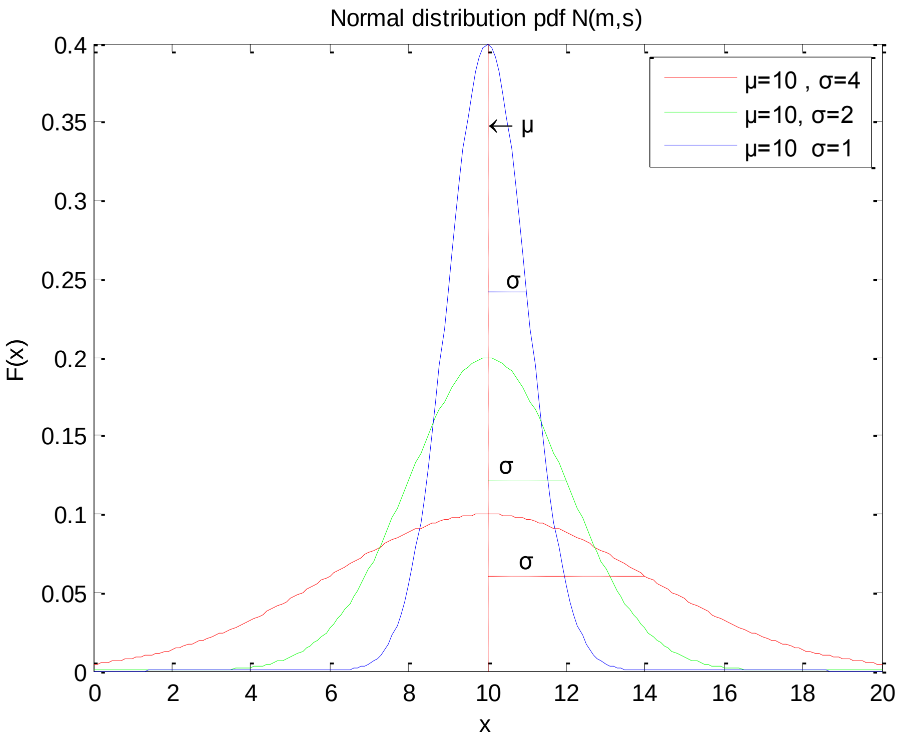
$$P_x(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2 \right] \qquad -\infty < x < \infty$$

$$\sigma^2 > 0 \qquad \kappa\alpha\iota \quad -\infty < \mu < \infty$$

Normal Distribution

- Στο matlab η εντολή είναι normpdf(X,m,s)
  Το matlab χρησιμοποιεί N(μ,σ)

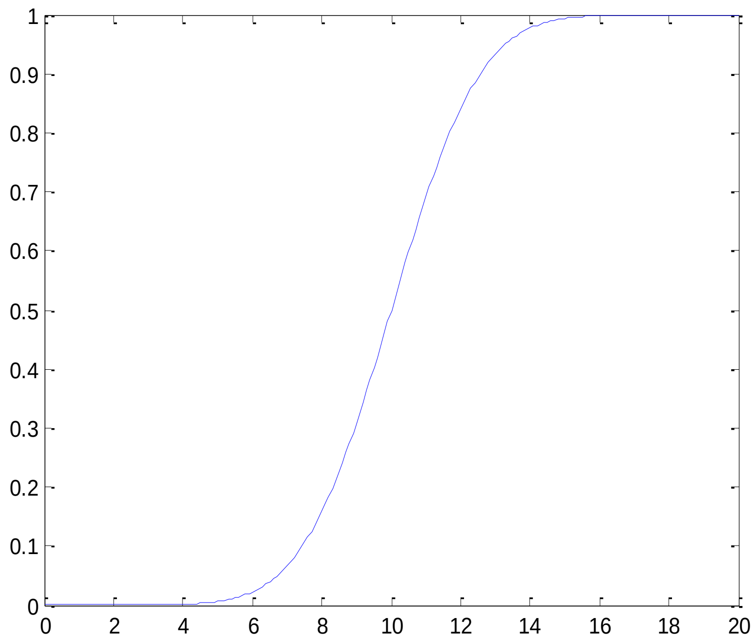
```
x=0:0.1:20;
y=normpdf(x,10,1);
plot(x,y);hold on
h=normpdf(x,10,4);
plot(x,h,'r')
k=normpdf(x,10,2.5);
plot(x,k,'g')
```



Normal Distribution

- Αθροιστική συνάρτηση κατανομής

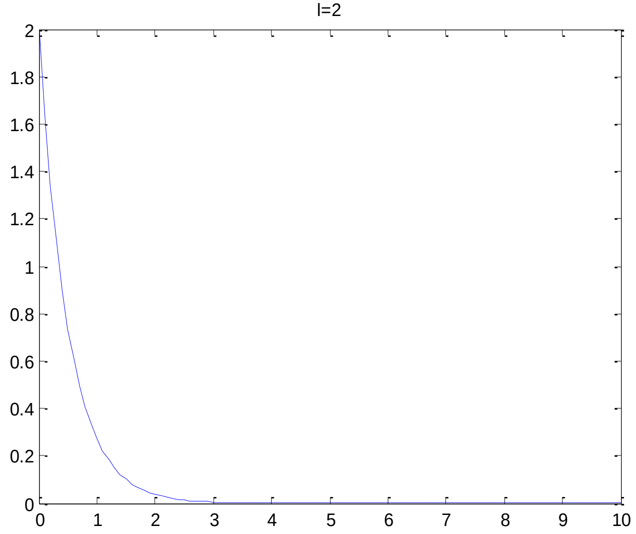
```
x=0:0.1:20;
y=normcdf(x,10,2);
plot(x,y)
```



Exponential Distribution

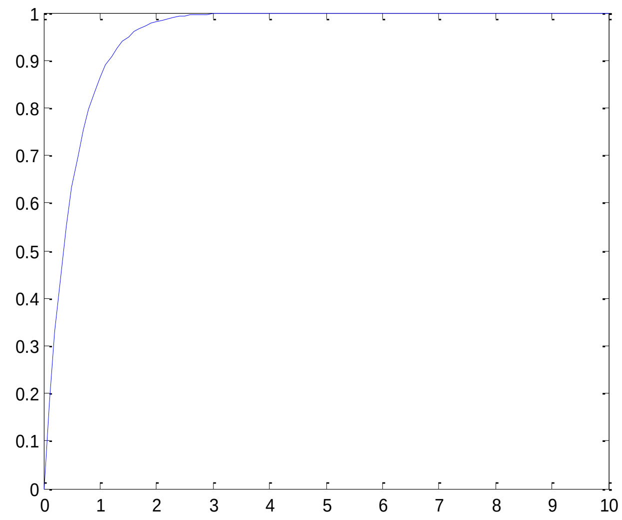
- Δίνεται από τον τύπο $f(x) = le^{-lx}, \quad m = \frac{1}{l}$
- Y = exppdf(X,mu)

```
x=0:0.1:10;
y=exppdf(x,1/2);
plot(x,y)
```



- P = expcdf(X,mu)

```
x=0:0.1:10;
y=exppdf(x,1/2);
plot(x,y)
```



Rayleigh

- Έστω ότι ρίχνουμε ένα ζάρι σε μία επιφάνεια την οποία θεωρούμε σαν xy-άξονα, και οι συντεταγμένες and suppose x και y του σημείου που έπεσε το ζάρι είναι ανεξάρτητες μεταξύ τους και έχουν μία κανονική κατανομή με μ=1 και σ= 1.

- Πως κατανέμεται η απόσταση από το σημείο (x,y) μέχρι το αρχικό σημείο ?

- Αν επιλέξουμε ένα ζευγάρι συντεταγμένων (x, y), τα οποία ανήκουν στην κανονική κατανομή με μ = 0 and σ= 1 , και υπολογίσουμε την απόσταση

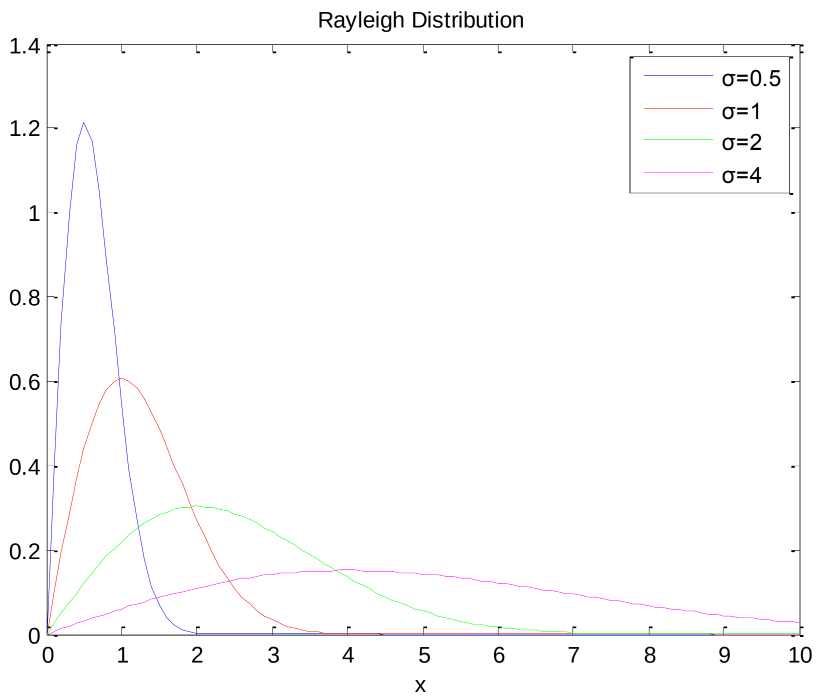
$$r = \sqrt{x^2 + y^2}$$

  από το σημείο (x,y) μέχρι το αρχικό σημείο για έναν αριθμό επαναλήψεων και παρουσιάσουμε το αποτέλεσμα σε ένα γράφημα ...

- Η r είναι μία τυχαία μεταβλητή με κατανομή Rayleigh

Rayleigh

```
x=0:0.1:10;
y=raylpdf(x,0.5);
y1=raylpdf(x,1);
y2=raylpdf(x,2);
y3=raylpdf(x,4);
plot(x,y);hold on
plot(x,y1,'r')
plot(x,y2,'g')
plot(x,y3,'m')
legend('σ=0.5','σ=1','σ=2','σ=4')
```



Poisson Approximation to Binomial

- Για εκείνες τις περιπτώσεις όπου έχουμε μεγάλο αριθμό επαναλήψεων ενός πειράματος, αλλά πολύ μικρό αριθμό πιθανών επιτυχιών (rare events) μπορεί να χρησιμοποιηθεί η κατανομή poisson για να προσεγγίζουμε την διωνυμική κατανομή

$$P(X) \cong \frac{e^{-np}(np)^{x}}{X!}$$

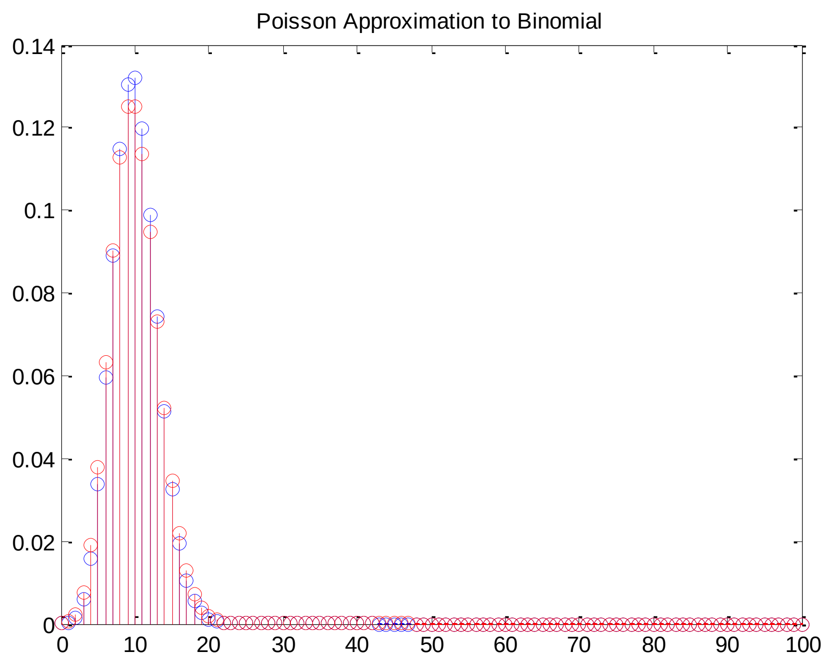
  - P(X)=πιθανότητα να έχουμε X επιτυχίες
  - n = αριθμός επαναλήψεων
  - p = πιθανότητα επιτυχίας
  - X = αριθμός των επιτυχιών στο δείγμα (1,2,…n)

- Ενδιαφέρον είναι ότι σε αυτή την περίπτωση έχουμε

$$\mu = \lambda = \mathrm{np}$$

Poisson Approximation to Binomial

```
n=100;
p=0.1;
l=10;
k=0:n;
y=binopdf(k,n,p);
h=poisspdf(k,l);
stem(k,y)
hold on
stem(k,h,'r')
```



Normal Approximation to Binomial

- Ας δούμε το ακόλουθο πρόβλημα
- Έχουμε ένα δείγμα 1500 Ελλήνων και θέλουμε να δούμε αν το ποσοστό των γυναικών στο δείγμα είναι ακριβές. Αν ξέρουμε ότι 12% των Ελλήνων είναι γυναίκες, περιμένουμε , το X ο αριθμός των γυναικών στο δείγμα να είναι 180.
- Ποια η πιθανότητα να έχουμε 170 ή λιγότερες γυναίκες στο δείγμα ?
- Αν εφαρμόσουμε τον τύπο της διωνυμικής κατανομής

$$P(X) = \binom{1500}{k}(0.12)^k(0.88)^{1500-k}, k = 0,1,\ldots,170$$

- Χρειαζόμαστε ένα πιο απλό τρόπο για τέτοιου είδους πειράματα

Normal Approximation to Binomial

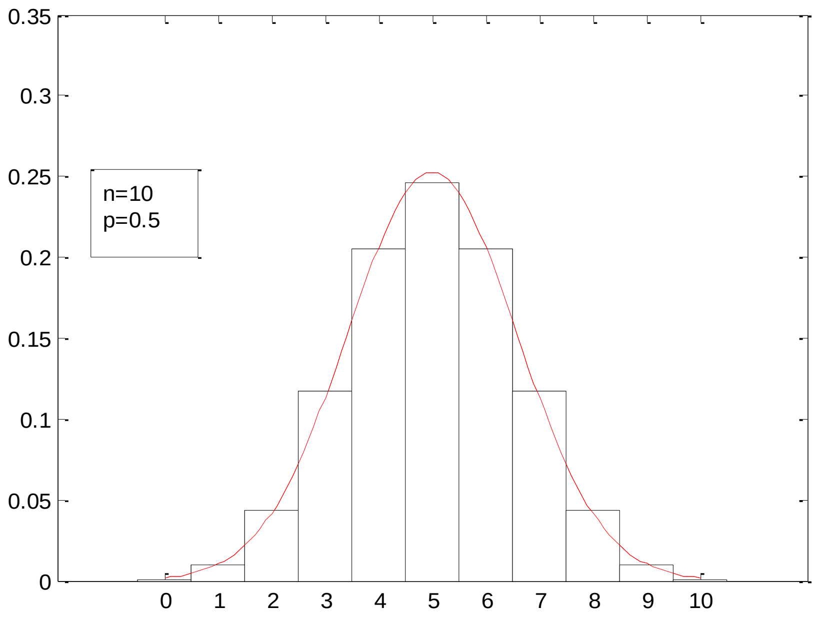
- Αποδεικνύεται ότι όσο το n μεγαλώνει η διωνυμική κατανομή «μοιάζει» με την κανονική κατανομή

- Ας το δούμε γραφικά

```
n=10;
k=0:n;
p=0.5;
y=binopdf(k,n,p);
x=0:0.1:n;
h=normpdf(x,n*p,sqrt(n*p*(1-p)));
bar(k,y,1,'w')
hold on;
plot(x,h,'r')
```

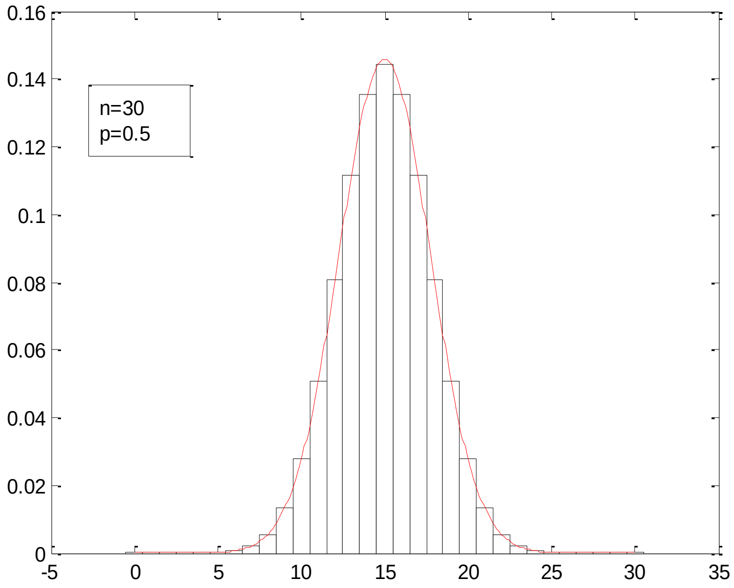
$$\mu = np$$

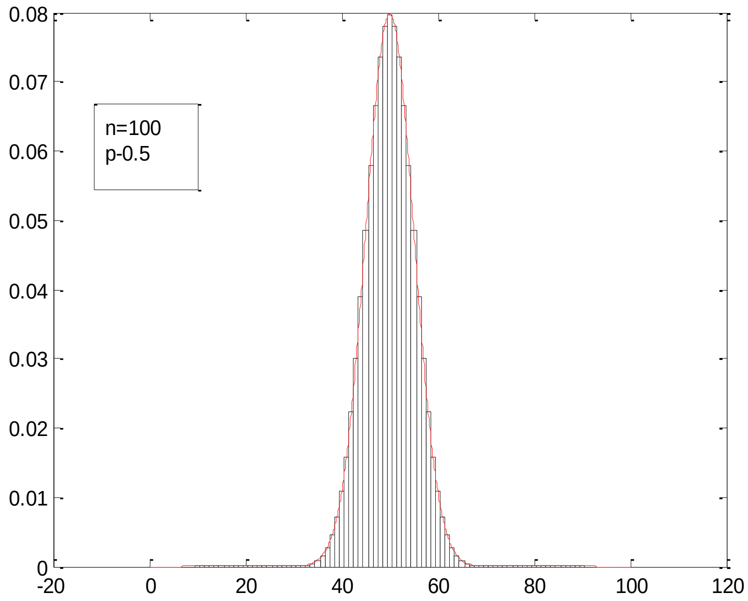
$$\sigma^2 = np(1-p)$$



Normal Approximation to Binomial

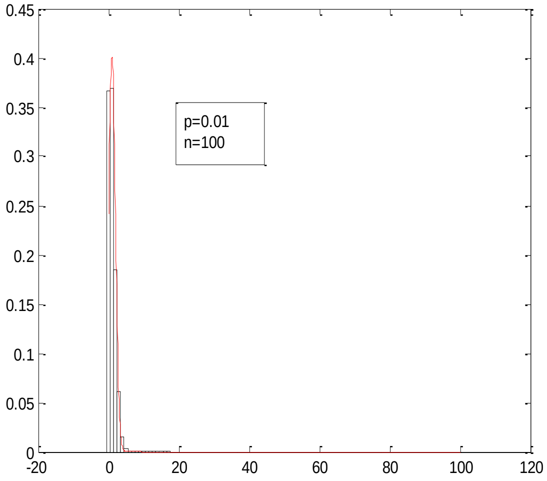
- Για διάφορα n

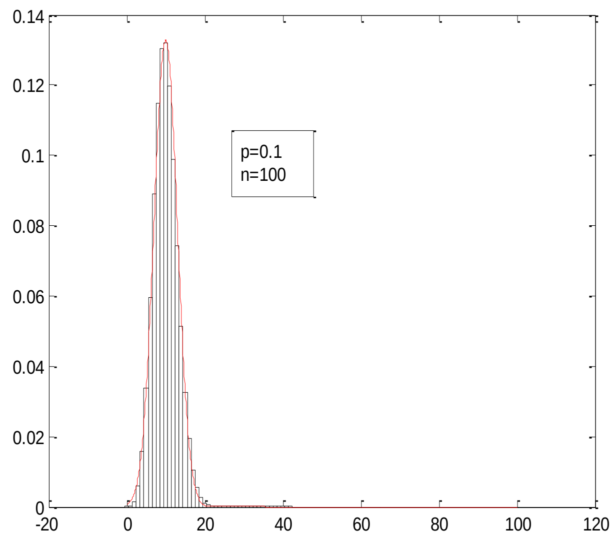


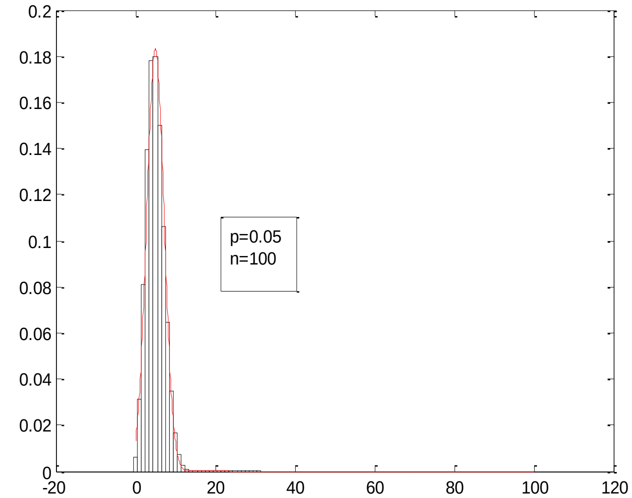


Normal Approximation to Binomial

- Ας δούμε τώρα τι γίνεται αν αλλάξουμε το p με n=100







Όσο πιο μακριά είναι το p από το 0.5 πρέπει η τιμή για το n να είναι μεγαλύτερη για να γίνει η προσέγγιση