Κατανομές διακριτών τυχαίων μεταβλητών

Ειδικές κατανομές διακριτών τ.μ.

- Με τον όρο **κατανομή** (***distribution***) ονομάζουμε ένα **μηχανισμό** που εφαρμόζεται πάνω σε μία μεταβλητή και περιγράφει την συμπεριφορά **τυχαιότητας των τιμών** της, δηλ. την **συχνότητα εμφάνισης** κάθε τιμής της μεταβλητής.

- Έτσι, κάθε τιμή της μεταβλητής ***x*** αποκτά μία **βαρύτητα** ίση με την **πιθανότητα** ***P(X=x)*** να εμφανιστεί (ή να την παρατηρήσουμε).

- Η κατανομή μιας τυχαίας μεταβλητής **περιγράφεται** πλήρως από:

  - την **συνάρτηση πυκνότητας** (ή **μάζας**) **πιθανότητας,** ***f(x)*** και
  - την **αθροιστική συνάρτηση κατανομής** **πιθανότητας,** ***F(x)***

  που ορίζονται πάνω στο πεδίο τιμών της μεταβλητής.

- Η γνώση (ο υπολογισμός) μιας εκ των δύο συναρτήσεων αρκεί για να καθοριστεί **πλήρως** η τυχαία μεταβλητή.

Ειδικές κατανομές διακριτών τ.μ.

- Γνωστές κατανομές **διακριτών** τυχαίων μεταβλητών που **περιγράφουν** μεταβλητές του «φυσικού» κόσμου.

- **Ομοιόμορφη** (***Uniform***) (***U***)
- ***Bernoulli***
- **Διωνυμική** (***Binomial***) (***Bin***)
- **Αρνητική Διωνυμική** (***Negative Binomial***) (***Nb***)
- **Γεωμετρική** (***Geometrical***) (***G***)
- ***Poisson*** (***P***)
- ….

(1) Ομοιόμορφη (Uniform) κατανομή

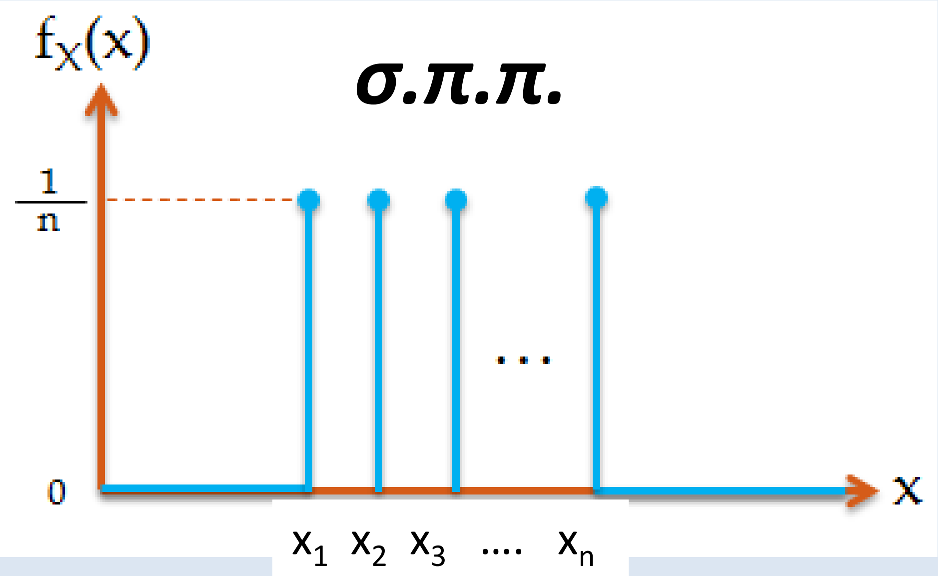
- **Σύνολο τιμών** της μεταβλητής:

$$\Omega_X = \{x_1, x_2, \ldots, x_n\} \quad ή \quad \Omega_X = \{1, 2, \ldots, n\}$$

- **Υπόθεση**: όλες οι τιμές της τυχαίας μεταβλητής είναι ισοπίθανες, δηλ. έχουν ίσο μέτρο πιθανότητας (π.χ. ρίψη νομίσματος ή ζαριού).

- **Συμβολισμός:** $\boxed{X \sim U(n)}$

(1) Ομοιόμορφη (**Uniform**) κατανομή (συν.)



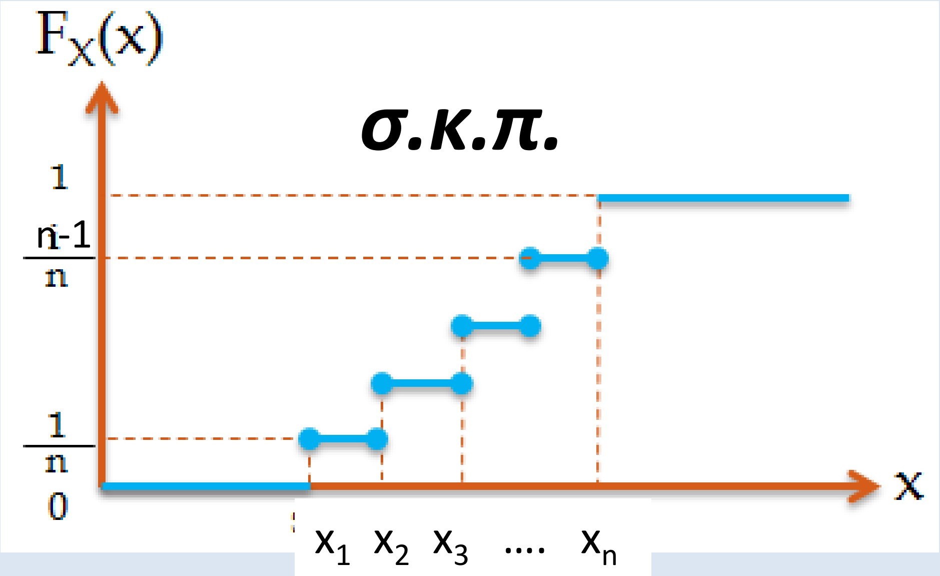
$$\sum_{i=1}^{n} f(x_i) = 1$$

- Συνάρτηση **πυκνότητας** πιθανότητας (**σ.π.π.**)

$$\boldsymbol{f(x_i)} = P(X = x_i) = \boldsymbol{\frac{1}{n}}, \quad \forall\ i = 1, \ldots, n$$

Ομοιόμορφη κατανομή

$$f\left(x_i\right)=\frac{1}{n}$$



- **Αθροιστική** συνάρτηση **κατανομής** πιθανότητας (**σ.κ.π.**)

$$F\left(x\right)=P\left(X\leq x\right)=\sum_{x_i\leq x} f\left(x_i\right)=\sum_{x_i\leq x}\frac{1}{n}=\frac{\text{πλήθος}\,\tau\iota\mu\omega\nu\,\text{στο}\,(-\infty,x]}{n}$$

- Για τιμές της μεταβλητής $\Omega_X = \{1, 2, \ldots, n\}$

- Αθροιστική συνάρτηση κατανομής πιθανότητας (σ.κ.π.)

$$\boldsymbol{F}(\boldsymbol{x}) = P(X \le x) = \sum_{x_i \le x} f(x_i) = \begin{cases} 0 & x < 1 \\ \dfrac{\lfloor x \rfloor}{n} & 1 \le x < n \\ 1 & x \ge n \end{cases}$$

$$f(x_i) = \frac{1}{n}$$

*ακέραιο μέρος του **x***

<u>**(2) *Bernoulli*** κατανομή</u> , *X~Bernoulli(ρ)*

- ***Bernoulli πείραμα ή δοκιμή***: τυχαίο πείραμα το αποτέλεσμα του οποίου είναι **δυαδικό**:
  - **επιτυχία** (**1**) με πιθανότητα ***P(X=1)=ρ*** και
  - **αποτυχία** (**0**) με πιθανότητα ***P(X=0)=1-ρ***.

- Σχετικό με μοντελοποίηση γεγονότων που είτε **συμβαίνουν** με *πιθανότητα επιτυχίας* ***ρ*** **,** είτε **δεν συμβαίνουν** με *πιθανότητα αποτυχίας* ***1-ρ***

- Παραδείγματα
  - Ρίψη νομίσματος (***κορώνα ή γράμματα***)
  - Κατάσταση καιρού (***ηλιοφάνεια ή όχι***)
  - Κατάσταση μηνύματος (***λήψη ή όχι***)
  - Αποτέλεσμα κάποιας εξέτασης (***επιτυχία ή αποτυχία***)
  - Επίτευξη στόχου (***επιτυχία ή αποτυχία***)
  - ….

<u>(2) **Bernoulli** κατανομή</u> , *X~Bernoulli(ρ)*

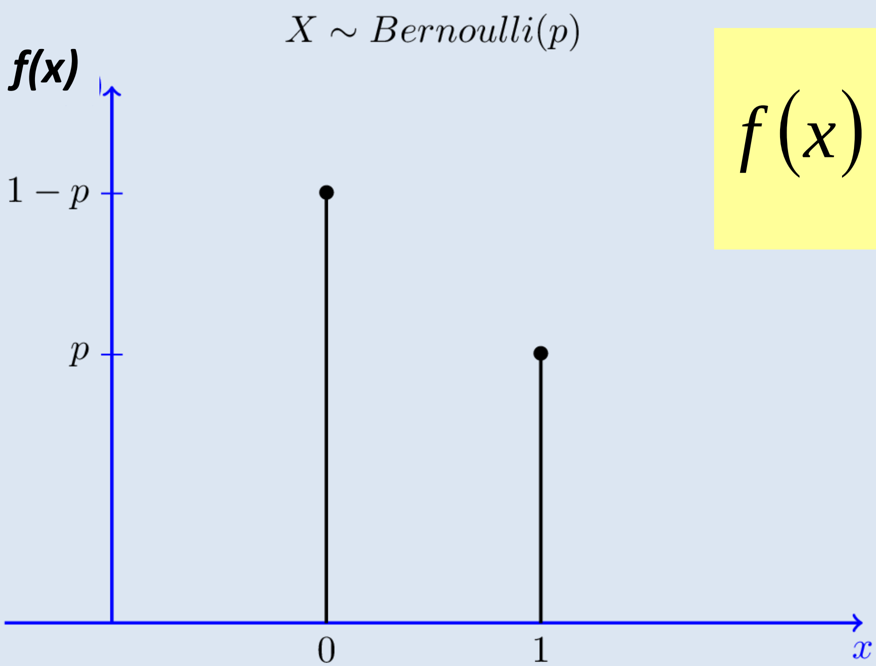
- ***Bernoulli πείραμα ή δοκιμή***: τυχαίο πείραμα το αποτέλεσμα του οποίου είναι δυαδικό: επιτυχία (**1**) με πιθανότητα ***P(X=1) = ρ*** και αποτυχία (**0**) με πιθανότητα ***P(X=0) = 1-ρ***.

✓ Σύνολο τιμών της μεταβλητής $\boldsymbol{\Omega_X = \{0,1\}}$

✓ **Συνάρτηση πυκνότητας (ή μάζας)** πιθανότητας **(σ.π.π.)**

$$f(x) = \begin{cases} \rho & x = 1 \\ 1 - \rho & x = 0 \end{cases} = \rho^x (1-\rho)^{1-x} \qquad x=\{0, 1\}$$

Η συνάρτηση πυκνότητας (ή μάζας) πιθανότητας f(x) της Bernoulli μεταβλητής



$$f(x)=\begin{cases}\rho & x=1\\ 1-\rho & x=0\end{cases}=\rho^{x}(1-\rho)^{1-x}$$

Bernoulli κατανομή , *X~Bernoulli(ρ)*

- ***Bernoulli πείραμα ή δοκιμή***: τυχαίο πείραμα το αποτέλεσμα του οποίου είναι δυαδικό: επιτυχία (**1**) με πιθανότητα ***P(X=1) = ρ*** και αποτυχία (**0**) με πιθανότητα ***P(X=0) = 1-ρ***.

- Σύνολο τιμών της μεταβλητής $\Omega_X = \{0,1\}$

✓ **σ.π.π.** $$f(x) = \begin{cases} \rho & x = 1 \\ 1-\rho & x = 0 \end{cases} = \rho^x (1-\rho)^{1-x}$$

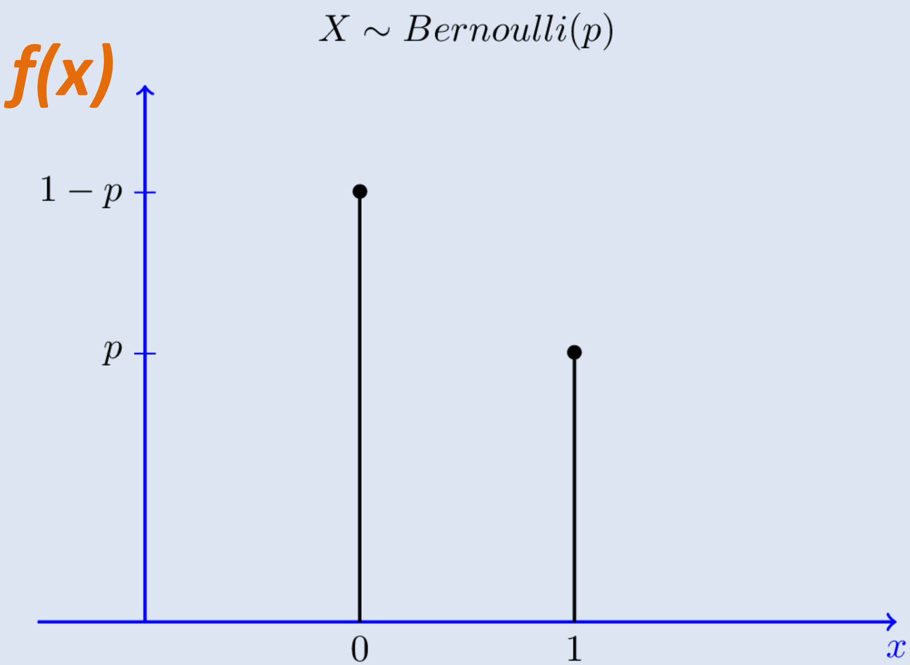
Η αθροιστική συνάρτηση κατανομής πιθανότητας

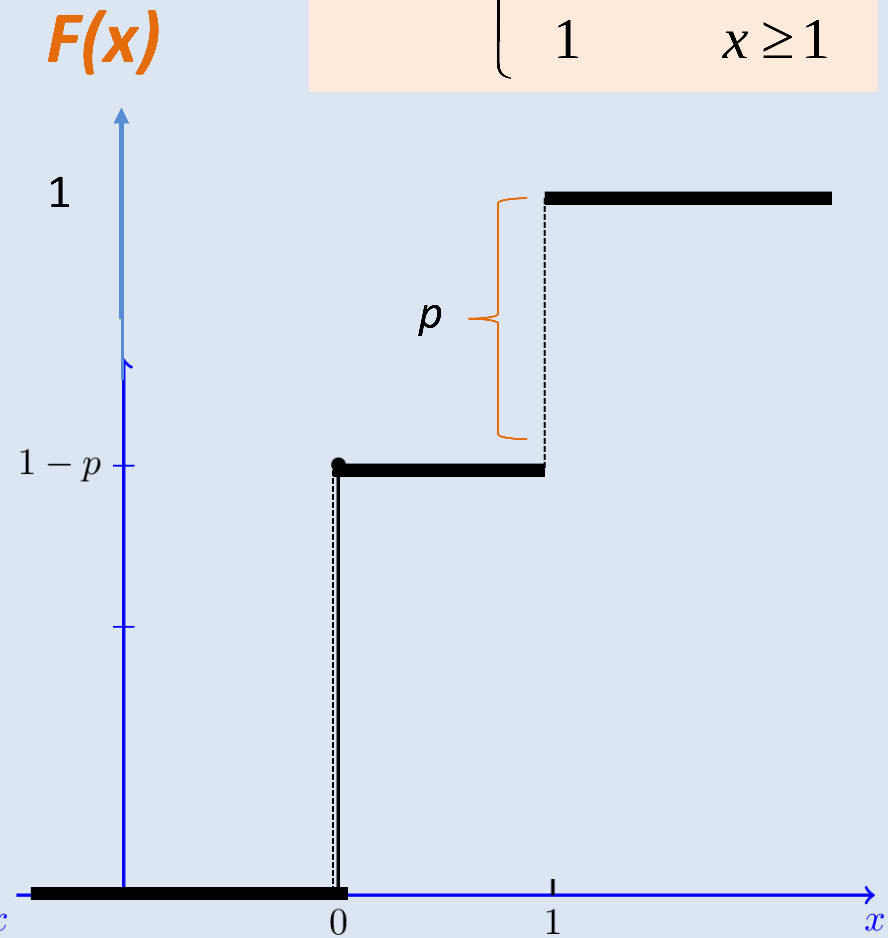
✓ **σ.κ.π.** $$F(x) = P(X \le x) = \begin{cases} 0 & x < 0 \\ 1-\rho & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases}$$

H Bernoulli κατανομή

$$f(x) = \begin{cases} \rho & x = 1 \\ 1-\rho & x = 0 \end{cases} = \rho^{x}(1-\rho)^{1-x}$$

$$F(x) = \begin{cases} 0 & x < 0 \\ 1-\rho & 0 \le x < 1 \\ 1 & x \ge 1 \end{cases}$$





(3) Διωνυμική (*Binomial*) κατανομή *X ~ Bin(n,ρ)*

- Εκτελούμε ***n*** **διαδοχικά ανεξάρτητα** *Bernoulli* πειράματα, δηλ. πειράματα όπου το αποτέλεσμα είναι **επιτυχία** (με πιθανότητα ***ρ***) ή **αποτυχία** (με πιθανότητα ***1-ρ***).

- **Τυχαία μεταβλητή** (***τ.μ.***): μετράει τη **συχνότητα εμφάνισης της επιτυχίας**, δηλ. το **πλήθος** των επιτυχιών στα ***n*** πειράματα.

(δεν μας ενδιαφέρει η διάταξη των αποτελεσμάτων)

(**3**) Διωνυμική (***Binomial***) κατανομή ***X ~ Bin(n,ρ)***

- **«Πλήθος επιτυχιών** σε ***n*** ανεξάρτητες *Bernoulli* δοκιμές με πιθανότητα επιτυχίας ***ρ*** ανά δοκιμή»

Συμβολίζουμε ως τ.μ. ***X ~ Bin(n,ρ)***

- Σύνολο τιμών της μεταβλητής $\boldsymbol{\Omega_X = \{0,1, \dots, n\}}$
- ***X*** **=** ***0:*** **καμία** επιτυχία σε ***n*** πειράματα
- ***X*** **=** ***n:*** **όλα** τα πειράματα είναι επιτυχημένα (τόσες επιτυχίες όσες και τα ***n*** πειράματα)

Υπολογισμός της σ.π.π. της διωνυμικής κατανομής *X~Bin(n,ρ)*

f(x)=P(X=x)=P('x επιτυχίες σε n ανεξ. πειράματα')

Έστω μία ακολουθία *010010…10* από *n* πειράματα με *x* επιτυχίες (*1*) και *n-x* αποτυχίες (*0*). Τότε, η πιθανότητα αυτής της συγκεκριμένης ακολουθίας είναι:

$$P(\text{'συγκεκριμενη ακολουθια'}) = \rho^x (1-\rho)^{n-x}$$

Υπάρχουν $\binom{n}{x}$ πλήθος τέτοιων ακολουθιών με ***x*** επιτυχίες.

Έτσι:

$$f(x) = P(X = x) = \binom{n}{x} \rho^x (1-\rho)^{n-x}$$

Υπολογισμός της σ.π.π. της διωνυμικής κατανομής *X~Bin(n,ρ)*

f(x)=P(X=x)=P(‘x επιτυχίες σε n ανεξ. πειράματα’)

Επαλήθευση

$$f(x) = P(X = x) = \binom{n}{x} \rho^x (1-\rho)^{n-x}$$

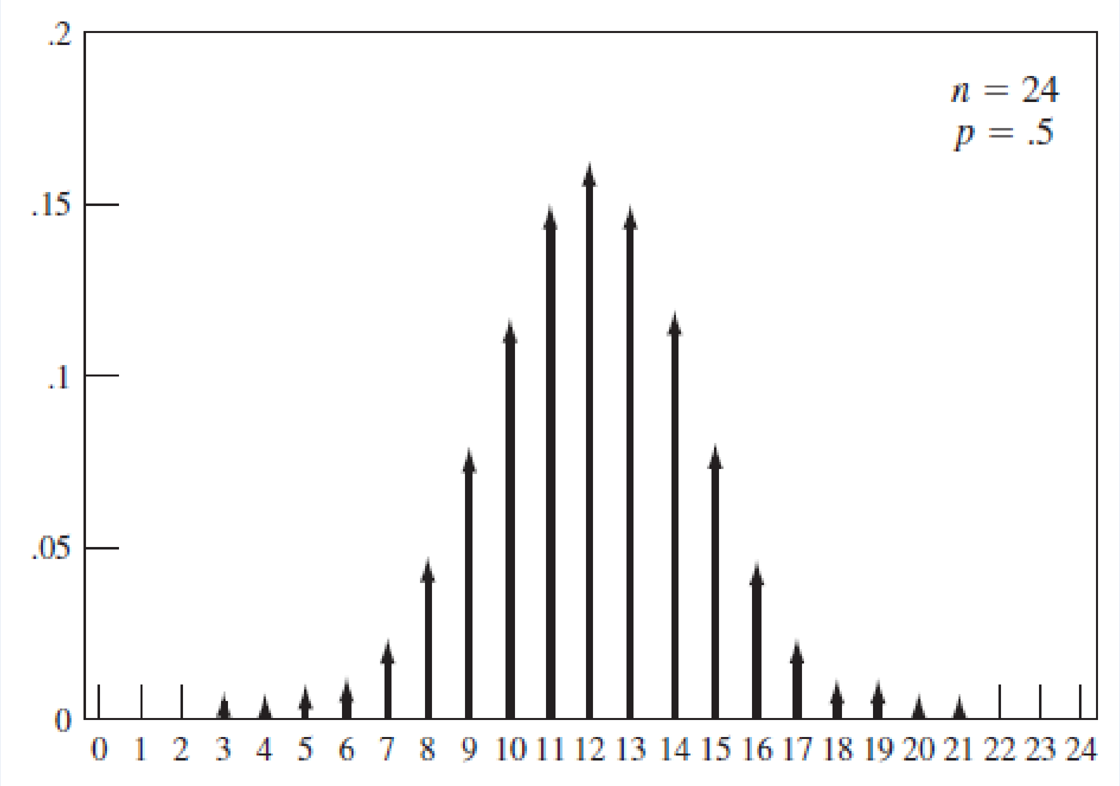
$$\sum_{x=0}^{n} f(x) = \sum_{x=0}^{n} \binom{n}{x} \rho^x (1-\rho)^{n-x} = (p + 1 - p)^n = 1$$

$$f(x)=\binom{n}{x}\rho^{x}(1-\rho)^{n-x}$$

Παρατηρήσεις

- *Αν **ρ=0.5** τότε η σ.π.π. είναι **συμμετρική** καθώς*

$$f(x)=f(n-x)=\binom{n}{x}2^{-n}$$



- ***Αναδρομικός τύπος***:

$$f(x)=\binom{n}{x}\rho^{x}(1-\rho)^{n-x}$$

- Εύρεση αναδρομικού τύπου ***f(x+1) = a * f(x)***

$$f(x+1)=\binom{n}{x+1}\rho^{x+1}(1-\rho)^{n-x-1}=$$

$$=\frac{n!}{(x+1)!(n-x-1)!}\rho\rho^{x}\frac{(1-\rho)^{n-x}}{(1-\rho)}=$$

$$=\frac{(n-x)\rho}{(x+1)(1-\rho)}\left[\frac{n!}{x!(n-x)!}\rho^{x}(1-\rho)^{n-x}\right]=$$

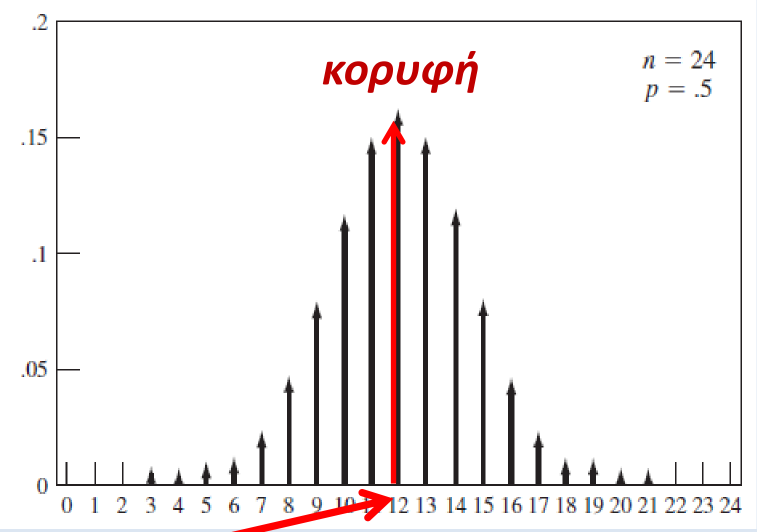
$$\boxed{\frac{(n-x)\rho}{(x+1)(1-\rho)}}f(x)=a(x)f(x)$$

➢ ***Αναδρομικός τύπος***:

$$\boldsymbol{f}(\boldsymbol{x}+\mathbf{1}) = \frac{(n-x)\rho}{(x+1)(1-\rho)} f(x) = \boldsymbol{a}(\boldsymbol{x}) f(x)$$

- Έτσι, αν ο συντελεστής ***α(x) ≥ 1 => x+1 ≤ (n+1) ρ***

τότε f(x+1)>f(x) ↑



- *Η **x*=[(n+1)ρ]** είναι η* **πιθανότερη τιμή (κορυφή) -** ***mode***

Ασκήσεις

1. Αν παίζεις ένα παιχνίδι με έναν ισοδύναμο αντίπαλο, ποιο ενδεχόμενο έχει μεγαλύτερη πιθανότητα: Να κερδίσεις τρεις από τις τέσσερις παρτίδες, ή να κερδίσεις έξη από τις οκτώ παρτίδες; Αιτιολογήστε την απάντησή σας.

2. Η πιθανότητα να αντέξει ένα μηχάνημα στην καταπόνηση είναι 0.8. Ποια είναι η πιθανότητα τουλάχιστον 2 μηχανήματα ενός συνόλου 8 μηχανημάτων να αντέξουν στην καταπόνηση.

3. Αν η γέννηση ενός κοριτσιού και ενός αγοριού είναι ίσης πιθανότητας, πόσα παιδιά πρέπει να αποκτήσει μια οικογένεια ώστε να έχει τουλάχιστον 1 αγόρι και 1 κορίτσι με πιθανότητα τουλάχιστον 90%?

(4) **Αρνητική Διωνυμική** (***Negative binomial***) ή **Pascal** κατανομή

- Έστω εκτέλεση διαδοχικών **Bernoulli** πειραμάτων με πιθανότητα επιτυχίας ρ (και αποτυχίας 1-ρ).

- Ορίζουμε **τυχαία μεταβλητή** ***X***: **πλήθος πειραμάτων** που απαιτούνται ώστε να εμφανιστούν ***n*** επιτυχίες.

συμβολίζουμε τ.μ. ***X ~ Nb(n,ρ)***

- **Σύνολο τιμών** της μεταβλητής $\Omega_X = \{n, n+1, n+2, \ldots\}$ (*απαιτούνται τουλάχιστον* ***n*** *προσπάθειες*)

(4) Αρνητική Διωνυμική κατανομή (συν.)

- **Υπολογισμός** της συνάρτησης πυκνότητας (σ.π.π.), ***f(x)*** :

f(x) = P(X = x) = P(«n επιτυχίες σε x προσπάθειες») =

*= P("ακριβώς **n-1 επιτυχίες** στα πρώτα **x-1** πειράματα και επιτυχία στο τελευταίο πείραμα") =*

=P("n-1 επιτυχίες σε x-1")P("επιτυχία στο τελευταίο") =

$$= \binom{x-1}{n-1} \rho^{n-1} (1-\rho)^{(x-1)-(n-1)} \times \rho$$

- *Αρα σ.π.π.* $f(x) = \binom{x-1}{n-1} \rho^{n} (1-\rho)^{x-n} \;,\; x \geq n$

(4) Αρνητική Διωνυμική κατανομή (συν.)

- *σ.π.π.* $$f(x) = \binom{x-1}{n-1} \rho^n (1-\rho)^{x-n} \quad , \quad x \geq n$$

- **Αναδρομικός τύπος:** *Παρόμοια με την Διωνυμική κατανομή*

$$f(x+1) = \frac{x(1-\rho)}{x-n+1} f(x)$$

α(x)

(5) **Γεωμετρική** (**Geometrical**) κατανομή

- Μετρά το **πλήθος** ***Bernoulli*** **δοκιμών** μέχρι την **πρώτη επιτυχία**, δηλ. ακολουθία παρατηρήσεων {**0 0 … 0 1**}

συμβολίζουμε τ.μ. $\boldsymbol{X \sim G(\rho)}$

- Ισχύει ότι $\boldsymbol{G(\rho) \equiv Nb(n=1,\rho)}$

δηλ. Αρνητική Διωνυμική με n=1

- **Σύνολο τιμών** της μεταβλητής $\boldsymbol{\Omega_X = \{1, 2, \ldots\}}$

(5) **Γεωμετρική (Geometrical)** κατανομή

Η Γεωμετρική κατανομή χρησιμοποιείται για:

- **Χρόνος αναμονής,** δηλ. χρόνος που απαιτείται μέχρι την εμφάνιση ενός ενδεχομένου, ή πόσο χρόνο θα πρέπει να περιμένουμε μέχρι να εμφανιστεί ένα φαινόμενο.

- **Διάρκεια ζωής**, δηλ. χρόνος που υπολείπεται μέχρι να σταματήσει η λειτουργία ενός αντικειμένου που δουλεύει χωρίς διακοπή.

(5) Γεωμετρική (**Geometrical**) κατανομή (συν.)

1. Υπολογισμός της συνάρτησης πυκνότητας (**σ.π.π.**)

- **Συγκεκριμένη** ακολουθία γεγονότων: $\underbrace{00\cdots0}_{x-1 \text{ fails}}1$

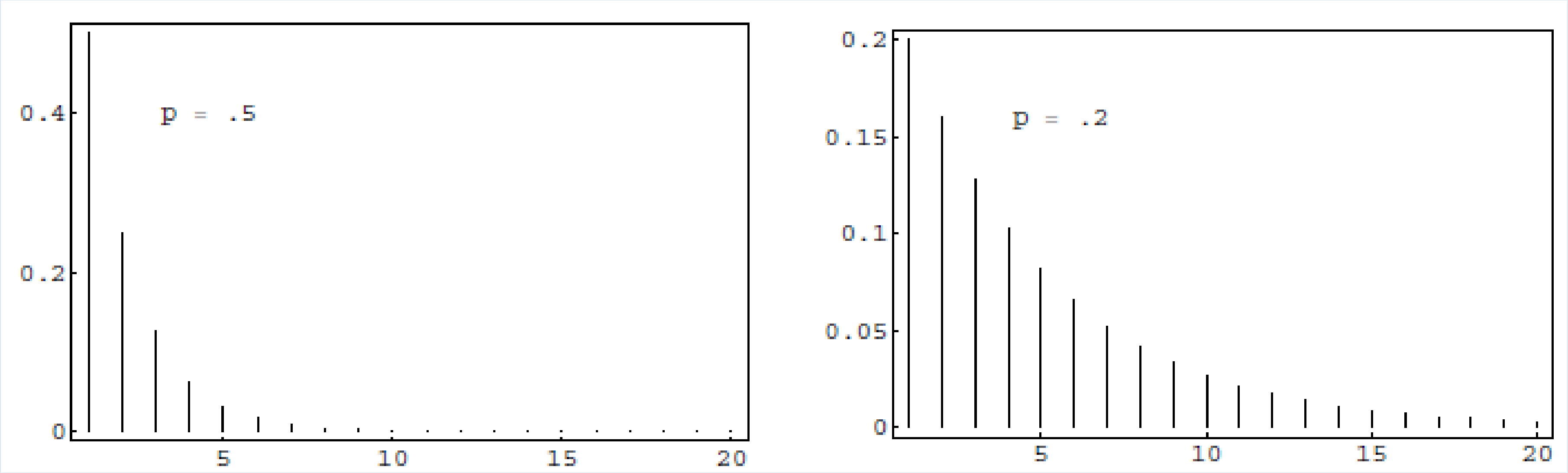
$$f(x) = P(X = x) = \rho(1-\rho)^{x-1} \quad x = \{1,2,\ldots\}$$

(5) Γεωμετρική (**Geometrical**) κατανομή (συν.)

1. Συνάρτηση πυκνότητας (**σ.π.π.**)

$$f(x) = P(X = x) = \rho(1-\rho)^{x-1} \quad x = \{1, 2, \ldots\}$$

Μέγιστη τιμή (κορυφή) της κατανομής για *x=1 : f(1)=ρ*



(5) Γεωμετρική (**Geometrical**) κατανομή (συν.)

1. Συνάρτηση πυκνότητας (**σ.π.π.**)

$$f(x) = P(X = x) = \rho(1-\rho)^{x-1} \quad x = \{1,2,\ldots\}$$

- Επαλήθευση:

$$\sum_{x=1}^{\infty} f(x) = \sum_{x=1}^{\infty} \rho(1-\rho)^{x-1} = \rho\sum_{k=0}^{\infty}(1-\rho)^k = \rho\frac{1}{\rho} = 1$$

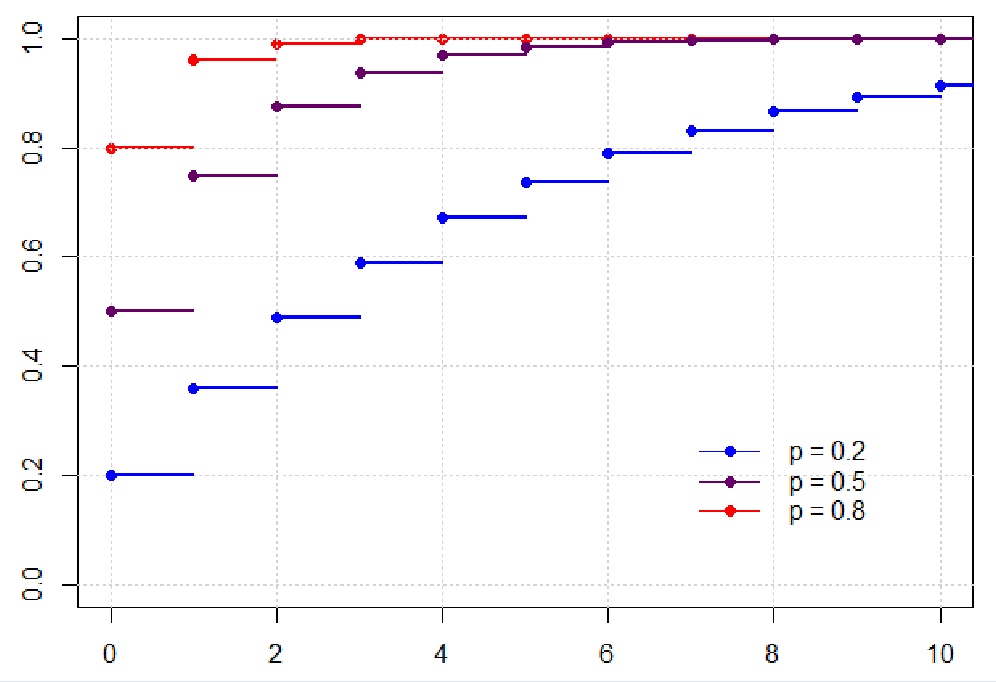
$$\text{Ισχυει οτι} \sum_{k=0}^{\infty} a^k = \frac{1}{1-a} \ , \ |a| < 1$$

2.Υπολογισμός της **αθροιστικής συνάρτησης κατανομής F(x)**

$$F(x)=P(X\le x)=\sum_{k=1}^{x} f(k)=\sum_{k=1}^{x}\rho(1-\rho)^{k-1}=\rho\sum_{k'=0}^{x-1}(1-\rho)^{k'}=$$

$$=\rho\frac{1-(1-\rho)^{x}}{1-(1-\rho)}=1-(1-\rho)^{x}$$

Ισχυει οτι $\sum_{k=0}^{n} a^{k}=\frac{1-a^{n+1}}{1-a}\ ,\ |a|<1$



3. Υπολογισμός **χρόνου αναμονής**

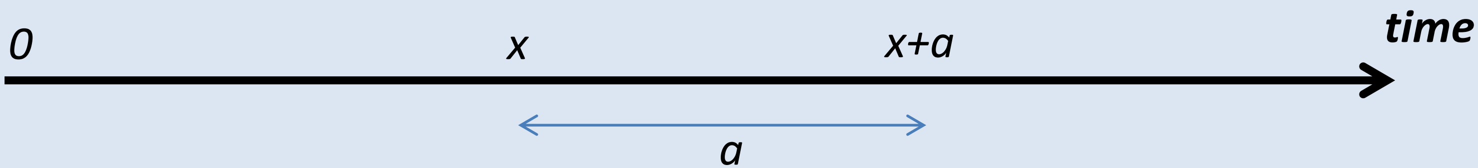
$$F(x) = P(X \leq x) = 1 - (1 - \rho)^x$$

- Πιθανότητα να **περιμένουμε** **τουλάχιστον *x* χρονικές στιγμές** μέχρι την εμφάνιση ενός ενδεχομένου

$$P(X > x) = 1 - P(X \leq x) =$$
$$= 1 - F(x) = (1 - \rho)^x$$

χρόνος αναμονής

4. Ιδιότητα της απώλειας μνήμης
(*memoryless property*)

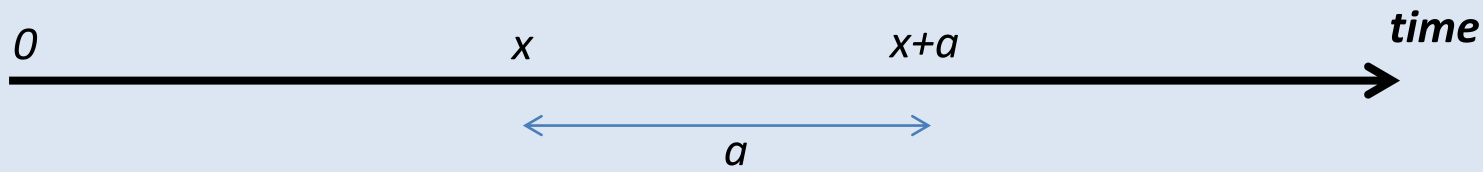


- Πιθανότητα να περιμένουμε **τουλάχιστον επιπλέον *α* χρονικές στιγμές** έως ότου εμφανιστεί κάποιο φαινόμενο, εάν έχουμε **ήδη περιμένει *x* χρονικές στιγμές** δίχως να έχει εμφανιστεί.

$$P\left(X > x+a \mid X > x\right) = \frac{P\left(X > x+a \ \cap \ X > x\right)}{P\left(X > x\right)} = \frac{P\left(X > x+a\right)}{P\left(X > x\right)} =$$

$$\boxed{P\left(X > x\right) = 1 - F\left(x\right) = \left(1-\rho\right)^{x}}$$

4. Ιδιότητα της απώλειας μνήμης (*memoryless property*)

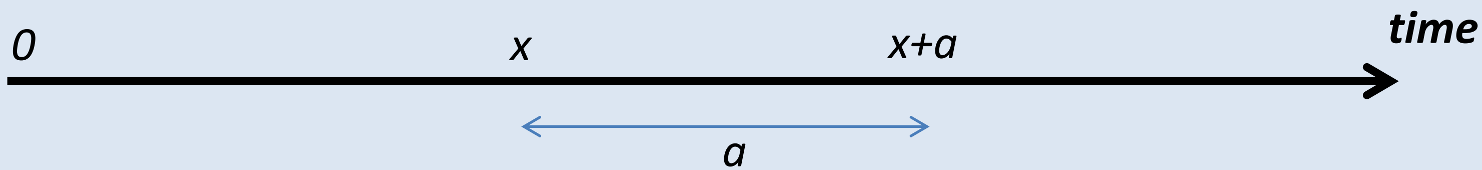


- Πιθανότητα να περιμένουμε **τουλάχιστον επιπλέον *α* χρονικές στιγμές** έως ότου εμφανιστεί κάποιο φαινόμενο, εάν έχουμε **ήδη περιμένει *x* χρονικές στιγμές** δίχως να έχει εμφανιστεί.

$$P(X > x + a \mid X > x) = \frac{P(X > x + a \cap X > x)}{P(X > x)} = \frac{P(X > x + a)}{P(X > x)} =$$

$$= \frac{(1-\rho)^{x+a}}{(1-\rho)^{x}} = (1-\rho)^{a} = P(X > a)$$

4. Ιδιότητα της απώλειας μνήμης

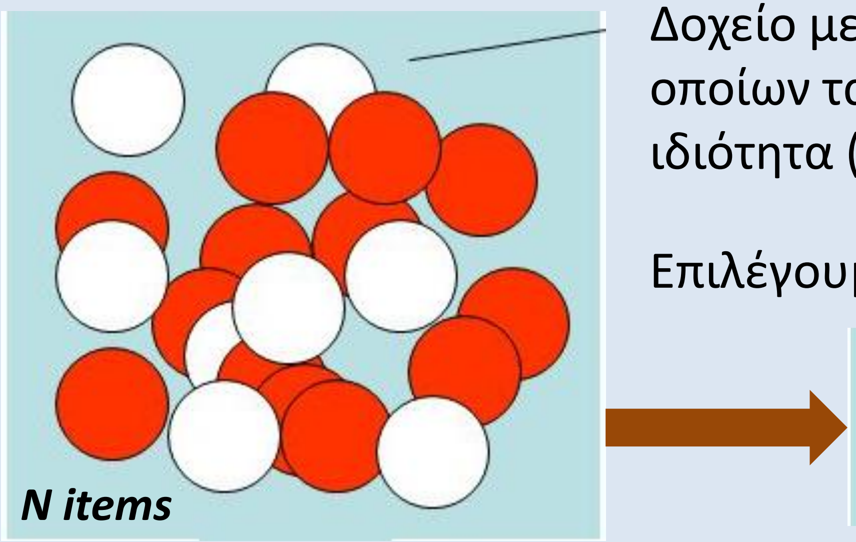


$$P\left(X > x+a \mid X > x\right) = P(X > a)$$

- **Συμπέρασμα:** Η πιθανότητα να περιμένουμε **επιπλέον α** χρονικές στιγμές είναι **ανεξάρτητη της γνώσης του παρελθόντος** και είναι ίση με την πιθανότητα να περιμέναμε **α** χρονικές στιγμές εξ' αρχής.

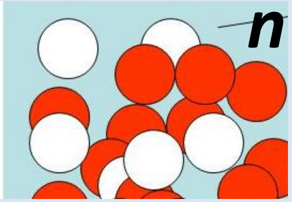
Έτσι υπάρχει **απώλεια μνήμης** στη διαδικασία.

(6) **Υπεργεωμετρική (Hyper-Geometrical)** κατανομή



Δοχείο με ***N*** αντικείμενα εκ των οποίων τα ***m*** έχουν μία κοινή ιδιότητα (π.χ. κόκκινο χρώμα)

Επιλέγουμε ***n*** αντικ. από τα *N*



- Η **Υπεργεωμετρική** τυχαία μεταβλητή (τ.μ.) ***X*** μετράει πόσα από τα ***n*** αντικείμενα έχουν την ιδιότητα των ***m*** αντικειμένων (π.χ. χρώμα) .

- Συμβολίζουμε *τ.μ. X ~ Hg(**N**,**m**,**n**)*

- Σύνολο τιμών της μεταβλητής $\boldsymbol{\Omega_X = \{0,1,\ \ldots,\ \min(n,m)\}}$

(6) Υπεργεωμετρική (**Hyper-Geometrical**) κατανομή (συν.)

- Συνάρτηση πυκνότητας πιθανότητας (σ.π.π.):

$$f(x) = P(X = x) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}} \quad , \quad x \in [0,1,\ldots,\min(n,m)]$$

- **Αναδρομικός τύπος**:

$$\frac{f(x)}{f(x-1)} = \frac{m-x+1}{x}\,\frac{n-x+1}{N-m-n-x}$$

(7) Poisson κατανομή

- Αναφέρεται στο πλήθος εμφάνισης ενός φαινομένου **σε ένα συγκεκριμένο χρονικό ή χωρικό διάστημα**.

  παραδείγματα:

  – αριθμός άφιξης πελατών **σε μια ημέρα**,
  – πλήθος αποσταλθέντων μηνυμάτων **σε μια εβδομάδα**,
  – αριθμός επισκέψεων ιστοσελίδας **σε ένα μήνα**,
  – πλήθος υπηρεσιών **σε ένα έτος**, κλπ.
  – αριθμός τυπογραφικών λαθών ανά σελίδα **σε ένα βιβλίο**,
  – αριθμός εμφάνισης της λέξης «πιθανότητα» **σε ένα άρθρο**,
  – αριθμός λακκουβών ανά 100 m που διαπιστώνεται κατά τον ποιοτικό έλεγχο ενός δρόμου,
  – αριθμός ελαττωματικών **pixel στην οθόνη** LCD συγκεκριμένης μάρκας.

(7) Poisson κατανομή

- Αναφέρεται στο πλήθος εμφάνισης ενός φαινομένου σε ένα **συγκεκριμένο** χρονικό ή χωρικό **διάστημα**.

- συμβολίζουμε τ.μ. $X \sim P(\lambda)$, $\lambda > 0$: **παράμετρος**

- σύνολο τιμών της μεταβλητής $\Omega_X = \{0,1, 2, \ldots\}$

(7) Poisson κατανομή (συν.)

- Συνάρτηση πυκνότητας πιθανότητας (**σ.π.π.**)

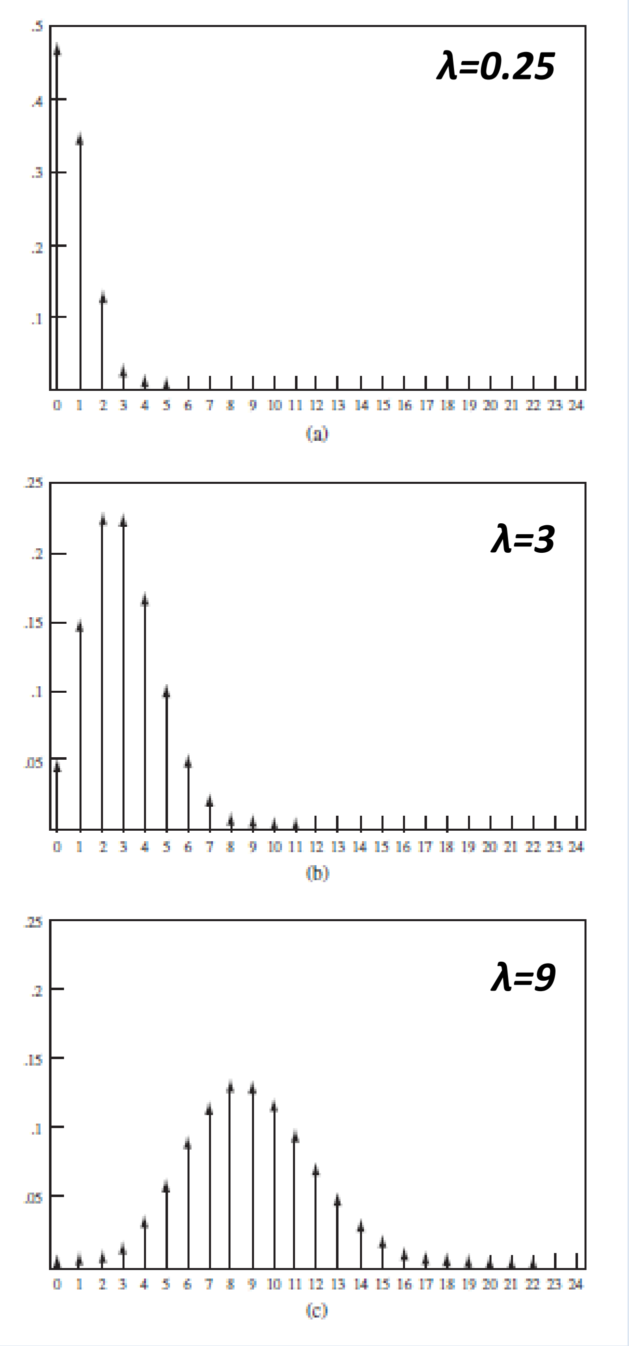
$$f(x) = P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

λ: *ρυθμός* εμφάνισης του φαινομένου

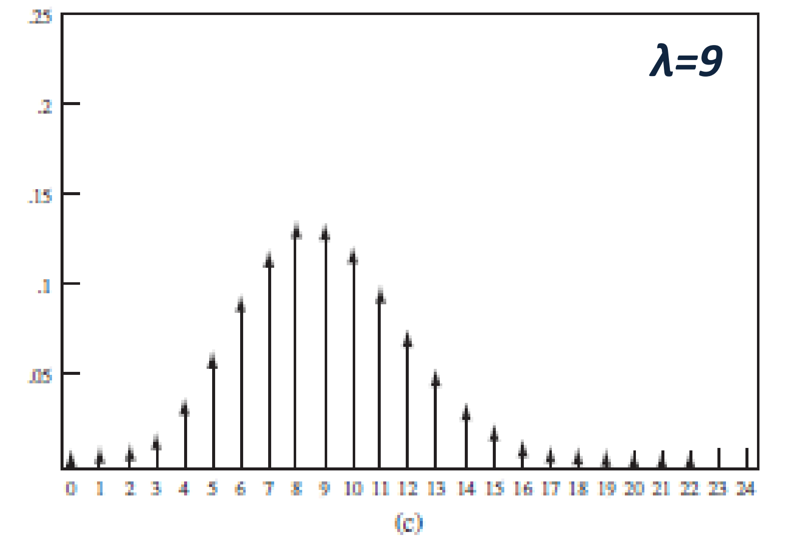
- **Επαλήθευση:**

$$\sum_{x=0}^{\infty} f(x) = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} e^{-\lambda} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1$$

ισχύει ότι: $\sum_{x=0}^{\infty} \frac{a^x}{x!} \to e^a$



Poisson κατανομή (συν.)



- αναδρομικός τύπος:

$$\frac{f(x+1)}{f(x)}=\frac{\dfrac{\lambda^{x+1}}{(x+1)!}e^{-\lambda}}{\dfrac{\lambda^{x}}{x!}e^{-\lambda}}=\frac{\lambda}{x+1}$$

Έτσι αν $$\alpha(x)=\frac{\lambda}{x+1}\geq 1\Rightarrow x+1\leq\lambda$$

τότε η ***f(x)*** είναι **αύξουσα** ↑

Προσέγγιση της Διωνυμικής κατανομής με **Poisson**

- Υποθέτουμε ότι $\boldsymbol{n \to \infty}$, $\boldsymbol{\rho \to 0}$ και ότι $\boldsymbol{\lambda = n\rho}$
- σ.π.π. της Διωνυμικής $f(x) = \binom{n}{x} \rho^x (1-\rho)^{n-x} = \frac{n!}{x!(n-x)!} \rho^x (1-\rho)^{n-x}$
- Αντικαθιστώντας $\rho = \frac{\lambda}{n}$ παίρνουμε

$$f(x) = \frac{n!}{x!(n-x)!} \frac{\lambda^x}{n^x} \left(1 - \frac{\lambda}{n}\right)^{n-x} =$$

$$= \frac{\lambda^x}{x!} \left[\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n\right] \left[\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{-x}\right] \left[\lim_{n\to\infty} \frac{n!}{(n-x)!\, n^x}\right] =$$

$$= \frac{\lambda^x}{x!} e^{-\lambda} \times 1 \times 1 \quad = \frac{\lambda^x}{x!} e^{-\lambda}$$

σ.π.π. της Poisson κατανομής

Προσέγγιση της Διωνυμικής κατανομής με **Poisson**

$$\lim_{n\to\infty} B\,in(n,\rho) = P(\lambda = n\rho)$$

- **Παράδειγμα:** «Η πιθανότητα σφάλματος στη μετάδοση ενός bit είναι 10^{-3}. Να βρεθεί η πιθανότητα σε ένα μπλοκ από 1000 bits να υπάρχουν περισσότερα από 4 σφάλματα»

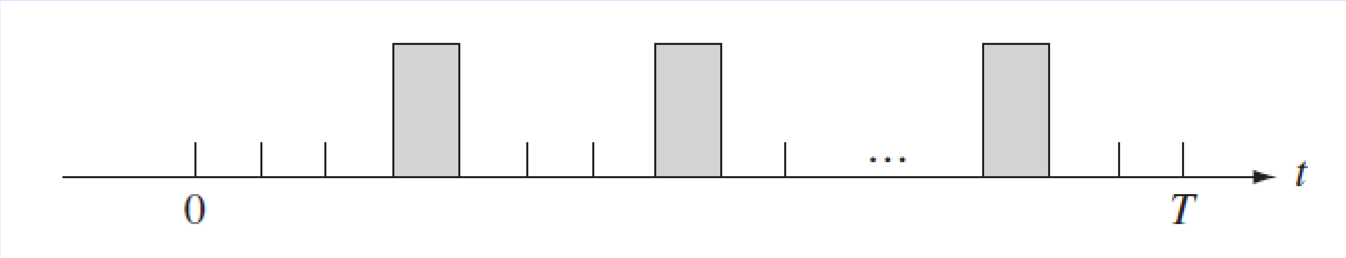
Λύση

Καθώς ***λ = np = 1,*** η τ.μ. *X*: πλήθος σφαλμάτων στο μπλοκ **προσεγγιστικά** είναι **Poisson**, δηλ. *X ~ P(λ=1).* Τότε:

$$P(X>4) = 1 - P(X \le 4) = 1 - \sum_{k=0}^{4} f(k) = 1 - \sum_{k=0}^{4} \frac{\lambda^k}{k!} e^{-\lambda} =$$

$$= 1 - e^{-1}\left(1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \frac{1}{4!}\right) = 0.00366$$

Διαδικασίες Poisson (*Poisson Processes*)



- Μετράμε τον αριθμό της εμφάνισης ενός φαινομένου σε ένα χρονικό **διάστημα Τ** με **ρυθμό εμφάνισης λ**.

- Διαιρούμε το διάστημα σε ***n*** **Bernoulli υποδιαστήματα** στα οποία:
  - Μπορεί να συμβεί **το πολύ ένα ενδεχόμενο**
  - Τα αποτελέσματα στα υποδιαστήματα είναι **ανεξάρτητα** μεταξύ τους
  - Η **πιθανότητα** να συμβεί ένα φαινόμενο σε ένα υποδιάστημα είναι **ρ=λ/n**

- Ο **αριθμός *N*** της εμφάνισης του φαινομένου στο διάστημα Τ ακολουθεί τη *Διωνυμική κατανομή* ***Bin(n,ρ=λ/n).***

- Καθώς n $\rightarrow\infty$, προσεγγιστικά έχουμε ότι ***N ~ Poisson(λ)***

- Η ακολουθία εμφάνισης γεγονότων στο χρόνο ονομάζεται **στοχαστική διαδικασία Poisson**, π.χ. άφιξη πελατών σε κατάστημα, εξυπηρέτηση μηνυμάτων σε έναν server, κλπ.

(8) Κατανομή **zipf** *(George Zipf - 1949)*

- Βαθμολογούμε τις λέξεις ενός κειμένου ως προς την συχνότητα εμφάνισής τους.

- **Νόμος του Zipf**: «Η συχνότητα της ***i*-οστής πιο συχνά** εμφανιζόμενης λέξης είναι ***1/i* φορές** την συχνότητα της πιο συχνής»

- Η τ.μ. X εκφράζει την βαθμολογία μιας λέξης και ακολουθεί την κατανομή *zipf* :

$$f(x) = \frac{1}{c_L}\frac{1}{x} \quad , \quad x = 1, 2, \ldots, L$$

όπου L είναι το πλήθος των λέξεων και c_L είναι μία σταθερά κανονικοποίησης:

$$c_L = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{L} = \sum_{x=1}^{L} \frac{1}{x}$$

Η 2η λέξη έχει πιθανότητα 1/2 της συχνότητας της πιο συχνής λέξης,
Η 3η λέξη έχει πιθανότητα 1/3 της συχνότητας της πιο συχνής λέξης, κ.ο.κ.

- **Εφαρμογές στο Internet (*επισκεψιμότητα ιστοσελίδων, ηλεκτρονικές πωλήσεις, ..*)**

Κυριότερες κατανομές Διακριτών τυχαίων μεταβλητών

Ονομασία Κατανομής	Συνάρτηση πυκνότητας πιθαν. *f(x)*	Συνάρτηση κατανομής πιθαν. *F(x)*	Φυσική σημασία
Ομοιόμορφη U(n)	$f(x_i) = \frac{1}{n}\ ,\forall\ \ i = 1,\ldots,n$	$F(x) = \begin{cases} 0 & x < 1 \\ \frac{\lfloor x \rfloor}{n} & 1 \le x < n \\ 1 & x \ge n \end{cases}$	**Ισοπίθανες τιμές** της τυχαίας μεταβλητής είναι
Διωνυμική Bin(n,ρ)	$f(x) = \binom{n}{x} \rho^x (1-\rho)^{n-x}\ ,$ $x = \{0,1,\ldots,n\}$		**Πλήθος επιτυχιών** σε *n* ανεξάρτητα *Bernoulli* πειράματα
Αρνητική Διωνυμική Nb(n,ρ)	$f(x) = \binom{x-1}{n-1} \rho^n (1-\rho)^{x-n},\ x \ge n$		**Πλήθος πειραμάτων** που απαιτούνται για να εμφανιστούν ***n*** επιτυχίες
Γεωμετρική G(ρ)	$f(x) = \rho(1-\rho)^{x-1}\ ,$ $x = \{1,2,\ldots\}$	$F(x) = 1-(1-\rho)^x$	**Χρόνος αναμονής** μέχρι την εμφάνιση **ενός** φαινομένου
Poisson P(λ)	$f(x) = \frac{\lambda^x}{x!} e^{-\lambda}\ ,\ x \ge 0$		**Συχνότητα φαινομένου** σε ένα χρονικό διάστημα

Παραδείγματα

(1) Στο τελικό των play-offs έχουν προκριθεί 2 ομάδες Α, Β οι οποίες παίζουν έως ότου συμπληρώσουν τις 3 νίκες. Η πιθανότητα να κερδίσει η ομάδα Α σε έναν αγώνα είναι p_A = 0.48 και η ομάδα Β είναι p_B=0.52 (δεν υπάρχει περίπτωση ισοπαλίας).

α) Ποια η πιθανότητα να χρειαστούν 5 αγώνες μέχρι την ανάδειξη νικητή?

β) Αν για την ανάδειξη του πρωταθλητή έχουν πραγματοποιηθεί 5 αγώνες, ποια είναι η πιθανότητα να είναι πρωταθλητής η ομάδα Α?

(2) Ένα ζάρι ρίχνεται συνέχεια μέχρις ότου να έρθει έξι. Ποια είναι η πιθανότητα να συμβεί αυτό

α) στην 10η ρίψη,

β) πριν από τη 10η ρίψη,

γ) μετά τη 10η ρίψη.

(3) Σε μία οικογένεια με 6 παιδιά τι είναι πιθανότερο,

(α) να υπάρχουν 3 αγόρια και 3 κορίτσια, ή

(β) να υπάρχουν 4 παιδιά από το ένα φύλο και 2 από το άλλο;

(Οι πιθανότητες για το φύλο κάθε παιδιού θεωρούνται μοιρασμένες, 50-50)