

4. Ειδικές Διακριτές, Συνεχείς Κατανομές

Στο προηγούμενο κεφάλαιο εισαγάγαμε την έννοια της τυχαίας μεταβλητής και είδαμε ότι σε κάθε τέτοια μεταβλητή, έστω X , αντιστοιχεί μία κατανομή. Είναι η κατανομή της συνολικής πιθανότητας που έχει κατανεμηθεί στο $\mathbb{R}$ (από τον Ω) μέσω της τ.μ. X . Η κατανομή αυτή εκφράζεται μέσω της συνάρτησης κατανομής F_X (ή εναλλακτικά μέσω της συνάρτησης πιθανότητας ή της συνάρτησης πυκνότητας πιθανότητας αν η X είναι διακριτή ή συνεχής τ.μ. αντίστοιχα). Δεν είναι λίγες οι περιπτώσεις όπου, τυχαίες μεταβλητές που προέρχονται από τη μελέτη φαινομενικά διαφορετικών μεταξύ τους στοχαστικών μοντέλων, ακολουθούν την ίδια κατανομή (έχουν την ίδια συνάρτηση κατανομής F). Υπάρχουν λοιπόν ορισμένες κατανομές οι οποίες εμφανίζονται αρκετά συχνά σε ένα ευρύ φάσμα εφαρμογών. Σε αυτό το κεφάλαιο θα μελετήσουμε τις κυριότερες και πιο εύχρηστες από αυτές τις κατανομές.

Ειδικές Διακριτές Κατανομές

4.1. Η ομοιόμορφη διακριτή κατανομή.

Η ομοιόμορφη διακριτή κατανομή εμφανίζεται στις περιπτώσεις όπου η υπό εξέταση τ.μ. X παίρνει πεπερασμένο πλήθος τιμές (π.χ. $X \in \{1, 2, \dots, n\}$) και όλες οι πιθανότητες $P(X = i)$ είναι ισοπίθανες. Π.χ. Αν X εκφράζει το αποτέλεσμα της ρίψης ενός ζαριού, τότε προφανώς

$$P(X = i) = \frac{1}{6}, i = 1, 2, \dots, 6$$

και η X ακολουθεί την ομοιόμορφη στο $\{1, 2, \dots, 6\}$ κατανομή. Γενικότερα η διακριτή ομοιόμορφη κατανομή ορίζεται ως εξής:

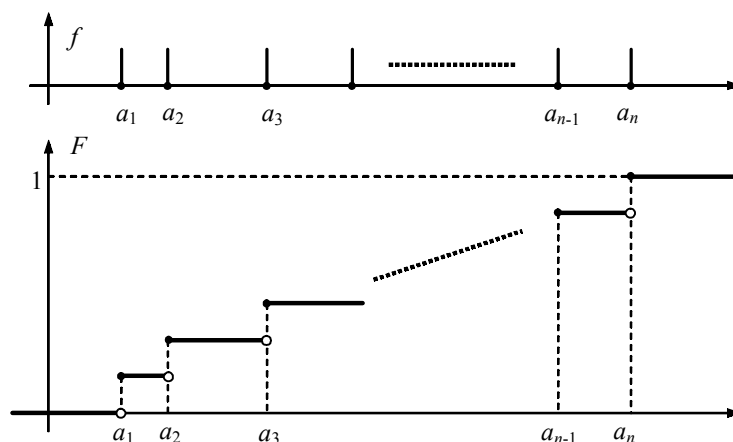
Ορισμός 4.1. Η κατανομή με σ.π.

$$f(a_i) = \frac{1}{n}, i = 1, 2, \dots, n$$

καλείται διακριτή ομοιόμορφη κατανομή στο σύνολο $\{a_1, a_2, \dots, a_n\}$.

Σύμφωνα με την κατανομή αυτή, η συνολική πιθανότητα κατανέμεται ομοιόμορφα στα σημεία $\{a_1, a_2, \dots, a_n\}$ του $\mathbb{R}$ (στο καθένα από αυτά κατανέμεται πιθανότητα $1/n$). Η συνάρτηση κατανομής της διακριτής ομοιόμορφης θα είναι

$$F(a_k) = \sum_{i=1}^k f(a_i) = \sum_{i=1}^k \frac{1}{n} = \frac{k}{n}, k = 1, 2, \dots, n.$$



Η μέση τιμή της (το «κέντρο βάρους» των n μαζών πιθανότητας) θα δίνεται από τη σχέση

$$E(X) = \sum_{i=1}^n a_i P(X = a_i) = \frac{1}{n} \sum_{i=1}^n a_i \equiv \bar{a},$$

δηλαδή θα είναι ίση με τον μέσο όρο των a_i (συμβ. με $\bar{a}$) ενώ επίσης

$$E(X^2) = \sum_{i=1}^n a_i^2 P(X = a_i) = \frac{1}{n} \sum_{i=1}^n a_i^2,$$

και άρα

$$V(X) = E(X^2) - E(X)^2 = \frac{1}{n} \sum_{i=1}^n a_i^2 - \bar{a}^2 = \frac{1}{n} \left(\sum_{i=1}^n a_i^2 - n\bar{a}^2 \right) = \frac{1}{n} \sum_{i=1}^n (a_i - \bar{a})^2.$$

Η τελευταία ισότητα ισχύει διότι

$$\sum_{i=1}^n (a_i - \bar{a})^2 = \sum_{i=1}^n (a_i^2 + \bar{a}^2 - 2a_i\bar{a}) = \sum_{i=1}^n a_i^2 + n\bar{a}^2 - 2\bar{a} \sum_{i=1}^n a_i = \sum_{i=1}^n a_i^2 + n\bar{a}^2 - 2n\bar{a}^2 = \sum_{i=1}^n a_i^2 - n\bar{a}^2.$$

Ας υποθέσουμε τώρα ότι μία τ.μ. X ακολουθεί τη διακριτή ομοιόμορφη στο $\{1, 2, \dots, n\}$ ($X \in \{1, 2, \dots, n\}$) δηλαδή, $a_i = i$ για $i = 1, 2, \dots, n$. Η συνάρτηση κατανομής της θα είναι

$$F(k) = \sum_{i=1}^k f(i) = \sum_{i=1}^k \frac{1}{n} = \frac{k}{n}, \quad k = 1, 2, \dots, n.$$

ενώ για τη μέση τιμή και τη διασπορά τώρα θα ισχύει

$$E(X) = \sum_{i=1}^n iP(X=i) = \frac{1}{n} \sum_{i=1}^n i = \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2},$$

και

$$E(X^2) = \sum_{i=1}^n i^2 P(X=i) = \frac{1}{n} \sum_{i=1}^n i^2 = \frac{1}{n} \frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6}$$

από όπου προκύπτει

$$V(X) = E(X^2) - E(X)^2 = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{2(n+1)(2n+1) - 3(n+1)^2}{12} = \frac{n^2 - 1}{12}.$$

4.2. Η Διωνυμική κατανομή.

Η διωνυμική κατανομή εμφανίζεται στις περιπτώσεις όπου μελετάμε το πλήθος των επιτυχιών σε μία ακολουθία ανεξάρτητων και ισόνομων δοκιμών. Ειδικότερα, έστω ότι εκτελούμε n όμοια και στοχαστικά ανεξάρτητα μεταξύ τους πειράματα (π.χ. n ρίψεις ενός νομίσματος). Έστω επίσης ότι το αποτέλεσμα κάθε ενός από τα n αυτά πειράματα μπορεί να είναι είτε E : επιτυχία είτε A : αποτυχία με πιθανότητες p και $q = 1-p$ αντίστοιχα (π.χ. είτε «κεφαλή» (E) είτε «γράμμα-τα» (A) με πιθανότητες 0.5 και 0.5 αντίστοιχα). Δηλαδή

$$i - \text{πείραμα} : \begin{cases} E \text{ με πιθ. } p \\ A \text{ με πιθ. } q = 1 - p \end{cases}, \quad i = 1, 2, \dots, n$$

Μία πραγματοποίηση αυτού του πειράματος μπορεί να είναι η ακόλουθη:

ΕΕΕΑΕΕΕΑΕΕΕΕΑΕΑΕΕΕΑΕΕΑ ($n=20$)

Θέτουμε X την τυχαία μεταβλητή που εκφράζει το πλήθος των επιτυχιών (E) σε αυτά τα n πειράματα (π.χ. $X = 13$ στην παραπάνω πραγματοποίηση). Προφανώς, $X \in \{0, 1, \dots, n\}$.

Ας δούμε ποια είναι η κατανομή αυτής της τ.μ. υπολογίζοντας την πιθανότητα $P(X=x)$, $x \in \{0, 1, \dots, n\}$. Ο δειγματικός χώρος Ω του πειράματος αποτελείται από όλες τις διατάξεις των 2 (Ε ή Α) στοιχείων ανά n με επανάληψη. Συνεπώς,

$$\Omega = \{(a_1, a_2, \dots, a_n) : a_i \in \{E, A\}\}, \quad |\Omega| = 2^n.$$

Το ενδεχόμενο $A_{n,x} = [X=x] = \{\text{ακριβώς } x \text{ επιτυχίες στις } n \text{ δοκιμές}\}$ θα είναι

$$A_{n,x} = \{\omega \in \Omega : X(\omega)=x\} = \{(a_1, a_2, \dots, a_n) \in \Omega : \text{ακριβώς } x \text{ από τα } a_1, a_2, \dots, a_n \text{ να είναι Ε}\}.$$

Πριν αναζητήσουμε την πιθανότητα του $A_{n,x}$ στην γενική περίπτωση, ας εξετάσουμε την απλή περίπτωση που $n=4$, $x=2$. Θα είναι

$$A_{4,2} = [X=2] = \{(E,E,A,A), (E,A,E,A), (E,A,A,E), (A,E,E,A), (A,E,A,E), (A,A,E,E)\}$$

και επομένως, για $n=4$ δοκιμές,

$$\begin{aligned} P(X=2) &= P(A_{4,2}) = P((E,E,A,A)) + P((E,A,E,A)) + \dots + P((A,A,E,E)) \\ &= p^2(1-p)^2 + p^2(1-p)^2 + \dots + p^2(1-p)^2 = 6 p^2(1-p)^2. \end{aligned}$$

Σε αυτό το σημείο είναι σημαντική η παρατήρηση ότι υπάρχει μία αντιστοίχιση μεταξύ των στοιχείων του $A_{4,2}$ και του συνόλου των δυνατών συνδυασμών των 4 στοιχείων $\{1,2,3,4\}$ ανά 2 (χωρίς επανάληψη) ως εξής:

$$\begin{aligned} (E,E,A,A) &\leftrightarrow \{1,2\}, (E,A,E,A) \leftrightarrow \{1,3\}, (E,A,A,E) \leftrightarrow \{1,4\}, \\ (A,E,E,A) &\leftrightarrow \{2,3\}, (A,E,A,E) \leftrightarrow \{2,4\}, (A,A,E,E) \leftrightarrow \{3,4\} \end{aligned}$$

(π.χ. το (E,E,A,A) αντιστοιχεί στο συνδυασμό $\{1,2\}$ διότι τα στοιχεία 1,2 του (E,E,A,A) είναι επιτυχίες). Επομένως $|A_{4,2}| = 4!/(2!(4-2)!) = 6$. Αυτή η παρατήρηση θα μας βοηθήσει να δούμε τι γίνεται και στην γενική περίπτωση.

Παρατηρούμε λοιπόν γενικά ότι το πλήθος των στοιχείων του ενδεχομένου $A_{n,x}$ είναι (ανάλογα με παραπάνω) ίσο με το πλήθος των συνδυασμών των n στοιχείων ανά x χωρίς επανάληψη. Αυτό συμβαίνει διότι κάθε στοιχείο $(a_1, a_2, \dots, a_n)$ του $A_{n,x}$ μπορεί να ληφθεί επιλέγοντας x από τα n το πλήθος a_i και θέτοντάς τα ίσα με Ε (ενώ τα υπόλοιπα $n-x$ τα θέτουμε Α). Κάθε τέτοια επιλογή αντιστοιχεί σε ένα συνδυασμό των n ανά x χωρίς επανάληψη¹. Έστω τώρα ένα στοιχείο $(a_1, a_2, \dots, a_n)$ του $A_{n,x}$. Αυτό θα έχει πιθανότητα

$$P((a_1, a_2, \dots, a_n)) = P(a_1 \text{ στην } 1^{\text{η}} \text{ δοκιμή}) \cdot \dots \cdot P(a_n \text{ στην } n\text{-οστή δοκιμή}) = p^x (1-p)^{n-x}$$

αφού ακριβώς x από τα a_i είναι Ε και ακριβώς $n-x$ είναι Α. Άρα τελικά, στην γενική περίπτωση,

$$P(X=x) = \sum_{(a_1, \dots, a_n) \in A_{n,x}} P((a_1, a_2, \dots, a_n)) = \sum_{(a_1, \dots, a_n) \in A_{n,x}} p^x (1-p)^{n-x} = |A_{n,x}| p^x (1-p)^{n-x} = \binom{n}{x} p^x (1-p)^{n-x},$$

για $x=0,1,2,\dots,n$. Η κατανομή αυτή εμφανίζεται αρκετές φορές στις εφαρμογές για αυτό και έχει ιδιαίτερη ονομασία.

Ορισμός 4.2. Η κατανομή με σ.π.

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x=0,1,\dots,n$$

¹ Αυστηρότερα, μπορούμε να πούμε ότι το σύνολο $B_{n,x} = \{\{b_1, \dots, b_x\} : b_i \in \{1,2,\dots,n\}, b_i \neq b_j\}$ (το οποίο περιέχει όλους τους συνδυασμούς των n στοιχείων ανά x χωρίς επανάληψη) είναι ισοπληθικό με το A_x διότι η απεικόνιση $g: A_x \rightarrow B_{n,x}$ έτσι ώστε $g((a_1, a_2, \dots, a_n)) = \{b_1, b_2, \dots, b_x : a_{b_1} = \dots = a_{b_x} = E\}$ είναι 1-1 και επί.

καλείται διωνυμική κατανομή με παραμέτρους $n > 0, p \in (0,1)$ (συμβολίζεται και με $B(n,p)$).

Στην ουσία, παραπάνω αποδείξαμε και την ακόλουθη πρόταση:

Πρόταση 4.1. Αν μία τ.μ. X μπορεί να θεωρηθεί ότι εκφράζει το πλήθος των επιτυχιών σε n ανεξάρτητες δοκιμές σε κάθε μία από τις οποίες εμφανίζεται επιτυχία με πιθαν. p και αποτυχία με πιθαν. $1-p$ τότε $X \sim B(n, p)$.

Η διωνυμική κατανομή είναι πράγματι κατανομή ($f \geq 0, \sum f(x) = 1$) διότι είναι γνωστό ότι (τύπος διωνύμου του Νεύτωνα)

$$(a+b)^n = \sum_{i=0}^n \binom{n}{i} a^i b^{n-i}, \quad a, b \in \mathbb{R}$$

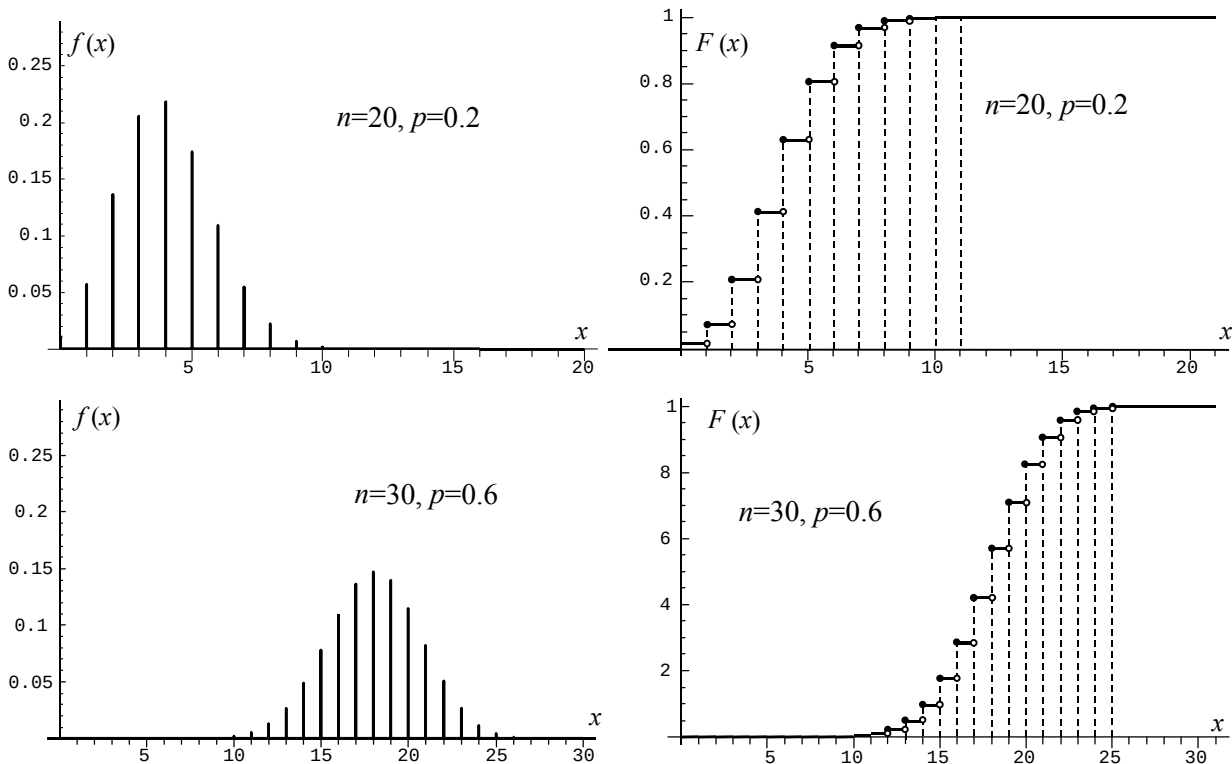
και άρα

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

Η συνάρτηση κατανομής της διωνυμικής κατανομής θα είναι

$$F(x) = P(X \leq x) = \sum_{i=0}^x f(i) = \sum_{i=0}^x \binom{n}{i} p^i (1-p)^{n-i}, \quad x = 0, 1, \dots, n.$$

Η συνάρτηση πιθανότητας και η συνάρτηση κατανομής της διωνυμικής κατανομής δίνονται στα επόμενα σχήματα για $n=20, p=0.2$ και $n=30, p=0.6$.



Έστω τώρα μία τ.μ. $X \sim B(n, p)$. Η μέση τιμή της θα είναι

$$\begin{aligned} E(X) &= \sum_{x=0}^n x f_X(x) = \sum_{x=1}^n x \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=1}^n \frac{x n!}{x!(n-x)!} p^x (1-p)^{n-x} = n \sum_{x=1}^n \binom{n-1}{x-1} p^x (1-p)^{n-x} \\ &= n \sum_{y=0}^{n-1} \binom{n-1}{y} p^{y+1} (1-p)^{n-y-1} = np \sum_{y=0}^{n-1} \binom{n-1}{y} p^y (1-p)^{n-1-y} = np. \end{aligned}$$

Έτσι αν π.χ. έχουμε 10 δοκιμές με πιθανότητα επιτυχίας $1/3$ στην κάθε μία, τότε η μέση τιμή είναι 3 επιτυχίες. Για την εύρεση της διασποράς της διωνυμικής τ.μ. X χρειαζόμαστε τον υπολογισμό της δεύτερης ροπής $E(X^2)$. Επειδή όμως ο απευθείας υπολογισμός της $E(X^2)$ είναι δύσκολος, εναλλακτικά υπολογίζουμε την $E(X(X-1))$ (η οποία ονομάζεται και δεύτερη παραγοντική ροπή της τ.μ. X) από την οποία θα προκύψει η $E(X^2)$. Θα είναι

$$\begin{aligned} E(X(X-1)) &= \sum_{x=0}^n x(x-1)f_X(x) = \sum_{x=2}^n x(x-1)\binom{n}{x}p^x(1-p)^{n-x} = \sum_{x=2}^n \frac{x(x-1)n!}{x!(n-x)!}p^x(1-p)^{n-x} \\ &= n(n-1)\sum_{x=2}^n \binom{n-2}{x-2}p^x(1-p)^{n-x} \underset{y=x-2}{=} n(n-1)\sum_{y=0}^{n-2} \binom{n-2}{y}p^{y+2}(1-p)^{n-2-y} = n(n-1)p^2 \end{aligned}$$

και επειδή $E(X(X-1)) = E(X^2 - X) = E(X^2) - E(X)$ προκύπτει ότι

$$E(X^2) = n(n-1)p^2 + E(X) = n(n-1)p^2 + np$$

Άρα τελικά,

$$V(X) = E(X^2) - E(X)^2 = n(n-1)p^2 + np - n^2p^2 = np - np^2 = np(1-p).$$

Για $n=1$ η κατανομή $B(1,p)$ είναι γνωστή και ως κατανομή Bernoulli. Αν $X \sim B(1,p)$ τότε $X \in \{0,1\}$ και

$$P(X=0) = \binom{1}{0}p^0(1-p) = 1-p \quad \text{και} \quad P(X=1) = \binom{1}{1}p^1(1-p)^{1-1} = p$$

ενώ $E(X) = 1 \cdot p = p$, $V(X) = 1 \cdot p(1-p) = p(1-p)$.

Άσκηση 4.1. Ένας ασφαλιστής ασφαλίζει 10 άτομα με την ίδια ηλικία και κατάσταση υγείας. Αν κάθε άτομο αυτής της κατηγορίας έχει πιθανότητα 60% να ζει μετά από 30 χρόνια τότε να υπολογιστεί η πιθανότητα να ζουν μετά από 30 χρόνια (α) κανένας, (β) το πολύ 3 άτομα, και (γ) τουλάχιστον 7 άτομα. Ποίος είναι ο μέσος αριθμός ατόμων που θα ζουν μετά από 30 χρόνια;

Λύση. Έστω X ο αριθμός των ατόμων από τα 10 που ζουν μετά από 30 χρόνια. Αν θεωρήσουμε ως i -δοκιμή (πείραμα) την καταγραφή της κατάστασης του i -ατόμου (μετά από 30 χρόνια) θεωρώντας ως «επιτυχία» την επιβίωση του, τότε η τ.μ. X μπορεί να θεωρηθεί ότι εκφράζει το πλήθος των επιτυχιών στις $n = 10$ αυτές ανεξάρτητες δοκιμές με πιθ. επιτυχίας $p = 60\%$ και αποτυχίας $1-p = 40\%$. Δηλαδή $X \sim B(10, 0.6)$.

α) Η πιθανότητα να μην ζει κανένας μετά από 30 χρόνια θα είναι

$$P(X=0) = \binom{n}{0}p^0(1-p)^{n-0} = (1-p)^{10} = 0.4^{10} \approx 0.000104$$

β) Η πιθανότητα να ζουν το πολύ 3 άτομα θα είναι

$$\begin{aligned} P(X \leq 3) &= \sum_{x=0}^3 P(X=x) = \sum_{x=0}^3 \binom{n}{x}p^x(1-p)^{n-x} = \\ &= \binom{n}{0}p^0(1-p)^{n-0} + \binom{n}{1}p^1(1-p)^{n-1} + \binom{n}{2}p^2(1-p)^{n-2} + \binom{n}{3}p^3(1-p)^{n-3} \\ &= 0.4^{10} + 10 \cdot 0.6 \cdot 0.4^9 + 45 \cdot 0.6^2 \cdot 0.4^8 + \frac{720}{6} 0.6^3 \cdot 0.4^7 = 0.0546. \end{aligned}$$

γ) Η πιθανότητα να ζουν τουλάχιστον 7 άτομα θα είναι

$$\begin{aligned} P(X \geq 7) &= \sum_{x=7}^{10} P(X=x) = \sum_{x=7}^{10} \binom{n}{x} p^x (1-p)^{n-x} = \\ &= \binom{n}{7} p^7 (1-p)^{n-7} + \binom{n}{8} p^8 (1-p)^{n-8} + \binom{n}{9} p^9 (1-p)^{n-9} + \binom{n}{10} p^{10} (1-p)^{n-10} = 0.3821 \end{aligned}$$

δ) Είναι γνωστό ότι όταν $X \sim B(n, p)$ τότε $E(X) = np$. Επομένως εδώ

$$E(X) = np = 10 \cdot 0.6 = 6$$

και άρα ο μέσος αριθμός ατόμων που θα ζουν μετά από 30 χρόνια είναι 6.

Άσκηση 4.2. Μια αεροπορική εταιρεία γνωρίζει ότι, κατά μέσο όρο, το 5% των ατόμων που κάνουν κράτηση για να ταξιδέψουν δεν εμφανίζονται. Αν η εταιρεία κάνει κράτηση για 52 άτομα σε μια πτήση που γίνεται με αεροσκάφος χωρητικότητας 50 ατόμων τότε ποια είναι η πιθανότητα να υπάρχει διαθέσιμο κάθισμα για κάθε ένα επιβάτη που εμφανίζεται για να ταξιδέψει;

Λύση. Έστω X ο αριθμός των ατόμων που θα εμφανιστούν από τα 52. Εδώ έχουμε να κάνουμε με 52 ανεξάρτητα πειράματα ή δοκιμές (κάθε άτομο εμφανίζεται ή όχι ανεξάρτητα από τα υπόλοιπα). Κάθε πείραμα συνίσταται από την εμφάνιση ή όχι του i -ατόμου, $i=1,2,\dots,52$. Αν ως επιτυχία θεωρήσουμε την εμφάνιση του ατόμου, η τ.μ. X μπορεί να θεωρηθεί ότι εκφράζει το πλήθος των επιτυχιών στις $n=52$ αυτές ανεξάρτητες δοκιμές με πιθαν. επιτυχίας $p=95\%$ και αποτυχίας $1-p=5\%$. Δηλαδή $X \sim B(52, 0.95)$. Η πιθανότητα να υπάρχει διαθέσιμο κάθισμα για κάθε ένα επιβάτη που εμφανίζεται για να ταξιδέψει είναι ίση με την πιθανότητα να εμφανιστούν το πολύ 50 άτομα. Δηλαδή,

$$P(X \leq 50) = \sum_{x=0}^{50} \binom{52}{x} p^x (1-p)^{52-x}$$

Επειδή είναι αρκετά επίπονο να υπολογίσουμε την πιθανότητα αυτή μέσα από το παραπάνω άθροισμα, χρησιμοποιούμε εναλλακτικά τη σχέση

$$\begin{aligned} P(X \leq 50) &= 1 - P(X > 50) = 1 - \sum_{x=51}^{52} \binom{52}{x} p^x (1-p)^{52-x} \\ &= 1 - 52 p^{51} (1-p) - p^{52} \approx 1 - 0.19 - 0.07 = 74\% . \end{aligned}$$

Άσκηση 4.3. Πόσα παιδιά πρέπει να αποκτήσει μια οικογένεια ώστε να έχει ένα τουλάχιστον αγόρι και ένα τουλάχιστον κορίτσι με πιθανότητα μεγαλύτερη του (α) 90%, (β) 95%, και (γ) 99%.

Λύση. Έστω ότι η οικογένεια έχει n παιδιά και X ο αριθμός των αγοριών. Όμοια και με τις προηγούμενες ασκήσεις, η τ.μ. X μπορεί να θεωρηθεί ότι εκφράζει το πλήθος των επιτυχιών σε n ανεξάρτητες δοκιμές με πιθαν. επιτυχίας $p=50\%$ (επιτυχία: γέννηση αγοριού). Δηλαδή $X \sim B(n, 0.50)$. Η οικογένεια θα έχει ένα τουλάχιστον αγόρι και ένα τουλάχιστον κορίτσι με πιθανότητα

$$\begin{aligned} P(0 < X < n) &= 1 - P(X=0) - P(X=n) = 1 - \binom{n}{0} p^0 (1-p)^n - \binom{n}{n} p^n (1-p)^{n-n} \\ &= 1 - (1-p)^n - p^n = 1 - \frac{1}{2^n} - \frac{1}{2^n} = 1 - \frac{1}{2^{n-1}} . \end{aligned}$$

(α) Αν $\alpha = 0.9$ θα είναι

$$1 - \frac{1}{2^{n-1}} > a \Leftrightarrow \frac{1}{2^{n-1}} < 1 - a \Leftrightarrow \frac{1}{1-a} < 2^{n-1} \Leftrightarrow -\ln(1-a) < (n-1)\ln 2 \Leftrightarrow n > 1 - \frac{\ln(1-a)}{\ln 2} = 4.32$$

και επομένως θα πρέπει $n = 5$.

(β) Όμοια, αν $\alpha=0.95$

$$n > 1 - \frac{\ln(1-a)}{\ln 2} = 1 - \frac{\ln(1-0.95)}{\ln 2} = 5.32$$

και επομένως θα πρέπει $n = 6$.

γ) Τέλος, αν $\alpha=0.99$

$$n > 1 - \frac{\ln(1-a)}{\ln 2} = 1 - \frac{\ln(1-0.99)}{\ln 2} = 7.74$$

και επομένως θα πρέπει $n = 8$.

4.3. Η γεωμετρική κατανομή

Η γεωμετρική κατανομή εμφανίζεται σε προβλήματα που έχουν σχέση με τον αριθμό των δοκιμών μέχρι την εμφάνιση της πρώτης επιτυχίας. Πιο συγκεκριμένα, ας υποθέσουμε ότι τα αποτελέσματα ενός πειράματος μπορεί να είναι είτε επιτυχία (E) είτε αποτυχία (A) με πιθανότητα p και $1-p$ αντίστοιχα. Το πείραμα αυτό επαναλαμβάνεται (έτσι ώστε το αποτέλεσμα κάθε πειράματος να μην εξαρτάται από το αποτέλεσμα των προηγούμενων) μέχρι την εμφάνιση της πρώτης επιτυχίας. Έστω X το πλήθος των δοκιμών (πειραμάτων) που χρειάστηκαν μέχρι και την εμφάνιση της πρώτης επιτυχίας. Το ενδεχόμενο $[X = n]$ είναι το ενδεχόμενο «στις $n-1$ πρώτες δοκιμές είχαμε αποτυχία ενώ στην n -οστή εμφανίστηκε επιτυχία». Συνεπώς,

$$\begin{aligned} P(X = n) &= P(A \text{ στην } 1^{\text{η}} \text{ δοκιμή}) \cdot \dots \cdot P(A \text{ στην } (n-1)\text{-οστή δοκιμή}) P(E \text{ στην } n\text{-οστή δοκιμή}) \\ &= (1-p)^{n-1} p, \quad n = 1, 2, \dots \end{aligned}$$

Θα έχουμε λοιπόν τον επόμενο ορισμό:

Ορισμός 4.3. Η κατανομή με σ.π.

$$f(x) = (1-p)^{x-1} p, \quad x = 1, 2, \dots$$

καλείται γεωμετρική κατανομή με παράμετρο p (συμβολίζεται και με $Ge(p)$).

Παραπάνω αποδείξαμε και την ακόλουθη πρόταση:

Πρόταση 4.2. Αν μία τ.μ. X μπορεί να θεωρηθεί ότι εκφράζει το πλήθος των δοκιμών μέχρι και την εμφάνιση της πρώτης επιτυχίας σε μία ακολουθία ανεξάρτητων δοκιμών με πιθανότητα επιτυχίας p τότε $X \sim Ge(p)$.

Η γεωμετρική κατανομή είναι πράγματι κατανομή ($f \geq 0$, $\sum f(x) = 1$) διότι

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

Η συνάρτηση κατανομής της διωνυμικής κατανομής θα είναι

$$F(x) = P(X \leq x) = \sum_{i=1}^x f(i) = \sum_{i=1}^x (1-p)^{i-1} p = p \sum_{i=0}^{x-1} (1-p)^i = p \frac{1-(1-p)^x}{1-(1-p)} = 1 - (1-p)^x, \quad x = 1, 2, \dots$$

Για την μέση τιμή της $X \sim Ge(p)$ θα είναι

$$E(X) = \sum_{x=1}^{\infty} x f_X(x) = \sum_{x=1}^{\infty} x(1-p)^{x-1} p = p \sum_{x=1}^{\infty} x(1-p)^{x-1}$$

Παραγωγίζοντας και τα δυο μέλη της γνωστής παράστασης $\sum_{x=0}^{\infty} a^x = (1-a)^{-1}$ προκύπτει ότι

$$\sum_{x=1}^{\infty} x a^{x-1} = \frac{1}{(1-a)^2}$$

και άρα τελικά

$$E(X) = p \sum_{x=1}^{\infty} x(1-p)^{x-1} = p \frac{1}{(1-(1-p))^2} = \frac{1}{p}.$$

Όσο λοιπόν η πιθανότητα επιτυχίας p μικραίνει, τόσο η μέση τιμή του πλήθους των δοκιμών μέχρι την εμφάνισή της μεγαλώνει. Για την εύρεση της διασποράς της τ.μ. $X \sim \text{Ge}(p)$ χρειαζόμαστε τον υπολογισμό της δεύτερης ροπής $E(X^2)$. Επειδή όμως ο απευθείας υπολογισμός της $E(X^2)$ είναι και εδώ δύσκολος, θα την υπολογίσουμε μέσω της δεύτερης παραγοντικής ροπής $E(X(X-1))$. Θα είναι

$$E(X(X-1)) = \sum_{x=1}^{\infty} x(x-1) f_X(x) = \sum_{x=2}^{\infty} x(x-1)(1-p)^{x-1} p$$

Παραγωγίζοντας δύο φορές την παράσταση $\sum_{x=0}^{\infty} a^x = (1-a)^{-1}$ προκύπτει ότι

$$\sum_{x=2}^{\infty} x(x-1) a^{x-2} = \frac{2}{(1-a)^3}$$

και άρα

$$E(X(X-1)) = p(1-p) \sum_{x=2}^{\infty} x(x-1)(1-p)^{x-2} = p(1-p) \frac{2}{(1-(1-p))^3} = \frac{2(1-p)}{p^2}$$

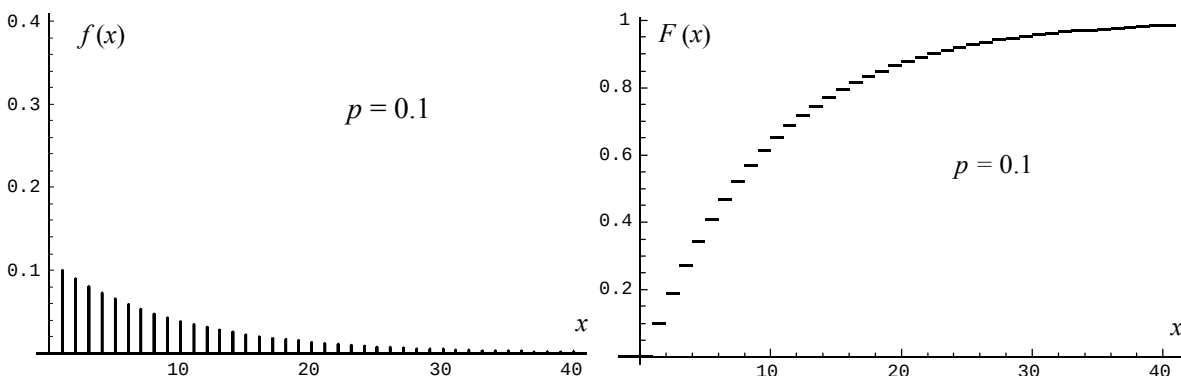
και επειδή $E(X(X-1)) = E(X^2 - X) = E(X^2) - E(X)$ προκύπτει ότι

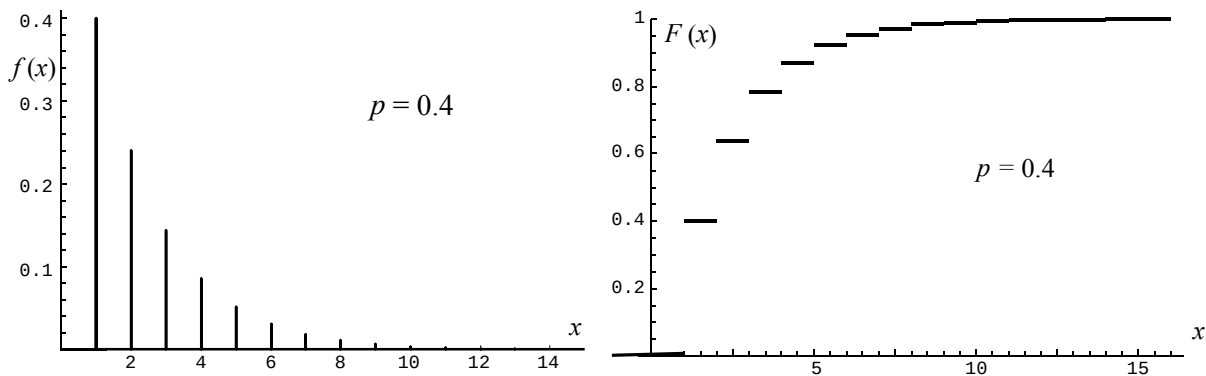
$$E(X^2) = \frac{2(1-p)}{p^2} + \frac{1}{p} = \frac{2-p}{p^2}$$

Άρα τελικά,

$$V(X) = E(X^2) - E(X)^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

Η συνάρτηση πιθανότητας και η συνάρτηση κατανομής της γεωμετρικής κατανομής δίνονται στα επόμενα σχήματα για $p = 0.1$ και $p = 0.4$.





Τέλος, αξίζει να παρατηρηθεί ότι η τ.μ. $Y = X-1$ εκφράζει το πλήθος των αποτυχιών (αντί των δοκιμών) μέχρι και την εμφάνιση της πρώτης επιτυχίας. Η τ.μ. Y θα έχει σ.π.

$$f_Y(x) = P(Y = x) = P(X - 1 = x) = P(X = x + 1) = f_X(x + 1) = (1 - p)^x p, \quad x = 0, 1, 2, \dots$$

Παρατηρούμε ότι $Y \in \{0, 1, 2, \dots\}$ ενώ $X \in \{1, 2, \dots\}$. Συνήθως λέγεται ότι η X ακολουθεί τη γεωμετρική από το 1 κατανομή ($Ge_1(p)$), ενώ η Y ακολουθεί τη γεωμετρική από το 0 κατανομή ($Ge_0(p)$). Θα είναι

$$F_Y(x) = P(Y \leq x) = P(X - 1 \leq x) = P(X \leq x + 1) = F_X(x + 1) = 1 - (1 - p)^{x+1}, \quad x = 0, 1, 2, \dots$$

και

$$E(Y) = E(X - 1) = E(X) - 1 = \frac{1}{p} - 1 = \frac{1-p}{p}, \quad V(Y) = V(X - 1) = V(X) = \frac{1-p}{p^2}.$$

Άσκηση 4.4. Ένα πείραμα εκτελείται συνεχώς έως ότου εκτελεστεί επιτυχώς για πρώτη φορά. Το κόστος εκτέλεσης του πειράματος την πρώτη φορά είναι 50 χιλ. δρχ, ενώ οι επόμενες εκτελέσεις του απαιτούν 10 χιλ. δρχ έκαστη. Αν η πιθανότητα επιτυχούς εκτέλεσης του πειράματος είναι ίση με $2/3$ (σταθερή σε κάθε επανάληψη του πειράματος) τότε να βρεθεί το αναμενόμενο συνολικό κόστος των πειραμάτων που θα χρειαστούν. Ποια η πιθανότητα να χρειαστούν περισσότερες από 80 χιλ. δρχ.

Λύση. Έστω X το πλήθος των πειραμάτων που θα εκτελεστούν μέχρι και την πρώτη επιτυχή εκτέλεση. Σύμφωνα με τα παραπάνω, η τ.μ. θα ακολουθεί γεωμετρική (από το 1) κατανομή με παράμετρο $p=2/3$. Το κόστος Y του πειράματος θα είναι

$$Y = 10(X - 1) + 50 = 10X + 40 \text{ χιλιάδες δρχ.}$$

Επομένως το αναμενόμενο κόστος $E(Y)$ του πειράματος θα είναι

$$E(Y) = E(10X + 40) = 10E(X) + 40 = 10 \frac{1}{p} + 40 = 10 \frac{3}{2} + 40 = 55 \text{ χιλιάδες δρχ.}$$

Η πιθανότητα να χρειαστούν περισσότερες από 80 χιλ. δρχ. θα είναι

$$P(Y > 80) = P(10X + 40 > 80) = P(X > 4) = 1 - P(X \leq 4) = 1 - F_X(4) = (1 - p)^4 = \frac{1}{3^4} = 0.0123.$$

Άσκηση 4.5. Κάποιος που παίζει ρουλέτα στο καζίνο του Monte Carlo ποντάρει συνέχεια στο μαύρο και η πιθανότητα να έρθει μαύρο σε κάθε γύρισμα της ρουλέτας είναι ίση με $1/2$. Ο παίκτης ξεκινάει ποντάροντας 1\$ και κάθε φορά που χάνει διπλασιάζει το ποντάρισμα στο επόμενο παιχνίδι ενώ σταματά όταν κερδίσει για πρώτη φορά. Να βρεθεί το αναμενόμενο κέρδος του παίκτη και το αναμενόμενο ποσό χρημάτων που θα χρειαστεί να ποντάρει.

Λύση. Έστω X το πλήθος των παιχνιδιών που ποντάρει ο παίκτης έως ότου κερδίσει για πρώτη φορά. Στο πρώτο παιχνίδι ποντάρει 1\$, στο δεύτερο 2\$, στο τρίτο 4\$ κ.ο.κ. Το κέρδος Y του παίκτη θα είναι

$$Y = 2^{X-1} - (1 + 2 + 2^2 + \dots + 2^{X-2}) = 2^{X-1} - \sum_{i=0}^{X-2} 2^i = 2^{X-1} - \frac{2^{X-1} - 1}{2 - 1} = 1\$$$

(στο i -παιχνίδι ο παίκτης ποντάρει 2^{i-1} \$ και άρα αν κερδίσει, θα έχει καθαρό κέρδος 2^{i-1} \$ μείον το ποσό που έχει ποντάρει στα πρώτα $i-1$ παιχνίδια που είναι $1+2+\dots+2^{i-2}$ \$). Επομένως ο παίκτης θα κερδίσει με βεβαιότητα 1 δολάριο. Άρα, αν ο παίκτης ακολουθήσει το ίδιο σύστημα n φορές θα κερδίσει με βεβαιότητα n δολάρια! Και τότε γιατί δεν ακολουθείται αυτό το σύστημα για να κερδίζει κανείς με βεβαιότητα; Εδώ υπάρχει μια ουσιώδης λεπτομέρεια. Για να κερδίσει ο παίκτης αυτό το 1\$ θα πρέπει να έχει να ποντάρει όσα θα χρήματα θα απαιτηθούν μέχρι να κερδίσει για πρώτη φορά. Αν τα χρήματα του τελειώσουν πριν κερδίσει τότε φυσικά δεν μπορεί να συνεχίσει το παιχνίδι. Αλλά πόσα κατά μέσο όρο χρήματα χρειάζεται να ποντάρει μέχρι να κερδίσει το δολάριο; Έστω Z το ποσό χρημάτων που θα χρειαστεί να ποντάρει. Θα ισχύει ότι

$$Z = 1 + 2^1 + 2^2 \dots + 2^{X-1} = \sum_{i=0}^{X-1} 2^i = \frac{2^X - 1}{2 - 1} = 2^X - 1$$

και άρα,

$$E(Z) = E(2^X - 1) = E(2^X) - 1 = \sum_{x=1}^{\infty} 2^x f_X(x) - 1 = \sum_{x=1}^{\infty} 2^x p(1-p)^{x-1} - 1 = \sum_{x=1}^{\infty} 2^x \frac{1}{2^x} - 1 = \sum_{x=1}^{\infty} 1 - 1 = \infty$$

Επομένως, για να είναι κάθε φορά σε θέση ο παίκτης να συνεχίζει το παιχνίδι μέχρι τέλους θα πρέπει να διαθέτει άπειρο χρηματικό ποσό! (πρακτικά, να διαθέτει ποσό μεγαλύτερο από αυτό που διαθέτει το καζίνο!). Κάτι τέτοιο φυσικά δεν είναι δυνατό οπότε θα υπάρχουν φορές που ο παίκτης θα αναγκαστεί να σταματήσει το παιχνίδι έχοντας χάσει όλα του τα χρήματα. Αλλά ως εξετάσουμε τι γίνεται στην περίπτωση που ο παίκτης διαθέτει ποσό $2^n - 1$ \$, δηλαδή μπορεί να ποντάρει μέχρι και n φορές. Σε αυτή την περίπτωση το κέρδος του παίκτη θα είναι

$$Y = \begin{cases} 1 & \text{αν } X \leq n \\ -(1 + 2 + \dots + 2^{n-1}) = -(2^n - 1) & \text{αν } X > n \end{cases}$$

και άρα

$$P(Y = 1) = P(X \leq n) = 1 - \frac{1}{2^n}, \quad P(Y = -(2^n - 1)) = P(X > n) = \frac{1}{2^n}$$

Το αναμενόμενο κέρδος του παίκτη θα είναι

$$E(Y) = 1P(Y = 1) - (2^n - 1)P(Y = -(2^n - 1)) = 1 - \frac{1}{2^n} - (2^n - 1)\frac{1}{2^n} = 0.$$

4.4. Η κατανομή Poisson.

Η κατανομή Poisson ανακαλύφθηκε από τον S.D. Poisson το 1837 ως προσέγγιση της διωνυμικής κατανομής. Η κατανομή αυτή συνήθως εμφανίζεται κατά τη μελέτη ενός πολύ μεγάλου αριθμού από σπάνια και ανεξάρτητα μεταξύ τους ενδεχόμενα. Πιο συγκεκριμένα, έστω, όπως και στην περίπτωση της διωνυμικής κατανομής, ότι εκτελούμε n όμοια και στοχαστικά ανεξάρτητα μεταξύ τους πειράματα το καθένα από τα οποία μπορεί να καταγραφεί ως επιτυχία με πιθανότητα p και αποτυχία με πιθανότητα $q = 1-p$. Ας συμβολίσουμε με X την τυχαία μεταβλητή που εκφράζει το πλήθος των επιτυχιών σε αυτά τα n πειράματα. Σύμφωνα με την Παράγραφο 4.2, η τ.μ. X θα ακολουθεί διωνυμική κατανομή $B(n, p)$. Είναι όμως αρκετά συνήθης η περίπτωση όπου το πλήθος των πειραμάτων n είναι αρκετά μεγάλο ενώ η πιθανότητα επιτυχίας p είναι πολύ μικρή.

Για παράδειγμα, έστω ότι μία ασφαλιστική εταιρία έχει n ασφαλισμένους και p είναι η πιθανότητα ένας ασφαλισμένος να ζητήσει αποζημίωση (επιτυχία: απαίτηση αποζημίωσης) για κάποιο ατύχημα που είχε (π.χ. κατά τη διάρκεια μίας συγκεκριμένης ημέρας). Το πλήθος X των αποζημιώσεων που θα κληθεί να καλύψει η εταιρία (κατά τη διάρκεια της συγκεκριμένης ημέρας) θα ακολουθεί διωνυμική κατανομή $B(n, p)$. Εδώ όμως, συνήθως ισχύει ότι το n είναι αρκετά μεγάλο ενώ το p είναι αρκετά μικρό. Σε ανάλογες περιπτώσεις (π.χ. πλήθος εισαχθέντων σε ένα νοσοκομείο από n κατοίκους, ο καθένας από τους οποίους εισάγεται με πιθαν. p , πλήθος πελατών που προσήλθαν σε ένα εμπορικό κατάστημα από τα n άτομα μιας περιοχής, το καθένα από τα οποία προσέρχεται με πιθαν. p κ.ο.κ.) όπου $n \rightarrow \infty, p \rightarrow 0$, η τ.μ. X μπορεί γενικότερα να θεωρηθεί ότι εκφράζει το πλήθος των πραγματοποιήσεων από ένα μεγάλο αριθμό «σπάνιων» ενδεχομένων. Είναι συνεπώς αρκετά χρήσιμη η μελέτη της κατανομής $B(n, p)$ κάτω από τις συνθήκες:

$$n \rightarrow \infty, p \rightarrow 0, E(X) = np \rightarrow \lambda \in (0, \infty)$$

Η τρίτη συνθήκη εξασφαλίζει ότι ο μέσος αριθμός επιτυχιών (πλήθος πραγματοποιήσεων των «σπάνιων» ενδεχομένων) δεν είναι 0 ή ∞ . Αν λοιπόν ισοδύναμα ισχύει ότι $n \rightarrow \infty, p = \lambda/n$ τότε

$$\begin{aligned} P(X = k) &= \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{1}{k!} n(n-1)\dots(n-k+1) \frac{\lambda^k}{n^k} \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{1}{k!} \frac{n}{n} \left(\frac{n-1}{n}\right) \dots \left(\frac{n-k+1}{n}\right) \lambda^k \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k}. \end{aligned}$$

Επειδή τώρα,

$$\frac{n}{n} \left(\frac{n-1}{n}\right) \dots \left(\frac{n-k+1}{n}\right) = 1 \cdot \left(1 - \frac{1}{n}\right) \dots \left(1 - \frac{k-1}{n}\right) \rightarrow_{n \rightarrow \infty} 1$$

και

$$\left(1 - \frac{\lambda}{n}\right)^n \rightarrow_{n \rightarrow \infty} e^{-\lambda}, \quad \left(1 - \frac{\lambda}{n}\right)^{-k} \rightarrow_{n \rightarrow \infty} (1-0)^{-k} = 1$$

προκύπτει τελικά ότι

$$\binom{n}{k} p^k (1-p)^{n-k} \xrightarrow[n \rightarrow \infty, p = \lambda/n]{} e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

Επομένως, καταλήγουμε στον επόμενο ορισμό.

Ορισμός 4.4. Η κατανομή με σ.π.

$$f(x) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \dots$$

καλείται κατανομή Poisson με παράμετρο λ (συμβολίζεται και με $Po(\lambda)$).

Παραπάνω αποδείξαμε και την ακόλουθη πρόταση:

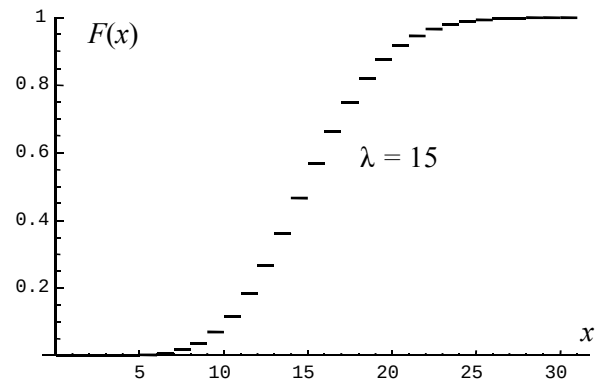
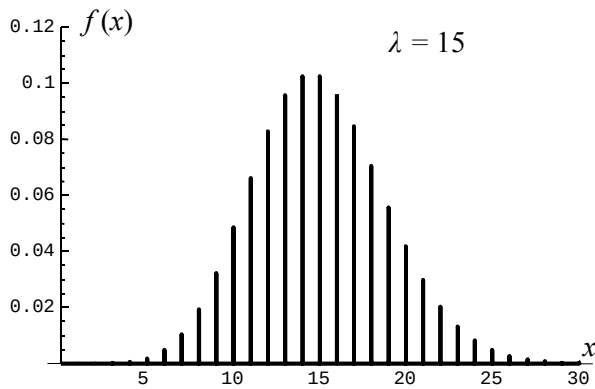
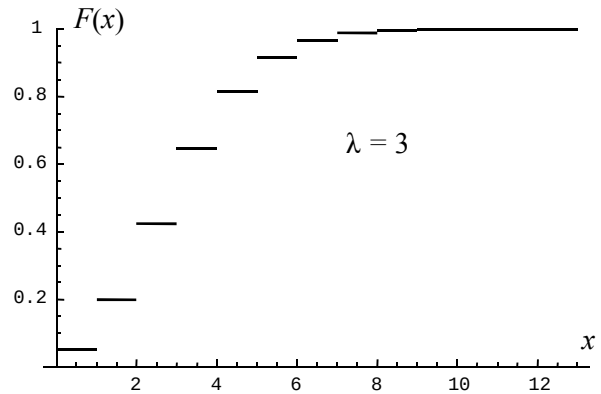
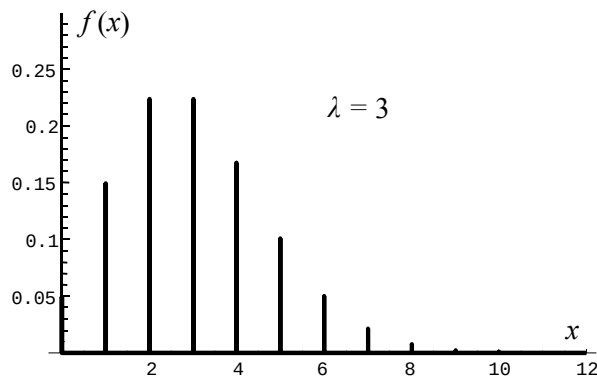
Πρόταση 4.3. Αν μία τ.μ. X μπορεί να θεωρηθεί ότι εκφράζει το πλήθος των πραγματοποιήσεων από ένα μεγάλο αριθμό «σπάνιων» ενδεχομένων (ή ισοδύναμα, το πλήθος των επιτυχιών σε ένα μεγάλο αριθμό δοκιμών ($n \rightarrow \infty$) με πιθανότητα επιτυχίας $p \rightarrow 0$ σε κάθε δοκιμή) τότε προσεγγιστικά $X \sim Po(\lambda)$.

Η κατανομή Poisson είναι πράγματι κατανομή ($f \geq 0, \sum f(x) = 1$) διότι

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

Η συνάρτηση κατανομής της κατανομής Poisson θα είναι

$$F(x) = P(X \leq x) = \sum_{i=0}^x f(i) = e^{-\lambda} \sum_{i=0}^x \frac{\lambda^i}{i!}, \quad x = 0, 1, 2, \dots$$



Έστω τώρα μία τ.μ. $X \sim \text{Po}(\lambda)$. Η μέση τιμή της θα είναι

$$E(X) = \sum_{x=0}^{\infty} x f_X(x) = \sum_{x=1}^{\infty} x e^{-\lambda} \frac{\lambda^x}{x!} = \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda.$$

Για την εύρεση της διασποράς της τ.μ. $X \sim \text{Po}(\lambda)$ χρειαζόμαστε τον υπολογισμό της $E(X^2)$. Είναι ευκολότερο να υπολογίσουμε την $E(X^2)$ μέσω της $E(X(X-1))$. Θα είναι

$$E(X(X-1)) = \sum_{x=0}^{\infty} x(x-1) e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} \lambda^2 \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} = e^{-\lambda} \lambda^2 e^{\lambda} = \lambda^2$$

και επειδή $E(X(X-1)) = E(X^2 - X) = E(X^2) - E(X)$ προκύπτει ότι

$$E(X^2) = \lambda^2 + \lambda$$

Άρα τελικά,

$$V(X) = E(X^2) - E(X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Άσκηση 4.6. Ο αριθμός X των τυπογραφικών λαθών σε μια σελίδα ενός βιβλίου ακολουθεί την κατανομή Poisson με παράμετρο $\lambda = 5$. Να υπολογιστεί το ποσοστό των σελίδων του βιβλίου που περιέχουν (α) 2 ακριβώς λάθη, και (β) το πολύ 2 λάθη.

Λύση. Παρατηρούμε ότι ο αριθμός X των τυπογραφικών λαθών σε μια σελίδα ενός βιβλίου μπορεί να θεωρηθεί ότι εκφράζει το πλήθος των πραγματοποιήσεων από ένα μεγάλο αριθμό «σπάνιων» ενδεχομένων (η σελίδα περιέχει έναν μεγάλο αριθμό γραμμάτων το καθένα από τα οποία γράφεται λάθος με πιθανότητα $p \rightarrow 0$). Επομένως η υπόθεση ότι $X \sim \text{Po}(\lambda)$ είναι αληθοφανής. Αρχικά θα υπολογίσουμε την πιθανότητα $P(X=2)$. Θα ισχύει ότι

$$P(X=2) = e^{-5} \frac{5^2}{2!} \approx 0.0842$$

Έστω τώρα $X_1, X_2, \dots, X_n$ το πλήθος των τυπογραφικών λαθών στην 1^η, 2^η, ..., n -οστή σελίδα αντίστοιχα. Προφανώς $X_i \sim \text{Po}(5)$. Αν θέσουμε

$$Y_i = \begin{cases} 1, & \text{αν } X_i = 2 \\ 0, & \text{αν } X_i \neq 2 \end{cases}, i=1, 2, \dots, n$$

τότε το ποσοστό των σελίδων (από τις n) που περιέχουν ακριβώς δύο λάθη θα είναι

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$$

Από το νόμο των μεγάλων αριθμών γνωρίζουμε ότι $\bar{Y} \xrightarrow{n \rightarrow \infty} E(Y)$. Αλλά

$$E(Y) = 0P(Y=0) + 1P(Y=1) = P(Y=1) = P(X=2) = 0.0842$$

Άρα τελικά, όταν το n είναι μεγάλο (έχουμε βιβλίο με πολλές σελίδες), το ποσοστό $\bar{Y}$ των σελίδων που περιέχουν ακριβώς δύο λάθη θα είναι περίπου 8.42%. Όμοια υπολογίζεται και το ποσοστό των σελίδων που περιέχουν το πολύ 2 λάθη.

Παρατήρηση. (Η πιθανότητα ως οριακή σχετική συχνότητα) Με την ευκαιρία της παραπάνω άσκησης, αξίζει να αναφέρουμε ένα γενικότερο αποτέλεσμα το οποίο είναι αρκετά σημαντικό. Έστω A ένα ενδεχόμενο (με πιθανότητα $P(A)$) που αφορά κάποιο πείραμα και έστω ότι εκτελούμε το πείραμα αυτό n φορές (ανεξάρτητες μεταξύ τους). Θεωρούμε τις τ.μ. $Y_1, Y_2, \dots, Y_n$ έτσι ώστε $Y_i=1$ αν το ενδεχόμενο A εμφανιστεί στο i -πείραμα, ενώ $Y_i=0$ αν το ενδεχόμενο A δεν εμφανιστεί στο i -πείραμα, $i=1, 2, \dots, n$. Ο μέσος αριθμός εμφανίσεων του ενδεχομένου A (=σχετική συχνότητα εμφάνισης του A) σε αυτά τα n πειράματα θα είναι

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i.$$

Επικαλούμενοι και πάλι το νόμο των μεγάλων αριθμών προκύπτει ότι $\bar{Y} \xrightarrow{n \rightarrow \infty} E(Y)$. Αλλά

$$E(Y) = 0P(Y=0) + 1P(Y=1) = P(Y=1) = P(A).$$

Δηλαδή, ο μέσος αριθμός εμφανίσεων του ενδεχομένου A σε ένα μεγάλο πλήθος επαναλήψεων του πειράματος θα είναι οριακά ίσος με την πιθανότητα $P(A)$ του ενδεχομένου A . Το γεγονός αυτό, αν και συνήθως θεωρείται (λανθασμένα) αυτονόητο, αποτελεί συνέπεια του νόμου των μεγάλων αριθμών ο οποίος με τη σειρά του βασίζεται στα αξιώματα Kolmogorov. Άρα, ο κανόνας υπολογισμού της πιθανότητας ως οριακής σχετικής συχνότητας από τον Von Mises (βλ. Κεφάλαιο 1) αποτελεί και αυτός (όπως και ο κατά Laplace ορισμός της πιθανότητας, όπως είδαμε στο 1^ο Κεφάλαιο) πόρισμα των αξιωμάτων Kolmogorov.

Άσκηση 4.7. Η πιθανότητα να έχει κάποιος ατύχημα σε ένα επικίνδυνο σημείο της εθνικής οδού είναι 0.0001. Αν κατά τη διάρκεια του Σαββατοκύριακου διέρχονται από το σημείο αυτό 1000 αυτοκίνητα ποιά είναι η πιθανότητα να γίνουν δύο ή περισσότερα ατυχήματα; Να γίνει υπολογισμός χρησιμοποιώντας (α) την ακριβή κατανομή, και (β) την προσέγγιση με την κατανομή Poisson.

Λύση. Έστω X το πλήθος των ατυχημάτων στο σημείο αυτό της εθνικής οδού κατά τη διάρκεια του Σαββατοκύριακου. Ας θεωρήσουμε ως πείραμα την παρατήρηση ενός αυτοκινήτου που διέρχεται από το συγκεκριμένο σημείο το Σαββατοκύριακο με δύο δυνατά αποτελέσματα: «επιτυχία» (πραγματοποίηση ατυχήματος) με πιθ. 0.0001 και «αποτυχία» (μη πραγματοποίηση ατυχήματος). Εδώ μπορεί να θεωρηθεί ότι έχουμε 1000 τέτοια ανεξάρτητα μεταξύ τους πειράματα (διέλευση 1000 αυτοκινήτων) και άρα η τ.μ. X εκφράζει το πλήθος των επιτυχιών σε $n = 1000$ ανεξάρτητες δοκιμές σε κάθε μία από τις οποίες εμφανίζεται επιτυχία με πιθ. $p = 0.0001$ ή αποτυχία με πιθ. $q = 1 - p = 0.9999$. Από την Πρόταση 4.1. έπεται ότι $X \sim B(n, p)$ και άρα η πιθανότητα να γίνουν δύο ή περισσότερα ατυχήματα είναι

$$\begin{aligned} P(X \geq 2) &= 1 - P(X < 2) = 1 - P(X \leq 1) = 1 - P(X = 0) - P(X = 1) \\ &= 1 - \binom{n}{0} p^0 (1-p)^{1000} - \binom{n}{1} p^1 (1-p)^{1000-1} = 1 - (1-p)^{1000} - np(1-p)^{999} = \\ &= 1 - 0.9999^{1000} - 1000 \cdot 0.0001 \cdot 0.9999^{999} \approx 1 - 0.90483 - 0.09049 = 0.00468 \end{aligned}$$

Επίσης, επειδή το n είναι αρκετά μεγάλο και p αρκετά μικρό μπορούμε να θεωρήσουμε ότι η τ.μ. X εκφράζει το πλήθος των πραγματοποιήσεων ενός μεγάλου αριθμού «σπανίων» ενδεχομένων και (Πρόταση 4.3) προσεγγιστικά θα είναι $X \sim \text{Po}(\lambda)$ με $\lambda = np = 1000 \cdot 0.0001 = 0.1$. Άρα θα ισχύει προσεγγιστικά ότι,

$$P(X \geq 2) = 1 - P(X = 0) - P(X = 1) \approx 1 - e^{-0.1} \frac{0.1^0}{0!} - e^{-0.1} \frac{0.1^1}{1!} = 1 - e^{-0.1} - 0.1e^{-0.1} \approx 0.00467$$

το οποίο είναι προσεγγιστικά ίσο με το αποτέλεσμα που προέρχεται από τη χρήση της διωνυμικής.

4.5. Η Υπεργεωμετρική κατανομή.

Ας δούμε ένα τυπικό παράδειγμα όπου εμφανίζεται η υπεργεωμετρική κατανομή. Έστω μία κάλπη η οποία περιέχει A λευκές και $N-A$ μαύρες σφαίρες. Αν επιλέξουμε στην τύχη n σφαίρες (χωρίς επανάθεση) από την κάλπη αυτή και συμβολίσουμε με X των αριθμό των λευκών σφαιρών στο δείγμα των n σφαιρών, τότε (βλ. π.χ. ασκ. 1.22.) θα ισχύει ότι

$$P(X = x) = \frac{\binom{A}{x} \binom{N-A}{n-x}}{\binom{N}{n}}$$

Πράγματι, αν αριθμήσουμε τις σφαίρες ώστε $\{1, 2, \dots, A\}$ είναι οι λευκές σφαίρες και $\{A+1, \dots, N\}$ οι μαύρες, τότε το σύνολο Ω των δυνατών αποτελεσμάτων περιέχει όλες τις n -άδες $\{a_1, a_2, \dots, a_n\}$ που αποτελούνται από n διαφορετικά στοιχεία του $\{1, 2, \dots, N\}$ (δεν μας ενδιαφέρει η διάταξη ενώ πρέπει κάθε n -άδα να έχει διαφορετικά στοιχεία). Συνεπώς το Ω αποτελείται από όλους τους συνδυασμούς των N ανά n και άρα

$$|\Omega| = \binom{N}{n}.$$

Το ενδεχόμενο $[X=x] = \{\text{ακριβώς } x \text{ λευκές σφαίρες στο δείγμα των } n \text{ σφαιρών}\} \subseteq \Omega$ αποτελείται από όλες τις n -άδες του Ω με x αριθμούς από το $\{1, 2, \dots, A\}$ και $n-x$ αριθμούς από το $\{A+1, \dots, N\}$. Το πλήθος των x -άδων από το $\{1, 2, \dots, A\}$ είναι ίσο με το πλήθος των συνδυασμών των A ανά x . Επίσης, για κάθε μία από αυτές τις x -άδες, οι υπόλοιποι $n-x$ αριθμοί μπορούν να επιλεγούν κατά

$\binom{N-A}{n-x}$ τρόπους. Από την πολλαπλασιαστική αρχή, έπεται ότι το ενδεχόμενο $[X=x]$ θα αποτελείται από $\binom{A}{x} \binom{N-A}{n-x}$ στοιχεία και τελικά,

$$P(X=x) = \frac{|[X=x]|}{|\Omega|} = \frac{\binom{A}{x} \binom{N-A}{n-x}}{\binom{N}{n}}.$$

Παρατηρούμε ότι η $P(X=x)$ μπορεί να πάρει θετικές τιμές όταν το x είναι τέτοιο ώστε $0 \leq x \leq A$ και $0 \leq n-x \leq N-A$ (ώστε οι συνδυασμοί στον αριθμητή να μην είναι 0). Δηλαδή το x πρέπει να είναι ένας ακέραιος μεταξύ των $\max\{0, n+A-N\}$, και $\min\{A, n\}$. Γενικά θα έχουμε τον επόμενο ορισμό:

Ορισμός 4.5. Η κατανομή με σ.π.

$$f(x) = \frac{\binom{A}{x} \binom{N-A}{n-x}}{\binom{N}{n}}, \quad x = \max\{0, n+A-N\}, \dots, \min\{A, n\}$$

καλείται υπεργεωμετρική κατανομή με παραμέτρους A, N, n ($A, n < N$). (συμβ. και με $HG(A, N, n)$).

Παραπάνω αποδείξαμε και την ακόλουθη πρόταση:

Πρόταση 4.4. Αν μία τ.μ. X μπορεί να θεωρηθεί ότι εκφράζει το πλήθος των λευκών σφαιρών ανάμεσα σε n τυχαία επιλεγμένες σφαίρες (χωρίς επανάθεση) από μία κάλπη με A λευκές και $N-A$ μαύρες τότε η τ.μ. X θα ακολουθεί την υπεργεωμετρική κατανομή με παραμέτρους A, N, n .

Αν μία τ.μ. X ακολουθεί την υπεργεωμετρική κατανομή με παραμέτρους A, N, n τότε αποδεικνύεται ότι

$$E(X) = n \frac{A}{N} \quad \text{και} \quad V(X) = n \frac{A}{N} \left(1 - \frac{A}{N}\right) \frac{N-n}{N-1}.$$

Παρατήρηση. Αν η επιλογή των σφαιρών δεν γίνεται χωρίς επανάθεση αλλά με επανάθεση (εκλέγεται η 1^η σφαίρα, καταγράφεται το χρώμα της και επιστρέφει στην κάλπη, στη συνέχεια εκλέγεται η 2^η σφαίρα κ.ο.κ.) τότε η κατανομή του αριθμού X των σφαιρών δεν ακολουθεί υπεργεωμετρική κατανομή. Συγκεκριμένα, αν κάθε σφαίρα επιστρέφει στην κάλπη μετά την εκλογή της, τότε μπορούμε να θεωρήσουμε τις n εκλογές σφαιρών σαν n ανεξάρτητα και όμοια πειράματα. Αν ως επιτυχία θεωρήσουμε την επιλογή λευκής σφαίρας, τότε η τ.μ. X μπορεί να θεωρηθεί ότι εκφράζει το πλήθος των επιτυχιών σε n ανεξάρτητες δοκιμές με πιθανότητα επιτυχίας $p = P(\text{επιλογή λευκής σφαίρας}) = A/N$ και με πιθανότητα αποτυχίας $1-A/N$. Επομένως, αν η τ.μ. X εκφράζει το πλήθος των λευκών σφαιρών ανάμεσα σε n τυχαία επιλεγμένες σφαίρες με επανάθεση από μία κάλπη με A λευκές και $N-A$ μαύρες τότε η τ.μ. X θα ακολουθεί τη διωνυμική κατανομή με παραμέτρους n και $p = A/N$. Δηλαδή,

$$P(X=x) = \binom{n}{x} \left(\frac{A}{N}\right)^x \left(1 - \frac{A}{N}\right)^{n-x}, \quad x = 0, 1, 2, \dots, n$$

Είναι διαισθητικά φανερό (αλλά μπορεί να αποδειχθεί και αυστηρά) ότι όταν το πλήθος των σφαιρών N στην κάλπη είναι αρκετά μεγάλο, τότε η επιλογή n σφαιρών (με n σχετικά μικρό) χωρίς επανάθεση είναι περίπου «ισοδύναμη» με την επιλογή n σφαιρών με επανάθεση. Αυτό γίνεται φανερό από το γεγονός ότι, όταν το N είναι αρκετά μεγάλο, η επιστροφή ή όχι π.χ. της πρώτης

σφαίρας στην κάλπη ελάχιστα επηρεάζει την πιθανότητα επιλογής μαύρης ή λευκής σφαίρας στη δεύτερη επιλογή σφαίρας κ.ο.κ. Επομένως, θα ισχύει ότι για μεγάλο N ,

$$HG(A, N, n) \approx B(n, A/N).$$

Δηλαδή,

$$\frac{\binom{A}{x} \binom{N-A}{n-x}}{\binom{N}{n}} = \binom{n}{x} \frac{(A)_x (N-A)_{n-x}}{(N)_n} \approx \binom{n}{x} \frac{A^x (N-A)^{n-x}}{N^n} = \binom{n}{x} \left(\frac{A}{N}\right)^x \left(1 - \frac{A}{N}\right)^{n-x}$$

Η άσκηση 1.22 και οι συμπληρωματικές ασκήσεις 7, 9, 12 του πρώτου κεφαλαίου μπορούν αμεσότερα να λυθούν παρατηρώντας ότι οι πιθανότητες που ζητούνται σε αυτές έχουν άμεση σχέση με την υπεργεωμετρική κατανομή. Για παράδειγμα, ας λύσουμε ξανά την άσκηση 1.22:

Άσκηση 4.8. (Άσκηση 1.22) Σε μία κλήρωση Lotto τοποθετούνται στην κληρωτίδα 49 σφαίρες αριθμημένες από το 1 έως το 49 και εκλέγονται στην τύχη 6 αριθμοί που κερδίζουν. i) Ποια είναι η πιθανότητα να κληρωθεί μία συγκεκριμένη εξάδα που έχουμε προσημειώσει; ii) Ποια είναι η πιθανότητα να περιέχονται

α) και οι 6 αριθμοί β) 5 αριθμοί γ) 4 αριθμοί που κερδίζουν ανάμεσα σε 20 αριθμούς που έχουμε προσημειώσει;

Λύση. i) Η κληρωτίδα περιέχει 49 σφαίρες από τις οποίες έχουμε προσημειώσει τις 6. Ας θεωρήσουμε αυτές τις 6 σφαίρες ως λευκές. Δηλαδή $A=6$ και $N=49$. Από την κάλπη τυχαία επιλέγουμε (κληρώνονται) $n=6$ σφαίρες και ζητείται η πιθανότητα να κληρωθούν και οι 6 σφαίρες που έχουμε προσημειώσει. Αν X είναι το πλήθος των προσημειωμένων («λευκών») σφαιρών στο δείγμα μεγέθους 6 τότε ζητείται η πιθανότητα (η εκλογή των σφαιρών προφανώς γίνεται χωρίς επανάθεση)

$$P(X=6) = \frac{\binom{A}{x} \binom{N-A}{n-x}}{\binom{N}{n}} = \frac{\binom{6}{6} \binom{49-6}{6-6}}{\binom{49}{6}} = \frac{1}{\binom{49}{6}}.$$

ii) Όμοια με το (i), η κληρωτίδα περιέχει 49 σφαίρες από τις οποίες έχουμε προσημειώσει τις 20. Ας θεωρήσουμε και πάλι αυτές τις 20 σφαίρες ως λευκές ($A=20$, $N=49$). Από την κάλπη τυχαία επιλέγουμε (κληρώνονται) $n=6$ σφαίρες και ζητείται η πιθανότητα να κληρωθούν και οι 6, β) 5, γ) 4 αριθμοί από τους προσημειωμένους. Αν X είναι και πάλι το πλήθος των προσημειωμένων («λευκών») σφαιρών στο δείγμα μεγέθους 6 τότε ζητούνται οι πιθανότητες

$$P(X=6) = \frac{\binom{A}{x} \binom{N-A}{n-x}}{\binom{N}{n}} = \frac{\binom{20}{6} \binom{49-20}{6-6}}{\binom{49}{6}} = \frac{\binom{20}{6}}{\binom{49}{6}},$$

$$P(X=5) = \frac{\binom{A}{x} \binom{N-A}{n-x}}{\binom{N}{n}} = \frac{\binom{20}{5} \binom{49-20}{6-5}}{\binom{49}{6}} = \frac{\binom{20}{5} \binom{29}{1}}{\binom{49}{6}},$$

$$P(X=4) = \frac{\binom{A}{x} \binom{N-A}{n-x}}{\binom{N}{n}} = \frac{\binom{20}{4} \binom{49-20}{6-4}}{\binom{49}{6}} = \frac{\binom{20}{4} \binom{29}{2}}{\binom{49}{6}}.$$

Άσκηση 4.9. Από τους 100 εργαζόμενους σε μία επιχείρηση, οι 40 είναι γυναίκες. Έστω ότι για κάποια συγκεκριμένη εργασία επιλέγονται τυχαία 5 εργαζόμενοι. Να υπολογιστεί η πιθανότητα όπως μεταξύ των 5 οι 2 είναι γυναίκες. Ποιος θα είναι ο αναμενόμενος αριθμός γυναικών ανάμεσα στους 5; Η παραπάνω πιθανότητα να υπολογισθεί και με τη χρήση κατάλληλης προσέγγισης.

Λύση. Έστω X ο αριθμός των γυναικών ανάμεσα στους 5 τυχαία επιλεγμένους. Η τ.μ. X μπορεί ισοδύναμα να θεωρηθεί ότι εκφράζει το πλήθος των λευκών σφαιρών ανάμεσα σε $n = 5$ τυχαία επιλεγμένες σφαίρες (χωρίς επανάθεση) από μία κάλπη με $A = 40$ λευκές και $N-A = 60$ μαύρες σφαίρες. Από την Πρόταση 4.4., η τ.μ. X θα ακολουθεί υπεργεωμετρική κατανομή με παραμέτρους $A = 40$, $N = 100$, $n = 5$. Ζητείται η πιθανότητα

$$\begin{aligned} P(X=2) &= \frac{\binom{A}{x} \binom{N-A}{n-x}}{\binom{N}{n}} = \frac{\binom{40}{2} \binom{100-40}{5-2}}{\binom{100}{5}} = \frac{\binom{40}{2} \binom{60}{3}}{\binom{100}{5}} \\ &= \frac{40 \cdot 39}{2!} \cdot \frac{60 \cdot 59 \cdot 58}{3!} \cdot \frac{5!}{100 \cdot 99 \cdot 98 \cdot 97 \cdot 96} = \frac{780 \cdot 34220}{75287520} = 0.3545. \end{aligned}$$

Ο αναμενόμενος αριθμός γυναικών ανάμεσα στους 5 τυχαία επιλεγμένους θα είναι

$$E(X) = n \frac{A}{N} = 5 \frac{40}{100} = 2.$$

Τέλος, σύμφωνα με παραπάνω παρατήρηση, μπορούμε να προσεγγίσουμε την $HG(A, N, n)$ με την $B(n, A/N)$ (το $N=100$ είναι σχετικά μεγάλο σε σχέση με το $n = 5$). Ειδικότερα, θα ισχύει προσεγγιστικά ότι

$$P(X=2) \approx \binom{n}{x} \left(\frac{A}{N}\right)^x \left(1 - \frac{A}{N}\right)^{n-x} = \binom{5}{2} \left(\frac{40}{100}\right)^2 \left(1 - \frac{40}{100}\right)^{5-2} = 10 \cdot 0.4^2 \cdot 0.6^3 = 0.3456.$$

Ειδικές Συνεχείς Κατανομές

4.6. Η ομοιόμορφη συνεχής κατανομή.

Η ομοιόμορφη συνεχής κατανομή εμφανίζεται σε περιπτώσεις όπου η υπό εξέταση τ.μ. X παίρνει τιμές σε ένα διάστημα $[a, b] \subseteq \mathbb{R}$ ενώ πιθανότητες της μορφής $P(X \in [x_1, x_2])$ είναι ανάλογες του μήκους $x_2 - x_1$ του διαστήματος $[x_1, x_2]$ ($a \leq x_1 < x_2 \leq b$).

Ορισμός 4.6. Η συνεχής κατανομή με σ.π.π.

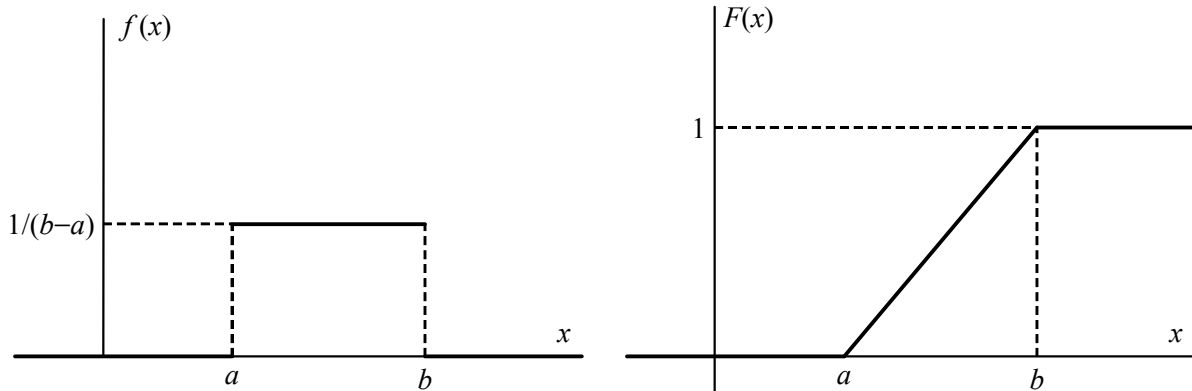
$$f(x) = \frac{1}{b-a}, \quad x \in [a, b]$$

($f(x) = 0$, $x \notin [a, b]$) καλείται συνεχής ομοιόμορφη κατανομή στο διάστημα $[a, b]$.

Η συνάρτηση κατανομής της συνεχούς ομοιόμορφης κατανομής θα είναι

$$F(x) = \int_{-\infty}^x f(t)dt = \int_a^x \frac{1}{b-a} dt = \frac{x-a}{b-a}, \quad x \in [a, b]$$

ενώ $F(x) = 0$ αν $x < a$, $F(x) = 1$ αν $x > b$.



Σύμφωνα με αυτή την κατανομή, η συνολική πιθανότητα έχει «απλωθεί» σε όλο το διάστημα $[a, b]$ με ομοιόμορφο τρόπο (βλέπουμε ότι η πυκνότητα της πιθανότητας είναι σταθερή σε όλο το $[a, b]$). Η μέση τιμή και διασπορά της συνεχούς ομοιόμορφης κατανομής θα είναι

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_a^b x \frac{1}{b-a} dx = \frac{1}{b-a} \left[\frac{x^2}{2} \right]_a^b = \frac{1}{b-a} \frac{b^2 - a^2}{2} = \frac{a+b}{2}$$

Όπως λοιπόν αναμενόταν το «κέντρο βάρους» της κατανομής αυτής είναι το μέσο του διαστήματος $[a, b]$. Για τη διασπορά θα είναι

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{b-a} \left[\frac{x^3}{3} \right]_a^b = \frac{1}{b-a} \frac{b^3 - a^3}{3} = \frac{a^2 + ab + b^2}{3}$$

και άρα

$$V(X) = E(X^2) - E(X)^2 = \frac{a^2 + ab + b^2}{3} - \frac{(a+b)^2}{4} = \frac{(a-b)^2}{12}.$$

Άσκηση 4.10 Ένας αριθμός X εκλέγεται τυχαία από το διάστημα $[0,1]$. Να υπολογιστούν οι πιθανότητες (α) το πρώτο δεκαδικό ψηφίο του αριθμού X να είναι το 2 ή το 5, και (β) Αν γνωρίζουμε ότι $X > 0.5$, ποια είναι η πιθανότητα τελικά να είναι $X < 0.6$;

Λύση. Η συνάρτηση κατανομής της X θα είναι

$$F(x) = \frac{x-0}{1-0} = x, \quad x \in [0,1]$$

ενώ $F(x)=0$ αν $x<0$, $F(x)=1$ αν $x>1$.

(α) Στην ουσία ζητείται η πιθανότητα

$$p = P(0.2 \leq X < 0.3 \text{ ή } 0.5 \leq X < 0.6) = P([0.2 \leq X < 0.3] \cup [0.5 \leq X < 0.6])$$

Πρόκειται για την πιθανότητα μιας ένωσης δύο ξένων ενδεχομένων. Επομένως,

$$\begin{aligned} p &= P(0.2 \leq X < 0.3) + P(0.5 \leq X < 0.6) = F(0.3) - F(0.2) + F(0.6) - F(0.5) \\ &= 0.3 - 0.2 + 0.6 - 0.5 = 0.2. \end{aligned}$$

(β) Θα είναι

$$P(X < 0.6 | X > 0.5) = \frac{P(0.5 < X < 0.6)}{P(X > 0.5)} = \frac{F(0.6) - F(0.5)}{1 - F(0.5)} = \frac{0.6 - 0.5}{1 - 0.5} = 0.2.$$

Άσκηση 4.11. Κάθε 40 λεπτά αναχωρεί τρένο από Αθήνα για Χαλκίδα. Κάποιος που δεν γνωρίζει την ώρα που αναχωρεί το τρένο εισέρχεται στο σταθμό για να ταξιδέψει με αυτό. Να υπολογιστεί ο μέσος χρόνος που χρειάζεται να περιμένει έως ότου αναχωρήσει το τρένο. Αν περιμένει ήδη 20 λεπτά χωρίς να έχει αναχωρήσει τρένο, ποια είναι η πιθανότητα να περιμένει ακόμη περισσότερο από 10 λεπτά;

Λύση. Έστω X ο χρόνος μεταξύ της άφιξης του ταξιδιώτη A και της επόμενης αναχώρησης τρένου. Η X είναι συνεχής τ.μ. με τιμές στο διάστημα $[0 \text{ min}, 40 \text{ min}]$. Ο ταξιδιώτης A εισέρχεται στο σταθμό σε τυχαίο χρόνο και άρα η πιθανότητα άφιξης του ταξιδιώτη σε κάθε απειροστό διάστημα χρόνου μεταξύ των δύο αναχωρήσεων τρένου είναι η ίδια (π.χ. η πιθανότητα άφιξης του A σε οποιοδήποτε δευτερόλεπτο μεταξύ δύο αναχωρήσεων είναι η ίδια). Συνεπώς, η τ.μ. X θα ακολουθεί την ομοιόμορφη στο $[0, 40]$ κατανομή και άρα ο μέσος χρόνος αναμονής έως ότου αναχωρήσει το τρένο θα είναι

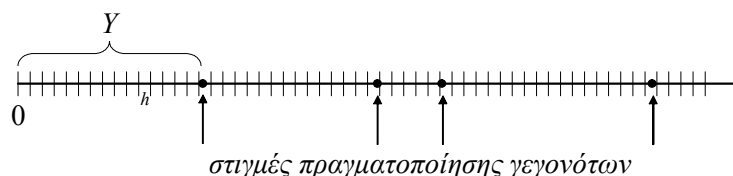
$$E(X) = \frac{0 + 40}{2} = 20 \text{ min}$$

Αν ο A περιμένει ήδη 20 λεπτά χωρίς να έχει αναχωρήσει τρένο, η πιθανότητα να περιμένει ακόμη περισσότερο από 10 λεπτά θα είναι

$$P(X > 30 | X > 20) = \frac{P(X > 30, X > 20)}{P(X > 20)} = \frac{P(X > 30)}{P(X > 20)} = \frac{1 - 30/40}{1 - 20/40} = 0.5 = 50\%.$$

4.7. Η εκθετική κατανομή.

Η εκθετική κατανομή εμφανίζεται συνήθως σε περιπτώσεις όπου μελετάμε *το χρόνο αναμονής μέχρι την πραγματοποίηση ενός γεγονότος*. Ας δούμε πως εμφανίζεται αυτή η κατανομή μέσα από ένα παράδειγμα. Έστω ότι, μία ακολουθία από γεγονότα πραγματοποιούνται σε *τυχαίες* στιγμές στο χρόνο και *ανεξάρτητα* το ένα από το άλλο (π.χ. αφίξεις πελατών σε ένα κατάστημα). Ειδικότερα, υποθέτουμε ότι σε κάθε *πολύ μικρό* χρονικό διάστημα μήκους h , η πραγματοποίηση ενός γεγονότος συμβαίνει με πιθανότητα *περίπου* λh ενώ είναι ανεξάρτητη από το τι έχει συμβεί στα υπόλοιπα διαστήματα (π.χ. έχουμε σταθερό «ρυθμό» άφιξης πελατών, δηλαδή η πιθανότητα άφιξης ενός πελάτη κατά τη διάρκεια οποιουδήποτε δευτερολέπτου είναι σταθερή και ίση π.χ. με 0.001 ενώ επίσης οι πελάτες προσέρχονται ανεξάρτητα ο ένας από τον άλλον). Έστω Y ο χρόνος αναμονής μέχρι την εμφάνιση του πρώτου γεγονότος (π.χ. του πρώτου πελάτη). Σχηματικά:



Το παραπάνω μοντέλο είναι αρκετά σύνηθες στις εφαρμογές και συνεπώς είναι χρήσιμο να γνωρίζουμε, κάτω από τις παραπάνω συνθήκες, την κατανομή του χρόνου αναμονής. Έστω λοιπόν ότι αναζητούμε την $P(Y \leq t)$, $t > 0$. Χωρίζοντας το χρονικό διάστημα $(0, t]$ σε n μικρά διαστήματα μήκους t/n το καθένα παρατηρούμε ότι

$$\begin{aligned} P(Y > t) &= P(\text{σε κανένα από τα } n \text{ διαστήματα δεν συνέβη το γεγονός}) \\ &= P(\text{δεν συνέβη στο } 1^\circ \text{ διάστημα}) \times \dots \times P(\text{δεν συνέβη στο } n\text{-οστό διάστημα}) \end{aligned}$$

$$\approx \left(1 - \lambda \frac{t}{n}\right) \cdot \dots \cdot \left(1 - \lambda \frac{t}{n}\right) = \left(1 - \lambda \frac{t}{n}\right)^n$$

Αν χωρίσουμε το διάστημα σε πάρα πολλά απειροστά διαστήματα (δηλ. $n \rightarrow \infty$), τότε προκύπτει τελικά ότι

$$P(Y > t) = \lim_{n \rightarrow \infty} \left(1 - \lambda \frac{t}{n}\right)^n = e^{-\lambda t}$$

και επομένως η συνάρτηση κατανομής και συνάρτηση πυκνότητας πιθανότητας του χρόνου Y α-ναμονής μέχρι την εμφάνιση του πρώτου γεγονότος (π.χ. πελάτη) θα είναι

$$F_Y(x) = 1 - P(Y > x) = 1 - e^{-\lambda x}, \quad x > 0, \quad f_Y(x) = \frac{d}{dx} F_Y(x) = \frac{d}{dx} (1 - e^{-\lambda x}) = \lambda e^{-\lambda x}, \quad x > 0$$

Επομένως, καταλήγουμε στον επόμενο ορισμό.

Ορισμός 4.7. Η συνεχής κατανομή με σ.π.π.

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0 \quad (\lambda > 0)$$

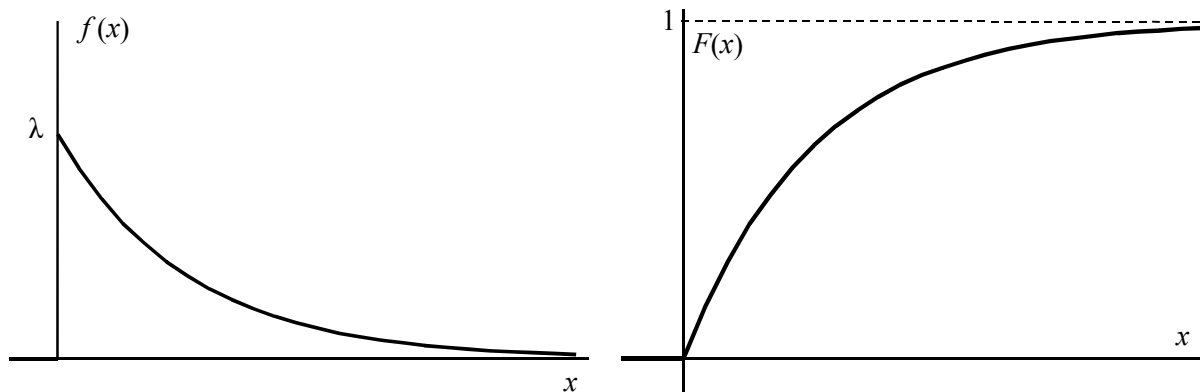
($f(x) = 0, x < 0$) καλείται εκθετική κατανομή με παράμετρο λ .

Η συνάρτηση κατανομής της εκθετικής κατανομής θα είναι

$$F(x) = \int_{-\infty}^x f(t) dt = \int_0^x \lambda e^{-\lambda t} dt = \lambda \left[\frac{e^{-\lambda t}}{-\lambda} \right]_0^x = 1 - e^{-\lambda x}, \quad x \geq 0$$

ενώ $F(x) = 0$ αν $x < 0$. Η μέση τιμή και διασπορά της εκθετικής κατανομής από τις ασκήσεις 3.7, 3.9.γ θα είναι

$$E(X) = \frac{1}{\lambda} \quad \text{και} \quad V(X) = \frac{1}{\lambda^2}.$$



Άσκηση 4.12. Ο χρόνος ζωής (σε ώρες) μιας ηλεκτρικής συσκευής ακολουθεί την εκθετική κατανομή με παράμετρο $\lambda > 0$. Η εταιρεία πουλάει τη συσκευή με καθαρό κέρδος k δρχ και δίνει στους πελάτες της εγγύηση α ωρών λειτουργίας. Σε περίπτωση που η συσκευή παρουσιάσει βλάβη πριν τη λήξη της εγγύησης, επισκευάζεται δωρεάν από την εταιρεία η οποία και επιβαρύνεται με κόστος επισκευής k_0 δρχ. Αν Y είναι το κέρδος της εταιρείας ανά συσκευή, να υπολογιστεί η σ.π. της Y και ναδειχθεί ότι το μέσο κέρδος δίνεται από τον τύπο

$$E(Y) = (k - k_0) + k_0 e^{-\lambda \alpha}.$$

Εφαρμογή. Έστω ότι $\lambda = 1/50000$, $k = 20000$ δρχ., $k_0 = 10000$ δρχ. Αν η εταιρεία θέλει να έχει μέσο κέρδος 15000 δρχ (ανά συσκευή), ποιος είναι ο χρόνος εγγύησης που θα πρέπει να δίνει στους αγοραστές της συσκευής;

Λύση. Έστω X ο χρόνος ζωής της ηλεκτρικής συσκευής. Το κέρδος της εταιρείας ανά συσκευή θα είναι

$$Y = \begin{cases} k, & X > a \\ k - k_0, & X \leq a \end{cases}$$

και επομένως η Y είναι μία διακριτή τ.μ. που παίρνει δύο τιμές, k , $k - k_0$. Η συνάρτηση πιθανότητας της τ.μ. Y θα είναι

$$P(Y = k) = P(X > a) = 1 - F_X(a) = e^{-\lambda a}, \quad P(Y = k - k_0) = P(X \leq a) = F_X(a) = 1 - e^{-\lambda a}$$

και

$$E(Y) = kP(Y = k) + (k - k_0)P(Y = k - k_0) = ke^{-\lambda a} + (k - k_0)(1 - e^{-\lambda a}) = (k - k_0) + k_0e^{-\lambda a}$$

Αν a_0 ο χρόνος εγγύησης που θα πρέπει η εταιρία να δίνει στους αγοραστές της συσκευής, θα πρέπει

$$E(Y) = (k - k_0) + k_0e^{-\lambda a_0} = c$$

όπου $c = 15000$ και άρα ισοδύναμα,

$$a_0 = \frac{1}{\lambda} \ln \frac{k_0}{c - (k - k_0)} = 50000 \ln \frac{10000}{15000 - (20000 - 10000)} \approx 34657.4 \approx 1444 \text{ ημέρες.}$$

Άσκηση 4.13. Ας υποθέσουμε ότι η διάρκεια σε λεπτά των υπεραστικών τηλεφωνικών συνδιαλέξεων ακολουθεί την εκθετική κατανομή με μέση τιμή 2 λεπτά. Να βρεθούν οι πιθανότητες η διάρκεια μιας υπεραστικής συνδιάλεξης (α) να υπερβεί τα 6 λεπτά, (β) να είναι μεταξύ 4 και 6 λεπτών (γ) να είναι μικρότερη από 4 λεπτά, και (δ) να είναι μικρότερη από 6 λεπτά δεδομένου ότι ήταν μεγαλύτερη από 4 λεπτά.

Λύση. Αν X είναι η διάρκεια σε λεπτά των υπεραστικών τηλεφωνικών συνδιαλέξεων τότε από την εκφώνηση θα ισχύει ότι

$$E(X) = \frac{1}{\lambda} = 2$$

και άρα,

$$F_X(x) = P(X \leq x) = 1 - e^{-\lambda x} = 1 - e^{-x/2}, \quad x \geq 0.$$

(α) Η πιθανότητα ώστε η διάρκεια μιας υπεραστικής συνδιάλεξης να υπερβεί τα 6 λεπτά είναι

$$P(X > 6) = 1 - F_X(6) = e^{-6/2} \approx 0.0498.$$

Αντίστοιχα,

$$(\beta) \quad P(4 < X < 6) = F_X(6) - F_X(4) = 1 - e^{-6/2} - 1 + e^{-4/2} = e^{-4/2} - e^{-6/2} \approx 0.0855,$$

$$(\gamma) \quad P(X < 6) = F_X(6) = 1 - e^{-6/2} \approx 0.8646,$$

$$(\gamma) \quad P(X < 6 | X > 4) = \frac{P(4 < X < 6)}{P(X > 4)} = \frac{F_X(6) - F_X(4)}{1 - F_X(4)} = \frac{e^{-4/2} - e^{-6/2}}{e^{-4/2}} \approx 0.6321.$$

(Περισσότερες και σημαντικότερες συνεχείς κατανομές θα μελετηθούν στα πλαίσια του μαθήματος Στατιστική ΙΙΙ)