ΚΕΦΑΛΑΙΟ 5

ΔΕΣΜΕΥΜΕΝΕΣ ΚΑΤΑΝΟΜΕΣ ΚΑΙ ΔΙΑΤΕΤΑΓΜΕΝΕΣ ΣΤΑΤΙΣΤΙΚΕΣ

Σε αυτό το κεφάλαιο θα αναπτύξουμε τη θεωρία των δεσμευμένων κατανομών. Επιπλέον, θα μελετήσουμε τις διατεταγμένες στατιστικές και θα ασχοληθούμε με την εύρεση της κατανομής συναρτήσεων τυχαίων μεταβλητών.

5.1 Δεσμευμένη συνάρτηση πιθανότητας και δεσμευμένη συνάρτηση πυκνότητας

Έστω (X,Y) μία διδιάστατη διακριτή διανυσματική τυχαία μεταβλητή και $p(x,y)$ η από κοινού συνάρτηση πιθανότητάς της. Για κάθε δυνατή τιμή y της Y τέτοια ώστε $p_Y(y)=P(Y=y)\neq 0,$ η **δεσμευμένη συνάρτηση πιθανότητας (δ.σ.π.)** (conditional probability function) της X δοθέντος ότι $Y=y,$ ορίζεται ως εξής:

$$p_{X|Y}(x\mid y):=P(X=x\mid Y=y)=\frac{P(X=x,Y=y)}{P(Y=y)}=\frac{p(x,y)}{p_Y(y)},$$

όπου $p_Y(y)$ είναι η συνάρτηση πιθανότητας της τυχαίας μεταβλητής $Y.$

Παρατηρούμε ότι, αν οι τυχαίες μεταβλητές X,Y είναι ανεξάρτητες, τότε η δ.σ.π. της X δοθέντος ότι $Y=y$ συμπίπτει με τη (μη δεσμευμένη) συνάρτηση πιθανότητας $p_X(x)$ της $X,$ διότι

$$p_{X|Y}(x\mid y)=\frac{P(X=x,Y=y)}{P(Y=y)}=\frac{P(X=x)P(Y=y)}{P(Y=y)}=p_X(x).$$

Η δεύτερη ισότητα είναι συνέπεια της ανεξαρτησίας των τυχαίων μεταβλητών X και $Y.$

Για κάθε συγκεκριμένη τιμή $y\in S_Y,$ όπου S_Y είναι ο φορέας της $Y,$ ορίζεται μία συνάρτηση πιθανότητας $g(x)=p_{X|Y}(x\mid y)$ στην οποία αντιστοιχεί μία αθροιστική συνάρτηση κατανομής $G(x)=\sum_{t\le x}g(t).$

Η συνάρτηση $G(x)$ καλείται **δεσμευμένη αθροιστική συνάρτηση κατανομής** (conditional distribution function) της X δοθέντος ότι $Y=y.$

Μπορούμε να γράψουμε $G(x)=P_{X|Y}(x\mid y)=P(X\le x\mid Y=y)=\sum_{t\le x}p_{X|Y}(t\mid y).$

Έστω (X,Y) μία διδιάστατη συνεχής διανυσματική τυχαία μεταβλητή και $f(x,y)$ η από κοινού συνάρτηση πυκνότητάς της. Για μία συγκεκριμένη τιμή y τέτοια ώστε $f_Y(y)\neq 0,$ η **δεσμευμένη συνάρτηση πυκνότητας (δ.σ.π.)** (conditional density function) της X δοθέντος ότι $Y=y,$ ορίζεται ως εξής: $f_{X|Y}(x\mid y):=\dfrac{f(x,y)}{f_Y(y)},$

όπου $f_Y(y)$ είναι η συνάρτηση πυκνότητας της τυχαίας μεταβλητής $Y.$

Παρατηρούμε ότι, αν οι τυχαίες μεταβλητές X,Y είναι ανεξάρτητες, τότε η δ.σ.π. της X δοθέντος ότι $Y=y$ συμπίπτει με τη (μη δεσμευμένη) συνάρτηση πυκνότητας $f_X(x)$ της X, διότι

$$f_{X|Y}(x\mid y)=\frac{f(x,y)}{f_Y(y)}=\frac{f_X(x)f_Y(y)}{f_Y(y)}=f_X(x).$$

Η δεύτερη ισότητα είναι συνέπεια της ανεξαρτησίας των τυχαίων μεταβλητών X και Y.

Για κάθε συγκεκριμένη τιμή $y\in S_Y$ τέτοια ώστε $f_Y(y)\neq 0$ ορίζεται μία συνάρτηση πυκνότητας $g(x)=f_{X|Y}(x\mid y)$ στην οποία αντιστοιχεί μία αθροιστική συνάρτηση κατανομής $G(x)=\int_{-\infty}^{x}g(t)dt.$

Η συνάρτηση $G(x)$ καλείται **δεσμευμένη αθροιστική συνάρτηση κατανομής** της X δοθέντος ότι $Y=y$.

Μπορούμε να γράψουμε $G(x)=F_{X|Y}(x\mid y)=P(X\le x\mid Y=y)=\int_{-\infty}^{x}f_{X|Y}(t\mid y)dt.$

Παράδειγμα 5.1 Υποθέτουμε ότι η από κοινού συνάρτηση πιθανότητας της διδιάστατης διακριτής δ.τ.μ. (X,Y) δίνεται από τις σχέσεις: $p(0,0)=0.4,\quad p(0,1)=0.2,\quad p(1,0)=0.1,\quad p(1,1)=0.3.$ Υπολογίστε τη δεσμευμένη συνάρτηση πιθανότητας της X, δοθέντος ότι $Y=1$.

Λύση. Ισχύει ότι $p_Y(1)=P(Y=1)=\sum_{x=0}^{1}p(x,1)=0.2+0.3=0.5.$ Επομένως, $p_{X|Y}(0\mid 1)=\dfrac{p(0,1)}{p_Y(1)}=\dfrac{2}{5}$ και

$p_{X|Y}(1\mid 1)=\dfrac{p(1,1)}{p_Y(1)}=\dfrac{3}{5}.$

Παράδειγμα 5.2 Έστω ότι $X\sim \text{Poisson}(\lambda_1)$ και $Y\sim\text{Poisson}(\lambda_2)$. Υποθέτουμε ότι οι τυχαίες μεταβλητές X,Y είναι ανεξάρτητες. Αν $Z:=X+Y$, δείξτε ότι η δεσμευμένη κατανομή της X δοθέντος ότι $Z=X+Y=z$ είναι διωνυμική με παραμέτρους z και $p:=\dfrac{\lambda_1}{\lambda_1+\lambda_2}$,

δηλαδή $X\mid Z=X+Y=z\sim Bin(z,p:=\dfrac{\lambda_1}{\lambda_1+\lambda_2})$ και $p_{X|Z}(x\mid z)=\dbinom{z}{x}p^x(1-p)^{z-x},\ \ x=0,1,\ldots,z.$

Λύση. Από το Παράδειγμα 4.13 έχουμε ότι $Z=X+Y\sim\text{Poisson}(\lambda_1+\lambda_2)$. Ισχύει ότι

$$p_{X|Z}(x\mid z)=\frac{P(X=x,Z=z)}{P(Z=z)}=\frac{P(X=x,X+Y=z)}{P(Z=z)}=\frac{P(X=x,Y=z-x)}{P(Z=z)}=\frac{P(X=x)P(Y=z-x)}{P(Z=z)}$$

$$= \frac{\frac{e^{-\lambda_1}\lambda_1^x}{x!}\cdot\frac{e^{-\lambda_2}\lambda_2^{z-x}}{(z-x)!}}{\frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^z}{z!}} = \binom{z}{x}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^x\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{z-x} = \binom{z}{x}p^x(1-p)^{z-x}.$$

Η τέταρτη ισότητα είναι συνέπεια της ανεξαρτησίας των τυχαίων μεταβλητών X και Y.

Παράδειγμα 5.3 Η από κοινού συνάρτηση πυκνότητας της διδιάστατης συνεχούς δ.τ.μ. (X,Y) δίνεται από τον τύπο $f(x,y)=\frac{12}{5}x(2-x-y),$ αν $0<x,y<1$ και $f(x,y)=0,$ διαφορετικά. Υπολογίστε τη δεσμευμένη συνάρτηση πυκνότητας της X δοθέντος ότι $Y=y,$ όπου $0<y<1.$

Λύση. Για $0<x<1,$ $0<y<1,$ έχουμε

$$f_{X|Y}(x\mid y)=\frac{f(x,y)}{f_Y(y)}=\frac{f(x,y)}{\int_{-\infty}^{\infty}f(x,y)dx}=\frac{x(2-x-y)}{\int_0^1 x(2-x-y)dx}=\frac{x(2-x-y)}{\frac{2}{3}-\frac{y}{2}}=\frac{6x(2-x-y)}{4-3y}.$$

5.2 Δεσμευμένη μέση τιμή

Αν X και Y είναι δύο **διακριτές τυχαίες μεταβλητές**, τότε για κάθε δυνατή τιμή y της $Y,$ τέτοια ώστε $p_Y(y)=P(Y=y)\neq 0,$ η **δεσμευμένη μέση τιμή** (conditional expectation) της $X,$ δοθέντος ότι $Y=y,$ ορίζεται ως εξής:

$$E[X\mid Y=y]=\sum_{x\in S_X}xP[X=x\mid Y=y]=\sum_{x\in S_X}xp_{X|Y}(x\mid y),$$

όπου S_X είναι το σύνολο των δυνατών τιμών (φορέας) της X και $p_{X|Y}(x\mid y)$ είναι η δεσμευμένη συνάρτηση πιθανότητας της $X,$ δοθέντος ότι $Y=y.$

Αν X και Y είναι δύο **συνεχείς τυχαίες μεταβλητές**, τότε για κάθε δυνατή τιμή y της $Y,$ τέτοια ώστε $f_Y(y)\neq 0,$ η **δεσμευμένη μέση τιμή** της $X,$ δοθέντος ότι $Y=y,$ ορίζεται ως εξής:

$$E[X\mid Y=y]=\int_{-\infty}^{\infty}xf_{X|Y}(x\mid y)dx,$$

όπου $f_{X|Y}(x\mid y)$ είναι η δεσμευμένη συνάρτηση πυκνότητας της X δοθέντος ότι $Y=y.$

Παράδειγμα 5.4 Η από κοινού συνάρτηση πιθανότητας μιας διδιάστατης διακριτής διανυσματικής τυχαίας μεταβλητής (X,Y) δίνεται από τον τύπο $p(x,y)=\frac{1}{15}(x+y),$ $x=0,1,2$ και $y=1,2.$ Να υπολογιστούν οι δεσμευμένες μέσες τιμές $E[X\mid Y=1],$ $E[Y\mid X=1]$ και να συγκριθούν μεταξύ τους.

Λύση. Είναι $E[X \mid Y=1]=\sum_{x=0}^{2} x p_{X|Y}(x \mid 1)$. Όμως, $p_{X|Y}(x \mid y)=\dfrac{p(x,y)}{p_Y(y)}$ και $p_Y(y)=\sum_{x=0}^{2} p(x,y)=\dfrac{1}{5}y+\dfrac{1}{5}$.

Άρα, $p_{X|Y}(x \mid y)=\dfrac{1}{3}\dfrac{(x+y)}{(y+1)}$ και $p_{X|Y}(x \mid 1)=\dfrac{p(x,1)}{p_Y(1)}=\dfrac{x+1}{6}$. Είναι $E[X \mid Y=1]=\sum_{x=0}^{2} x\left(\dfrac{x+1}{6}\right)=\dfrac{8}{6}$.

Είναι $E[Y \mid X=1]=\sum_{y=1}^{2} y p_{Y|X}(y \mid 1)$. Όμως, $p_{Y|X}(y \mid x)=\dfrac{p(x,y)}{p_X(x)}=\dfrac{x+y}{2x+3}$

και $p_X(x)=\sum_{y=1}^{2} p(x,y)=\dfrac{1}{15}(2x+3)$. Άρα, $p_{Y|X}(y \mid 1)=\dfrac{1+y}{5}$.

Είναι $E[Y \mid X=1]=\sum_{y=1}^{2} y p_{Y|X}(y \mid 1)=\sum_{y=1}^{2}\dfrac{y(1+y)}{5}=\dfrac{8}{5}$. Επομένως, $E[Y \mid X=1]>E[X \mid Y=1]$.

Παράδειγμα 5.5 Η από κοινού συνάρτηση πυκνότητας της διδιάστατης συνεχούς δ.τ.μ. (X,Y) δίνεται από τον τύπο $f(x,y)=\dfrac{e^{-\frac{x}{y}}e^{-y}}{y}$, $0<x,y<\infty$ και $f(x,y)=0$, διαφορετικά. Υπολογίστε τη δεσμευμένη μέση τιμή $E(X \mid Y=y)$, όπου $y>0$.

Λύση. Αρχικά, θα υπολογίσουμε τη δεσμευμένη συνάρτηση πυκνότητας της X, δοθέντος ότι $Y=y$, όπου $y>0$. Διαδοχικά έχουμε

$$f_{X|Y}(x \mid y)=\frac{f(x,y)}{f_Y(y)}=\frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y)dx}=\frac{\dfrac{e^{-\frac{x}{y}}e^{-y}}{y}}{e^{-y}\int_0^{\infty}\left(\dfrac{1}{y}\right)e^{-\frac{x}{y}}dx}=\frac{\dfrac{1}{y}e^{-\frac{x}{y}}}{\int_0^{\infty}\dfrac{1}{y}e^{-\frac{x}{y}}dx}=\frac{\left(\dfrac{1}{y}\right)e^{-\frac{x}{y}}}{[-e^{-\frac{x}{y}}]_0^{\infty}}=\frac{1}{y}e^{-\frac{x}{y}},\ x>0.$$

Παρατηρούμε ότι $X \mid Y=y \sim$ Εκθετική$\left(\dfrac{1}{y}\right)$. Επομένως, $E(X \mid Y=y)=\int_0^{\infty}\dfrac{x}{y}e^{-\frac{x}{y}}dx=y$.

Παράδειγμα 5.6 Η από κοινού συνάρτηση πυκνότητας της διδιάστατης συνεχούς διανυσματικής τυχαίας μεταβλητής (X,Y) δίνεται από τον τύπο $f(x,y)=4xy$, $0<x,y<1$ και $f(x,y)=0$, αλλού.

(α) Να υπολογιστούν οι συναρτήσεις $f_X(x)$, $f_Y(y)$, $f_{X|Y}(x \mid y)$ και $f_{Y|X}(y \mid x)$.

(β) Να υπολογιστούν οι ποσότητες $E\left[X \mid Y=\dfrac{1}{2}\right]$ και $E\left[Y \mid X=\dfrac{1}{3}\right]$.

Λύση. (α) Είναι $f_X(x)=\int_{-\infty}^{\infty} f(x,y)dy=\int_0^1 4xydy=4x\left[\dfrac{y^2}{2}\right]_0^1=2x,\ 0<x<1$.

$$f_Y(y)=\int_{-\infty}^{\infty} f(x,y)dx=\int_0^1 4xydx=4y\left[\frac{x^2}{2}\right]_0^1=2y,\ \ 0<y<1.$$

$$f_{X|Y}(x\,|\,y)=\frac{f(x,y)}{f_Y(y)}=\frac{4xy}{2y}=2x,\ \ 0<x<1.$$

$$f_{Y|X}(y\,|\,x)=\frac{f(x,y)}{f_X(x)}=\frac{4xy}{2x}=2y,\ \ 0<y<1.$$

(β) Είναι $E\left[X\,|\,Y=\frac{1}{2}\right]=\int_{-\infty}^{\infty} xf_{X|Y}(x\,|\,y)dx=\int_0^1 2x^2dx=\frac{2}{3}.$

$$E\left[Y\,|\,X=\frac{1}{3}\right]=\int_{-\infty}^{\infty} yf_{Y|X}(y\,|\,x)dy=\int_0^1 2y^2dy=\frac{2}{3}.$$

Για την εύρεση της δεσμευμένης μέσης τιμής μιας συνάρτησης $g(X)$ της τυχαίας μεταβλητής X, δοθέντος ότι $Y=y$, ισχύουν οι εξής τύποι: $E[g(X)\,|\,Y=y]=\sum_x g(x)p_{X|Y}(x\,|\,y)$, αν οι X και Y είναι διακριτές τυχαίες μεταβλητές και $E[g(X)\,|\,Y=y]=\int_{-\infty}^{\infty} g(x)f_{X|Y}(x\,|\,y)dx$, αν οι X και Y είναι συνεχείς τυχαίες μεταβλητές.

Αν θέσουμε $E[X]=\mu$ και θεωρήσουμε τη συνάρτηση $h(x)=(x-\mu)^2$, $x\in\Re$, τότε, από τους παραπάνω τύπους, η **δεσμευμένη διακύμανση** (conditional variance) της X δοθέντος ότι $Y=y$, είναι $Var[X\,|\,Y=y]=E[h(X)\,|\,Y=y]$ και δίνεται από τον τύπο $Var[X\,|\,Y=y]=E[X^2\,|\,Y=y]-(E[X\,|\,Y=y])^2$.

Συμβολίζουμε με $E[X\,|\,Y]$ τη συνάρτηση της τυχαίας μεταβλητής Y, της οποίας η τιμή όταν $Y=y$ είναι $E[X\,|\,Y=y]$. Παρατηρούμε ότι η ποσότητα $E[X\,|\,Y]$ είναι επίσης μία τυχαία μεταβλητή. Για τη συνάρτηση $Var(X\,|\,Y)$ ισχύει η σχέση $Var(X\,|\,Y)=E(X^2\,|\,Y)-(E(X\,|\,Y))^2$. Η παρακάτω πρόταση είναι σημαντικότατη και έχει πολλές εφαρμογές.

Πρόταση 5.1 Έστω δύο τυχαίες μεταβλητές X,Y. Ισχύει ότι $E[X]=E[E[X\,|\,Y]]$. (5.1)

(α) Αν η Y είναι μία διακριτή τυχαία μεταβλητή τότε η σχέση (5.1) γράφεται ως εξής:

$$E[X]=\sum_y E[X\,|\,Y=y]P[Y=y].$$

(β) Αν η Y είναι μία συνεχής τυχαία μεταβλητή με συνάρτηση πυκνότητας f_Y τότε η σχέση (5.1) γράφεται ως εξής: $E[X]=\int_{-\infty}^{\infty} E[X\,|\,Y=y]f_Y(y)dy.$

Στη συνέχεια θα αποδείξουμε την Πρόταση 5.1 για την περίπτωση κατά την οποία και οι δύο τυχαίες μεταβλητές X,Y είναι διακριτές.

Απόδειξη. Διαδοχικά, έχουμε

$$E[E[X\mid Y]]=\sum_y E[X\mid Y=y]P[Y=y]=\sum_y\left[\sum_x xP[X=x\mid Y=y]\right]P[Y=y]$$

$$=\sum_y\sum_x x\frac{P(X=x,Y=y)}{P(Y=y)}P(Y=y)=\sum_x\sum_y xP(X=x,Y=y)=\sum_x x\sum_y P(X=x,Y=y)$$

$$=\sum_x xP(X=x)=E[X].\ \blacksquare$$

Η Πρόταση 5.1 ισχύει και όταν μία από τις τυχαίες μεταβλητές X,Y είναι διακριτή και η άλλη είναι συνεχής.

Παράδειγμα 5.7 Αν A είναι ένα ενδεχόμενο και Y είναι μία τυχαία μεταβλητή, τότε

$P(A)=\sum_y P(A\mid Y=y)P(Y=y),$ αν η Y είναι μία διακριτή τυχαία μεταβλητή και

$P(A)=\int_{-\infty}^{\infty}P(A\mid Y=y)f_Y(y)dy,$ αν η Y είναι μία συνεχής τυχαία μεταβλητή με σ.π. f_Y.

Λύση. Έστω η τυχαία μεταβλητή X τέτοια ώστε $X=1,$ αν το ενδεχόμενο A συμβαίνει και $X=0,$ αν το ενδεχόμενο A δεν συμβαίνει. Ισχύει ότι $P(A)=E[X]$ και $P(A\mid Y=y)=E[X\mid Y=y].$ Εφαρμόζοντας την Πρόταση 5.1 έχουμε

$$P(A)=E[X]=E[E[X\mid Y]]=\begin{cases}\sum_y E[X\mid Y=y]P(Y=y)\\ \int_{-\infty}^{\infty}E[X\mid Y=y]f_Y(y)dy\end{cases}=\begin{cases}\sum_y P(A\mid Y=y)P(Y=y)\\ \int_{-\infty}^{\infty}P(A\mid Y=y)f_Y(y)\end{cases}.$$

Το άνω σκέλος των τελευταίων δύο ισοτήτων αντιστοιχεί στην περίπτωση που η Y είναι διακριτή και το κάτω σκέλος αντιστοιχεί στην περίπτωση που η Y είναι συνεχής με σ.π. f_Y.

Παράδειγμα 5.8 Οι τυχαίες μεταβλητές $X_1,\ldots,X_N$ ακολουθούν τη διωνυμική κατανομή με παραμέτρους $m,$ p και η τυχαία μεταβλητή N είναι ανεξάρτητη των $X_1,\ldots,X_N$ και ακολουθεί την κατανομή Poisson με παράμετρο $\lambda.$ Υπολογίστε τη μέση τιμή $E\left(\sum_{i=1}^{N}X_i\right).$

Λύση. Ισχύει ότι $E\left(\sum_{i=1}^{N}X_i\mid N=n\right)=E\left(\sum_{i=1}^{n}X_i\mid N=n\right)=E\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}E[X_i]=nmp.$

Η δεύτερη ισότητα είναι συνέπεια της ανεξαρτησίας των τυχαίων μεταβλητών N και $X_i, i=1,\dots,n.$ Άρα, $E\left(\sum_{i=1}^{N} X_i \mid N\right) = mpN.$

Συνεπώς, $E\left(\sum_{i=1}^{N} X_i\right) = E\left[E\left(\sum_{i=1}^{N} X_i \mid N\right)\right] = E(mpN) = mpE(N) = mp\lambda.$ Η τελευταία ισότητα προκύπτει διότι $N \sim \text{Poisson}(\lambda).$

Παράδειγμα 5.9 Υποθέτουμε ότι ο αριθμός των ταξιδιωτών που φθάνουν σε ένα σιδηροδρομικό σταθμό μέχρι τη χρονική στιγμή $t,$ ακολουθεί την κατανομή $\text{Poisson}(\lambda t),$ όπου λ είναι ο ρυθμός με τον οποίον οι ταξιδιώτες φθάνουν στο σιδηροδρομικό σταθμό. Αν η χρονική στιγμή άφιξης του τρένου στο σταθμό ακολουθεί την Ομοιόμορφη κατανομή στο διάστημα $(0,T)$ και είναι ανεξάρτητη των αφίξεων των ταξιδιωτών στο σταθμό, ποια είναι η μέση τιμή του αριθμού των ταξιδιωτών που επιβιβάζονται στο τρένο;

Λύση. Έστω η τυχαία μεταβλητή $N(t)$ που αναπαριστά τον αριθμό των αφίξεων των ταξιδιωτών στο σταθμό μέχρι τη χρονική στιγμή $t,$ $t \geq 0.$ Έστω ότι η τυχαία μεταβλητή Y αναπαριστά τη χρονική στιγμή άφιξης του τρένου στο σταθμό. Ισχύει ότι $N(t) \sim \text{Poisson}(\lambda t)$ και $Y \sim \text{Ομοιόμορφη}(0,T).$ Ζητάμε τη μέση τιμή $E[N(Y)].$

Ισχύει ότι $E\big[N(Y) \mid Y = t\big] = E\big[N(t) \mid Y = t\big] = E[N(t)] = \lambda t.$ Η δεύτερη ισότητα είναι συνέπεια της ανεξαρτησίας των τυχαίων μεταβλητών Y και $N(t).$ Η τελευταία ισότητα προκύπτει διότι $N(t) \sim \text{Poisson}(\lambda t).$

Άρα, $E[N(Y) \mid Y] = \lambda Y.$ Συνεπώς, $E[N(Y)] = E\big[E[N(Y) \mid Y]\big] = E(\lambda Y) = \lambda E(Y) = \dfrac{\lambda T}{2}.$

Η τελευταία ισότητα προκύπτει διότι $Y \sim \text{Ομοιόμορφη}(0,T).$

Παράδειγμα 5.10 Αν $N \sim Bin(n, X),$ όπου η πιθανότητας επιτυχίας X είναι μία τυχαία μεταβλητή που ακολουθεί την Ομοιόμορφη κατανομή στο $(0,1),$ δείξτε ότι $P(N = k) = \dfrac{1}{n+1},$ $0 \leq k \leq n.$

Λύση. Για $0 \leq k \leq n,$ διαδοχικά έχουμε

$$P(N=k) = \int_0^1 P(N=k \mid X=x) f_X(x)dx = \int_0^1 P(N=k \mid X=x)dx = \int_0^1 \binom{n}{k} x^k (1-x)^{n-k}\,dx$$

$$= \binom{n}{k}\int_0^1 x^k (1-x)^{n-k}\,dx = \binom{n}{k} B(k+1, n-k+1) = \binom{n}{k}\frac{\Gamma(k+1)\Gamma(n-k+1)}{\Gamma(n+2)} = \frac{n!}{k!(n-k)!}\cdot\frac{k!(n-k)!}{(n+1)!} = \frac{1}{n+1}.$$

Η δεύτερη ισότητα προκύπτει διότι $f_X(x) = 1,$ $x \in (0,1).$ $B(a,b)$ και $\Gamma(a)$ είναι οι συναρτήσεις Βήτα και Γάμα, αντίστοιχα.

Παράδειγμα 5.11 Ένας μεταλλωρύχος είναι παγιδευμένος σε ένα ορυχείο το οποίο έχει τρεις πόρτες. Η πρώτη πόρτα οδηγεί σε ένα τούνελ το οποίο τον απελευθερώνει μετά από πορεία τριών ωρών. Η δεύτερη πόρτα οδηγεί σε ένα τούνελ το οποίο τον ξαναφέρνει στο ορυχείο μετά από πορεία πέντε ωρών. Η τρίτη πόρτα οδηγεί σε ένα τούνελ το οποίο τον ξαναφέρνει στο ορυχείο μετά από πορεία επτά ωρών. Αν υποθέσουμε ότι πάντοτε ο μεταλλωρύχος διαλέγει κάποια από τις τρεις πόρτες με την ίδια πιθανότητα, ποιος είναι ο αναμενόμενος χρόνος μέχρι την απελευθέρωσή του;

Λύση. Έστω Y η τυχαία μεταβλητή που αναπαριστά το χρόνο μέχρι την απελευθέρωση και X η τυχαία μεταβλητή που αναπαριστά την πόρτα που διαλέγει αρχικά ο μεταλλωρύχος. Ζητάμε τη μέση τιμή $E[Y]$. Ισχύει ότι $E[Y] = E[Y \mid X=1]P[X=1] + E[Y \mid X=2]P[X=2] + E[Y \mid X=3]P[X=3]$

$$= \frac{1}{3}(E[Y \mid X=1] + E[Y \mid X=2] + E[Y \mid X=3]).$$

Όμως, εξ υποθέσεως, ισχύει ότι $E[Y \mid X=1] = 3,\ E[Y \mid X=2] = 5 + E[Y],\ \ E[Y \mid X=3] = 7 + E[Y].$

Άρα, $E[Y] = \frac{1}{3}(3 + 5 + E[Y] + 7 + E[Y]) \Rightarrow E[Y] = 15.$

5.3 Διατεταγμένες στατιστικές

Έστω $X_1, \ldots, X_n$ ανεξάρτητες συνεχείς τυχαίες μεταβλητές με κοινή σ.π. f και συνάρτηση κατανομής F. Ορίζουμε:

$X_{(1)} =$ η μικρότερη των $X_1, \ldots, X_n$

$X_{(2)} =$ η δεύτερη στη σειρά μικρότερη των $X_1, \ldots, X_n$

$\vdots$

$X_{(j)} =$ η $j-$οστή μικρότερη των $X_1, \ldots, X_n$

$\vdots$

$X_{(n)} =$ η μεγαλύτερη των $X_1, \ldots, X_n$

δηλαδή ισχύει ότι $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$. Οι τυχαίες μεταβλητές $X_{(1)}, \ldots, X_{(n)}$ είναι γνωστές ως **διατεταγμένες στατιστικές** (order statistics) που αντιστοιχούν στις τυχαίες μεταβλητές $X_1, \ldots, X_n$.

Για παράδειγμα, σε έναν αγώνα δρόμου εκατό μέτρων συμμετέχουν πέντε δρομείς. Αν υποθέσουμε ότι οι χρόνοι που επιτυγχάνουν οι δρομείς περιγράφονται από τις ανεξάρτητες τυχαίες μεταβλητές $X_1, \ldots, X_5$, οι οποίες επιπλέον ακολουθούν την ίδια κατανομή, τότε η τυχαία μεταβλητή $X_{(1)} = \min(X_1, \ldots, X_5)$ αντιστοιχεί στον καλύτερο χρόνο που επιτεύχθηκε στη κούρσα, η τυχαία μεταβλητή $X_{(5)} = \max(X_1, \ldots, X_5)$ αντιστοιχεί στον χειρότερο χρόνο που επιτεύχθηκε στην κούρσα, οι τυχαίες μεταβλητές $X_{(2)}, X_{(3)}$ αντιστοιχούν στους χρόνους των αθλητών που πήραν το ασημένιο και το χάλκινο μετάλλιο, αντίστοιχα και η τυχαία μεταβλητή $X_{(2)} - X_{(1)}$ εκφράζει τη διαφορά στους χρόνους άφιξης των δύο πρώτων αθλητών στο τέρμα της κούρσας. Αν οι χρόνοι (σε δευτερόλεπτα) που πέτυχαν οι πέντε αθλητές σε μία συγκεκριμένη κούρσα ήταν: $X_1 = 9.9$, $X_2 = 10.5$, $X_3 = 9.5$, $X_4 = 11$ και $X_5 = 10.3$, οι αντίστοιχες τιμές των τυχαίων μεταβλητών $X_{(1)}, \dots, X_{(5)}$ θα είναι: $X_{(1)} = 9.5$, $X_{(2)} = 9.9$, $X_{(3)} = 10.3$, $X_{(4)} = 10.5$ και $X_{(5)} = 11$.

Η από κοινού συνάρτηση πυκνότητας της $n-$διάστατης συνεχούς δ.τ.μ. $(X_{(1)}, \dots, X_{(n)})$ δίνεται από τον τύπο:

$$f_{(X_{(1)},\dots,X_{(n)})}(x_1, \dots, x_n) = n!\, f(x_1)\cdots f(x_n), \quad x_1 < x_2 < \cdots < x_n. \tag{5.2}$$

Η σχέση (5.2) μπορεί να εξηγηθεί ως εξής: Για να είναι ίση η δ.τ.μ. $(X_{(1)}, \dots, X_{(n)})$ με $(x_1, \dots, x_n)$ πρέπει και αρκεί η δ.τ.μ. $(X_1, \dots, X_n)$ να είναι ίση με κάποια από τις $n!$ μεταθέσεις της $n-$άδας των τιμών $(x_1, \dots, x_n)$. Δεδομένου ότι η από κοινού σ.π. της δ.τ.μ. $(X_1, \dots, X_n)$ είναι ίση με $f(x_{i_1})\cdots f(x_{i_n})$ για οποιαδήποτε μετάθεση $x_{i_1}, \dots, x_{i_n}$ των τιμών $x_1, \dots, x_n$, η σχέση (5.2) προκύπτει άμεσα.

Παράδειγμα 5.12 Σε ένα δρόμο μήκους ενός χιλιομέτρου, τρεις άνθρωποι βρίσκονται σε τρία τυχαία σημεία του. Βρείτε την πιθανότητα να μην υπάρχουν δύο άνθρωποι, η απόσταση των οποίων να είναι μικρότερη από d χιλιόμετρα, όπου $d \le 1/2$.

Λύση. Έστω οι τυχαίες μεταβλητές X_i, $i = 1, 2, 3$ που αναπαριστούν τη θέση του $i-$οστού ανθρώπου στο δρόμο. Το γεγονός ότι η θέση του κάθε ανθρώπου στο δρόμο είναι εντελώς τυχαία μας επιτρέπει να θεωρήσουμε τις τυχαίες μεταβλητές X_i, $i = 1, 2, 3$ ως ανεξάρτητες. Επιπλέον μπορούμε να θεωρήσουμε ότι οι X_i, $i = 1, 2, 3$ είναι ομοιόμορφα κατανεμημένες στο διάστημα $(0, 1)$ (δηλαδή κατά μήκος του δρόμου ενός χιλιομέτρου). Η ζητούμενη πιθανότητα είναι ίση με $P\{X_{(i)} > X_{(i-1)} + d, i = 2, 3\}$, όπου $X_{(i)}$, $i = 1, 2, 3$ είναι οι διατεταγμένες στατιστικές που αντιστοιχούν στις τυχαίες μεταβλητές X_i, $i = 1, 2, 3$. Από τη σχέση (5.2) προκύπτει ότι η από κοινού συνάρτηση πυκνότητας των διατεταγμένων στατιστικών $X_{(i)}$, $i = 1, 2, 3$ είναι $f_{(X_{(1)},X_{(2)},X_{(3)})}(x_1, x_2, x_3) = 3!$, $0 < x_1 < x_2 < x_3 < 1$. Διαδοχικά έχουμε

$$P\{X_{(i)} > X_{(i-1)} + d, i = 2, 3\} = \iiint\limits_{\substack{x_i > x_{i-1}+d \\ i=2,3}} f_{(X_{(1)},X_{(2)},X_{(3)})}(x_1, x_2, x_3)\,dx_1 dx_2 dx_3 = 3! \int_0^{1-2d} \int_{x_1+d}^{1-d} \int_{x_2+d}^{1} dx_3 dx_2 dx_1$$

$$= 6 \int_0^{1-2d} \int_{x_1+d}^{1-d} (1 - d - x_2)\,dx_2 dx_1 = 6 \int_0^{1-2d} \int_0^{1-2d-x_1} y_2\,dy_2 dx_1 = 6 \int_0^{1-2d} \frac{(1-2d-x_1)^2}{2} dx_1 = 6 \int_0^{1-2d} \frac{y_1^2}{2} dy_1 = (1-2d)^3.$$

Στην τέταρτη ισότητα κάναμε την αντικατάσταση $y_2 = 1 - d - x_2$ και στην έκτη ισότητα κάναμε την αντικατάσταση $y_1 = 1 - 2d - x_1$. Συνεπώς η πιθανότητα να μην υπάρχουν δύο άνθρωποι, η απόσταση των οποίων να είναι μικρότερη από d χιλιόμετρα, όπου $d \le 1/2$, όταν υπάρχουν τρεις άνθρωποι σε ένα δρόμο ενός χιλιομέτρου οι θέσεις των οποίων είναι ανεξάρτητες τυχαίες μεταβλητές ομοιόμορφα κατανεμημένες στο διάστημα $(0,1)$, είναι ίση με $(1-2d)^3$. Αν το πλήθος των ανθρώπων είναι ίσο με n, τότε η ζητούμενη πιθανότητα είναι ίση με $[1-(n-1)d]^n$, όπου $d \le \dfrac{1}{n-1}$.

Η περιθώρια συνάρτηση πυκνότητας της $j-$οστής διατεταγμένης στατιστικής $X_{(j)}$ προκύπτει, αν ολοκληρώσουμε την από κοινού συνάρτηση πυκνότητας $f_{(X_{(1)},\ldots,X_{(n)})}(x_1,\ldots,x_n)$ ως προς $x_1,\ldots,x_{j-1},x_{j+1},\ldots,x_n$. Ένας άμεσος συλλογισμός που μας οδηγεί στην περιθώρια συνάρτηση πυκνότητας της διατεταγμένης στατιστικής $X_{(j)}$ είναι ο εξής. Για να είναι ίση με x η τυχαία μεταβλητή $X_{(j)}$ θα πρέπει $j-1$ από τις n τιμές των τυχαίων μεταβλητών $X_1,\ldots,X_n$ να είναι μικρότερες από x, $n-j$ τιμές να είναι μεγαλύτερες από x και ακριβώς μία τιμή να είναι ίση με x. Η συνάρτηση πυκνότητας οποιουδήποτε συνόλου τιμών των τυχαίων μεταβλητών $X_1,\ldots,X_n$ τέτοιου ώστε $j-1$ τιμές να είναι όλες μικρότερες από x, $n-j$ τιμές να είναι όλες μεγαλύτερες από x και ακριβώς μία τιμή να είναι ίση με x, είναι $[F(x)]^{j-1}[1-F(x)]^{n-j} f(x)$.

Επειδή υπάρχουν $\dbinom{n}{j-1,n-j,1} = \dfrac{n!}{(n-j)!(j-1)!1!} = \dfrac{n!}{(n-j)!(j-1)!}$ διαφορετικοί τρόποι έτσι ώστε ένα σύνολο n τιμών των τυχαίων μεταβλητών $X_1,\ldots,X_n$ να είναι τέτοιο ώστε $j-1$ τιμές του να είναι όλες μικρότερες από x, $n-j$ τιμές του να είναι όλες μεγαλύτερες από x και ακριβώς μία τιμή του να είναι ίση με x, καταλήγουμε ότι η περιθώρια συνάρτηση πυκνότητας της τυχαίας μεταβλητής $X_{(j)}$ είναι

$$f_{X_{(j)}}(x) = \frac{n!}{(n-j)!(j-1)!}[F(x)]^{j-1}[1-F(x)]^{n-j} f(x) \tag{5.3}$$

Παράδειγμα 5.13 Έστω ότι παρατηρούμε ένα δείγμα μεγέθους $2n+1$, δηλαδή έστω ότι παρατηρούμε $2n+1$ ανεξάρτητες τυχαίες μεταβλητές οι οποίες ακολουθούν την ίδια κατανομή. Η **διάμεσος** (median) του δείγματος είναι η $(n+1)-$οστή μικρότερη από αυτές τις τυχαίες μεταβλητές. Αν ένα δείγμα μεγέθους τρία από την Ομοιόμορφη κατανομή στο διάστημα $(0,1)$ παρατηρείται, υπολογίστε την πιθανότητα η διάμεσος του δείγματος να βρίσκεται ανάμεσα στους αριθμούς $1/4$ και $3/4$.

Λύση. Από τη σχέση (5.3) η συνάρτηση πυκνότητας της διατεταγμένης στατιστικής $X_{(2)}$ που αντιστοιχεί στο δείγμα X_1, X_2, X_3 δίνεται από τον τύπο

$$f_{X_{(2)}}(x) = \frac{3!}{1!1!}x(1-x),\ \ 0 < x < 1.\ \text{Άρα,}\ P\left(\frac{1}{4} < X_{(2)} < \frac{3}{4}\right) = 6\int_{1/4}^{3/4} x(1-x)dx = 6\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_{1/4}^{3/4} = \frac{11}{16}.$$

Η συνάρτηση κατανομής της $j-$οστής διατεταγμένης στατιστικής $X_{(j)}$ μπορεί να εξαχθεί ολοκληρώνοντας την εξίσωση (5.3). Ισχύει ότι

$$F_{X_{(j)}}(y) = \frac{n!}{(n-j)!(j-1)!}\int_{-\infty}^{y}[F(x)]^{j-1}[1-F(x)]^{n-j} f(x)dx \tag{5.4}$$

Ένας εναλλακτικός τρόπος για να εξαχθεί άμεσα η συνάρτηση κατανομής της τυχαίας μεταβλητής $X_{(j)}$ είναι ο εξής. Η τιμή της $j-$οστής διατεταγμένης στατιστικής $X_{(j)}$ είναι μικρότερη ή ίση με y αν και μόνο αν υπάρχουν j ή περισσότερες των τυχαίων μεταβλητών $X_1,\ldots,X_n$ οι οποίες είναι μικρότερες ή ίσες με y. Επειδή ο αριθμός των τυχαίων μεταβλητών $X_1,\ldots,X_n$ που είναι μικρότερες ή ίσες του y είναι μία τυχαία μεταβλητή που ακολουθεί τη διωνυμική κατανομή με παραμέτρους n και $p = F(y)$, έπεται ότι

$$F_{X_{(j)}}(y) = P[X_{(j)} \le y] = P[j \text{ ή περισσότερες των } X_1,\ldots,X_n \le y] = \sum_{k=j}^{n}\binom{n}{k}[F(y)]^k[1-F(y)]^{n-k} \tag{5.5}$$

Αν στις σχέσεις (5.4) και (5.5) θεωρήσουμε ως F τη συνάρτηση κατανομής της Ομοιόμορφης κατανομής στο διάστημα $(0,1)$, δηλαδή θεωρήσουμε ότι $F(x) = x,\ 0 < x < 1$, προκύπτει η ακόλουθη ενδιαφέρουσα ταυτότητα

$$\sum_{k=j}^{n}\binom{n}{k}y^k(1-y)^{n-k} = \frac{n!}{(n-j)!(j-1)!}\int_0^y x^{j-1}(1-x)^{n-j}dx,\ \ 0 \le y \le 1. \tag{5.6}$$

Χρησιμοποιώντας παρόμοιους συλλογισμούς με αυτούς με τους οποίους καταλήξαμε στη σχέση (5.3), προκύπτει ότι η από κοινού συνάρτηση πυκνότητας της διδιάστατης δ.τ.μ. $(X_{(i)}, X_{(j)})$ με $i < j$, δίνεται από τον τύπο

$$f_{(X_{(i)},X_{(j)})}(x_i, x_j) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F(x_i)]^{i-1}[F(x_j)-F(x_i)]^{j-i-1}[1-F(x_j)]^{n-j} f(x_i)f(x_j) \tag{5.7}$$

για όλα τα $x_i < x_j$.

Παράδειγμα 5.14 (Κατανομή του εύρους ενός τυχαίου δείγματος). Υποθέτουμε ότι παρατηρούμε n ανεξάρτητες τυχαίες μεταβλητές $X_1,\ldots,X_n$ οι οποίες ακολουθούν την ίδια κατανομή. Η τυχαία μεταβλητή $R := X_{(n)} - X_{(1)}$ καλείται **εύρος** (range) των παρατηρούμενων τυχαίων μεταβλητών. Αν οι τυχαίες μεταβλητές

$X_1, \dots, X_n$ έχουν κοινή συνάρτηση κατανομής F και κοινή συνάρτηση πυκνότητας f, τότε η συνάρτηση κατανομής της τυχαίας μεταβλητής R μπορεί να εξαχθεί χρησιμοποιώντας τη σχέση (5.7) ως εξής:
Για $a \geq 0$, διαδοχικά έχουμε

$$P\{R \leq a\} = P\{X_{(n)} - X_{(1)} \leq a\} = \iint\limits_{x_n - x_1 \leq a} f_{(X_{(1)}, X_{(n)})}(x_1, x_n) dx_1 dx_n = \int_{-\infty}^{\infty} \int_{x_1}^{x_1 + a} \frac{n!}{(n-2)!} [F(x_n) - F(x_1)]^{n-2} f(x_1) f(x_n) dx_n dx_1$$

Κάνουμε την αντικατάσταση $y = F(x_n) - F(x_1)$, $dy = f(x_n) dx_n$ και έχουμε

$$\int_{x_1}^{x_1 + a} [F(x_n) - F(x_1)]^{n-2} f(x_n) dx_n = \int_0^{F(x_1 + a) - F(x_1)} y^{n-2} dy = \frac{1}{n-1} [F(x_1 + a) - F(x_1)]^{n-1}.$$

Συνεπώς, $P\{R \leq a\} = n \int_{-\infty}^{\infty} [F(x_1 + a) - F(x_1)]^{n-1} f(x_1) dx_1.$ (5.8)

Στην ειδική περίπτωση κατά την οποία οι τυχαίες μεταβλητές $X_1, \dots, X_n$ είναι ομοιόμορφα κατανεμημένες στο διάστημα $(0,1)$, για $0 < a < 1$, από τη σχέση (5.8), προκύπτει ότι

$$P\{R \leq a\} = n \int_0^1 [F(x_1 + a) - F(x_1)]^{n-1} f(x_1) dx_1 = n \int_0^{1-a} a^{n-1} dx_1 + n \int_{1-a}^1 (1 - x_1)^{n-1} dx_1 = n(1-a) a^{n-1} + a^n.$$

Αν παραγωγίσουμε ως προς a, παίρνουμε τη συνάρτηση πυκνότητας του εύρους ως εξής:

$f_R(a) = n(n-1) a^{n-2} (1-a)$, $0 \leq a \leq 1$ και $f_R(a) = 0$, διαφορετικά.

5.4 Εύρεση της κατανομής συναρτήσεων τυχαίων μεταβλητών

Έστω (X_1, X_2) μία διδιάστατη συνεχής δ.τ.μ. με από κοινού συνάρτηση πυκνότητας $f_{(X_1, X_2)}(x_1, x_2)$. Έστω τυχαίες μεταβλητές Y_1, Y_2 τέτοιες ώστε $Y_1 = g_1(X_1, X_2)$ και $Y_2 = g_2(X_1, X_2)$, για κάποιες συναρτήσεις g_1, g_2 που ικανοποιούν τις εξής δύο συνθήκες.

Συνθήκη 1. Οι εξισώσεις $y_1 = g_1(x_1, x_2)$ και $y_2 = g_2(x_1, x_2)$ μπορούν κατά μοναδικό τρόπο να λυθούν ως προς x_1 και x_2 συναρτήσει των y_1 και y_2, με λύση: $x_1 = h_1(y_1, y_2)$, $x_2 = h_2(y_1, y_2)$, για κάποιες συναρτήσεις h_1 και h_2.

Συνθήκη 2. Οι συναρτήσεις g_1, g_2 έχουν συνεχείς μερικές παραγώγους σε όλα τα σημεία (x_1, x_2) και είναι τέτοιες ώστε η παρακάτω 2×2 ορίζουσα να είναι διάφορη του μηδενός σε όλα τα σημεία (x_1, x_2).

$$J(x_1, x_2) = \begin{vmatrix} \frac{\vartheta g_1}{\vartheta x_1} & \frac{\vartheta g_1}{\vartheta x_2} \\ \frac{\vartheta g_2}{\vartheta x_1} & \frac{\vartheta g_2}{\vartheta x_2} \end{vmatrix} = \frac{\vartheta g_1}{\vartheta x_1} \cdot \frac{\vartheta g_2}{\vartheta x_2} - \frac{\vartheta g_1}{\vartheta x_2} \cdot \frac{\vartheta g_2}{\vartheta x_1}$$

Αν ισχύουν οι παραπάνω δύο συνθήκες, αποδεικνύεται ότι η από κοινού συνάρτηση πυκνότητας της διδιάστατης δ.τ.μ. (Y_1, Y_2) δίνεται από τον τύπο

$$f_{(Y_1,Y_2)}(y_1, y_2) = f_{(X_1,X_2)}(x_1, x_2) \cdot |J(x_1, x_2)|^{-1},$$

όπου $x_1 = h_1(y_1, y_2)$ και $x_2 = h_2(y_1, y_2)$, για κάποιες συναρτήσεις h_1 και h_2.

Παράδειγμα 5.15 Έστω (X_1, X_2) διδιάστατη συνεχής δ.τ.μ. με από κοινού συνάρτηση πυκνότητας $f_{(X_1,X_2)}(x_1, x_2)$. Έστω τυχαίες μεταβλητές $Y_1 = X_1 + X_2$ και $Y_2 = X_1 - X_2$. Υπολογίστε την από κοινού συνάρτηση πυκνότητας της διδιάστατης συνεχούς δ.τ.μ. (Y_1, Y_2) συναρτήσει της $f_{(X_1,X_2)}$.

Λύση. Έστω $g_1(x_1, x_2) = x_1 + x_2$ και $g_2(x_1, x_2) = x_1 - x_2$. Τότε $J(x_1, x_2) = -2 \neq 0$. Οι εξισώσεις $y_1 = x_1 + x_2$ και $y_2 = x_1 - x_2$ έχουν ως λύση $x_1 = \dfrac{(y_1 + y_2)}{2}$ και $x_2 = \dfrac{(y_1 - y_2)}{2}$.

Επομένως, $f_{(Y_1,Y_2)}(y_1, y_2) = \dfrac{1}{2} f_{(X_1,X_2)}\left(\dfrac{y_1 + y_2}{2}, \dfrac{y_1 - y_2}{2}\right)$.

Αν, για παράδειγμα, οι τυχαίες μεταβλητές X_1, X_2 ακολουθούν την Ομοιόμορφη κατανομή στο διάστημα $(0,1)$ και είναι ανεξάρτητες, τότε: $f_{(Y_1,Y_2)}(y_1, y_2) = \dfrac{1}{2}$, αν $0 \le y_1 + y_2 \le 2$, $0 \le y_1 - y_2 \le 2$ και $f_{(Y_1,Y_2)}(y_1, y_2) = 0$, διαφορετικά.

Αν, για παράδειγμα, η X_1 ακολουθεί την Εκθετική κατανομή με παράμετρο λ_1 και η X_2 ακολουθεί την Εκθετική κατανομή με παράμετρο λ_2 και είναι ανεξάρτητες, τότε:

$$f_{(Y_1,Y_2)}(y_1, y_2) = \frac{\lambda_1 \lambda_2}{2} \exp\left\{-\lambda_1\left(\frac{y_1 + y_2}{2}\right) - \lambda_2\left(\frac{y_1 - y_2}{2}\right)\right\}, \text{ αν } y_1 + y_2 \ge 0, \ y_1 - y_2 \ge 0 \text{ και } f_{(Y_1,Y_2)}(y_1, y_2) = 0,$$

διαφορετικά.

Αν, για παράδειγμα, οι τυχαίες μεταβλητές X_1, X_2 ακολουθούν την τυπική κανονική κατανομή, δηλαδή $X_1, X_2 \sim N(0, 1)$ και είναι ανεξάρτητες, τότε:

$$f_{(Y_1,Y_2)}(y_1, y_2) = \frac{1}{4\pi} \exp\left\{-\left(\frac{(y_1 + y_2)^2}{8}\right) + \left(\frac{(y_1 - y_2)^2}{8}\right)\right\} = \frac{1}{4\pi} \exp\left\{-\frac{(y_1^2 + y_2^2)}{4}\right\} = \frac{1}{\sqrt{4\pi}} e^{-\frac{y_1^2}{4}} e^{-\frac{y_2^2}{4}}, \ y_1, y_2 \in \Re.$$

Στην τελευταία περίπτωση προκύπτει επιπλέον το ενδιαφέρον αποτέλεσμα ότι οι τυχαίες μεταβλητές $Y_1 = X_1 + X_2$ και $Y_2 = X_1 - X_2$, είναι (επίσης) ανεξάρτητες.

Παράδειγμα 5.16 Αν $X \sim$ Γάμα(α, λ), $Y \sim$ Γάμα(β, λ) και X, Y είναι ανεξάρτητες τυχαίες μεταβλητές, υπολογίστε την από κοινού συνάρτηση πυκνότητας της διδιάστατης συνεχούς δ.τ.μ. (U, V), όπου $U = X + Y$ και $V = X/(X+Y)$.

Λύση. Η από κοινού συνάρτηση πυκνότητας της διδιάστατης συνεχούς δ.τ.μ. (X, Y) δίνεται από τον τύπο:

$$f_{(X,Y)}(x,y) = \frac{\lambda^{\alpha} x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)} \cdot \frac{\lambda^{\beta} y^{\beta-1} e^{-\lambda y}}{\Gamma(\beta)} = \frac{\lambda^{\alpha+\beta}}{\Gamma(\alpha)\Gamma(\beta)} e^{-\lambda(x+y)} x^{\alpha-1} y^{\beta-1}.$$

Έστω $g_1(x,y) = x + y$, $g_2(x,y) = x/(x+y)$. Τότε $\frac{\vartheta g_1}{\vartheta x} = \frac{\vartheta g_1}{\vartheta y} = 1$, $\frac{\vartheta g_2}{\vartheta x} = \frac{y}{(x+y)^2}$, $\frac{\vartheta g_2}{\vartheta y} = -\frac{x}{(x+y)^2}$.

Άρα, $J(x,y) = -\frac{1}{(x+y)}$. Οι εξισώσεις $u = x + y$, $v = x/(x+y)$ έχουν ως λύση $x = uv$, $y = u(1-v)$.

Επομένως, $f_{(U,V)}(u,v) = u \cdot f_{(X,Y)}(uv, u(1-v)) = \frac{\lambda e^{-\lambda u} (\lambda u)^{\alpha+\beta-1}}{\Gamma(\alpha+\beta)} \cdot \frac{v^{\alpha-1}(1-v)^{\beta-1}\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$.

Συνεπώς, οι τυχαίες μεταβλητές $U = X + Y$ και $V = X/(X+Y)$ είναι (επίσης) ανεξάρτητες.

Επιπλέον ισχύει ότι $X + Y \sim$ Γάμα$(\alpha + \beta, \lambda)$ και $X/(X+Y) \sim$ Βήτα(α, β).

Παράδειγμα 5.17 Αν $X \sim N(0,1)$, $Y \sim X_n^2$ και X, Y είναι ανεξάρτητες τυχαίες μεταβλητές, δείξτε ότι $Z := \frac{X}{\sqrt{\frac{Y}{n}}} \sim t_n$.

Λύση. Έστω η βοηθητική τυχαία μεταβλητή $W := \sqrt{\frac{Y}{n}}$. Τότε $X = ZW$ και $Y = nW^2$. Έστω $g_1(x,y) = \frac{x}{\sqrt{\frac{y}{n}}}$,

$g_2(x,y) = \sqrt{\frac{y}{n}}$. Τότε $\frac{\vartheta g_1}{\vartheta x} = \left(\frac{n}{y}\right)^{1/2}$, $\frac{\vartheta g_1}{\vartheta y} = -\frac{\sqrt{n}}{2} x y^{-3/2}$, $\frac{\vartheta g_2}{\vartheta x} = 0$, $\frac{\vartheta g_2}{\vartheta y} = \frac{1}{2}(ny)^{-1/2}$.

Ισχύει ότι $|J(x,y)| = \frac{1}{2y} = \frac{1}{2nw^2}$.

Άρα, $f_{(Z,W)}(z,w) = 2nw^2 \phi(zw) \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} (nw^2)^{\frac{n}{2}-1} e^{-\frac{nw^2}{2}}$, όπου ϕ είναι η συνάρτηση πυκνότητας της τυπικής κανονικής κατανομής και $\Gamma(n)$ είναι η συνάρτηση Γάμα.

Έχουμε ότι $f_{(Z,W)}(z,w) = \dfrac{2n^{\frac{n}{2}}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)\sqrt{2\pi}} w^n \exp\left\{-\frac{1}{2}(n+z^2)w^2\right\}, \ z \in \Re, \ w > 0.$

Όμως, $f_Z(z) = \int_0^\infty f_{(Z,W)}(z,w)dw = \dfrac{\left(1+\frac{z^2}{n}\right)^{-\frac{n+1}{2}}}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} 2\int_0^\infty \xi^n e^{-\xi^2} d\xi.$ Στη δεύτερη ισότητα θέσαμε $\xi = w \cdot \sqrt{\dfrac{(n+z^2)}{2}}.$

Αν θέσουμε $r = \xi^2$, έχουμε $2\int_0^\infty \xi^n e^{-\xi^2} d\xi = \int_0^\infty r^{\frac{n+1}{2}-1} e^{-r} dr = \Gamma\left(\frac{n+1}{2}\right).$

Συνεπώς, $f_Z(z) = \dfrac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}}\left(1+\frac{z^2}{n}\right)^{-\frac{n+1}{2}}, \ z \in \Re$, δηλαδή $Z \sim t_n$.

Παράδειγμα 5.18 Έστω (X,Y) οι συντεταγμένες ενός τυχαίου σημείου στο επίπεδο. Υποθέτουμε ότι $X,Y \sim N(0,1)$ και ότι X,Y είναι ανεξάρτητες τυχαίες μεταβλητές. Έστω (R,Θ) οι πολικές συντεταγμένες του σημείου. Υπολογίστε την από κοινού συνάρτηση πυκνότητας της διδιάστατης συνεχούς δ.τ.μ. (R,Θ).

Λύση. Θέτουμε $r = g_1(x,y) = \sqrt{x^2+y^2}, \ \theta = g_2(x,y) = \tau o\xi\varepsilon\phi \dfrac{y}{x}.$

Τότε $\dfrac{\vartheta g_1}{\vartheta x} = \dfrac{x}{\sqrt{x^2+y^2}}, \dfrac{\vartheta g_1}{\vartheta y} = \dfrac{y}{\sqrt{x^2+y^2}}, \ \dfrac{\vartheta g_2}{\vartheta x} = -\dfrac{y}{x^2+y^2}, \ \dfrac{\vartheta g_2}{\vartheta y} = \dfrac{x}{x^2+y^2}.$

Άρα, $J(x,y) = \dfrac{x^2}{(x^2+y^2)^{3/2}} + \dfrac{y^2}{(x^2+y^2)^{3/2}} = \dfrac{1}{\sqrt{x^2+y^2}} = \dfrac{1}{r}.$

Η από κοινού συνάρτηση πυκνότητας της διδιάστατης δ.τ.μ. (X,Y) είναι

$$f_{(X,Y)}(x,y) = \frac{1}{2\pi} e^{-\frac{(x^2+y^2)}{2}}.$$

Επομένως, η από κοινού συνάρτηση πυκνότητας της διδιάστατης συνεχούς διανυσματικής τυχαίας μεταβλητής $\left(R = \sqrt{X^2+Y^2}, \ \Theta = \tau o\xi\varepsilon\phi \dfrac{y}{x}\right)$ δίνεται από τον τύπο $f_{(R,\Theta)}(r,\theta) = \dfrac{1}{2\pi} r e^{-r^2/2}, \ r > 0, \ 0 < \theta < 2\pi.$ Οι περιθώριες συναρτήσεις πυκνότητας των τυχαίων μεταβλητών R και Θ είναι

$$f_R(r) = \int_0^{2\pi} f_{(R,\Theta)}(r,\theta)d\theta = re^{-\frac{r^2}{2}}, \ r > 0 \ \text{ και } \ f_\Theta(\theta) = \int_0^\infty f_{(R,\Theta)}(r,\theta)dr = \frac{1}{2\pi}, \ 0 < \theta < 2\pi.$$

Η κατανομή της τυχαίας μεταβλητής R καλείται **κατανομή Rayleigh** και η κατανομή της τυχαίας μεταβλητής Θ είναι η Ομοιόμορφη στο διάστημα $(0, 2\pi)$. Παρατηρούμε ότι οι τυχαίες μεταβλητές R και Θ είναι (επίσης) ανεξάρτητες.

ΚΕΦΑΛΑΙΟ 6
Ο ΙΣΧΥΡΟΣ ΝΟΜΟΣ ΤΩΝ ΜΕΓΑΛΩΝ ΑΡΙΘΜΩΝ ΚΑΙ ΤΟ ΚΕΝΤΡΙΚΟ ΟΡΙΑΚΟ ΘΕΩΡΗΜΑ

Τα πιο σημαντικά θεωρητικά αποτελέσματα της Θεωρίας Πιθανοτήτων είναι τα Οριακά Θεωρήματα τα οποία ταξινομούνται σε δύο κατηγορίες. Στη πρώτη κατηγορία συμπεριλαμβάνονται οι Νόμοι των Μεγάλων Αριθμών οι οποίοι δίνουν τη θεωρητική βάση της διαισθητικής και πειραματικής διαπίστωσης των Μαθηματικών του 18ου-19ου αιώνα σύμφωνα με την οποία, αν επαναλάβουμε ένα πείραμα τύχης πολλές φορές τότε ο στατιστικός ορισμός της πιθανότητας που αποδίδεται στον Von Mises είναι αληθής. Επιπλέον οι Νόμοι των Μεγάλων Αριθμών συνιστούν ένα ισχυρό εργαλείο για τη μελέτη διάφορων θεωρητικών και εφαρμοσμένων προβλημάτων Πιθανοτήτων και Στατιστικής. Στη δεύτερη κατηγορία συμπεριλαμβάνονται τα Κεντρικά Οριακά Θεωρήματα σύμφωνα με τα οποία, κάτω από αρκετά γενικές συνθήκες, η κατανομή του αθροίσματος ενός μεγάλου αριθμού τυχαίων μεταβλητών μπορεί να προσεγγιστεί ικανοποιητικά από μία κανονική κατανομή. Σε αυτό το κεφάλαιο θα παρουσιάσουμε τον Ισχυρό Νόμο των Μεγάλων Αριθμών και το Κεντρικό Οριακό Θεώρημα.

6.1 Όρια ακολουθιών τυχαίων μεταβλητών

Στο παρόν εδάφιο παραθέτουμε κάποιους ορισμούς που θα μας βοηθήσουν να παρουσιάσουμε τον Ισχυρό Νόμο των Μεγάλων Αριθμών και το Κεντρικό Οριακό Θεώρημα.

Θεωρούμε μία ακολουθία τυχαίων μεταβλητών $X_n,\ n = 1, 2, \dots$ που ορίζονται στον ίδιο πιθανοθεωρητικό χώρο $(\Omega, \Im, P)$. Δίνουμε τους ακόλουθους ορισμούς.

Ορισμός 6.1 Λέμε ότι έχουμε μία ακολουθία **ανεξάρτητων** (independent) τυχαίων μεταβλητών $X_n,\ n = 1, 2, \dots$ αν ισχύει ότι $P(X_1 \in B_1, \dots, X_n \in B_n) = P(X_1 \in B_1) \cdots P(X_n \in B_n)$, όπου $B_i, i = 1, \dots, n$ είναι οποιαδήποτε υποσύνολα του $\Re$.

Ορισμός 6.2 Λέμε ότι έχουμε μία ακολουθία **ισόνομων** (identically distributed) τυχαίων μεταβλητών X_n, $n=1,2,\ldots$ αν όλες οι τυχαίες μεταβλητές ακολουθούν την ίδια κατανομή.

Στη συνέχεια παραθέτουμε δύο έννοιες συγκλίσεως μιας ακολουθίας τυχαίων μεταβλητών X_n, $n=1,2,\ldots$.

Ορισμός 6.3 (α) Λέμε ότι η ακολουθία τυχαίων μεταβλητών X_n, $n=1,2,\ldots$ **συγκλίνει με πιθανότητα 1 (ή σχεδόν βεβαίως ή σχεδόν παντού)** (converges with probability 1) στην τυχαία μεταβλητή X αν και μόνο αν $P\{\omega\in\Omega \mid X_n(\omega)\underset{n\to\infty}{\to} X(\omega)\}=1.$

Ισοδύναμα, $\lim\limits_{n\to\infty} P(|X_n - X| > \varepsilon$ για ένα τουλάχιστον r με $r \geq n) = 0,$ για κάθε $\varepsilon > 0.$

(β) Λέμε ότι η ακολουθία τυχαίων μεταβλητών X_n, $n=1,2,\ldots$ **συγκλίνει κατά κατανομή** (converges in distribution) στην τυχαία μεταβλητή X αν και μόνο αν $\lim\limits_{n\to\infty} F_{X_n}(x) = F_X(x),$ για κάθε σημείο συνέχειας της συνάρτησης κατανομής F_X της X, όπου F_{X_n} είναι η συνάρτηση κατανομής της ακολουθίας X_n, $n=1,2,\ldots$.

Η σχεδόν βέβαια σύγκλιση συνεπάγεται ότι, σε ένα σύνολο του οποίου η πιθανότητα εμφάνισης είναι ίση με 1, για κάθε δειγματικό σημείο ω η διαφορά $X_n(\omega)-X(\omega)$ γίνεται αυθαίρετα μικρή όταν το n αυξάνεται απεριόριστα. Αποδεικνύεται ότι, αν η ακολουθία τυχαίων μεταβλητών X_n, $n=1,2,\ldots$ συγκλίνει σχεδόν παντού στην τυχαία μεταβλητή X, τότε συγκλίνει και κατά κατανομή στην τυχαία μεταβλητή X.

Παράδειγμα 6.1 Έστω $X_1, X_2, \ldots$ ανεξάρτητες και ισόνομες τυχαίες μεταβλητές που ακολουθούν την Ομοιόμορφη κατανομή στο διάστημα $(0,1)$. Έστω $Y_n := \max(X_1,\ldots,X_n)$. Δείξτε ότι $Z_n := n(1-Y_n)$ συγκλίνει, καθώς $n\to\infty$, κατά κατανομή σε μία τυχαία μεταβλητή Z η οποία ακολουθεί την Εκθετική κατανομή με παράμετρο τη μονάδα.

Λύση. Είναι $F_{Z_n}(z) = P[n(1-Y_n)\leq z] = P\left[Y_n \geq 1-\dfrac{z}{n}\right] = 1-P\left[Y_n < 1-\dfrac{z}{n}\right] = 1-\prod\limits_{i=1}^{n} P\left[X_i < 1-\dfrac{z}{n}\right]$

$= 1-\left[P\left(X_1 < 1-\dfrac{z}{n}\right)\right]^n = 1-\left(1-\dfrac{z}{n}\right)^n, 0<z<1.$ Η τέταρτη ισότητα είναι συνέπεια της ανεξαρτησίας των τυχαίων μεταβλητών $X_1,\ldots,X_n$. Ισχύει ότι $F_{Z_n}(z)\to 1-e^{-z} = F_Z(z),$ όπου $Z\sim$ Εκθετική(1), δηλαδή η τυχαία μεταβλητή Z_n συγκλίνει, καθώς $n\to\infty$, κατά κατανομή στην τυχαία μεταβλητή Z.

6.2 Ο Ισχυρός Νόμος των Μεγάλων Αριθμών

Ο Ισχυρός Νόμος των Μεγάλων Αριθμών (The Strong Law of Large Numbers) είναι ένα από τα πιο γνωστά αποτελέσματα της Θεωρίας Πιθανοτήτων. Σύμφωνα με το παρακάτω θεώρημα, το οποίο παρατίθεται χωρίς απόδειξη, κάτω από κατάλληλες υποθέσεις, ο δειγματικός μέσος μιας ακολουθίας ανεξάρτητων τυχαίων μεταβλητών που ακολουθούν μία κοινή κατανομή συγκλίνει σχεδόν βεβαίως προς τον θεωρητικό μέσο (τη μέση τιμή) της κατανομής.

Θεώρημα 6.1 (Ισχυρός Νόμος των Μεγάλων Αριθμών, Ι.Ν.Μ.Α.). Έστω μία ακολουθία ανεξάρτητων και ισόνομων τυχαίων μεταβλητών $X_n,\ n=1,2,\ldots$ τέτοιων ώστε να υπάρχει η μέση τιμή και η διακύμανση της τυχαίας μεταβλητής X_1 (δηλαδή να ισχύει ότι $\mu = EX_1 < \infty$ και $\sigma^2 = Var(X_1) < \infty$). Τότε η τυχαία μεταβλητή $\overline{X_n} = \frac{\sum_{i=1}^{n} X_i}{n}$ (δηλαδή η ακολουθία των δειγματικών μέσων) συγκλίνει με πιθανότητα 1 (ή σχεδόν βεβαίως ή σχεδόν παντού) στη θεωρητική μέση τιμή μ, καθώς $n \to \infty$, δηλαδή $P(\lim_{n\to\infty} \overline{X}_n = \mu) = 1.$

Ο Ι.Ν.Μ.Α. αρχικά αποδείχθηκε από τον Γάλλο Μαθηματικό Borel στην ειδική περίπτωση κατά την οποία η ακολουθία $X_n,\ n=1,2,\ldots$ είναι μία ακολουθία τυχαίων μεταβλητών Bernoulli. Η γενική μορφή του Ι.Ν.Μ.Α., όπως παρουσιάζεται στο Θεώρημα 6.1, αποδείχθηκε από τον Ρώσο Μαθηματικό Kolmogorov.

Παράδειγμα 6.2 Θεωρούμε μία ακολουθία ανεξάρτητων δοκιμών ενός τυχαίου πειράματος. Έστω A ένα συγκεκριμένο ενδεχόμενο του πειράματος και $P(A)$ η πιθανότητα να συμβεί το ενδεχόμενο A σε μία οποιαδήποτε δοκιμή. Έστω η τυχαία μεταβλητή X_i τέτοια ώστε $X_i = 1,$ αν το ενδεχόμενο A συμβαίνει κατά την $i-$οστή δοκιμή και $X_i = 0,$ αν το ενδεχόμενο A δεν συμβαίνει κατά την $i-$οστή δοκιμή. Ισχύει ότι $EX_i = 1P(X_i = 1) + 0P(X_i = 0) = P(A).$ Από τον Ι.Ν.Μ.Α, η τυχαία μεταβλητή $\overline{X_n} = \frac{X_1 + \cdots + X_n}{n}$ συγκλίνει σχεδόν παντού στη $EX_1 = P(A).$ Το αποτέλεσμα του παραδείγματος μπορεί να ερμηνευτεί ως εξής. Αφού η τυχαία μεταβλητή $X_1 + \cdots + X_n$ αναπαριστά τον αριθμό των φορών κατά τις οποίες το ενδεχόμενο A συμβαίνει στις πρώτες n δοκιμές του πειράματος, η πιθανότητα $P(A)$ είναι το οριακό ποσοστό των φορών κατά τις οποίες συμβαίνει το ενδεχόμενο A.

Παράδειγμα 6.3 Έστω $X_1, X_2, \ldots$ μία ακολουθία ανεξάρτητων συνεχών τυχαίων μεταβλητών με συνάρτηση πυκνότητας $f(x) = cx^2(2-x)^2, \ 0 \le x \le 2,$ όπου c κατάλληλη πραγματική σταθερά. Να βρεθεί το όριο της ακολουθίας των δειγματικών μέσων $\overline{X}_n, n = 1, 2, \ldots$ με την έννοια της σχεδόν βεβαίας σύγκλισης.

Λύση. Από την ισότητα $1 = \int_{-\infty}^{\infty} f(x)dx = c\int_0^2 x^2(4-4x+x^2)dx = \dfrac{16c}{15}$ βρίσκουμε $c = \dfrac{15}{16}$. Σύμφωνα με τον Ισχυρό Νόμο των Μεγάλων Αριθμών, αν η μέση τιμή μ και η διακύμανση σ^2 της ακολουθίας $X_1, X_2, \ldots$ είναι πεπερασμένες, η ακολουθία των δειγματικών μέσων $\overline{X}_n, n = 1, 2, \ldots$ θα συγκλίνει με πιθανότητα 1 (σχεδόν βεβαίως) στη μέση τιμή μ. Είναι $\mu = E[X_1] = \int_0^2 \dfrac{15}{16} x^3(4-4x+x^2)dx = 1 < \infty$.

Επιπλέον, $E[X_1^2] = \dfrac{15}{16}\int_0^2 x^4(4-4x+x^2)dx = \dfrac{8}{7}$. Άρα, $Var[X_1] = E[X_1] - (E[X_1])^2 = \dfrac{1}{7} < \infty$.

Συνεπώς, $P(\lim_{n\to\infty} \overline{X}_n = 1) = 1$.

6.3 Το Κεντρικό Οριακό Θεώρημα

Το Κεντρικό Οριακό Θεώρημα (The Central Limit Theorem) είναι ένα από τα πιο αξιοσημείωτα αποτελέσματα της Θεωρίας Πιθανοτήτων. Σύμφωνα με το παρακάτω θεώρημα, το άθροισμα ενός μεγάλου αριθμού ανεξάρτητων τυχαίων μεταβλητών ακολουθεί μία κατανομή η οποία προσεγγίζει την κανονική κατανομή. Όπως θα δούμε στα Παραδείγματα 6.4 και 6.5, το ακόλουθο θεώρημα παρέχει, επιπλέον, μία απλή μέθοδο για να υπολογίζουμε, κατά προσέγγιση, πιθανότητες αθροισμάτων ανεξάρτητων τυχαίων μεταβλητών.

Θεώρημα 6.2 (Κεντρικό Οριακό Θεώρημα, Κ.Ο.Θ.). Έστω μία ακολουθία ανεξάρτητων και ισόνομων τυχαίων μεταβλητών $X_n, \ n = 1, 2, \ldots$ τέτοιων ώστε $\mu = EX_1$ και $\sigma^2 = Var(X_1) < \infty$.

Τότε, η τυχαία μεταβλητή $Z_n = \dfrac{(X_1 + X_2 + \cdots + X_n) - n\mu}{\sigma\sqrt{n}}$ συγκλίνει κατά κατανομή, καθώς $n \to \infty$, στη συνάρτηση κατανομής της τυποποιημένης κανονικής κατανομής Φ,

δηλαδή ισχύει ότι $\lim_{n\to\infty} P(Z_n \le x) = \Phi(x) = \dfrac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt$, για κάθε $x \in \Re$.

Εναλλακτικά, η τυχαία μεταβλητή Z_n μπορεί να γραφεί στη μορφή: $Z_n = \dfrac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma}$, όπου $\overline{X}_n := \dfrac{\sum_{i=1}^{n} X_i}{n}$. Η τυχαία μεταβλητή Z_n είναι ο γενικός όρος της ακολουθίας των τυποποιημένων δειγματικών μέσων.

Η πρώτη μορφή του Κεντρικού Οριακού Θεωρήματος αποδείχθηκε από τον DeMoivre γύρω στα 1733 μ.Χ. για την ειδική περίπτωση που οι τυχαίες μεταβλητές X_i, $i=1,\dots,n$ είναι δίτιμες και ακολουθούν την κατανομή Bernoulli με πιθανότητα επιτυχίας $p = 1/2$. Η απόδειξη του DeMoivre γενικεύτηκε από τον Laplace για οποιαδήποτε τιμή της πιθανότητας p. Μία αυστηρή απόδειξη του Κ.Ο.Θ., όπως αυτό διατυπώνεται στο Θεώρημα 6.2, δόθηκε από τον Ρώσο Μαθηματικό Lyapunov, μαθητή του Chebyshev, την περίοδο 1901-1902. Στη συνέχεια παραθέτουμε κάποιες εφαρμογές του Κ.Ο.Θ. Για την απόδειξη του Κ.Ο.Θ., παραπέμπουμε στο βιβλίο του Μ. Κούτρα, Εισαγωγή στις Πιθανότητες, Μέρος ΙΙ, σελ. 335-337.

Το Κ.Ο.Θ. μας επιτρέπει να κατανοήσουμε το λόγο για τον οποίον η κανονική κατανομή κατέχει εξέχουσα θέση στη Θεωρία Πιθανοτήτων και στη Στατιστική. Τα χαρακτηριστικά (π.χ. βάρος, ύψος) των ατόμων ενός υπό εξέταση πληθυσμού περιγράφονται από κατάλληλες τυχαίες μεταβλητές που μπορούν να θεωρηθούν ως αποτέλεσμα άθροισης πολλών μικρών τυχαίων παραγόντων. Η συσσώρευση πολλών τέτοιων μικρών τυχαίων παραγόντων οδηγεί σε μία (μη αμελητέα) τυχαία ποσότητα, η οποία σύμφωνα με το Κ.Ο.Θ., θα ακολουθεί κατά προσέγγιση την κανονική κατανομή.

Παράδειγμα 6.4 Οι καθημερινές διακυμάνσεις μιας μετοχής στο Χρηματιστήριο Αθηνών ακολουθούν, αυτό το χρονικό διάστημα, κάποια άγνωστη κατανομή με γνωστή μέση τιμή $\mu=-5$ και διασπορά $\sigma^2 = 225$. Αν η τιμή μιας μετοχής σήμερα είναι 3000 ευρώ ποια είναι η πιθανότητα σε ένα μήνα (τριάντα ημέρες) από σήμερα η τιμή της μετοχής να βρίσκεται κάπου ανάμεσα στα 2700 και 3050 ευρώ;

Λύση. Έστω $X_1,\dots,X_{30}$ ανεξάρτητες και ισόνομες τυχαίες μεταβλητές που αναπαριστούν τις καθημερινές διακυμάνσεις της μετοχής. Ζητάμε την πιθανότητα p.

$$p = P\left(2700 \le 3000 + \sum_{i=1}^{30} X_i \le 3050\right) = P\left(-300 \le \sum_{i=1}^{30} X_i \le 50\right) = P\left(-10 \le \overline{X}_{30} \le \frac{5}{3}\right)$$

$$= P\left(-\frac{5\sqrt{30}}{15} \le \frac{(\overline{X}_{30}+5)\sqrt{30}}{15} \le \frac{20\sqrt{30}}{3\cdot 15}\right).$$

Από το Κ.Ο.Θ. η τυχαία μεταβλητή $Z := \dfrac{\sqrt{30}(\overline{X}_{30}+5)}{15}$ συγκλίνει κατά κατανομή στη συνάρτηση κατανομής Φ της τυπικής κανονικής κατανομής.

Άρα, $p = \Phi\left(\frac{20\sqrt{30}}{3\cdot 15}\right) - \Phi\left(-\frac{5\sqrt{30}}{15}\right) = \Phi\left(\frac{20\sqrt{30}}{3\cdot 15}\right) - \left(1-\Phi\left(\frac{5\sqrt{30}}{15}\right)\right) = \Phi\left(\frac{20\sqrt{30}}{3\cdot 15}\right) + \Phi\left(\frac{5\sqrt{30}}{15}\right) - 1.$

Η δεύτερη ισότητα είναι συνέπεια της Πρότασης 2.2.

Παράδειγμα 6.5 Έστω $X_i, i=1,\ldots,10$ ανεξάρτητες τυχαίες μεταβλητές οι οποίες είναι ομοιόμορφα κατανεμημένες στο διάστημα $(0,1)$. Υπολογίστε, κατά προσέγγιση, την πιθανότητα $P\left\{\sum_{i=1}^{10} X_i > 6\right\}$.

Λύση. Ισχύει ότι $E(X_i)=\dfrac{1}{2}$ και $Var(X_i)=\dfrac{1}{12}$. Από το Κ.Ο.Θ. έχουμε

$$P\left\{\sum_{i=1}^{10} X_i > 6\right\} = P\left\{\frac{\sum_{i=1}^{10} X_i - 5}{\sqrt{10\left(\frac{1}{2}\right)}} > \frac{6-5}{\sqrt{10\left(\frac{1}{2}\right)}}\right\} \approx 1-\Phi(\sqrt{1.2}) \approx 0.16.$$

Παράδειγμα 6.6 Έστω ότι ρίχνουμε ένα αμερόληπτο ζάρι δέκα φορές. Υποθέτουμε ότι οι ρίψεις είναι ανεξάρτητες. Βρείτε κατά προσέγγιση την πιθανότητα να είναι το άθροισμα των εμφανιζόμενων αριθμών μεγαλύτερο του 30 και μικρότερο ή ίσο του 40.

Λύση. Έστω ότι η τυχαία μεταβλητή X_i αναπαριστά το αποτέλεσμα του $i-$οστού ζαριού, $i=1,\ldots,10$. Ισχύει ότι $E(X_i)=\dfrac{7}{2}$ και $Var(X_i)=\dfrac{35}{12}$. Από το Κ.Ο.Θ. έχουμε

$$P\left(30<\sum_{i=1}^{10} X_i \le 40\right) = P\left(\sqrt{10}\left(3-\frac{7}{2}\right)\sqrt{\frac{12}{35}} < \sqrt{10}\left(\frac{1}{10}\sum_{i=1}^{10} X_i - \frac{7}{2}\right)\sqrt{\frac{12}{35}} \le \sqrt{10}\left(4-\frac{7}{2}\right)\sqrt{\frac{12}{35}}\right) \approx 2\Phi\left(\sqrt{\frac{6}{7}}\right)-1$$

$$\approx 0.65.$$

Παράδειγμα 6.7 Δείξτε ότι $e^{-n}\sum_{k=0}^{n}\dfrac{n^k}{k!} \underset{n\to\infty}{\longrightarrow} \dfrac{1}{2}$.

Λύση. Έστω $X_n, n=1,2,\ldots$ μία ακολουθία ανεξάρτητων και ισόνομων τυχαίων μεταβλητών που ακολουθούν την κατανομή Poisson με παράμετρο ίση με τη μονάδα. Από το Παράδειγμα 4.13 έπεται ότι $n\overline{X}_n \sim \text{Poisson}(n)$, όπου $\overline{X}_n = \dfrac{\sum_{i=1}^{n} X_i}{n}$. Από το Κ.Ο.Θ. έχουμε

$$e^{-n}\sum_{k=0}^{n}\frac{n^k}{k!} = P(n\overline{X}_n \le n) = P(\overline{X}_n \le 1) = P\left[\frac{\sqrt{n}(\overline{X}_n - EX_1)}{\sqrt{\Delta X_1}} \le \frac{\sqrt{n}(1-1)}{\sqrt{1}} = 0\right] \underset{n\to\infty}{\longrightarrow} \Phi(0)=\frac{1}{2},$$

όπου Φ είναι η συνάρτηση κατανομής της τυπικής κανονικής κατανομής.

Παράδειγμα 6.8 Ο αριθμός των φοιτητών που διαλέγουν το μάθημα της Ψυχολογίας είναι μία τυχαία μεταβλητή που ακολουθεί την κατανομή Poisson με παράμετρο ίση με 100. Ο καθηγητής έχει αποφασίσει ότι, αν ο αριθμός αυτός είναι μεγαλύτερος ή ίσος του 120, θα διαιρέσει τους φοιτητές σε δύο τμήματα, ενώ αν είναι μικρότερος του 120, θα διδάξει το μάθημα σε ένα τμήμα αποτελούμενο από όλους τους φοιτητές. Χρησιμοποιώντας το Κ.Ο.Θ. υπολογίστε, κατά προσέγγιση, την πιθανότητα να διαιρέσει ο καθηγητής τους φοιτητές σε δύο τμήματα.

Λύση. Έστω η τυχαία μεταβλητή X που αναπαριστά τον αριθμό των φοιτητών που διαλέγουν το μάθημα της Ψυχολογίας. Ισχύει ότι $X \sim \text{Poisson}(100)$. Η τυχαία μεταβλητή X μπορεί να γραφεί ως εξής: $X = X_1 + \cdots + X_{100}$, όπου X_i, $i = 1,\ldots,100$ είναι ανεξάρτητες και ισόνομες τυχαίες μεταβλητές που ακολουθούν την κατανομή Poisson με παράμετρο ίση με τη μονάδα.

Από το Κ.Ο.Θ. έχουμε

$$P(X \geq 120) = P\left[\sqrt{100}\left(\frac{1}{100}\sum_{i=1}^{100} X_i - 1\right) \geq \sqrt{100}\left(\frac{120}{100} - 1\right)\right] \approx 1 - \Phi(2) = 0.0228,$$

όπου Φ είναι η συνάρτηση κατανομής της τυπικής κανονικής κατανομής.

Παράδειγμα 6.9 Δύο παίκτες Α και Β παίζουν το εξής παιχνίδι. Ρίχνουν διαδοχικά ένα ζάρι και αν η ένδειξη είναι 1, 3 ή 5, ο παίκτης Α δίνει στον παίκτη Β ποσό 1, 2 ή 3 ευρώ, αντίστοιχα. Αν η ένδειξη είναι 2, 4 ή 6, ο παίκτης Β δίνει στον παίκτη Α ποσό 3, 2 ή 1 ευρώ, αντίστοιχα. Να υπολογισθεί η πιθανότητα σε 60 ρίψεις

(α) ο παίκτης Α να κερδίσει τουλάχιστον 17 ευρώ.

(β) ο παίκτης Α να κερδίσει το πολύ 19 ευρώ.

Λύση. Έστω X_i το κέρδος του παίκτη Α κατά την $i-$οστή ρίψη του ζαριού, $i = 1,\ldots,60$.

Έχουμε $P(X_i = j) = \dfrac{1}{6}$, $j = -3, -2, -1, 1, 2, 3$. Το συνολικό κέρδος του παίκτη Α σε $n = 60$ ρίψεις του ζαριού θα δίνεται από την τυχαία μεταβλητή $S = \sum_{i=1}^{60} X_i$. Σύμφωνα με το Κ.Ο.Θ. η τυποποιημένη τυχαία μεταβλητή $Z = \dfrac{S - E(S)}{\sqrt{Var(S)}}$ ακολουθεί κατά προσέγγιση την τυπική κανονική κατανομή $N(0,1)$.

Όμως, $E(X_i) = \sum_j jP(X_i = j) = \dfrac{1}{6}[(-3) + (-2) + (-1) + 1 + 2 + 3] = 0.$

$$Var(X_i) = E(X_i^2) - (E(X_i))^2 = E(X_i^2) = \sum_j j^2 P(X_i = j) = \frac{14}{3}.$$

Επομένως, $E(S) = \sum_{i=1}^{60} E(X_i) = 0$, $Var(S) = \sum_{i=1}^{60} Var(X_i) = 60 \cdot \dfrac{14}{3} = 280.$

(α) Η ζητούμενη πιθανότητα είναι $P(S \geq 17) = P\left(\frac{S-0}{\sqrt{280}} \geq \frac{17-0}{\sqrt{280}}\right) = P(Z \geq 1.02) \cong 1-\Phi(1.02) = 0.1539.$

(β) Ομοίως, $P(S \leq 19) = P\left(\frac{S-0}{\sqrt{280}} \leq \frac{19-0}{\sqrt{280}}\right) = P(Z \leq 1.14) \cong \Phi(1.14) = 0.8729.$