Κεφάλαιο 1

Στοιχεία Πιθανοτήτων

1.1 Διακριτές τυχαίες μεταβλητές

Έστω X διακριτή τ.μ. με πεδίο τιμών $R_X = \{x_0, x_1, \ldots\}$ ή $R_X = \{x_0, x_1, \ldots, x_n\}$.

Συνάρτηση πιθανότητας για μία διακριτή τ.μ. X: $P_X(x) = P(X = x)$, $x \in R_X$

Ιδιότητες: (i) $0 \le P_X(x) \le 1$, $x \in R_X$

(ii) $\sum_{x \in R_X} P_X(x) = 1$.

Αθροιστική Συνάρτηση Κατανομής

$$F_X(x) = P(X \le x) = \sum_{y \le x} P(X = y)$$

$$= \sum_{y \le x} P_X(y), \; x \in \mathbb{R}$$

Μέση τιμή. Αν $\sum_{x \in R_X} |x| P_X(x) < \infty$,

τότε: $E(X) = \sum_{x \in R_X} x \cdot P_X(x)$.

Διασπορά. Αν $\sum_{x\in R_X} x^2 \cdot P_X(x) < \infty$,

τότε:

$$Var(X) = E[X-E(X)]^2 = E(X^2) - (E(X))^2$$

Ισχύει ότι:

$$E[g(X)] = \sum_{x\in R_X} g(x) P_X(x).$$

Έστω X, Y δύο τ.μ. Οι τ.μ. X, Y είναι ανεξάρτητες $\underline{\underline{\text{ανν}}}$ $P(X\in A, Y\in B) = P(X\in A)P(Y\in B)$ $\forall A, B \subseteq \mathbb{R} \Leftrightarrow P_{X,Y}(x,y) = P_X(x)\cdot P_Y(y)$. $\forall x \in R_X, \forall y \in R_Y$.

Ειδικές Διακριτές Κατανομές

- Κατανομή Bernoulli – Bernoulli(p)

$p\in(0,1)$. $X = \begin{cases} 0, & \text{αποτυχία σε μία δοκιμή} \\ 1, & \text{επιτυχία σε μία δοκιμή} \end{cases}$

$P_X(x) = p^x(1-p)^{1-x}$, $x\in\{0,1\}$

$E(X) = p$, $Var(X) = p(1-p)$

$X, Y \sim Bernoulli(p)$ ανεξάρτητες

$\Rightarrow X+Y \sim Bin(2,p)$.

- Διωνυμική Κατανομή - $Bin(n,p)$, $p\in(0,1)$.

  $X:=$ πλήθος επιτυχιών σε n δοκιμές

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

  $E(X)=np$, $Var(X)=np(1-p)$

  $X\sim Bin(n,p)$, $Y\sim Bin(m,p)$ ανεξάρτητες

  $\Rightarrow X+Y\sim Bin(n+m,p)$.

- Γεωμετρική Κατανομή - $Geom(p)$, $p\in(0,1)$.

  $X:=$ πλήθος δοκιμών μέχρι την πρώτη επιτυχία

  $$p_X(x)=p(1-p)^{x-1},\ x\in\{1,2,\dots\}.$$

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

  $X,Y\sim Geom(p)$ ανεξάρτητες $\Rightarrow$

  $X+Y\sim NegBin(2,p)$

- Αρνητική Διωνυμική Κατανομή - $NegBin(r,p)$

  $r\in\mathbb{N}$, $p\in(0,1)$.

  $X:=$ πλήθος δοκιμών ως την r-οστή επιτυχία, $r>1$.

$$P_X(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r},$$
$$x \in \{r, r+1, \dots\}$$

$$E(X) = \frac{r}{p}, \ Var(X) = \frac{rq}{p^2}.$$

$X \sim NegBin(r,p)$, $Y \sim NegBin(m,p)$
ανεξάρτητες $\Rightarrow X+Y \sim NegBin(r+m,p)$.

• Κατανομή Poisson – $Poisson(\lambda)$, $\lambda > 0$.
$X :=$ πλήθος γεγονότων εντός συγκεκριμένου χρονικού διαστήματος

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

$$E(X) = Var(X) = \lambda.$$

$X \sim Poisson(\lambda)$, $Y \sim Poisson(\mu)$, ανεξάρτητες
$\Rightarrow X+Y \sim Poisson(\lambda+\mu)$.

1.2 Συνεχείς Τυχαίες Μεταβλητές

Έστω X τ.μ. συνεχής με πεδίο τιμών ένα διάστημα του $\mathbb{R}$.

Συνάρτηση πυκνότητας πιθανότητας:
$f_X : \mathbb{R} \mapsto [0,\infty)$ με $P(X \in A) = \int_A f_X(x)dx$
$\forall A \subseteq \mathbb{R}$.

Αθροιστική Συνάρτηση Κατανομής

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(y)\,dy, \quad x \in \mathbb{R}.$$

Ιδιότητες Συνάρτησης πυκνότητας πιθανότητας και Κατανομής

(i) $f_X(x) \geq 0, x \in \mathbb{R}$ (ii) $\int_{\mathbb{R}} f_X(x)\,dx = 1$

(iii) $\int_a^b f_X(x)\,dx = P(a < X < b) = P(a \leq X \leq b)$
$= P(a < X \leq b) = P(a \leq X < b)$

(iv) $P(X = x) = 0, \forall x \in \mathbb{R}.$

(v) $f_X(x) = F_X'(x) \ \forall x \in \mathbb{R}.$

(vi) $\lim_{x \to -\infty} F_X(x) = 0, \ \lim_{x \to +\infty} F_X(x) = 1$

$F_X(x)$ γνησίως αύξουσα στο σύνολο

$$S' = \{x \in \mathbb{R} \mid f_X(x) > 0\}.$$

$F_X(x)$ δεξιά συνεχής $\forall x \in \mathbb{R}.$
δηλ. $\forall x_0 \in \mathbb{R}: \lim_{x \to x_0^+} F_X(x) = F_X(x_0).$

Μέση τιμή. Αν $\int_{\mathbb{R}} |x| f_X(x)\,dx < \infty$, τότε:

$$E(X) = \int_{\mathbb{R}} x f_X(x)dx.$$

Διασπορά. Αν $\int_{\mathbb{R}} x^2 f_X(x)dx < \infty$, τότε:

$$Var(X) = E[X - E(X)]^2 = E(X^2) - (E(X))^2.$$

Ισχύει ότι: $E[g(X)] = \int_{\mathbb{R}} g(x) f_X(x) dx.$

Ανεξαρτησία. X, Y ανεξάρτητες <u>ανν</u>

$$P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B)$$

$$\forall A, B \subseteq \mathbb{R} \Leftrightarrow f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$$

$$\forall x, y \in \mathbb{R}.$$

Συνάρτηση Gamma

$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-t} dt, \ a > 0$$

Ιδιότητες συνάρτησης Gamma

(i) $\Gamma(a) = (a-1)\Gamma(a-1), \ a > 1.$

(ii) $\Gamma(n) = (n-1)!, \ n \in \mathbb{N}.$

Ειδικές Συνεχείς κατανομές

- Συνεχής Ομοιόμορφη Κατανομή - $U[a,b]$, $a < b$. $X :=$ τυχαία επιλογή αριθμού στο $[a,b]$.

$$f_X(x)=\frac{1}{b-a},\ x\in[a,b],\ F_X(x)=\begin{cases}0, & x<a\\ \frac{x-a}{b-a}, & a\le x\le b\\ 1, & x>b\end{cases}$$

$$E(X)=\frac{a+b}{2},\ Var(X)=\frac{(b-a)^2}{12}.$$

Αν $X\sim \mathcal{U}[a,b]$ τότε $\mathcal{U}=\frac{X-a}{b-a}\sim \mathcal{U}[0,1]$.

- Εκθετική Κατανομή $-Exp(\lambda),\ \lambda>0$.

$X:=$ χρόνος μεταξύ δύο γεγονότων.

$$f_X(x)=\lambda e^{-\lambda x},\ x>0.$$

$$F_X(x)=1-e^{-\lambda x},\ x>0,\ E(X)=\frac{1}{\lambda},$$
$$Var(X)=\frac{1}{\lambda^2}.$$

Αν $X\sim Exp(\lambda)\Rightarrow bX\sim Exp\left(\frac{\lambda}{b}\right),\ b>0.$

Αν $X\sim Exp(\lambda),\ Y\sim Exp(\mu)$, ανεξάρτητες

$\Rightarrow \min\{X,Y\}\sim Exp(\lambda+\mu)$

Αν $X,Y\sim Exp(\lambda)$ ανεξάρτητες $\Rightarrow$

$X+Y\sim Gamma(2,\lambda)$.

- Κατανομή Gamma $-Gamma(a,\lambda),\ a>0,\ \lambda>0$

$$f_X(x)=\frac{\lambda^a}{\Gamma(a)}x^{a-1}e^{-\lambda x},\ x>0.$$

$E(X)=\frac{a}{\lambda}$, $Var(X)=\frac{a}{\lambda^2}$.

Αν $X\sim Gamma(a,\lambda) \Rightarrow bX\sim Gamma(a,\frac{\lambda}{b})$
$b>0$.

Αν $X\sim Gamma(a,\lambda)$, $Y\sim Gamma(b,\lambda)$
ανεξάρτητες $\Rightarrow X+Y\sim Gamma(a+b,\lambda)$.

• Κανονική κατανομή $-N(\mu,\sigma^2)$, $\mu\in\mathbb{R}$, $\sigma^2>0$.

$$f_X(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\cdot\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, x\in\mathbb{R}.$$

$E(X)=\mu$, $Var(X)=\sigma^2$.

Αν $X\sim N(\mu,\sigma^2)\Rightarrow Z=\frac{X-\mu}{\sigma}\sim N(0,1)$.

Αν $Z\sim N(0,1)\Rightarrow X=\sigma Z+\mu\sim N(\mu,\sigma^2)$

Αν $X\sim N(\mu_1,\sigma_1^2)$, $Y\sim N(\mu_2,\sigma_2^2)$, ανεξάρτητες $\Rightarrow X+Y\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$.

• Κατανομή Beta $-Beta(a,b)$, $a>0$, $b>0$.

$$f_X(x)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}, x\in(0,1)$$

$$E(X)=\frac{a}{a+b},\ Var(X)=\frac{ab}{(a+b+1)(a+b)^2}$$

Αν $X\sim Beta(a,b)\Rightarrow 1-X\sim Beta(b,a)$.

Αν $X \sim Beta(a,1) \Rightarrow Y = -\log X \sim Exp(a)$.

Αν $X \sim Beta(1,a) \Rightarrow Y = -\log(1-X) \sim Exp(a)$

Αν $Y \sim Exp(a) \Rightarrow X_1 = e^{-Y} \sim Beta(a,1)$

και $X_2 = 1 - e^{-Y} \sim Beta(1,a)$.

Για την απόδειξη όλων των παραπάνω αποτελεσμάτων για τις διάφορες κατανομές, είναι χρήσιμος ο ορισμός της ροπογεννήτριας συνάρτησης, για μία τ.μ. X.

$$M_X(t) = E(e^{tX}) = \begin{cases} \sum_{x \in R_X} e^{tx} \cdot p_X(x) \\ \int_{\mathbb{R}} e^{tx} \cdot f_X(x)\,dx \end{cases}.$$

π.χ. Έστω X τ.μ.

$Bernoulli(p) \Rightarrow M_X(t) = pe^t + 1 - p$, $t \in \mathbb{R}$.

$Bin(n,p) \Rightarrow M_X(t) = (pe^t + 1 - p)^n$, $t \in \mathbb{R}$.

$Geom(p) \Rightarrow M_X(t) = \dfrac{pe^t}{1-(1-p)e^t}$, $t < -\log(1-p)$

$NegBin(r,p) \Rightarrow M_X(t) = \left[\dfrac{pe^t}{1-(1-p)e^t}\right]^r$, $t < -\log(1-p)$.

$\text{Poisson}(\lambda) \Rightarrow M_X(t) = \exp\{\lambda(e^t - 1)\}$

$\mathcal{U}[a,b] \Rightarrow M_X(t) = \begin{cases} \dfrac{e^{bt} - e^{at}}{t(b-a)}, & t \neq 0 \\ 1, & t = 0 \end{cases}$

$\text{Exp}(\lambda) \Rightarrow M_X(t) = \dfrac{\lambda}{\lambda - t}, \ t < \lambda.$

$\text{Gamma}(a,\lambda) \Rightarrow M_X(t) = \left(\dfrac{\lambda}{\lambda - t}\right)^a, \ t < \lambda.$

$N(\mu,\sigma^2) \Rightarrow M_X(t) = \exp\left\{\mu t + \dfrac{\sigma^2 t^2}{2}\right\}, \ t \in \mathbb{R}.$

Χρήσιμη στη Μαθηματική Στατιστική είναι η δείκτρια συνάρτηση:

$I_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A. \end{cases}$

Επίσης έχουμε και τη δείκτρια τ.μ.

$X = I_A(Y) = \begin{cases} 1, & \text{αν } Y \in A \\ 0, & \text{αν } Y \notin A \end{cases}$

Τότε: $E(X) = 1 \cdot P(Y \in A) + 0 \cdot P(Y \notin A)$
$= P(Y \in A).$

1.3 Ορισμοί και Ιδιότητες

Συνδιακύμανση. Αν $E(XY) < \infty$, τότε

$Cov(X,Y) = E[(X - E(X))(Y - E(Y))]$

$= E(XY) - E(X) \cdot E(Y)$

Ιδιότητες Μέσης Τιμής

(i) $E(aX+b) = aE(X)+b$

(ii) $E(aX+bY) = aE(X)+bE(Y)$

(iii) Αν X, Y ανεξάρτητες $\Rightarrow E(XY) = E(X) \cdot E(Y)$

(iv) $a \leq X \leq b \Rightarrow a \leq E(X) \leq b$

(v) $X \leq Y \Rightarrow E(X) \leq E(Y)$

Ιδιότητες Διασποράς

(i) $Var(X) \geq 0$

(ii) $Var(aX+b) = a^2 Var(X)$

(iii) $Var(aX+bY) = a^2 Var(X) + b^2 Var(Y) + 2ab\,Cov(X,Y)$

(iv) Αν X, Y ανεξάρτητες

$\Rightarrow Var(aX+bY) = a^2 Var(X) + b^2 Var(Y)$

(v) Αν X, Y ανεξάρτητες

$\Rightarrow Var(XY) = E(X^2)E(Y^2) - (E(X)E(Y))^2$

Ιδιότητες Συνδιακύμανσης

(i) $Cov(X,a) = 0$, (ii) $Cov(X,X) = Var(X)$

(iii) $Cov(Y,X) = Cov(X,Y)$

(iv) $Cov(aX+b, cY+d) = ac\,Cov(X,Y)$

(v) $Cov(X+Y, Z+W)$
$= Cov(X,Z) + Cov(X,W) + Cov(Y,Z)$
$+ Cov(Y,W)$.

(vi) Αν X, Y ανεξάρτητες $\Rightarrow Cov(X,Y)=0$.

Δεσμευμένη μέση τιμή της X δοθέντος ότι $Y=y$.

$$m_{X|Y}(y) = E(X|Y=y)$$

$$= \begin{cases} \sum_{x \in R_X} x \cdot P_{X|Y}(x|y), & X, Y \text{ διακριτές} \\ \int_{\mathbb{R}} x f_{X|Y}(x|y), & X, Y \text{ συνεχείς} \end{cases}$$

όπου $P_{X|Y}(x|y) = \dfrac{P_{X,Y}(x,y)}{P_Y(y)}$, X, Y διακριτές

και $f_{X|Y}(x|y) = \dfrac{f_{X,Y}(x,y)}{f_Y(y)}$, X, Y συνεχείς.

Η δεσμευμένη μέση τιμή της X δοθείσης της Y: $E(X|Y) = m_{X|Y}(Y)$.

Ισχύει ότι:

$$E(X) = E[E(X|Y)] = E[m_{X|Y}(Y)].$$

$$= \begin{cases} \sum\limits_{y \in R_Y} m_{X|Y}(y) P_Y(y), & Y \text{ διακριτή} \\ \int\limits_{\mathbb{R}} m_{X|Y}(y) f_Y(y)\,dy, & Y \text{ συνεχής} \end{cases}$$

Κάποιες από τις ιδιότητες των ροπογεννητριών συνοψίζονται παρακάτω:

(i) $M_X(t) = M_Y(t) \ \forall t \in \mathbb{R}$ $\underline{\underline{\text{ανν}}}$ X, Y ισόνομες (προέρχονται από την ίδια κατανομή με τις ίδιες τιμές παραμέτρων).

(ii) $M_X(0) = 1$

(iii) $M_X^{(k)}(0) = E(X^k)$, $k \in \mathbb{N}$.

(iv) Αν X, Y ανεξάρτητες $\Rightarrow M_{X+Y}(t) = M_X(t) M_Y(t)$

Βασικές Ανισότητες:

(i) Markov: $X \geq 0 \Rightarrow P(X > a) \leq \dfrac{E(X)}{a}$, $a > 0$

(ii) Chebyshev: $P(|X - E(X)| > a) \leq \dfrac{Var(X)}{a^2}$, $a > 0$

(iii) Ανισότητα Cauchy-Schwarz:

$$[E(XY)]^2 \le E(X^2)\cdot E(Y^2)$$

(iv) $[Cov(X,Y)]^2 \le Var(X)\cdot Var(Y)$. ⊠

1.4 Αποτελέσματα από Θεωρία Πιθανοτήτων

Ισχύουν τα ακόλουθα:

(1) Αν $X_1,\dots,X_n$ ανεξάρτητες τ.μ. τότε:

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i).$$

(2) Αν $X_1,\dots,X_n$ ανεξάρτητες τ.μ. τότε:

Αν $X_i \sim N(\mu_i, \sigma_i^2)$ και $a_i \in \mathbb{R}$, $i=1,\dots,n$ σταθερές τότε $\sum_{i=1}^{n} a_i X_i \sim N\left(\sum_{i=1}^{n} a_i\mu_i, \sum_{i=1}^{n} a_i^2\sigma_i^2\right)$.

Αν $X_i \sim Bin(m,p) \Rightarrow \sum_{i=1}^{n} X_i \sim Bin(nm, p)$

Αν $X_i \sim Poisson(\lambda_i)$, $i=1,\dots,n$ $\Rightarrow \sum_{i=1}^{n} X_i \sim Poisson\left(\sum_{i=1}^{n}\lambda_i\right)$.

(3) Αν $X \sim Gamma(a,b) \Rightarrow cX \sim Gamma\left(a, \frac{b}{c}\right)$, $c>0$ σταθερά.

Αν $X_i \sim Gamma(a_i, b)$, $i=1,\dots,n$, τότε:

$\sum_{i=1}^{n} X_i \sim Gamma\left(\sum_{i=1}^{n} a_i, b\right)$, X_i ανεξάρτητες.

(4) Αν $X_1,\dots,X_n$ ανεξάρτητες με $X_i \sim N(\mu,\sigma^2)$ $i=1,\dots,n$ τότε $\sum_{i=1}^{n}(X_i-\mu)^2/\sigma^2 \sim \chi^2_n$ και

$$\sum_{i=1}^{n}(X_i-\bar{X})^2/\sigma^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1},$$ όπου

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2, \quad \bar{X} = \sum_{i=1}^{n} X_i/n.$$ ⊠