Distribuţii de probabilitate clasice

Principalele legi de probabilitate

Radu Trîmbiţaş

December 2012

Conţinut

Cuprins

1 Distribuţii discrete

1.1 Repartiţia binomială

Repartiţia binomială

- Variabila aleatoare X având distribuţia

  $$X : \begin{pmatrix} k \\ \binom{n}{k} p^k q^{n-k} \end{pmatrix}_{k=\overline{1,n}}, \ q = 1 - p$$

  se numeşte variabilă aleatoare ***binomială***. Repartiţia astfel determinată se numeşte ***repartiţie binomială*** de ordinul n şi parametru p.

- Mulţimea tuturor variabilelor aleatoare binomiale de ordinul n şi parametru p se va nota cu $b(p, n)$.

Observaţia 1. *Dacă*

$$X_k : \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}$$

atunci $Y = X_1 + X_2 + \cdots + X_n \in b(p, n)$ *(*Y *este o variabilă aleatoare care ia valoarea* k *dacă evenimentul* $A := (X_k = 1)$ *s-a produs de* k *ori).*

Teorema 2. *Valoarea medie şi dispersia unei variabile aleatoare binomiale de ordin* n *şi parametru* p *sunt*

$$M(X) = np, \quad D^2(X) = npq.$$

Demonstraţie. Demonstraţia se bazează pe observaţia 1. Dacă

$$X_k : \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix},$$

atunci $M(X_k) = p$, $D^2(X_k) = pq$ şi deoarece X_k sunt independente, avem

$$M(X) = M(Y) = M\left(\sum_{k=1}^{n} X_k\right) = np,$$

$$D^2(X) = D^2(Y) = D^2\left(\sum_{k=1}^{n} X_k\right) = npq.$$

■

Funcţia caracteristică a repartiţiei binomiale este

$$\begin{aligned} g(t) = M\left(e^{itX}\right) &= \sum_{k=0}^{n} e^{itk} \binom{n}{k} p^k q^{n-k} \\ &= \sum_{k=0}^{n} \binom{n}{k} (pe^{it})^k q^{n-k} = (pe^{it} + q)^n. \end{aligned}$$

1.2 Repartiţia Poisson

Repartiţia Poisson

- Repartiţia discretă determinată de probabilităţile

$$P(k; \lambda) = \frac{\lambda^k}{k!} e^{-\lambda}, \tag{1}$$

unde $k \in \mathbb{N}$ şi $\lambda > 0$ se numeşte ***repartiţie Poisson*** de parametru λ. Variabila aleatoare cu distribuţia

$$X : \begin{pmatrix} k \\ \frac{\lambda^k}{k!} e^{-\lambda} \end{pmatrix}_{k \in \mathbb{N}} \tag{2}$$

se numeşte ***variabilă aleatoare Poisson***.

- Vom nota cu $Po(\lambda)$ mulţimea variabilelor aleatoare Poisson de parametru λ.
- Relaţia (1) se poate obţine pornind de la o repartiţie binomială. Punem $np = \lambda$ şi trecând la limită pentru $n \to \infty$ (menţinând λ constant) avem

$$\lim_{n\to\infty} P(n;\lambda) = \lim_{n\to\infty} \frac{n(n-1)\dots(n-k+1)}{k!} \cdot \frac{\lambda^k}{n^k}(1-\frac{\lambda}{n})^{n-k} = \\ = \frac{\lambda^k}{k!}e^{-\lambda}$$

căci

$$\lim_{n\to\infty} \frac{n(n-1)\dots(n-k+1)}{n^k} = 1 \text{ şi } \lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n-k} = e^{-\lambda}.$$

- Din acest motiv repartiţia Poisson se mai numeşte şi *legea evenimentelor rare*.

Teorema 3. *Valoarea medie şi dispersia unei variabile aleatoare Poisson de parametru λ sunt $M(X) = \lambda$ şi $D^2(X) = \lambda$.*

Demonstraţie. Deoarece

$$M(X) = \sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} = \lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda}e^{\lambda} = \lambda$$

şi

$$M(X^2) = \sum_{k=0}^{\infty} k^2\frac{\lambda^k}{k!}e^{-\lambda} = \sum_{k=0}^{\infty}(k^2-k+k)\frac{\lambda^k}{k!}e^{-\lambda} = \\ = \sum_{k=1}^{\infty} k(k-1)\frac{\lambda^k}{k!}e^{-\lambda} + \sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda} = \lambda^2 e^{-\lambda}e^{\lambda} + \lambda = \lambda^2 + \lambda,$$

obţinem $D^2(X) = M(X^2) - [M(X)]^2 = \lambda$. ∎

- Funcţia caracteristică a unei variabile aleatoare Poisson are expresia

$$g(t) = \sum_{k=0}^{\infty} e^{ikt}\frac{\lambda^k}{k!}e^{-\lambda} = e^{\lambda(e^{it}-1)}.$$

- Cu ajutorul funcţiei caracteristice se verifică uşor că suma a două variabile aleatoare Poisson de parametru λ şi respectiv μ este o variabilă aleatoare Poisson de parametru $\lambda + \mu$.

2 Distribuţii continue

2.1 Repartiţia uniformă

Repartiţia uniformă

- O variabilă aleatoare X având densitatea de probabilitate

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{pentru} \quad x \in [a,b] \\ 0 & \text{pentru} \quad x \notin [a,b] \end{cases} \tag{3}$$

  se numeşte **variabilă aleatoare uniformă** pe $[a,b]$, iar densitatea de probabilitate şi funcţia de repartiţie corespunzătoare se numesc densitate de probabilitate şi respectiv funcţie de repartiţie uniformă pe $[a,b]$.

- Mulţimea variabilelor aleatoare uniforme pe $[a,b]$ se notează cu $U[a,b]$.

- Se verifică imediat că funcţia de repartiţie uniformă pe $[a,b]$ este

$$F(x) = \begin{cases} 0, & \text{pentru} \quad x \leq a \\ \frac{x-a}{b-a}, & \text{pentru} \quad a < x \leq b \\ 1, & \text{pentru} \quad x > b. \end{cases}$$

  Graficele lui f şi F apar în figura 1.

Teorema 4. *Valoarea medie şi dispersia unei variabile aleatoare uniforme sunt*

$$M(X) = \frac{a+b}{2}, \;\; D^2(X) = \frac{(b-a)^2}{12}.$$

Demonstraţie.

$$M(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{-\infty}^{a} xf(x)dx + \int_{a}^{b} xf(x)dx + \int_{b}^{\infty} xf(x)dx =$$

$$= \int_{a}^{b} xf(x)dx = \frac{1}{b-a}\frac{x^2}{2}\bigg|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}.$$

De asemenea

$$M(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \frac{1}{b-a}\int_a^b x^2 dx = \frac{a^2+ab+b^2}{3},$$

de unde

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

■

- Ţinând cont că $b > a$ avem $D(X) = \frac{b-a}{2\sqrt{3}}$.

- Funcţia caracteristică pentru $X \in U[a,b]$ este

$$g(t) = \frac{e^{itb} - e^{ita}}{it(b-a)}.$$

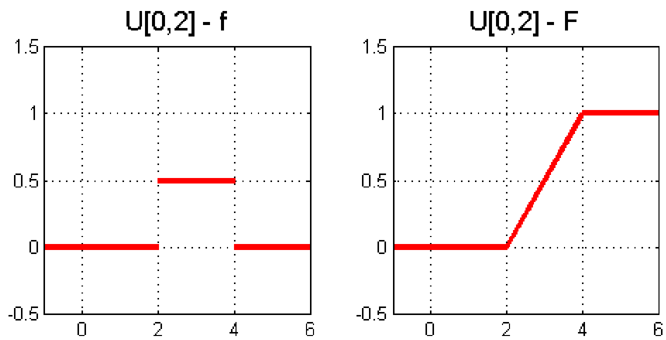


Figura 1: Densitatea de probabilitate şi funcţia de repartiţie pentru $U[0,2]$

2.2 Repartiţia normală

Repartiţia normală

- Spunem că variabila aleatoare X urmează ***legea normală*** de parametri m şi σ^2 (sau uneori m şi σ), $m \in \mathbb{R}$, $\sigma > 0$ dacă densitatea sa de probabilitate este

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}. \tag{4}$$

- Mulţimea variabilelor aleatoare ce urmează legea normală de parametri m şi σ^2 se va nota cu $N(m, \sigma^2)$.

- Funcţia de repartiţie este

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-m)^2}{2\sigma^2}} dt.$$

- Dacă $m = 0$ şi $\sigma^2 = 1$ se obţine

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \tag{5}$$

şi spunem că X urmează ***repartiţia normală standard*** sau ***redusă***, notată cu $N(0,1)$.

Teorema 5. *Media şi dispersia unei variabile aleatoare normale reduse sunt $M(X) = 0$ şi $D^2(X) = 1$.*

Demonstraţie.

$$M(X) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} x e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{x^2}{2}} dx = 0,$$

deoarece funcţia de integrat este impară.

$$D^2(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-m)^2 e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2}} dx.$$

Integrala se calculează prin părţi, luând $u = x$ şi $dv = xe^{-\frac{x^2}{2}}$, obţinându-se

$$D^2(X) = \frac{1}{\sqrt{2\pi}} \left[-xe^{-\frac{x^2}{2}} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \right] = 1,$$

deoarece

$$\int_{0}^{\infty} e^{-x^2} dx = \sqrt{\pi}. \qquad \text{(integrala lui Poisson)}$$

■

Corolarul 6. *Media şi dispersia unei variabile aleatoare normale, de parametri m şi σ^2 sunt $M(X) = m$ şi $D^2(X) = \sigma^2$.*

Demonstraţie. Temă. ■

Observaţia 7. *Parametri m şi σ^2 ai repartiţiei normale reprezintă valoarea medie şi respectiv dispersia unei variabile aleatoare repartizată $N(m, \sigma^2)$. Rezultă totodată că funcţia de repartiţie a unei variabile aleatoare normale este perfect determinată de valoarea medie şi dispersia variabilei.*

Momentele de ordin impar ale repartiţiei normale standard sunt nule, deoarece funcţia de integrat care intervine la calculul lor este impară.

$$M_{2k+1}(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{2k+1} e^{-\frac{x^2}{2}} dx = 0.$$

Pentru momentele de ordin par avem

$$\begin{aligned} M_{2k}(X) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{2k} e^{-\frac{x^2}{2}} dx = -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^{2k-1} \left(e^{-\frac{x^2}{2}} \right)' dx = \\ &= (2k-1) M_{2k-2}(x). \end{aligned}$$

Deci am obţinut

$$M_{2k+1}(X) = 0 \tag{6}$$
$$M_{2k}(X) = (2k-1)!! = \frac{(2k)!}{2^k k!}.$$

Asimetria şi excesul pentru o variabilă aleatoare normală X sunt

$$As = \frac{M_3(X)}{\sigma^3} = 0,\ E = \frac{M_4(X)}{\sigma^4} - 3 = 0.$$

În faptul că asimetria şi excesul repartiţiei $N(m, \sigma^2)$ sunt nule îşi are originea procedeul statisticii descriptive de a considera aceste caracteristici drept criterii de stabilire a normalităţii. Totuşi, condiţia $As = E = 0$ este o condiţie necesară, dar nu suficientă pentru normalitate.

Funcţia de repartiţie a unei variabile aleatoare normale standard este

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt \tag{7}$$

şi se numeşte *funcţia lui Laplace*. Deoarece densitatea de probabilitate normală standard este simetrică faţă de dreapta $x = 0$ rezultă $\Phi(-x) = 1 - \Phi(x)$.

Cu ajutorul funcţiei Φ, prin schimbarea de variabilă $u = \frac{x-m}{\sigma}$ se determină probabilităţilor ce privesc orice variabilă aleatoare normală:

$$\begin{aligned} P(a < X < b) &= \int_a^b f(x)dx = \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx = \\ &= \frac{1}{\sqrt{2\pi}} \left[\int_{-\infty}^{\frac{b-m}{\sigma}} e^{-\frac{u^2}{2}} du - \int_{-\infty}^{\frac{a-m}{\sigma}} e^{-\frac{u^2}{2}} du \right] = \\ &= \Phi\left(\frac{b-m}{\sigma}\right) - \Phi\left(\frac{a-m}{\sigma}\right). \end{aligned}$$

Reprezentarea grafică

Interpretarea geometrică a funcţiei lui Laplace

Regula celor 3 sigma

Regula celor 3 σ pentru variabile aleatoare normale.

$$\begin{aligned} P\left(|x-m| < \varepsilon\right) &= P\left(-\varepsilon + m < X < \varepsilon + m\right) = \Phi\left(\frac{\varepsilon}{\sigma}\right) - \Phi\left(-\frac{\varepsilon}{\sigma}\right) = \\ &= 2\Phi\left(\frac{\varepsilon}{\sigma}\right) - 1. \end{aligned}$$

Luând $\varepsilon = 3\sigma$, obţinem

$$p = P\left(|x-m| < \varepsilon\right) = 2\Phi(3) - 1.$$

În tabele găsim $\Phi(3) = 0.9987$, de unde $p = 0.9974$. Cu alte cuvinte, probabilitatea ca abaterea în valoare absolută să depăşească 3σ este $1 - 0.9974 = 0.0026$, adică practic 0.

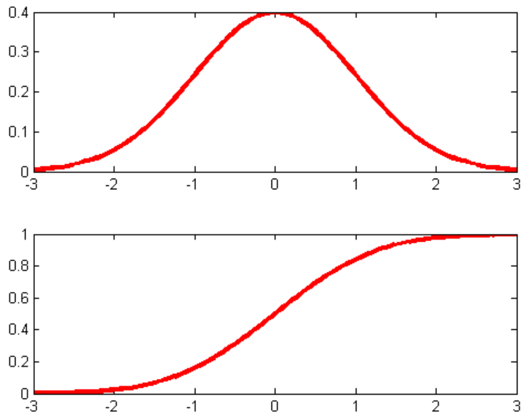
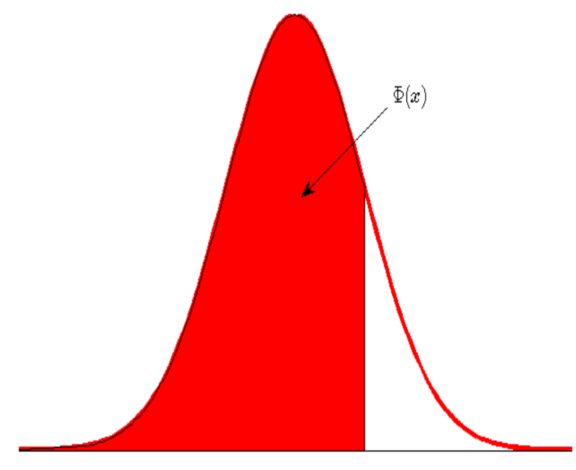


Figura 2: Densitatea de probabilitate (sus) şi funcţia de repartiţie $N(0, 1)$



Funcţia caracteristică

Vom calcula întâi funcţia caracteristică a unei variabile aleatoare normale standard. Avem

$$g(t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{itx} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \sum_{n=0}^{\infty} \frac{t^n i^n}{n!} x^n e^{-\frac{x^2}{2}} dx.$$

Deoarece seria din membrul drept este uniform convergentă putem permuta suma cu integrala, obţinând

$$g(t) = \sum_{n=0}^{\infty} \frac{t^n i^n}{n!} \cdot \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x^n e^{-\frac{x^2}{2}} dx = \sum_{n=0}^{\infty} \frac{t^n i^n}{n!} M_n(X).$$

Ţinând cont de expresiile (6) ale momentelor repartiţiei normale standard avem în final

$$g(t) = \sum_{m=0}^{\infty} \frac{t^{2m} i^{2m}}{(2m)!} \cdot \frac{(2m)!}{2^m m!} = \sum_{m=0}^{\infty} \frac{(-t^2/2)^m}{m!} = e^{-\frac{t^2}{2}}.$$

Pentru o variabilă aleatoare $X \in N(m, \sigma^2)$, deoarece $Y = \frac{x-m}{\sigma}$ este normală standard obţinem

$$g_X(t) = g_{\sigma Y + m}(t) = e^{itm} g_Y(\sigma t) = e^{itm - \frac{\sigma^2 t^2}{2}}.$$

Cu ajutorul funcţiei caracteristice se poate demonstra:

Teorema 8. *Dacă $a_1, \dots, a_n$ sunt constante şi $X_1, \dots, X_n$ sunt variabile aleatoare independente, repartizate $N(m_k, \sigma_k^2)$, $k = \overline{1,n}$, atunci variabila aleatoare $Y = a_1 X_1 + \dots + a_n X_n$ este repartizată $N\left(\sum_{k=1}^n a_k m_k, \sum_{k=1}^n a_k^2 \sigma_k^2\right)$.*

2.3 Familia de repartiţii gama

Familia de repartiţii gama

- Spunem că variabila aleatoare X urmează ***legea gama*** de parametri $r > 0$ şi $\lambda > 0$ dacă densitatea sa de probabilitate este

$$f(x; r, \lambda) = \begin{cases} \frac{\lambda e^{-\lambda x} (\lambda x)^{r-1}}{\Gamma(r)}, & \text{pentru} \quad x > 0 \\ 0, & \text{pentru} \quad x \leq 0, \end{cases} \tag{8}$$

  unde $\Gamma(r)$ reprezintă valoarea funcţiei gama a lui Euler în r.

- Mulţimea variabilelor aleatoare ce urmează legea gama de parametri r şi λ se va nota cu $\gamma(r, \lambda)$.

- Pentru $\lambda = 1$ se obţine repartiţia ***gama standard*** de parametru r, având densitatea de probabilitate

$$f(x; r) = \begin{cases} \frac{e^{-x} x^{r-1}}{\Gamma(r)}, & \text{pentru} \quad x > 0 \\ 0, & \text{pentru} \quad x \leq 0. \end{cases}$$

- Mulţimea variabilelor aleatoare ce urmează legea gama standard de parametru r se va nota cu $\gamma^*(r)$.

Observaţia 9. *Are loc* $X \in \gamma(r, \lambda) \Rightarrow aX \in \gamma(r, \frac{\lambda}{a})$.

Să determinăm funcţia caracteristică a unei variabile aleatoare $X \in \gamma^*(r)$.

$$
\begin{aligned}
g(t) &= \frac{1}{\Gamma(r)} \int_0^\infty e^{itx} e^{-x} x^{r-1} dx = \frac{1}{\Gamma(r)} \int_0^\infty \left(\sum_{n=0}^{\infty} \frac{(itx)^n}{n!} \right) e^{-x} x^{r-1} dx \\
&= \frac{1}{\Gamma(r)} \sum_{n=0}^{\infty} \frac{(it)^n}{n!} \int_0^\infty e^{-x} x^{n+r-1} dx = \frac{1}{\Gamma(r)} \sum_{n=0}^{\infty} \frac{\Gamma(n+r)}{n!} (it)^n \\
&= 1 + \sum_{n=0}^{\infty} \frac{r(r+1)\dots(r+n+1)}{n!} (it)^n = (1-it)^{-r}.
\end{aligned}
$$

Deci $g(t) = (1-it)^{-r}$.

Dacă $X \in \gamma(r, \lambda)$, atunci aX are densitatea de probabilitate $f(\frac{x}{a})$; deci conform observaţiei 9 $aX \in \gamma\left(r, \frac{\lambda}{a}\right)$ şi

$$
g_X(t) = \left(1 - \frac{it}{\lambda}\right)^{-r} = \frac{\lambda^r}{(\lambda - it)^r} \,,
$$

căci dacă $X \in \gamma(r, \lambda)$, $\lambda X \in \gamma^*(r)$.

Momentele de ordinul s ale lui $X \in \gamma(r, \lambda)$ sunt

$$
M_s(X) = \frac{1}{\Gamma(r)} \int_0^\infty x^s \lambda e^{-\lambda x} (\lambda x)^{r-1} dx = \frac{1}{\lambda^s \Gamma(r)} \int_0^\infty e^{-y} y^{r+s+1} dy
$$

şi deci

$$
M_s(X) = \frac{\Gamma(r+s)}{\lambda^s \Gamma(r)},
$$

de unde

$$
M(X) = \frac{r}{\lambda}, \;\; M_2(X) = \frac{r(r+1)}{\lambda^2}, \;\; D^2(X) = \frac{r}{\lambda^2}.
$$

Observaţia 10. *Pentru* $r = 1$, $\gamma(1, \lambda)$ *devine* ***legea exponenţială*** *de parametru* λ. *Densitatea sa de probabilitate este*

$$
f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x}, & \text{pentru} \quad x > 0 \\ 0, & \text{pentru} \quad x \leq 0. \end{cases}
$$

Prin calcul se obţine

$$
F(x; \lambda) = \begin{cases} 1 - e^{-\lambda x}, & \text{pentru} \quad x > 0 \\ 0, & \text{pentru} \quad x \leq 0, \end{cases}
$$

iar prin particularizarea rezultatelor de mai sus

$$
M(X) = M_1(X) = \frac{1}{\lambda}, \; M_2(X) = \frac{2}{\lambda^2}, \; D^2(X) = \frac{1}{\lambda^2}, \; g(x) = \frac{\lambda}{\lambda - it}.
$$

Teorema 11. *Dacă* $X \in N(m, \sigma^2)$ *atunci* $Y = \frac{(X-m)^2}{2\sigma^2} \in \gamma^*(\frac{1}{2})$.

Demonstraţie. Avem

$$P(Y < x) = P\left(-\sqrt{x} < \frac{X-m}{\sigma\sqrt{2}} < \sqrt{x}\right) = P\left(-\sqrt{2x} < \frac{X-m}{\sigma} < \sqrt{2x}\right).$$

Deoarece $\frac{X-m}{\sigma} \in N(0,1)$, obţinem

$$P(Y < x) = \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{2x}}^{\sqrt{2x}} e^{-\frac{u^2}{2}} du = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{2x}} e^{-\frac{u^2}{2}} du.$$

Efectuând schimbarea de variabilă $u^2 = 2v$ şi ţinând cont că $\Gamma(1/2) = \sqrt{\pi}$, rezultă că

$$P(Y < x) = \frac{1}{\Gamma(\frac{1}{2})} \int_0^x e^{-v} v^{-\frac{1}{2}} dv = F_Y(x),$$

adică

$$f_Y(x) = \begin{cases} \dfrac{e^{-x} x^{\frac{1}{2}-1}}{\Gamma(\frac{1}{2})}, & \text{pentru} \quad x > 0 \\ 0, & \text{pentru} \quad x \leq 0 \end{cases} = f_{\gamma^*}(x; \frac{1}{2}).$$

■

2.4 Repartiţia hi-pătrat

Repartiţia hi-pătrat

Spunem că variabila aleatoare X urmează **legea** χ^2 (se citeşte **hi-pătrat**) de parametri $\nu > 0$(număr de grade de libertate) şi $\sigma^2 > 0$ dacă $X \in \gamma(\frac{\nu}{2}, \frac{1}{2\sigma^2})$. Pentru densitatea de probabilitate din (8) se obţine

$$f(x; m, \sigma^2) = \begin{cases} \dfrac{e^{-\frac{x}{2\sigma^2}} x^{\frac{\nu}{2}-1}}{(2\sigma^2)^{\frac{\nu}{2}} \Gamma(\frac{\nu}{2})}, & \text{pentru} \quad x > 0 \\ 0, & \text{pentru} \quad x \leq 0. \end{cases}$$

Această repartiţie a fost descoperită de astronomul Helmert în 1876 şi pusă în valoare 30 de ani mai târziu de Karl Pearson, motiv pentru care se mai numeşte şi **repartiţia Helmert–Pearson**. Pentru $\sigma^2 = 1$ se obţine repartiţia χ^2 standard cu $\nu > 0$ grade de libertate, având densitatea de probabilitate

$$f(x; \nu) = \begin{cases} \dfrac{e^{-\frac{x}{2}} x^{\frac{\nu}{2}-1}}{2^{\frac{\nu}{2}} \Gamma(\frac{\nu}{2})}, & \text{pentru} \quad x > 0 \\ 0, & \text{pentru} \quad x \leq 0. \end{cases}$$

Mulţimea variabilelor aleatoare ce urmează legea χ^2 de parametri ν şi σ^2 se va nota cu $\chi^2(\nu, \sigma^2)$ sau cu $X^2(\nu, \sigma^2)$. În figura 3 apar mai multe exemple de distribuţii χ^2 standard, cu diverse grade de libertate.Funcţia caracteristică este

$$g(x) = (1 - 2\sigma^2 ti)^{-\frac{\nu}{2}}$$

şi se obţine din funcţia caracteristică a repartiţiei gama (ec. 8) pentru $r = \frac{\nu}{2}$ şi $\lambda = \frac{1}{2\sigma^2}$.

Momentele de ordinul s sunt

$$M_s(X) = \frac{(2\sigma^2)^s \Gamma(s + \frac{\nu}{2})}{\Gamma(\frac{\nu}{2})}$$

şi deci $M(X) = \nu\sigma^2,\ M_2(X) = \nu(\nu+2)\sigma^4,\ D^2(X) = 2\nu\sigma^4$.

Teorema 12. *Dacă* $X_1, \ldots, X_\nu \in N(0, \sigma^2)$ *şi ele sunt independente, atunci*

$$Y = X_1^2 + \cdots + X_\nu^2 \in X^2(\nu, \sigma^2).$$

Demonstraţie. Conform teoremei 11 $X_j \in N(0, \sigma^2) \Rightarrow Y_j = \frac{X_j^2}{2\sigma^2} \in \gamma^*(\frac{1}{2}) \Leftrightarrow X_j^2 \in \gamma(\frac{1}{2}, \frac{1}{2\sigma^2}) \Leftrightarrow X_j^2 \in X^2(1, \sigma^2)$. Deoarece $X_1, \ldots, X_\nu$ sunt independente

$$g_Y(t) = g_{x_1^2}(t) \cdots g_{x_\nu^2}(t) = (1 - 2\sigma^2 ti)^{-\frac{\nu}{2}}$$

adică $Y \in X^2(\nu, \sigma^2)$. ■

Teorema 13. *Fie* $X \in X^2(\nu, \sigma^2)$. *Pentru* $\nu \to \infty$, $X^* = \frac{X - M(X)}{D(X)}$ *urmează legea normală standard* $N(0, 1)$.

Demonstraţie. Avem $M(X) = 2\nu\sigma^2$, $D^2(X) = 2\nu\sigma^4$şi deci

$$X^* = \frac{X - \nu\sigma^2}{\sqrt{2\nu\sigma^4}} = \frac{X}{\sigma^2\sqrt{2\nu}} - \sqrt{\frac{\nu}{2}},$$

din care se obţine

$$g_{X^*}(t) = e^{-it\sqrt{\frac{n}{2}}} g_X\left(\frac{t}{\sigma^2\sqrt{2\nu}}\right) = e^{-it\sqrt{\frac{\nu}{2}}}\left(1 - it\sqrt{\frac{2}{\nu}}\right)^{-\frac{\nu}{2}} =$$

$$= \left(e^{it\sqrt{\frac{2}{\nu}}}\left(1 - it\sqrt{\frac{2}{\nu}}\right)\right)^{-\frac{\nu}{2}}.$$

Dezvoltăm în serie primul factor

$$e^{it\sqrt{\frac{2}{\nu}}} = e^{\frac{it}{\sqrt{\nu/2}}} = 1 + it\sqrt{\frac{2}{\nu}} - \frac{t^2}{\nu} + \Theta\left(\nu^{-3/2}\right),$$

înlocuim în al doilea şi obţinem

$$g_{X^*}(t) = 1 + \frac{t^2}{\nu} + \Theta\left(\nu^{-3/2}\right),$$

de unde

$$g_{X^*}(t) = \exp\left(\lim_{\nu\to\infty}\left(-\frac{t^2}{2} - \Theta\left(\nu^{-1/2}\right)\right)\right),$$

adică funcţia caracteristică pentru $N(0, 1)$. ■

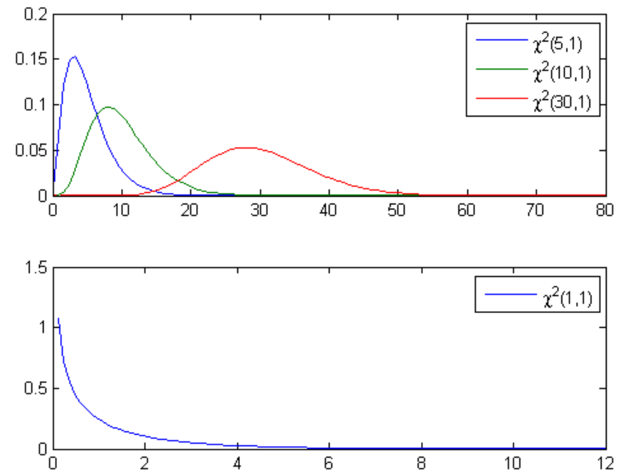


Figura 3: Densităţi de probabilitate χ^2

Reprezentarea grafică

2.5 Repartiţia Student

Repartiţia Student

- Spunem că variabila aleatoare X urmează ***legea Student*** sau ***legea T***cu $\nu > 0$ grade de libertate dacă densitatea sa de probabilitate este

$$f(x|\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}. \tag{9}$$

- Mulţimea variabilelor aleatoare ce urmează legea Student cu ν grade de libertate se notează cu $T(\nu)$ sau $S^*(\nu)$.

- Deoarece f este o funcţie pară momentele de ordin impar $M_{2k+1}(X)$ sunt nule, $M(X) = 0$, iar momentele centrate coincid cu momentele obişnuite. Pentru calculul momentelor de ordin par efectuăm substituţia $x^2 = \nu y$ şi obţinem

$$M_{2k}(X) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)} \int_{-\infty}^{\infty} x^{2k}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} dx = \tag{10}$$

$$= \frac{\nu^k\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \int_{0}^{\infty} y^{k-\frac{1}{2}}(1+y)^{-\frac{\nu+1}{2}} dy =$$

$$= \frac{\nu^k\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{\nu}{2}\right)} B\left(\frac{\nu}{2}-k, k+\frac{1}{2}\right) =$$

$$= \frac{\nu^k\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \frac{\Gamma\left(\frac{\nu}{2}-k\right)\Gamma\left(k+\frac{1}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)},\ 0 < k < \frac{\nu}{2} \tag{11}$$

Dar

$$\Gamma\left(\frac{\nu}{2}\right) = \left(\frac{\nu}{2}-1\right)\cdots\left(\frac{\nu}{2}-k\right)\Gamma\left(\frac{\nu}{2}-k\right)$$

şi

$$\Gamma\left(k+\frac{1}{2}\right) = \left(k-\frac{1}{2}\right)\cdots\frac{1}{2}\Gamma\left(\frac{1}{2}\right),$$

de unde

$$M_{2k}(X) = \frac{1\cdot 3\cdots(2k-1)}{(\nu-2)(\nu-4)\cdots(\nu-2k)}\nu^k \tag{12a}$$

cu condiţia să avem $k < \frac{\nu}{2}$.

- Deci, primele două momente de ordin par, dacă există sunt

$$M_2(X) = m_2(X) = \frac{\nu}{\nu-2},\ M_4(X) = m_4(X) = \frac{3\nu^2}{(\nu-2)(\nu-4)}.$$

- Din relaţia (12a) obţinem

$$\lim_{n\to\infty} M_{2k}(X) = (2k-1)!!$$

şi deoarece $M_{2k+1}(X) = 0$, rezultă că momentele repartiţiei Student tind la momentele repartiţiei $N(0,1)$ când $n \to \infty$. Am demonstrat astfel teorema:

Teorema 14. *Dacă $X \in T(\nu)$, atunci X este asimptotic normală standard (când $\nu \to \infty$).*

În figura 4 apar mai multe grafice de densităţi de probabilitate Student pentru diverse grade de libertate.

Reprezentarea grafică

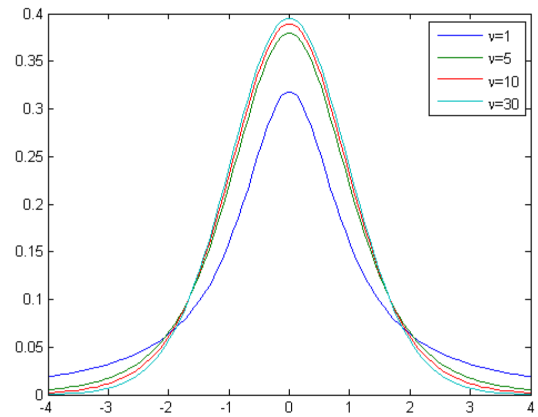


Figura 4: Densităţi de probabilitate Student pentru $\nu = 1, 5, 10, 30$ grade de libertate

Repartiţia Student

Teorema 15. *Dacă $X \in N(0, \sigma^2)$ şi $Y \in X^2(\nu, \sigma^2)$ sunt variabile aleatoare independente, atunci variabila aleatoare*

$$Z = \frac{X}{\sqrt{\frac{Y}{\nu}}}$$

urmează legea $T(\nu)$.

Demonstraţie. Densitatea de probabilitate $f(x, y)$ a vectorului aleator (X, Y) este $f(x, y) = f_X(x) f_Y(y)$. Deoarece $x = z\sqrt{\frac{y}{\nu}}$ şi $y = y$ avem

$$\frac{D(x, y)}{D(z, y)} = \begin{vmatrix} \sqrt{\frac{y}{\nu}} & \frac{z}{2\sqrt{\nu y}} \\ 0 & 1 \end{vmatrix} = \sqrt{\frac{y}{\nu}},$$

de unde rezultă că

$$f(z, y) = f_X\left(z\sqrt{\frac{y}{\nu}}\right) f_Y(y) \sqrt{\frac{y}{\nu}}.$$

Densitatea de probabilitate a lui Z este

$$f(z) = \int_0^\infty f(z, y)dy = \frac{1}{\sqrt{2\pi\nu}\Gamma\left(\frac{\nu}{2}\right)(2\sigma^2)^{\frac{\nu}{2}}} \int_0^\infty e^{-\frac{y}{2\sigma^2}\left(1+\frac{z^2}{\nu}\right)} y^{\frac{\nu-1}{2}} dy;$$

efectuând schimbarea de variabilă $y = \frac{2\sigma^2}{1+z^2/\nu}$, vom avea în final

$$f(z) = \frac{\left(1 + \frac{z^2}{\nu}\right)^{\frac{\nu+1}{2}}}{\sqrt{\nu\pi}\Gamma\left(\frac{\nu}{2}\right)} \int_0^\infty e^{-t} t^{\frac{\nu+1}{2}-1} dt = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{z^2}{\nu}\right)^{-\frac{\nu+1}{2}}.$$

■

Din teoremele 15 şi 12 rezultă

Teorema 16. *Dacă* $X_0, X_1, \ldots, X_\nu$ *sunt variabile aleatoare independente de clasă* $N(0, \sigma^2)$*, atunci variabila aleatoare*

$$Y = \frac{X_0}{\sqrt{\dfrac{X_1^2 + \cdots + X_\nu^2}{\nu}}}$$

este de clasă $T(\nu)$.

Un alt rezultat util de acest tip este

Teorema 17. *Dacă* $X_1, \ldots, X_\nu$ *sunt variabile aleatoare independente de clasă* $N(0, \sigma^2)$ *şi*

$$Y = \frac{X_1 + \cdots + X_\nu}{\nu}, \tag{13}$$

atunci variabila aleatoare

$$Z = \frac{Y}{\sqrt{\dfrac{\sum\limits_{i=1}^{\nu}(X_i - Y)^2}{\nu(\nu-1)}}} \tag{14}$$

aparţine clasei $T(\nu - 1)$.

Demonstraţie. Fie variabilele aleatoare

$$Y_i = \sum_{j=1}^{\nu} a_{ij} X_j, \quad i = \overline{1, \nu - 1} \tag{15}$$

şi

$$Y_\nu = \frac{1}{\sqrt{\nu}} (X_1 + \cdots + X_\nu) \tag{16}$$

cu coeficienţii aleşi astfel încât matricea $A = (a_{ij})$ să fie ortogonală. Deoarece $X_1, \ldots, X_\nu$ sunt variabile aleatoare independente de clasă $N(0, \sigma^2)$, rezultă că şi $Y_1, \ldots, Y_\nu$ sunt independente şi de clasă $N(0, \sigma^2)$. Mai mult

$$Y_1^2 + \cdots + Y_\nu^2 = X_1^2 + \cdots + X_\nu^2 \tag{17a}$$

de unde

$$\begin{aligned} Y_1^2 + \cdots + Y_{\nu-1}^2 &= X_1^2 + \cdots + X_\nu^2 - \frac{X_1^2 + \cdots + X_\nu^2}{\nu} = \\ &= X_1^2 + \cdots + X_\nu^2 - \nu Y^2 = \sum_{i=1}^{\nu} (X_i - Y)^2 \end{aligned} \tag{18}$$

Utilizând (13), (14), (16) şi (18) obţinem

$$Z = \frac{\frac{1}{\nu}\sqrt{\nu(\nu-1)}(X_1+\cdots+X_\nu)}{\sqrt{\sum_{i=1}^{\nu}(X_i-Y)^2}} = \frac{\frac{1}{\nu}\sqrt{\nu(\nu-1)}\sqrt{\nu}Y_\nu}{\sqrt{\sum_{i=1}^{\nu}Y_i^2}} =$$

$$= \frac{Y_\nu}{\sqrt{\frac{Y_1^2+\cdots+Y_{\nu-1}^2}{\nu-1}}}$$

şi conform teoremei precedente $Z \in T(\nu-1)$. ■

În figura 5 apar pe acelaşi grafic o distribuţie $N(0,1)$ şi o distribuţie $T(5)$.

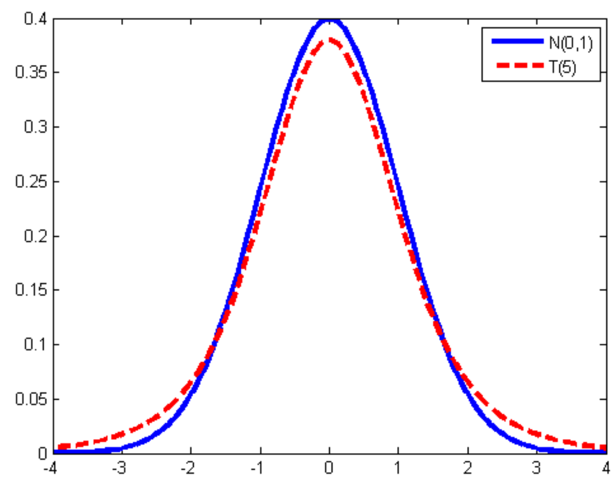


Figura 5: Comparaţie între $N(0,1)$ şi $T(5)$

2.6 Repartiţia F

Repartiţia F

Spunem că o variabilă aleatoare are **repartiţia F** sau **repartiţia Snedecor-Fisher** cu $\nu_1 > 0$ şi $\nu_2 > 0$ grade de libertate dacă densitatea sa de probabilitate este

$$f(x;\nu_1,\nu_2) = \begin{cases} \left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} \dfrac{x^{\frac{\nu_1}{2}-1}}{B\left(\frac{\nu_1}{2},\frac{\nu_2}{2}\right)\left(1+\frac{\nu_1}{\nu_2}x\right)^{\frac{\nu_1+\nu_2}{2}}} & \text{pentru} \quad x > 0 \\ 0 & \text{pentru} \quad x \le 0. \end{cases} \tag{19}$$

Mulţimea variabilelor cu această repartiţie se va nota cu $F(\nu_1, \nu_2)$. Momentele de ordinul r ale unei variabile aleatoare $X \in F(\nu_1, \nu_2)$ sunt

$$M_r(X) = \left(\frac{\nu_1}{\nu_2}\right)^r \frac{\Gamma\left(\frac{\nu_1}{2} + r\right)\Gamma\left(\frac{\nu_2}{2} - r\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)}, \tag{20}$$

cu condiţia ca $0 < 2r < \nu_2$. Pentru r întreg se obţine:

$$M_r(X) = \left(\frac{\nu_1}{\nu_2}\right)^r \frac{\nu_1(\nu_1+2)\dots(\nu_1+2r-2)}{(\nu_2-2)(\nu_2-4)\dots(\nu_2-2r)}.$$

În particular, au loc formulele

$$M(X) = \frac{\nu_2}{\nu_2-2}; \;\; M_2(X) = \frac{\nu_2^2}{\nu_1} \cdot \frac{\nu_1+2}{(\nu_2-2)(\nu_2-4)},$$

valabile pentru $\nu_2 > 2$ şi respectiv $\nu_2 > 4$. Dacă $\nu_2 > 4$, atunci

$$D^2(X) = \frac{2\nu_2^2(\nu_1+\nu_2-2)}{\nu_2-2}.$$

În figura 6 apare graficul unei densităţi de probabilitate de tip F.

Cele două teoreme care urmează precizează legătura între distribuţia F şi alte repartiţii.

Teorema 18. *Dacă $X_1 \in X^2(\nu_1, \sigma^2)$ şi $X_2 \in X^2(\nu_2, \sigma^2)$, atunci variabila aleatoare*

$$Y = \frac{\frac{1}{\nu_1}X_1}{\frac{1}{\nu_2}X_2}$$

aparţine clasei $F(\nu_1, \nu_2)$.

Demonstraţie. Deoarece X_1 şi X_2 sunt independente, densitatea de probabilitate a vectorului aleator (X_1, X_2) este

$$f(x_1, x_2) = f_{\chi^2}(x_1; \nu_1, \sigma^2) f_{\chi^2}(x_2; \nu_2, \sigma^2).$$

Densitatea de probabilitate a lui (Y, X_2) se obţine observând că

$$x_1 = \frac{\nu_1}{\nu_2} y x_2; \;\; x_2 = x_2$$

şi deci

$$\frac{D(x_1, x_2)}{D(y, x_2)} = \begin{vmatrix} \frac{\nu_1}{\nu_2}x_2 & \frac{\nu_1}{\nu_2}y \\ 0 & 1 \end{vmatrix} = \frac{\nu_1}{\nu_2}x_2,$$

de unde rezultă

$$f(y, x_2) = f(x_1, x_2)\frac{\nu_1}{\nu_2}x_2 = f_{\chi^2}(\frac{\nu_1}{\nu_2}yx_2; \nu_1, \sigma^2) f_{\chi^2}(x_2; \nu_2, \sigma^2)\frac{\nu_1}{\nu_2}x_2.$$

De aici deducem că densitatea de probabilitate a lui Y este

$$f(y) = \int_{-\infty}^{\infty} f(y, x_2) dx_2 = \frac{\nu_1}{\nu_2} \int_0^{\infty} f_{\chi^2}(\frac{\nu_1}{\nu_2} y x_2; \nu_1, \sigma^2) f_{\chi^2}(x_2; \nu_2, \sigma^2) \frac{\nu_1}{\nu_2} x_2 dx_2.$$

După efectuarea calculelor se obţine $f(y) = f(y; \nu_1, \nu_2)$. ■

Din ultima teoremă şi din teorema 12 rezultă:

Theorem 19. *Dacă $X_1, \ldots, X_{\nu_1}$ şi $Y_1, \ldots, Y_{\nu_2}$ sunt variabile aleatoare independente din clasa $N(0, \sigma^2)$, atunci variabila aleatoare*

$$Z = \frac{\nu_2}{\nu_1} \cdot \frac{X_1^2 + \cdots + X_{\nu_1}^2}{Y_1^2 + \cdots Y_{\nu_2}^2}$$

este de clasă $F(\nu_1, \nu_2)$.

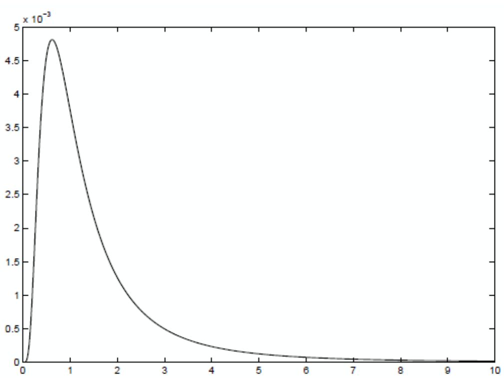


Figura 6: Graficul densităţii de probabilitate $F(15, 5)$