VALORI APROXIMATE

$x_1, \dots, x_n$ - date statistice

$\overline{X_n} = \frac{1}{n}(x_1 + \dots + x_n)$ = val. medie de selecție, nedeplasat & consistent pt medie

$S_n^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \overline{X_n})^2$ = val. varianței de selecție, nedeplasat și consistent pt var

S_n - abatere standard aproximată $= \sqrt{S_n^2}$, consistent pt deviație std.

$m_n = \frac{1}{n}\sum_{k=1}^{n}(x_k - \overline{X_n})^2$ = val mom. central de selecție, consistent pt. varianță

funcția de repartiție empirică

$F_n(x) = \frac{\# \{i \in \{1, \dots, n\} : x_i \le x\}}{n}$

nedeplasat și consistent pt $F(x)$

nedeplasat: $E(g(x_1, \dots, x_n)) = \theta$

consistent: $g(x_1 \dots x_n) \xrightarrow{a.s} \theta$

INTERVALE DE ÎNCREDERE

* pt. medie când $V(x) = \sigma^2$ cunoscut

$\left(\overline{X_n} - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}, \overline{X_n} + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}\right)$ bilateral

unilateral

$\left(-\infty, \overline{X_n} - \frac{\sigma}{\sqrt{n}} \cdot z_\alpha\right)$ sau $\left(\overline{X_n} - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\alpha}, \infty\right)$

unde $\alpha \to$ niv. semnificație, $1-\alpha \to$ niv încredere

* pt. medie când $V(x) = \sigma^2$ necunoscut

$\left(\overline{X_n} - \frac{S_n}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}}, \overline{X_n} + \frac{S_n}{\sqrt{n}} \cdot t_{1-\frac{\alpha}{2}}\right)$ bilateral

$\left(-\infty, \overline{X_n} - \frac{S_n}{\sqrt{n}} \cdot t_\alpha\right)$ sau $\left(\overline{X_n} - \frac{S_n}{\sqrt{n}} \cdot t_{1-\alpha}, \infty\right)$ unilateral

unde $S_n \to$ abatere standard

* pt. varianță

$\left(\frac{n-1}{c_{1-\frac{\alpha}{2}}} \cdot S_n^2, \frac{n-1}{c_{\frac{\alpha}{2}}} \cdot S_n^2\right)$ bilateral

unilateral

$\left(0, \frac{n-1}{c_\alpha} \cdot S_n^2\right)$ sau $\left(\frac{n-1}{c_{1-\alpha}} \cdot S_n^2, \infty\right)$

unde $S_n^2 \to$ varianța

* pt. abatere standard e ca și pt. varianță dar pun $\sqrt{\ }$ mare peste fiecare din cele 2 componente

TESTE STATISTICE

* test pt. media (m) când $\sigma^2 = V(x)$ cunoscut

$z = \frac{\overline{X_n} - m_0}{\frac{\sigma}{\sqrt{n}}} \to$ (verificați dacă $m_0 = \dots$)

concluzii	$H_0 : m = m_0$ $H_1 : m \neq m_0$	$H_0 : m \ge m_0$ $H_1 : m < m_0$	$H_0 : m \le m_0$ $H_1 : m > m_0$
se acc. H_0	$\lvert z \rvert < z_{1-\frac{\alpha}{2}}$	$z > z_\alpha$	$z < z_{1-\alpha}$

altfel se respinge H_0 și se ia H_1

* test pt. media (m) când $\sigma^2 = V(x)$ necunoscut

$t = \frac{\overline{X_n} - m_0}{\frac{S_n}{\sqrt{n}}}$ → concluzii ca sus, dar pun t în loc de z peste tot (și la cuantile)

* test pt. abatere standard $\sigma = \sqrt{V(x)}$

$c = \frac{n-1}{\sigma_0^2} \cdot S_n^2$ → varianța, S_n - abatere standard → dată statistică (verificați dacă $\sigma_0 = \dots$)

$H_0 : \sigma = \sigma_0$ $H_1 : \sigma \neq \sigma_0$	$H_0 : \sigma \ge \sigma_0$ $H_1 : \sigma < \sigma_0$	$H_0 : \sigma \le \sigma_0$ $H_1 : \sigma > \sigma_0$
$c_{\frac{\alpha}{2}} < c < c_{1-\frac{\alpha}{2}}$	$c > c_\alpha$	$c < c_{1-\alpha}$

se acceptă H_0, altfel se respinge H_0 și se acceptă H_1

z_α = norminv$(\alpha, 0, 1)$

t_α = tinv(α, m)

c_α = chi2inv(α, m)

unde m = nr eșantion -1

METODE

METODA MOMENTELOR

facem media (teoretic) = media (empiric) pt. estimare

$E(x) = \overline{X_n}$

ex. la exp $M[X] = \frac{1}{\lambda}$, $\overline{X_n} = \frac{x_1 + \dots + x_n}{n}$ $\Rightarrow \hat{\lambda} = \frac{n}{x_1 + \dots + x_n}$ → pt. că e aproximat

METODA VEROSIMILITĂȚII MAXIME

$L(X, \theta) = \prod_{i=1}^{n} f(x_i, \theta)$ / ln și ()'$_\theta$ = 0

$\to \sum_{i=1}^{n} \frac{\partial \ln f(x_i, \theta)}{\partial \theta} = 0$

BASICS 1

$f(x)$ - f. de densitate (pe caz discret e un tabel)

$F(x)$ - f. de repartiție $= \int_{-\infty}^{x} f(t)dt$ (pe caz discret e suma cumulativă)

! $F(x) = 0 \ \forall x < a$, $F(x) = 1 \ \forall x > b$

$f(x) = (F(x))'$

$\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$, e cont. la dreapta $\Leftrightarrow \lim_{x \searrow a} F(x) = F(a)$

suma tuturor e 1 $\Leftrightarrow \int_{\mathbb{R}} f(t)\,dt = 1$ (și pe caz discret!) Σ din tabel simplu !

$E(x)$ = val. medie $= \int_{\mathbb{R}} x \cdot f(x)dx$

→ pe caz discret $\sum_{k=0}^{n} k \cdot P(x = k)$ basically suma ponderată din tabel

$E(x^2) = \int_{\mathbb{R}} x^2 \cdot f(x)\,dx$

$V(x) = E(x^2) - E^2(x)$ varianță/dispersie $= \sigma^2$!

! σ - abatere/deviație standard

varianța = abatere$^2 \Leftrightarrow$ abatere = $\sqrt{\text{varianța}}$

$V(ax + b) = a^2 V(x) \ \forall a, b \in \mathbb{R}$

BASICS 2

proprietăți ale lui P

$P(A \setminus B) = P(A) - P(A \cap B)$

$A \subseteq B \Rightarrow P(A) \le P(B)$

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$!

$P(A \cap B) = P(A) \cdot P(B)$ dacă A, B independente

$P(A|B) = \frac{P(A \cap B)}{P(B)}$

pe caz continuu: $P(X = a) = 0 \ \forall a \in \mathbb{R}$

perm. cu repetiție: $\frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_k!}$

comb. cu repetiție: $C_{n+k-1}^{k} = \frac{(n+k-1)!}{k!(n-1)!}$, n - nr. obiecte distincte

n culori și bila NEreturnată: n_i - bile de cul. i, k_i - extr. de culoarea i, n - extrageri

$\frac{C_{n_1}^{k_1} \cdot C_{n_2}^{k_2} \cdot \dots \cdot C_{n_r}^{k_r}}{C_{n_1 + n_2 + \dots + n_r}^{n}}$

cu rst dacă am probabilități: $\frac{n!}{k_1! \cdot k_2! \cdot \dots \cdot k_r!} \cdot p_1^{k_1} \cdot \dots \cdot p_r^{k_r}$

LTNM $\frac{x_1 + \dots + x_n}{n} \xrightarrow{a.s} m$ (unde $m = E(X_i)$) → oricare din ele

BAYES $P(E) = \sum_{i=1}^{n} P(E|H_i) \cdot P(H_i)$ unde H_i - partiție

$P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)}$ → pt $\forall H$; $P(E)$ → de sus

PT. DISCRETE

* $X \sim$ Unid (n) discretă uniformă

$X \sim \begin{pmatrix} 1 & 2 & \dots & n \\ \frac{1}{n} & \frac{1}{n} & & \frac{1}{n} \end{pmatrix}$

* Bernoulli (p), $p \in (0, 1)$

$X \sim \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$ dacă $X \sim$ Bernoulli(p), $\sum_{i=1}^{n} x_i$ = Bino(n, p)

* Bino (n, p) → adică $X \sim \begin{pmatrix} k \\ C \dots \end{pmatrix}$

$P(X = k) = C_n^k p^k (1-p)^{n-k}$ ← CU RETURNARE 2 culori

nr. de succese din n extrageri

p → prob. succes $E(X) = n \cdot p$ ~ media la Bino

* hipergeometrică (n, n_1, n_2) → nr aruncări, p. succes

n → nr. extrageri, n_1 → nr succese, n_2 → nr eșecuri

$P(X = k) = \frac{C_{n_1}^{k} \cdot C_{n_2}^{n-k}}{C_{n_1+n_2}^{n}} \Rightarrow X \sim \begin{pmatrix} k \\ C \dots \end{pmatrix}$

k → extrageri bune din n_1

FĂRĂ RETURNARE 2 culori

PT. CONTINUE

* unif [a, b]

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

* exp(λ) media $= \frac{1}{\lambda}$

$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & \text{altfel} \end{cases}$

* normală (μ, σ^2) → μ media, σ^2 → varianță

$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ → aka $e^{\wedge}$

* geometrică $X \sim$ Geo(p)

$P(X = k) = p(1-p)^k$ → prob. succes

k → nr. insuccese până la succes

VALORI APROXIMATE

$x_1, \dots, x_m$ - date statistice

$\overline{X_m} = \frac{1}{m}(x_1 + \dots + x_m)$ = val. medie de selecție mediplasat & consistent pt medie

$S_m^2 = \frac{1}{m-1}\sum_{k=1}^{m}(x_k - \overline{x_m})^2$ = val. varianței de selecție nedeplasat și consistent pt var

S_m = abatere standard aproximată $= \sqrt{S_m^2}$ consistent pt deviație std.

$m_m = \frac{1}{m}\sum_{k=1}^{m}(x_k - \overline{x_m})^2$ = val mom. central de selecție, consistent pt. varianță

funcția de repartiție empirică

$$F_m(x) = \frac{\#\{i \in \{1, \dots, m\} : x_i \le x\}}{m}$$

nedeplasat și consistent pt $F(x)$

nedeplasat: $E(g(x_1, \dots, x_m)) = \theta$

consistent: $g(x_1 \dots x_m) \xrightarrow{a.s} \theta$

INTERVALE DE ÎNCREDERE

* pt. medie când $V(x) = \sigma^2$ cunoscut

$\left(\overline{X_m} - \frac{\sigma}{\sqrt{x}} \cdot z_{1-\frac{\alpha}{2}}, \overline{X_m} + \frac{\sigma}{\sqrt{m}} \cdot z_{1-\frac{\alpha}{2}}\right)$ bilateral

$\left(-\infty, \overline{X_m} - \frac{\sigma}{\sqrt{m}} \cdot z_\alpha\right)$ sau $\left(\overline{X_m} - \frac{\sigma}{\sqrt{m}} \cdot z_{1-\alpha}, \infty\right)$ unilateral

unde α -> niv. semnificație, $1-\alpha$ -> niv încredere

* pt. medie când $V(x) = \sigma^2$ necunoscut

$\left(\overline{X_m} - \frac{S_m}{\sqrt{m}} \cdot t_{1-\frac{\alpha}{2}}, \overline{X_m} + \frac{S_m}{\sqrt{m}} \cdot t_{1-\frac{\alpha}{2}}\right)$ bilateral

$\left(-\infty, \overline{X_m} - \frac{S_m}{\sqrt{m}} \cdot t_\alpha\right)$ sau $\left(\overline{X_m} - \frac{S_m}{\sqrt{m}} \cdot t_{1-\alpha}, \infty\right)$ unilateral

unde S_m -> abatere standard

* pt. varianță

$\left(\frac{m-1}{C_{1-\frac{\alpha}{2}}} \cdot S_m^2, \frac{m-1}{C_{\frac{\alpha}{2}}} \cdot S_m^2\right)$ bilateral

$\left(0, \frac{m-1}{C_\alpha} \cdot S_m^2\right)$ sau $\left(\frac{m-1}{C_{1-\alpha}} \cdot S_m^2, \infty\right)$ unilateral

unde S_m^2 -> varianța

* pt. abatere standard e ca și pt. varianță dar pun $\sqrt{}$ mare peste fiecare din cele 2 componente

TESTE STATISTICE

* test pt. media (m) când $\sigma^2 = V(x)$ cunoscut

$z = \frac{\overline{X_m} - m_0}{\frac{\sigma}{\sqrt{m}}}$ -> (verificați dacă $m_0 = \dots$)

concluzii	$H_0 : m = m_0$, $H_1 : m \ne m_0$	$H_0 : m \ge m_0$, $H_1 : m < m_0$	$H_0 : m \le m_0$, $H_1 : m > m_0$
se acc. H_0	$\lvert z \rvert < z_{1-\frac{\alpha}{2}}$	$z > z_\alpha$	$z < z_{1-\alpha}$

altfel se respinge H_0 și se ia H_1

* test pt. media (m) când $\sigma^2 = V(x)$ necunoscut

$t = \frac{\overline{X_m} - m_0}{\frac{S_m}{\sqrt{m}}}$ -> concluzii ca sus, dar pun t în loc de z peste tot (și la cuantile)

* test pt. abatere standard $\sigma = \sqrt{V(x)}$

$c = \frac{m-1}{\sigma_0^2} \cdot S_m^2$ -> varianța S_m - abatere standard

-> dată statistică (verificați dacă $\sigma_0 = \dots$)

$H_0 : \sigma = \sigma_0$, $H_1 : \sigma \ne \sigma_0$	$H_0 : \sigma \ge \sigma_0$, $H_1 : \sigma < \sigma_0$	$H_0 : \sigma \le \sigma_0$, $H_1 : \sigma > \sigma_0$
$C_{\frac{\alpha}{2}} < c < C_{1-\frac{\alpha}{2}}$	$c > C_\alpha$	$c < C_{1-\alpha}$

se acceptă H_0, altfel se respinge H_0 și se acceptă H_1

z_α = norminv(α, 0, 1)

t_α = tinv(α, m) > unde m = nr eșantion - 1

c_α = chi2inv(α, m)

PT. CONTINUE

METODE

METODA MOMENTELOR

facem media (teoretic) = media (empiric) pt. estimare

$E(X) = \overline{X_n}$

pt. că e aproximat

ex. la exp $M[X] = \frac{1}{\lambda}$

$\left.\begin{array}{l} M[X] = \frac{1}{\lambda} \\ \overline{X_n} = \frac{X_1 + \dots + X_n}{n} \end{array}\right\} \Rightarrow \hat{\lambda} = \frac{n}{X_1 + \dots + X_n}$

METODA VEROSIMILITĂȚII MAXIME

$L(X, \theta) = \prod_{i=1}^{n} f(x_i, \theta)$ / ln și $()'_\theta = 0$

$\Rightarrow \sum_{i=1}^{n} \frac{\partial \ln f(x_i, \theta)}{\partial \theta} = 0$

f(x) - f. de densitate (pe caz discret e un tabel)

F(x) - f. de repartiție = $\int_{-\infty}^{x} f(t)dt$ (pe caz discret e suma cumulativă)

! $F(x) = 0 \ \forall x<a$

$F(x) = 1 \ \forall x>b$

$f(x) = (F(x))'$

$\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$, e cont. la dreapta ⇔ $\lim_{x\searrow a} F(x) = F(a)$

suma tuturor e 1 ⇔ $\int_{\mathbb{R}} f(t)dt = 1$ (și pe caz discret!) ! Σ dintre toate simplu

E(x) = val. medie = $\int_{\mathbb{R}} x\cdot f(x)dx$

→ pe caz discret $\sum_{k=0}^{\infty} k\cdot P(X=k)$ basically suma ponderată din tabel

$E(x^2) = \int_{\mathbb{R}} x^2\cdot f(x)dx$

$V(x) = E(x^2) - E^2(x)$ varianță/dispersie $= \sigma^2$!

! σ - abatere/deviație standard

varianța = abatere^2 ⇔ abatere = $\sqrt{\text{varianță}}$

$V(ax+b) = a^2 V(x) \ \forall a,b\in\mathbb{R}$

BASICS 1

BASICS 2

proprietăți ale lui P

$P(A\setminus B) = P(A) - P(A\cap B)$

$A\subseteq B \Rightarrow P(A)\le P(B)$

$P(A\cup B) = P(A)+P(B)-P(A\cap B)$!

$P(A\cap B) = P(A)\cdot P(B)$ dacă A, B independente

$P(A|B) = \frac{P(A\cap B)}{P(B)}$

pe caz continuu: $P(X=a) = 0 \ \forall a\in\mathbb{R}$

perm. cu repetiție: $\frac{n!}{n_1!\cdot n_2!\cdot\ldots\cdot n_k!}$

comb. cu repetiție: $C_{n+k-1}^{k} = \frac{(n+k-1)!}{k!(n-1)!}$

n - nr. obiecte distincte

n culori și bila NE returnată: m_i - bile de cul. i, k_i - extr. de culoarea i, m - extrageri

$\frac{C_{m_1}^{k_1}\cdot C_{m_2}^{k_2}\cdot\ldots\cdot C_{m_n}^{k_n}}{C_{m_1+m_2+\ldots+m_n}^{m}}$

cu ret dacă am probabilități: $\frac{m!}{k_1!\cdot k_2!\cdot\ldots\cdot k_n!}\cdot p_1^{k_1}\cdot\ldots\cdot p_n^{k_n}$

LTNM $\frac{x_1+\ldots+x_m}{m} \xrightarrow{a.s.} m$ (unde $m = E(X_i)$) oricare din ele

BAYES $P(E) = \sum_{i=1}^{m} P(E|H_i)\cdot P(H_i)$ unde H_i - partiție

$P(H|E) = \frac{P(E|H)\cdot P(H)}{P(E)}$ → pt ∀H; P(E) → de sus

$P(Z) \to$ de sus

PT. DISCRETE

* $X \sim$ Unid (n) discretă uniformă

$$X \sim \begin{pmatrix} 1 & 2 & \dots & n \\ \frac{1}{n} & \frac{1}{n} & & \frac{1}{n} \end{pmatrix}$$

* Bernoulli (p), $p \in (0,1)$

$$X \sim \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$$ dacă $X \sim$ Bernoulli(p) $\sum_{i=1}^{n} X_i =$ Bino(n, p)

* Bino (n,p) $\to$ adică $X \sim \begin{pmatrix} k \\ C \dots \end{pmatrix}$

$P(X=k) = C_n^k p^k (1-p)^{n-k}$

CU RETURNARE
2 culori

nr. de succese din n extrageri

$p \to$ prob. succes $E(X) = \underbrace{n \cdot p}$ ~ media la Bino
nr aruncări
p. succes

* hipergeometrică (n, n_1, n_2)

$n \to$ nr. extrageri, $n_1 \to$ nr succese, $n_2 \to$ nr eșecuri

$$P(X=k) = \frac{C_{n_1}^{k} \cdot C_{n_2}^{n-k}}{C_{n_1+n_2}^{n}} \Rightarrow X \sim \begin{pmatrix} k \\ C \dots \end{pmatrix}$$

$k \downarrow$ extrageri bune din n_1

FĂRĂ RETURNARE
2 culori

PT. CONTINUE

* unif [a, b]

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

* exp(λ) media = $\frac{1}{\lambda}$

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & \text{altfel} \end{cases}$$

* normală (μ, σ^2)

$\mu \downarrow$ media; $\sigma^2 \to$ varianță

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

exp $\downarrow$ aka e^

* geometrică $X \sim$ Geo(p)

$P(X=k) = p(1-p)^k$, $p \to$ prob. succes

$k \downarrow$ nr. insuccese până la succes