Pravd.prostor $(\Omega, \mathcal{F}, P), \Omega \neq \emptyset, \mathcal{F} \subseteq P(\Omega), P: \mathcal{F} \to [0,1]$. $P: \mathcal{F} \to [0,1]; P(\Omega) = 1, P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ pro $A_1, A_2, \ldots$ po dvou disj.
$(\Omega, \mathcal{F}, P):$ $(1) P(A) + P(A^C) = 1$ $(2) A \subseteq B \implies P(A) \leq P(B)$ $(3) P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (4) $\mathrm{P}(\mathrm{A}_1 \cup A_2 \cup \ldots) \leq \sum P(A_i)$
Dk:$(1)\Omega = A \cup A^C \implies P(\Omega) = P(A) + P(B)$ $(2) B = A \cup (A - B) : P(B) = P(A) + P(B - A) \implies \ldots$ $(3) P(A \cup B) = P(A - B \cup A \cap B \cup B - A) + P(A \cap B) - P(A \cap B) = \ldots$
(4)trik zdisjunk.$B_1 = A, B_2 = A_2 - A_1, B_i = A_i - \bigcup_{j=1}^{i-1} : P(\bigcup A_i) = P(\bigcup B_i) = \sum P(B_i) = \sum P(A_i) \square$
O průnicích $P(A_1 \cap \ldots \cap A_n) = P(A_1) P(A_2 | A_1) \ldots P(A_n | A_1 \cap \ldots \cap A_{n-1})$ $Dk: P(A_1) \frac{P(A_1 \cap A_2)}{P(A_1)} \ldots \frac{P(A_1 \cap \ldots \cap A_n)}{P(A_1 \cap \ldots \cap A_{n-1})} = P(A_1 \cap \ldots \cap A_n) \square$
Úplná pravd. $P(A) = \sum P(B_i) P(A|B_i). A, B_i \in \mathcal{F}, B_i$ rozklad A. Pokud $P(B_i) = 0, = 0. Dk: P(A) = \sum P(A \cap B_i) = \sum P(A) P(A|B_i)$
Bayes $P(B|A) = \frac{P(A|B)P(B)}{P(A)}$ $Dk: triv.$ (jmenovatel podle ^) Df: Nezávislost $P(A \cap B) = P(A)P(B)$ $P(\bigcap_{j \in I' \subseteq I} A_j) = \prod_{j \in I'} P(A_j)$
Df: Náhodná dis. vel. X: $X: \Omega \to \mathbb{R}, Im(X)$ je spočetná $. \forall x \in \mathbb{R}: \{\omega \in \Omega : X(\omega) = x\} \in \mathcal{F}$ Df: Pravd.fce: $p_X(x_i) = P(X = x_i) : \mathbb{R} \to [0,1]$
Df: $\mathbb{E}(X) = \sum_{x \in Im(X)} x p_x(x)$ pokud má souč. smysl. Pozn.: $\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega) P(\omega)$ $Dk := \sum_{x \in Im(X)} \sum_{\omega : X(\omega) = x} X(\omega) P(\{\omega\}) = \ldots$
LOTUS: $\mathbb{E}(g(x)) = \sum_{x \in Im(X)} g(x) P(X = x)$ $Dk:$ $\mathbb{E}(Y) = \sum_{y \in Im(y)} y P(\underbrace{Y = y}) : \{\omega : g(X(\omega)) = y\} = \bigcup_{x \in Im(X) g(x) = y} \{\omega : X(\omega) = x\}$
(disj. sjed.) : $\ldots = \sum_{y \in Im(y)} y \sum_{x \in Im(X) g(x) = y} P(X = x).(Nyní(y = g(x)))\square$ Vlastnosti $\mathbb{E}: (1) P(X \geq 0) = 1 \& \mathbb{E}(X) = 0 \implies P(X = 0) = 1$
$(2)\mathbb{E}(X) \geq c \implies P(X \geq c) > 0$ $(3)\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$ $(4)\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ Dk: $(3) g(x) = ax + b(+LOTUS + linearita)$
(4)pouze pro disk.pr.: linearita: $= \sum_{\omega \in \Omega} (X(\omega) + Y(\omega)) P(\omega)$ $(1)\mathbb{E}(X) = \sum_{x \in Im(X)} x(P = x) \geq 0, \mathbb{E}(X) = 0 : \implies xP(X = x) = 0 \implies x = 0 \vee P(X = x) = 0.$
$(2) Sporem: c \leq \mathbb{E}(X) = \sum x P(X = x). P(X \geq c) = 0 \implies \forall x \geq c : P(X = x) = 0 \implies \sum x P(X = x) < \sum c P(X = x) = c \sum P(X = x) = c \implies c < c \implies$ spor
Df: Podm. $\mathbb{E}: \mathbb{E}(X|B) = \sum_{x \in Im(X)} x P(X = x|B)$ O celkové $\mathbb{E}: B_1, \ldots,$ rozklad Ω, X $d.n.v.: \mathbb{E} = \sum P(B_i) \mathbb{E}(X|B_i) Dk: \mathbb{E}(X) = \sum x P(X = x) = \sum_x x \sum_i P(B_i) P(X = x|B_i) = \sum_i \sum_x x P(B_i) P(X = x|B_i) = \sum_i P(B_i) \mathbb{E}(X = x|B_i) \square$
Df: $var(X) = \mathbb{E}[(X - \mathbb{E})^2]. \sigma_X = \sqrt{var(X)}$ směr. odch. $var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$ $Dk: \mu = \mathbb{E}(X); \mathbb{E}[(X - \mu)^2] = \mathbb{E}[X^2 - 2X\mu + \mu^2] = \ldots linearita. \square$
$\mathbb{E}(X) = \sum_{n=0}^{\infty} P(X > n)$ když $Im(X) \subseteq \mathbb{N}_0$ Df. Sdružená pstní fce vel X,Y: $P_{X,Y}(x, y) = P(X = x \& Y = y) : \mathbb{R}^2 \to [0,1],$ pro $\in \mathcal{F}$
$var(aX = b) = a^2 var(X)$ Df: Nez. náh. vel.: $\forall x, y \{X = x\}, Y = y$ jsou nezav. Df. Multinomiální rozd. $P_{X_1, \ldots, X_k}(k_1, \ldots, k_n) = \binom{n}{k_1, \ldots, k_n} p_1^{k_1} \ldots p_n^{k_n}$
Konvoluce $P(X + Y = z) = \sum_{x \in Im(X)} P(X = x \& Y = y - x) = \sum_{x \in Im(X)} p_X(x) p_Y(z - x)$. $\mathbb{E}(g(X, Y)) = \sum_{x \in Im(X), y \in Im(Y)} g(x, y) P_{X,Y}(x, y)$
(<máli součet smysl.). Dk: $P(Z = z) = \sum_{x \in Im(X)} P(X = x) P(Z = z|X = x)$.Dle df Z: $P(Z = z|X = x) = P(Y = z - x|X = x). P(Y = z - x|X = x) =$
(nez.)=$P(Y = z - x)\square$ Pro n.n.v$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ $Dk: \mathbb{E}(XY) = \sum_{x, y \in Im} (xy) P(X = x \& Y = y) = \sum xy P(X = x) P(Y = y) = \mathbb{E}(X)\mathbb{E}(Y) \square$
$var(X + Y) = var(X) + var(Y) + 2cov(X, Y)$ Df: kovariance $cov(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$. $cov(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))$
$cov(X, Y) \leq \sqrt{var(X) var(Y)} \implies -\sigma_X \sigma_Y \leq cov(X, Y) \leq \sigma_X \sigma_Y$ Df: Korelace $cor(X, Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$ Věta $var(\sum^n X_i) = \sum_{i=1, j=1}^{n,n} cov(X_i, X_j) =$
$= \sum_{i=1}^n var(X_i) + \sum_{i \neq j}^n cov(X_i, X_j) = ($ pro po 2 nezav. $) = \sum^n var(X_i)$ Dk: $\mathbb{E}(\sum X_i)^2 = \mathbb{E}(\sum_{i,j} X_i X_j) = \sum_{i,j} \mathbb{E}(X_i X_j)$
$(\mathbb{E}(\sum X_i))^2 = \sum_{i,j} \mathbb{E}(X_i)\mathbb{E}(X_j) \implies var(\sum X_i) = \mathbb{E}(\sum X_i)^2 - (\mathbb{E}(\sum X_i))^2 = \sum_{i,j} \mathbb{E}(X_i X_j) - \mathbb{E}(X_i)\mathbb{E}(X_j) = \sum_{i,j} cov(X_i, X_j)(1. =)$
$cov(X, X) = var(X)($ ze vzorce $)(2. =)$ $NEZ \implies cov(X_i, X_j) = 0(3. =)\square$ $P_{X|A}(x) = P(X = x|A); P_{X|Y}(x|y) = P(X = x|Y = y)$
Df. Obecná n.v. $X: \Omega \to \mathbb{R}; \forall x \in \mathbb{R} : \omega \in \Omega : X(\omega) \leq x \in \mathcal{F}$ Df.: Distribuční funkce $F_X(x) = P(X \leq x)$; poz: neklesající, zprava spojitá, lim 0,1
Df. X je spojitá n.v.: $\exists f_X : \mathbb{R} \to [0, \infty] : F_X(x) = \int_{-\infty}^x f_X(t) dt$ má hustotu. Věta: $(1) P(X = x) = 0 \forall x \in \mathbb{R}$ $(2) P(a \leq X \leq b) = \int_a^b f_X(t) dt$
$(3)\forall$ rozumn. množ. $A: P(X \in A) = \int_A f_X(t) dt$ Dk: $(1), (3)$ bez dk $(2) P(a \leq X \leq b) = P(a < X \leq b) = P(X \leq b) - P(X \leq a) = \int_{-\infty}^b f_X - \int_{-\infty}^a f_X = \int_a^b f_X \square$
Df: $\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x) dx$ pokud má smysl. LOTUS2: $\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x) dx$ Bez dk.
Df.: Kvantilová fce X $n.v., Q_X : [0,1] \to \mathbb{R} : Q_X(p) = min\{x \in \mathbb{R} : p \leq F_X(X)\}$ Pokud X je spoj. rostoucí: $Q_X = F_X^{-1}$ Df.: medián $Q_X(\frac{1}{2})$, kvartil,
Veta: X n.v. s dist $F_x = F$, F je spoj. rost.: $F(X) \sim U(0,1)$. Dk: $Y := F(X). F_y(y) = P(Y \leq y) = P(F(X) \leq y) = (pro\ y = F(x)) = P(X \leq x)$
$pro\ y \in (0,1) \ldots F(x) = y$, pro $y > 1 \ldots 1$, pro $y < 0 \ldots 0 \square$. Veta: F je fce “typu dist. fce”, Q je odp. kvant. fce, $U \sim U(0,1), X = Q(U)$, pak $F_X = F$
Dk: $P(Q(U) \leq x) = P(U < F(x)) = F(x) \square$
Df: Náhod. Vektor (X,Y) na Pravd.prost. je dvojice n.v.: $\forall x, y \in \mathbb{R}$ je def. $F_{X,Y}(x, y) = P(X \leq x \& Y \leq y)(\in \mathcal{F})$ sdružená dist. fce. Df: pokud
$\forall x, y F_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(s, t) ds dt$ tak X,Y jsou sdruženě spojité a $f_{X,Y}$ je sdružená hustota. Nezávislost $\iff$ $\{X \leq x\}, \{Y \leq y\}$ nezav.
$F_{X,Y}(x, y) = F_X(x) F_Y(y) \iff f_{X,Y}(x, y) = f_X(x) f_Y(y)$. Konvoluce: X,Y n.n.v spojité => Z=X+Y taky spojitá, $f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) dx$
Normální rozd. $\varphi(x) = 1/\sqrt{2\pi} exp(-x^2/2)$. $\varphi_X(x) = \frac{1}{\sigma} \varphi(\frac{x - \mu}{\sigma})$. Vícerozměrné normální rozd. $\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}. f(t_1, \ldots, t_n) = \varphi(t_1) \ldots \varphi(t_n)$
$f(t_1, \ldots, t_n) = (2\pi)^{n/2} exp(-\frac{(t_1^2 + \ldots + t_n^2)^2}{2})$. Df. Podm. hustota: $f_{X|Y}(x, y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$ $pokud$ $f_Y(y) > 0$. Marginální hust: $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx$
Markovova nerovnost: X je n.v $X \geq 0, a > 0 : P(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$ Dk: $\mathbb{E}(X) = P(X \geq A) \mathbb{E}(X|X \geq a) + P(X < A) \mathbb{E}(X|X < a)$ (druhý větší než 0)
Čebyšev: X je nv. $\mathbb{E}(X) = \mu < \infty, var(X) = \sigma^2 < \infty : P(|X - \mu| \geq a\sigma) \leq \frac{1}{a^2}$ Dk: $Y = (X - \mu)^2 \geq 0$. $P(Y > a^2 \sigma^2) \leq \frac{\mathbb{E}}{a^2 \sigma^2} = \frac{var(X)}{a^2 var(X)} \square$ (Markov.)
Chernoff $X = \sum X_i . X_i$ $n.n.v. = \pm 1(50\%), t > 0 : P(X \geq t) = P(X \leq -t) \leq exp(-\frac{t^2}{2\sigma^2}). Kde$ $\sigma = \sqrt{n}$ Bez dk.
Silný. zák. velk. č.: $X_1, X_2, \ldots$ Jsou stejně rozd. n.n.v s $var = \sigma^2, \mathbb{E} = \mu$. $S_n = \frac{X_1 + \ldots + X_n}{n}$, (vyb. prům.)pak $\lim_{n \to \infty} S_n = \mu$ skoro jistě (P = 1).
Pozn: Důsl: je dobré opakovat měření. Aplik: Monte Carlo integrace. (konec dk. slabý) $\implies P(|S_n - \mu| > \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$. $\lim_{n \to infty} \frac{\sigma^2}{n\varepsilon^2} = 0 \square$
Slabý z.v.č: (stejné př. jako silný). $\forall \varepsilon > 0 : \lim_{n \to \infty} P(|S_n - \mu| > \varepsilon) = 0$ Dk: (Čebyšev pro $X = S_n, a = \sqrt{n}\varepsilon/\sigma$) $\mathbb{E}(S_n) = \frac{\mathbb{E}(X_1) + \ldots + \mathbb{E}(X_n)}{n} = \mu$
$var(S_n) = var(\frac{(X_1) + \ldots + (X_n)}{n}) = \frac{var(X_1) + \ldots + var(X_n)}{n^2} = \frac{\sigma^2}{n}$. Čebyšev: $P(|S_n - \mu| > \varepsilon) \leq \frac{1}{a^2} \implies P(|S_n - \mu| > a \frac{\sigma}{\sqrt{n}}) \leq \frac{1}{a^2}$ (konec ^)
Gambler's fallacy: if a particular event occurs more frequently than normal during the past, it is less likely to happen in the future(není limita)
CLV: $X_1, X_2, \ldots$ Jsou stejně rozd. n.n.v s $var = \sigma^2, \mathbb{E} = \mu$. $Y_n = \frac{(X_1 + \ldots + X_n) - n\mu}{\sigma\sqrt{n}}. Pak$ $Y_n \xrightarrow{d} N(0,1). Neboli$ $pokud F_n = F_{Y_n} : \forall x \in \mathbb{R} : \lim_{x \to \infty} F_n(x) = \Phi(x)$
Význam: umožnuje aproximovat soucet mnoha náhodných velicin pomocí jednoho, dobre známého rozdelení. Takové soucty se casto vyskytují – napr Bin(n; p) je soucet n velicin Ber(p), ve statistice takové soucty casto potkáme a mnoho fyzikálních jevu je (približne) popsáno jako soucet nez náh vel. To vysvetluje, proc binomická císla, stejne jako mnoho v praxi se vyskytujících velicin, mají približne normální rozdelení
-Testovani hypotez: Testujeme platnost tvrzení, např. mince spravedlivá. Označíme H0 očekávané, tzv. nulovou hypotézu. Alternativní hypotéza H1. Vytvoříme statistický model situace (pokud házíme n-krát mincí, typicky jednotlivé hody jsou nezávislé, rozdělení $Bin(n; \vartheta)$. H0 říká $\vartheta = \frac{1}{2}$. Před naměřením rozhodneme, jaký bude kritický obor, množina W možných měření, při kterých H0 zamítneme.Chceme zajistit, aby byla předem daná prvd. chyby 1. druhu: chybného zamítnutí. Omezení pravděpodobnosti chyby 1. d. značíme α Zároven se snažíme min. pravdepodobnost chyby 2. druhu, chybného přijetí (značíme β). Pravděpodobnost, že chybnou hypotézu zamítneme nazýváme síla testu.
-Test dobré shody: Opakujeme n-krát pokus, který muže dopadnout jedním z k výsledku,opakování jsou nezav. Testujeme nulovou hyp., že i-tý výsledek nastává s pravd. p_i. Pro $i = 1, \ldots, k$ oznacme X_i pocet pokusu s i-tým výsledkem a $E_i = p_i n$ strední hodnotu X_i. Dále spocteme Pearsonovu statistiku $T = \sum_{i=1}^k \frac{(E_i - X_i)^2}{E_i}$ kritická hod. γ je kvantilová funkce χ^2 rozdelení s k-1 stupni volnosti v bode 1 - α , tj. $\gamma = Q(1 - \alpha)$, kde Q je kvant. funkce pro χ^2_{k-1}. H0 zamítneme, pokud $T > \gamma$. Pri takové volbe je pravdepodobnost chyby prvního druhu približne rovna α.
-jak generovat náhodnou veličinu s danou dist. fci: Inverse transformation: Necht F je funkce „typu distribucní funkce“: neklesající zprava spojitá funkce s $\lim_{x \to -\infty} F(x) = 0$ a $\lim_{x \to \infty} F(x) = 1$. Nechť Q je odpov. kvantilova. fce. Necht $U \sim U(0,1), X = Q(U)$. Pak X má dist. fci F.
Zamítací metoda: Chceme vygenerovat n.v. s hustotou f. Umíme vygenerovat n.v. s hustotou g, „podobná“, $\frac{f(y)}{g(y)} \leq c$ pro nějakou konst. c.
Postup: 1. vygenerujeme Y s hust. g, a $U \sim U(0,1)$. 2. Pokud $U \leq \frac{f(Y)}{cg(Y)}, tak$ $X := Y$. 3. Jinak hodnotu Y,U zamítneme a opakujeme od bodu 1.
-Empiricka dist.fce: Pro náh. výběr: $X_1, \ldots, X_n$: e.d.f: $\hat{F}_n(x) = \frac{\sum_{i=1}^n I(X_i \leq x)}{n}. I(X_i \leq x) = 1$ $pokud$ $X_i \leq x, 0$ $jinak$.

-Metoda max. věrohodnosti: $X_1, ..., X_n$ vyběr, $x_1, ..., x_2$ realizace, diskrétní $p_X(x, \vartheta) = \prod p_{X_i}(x_i, \vartheta)$, spojité $p_X(x, \vartheta) = \prod p_{X_i}(x_i, \vartheta)$ = Likelihood $L(x, \vartheta)$. Zvolíme takové ϑ, aby L bylo maximální, případně stačí max. $l = \log L$. Př: Počítání leváků: k z n lidi jsou leváci: $X_1, ..., X_n \sim Bern(\vartheta)$ $L(x, \vartheta) = \vartheta^k (1-\vartheta)^{n-k}, l = k \log \vartheta + (n-k) \log(1-\vartheta)$. Najdeme maximum derivací.Př2: Kostka, 6=p=x, hodil 2,3,6. $p^x(\frac{1-p}{5})^y$minimalizujeme,$\frac{1}{3}$
-Bodový odhad $\hat{\Theta}_n$ pro nah. vyb. $X, \ldots, X_n \sim F_\vartheta$ a libovolnou fci g je: Nevychýlený, pokud $\mathbb{E}(\hat{\Theta}_n) = g(\vartheta)$. Asympt. Nevychýchelný, $\lim_{n\to\infty} \mathbb{E}(\hat{\Theta}_n) = g(\vartheta)$ Konzistentní: $(\hat{\Theta}_n) \xrightarrow{p} g(\vartheta)$. Vychýlení $bias(\hat{\Theta}_n) = \mathbb{E}(\hat{\Theta}_n) - g(\vartheta)$. Stred. kvad. odchyl. $MSE(\hat{\Theta}_n) = \mathbb{E}(((\hat{\Theta}_n) - g(\vartheta)^2))$.
$\overline{X}_n = \frac{1}{n}\sum^n X_i$ vyběrový prů. $\overline{S}_n^2 = \frac{1}{n}\sum^n (X_n - \overline{X}_n)^2$ výb.rozptyl. - konz. asympt. nestran. $\hat{S}_n^2 = \frac{1}{n-1}\sum^n (X_n - \overline{X}_n)^2$ rozptyl, konz. nestranný.
-Metoda momentů: r-tý moment pro $X \sim F_\vartheta$ je $m_r(\vartheta) = \mathbb{E}(X_i^r)$. rtý výběrový moment pro náh. výběr: $\widehat{m_r(\vartheta)} = \frac{1}{n}\sum^n X_i^t$. $\widehat{m_r(\vartheta)}$ je nestranny konz. odhad $m_r(\vartheta)$ Tedy volíme \vartheta řešící soustavu rovnic $\widehat{m_r(\vartheta)} = m_r(\vartheta)$ pro 1...k – typicky $k = \#\mathbb{R}$ císel tvořici parametr ϑ Př: náhodné výšky lidí $\sim N(\mu, \sigma^2), \vartheta = (\mu, \sigma^2)$. $m_1(\mu) = \vartheta, m_2(\mu) = \mathbb{E}(X^2) = var(X) + \mathbb{E}(X)^2 = \sigma^2 + \mu^2$. Mame polozit $\mu = m_1(\vartheta) = \widehat{m_1(\vartheta)} = \overline{X}_n$.
$\sigma^2 + \mu^2 = m_1(\vartheta) = \widehat{m_1(\vartheta)} = \overline{X}_n = \frac{X_1^2 + ... + X_n^2}{n}$. Řešení: $\hat{\mu} = \overline{X}_n, \hat{\sigma} = \sqrt{\frac{X_1^2 + ... + X_n^2}{n} - \overline{X}_n^2}$. Př.2: Kostka, pravd. padne 6=p, jinak rovnomerne. Padlo 2,3,6. Vypocteme $\mathbb{E} = 3 + 3p$ pak $\hat{p} = \mu/3 - 1$
-Studentův t-test: Student. t-rozdělení s n-1 stupni volnosti $\frac{\overline{X}^n - \mu}{\hat{S}_n/\sqrt{n}} = \Psi_{n-1}$. Odhad. pro N(0,1) když nez. odch.$\alpha \in (0,1) : \Psi_{n-1}(t_{\alpha/2}) = 1 - \frac{\alpha}{2}$
$C_n = \langle \overline{X}_n - \delta; \overline{X}_n + \delta \rangle \ pro \ \delta = t_{\alpha/2}\frac{\hat{S}_n}{\sqrt{n}}$. Pak $P(\mu \in C_n) = 1 - \alpha$
-Jednovyberovy test (z-test): $X_1, ..., X_n$ náhodný výběr, σ zname ze zkusenosti $H_0 = \mu$, Zvolime $z_{\alpha/2} = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$ hranici, pak P(I.chyba)$= \alpha$
-Dvouvyběrový test: $H_0 : \vartheta_X = \vartheta_Y$, volíme $z = \frac{\overline{X}_n - \overline{Y}_m}{\sigma\sqrt{1/n + 1/m}}$. $S = \overline{X}_n - \overline{Y}_m, \mathbb{E}(S) = 0$. W stejně jako u 1-vyberoveho. Pokud nezname rozptyl, pouzijeme vyberovy rozptyl. P-hodnota: Říká, kolik % měření je horších.(the probability of obtaining test results at least as extreme as the result actually observed)
-Intervalove odhady:$X_1, ..., X_n$ náh. vyb. z $N(\vartheta, \sigma^2)$ $\alpha \in (0,1)$. Pak $\Phi(z_{\alpha/2}) = 1 - \alpha/2$, $S_n = (X_1 + ... + X_n)/n$ (výběrový průměr)
$C_n = [\overline{X}_n - \delta'; \overline{X}_n + \delta'], \delta' = z'_{\alpha/2}\frac{\hat{S}_n}{\sqrt{n}} = z'_{\alpha/2}\frac{\sqrt{\frac{1}{n-1}\sum\cdots}}{\sqrt{n}}$ Pak $P(\vartheta \in C_n) = 1 - \alpha$ Dk. Pomocí CLV.
-Linearni regrese: data: $(x_i; y_i)$ pro $i = 1...n$, cil: $y = \vartheta_1 x + \vartheta_0$ merime pomocí kvadratické odchylky $\sum^n (y_i - (\vartheta_0 + \vartheta_1 x))^2$ - minimalizujeme.
-Permutační test: Máme k dispozici dve sady nezávislych nahodnych velicin $X_1...X_n \sim F_X, Y_1...Y_n \sim F_Y$, chceme rozhodnout, zda plati $H_0 : F_X = F_Y, nebo : H_1 : F_X \neq F_Y$. Napr: doba behu programu pred/po vylepseni. Hladina cholesterolu u lidi, co jedi/nejedi pilulku. Frekvence kratkych slov v textu autora X a Y. Nevime nic o vlast. F_X, F_Y, nečekáme, že jsou norm. Postup: Zvolíme vhodnou statistiku napr: $T(X_1, ..., X_n, Y_1, ..., Y_m) = |\overline{X}_n - \overline{Y}_m|$. Za predpokladu H_0 jsou "sechny permutace stejne" X_i, X_j se generovaly ze stejneho rozdeleni. $t_{obs} := T(X_1, ..., X_n, Y_1, ..., Y_m)$Nahodne zpermutujeme zadanych m+n cisel a pro kazdou permutaci vycislime T, dostaneme $T_1, T_2, ..., T_{(n+m)!}$. Jako p-hodnotu vzit pravd., ze $T > t_{obs}$. Neboli $p = \frac{1}{(m+n)!}\sum_j I(T_j > t_{obs})$. Tj pravd. I. chyby, H_0 zamitneme, pokud $p < \alpha = 0.05$ (α zvolime)
-Bootstrap: Z naměřených dat $x_1, \ldots, x_n \sim F$ vytvoříme $\hat{F}_n$. Další data můžeme samplovat z $\hat{F}_n$. To se dělá tak, že vybereme uniformně náhodne $i \in \{1, \ldots, n\}$ a řekneme x_i. Základní použití: $T_n = g(X_1, \ldots, X_n)$ nějaká statistika. Chceme odhadnout $var(T_n)$. Nasamplujeme $X_1^*, \ldots, X_n^* \sim \hat{F}_n$. Spočteme $T_n^* = g(X_1^*, \ldots, X_n^*)$. Opakujeme B-krát, dostaneme $T_{n,1}^*, ..., T_{n,B}^*$. Odhad rozptylu: $\frac{1}{B}\sum_{b=1}^B (T_{n,b}^* - \frac{1}{B}\sum_{k=1}^B T_{n,k}^*)^2$
-Bayesovská statistika: Srovnání: Frekventistický/klasický přístup: ∘Pravděpodobnost je frekvence (z 6000 hodu kostkou padla šestka 1026-krát). Je to objektivní vlastnost reálného sveta. ∘Parametry jsou pevné, neznámé konstanty. Nelze o nich říkat smysluplné pravdepodobnostní výroky. ∘Navrhujeme statistické procedury tak, aby mely žádané dlouhodobé vlastnosti. Napr. 95 % z našich intervalových odhadů pokryje neznámý parametr. |Vs. Bayesovský prístup: Pravd. popisuje, jak moc veríme nejakému jevu, jak moc jsme ochotní se vsadit. (Pravdepodobnost, že Thomas Bayes m el 18. prosince 1760 šálek caje, je 90 %.) Mužeme vyslovovat pravdep. výroky i o parametrech (trebaže jsou to pevné konstanty). Spocítáme distribuci ϑ a z ní tvoríme bodové a intervalové odhady, atd. Základní popis Bayes. metody: ∘neznámý parametr považujeme za náhodnou velicinu Θ. ∘Zvolíme apriorní distribuci (prior distribution), neboli hustotu pravdepodobnosti $f_\Theta(\vartheta)$ nezávislou na datech. ∘zvolíme statistický model $f_{\Theta|X}(\vartheta|x)$, který popisuje, co nameríme (s jakou pravdepodobností), v závislosti na hodnote parametru ∘poté, co pozorujeme hodnotu X = x, spocítáme posteriorní distribuci (posterior distribution) $f_{\Theta|X}(\vartheta|x)$ ∘z té pak odvodíme, co potrebujeme napr najdeme a, b, aby $\int_a^b f_{\Theta|X}(\vartheta|x)d\vartheta \geq 1 - \alpha$ | Bayesovské bodové odhady: MAP – Maximum A-Posteriori: Volíme $\hat{\vartheta}$ tak, aby maximalizovalo $p_{\Theta|X}(\vartheta|x)$ v diskrétním případě, $f_{\Theta|X}(\vartheta|x)$ ve spojitém. LMS – Least Mean Square: Též metoda podmíněné strední hodnoty: Volíme $\hat{\vartheta} = \mathbb{E}(\Theta|X = x)$. Př.: Bayes. klasifikátor spamů: 1.vytvoríme seznam podezrelých slov (money, win, pharmacy, . . .) 2. N.v. X_i popisuje, zda email obsahuje podezřelé slovo w_i. 3. N.v. Θ popisuje, zda email je spam $\Theta = 1$ nebo ne $\Theta = 0$. 4. Z pˇredchozích email°u získáme odhady $p_{X|\Theta}$ a p_Θ. Použijeme Bayesovu vetu na výpocet $p_{X|\Theta}$.

$$cov(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}\left[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right]$$
$$var(x) = \mathbb{E}\left[(X - \mathbb{E}(X))^2\right] = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$
$$var(aX + b) = a^2 var(X)$$
$$var(X + Y) = var(X) + var(Y) + 2cov(X,Y)$$
$$\mathbb{E}(g(X)) = \sum_{x \in Im(X)} g(x)P(X = x)$$

Pro X, Y nezávislé: $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$
Cauchy $cov(X,Y) \leq \sqrt{var(X) \cdot var(Y)}$
neboli $-\sigma_X \cdot \sigma_Y \leq cov(X,Y) \leq \sigma_X \cdot \sigma_Y$

Diskrétní rozdělení (P, E, var)
$Bern(p), p_X(1) = p, p, p(p-1)$
$Binom(n,p), \binom{n}{k}p^k(1-p)^{n-k}, pn, pn(1-p)$
$Geom(p), p_X(x) = (1-p)^{x-1}p, \frac{1}{p}, \frac{1-p}{p^2}$
$Hypergeom(N, K, n), p_X(k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}, n\frac{K}{N}, -$
$Pois(\lambda), p_X(k) = e^{-\lambda}\frac{\lambda^k}{k!}, \lambda, \lambda \quad (\lim Binom(n, \frac{\lambda}{n}))$
$U(a,b), \frac{1}{b-a+1}, \frac{a+b}{2}, -$
Odhad Bin pomocí Pois: $Bin(n;p) \to Pois(np)$

Spojité rozdělení (f, F, E, var)
$U(a,b), \frac{1}{b-a}, \frac{x-a}{b-a}, \frac{a+b}{2}, \frac{(b-a)^2}{12}$
$Exp(\lambda), \lambda e^{-\lambda x}, 1 - e^{-\lambda x}, \frac{1}{\lambda}, \frac{1}{\lambda^2}$
$N(\mu, \sigma^2), \frac{1}{\sigma}\varphi(\frac{x-\mu}{\sigma}) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \Phi(\frac{x-\mu}{\sigma}), \mu, \sigma^2$
$X_i \sim N(\mu_i, \sigma_i^2) \implies X_1 + ... + X_n \sim N(\mu_1 + ... + \mu_n, \sigma_1^2 + ...$
$\Phi(-x) = 1 - \Phi(x)$

x	0	1	2	3	4
$\Phi(x)$	0.5	0.84135	0.97725	0.99865	0.99997

$Cauchy, \frac{1}{\pi(x+1)^2}, \frac{1}{\pi}\arctan x + \frac{1}{2}, undef, undef$

Náhodné vektory
> Hustota a distribuční funkce
$F_{X,Y}(x,y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(s,t)dtds$
$f_{X,Y}(x,y) = \int_{-\infty}^x = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}$
> Marginální hustota
$f_X(x) = \int_{-\infty}^\infty f_{X,Y}(x,y)dy$
> Množiny
$P((X,Y) \in A) = \int_A f_{X,Y}(x,y)dxdy$
> Nezávislost X,Y
$\iff f_{X,Y}(x,y) = f_X(x)f_Y(y) \iff F_{X,Y}(x,y) = F_X(x)F_Y(y)$