Poznámky z přednášek
Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

Pravděpodobnost a statistika
BONUS

Peter Černo, 2010
petercerno@gmail.com

Garant: prof. RNDr. Jaromír Antoch, CSc.
E-mail: Jaromir.Antoch@mff.cuni.cz
Domácí stránka: `http://www.karlin.mff.cuni.cz/~antoch/`

Anotace: Zavedení základních pojmů a metod teorie pravděpodobnosti a matematické statistiky a příklady jejich aplikací. Jedná se zejména o pojem pravděpodobnosti, náhodné veličiny a jejího rozdělení, nezávislosti, náhodného výběru a jeho popisných charakteristik, konstrukci odhadů, testování hypotéz, náhodné generátory. Důraz je kladen na praktické použití metod s využitím dostupného statistického software.

Sylabus:

1. Základní pojmy teorie pravděpodobnosti - náhodné jevy, pravděpodobnost, podmíněná pravděpodobnost, věta o úplné pravděpodobnosti a Bayesova věta, nezávislost náhodných jevů

2. Náhodné veličiny a jejich rozdělení - náhodná veličina, diskrétní náhodná veličina, charakteristiky diskrétní náhodné veličiny, některé typy diskrétně rozdělených náhodných veličin, spojitá náhodná veličina, charakteristiky spojité náhodné veličiny, některé typy spojitě rozdělených náhodných veličin, centrální limitní věta

3. Náhodné vektory a jejich rozdělení - náhodný vektor, charakteristiky rozdělení náhodného vektoru, nezávislost náhodných vektorů, charakteristiky lineární kombinace náhodných veličin, vícerozměrné normální rozdělení

4. Úvod do matematické statistiky - náhodný výběr, uspořádaný výběr, přehled běžně užívaných popisných statistik

5. Teorie odhadu - bodové odhady, bodové odhady parametrů pro vybraná rozdělení, intervaly spolehlivosti

6. Teorie testování hypotéz - úvod do testování hypotéz, jedno- a dvouvýběrová analýza pro normální rozdělení, párový test, testy nulovosti korelačního koeficientu, test chí-kvadrát dobré shody

7. Regrese - lineární regrese s jednou vysvětlující proměnnou, lineární regrese s více vysvětlujícími proměnnými

8. Simulace - generátory náhodných čísel a základy simulací Monte Carlo

Cíl předmětu: Studenti se seznámí se základy teorie pravděpodobnosti a matematické statistiky. Hlavním cílem je pochopení podstaty statistických a pravděpodobnostních postupů prezentovaných v dalších přednáškách.

Literatura:

1. Anděl J., Statistické metody, MATFYZPRESS, Praha 1998.
2. Bartoszynski R. and Niewiadomska-Budaj M., Probability and Statistical Inference, J. Wiley, 1996.
3. Jarušková D., Matematická statistika, skriptum ČVUT, Praha 2000.
4. Zvára K. a Štěpán J., Pravděpodobnost a matematická statistika, MATFYZPRESS, Praha 1997.

This page is intentionally left blank.

1/1

02.12.2007 Su
PETER ČERNO

MAI059 PRAVDEPODOBNOSŤ A ŠTATISTIKA

MAXWELLOV-BOLTZMANOV MODEL

r ČASTÍC n PRIEHRADOK

ČASTICE SÚ ROZLÍŠITEĽNÉ

$|\Omega| = n^r$, $P(K_1=k) = \binom{r}{k}\left(\frac{1}{n}\right)^k\left(\frac{n-1}{n}\right)^{r-k}$

$\lim_{n\to\infty} P(K_1=k) = \frac{\lambda^k}{k!}e^{-\lambda}$ $\quad \frac{r}{n}\to\lambda$

(POISSONOVE ROZDELENIE)

$P(\exists\,\emptyset\ \text{PRIEHR.}) = P\left(\bigcup_{i=1}^n B_i\right) =$

$\binom{n}{1}\left(\frac{n-1}{n}\right)^r - \binom{n}{2}\left(\frac{n-2}{n}\right)^r + \dots + (-1)^n\binom{n}{n-1}\left(\frac{1}{n}\right)^r$

BOSEOV-EINSTEINOV MODEL

r ČASTÍC, n PRIEHRADOK

ČASTICE SÚ NEROZLÍŠITEĽNÉ

$|\Omega| = \binom{n+r-1}{r}$ $\quad P(K_1=k) = \dfrac{\binom{n-1+(r-k)-1}{r-k}}{\binom{n+r-1}{r}}$

$\lim_{n\to\infty} P(K_1=k) = \left(\frac{1}{1+\lambda}\right)\left(1-\frac{1}{1+\lambda}\right)^k$ $\quad \frac{r}{n}\to\lambda$

(GEOMETRICKÉ ROZDELENIE)

$P(\exists\,\emptyset\ \text{PR.}) = 1 - \binom{n+(r-n)-1}{r-n}\Big/\binom{n+r-1}{r}$

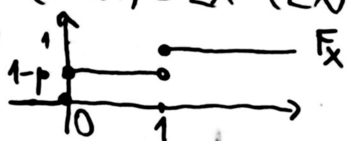
DISKRÉTNE NÁH. VELIČINY, ROZDELENIA

1. ALTERNATÍVNE $X\sim Alt(p)$ $\quad X\in\{0,1\}$ $\quad P(X=1)=p$, $P(X=0)=1-p=q$

$A_{Alt}(t)=q+pt$, $EX=A'(1)=p$, $\operatorname{var}X = E(X-EX)^2 = EX^2-(EX)^2 =$
$= A''(1)+A'(1)-(A'(1))^2 = p(1-p)$

PRE X, Y NEZÁVISLÉ
$A_{X+Y}(t)=A_X(t).A_Y(t)$

HOD MINCOU, ÁNO/NIE [ROVNOMERNÉ ROZDELENIE]



2. BINOMICKÉ $X\sim Bin(n,p)$ $\quad X\in\{0,\dots,n\}$ $\quad P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$

$A_{Bin}(t) = (q+pt)^n$, $EX=np$, $\operatorname{var}X=np(1-p)$ $\quad | \quad np_n\to\lambda \Rightarrow P(X=k) \to \frac{e^{-\lambda}\lambda^k}{k!}$

POČET ÚSPECHOV, $\sum_{i=1}^n X_i$ $\quad X_i\sim Alt(p)$ NEZÁV.

3. POISSONOVE $X\sim Poiss(\lambda)$ $\quad X\in\{0,1,\dots\}$ $\quad P(X=k)=\frac{e^{-\lambda}\lambda^k}{k!}$

$A_{Poiss}(t)=\sum_{k=0}^\infty \frac{e^{-\lambda}\lambda^k}{k!}t^k = e^{-\lambda}\cdot e^{\lambda t}$, $EX=\lambda$, $\operatorname{var}X=\lambda$

POČTY NIEČOHO

4. GEOMETRICKÉ $X\sim Geom(p)$ $\quad X\in\{0,1,\dots\}$ $\quad P(X=k)=p(1-p)^k$

$A_{Geom}(t) = \frac{p}{1-qt}$ $\quad A'_{Geom}(t) = \frac{pq}{(1-qt)^2}$, $EX=\frac{q}{p}$, $\operatorname{var}X = \frac{q^2}{p^2} + \frac{q}{p}$

5. HYPERGEOMETRICKÉ $X\sim \overset{HPG}{HyperGeom}(a,b,n)$ $\quad 0\le n\le a+b$

$X\in\{0,1,\dots,n\}$ VÝBER n PRVKOV (S VRACANÍM), a DOBRÝCH, b ZLÝCH

$P(X=k) = \binom{a}{k}\binom{b}{n-k}\Big/\binom{a+b}{n}$ $\quad \max(0,n-b)\le k\le \min(a,n)$

6. NEGATÍVNE BINOMICKÉ $X\sim \overset{NBin}{NegBin}(r,p)$

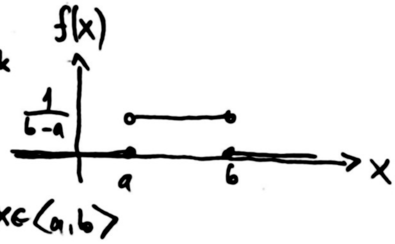
$P(X=k) = \binom{k-1}{k-r} p^r(1-p)^{k-r}$ $\quad$ PRAVDEPODOBNOSŤ r-TÉHO ZDARU PRI k-TOM POKUSE

$P(X=k) = \binom{k+r-1}{k} p^r (1-p)^k$

7. MULTINOMICKÉ $X\sim Multi(n,p_0,\dots,p_n)$ $\quad P(X_1=x_1,\dots,X_n=x_n) = \binom{n}{x_1,\dots,x_n} p_1^{x_1}\cdots p_n^{x_n}$

SPOJITÉ NÁHODNÉ VELIČINY, ROZDELENIA

1. ROVNOMERNÉ, UNIFORMNÉ $X\sim Unif(a,b)$



$f(x)=1/(b-a)$ $a<b$, $x\in(a,b)$, $\mu_k=E(X-EX)^k$

$2\mid k$ $\mu_k=(b-a)^k/2^k(k+1)$ $\quad 2\nmid k$ $\mu_k=0$

$EX=(a+b)/2$, $\operatorname{var}X=(b-a)^2/12$, $F(x)=\frac{x-a}{b-a}$ $x\in\langle a,b\rangle$

2. EXPONENCIÁLNE $X\sim Exp(\lambda)$

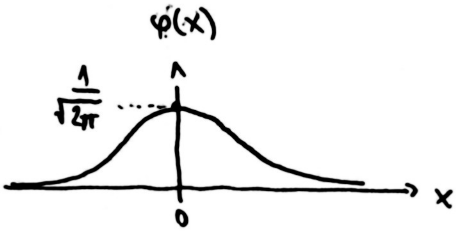
$f(x)=\frac{1}{\lambda}e^{-x/\lambda}$ $x>0$

GAMMA FCIA. $\Gamma(a)=\int_0^\infty x^{a-1}e^{-x}dx$, $a\in\mathbb{R}$

$EX^k=\lambda^k\Gamma(k+1)$, $EX^k=k!\lambda^k$ $k\in\mathbb{N}$

$EX=\lambda$, $\operatorname{var}X=\lambda^2$, $F(x)=1-e^{-\frac{x}{\lambda}}$

3. NORMÁLNE $X\sim N(\mu,\sigma^2)$



$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $\quad\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$

$X\sim N(\mu,\sigma^2), Y\sim N(\nu,\tau^2)\Rightarrow aX\sim(a\mu,a^2\sigma^2),\ X+Y\sim N(\mu+\nu,\sigma^2+\tau^2)$

Φ NEMOŽNO VYJADRIŤ ANALYTICKY $\quad\Phi(-x)=1-\Phi(x)$

$\Phi^{-1}(0,95)\doteq 1,64$, $\Phi^{-1}(0,975)\doteq 2$

ČEBYŠEVOVA NEROVNOSŤ: $\forall\varepsilon>0\ P(|X-\mu|\geq\varepsilon)\leq\frac{\sigma^2}{\varepsilon^2}$ $\quad\mu=EX$, $\sigma^2=\operatorname{var}X$

X_i SÚ NEZ. S ROVN. ROZD. $P(|\bar{X}_n-\mu|\geq\varepsilon)\leq\frac{\sigma^2}{n\varepsilon^2}$ $\quad(\operatorname{var}X_1<\infty)$, $\mu=EX_1$, $\sigma^2=\operatorname{var}X_1$

SILNÝ ZÁKON VEĽKÝCH ČÍSIEL $P(\bar{X}_n\to\mu)=1$.

CENTRÁLNA LIMITNÁ VETA: $P\left(\frac{S_n-ES_n}{\sqrt{\operatorname{var}S_n}}\leq x\right)\approx\Phi(x)$

$\mu_k'=EX^k$ $\quad\mu_k=E(X-EX)^k$

ŠIKMOSŤ $\alpha_3=\frac{\mu_3}{\sigma^3}$, ŠPICATOSŤ $\alpha_4=\frac{\mu_4}{\sigma^4}$,

$\operatorname{cov}(X,Y)=E(X-EX)(Y-EY)$

$\operatorname{corr}(X,Y)=\operatorname{cov}(X,Y)/\sqrt{\operatorname{var}X\cdot\operatorname{var}Y}$

$\psi(t)=Ee^{tX}=\sum_{k=0}^{\infty}e^{tx_k}p_k$

$\psi'(0)=EX$, $\psi''(0)=EX^2$

02.02.2008 So
PETER ČERNO

MAI059 PRAVDEPODOBNOSŤ A ŠTATISTIKA - SKÚŠKA

SYSTÉM $\mathcal{A}$ SA NAZÝVA **ALGEBRA**, KEĎ PLATÍ

$A,B \in \mathcal{A} \Rightarrow A\cup B \in \mathcal{A},\ A\cap B \in \mathcal{A},\ A^c \in \mathcal{A}$

$\Omega \in \mathcal{A},\ \emptyset \in \mathcal{A}$

$\omega \in \Omega$ = **ELEMENTÁRNY JAV**

FUNKCIA $P(A)$ JE **PRAVDEPODOBNOSŤ**, AK:

$A\in\mathcal{A} \Rightarrow P(A)\ge 0$

$A,B\in\mathcal{A},\ A\cap B=\emptyset \Rightarrow P(A\cup B)=P(A)+P(B)$

$P(\Omega)=1,\ P(\emptyset)=0$

TROJICA $(\Omega,\mathcal{A},P)$ JE **KLASICKÝ PRAVD. PRIESTOR**, AK [Ω JE KONEČNÁ $\neq\emptyset$, $\mathcal{A}=\mathcal{P}(\Omega)$, $P(A)=\frac{|A|}{|\Omega|}$]

NÁHODNÁ VELIČINA NA KLAS. PRAV. PRIEST. JE ĽUBOVOĽNÁ $X:\Omega\to\mathbb{R}$

FUNKCIA $P[X=x]$ $x\in\mathbb{R}$ JE **ROZDELENIE N.V.**

$EX=\frac{1}{|\Omega|}\sum_{\omega\in\Omega}X(\omega)=\sum_x x\, P[X=x]$ **STR. HOD.**

KOLMOGOROVOVA DEFINÍCIA PRAVDEPODOBNOSTI

NECH Ω JE ĽUBOVOĽNÁ MNOŽINA $\neq\emptyset$. **NEPRÁZDNY** SYSTÉM $\mathcal{A}$ PODMNOŽÍN MNOŽINY Ω SA NAZÝVA **σ-ALGEBRA**, AK:

$A\in\mathcal{A} \Rightarrow A^c\in\mathcal{A}$

$A_i\in\mathcal{A}\ \ i\in\mathbb{N} \Rightarrow \bigcup_{i=1}^{\infty}A_i\in\mathcal{A}$

PRAVDEPODOBNOSŤOU SA NAZÝVA REÁLNA FUNKCIA (NEZÁPORNÁ σ-ADITÍVNA FUNKCIA) $P(A)$ DEFINOVANÁ NA $\mathcal{A}$, KTORÁ PRE $A\in\mathcal{A}$, $A_1,A_2,\ldots\in\mathcal{A}$ $A_i\cap A_j=\emptyset$ $i\neq j$:

$P(\Omega)=1,\ P(A)\ge 0,$

$P\left(\bigcup_{i=1}^{\infty}A_i\right)=\sum_{i=1}^{\infty}P(A_i)$

TROJICA $(\Omega,\mathcal{A},P)$ JE **PRAVD. PRIESTOR**

NECH $P(B)>0$, POTOM **PODMIENENÁ PRAVD.**: $P(A|B)=\frac{P(A\cap B)}{P(B)}$.

$A_1,A_2,\ldots\in\mathcal{A}$ **ÚPLNÝ SYSTÉM JAVOV** AK:

$A_i\cap A_j=\emptyset\ \ i\neq j,\quad \bigcup_{i=1}^{\infty}A_i=\Omega$

AK $P(A_i)>0\ \forall i\in\mathbb{N}$, POTOM $P(B)=\sum_{i=1}^{\infty}P(B|A_i)P(A_i)$.

BAYESOVA VETA: $P(A_j|B)=\frac{P(B|A_j).P(A_j)}{\sum_{i=1}^{\infty}P(B|A_i).P(A_i)}$

JAVY A,B SÚ **NEZÁVISLÉ**, AK:
$P(A\cap B)=P(A).P(B)$

$A_1,A_2,\ldots$ SÚ **NEZÁVISLÉ**, AK: $P\left(\bigcap_{j=1}^{k}A_{i_j}\right)=\prod_{j=1}^{k}P(A_{i_j})$

(**PO DVOCH NEZÁVISLÉ** - KAŽDÉ DVA SÚ NEZ.)

[NEUSPORIADANÝ / USPORIADANÝ] **VÝBER** [BEZ VRACANIA / S VRACANÍM]

MAXWELLOV - BOLTZMANOV MODEL

KU KAŽDEJ ČASTICI INDEX PRIEHRADKY

r ROZLÍŠITEĽNÝCH ČASTÍC, n PRIEHRADOK

KLASICKÝ PRAVD. PRIESTOR, ELEM. JAVY $\omega = (s_1, \dots, s_r)$ $1 \le s_i \le n$

$|[K_1 = k]| = \binom{r}{k}(n-1)^{r-k}$, $p_{k,n,r} = \binom{r}{k}\frac{(n-1)^{r-k}}{n^r} = \binom{r}{k}\left(\frac{1}{n}\right)^k\left(1-\frac{1}{n}\right)^{r-k}$

AK $\lim_{n\to\infty}\frac{r_n}{n} = \lambda \Rightarrow \lim_{n\to\infty} p_{k,n,r_n} = \frac{\lambda^k}{k!}e^{-\lambda}$ (POISSONOVE ROZDELENIE)

$EK_1 = \frac{r}{n}$

$P([K_1 = r_1, \dots, K_n = r_n]) = \frac{1}{n^r}\frac{r!}{r_1! \dots r_n!}$ $r_1 + \dots + r_n = n$

BOSEOV - EINSTEINOV MODEL

KU KAŽDEJ PRIEHR. POČET ČASTÍC

r NEROZLÍŠITEĽNÝCH ČASTÍC, n PRIEHRADOK, $\omega = (r_1, \dots, r_n)$

STAV SYSTÉMU JE URČENÝ POČTAMI ČASTÍC V PRIEHR., V STAVY SÚ ROVN. PRAV.

$|\Omega| = \binom{r+n-1}{r}$

$p_{k,n,r} = P[K_1 = k] = \binom{r-k+n-2}{r-k} / \binom{r+n-1}{r}$

AK $\lim_{n\to\infty}\frac{r_n}{n} = \lambda \Rightarrow \lim_{n\to\infty} p_{k,n,r_n} = \frac{\lambda^k}{(1+\lambda)^{k+1}}$

FERMIOV - DIRACOV MODEL

= BOSEOV - EINSTEINOV MODEL, V KAŽDEJ PRIEHRADKE NAJVIAC 1 ČASTICA

PÓLYOVE URNOVÉ SCHÉMA

$a, b, n \in \mathbb{N}$, $\Delta \in \mathbb{Z}$. V URNE a ČIERNYCH, b BIELYCH GÚĽ, n-KRÁT ŤAHÁME, VRÁTIME A PRIDÁME Δ GÚĽ ROVNAKEJ FARBY

$x^{[k]} \overset{def}{:=} x(x+\Delta)\dots(x+(k-1)\Delta)$ AK $x+(k-1)\Delta > 0$, INAK $= 0$. $x^{[0]} = 1$

$D_k \overset{def.}{:=}$ [V n ŤAHOCH BOLO VYTIAHNUTÝCH k BIELYCH GÚĽ] $k = 0, 1, \dots$

$|D_k| = \binom{n}{k} b^{[k]} a^{[n-k]}$, $1 = \sum_{k=0}^{n} P(D_k) = \sum_{k=0}^{n}\binom{n}{k}\frac{b^{[k]}a^{[n-k]}}{(a+b)^{[n]}}$

$B_i :=$ V i-TOM ŤAHU BOLA ŤAHANÁ BIELA GUĽA

$P(B_i) = \frac{b}{b+a}$

$\Delta = 0$: BERNOULLIHO SCHÉMA

$\Delta = -1$: PEARSONOVO SCHÉMA

2/5

02.02.2008 Sa
PETER ČERNO

MAI059 PRAVDEPODOBNOSŤ A ŠTATISTIKA - SKÚŠKA

NÁHODNÁ PRECHÁDZKA

$S_0 = 0, \quad P[S_{k+1} - S_k = 1] = P[S_{k+1} - S_k = -1] = \frac{1}{2}$

PRE POSTUPNOSŤ $S_0, \dots, S_n$ JE $\Omega = \{0,1\}^n$

$P[S_n = 2k-n] = \binom{n}{k}\frac{1}{2^n} = P[S_n = n-2k] \Rightarrow ES_n = 0$

$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n (1+\varepsilon_n), \ \lim_{n\to\infty} \varepsilon_n = 0$ (STIRLINGOV VZOREC)

$\Rightarrow P[S_{2n} = 0] = \frac{1}{\sqrt{\pi n}}(1+\delta_n), \ \lim_{n\to\infty} \delta_n = 0$

$x = p+q, \ y = p-q, \ N_{x,y} = \binom{p+q}{p} = \binom{x}{\frac{x+y}{2}}, \ x,y>0 \Rightarrow N^+_{x,y} = \frac{y}{x} N_{x,y}$

NÁHODNÁ VELIČINA

$X: (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{B})$ JE NÁHODNÁ VELIČINA, AK:

$\mathcal{B}$ JE BORELOVSKÁ σ-ALGEBRA (NAJMENŠIA σ-ALG., KTORÁ OBSAHUJE $\forall$ INTERVALY $(-\infty, x)$)

$[X<b] \overset{def}{:=} \{\omega \in \Omega : X(\omega) < b\}$

$[a<X<b] \overset{def}{:=} \{\omega \in \Omega : a < X(\omega) < b\}$ ...

$\forall x \in \mathbb{R} : [X<x] \in \mathcal{A}$ (X SA NAZÝVA MERATEĽNÁ)

$\Rightarrow \forall B \in \mathcal{B} \quad X^{-1}(B) \in \mathcal{A} \qquad P_X(B) \overset{def}{:=} P(X^{-1}(B))$

P_X NAZÝVA SA ROZDELENIE PRAVDEPODOBNOSTI NÁH. VELIČINY X

$F_X(x) \overset{def}{:=} P[X<x]$ JE DISTRIBUČNÁ FUNKCIA NÁH. VEL. X

(NIEKEDY TIEŽ $F_X(x) \overset{def}{:=} P[X \leq x]$)

DISKRÉTNE ROZDELENIA

VYTVÁR. FCIA. PRE DISKR. ROZD.:

$A(t) = \sum_{i=0}^{\infty} P[X=i]\, t^i$

$\exists x_0 < x_1 < x_2 < \dots \in \mathbb{R}, \quad \sum_{i=0}^{\infty} P[X = x_i] = 1,$

ZATIAĽ UVAŽUJME $x_i = i \ \forall i \in \mathbb{N}_0$

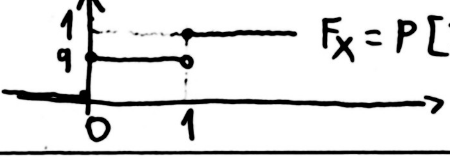
ALTERNATÍVNE ROZDELENIE

$X \sim Alt(p) \quad X \in \{0,1\}$

$P(X=0) = q = 1-p, \quad P(X=1) = p$ PRAVDEPODOBNOSŤ ZDARU

$A_{Alt}(t) = q + p\cdot t, \quad EX = A'(1) = p, \quad var X = EX^2 - (EX)^2 =$

$A''(1) + A'(1) - (A'(1))^2 = p(1-p)$



$F_X = P[X \leq x]$

HOD MINCOU, ÁNO/NIE

BINOMICKÉ ROZDELENIE $X \sim Bin(n, p)$ $X \in \{0, 1, \dots, n\}$

$P[X=k] = \binom{n}{k} p^k (1-p)^{n-k}$ PRAVD. k ZDAROV PRI n POKUSOCH

$A_{Bin}(t) = (q + pt)^n$, $EX = np$, $var\, X = np(1-p)$

POČET ZDAROV, $X_i \sim Alt(p)$ NEZÁVISLÉ, POTOM $X = \sum_{i=1}^{n} X_i$

$n \cdot p_n \to \lambda \Rightarrow P(X=k) \to \frac{\lambda^k}{k!} e^{-\lambda}$ (POISSONOVE ROZD.)

POISSONOVE ROZDELENIE $X \sim Poiss(\lambda)$ $X \in \{0, 1, 2, \dots\}$

$P[X=k] = \frac{\lambda^k}{k!} e^{-\lambda}$, APROXIMÁCIA $\binom{n}{k} p^k (1-p)^{n-k}$ PRE $\lambda = np$

$A_{Poiss}(t) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} \cdot t^k = e^{-\lambda} \cdot e^{\lambda t}$, $EX = \lambda$, $var\, X = \lambda$

λ SA NAZÝVA NIEKEDY INTENZITA VÝSKYTU UDALOSTÍ

GEOMETRICKÉ ROZDELENIE $X \sim Geom(p)$ $X \in \{0, 1, 2, \dots\}$

$P[X=k] = p \cdot (1-p)^k$ PRAVD. k NEZDAROV PRED PRVÝM ZDAROM

$A_{Geom}(t) = \frac{p}{1-qt}$ $A'_{Geom}(t) = \frac{pq}{(1-qt)^2}$, $EX = \frac{q}{p}$, $var\, X = \frac{q}{p^2}$

NEGATÍVNE BINOMICKÉ ROZD. $X \sim NGeom(r, p)$ $X \in \{0, 1, 2, \dots\}$

$P[X=k] = \binom{k+r-1}{r-1} p^{r-1} \cdot (1-p)^k \cdot p$ PRAV. k NEZDAROV PRED r-TÝM ZDAROM

$EX = r\frac{q}{p}$ $var\, X = r\frac{q}{p^2}$

HYPERGEOMETRICKÉ ROZDELENIE $X \sim HPG(A, \overset{A+B}{\tilde{N}}, n)$ $0 \le n \le N$

$X \in \{\max(0, \overbrace{A+n-N}^{n-B}), \dots, \min(A, n)\}$

N=A+B PREDMETOV, A MÁ VLASTNOSŤ V.

X UDÁVA POČET PREDMETOV S VL. V Z n VYBRANÝCH PREDMETOV

$P[X=k] = \frac{\binom{A}{k}\binom{N-A}{n-k}}{\binom{N}{n}}$

MULTINOMICKÉ ROZDELENIE $X \sim Multi(n, p_1, \dots, p_n)$

$P(X_1 = x_1, \dots, X_n = x_n) = \binom{n}{x_1, \dots, x_n} p_1^{x_1} \cdots p_n^{x_n}$.

02.02.2008 So
Peter Černo

MAI059 Pravdepodobnosť a štatistika - skúška

Spojité rozdelenia

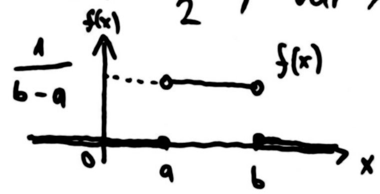
Náh. veličina X má spojité rozdelenie, keď existuje f_X:
$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$, f_X je <u>hustota rozdelenia</u> náh. vel. X

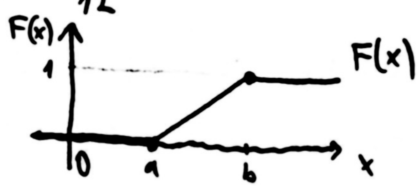
Rovnomerné, uniformné r. $X \sim Unif(a,b)$ $a<b$

$f(x) = \begin{cases} 1/(b-a) & x \in (a,b) \\ 0 & x \notin (a,b) \end{cases}$ pre $\mu_k = E(X-EX)^k$ je

$2 | k \Rightarrow \mu_k = \frac{(b-a)^k}{2^k(k+1)}$ $2 \nmid k \Rightarrow \mu_k = 0$

$EX = \frac{a+b}{2}$, $\mathrm{var}\, X = \frac{(b-a)^2}{12}$ $F(x) = \begin{cases} 0 & x \le a \\ \frac{x-a}{b-a} & x \in (a,b) \\ 1 & x \ge b \end{cases}$



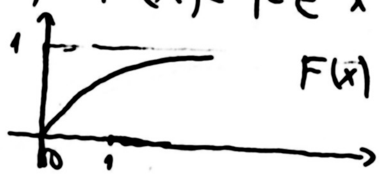


Exponenciálne rozdelenie $X \sim Exp(\lambda)$

$f(x) = \begin{cases} 0 & x \le 0 \\ \frac{1}{\lambda} e^{-x/\lambda} & x > 0 \end{cases}$

Gamma fcia. $\Gamma(a) = \int_0^\infty x^{a-1} \cdot e^{-x} dx$ $a>0$
$\Gamma(1/2) = \sqrt{\pi}$, $\Gamma(1) = 1$, $\Gamma(a+1) = a\Gamma(a)$, $\Gamma(n) = (n-1)!$

$EX^k = \lambda^k \Gamma(k+1)$ pre $k \in \mathbb{N}_0$ $= k!\lambda^k$
$EX = \lambda$, $\mathrm{var}\, X = \lambda^2$, $F(x) = 1 - e^{-\frac{x}{\lambda}}$

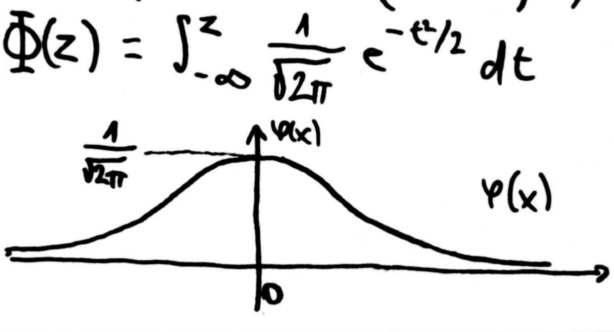


Normálne rozdelenie $X \sim N(\mu, \sigma^2)$

$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

$X \sim N(\mu, \sigma^2)$, $Y \sim (\nu, \tau^2) \Rightarrow aX \sim N(a\mu, a^2\sigma^2)$, $X+Y \sim N(\mu+\nu, \sigma^2+\tau^2)$

$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,dt$



ROZDELENIE χ^2 S 1 ST. VOLNOSTI $\quad X \sim \chi^2(1)$

NECH $Z \sim N(0,1)$. HĽADÁME HUSTOTU NÁH. VELIČINY $X = Z^2$.

$x > 0 \Rightarrow P[Z^2 < x] = P[-\sqrt{x} < Z < \sqrt{x}] = \int_{-\sqrt{x}}^{\sqrt{x}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz =$

$= \int_0^x \frac{1}{\sqrt{2\pi}} t^{-1/2} e^{-t/2} dt \quad \Rightarrow \quad f(x) = \frac{1}{\sqrt{2\pi}} x^{-1/2} e^{-x/2}$

NECH $F(x)$ JE DISTRIBUČNÁ FCIA. NÁH. VELIČINY X.
FUNKCIA $F^{-1}(u) \overset{def}{:=} \inf \{x \in \mathbb{R} : F(x) \geq u\} \quad 0 < u < 1$
SA NAZÝVA KVANTILOVÁ FUNKCIA NÁH. VEL. X.
$F^{-1}(\alpha)$ SA NAZÝVA α-KVANTIL. $F^{-1}(0{,}5)$ SA NAZÝVA MEDIÁN

$\Phi^{-1}(0{,}975) = 1{,}96 \Rightarrow P[Z < 1{,}96] = 0{,}975, \quad P[|Z| < 1{,}96] = 0{,}95$

AK $Z \sim N(0,1)$ POTOM PRE $\alpha \in (0,1)$ JE $z(\alpha) \overset{def}{:=} \Phi^{-1}(1-\alpha)$
TZV. KRITICKÁ HODNOTA PRE α. PLATÍ $P[Z > z(\alpha)] = \alpha$.

NÁHODNÝ VEKTOR $[\vec{X} < \vec{x}] = \{\omega \in \Omega : X_1(\omega) < x_1, \dots, X_n(\omega) < x_n\}$
$\mathcal{B}_n$ JE NAJMENŠIA σ-ALGEBRA NAD INTERVALMI
$(-\infty, x_1) \times \dots \times (-\infty, x_n) \quad \forall \vec{x} \in \mathbb{R}_n$.
A $\vec{x} \in \mathbb{R}_n \Rightarrow [\vec{X} < \vec{x}] \in \mathcal{A} \qquad \vec{X}(\omega) = (X_1(\omega), \dots, X_n(\omega))^T$
ZREJME $[\vec{X} < \vec{x}] = \bigcap_{i=1}^n [X_i < x_i]$

DISKRÉTNE ROZDELENIE $(X,Y)^T$ JE DANÉ ZOZNAMOM
HODNÔT $x_1, x_2, \dots, y_1, y_2, \dots$ A PRAVDEPODOBNOSŤAMI $P[X = x_1, Y = y_1]$
VZŤAH MEDZI ZDRUŽENÝM A MARGINÁLNYM ROZDELENÍM
$P[X = x_i] = \sum_{j=1}^{\infty} P[X = x_i, Y = y_j]$

SPOJITÉ ROZDELENIE $(X,Y)^T$
ZDRUŽENÁ DISTRIBUČNÁ FCIA. $F_{X,Y}(x,y) = P[X < x, Y < y]$,
ZDRUŽENÁ HUSTOTA $f_{X,Y}$: $F_{X,Y}(x,y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(u,v)\,dv\,du$.
MARGINÁLNA HUSTOTA X JE $\int_{-\infty}^{\infty} f_{X,Y}(x,v)\,dv$.

NEZÁVISLOSŤ NÁH. VELIČÍN: NÁH. VELIČINY $X_1, X_2, \dots$
SÚ (PO DVOCH) NEZÁVISLÉ $\Leftrightarrow \forall x_1, x_2, \dots \in \mathbb{R}$ SÚ
JAVY $[X_1 < x_1], [X_2 < x_2], \dots$ (PO DVOCH) NEZÁVISLÉ
$\Leftrightarrow$ PRE DISKR. $P[X = x_i, Y = y_j] = P[X = x_i] \cdot P[Y = y_j]$ $\qquad f_{X_1,\dots,X_n}(x_1,\dots,x_n) =$
PRE SPOJ. $F_{X_1,\dots,X_n}(x_1,\dots,x_n) = F_{X_1}(x_1) \dots F_{X_n}(x_n) \Rightarrow f_{X_1}(x_1) \dots f_{X_n}(x_n)$

4/5

02.02.2008 Sa
PETER ČERNO

MAI059 PRAVDEPODOBNOSŤ A ŠTATISTIKA - SKÚŠKA

NECH X JE NÁH. VELIČINA S DISKR. ROZDELENÍM, KTORÁ NADOBÚDA HODNÔT $x_1, x_2, \dots$ KEĎ $\sum_{i=1}^{\infty} x_i P[X = x_i]$ KONVERGUJE ABSOLÚTNE, NAZÝVAME JEJ HODNOTU <u>STREDNOU HODNOTOU</u> EX. ROZDELENIA X.

AK X MÁ SPOJITÉ ROZDELENIE S HUSTOTOU $f(x)$, POTOM AK JE $\int_{-\infty}^{\infty} |x| f(x)\,dx < \infty$, NAZÝVAME $\int_{-\infty}^{\infty} x f(x)\,dx$ <u>STREDNOU HODNOTOU</u> EX ROZDELENIA X.

NECH X JE NÁHODNÁ VELIČINA S DISKRÉTNYM ROZDELENÍM, KTORÁ NADOBÚDA HODNOTY $x_1, x_2, \dots$ PRE KAŽDÚ REÁLNU FUNKCIU g JE $Y = g(X)$ NÁH. VELIČINOU

$$P[Y = y_j] = P\{\omega \in \Omega : g(X(\omega)) = y_j\} = \sum_{i : g(x_i) = y_j} P[X = x_i].$$

PLATÍ $Eg(X) = \sum_{i=1}^{\infty} g(x_i) P[X = x_i]$

$E(X+Y) = EX + EY$,

KEĎ SÚ X, Y NEZÁVISLÉ, POTOM $EXY = EX \cdot EY$.

NECH X JE NÁHODNÁ VELIČINA S KONEČNOU STR. HODNOTOU. $\text{var}\, X = E(X - EX)^2$ SA NAZÝVA <u>ROZPTYL</u>, $\sqrt{\text{var}\, X}$ <u>SMERODATNÁ ODCHÝLKA</u>.

$\text{var}\, X = EX^2 - (EX)^2$

$\text{var}\,(a + bX) = b^2 \text{var}\, X$

<u>NORMOVANÁ</u> NÁH. VELIČINA $Z = \dfrac{X - EX}{\sqrt{\text{var}\, X}}$ MÁ NULOVÚ STREDNÚ HODNOTU A JEDNOTKOVÝ ROZPTYL

NECH PRE NÁH. VELIČINY X, Y EXISTUJÚ ROZPTYLY.

$\text{cov}(X,Y) = E(X - EX)(Y - EY)$ SA NAZÝVA <u>KOVARIANCIA</u> N.V. X, Y

$\text{cov}(X,Y) = EXY - (EX)(EY)$.

$\text{var}(X+Y) = \text{var}\, X + \text{var}\, Y + 2\,\text{cov}(X,Y)$.

<u>KORELAČNÝ KOEFICIENT</u> $\rho_{X,Y} = \text{cov}\left(\dfrac{X - EX}{\sqrt{\text{var}\, X}}, \dfrac{Y - EY}{\sqrt{\text{var}\, Y}}\right) = \dfrac{\text{cov}(X,Y)}{\sqrt{\text{var}\, X \cdot \text{var}\, Y}}$, $|\rho_{X,Y}| \leq 1$

VARIAČNÁ MATICA NÁH. VEKTORU $\vec{X} = (X_1, \dots, X_n)^T$

$\operatorname{var} \vec{X} = {}_{i\cdots}\left(\operatorname{cov}(X_i, X_j) \right)^{\vdots}_{j} = E(\vec{X} - E\vec{X})(\vec{X} - E\vec{X})^T$

$\operatorname{var} (\vec{a} + \vec{B}\vec{X}) = \vec{B} \cdot \operatorname{var} \vec{X} \cdot \vec{B}^T$

k-TY VŠEOBECNÝ MOMENT: $\mu'_k = EX^k$

k-TY CENTRÁLNY MOMENT: $\mu_k = E(X - EX)^k$

KOEFICIENT ŠIKMOSTI: $\gamma_1 = \dfrac{\mu_3}{(\sqrt{\operatorname{var} X})^3}$

KOEFICIENT ŠPIČATOSTI: $\gamma_2 = \dfrac{\mu_4}{(\sqrt{\operatorname{var} X})^4} - 3$

MOMENTOVÁ VYTVÁRAJÚCA FUNKCIA $M_X(t) = Ee^{tX}$ N. V. X

$$\mu'_r = \frac{d^r}{dt^r} M_X(t) \Big|_{t=0}$$

$$M_X(t) = \sum_{r=0}^{\infty} \frac{t^r}{r!} \cdot \mu'_r$$

$M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$ PRE NEZÁVISLÉ X, Y

PRE BINOMICKÉ ROZDELENIE $X \sim Bin(n, p)$, $P[X=k] = \binom{n}{k} p^k q^{n-k}$

JE $M_X(t) = \sum_{k=0}^{n} e^{t \cdot k} \binom{n}{k} p^k (1-p)^{n-k} = (p(e^t - 1) + 1)^n$

$\frac{d}{dt} M_X(t) = n (p(e^t - 1) + 1)^{n-1} e^t p \Rightarrow \frac{d}{dt} M_X(t) \Big|_{t=0} = \mu'_1 = np$.

PRE $Z \sim N(0,1)$ JE

$M_Z(t) = \int_{-\infty}^{\infty} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = \dots = e^{t^2/2}$

$\Rightarrow M'_Z(t) = te^{t^2/2}$

PRE $Z \sim N(\mu, \sigma^2)$ JE $M_Z(t) = e^{\mu t + \sigma^2 \frac{t^2}{2}}$.

02. 02. 2008 So
PETER ČERNO

MAI059 PRAVD. A ŠTATISTIKA – SKÚŠKA

KONVOLÚCIA AK X, Y SÚ NEZÁVISLÉ

a) $f_{X+Y}(w) = \int_{-\infty}^{\infty} f_X(x) f_Y(w-x)\,dx$

b) $P[X+Y=w_k] = \sum_{i=1}^{\infty} P[X=x_i]\,P[Y=w_k-x_i]$

ČEBYŠEVOVA NEROVNOSŤ

NECH $\operatorname{var} X < \infty$, $\varepsilon > 0$. POTOM $P[|X-EX| \geq \varepsilon] \leq \frac{\operatorname{var} X}{\varepsilon^2}$

BERNOULLI: $X_n \sim Bin(n,p)$ $n=1,2,\dots$ $p\in(0,1)$, $\varepsilon>0$

$\Rightarrow \lim_{n\to\infty} P\left[\left|\frac{X_n}{n} - p\right| \geq \varepsilon\right] = 0.$

(SLABÝ) ZÁKON VEĽKÝCH ČÍSIEL

NECH $X_1, X_2, \dots$ SÚ PO DVOCH NEZÁVISLÉ NÁH. VELIČINY TAKÉ, ŽE $EX_i = a$, $\operatorname{var} X_i < c$ $\forall i \in \mathbb{N}$, $|a|, c < \infty$

PRE $\forall \varepsilon > 0$ $\lim_{n\to\infty} P\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - a\right| < \varepsilon\right] = 1.$

BORELOV-CANTELLIOV NULA-JEDNOTKOVÝ PRINCÍP

NECH $F_1, F_2, \dots$ SÚ NÁH. JAVY V PRAVD. PRIESTORE $(\Omega, \mathcal{A}, P)$.

$[F_n, \infty \times n] \overset{def.}{:=} \{\omega \in \Omega : \omega$ JE PRVKOM ∞ VEĽA $F_n\}$

ZREJME $[F_n, \infty \times n] = \{\omega\in\Omega : \forall n\in\mathbb{N}\ \exists k(n) \geq n : \omega \in F_{k(n)}\} = \bigcap_{n=1}^{\infty} \bigcup_{k\geq n} F_k$

PLATÍ: $\sum_{k=1}^{\infty} P(F_k) < \infty \Rightarrow P[F_n, \infty \times n] = 0.$

KEĎ SÚ $F_1, F_2, \dots$ NEZÁVISLÉ, POTOM PLATÍ EKVIVALENCIA $\Leftrightarrow$

BORELOV SILNÝ ZÁKON VEĽKÝCH ČÍSIEL

NECH $X_1, X_2, \dots$ SÚ NEZÁVISLÉ, ROVNAKO ROZDELENÉ NÁH. VEL. S KONEČNOU STREDNOU HODNOTOU. $EX_1 = \mu$.

POTOM $P\left(\left\{\omega\in\Omega : \lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} X_j(\omega) = \mu\right\}\right) = 1.$

CENTRÁLNA LIMITNÁ VETA

NECH $Y_1, Y_2, \dots$ JE POSTUPNOSŤ NEZÁVISLÝCH NÁHODNÝCH VELIČÍN, PRE KTORÉ $EY_i = \mu$, $\operatorname{var} Y_i = \sigma^2 > 0$, $E|Y_i|^3 < \infty$.

PRE DISTRIBUČNÚ FUNKCIU NÁH. VEL.

$$S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{Y_i - \mu}{\sigma} \right)$$

PLATÍ $\lim_{n \to \infty} P[S_n < x] = \Phi(x)$.

BERRY-ESSÉNOVA NEROVNOSŤ:

NECH $Y_1, \dots, Y_n$ SÚ NEZÁVISLÉ, ROVNAKO ROZDELENÉ NÁH. VELIČINY, $EY_i = \mu$, $\operatorname{var} Y_i = \sigma^2 \in (0, \infty)$.

$$F_n(x) \overset{\text{def.}}{:=} P\left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{Y_i - \mu}{\sigma} < x \right]$$

POTOM $|F_n(x) - \Phi(x)| \leq 0{,}7975 \cdot \frac{1}{\sqrt{n}} \frac{E|Y_i - \mu|^3}{\sigma^3}$ $\quad x \in (-\infty, \infty)$

03.02.2008 Su
PETER ČERNO

MAI059 PRAVDEPODOBNOSŤ A ŠTATISTIKA - SKÚŠKA

ŠTATISTIKA

SÚBOR HODNÔT $x_1, \dots, x_n$, n = ROZSAH SÚBORU

$x_{(1)} \leq \dots \leq x_{(n)}$

MIERY POLOHY:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i \overset{\geq}{,} \bar{x}_G = \left(\prod_{i=1}^n x_i\right)^{1/n} \overset{\geq}{,} \bar{x}_H = \left(\frac{1}{n}\sum_{i=1}^n x_i^{-1}\right)^{-1}$$

VÝBEROVÝ PRIEM. GEOM. PRIEM. HARMONICKÝ PRIEM.

MEDIÁN $\tilde{x} = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & 2 \nmid n \\ \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}\right)+1}\right) & 2 \mid n \end{cases}$

$p \in (0,1)$ p-TÝ KVANTIL (PERCENTIL)

$$x_p = \begin{cases} x_{(\lfloor np \rfloor + 1)} & np \neq \lfloor np \rfloor \\ \frac{1}{2}\left(x_{\lfloor np \rfloor} + x_{\lfloor np \rfloor + 1}\right) & np = \lfloor np \rfloor \end{cases}$$

$Q_1 = x_{0,25}$ DOLNÝ KVARTIL

$Q_3 = x_{0,75}$ HORNÝ KVARTIL

MIERY VARIABILITY:

ROZPTYL $s_x^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^n x_i^2 - (\bar{x})^2$, s_x = SMERODATNÁ ODCHYLKA > 0

NIEKEDY SA V MENOVATELI POUŽÍVA $n-1$

ROZPÄTIE $R = x_{(n)} - x_{(1)}$

KVARTILOVÉ ROZPÄTIE $R_Q = Q_3 - Q_1$, $\frac{1}{2} R_Q$ KVART. ODCHYLKA

PRIEMERNÁ ODCHYLKA $d = \frac{1}{n}\sum_{i=1}^n |x_i - \bar{x}|$

DIAGRAMY: KRABICOVÝ, ROZPTYLOVÝ, MATICOVÝ ..

VÝBER

UVAŽUJME VEĽKÝ ŠTATISTICKÝ SÚBOR $x_1, \ldots, x_N$, OZNAČME
$\mu = \bar{x} = \frac{1}{N}\sum_{i=1}^N x_i$, $\sigma^2 = \frac{1}{N}\sum_{i=1}^N (x_i - \mu)^2$

VÝBER $\boxed{\text{BEZ VRACANIA}}$ Z KONEČNEJ POPULÁCIE

VÝBEROVÝ SÚBOR ($n < N$ ŠTATISTICKÝCH JEDNOTIEK)

NÁHODNÝ VÝBER $s \in \binom{\{1,\ldots,N\}}{n}$ MÁ PRAVD. $1/\binom{N}{n}$, TJ. $\Omega = \binom{\{1,\ldots,N\}}{n}$

ZAVEDIEME NÁH. VEKTOR $\vec{W} = (W_1, \ldots, W_N)^T$, KDE
$W_i(s) \overset{\text{def}}{:=} \begin{cases} 1, & i \in s \\ 0, & i \notin s \end{cases}$

ZREJME $P[W_i = 1] = \frac{n}{N}$, $EW_i = \frac{n}{N}$, $\operatorname{var} W_i = \frac{n}{N}\left(1 - \frac{n}{N}\right)$, $EW_iW_j = \frac{n(n-1)}{N(N-1)}$

VÝBEROVÝ PRIEMER: $\bar{X} := \frac{1}{n}\sum_{i=1}^N x_i W_i$

PLATÍ $E\bar{X} = \mu$, $\operatorname{var} \bar{X} = \frac{N-n}{N-1} \cdot \frac{\sigma^2}{n}$

VÝBEROVÝ ROZPTYL: $S^2 := \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$, KDE $X_i = x_{s_i}$, $s = \{s_1, \ldots, s_n\} \in \binom{\{1,\ldots,N\}}{n}$

PLATÍ $ES^2 = \frac{N}{N-1}\sigma^2$

NÁHODNÝ VÝBER $\boxed{\text{S VRACANÍM}}$

$n < N$, $\Omega = \{1, \ldots, N\}^n$, $|\Omega| = N^n$,

DEFINUJME $\vec{X} = (X_1, \ldots, X_n)^T$ PRE $s = (s_1, \ldots, s_n)^T \in \Omega$ JE $X_i(s) = x_{s_i}$

$X_1, \ldots, X_n$ SÚ ZREJME NEZÁVISLÉ NÁH. VEL.

$EX_i = \mu$, $\operatorname{var} X_i = \sigma^2$

VÝBEROVÝ PRIEMER: $\bar{X} := \frac{1}{n}\sum_{i=1}^n X_i$

PLATÍ $E\bar{X} = \mu$, $\operatorname{var} \bar{X} = \sigma^2/n$

VÝBEROVÝ ROZPTYL: $S^2 = \frac{1}{n-1}\sum_{i=1}^N (X_i - \bar{X})^2$

PLATÍ $ES^2 = \sigma^2$

PRE NÁHODNÝ VÝBER S VRACANÍM Z NORMÁLNEHO ROZDELENIA, TJ. $X_i \sim N(\mu, \sigma^2)$

PLATÍ: $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$.

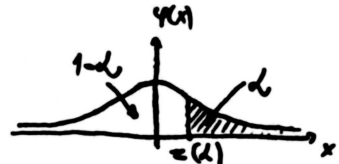
03.02.2008 Su
PETER ČERNO

MAI059 PRAVDEPODOBNOSŤ A ŠTATISTIKA - SKÚŠKA

NECH $X_1, \dots, X_n$ JE <u>NÁHODNÝ VÝBER Z NORMÁLNEHO ROZD.</u> $N(\mu, \sigma^2)$, KDE σ^2 POZNÁME

$\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$, PRIČOM $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

POLOŽME $Z := \dfrac{\bar{X} - E\bar{X}}{\sqrt{\operatorname{var}\bar{X}}} = \dfrac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \sim N(0,1)$

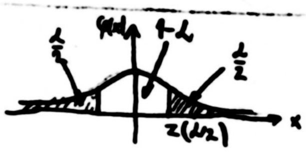


POTOM $P[Z > z(\frac{\alpha}{2})] = \frac{\alpha}{2} \Rightarrow$ KDE $z(\alpha) = \Phi^{-1}(1-\alpha)$

$P[|Z| > z(\frac{\alpha}{2})] = \alpha \Rightarrow$ $[P[Z \le x] = \Phi(x) \Rightarrow P[Z > x] = 1 - \Phi(x)]$

$1 - \alpha = P[|Z| < z(\frac{\alpha}{2})] =$

$= P[\bar{X} - \frac{\sigma}{\sqrt{n}} z(\frac{\alpha}{2}) < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}} z(\frac{\alpha}{2})]$



$I = (\bar{X} - \frac{\sigma}{\sqrt{n}} z(\frac{\alpha}{2}), \bar{X} + \frac{\sigma}{\sqrt{n}} z(\frac{\alpha}{2}))$ JE <u>INTERVAL SPOĽAHLIVOSTI</u> PRE μ S <u>KOEFICIENTOM SPOĽAHLIVOSTI</u> $1-\alpha$

$\frac{\sigma}{\sqrt{n}}$ JE <u>ŠTANDARDNÁ</u> (STREDNÁ) CHYBA PRIEMERU

$\bar{X}$ JE <u>BODOVÝ ODHAD</u> μ, I JE <u>INTERVALOVÝ ODHAD</u> μ

ODHAD PARAMETROV MET. MAXIMÁLNEJ VIEROHODNOSTI

NECH $X_1, \dots, X_n$ JE NÁHODNÝ VÝBER Z ROZDELENIA, KT. ZÁVISÍ NA $\vec{\Theta}$.

FUNKCIA $\vec{T}(X_1, \dots, X_n)$ JE <u>BODOVÝ ODHAD</u> Θ, $\vec{b} = E\vec{T} - \vec{\Theta}$ JE <u>VYCHÝLENIE</u>

<u>NESTRANNÝ ODHAD</u>: $\forall \vec{\Theta} : E\vec{T} = \vec{\Theta}$

<u>NAJLEPŠÍ ODHAD</u> $\vec{T}_0 \in \mathcal{J}$: $\forall \vec{T} \in \mathcal{J} : \operatorname{var} \vec{T}_0 \le \operatorname{var} \vec{T}$

<u>KONZISTENTNÝ ODHAD</u> $\vec{T} = \vec{T}_n$: $\forall \varepsilon > 0 : \lim_{n \to \infty} P[\|\vec{T}_n - \vec{\Theta}\| < \varepsilon] = 1$

$(\vec{T}_L, \vec{T}_U)$ JE <u>INTERVAL SPOĽAHLIVOSTI</u> PRE $\vec{\Theta}$ S KOEF. $1-\alpha$, AK $P[\vec{T}_L \le \Theta \le \vec{T}_U] = 1 - \alpha$ (TZV. <u>INTERVALOVÝ ODHAD</u>)

UVAŽUJME ROZDELENIE S HUSTOTOU $f(x, \vec{\Theta})$

$\Rightarrow$ ZDRUŽENÁ HUSTOTA $(X_1, \dots, X_n)^T$ JE POTOM $f(\vec{x}, \vec{\Theta}) = \prod_{i=1}^n f(x_i, \vec{\Theta})$ KVÔLI NEZÁVISLOSTI

VEKTOR $\vec{x}$ SÚ NAPR. NAMERANÉ HODNOTY $\to$ TZV. VIEROHODNOST. FCIA.

<u>LOGARITMICKÁ VIEROHODNOSTNÁ FUNKCIA</u>

$\ell(\vec{\Theta}) = \ln f(\vec{x}, \vec{\Theta}) = \sum_{i=1}^n \ln f(x_i, \vec{\Theta})$, POPR.

$\ell(\vec{\Theta}) = \sum_{i=1}^n \ln P[X_i = x_i, \vec{\Theta}]$ PRE DISKRÉTNE ROZDELENIA

ODHAD $\hat{\Theta}$ METÓDOU <u>MAXIMÁLNEJ VIEROHODNOSTI</u> JE $\hat{\Theta}$ TAKÉ, ŽE

$\forall \vec{\Theta} : \ell(\hat{\Theta}) \ge \ell(\Theta)$ .. MAXIMALIZUJEME PRAVDEP. NÁH. VÝBERU $\vec{x}$!!

TESTOVANIE HYPOTÉZ

NECH $\vec{X} = (X_1, \dots, X_n)^T$ JE NÁHODNÝ VEKTOR SO ZDRUŽ. DISTR. F. $F_{\vec{X}}$

HYPOTÉZA = TVRDENIE O ROZDELENÍ

NULOVÁ HYPOTÉZA H_0, ALTERNATÍVNA HYPOTÉZA H_A

ROZHODNUTIE : NULOVÚ HYPOTÉZU ZAMIETNEME / NEZAMIETNEME (NA ZÁKLADE REALIZÁCIE NÁHODNÉHO VEKTORU)

CHYBA PRVÉHO DRUHU : ZAMIETNEME PLATNÚ HYPOTÉZU H_0

CHYBA DRUHÉHO DRUHU: NEZAMIETNEME NEPLATNÚ HYPOTÉZU H_0

KRITICKÝ OBOR W = MNOŽINA VÝSLEDKOV POKUSU, KEDY H_0 ZAMIETNEME

TVAR - VOLÍME TAK, ABY $\vec{X}$ PADOL DO W ZA PLATNOSTI H_A ČASTO

VEĽKOSŤ - VOLÍME TAK, ABY SME PLATNÚ H_0 ZAMIETLI S PRAVD. $\leq \alpha$

HLADINA TESTU α = MAX. PRAVD. CHYBY 1. DRUHU (=HLADINA VÝZNAM.)

DOSIAHNUTÁ HLADINA TESTU (p-HODNOTA) = PRAVDEPODOBNOSŤ, ŽE DOSTANEME NAŠU REAL. $\vec{x}$ ALEBO HORŠIU, ZA PREDP. PLATNOSTI H_0

SILA TESTU = PRAVDEPODOBNOSŤ, S AKOU ZAMIETNEME H_0, KEĎ PLATÍ H_A

MAJME NÁHODNÝ VÝBER $X_1, \dots, X_n$ Z ROZDELENIA $N(\mu, \sigma^2)$ SO ZNÁMYM ROZPTYLOM σ^2. NA HLADINE VÝZNAMNOSTI TESTUJEME $H_0 : \mu = \mu_0$, KDE μ_0 JE DANÁ KONŠTANTA OPROTI

$$Z := \frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/n}}, \quad \text{KDE } \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

a) $H_A : \mu > \mu_0$ POUŽIJEME KRITICKÝ OBOR $Z \geq z(\alpha)$

b) $H_A : \mu \neq \mu_0$ POUŽIJEME KRITICKÝ OBOR $|Z| \geq z(\frac{\alpha}{2})$

a) INTERVAL SPOĽAHLIVOSTI SO SPOĽ. $1-\alpha$ JE PRE $Z \geq z(\alpha)$

$$\left(-\infty ; \ \bar{X} + \frac{\sigma}{\sqrt{n}} z(\alpha)\right)$$

b) INT. SPOĽ. SO SPOĽ. $1-\alpha$ JE PRE $|Z| \geq z(\frac{\alpha}{2})$

$$\left(\bar{X} - \frac{\sigma}{\sqrt{n}} z\left(\frac{\alpha}{2}\right), \bar{X} + \frac{\sigma}{\sqrt{n}} z\left(\frac{\alpha}{2}\right)\right)$$

03.02.2008 Su
PETER ČERNO

MAI059 PRAVDEPODOBNOSŤ A ŠTATISTIKA - SKÚŠKA

JEDNOVÝBEROVÝ t TEST

NECH MAJÚ NEZÁVISLÉ NÁHODNÉ VELIČINY $Y_1, \dots, Y_n$ NORMÁLNE ROZDELENIE $N(\beta, \sigma^2)$

$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$, $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$

ŠTATISTIKA $T = \frac{\bar{Y} - \beta_0}{S}\sqrt{n}$

MÁ ZA PLATNOSTI $H_0 : \beta = \beta_0$ ROZDELENIE $t(n-1)$

H_0 MOŽNO TESTOVAŤ OPROTI ALTERN. HYPOTÉZE $H_1 : \beta \neq \beta_0$

H_0 ZAMIETNEME NA HLADINE α, KEĎ $|T| \geq t_{n-1}(\alpha)$

INTERVAL SPOĽAHLIVOSTI PRE STR. HODNOTU β SO SPOĽ. $1-\alpha$ JE $\left(\bar{Y} - \frac{S}{\sqrt{n}} t_{n-1}(\alpha),\ \bar{Y} + \frac{S}{\sqrt{n}} t_{n-1}(\alpha)\right)$

PÁROVÝ t TEST

MÁME NEZÁVISLÉ DVOJICE $(W_{11}, W_{21}), \dots, (W_{1n}, W_{2n})$ NÁHODNÝCH VELIČÍN, KTORÝCH ROZDIELY $Y_i = W_{1i} - W_{2i}$ MAJÚ ROZDELENIE $N(\beta, \sigma^2)$.

TESTUJEME $H_0 : EW_{1i} = EW_{2i}$ TJ. $\beta = 0$

ŠTATISTIKA $T = \frac{\bar{W}_1 - \bar{W}_2}{S}\sqrt{n}$

LINEÁRNY MODEL

MÁME NÁHODNÝ VEKTOR $\vec{Y} = \vec{X}\vec{\beta} + \sigma\vec{Z}$, KDE
$\vec{X}$ JE MATICA ZNÁMYCH KONŠTÁNT $n \times k$ HODNOSTI k
$\vec{\beta}$ JE VEKTOR NEZNÁMYCH k PARAMETROV, $\sigma > 0$ NEZN. SMER. ODCH.
$\vec{Z} \sim N(\vec{0}, \vec{1})$

ODHAD VEKTORU $\vec{X}\vec{\beta}$ METÓDOU NAJMENŠÍCH ŠTVORCOV
JE $\hat{Y} = \vec{X}\vec{b}$: $\|\vec{Y} - \vec{X}\vec{\beta}\|$ JE MINIMÁLNE PRE $\vec{\beta} = \vec{b}$
(PREDP., ŽE POZNÁME REALIZÁCIU NÁH. VEKTORU $\vec{Y}$)

KEĎ STĹPCE $\vec{P}$ TVORIA ORTONORMÁLNU BÁZU $M(\vec{X})$,
POTOM $\hat{Y} = \vec{P}\vec{P}^T\vec{Y}$.
KEĎŽE $\vec{X}^T(\vec{Y} - \hat{Y}) = \vec{0}$, POTOM $\vec{X}\vec{b} = \hat{Y} \Rightarrow \vec{X}^T\vec{X}\vec{b} = \vec{X}^T\vec{Y}$
$\Rightarrow b = (\vec{X}^T\vec{X})^{-1}\vec{X}^T\vec{Y}$

<u>REZIDUÁLNY SÚČET ŠTVORCOV</u>: $RSS = \|\vec{Y} - \hat{Y}\|^2$
<u>REZIDUÁLNY ROZPTYL</u> : $S^2 = \|\vec{Y} - \hat{Y}\|^2/(n-k) = \frac{RSS}{n-k}$

NIEKOĽKO VÝBEROV

MÁME $k \geq 2$ NEZÁVISLÝCH VÝBEROV
$Y_{11}, \dots, Y_{1n_1} \sim N(\mu_1, \sigma^2)$
...
$Y_{k1}, \dots, Y_{kn_k} \sim N(\mu_k, \sigma^2)$

LINEÁRNY MODEL:

$$\begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ \vdots \\ Y_{k1} \\ \vdots \\ Y_{kn_k} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \vdots & & & \\ 1 & 0 & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_k \end{pmatrix} + \sigma\vec{Z}$$

$n = \sum_{i=1}^{k} n_i$

AKO ODHADY μ_i DOSTANEME
$\bar{Y}_{i\cdot} = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}$ (VIĎ $\vec{b} = \dots$)
$\Rightarrow \hat{Y} = \vec{X}\vec{b}$, T.J. $\hat{Y}_{ij} = \bar{Y}_{i\cdot}$
$\Rightarrow RSS = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{i\cdot})^2$
$\bar{Y}_{\cdot\cdot} = \frac{1}{n}\sum\sum Y_{ij}$
PRE TESTOVANIE HYPOTÉZY:
$RSS^0 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{\cdot\cdot})^2$

$$F = \frac{(RSS^0 - RSS)/(k-1)}{RSS/(n-k)} \sim F(k-1, n-k)$$

03.02.2008 Su.
PETER ČERNO

MAI059 PRAVDEPODOBNOSŤ A ŠTATISTIKA - SKÚŠKA

ANALÝZA ROZPTYLU

MODEL NIEKOĽKÝCH VÝBEROV MOŽNO ZAPÍSAŤ AKO:

$$Y_{ij} = \mu + \alpha_i + e_{ij} \quad 1 \leq i \leq k \quad 1 \leq j \leq n_i$$

KDE $e_{ij} \sim N(0, \sigma^2)$ SÚ NEZÁVISLÉ NÁH. VELIČINY

NA EFEKTY α_i KLADIEME PODMIENKU $\sum_{i=1}^{k} \alpha_i = 0$

PRETOŽE α_i SÚ PEVNÉ NENÁHODNÉ, SÚ TO PEVNÉ EFEKTY

NULOVÁ HYPOTÉZA $H_0 : \alpha_1 = \ldots = \alpha_k \; (=0)$

ZNAK, PODĽA KTORÉHO TRIEDIME VELIČINY Y DO JEDN. VÝBEROV SA NAZÝVA FAKTOR

REGRESNÁ PRIAMKA

ŠPECIÁLNY PRÍPAD LIN. MODELU

LINEÁRNA REGRESIA: NEZÁVISLÉ NÁH. VELIČINY $Y_1, \ldots, Y_n$ MAJÚ ROZDELENIE $N(\beta_0 + \beta_1 x_i, \sigma^2)$, KDE x_i SÚ DANÉ KONŠT.

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ \vdots & \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \sigma Z$$

$$\vec{b} = (\vec{X}^T \vec{X})^{-1} \vec{X}^T \vec{Y}$$

$$\vec{X}^T \vec{X} = \begin{pmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^{n} x_i^2 \end{pmatrix}, \quad \vec{X}^T \vec{Y} = \begin{pmatrix} n\bar{Y} \\ \sum_{i=1}^{n} Y_i \end{pmatrix}$$

$$\hat{Y} = \vec{X}\vec{b}, \quad RSS = \| \vec{Y} - \hat{Y} \|^2$$

08.04.2008 Tu
PETER ČERNO

NMAI059 PRAVD. A ŠTATISTIKA - SKÚŠKA

SENZITIVITA TESTU = PRAVDEPODOBNOSŤ, ŽE TEST JE POZITÍVNY U SKRYTO NEMOCNEJ OSOBY

SPECIFICITA TESTU = PRAVDEPODOBNOSŤ, ŽE TEST JE NEGATÍVNY U ZDRAVEJ OSOBY

INCIDENCIA NEMOCI = PERCENTO NAKAZENEJ POPULÁCIE

ROZDIEL MEDZI NEZLÚČITEĽNÝMI JAVMI A NEZÁVISLÝMI JAVMI:

a) NEZLUČITEĽNÉ: $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$

b) NEZÁVISLÉ JAVY: $P\left(\bigcup_{i=1}^{n} A_i\right) = 1 - \prod_{i=1}^{n} \left(1 - P(A_i)\right)$

$C'(M,m)$ = [NEKLESAJÚCE POSTUPNOSTI PRVKOV MNOŽINY $\{1,\dots,M\}$ DĹŽKY m]

a)

$$a_0 = \underbrace{1 \le a_1}_{d_1 = a_1 - 1} \le a_2 \le \dots \le \underbrace{a_m \le M = a_{m+1}}_{d_{m+1} = M - a_m} \qquad \sum_{i=1}^{m+1} d_i = M-1, \quad d_i \ge 0$$

$$d_i = a_i - a_{i-1}$$

SPOLU $\binom{m+M-1}{m}$ MOŽNOSTÍ

b) $c_1 + \dots + c_M = m, \quad 0 \le c_j \le m \qquad \binom{M-1+m}{M-1} = \binom{M-1+m}{m}$ MOŽNOSTÍ

$$\binom{r_n}{k} \frac{(n-1)^{r_n-k}}{n^{r_n}} = \frac{r_n(r_n-1)\dots(r_n-k+1)}{(n-1)^k} \frac{1}{k!} \left(1 - \frac{\frac{r_n}{n}}{n}\right)^{r_n}$$

$$\binom{r_n-k+n-2}{r_n-k} \Big/ \binom{r_n+n-1}{r_n} = \frac{\cancel{(r_n-k+n-2)!}\ \overset{n-1}{\cancel{(n-1)!}}\ \cancel{r_n!}}{\cancel{(r_n-k)!(n-2)!}\ \cancel{(r_n+n-1)!}} \frac{r_n \dots (r_n-k+1)}{(r_n+n-1)\dots(r_n+n-k-1)}$$

$$N^+_{x,y} = [(1,1) \to (x,y)] - [(1,-1) \to (x,y)] =$$

$$= N_{x-1,y-1} - N_{x-1,y+1} = \frac{y}{x} N_{x,y}$$

$$P_{2n} = P[n_0^* = 2n] = 2 \cdot \frac{1}{2} P[S_1 = 1, S_2 > 0, \dots, S_{2n-2} > 0, S_{2n-1} = 1) = \frac{1}{2n} P[S_{2n-2} = 0]$$

09.04.2008 We
PETER ČERNO

NMAI059 PRAVDEPODOBNOSŤ A ŠTATISTIKA - SKÚŠKA

$X_n \sim Bin(n, p_n)$ $\quad p_n \in (0,1)$, $\lim\limits_{n\to\infty} n p_n = \lambda \in (0,\infty)$, $X \sim Poiss(\lambda)$

$$P[X_n = k] = \binom{n}{k} p_n^k (1-p_n)^{n-k} =$$

$$= \frac{1}{k!} \frac{n p_n (n-1) p_n \dots (n-k+1) p_n}{(1-p_n)^k} \left(1 - \frac{n p_n}{n}\right)^n$$

$Z \sim N(0,1)$

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

NECH $\mu, \sigma \in \mathbb{R}$, $\sigma > 0$, $X = \mu + \sigma Z \sim (\mu, \sigma^2)$

$$F_X(x) = P[X < x] = P[\mu + \sigma Z < x] = \Phi\left(\frac{x-\mu}{\sigma}\right) =$$

$$= \int_{-\infty}^{(x-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(z-\mu)^2}{2\sigma^2}} dz$$

$$t = \frac{z-\mu}{\sigma} \quad dt = \frac{dz}{\sigma},$$

$$f_X(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(t-\mu)^2}{2\sigma^2}}$$

$Z \sim N(0,1)$ $\qquad X = Z^2 \sim \chi^2(1)$

$$F_X(x) = P[X < x] = P[Z^2 < x] = P[-\sqrt{x} < Z < \sqrt{x}] =$$

$$= \int_{-\sqrt{x}}^{\sqrt{x}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = 2 \int_{0}^{\sqrt{x}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \qquad t = \sqrt{u} \quad dt = \frac{1}{2} \frac{du}{\sqrt{u}}$$

$$= \int_{0}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{u}{2}} \cdot u^{-\frac{1}{2}} du$$

$$f_X(x) = \frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} e^{-\frac{x}{2}} \qquad x > 0, \qquad \text{INAK} = 0$$

MOIVREOVA - LAPLACEOVA LOKÁLNA VETA $\quad p \in (0,1),\ q = 1-p$

$$\max_{0 \le k \le n} \left| \binom{n}{k} p^k q^{n-k} - \frac{1}{\sqrt{npq}} \varphi\left(x_{n,k}\right) \right| = o\left(n^{-\frac{1}{2}}\right)$$

$$x_{n,k} = \frac{k - np}{\sqrt{npq}}$$

MOIVREOVA - LAPLACEOVA INTEGRÁLNA VETA

$p \in (0,1),\ q = 1-p,\ X_n \sim Bin(n,p) \quad -\infty < a < b < +\infty$

$$\lim_{n \to \infty} P\left[a < \frac{X_n - np}{\sqrt{npq}} < b \right] = \Phi(b) - \Phi(a)$$

$$P\left[\left| \frac{X_n}{n} - p \right| < \delta \right] = P\left[-\delta < \frac{X_n - np}{n} < +\delta \right] =$$

$$= P\left[-\delta \sqrt{\frac{n}{pq}} < \frac{X_n - np}{\sqrt{npq}} < +\delta \sqrt{\frac{n}{pq}} \right] =$$

$$= \underline{\underline{\Phi\left(\frac{n\delta}{\sqrt{npq}}\right) - \Phi\left(-\frac{n\delta}{\sqrt{npq}}\right)}} = \underline{\underline{2\,\Phi\left(\frac{n\delta}{\sqrt{npq}}\right) - 1}}$$

$$P\left[\left| \frac{X_n}{n} - p \right| < \delta \right] \ge 1 - \beta$$

$$\frac{n\delta}{\sqrt{npq}} \ge \Phi^{-1}\left(1 - \frac{\beta}{2}\right) = z\left(\frac{\beta}{2}\right). \quad pq \le \frac{1}{4} \Rightarrow \underline{\underline{n \ge}} \left(\frac{z(\beta/2)}{2\delta}\right)^2$$

12. 04. 2008 Sa
PETER ČERNO

NMAI059 PRAVDEPODOBNOSŤ A ŠTATISTIKA - SKÚŠKA

MAJME $Z \sim N(0,1)$ NORMÁLNE ROZDELENIE

$$\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \qquad \Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

POLOŽME $Y = \mu + \sigma Z$

$$F_Y(x) = P[\mu + \sigma Z < x] = P\left[Z < \frac{x-\mu}{\sigma}\right] =$$

$$= \int_{-\infty}^{(x-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \frac{dy}{\sigma} =$$

$$t = \frac{y-\mu}{\sigma} \quad dt = \frac{dy}{\sigma}$$

$$= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy \Rightarrow f_Y(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(t-\mu)^2}{2\sigma^2}}$$

ZREJME $M_Z(t) = Ee^{tZ} = \int_{-\infty}^{\infty} e^{tx} \varphi(x)\, dx =$

$$= \int_{-\infty}^{\infty} e^{tx} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2+2tx-t^2+t^2}{2}} dx =$$

$$= e^{\frac{t^2}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-t)^2}{2}} dx = e^{t^2/2}$$

PLATÍ: AK $Y = \mu + \sigma X$, POTOM $M_Y(t) = e^{\mu t} \cdot M_X(\sigma t)$

$$M_Y(t) = Ee^{tY} = Ee^{t(\mu+\sigma X)} = Ee^{t\mu} e^{t\sigma X} =$$

$$= e^{\mu t} \cdot Ee^{\sigma t X} = e^{\mu t} M_X(\sigma t)$$

ĎALEJ AK X, Y SÚ NEZÁVISLÉ NÁH. VELIČINY, POTOM

$$M_{X+Y}(t) = Ee^{t(X+Y)} = Ee^{tX} \cdot e^{tY} = M_X(t) \cdot M_Y(t)$$

ĎALEJ $\mu'_r = \frac{d^r}{dt^r} M_X(t) \Big|_{t=0}$

NECH TEDA $Z \sim N(0,1)$, $M_Z(t) = e^{\frac{t^2}{2}}$

$M_Z'(t) = t e^{t^2/2}$, $M_Z''(t) = e^{t^2/2} + t^2 e^{t^2/2}$

$\Rightarrow EZ = \mu_1' = M_Z'(0) = 0$, $EZ^2 = \mu_2' = M_Z''(0) = 1$

$\operatorname{var} Z = EZ^2 - (EZ)^2 = \mu_2' - (\mu_1')^2 = 1$

$Y = \mu + \sigma Z \Rightarrow EY = \mu$, $\operatorname{var} Y = \sigma^2$, PÍŠEME $Y \sim N(\mu, \sigma^2)$.

$$M_Y(t) = e^{\mu t} \cdot M_Z(\sigma t) = e^{\mu t} \cdot e^{\frac{\sigma^2 t^2}{2}} = \underline{\underline{e^{\mu t + \sigma^2 \cdot \frac{t^2}{2}}}}$$

MAJME $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$ NEZÁVISLÉ

POTOM $M_{X+Y}(t) = e^{\mu_1 t + \sigma_1^2 t^2/2} \cdot e^{\mu_2 t + \sigma_2^2 t^2/2} =$

$$= \underline{\underline{e^{(\mu_1+\mu_2)t + (\sigma_1^2+\sigma_2^2) t^2/2}}} \Rightarrow \underline{\underline{X+Y}} \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$$

NECH $Z \sim N(0,1)$. POLOŽME $X = Z^2$

POTOM $F_X(x) = P[X < x] = P[Z^2 < x] =$

$$= P[-\sqrt{x} < Z < \sqrt{x}] = \int_{-\sqrt{x}}^{\sqrt{x}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt = 2 \int_0^{\sqrt{x}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt =$$

$t = \sqrt{y}$ $dt = \frac{1}{2\sqrt{y}}\, dy$

$$= 2 \int_0^x \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{y}{2}} \frac{1}{2} y^{-\frac{1}{2}}\, dy = \underline{\underline{\int_0^x \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{y}{2}}\, dy}}$$

$$\underline{\underline{f_X(t) = \frac{1}{\sqrt{2\pi}} t^{-\frac{1}{2}} e^{-\frac{t}{2}}}}$$

X MÁ ROZDELENIE $\chi^2(1)$..

12.04.2008 So
PETER ČERNO

NMAI059 PRAVDEPODOBNOSŤ A ŠTATISTIKA - SKÚŠKA

GAMMA FUNKCIA $\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx \quad a>0$

PLATÍ $\Gamma(1) = 1$, $\Gamma(a+1) = a \cdot \Gamma(a)$

$\Gamma(\frac{1}{2}) = \sqrt{\pi}$ $\quad \Gamma(n+1) = n! \quad n \in \mathbb{N}$

UVAŽUJME HUSTOTU $f(x) = c x^{a-1} e^{-bx} \quad a,b,c>0, \quad x>0$

MUSÍ PLATIŤ $\int_0^\infty c x^{a-1} e^{-bx} dx = 1$, TJ.

$$\int_0^\infty c x^{a-1} e^{-bx} dx \underset{x=\frac{t}{b} \quad dx=\frac{dt}{b}}{=} \int_0^\infty c \left(\frac{t}{b}\right)^{a-1} \cdot e^{-t} \frac{dt}{b} =$$

$$= \int_0^\infty c \cdot \frac{1}{b^a} \cdot t^{a-1} e^{-t} dt = \frac{c}{b^a} \cdot \Gamma(a) \Rightarrow \quad c = \frac{b^a}{\Gamma(a)}$$

GAMMA ROZDELENIE $\Gamma(a,b)$ $\quad a,b>0$ S HUSTOTOU

$$f(x) = \frac{b^a}{\Gamma(a)} \cdot x^{a-1} e^{-bx}$$

NECH $X \sim \Gamma(a,b)$. POTOM PRE $0<t<b$:

$$M_X(t) = E e^{tX} = \int_0^\infty e^{tx} \frac{b^a}{\Gamma(a)} \cdot x^{a-1} e^{-bx} dx =$$

$$= \frac{b^a}{(b-t)^a} \int_0^\infty \frac{(b-t)^a}{\Gamma(a)} x^{a-1} \cdot e^{-(b-t)x} dx = \left(\frac{b}{b-t}\right)^a .$$

AK $X \sim \Gamma(a_1, b)$ $\quad Y \sim \Gamma(a_2, b)$ SÚ NEZÁVISLÉ, POTOM

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t) = \left(\frac{b}{b-t}\right)^{a_1+a_2} \Rightarrow X+Y \sim \Gamma(a_1+a_2, b).$$

VŠIMNIME SI, ŽE $\chi^2(1)$ JE VLASTNE $\Gamma(\frac{1}{2}, \frac{1}{2})$.

NECH $Z_1, \dots, Z_n \sim N(0,1)$ NEZÁVISLÉ, $X = Z_1^2 + \dots + Z_n^2$.

POTOM ZREJME $X \sim \Gamma(\frac{n}{2}, \frac{1}{2})$ JE ROZDELENIE $\chi^2(n)$,

$$f_X(t) = \frac{1}{2^{n/2} \Gamma(\frac{n}{2})} x^{n/2-1} e^{-\frac{1}{2}x} \qquad M_X(t) = \left(\frac{1/2}{1/2-t}\right)^{n/2} .$$

NECH X, Y SÚ NEZÁVISLÉ NÁHODNÉ VELIČINY S HUSTOTAMI f_X, f_Y, KDE $f_Y(y) = 0$ PRE $y \leq 0$, TJ. $P[Y>0]=1$. NECH c JE KONŠTANTA.

POLOŽME $U = \frac{cX}{Y}$.

$$F_U(u) = P[U<u] = P\left[X < \frac{u}{c}\cdot Y\right] = \int\limits_{\substack{x,y\in\mathbb{R} \\ y>0,\, x<\frac{u}{c}y}} \overbrace{f_{X,Y}(x,y)}^{f_X(x)\,.\,f_Y(y) \text{ Z NEZÁVISLOSTI}} \,dx\,dy =$$

$$\int_0^\infty \int_{-\infty}^{(u/c)y} f_X(x)\, f_Y(y)\, dx\, dy =$$

$$x = \left(\frac{y}{c}\right)\cdot t, \quad dx = \frac{y}{c}\,dt$$

$$= \int_0^\infty \int_{-\infty}^{u} f_X\left(\frac{y}{c}t\right) . f_Y(y) . \frac{y}{c}\, dt\, dy = \int_{-\infty}^{u} \left(\frac{1}{c}\int_0^\infty y\, f_X\left(\frac{yt}{c}\right) . f_Y(y)\, dy\right) dt$$

$$\Rightarrow \quad \underline{\underline{f_U(t) = \frac{1}{c}\int_0^\infty y\, f_X\left(\frac{yt}{c}\right) . f_Y(y)\, dy}}$$

F - ROZDELENIE (FISHEROVO - SNEDECOROVO ROZD.) $F(k,m)$

NECH X, Y SÚ DVE NEZÁVISLÉ NÁHODNÉ VELIČINY S ROZDELENÍM $\chi^2(k)$, $\chi^2(m)$.

POTOM $U = \frac{X/k}{Y/m}$ MÁ ROZDELENIE $F(k,m)$

$$f_U(t) = \frac{1}{(m/k)} \cdot \int_0^\infty y \cdot \frac{1}{2^{k/2}\,\Gamma\left(\frac{k}{2}\right)} \left(\frac{ykt}{m}\right)^{k/2-1} e^{-\frac{ykt}{2m}} \cdot \frac{1}{2^{m/2}\,\Gamma\left(\frac{m}{2}\right)} \cdot y^{m/2-1} e^{-\frac{y}{2}}\, dy =$$

$$= \left(\frac{k}{m}\right)^{k/2} \frac{1}{2^{(k+m)/2}} \frac{1}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{m}{2}\right)} \cdot \int_0^\infty y^{\frac{k+m}{2}-1} . t^{\frac{k}{2}-1} e^{-\frac{y(kt+m)}{2m}}\, dy =$$

$$= \left(\frac{k}{m}\right)^{k/2} t^{k/2-1} \frac{\Gamma\left(\frac{k}{2}+\frac{m}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{m}{2}\right)} \cancel{\frac{1}{2^{(k+m)/2}}} \left(\frac{\cancel{2}m}{kt+m}\right)^{\frac{k+m}{2}} =$$

$$= \underline{\underline{\frac{\Gamma\left(\frac{k+m}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{m}{2}\right)} \cdot \left(\frac{k}{m}\right)^{k/2} . t^{k/2-1} . \left(1+\frac{k}{m}t\right)^{-\frac{k+m}{2}}}}$$

12.04.2008 Sa
PETER ČERNO

NMAI059 PRAVDEPODOBNOSŤ A ŠTATISTIKA - SKÚŠKA

NECH $X \sim \chi^2(n)$. POTOM $Y = \sqrt{X}$ MÁ ROZDELENIE $\chi(n)$.

$$F_Y(y) = P[Y < y] = P[\sqrt{X} < y] = P[X < y^2] =$$

$$= \int_0^{y^2} \frac{1}{2^{n/2}\,\Gamma(\frac{n}{2})} t^{\frac{n}{2}-1} e^{-\frac{t}{2}}\, dt = \qquad t = u^2 \quad dt = 2u\,du$$

$$= \int_0^{y} \frac{1}{2^{n/2}\,\Gamma(\frac{n}{2})} \cdot 2 \cdot u^{n-1} \cdot e^{-\frac{u^2}{2}}\, du$$

$$\Rightarrow f_Y(t) = \frac{1}{2^{\frac{n}{2}-1}\,\Gamma(\frac{n}{2})} \cdot t^{n-1} e^{-\frac{t^2}{2}}$$

STUDENTOVO t-ROZDELENIE $T(n)$

NECH $Z \sim N(0,1)$, $X \sim \chi^2(n)$, POTOM

$T = \frac{Z}{\sqrt{X/n}}$ MÁ STUDENTOVO ROZDELENIE S n ST. VOĽNOSTI

POLOŽME $Y = \sqrt{X}$, $Y \sim \chi(n)$, $c = \sqrt{n}$.

POTOM

$$f_T(t) = \frac{1}{\sqrt{n}} \int_0^\infty y \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{\left(\frac{yt}{\sqrt{n}}\right)^2}{2}} \cdot \frac{1}{2^{\frac{n}{2}-1}\,\Gamma(\frac{n}{2})} y^{n-1} e^{-\frac{y^2}{2}}\, dy =$$

$$= \frac{1}{\sqrt{2\pi n}} \cdot \frac{1}{2^{\frac{n}{2}-1}\,\Gamma(\frac{n}{2})} \int_0^\infty y^n \cdot e^{-\frac{(yt)^2}{2n} - \frac{y^2}{2}}\, dy =$$

$$= \frac{1}{\sqrt{2\pi n}} \cdot \frac{1}{2^{n/2-1}\,\Gamma(\frac{n}{2})} \int_0^\infty y^n \cdot e^{-y^2\left(\frac{t^2}{2n} + \frac{1}{2}\right)}\, dy =$$

$$Y = \sqrt{X}, \quad dy = \frac{dx}{2\sqrt{x}}$$

$$= \frac{1}{\sqrt{\pi n}} \frac{1}{2^{\frac{n-1}{2}}\,\Gamma(\frac{n}{2})} \int_0^\infty x^{\frac{n+1}{2}-1} \cdot e^{-x\left(\frac{1+t^2/n}{2}\right)} \frac{1}{2}\, dx = \frac{\Gamma(\frac{n+1}{2}) \cdot \cancel{2^{\frac{n+1}{2}}}}{\sqrt{\pi n}\, \cancel{2^{\frac{n-1}{2}}} \cdot \Gamma(\frac{n}{2})} \cdot \frac{1}{\cancel{2}} \cdot \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$$

NECH $X \sim \chi^2(n)$, TJ. $f_X(x) = \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} x^{n/2-1} e^{-x/2}$ $x > 0$

$$EX = \int_0^\infty x f_X(x)\,dx = \int_0^\infty \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} x^{\frac{n+2}{2}-1} e^{-\frac{x}{2}}\,dx =$$

$$= \frac{2^{\frac{n}{2}+1}\Gamma(\frac{n}{2}+1)}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} \int_0^\infty \frac{1}{2^{\frac{n+2}{2}}\Gamma(\frac{n+2}{2})} x^{\frac{n+2}{2}-1} e^{-\frac{x}{2}}\,dx =$$

$$= 2\cdot\left(\frac{n}{2}\right) = \underline{\underline{n}}$$

$$EX^2 = \int_0^\infty x^2 f_X(x)\,dx = \int_0^\infty \frac{1}{2^{n/2}\Gamma(\frac{n}{2})} \cdot x^{\frac{n+4}{2}-1} e^{-\frac{x}{2}}\,dx =$$

$$= \frac{2^{\frac{n}{2}+2}\cdot\Gamma(\frac{n}{2}+2)}{2^{\frac{n}{2}}\cdot\Gamma(\frac{n}{2})} \int_0^\infty \frac{1}{2^{\frac{n+4}{2}}\Gamma(\frac{n+4}{2})} \cdot x^{\frac{n+4}{2}-1} e^{-\frac{x}{2}}\,dx =$$

$$= 4\cdot\left(\frac{n}{2}+1\right)\left(\frac{n}{2}\right) = \underline{\underline{n^2+2n}}$$

$$\Rightarrow \operatorname{var} X = EX^2 - (EX)^2 = n^2+2n-n^2 = \underline{\underline{2n}}$$

AK $X_1, \dots, X_n$ JE NÁHODNÝ VÝBER Z $N(\mu, \sigma^2)$, POTOM PRE $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ PLATÍ

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

$$\Rightarrow \operatorname{var}\left(\frac{n-1}{\sigma^2} S^2\right) = 2(n-1) \Rightarrow \operatorname{var} S^2 = \frac{2\sigma^4}{n-1}$$

14.04.2008 Mo
PETER ČERNO

NMAI059 PRAVDEPODOBNOSŤ A ŠTATISTIKA - SKÚŠKA

VÝBER BEZ VRACANIA Z KONEČNEJ POPULÁCIE

POPULÁCIA (ZÁKLADNÝ SÚBOR) $x_1, \dots, x_N$

$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$

VÝBER (VÝBEROVÝ SÚBOR) VEĽKOSTI $n < N$

$\Omega = \binom{\{1, \dots, N\}}{n}$

ZAVEDIEME $\vec{W} = (W_1, \dots, W_N)^T \quad W_i = \begin{cases} 1 & i \in s \\ 0 & i \notin s \end{cases} \qquad s \in \Omega$

VÝBEROVÝ PRIEMER $\bar{X} = \frac{1}{n}\sum_{i \in s} x_i = \frac{1}{n}\sum_{i=1}^{N} W_i$

PLATÍ $E\bar{X} = \mu, \quad \operatorname{var}\bar{X} = \frac{N-n}{N-1}\frac{\sigma^2}{n}$ $\qquad \operatorname{var}(\bar{X}) = E(\bar{X} - E\bar{X})^2 = E\left(\frac{1}{n}\sum_{i=1}^{N}(x_i - \mu)W_i\right)^2$

VÝBEROVÝ ROZPTYL $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

PLATÍ $ES^2 = \frac{N}{N-1}\sigma^2$

IDENTITA $\boxed{\sum_{i=1}^{n}(X_i - \mu)^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2 + n(\bar{X} - \mu)^2}$

NÁHODNÝ VÝBER

NÁHODNÝ VÝBER V ROZSAHU n JE n-TICA NEZÁVISLÝCH NÁH. VELIČÍN $X_1, \dots, X_n$ S ROVNAKÝM ROZDELENÍM DANÝM F_X

$EX = \mu, \quad \operatorname{var} X = \sigma^2$

VÝBEROVÝ PRIEMER $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

ZREJME $E\bar{X} = \mu, \quad \operatorname{var}\bar{X} = \frac{\sigma^2}{n}$

VÝBEROVÝ ROZPTYL $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

PLATÍ $ES^2 = \sigma^2$

NÁHODNÝ VÝBER Z NORMÁLNEHO ROZDELENIA

$X_1, \dots, X_n$ NEZÁV. NÁH. VELIČINY Z $N(\mu, \sigma^2)$

POTOM $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

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

$\bar{X}, S^2$ SÚ NEZÁVISLÉ NÁH. VELIČINY

ZREJME $\frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}} \sim N(0,1)$

$$\Rightarrow \frac{\frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}/(n-1)}} = \boxed{\frac{\bar{X}-\mu}{S}\sqrt{n} \sim t(n-1)}$$

AK $X \sim \chi^2(n) \Rightarrow \operatorname{var} X = 2n$

$$\Rightarrow \operatorname{var} S^2 = \frac{\sigma^4}{(n-1)^2}\cdot 2(n-1) = \underline{\underline{\frac{2\sigma^4}{(n-1)}}}$$

NECH $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ NEZÁVISLÉ, σ^2 POZNÁME

POTOM $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Rightarrow \frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}} \sim N(0,1)$

$$\Rightarrow 1-\alpha = P\left[\left|\frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}}\right| < z(\alpha/2)\right] =$$

$$= P\left[\underbrace{\bar{X} - \frac{\sigma}{\sqrt{n}}z\left(\frac{\alpha}{2}\right) < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}}z\left(\frac{\alpha}{2}\right)}\right]$$

INTERVAL SPOĽAHLIVOSTI PRE μ S KOEF. SPOĽ. $1-\alpha$

AK NEPOZNÁME ANI σ^2:

$$\frac{\bar{X}-\mu}{S}\sqrt{n} \sim t(n-1)$$

$$1-\alpha = P\left[\left|\frac{\bar{X}-\mu}{S}\sqrt{n}\right| < t_{n-1}(\alpha)\right] = \underline{\underline{P\left[\bar{X} - \frac{S}{\sqrt{n}}t_{n-1}(\alpha) < \mu < \bar{X} + \frac{S}{\sqrt{n}}t_{n-1}(\alpha)\right]}}$$

14.04.2008 Mo
PETER ČERNO

NMAI059 PRAVD. A ŠTATISTIKA - SKÚŠKA

ODHAD PARAMETROV METÓDOU MAX. VIEROHODNOSTI

NÁHODNÝ VÝBER $X_1, \dots, X_n \sim F_{\vec{\theta}}$

BODOVÝ ODHAD: $\vec{T}(X_1, \dots, X_n)$

NESTRANNÝ $\Leftrightarrow$ $E\vec{T} = \vec{\Theta}$ $\forall \vec{\Theta}$

NAJLEPŠÍ, AK MINIMALIZUJE $\operatorname{var} \vec{T}$

KONZISTENTNÝ $\vec{T}_n$ AK $\lim_{n\to\infty} P[\|T_n - \vec{\Theta}\| < \varepsilon] = 1$

TJ. $\vec{T}_n$ KONVERGUJE K Θ PODĽA PRAVDEPODOBNOSTI

AK $\lim_{n\to\infty} ET_n = \Theta$, $\lim_{n\to\infty} \operatorname{var} T_n = 0$ $\Rightarrow$ $\lim_{n\to\infty} P[|T_n - \Theta| < \varepsilon] = 1$

INTERVALOVÝ ODHAD: $\vec{T}_L, \vec{T}_U$

$P[\vec{T}_L < \vec{\Theta} < \vec{T}_U] = 1 - \alpha$

NECH ROZDELENIE JE CHARAKT. HUSTOTOU $f(x, \vec{\Theta})$

$\Rightarrow$ ZDRUŽENÁ HUSTOTA $f(x_1, \dots, x_n, \vec{\theta}) = \prod_{i=1}^n f(x_i; \vec{\Theta})$

$v(\vec{\theta}) = f(\vec{x}, \vec{\theta})$ JE VIEROHODNOSTNÁ FUNKCIA

$\ell(\vec{\theta}) = \ln f(\vec{x}, \vec{\theta})$ JE LOG. VIEROH. FUNKCIA

HĽADÁME $\vec{\Theta}$, KT. MAXIMALIZUJE $v(\vec{\Theta})$

TESTOVANIE HYPOTÉZ

KRITICKÝ OBOR W = MNOŽINA VÝSLEDKOV, PRI KTORÝCH BUDEME H_0 ZAMIETAŤ

VEĽKOSŤ VOLÍME TAK, ABY SME PLATNÚ HYP. ZAMIETALI S PRAVD. α = HLADINA TESTU

P-HODNOTA = PRAVDEPODOBNOSŤ ZA PLATNOSTI H_0, ŽE DOSTANEME NÁŠ VEKTOR X ALEBO EŠTE HORŠÍ
= NAJMENŠIA HLADINA, NA KTOREJ ZAMIETAME

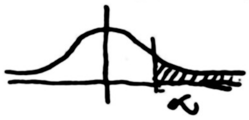
1) MAJME $X_1, \dots, X_n$ Z NORMÁLNEHO ROZDELENIA SO ZNÁMYM ROZPTYLOM σ^2.

$H_0: \mu = \mu_0$
$H_1: \mu > \mu_0$

ZA PLATNOSTI H_0 JE

$$Z = \frac{\bar{X} - \mu_0}{\sigma}\sqrt{n} \sim N(0,1)$$

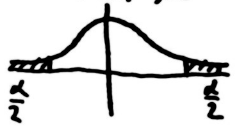
HYPOTÉZU H_0 ZAMIETNEME NA HLADINE α AK $Z \geq z(\alpha)$



$H_0: \mu = \mu_0$
$H_1: \mu \neq \mu_0$

HYPOTÉZU H_0 ZAMIETNEME AK:

$$|Z| \geq z\left(\frac{\alpha}{2}\right)$$



2) MAJME $Y_1, \dots, Y_n$ Z NORMÁLNEHO ROZDELENIA μ ANI σ^2 NEPOZNÁME

$H_0: \mu = \mu_0$
$H_1: \mu \neq \mu_0$

ZA PLATNOSTI H_0 MÁ ŠTATISTIKA

$$T = \frac{\bar{Y} - \mu_0}{S}\sqrt{n}$$

ROZDELENIE $t(n-1)$

HYPOTÉZU H_0 ZAMIETNEME NA HLADINE α AK

$$|T| \geq t_{n-1}(\alpha)$$

14. 04. 2008 Mo
PETER ČERNO

NMAI 059 PRAVDEPODOBNOSŤ A ŠTATISTIKA - SKÚŠKA

LINEÁRNY MODEL

METÓDA NAJMENŠÍCH ŠTVORCOV

MAJME NÁHODNÝ VEKTOR:

$\vec{Y} = \vec{X}\vec{\beta} + \sigma\vec{Z}$ $\quad \vec{X} \in \mathbb{R}^{n \times k}$ $k < n$ $\text{rank}(\vec{X}) = k$, $\vec{Z} \sim N(\vec{0}, \vec{1})$

$\vec{\beta} \in \mathbb{R}^k$, $\sigma \in \mathbb{R}$ SÚ NEZNÁME

$\equiv$ <u>LINEÁRNY MODEL</u> (S ÚPLNOU HODNOSŤOU)

ODHAD VEKTORU $\vec{X}\vec{\beta}$ <u>METÓDOU NAJMENŠÍCH ŠTVORCOV</u>
JE $\hat{Y} = \vec{X}\vec{b} \in \mathcal{L}(\vec{X})$: $\|\vec{Y} - \hat{Y}\|$ JE MINIMÁLNE

ZREJME $\hat{Y} = \vec{P}\vec{P}^T\vec{Y}$

PLATÍ: $\boxed{\dfrac{\|\vec{Y} - \hat{Y}\|^2}{\sigma^2} \sim \chi^2(n-k)}$

ZREJME $\vec{X}^T(\vec{Y} - \hat{Y}) = \vec{0}$

$\vec{X}\vec{b} = \hat{Y} \Rightarrow \vec{X}^T\vec{X}\vec{b} = X^T\hat{Y} = X^TY \Rightarrow \boxed{\vec{b} = (\vec{X}^T\vec{X})^{-1}\vec{X}^T\vec{Y}}$

HODNOTU $\|\vec{Y} - \hat{Y}\|^2$ NAZÝVAME <u>REZIDUÁLNY SÚČET ŠTVORCOV</u>
A ZNAČÍME RSS.

AKO ODHAD ROZPTYLU σ^2 POUŽÍVAME <u>REZIDUÁLNY ROZPTYL</u>

$\boxed{S^2 = \dfrac{\|\vec{Y} - \hat{Y}\|^2}{n-k} = \dfrac{RSS}{n-k}}$

PLATÍ $\boxed{\vec{b} \sim N(\vec{\beta}, \sigma^2(\vec{X}^T\vec{X})^{-1})}$

$\boxed{\dfrac{(n-k)S^2}{\sigma^2} \sim \chi^2(n-k)}$

$\boxed{ES^2 = \sigma^2}$

OZNAČME $\vec{C} = (\vec{X}^T\vec{X})^{-1}$

POTOM $\boxed{b_j \sim N(\beta_j, \sigma^2 c_{jj})}$

$\boxed{T_j = \dfrac{b_j - \beta_j}{S\sqrt{c_{jj}}} \sim t(n-k)}$

JEDEN VÝBER

$$\boxed{\begin{pmatrix}Y_1\\ \vdots \\ Y_n\end{pmatrix} = \begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}\beta + \sigma Z} \qquad k=1 \qquad C=(X^TX)^{-1}=\frac{1}{n}$$

$$b=(X^TX)^{-1}X^TY=\frac{1}{n}\sum_{i=1}^n Y_i=\bar{Y} \quad P=\begin{pmatrix}1/\sqrt{n}\\ \vdots \\ 1/\sqrt{n}\end{pmatrix} \quad PP^TY=\begin{pmatrix}\frac{1}{n}\sum_{i=1}^n Y_i\\ \vdots \\ \frac{1}{n}\sum_{i=1}^n Y_i\end{pmatrix}$$

$$S^2=\frac{RSS}{n-1}=\frac{\|\vec{Y}-\hat{Y}\|^2}{n-1}=\frac{1}{n-1}\sum_{i=1}^n (Y_i-\bar{Y})^2$$

$$b\sim N\left(\beta,\frac{\sigma^2}{n}\right), \quad \frac{(n-1)S^2}{\sigma^2}\sim \chi(n-1)$$

$$T=\frac{b-\beta}{S\sqrt{\frac{1}{n}}}\sim t(n-1)$$

JEDNOVÝBEROVÝ t-TEST: $H_0: \beta=\beta_0$, $H_1: \beta\neq\beta_0$ ZAMIETAME PRI $|T|\geq t_{n-1}(\alpha)$

PODOBNE PÁROVÝ t-TEST: $Y_i = W_{i1}-W_{i2}$ $\quad T=\frac{\bar{W}_1-\bar{W}_2}{S}\sqrt{n}$ $\quad \beta_0=0$

14.04.2008 Mo
PETER ČERNO

NMAI059 PRAVDEPODOBNOSŤ A ŠTATISTIKA - SKÚŠKA

DVA VÝBERY

$Y_{1i} \sim N(\mu_1, \sigma^2)$, $Y_{2i} \sim N(\mu_2, \sigma^2)$

$\beta_1 = \mu_1$, $\beta_2 = \mu_2 - \mu_1$

$$\begin{pmatrix} Y_{11} \\ \vdots \\ Y_{1n_1} \\ Y_{21} \\ \vdots \\ Y_{2n_2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \sigma Z \qquad k=2$$

$$X^TX = \begin{pmatrix} n_1+n_2 & n_2 \\ n_2 & n_2 \end{pmatrix} \qquad C = (X^TX)^{-1} = \frac{1}{n_1 n_2} \begin{pmatrix} n_2 & -n_2 \\ -n_2 & n_1+n_2 \end{pmatrix}$$

$$P = \begin{pmatrix} 1/\sqrt{n_1} & 0 \\ \vdots & \vdots \\ 1/\sqrt{n_1} & 0 \\ 0 & 1/\sqrt{n_2} \\ \vdots & \vdots \\ 0 & 1/\sqrt{n_2} \end{pmatrix} \qquad PP^T = \begin{pmatrix} 1/n_1 & \cdots & 1/n_1 & 0 & \cdots & 0 \\ \vdots & & & \vdots & & \vdots \\ 1/n_1 & \cdots & 1/n_1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1/n_2 & \cdots & 1/n_2 \\ \vdots & & \vdots & \vdots & & \\ 0 & \cdots & 0 & 1/n_2 & \cdots & 1/n_2 \end{pmatrix}$$

$$b = (X^TX)^{-1}X^TY = \frac{1}{n_1 n_2} \begin{pmatrix} n_2 & -n_2 \\ -n_2 & n_1+n_2 \end{pmatrix} \begin{pmatrix} n_1\bar{Y}_1 + n_2\bar{Y}_2 \\ n_2\bar{Y}_2 \end{pmatrix} = \begin{pmatrix} \bar{Y}_1 \\ \bar{Y}_2 - \bar{Y}_1 \end{pmatrix}$$

$$\hat{Y} = PP^TY = (\underbrace{\bar{Y}_1, \dots, \bar{Y}_1}_{n_1}, \underbrace{\bar{Y}_2, \dots, \bar{Y}_2}_{n_2})^T$$

$$\|Y - \hat{Y}\|^2 = \sum_{i=1}^{n_1} (Y_{1i} - \bar{Y}_1)^2 + \sum_{i=1}^{n_2} (Y_{2i} - \bar{Y}_2)^2 = RSS$$

$$S^2 = \frac{RSS}{n_1+n_2-2}, \qquad c_{11} = \frac{1}{n_1}, \qquad c_{22} = \frac{1}{n_1} + \frac{1}{n_2}$$

$$b_1 \sim N\left(\beta_1, \frac{\sigma^2}{n_1}\right)$$
$$b_2 \sim N\left(\beta_2, \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right)$$

$$T_2 = \frac{b_2 - \beta_2}{S\sqrt{c_{22}}} = \frac{(\bar{Y}_2 - \bar{Y}_1) - \beta_2}{S} \sqrt{\frac{n_1 n_2}{n_1+n_2}} \sim t(n_1+n_2-2)$$

REGRESNÁ PRIAMKA

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ \vdots & \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \sigma Z \qquad k=2 \qquad \text{rank}(X)=2$$

$$X^T X = \begin{pmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^n x_i^2 \end{pmatrix} \cdot \begin{pmatrix} \sum x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{pmatrix} = \begin{pmatrix} \overbrace{n\sum x_i^2 - n^2\bar{x}^2}^{n\sum x_i^2 - (\sum x_i)^2 = n\sum(x_i-\bar{x})^2} & 0 \\ 0 & n\sum x_i^2 - n^2\bar{x}^2 \end{pmatrix}$$

$$\Rightarrow \quad C = (X^T X)^{-1} = \frac{1}{\sum(x_i-\bar{x})^2} \cdot \begin{pmatrix} \frac{1}{n}\sum x_i^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}$$

$$b = (X^T X)^{-1} X^T Y = \frac{1}{\sum(x_i-\bar{x})^2} \cdot \begin{pmatrix} \frac{1}{n}\sum x_i^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix} \cdot \begin{pmatrix} n\bar{Y} \\ \sum x_i Y_i \end{pmatrix}$$

$$\hat{Y} = Xb$$

$$RSS = \|Y - \hat{Y}\|^2, \qquad S^2 = \frac{RSS}{n-2}$$

$$c_{11} = \frac{1}{\sum(x_i-\bar{x})^2} \frac{1}{n} \sum x_i^2, \qquad c_{22} = \frac{1}{\sum(x_i-\bar{x})^2}$$

14. 04. 2008 Mo
PETER ČERNO

NMAI059 PRAVDEPODOBNOSŤ A ŠTATISTIKA – SKÚŠKA

MULTINOMICKÉ ROZDELENIE

n NEZÁVISLÝCH POKUSOV, V KAŽDOM m MOŽNÝCH VÝSLEDKOV $A_1, \dots, A_m$ S PRAVDEP. $p_1, \dots, p_m$

ZAVEĎME NÁH. VEKTORY $\vec{Y}_1, \dots, \vec{Y}_n$ (NEZÁVISLÉ)

$$\vec{Y}_i = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{im} \end{pmatrix} \qquad Y_{ij} = 1 \Leftrightarrow \text{V } i\text{-TOM POKUSE NASTAL JAV } A_j$$

$$P[Y_{ij} = 1] = p_j, \quad P[Y_{ij} = 0] = 1 - p_j$$

$$E Y_{ij} = E Y_{ij}^2 = p_j$$

$$E Y_{ij} Y_{iq} = 0 \qquad j \neq q$$

$$\Rightarrow \operatorname{cov}(Y_{ij}, Y_{iq}) = E Y_{ij} Y_{iq} - E Y_{ij} E Y_{iq} = \begin{cases} p_j - p_j p_q & j = q \\ 0 - p_j p_q & j \neq q \end{cases} = p_j(\delta_{jq} - p_q)$$

ZAVEĎME $X_j = \sum_{i=1}^n Y_{ij}$. ZREJME $EX_j = np_j$

$\vec{X} = (X_1, \dots, X_m)^T$ MÁ MULTINOMICKÉ ROZDELENIE S PARAMETRAMI $n, p_1, \dots, p_m$

A PLATÍ $$P[X_1 = n_1, \dots, X_m = n_m] = \frac{n!}{n_1! \dots n_m!} p_1^{n_1} \dots p_m^{n_m}$$

PLATÍ: $$\boxed{\sum_{j=1}^m \frac{(X_i - np_j)^2}{np_j} \sim \chi^2(m-1)}$$

χ^2 TEST DOBREJ ZHODY

H_0: $P(A_j) = p_j^0 \quad \forall j$

$$X^2 = \sum_{j=1}^m \frac{(X_j - np_j^0)^2}{np_j^0}$$

H_0 ZAMIETNEME AK $X^2 \geq \chi^2_{m-1}(\alpha)$.