analogie s obecným členem binomické věty:

$$(a+b)^n = \sum_{i=0}^{n} \binom{n}{i} a^i b^{n-i}$$

Platí: $$\sum_{k=0}^{n} P_n(k) = 1$$

$$P_n(k \le m) = \sum_{k=0}^{m} P_n(k) \;;\; \text{pro } m \le n$$

vyjadřuje pravděpodobnost, že jev A se v pozorování vyskytne nejvýše m-krát.

resp. $$P(X=k) = \binom{n}{k}[P(A)]^k[1-P(A)]^{n-k}$$

$$\bar{X} = n\cdot P(A) \qquad \bar{X} = \bar{k}$$

$$\sigma^2 = \bar{X}[1-P(A)]$$

Z definice: $$\bar{X} = \sum_{k=0}^{n} k\cdot\binom{n}{k}[P(A)]^k[1-P(A)]^{n-k} \;; \ldots$$

Pomocí charakteristické funkce: $$P(X=k) = \binom{n}{k} p^k q^{n-k}$$

$$\Theta(v) = \sum_{k=0}^{n} P(X=k)\, e^{jvk} = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} e^{jvk}$$

Využijeme-li: $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k$, $\Theta(v) = (p\cdot e^{jv} + q)^n$

$$\Theta'(v=0) = j\cdot n\cdot p$$

$$\Theta''(v=0) = (-np^2 + n^2p^2 + np)\, j^2 = np^2 - n^2p^2 - np$$

$$\bar{X} = \frac{1}{j}\Theta'(v=0) = n\cdot p$$

$$\overline{X^2} = \frac{1}{j^2}\Theta''(v=0) = np + n^2p^2 - np^2 \Rightarrow \sigma^2 = \overline{X^2} - \bar{X}^2 = np - np^2 = n\cdot p(1-p) = n\cdot p\cdot q$$

PŘÍKLAD:

Průměrný počet volání účastníků na telef. ústřednu je 120 / hod.

a) ? je rozložení pravděpodobnosti počtu volání k,

b) ? je pravděpodobnost, že se za 1 min nevyskytne žádné volání

c) ? za 1 minutu se vyskytne právě 1 volání.

Počátky volání tvoří posloupnost okamžiků: $t_1, t_2, t_3 \dots$

Během hlavní provozní hodiny je tok volání homogenní: počet volání ve stejných intervalech je stejný a úměrný délce časového intervalu.

Volání jsou nezávislá, pp. výskytu více než 1 volání v intervalu Δt je zanedbatelně malá.

a) → POISSONOVSKÉ ROZDĚLENÍ.

$P(k, \tau)$ $\nu = 120 / \text{hod.}$

$\tau_1 = 1/60$ hod ... interval 1 minuty

$$\lambda_1 = \nu \tau_1 = 120 \cdot \frac{1}{60} = 2.$$

b) $\underline{P(0,\tau_1) = e^{-\lambda_1} = e^{-2} \doteq 0.135}$

c) $\underline{P(1,\tau_1)} = \lambda_1 e^{-\lambda_1} = \underline{0.27}$

alespoň 1 volání: $\underline{P(k \geq 1, \tau_1)} = 1 - P(0, \tau_1) = \underline{0.865}$

CENTRÁLNÍ LIMITNÍ VĚTA

Nechť $X_1, X_2, \dots X_n$ jsou nezávislé náhodné proměnné, které mají stejnou hustotu pravděpodobnosti, stejné střední hodnoty $\overline{X_i} = \overline{X_1}$ a disperze $\mathrm{var}\, X_i = \mathrm{var}\, X_1 = \sigma^2$

Potom $Y = \sum_{k=1}^{n} X_k$ je asymptoticky normální (při velkých n) náhodná veličina se střední hodnotou $\overline{Y} = n \cdot \overline{X_1}$ a disperzí $\sigma^2 \cdot n$ pokud σ existuje.

Prakticky pro všechny šumové jevy je experimentální určení hustoty pravděpodobnosti získem velmi malé informace, protože je nejpravděpodobnější, že se toto rozdělení ukáže být NORMÁLNÍM.

VĚTA O DISPERZI (~~BE~~BORGESS) → (R.E. BURGESS)

zašmodrcho!

posloupnost N událostí v časovém intervalu τ

Nechť a_i náleží každé události (realizace)

Nechť $n = \sum_{i=1}^{N} a_i$

Předpokládejme, že N i a_i fluktuují; $\overline{a_i} = \overline{a}$, $\overline{a_i^2} = \overline{a^2}$ nezávisle na i

Potom platí: $\overline{n} = \overline{N}\,\overline{a}$; $\mathrm{var}\, n = (\overline{a})^2 \mathrm{var}\, N + \overline{N}\, \mathrm{var}\, a$

D: středování ve 3 etapách ve statistickém souboru:

a) rozdělíme soubor na podsoubory se stejnými hodnotami N

b) středujeme nejdříve v každém podsouboru odděleně: $\overline{}^{s}$

c) středujeme přes všechny podsoubory

1. t.j. $\overline{n} = \overline{\overline{n}^{s}} = \overline{N \cdot \overline{a}} = \overline{N}\,\overline{a}$

2. $n^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j \Rightarrow \overline{n^2}^{s} = \sum_{i=1}^{N} \sum_{j=1}^{N} \overline{a_i a_j}^{s} = N(N-1)(\overline{a})^2 + N\overline{a^2}$

Σ obsahuje :

N sčítanců $a_i a_j$ s $i=j$, pro které $\overline{a_i a_j}^s = \overline{a^2}$

a $N(N-1)$ sčítanců $a_i a_j$ s $i \neq j$, pro které $\overline{a_i a_j}^s = (\bar{a})^2$

$$\overline{n^2}^s = N^2 (\bar{a})^2 + N \operatorname{var} a$$

t.j. $\overline{n^2} = \overline{\overline{n^2}^s} = \overline{N^2} (\bar{a})^2 + \bar{N} \operatorname{var} a$

teda

$$\operatorname{var} n = \overline{n^2} - (\bar{n})^2 = (\bar{a})^2 \left[\overline{N^2} - (\bar{N})^2 \right] + \bar{N} \operatorname{var} a .$$

STATISTICKÉ ZPRACOVÁNÍ DAT

histogram
bodový odhad parametrů
(nestrannost, konzistence, eficience)
intervalový odhad stř. hodnoty

Teoretické charakteristiky:

Střední hodnota: $\bar{x}$

Rozptyl: σ_x^2

Kovariance: K_{xy}

Koeficient korelace: r

Výběrové charakteristiky souboru o rozsahu n:

Výběrový průměr: $m_x = \frac{1}{n}\sum_{i=1}^{n} x_i$

Výběrový rozptyl: $s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - m_x)^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - m_x^2$

Výběrová kovariance: $S_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - m_x)(y_i - m_y)$

Výběrový koeficient korelace: $\rho = \frac{S_{xy}}{s_x s_y}$.

Bodový odhad teoretických charakteristik:

Odhad střední hodnoty: $\tilde{\bar{x}} = m_x$;

Odhad rozptylu: $\tilde{\sigma}_x^2 = \frac{n}{n-1} s_x^2$;

Odhad kovariance: $\tilde{K}_{xy} = \frac{n}{n-1} S_{xy}$;

Odhad koeficientu korelace: $\tilde{r} = \rho$.

Intervalový odhad teoretických charakteristik:

Interval spolehlivosti pro střední hodnotu: $m - \varepsilon < \bar{x} < m + \varepsilon$.

Přesnost odhadu při známém rozptylu: $\varepsilon = z_p \sigma \cdot \frac{1}{\sqrt{n}}$.

~" při neznámém ~" : $\varepsilon = t_p s \frac{1}{\sqrt{n-1}}$.

Koeficient spolehlivosti α:
(pravděpodobnost pokrytí teoretické charakteristiky intervalem spolehlivosti)

Obvykle se volí $\alpha = 0.99$ nebo 0.95

Hladina významnosti p:
(pravděpodobnost, že teor. char. nebude pokryta intervalem spolehlivosti)

$p = 1 - \alpha$

Činitel z_p pro zvolenou hladinu významnosti p: (resp. z α) určíme pro daný $\Phi(z_p)$ z tabulky normálního rozdělení)

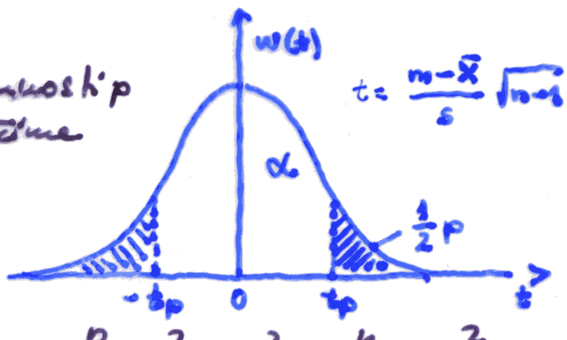
$$\Phi(z_p) = \frac{1}{2}(\alpha + 1)$$

Pro obvyklé hodnoty:

$\alpha = 0.99 \rightarrow z_{0.01} = 2.58$

$\alpha = 0.95 \rightarrow z_{0.05} = 1.96$

Činitel t_p pro zvolenou hladinu významnosti p a daný počet stupňů volnosti $k = n-1$ určíme z tabulky Studentova rozdělení.



Interval spolehlivosti pro rozptyl:

$$\frac{n}{\lambda_p} s^2 < \sigma^2 < \frac{n}{\lambda_{1-p}} s^2$$

$$\lambda = \frac{n}{\sigma^2 s^2}$$

Hladina významnosti p
(Pp., že teoretická charakteristika nebude pokryta intervalem spolehlivosti je $2p$)

$$p = \frac{1}{2}(1-\alpha)$$

Činitele λ_p a λ_{1-p} pro zvolenou hladinu významnosti p (resp. pro daný koeficient spolehlivosti α) a daný počet stupňů volnosti $k = n-1$ určíme z rozdělení χ^2.

Interval spolehlivosti pro koeficient korelace: $\rho - \varepsilon < r < \rho + \varepsilon$

Přesnost odhadu: $\varepsilon = z_p \frac{1-\rho^2}{\sqrt{n}}$

Hladina významnosti p: $p = 1 - \alpha$

Koeficient spolehlivosti α: $\alpha = 0.99$ nebo $\alpha = 0.95$

Činitel z_p pro zvolené p určíme pro daný $\Phi(z_p)$ z tabulky normálního rozdělení

$$\Phi(z_p) = \frac{1}{2}(\alpha + 1)$$