2 ŠTATISTICKÉ CHARAKTERISTIKY SYSTÉMOV

Pri identifikácii sa zväčša opierame o **informácie**, ktoré sa **získavajú experimentálne**. Výsledky experimentov bývajú **skresľované** rôznymi **rušivými faktormi**. Na účinné **potláčanie vplyvu takýchto negatívnych faktorov** a na **optimálnu selekciu užitočnej informácie** zo stochastického prostredia sa pri spracúvaní nameraných údajov používajú metódy **matematickej štatistiky**.

2.1. Základné pojmy

Deterministická časová funkcia – v každom čase má **jednoznačne** priradenú hodnotu.

Náhodná (**stochastická**) funkcia – ku každému časovému okamihu je priradená **náhodná** hodnota z určitej množiny možných hodnôt.

Náhodný jav A je charakterizovaný **pravdepodobnosťou** $P(A)$ (či pri splnení určitých podmienok môže jav nastať), $0 < P(A) < 1$.

Jav istý $\Rightarrow$ $P(A) = 1$.

Jav nemožný $\Rightarrow$ $P(A) = 0$.

2.2. Náhodné veličiny a náhodné procesy

2.2.1. Náhodné veličiny

Náhodná veličina ξ – premenná, ktorej hodnota je jednoznačne určená výsledkom náhodného pokusu.

Diskrétna náhodná veličina ξ – množina hodnôt, ktoré nadobúda, je konečná postupnosť čísel $\{x_1, x_2, \ldots, x_n\}$. Diskrétna náhodná veličina je určená vtedy, ak sú známe jej číselné hodnoty ktoré nadobúda ako aj pravdepodobnosť $P_i = P(\xi = x_i)$ výskytu každej z nich.

Spojitá náhodná veličina – môže nadobúdať všetky hodnoty v ohraničenom intervale (a, b) alebo v intervale $(-\infty, \infty)$ a distribučná funkcia je v týchto intervaloch spojitá.

Budeme ďalej **uvažovať spojité náhodné veličiny**. Ak je ξ náhodná veličina a x je jej realizácia (konkrétna hodnota), potom:

Distribučná funkcia veličiny ξ je pravdepodobnosť javu $\xi < x$ $\Rightarrow$ $F(x) = P(\xi < x)$

pričom $\lim\limits_{x\to\infty} F(x) = 1$ a $\lim\limits_{x\to-\infty} F(x) = 0$

Pravdepodobnosť, že spojitá náhodná veličina je v intervale (a, b) $\Rightarrow$

$$P(a \le \xi < b) = F(b) - F(a)$$

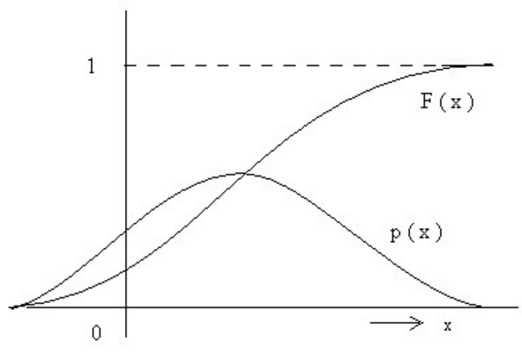
Pravdepodobnosť, že spojitá náhodná veličina bude medzi hodnotami x a x+dx $\Rightarrow$

$$P(x \le \xi < x + dx) = \frac{\partial F(x)}{\partial x} dx \qquad \text{pričom} \quad 0 \le F(x) \le 1$$

Hustota pravdepodobnosti $\Rightarrow$ $\boxed{p(x) = \dfrac{dF(x)}{dx} \ge 0}$

pričom $\boxed{F(x) = \int\limits_{-\infty}^{x} p(x) dx}$

$$P(a \le \xi < b) = \int\limits_a^b p(x) dx$$



$$\lim_{x\to\infty}\int_{-\infty}^{x} p(x)dx = \int_{-\infty}^{\infty} p(x)dx = 1$$

Začiatočný moment k-teho stupňa náhodnej veličiny $\xi \Rightarrow$ $\boxed{\mu_\xi^k = E\{\xi^k\} = \int_{-\infty}^{\infty} x^k\, p(x)dx}$

Prvý začiatočný moment – stredná hodnota (matematická nádej) spojitej náhodnej veličiny ξ (číslo okolo ktorého sa sústreďujú hodnoty náhodnej veličiny)

$$\Rightarrow \boxed{\mu_\xi = E\{\xi\} = \int_{-\infty}^{\infty} x\, p(x)dx}$$

Centrálny moment k-teho stupňa $\Rightarrow$ $\boxed{\overset{\bullet}{\mu}{}_\xi^k = E\{(\xi-\mu_\xi)^k\} = \int_{-\infty}^{\infty} (x-\mu_\xi)^k\, p(x)dx}$

Prvý centrálny moment je nulový

$$\Rightarrow \overset{\bullet}{\mu}_\xi = E\{\xi-\mu_\xi\} = \int_{-\infty}^{\infty} (x-\mu_\xi)p(x)dx = \int_{-\infty}^{\infty} x\,p(x)dx - \mu_\xi = \mu_\xi - \mu_\xi = 0$$

Druhý centrálny moment – rozptyl (variancia, disperzia) spojitej náhodnej veličiny ξ (ako ďaleko od strednej hodnoty sú jednotlivé hodnoty náhodnej veličiny rozptýlené)

$$\Rightarrow \boxed{\overset{\bullet}{\mu}{}_\xi^2 = \sigma_\xi^2 = D(\xi) = E\{(\xi-\mu_\xi)^2\} = \int_{-\infty}^{\infty} (x-\mu_\xi)^2 p(x)dx}$$

Stredná kvadratická odchýlka (**smerodajná odchýlka**) náhodnej veličiny (druhá odmocnina z rozptylu) $\Rightarrow$ $\boxed{\sigma_\xi = \sqrt{E(\xi^2) - E\{\xi\}^2}}$

Normálny zákon rozdelenia spojitej náhodnej veličiny $N(\mu_\xi, \sigma_\xi^2)$ (**Gaussovo rozdelenie**) – jeho hustota pravdepodobnosti je $p(x) = \dfrac{1}{\sigma_\xi\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu_\xi)^2}{\sigma_\xi^2}}$ $\Rightarrow$ označenie $\boxed{\xi \in N(\mu_\xi, \sigma_\xi^2)}$

Normovaná centrálna náhodná veličina $\xi_n = \dfrac{\xi-\mu_\xi}{\sigma_\xi}$ sa riadi štandardizovaným normálnym rozdelením N(0,1). $\Rightarrow$ označenie $\boxed{\xi_n \in N(0,1)}$

Majme **dve spojité náhodné veličiny** ξ_1, ξ_2 definované na **tom istom pravdepodobnostnom intervale**.

Združená hustota pravdepodobnosti dvoch náhodných veličín $\Rightarrow$ ak ξ_1 a ξ_2 sú nezávislé, potom $p(x_1, x_2) = p_1(x_1)p_2(x_2)$

Začiatočný moment dvoch náhodných veličín $\Rightarrow$ $E\{\xi_1^r, \xi_2^s\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1^r x_2^s\, p(x_1, x_2)dx_1dx_2$

Centrálny moment dvoch náhodných veličín $\Rightarrow$

$$E\{(\xi_1-\mu_1)^r(\xi_2-\mu_2)^s\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x_1-\mu_1)^r(x_2-\mu_2)^s p(x_1, x_2)dx_1dx_2$$

Kovariancia (korelačný moment) **náhodných veličín** $\Rightarrow$

$$\operatorname{cov}(\xi_1, \xi_2) = E\{(\xi_1-\mu_1)(\xi_2-\mu_2)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x_1-\mu_1)(x_2-\mu_2)p(x_1, x_2)dx_1dx_2$$

Koeficient korelácie náhodných veličín $\Rightarrow$ $r(\xi_1,\xi_2)=\dfrac{\mathrm{cov}(\xi_1,\xi_2)}{\sigma_1\sigma_2}$

Náhodné veličiny ξ_1 a ξ_2 sa nazývajú **nekorelované** ak $\mathrm{cov}(\xi_1,\xi_2)=0$

alebo $E\{\xi_1,\xi_2\}=E\{\xi_1\}E\{\xi_2\}$

Majme **n náhodných veličín**, ktoré sú prvkami vektora $\xi=(\xi_1,\xi_2,\ldots,\xi_n)$

Stredná hodnota $E\{\xi\}$ **vektora** ξ je vektor stredných hodnôt

Kovariančná matica $\mathrm{cov}(\boldsymbol{\xi},\boldsymbol{\xi})$ **vektora** ξ $\Rightarrow$ $\mathrm{cov}(\boldsymbol{\xi},\boldsymbol{\xi})=E\{(\boldsymbol{\xi}-E(\boldsymbol{\xi}))(\boldsymbol{\xi}-E(\boldsymbol{\xi}))^T\}$

je to **symetrická kladne definitná** matica

v **i-tom riadku a j-tom stĺpci** má **kovarianciu náhodných veličín** ξ_i a ξ_j $\Rightarrow$

$$\mathrm{cov}(\xi_i,\xi_j)=E\{(\xi_i-E(\xi_i))(\xi_j-E(\xi_j))^T\}=\mathrm{cov}(\xi_j,\xi_i)$$

na **hlavnej diagonále** sú **rozptyly** náhodných veličín ξ_i $\Rightarrow$

$$\mathrm{cov}(\xi_i,\xi_i)=E\{(\xi_i-E(\xi_i))(\xi_i-E(\xi_i))^T\}=\sigma_i^2$$

Korelačná matica (druhý začiatočný moment) $\Rightarrow$ $E\{\boldsymbol{\xi}\boldsymbol{\xi}^T\}=R_{\boldsymbol{\xi}\boldsymbol{\xi}}=\int_{-\infty}^{\infty}\mathbf{x}\mathbf{x}^T p(\mathbf{x})d\mathbf{x}$

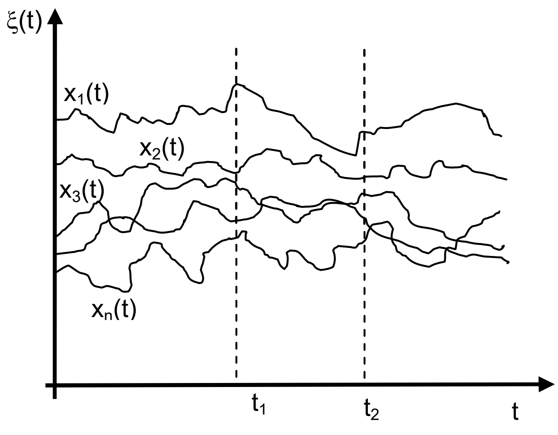
je to **symetrická** matica

v **i-tom riadku a j-tom stĺpci** má **koeficienty korelácie náhodných veličín** ξ_i a ξ_j $\Rightarrow$

$$\rho_{ij}=r(\xi_i,\xi_j)$$

na **hlavnej diagonále** sú **jednotky** $\Rightarrow$ $\rho_{ii}=1$

2.2.2. <u>Náhodné procesy</u>

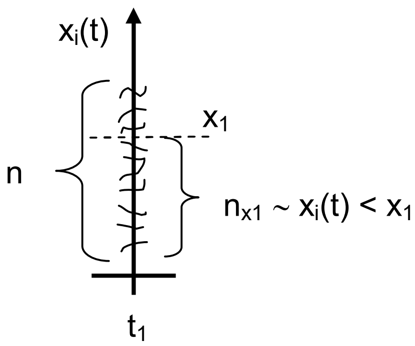


Náhodný proces $\xi(t)$ – množina náhodných veličín závislých od času (funkcia času, ktorej hodnoty sú v každom čase náhodnou veličinou).

Realizácia náhodného procesu $x(t)$ – náhodná veličina, zodpovedajúca podmienkam jednotlivého pokusu a meniaca sa v čase, ktorá patrí do množiny náhodných veličín $\xi(t)$.

Náhodný proces môže byť určený **súborom hustôt pravdepodobností**, ktoré zodpovedajú spojitým náhodným veličinám $\xi(t_1),\xi(t_2),\ldots,\xi(t_n)$ pre okamihy času $t_1,t_2,\ldots,t_n$:

Distribučná funkcia $\Rightarrow$ $F_1(x_1,t_1)=P(\xi(t_1)<x_1)=\dfrac{n_{x1}(t_1)}{n}$



kde $n_{x1}(t_1)$ je počet realizácií v čase t_1, ktoré spĺňajú podmienku $x_i(t_1)\le x_1$.

Jednorozmerová hustota pravdepodobnosti $\Rightarrow$

$$p_1(x_1,t_1)=\frac{dF_1(x_1,t_1)}{dx_1}$$

Dvojrozmerová hustota pravdepodobnosti $\Rightarrow$

$$p_2(x_1,t_1,x_2,t_2)=\frac{\partial^2 F_2(x_1,t_1,x_2,t_2)}{\partial x_1 \partial x_2} \qquad F_2(x_1,t_1,x_2,t_2)=P(\xi(t_1)<x_1,\xi(t_2)<x_2)$$

Úplne náhodný proces (biely šum) – všetky hodnoty náhodnej veličiny v jednotlivých okamihoch času sú navzájom nezávislé $\Rightarrow$ $p_2(x_1,t_1;x_2,t_2)=p(x_1,t_1)p(x_2,t_2)$

až $p_n(x_1,t_1;x_2,t_2;\ldots;x_n,t_n)=p(x_1,t_1)p(x_2,t_2)\ldots p(x_n,t_n)$

Stredná hodnota náhodného procesu – je funkciou času a je strednou hodnotou pre všetky realizácie náhodného procesu $\Rightarrow$ $\mu_\xi(t)=E\{\xi(t)\}=\int\limits_{-\infty}^{\infty} x\, p_1(x,t)dx$

Rozptyl náhodného procesu – ako sú realizácie náhodného procesu rozptýlené vzhľadom na strednú hodnotu $\Rightarrow$ $D(\xi(t))=\int\limits_{-\infty}^{\infty}(x-\mu_\xi(t))^2 p_1(x,t)dx$

$$D(\xi(t))=E\{\xi^2(t)\}-(E\{\xi(t)\})^2$$

Autokorelačná funkcia – väzba medzi hodnotami náhodného procesu v okamihoch t_1 a t_2

$$\Rightarrow \qquad R_{\xi\xi}(t_1,t_2)=E\{\xi(t_1)\xi(t_2)\}=\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} x_1 x_2 p_2(x_1,t_1;x_2,t_2)dx_1dx_2$$

Autokovariančná funkcia $\Rightarrow$

$$cov_{\xi\xi}(t_1,t_2)=E\{(\xi(t_1)-\mu_\xi(t_1))(\xi(t_2)-\mu_\xi(t_2))\}=\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}(x_1-\mu_\xi(t_1))(x_2-\mu_\xi(t_2))p_2(x_1,t_1;x_2,t_2)dx_1dx_2$$

Ak $t_1=t_2=t$, potom $cov_\xi(t,t)=\{(\xi(t)-\mu_\xi(t))^2\}$ – autokovariančná funkcia je rovná rozptylu.

Vzájomná korelačná funkcia – vzájomný vzťah dvoch náhodných procesov $\xi(t)$ a $\eta(t)$

$$\Rightarrow R_{\xi\eta}(t_1,t_2)=E\{\xi(t_1)\eta(t_2)\}$$

Vzájomná kovariančná funkcia $\Rightarrow$ $cov_{\xi\eta}(t_1,t_2)=E\{(\xi(t_1)-\mu_\xi(t_1))(\eta(t_2)-\mu_\eta(t_2))\}$

V prípade nulových stredných hodnôt sa korelačná a kovariančná funkcia rovnajú.

Stacionárny náhodný proces – všetky hustoty pravdepodobnosti $p_1,p_2,\ldots,p_n$ nezávisia od času t, ale iba od rozdielu časov konečného a začiatočného a jednorozmerová hustota pravdepodobnosti vo všeobecnosti nezávisí od času, t.j. stredná hodnota a rozptyl nie sú funkciou času.

$$p_1(x_1,t_1)=p_1(x_1)$$

$$p_2(x_1,t_1,x_2,t_2)=p_1(x_1,x_2,t_2-t_1)$$

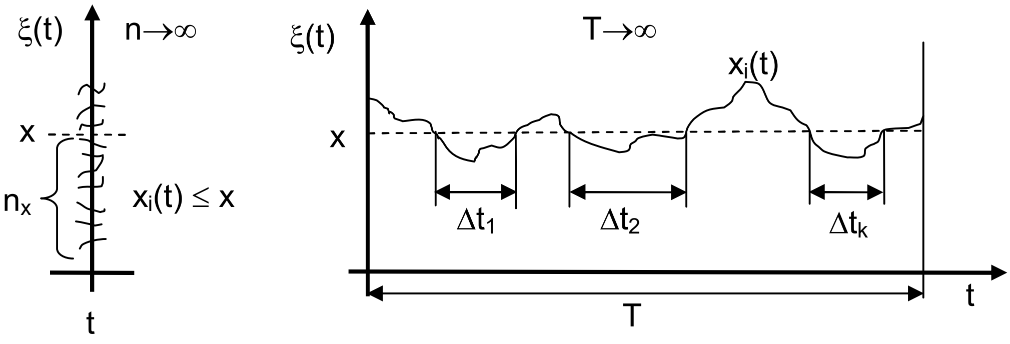
Overenie stacionárnosti: opakovaný výpočet charakteristík zo súboru realizácií pre rôzne časy, ak sú konštantné, proces je stacionárny.

Ergodický náhodný proces – s pravdepodobnosťou rovnou jednej sa jeho charakteristiky získané zo súboru realizácií náhodného procesu rovnajú charakteristikám získaným v čase, t.j. z jednej dostatočne dlhej realizácie.

Napr. pre strednú hodnotu platí

$$\mu=\int\limits_{-\infty}^{\infty}x(t)p(x)dx=\overline{\mu}=\lim_{T\to\infty}\frac{1}{2T}\int\limits_{-T}^{T}x(t)dt$$

Overenie ergodickosti: napr. vyšetrenie distribučnej funkcie



Distribučná funkcia zo súboru realizácií $x_i(t)$, t=1, 2, ..., n v čase t

$$F_n(x) = P\{x_i(t) \le x\} = \frac{n_x(t)}{n}$$

Distribučná funkcia z jednej realizácie dĺžky T (náhodne vybraná realizácia x_i)

$$F_x(x) = \frac{\sum_{i=1}^{k} \Delta t_k}{T}$$

Δt_1, Δt_2, ... Δt_k sú časové intervaly, pre ktoré platí $x_i(t) \le x$

Podmienka ergodicity

$$\lim_{n\to\infty} F_n(x) = \lim_{T\to\infty} F_x(x)$$

Na skúmanie vlastností stacionárneho a ergodického procesu postačuje skúmať jednu realizáciu náhodného procesu v dostatočne veľkom časovom intervale. V praxi sa obyčajne predpokladá, že procesy sú stacionárne a ergodické.

Pre stacionárny proces dvojrozmerová hustota pravdepodobnosti p_2 nezávisí od okamihov časov t_1 a t_2, ale závisí od $\tau = t_2 - t_1$, potom autokorelačná funkcia môže byť vyjadrená v tvare

$$R_{\xi\xi}(\tau) = E\{\xi(t_1)\xi(t_2)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 p_2(x_1, x_2, \tau)\,dx_1 dx_2$$

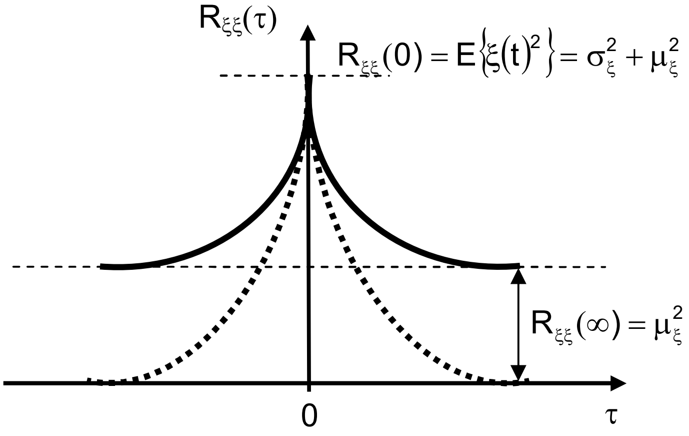
Pre stacionárny a ergodický proces je autokorelačná funkcia

$$R_{\xi\xi}(\tau) = E\{\xi(t)\xi(t-\tau)\} = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} x(t)x(t-\tau)\,dt$$

platí $R_{\xi\xi}(-\tau) = R_{\xi\xi}(\tau)$

$$R_{\xi\xi}(0) = E\{\xi(t)^2\} \ge R_{\xi\xi}(\tau)$$

$$R_{\xi\xi}(\infty) = \mu_\xi^2 = (\overline{\mu}_\xi)^2$$



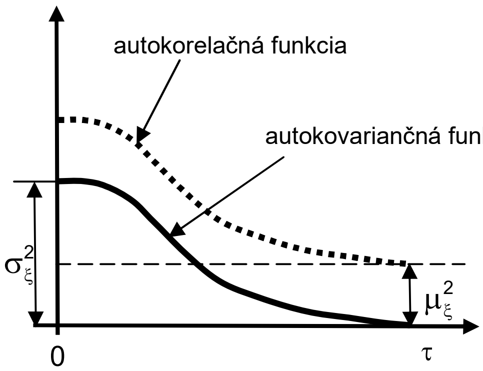
Čím rýchlejšie funkcia klesá pre rastúce $|\Delta\tau|$, tým menšia je štatistická súvislosť po sebe idúcich hodnôt signálu ξ(t).

Pozvoľnejší priebeh znamená väčšiu zotrvačnosť systému.

Autokovariančná funkcia

$$\mathrm{cov}_{\xi\xi}(\tau) = V_{\xi}(\tau) = E\{(\xi(t)-\mu)(\xi(t-\tau)-\mu)\} =$$

$$= \lim_{T\to\infty} \frac{1}{2T} \int_{-\infty}^{\infty} (x(t)-\mu)(x(t-\tau)-\mu)\ dt$$



Výkonová spektrálna hustota náhodného procesu – <u>Fourierov obraz autokorelačnej funkcie</u>

$$\underbrace{S_{\xi\xi}(\omega) = \int_{-\infty}^{\infty} R_{\xi\xi}(\tau) e^{-j\omega\tau} d\tau \quad \text{potom} \quad R_{\xi\xi}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{\xi\xi}(\omega) e^{j\xi\tau} d\omega}_{\textbf{Wiener – Chinčinove rovnice}}$$

platí $\quad S_{\xi\xi}(-\omega) = S_{\xi\xi}(\omega)$

$$S_{\xi\xi}(\omega) = 2\int_{0}^{\infty} R_{\xi\xi}(\tau)\cos(\omega\tau)d\tau \quad \text{a} \quad R_{\xi\xi}(\tau) = \frac{1}{\pi}\int_{0}^{\infty} S_{\xi\xi}(\omega)\cos(\omega\tau)d\omega$$

Vzájomná výkonová spektrálna hustota dvoch ergodických náhodných procesov s nulovými strednými hodnotami – <u>Fourierov obraz ich vzájomnej korelačnej funkcie</u>

$$S_{\xi\eta}(j\omega) = \int_{-\infty}^{\infty} R_{\xi\eta}(\tau) e^{-j\omega\tau} d\tau \qquad \text{potom} \quad R_{\xi\eta}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_{\xi\eta}(j\omega) e^{j\omega\tau} d\omega$$

$$R_{\xi\eta}(\tau) = E\{\xi(t)\eta(t-\tau)\} = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} x(t)y(t-\tau)dt$$

Ak označíme $\boldsymbol{\xi}(j\omega)$ a $\boldsymbol{\eta}(j\omega)$ Fourierove obrazy vektorov $\boldsymbol{\xi}(t)$ a $\boldsymbol{\eta}(t)$, potom platí

$$S_{\boldsymbol{\xi\xi}}(\omega) = \boldsymbol{\xi}(j\omega)\boldsymbol{\xi}^{T}(-j\omega)$$

$$S_{\boldsymbol{\xi\eta}}(\omega) = \boldsymbol{\xi}(j\omega)\boldsymbol{\eta}^{T}(-j\omega)$$

Biely šum – stacionárny náhodný proces, pre ktorý platí, že pre všetky frekvencie má rovnakú hodnotu výkonovej spektrálnej hustoty $S_{\xi\xi}(\omega) = V$, potom

$$R_{\xi\xi}(\tau) = \frac{1}{\pi}\int_{0}^{\infty} V\cos(\omega\tau)d\omega = V\delta(\tau) \qquad \text{kde} \qquad \delta(\tau) = \frac{1}{\pi}\int_{0}^{\infty}\cos(\omega\tau)d\omega \ \text{ je Diracova funkcia.}$$

Diskrétny biely šum $\xi(t)$ je signál **s nulovou strednou hodnotou, konečným rozptylom a nekorelovanými prvkami**

$$E\{\xi(t)\} = 0 \qquad \text{a} \qquad E\{\xi(t)\xi^{t}(t+j)\} = \begin{cases} \sigma^2, & j = 0 \\ 0, & j \neq 0 \end{cases}$$

2.3. <u>Typy a vlastnosti odhadov</u>

2.3.1. <u>Typy odhadov</u>

1. **Bodový dohad** (*point estimate*) – parameter aproximujeme jedným číslom
2. **Intervalový odhad** (*confidence interval*) – parameter aproximujeme intervalom, v ktorom jeho hodnota leží s určitou pravdepodobnosťou

2.3.2. Vlastnosti bodového odhadu

Cieľ identifikácie: získať **odhad** $\hat{\boldsymbol{\theta}}$ vektora parametrov $\boldsymbol{\theta}$.

Označme $\hat{\boldsymbol{\theta}}_N$ odhad urobený na základe N meraní.

1. Odhad nazývame **nevychýlený** (nestranný, neskreslený, *unbiased estimate*), ak pre každé N platí, že **stredná hodnota odhadu** je priamo **rovná skutočnej hodnote** $\boldsymbol{\theta}$

   $$E\{\hat{\boldsymbol{\theta}}_N\} - \boldsymbol{\theta} = \lambda(\hat{\boldsymbol{\theta}}_N) = 0 \qquad \lambda(\hat{\boldsymbol{\theta}}_N) \text{ je } \textbf{výchylka odhadu}.$$

2. Odhad nazývame **výdatný** (efektívny), ak spomedzi všetkých nevychýlených odhadov $\bar{\boldsymbol{\theta}}$ bude jeho **kovariancia najmenšia**

   $$\operatorname{cov}\{\bar{\boldsymbol{\theta}}\} - \operatorname{cov}\{\hat{\boldsymbol{\theta}}_N\} \geq 0$$

   t.j. rozdiel pozitívne definitných matíc na ľavej strane musí byť pozitívne aspoň semidefinitnou maticou. V **skalárnom prípade** je výdatný ten odhad, ktorý má **minimálny rozptyl**.

3. Odhad nazývame **asymptoticky nevychýlený**, ak jeho výchylka asymptoticky konverguje k nule s rastom rozsahu výberového súboru $N \to \infty$.

4. Odhad nazývame **konzistentný,** ak jeho **vychýlenie aj rozptyl konvergujú k nule pre počet pozorovaní idúci do nekonečna**, t.j. so zvyšovaním počtu meraní sa stále viac koncentruje okolo skutočnej hodnoty.

2.4. Bodové odhady parametrov úplného súboru s normálnym rozdelením pravdepodobnosti

Matematická štatistika sa zaoberá **štúdiom pravdepodobnostných modelov pre úplné súbory** na základe výsledkov **skúmania na výberovom súbore** rozsahu N (počet **štatistických jednotiek** = **realizovaných experimentov**).

Neznáme parametre rozdelenia pravdepodobnosti, ktorým sa **riadi úplný súbor** náhodnej premennej **odhadujeme na základe hodnôt náhodného výberu (experimentov)** pomocou vhodne zvolenej funkcie – **štatistiky**.

Metódy hľadania bodových odhadov:

- **Metóda maximálnej vierohodnosti** (maximum likelihood method) – je založená na vlastnostiach funkcie združenej hustoty pravdepodobnosti
- **Metóda momentov** – je založená na rovnosti výberových momentov a momentov rozdelenia

2.4.1. Výberové charakteristiky

Nech $x_1, x_2, \ldots, x_N$ je **náhodný výber,** potom definujeme:

k-ty výberový začiatočný moment $\Rightarrow$ $\boxed{m_x^k = \frac{1}{N}\sum_{i=1}^{N} x_i^k}$

výberový priemer (k=1) $\Rightarrow$ $\boxed{\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i}$

k-ty výberový centrálny moment $\Rightarrow$ $\boxed{\dot{m}_x^k = \frac{1}{N}\sum_{i=1}^{N} (x_i - \bar{x})^k}$

prvý výberový centrálny moment je nulový.

druhý výberový centrálny moment $\Rightarrow$ $\boxed{\overset{\bullet}{m}{}_x^2 = s_x^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2}$

Autokorelačná funkcia...$\Rightarrow$... $\boxed{R_{xx}(\tau) = R_{xx}(kT_{vz}) = \frac{1}{N-k}\sum_{i=1}^{N-k} x(i)x(i+k)}$

Výpočet opakujeme pre rôzne τ až po τ_{max}, pričom by malo platiť $10\tau_{max} \le NT_{vz}$.

2.4.2. Bodový odhad strednej hodnoty a disperzie

a) Odhad strednej hodnoty μ_x

Hypotéza: odhadom strednej hodnoty je výberový priemer $\overline{x}$

$$E\{\overline{x}\} = E\left\{\frac{1}{N}\sum_{i=1}^{N} x_i\right\} = \frac{1}{N}\sum_{i=1}^{N} E\{x_i\} = \frac{1}{N}\sum_{i=1}^{N}\mu_x = \mu_x$$

Výberový priemer $\overline{x}$ je nevychýleným odhadom strednej hodnoty μ_x.

$$D\{\overline{x}\} = E\left\{(\overline{x} - \mu_x)^2\right\} = E\left\{\left(\frac{1}{N}\sum_{i=1}^{N} x_i - \mu_x\right)^2\right\} = E\left\{\left(\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_x)\right)^2\right\} =$$

$$= \frac{1}{N^2}\sum_{i=1}^{N} E\left\{(x_i - \mu_x)^2\right\} = \frac{1}{N^2}\sum_{i=1}^{N}\sigma_x^2 = \frac{\sigma_x^2}{N}$$

Ak **základný súbor podlieha normálnemu rozdeleniu** $x \in N(\mu_x, \sigma_x^2)$, potom **aritmetický priemer** $\overline{x}$ z N-prvkového výberového súboru **podlieha normálnemu rozdeleniu** $\overline{x} \in N\left(\mu_x, \frac{\sigma_x^2}{N}\right)$.

b) Odhad disperzie σ_x^2

Hypotéza: odhadom disperzie je druhý výberový centrálny moment s_x^2

$$s_x^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2 = \frac{1}{N}\sum_{i=1}^{N}\left[(x_i - \mu_x) - (\overline{x} - \mu_x)\right]^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_x)^2 - 2(\overline{x} - \mu_x)\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_x) + (\overline{x} - \mu_x)^2 =$$

$$= \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_x)^2 - 2(\overline{x} - \mu_x)\left(\frac{1}{N}\sum_{i=1}^{N} x_i - \mu_x\right) + (\overline{x} - \mu_x)^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_x)^2 - (\overline{x} - \mu_x)^2$$

$$E\{s_x^2\} = \frac{1}{N}\sum_{i=1}^{N} E\left\{(x_i - \mu_x)^2\right\} - E\left\{(\overline{x} - \mu_x)^2\right\} = \sigma_x^2 - \frac{\sigma_x^2}{N} = \frac{N-1}{N}\sigma_x^2 \quad \Rightarrow \quad \lim_{N\to\infty} E\{s_x^2\} = \lim_{N\to\infty}\frac{N-1}{N}\sigma_x^2 = \sigma_x^2$$

Štatistika s_x^2 je **iba asymptoticky nevychýleným** odhadom disperzie σ_x^2. **Nevychýleným odhadom disperzie** σ_x^2 preto bude **výberová disperzia**

$$\boxed{S_x^2 = \frac{N}{N-1}s_x^2 = \frac{N}{(N-1)}\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2}$$

2.5. Intervalové odhady

Ak na odhad parametrov systému použijeme metódu momentov, ktorá je založená na princípe náhrady začiatočných resp. centrálnych momentov im zodpovedajúcimi výberovými začiatočnými resp. centrálnymi momentmi, takto získané odhady spravidla **nebývajú výdatné**. Okrem **strednej hodnoty** týchto štatistík je **žiaduce poznať aj veľkosť rozptylu**, aby bolo možné urobiť závery o **kvalite odhadu**.

Intervalový odhad $\Rightarrow$ namiesto jedného vektora parametrov $\hat{\boldsymbol{\theta}}^*$ sa určuje tzv. **dolná a horná hranica intervalu spoľahlivosti (konfidenčný interval)**, v ktorom leží jeho hodnota s určitou pravedpodobnosťou – **spoľahlivosť odhadu**.

$$\boxed{P(\boldsymbol{\theta}_d \le \boldsymbol{\theta} \le \boldsymbol{\theta}_h) = 1-\alpha}$$

α sa nazýva hladina významnosti, najčastejšie $(1-\alpha) = 0.95$ alebo 0.99.

Väčšinou sa konfidenčné hranice volia tak, aby platilo:

$$P(\boldsymbol{\theta} < \boldsymbol{\theta}_d) = \frac{\alpha}{2} = P(\boldsymbol{\theta} > \boldsymbol{\theta}_h)$$

2.5.1. Rozdelenia pravdepodobnosti

Pri určovaní intervalových odhadov budeme využívať **kvantily rozdelení pravdepodobnosti**.

Kvantilová funkcia $\Rightarrow$ $Q(x) = F(x)^{-1}$ **inverzná funkcia k distribučnej funkcii**

α**-kvantil** je také číslo x_α, pre ktoré distribučná funkcia dosahuje práve hodnotu α

$$\boxed{F(x_\alpha) = \int_{-\infty}^{x_\alpha} p(x)dx = P(\xi < x_\alpha) = \alpha}$$

Špeciálne označenie kvantilov:

- $x_{0,5}$ – medián
- $x_{0,25}$ – dolný kvartil
- $x_{0,75}$ – horný kvartil
- $x_{100.\,\alpha}$ – percentil

Využitie kvantilov pri výpočte intervalu spoľahlivosti:

$$1-\alpha = 1-\frac{\alpha}{2}-\frac{\alpha}{2} = \left(1-\frac{\alpha}{2}\right)-\frac{\alpha}{2} = F\left(x_{1-\frac{\alpha}{2}}\right)-F\left(x_{\frac{\alpha}{2}}\right) = \qquad \boxed{P_{1-\alpha} = P\left(x_{\frac{\alpha}{2}} \le x \le x_{1-\frac{\alpha}{2}}\right) = 1-\alpha}$$

$P_{1-\alpha}$ je pravdepodobnosť, s akou sa bude náhodná veličina nachádzať v intervale $\left[x_{\frac{\alpha}{2}}; x_{1-\frac{\alpha}{2}}\right]$.

Výpočet intervalu spoľahlivosti spočíva v tom, že za náhodnú veličinu *x* dosadíme jej definičný vzorec a výraz upravíme tak, aby sme medzi znamienkami ≤ osamostatnili odhadovaný parameter. Použijeme kvantily toho rozdelenia, ktorým sa náhodná veličina riadi.

a. Štandardizované normálne rozdelenie (Gaussovo)

Ak náhodná veličina ξ podlieha **normálnemu zákonu rozdelenia** $N(\mu_\xi, \sigma_\xi^2)$, potom sa **transformovaná štandardizovaná** náhodná premenná $r = \dfrac{\xi - \mu_\xi}{\sigma_\xi}$ riadi rozdelením N(0,1).

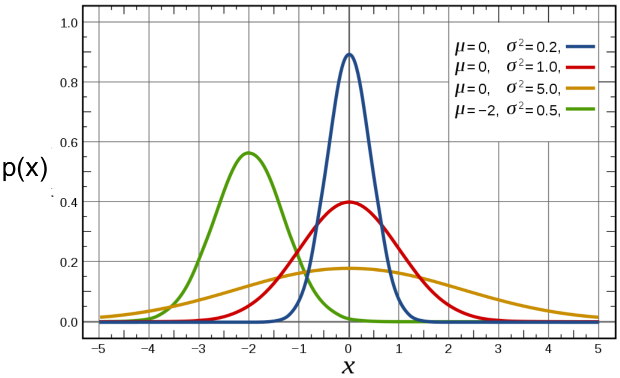
Zodpovedajúca **hustota pravdepodobnosti** $\Rightarrow$ $p(r) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{r^2}{2}}$

α-kvantil r_α $\Rightarrow$ $\alpha = \int_{-\infty}^{r_\alpha} p(r)dr$

Funkcia $p(r)$ je párna, preto platí

$$\boxed{r_{\frac{\alpha}{2}} = -r_{1-\frac{\alpha}{2}}}$$

Poznámka: hodnoty kvantilov sú dané v tabuľke.



Aritmetický priemer $\bar{x}$ z N-prvkového výberového súboru **podlieha normálnemu rozdeleniu**

$$\bar{x} \in N\left(\mu_x, \frac{\sigma_x^2}{N}\right)$$

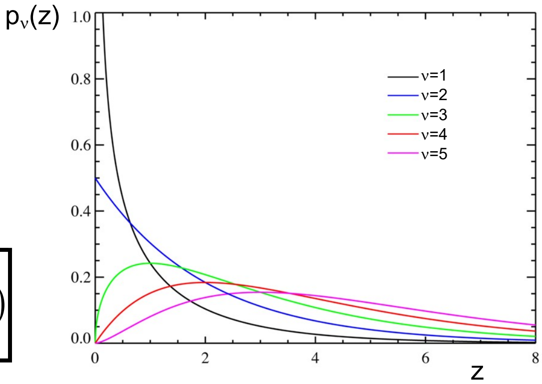
potom **normovaná náhodná veličina** sa riadi rozdelením N(0,1) $\boxed{r = \dfrac{\bar{x} - \mu_x}{\sigma_x}\sqrt{N} \in N(0,1)}$

b. Rozdelenie $\chi^2(\nu)$ (chí-kvadrát s ν-stupňami voľnosti)

Rieši sa ním náhodná premenná $\boxed{z = \sum_{i=1}^{N}\left(\dfrac{\xi_i - \mu_\xi}{\sigma_\xi}\right)^2 = \sum_{i=1}^{N} r_i^2}$,

kde premenné r_i sa riadia rozdelením N(0,1). Ak spomedzi N premenných je $N - \nu$ závislých od zvyšných premenných, riadi sa ich súčet štvorcov rozdelením $\chi^2(\nu)$ s hustotou pravdepodobnosti

$$p_\nu(z) = \frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} z^{\frac{\nu}{2}-1} e^{-\frac{z}{2}} \qquad 0 \le z < \infty$$



Rozdelením $\chi^2(N-1)$ sa riadi náhodná premenná

$$\boxed{(N-1)\left(\frac{S_x}{\sigma_x}\right)^2 = \frac{1}{\sigma_x^2}\sum_{i=1}^{N}(x_i - \bar{x})^2 = \sum_{i=1}^{N}\left(\frac{x_i - \bar{x}}{\sigma_x}\right)^2 = z \in \chi^2(N-1)}$$

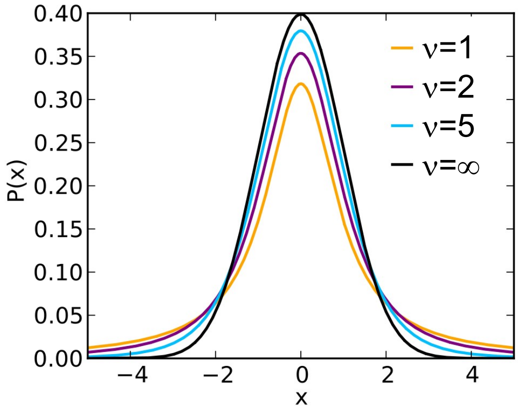
Ak sa $\chi^2(\nu_i)$ rozdelením riadi k navzájom nezávislých veličín z_i, potom aj ich suma $z = \sum_{i=1}^{k} z_i$ podlieha rozdeleniu $\chi^2(\nu)$, kde $\nu = \sum_{i=1}^{k} \nu_i$.

α-kvantil z_α $\Rightarrow$ $\alpha = \int_{-\infty}^{z_\alpha} p_\nu(z)dz$

c. Studentovo rozdelenie

Keď r je náhodná premenná s rozdelením N(0,1) a z je od nej nezávislá náhodná veličina s rozdelením $\chi^2(\nu)$, potom premenná

$$\boxed{t = \frac{r}{\sqrt{\frac{z}{\nu}}}}$$



sa riadi Studentovým rozdelením s ν-stupňami voľnosti a s hustotou pravdepodobnosti

$$p_\nu(t) = \frac{1}{\sqrt{\nu\pi}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{\frac{-\nu+1}{2}} \qquad -\infty < t < \infty$$

Pre $\nu \to \infty$ prechádza Studentovo rozdelenie na rozdelenie N(0,1).

Pretože aritmetický priemer $\bar{x}$ z N-prvkového výberového súboru je náhodná veličina s rozdelením $N\left(\mu_x, \frac{\sigma_x^2}{N}\right)$, môžeme ju štandardizovať $\quad r = \frac{\bar{x} - \mu_x}{\sigma_x}\sqrt{N} \in N(0,1)$

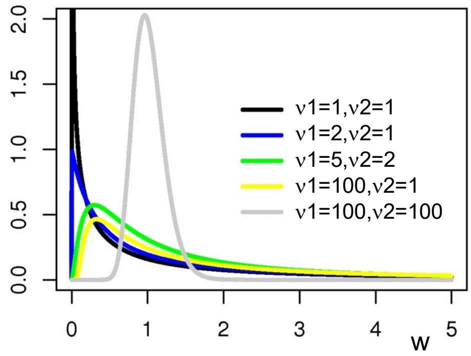
Od nej nezávislá bude premenná $\quad z = (N-1)\left(\frac{S_x}{\sigma_x}\right)^2 \in \chi^2(N-1)$

Potom
$$\boxed{t = \frac{r}{\sqrt{\frac{z}{\nu}}} = \frac{\frac{\bar{x} - \mu_x}{\sigma_x}\sqrt{N}}{\frac{S_x}{\sigma_x}} = \frac{\bar{x} - \mu_x}{S_x}\sqrt{N} \in St(N-1)}$$

d. Fisher-Snedecorovo rozdelenie

Keď z_1 je náhodná premenná s rozdelením $\chi^2(\nu_1)$ a z_2 je od nej nezávislá náhodná veličina s rozdelením $\chi^2(\nu_2)$, potom premenná

$$\boxed{w = \frac{\frac{z_1}{\nu_1}}{\frac{z_2}{\nu_2}}}$$



sa riadi Fisherovým-Snedecorovým rozdelením F,S_{ν_1,ν_2} s hustotou pravdepodobnosti

$$p_{\nu_1,\nu_2}(w) = \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\Gamma\left(\frac{\nu_2}{2}\right)} \left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} \frac{w^{\frac{\nu_1-2}{2}}}{\left(1 + \frac{\nu_1}{\nu_2} w\right)^{\frac{\nu_1+\nu_2}{2}}} \qquad 0 \le w \le \infty$$

Ak zo základného súboru s rozdelením $N(\mu_x, \sigma_x^2)$ urobíme náhodný výber rozsahu N_x a z iného základného súboru s rozdelením $N(\mu_y, \sigma_y^2)$ urobíme náhodný výber rozsahu N_y,

$$\bar{x} = \frac{1}{N_x}\sum_{i=1}^{N_x} x_i \in N\left(\mu_x, \frac{\sigma_x^2}{N_x}\right) \qquad \bar{y} = \frac{1}{N_y}\sum_{i=1}^{N_y} y_i \in N\left(\mu_y, \frac{\sigma_y^2}{N_y}\right)$$

$$z_x = (N_x - 1)\left(\frac{S_x}{\sigma_x}\right)^2 = \sum_{i=1}^{N_x}\left(\frac{x_i - \bar{x}}{\sigma_x}\right)^2 \in \chi^2(N_x - 1)$$

$$z_y = (N_y - 1)\left(\frac{S_y}{\sigma_y}\right)^2 = \sum_{i=1}^{N_y}\left(\frac{y_i - \bar{y}}{\sigma_y}\right)^2 \in \chi^2(N_y - 1)$$

potom pomocou týchto štatistík môžeme vytvoriť náhodnú veličinu

$$\boxed{w = \frac{\dfrac{(N_x - 1)S_x^2}{\sigma_x^2(N_x - 1)}}{\dfrac{(N_y - 1)S_y^2}{\sigma_y^2(N_y - 1)}} = \frac{S_x^2\sigma_y^2}{S_y^2\sigma_x^2} \in F, S_{N_x - 1, N_y - 1}}$$

ZHRNUTIE

Do vzťahu **pre výpočet intervalu spoľahlivosti**

$$\boxed{P_{1-\alpha} = P\left(x_{\frac{\alpha}{2}} \le x \le x_{1-\frac{\alpha}{2}}\right) = 1 - \alpha}$$

dosadíme za x náhodnú veličinu r, z, t alebo w a kvantily príslušného rozdelenia (napr. z tabuliek) a **upravíme** tak, aby **medzi znamienkami ≤ bol osamostatnený parameter**, pre ktorý **hľadáme intervalový odhad**.

$$\boxed{r = \frac{\bar{x} - \mu_x}{\sigma_x}\sqrt{N} \in N(0,1)} \qquad \boxed{z = (N-1)\left(\frac{S_x}{\sigma_x}\right)^2 \in \chi^2(N-1)}$$

$$\boxed{t = \frac{\bar{x} - \mu_x}{S_x}\sqrt{N} \in St(N-1)} \qquad \boxed{w = \frac{S_x^2\sigma_y^2}{S_y^2\sigma_x^2} \in F, S_{N_x - 1, N_y - 1}}$$

2.5.2. Intervalový odhad strednej hodnoty μ_x pomocou výberového súboru rozsahu N

a. Poznáme rozptyl základného súboru σ_x^2

Použijeme premennú: $\boxed{r = \frac{\bar{x} - \mu_x}{\sigma_x}\sqrt{N} \in N(0,1)}$

Pre zvolenú hodnotu koeficientu spoľahlivosti $1 - \alpha$ $(1\% \le \alpha \le 5\%)$ nájdeme v tabuľkách normálneho rozdelenia kvantil $r_{1-\frac{\alpha}{2}}$, potom môžeme **vyjadriť interval spoľahlivosti** vzťahom

$$P\left(-r_{1-\frac{\alpha}{2}} \le r \le r_{1-\frac{\alpha}{2}}\right) = P\left(-r_{1-\frac{\alpha}{2}} \le \frac{\bar{x} - \mu_x}{\sigma_x}\sqrt{N} \le r_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

po úprave dostaneme $\boxed{P\left(\bar{x}-\frac{\sigma_x}{\sqrt{N}}r_{1-\frac{\alpha}{2}} \le \mu_x \le \bar{x}+\frac{\sigma_x}{\sqrt{N}}r_{1-\frac{\alpha}{2}}\right)=1-\alpha}$

Odhadovaná stredná hodnota μ_x sa nachádza s pravdepodobnosťou $1-\alpha$ v pásme okolo $\bar{x}$, ktorého šírka je dvojnásobkom tzv. **prípustnej chyby odhadu**

$$\boxed{\delta_x = \frac{\sigma_x}{\sqrt{N}}r_{1-\frac{\alpha}{2}}}$$

Príklad → cvičenia

b. <u>Rozptyl základného súboru nepoznáme</u>

Nahradíme ho výberovým rozptylom S_x^2 a budeme pracovať s premennou

$$\boxed{t = \frac{\bar{x}-\mu_x}{S_x}\sqrt{N} \in St(N-1)}$$

dosadíme ju do intervalu spoľahlivosti $P\left(-t_{1-\frac{\alpha}{2}} \le \frac{\bar{x}-\mu_x}{S_x}\sqrt{N} \le t_{1-\frac{\alpha}{2}}\right)=1-\alpha$

a upravíme $\boxed{P\left(\bar{x}-\frac{S_x}{\sqrt{N}}t_{1-\frac{\alpha}{2}} \le \mu_x \le \bar{x}+\frac{S_x}{\sqrt{N}}t_{1-\frac{\alpha}{2}}\right)=1-\alpha} \Rightarrow \boxed{\delta_x = \frac{S_x}{\sqrt{N}}t_{1-\frac{\alpha}{2}}}$

kde $t_{1-\frac{\alpha}{2}}$ je kvantil Studentovho rozdelenia s N-1 stupňami voľnosti

Príklad → cvičenia

2.5.3. <u>Intervalový odhad rozptylu σ_x^2 pomocou výberového súboru rozsahu N</u>

Použijeme premennú $\boxed{z = (N-1)\left(\frac{S_x}{\sigma_x}\right)^2 \in \chi^2(N-1)}$

$$P\left(z_{\frac{\alpha}{2}} \le (N-1)\left(\frac{S_x}{\sigma_x}\right)^2 \le z_{1-\frac{\alpha}{2}}\right)=1-\alpha \quad \Rightarrow \quad P\left(S_x\sqrt{\frac{N-1}{z_{1-\frac{\alpha}{2}}}} \le \sigma_x \le S_x\sqrt{\frac{N-1}{z_{\frac{\alpha}{2}}}}\right)=1-\alpha$$

dosadenie do intervalu spoľahlivosti potom pre smerodajnú odchýlku platí

a pre rozptyl platí $\boxed{P\left(\frac{N-1}{z_{1-\frac{\alpha}{2}}}S_x^2 \le \sigma^2 \le \frac{N-1}{z_{\frac{\alpha}{2}}}S_x^2\right)=1-\alpha}$

kde $z_{1-\frac{\alpha}{2}}$ je kvantil $\chi^2(N-1)$ rozdelenia

Príklad → cvičenia

Poznámka: **Intervalový odhad parametrov rozdelení je možné v Matlabe vypočítať pomocou príkazu *paramci*.**

2.6. Testovanie štatistických hypotéz

Každý test hypotézy je zameraný na **overenie** určitej **základnej hypotézy**, tzv. **nulovej hypotézy H_0**. Oproti nulovej hypotéze stojí **alternatívna hypotéza H_1**. Cieľom testovania hypotéz je **rozhodnutie o prijatí alebo zamietnutí základnej hypotézy**. Ak na základe testu zamietame nulovú hypotézu, prijímame alternatívnu hypotézu.

Pri testovaní H_0 oproti H_1 sa môžeme **dopustiť jednej z dvoch chýb**:

- **chyba 1. druhu**: H_0 zamietneme, aj keď v skutočnosti platí,
- **chyba 2. druhu**: H_0 nezamietneme, aj keď v skutočnosti neplatí.

skutočnosť	**rozhodnutie**	
	H_0 nezamietame	**H_0 zamietame**
H_0 platí	správne rozhodnutie	**chyba 1. druhu**
H_0 neplatí	**chyba 2. druhu**	správne rozhodnutie

Pravdepodobnosť chyby 1. druhu sa označuje α a nazýva sa hladina významnosti testu (väčšinou býva α=0,05 menej často α=0,1 alebo α=0,01).

Testovaním H_0 oproti H_1 rozumieme rozhodovací postup založený na náhodnom výbere, pomocou ktorého zamietame alebo nezamietame platnosť nulovej hypotézy.

Najčastejšie sa preveruje pravdivosť hypotézy o rovnosti dvoch stredných hodnôt alebo dvoch rozptylov $\Rightarrow$ všeobecne **rovnosť ľubovoľnej dvojice parametrov** $\beta_1 = \beta_0$. Predpokladáme, že parameter β_1 nepoznáme a určujeme ho prostredníctvom jeho odhadu b_1. Preto ani v prípade pravdivého výroku nemusí byť nulový rozdiel

$$\Delta = b_1 - \beta_0 = (\beta_1 - \beta_0) - (\beta_1 - b_1)$$

čo môže byť spôsobené

a. **náhodnou chybou** odhadu $\Delta_n = (\beta_1 - b_1)$, ktorú do istej hodnoty tolerujeme a **hypotézu prijímame**

b. **systematickou** chybou $\Delta_s = (\beta_1 - \beta_0)$, pri ktorej **hypotézu zamietame**.

Tieto chyby však nevieme od seba oddeliť, preto sa stanovuje **prahová hodnota** Δ_k, pri **prekročení ktorej sa hypotéza odmieta**. **Malé hodnoty** Δ pripisujeme chybe **náhodnej** a **väčšie systematickej**. **Prahovú hodnotu** Δ_k **stanovujeme** pomocou **kvantilu** rozdelenia, ktorým sa štatistika riadi.

2.6.1. Testovanie pomocou intervalu spoľahlivosti

Prahovú hodnotu stanovujeme pomocou kvantilu rozdelenia, ktorým sa štatistika riadi. **Zostrojíme** pre príslušnú náhodnú veličinu **(1-α) interval spoľahlivosti a ak pokryje** tento interval **hodnotu náhodnej veličiny**, potom **nulovú hypotézu na hladine významnosti α nezamietame**.

a) t-test (Studentov test) – overenie strednej hodnoty

Overujeme hypotézu, či **normálne rozdelenie, z ktorého pochádza náhodný výber má určitú konkrétnu strednú hodnotu μ_0, pričom rozptyl je neznámy**.

Nulová hypotéza: H_0: $\mu=\mu_0$ alternatívna hypotéza H_1: $\mu\neq\mu_0$.

Použijeme štatistiku $\boxed{t=\dfrac{\overline{x}-\mu_x}{S_x}\sqrt{N}\in St(N-1)}$ pre ktorú platí $\boxed{P\left(-t_{1-\frac{\alpha}{2}}\leq t\leq t_{1-\frac{\alpha}{2}}\right)=1-\alpha}$

Prahovou hodnotou je kvantil Studentovho rozdelenia s N-1 stupňami voľnosti $t_{1-\frac{\alpha}{2}}$.

Nulovú hypotézu H_0 prijímame, ak platí nerovnosť $\boxed{|t|\leq t_{1-\frac{\alpha}{2}}}$

b) F-test – porovnanie rozptylov

Vychádzame **z údajov 2 výberových súborov** a **overujeme hypotézu, či majú rovnaký rozptyl**.

$$x\in N\left(\mu_x,\sigma_x^2\right) \text{ a } y\in N\left(\mu_y,\sigma_y^2\right) \quad\Rightarrow\quad \sigma_x^2=\sigma_y^2 \ \textbf{?}$$

Nulová hypotéza: $H_0:\sigma_x^2=\sigma_y^2$

Použijeme premennú $\boxed{w=\dfrac{\dfrac{(N_x-1)S_x^2}{\sigma_x^2(N_x-1)}}{\dfrac{(N_y-1)S_y^2}{\sigma_y^2(N_y-1)}}=\dfrac{S_x^2\sigma_y^2}{S_y^2\sigma_x^2}=\dfrac{S_x^2}{S_y^2}\in F,S_{\nu_x,\nu_y}}$

$$\nu_x=N_x-1$$
$$\nu_y=N_y-1$$

Platí $\boxed{P\left(\dfrac{S_x^2}{S_y^2}\leq w_{1-\alpha}(\nu_x,\nu_y)\right)=1-\alpha}$

Prahovou hodnotou je kvantil Fisher-Snedecorovho rozdelenia $w_{1-\alpha}(\nu_x,\nu_y)$.

Hypotézu o rovnosti rozptylov prijímame ak $\boxed{\dfrac{S_x^2}{S_y^2}\leq w_{1-\alpha}(\nu_x,\nu_y)}$

Príklad → priklad_statistika.pdf + cvičenia

Poznámky:

Hodnoty kvantilov je možné vypočítať v Matlabe pomocou funkcií (Statistics toolbox):

- ***norminv*** – pre štandardizované normálne rozdelenie,
- ***chi2inv*** – pre chí-kvadrát rozdelenie,
- ***tinv*** – pre Studentovo rozdelenie,
- ***finv*** – pre Fisher-Snedecorovo rozdelenie.

alebo napr. v Exceli 2010 pomocou štatistických funkcií:

- ***NORM.INV*** – pre štandardizované normálne rozdelenie,
- ***CHISQ.INV*** – pre chí-kvadrát rozdelenie,
- ***T.INV*** – pre Studentovo rozdelenie,
- ***F.INV*** – pre Fisher-Snedecorovo rozdelenie.

Testovanie hypotéz v Matlabe:

- ***ttest*** – t-test
- ***vartest2*** – F-test
- ***vartest*** – overenie hypotézy, či normálne rozdelenie, z ktorého pochádza náhodný výber má určitú konkrétnu hodnotu rozptylu.

2.6.2. Testovanie pomocou p-hodnoty

Šírka intervalu spoľahlivosti závisí od hladiny významnosti α – čím je α menšie, tým väčší je interval (α sa zmenšuje, pravdepodobnosť, že hodnota padne do intervalu sa zvyšuje, teda aj interval musí byť širší, aby sme mali väčšiu istotu, že do neho hodnota padne).

p-hodnota je taká hodnota α, pre ktorú je interval tak dlhý, že jeho okraj je na nameranej hodnote charakteristiky.

Napr. ak je nameraná hodnota charakteristiky -2, potom p=0,045. Ak je nameraná hodnota charakteristiky -1,5, potom p=0,133 (viď obrázok nižšie).

p-hodnota udáva najnižšiu možnú hladinu významnosti pre zamietnutie nulovej hypotézy. Ak je **p-hodnota menšia alebo rovná α, potom H_0 zamietame na hladine významnosti α**, ak je p-hodnota väčšia ako α, potom H_0 nezamietame na hladine významnosti α.

Ak je teda napr. p-hodnota menšia ako 0,05, nulovú hypotézu zamietame a hovoríme o štatisticky významnom výsledku na hladine významnosti α = 0,05.

Štatistické programové systémy poskytujú vo svojich výstupoch p-hodnotu. Jej výpočet vyžaduje znalosť distribučnej funkcie rozdelenia, ktorým sa riadi testovacie kritérium, ak je H_0 pravdivá.