Oktatási segédanyag a

Matematikai statisztika

c. tárgyhoz

Maricza István

2002

Tartalomjegyzék

1. fejezet

Valószínűségelméleti alapok

1.1. Alapfogalmak

Eloszlásfüggvény:

$$F(x) = \mathrm{P}\{\omega : X(\omega) < x\} = \mathrm{P}\{X < x\}$$

Farok: Egyoldali: $\mathrm{P}\{X > x\}$, $\mathrm{P}\{X < -x\}$
Kétoldali: $\mathrm{P}\{|X| > x\}$

Sűrűségfüggvény: $f(x) = F'(x)$ (ha létezik)

Várható érték:

$$\mathrm{E}[X] = \int_\Omega X\, dP = \int_{-\infty}^{\infty} x dF(x)$$

Nem mindig létezik (pl. a Cauchy-eloszlás esetén), és lehet $\pm\infty$ is
A „tudatlan statisztikus tétele”:

$$\mathrm{E}[h(X)] = \int_{-\infty}^{\infty} h(x) dF(x)$$

Ha $X \geq 0$, akkor $\mathrm{E}[X] = \int_0^\infty [1 - F(x)]dx$

Markov-egyenlőtlenség:

$$\mathrm{P}\{|X| > \varepsilon\} \leq \frac{\mathrm{E}[|X|]}{\varepsilon}$$

Bizonyítás:

$$\mathrm{E}[|X|] = \int_\Omega |X| dP \geq \int_{\omega:|X(\omega)|>\varepsilon} |X| dP \geq \varepsilon \int_{|X|>\varepsilon} dP = \varepsilon \mathrm{P}\{|X| > x\}$$

Függetlenség: X és Y függetlenek, ha

$$\forall x, y \quad \mathrm{P}\{X < x, Y < y\} = \mathrm{P}\{X < x\}\,\mathrm{P}\{Y < y\}$$

Szórásnégyzet:

$$D^2[X] = \mathrm{E}\left[X - \mathrm{E}\left[X\right]\right]^2 = \mathrm{E}\left[X^2\right] - \left(\mathrm{E}\left[X\right]\right)^2$$

Ha X és Y függetlenek, akkor $D^2[X+Y] = D^2X + D^2Y$ (a szórásnégyzet a második kumuláns)

Csebisev-egyenlőtlenség:

$$\mathrm{P}\left\{|X - \mathrm{E}\left[X\right]| > \varepsilon\right\} \leq \frac{D^2[X]}{\varepsilon^2}$$

Kovariancia

$$\mathrm{Cov}\left[X,Y\right] = \mathrm{E}\left[(X - \mathrm{E}\left[X\right])(Y - \mathrm{E}\left[Y\right])\right] = \mathrm{E}\left[XY\right] - \mathrm{E}\left[X\right]\mathrm{E}\left[Y\right]$$

Ha X és Y függetlenek, akkor $\mathrm{Cov}\left[X,Y\right] = 0$ (X és Y korrelálatlanok). Fordítva nem mindig igaz.

Korreláció

$$\rho(X,Y) = \frac{\mathrm{Cov}\left[X,Y\right]}{\sqrt{D^2X}\sqrt{D^2Y}}$$

$\rho(\cdot,\cdot)$ skalárszorzat az $L^2(P) = \left\{X : \Omega \to \mathbb{R} \mid \int X\,dP = 0, \int X^2\,dP < \infty\right\}$ Hilbert-térben.

Standardizálás Ha $\mu = \mathrm{E}\left[X\right], \sigma^2 = D^2X$, akkor

$$X' = \frac{X - \mu}{\sigma}$$

a *standardizált* változó. Például $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \sqrt{n}\frac{\bar{x}-\mu}{\sigma} \sim N(0,1)$. Amennyiben a μ ismert, de a σ^2 nem, akkor helyette a becsült értéket használjuk:

$$t = \sqrt{n}\,\frac{\bar{x} - \mu}{s_n^*} = \sqrt{n}\frac{\bar{x} - \mu}{\sigma}\sqrt{\frac{\sigma^2}{s_n^{*\,2}}} \sim \frac{N(0,1)}{\sqrt{\frac{\chi^2_{n-1}}{n-1}}}$$

Ez utóbbi eloszlást $n-1$ szabadsági fokú t-eloszlásnak nevezzük, sűrűségfüggvénye

$$f(x) = \frac{1}{\sqrt{2n\pi}}\frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})}\frac{1}{(\frac{x^2}{n}+1)^{\frac{n+1}{2}}}$$

$n \to \infty$ esetén elég gyorsan tart $N(0,1)$-hez. Az 1 paraméterű t_1 eloszlás megegyezik a Cauchy-eloszlással, tehát nincs várható értéke. $\mathrm{E}\left[t_n\right] = 0$, ha $n > 1$ és $D^2[t_n] = \frac{n}{n-2}$, ha $n > 2$.

Nevezetes eloszlások

Diszkrét eloszlások	Eloszlás	Karakterisztikus függvény	Várható érték	Szórásnégyzet
Poisson $\lambda > 0$	$e^{-\lambda}\frac{\lambda^k}{k!}$	$e^{\lambda(e^{iu}-1)}$	λ	λ
Binomiális $n = 1, 2, \ldots$ $0 < p < 1$, $q = 1-p$	$\binom{n}{k} p^k\, q^{n-k}$	$(pe^{iu}+q)^n$	np	$np\,q$
Pascal $0 < p < 1$, $q = 1-p$	pq^k	$\frac{p}{1-qe^{iu}}$	$\frac{q}{p}$	$\frac{q}{p^2}$
Negatív binomiális $r = 1, 2, \ldots$ $0 < p < 1$, $q = 1-p$	$\binom{r+k-1}{k} p^r q^k$	$\left(\frac{p}{1-qe^{iu}}\right)^r$	$\frac{rq}{p}$	$\frac{rq}{p^2}$

Folytonos eloszlások	Sűrűségfüggvény	Karakterisztikus függvény	Várható érték	Szórásnégyzet
Exponenciális $\lambda > 0$	$\lambda e^{-\lambda x}$ ha $x \geq 0$, 0 különben	$\frac{\lambda}{\lambda - iu}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma $r > 0$, $\lambda > 0$	$\lambda e^{-\lambda x}\frac{(\lambda x)^{r-1}}{\Gamma(r)}$ ha $x \geq 0$ 0 különben	$\left(\frac{\lambda}{\lambda - iu}\right)^r$	$\frac{r}{\lambda}$	$\frac{r}{\lambda}$
Egyenletes $[a, b]$-n	$\frac{1}{b-a}$ ha $a < x < b$ 0 különben	$\frac{e^{iub}-e^{iua}}{iu(b-a)}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Normális $\mu \in \mathbb{R}$, $\sigma^2 > 0$	$\frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(x-\mu)^2/2\sigma^2}$	$e^{iu\mu-\sigma^2u^2/2}$	μ	σ^2

Statisztikai eloszlások	Sűrűségfüggvény	Előállítás	Várható érték	Szórásnégyzet
Béta $\alpha, \beta > 0$	$\frac{x^{\alpha-1}x^{\beta-1}}{B(\alpha,\beta)}$ $x \in [0,1]$ 0 különben	$\frac{\Gamma(1,a)}{\Gamma(1,a)+\Gamma(1,b)}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
χ^2 $n \in \mathbb{Z}^+$	$\frac{x^{\frac{n}{2}-1}\exp\left(-\frac{x}{2}\right)}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}$ $x > 0$ 0 különben	$\sum_{i=1}^{n} N(0,1)$	n	$2n$
Student, t $n \in \mathbb{Z}$	$\frac{1}{\sqrt{2n\pi}}\frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})}\frac{1}{(\frac{x^2}{n}+1)^{\frac{n+1}{2}}}$	$\frac{N(0,1)}{\sqrt{\frac{\chi^2_{n-1}}{n-1}}}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
Fisher, F $m, n \in \mathbb{Z}^+$	-	$\frac{\chi^2_{n-1}/(n-1)}{\chi^2_{m-1}/(m-1)}$	$\frac{n}{n-1}$, ha $n > 2$	$\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$ ha $n > 4$

1.2. A nagy számok erős törvénye

1. Definíció. *Azt mondjuk, hogy valószínűségi változók egy ξ_n sorozata* sztochasztikusan *tart egy ξ változóhoz, amennyiben*

$$\forall \varepsilon > 0 \quad P\left(|\xi_n - \xi| > \varepsilon\right) \to 0.$$

2. Definíció. *Azt mondjuk, hogy valószínűségi változók egy ξ_n sorozata* majdnem mindenütt, *vagy más szóval* 1 valószínűséggel *tart egy ξ változóhoz ($\xi_n \to \xi$ m.m., ill. 1 val.), amennyiben*

$$P\left(\omega \in \Omega : \lim_{n\to\infty} \xi_n(\omega) = \xi(\omega)\right) = 1$$

Könnyen látható, hogy az 1 valószínűségű konvergencia maga után vonja a sztochasztikus konvergenciát. A definíció alapján ugyanis minden $\varepsilon > 0$ esetén a $\chi_{\{|\xi_n - \xi| > \varepsilon\}}$ indikátor-változó m.m. 0-hoz tart. Ez a sorozat korlátos, ezért alkalmazható a *Lebesgue*-tétel, vagyis

$$E\chi_{\{|\xi_n - \xi| > \varepsilon\}} = P\left(|\xi_n - \xi| > \varepsilon\right) \to 0.$$

1. Tétel (Kolmogorov). *Ha $X_1, X_2, \ldots$ független, azonos eloszlású változók, melyeknek létezik és véges a közös várható értékük: $EX_i = \mu < \infty$, akkor*

$$\frac{X_1 + \ldots + X_n}{n} \to \mu \ \ 1 \text{ val.}$$

A továbbiakban ennek a tételnek egy kicsit módosított alakját bizonyítjuk: feloldjuk az azonos eloszlásra vonatkozó megszorítást, de cserébe feltételezzük, hogy a várható értékek mellett a szórások is léteznek és végesek, sőt egy szummabilitási kritériumot is előírunk.

2. Tétel (Nagy számok erős törvénye). *Az $X_1, X_2, \ldots$ független valószínűségi változóknak létezik és véges az első két momentuma, valamint teljesül az alábbi azonosság:*

$$\sum_{n=1}^{\infty} \frac{D^2(X_n)}{n^2} < \infty.$$

Ekkor igaz a Nagy Számok Erős Törvénye*:*

$$\frac{1}{n}\sum_{k=1}^{n} (X_k - EX_k) \to 0 \ \ 1 \text{ val.}$$

A továbbiakban az egyszerűség kedvéért (és az általánosság csorbítása nélkül) feltesszük, hogy minden k-ra $EX_k = 0$. Első lépésként bebizonyítunk egy elemi állítást, az ún. *Kronecker*-lemmát.

Lemma: Ha q_n végtelenhez tartó monoton nemnegatív számsorozat és a $\sum_{n=1}^{\infty} a_n$ sor konvergens, akkor $\frac{1}{q_n}\sum_{k=1}^{n} q_k a_k \to 0$.

Bizonyítás: Legyen $R_n = \sum_{k=n}^{\infty} a_k$ a sor maradéktagja. A feltevés miatt $R_n \to 0$. Ekkor minden pozitív ϵ-hoz létezik egy N küszöbindex úgy, hogy $\forall n \geq N$ esetén $|R_n| \leq \epsilon$. Felírjuk a kiszámítandó hányados számlálóját:

$$\sum_{k=1}^{n} q_k\, a_k = \sum_{k=1}^{N-1} q_k\, a_k + \sum_{k=N}^{n} q_k\, a_k = \sum_{k=1}^{N-1} q_k\, a_k + \sum_{k=N}^{n} q_k\, (R_k - R_{k+1})$$

A legutolsó összeget a szokásos módon átalakítjuk:

$$\begin{aligned}\textstyle\sum_{k=N}^{n} q_k\, (R_k - R_{k+1}) = q_N\, R_N - q_N\, R_{N+1} + q_{N+1}\, R_{N+1} - q_{N+1}\, R_{N+2} + \\ + \ldots + q_n\, R_n - q_n\, R_{n+1} = q_N\, R_N + \textstyle\sum_{k=N+1}^{n} R_k\, (q_k - q_{k-1}) - q_n\, R_{n+1}\end{aligned}$$

A számláló abszolút értékét felülről becsüljük:

$$\left|\sum_{k=1}^{n} q_k\, a_k\right| \leq \left|\sum_{k=1}^{N-1} q_k\, a_k + q_N\, R_N\right| + \sum_{k=N+1}^{n} |R_k|\, (q_k - q_{k-1}) + q_n\, |\, R_{n+1}| \leq$$

$$\leq K + \epsilon\, (q_n - q_{N+1}) + q_n \epsilon.$$

Elosztva q_n-nel és n-nel végtelenhez tartva azt kapjuk, hogy

$$\limsup_{n\to\infty} \frac{1}{q_n} \sum_{k=1}^{n} q_k\, a_k \leq 2\epsilon.$$

De ϵ tetszőleges volt, ezért a limesz létezik és 0-val egyenlő.

Alkalmazva a lemmát a $q_n = n$ és $a_n = X_n/n$ helyettesítésekkel azt kapjuk, hogy

$$\sum_{k=1}^{\infty} \frac{X_k}{k} < \infty \text{ 1 val.} \implies \frac{X_1 + \ldots + X_n}{n} \to 0 \text{ 1 val.}$$

Végül belátjuk, hogy a tétel feltevése elégséges feltétel ezen implikáció premisszájának teljesülésére. Ehhez azt kell igazolnunk, hogy ha $Ea_n = 0$ és $\sum_{k=1}^{\infty} Ea_k^2 < \infty$, akkor $S_n = \sum_{k=1}^{n} a_k$ 1 valószínűséggel konvergens. Ez pontosan akkor áll fenn, ha $\forall \epsilon > 0\, \exists N = N(\epsilon)$ úgy, hogy $\forall n > m \geq N$ esetén $|S_n - S_m| < \epsilon$. Felírjuk a komplementer eseményt:

$$\exists \epsilon > 0 \text{ ú.h. } \forall N\text{-re } \exists n > m \geq N \text{ , melyre } |S_n - S_m| \geq \epsilon \tag{1.1}$$

A Csebisev-egyenlőtlenség miatt rögzített n, m és ϵ mellett

$$P\,(|S_n - S_m| \geq \epsilon) \leq \frac{D^2\,(S_n - S_m)}{\epsilon^2} = \frac{\sum_{k=m+1}^{n} Ea_n^2}{\epsilon^2}$$

Ha $n \to \infty$, akkor a bal oldalon a $\{\exists n : |S_n - S_m| \geq \epsilon\}$ esemény valószínűségét, a jobb oldalon a konvergens $\sum Ea_n^2$ sor 0-hoz tartó maradéktagját kapjuk, ezért az $m \to \infty$ határátmenettel az adódik, hogy az $A(\epsilon) = \{\forall N\, \exists n, m \geq N : |S_n - S_m| \geq \epsilon\}$ esemény valószínűsége 0. Ez minden fix ϵ-ra fennáll, amiből könnyen látható, hogy az (1.1) esemény is 0 valószínűségű.

□

1.3. A centrális határeloszlástétel

A nagy számok erős törvénye miatt független, azonos eloszlású X_i változókra $d_n = \frac{X_1+\cdots+X_n}{n} - \mu$ 1 valószínűséggel 0-hoz tart. A d_n mennyiség várható értéke 0 és szórásnégyzete $\frac{\sigma^2}{n}$, ezért a

$$\frac{\sqrt{n}}{\sigma}\left(\frac{X_1+\cdots+X_n}{n}-\mu\right)$$

mennyiség szórásnégyzete minden n-re 1. Ennek a véletlen sorozatnak az eloszlását vizsgáljuk nagy n-re.

3. Definíció. $\xi_n \xrightarrow{d} \xi$ *(eloszlásban), ha*

$$F_{\xi_n}(x) \to F_\xi(x), \forall x \in C(F_\xi),$$

ahol $C(F_\xi)$ *az* F_ξ *folytonossági pontjainak halmaza.*

Cramér-Szlutckij lemma: Ha $X_n \to X$ gyengén és X_n és Y_n távolsága 0-hoz tart, akkor $Y_n \to X$ gyengén.

4. Definíció. Karakterisztikus függvény

$$\varphi_X(t) = \mathrm{E}\left[e^{itX}\right] = \int_{-\infty}^{\infty} e^{itx}\,\mathrm{d}F(x)$$

3. Tétel. Folytonossági tétel

$$\xi_n \xrightarrow{d} \xi \Leftrightarrow \varphi_{\xi_n}(t) \to \varphi_\xi(t)$$

A standard normális eloszlás karakterisztikus függvénye $\varphi_{N(0,1)}(t) = e^{-t^2/2}$.

A 0 körüli sorfejtéssel kifejezhetjük a karakterisztikus függvényt a momentumok segítségével:

$$\varphi_X(t) = \sum_{k=0}^{n} \frac{t^k}{k!}\,\varphi_X^{(k)}(0) + o(t^n)$$

$$\varphi_X'(t)|_{t=0} = \mathrm{E}\left[iX\right] e^{itX}|_{t=0} = i\mathrm{E}\left[X\right], \text{ illetve általában}$$

$$\varphi_X^{(k)}(t)|_{t=0} = \mathrm{E}\left[iX\right]^k e^{itX}|_{t=0} = i^k\mathrm{E}\left[X^k\right].$$

Tehát

$$\varphi_X(t) = \sum_{k=0}^{\infty} \frac{(it)^k}{k!}\mathrm{E}\left[X^k\right] + o(t^{n+1}).$$

Speciálisan ha $\mu = 0, \sigma = 1, \Psi(t) = Ee^{itX_1}$, akkor

$$\begin{aligned}
\varphi_{\frac{X_1+\cdots+X_n}{\sqrt{n}}}(t) &= Ee^{it\frac{X_1+\cdots+X_n}{\sqrt{n}}} = \Psi\left(\frac{t}{\sqrt{n}}\right)^n = \\
&= \left[1 + i\frac{t}{\sqrt{n}}\mathrm{E}\left[X_1\right] - \left(\frac{t}{\sqrt{n}}\right)^2 \frac{1}{2}\mathrm{E}\left[X_1\right]^2 + o(t^2)\right]^n = \\
&= \left[1 - \frac{t^2}{2n} + o(t^2)\right]^n \to e^{-t^2/2}
\end{aligned}$$

Ezzel bebizonyítottuk a *Centrális Határeloszlástételt*:

4. Tétel. *Ha független, azonos eloszlású X_i változókra* $\mathrm{E}[X_1] = \mu < \infty$ *és* $D^2[X_1] = \sigma^2 < \infty$, *akkor*

$$\frac{\sqrt{n}}{\sigma}\left(\frac{X_1 + \cdots + X_n}{n} - \mu\right) \xrightarrow{d} \mathrm{N}(0,1)$$

1.4. A multinormális eloszlás

Egy X valós értékű valószínűségi változót μ és σ^2 paraméterű normális eloszlásúnak nevezünk (röviden $X \sim N\left(\mu, \sigma^2\right)$), ha eloszlása abszolút folytonos és sűrűségfüggvénye

$$\frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad \forall x \in \mathbb{R}.$$

Könnyen látható, hogy $EX = \mu$ és $D^2(X) = \sigma^2$. A továbbiakban ezen eloszlás többdimenziós általánosításával foglalkozunk. Induljunk ki egy független, azonos $N(0,1)$ eloszlású változókból álló $\mathbf{X} = (X_1, \ldots, X_n)^T$ n dimenziós vektorváltozóból és legyenek $A \in \mathbb{R}^{k\times n}$ és $\boldsymbol{\mu} \in \mathbb{R}^k$ tetszőleges determinisztikus mátrixok ($k \le n$).

5. Definíció. *Az* $\mathbf{Y} = A\mathbf{X} + \boldsymbol{\mu}$ *alakú valószínűségi változókat* k dimenziós normális eloszlású*nak, másszóval* multinormális*nak nevezzük.*

A definícióból következik, hogy $E\mathbf{Y} = \boldsymbol{\mu}$ és $\mathrm{cov}(\mathbf{Y}) = AA^T =: \Sigma$, ahol a Σ kovariancia-mátrix $k\times k$-dimenziós. Hasonlóan egyszerűen adódnak a következő tulajdonságok:

1. Ha $\mathbf{Y}$ normális és $B \in \mathbb{R}^{l\times k}$, $\mathbf{b} \in \mathbb{R}^l$, akkor $B\mathbf{Y}+\mathbf{b}$ is normális. Speciálisan minden k dimenziós c vektorra $c^T\mathbf{Y}$ normális eloszlású. Érdekes módon ez a tulajdonság jellemzi a multinormális eloszlást: ha minden c vektorra $c^T\mathbf{Y}$ normális, akkor $\mathbf{Y}$ multinormális.

2. Ha adott egy $\boldsymbol{\mu} \in \mathbb{R}^k$ vektor és egy $\Sigma \in \mathbb{R}^{k\times k}$ szimmetrikus pozitív szemidefinit mátrix, akkor létezik olyan $\mathbf{Y}$ k-dimenziós normális változó, melyre $E\mathbf{Y} = \boldsymbol{\mu}$ és $\mathrm{cov}(\mathbf{Y}) = \Sigma$.
   Bizonyítás: Az állítás feltételei miatt Σ sajátértékei nemnegatív valós számok. Diagonalizálva a Σ-t azt kapjuk, hogy $\Sigma = U^T\Lambda U$, ahol U unitér mátrix és $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_k)$. Legyen $A = U^T\sqrt{\Lambda}U$, ahol $\sqrt{\Lambda} = \mathrm{diag}\left(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_k}\right)$. Ekkor $AA^T = \Sigma$ és az $\mathbf{Y} = A\mathbf{X} + \boldsymbol{\mu}$ választással megkapjuk a kívánt változót.

3. Ha $\mathrm{rang}(A) = k$, akkor $\mathbf{Y}$ abszolút folytonos eloszlású és a sűrűségfüggvénye

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{(2\pi)^{k/2}\left(\det AA^T\right)^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^T\left(AA^T\right)^{-1}(\mathbf{y}-\boldsymbol{\mu})\right\},$$

   ahol $\mathbf{y} = (y_1, \ldots, y_k)$.
   A **bizonyítást** először a $k = n$ esetben végezzük el. Felírjuk az $\mathbf{X} =$

$(X_1, \ldots, X_n)^T$ vektorváltozó többdimenziós sűrűségfüggvényét:

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{n} f_{X_i}(x_i) = \frac{1}{\left(\sqrt{2\pi}\right)^k} e^{-\frac{1}{2}\sum_{i=1}^{n} x_i^2} = \frac{1}{(2\pi)^{k/2}} e^{-\frac{1}{2}\mathbf{x}^T\mathbf{x}}, \ \mathbf{x} \in \mathbb{R}^n$$

Általában ha φ: $\mathbb{R}^n \to \mathbb{R}^k$ „kellően szép" függvény, akkor az $\mathbf{Y} = \varphi(\mathbf{X})$ változó sűrűségfüggvényét a többdimenziós integrálok helyettesítéses integrál-formulájának segítségével a következőképpen lehet felírni (ld. például *Vetier: Szemléletes mérték- és valószínűségelmélet, 226. o.*):

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}\left(\varphi^{-1}(\mathbf{y})\right) \cdot \left|\det\left(\varphi^{-1}\right)'(\mathbf{y})\right|$$

Speciálisan az $\mathbf{Y} = A\mathbf{X} + \boldsymbol{\mu}$ esetben $X^T X = (\mathbf{Y} - \boldsymbol{\mu})^T \left(AA^T\right)^{-1} (\mathbf{Y} - \boldsymbol{\mu})$, ezért

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{\left(\sqrt{2\pi}\right)^k |\det A|} \exp\left\{-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^T \left(AA^T\right)^{-1} (\mathbf{y} - \boldsymbol{\mu})\right\}.$$

De $n = k$ esetén $\det A = \left(\det A \cdot \det A^T\right)^{1/2}$, vagyis az állításban szereplő képletet kaptuk.
A $k < n$ esetben az A mátrixot bővítsük ki további $n - k$ darab sorral (ezek összességét nevezzük B mátrixnak), mégpedig úgy, hogy A és B sorai legyenek páronként egymásra merőleges vektorok (ez mindig megtehető a *Gram-Schmidt ortogonalizáció* segítségével): $AB^T = BA^T = 0$. Az egyszerűség kedvéért tegyük fel, hogy $\boldsymbol{\mu} = \mathbf{0}$.

Ekkor a $C = \begin{bmatrix} A \\ B \end{bmatrix}$ mátrix $n \times n$ dimenziós teljes rangú mátrix, ezért az $\tilde{\mathbf{Y}} = \begin{bmatrix} \mathbf{Y} \\ \bar{\mathbf{Y}} \end{bmatrix} = \begin{bmatrix} A \\ B \end{bmatrix} \mathbf{X}$ vektor sűrűségfüggvénye

$$f_{\tilde{\mathbf{Y}}}(\tilde{\mathbf{y}}) = \frac{1}{\left(\sqrt{2\pi}\right)^k (\det CC^T)^{1/2}} \exp\left\{-\frac{1}{2}\tilde{\mathbf{y}}^T \left(CC^T\right)^{-1} \tilde{\mathbf{y}}\right\},$$

ahol $\tilde{\mathbf{y}} = (\mathbf{y}, \bar{\mathbf{y}}) \in \mathbb{R}^n$. De a merőlegesség miatt $CC^T = \begin{bmatrix} AA^T & 0 \\ 0 & BB^T \end{bmatrix}$ és

$$\left(CC^T\right)^{-1} = \begin{bmatrix} \left(AA^T\right)^{-1} & 0 \\ 0 & \left(BB^T\right)^{-1} \end{bmatrix},$$

valamint $\det CC^T = \det AA^T \cdot \det BB^T$. Azt kapjuk, hogy

$$\tilde{\mathbf{y}}^T \left(CC^T\right)^{-1} \tilde{\mathbf{y}} = \mathbf{y}^T \left(AA^T\right)^{-1} \mathbf{y} + \bar{\mathbf{y}}^T \left(BB^T\right)^{-1} \bar{\mathbf{y}},$$

vagyis az $\tilde{\mathbf{Y}}$ sűrűségfüggvénye szorzótényezőkre bomlik: $f_{\tilde{\mathbf{Y}}}(\tilde{\mathbf{y}}) = f_{\mathbf{Y}}(\mathbf{y}) \cdot f_{\bar{\mathbf{Y}}}(\bar{\mathbf{y}})$. Innen beazonosíthatjuk az eredeti $\mathbf{Y}$ változó sűrűségfüggvényét:

$$f_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{\left(\sqrt{2\pi}\right)^k (\det AA^T)^{1/2}} \exp\left\{-\frac{1}{2}\mathbf{y}^T \left(AA^T\right)^{-1} \mathbf{y}\right\}$$

A fenti előállításból az is következik, hogy a k dimenziós normális eloszlást meghatározza a $\boldsymbol{\mu}$ várható érték vektora és a Σ kovariancia-mátrixa.

Az eddigiek alapján látható, hogy a multinormális eloszlás peremeloszlásai, speciálisan az egydimenziós eloszlások is normálisak. Természetesen ha egy kétdimenziós vektorváltozó peremeloszlásai normálisak, abból még nem következik, hogy együttesen is normálisak lennének. Például ha $U, V \sim N(0,1)$ függetlenek és $Z = \begin{cases} +U, & \text{ha } U \cdot V \geq 0 \\ -U, & \text{ha } U \cdot V < 0 \end{cases}$, akkor U és Z azonos normális eloszlásúak, de együttesen az eloszlásuk nem normális.

A továbbiakban megvizsgáljuk az együttesen normális változók összefüggési tulajdonságait.

Tétel: Legyen $\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}$ normális $E\begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}$ várható érték vektorral és $\operatorname{cov}(X) = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$ kovariancia-mátrixszal. (Itt a mátrixok elemei maguk is mátrixok, tehát ún. blokkmátrixokkal dolgozunk. Ezekkel ugyanúgy lehet számolni, mint a közönséges mátrixokkal.) Ekkor $\mathbf{X_1}$ és $\mathbf{X_2}$ pontosan akkor függetlenek, ha $\Sigma_{12} = 0$, vagyis a változók korrelálatlanok.
A **bizonyítás** az eddigiek alapján nagyon egyszerű. Ha $\Sigma > 0$ (pozitív definit), akkor van sűrűségfüggvény, ami Σ blokkdiagonális volta miatt ($\Sigma_{12} = \Sigma_{21} = 0$) szorzat alakú, tehát $\mathbf{X_1}$ és $\mathbf{X}_2$ függetlenek. Ha Σ nem teljes rangú, akkor $\mathbf{X}_1$, illetve $\mathbf{X}_2$ lineárisan független koordinátáinak vektorait jelölje rendre $\mathbf{Y}_1$ és $\mathbf{Y}_2$. A többi komponens ezen koordináták lineáris kombinációja: $\mathbf{X}_i = A_i\mathbf{Y}_i + \mathbf{b}_i$, $i = 1, 2$. Innen könnyen látható, hogy $\mathbf{X}_1$-re és $\mathbf{X}_2$-re pontosan akkor igaz a tétel, ha $\mathbf{Y}_1$-re és $\mathbf{Y}_2$-re igaz. Ez a feltétel pedig teljesül, ha $\Sigma_{\mathbf{Y}} = \operatorname{cov}(\mathbf{Y}_1, \mathbf{Y}_2)$ pozitív definit. Ennek belátását az Olvasóra bízzuk.

A következő lemma a multinormális változók feltételes eloszlásainak vizsgálatakor nyújt segítséget.

Schur lemmája: Legyen $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$ olyan blokkmátrix, melyre Σ_{22} invertálható. Ekkor Σ „diagonalizálható”:

$$\begin{bmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{bmatrix} \cdot \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \cdot \begin{bmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{bmatrix}^T = \begin{bmatrix} \Sigma_{11\cdot 2} & 0 \\ 0 & \Sigma_{22} \end{bmatrix},$$

ahol $\Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$ és $J = \begin{bmatrix} I & -\Sigma_{12}\Sigma_{22}^{-1} \\ 0 & I \end{bmatrix}$ ortogonális mátrix.

A lemma egyszerű következménye a következő

Állítás: Ha $\mathbf{X}_1$ és $\mathbf{X}_2$ együttesen normális, akkor $\mathbf{X}_2$ és $\mathbf{X}_1 - \Sigma_{12}\Sigma_{22}^{-1}\mathbf{X}_2$ függetlenek.

Bizonyítás: Ha $\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ kovariancia-mátrixa Σ, akkor $J\mathbf{X}$ kovariancia-mátrixa a Schur-lemma miatt blokk-diagonális, vagyis $J\mathbf{X}$ normális eloszlású komponensei korrelálatlanok, tehát függetlenek.

Az utolsó tételünk az $\mathbf{X}_1$ k dimenziós változó az $\mathbf{X}_2$ $n-k$ dimenziós változóra vonatkozó feltételes eloszlásáról szól:

Tétel: Ha $\mathbf{X}$ normális és a Σ kovariancia-mátrix pozitív definit, akkor $\mathbf{X}_1$ feltételes eloszlása az $\mathbf{X}_2 = \mathbf{x}_2$ feltétel mellett $N_k\left(\boldsymbol{\mu}_{1\cdot 2}, \Sigma_{11\cdot 2}\right)$ eloszlású, ahol

$$\boldsymbol{\mu}_{1\cdot 2} = \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}\left(\mathbf{x}_2 - \boldsymbol{\mu}_2\right).$$

Bizonyítás: Az előző állítás szerint $\mathbf{X}_2 - \boldsymbol{\mu}_2$ és $\mathbf{X}_1 - \boldsymbol{\mu}_1 - \Sigma_{12}\Sigma_{22}^{-1}\left(\mathbf{X}_2 - \boldsymbol{\mu}_2\right)$ függetlenek, ezért $\mathbf{X}_1 - \boldsymbol{\mu}_{1\cdot 2}$ eloszlása az $\mathbf{X}_2 = \mathbf{x}_2$ feltétel mellett ugyanaz, mint a feltétel nélkül. $\begin{bmatrix} \mathbf{X}_1 - \boldsymbol{\mu}_{1\cdot 2} \\ \mathbf{X}_2 - \boldsymbol{\mu}_2 \end{bmatrix} = J \cdot (\mathbf{X} - \boldsymbol{\mu})$, ezért $\operatorname{cov}\left(\mathbf{X}_1 - \boldsymbol{\mu}_{1\cdot 2}\right) = \Sigma_{11\cdot 2}$.

1.5. A multinomiális eloszlás

Bevezetésként emlékeztetünk a valószínűségszámításból jól ismert binomiális eloszlásra. Egy kísérletet elvégzésénél a siker valószínűsége legyen p. Végezzünk n darab független kísérletet és jegyezzük fel a sikerek számát, ez legyen Q. Ennek eloszlása

$$\mathrm{P}\{Q = k\} = \frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}, \quad k = 0, \ldots, n$$

A *multinomiális* (másnéven *polinomiális*) eloszlás ennek általánosítása arra az esetre, amikor a kísérletnek kettőnél több lehetséges kimenetele van.

Legyen az $A_1, \ldots, A_r$ teljes eseményrendszer egy kísérlet lehetséges kimeneteleinek halmaza és jelölje $X_i^{(n)} \quad (i = 1, \ldots, r)$ azt, hogy n kísérletből hányszor következett be az A_i esemény, melynek valószínűsége p_i. Nyilván $X_1^{(n)} + \cdots + X_r^{(n)} = n$. (A továbbiakban a felső indexet az egyszerűség kedvéért elhagyjuk.)

Az X_i változók együttes eloszlása kombinatorikai megfontolások alapján könnyen meghatározható. Ha $k_1 + \cdots + k_r = n$, akkor

$$\mathrm{P}\{X_1 = k_1; \ldots; X_r = k_r\} = \frac{n!}{k_1! \ldots k_r!}\, p_1^{k_1} \cdots p_r^{k_r}, \tag{1.2}$$

illetve különben 0. Az együttes eloszlásból visszakapjuk, hogy a marginálisok binomiális eloszlásúak és első két kumulánsuk $EX_i = np_i$ és $D^2X_i = np_i(1-p_i)$. Hasonlóan adódik a párok együttes eloszlása is. $i \neq j$ esetén

$$\mathrm{P}\{X_i = k_i; X_j = k_j\} = \frac{n!}{k_i!\, k_j!\, (n-k_i-k_j)!}\, p_i^{k_i}\; p_j^{k_j}\; (1-p_i-p_j)^{n-k_i-k_j} \tag{1.3}$$

Ennek alapján könnyen kiszámítható a szorzat várható értéke: $EX_iX_j = n(n-1)p_ip_j$, vagyis

$$\operatorname{Cov}(X_i, X_j) = -np_ip_j \tag{1.4}$$

A negatív korreláltság nem meglepő, hiszen minél többször fordul elő egy kimenetel, annál kevesebbszer fordulhat elő egy másik (mivel az összes kísérlet száma adott) és a kovariancia arányos a várható előfordulási számmal.

Képezzük az $\mathbf{X} = (X_1, \ldots, X_r)^T$ vektorváltozót. Ennek várható értéke $E\mathbf{X} = n(p_1, \ldots, p_r)^T$ és kovariancia-mátrixa

$$\mathrm{Cov}(\mathbf{X}) = n \cdot \begin{pmatrix} p_1(1-p_1) & -p_1p_2 & \ldots & -p_1p_r \\ -p_2p_1 & p_2(1-p_2) & \ldots & -p_2p_r \\ . & . & . & . \\ . & . & . & p_r(1-p_r) \end{pmatrix} \tag{1.5}$$

Könnyen látható, hogy a standardizált $\mathbf{Z}_n = \frac{1}{\sqrt{n}}(\mathbf{X}^{(n)} - E\mathbf{X}^{(n)})$ vektorváltozó $0 \in \mathbb{R}^r$ várható értékű és kovariancia-mátrixa a (1.5) egyenletben n mellett szereplő mátrix, melyet jelöljünk S-sel.
A $\mathbf{Z}_n$ karakterisztikus függvényének sorfejtésével bebizonyítható, hogy $n \to \infty$ esetén $\mathbf{Z}_n$ eloszlása $(0, \mathrm{S})$ paraméterű r-dimenziós multinomális eloszláshoz tart.

Az S mátrix struktúrája nagyon egyszerű:

$$S = \mathrm{diag}(p_1, \ldots, p_r) - (p_1, \ldots, p_r)^T(p_1, \ldots, p_r) \tag{1.6}$$

A rang meghatározásához szorozzuk meg előlről és hátulról az

$$A = \mathrm{diag}\left(\frac{1}{\sqrt{p}_1}, \ldots, \frac{1}{\sqrt{p}_r}\right)$$

invertálható mátrixszal:

$$ASA^T = I - \mathbf{a}\mathbf{a}^T, \tag{1.7}$$

ahol I az $r \times r$-es egységmátrix és $\mathbf{a} = (\sqrt{p_1}, \ldots, \sqrt{p_r})^T$ olyan vektor, melyre $\mathbf{a}^T\mathbf{a} = 1$. Így

$$\frac{1}{\sqrt{n}}\, A\,(\mathbf{X^{(n)}} - E\mathbf{X^{(n)}}) = \left(\frac{X_1 - np_1}{\sqrt{np_1}}, \ldots, \frac{X_r - np_r}{\sqrt{np_r}}\right) \xrightarrow{d} N_r(0,\ I - \mathbf{a}\mathbf{a}^T) \tag{1.8}$$

A kovariancia-mátrixról könnyen belátható, hogy a 0 egyszeres, az 1 $(r-1)$-szeres sajátértéke, vagyis felírható

$$I - \mathbf{a}\mathbf{a}^T = U\,J\,U^T$$

diagonalizált alakban, ahol U ortogonális mátrix és

$$J = \begin{pmatrix} 1 & \ldots & 0 & 0 \\ & \ddots & & \\ 0 & \ldots & 1 & 0 \\ 0 & \ldots & 0 & 0 \end{pmatrix}$$

A (1.8) egyenletben szereplő vektort elforgatva, vagyis megszorozva az U^T ortogonális mátrixszal az eredetivel azonos hosszúságú, de 0 várható értékű és J aszimptotikus kovariancia-mátrixú normális eloszlású vektort kapunk. A multinormális eloszlás tulajdonságai alapján ez a vektor $(\nu_1, \ldots, \nu_{r-1}, 0)^T$ alakú, ahol $\nu_1, \ldots, \nu_{r-1}$ *független* egydimenziós standard normális eloszlású változók. Következésképpen

$$\sum_{j=1}^{r} \frac{(X_j - np_j)^2}{np_j} \stackrel{d}{=} \nu_1^2 \cdots + \nu_{r-1}^2 \stackrel{d}{=} \chi^2_{r-1} \tag{1.9}$$

2. fejezet

A statisztika alapfogalmai

A statisztika a tömegjelenségek leírására valószínűségelméleti modellt használ. Tegyük fel, hogy egy megfigyelt valós értékű X mennyiség több nem feltétlenül ismert tényezőtől függhet, melyeknek összes lehetséges kimenetele egy Ω halmazt alkot. Ekkor a mennyiség úgy tekinthető, mint egy

$$X : \Omega \to \mathbb{R}$$

leképezés. Ha a megfigyelés többször megismételhető, akkor érdemes az Ω halmazt valószínűségi struktúrával ellátni. Ekkor az X leképezés ún. *valószínűségi változó*. A változó különböző realizációi az

$$X(\omega_1),\ X(\omega_2), \dots\ X(\omega_n)$$

megfigyelések. A valószínűségszámításban felvetett kérdések nagy része a változók eloszlására vonatkozik, ezért kikerülhetők az Ω alaphalmazra való utalások. Az egyik alapvető technika az, hogy egy változó több különböző realizációját helyettesítjük az eredeti változó több független, azonos eloszlású kópiájának egyetlen realizációjával. Szemléletesen: ha egy pénzdarabot feldobunk ezerszer, az ugyanaz, mint ha ezer egyforma és egymástól független pénzdarabot egyszerre feldobunk egyszer.

A statisztika alapmodellje a *független, azonos eloszlású minta*:

$$X_1(\omega),\ X_2(\omega), \dots, X_n(\omega)$$

Ha magát a tömegjelenséget vizsgáljuk, akkor a realizáció helyett a valószínűségi változók sorozatát tekintjük.

6. Definíció. *Az n elemű* rendezett minta *a minta nagyság szerint rendezett transzformáltja:*

$$X_1^* \le X_2^* \le \dots \le X_n^*$$

Ez a művelet az eredeti mintától függő véletlen permutációját adja a mintának és erős összefüggőséget visz a független mintába (hiszen pl. ha $X_1^* \ge t$, akkor 1 valószínűséggel $X_n^* \ge t$). Ha X_i eloszlása F, akkor

$$\mathrm{P}\{X_n^* < x\} = \mathrm{P}\{X_1 < x, ..., X_n < x\} = F(x)^n$$
$$\mathrm{P}\{X_1^* < x\} = 1 - \mathrm{P}\{X_1^* \geq x\} = 1 - [1 - F(x)]^n$$

Belátható, hogy ha X_i egyenletes eloszlású a $[0,1]$ intervallumon, akkor X_k^* eloszlása Béta$(k, n-k+1)$.

7. Definíció. *Statisztikának nevezzük az $X_1, \ldots, X_n$ minta $T_n(X_1, \ldots, X_n)$ függvényeit.*

Ezek közül a legegyszerűbb a mintaátlag:

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}$$

A Nagy Számok Törvénye miatt ha $\mathrm{E}[X] = \mu$, akkor

$$\bar{X}_n \to \mu \quad 1 \text{ valószínűséggel}$$

A statisztika és a valószínűségelmélet közötti párhuzamokat az alábbi táblázat foglalja össze:

Valószínűségelmélet	Statisztika
Populáció	Minta
Valószínűség	Relatív gyakoriság
Várható érték	Mintaátlag

2.1. Tapasztalati eloszlásfüggvény

Az eloszlásfüggvény statisztikai megfelelője a *tapasztalati eloszlásfüggvény*:

$$F_n^*(x) = \frac{1}{n} \# \{1 \leq k \leq n : X_k < x\}$$

A rendezett minta segítségével ez az alábbi alakba írható:

$$F_n^*(x) = \begin{cases} 0, & \text{ha } x \leq X_1^* \\ \vdots & \vdots \\ k/n, & \text{ha } X_k^* < x \leq X_{k+1}^* \\ \vdots & \vdots \\ 1, & \text{ha } X_n^* < x \end{cases}$$

$F_n^*(x)$ véletlen eloszlásfüggvény, méghozzá $nF_n^*(x) : \Omega \to \{0, 1, \ldots, n\}$ binomiális eloszlású változó:

$$\begin{aligned} \mathrm{P}\{nF_n^*(x) = k\} &= \binom{n}{k} F(x)^k [1 - F(x)]^{n-k} \\ \mathrm{E}[nF_n^*(x)] &= nF(x) \\ D^2[nF_n^*(x)] &= nF(x)[1 - F(x)] \end{aligned}$$

A fentiek alapján $\mathrm{E}\left[F_n^*(x)\right] = F(x)$ és a Nagy Számok Erős Törvénye miatt 1 valószínűséggel

$$F_n^*(x) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i<x\}} \overset{n\to\infty}{\longrightarrow} \mathrm{E}\left[I_{\{X_i<x\}}\right] = \mathrm{P}\left\{X_i < x\right\} = F(x)$$

Ezt úgy fejezzük ki a statisztika nyelvén, hogy a tapasztalati eloszlásfüggvény *torzítatlan* és *erősen konzisztens* becslése az eloszlásfüggvénynek.

Általában, ha $T_n(X_1,\ldots,X_n)$ a θ paraméter becslése, akkor az alábbi tulajdonságokról beszélünk:

Torzítatlanság	$\mathrm{E}\left[T_n(X_1,\ldots,X_n)\right] = \theta$
Konzisztencia	$T_n(X_1,\ldots,X_n) \to \theta$ sztochsztikusan
Erős konzisztencia	$T_n(X_1,\ldots,X_n) \to \theta$ 1 valószínűséggel

5. Tétel. A matematikai statisztika alaptétele (Glivenko-Cantelli)

$$\sup_{x\in\mathbb{R}} |F_n^*(x) - F(x)| \to 0 \quad \textit{1 valószínűséggel}$$

2.2. Tapasztalati szórásnégyzet

A szórásnégyzet becslésére az

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

mennyiséget használjuk. Látható, hogy

$$\mathrm{E}\left[ns_n^2\right] = \sum_{i=1}^{n}\mathrm{E}\left[X_i-\mu\right]^2 - n\mathrm{E}\left[\bar{x}-\mu\right]^2 = n\sigma^2 - n\frac{\sigma^2}{n},$$

ezért $\mathrm{E}\left[s_n^2\right] = (1-\frac{1}{n})\sigma^2 < \sigma^2$, vagyis tapasztalati szórásnégyzet torzított becslése σ^2-nek. Ennek kiküszöbölésére használják az úgynevezett *korrigált tapasztalati szórásnégyzet*et:

$$s_n^{*2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

6. Tétel. Steiner-formula

$$ns_n^2 = \sum_{i=1}^{n}(X_i - a)^2 - n(\bar{X}-a)^2$$

Speciálisan

$$a = 0 \quad \Rightarrow \quad ns_n^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2$$

$$a = \mu \quad \Rightarrow \quad ns_n^2 = \sum_{i=1}^{n}(X_i-\mu)^2 - n(\bar{X}-\mu)^2$$

7. Tétel. Fisher-Bartlett tétel $X_1, ..., X_n \sim N(\mu, \sigma^2) \Rightarrow \bar{X}$ *és* s_n^2 *függetlenek*

A bizonyításhoz az $X = (X_1, \ldots, X_n) \in \mathbb{R}^n$ vektort szorozzuk meg az

$$U = \begin{pmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{n}} & \cdots & & \\ \vdots & & & \\ \frac{1}{\sqrt{n}} & \cdots & & \end{pmatrix}$$

mátrixszal, amelyben az első sor és oszlop adott, a többi elem tetszőleges, feltéve, hogy a mátrix *ortogonális*. Legyen $Y = UX$ az elforgatott vektor. Ekkor $Y_1 = \frac{1}{\sqrt{n}}(X_1 + ... + X_n) = \sqrt{n}\bar{X}$ és $X_1^2 + \ldots + X_n^2 = Y_1^2 + \ldots + Y_n^2$, mivel a forgatás nem változtatja meg a vektorok hosszát. X kovariancia-mátrixa az identitás-mátrix, akárcsak az Y vektoré, hiszen ez sem változik a forgatástól. Az Y multinormális eloszlású, ezért Y koordinátái *függetlenek*. Továbbá

$$ns_n^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2 = \sum_{i=1}^{n} Y_i^2 - Y_1^2 = \sum_{i=2}^{n} Y_i^2$$

Vagyis $\bar{X}$-t Y_1-ből, s_n^2-et $Y_2, \ldots, Y_n$-ből számoljuk, tehát a két statisztika független. Az is látszik, hogy $\frac{ns_n^2}{\sigma^2} \sim \vartheta_1^2 + ... + \vartheta_{n-1}^2$, ahol $\vartheta_i \sim N(0,1)$.

Következmény: Ha $X_1, ..., X_n \sim N(\mu, \sigma^2)$, akkor

$$\begin{aligned} \bar{X}_n &\sim N(\mu, \sigma^2/n) \\ \frac{ns_n^2}{\sigma^2} &\sim \mathcal{X}_{n-1}^2 \end{aligned}$$

A χ^2 eloszlás tulajdonságai:

1. χ_n^2 és χ_m^2 független $\Rightarrow \mathcal{X}_n^2 + \mathcal{X}_m^2 \sim \mathcal{X}_{n+m}^2$
2. $\mathrm{E}\left[\chi_n^2\right] = n$ és $D^2[\chi_n^2] = 2n$
3. $\chi_n^2 = \Gamma(\frac{n}{2}, \frac{1}{2})$, vagyis Gamma eloszlású

2.3. Paraméteres becslési eljárások

2.3.1. Momentummódszer

Ezzel a módszerrel úgy becslünk meg paramétereket, hogy a minta és a becsült eloszlás első momentumai megegyezzenek. Tehát a becsült k. momentum

$$\hat{\mu}_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k.$$

Például ha egy (a, b) intervallumon egyenletes eloszlás két paraméterét kell becsülni az (X_1, X_2) mintából, akkor így járunk el:

$$\mathrm{E}\left[E(a,b)\right] = \frac{a+b}{2}, D^2E(a,b) = \frac{(b-a)^2}{12},$$

$$\frac{X_1+X_2}{2} = \frac{a+b}{2}, \frac{(X_1-X_2)^2}{2} = \frac{(b-a)^2}{12},$$

$$X_1+X_2 = a+b, 6(X_1-X_2)^2 = (b-a)^2.$$

Ezt az egyenletrendszert kell megoldanunk az (a,b) ismeretlenekre.

A momentummódszer általánosítható olyan irányban, hogy a momentumok helyett a minta és a paraméteres eloszlás más funkcionáljait válasszuk. Ez az általánosított momentummódszer erősen konzisztens, de nem biztos, hogy torzítatlan.

2.3.2. Maximum likelihood módszer

Ez a módszer azt a paraméterértéket keresi meg, amely az adott minta mellett a legvalószínűbb, illetve folytonos esetben azt, amelynek a sűrűsége a legnagyobb. Ehhez a likelihood függvényt (jelölésben $L(p,X)$) használjuk, amely független, azonos eloszlású minta esetén folytonos eloszlás esetén az együttes sűrűségfüggvény, diszkrét esetben az együttes eloszlás.

A maximumkeresést általában deriválással végezzük, azonban ezt ellenőrizni kell, hogy valóban globális maximumot ad-e az $\frac{\partial L(p,X)}{\partial p} = 0$ egyenlet megoldása.

Folytonos esetben az együttes sűrűségfüggvény szorzat alakban áll elő. Ennek kezelése helyett általában a loglikelihood függvényt (jelölésben $l(p,X) = \log\left(L(p,X)\right)$) deriváljuk, mert a két függvény globális maximuma úgyis egybeesik, és a szorzat összegre esik szét, amely tagonként deriválható.

Például ha egy $X_1,\ldots,X_n$ minta λ paraméterű Poisson eloszlásból származik, akkor a paramétert a következő módon becsüljük:

$$L(\lambda,X) = \mathrm{P}\left\{\mathrm{Poisson}(\lambda) = X_1,\ldots,\mathrm{Poisson}(\lambda) = X_n\right\} =$$
$$\mathrm{P}\left\{\mathrm{Poisson}(\lambda) = X_1\right\}\ldots\mathrm{P}\left\{\mathrm{Poisson}(\lambda) = X_n\right\} =$$
$$= e^{-\lambda}\frac{\lambda^{X_1}}{X_1!}\ldots e^{-\lambda}\frac{\lambda^{X_n}}{X_n!} = e^{-\lambda n}\frac{\lambda^{\sum_{j=1}^n X_j}}{X_1!\ldots X_n!}.$$

$$l(\lambda,X) = -n\lambda + \log(\lambda)\sum_{j=1}^n X_j - \sum_{j=1}^n \log X_j!$$

$$0 = \frac{\partial l(\hat{\lambda},X)}{\partial\hat{\lambda}} = -n + \frac{\sum_{j=1}^n X_j}{\hat{\lambda}}, \text{ azaz}$$

$$\hat{\lambda} = \frac{\sum_{j=1}^n X_j}{n} = \bar{X}_n$$

Exponenciális mintánál a likelihood függvény az együttes sűrűségfüggvény:

$$L(\lambda,X) = \prod_{j=1}^n \lambda e^{-\lambda X_j}$$

$$l(\lambda, X) = \sum_{j=1}^{n} (\log(\lambda) - \lambda X_j) = n \log(\lambda) - \lambda \sum_{j=1}^{n} X_j$$

$$\frac{\partial l(\hat{\lambda}, X)}{\partial \hat{\lambda}} = \frac{n}{\hat{\lambda}} - \sum_{j=1}^{n} X_j = 0.$$

Ebből a becsült paraméter

$$\hat{\lambda} = \frac{n}{\sum_{j=1}^{n} X_j} = \frac{1}{\bar{X}_n}$$

2.4. Elégséges statisztikák

8. Definíció. *Egy statisztika elégséges, ha minden információt tartalmaz a mintáról, azaz a minta egy halmazba esésének valószínűsége nem függ a paramétertől, feltéve, ha ismerjük ezt a statisztikát.*

8. Tétel (Neyman-féle faktorizációs tétel). *Egy $T(X)$ statisztika elégséges a p paraméterre, ha a likelihood függvény felbomlik*

$$L(p, X) = g_p(T(X)) \cdot h(X)$$

alakban, ahol a h függvény nem függ a paramétertől.

9. Tétel (Rao-Blackwell). *Ha adott egy T torzítatlan becslése egy p paraméternek, akkor az S elégséges statisztikára vett $\mathrm{E}[T|S]$ feltételes várható érték is torzítatlan becslése lesz p-nek, szórásnégyzete nem lesz nagyobb, mint a T becslésé, továbbá a feltételben szereplő elégséges statisztika miatt nem függ p-től.*

Például a Poisson eloszlás fentebb kiszámolt likelihood függvényéből

$$L(\lambda, X) = e^{-\lambda n} \frac{\lambda^{\sum_{j=1}^{n} X_j}}{X_1! \dots X_n!}$$

megállapítható, hogy elégséges statisztika lesz például a $\sum_{j=1}^{n} X_j$ vagy a mintaátlag $\bar{X}_n$.

3. fejezet

Hipotézisvizsgálat

3.1. Konfidencia-intervallumok

Amikor egy paramétert a minta alapján megbecsülünk, a becslés torzítatlansága mellett azt is tudni szeretnénk, hogy mekkora a szórás, vagyis a nagy átlagban elért várható értéktől mennyire térhet el a becslés. (Sokszor a gyakorlatban $\mu \pm \sigma$ alakban adják meg ezeket a leírókat.) Minél kisebb a szórás, a becslés annál jobban lokalizált. Az ún. *konfidencia-intervallumok*, másnéven *intervallum-becslések* ezt az összefüggést számszerűsítik.

9. Definíció. *A (T_1, T_2) véletlen intervallumot a ϑ paraméter α-szintű konfidencia-intervallumának nevezzük ha*

$$\mathrm{P}\left\{T_1 < \vartheta < T_2\right\} = 1 - \alpha$$

A konfidencia-intervallumok megadásának általános módszere:

1. Keresünk egy ismert eloszlású statisztikát
2. Felírunk egy intervallumot, ahová ez nagy valószínűséggel esik
3. Kifejezzük a paramétert

Példák

1. $X_1, \ldots, X_n$ $N(\mu, \sigma^2)$ eloszlású minta, σ^2 ismert. μ-re szeretnénk konfidencia-intervallumot adni. Tudjuk, hogy

$$\sqrt{n}\,\frac{\bar{X} - \mu}{\sigma} \sim N(0,1). \tag{3.1}$$

Szimmetrikus intervallumot keresünk, mivel a normális eloszlás szimmetrikus. Valamely u-ra

$$\mathrm{P}\left\{-u < \sqrt{n}\,\frac{\bar{X} - \mu}{\sigma} < u\right\} = 1 - \alpha$$

Jelölje Φ a normális eloszlásfüggvény eloszlásfüggvényét. Ekkor a baloldal

$$\Phi(u) - \underbrace{\Phi(-u)}_{1-\Phi(u)} = 2\Phi(u) - 1,$$

vagyis $\Phi(u) = 1 - \frac{\alpha}{2}$, melynek megoldása

$$u_{\alpha/2} = \Phi^{-1}\left[1 - \frac{\alpha}{2}\right],$$

az $\alpha/2$-höz tartozó *kritikus érték*. Innen kifejezve a μ-t kapjuk az $1-\alpha$ szintű konfidencia-intervallumot:

$$\mathrm{P}\left\{\bar{X} - u_{\alpha/2}\,\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + u_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}\right\} = 1 - \alpha$$

A 95%-os szinthez tartozó együttható például $u_{0.025} \approx 1.96$.

2. Ha az előző példában nem tesszük fel a normalitást, az (3.1)egyenletben egyenlőség helyett a baloldal a centrális határeloszlástétel miatt tart a jobboldalhoz, ha $n \to \infty$. Ekkor ún. *aszimptotikus* konfidencia-intervallumot kapunk, amely nagy n-ekre, $1-\alpha$ valószínűséggel tartalmazza az ismeretlen paramétert.

3. Legyen most az első példában a μ ismert és a σ^2 ismeretlen. Ez utóbbira keressünk konfidencia-intervallumot. Tudjuk, hogy $\dfrac{X_i - \mu}{\sigma} \sim N(0,1)$, ezért

$$\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \mathcal{X}_n^2$$

Felírunk egy intervallumot:

$$\mathrm{P}\left\{\underline{X}_\alpha^2 < X_\alpha^2 < \overline{X}_\alpha^2\right\} = 1 - \alpha$$

Innen kifejezhetjük a σ^2-et:

$$\mathrm{P}\left\{\underline{X}_\alpha^2 < \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2 < \overline{X}_\alpha^2\right\} =$$

$$= \mathrm{P}\left\{\frac{1}{\overline{X}^2}\sum_{i=1}^{n}(X_i - \mu)^2 < \sigma^2 < \frac{1}{\underline{X}^2}\sum_{i=1}^{n}(X_i - \mu)^2\right\} = 1 - \alpha$$

Hogyan határozzuk meg a $\underline{X}_\alpha^2$ és $\overline{X}_\alpha^2$ határokat? Legjobb lenne úgy, hogy a legszűkebb legyen, vagyis $\mathrm{E}\left[T_2 - T_1\right] \to \min$. Ennél egyszerűbb, ha úgy választjuk, hogy a két farok egyenlő súlyú legyen:

$$\mathrm{P}\left\{\chi^2 < \underline{\chi}_\alpha^2\right\} = \mathrm{P}\left\{\chi^2 > \overline{\chi}_\alpha^2\right\} = \alpha/2$$

4. Tegyük fel, hogy az első példában mindkét paraméter ismeretlen és μ-re keresünk konfidencia-intervallumot. Ekkor σ^2 ún. „zavaró” paraméter (nuisance parameter). Ezen feltételekkel

$$\sqrt{n}\,\frac{\bar{X} - \mu}{s_n^*} \sim t_{n-1}$$

Ekkor az első példához hasonlóan

$$\mathrm{P}\left\{-t_{\alpha/2} < \frac{\bar{X}-\mu}{s_n^*}\sqrt{n} < t_{\alpha/2}\right\} = 1-\alpha,$$

vagyis

$$\mathrm{P}\left\{\mu \in [\bar{X} - t_{\alpha/2}\frac{s_n^*}{\sqrt{n}}, \bar{X} + t_{\alpha/2}\frac{s_n^*}{\sqrt{n}}]\right\} = 1-\alpha$$

3.2. Statisztikai próbák alapfogalmai

A hipotézisvizsgálat a minta eloszlására (annak paraméterére vagy alakjára) vonatkozó feltevés ellenőrzése az adatok alapján. A felhasznált matematikai modell a *statisztikai mező*: $(\Omega, \mathcal{A}, P_\vartheta)$, $\vartheta \in \Theta$ az ismeretlen paraméter.

Van egy elképzelésünk a valódi paraméterekről: $\vartheta \in \Theta_n \subset \Theta$, a null-hipotézis ($H_0$). Azt vizsgáljuk, hogy ez összhangban van-e az adatokkal. A próbákat úgy konstruálják, hogy ha csak lehet, elfogadják a nullhipotézist és csak akkor utasítják el, ha az adatok „nagyon tiltakoznak" ellene. Ezért célszerű nullhipotézisként olyasmit feltenni, amiről reméljük, hogy a próba el fogja utasítani. Ez ugyanis a szignifikáns eredmény.

A minta alapján el kell tudnunk dönteni, hogy a null-hipotézist elfogadjuk-e vagy pedig elvetjük. *Ellen-hipotézis* (H_1) a paramétertérnek az a része, melynél azt szeretnénk, hogy a próba elutasító választ adjon. Ez gyakran a H_0 komplementere, de nem mindig. A próba tehát egy $t : \mathbb{R}^n \to 0, 1$ leképezés, amely minden n elemű mintára megmondja, hogy annak alapján a null-hipotézist elfogadjuk, avagy elvetjük. A minták tere két diszjunkt halmaz úniójára bomlik:

$$\mathbb{R}^n = C_0 \cup C_1,$$

ahol $C_0 \subseteq \mathbb{R}^n$ az *elfogadási tartomány* és $C_1 \subseteq \mathbb{R}^n$ a *kritikus tartomány*.

A próbák legtöbbször olyanok, hogy egy ún. *próbastatisztika* alapján döntünk:

$$\begin{aligned} C_0 &= \{(X_1,\ldots,X_n) \in \mathbb{R}^n \quad : \quad T_n(X_1,\ldots,X_n) \in [a,b]\} \\ C_1 &= \{(X_1,\ldots,X_n) \in \mathbb{R}^n \quad : \quad T_n(X_1,\ldots,X_n) \notin [a,b]\} \end{aligned}$$

A hipotézisvizsgálat során a következő esetek fordulhatnak elő:

	H_0-t elfogadjuk	H_0-t elvetjük
H_0 igaz	$\emptyset$	I. fajú hiba
H_1 igaz	II. fajú hiba	Szignifikáns eredmény

A próbákat úgy konstruáljuk meg, hogy pozitívnak nevezzük azt az eredményt, hogy a nullhipotézist elvetjük. Ekkor a következőket mondhatjuk:

1. Ha H_0 igaz és elfogadjuk, akkor hibát ugyan nem követünk el, de az eredmény nem szignifikáns.

2. Az elsőfajú hiba elkövetésével álpozitív eredményt kapunk. Ezt igyekszünk elsősorban elkerülni úgy, hogy csak olyan próbákat tekintünk, melyekre ennek valószínűsége kicsi.

3. A másodfajú hiba elkövetésével álnegatív eredményt kapunk, vagyis nem vesszük észre a pozitív eredményt. Minél kevésbé követi el egy próba ezt a fajta hibát, annál erősebbnek nevezzük.

4. Szignifikáns eredményt kapunk, ha helyesen elvetjük a nullhipotézist.

Olyan próbákat keresünk tehát, hogy egy előre rögzített kis α esetén

$$\mathrm{P}\{C_1|H_0\} \leq \alpha,$$

és ezen feltétellel $\mathrm{P}\{C_0|H_1\}$ minimális.

10. Definíció. *A próba terjedelme a legnagyobb megengedett elsőfajú hiba:*

$$\sup_{\vartheta \in \Theta_0} \mathrm{P}_\vartheta\{C_1|H_0\} = \alpha$$

A próba szintje (gyakran %-ban kifejezve): $1 - \alpha$
Erőfüggvény:

$$\gamma(\theta) = \mathrm{P}\{H_0\text{-}t \text{ elvetjük} \mid H_0 \text{ hamis és a paraméter értéke } \theta\}$$

A próbával szemben támasztott követelmények:

torzítatlanság $\gamma(\theta) \geq \alpha$, vagyis nagyobb valószínűséggel vessük el a nullhipotézist akkor, ha hamis, mint akkor, ha igaz

konzisztencia $\lim_{n\to\infty} \gamma_n(\theta) = 1$, vagyis a minta növekedésével egyre biztosabban döntsünk

Minél kisebb a terjedelem, annal nagyobb az elfogadási tartomány, mert hiszen ha nagyon megszorítjuk az elsőfajú hibát, akkor a próba „nem meri" elvetni a nullhipotézist, vagyis mindent elfogad. Előfordulhat, hogy egy adott szinten elfogadjuk, egy másik szinten elutasítjuk a nullhipotézist. Van egy olyan terjedelem, amelynél a próbastatisztika pontosan az elfogadási tartomány határán található. Ezt nevezzük *p-érték*nek. Általában az egyes alkalmazásoknak egy tipikus terjedelemmel dolgoznak (pl. 0.01, 0.05 vagy 0.1). A számítógépes gyakorlatban általában egy p-értéket kapunk és a következő szabály szerint döntünk:

$\alpha < p$	elfogadunk
$\alpha \geq p$	elutasítunk

Ha a p-érték nagyon kicsi (vagy nagyon nagy), akkor mondhatjuk, hogy minden használatos szinten elutasítunk (illetve elfogadunk).

3.3. Klasszikus paraméteres próbák

A statisztikai próbák megadásának általános módszere:

1. Keresünk egy statisztikát (ez az ún. *próbastatisztika*), melynek H_0 teljesülése esetén ismert az eloszlása és az lehetőleg markánsan különbözik a H_1 teljesülése esetén érvényes eloszlástól
2. Felírunk egy (esetleg félig végtelen) intervallumot, ahová ez $1-\alpha$ valószínűséggel esik. Az intervallum komplementere feleljen meg az ellenhipotézisben szereplő paramétertartománynak
3. α lesz a próba terjedelme, a kapott intervallum az elfogadási tartomány

3.3.1. Egymintás u-próba

Hasonlóan a konfidencia-intervallumokhoz elsőként tekintsünk egy független normális eloszlású mintát ismert szórásnégyzettel, vagyis a statisztikai mező legyen $N(\mu, \sigma_0^2)$. A nullhipotézis legyen az, hogy $\mu = \mu_0$. Kétféle ellenhipotézist tekinthetünk: μ különbözik a feltett μ_0-tól (vagyis H_1 a H_0 komplementere) vagy pedig nagyobb nála. Az első esetben *kétoldali*, a másodikban *egyoldali* ellenhipotézisről beszélünk:

$$\begin{cases} H_0: & \mu = \mu_0 \\ H_1: & \mu \neq \mu_0 \end{cases} \qquad \begin{cases} H_0: & \mu = \mu_0 \\ H_1^*: & \mu > \mu_0 \end{cases}$$

A vizsgált statisztika:

$$u = \sqrt{n}\,\frac{\bar{X}_n - \mu_0}{\sigma_0} \overset{H_0}{\sim} N(0,1)$$

Felírjuk, hogy mi az eloszlás, ha $\mu \neq \mu_0$: ???

$$\sqrt{n}\,\frac{\bar{X}_n - \mu_0}{\sigma_0} = \underbrace{\sqrt{n}\,\frac{\bar{X}_n - \mu}{\sigma_0}}_{N(0,1)} + \underbrace{\sqrt{n}\,\frac{\mu - \mu_0}{\sigma_0}}_{d_\mu} \overset{H_\mu}{\sim} N(d_\mu, 1)$$

Felírjuk az $1-\alpha$ szintű konfidencia-intervallumot a H_1 kétoldali ellenhipotézis esetén:

$$\mathrm{P}\left\{-u_{\alpha/2} < \sqrt{n}\,\frac{\bar{X}_n - \mu_0}{\sigma_0} < u_{\alpha/2} \mid H_0\right\} = 1-\alpha \tag{3.2}$$

Ekkor azt kapjuk, hogy

$$\mathrm{P}\left\{\text{I. fajú hiba}\right\} = \mathrm{P}\left\{C_1|H_0\right\} = \mathrm{P}\left\{|u| > u_{\alpha/2}|H_0\right\} = \alpha$$

és az elfogadási tartomány

$$C_0 = \left\{(X_1, \ldots, X_n) : \left|\sqrt{n}\,\frac{\bar{X}_n - \mu_0}{\sigma_0}\right| < u_{\alpha/2}\right\}$$

A másodfajú hibát μ függvényében számítjuk ki:

$$\beta(\mu) = \mathrm{P}\left\{C_0|H_\mu\right\} = \mathrm{P}\left\{-u_{\alpha/2} < N(d_\mu, 1) < u_{\alpha/2} \mid H_\mu\right\} =$$

$$= \Phi(u_{\alpha/2} - d_\mu) - \Phi(-u_{\alpha/2} - d_\mu)$$

Ha $\mu \to \pm\infty$, akkor $d_\mu \to \pm\infty$, ezért

$$\lim_{\mu \to \pm\infty} \beta(\mu) = 0$$

vagyis az erőfüggvény 1-hez tart. Hasonlóképpen az is látható, hogy $n \to \infty$ esetén fix μ-re is 1-hez tart az erőfüggvény. Ha ez a tulajdonság teljesül, akkor azt mondjuk, hogy a próba *konzisztens*.

Az elfogadási tartomány az $\{|u| \leq u_{\alpha/2}\}$ intervallum, ahol a kritikus értéket a $\mathrm{P}\left\{|u| \leq u_{\alpha/2}\right\} = 1 - \alpha$ egyenletből kapjuk, vagyis

$$u_{\alpha/2} = \Phi^{-1}(1 - \frac{\alpha}{2})$$

Egyoldalú ellenhipotézis esetén a (3.2) egyenlet helyett használt konfidencia-intervallum

$$\mathrm{P}\left\{\sqrt{n}\,\frac{\bar{X}_n - \mu}{\sigma_0} < u_\alpha\right\} = 1 - \alpha,$$

ahol $u_\alpha = \Phi^{-1}(1 - \alpha)$. Ekkor az elfogadási tartomány $\{u < u_\alpha\}$, vagyis csak a próbastatisztika nagy értékeire utasítunk el.

3.3.2. Kétmintás u-próba

A kétmintás u-próbánál két egymástól független minta várható értékeit hasonlítjuk össze:

$$\left.\begin{array}{rcl} X_1, \ldots X_n & \sim & N(\mu_1, \sigma_1^2) \\ Y_1, \ldots, Y_m & \sim & N(\mu_2, \sigma_2^2) \end{array}\right\} \quad \sigma_1^2, \sigma_2^2 \text{ ismert}$$

A hipotézisek: $H_0 : \mu_1 = \mu_2; H_1 : \mu_1 \neq \mu_2$. A próbastatisztika felírásához vegyük észre, hogy $\mathrm{E}\left[\bar{X}_n - \bar{Y}_m\right] \overset{H_0}{=} 0$ és $D^2(\bar{X}_n - \bar{Y}_m) = \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}$. Következésképpen a próbastatisztika

$$u = \frac{\bar{X}_n - \bar{Y}_m}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \overset{H_0}{\sim} N(0,1)$$

3.3.3. Egymintás t-próba

Ez a próba abban különbözik az u-próbától, hogy nem ismerjük a szórásnégyzetet:

$$X_1, \ldots, X_n \sim N(\mu, \sigma^2), \quad H_0 : \mu = \mu_0,\ H_1 : \mu \neq \mu_0 \text{ vagy } H_1^* : \mu > \mu_0$$

A próbastatisztika ebben az esetben (lásd 6. oldal):

$$t = \sqrt{n}\,\frac{\bar{X}_n - \mu_0}{s_n^{*\,2}} \overset{H_0}{\sim} t_{n-1}$$

Az kritikus érték függését a mintaelemszámtól az alábbi táblázat szemlélteti ($\alpha = 95\%$):

$n = f + 1$	kritikus érték $(t_{\alpha/2})$
2	12.7
4	3.18
10	2.26
$\vdots$	$\vdots$
∞	1.96

3.3.4. Kétmintás t-próba

A kétmintás u-próbához hasonlóan két független, ismeretlen szórásnégyzetű minta várható értékét hasonlítjuk össze. A minták méretének függvényében több esetet különböztetünk meg:

1. $n = m$
   Legyen $Z_i = X_i - Y_i$, $i = 1, \ldots, n$ új minta, melyre $D^2(Z_i) = \sigma_1^2 + \sigma_2^2$. A problémát visszavezethetjük egy egyszerű t-próbára:
   $$H_0^{'} : \mu = 0, \quad H_1^{'} : \mu \neq 0$$

2. $n \neq m$, de $\sigma_1 = \sigma_2$ (ezt előre teszteljük a később ismertetendő F-próbával)
   Az $(n-1){s_n^*}^2 + (m-1){s_m^*}^2 = \sum_{i=1}^{n}(X_i - \bar{X}_n)^2 + \sum_{i=1}^{m}(Y_i - \bar{Y}_m)^2$ mennyiséget az Steiner-formula (ld. 19. old.) alapján átalakítva az kapjuk, hogy
   $$(n-1){s_n^*}^2 + (m-1){s_m^*}^2 =$$
   $$\sum_{i=1}^{n}(X_i - \mu)^2 + \sum_{i=1}^{m}(Y_i - \mu)^2 - n(\bar{X}_n - \mu)^2 - m(\bar{Y}_m - \mu)^2$$
   eloszlása χ^2_{m+n} (két szabadsági fokot veszítünk el). Az egymintás esethez hasonlóan itt is felírhatjuk, hogy
   $$\mathrm{E}\left[\bar{X}_n - \bar{Y}_m\right] \overset{H_0}{=} 0 \text{ és } D^2(\bar{X}_n - \bar{Y}_m) = \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m} = \sigma^2\left(\frac{1}{n} + \frac{1}{m}\right)$$
   Ebből adódik a próbastatisztika:
   $$t = \frac{\bar{X}_n - \bar{Y}_m}{\sqrt{\frac{1}{n} + \frac{1}{m}}\sqrt{\frac{(n-1){s_n^*}^2 + (m-1){s_m^*}^2}{n+m-2}}} \overset{H_0}{\sim} t_{n+m-2}$$

3.3.5. F-próba

Az F-próbánál független minták szórásnégyzeteit hasonlítjuk össze:

$$\begin{aligned} X_1, \ldots, X_n &\sim N(\mu_1, \sigma_1^2) \\ Y_1, \ldots, Y_m &\sim N(\mu_2, \sigma_2^2) \end{aligned}$$

A hipotézisek: $H_0 : \sigma_1^2 = \sigma_2^2$, $H_1 : \sigma_1^2 \neq \sigma_2^2$ vagy $H_1^* : \sigma_1^2 > \sigma_2^2$. A tapasztalati szórásnégyzet tulajdonságai alapján (lásd 19. oldal)

$$\frac{n-1}{\sigma_1^2}s_n^{*2} \sim \chi^2_{n-1} \quad \text{és} \quad \frac{m-1}{\sigma_2^2}s_m^{*2} \sim \chi^2_{m-1}$$

H_0 teljesülése esetén

$$\frac{s_n^{*2}}{s_m^{*2}} \sim \frac{\chi^2_{n-1}/(n-1)}{\chi^2_{m-1}/(m-1)} \sim F_{n-1,m-1}$$

Egyoldalú ellenhipotézis esetén ($H_1 : \sigma_1^2 > \sigma_2^2$) mindig elvetjük H_1-et, ha $s_n^{*2} < s_m^{*2}$. Amennyiben $s_n^{*2} \geq s_m^{*2}$, akkor a kritikus tartomány

$$\frac{s_n^{*2}}{s_m^{*2}} > F_{n-1,m-1}(\alpha),$$

vagyis a hányados nagyobb, mint az F eloszlás α-hoz tartozó kritikus értéke.

Ha az ellenhipotézis kétoldali, akkor kihasználjuk, hogy

$$1/F_{n-1,m-1} = F_{m-1,n-1}, \text{ ezért}$$
$$\mathrm{P}\{F > x\} = \alpha/2 \quad \text{és} \quad \mathrm{P}\{1/F < 1/x\} = \alpha/2$$

Legyen a próbastatisztika

$$F^* = \max\left\{\frac{s_n^{*2}}{s_m^{*2}}, \frac{s_m^{*2}}{s_n^{*2}}\right\} =\geq 1$$

Ekkor az elutasítási tartomány $F^* \geq F_{n-1,m-1}(\alpha/2)$.

3.4. Nemparaméteres próbák

Nemparaméteres próbák esetén a feltevés az eloszlások egészére vonatkozik, nem pedig az eloszlásokat leíró paraméterek számértékére. Például ebbe a kategóriába tartozik annak eldöntése, hogy adott események függetlenek-e egymástól, vagy az, hogy a két mintánk azonos eloszlásból származik-e.

3.4.1. χ^2-próbák

Ezekben a próbákban egy $A_1, \ldots, A_r$ teljes eseményrendszerből indulunk ki, ahol $\forall i \in [1..r]$-re $\mathrm{P}\{A_i\} = p_i$. Az N mintaelemet N független kísérletnek tekintjük, és az A_i bekövetkezéseinek számát X_i-vel jelöljük. Ennek eloszlása multinomiális (polinomiális):

$$\mathrm{P}\{X_1 = k_1, \ldots, X_r = k_r\} = \begin{cases} \frac{N!}{k_1!\ldots k_n!} p_1^{k_1} \ldots p_r^{k_r} & \text{ha } k_1 + \ldots k_r = N \\ 0 & \text{különben.} \end{cases}$$

$X_1 + \cdots + X_r = N$ miatt a változók összefüggőek.

Ha csak maximum 3 osztályt tekintünk, akkor

$$\begin{aligned} \mathrm{P}\{X_1 = k_1\} &= \tfrac{N!}{k_1!(N-k_1)!} p_1^{k_1}(1-p_1)^{N-k_1} \\ \mathrm{P}\{X_1 = k_1, X_2 = k_2\} &= \tfrac{N!}{k_1!k_2!(N-k_1-k_2)!} p_1^{k_1} p_2^{k_2}(1-p_1-p_2)^{N-k_1-k_2} \end{aligned}$$

Ezek segítségével a változók kovariancia-struktúrája kifejezhető:

$$\begin{array}{ll} \mathrm{E}[X_1] & = Np_1 \\ D^2[X_1] & = Np_1(1-p_1) \\ \mathrm{E}[X_1X_2] & = N(N-1)p_1p_2 \\ \mathrm{Cov}[X_i,X_j] & = -Np_ip_j \end{array}$$

Bevezetjük a következő jelöléseket, és tovább számolunk:

$$\vec{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_r \end{pmatrix}, \quad \mathrm{E}\left[\vec{X}\right] = N \begin{pmatrix} p_1 \\ \vdots \\ p_r \end{pmatrix} = N\vec{p}$$

$$\begin{aligned} Var\vec{X} &= \mathrm{E}\left[\vec{X}\vec{X}^T\right] - \mathrm{E}\left[\vec{X}\right]\mathrm{E}\left[\vec{X}\right]^T = \\ &= N \begin{pmatrix} p_1(1-p_1) & -p_1p_2 & \dots & -p_1p_r \\ -p_2p_1 & p_2(1-p_2) & \dots & -p_2p_r \\ \vdots & \vdots & \ddots & \vdots \\ -p_rp_1 & -p_rp_2 & \dots & p_r(1-p_r) \end{pmatrix}. \end{aligned}$$

Jelölje itt a mátrixot $\mathbb{S} := \mathrm{diag}(p_1,\dots,p_r) - [p_ip_j]_{1\le i,j\le r}$! A centrális határeloszlás tétel alapján az aszimptotikus eloszlás számítható:

$$Z_N := \frac{1}{\sqrt{N}}\left(\vec{X} - \mathrm{E}\left[\vec{X}\right]\right) \xrightarrow{d} N(0,\mathbb{S}).$$

Ebből következik, hogy $\forall A \in \mathbb{R}^{r\times r}$ mátrixra

$$\frac{1}{\sqrt{N}} A\left(\vec{X} - \mathrm{E}\left[\vec{X}\right]\right) \xrightarrow{d} N(0, A\mathbb{S}A^T).$$

Speciálisan $A = \mathrm{diag}(\frac{1}{\sqrt{p_1}},\dots\frac{1}{\sqrt{p_r}})$ esetén $\mathbb{S} = \mathrm{diag}(p_1,\dots,p_r) - \vec{p}\vec{p}^T$ miatt a szórásnégyzet a következő alakú:

$$A\mathbb{S}A = \mathbb{I} - (A\vec{p})(A\vec{p})^T = \mathbb{I} - aa^T,$$

ahol $\mathbb{I}$ az identitásmátrix és $a = \begin{pmatrix} \sqrt{p_1} \\ \vdots \\ \sqrt{p_r} \end{pmatrix}$ és emiatt $a^Ta = 1$.

Legyen $Y_j := \frac{X_j - Np_j}{\sqrt{Np_j}}$! Ekkor $\|Y\|_{L_2} = \sum_{j=1}^r \frac{(X_j-Np_j)^2}{Np_j} \to W^TW$, ahol $W \sim N(0, \mathbb{I}-aa^T)$. Ez a mennyiség nem független normális eloszlások négyzetösszege. Az $\mathbb{I} - aa^T$ mátrixnak 0 sajátértéke az a sajátvektorral, mivel $(\mathbb{I} - aa^T)a = a - a = 0$. Ezen felül minden $v \perp a$ vektorra $a^Tv = 0$ miatt $(\mathbb{I} - aa^T)v = v$, azaz ezek a vektorok az 1 sajátértékhez tartozó sajátvektorok. Így ha W szórásmátrixát diagonalizáljuk, akkor $UW = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix}$ alakú lesz.

Azaz a $W^TW = W^TU^TUW$ mátrix $(n-1)$ db $N(0,1)$ eloszlású független változó négyzetösszege, azaz

$$\sum_{j=1}^{r} \frac{(X_j - Np_j)^2}{Np_j} \xrightarrow{d} \chi^2_{r-1}.$$

Megjegyezzük, hogy a próba akkor alkalmazható hatékonyan, ha minden részhalmazba legalább 6 mintaelem esik. Ellenkező esetben célszerű osztályokat összevonni.

Hipotézisvizsgálat

$$H_0 : \mathrm{P}\{A_i\} = p_i$$

$$\chi^2 = \sum_{i=1}^{r} \frac{(k_i - Np_i)^2}{Np_i} = \sum_{i=1}^{r} \frac{(\text{megfigyelt} - \text{várt})^2}{\text{várt}} \xrightarrow{d} \chi^2_{r-1},$$

ahol az A_i esemény k_i-szer következett be (megfigyelt érték), míg várható értékben Np_i-szer kellett volna ennek történnie (várt).

Ha ez a χ^2 próbastatisztika kis értéket vesz fel, a nullhipotézist elfogadjuk, ellenkező esetben elutasítjuk. Ennek eldöntéséhez a megfelelő szabadsági fokú χ^2 eloszlás táblázatát használjuk, amivel a próbastatisztika aszimptotikusan azonos eloszlású.

Példa:

Egy dobókockával 600–szor dobunk. Az eredményeink:

1	2	3	4	5	6
83	91	122	107	74	123

Kérdés: Szabályos-e a kocka?
Megoldás:

$$\chi^2 = \frac{(83-100)^2}{100} + \cdots + \frac{(123-100)^2}{100} = 21,08$$

Az 5 szabadsági fokú χ^2 eloszlás táblázatában 0,001-es szinten 20,5, 0,0005-ös szinten 22,1 szerepel, így a nullhipotézist 0,001 és e feletti szinteken fogadjuk el.

Tiszta illeszkedésvizsgálat

Az a kérdés, hogy mintánk egy előre adott eloszlásból származik-e?

$$H_0 : X \sim F(x)$$

Válasszunk $A_i := \{X \in C_i\}$ eseményeket, ahol a C_i-k diszjunkt intervallumok, melyek uniója $\mathbb{R}$!

Példa:

Százelemű mintáról döntsük el, hogy 2 paraméterű Poisson eloszlásból származik-e! A választott események, és a minta eloszlása ezek között:

A_0	A_1	A_2	A_3	A_4
$\{X=0\}$	$\{X=1\}$	$\{X=2\}$	$\{X=3\}$	$\{X\geq 4\}$
12	32	25	21	10

A fenti események nullhipotézis melletti valószínűségeit könnyen számíthatjuk:

A_0	A_1	A_2	A_3	A_4
$0,135$	$0,270$	$0,270$	$0,180$	$0,145$

$$\chi^2 = \frac{(12-13,5)^2}{13,5} + \cdots + \frac{(10-14,5)^2}{14,5} \approx 3,316.$$

A kritikus érték pedig 0,9-es szinten 7,77. A nullhipotézist elfogadjuk.

Becsléses illeszkedésvizsgálat

A becsléses illeszkedésvizsgálat feladata abban különbözik a tiszta illeszkedésvizsgálattól, hogy nem tudjuk pontosan, melyik eloszlással egyezhet meg a minta, csak a paraméteres eloszlás adott, így először egy paraméterbecslést kell végezni.

$$H_0 : \mathrm{P}\{X < t\} = F(t; \theta_1, \ldots \theta_k)$$

Az r elemű minta alapján először a k paraméterre adunk maximum likelihood becslést, majd egy tiszta illeszkedésvizsgálatot végzünk a következő nullhipotézissel:

$$H_0' : \mathrm{P}\{X < t\} = F(t; \hat{\theta_1}, \ldots, \hat{\theta_k})$$

Ekkor az előzetes becslések miatt a $\sum_{j=1}^{r} \frac{(X_j - Np_j)^2}{Np_j}$ próbastatisztika aszimptotikus eloszlása csak $r-k-1$ szabadsági fokú χ^2 eloszlás lesz!

Függetlenségvizsgálat

Adott két teljes eseményrendszer: $A_1, \ldots, A_r$ és $B_1, \ldots, B_s$. Döntsünk arról a hipotézisről, hogy minden i-re és j-re az A_i és B_j események függetlenek!

$$H_0 : \mathrm{P}\{A_i \cap B_j\} = \mathrm{P}\{A_i\}\,\mathrm{P}\{B_j\}$$

Amennyiben a $\mathrm{P}\{A_i\} = p_i$ és a $\mathrm{P}\{B_j\} = q_j$ értékek ismertek, az aszimptotikus határeloszlás szabadsági fokainak száma $rs-1$ lesz.

Becsléses vizsgálat esetén egy ún. kontingenciatáblázatot írunk fel:

	B_1	B_2	$\dots$	B_s	$\sum$
A_1	$k_{1,1}$	$k_{1,2}$	$\dots$	$k_{1,s}$	$k_{1,\bullet}$
A_2	$k_{2,1}$	$k_{2,2}$	$\dots$	$k_{2,s}$	$k_{2,\bullet}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_r	$k_{r,1}$	$k_{r,2}$	$\dots$	$k_{r,s}$	$k_{r,\bullet}$
$\sum$	$k_{\bullet,1}$	$k_{\bullet,2}$	$\dots$	$k_{\bullet,s}$	$N = k_{\bullet,\bullet}$

Itt $\forall i \in \{1,\dots,r\}, \forall j \in \{1,\dots,s\} \quad k_{i,j} = \#\{X_l \mid X_l \in A_i \cap B_j\}$, $\forall i \in \{1,\dots,r\} \quad k_{i,\bullet} = \#\{X_l \mid X_l \in A_i\}$, $\forall j \in \{1,\dots,s\} \quad k_{\bullet,j} = \#\{X_l \mid X_l \in B_j\}$.

A táblázat segítségével a p_i és q_j valószínűségeket a $\frac{k_{i,\bullet}}{N}$ és a $\frac{k_{\bullet,j}}{N}$ mennyiségekkel becsülhetjük, így a $\mathrm{P}\{A_i \cap B_j\}$ várt értéke $N\frac{k_{i,\bullet}}{N}\frac{k_{\bullet,j}}{N} = \frac{k_{i,\bullet}k_{\bullet,j}}{N}$ lesz. Tehát a következő statisztikát alkalmazzuk:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{(k_{i,j} - \frac{k_{i,\bullet}k_{\bullet,j}}{N})^2}{\frac{k_{i,\bullet}k_{\bullet,j}}{N}} \to \chi^2_{\text{szabadsági fok}}$$

A szabadsági fokok száma pedig: $rs - 1 - (r-1) - (s-1) = (r-1)(s-1)$, mivel a p_i-k és a q_j-k között van egy-egy szabad összefüggés ($\sum_{i=1}^{r} p_i = 1$).

Ha két valószínűségi változó (X, Y) függetlenségére vagyunk kíváncsiak, akkor az eseményeink a következők lesznek:

$$\begin{aligned} A_i &= \{a_{i-1} \le X < a_i\} \quad i = 1,\dots,r \\ B_j &= \{b_{i-1} \le Y < b_i\} \quad j = 1,\dots,s \end{aligned}$$

Példa:

Egy kétszáz fős minta alapján vizsgáljuk meg a szem és a hajszín függetlenségét!

	szőke	barna	fekete	$\sum$
kék szem	42	28	3	73
barna szem	17	89	21	117
$\sum$	59	117	24	200

A χ^2 statisztika értéke 49,11, a szabadsági fokok száma 2, ennek az eloszlásnak a kritikus értéke még $\alpha = 0,001$ esetén is 13,8, azaz elvetjük a függetlenséget, a haj- és a szemszínek összefüggnek!

3.4.2. Kolmogorov-Szmirnov próba

A próba eloszlások egyenlőségének igazolására szolgál. Első esetben legyen adott két minta, $X_1,\dots,X_n$ és $Y_1,\dots,Y_n$, amelyek az F és G eloszlásfüggvényű eloszlásokból származnak.

$$H_0 : F = G$$

H_0 esetén $F_n^*(x)$ és $G_n^*(x)$ „közel" van egymáshoz, és a $D_{n,n}^+ := F_n^*(x) - G_n^*(x)$ ugrófüggvény $\pm\frac{1}{n}$ ugrásokkal, $\lim_{x\to\pm\infty} D_{n,n}^+ = 0$. Feltehető, hogy az ugrások nem esnek egybe (ez 0 valószínűségű), azaz $2n$ ugráspont van.

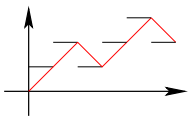
Ki fogjuk számítani a következő esemény valószínűségét:

$$\left\{\sqrt{\frac{n}{2}}D_{n,n}^+ < z\right\} = \left\{\sqrt{2n}\sqrt{\frac{n}{2}}D_{n,n}^+ < \sqrt{2n}z\right\}.$$

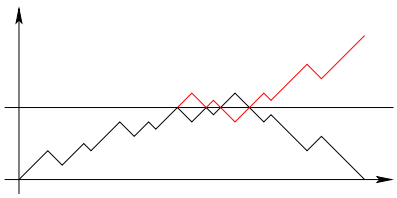
Ennek komplementer eseménye

$$\left\{nD_{n,n}^+ \geq \lceil z\sqrt{2n}\rceil\right\}.$$

Az $nD_{n,n}^+ = n\left(F_n^*(x) - G_n^*(x)\right)$ egységlépcsős függvény, a legnagyobb 0-tól való eltérésének eloszlását szeretnénk kiszámolni. Mivel minket csak az eltérés maximuma érdekel, így az ugrófüggvényt helyettesíthetjük egy bolyongással, mivel a két függvény supremuma megegyezik.



Kérdésünk tehát úgy fogalmazható át, hogy mi a valószínűsége annak, hogy a $(0,0)$-ból induló $(2n,0)$-ba érkező töröttvonal érinti a c szintet.



Az összes c szintet érintő vonal számát (itt $c = \lceil z\sqrt{2n}\rceil$ lesz) tükrözési módszerrel határozzuk meg. Minden vonal első c szintet érintő pontja utáni részét tükrözzük a c szintre. Ekkor egy $(0,0)$-ból induló $(2n,2c)$-be érkező töröttvonalat kapunk. Minden c-t érintő első típusú töröttvonalból kaptunk egy ilyet egyértelműen, és visszatükrözéssel mindegyikből egyértelműen visszakaphatjuk az eredetit, tehát a $(0,0)$-ból induló $(2n,0)$-ba érkező c-t érintő töröttvonalak száma megegyezik a $(0,0)$-ból induló $(2n,2c)$-be érkező töröttvonalak számával.

Az összes visszatérő vonal száma: $\binom{2n}{n}$.

A visszatérő c-t érintő töröttvonalak száma: $\binom{2n}{n-c}$.

$$\mathrm{P}\left\{\sqrt{\frac{n}{2}}\sup\left(F_n^*(x) - G_n^*(x)\right) < z\right\} = \begin{cases} 0 & \text{ha } 0 \geq z \\ 1 - \left\lfloor \frac{\binom{2n}{n-c}}{\binom{2n}{n}} \right\rfloor & \text{ha } 0 < z \leq \sqrt{\frac{n}{2}} \\ 1 & \text{ha } z > \sqrt{\frac{n}{2}} \end{cases}$$

$$\frac{\binom{2n}{n-c}}{\binom{2n}{n}} = \frac{(2n)!}{(n-c)!(n+c)!} \Big/ \frac{(2n)!}{n!n!} = \frac{n!n!}{(n-c)!(n+c)!} \approx$$

$$\frac{2\pi n\left(\frac{n}{e}\right)^{2n}}{\sqrt{2\pi(n-c)}\left(\frac{n-c}{e}\right)^{n-c}\sqrt{2\pi(n+c)}\left(\frac{n+c}{e}\right)^{n+c}} =$$

$$\frac{n}{\sqrt{n^2-c^2}}\frac{n^{2n}}{(n-c)^{n-c}(n+c)^{n+c}} = \frac{n}{\sqrt{n^2-c^2}}\frac{n^{2n}}{(n^2-c^2)^{n-c}(n+c)^{2c}} =$$

$$\frac{n}{\sqrt{n^2-c^2}}\left(\frac{n^2}{n^2-c^2}\right)^{n-c}\left(\frac{n}{n+c}\right)^{2c}=$$

$$\sqrt{\left(1+\frac{c^2}{n^2-c^2}\right)}\left(1+\frac{c^2}{n^2-c^2}\right)^{n-c}\left(1-\frac{c}{n+c}\right)^{2c}\overset{c\approx z\sqrt{2n}}{\approx}$$

$$\sqrt{\left(1+\frac{2z^2}{n-2z^2}\right)}\left(1+\frac{2z^2}{n-2z^2}\right)^{n-z\sqrt{2n}}\left(1-\frac{z\sqrt{2n}}{n+z\sqrt{2n}}\right)^{2z\sqrt{2n}}\overset{n\to\infty}{\longrightarrow}$$

$$\to e^{2z^2}e^{-4z^2}=e^{-2z^2}$$

Tehát $n\to\infty$ esetén belátható, hogy a $\mathrm{P}\left\{\sqrt{\frac{n}{2}}\sup\left(F_n^*(x)-G_n^*(x)\right)<z\right\}$ negatív z-kre 0 pozitívakra pedig aszimptotikusan $1-e^{-2z^2}$. Valójában a tapasztalati eloszlásfüggvények különbségének abszolútértékét szoktuk venni, és ennek supremumát számítjuk. Ennek aszimptotikus eloszlását egy jóval bonyolultabb modell segítségével kaphatjuk:

$$\mathrm{P}\left\{\sqrt{\frac{n}{2}}\sup|F_n^*(x)-G_n^*(x)|<z\right\}\to K(z)=\sum_{k=-\infty}^{\infty}(-1)^k e^{-2k^2z^2},$$

ahol az itt definiált $K(z)$ függvényt Kolmogorov függvénynek hívjuk.

Különböző elemszámú minták (n és m) esetén a $\sqrt{\frac{n}{2}}$ helyett $\sqrt{\frac{mn}{m+n}}$ használandó.

Egymintás próbában a mintát egy ismert eloszlás eloszlásfüggvényével hasonlítjuk. Ekkor az igaz, hogy

$$D_n:=\sup|F_n^*(x)-F(x)|$$
$$\mathrm{P}\left\{\sqrt{n}D_n<z\right\}\to K(z)$$

4. fejezet

Lineáris modellek

4.1. Kovariancia, korreláció

Két változó közötti összefüggés felállítása a megfigyelt értékek alapján

$$\operatorname{Cov}[X,Y] = \operatorname{E}[(X - \operatorname{E}[X])(Y - \operatorname{E}[Y])] = \operatorname{E}[XY] - \operatorname{E}[X]\operatorname{E}[Y]$$

Ha $\operatorname{Cov}[X,Y] = 0$, akkor a két változót korrelálatlannak nevezzük. Ha két együttesen normális változó korrelálatlan, akkor függetlenek is.

A korrelációt pedig a következőképpen definiáljuk:

$$\rho(X,Y) := \frac{\operatorname{Cov}[X,Y]}{D(X)D(Y)}$$

Az $\operatorname{E}[X] = \operatorname{E}[Y] = 0$ speciális esetben $\rho(X,Y) = \frac{\operatorname{E}[XY]}{\sqrt{\operatorname{E}[X^2]\operatorname{E}[Y^2]}}$

A Cauchy-Bunyakovszkij-Schwarz egyenlőtlenség miatt bármely két változó korrelációs együtthatója abszolútértékben kisebb 1-nél. Ha pedig $|\rho(X,Y)| = 1$, akkor $Y = aX + b$. Ha viszont $|\rho(X,Y)| \neq 1$, akkor is lehet $Y \approx aX + b$.

Két minta tapasztalati korrelációs együtthatóját a következőképpen számoljuk:

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

skalárszorzat, norma megközelítés

4.2. Regressziószámítás

4.2.1. Elméleti regresszió

Elméleti (lineáris) regressziós feladat: adott X és Y valószínűségi változók, amelyekre feltételezhető, hogy közel lineárisan függnek egymástól. Keressük meg a linearitás paramétereit!

$$\|Y - (aX + b)\|^2 \xrightarrow{a,b=?} \min$$

Az általános regressziós feladatban pedig azt az f függvényt (regressziós görbe) keressük, amely minimalizálja a $\mathrm{E}\left[(Y - f(X))^2\right]$ mennyiséget (legkisebb négyzetek módszere).

$$\mathrm{E}\left[(Y - f(X))^2\right] = \int\int_{\mathbb{R}^2} [y - f(x)]^2 h(x,y)dxdy,$$

ahol $h(x,y)$ az együttes sűrűségfüggvény. Ennek megoldása az $\mathrm{E}[Y|X]$ feltételes várható érték lesz, mivel:

$$\mathrm{E}\left[(Y - f(X))^2\right] = \mathrm{E}\left[[(Y - \mathrm{E}[Y|X]) + (\mathrm{E}[Y|X] - f(X))]^2\right] =$$

$$= \mathrm{E}\left[[Y - \mathrm{E}[Y|X]]^2\right] + \mathrm{E}\left[[\mathrm{E}[Y|X] - f(X)]^2\right] +$$

$$+2\mathrm{E}\left[(Y - \mathrm{E}[Y|X])(\mathrm{E}[Y|X] - f(X))\right] =$$

$$= \mathrm{E}\left[[Y - \mathrm{E}[Y|X]]^2\right] + \mathrm{E}\left[[\mathrm{E}[Y|X] - f(X)]^2\right], \text{ mert}$$

$$\mathrm{E}\left[(Y - \mathrm{E}[Y|X])(\mathrm{E}[Y|X] - f(X))\right] =$$

$$\mathrm{E}\left[\mathrm{E}\left[(Y - \mathrm{E}[Y|X])(\mathrm{E}[Y|X] - f(X))\,|X\right]\right] =$$

$$\mathrm{E}\left[(\mathrm{E}[Y|X] - \mathrm{E}[Y|X])(\mathrm{E}[Y|X] - \mathrm{E}[f(X)|X])\right] = 0$$

Tehát a fenti kifejezés a minimumát épp a feltételes várható érték helyen veszi fel, azaz a regressziós görge

$$f(x) = \mathrm{E}[Y|X = x] = \frac{\int_{\mathbb{R}} yh(x,y)dy}{\int_{\mathbb{R}} h(x,y)dy}$$

Általában ezt a regressziós görbét helyettesítjük a legjobban közelítő egyenessel.

Ha az (X,Y) valószínűségi változó multinormális eloszlású, akkor a regressziós görbe egy egyenes lesz, ugyanis konstruálhatunk egy X-től független Z valószínűségi változót a következő módon:

Legyen $Z(k) := (Y - \mathrm{E}[Y]) + k(X - \mathrm{E}[X])$! $Z(k)$ minden k-ra 0 várható értékű, de vajon létezik-e olyan k, amelyre független X-től?

$$\mathrm{Cov}[Z(k), X] = \mathrm{E}[(Z(k) - 0)(X - \mathrm{E}[X])] = \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])] +$$

$$+k\mathrm{E}\left[(X - \mathrm{E}[X])^2\right] = \mathrm{Cov}[X,Y] + kD^2(X),$$

amely kifejezés pontosan akkor 0, ha

$$k = \frac{-\mathrm{Cov}[X,Y]}{D^2(X)}, \text{ így}$$

$$0 = \mathrm{E}[Z] = \mathrm{E}[Z|X = x] = \mathrm{E}[Y|X = x] - \mathrm{E}[Y] - \frac{\mathrm{Cov}[X, Y]}{D^2(X)}(x - \mathrm{E}[X]).$$

Ez a képlet megadja a $\mathrm{E}[Y|X = x]$ a regressziós görbe egyenletét, ami egy egyenes.

4.2.2. Tapasztalati regresszió

Adott (X, Y) n elemű minta. Ezek tapasztalati korrelációja

$$\rho_{X,Y} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})\sum_{i=1}^{n}(Y_i - \bar{Y})}}$$

Az optimalizálás elvégzése után kapott tapasztalati regressziós egyenes egyenlete ennek segítségével írható fel:

$$\frac{y - \bar{Y}}{s_Y} = \rho_{X,Y}\frac{x - \bar{X}}{s_X}$$

4.2.3. Parciális korreláció

Két valószínűségi változó parciális korrelációja az a szám, amely megmutatja a két valószínűségi változó összefüggésének mértékét valameny más változó(k) hatásának kiszűrése után.

Például a valós életben azt tapasztaljuk, hogy a testmagasság és a hajhossz erősen negatívan korrelált az embereknél, ami csak ezt a két mennyiséget figyelembe véve nem indokolható. A jelenség oka az, hogy a férfiak általában magasabbak, és rövid hajuk van, a nők pedig általában alacsonyabbak hosszú hajjal. Tehát e harmadik valószínűségi változó (a nem) okozza a nagy negatív korrelációt, ennek hatását kiszűrve a parciális korreláció már nem lesz ilyen szélsőséges.

Adott $X_1, \ldots, X_n$ valószínűségi változók esetén $\rho_{i,j\bullet k_1,\ldots,k_l}$-lel jelöljük az az i. és a j. parciális korrelációját, ha az $X_{k_1}, \ldots, X_{k_l}$ valószínűségi változók hatását számítjuk le.

Kiszámításakor valójában az történik, hogy az X_i-t és az X_j-t levetítjük az $X_{k_1}, \ldots, X_{k_l}$ által kifeszített altérre (regresszió), és az így kapott változók (minták) korrelációját számítjuk.

Egy valószínűségi változó hatásának kiszűrésekor lineáris regressziót kell végezni, ennek eredménye egyszerű alakban a következőképpen írható:

$$\rho_{i,j\bullet k} = \frac{\rho_{i,j} - \rho_{i,k}\rho_{j,k}}{\sqrt{(1 - \rho_{i,k}^2)(1 - \rho_{j,k}^2)}}.$$

Igaz továbbá a következő két formula:

$$\rho_{i,j\bullet k,l} = \frac{\rho_{i,j\bullet k} - \rho_{i,k\bullet l}\rho_{j,k\bullet l}}{\sqrt{(1 - \rho_{i,k\bullet l}^2)(1 - \rho_{j,k\bullet l}^2)}}.$$

tobbszoros korrelacio

4.2.4. Többváltozós regresszió

Az Y valószínűségi változót vizsgáljuk néhány $X_1, \ldots, X_k$ (ún. faktor) függvényében, azaz $Y = f(X_1, \ldots, X_k)$. Ezt lineárisan a $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$ egyenlettel közelítjük, ahol $\underline{y} \in \mathbb{R}^n$ egy n elemű minta, $X \in \mathbb{R}^n \times \mathbb{R}^p$ mátrix ($n \gg p$), amelynek első oszlopa csupa egyesből áll, többi oszlopa a faktorok megfigyelt értékeit tartalmazza, $\underline{\varepsilon}$ a zajvektor, célunk pedig a faktorok súlyának, a $\underline{\beta}$ vektornak optimális meghatározása.

$$\|\varepsilon\|^2 = (y - X\beta)^T(y - X\beta) = (y^T - \beta^T X^T)(y - X\beta) =$$
$$y^T y - \beta^T X^T y - y^T X\beta + \beta^T X^T X\beta = y^T y - 2\beta^T X^T y + \beta^T X^T X\beta \to \min$$

Ennek megoldását szemléletesen vektor szerinti derivált formájába írhatjuk:

$$\frac{\partial \varepsilon^T \varepsilon}{\partial \beta} = -2X^T y + 2X^T X\beta = 0.$$

Itt alkalmaztuk a vektor szerinti deriválás következő szabályait:

$$\frac{\partial}{\partial \underline{v}} M\underline{v} = M^T,$$

$$\frac{\partial}{\partial \underline{v}} \underline{v}^T S\underline{v} = 2S\underline{v}, \text{ ahol } S \text{ szimmetrikus mátrix.}$$

Ezek segítségével kapható meg az ún normálegyenletrendszer:

$$X^T X\underline{\beta} = X^T \underline{y}.$$

Speciálisan egy változó esetén a következőket kapjuk:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \text{ azaz}$$
$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}.$$

$$X^T X = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ x_1 & x_2 & \ldots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} = \begin{pmatrix} n & n\bar{x} \\ -n\bar{x} & \sum_{i=1}^n x_i^2 \end{pmatrix}$$

$$\left(X^T X\right)^{-1} = \frac{1}{n\left(\sum_{i=1}^n x_i^2 - n\bar{x}^2\right)} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -n\bar{x} \\ n\bar{x} & n \end{pmatrix}.$$

Ezek segítségével a legkisebb négyzetes közelítés a normálegyenletrendszerből számítható:

$$\hat{\beta} = \left(X^T X\right)^{-1} X^T y = \frac{1}{n\left(\sum_{i=1}^n x_i^2 - n\bar{x}^2\right)} \begin{pmatrix} \sum_{i=1}^n x_i^2 & -n\bar{x} \\ n\bar{x} & n \end{pmatrix} \cdot$$

$$\cdot \begin{pmatrix} n\bar{y} \\ \sum_{i=1}^n x_i y_i \end{pmatrix} = \frac{1}{n\left(\sum_{i=1}^n x_i^2 - n\bar{x}^2\right)} \begin{pmatrix} \bar{y}\sum_{i=1}^n x_i^2 - \bar{x}\sum_{i=1}^n x_i y_i \\ -n\bar{x}\bar{y} + \sum_{i=1}^n x_i y_i \end{pmatrix}$$

Ez tehát egy általános lineáris modell, mivel az $\underline{y}$ vektor a $\underline{\beta}$ vektor mátrixszorzatával kifejezhető.

4.3. Szórásanalízis

Célunk most bizonyos tényezők hatásának vizsgálata a szórás azonosítása révén. A különböző faktoroknak (a kísérlet kimenetelét befolyásoló tényezőknek) szintjeik vannak, amelyek minőségi és mennyiségi jellemzőket jelentenek, a kérdés pedig az, hogy szignifikáns-e a faktor hatása az adatokra?

Az egytényezős modellben p darab sokaság van, az i. mérete ebből legyen n_i, $n = \sum_{i=1}^p n_i$, valamint jelölje $Y_{i,j}$ az i. sokaság j. elemét ($i = 1 \dots p$, $j = 1 \dots n_i$)! Feltevésünk szerint ekkor $Y_{i,j} = \beta_i + e_{i,j}$, ahol $e_{i,j} \sim N(0, \sigma^2)$ hibák.

A hibák oka lehet a szinteken belüli ingadozás (az adott faktortól függetlenül) vagy az, hogy a faktor szignifikáns.

Definiáljuk a β_i-k eltérését a teljes átlagtól:

$$\beta_i = \mu + \alpha_i, \text{ ahol } \mu = \frac{1}{n}\sum_{i=1}^p n_i \beta_i, \text{ azaz}$$

$$Y_{i,j} = \mu + \alpha_i + e_{i,j}$$

Nullhipotézisünk az, hogy a faktornak nincs hatása, azaz

$$H_0 : \beta_1 = \dots = \beta_p = \mu, \text{ vagyis } \alpha_1 = \dots = \alpha_p = 0$$

Az átlagokat a szokásos módon jelöljük:

$$\bar{Y}_{\bullet,\bullet} := \frac{1}{n}\sum_{i=1}^p \sum_{j=1}^{n_i} Y_{i,j},$$

$$\bar{Y}_{i,\bullet} := \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{i,j}.$$

Most Steiner tételét alkalmazzuk:

$$Q := \sum_{i=1}^p \sum_{j=1}^{n_i} \left(Y_{i,j} - \bar{Y}_{\bullet,\bullet}\right)^2 = \underbrace{\sum_{i=1}^p \sum_{j=1}^{n_i} \left(Y_{i,j} - \bar{Y}_{i,\bullet}\right)^2}_{Q_{csb}} +$$

$$\underbrace{+\sum_{i=1}^{p}\sum_{j=1}^{n_i}\left(\bar{Y}_{i,\bullet}-\bar{Y}_{\bullet,\bullet}\right)^2}_{=\sum_{i=1}^{p} n_i\left(\bar{Y}_{i,\bullet}-\bar{Y}_{\bullet,\bullet}\right)^2=Q_{csk}} = Q_{csb}+Q_{csk}, \text{ ahol}$$

Q_{csk} a csoportok közötti, és Q_{csb} a csoportokon belüli négyzetösszegek, tehát

$$Q=\sum_{i=1}^{p}\sum_{j=1}^{n_i} Y_{i,j}^2 - n\bar{Y}_{\bullet,\bullet}.$$

A belső négyzetösszeg $(n-p)$, a külső négyzetösszeg $(p-1)$ szabadsági fokú, így az összeg szabadsági fokainak száma $(n-1)$.

Lemma: Ha az $X_1,\ldots,X_n$ valószínűségi változók függetlenek, és $\mathrm{E}[X_i]=\mu_i$ valamint $D^2[X_i]=\sigma^2$, akkor $\mathrm{E}\left[(X_i-\bar{X})^2\right]=(\mu_i-\mu)^2+\frac{n-1}{n}\sigma^2$, ahol $\bar{\mu}=\frac{1}{n}\sum_{i=1}^{n}\mu$.

Az állítás a következő két egyenlőség összevetéséből következik:

$$\mathrm{E}\left[(X_i-\bar{X})^2\right]=\underbrace{\left(\mathrm{E}\left[X_i-\bar{X}\right]\right)^2}_{(\mu_i-\bar{\mu})^2}+D^2(X_i-\bar{X})$$

$$D^2(X_i-\bar{X})=D^2(X_i)+D^2(\bar{X})-2\mathrm{Cov}\left[X_i,\bar{X}\right]=$$

$$=\sigma^2+\frac{\sigma^2}{n}-2\frac{1}{n}\sigma^2=\frac{n-1}{n}\sigma^2$$

E lemmának következménye, hogy $\frac{Q_{csb}}{n-p}$ torzítatlan becslés σ^2-re, mivel

$$\mathrm{E}\left[Q_{csb}\right]=\mathrm{E}\left[\sum_{i=1}^{p}\sum_{j=1}^{n_i}\left(Y_{i,j}-\bar{Y}_{i,\bullet}\right)^2\right]=\sum_{i=1}^{p}\sum_{j=1}^{n_i}\left(0+\frac{n_i-1}{n_i}\sigma^2\right)=$$

$$=\sum_{i=1}^{p}(n_i-1)\sigma^2=(n-p)\sigma^2.$$

Továbbá a csoportok közötti négyzetösszegre:

$$\mathrm{E}\left[Q_{csk}\right]=\sum_{i=1}^{p}\sum_{j=1}^{n_i}\mathrm{E}\left[(\bar{Y}_{i,\bullet}-\bar{Y}_{\bullet,\bullet})^2\right]=$$

$$=\sum_{i=1}^{p}\sum_{j=1}^{n_i}\left((\mathrm{E}\left[\bar{Y}_{i,\bullet}\right]-\mathrm{E}\left[\bar{Y}_{\bullet,\bullet}\right])^2+\frac{p-1}{p}D^2\bar{Y}_{i,\bullet}\right)=$$

$$=\sum_{i=1}^{p}\sum_{j=1}^{n_i}\left(\alpha_i^2+\frac{p-1}{p}\frac{\sigma^2}{n_i}\right)=\sum_{i=1}^{p}n_i\alpha_i^2+(p-1)\sigma^2$$

Így az $\alpha_1=\cdots=\alpha_p=0$ nullhipotézis mellett $\frac{Q_{csk}}{p-1}$ is torzítatlan becslése σ^2-nek. Tehát H_0 mellett $\frac{Q_{csk}}{p-1}$ és $\frac{Q_{csb}}{n-p}$ két független χ^2_{n-p} és χ^2_{p-1} eloszlású

miért

statisztika. Tehát e kettő hányadosa $F_{p-1,n-p}$ eloszlású lesz, ezt használjuk próbastatisztikának.

Példa: négy különböző búzafajtát termelünk, mindegyik fajtából 10-10 parcellányit. nincsenek meg

$p = 4$, $n_i = 10$, $F_{3,36}$ kritikus értéke 95%-on 2.87. az adatok

Ha H_0-t elvetjük, akkor adjunk becslést α_i-kre! A naiv pontbecslésnél jobb az intervallumbecslés:

Tudjuk: $S^2_{csb} = \frac{Q_{csb}}{n-p}$ és Q_{csb} eloszlása χ^2_{n-p} valamint ez független $\bar{Y}_{i,\bullet}$-től, mert $Q_{csb} = \sum_{i=1}^{p}\sum_{j=1}^{n_i}\left(Y_{i,j} - \bar{Y}_{i,\bullet}\right)^2$ és $\bar{Y}_{i,\bullet}$ független $\sum_{j=1}^{n_i}\left(Y_{i,j} - \bar{Y}_{i,\bullet}\right)^2$-től (Fisher-Bartlett). Azaz megállapítható, hogy

$$\frac{\bar{Y}_{i,\bullet} - \alpha_i}{S_{csb}} \sim t_{n-p}, \text{ azaz}$$

$$\mathrm{P}\left\{-t_{n-p}(\alpha/2) \leq \frac{\bar{Y}_{i,\bullet} - \alpha_i}{S_{csb}} \leq t_{n-p}(\alpha/2)\right\} = 1 - \alpha$$

5. fejezet

Idősorok

5.1. Alapfogalmak, definíciók

A továbbiakban olyan folyamatokat vizsgálunk, amelyeknél $X_1, \ldots, X_n$ nem független, azonos eloszlású változók.

11. Definíció. Véges dimenziós eloszlások $X_t : t \in T$
$t_1, \ldots, t_n \in T : (X_{t_1}, \ldots, X_{t_n})$
$\mathrm{P}\{X_{t_1} < x_1, \ldots, X_{t_n} < x_n\}$
$\mathrm{P}\{(X_{t_1}, \ldots, X_{t_n}) \in B\}, B \in \mathbb{R}^n$

10. Tétel (Kolmogorov). *Ha adott véges dimenziós eloszlásoknak egy kompatibilis rendszere, akkor létezik egy valószínűségi mező és azon egy sztochasztikus folyamat, amelynek pont ezek a véges dimenziós eloszlásai (vagyis véges dimenziós eloszlásai meghatározzák a folyamatot).*

12. Definíció. Gauss folyamat *Akkor nevezünk egy folyamatot* Gauss folyamat*nak, ha minden véges dimenziós eloszlása normális.*

5.1.1. Összefüggőségi struktúrák

1. véges rendű Markov-i összefüggés
2. *martingál tulajdonság*: $\mathrm{E}[X_{n+1}|X_1, \ldots, X_n] = X_n$
3. *stacionaritás*

13. Definíció. Erős stacionaritás *Egy sztochasztikus folyamatot akkor nevezünk* erősen stacionárius*nak ha*
$\forall t_1, \ldots, t_n, s : (X_{t_1}, \ldots, X_{t_n}) \stackrel{d}{=} (X_{t_1+s}, \ldots, X_{t_n+s})$

14. Definíció. Gyenge stacionaritás *Egy sztochasztikus folyamatot akkor nevezünk* gyengén stacionárius*nak ha*

1. $\forall t : \mathrm{E}[X_t] = \mathrm{E}[X_1]$
2. $\mathrm{Cov}(X_t, X_s) = \gamma(t-s)$

A γ függvényt a folyamat autokovariancia *függvényének nevezzük.*

5.1.2. Az autokovariancia függvény (γ) tulajdonságai

1. $\gamma(0) = D^2[X_t] \geq 0$ (ha létezik)
2. $|\gamma(h)| \leq \gamma(0)$

   Bizonyítás: Cauchy-Schwartz: $\langle a, b\rangle^2 \leq ||a||^2 ||b||^2$.

$$\begin{aligned} \gamma(h) = \quad & \operatorname{Cov}(X_t, X_{t+h}) \\ & |\operatorname{Cov}(X_t, X_{t+h})|^2 \quad \leq D^2[X_t] D^2[X_{t+h}] \end{aligned}$$

3. $\gamma(h) = \gamma(-h), \forall h \in \mathbb{Z}$
4. pozitív szemidefinit

15. Definíció (Pozitív Szemidefinit). *Az $M \in \mathbb{R}^{n\times n}$ mátrix* pozitív szemidefinit *ha $\forall a \in \mathbb{R}^n$ esetén $a^T M a \geq 0$.*

Ha egy egész számokon értelmezett u függvényre igaz, az, hogy $\forall t_1, \ldots, t_n \in \mathbb{Z}$ esetén a $[u(t_i - t_j)]_{1\leq i\leq n, 1\leq j\leq n}$ mátrix pozitív szemidefinit akkor a függvényt pozitív szemidefinit*nek nevezzük.*

11. Tétel. *Gyengén stacionárius folyamat autkovariancia függvénye pozitív szemidefinit.*

Bizonyítás: Legyen $Z := (X_{t_1} - \mathrm{E}[X], X_{t_2} - \mathrm{E}[X], \ldots, X_{t_n} - \mathrm{E}[X])^T$. Tudjuk, hogy $\mathrm{E}[Z] = 0$, ezért $\forall a \in \mathbb{R}^n$: $\mathrm{E}\left[a^T Z\right] = 0$. Ezért:

$$D^2[a^T Z] = \mathrm{E}\left[a^T Z a^T Z\right].$$

Mivel $a^T Z$ egy skalár, ezért egyenlő önmaga transzponáltjával, így:

$$D^2[a^T Z] = \mathrm{E}\left[a^T Z Z^T a\right] = a^T \mathrm{E}\left[Z Z^T\right] a.$$

A ZZ^T mátrix egy olyan (diadikus) mátrix melynek i,j-edik eleme $(X_{t_i} - \mathrm{E}[X])(X_{t_j} - \mathrm{E}[X])$, így az $\mathrm{E}\left[ZZ^T\right]$ mátrix i,j-edik eleme $\mathrm{E}[X_{t_i} - \mathrm{E}[X]](X_{t_j} - \mathrm{E}[X]) = \gamma(t_i - t_j)$. Tudjuk azonban, hogy mivel $D^2[a^T Z]$ egy valószínűségi változó szórását jelöli ezért nem lehet negatív, így $\forall a \in \mathbb{Z}^n$:

$$0 \leq D^2[a^T Z] = a^T \mathrm{E}\left[Z Z^T\right] a.$$

Vagyis az $\mathrm{E}\left[ZZ^t\right]$ mátrix pozitív szemidefinit és így a γ függvény is.

12. Tétel (Herglotz). *Legyen $X_1, \ldots, X_n, \ldots$ komplex értékű stacioner folyamat $\gamma(h) = \operatorname{Cov}(X_t, \bar{X_{t+h}})$ autokovariancia-függvénnyel. Ekkor*

$$\gamma(h) = \int_{-\pi}^{\pi} e^{ih\nu} dF(\nu),$$

ahol $F(\nu)$ a spektrálfüggvény, amelyre igaz, hogy $F(-\pi) = 0$, jobbról folytonos, korlátos.

5.2. Idősorok transzformációja

Klasszikus dekompozíció:

$$X_t = m_t + \Delta_t + Y_t,$$

ahol

m_t**:** trend, lassan változó determinisztikus,

Δ_t**:** szezonalitás, periodikus függvény,

Y_t**:** stacionárius folyamat.

5.2.1. Nincs periodikus komponens

Kiindulunk egy ismert trendfüggvényből: $m_t = a + bt$. A minták alapján a-t, és b-t úgy határozzuk meg, hogy a $\sum_{t=1}^{n}[X_t - (a+bt)]^2$ négyzetes eltérés minimális legyen.

Mozgó átlagos simítás (moving average smoothing)

$W_t = \frac{1}{2q+1}\sum_{j=-q}^{q} X_{t+j}$ a simított folyamat.

$$X_t \quad \to \quad \underbrace{\frac{1}{2q+1}\sum_{j=-q}^{q} m_{t+j}}_{\approx m_t} \quad + \quad \underbrace{\frac{1}{2q+1}\sum_{j=-q}^{q} Y_{t+j}}_{\approx 0}$$

$W_t = \hat{m}_t$ a trend becslése, feltéve, ha az *lineáris*! $\hat{Y}_t = X_t - \hat{m}_t$.

Exponenciális simítás (exponential smoothing)

Legyen $a \in (0,1)$.

$$\begin{aligned} \hat{m}_1 &:= X_1 \\ \hat{m}_t &:= aX_t + (1-a)\hat{m}_{t-1} \\ \hat{m}_t &= \sum_{j=0}^{t-2} a(1-a)^j X_{t-j} + (1-a)^t X_1 \end{aligned}$$

Különbségképzés (differencing)

Definiáljuk a különbségképző (differencing, ∇), illetve backward shift (B) operátorokat a következőképpen:

$$B : \mathbb{R}^{\mathbb{Z}} \mapsto \mathbb{R}^{\mathbb{Z}}, BX_t = X_{t-1},$$

illetve

$$\nabla : \mathbb{R}^{\mathbb{Z}} \mapsto \mathbb{R}^{\mathbb{Z}}, \nabla = 1 - B$$
$$\nabla X_t = (1-B)X_t = X_t - X_{t-1}$$

Ezek felhasználásával: $\nabla(at+b) = a$, illetve lineáris $m_t(= at+b)$ trend estén $\nabla(m_t + Y_t) = a + \nabla Y_t$.

Hasonlóképp definiálhatjuk a ∇^k operátort is:

$$\nabla^k = (1-B)^k = 1 - kB + k(k-1)B^2 + \cdots + (-B)^k.$$

Például : $\nabla^2 X_h = X_h - 2X_{h-1} + X_{h-2}$, illetve: $\nabla^k(a_k t^k + \cdots + a_0) = k! a_k$.

5.2.2. Trend és szezonalitás

Lassú trend

Legyen

$$X_{j,k} = Y_{k+12(j-1)},$$

ahol j jelentheti pl. az évet és k a hónapot. Feltesszük, hogy egy éven belül a trend konstans: m_j.

$$\hat{m}_j := \frac{1}{12}\sum_{k=1}^{12} X_{j,k}.$$

Szezonalitás becslése:

$$\hat{s}_k := \frac{1}{20}\sum_{j=1}^{20}(X_{j,k} - \hat{m}_j).$$

$$\sum_{k=1}^{12} \hat{s}_k = 0$$

A szezonalitás periódusát ismernünk kell! Ennek meghatározásához hasz-
ref nálhatjuk például a periodogram módszert.

Mozgó átlagos simítás

Trend $(\hat{m}_t)$ becslése:

$$\hat{m}_t = \begin{cases} \frac{1}{2q+1}\sum_{i=-q}^{q} X_{t+i} & : d = 2q+1 \\ \frac{1}{2q}(0.5(X_{t-q} + X_{t+q}) + \sum_{i=-q+1}^{q-1} X_{t+i}) & : d = 2q \end{cases}$$

$$\nabla_d X_t := X_t - X_{t-d} = X_t - B^d X_t = (1 - B^d)X_t$$

5.3. Tapasztalati autokovariancia és autokorreláció

Adott egy n elemű minta. Ekkor a *tapasztalati autokovariancia függvény*t a következőképpen definiáljuk :

$$\hat{\gamma}(h) := \frac{1}{n}\sum_{j=1}^{n-h}(X_{j+h} - \bar{X})(X_j - \bar{X}), \tag{5.1}$$

ahol $\bar{X}$ jelöli a mintaátlagot.

Bizonyítható, hogy az így definiált empirikus autokovariancia tagokból képzett $\Sigma = [\hat{\gamma}(i-j)]_{1\leq i,j\leq n}$ mátrix pozitív szemidefinit. Amennyiben a 5.1 egyenletben n helyett $(n-r)$-rel normálnánk, úgy a kapott empirikus autokovariancia mátrixra ez nem teljesülne.

Hasonlóképp értelmezhetjük az *emprikus autokorrelációs függvény*t is:

$$\hat{\rho}(h) := \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}.$$

Box és Jenkins ökölszabálya szerint $n \geq 50$ és $h \leq n/4$ esetén van értelme az idősor analízisével foglalkozni.

Fontos megjegyezni, hogy az empirikus autokovariancia ill. autokorreláció függvényeket nem-stacionárius esetben is ki tudjuk számítani, ezért a kapott eredmények értelmezésénél ezt figyelembe kell venni.

Ha például $\hat{\rho}(h)$ függvény lecsengése lassú, hatványfüggvény jellegű, az az erős függés helyett jelentheti lassú determinisztikus trend jelenlétét is.

Amennyiben a folyamatban szezonalitás van jelen ez a $\hat{\rho}(h)$ periodicitását eredményezi.

5.4. Parciális autokovariancia függvény

$$\begin{aligned} \alpha(1) &= \mathrm{Cov}(X_1, X_2) \\ \alpha(h) &= \mathrm{Cov}(X_1 - \mathrm{E}\left[X_1|X_2,\ldots,X_h\right], X_{h+1} - \mathrm{E}\left[X_{h+1}|X_2,\ldots,X_h\right]) \end{aligned}$$

$$\alpha(k) = \frac{\det R_k^*}{\det R_k},$$

ahol R_k az autokorreláció mátrix, $R_k = [\rho(i-j)]_{1\leq i,j\leq k}$, R_k^*-t pedig úgy kapjuk R_k-ból, hogy annak utolsó sorát a $[\rho_1,\ldots,\rho_k]$ vektorra cseréjük.

5.5. Fehér zaj

16. Definíció. Fehér zaj Fehér zaj*nak hívjuk, és* $WN(0,\sigma^2)$*-tel jelöljük azokat a folyamatokat, melyre*

$$\gamma(k) = \begin{cases} \sigma^2, & k = 0 \\ 0, & k \neq 0 \end{cases}$$

A fehér zaj spektrálfüggvénye konstans.

$$f(\lambda) = \frac{\sigma^2}{2\pi}, \lambda \in [-\pi, \pi]$$

5.6. Mozgóátlag (MA) folyamatok

17. Definíció. Mozgó átlag folyamat *Akkor nevezünk egy* X_t *folyamatot* mozgó átlag folyamat*nak ha felírható*

$$X_t = \Theta_0 e_t + \Theta_1 e_{t-1} + \cdots + \Theta_q e_{t-q}$$

alakban, ahol $e_t \sim WN(0,1)$ *fehér zaj,* $\Theta_i, 0 \leq i \leq q$ *pedig konstansok.*

X_t és e_t közötti összefüggést felírhatjuk a B backshift operátor segítségével:

$$\begin{aligned} X_t &= \Theta_0 e_t + \Theta_1 B e_t + \cdots + \Theta_q B^q e_t \\ &= (\Theta_0 + \Theta_1 B + \cdots + \Theta_q B^q) e_t \end{aligned}$$

Formálisan definiálhatjuk a $\Theta(z)$ polinomot a következőképp: $\Theta(z) := \Theta_0 + \Theta_1 z + \cdots + \Theta_q z^q$. Ezzel a jelöléssel:

$$X_t = \Theta(B) e_t$$

Az algebra alaptétele szerint egy n változós polinomnak pontosan n darab nem feltétlenül különböző gyöke van. Ennek alapján felírhatjuk $\Theta(z)$ gyöktényezős alakját:

$$\Theta(z) = \Theta_q \prod_{i=1}^{q} (z - z_i).$$

$$\mathrm{E}\left[X_r e_s\right] = \begin{cases} \Theta_{r-s} & : r - q \leq s \leq r \\ 0 & : \text{különben} \end{cases}$$

$$\mathrm{E}\left[X_{t+k} X_t\right] = \mathrm{E}\left[X_{t+k} \sum_{i=0}^{q} \Theta_i e_{t-i}\right] = \sum_{i=0}^{q} \Theta_i E[X_{t+k} e_{t-i}] = \sum_{i=0}^{q-k} \Theta_i \Theta_{k+i}$$

Mivel $\mathrm{E}\left[X_{t+k} X_t\right]$ t-től nem, csak k-tól függ, ezért X_t gyengén stacionárius, és az autokovariancia függvénye:

$$\gamma(k) = \begin{cases} \sum_{i=0}^{q-k} \Theta_i \Theta_{k+i} & : k \leq q \\ 0 & : k > q \end{cases} \tag{5.2}$$

Az előző egyenletben kifejeztük $\gamma(k)$ értékét a Θ_i értékek segítségével. Felvetődik a kérdés, hogy lehet-e ugyanezt visszefelé, illetve mi annak a feltétele, hogy adott $\gamma(k)$, $k = 0,1,2 \ldots, q$ autokovariancia függvényhez létezzenek a 5.2 egyenletet kielégítő Θ_i értékek.

Vagyis a kérdés: adott $\gamma(k)$, $k = 0,1,2,\ldots,q$ esetén megoldható-e $\gamma(k) = b_0 b_k + b_1 b_{k+1} + \ldots + b_{g-k} b_q$, $k = 0,1,2,\ldots,q$ egyenletrendszer.

A válasz pedig, hogy a megoldhatóság feltétele, hogy a

$$\Gamma(s) := \gamma(0) + \sum_{k=1}^{q} \gamma(k)(s^k + s^{-k})$$

függvénynek az az $|s| = 1$ egységkörön csak páros multiplicitású gyökei legyenek.

5.7. Autoregresszív (AR) folyamatok

18. Definíció. Autoregresszív folyamatok *Akkor nevezünk egy (X_t) folyamatot* autoregresszív*nek, ha felírható*

$$X_t = \Phi_1 X_{t-1} + \cdots + \Phi_p X_{t-p} + e_t$$

alakban, ahol e_t fehér zaj.

A mozgó átlag folyamatoknál definiált Θ függvényhez hasonlóan definiálhatjuk a $\Phi(z) := \Phi_0 + \Phi_1 z + \cdots + \Phi_p z^p$ polinomot. Így

$$\Phi(B) X_t = e_t.$$

5.7.1. Példa, AR(1) folyamat

$$X_t := \Lambda X_{t-1} + e_t = e_t + \Lambda(\Lambda X_{t-2} + e_{t-1}) = \sum_{j=0}^{k} \Lambda^j e_{t-j} + \Lambda^{k+1} X_{t-k-1}$$

$$\left|\left| X_t - \sum_{j=0}^{k} \Lambda^j e_{t-j} \right|\right|^2 = \Lambda^{2(k+1)} \, ||X_{t-k-1}||^2 \to 0$$

ha $|\Lambda| < 1$. Ebben az esetben azt mondjuk, hogy az AR folyamat *kauzális*. Ebben az esetben felírhatjuk az X_t-t

$$X_t = \sum_{j=0}^{\infty} \Lambda^j e_{t-j}$$

alakban is, ami egy MA(∞) folyamatnak felel meg. Ezt nevezzük az AR(1) folyamat kauzális MA(∞) előállításának.

Észrevehetjük, hogy amennyiben az AR(1) folyamat kauzális, azaz létezik MA(∞) előállítása, akkor a

$$\Phi(z) = 1 - \Lambda z$$

polinomnak az egyedüli gyöke az egységkörön kívül helyezkedik el. Általánosságban is igaz a következő

13. Tétel. *Egy AR(p) folyamatnak akkor és csak akkor létezik kauzális MA(∞) előállítása ha a $\Phi(z) = 0$ egyenletnek nincsen a $|z| \leq 1$ egységkörön belül gyöke.*

Tekintsük most azt az esetet amikor $|\Lambda| > 1$!

$$\begin{aligned} X_{t+1} &= \Lambda X_t + e_{t+1} \\ X_t &= \frac{1}{\lambda} X_{t+1} - \frac{1}{\Lambda} e_{t+1} = \\ &= -\sum_{j=1}^{k} \frac{1}{\Lambda^j} e_{t+j} + \frac{1}{\Lambda} x_{t+k} \end{aligned}$$

$|\Lambda| > 1$ esetén

$$X_t = -\sum_{j=1}^{\infty} \frac{1}{\Lambda^j} e_{t+j}$$

Vagyis $|\Lambda| > 1$ esetén is létezik MA(∞) előállítás, ez azonban nem kauzális.

AR(1) folyamatok autokovariancia - függvényét könnyen kifejezhetjük az MA(∞) előállításuk segítségével.

$$\mathrm{Cov}(X_{t+h}, X_t) = \lim_{n\to\infty} \mathrm{E}\left[\sum_{j=0}^{n} \Lambda^j e_{t+k-j} \sum_{k=0}^{n} \Lambda^k e_{t-k} \right] = \Lambda^h \sum_{j=0}^{\infty} \Lambda^{2j} = \frac{\Lambda^h}{1-\Lambda^2}$$

5.7.2. Yule-Walker egyenletek

$$\begin{aligned} X_t &= \Phi_1 X_{t-1} + \cdots + \Phi_p X_{t-p} + e_t \\ X_{t-k} X_t &= \Phi_1 X_{t-k} X_{t-1} + \cdots + \Phi_p X_{t-k} X_{t-p} + X_{t-k} e_t \\ \gamma(k) = \mathrm{E}\left[X_{t-k} X_t\right] &= \Phi_1 \gamma(k-1) + \cdots + \Phi_p \gamma(k-p) \\ \rho(k) &= \Phi_1 \rho(k-1) + \cdots + \Phi_p \rho(k-p) \end{aligned}$$

Ezen utóbbi egyenletet $k = 1,2,\ldots,p$ értékekre felírva és mátrix alakba rendezve kapjuk a *Yule-Walker* egyenletrendszert.

$$\begin{pmatrix} \rho_0 & \rho_1 & \cdots & \rho_{p-1} \\ \rho_1 & \rho_0 & \cdots & \rho_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p-1} & \cdots & \rho_1 & \rho_0 \end{pmatrix} \begin{pmatrix} \Phi_1 \\ \Phi_2 \\ \vdots \\ \Phi_p \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{pmatrix}$$

Ennek segítségével találhatunk adott $\rho(1),\ldots,\rho(p)$ értékekhez olyan $\Phi_1,\ldots,\Phi_p$ együtthatókat, hogy a kapott AR(p) folyamat autokorreláció - függvényének első p eleme az adott $\rho(1),\ldots,\rho(p)$-val egyezzen. Ugyanez az egyenlet használható a $\hat{\Phi}_1,\ldots,\hat{\Phi}_p$ értékek becslésére is a $\hat{\rho}_1,\ldots,\hat{\rho}_p$ segítségével.

5.8. Autoregresszív - mozgóátlag (ARMA) folyamatok

19. Definíció. *Az X_t folyamatot* ARMA(p,q) *folyamatnak nevezzük, ha*

$$\Phi(B) X_t = \Theta(B) e_t, \tag{5.3}$$

ahol $\Phi(B)$ p-ed és $\Theta(B)$ q-ad fokú polinomok.

Akkor nevezzük a folyamatot kauzálisnak, ha létezik MA(∞) előállítása. Egy kauzális ARMA folyamat autokovariancia - függvényét az MA(∞) előállításában szereplő együtthatókkal a következő tétel segítségével tudjuk kifejezni:

14. Tétel. *Amennyiben ξ_t stacionárius, és $\sum_{j=-\infty}^{\infty} |\Psi_j| < \infty$ akkor az $\eta_t := \sum_{j=-\infty}^{\infty} \Psi_j \xi_{t-j}$ függvény is stacionárius, és autokovariancia - függvénye:*

$$\gamma_\eta(h) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \Psi_j \Psi_k \gamma_\xi(h-j+k)$$

5.8.1. A kauzalitás szükséges és elégséges feltétele

Tegyük fel, hogy a $\Phi(z)$ és $\Theta(z)$ polinomoknak nincs közös gyöke. Ezt nyugodtan feltehetjük, mivel ellenkező esetben a (5.3) egyenletben ezzel a közös gyökkel egyszerűsíthetünk.

15. Tétel. *Az X_t ARMA(p,q) folyamat kauzalitásának szükséges és elégséges feltétele, hogy a $\Phi(z) = 0$ egyenletnek ne legyen a $|z| \leq 1$ egységkörön belül gyöke.*

$$\sum_{j=0}^{\infty} \Psi_j z^j = \frac{\Theta(z)}{\Phi(z)}, |z| \leq 1$$

20. Definíció. Invertálhatóság *Az X_t folyamatot* invertálhatónak *nevezzük, ha* $\exists \Pi_j : \sum_{j=0}^{\infty} |\Pi_j| < \infty$ *és* $e_t = \sum_{j=0}^{\infty} \Pi_j X_{t-j}$.

16. Tétel. *Az X_t ARMA(p,q) folyamat invertálhatóságának szükséges és elégséges feltétele, hogy a $\Theta(z) = 0$ egyenletnek ne legyen a $|z| \leq 1$ egységkörön belül gyöke.*

$$\sum_{j=0}^{\infty} \Pi_j z^j = \frac{\Phi(z)}{\Theta(z)}, |z| \leq 1$$

5.9. Az átlag és az autokovariancia becslései

$$\mathrm{E}[X_t] = \mu, \mathrm{Cov}[X_t, X_{t+n}] = \gamma(n)$$

Az átlag természetes becslése

$$\bar{X_n} = \frac{X_1 + \ldots X_n}{n}, \text{ erre } \mathrm{E}\left[\bar{X_n}\right] = \mu$$

Az átlag szórásnégyzete

$$nD^2(\bar{X_n}) = \frac{1}{n} \sum_{i,j=1}^{n} \mathrm{Cov}[X_i, X_j] = \sum_{i,j=1}^{n} \frac{1}{n} \gamma(i-j) =$$

$$= \sum_{h=-n+1}^{n-1} \sum_{j=1}^{n-|h|} \frac{1}{n} \gamma(h) = \sum_{|h|<n} \frac{n-|h|}{n} \gamma(h) \leq \sum_{|h|<n} |\gamma(h)|$$

Ha $\gamma(n) \to 0$, akkor

$$D^2(\bar{X_n}) \leq \sum_{|h|<n} |\gamma(h)| \to 0.$$

Ha pedig még az is igaz, hogy a $\gamma(n)$ sorozat abszolút konvergens, akkor az átlag szórásnégyzetére az alábbi aszimptotikát adhatjuk

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty \Rightarrow nD^2(\bar{X_n}) \to \sum_{h=-\infty}^{\infty} \gamma(h)$$

5.9.1. A spektrálfüggvény és az autokovariancia kapcsolata

17. Tétel (Inverz Fourier transzformált).

$$Ha \sum_{n=-\infty}^{\infty} |K(n)| < \infty, \text{ akkor } K(h) = \int_{-\pi}^{\pi} e^{ih\nu} f(\nu) d\nu, \text{ ahol}$$

$$f(\lambda) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{in\lambda} K(n), \text{ ugyanis}$$

$$\int_{-\pi}^{\pi} e^{ih\nu} f(\nu) d\nu = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} K(n) \int_{-\pi}^{\pi} e^{i(h-n)\nu} = K(h).$$

Ennek egyszerű következménye az alábbi állítás

$$\text{Egy } \sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty \text{ tulajdonságú } \gamma \text{ függvény autokovariancia függvény } \Leftrightarrow$$

$$\Leftrightarrow f(\lambda) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{-in\lambda} \gamma(n) \geq 0.$$

Ennek alkalmazásaként számítsuk ki, mikor lesz az alábbi alakú K függvény autokovariancia függvény!

$$K(h) = \begin{cases} 1, \text{ ha } h = 0 \\ \rho, \text{ ha } h = \pm 1 \\ 0 \text{ különben.} \end{cases}$$

$$f(\lambda) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{-in\lambda} K(n) = \frac{1}{2\pi} (1 + 2\rho \cos \lambda) \geq 0 \Rightarrow |\rho| \leq \frac{1}{2}$$

5.9.2. Aszimptotikus normalitás

Legyen például

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \Psi_j Z_{t-j}, \text{ ahol}$$

Z_t független, azonos eloszlású 0 várható értékkel és σ^2 szórásnégyzettel, továbbá

$$\sum_{j=-\infty}^{\infty} |\Psi_j| < \infty, \text{ de } \sum_{j=-\infty}^{\infty} \Psi_j \neq 0.$$

Ekkor

$$\bar{X_n} \xrightarrow{d} N\left(\mu, \frac{1}{n} \sum_{j=-\infty}^{\infty} \Psi(n)\right)$$

5.9.3. $\gamma(n)$ becslése

$$\widehat{\gamma(n)} = \frac{1}{n}\sum_{t=1}^{n-h}\left(X_t - \bar{X_n}\right)\left(X_{t+h} - \bar{X_n}\right), \text{ ahol } 0 \leq h \leq n-1$$

Ez általánosságban torzított becslés (bár bizonyos további feltételek mellett aszimptotikusan torzítatlan), de viszont a $\widehat{\Gamma_n} = \left[\widehat{\gamma(i-j)}\right]_{1\leq i,j\leq n}$ mátrixa pozitív szemidefinit.

Ennek bizonyításához elég, hogy $\widehat{\Gamma_n} = \frac{1}{n}MM^T$ a következő az $Y_i = X_i - \bar{X_n}$ jelöléssel kifejezett M mátrixszal

$$M = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 & Y_1 & Y_2 & \dots & Y_{n-1} & Y_n \\ 0 & 0 & \dots & 0 & Y_1 & Y_2 & Y_3 & \dots & Y_n & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & Y_1 & \dots & Y_{n-2} & Y_{n-1} & Y_n & 0 & \dots & 0 & 0 \end{bmatrix}.$$

Általános ökölszabályként elmondhatjuk, hogy $\rho(h)$ becslése $\left(\widehat{\rho(h)} = \frac{\widehat{\rho(h)}}{\widehat{\rho(0)}}\right)$ akkor jó, ha $n \geq 50$ és $h \leq \frac{n}{4}$.

5.9.4. Az autokorrelációk mikor különböznek szignifikánsan 0-tól?

Ha a következő alakú szűrt független zajra

$$X_t - \mu = \sum_{j=-\infty}^{\infty}\Psi_j Z_{t-j}, \text{ és}$$

Z_t független, azonos eloszlású 0 várható értékkel és σ^2 szórásnégyzettel, továbbá

$$\sum_{j=-\infty}^{\infty}|\Psi_j| < \infty, \text{ és } \mathrm{E}\left[Z_i^4\right] < \infty, \text{ akkor}$$

$$\left[\widehat{\rho(1)}, \dots \widehat{\rho(h)}\right] \xrightarrow{d} N\left(\left[\rho(1), \dots \rho(h)\right], \frac{1}{n}W\right), \text{ ahol } W \text{ az ún Bartlett mátrix}$$

$$W_{i,j} = \sum_{k=1}^{\infty}\left[\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k)\right]\left[\rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k)\right].$$

Például a független fehér zaj folyamatra $\rho(l) \neq 0$, ha $l \neq 0$, azaz

$$W_{i,j} = \begin{cases} 1, \text{ ha } i = j \\ 0, \text{ ha } i \neq j \end{cases} \quad \text{vagyis}$$

$$\widehat{\rho(1)}, \dots, \widehat{\rho(h)} \approx \text{i.i.d } N\left(0, \frac{1}{n}\right),$$

ennek konfidenciaintervalluma $\pm 1.96\frac{1}{\sqrt{n}}$, amely értéket a normális eloszlás táblázatából olvashatunk ki.

5.10. *ARMA* modellek becslései

Amikor egy folyamatot *ARMA* modellel közelítünk, a következő lépések szerint járunk el:

1. megbecsüljük p-t és q-t, az *ARMA* folyamathoz tartozó két polinom fokszámát
2. megbecsüljük a polinomok együtthatóit
3. megbecsüljük a szórásnégyzetet

5.10.1. Ismert p és q

Tiszta autoregresszív esetben felírhatjuk a Yule-Walker egyenleteket:

$$X_t - \Phi_1 X_{t-1} - \cdots - \Phi_p X_{t-p} = e_t, \text{ ahol } e_t \sim WN(0, \sigma^2)$$

$$\Gamma_p \Phi = \gamma_p, \text{ ahol}$$

$$\Gamma_p = [\gamma(i-j)]_{i,j=1}^{p}$$
$$\gamma_p^T = [\gamma(1), \ldots, \gamma(p)]$$
$$\Phi^T = [\Phi(1), \ldots, \Phi(p)].$$

Továbbá

$$\sigma^2 = D^2 e_t = D^2(X_t - \Phi_1 X_{t-1} - \cdots - \Phi_p X_{t-p}) = \gamma(0) - \Phi^T \gamma_p.$$

Így a Yule-Walker becslések a következő alakúak lesznek:

$$\widehat{\Gamma_p}\widehat{\Phi} = \widehat{\gamma_p} \text{ és}$$
$$\widehat{\sigma^2} = \widehat{\gamma(0)} - \widehat{\Phi}^T \widehat{\gamma_p}.$$

5.10.2. Ismeretlen p

Ha p nem ismert és $AR(m)$-et próbálunk illeszteni, akkor azt várjuk, hogy $\widehat{\Phi_{m,m}}$ kicsi lesz.

$$\sqrt{n}\left(\widehat{\Phi_m} - \Phi_m\right) \xrightarrow{d} N_m(0, \sigma^2 \Gamma_m^{-1}),$$

ahol Φ_m az X_{m+1} legjobb lineáris közelítésének együtthatóvektora:

$$\left\| X_{m+1} - \Phi_m^T (X_1, \ldots, X_m) \right\| \to min.$$

5.10.3. A Durbin-Levinson algoritmus

A mátrixinvertálás kikerülésére a Durbin-Levinson algoritmust használjuk. Legyenek az m-ed rendű illesztés együtthatói

$$\widehat{\Phi}_m = \left(\widehat{\Phi}_{m,1}, \widehat{\Phi}_{m,2}, \dots \widehat{\Phi}_{m,m}\right) = \widehat{R}_m^{-1}\widehat{\rho}_m \text{ és}$$

$$\widehat{v}_m = \widehat{\gamma}(0)\left[1 - \widehat{\rho}_m^T \widehat{R}_m^{-1}\widehat{\rho}_m\right].$$

Ekkor $\widehat{\Phi}_{1,1} = \widehat{\rho}(1)$ és $\widehat{v}_1 = \widehat{\gamma}(0)\left[1 - \widehat{\rho}^2(1)\right]$. Továbbá a becsült parciális autokovariancia függvény

$$\widehat{\Phi}_{m,m} = \left[\widehat{\gamma}(m) - \sum_{j=1}^{m-1} \widehat{\Phi}_{m-1,j}\widehat{\gamma}(m-j)\right] / \widehat{v}_{m-1}$$

$$\begin{bmatrix} \widehat{\Phi}_{m,1} \\ \vdots \\ \widehat{\Phi}_{m,m-1} \end{bmatrix} = \widehat{\Phi}_{m-1} - \widehat{\Phi}_{m,m} \begin{bmatrix} \widehat{\Phi}_{m-1,m-1} \\ \vdots \\ \widehat{\Phi}_{m-1,1} \end{bmatrix}, \text{ és}$$

$$\widehat{v}_m = \widehat{v}_{m-1}\left(1 - \widehat{\Phi}_{m,m}^2\right).$$

Elméletileg $\alpha(m) = \Phi_{m,m} = 0$, ha $m > p$, gyakorlatilag $\sqrt{n}\widehat{\Phi}_{m,m} \to N(0,1)$, azaz $\mathrm{P}\left\{-1.96\frac{1}{\sqrt{n}} < \widehat{\Phi}_{m,m} < 1.96\frac{1}{\sqrt{n}}\right\} = 0.95$. A rendre ezzel előzetes becslést adhatunk: $\widehat{p} = \min\left\{r : \forall m > r \;\; |\widehat{\Phi}_{m,m}| < 1.96\frac{1}{\sqrt{n}}\right\}$.

5.10.4. Az innovációs algoritmus

A Gram-Schmidt ortogonalizációs eljárással független vektorokból ortogonális rendszert készíthetünk vetítésekkel. Az eljárást idősorokra is alkalmazhatjuk a következő módon:

$$\mathrm{E}[X]_t = 0, \text{ és } \kappa(i,j) := \mathrm{E}[X_i X_j]$$

$$H_n := \langle X_1, \dots, X_n\rangle = \langle X_1 - \hat{X}_1, \dots, X_n - \hat{X}_n\rangle, \text{ ahol } \hat{X}_{n+1} := pr_{H_n} X_{n+1}.$$

$$\hat{X}_{n+1} = \begin{cases} 0, & n = 0 \\ \sum_{j=0}^{n} \Theta_{n,j}\left(X_{n-j+1} - \hat{X}_{n-j+1}\right), & n \neq 0 \end{cases}.$$

A Θ együtthatók rekurzív kiszámítását a v segédváltozóval (szórásnégyzet) a következő rendszer adja meg:

$$v_n := \left\|X_{n+1} - \hat{X}_{n+1}\right\|^2, \text{ így } v_0 = \kappa(1,1)$$

$$\Theta_{n,n-k} = \frac{1}{v_k}\left[\kappa(n+1,k+1) - \sum_{j=0}^{k-1} \Theta_{k,k-j}\Theta_{n,n-j}v_j\right], \text{ ahol } k = 0..n-1$$

$$v_n = \kappa(n+1, n+1) - \sum_{j=0}^{n-1} \Theta_{n,n-j}^2 v_j$$

Ennek bizonyítását úgy kezdjük, hogy $0 \leq k \leq n$ esetén $X_1 - \hat{X}_1, \ldots, X_n - \hat{X}_n$ ortogonális, hiszen $X_i - \hat{X}_i \in H_{j-1}$, ha $i < j$, és $X_j - \hat{X}_j \perp H_{j-1}$. Tehát $\hat{X}_{n+1}$ definícióját használva

$$\langle \hat{X}_{n+1}, X_{k+1} - \hat{X}_{k+1} \rangle = \Theta_{n,n-k} v_k,$$

amely egyenlethez a $X_{n+1} - \hat{X}_{n+1} \perp X_{k+1} - \hat{X}_{k+1}$ azonosságot adva kapjuk, hogy

$$\langle X_{n+1}, X_{k+1} - \hat{X}_{k+1} \rangle = \Theta_{n,n-k} v_k.$$

Ebbe $\hat{X}_{k+1}$ defnícióját írva

$$\Theta_{n,n-k} = \frac{1}{v_k} \langle X_{n+1}, X_{k+1} - \sum_{j=0}^{k-1} \Theta_{k,k-j} \left(X_{j+1} - \hat{X}_j + 1 \right) \rangle =$$

$$\frac{1}{v_k} \left[\kappa(n+1, k+1) - \sum_{j=0}^{k-1} \Theta_{k,k-j} \Theta_{n,n-j} v_j \right], \text{ továbbá}$$

$$v_n = \left| X_{n+1} - \hat{X}_{n+1} \right|^2 = |X_{n+1}|^2 + \left| \hat{X}_{n+1} \right|^2 =$$

$$\kappa(n+1, k+1) - \sum_{j=0}^{n-1} \Theta_{n,n-j}^2 v_j$$

Például a MA(1) folyamat predikcióját így adhatjuk meg:

$$X_t = Z_t + \Theta Z_{t-1}, \text{ ahol } Z_t \sim WN(0, \sigma^2)$$

$$\kappa(i,j) = \begin{cases} \sigma^2(1+\Theta^2) & i = j \\ \Theta\sigma^2 & i = j+1 \\ 0 & \text{különben} \end{cases}$$

$$v_0 = \kappa(1,1) = \sigma^2(1+\Theta^2),$$

$$\Theta_{n,j} = \begin{cases} \frac{1}{v_{n-1}} \Theta\sigma^2 & j = 1 \\ 0 & 2 \leq j \leq n \end{cases}$$

$$v_n = (1+\Theta^2)\sigma^2 - \frac{1}{v_{n-1}^2} \Theta^2 \sigma^4,$$

$$r_n := \frac{v_n}{\sigma^2} = (1+\Theta^2) - \frac{1}{v_{n-1}^2} \Theta^2 \sigma^2.$$

Tehát a predikció:

$$\hat{X}_{n+1} = \frac{\Theta}{r_{n-1}} (X_n - \hat{X}_n)$$

5.10.5. Mozgóátlag folyamatok becslései

Az $X_1, \ldots, X_n$ adatokra a következő előfeltevést tesszük annak analógiájára, hogy az $X\hat{X}$ mennyiségek voltak a hibák:

$$X_t = Z_t + \hat{\Theta}_{m,1} Z_{t-1} + \cdots + \hat{\Theta}_{m,m} Z_{t-m}, \text{ ahol } Z_t \sim WN(0, \hat{v}_m^2).$$

Ha $\hat{\gamma}(0) > 0$, akkor vezessük be a becsült együtthatók vektorára a $\hat{\Theta}_m = \left(\hat{\Theta}_{m,1}, \ldots, \hat{\Theta}_{m,m}\right)$ jelölést! Ezekre a következő rekurzív becslés érvényes:

$$\hat{v}_0 = \hat{\gamma}(0) \text{ és}$$

$$\hat{\Theta}_{m,m-k} = \hat{v}_k^{-1} \left[\hat{\gamma}(m-k) - \sum_{j=0}^{k-1} \hat{\Theta}_{m,m-j} \hat{\Theta}_{k,k-j} \hat{v}_j \right]$$

$$\hat{v}_m = \hat{\gamma}(0) - \sum_{j=0}^{k-1} \hat{\Theta}_{m,m-j}^2 \hat{v}_j, \text{ ahol } k = 0, \ldots, m-1$$

5.10.6. Aszimptotikus viselkedés ARMA folyamatok esetén

A jelölések rövid leírása a következő:

$$\Phi(B) X_t = \Theta(B) Z_t, \text{ ahol } Z_t \sim IID(0, \sigma^2) \text{ és } \mathrm{E}\left[Z_t^4\right] < \infty$$

$$\Psi(z) = \sum_{j=0}^{\infty} \frac{\Theta(z)}{\Phi(z)}, \ |z| \leq 1, \ \Psi_0 = 1$$

Ekkor minden k-ra

$$\sqrt{n} \left[\hat{\Theta}_{m,1} - \Psi_1, \ldots, \hat{\Theta}_{m,k} - \Psi_k \right] \xrightarrow{d} N(0, A), \text{ ahol}$$

$$A_{i,j} = \sum_{r=1}^{min(i,j)} \Psi_{i-r} \Psi_{j-r}, \text{ továbbá}$$

$$m(n) \to \infty \text{ úgy, hogy } m(n) = o(\sqrt[3]{n}), \text{ és } \hat{v}_m \xrightarrow{p} \sigma^2.$$

Itt jegyezzük meg, hogy $AR(p)$ esetben a Durbin-Levinson algoritmus által a Φ_p-re adott $\hat{\Phi}_p = \left(\hat{\Phi}_{p,1}, \ldots, \hat{\Phi}_{p,p}\right)$ becslés konzisztens, ha $n \to \infty$. Viszont $MA(k)$ esetben az innovációs algoritmus által adott $\hat{\Theta}_q = \left(\hat{\Theta}_{q,1}, \ldots, \hat{\Theta}_{q,q}\right)$ becslés nem konzisztens, viszont a $\left(\hat{\Theta}_{m,1}, \ldots, \hat{\Theta}_{m,q}\right)$ már az.

A gyakorlatban $MA(q)$ esetben tudjuk, hogy $\rho(m) = 0$, ha $m > q$, és Bartlett tétele miatt

$$\hat{\rho}(m) \overset{???}{\to} N\left(0, \frac{1}{n} \sum_{i=-q}^{???} \rho(i)\right).$$

*

5.10.7. Maximum likelihood becslések

$\mathrm{E}[X_t] = 0$ tulajdonságú Gauss folyamat esetén a $\Gamma_n = \mathrm{E}\left[\underline{X}_n \underline{X}_n^T\right]$ jelöléssel a likelihood függvény a következő:

$$L(\Gamma_n) = \frac{1}{(2\pi)^{n/2}} \frac{1}{(\det \Gamma_n)^{1/2}} \exp\left(-\frac{1}{2} \underline{X}_n^T \Gamma_n^{-1} \underline{X}_n\right).$$

A mátrixinvertálás és a determinánsszámítás kikerülésére a következő algoritmus javasolt: $k = 0, \ldots, n-1$ esetén

$$\hat{X}_{k+1} = \sum_{j=0}^{k-1} \Theta_{k,k-j} \left(X_{j+1} - \hat{X}_{j+1}\right), \text{ azaz}$$

$$\begin{pmatrix} \hat{X}_1 \\ \vdots \\ \hat{X}_n \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 \\ \Theta_{1,1} & 0 & 0 & \ldots & 0 \\ \Theta_{2,2} & \Theta_{2,1} & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Theta_{n-1,n-1} & \Theta_{n-1,n-2} & \Theta_{n-1,n-3} & \ldots & 0 \end{pmatrix} \begin{pmatrix} X_1 - \hat{X}_1 \\ \vdots \\ X_n - \hat{X}_n \end{pmatrix}.$$

A képletben szereplő mártixot jelöljük a következőképpen

$$\bar{C} = [\Theta_{i,i-j}]_{i,j=0}^{n-1}, \text{ ahol } j \leq 0 \text{ esetén } \Theta_{i,j} = 0.$$

Ezt a mátrisot módosítsuk úgy, hogy a főátlóba 1-eket írunk: $C := \bar{C} + Id$. Ekkor

$$C\left(\underline{X}_n - \underline{\hat{X}}_n\right) = (\bar{C} + Id)\left(\underline{X}_n - \underline{\hat{X}}_n\right) = \underline{\hat{X}}_n + \underline{X}_n - \underline{\hat{X}}_n = \underline{X}_n, \text{ azaz}$$

$$\Gamma_n = \mathrm{E}\left[X_n X_n^T\right] = C\mathrm{E}\left[\left(\underline{X}_n - \underline{\hat{X}}_n\right)\left(\underline{X}_n - \underline{\hat{X}}_n\right)^T\right] C^T = CDC^T, \text{ ahol}$$

$D = \mathrm{diag}(v_0, v_1, \ldots, v_{n-1})$, azaz a determináns egyszerűen így számítandó:

$$\det \Gamma_n = (\det C)^2 \det D = v_0 v_1 \ldots v_{n-1}.$$

A kitevő is egyszerűbb alakra hozható:

$$\underline{X}_n^T \Gamma_n^{-1} \underline{X}_n = \left(\underline{X}_n - \underline{\hat{X}}_n\right)^T C^T \Gamma^{-1} C \left(\underline{X}_n - \underline{\hat{X}}_n\right) =$$

$$\left(\underline{X}_n - \underline{\hat{X}}_n\right)^T C^T C^{T^{-1}} D^{-1} C^{-1} C \left(\underline{X}_n - \underline{\hat{X}}_n\right) =$$

$$\left(\underline{X}_n - \underline{\hat{X}}_n\right)^T D^{-1} \left(\underline{X}_n - \underline{\hat{X}}_n\right) = \sum_{j=1}^{n} \frac{\left(X_j - \hat{X}_j\right)^2}{v_{j-1}}$$

Tehát végeredményképpen a likelihood függvény legegyszerűbb alakja:

$$L(\Gamma_n) = (2\pi)^{-n/2}(v_0 v_1 \dots v_{n-1})^{-1/2} \exp\left[-\frac{1}{2}\sum_{j=1}^{n}\frac{\left(X_j - \hat{X}_j\right)^2}{v_{j-1}}\right].$$