1. Mérték és integrálelmélet

1.1. Mérhető tér, mérhető leképzések, mérték

1.1.1. Alapfogalmak

Nemüres halmazokból képező függvények, a nulla eleme a halmazoknak

- Egy α halmazfüggvény **Monoton**, ha $A, B \subset \mathcal{A}$ és $A \subset B$ akkor $\alpha(A) < < \alpha(B)$
- Egy α halmazfüggvény **Additív**, ha diszjunkt $A, B \subset \mathcal{A}$ ahol $A \cup B \subset \mathcal{A}$ esetében $\alpha(A \cup B) = \alpha(A) + \alpha(B)$
- Egy α halmazfüggvény σ**-Additív**, ha páronként diszjunkt $A_1, A_2, \ldots \subset \subset \mathcal{A}$ ahol $\cup_{k=1}^{\infty} A_k \subset \mathcal{A}$ esetében $\alpha(\cup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} \alpha(A_k)$

Algebrai rendszerek:

- Az A rendszer **Modulus**, ha $A, B \in \mathcal{A}$ esetén $A \backslash B \in \mathcal{A}$
- Az A rendszer **Félgyűrű**, ha $A, B \in \mathcal{A}$ esetén $A \cap B \in \mathcal{A}$
- Az A rendszer **Algebra**, ha $A \in \mathcal{A}$ esetén $X \backslash A \in \mathcal{A}$
- Az A rendszer σ**Algebra**, ha $\emptyset, X \in \mathcal{A}$ és ha $A \in \mathcal{A}$ Akkor $A^C \in \mathcal{A}$ és megszámlálható $A_1, \ldots \in \mathcal{A}$-ra $\cup_{i=1}^{\infty} \in \mathcal{A}$

1.1. Definíció (Mérhető tér)**.**

1.2. Definíció (Mérhető leképzés)**.**

1.3. Definíció (Monoton halmazfüggvény)**.**

1.4. Definíció (Additív halmazfüggvény)**.**

1.5. Definíció (σ-Additív halmazfüggvény)**.**

1.2. Integrál, függvénysorozatok, függvénysorok és integráltjaik

1.3. Mértékek kiterjesztése

1.4. Lebesgue- és Lebesgue-Stieltjes mérték

1.5. Riemann- és Riemann-Stieltjes integrál, modern kontextusban, Mértéktartó leképezések

1.6. Előjeles mértékek és variációik, felbontások

1.7. Abszolút folytonos és szinguláris mértékek, Lebesgue-felbontás, Radon-Nikodym tétel

1.8. Mértékek differenciálása

1.9. Korlátos változású, abszolút folytonos és szinguláris függvények, felbontásuk, differenciálásuk

1.10. Mértékterek szorzata

2. Valószínűségszámítás

2.1. Valószínűségi mező, változók, alapvető objektumok. Eloszlások, sűrűségfüggvények, transzformációk speciális esetei. Doob-lemma

2.1. Definíció (Valószínűségi tér)**.** $(\Omega, \mathcal{A}, P)$ hármast valószínűségi térnek hívjuk. Ezek a következőek:

$$(\underbrace{\Omega}_{\text{Elemi események tere}}, \underbrace{\mathcal{A}}_{\text{Eseménytér}}, \underbrace{P}_{\text{Valószínűségi mérték}})$$

Fontos, hogy $P(\Omega) = 1$

2.2. Definíció (Valószínűségi változó)**.**

$$X : \Omega \underset{\text{mérhető fv.}}{\longrightarrow} (\mathfrak{X}, \mathcal{B})$$

2.3. Definíció (Valószínűségi változó eloszlása x-re)**.**

$$Q_x(B) = P(X^{-1}(B)),\ \forall B \in \mathcal{B}$$

2.1. Lemma. $(\mathfrak{X}, \mathcal{B}, Q_x)$ mértéktér, és valószínűségi tér is.

$$Q_x(\mathfrak{X}) = P(X^{-1}(\mathfrak{X})) = P(\Omega) = 1$$

Több valószínűségi változó esetében mértékterek szorzata a valószínűségi változó.

Vannak $(\Omega_i, \mathcal{A}_i, P_i),\ i \in I$ valószínűségi tereink és $X_i : \Omega_i \to (\mathfrak{X}_i, \mathcal{B}_i)$ valószínűségi változóink. Ekkor a valószínűségi terünk $(\Omega, \mathcal{A}, P) = \bigtimes_{i \in I}(\Omega_i, \mathcal{A}_i, P_i)$ És a $(\mathfrak{X}, \mathcal{B}) = \bigtimes_{i \in I}(\mathfrak{X}_i, \mathcal{B}_i)$

2.2. Lemma.

$$X \text{ mérhető } \Leftrightarrow \forall i\ X_i \text{ mérhető}$$

2.3. Lemma (Marginálisok)**.** $X : \Omega \to \mathfrak{X}$ valószínűségi változóból indulva $X_i = \prod_i(X)$ (csak az i indexet választja ki)

$$X_i : \Omega_i \to \mathfrak{X}_i$$

valószínűségi változó lesz, neve marginális.

2.4. Definíció (Valószínűségi változó eloszlása)**.** Valós, $\mathbb{R}^n$ -ben.
$n = 1$ esetében:

$$Q_x((-\infty, t)) = \underbrace{P(x < t)}_{P(\underbrace{\{\omega : X(\omega) < t\}}_{\in \mathcal{A}})} =: F_x(t)$$

$n > 1$ esetében:

$$Q_x((-\infty, t_1) \times \ldots (-\infty, t_n)) = P(X_1 < t_1, \ldots, X_n < t_n) =: P_x(t_1, \ldots, t_n)$$

Tulajdonságai

- $n = 1$
  - monoton növő
  - $\lim_{x\to-\infty} F_x(x) = 0$
  - $\lim_{x\to\infty} F_x(x) = 1$
  - balról folytonos
- $n > 1$
  - monoton növő minden koordinátában
  - $\lim_{\min(x_i)\to-\infty} F_x(\underline{x}) = 0$
  - $\lim_{\min(x_i)\to\infty} F_x(\underline{x}) = 1$
  - balról folytonos minden koordinátában
  - $\forall a < b \in \mathbb{R}^n$ (minden koordinátában)

2.5. Definíció (Sűrűségfüggvény)**.** $X : \Omega \to \mathfrak{X}$, $(\mathfrak{X}, \mathcal{B}, \mu)$ μ : σ-véges X a μ-re abszolút folytonos, ha $Q_x << \mu$, ekkor a sűrűségfüggvénye

$$f = \frac{dQ_x}{d\mu}$$

Tulajdonságai: $f \geq 0$ m.m., $\int_{\mathfrak{X}} f d\mu = 1$

2.4. Lemma (Doob-lemma)**.** $h : \mathfrak{X} \to y$ és $g : \mathfrak{X} \to \mathbb{R}$ mérhető függvények, g mérhető $\sigma(h)$-ra (a legkisebb σ-algebra, amire h mérhető)

$$\exists l : y \to \mathbb{R} \text{ mérhető, } g = l \circ h$$

2.2. Függetlenség. Kapcsolat eloszlással. Aszimptotikus állítások, Kolmogorov 0-1 törvény, Borel-Cantelli lemma. Konvolúció

2.6. Definíció (Függetlenség)**.** – $A, B \in \mathcal{A} fggetlen, ha P(A \cap B) = P(A) \cdot P(B)$

- $A_1, \ldots, A_n \in \mathcal{A}$ (együttes) függetlenek, ha $P(A_1 \cap A_2 \ldots A_N) = \prod_j P(A_j)$ minden $\{l_1, \ldots, l_j\} \subseteq 1, n$-re
- R_γ eseményhalmazok rendszere ($p \in \Gamma$) függetlenek, ha $\forall A_{\gamma_1}, A_{\gamma_n}$ független, $A_{\gamma_i} \in R_{\gamma_i} \quad \{\gamma_1, \ldots, \gamma_k\} \subseteq \Gamma$
- $X_{\gamma_i} : \Omega \to \mathfrak{X}_\gamma$ v.v. függetlenek, ha $\underbrace{P(X_\gamma)}_{R_\gamma}$ függetlenek.

2.5. Tétel. $X = (X_1, \ldots, X_n), \omega \to \mathbb{R}^n$ esetén ekvivalensek:

- $X_1, \ldots, X_n$ függetlenek

– $Q_x = Q_{x_1} \times \ldots \times Q_{x_n}$

– $F_x(t_1, \ldots, t_n) = F_{x_1}(t_1) \cdot \ldots \cdot F_{x_n}(t_n)$

2.6. Tétel. X, mint előbb T.f.h, abszolút folytonos $\mathbb{R}^n$-en $\Rightarrow \forall X_i$ abszolút folytonos. Ekkor:

$$f_x(t_1, \ldots, t_n) = f_{x_1}(t_1) \cdot \ldots \cdot f_{x_n}(t_n)$$

Megjegyzés: visszafele nem igaz!

2.3. Várható érték, magasabb momentumok. Kapcsolódó egyenlőtlenségek (Jensen és társai)

2.4. Valószínűségi változók konvergenciája, 1 valószínűségű, L_p-beli és sztochasztikus konvergencia, kapcsolatuk és eszközeik.

2.5. Valószínűségi mértékek gyenge konvergeniája. Valószínűségi változók eloszlásbeli konvergenciája. Karakterisztikus függvény

2.6. Nagy számok gyenge és erős törvényei

2.7. Centrális határeloszlás tétel változatai.

2.7. Tétel (Centrális határeloszlás)**.** $X_1, X_2, \ldots$ függelten azonos eloszlású valószínűségi változók, $E(X) = m$, $D^2(X) = \sigma^2$, $S_n := X_1 + \cdots + X_n$ ekkor:

$$\underbrace{\frac{S_n - nm}{\sigma\sqrt{n}}}_{Z_n} \to \mathcal{N}(0{,}1) \text{ eloszlásban} \tag{1}$$

Bizonyítás.

$$\varphi_{\frac{X_1 - m}{\sigma\sqrt{n}}}(t) = \varphi_{X_1-m}\left(\frac{t}{\sigma\sqrt{n}}\right)$$

A karakterisztikus függvényt a Taylor sorára bontjuk

$$\varphi_{X_1-m}(s) \underbrace{=}_{0 \text{ körül}} 1 + \underbrace{0}_{E(\varphi(s))} s - \frac{\sigma^2 s^2}{2} + \mathcal{O}(s^2)$$

($\mathcal{O}(s^2)$ az s^2-nél kisebb nagyságrendű tagok)

$$\varphi_{Z_n}(t) = (1 - \frac{\sigma^2}{2} \cdot \frac{t^2}{\sigma^2 n} + \mathcal{O}(\frac{t^2}{n}))^n == (1 - \frac{t^2}{2n} + \mathcal{O}(\frac{t^2}{n}))^n \to e^{\frac{-t^2}{2}} \tag{2}$$

□

2.8. A feltételes várhatóérték és tulajdonságai. A feltételes eloszlás

2.9. Martingálok tulajdonságai, Doob-felbontása, kanonikus növekvő folyamata, Megállási idő

- $F_1 \subseteq F_2 \subseteq \dots$ σ-algebrák $\Rightarrow$ **filtráció**
- $X_1, X_2, \dots$ X_i F_i mérhető, **adaptált sorozat**

2.7. Definíció (Martingál)**.** A martingál (X_n, F_n) sorozat (páros), tulajdonságai:

- X_n adaptált az F_n filtrációra
- $|EX_1| < \infty$
- $E(X_n|F_{n-1}) = X_{n-1}$

2.8. Definíció (Szubmartingál)**.** Martingál tulajdonságai, kivéve, hogy $E(X_n|F_{n-1}) \geq$ $\geq X_{n-1}$

2.9. Definíció (Supermartingál)**.** Martingál tulajdonságai, kivéve, hogy $E(X_n|F_{n-1}) \leq$ $\leq X_{n-1}$

2.10. Definíció (Martingál differencia)**.** $d_n = X_n - X_{n-1},\ d_1 = X_1$

2.8. Lemma. Ha $d_1, d_2, \dots,\ E(d_i) = 0, \forall i \in \mathbb{N}$ függelten sorozat, akkor $\Rightarrow$ martingál differenciák (adott F-el)

Ha $d_1, d_2, \dots$ martingál differenciák, $Ed_n^2 < \infty \Rightarrow Ed_id_j \underset{i \neq j}{=} 0$

2.11. Definíció (Megállási idő)**.**

$$\nu : \Omega \to \mathbb{Z}^+ \cup \{\infty\} \text{ valószínűségi változó },\{\nu = m\} \in F_m$$

2.12. Definíció (Előrejelezhető folyamat)**.** $X_1, X_2, \dots, X_n$ F_{n-1} mérhető $(F_0 =$ $= \{\emptyset, \Omega\})$

Példa:
Tetszőleges B halmaz, $\nu_B = \min\{n : X_n \in B\}$ elérési idő

$$X'_n = X_{n \wedge \nu} = X_{\min(n,\nu)}$$

2.9. Lemma. (X'_n, F_n) (szub)martingál

$$X_n = d_n + d_{n-1} + \dots + d_2 + d_1$$
$$X^*_n = \max(x_1, \dots, x_n)$$

2.10. Lemma. (X_n, F_n) Martingál, $\sup_n E(X_n^2) < \infty$ (L_2 korlátos) $\Rightarrow X_n$ konvergál L_2-ben és 1 valószínűséggel.

2.11. Tétel (Doob-egyenlőtlenség)**.** – szubmartingál $x_n \geq 0 \Rightarrow E((X_n^*)^p) \leq$
$\leq (\frac{p}{p-1})^p \cdot E(X_n^p)$

– martingál $\Rightarrow ||X_n^*||_p \leq \frac{p}{p-1}||X_n||_p$

2.12. Tétel (Doob-felbontás)**.** (X_n, F_n) szubmartingál, előáll $X_n = M_n + A_n$, ahol M_n martingál, A_n előrejelezhető növekvő (m.m) nemnegatív.

$A_n - A_{n-1} = E(\underbrace{X_n - X_{n-1}}_{d_n}|F_{n-1})$

$A_n = \sum_{k=1}^n E(d_k|F_{k-1})$

$\underbrace{M_n}_{\text{Doob martingál}} = \sum_{k=1}^n (d_k - E(d_k|F_{k-1}))$

2.13. Lemma (Négyzetesen integrálható martingál kanonikus növekvő folyamata)**. Nem találtam**

2.14. Tétel (Krickeberg-felbontás)**.** Ha a martingál L_1 korlátos $\Rightarrow X_n = P_n -$
$- N_n$ két nemnegatív martingál különbségére felbontható.

2.10. Martingálok konvergencia-tételei. Nagy-eltérés tételek (Bernstein és társai)

2.15. Tétel. Ha egy martingál L_1 korlátos $\Rightarrow$ 1 valószínűséggel konvergál.

2.13. Definíció (Reguláris martingál)**.**

$$X_n = E(X|F_n)$$

Akkor a martingál reguláris.

2.16. Lemma. Ekvivalensek (sima martingálra)

– L_1-ben konvergens

– egyenletesen integrálható

– reguláris

2.17. Lemma. $p > 1$ esetében ekvivalensek:

– L_p-ben korlátos

– L_p-ben konvergens

– reguláris, $X \in L_p$

Becslések

2.18. Tétel (Bernstein)**.** (X_n, F_n) szubmartingál, $t \geq 0$ λ bármi, t tetszőleges, ekkor:

$$\underbrace{P(X_n^* \geq \lambda)}_{X_n\text{-ig elért maximum}} \leq e^{-t\lambda} \cdot E(e^{tX_1}) \quad (3)$$

2.19. Tétel (Azuma-Hoeffding)**.**
(X_n, F_n) martingál, korlátos növekmény: $|X_n - X_{n-1}| \leq S_n$ m.m, $X_0 = 0$. Ekkor:

$$\begin{aligned} &P(X_n \geq \lambda) \leq \exp(-\tfrac{\lambda^2}{S_1^2+\cdots+S_n^2}) \\ &P(|X_n| \geq \lambda) \leq 2\cdot\exp(-\tfrac{\lambda^2}{S_1^2+\cdots+S_n^2}) \\ &\text{Speciális eset, ha} S_n = 1 \\ &P(X_n \geq \lambda) \leq \exp(-\tfrac{\lambda^2}{n}) \end{aligned} \tag{4}$$

2.20. Tétel (Talagrand)**.** $X_1, \dots X_n$ függelten azonos eloszlásúak [0,1]-en $f : \mathbb{R}^{\ltimes} \to \mathbb{R}$ konvex, Lipsitz ($L = 1$), ekkor:

$$P(|f(X_1, \dots, X_n) - \underbrace{m_f}_{f \text{ mediánja}} \geq \lambda) \leq 4 \cdot \exp(\frac{-\lambda^2}{4}) \tag{5}$$

3. Statisztika

3.1. Statisztikai mező, alapfogalmak, Tapasztalati becslések

3.1. Definíció (Statisztikai mező)**.** Statisztikai mezőn egy $(\Omega, \mathcal{A}, \mathcal{P})$ hármast értünk, ahol

- Ω: a kisérlet lehetséges kimeneteleinek, az elemi eseményeinek a halmaza,
- $\mathcal{A}$: az Ω részhalmazaiból álló, a kísérlettel kapcsolatban megfigyelhető események σ-algebrája
- $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$ az $\mathcal{A}$-n értelmezett valószínűségeknek egy családja, melynek elemei közül kerül ki a véletlent elíró valószínűség.

Speciális eset: Paraméteres statisztikai mező

$$\mathcal{P} = \{\mathcal{P}_\theta : \theta \in \Theta\}$$

Például: $\Theta \subseteq \mathbb{R}^p$

3.2. Definíció (Mintatér)**.** Az $\mathfrak{X} = \mathbb{R}^n$ halmazt, amelybe a megfigyelhető mérési eredmények esnek **mintatérnek nevezzük**, elemei a mintaértékek.

3.3. Definíció (Minta)**.** $\mathbf{X} : (\Omega, \mathcal{A}, \mathcal{P}) \to (\mathfrak{X}, \mathfrak{B})$ mérhető függvény
$P \in \mathcal{P}$ mellett eloszlása, $Q = P \circ \mathbf{X}^{-1}$
Q-kat összegyűjtve $(\mathcal{X}, \mathfrak{B}, \mathcal{Q})$ statisztikai mező.
Megjegyzés: Sok esetben $\mathbf{X} = (X_1, \ldots, X_n)$ függelten azonos eloszlású.

3.4. Definíció (Statisztika)**.** $\mathbf{X} : \Omega \to \mathfrak{X}$ minta esetén az $\mathfrak{X}$ mintatéren értelmezett $\mathbf{T} : \mathfrak{X} \to \mathbb{R}^p$ függvényt *statisztikának* nevezzük.
Példák:

- $\overline{x} = (\sum_{i=1}^n X_i)/n$ (mintaáltag)
- $(X_1^*, X_2^*, \ldots, X_n^*)$ (rendezett minta)
- $X_n^* - X_1^*$ (range)
- $\frac{X_1^* + X_n^*}{2}$ (midrange)
- tapasztalati medián

3.5. Definíció (Tapasztalati eloszlás)**.** $(\mathfrak{X}, \mathfrak{B})$-n tetszőleges $B \in \mathfrak{B}$-re

$$Q_n^*(B) = \frac{1}{n} \sum_{i=1}^n \chi_B(X_i) = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}(B)$$

Állítások:

- $EQ_n^*(B) = Q(B)$ fix Q-nál
- $Q_n^*(B) \to Q(B)$ 1 valószínűséggel (NSZET)
- $D^2(Q_n^*(B)) = \frac{1}{n} \cdot Q(B) \cdot (1 - Q(B))$ 1 valószínűséggel (NSZET)
- $\sqrt{n} \cdot (Q_n^*(B) - Q(B)) \to \mathcal{N}(0, Q(B) \cdot (1 - Q(B))$ eloszlásban (CHT)

3.6. Definíció (Tapasztalati eloszlás függvény)**.** ($\mathfrak{X} = \mathbb{R}$)

$$F_n^*(x) = Q_n^*((-\infty, x)) = \frac{1}{n} \sum_{i=1}^{n} \chi(X_i < x) = \begin{cases} 0, & \text{ha } x \leq X_1^* \\ \frac{k}{n}, & \text{ha } X_k^* < x \leq X_1^* \\ 1, & \text{ha } X_n^* < x \end{cases}$$

Ha $\mathfrak{X} = \mathbb{R}^k$ akkor

$$F_n^*((x_1, \dots, x_k)) = Q_n^*((-\infty, x_1) \times (-\infty, x_2) \times \ldots \times (-\infty, x_k))$$

3.1. Tétel (Glivenko-Cantelli (matematikai statisztika alaptétele))**.** $X_1, \dots, X_n$ legyenek FAE valószínűségi változók. A belőlük, mint mintából képzett tapasztalati eloszlásfüggvények sorozatára igaz, hogy:

$$P(\lim_{n \to \infty} \sup_x |F_n(x) - F(x)| = 0) = 1 \tag{6}$$

3.2. Dominált statisztikai mező. Elégséges statisztikák

3.7. Definíció (Dominált statisztikai mező)**.** $\exists \mu$ mérték, σ-véges,

$$P << \mu, \ \forall P \in \mathcal{P}$$

$P \hookrightarrow f = \frac{dP}{d\mu}$ lesz egy paraméteres/nem paraméteres sűrűségfüggvény család

3.8. Definíció (Ekvivalens mértékek)**.** Két mérték P, Q, $P \sim Q$ (ekvivalens), ha $P << Q$ és $P >> Q$.

Azaz $\forall B$ mérhető halmazra $P(B) = 0 \Longleftrightarrow Q(B) = 0$
Mérték osztályra $\mathcal{P} \sim \mathcal{Q}$, ha $\forall B$ mérhető halmazra:

$$\forall P \in \mathcal{P}: \quad P(B) = 0 \Longleftrightarrow Q(B) = 0, \quad \forall Q \in \mathcal{Q}$$

3.2. Tétel (Halmos-Savage)**.** $\mathcal{P}$ mértékosztály, ekvivalensek az alábbiak:

- $\mathcal{P}$ dominált
- van $\mathcal{P}' \subseteq \mathcal{P}$, hogy $\mathcal{P}'$ megszámlálható és $\mathcal{P}' \sim \mathcal{P}$
- a domináló μ kikeverhető $\mathcal{P}$-ből: $\mu = \sum_{i=1}^{\infty} \alpha_i P_i$
  ahol: $\alpha_i \geq 0, \quad \sum_{i=1}^{\infty} \alpha_i = 1, \quad P_i \in \mathcal{P}$

3.9. Definíció (Elégséges statisztika)**.** T statisztika elégséges, ha:
$P_\vartheta(X \in B|T)$ függelten ϑ-tól

Példa:
$T = (X_1^*, \dots, X_n^*), \quad T : \mathbb{R}^n \to \mathbb{R}^n = X : X^* +$ sorrend
$P_\vartheta(X|T) =$ "$\Pi \cdot T$" (Permutációk (sorrend), ϑ mindegy) $\Rightarrow T$ elégséges

3.3. Tétel (Neyman-Fisher faktorizáció)**.** P_ϑ eloszlás, f_ϑ függvények, T statisztika
Ekvivalens:

- T elégséges
- $\exists \mu$ keverék $\mathcal{P}$-ből ($\mu = \sum \alpha_i P_i$) amire nézve $f_\vartheta(x) = g_\vartheta(T(X))$
- Tetszőleges λ domináló mértékre $f_\vartheta(x) = h(x) \cdot g_\vartheta(T(X))$
  ($h(x)$ ϑ-tól nem függ, csak x-től, ϑ-tól függő rész csak a statisztikán keresztül függ)

Példa: Poisson (λ) dominálva $\mathbb{Z}$-n számláló, $\lambda \in \mathbb{R}^+$, n elemű minta

$$f_\lambda((X_1, \dots, X_n)) = e^{-\lambda}\frac{\lambda^{X_1}}{X_1!} \cdot \ldots e^{-\lambda}\frac{\lambda^{X_n}}{X_n!} =$$

$$= e^{-\lambda n} \cdot \frac{\lambda^{\sum_{i=1}^n X_i}}{\prod_{i=1}^n X_i!} = \underbrace{\frac{1}{\prod_{i=1}^n X_i!}}_{h(X)} \cdot \underbrace{e^{-\lambda n} \cdot \lambda^{\sum_{i=1}^n X_i}}_{g_\lambda(\sum_i=1^n X_i} \Rightarrow$$

$$\underset{(N-F)}{\Longrightarrow} \sum_{i=1}^n X_i \quad \text{Elégséges} \tag{7}$$

3.4. Tétel (Hányados kritérium)**.** ($\mu \times \mu$ domináló mérték)

$$S \text{ elégséges} \iff \text{m.m } (x, y) \in \mathfrak{X}^2 \quad S(x) = S(y) \Rightarrow \frac{f_\vartheta(x)}{f_\vartheta(y)} \text{ független } \vartheta\text{-tól} \tag{8}$$

($\mu \times \mu$ domináló mérték)

$$S \text{ minimális elégséges} \iff \text{m.m } (x, y) \in \mathfrak{X}^2 \quad S(x) = S(y) \iff \frac{f_\vartheta(x)}{f_\vartheta(y)} \text{ független } \vartheta\text{-tól} \tag{9}$$

3.3. Fisher-információ és regularitási feltételek

3.3.1. Fisher-információ

Kérdés: $X \hookrightarrow \hat{\vartheta}$ becslés mennyire lesz pontos, mekkora hibára számíthatunk?
Dominált paraméteres mező, van $\{f_\vartheta : \vartheta \in \Theta\}$ sűrűségfüggvény család

3.10. Definíció (Likelihood, log-likelihood függvények)**.** Fix X-re

$$\begin{cases} \vartheta \to f_\vartheta(x) & \text{likelihood függvény } (L(x)) \\ \vartheta \to \log f_\vartheta(x) & \text{log-likelihood függvény } (l(x)) \end{cases} \tag{10}$$

Ezek nem sűrűségfüggvények, x nem fut, hanem ϑ!

3.11. Definíció (Fisher-információ)**.**

$$I(\vartheta) = E_\vartheta([\delta l(\vartheta)]^T \delta l(\vartheta)) = \int_{\mathfrak{X}} \delta l^T \delta l f_\vartheta d\mu = 4 \cdot \int_{\mathfrak{X}} [\delta \sqrt{f_\vartheta(x)}]^T \cdot [\delta \sqrt{f_\vartheta(x)}] d\mu(x) \tag{11}$$

A Fisher-információ.

Itt: $\delta : \frac{\delta}{\delta\vartheta}$ deriválás, $\delta l(p) : 1 \times p$-es, Tfh. $\Theta \subset \mathbb{R}^p$
$E_\vartheta : P_\vartheta$ szerint választjuk x-et

Tuladjonságai:

$$\mu \hookrightarrow \mu', \quad f_\vartheta \hookrightarrow f_\vartheta \cdot \frac{d\mu}{d\mu'}$$

domináló mértéktől nem függ: $\log f_\vartheta \hookrightarrow \log f_\vartheta + \log \frac{d\mu}{d\mu'} \Rightarrow \delta \log f_\vartheta \hookrightarrow \delta \log f_\vartheta$

$I(\vartheta)$ pozitív szemidefinit (PSD)

$$A \text{ PSD } \forall x \quad x^T A x \geq 0$$

$$B \geq C, \text{ ha } B - C \text{ PSD}$$

3.3.2. Regularitási feltételek

3.12. Definíció (Regularitási tételek)**.**
Gyenge (R)

- μ m.m $\vartheta \to \sqrt{f_\vartheta(x)}$ folytonosan deriválható
- $I(\vartheta)$ véges, ϑ-folytonos, nem szinguláris

Erős (RR) előzőken felül

- μ m.m $l(\vartheta)$ 2x folytonosan deriválható
- $||\delta^2_\vartheta l(\vartheta)|| < M \hookrightarrow \sup_\vartheta ||\delta^2 \vartheta l_x(\vartheta)|| \leq M(x), \quad \sup_\vartheta E_\vartheta M^2(x) < \infty$

3.5. Lemma (RR)**.** Erősen reguláris, ha:

$$\int_{\mathfrak{X}} \delta^2 f_\vartheta(x) d\mu(x) = 0$$

(és) vagy

$$I(\vartheta) = -E_\vartheta(\delta^2 l(\vartheta))$$

3.6. Lemma (R)**.** Gyengén reguláris T statisztika, $\mathfrak{X} \to \mathbb{R}^k$, lokális korlátos az $E_\vartheta ||T(x)||^2$, $\quad g(\vartheta) = E_\vartheta(T(x)) \Rightarrow g$ folytonosan differenciálható $\delta g(\vartheta) =$ $= E_\vartheta(T(x)\delta l_x(\vartheta))$

$$g(\vartheta) = \int T(x) f\vartheta(x) d\mu(x) \Rightarrow$$

$$\delta g(\vartheta) = \int T(x) \cdot \underbrace{f'\vartheta(x)}_{\delta l \cdot f_\vartheta} d\mu(x)$$

3.7. Tétel (Függetlenek információja összeadódik)**.** $x \in \mathfrak{X}, y \in \mathfrak{Y}$ függetlenek
f_ϑ, g_ϑ x és y sűrűségfüggvényei
(x,y) az $\mathfrak{X} \times \mathfrak{Y}$-on $\mu \times \nu$ dominálja $h_\vartheta = f_\vartheta \cdot g_\vartheta$ sűrűségfüggvény és

$$I_{(x,y)}(\vartheta) = I_{(x)}(\vartheta) + I_{(y)}(\vartheta)$$

(ahol x, y-ra **(R)**, ekkor (x, y)-ra is)

3.8. Lemma. (Statisztikából új statisztika készítésével nem nyerünk információt)
S statisztika $\mathfrak{X} \to \mathfrak{Y}$ (ugyanúgy $(Y, P, \mathfrak{Q}_y)$ statisztikai mező, (R)-t Y,S-re felteszszük)
Ekkor: $I_x(\vartheta) \geq I_s(\vartheta)$

3.9. Lemma. S elégséges, **(R)** X-re, ekkor:

$$I_x(\vartheta) = I_s(\vartheta) \tag{12}$$

(R) S-re is következik.

3.10. Lemma (S elégségesség)**.** S statisztika, minden P_ϑ ekvivalens

$$\text{Ha } \forall\vartheta \in \Theta, \quad I_x(\vartheta) = I_s(\vartheta) \Rightarrow \text{ S elégséges} \tag{13}$$

3.4. Többdimenziós normális eloszlás

3.11. Lemma. $X_1, \ldots, X_n$ független (skalár) standard normális eloszlások.
$\mathbf{X} = (X_1, \ldots, X_n)$ n-dimenziós standard normális eloszlású
Ha $A \subseteq \mathbb{R}^{n \times k}$, $a \in \mathbb{R}^k$-ra $Y = A\mathbf{X} + a$ k-dimenziós normális eloszlású.
Sűrűségfüggvénye:

$$f(\mathbf{x}) = (2\pi)^{-k/2} \det(\boldsymbol{\Sigma})^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right), \tag{14}$$

Az eloszlás gömbszimmetrikus

3.12. Tétel (Állítások több dimenziós norm. elo-ról.)**.**

– $Ey = n????$

- $\Sigma = AA^T$
- Több dimenziós normális eloszlás egy kisebb dimenziós része is normális eloszlású

3.13. Tétel. Állítás: Minden $n \in \mathbb{R}^n, \Sigma \in \mathbb{R}k \times k$ PSD-re
Van olyan normális Y, hogy $EY = m$, $\Sigma y = \Sigma$, $\mathcal{N}(m, \Sigma)$

3.14. Tétel. Állítás: Y normális, Y_1, Y_2 függetlenek $\longleftrightarrow$ $\text{cov}(Y_1, Y_2) = 0 = \Sigma_{1,2}$

3.15. Tétel.

$$Y \sim \mathcal{N}(m, \Sigma), AY + n, BY + b \text{ független} \iff A\Sigma B^T = 0$$

3.16. Tétel. Σ nem szinguláris, Y normális eloszlású, ekkor

$$Y_2, Y_1 - \Sigma_{1,2}\Sigma_{2,2}^{-1}Y_2 \text{ függetlenek}$$

($\sim$ orthogonalizáljuk a két valváltozót)

3.17. Tétel. Y normális, ekkor $P(Y_1|Y_2) \sim \mathcal{N}(m_{1|2}, \Sigma_{1|2})$
Ahol:

$$m_{1|2} = m_1 + \Sigma_{1,2}\Sigma_{2,2}^{-1}(Y_2 - m_2)$$
$$\Sigma_{1|2} = \Sigma_{1,1} - \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1}$$

3.5. Becslések elmélete. Blackwellizálás

3.5.1. Becslések elmélete

3.13. Definíció (Becslés).
$g : \Theta \to y$ a paraméteres mező paraméterfüggvénye (mérhető)
$T : \mathfrak{X} \to Y$ $g(\vartheta)$-t a $T(X)$-el becsüljük
Példa: $g(\vartheta) = \vartheta$

3.14. Definíció (Torzítatlan becslés). $x = \mathbb{R}^k, g : \Theta \to \mathbb{R}^k, T : \mathfrak{X} \to \mathbb{R}^k$
T torzítatlan becslése a g-nek, ha $E_\vartheta(T(X)) = g(\vartheta), \forall \vartheta \in \Theta$

Ellenkező esetben a **torzítás** $b_T(\vartheta) = E_\vartheta(T(X)) - g(\vartheta)$

3.15. Definíció (Veszteségfüggvény).

$$\begin{array}{l} W : \mathbb{R}^k \times \Theta \to \mathbb{R} \\ W : \mathbb{R}^k \times \Theta \to \mathbb{R}^{k \times k} \end{array} \tag{15}$$

($W(T(X), \vartheta)$ mennyire torz)

Tulajdonságai:

- $W \geq 0$ (skalár/PSD)

- Konvex első változójában ($\rightarrow$ ha 2x akkorát hibázom, legalább 2x akkora legyen a veszteség is)
- $0 = W(g(\vartheta), \vartheta)$

Speciális esetei:

- $W(T(X), \vartheta) = \tilde{W}(T(X) - g(\vartheta))$ (1 változós esetben)
- $W_1(\nu) = \nu\nu^T \in \mathbb{R}^{k \times k}$
- $W_2(\nu) = \nu^T\nu = ||v||_2^2 = Tr(W_1(\nu)) \in \mathbb{R}$

3.16. Definíció (Rizikófüggvény)**.** T rizikófüggvénye:

$$R_T(\vartheta) = E_\vartheta(W(T(X)), \vartheta) \tag{16}$$

(A veszteségfüggvény várható értéke)

Becslések összehasonlítása

3.17. Definíció (Ekvivalens becslések)**.** Ha T_1, T_2 -re $R_{T_1} = R_{T_2}$ akkor $T_1 \sim T_2$ (ekvivalens becslések)

3.18. Definíció (Jobb becslés)**.** Ha T_1, T_2 becslések és $\forall \vartheta \in \Theta$-ra $R_{T_1} \geq R_{T_2}$, akkor T_1 jobb, mint T_2

3.19. Definíció (Megengedhető, optimális, hatásos és minimax becslés)**.**
$\mathcal{D}$ becslések egy családja

- $T \in \mathcal{D}$ megengedhető, ha nincs nála jobb nem ekvivalens
- $T \in \mathcal{D}$ optimális, ha legjobb, mindegyikkel összehasonlítható, és jobb
- $T \in \mathcal{D}$ minimax, ha $\sup_\vartheta R_T(\vartheta)$-t minimalizálja $\mathcal{D}$-n
- $\mathcal{D}$ torzítatlan W_1-re, akkor hatásos

3.18. Tétel. Optimális $T \Rightarrow$ megengedhető és minimax

3.19. Tétel.
W_1-re $R_T(\vartheta) = E_\vartheta([T(X) - g(\vartheta)][T(X) - g(\vartheta)]^T) = \Sigma_T(\vartheta) + b_T(\vartheta) \cdot b_T(\vartheta)^T$

3.20. Tétel. – $\mathcal{D}$ konvex (halmaz)

- W szigorúan konvex (1 változóban) $\Rightarrow$
  $\exists$ optimális becslés, P m.m. egyértelmű

3.5.2. Blackwellizálás

3.20. Definíció (Konzisztens és erősen konzisztens)**.** $T_n : \mathfrak{X}^n \to \mathbb{R}^k$ n elemű becsléssorozat

- T_n **konzisztens**, ha $T_n((X_1, \dots, X_n)) \to g(\vartheta)$ sztochasztikusan $(\forall \vartheta \in \Theta)$
- T_n **erősen konzisztens**, ha $T_n((X_1, \dots, X_n)) \to g(\vartheta)$ 1 valószínűséggel $(\forall \vartheta \in \Theta)$

3.21. Tétel (Blackwell-Rao)**.** T statisztika, $g(\vartheta)$ becslése, S elégséges.
Legyen $\tilde{T} = E_\vartheta(T|S) = E(T|S)$ (elégségesség miatt)
Ekkor: $b_T(\vartheta) = b_{\tilde{T}}(\vartheta), \quad R_{\tilde{T}}(\vartheta) \leq R_T(\vartheta)$

Példa: Poisson (λ), $T(X) = X_1$, $g(\lambda) = \lambda$
$S(X) = \sum X_i \hookrightarrow \tilde{T} = \overline{X}$

3.21. Definíció (Teljes statisztika)**.** T statisztika teljes, ha $\forall \varphi$ függvényre, amire $\forall \vartheta$-ra $E_\vartheta(\varphi(T)) = 0$, azokra $\varphi(T) = 0$ $\mathcal{P}$-m.m.

3.22. Tétel. Ha T teljes, akkor egyetlen torzítatlan becslése lehet T függvényeként g-nek

3.23. Tétel. Teljes, elégséges $\Rightarrow$ minimális elégséges

Kombinálva:
T tetszőleges torzítatlan, S teljes elégséges

- $R_{\tilde{T}} \leq R_T$
- Minden T-re megy
- $\tilde{T}$, T-től független

Ha fennállnak, akkor $\tilde{T}$ **optimális** becslés a torzítatlanok között.

3.6. Információs határ. Momentum módszer és maximum likelihood módszer

3.6.1. Információs határ

3.24. Tétel (Cramer-Rao)**.** Legyen $(X_1, X_2, \dots, X_n)$ egy $g_\vartheta(x)$, $\vartheta \in \Theta$ sűrűségfüggvényű eloszlásból származó minta. Tegyük fel, hogy a minta FAE és véges a Fisher-Információja.

Legyen $T(X)$ statisztika torzítatlan becslése a ϑ paramérternek, és legyen $G := \delta g \in \mathbb{R}^{k \times p}, \quad \Theta \in \mathbb{R}^p$

Ekkor
$$\Sigma_T(\vartheta) \geq G(\vartheta) \cdot \underbrace{I(\vartheta)^{-1}}_{\text{Fisher-inf.}} \cdot G(\vartheta)^T \tag{17}$$

Speciális $g(\vartheta) = \vartheta$ esetben
$$\Sigma_T(\vartheta) \geq I(\vartheta)^{-1} \tag{18}$$

3.25. Tétel (Cramer-Rao, Torzított)**.** Feltételek, mint Cramer-Rao, de T nem torzítatlan. Mivel g folytonosan differenciálható W_1-re nézve, ezért $b_T(\vartheta)$ is differenciálható, jelölje $B := \delta b_T$

Ekkor

$$R_T(\vartheta) \geq (G(\vartheta) + B(\vartheta)) \cdot I(\vartheta)^{-1} \cdot (G(\vartheta) + B(\vartheta))^T + b_T(\vartheta) \cdot b_T(\vartheta)^T \tag{19}$$

3.7. Hipotézisvizsgálat alapjai. A likelihood-hányados próba egyszerű hipotéziseknél

3.8. A normális eloszlás paramétereire vonatkozó próbák áttekintése. Lineáris modell paraméterbecslése

3.8.1. Hipotézisvizsgálathoz szükséges eloszlások

3.22. Definíció (χ^2 eloszlás)**.** $X \sim \mathcal{N}(0, I_n)$ n-dimenziós standard normális eloszlás, ekkor $Y = ||X||_2^2$ eloszlása n-szabadságfokú χ^2 eloszlás.

(n darab standard normális eloszlás négyzeteinek összege)

3.23. Definíció (t-eloszlás)**.** $Y \sim \chi^2, \quad X \sim \mathcal{N}(0{,}1)$, függetlenek. Ekkor $Z = \frac{X}{\sqrt{Y/n}}$ eloszlása t_n, azaz n szabadságfokú t-eloszlás.

3.26. Tétel (Cauchy eloszlás)**.** $t_1 = \frac{X_1}{X_2}$, azaz az 1 szabadságfokú t-eloszlást Cauchy eloszlásnak is hívjuk, sűrűségfüggvénye: $f(x) = \frac{1}{\pi(1+x^2)}$

3.27. Tétel (t-eloszlás és std. normális eloszlás kapcsolata)**.** $t_n, n \to \infty$ $\mathcal{N}(0{,}1)$-hez tart.

3.24. Definíció (F-eloszlás)**.** $Y_n \sim \chi_n^2,\ Y_m \sim \chi_m^2,\ Z = \frac{(1/n)Y_n}{(1/m)Y_m}$ eloszlást $F_{n,m}$-el jelöljük, neve F-eloszlás.

3.28. Tétel (F-eloszlással kapcsolatos összefüggések)**.**

$$F_{n,m} = 1/F_{m,n}$$

$$F_{1,m} = t_m^2$$

$$\text{fix } n \text{ mellett, } F_{n,m} \underset{m\to\infty}{\longrightarrow} \chi_n^2$$

3.29. Tétel (Fisher-Bartlett)**.** $(X_1, \ldots, X_n) \sim \mathcal{N}(\mu, \sigma^2)$ független, ebből becslünk $(\overline{x}, S_n^*)$-ot.

Ekkor:

- Ez elégséges statisztika (párban)
- Függelten a kettő, $\overline{x} \sim \mathcal{N}(\mu, \sigma^2/n)$, $S_n^{*2} \sim \frac{\sigma^2}{n-1}\chi_{n-1}^2$

3.8.2. Hipotéziscsaládok

Várható értékre

- Ismert szórás?
  - Igen → u (vagy z)
  - Nem → t
- Egyezés, vagy egyenlőtlenség
  - Egyezés → kétirányú
  - Egyenlőtlenség → egyirányú
- Mihez hasonlítunk:
  - Referenciához → egymintás
  - Másik mintához → kétmintás

Példa:
u, kétmintás, egyoldali
$x_1, \ldots, x_n \sim \mathcal{N}(\mu_1, \sigma_1^2)$
$y_1, \ldots, y_m \sim \mathcal{N}(\mu_2, \sigma_2^2)$
$u = \frac{\overline{x}-\overline{y}}{\sqrt{\sigma_1^2/n+\sigma_2^2/m}} \sim \mathcal{N}(0{,}1),$ ha $\mu_1 = \mu_2$

Szórásra (F variánsai) Példa
F próba, 2 minta, 2 oldali
$x_1, \ldots, x_n \sim \mathcal{N}(\mu_1, \sigma_1^2)$
$y_1, \ldots, y_m \sim \mathcal{N}(\mu_2, \sigma_2^2)$
$H_0: \quad \sigma_1 = \sigma_2$
$H_1: \quad \sigma_1 \neq \sigma_2$

$$F = \frac{S_n^{*2}}{S_m^{*2}} \sim F_{n-1,m-1} \text{ ha } \sigma_1 = \sigma_2$$

$$\mathfrak{X}_k = \{F < F_{n-1,m-1}(1-\alpha/2) \text{ vagy } F > F_{n-1,m-1}(\alpha/2)\}$$

$$\frac{1}{F} > F_{m-1,n-1}(\alpha/2)$$

3.9. χ^2 próbák illeszkedésvizsgálata és következményeik. Folytonos illeszkedésvizsgálat

"Valami kategorizálás/hányan vannak teszt"

3.30. Tétel. $A_1, \ldots, A_m$ teljes eseményrendszer, $P(A_i) = P_i$, mintában A_i n_i-szer teljesült. $\sum_{i=1}^{m} n_i = n$ Ekkor:

$$\chi^2 = \sum_{i=1}^{m} \frac{(np_i - n_i)^2}{np_i} \underset{n\to\infty}{\longrightarrow} \chi^2_{m-1} \text{ eloszlásban} \tag{20}$$