Valszám vizsga

1.Feladat:

Nevezetes abszolút folytonos eloszlások:

Eloszlás neve	Jelölése	Eloszlásfüggvény	Sűrűségfüggvény	EX	D^2X
Egyenletes	$E(a,b)$	$\begin{cases} 0 & \text{ha } x \le a \\ \frac{x-a}{b-a} & \text{ha } a < x \le b \\ 1 & \text{ha } b < x \end{cases}$	$\begin{cases} \frac{1}{b-a} & \text{ha } a < x \le b \\ 0 & \text{különben} \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponenciális	$\text{Exp}(\lambda)$	$\begin{cases} 1-e^{-\lambda x} & \text{ha } x \ge 0 \\ 0 & \text{különben} \end{cases}$	$\begin{cases} \lambda e^{-\lambda x} & \text{ha } x \ge 0 \\ 0 & \text{különben} \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Standard normális	$N(0,1^2)$	$\Phi(x) = \ldots$	$\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \quad x \in \mathbb{R}$	0	1
Normális	$N(m,\sigma^2)$	...	$\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}} \quad x \in \mathbb{R}$	m	σ^2

Állítás. **Val.változó függvényének várható értéke**

Legyen X val. változó; g: $\mathbb{R} \to \mathbb{R}$ függvény.
Ekkor

- $E(g(X)) = \sum_k g(x_k)p_k$, ha X diszkrét

Állítás. Legyenek $X, Y, X_1, ..., X_n$ valószínűségi változók; $c, c_i, a, b \in \mathbb{R}$. Ekkor

- $E(X+Y) = EX + EY$;
- $E(cX) = cEX$;
- $E\sum_{i=1}^{n} c_i X_i = \sum_{i=1}^{n} c_i EX_i$;
- $D^2(aX+b) = a^2 D^2 X$.

pl: X -2Y => D^2(X-2Y) =D^2(X) + D^2(2Y) = D^2(X) + 4*D^2(Y)

Definíció. **X val.változó abszolút folytonos**, ha létezik olyan f(x) függvény, amelyre $F(x) = \int_{-\infty}^{x} f(t)\,dt$. Ilyenkor f(x)-et **sűrűségfüggvény**nek hívjuk.

Állítás. Legyen X abszolút folytonos eloszlású. Ekkor

- f(x)=F'(x);
- f(x) ≥ 0;
- $\int_{-\infty}^{\infty} f(x)\,dx = 1$;
- $P(X = x) = 0 \quad \forall x$-re;
- $P(a < X \le b) = P(a \le X < b) = F(b) - F(a)$.

Definíció. X **és** Y **kovarianciája:** $\text{Cov}(X,Y) = E\left[(X-EX)(Y-EY)\right]$.
Köv.: $\text{Cov}(X,Y) = E(XY) - EXEY$.

Definíció. X **és** Y **korrelációja:** $R(X,Y) = \frac{\text{Cov}(X,Y)}{DXDY}$.

Definíció. **Val.változók konvolúciója:** Legyenek X és Y független val.változók. X és Y konvolúciójának (jel. X∗Y) az X+Y val.változót nevezzük.

Állítás. **A konvolúció eloszlásának meghatározása**

- Diszkrét eset: $P(X+Y=k)=\sum\limits_{l=0}^{\infty}P(X=l)\cdot P(Y=k-l)$
- Folytonos eset: $f_{X+Y}(z)=\int\limits_{-\infty}^{\infty}f_X(u)f_Y(z-u)du$

Állítás. Legyenek $X_1,...,X_n$, X és Y független val. változók

- $X_1\sim\text{Ind}(p),...,X_n\sim\text{Ind}(p) \quad\Rightarrow\quad X_1+...+X_n\sim\text{Bin}(n,p)$
- $X\sim\text{Bin}(n,p),\ Y\sim\text{Bin}(m,p) \quad\Rightarrow\quad X+Y\sim\text{Bin}(n+m,p)$
- $X_1\sim\text{Geo}(p),...,X_n\sim\text{Geo}(p) \quad\Rightarrow\quad X_1+...+X_n\sim\text{NegBin}(n,p)$
- $X\sim\text{Poi}(\lambda_1),\ Y\sim\text{Poi}(\lambda_2) \quad\Rightarrow\quad X+Y\sim\text{Poi}(\lambda_1+\lambda_2)$
- $X\sim N(m_1,\sigma_1^2),\ Y\sim N(m_2,\sigma_2^2) \quad\Rightarrow\quad X+Y\sim N(m_1+m_2,\sigma_1^2+\sigma_2^2)$

2.Feladat:

Eloszlás neve	Jelölése	Eloszlása	EX	D^2X
Karakterisztikus (indikátorvált.)	$\text{Ind}(p)$	$P(X=1)=p$ $P(X=0)=1-p$	p	$p(1-p)$
Geometriai (Pascal)	$\text{Geo}(p)$	$P(X=k)=p(1-p)^{k-1}$ k=1,2,...	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Hipergeometriai	$\text{Hipgeo}(N,M,n)$	$P(X=k)=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$ k=0,1,...,n	$n\frac{M}{N}$	$n\frac{M}{N}\left(1-\frac{M}{N}\right)\left(1-\frac{n-1}{N-1}\right)$
Binomiális	$\text{Bin}(n,p)$	$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$ k=0,1,...,n	np	$np(1-p)$
Negatív binomiális	$\text{NegBin}(n,p)$	$P(X=k)=\binom{k-1}{n-1}p^n(1-p)^{k-n}$ k=n,n+1,...	$\frac{n}{p}$	$\frac{n(1-p)}{p^2}$
Poisson	$\text{Poi}(\lambda)$	$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$ k=0,1,...	λ	λ

Előfordulásuk:

- Indikátor változó: egy p valószínűségű esemény bekövetkezik-e vagy sem
- Geometriai: hányadikra következik be először egy p valószínűségű esemény
- Hipergeometriai: visszatevés nélküli mintavétel
- Binomiális: visszatevéses mintavétel
- Negatív binomiális: hányadikra következik be n. alkalommal egy p valószínűségű esemény

λ -> várhatóan hány cucc történik a vizsgált intervallumon

3.Feladat:

13.4. Tétel. Nagy számok gyenge törvénye:
Legyenek X_i-k független, azonos eloszlású valószínűségi változók, továbbá legyen $E(X_i)=m$; $D(X_i)=\sigma$ (tehát léteznek).
Ekkor $S_n=X_1+\cdots+X_n \Rightarrow \frac{S_n}{n}\longrightarrow m$ sztochasztikusan $(n\to\infty)$

- Megadott eloszlás szerint kiszámoljuk a várható értéket.
- Ha a sorozat meg van szorozva egy konstanssal, akkor megszorozzuk vele a várható értéket

Tétel. **Centrális határeloszlás tétel (CHT):**
Legyenek $X_1, X_2, \ldots$ i.i.d. val. változók, $\mathrm{E}X_1 = m$, $D^2(X_1) = \sigma^2 < \infty$.
Ekkor
$\frac{X_1+\ldots+X_n-nm}{\sqrt{n}\sigma} \overset{n\to\infty}{\to}$ N(0,1) gyengén, azaz $P\left(\frac{X_1+\ldots+X_n-nm}{\sqrt{n}\sigma} < x\right) \overset{n\to\infty}{\to} \Phi(x)$.

4-5.Feladat:
- exponenciális eloszlás: Abszolút folytonos örökifjú tulajdonságú eloszlás
- nevezetes eloszlások eloszlásfüggvénye: lásd első feladat táblázat
-Abszolút folytonos eloszlás egy diszkrét értéket 0 valószínűséggel vesz fel

Definíció. **X szórásnégyzete** : $\mathrm{D}^2\mathrm{X}=\mathrm{E}[(\mathrm{X}-\mathrm{EX})]^2 = \mathrm{EX}^2\text{-}\mathrm{E}^2\mathrm{X}$.

Definíció. **X szórása** : $\mathrm{DX}=\sqrt{D^2X}$.

Állítás. **Val.változó függvényének várható értéke**

Legyen X val. változó; g: $\mathbb{R} \to \mathbb{R}$ függvény.
Ekkor
- $E(g(X)) = \sum\limits_k g(x_k)p_k$, ha X diszkrét
- $E(g(X)) = \int\limits_{-\infty}^{\infty} g(x)f(x)\,dx$, ha X abszolút folytonos

Mindkét esetben a várható érték létezéséhez a szumma/integrál abszolút konvergenciájára van szükség.

6.Feladat:
4.2. Tétel. (**Teljes valószínűség tétele**) Legyen $A \in \mathcal{A}$ tetszőleges esemény, és $B_1, B_2, \ldots$ TER. Ekkor

$$P(A) = \sum_{i=1}^{\infty} P(A|B_i) \cdot P(B_i).$$

4.3. Tétel. (**Bayes tétele**) Legyen $A \in \mathcal{A}$ esemény, és $B_1, B_2, \ldots$ teljes eseményrendszer. Ekkor

$$P(B_k|A) = \frac{P(A|B_k) \cdot P(B_k)}{\sum\limits_{i=1}^{\infty} P(A|B_i) \cdot P(B_i)}.$$

4.3. Példa. Tegyük fel, hogy egy hallgató a feltett kérdésre $\frac{3}{4}$ valószínűséggel tudja a választ. Ha nem tudja, akkor tippel, és $\frac{1}{3}$ valószínűséggel találja el a helyes választ. a) Mennyi az esélye, hogy a hallgató helyesen válaszol? b) Ha a hallgató helyesen válaszolt, mennyi a valószínűsége, hogy tudta is a választ?
Legyen A: a hallgató helyesen válaszol, B_1: tudja a választ, B_2: nem tudja a választ. Ekkor B_1, B_2 TER, $P(B_1) = \frac{3}{4}$, $P(B_2) = \frac{1}{4}$, $P(A|B_1) = 1$, $P(A|B_2) = \frac{1}{3}$. Ebből
a) $P(A) = P(A|B_1) \cdot P(B_1) + P(A|B_2) \cdot P(B_2) = 1 \cdot \frac{3}{4} + \frac{1}{3} \cdot \frac{1}{4} = \frac{10}{2} = \underset{\sim}{\frac{5}{6}}$.

b) $P(B_1|A) = \dfrac{P(A|B_1) \cdot P(B_1)}{P(A|B_1) \cdot P(B_1) + P(A|B_2) \cdot P(B_2)} = \dfrac{1 \cdot \frac{3}{4}}{1 \cdot \frac{3}{4} + \frac{1}{3} \cdot \frac{1}{4}} = \underset{\sim}{\dfrac{9}{10}}$. ■

7.Feladat:
- Valamelyik nevezetes eloszlás (hányszor kell húznunk – Geometriai, hány darab pirosat húzunk – Binomiális)
lásd: fenti táblázat

Tönkremenési feladat:

Péter és Gábor úgy játszanak, hogy mindketten $\frac{1}{2}$ valószínűséggel nyernek egy játszmában a másiktól 1 forintot. A játék addig megy, amíg valaki a másik összes pénzét el nem nyeri. Péternél a játék kezdeténél 10 forint, Gábornál pedig 16 forint van. Mekkora valószínűséggel megy tönkre Péter?
Legyen $k = 10$ és $n = 26$. Ekkor $n - k = 16$.

$A := \{$Péter k forintról tönkremegy $\}$
$p(k) := P(A)$.
Ekkor $p(0) = 1$ és $p(n) = 0$.
B_1 : az első lépésben Péter nyer,
B_2 : az első lépésben Gábor nyer.
Ekkor
$P(A) = P(A|B_1) \cdot P(B_1) + P(A|B_2) \cdot P(B_2) = p(k+1) \cdot \frac{1}{2} + p(k-1) \cdot \frac{1}{2}$,
ahol $1 \leq k \leq n - 1$.
$\Rightarrow 2 \cdot p(k) = p(k+1) + p(k-1)$
$\Rightarrow p(k+1) - p(k) = p(k) - p(k-1) = d$

$p(k) = p(k) - p(k-1) + p(k-1) - p(k-2) + \ldots + p(1) - p(0) + p(0) = 1 + k \cdot d$.
$0 = p(n) = 1 + nd \Rightarrow d = -\frac{1}{n}$.
$p(k) = 1 - \frac{k}{n}, 1 - p(k) = \frac{k}{n}$
$k = 10, n = 26 \Rightarrow P(A) = 1 - \frac{10}{26} = \frac{8}{13}$.

- P(A) = 1-k/n

8.Feladat:

Gyakran előforduló statisztikai mennyiségek:

Legyenek $x_1, x_2, \ldots, x_n \in \mathbb{R}$ az $X_1, X_2, \ldots, X_n$ független, azonos eloszlású valószínűségi változók *realizáció*inak értékei. (Azaz egy kísérlet n-szeri, egymástól függetlenül, azonos körülmények közti elvégzésének eredményei.)

1. *Mintaközép* (más szóval *tapasztalati várható érték*):

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

2. k-adik *tapasztalati momentum* ($k \geq 1$ egész):

$$\widehat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k$$

3. *Tapasztalati szórásnégyzet*:

$$s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \widehat{\mu}_2 - \widehat{\mu}_1^2$$

4. *Korrigált tapasztalati szórásnégyzet*:

$$s_n^{*\,2} = \frac{n}{n-1} s_n^2$$

5. *Tapasztalati eloszlásfüggvény*:

$$F_n^*(t) = \frac{1}{n}\sum_{i=1}^{n} \chi_{(x_i < t)} \qquad (t \in \mathbb{R})$$

- *Tapasztalati eloszlásfüggvény*: $F_n(x) = \frac{\sum_{i=1}^{n} I(X_i < x)}{n}$
  ahol $I(X_i < x) = \begin{cases} 1 & \text{ha } X_i < x \\ 0 & \text{ha } X_i \geq x \end{cases} \rightsquigarrow$ karakterisztikus függvény
- Terjedelem: Legnagyobb elem – Legkisebb elem
- Nagyság: Minta elemszáma

Maximum Likelihood becslés Nevezetes eloszlásokra:

Eloszlás	Becslés
Binomiális	$\overline{X}$ / n
Geometria	$1/\overline{X}$
Poisson	$\overline{X}$
Normál	$\overline{X}$
Exponenciális	$1/\overline{X}$

9.Feladat:

Definíció. **Torzítatlan becslés:**
$\mathrm{T}(\mathbf{X})$ statisztika torzítatlan becslése θ-nak, ha $E_\theta T(\mathbf{X}) = \theta \quad \forall\theta$-ra.

Torzítatlan becslésekre: T_1 hatásosabb becslése a θ paraméternek a T_2-nél, ha

$$D_\theta^{\ 2}\left(T_1(\underline{X})\right) \leq D_\theta^{\ 2}\left(T_2(\underline{X})\right)$$

teljesül minden θ paraméterértékre.

10.Feladat:

Hipotézis $\sim$ valami állítás, aminek igazságát vizsgálni szeretnénk

Paramétertér: $\Theta = \Theta_0 \cup^* \Theta_1 \longrightarrow$ "valóság"
Mintatér: $\mathcal{X} = \mathcal{X}_e \cup^* \mathcal{X}_k \longrightarrow$ "látszat" - MINTÁBÓL
$\mathcal{X}_k$: kritikus tartomány - azon $\underline{X}$ megfigyelések halmaza, amikre *elutasítjuk* a nullhipotézist
$\mathcal{X}_e$: elfogadási tartomány - azon $\underline{X}$ megfigyelések halmaza, amikre *elfogadjuk* a nullhipotézist

Hipotézisvizsgálati feladat:
$H_0: \ \theta \in \Theta_0 \quad \rightsquigarrow$ nullhipotézis
$\underline{H_1: \ \theta \in \Theta_1} \quad \rightsquigarrow$ ellenhipotézis

Tehát ha $\underline{X} \in \mathcal{X}_e$, akkor elfogadjuk H_0-t; ha $\underline{X} \in \mathcal{X}_k$, akkor pedig elutasítjuk H_0-t.
Amennyiben a Θ_0 halmaz egyelemű, akkor azt mondjuk, hogy H_0 egyszerű. H_1-re ugyanígy.

Az $\mathcal{X}$ mintatér felosztását általában egy statisztika (neve: próbastatisztika) segítségével végezzük el:
legyen T: $\mathcal{X} \to \mathbb{R}$, $\quad \mathcal{X}_k = \{\underline{x} \in \mathcal{X} : \mathrm{T}(\underline{x}) > c\}$ $\qquad c$ neve: kritikus érték
$\mathcal{X}_e = \{\underline{x} \in \mathcal{X} : \mathrm{T}(\underline{x}) \leq c\}$

döntés / "valóság"	H_0-t elfogadjuk ($\mathcal{X}_e$)	H_0-t elutasítjuk ($\mathcal{X}_k$)
H_0 teljesül (Θ_0)	helyes döntés	elsőfajú hiba
H_0 nem teljesül (Θ_1)	másodfajú hiba	helyes döntés

P(elsőfajú hiba)=$\alpha(\theta)$=$\mathrm{P}_\theta(\mathcal{X}_k)$, ahol $\theta \in \Theta_0$
P(másodfajú hiba)=$\beta(\theta)$=$\mathrm{P}_\theta(\mathcal{X}_e)$, ahol $\theta \in \Theta_1$
Erőfüggvény: ψ: $\Theta_1 \to \mathbb{R}$, $\psi(\theta) = P_\theta(\mathcal{X}_k)$
Terjedelem: $\alpha = \sup\,\{\alpha(\theta)\text{: } \theta \in \Theta_0\}$

Azt mondjuk, hogy az 1-es próba *erősebb* a 2-es próbánál, ha $\alpha_1 = \alpha_2$ és $\psi_1(\theta) \geq \psi_2(\theta)\ \forall\theta \in \Theta_1$.
Próbafüggvény: φ: $\mathcal{X} \to [0,1] \rightsquigarrow$ ennyi valószínűséggel vetem el a H_0-t a minta alapján
$\underline{x} \in \mathcal{X}_k \Rightarrow \varphi(\underline{x}) = 1$
$\underline{x} \in \mathcal{X}_e \Rightarrow \varphi(\underline{x}) = 0$
p-érték: az az α terjedelem, ami esetén a próbastatisztika értéke egyenlő a kritikus értékkel : $T(\underline{x}) = c_\alpha$.
A p-érték a legkisebb terjedelem, amire még elutasítjuk a H_0-t. Ha egy próbát számítógép segítségével végzünk el, rendszerint a p-érték révén tudunk dönteni: ha (p-érték)$<\alpha$, akkor elvetjük H_0-t.

11.Feladat:

Állítás. **Normálás**
Legyen X$\sim N(m, \sigma^2)$. Ekkor $\frac{X-m}{\sigma} \sim N(0,1)$.

Állítás. $\Phi(-x) = 1 - \Phi(x)$

Állítás. $\Phi^{-1}(q) = -\Phi^{-1}(1-q) \quad 0 < q < 1$

Definíció. **Konfidencia intervallum:** Adott α-hoz legalább $(1-\alpha)$ valószínűséggel tartalmazza az adott paramétert (vagy annak egy függvényét): $P_\theta\left(T_1(\mathbf{X}) < \hat{\theta} < T_2(\mathbf{X})\right) \geq 1-\alpha$.
Gyakran keresünk szimmetrikus konfidencia intervallumot, ilyenkor $T_1 = T_2 =: \Delta$, és az intervallum $\hat{\theta} \pm \Delta$ alakba írható.

Legyen $X_1, ..., X_n \sim N(m,\sigma)$ i.i.d. minta

- m-re konfidencia intervallum
  - ha σ ismert, akkor $\overline{x} \pm u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$
  - ha σ ismeretlen, akkor $\overline{x} \pm t_{n-1,\frac{\alpha}{2}} \frac{s_n^*}{\sqrt{n}}$

- σ^2-re konfidencia intervallum: $\left[\frac{(n-1)\cdot(s_n^*)^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}; \frac{(n-1)\cdot(s_n^*)^2}{\chi^2_{n-1,\frac{\alpha}{2}}}\right]$

Konfidencia intervallum a valószínűségre (p) nagy minta esetén, ha normális eloszlással közelítünk: $\hat{p} \pm u_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

u-próba

$X_1, \ldots, X_n \sim$N(m,σ^2), ahol σ ismert, m paraméter

a.) $H_0: m = m_0$, $\underline{H_1: m \neq m_0}$ b.) $H_0: m = m_0$, $\underline{H_1: m > m_0}$ c.) $H_0: m = m_0$, $\underline{H_1: m < m_0}$

A próbastatisztika mindhárom esetben ugyanaz: T($\underline{\text{X}}$)$=u = \sqrt{n}\frac{\overline{X}-m_0}{\sigma} \overset{H_0 \text{ esetén}}{\sim}$N(0,1)

A kritikus tartományok:

a.) $\mathcal{X}_k = \{\underline{x} : |u| > u_{\alpha/2}\}$
b.) $\mathcal{X}_k = \{\underline{x} : u > u_\alpha\}$
c.) $\mathcal{X}_k = \{\underline{x} : u < -u_\alpha\}$

- Feladatban megadott 10%, 2% stb. -> α=0,10 $u_\alpha = \varphi^{-1}(1-\alpha)$ (vagy $\alpha/2$)

t-próba

$X_1, \ldots, X_n \sim$N(m,σ^2), ahol σ, m paraméter

a.) $H_0: m = m_0$, $\underline{H_1: m \neq m_0}$ b.) $H_0: m = m_0$, $\underline{H_1: m > m_0}$ c.) $H_0: m = m_0$, $\underline{H_1: m < m_0}$

A próbastatisztika mindhárom esetben ugyanaz: T($\underline{\text{X}}$)$=t = \sqrt{n}\frac{\overline{X}-m_0}{s_n^*} \overset{H_0 \text{ esetén}}{\sim} t_{n-1}$

A kritikus tartományok:

a.) $\mathcal{X}_k = \{\underline{x} : |t| > t_{n-1,\alpha/2}\}$
b.) $\mathcal{X}_k = \{\underline{x} : t > t_{n-1,\alpha}\}$
c.) $\mathcal{X}_k = \{\underline{x} : t < -t_{n-1,\alpha}\}$

- t- próbás táblázatból kikeresés

\- Ha a kritikus tartományba esik az érték, akkor elutasítjuk a nullhipotézist

2.) **Kétmintás próbák**

$X_1, \ldots, X_n \sim \mathrm{N}(m_1, \sigma_1^2)$
$Y_1, \ldots, Y_m \sim \mathrm{N}(m_2, \sigma_2^2)$
Az elvégzendő próbák $H_0 : m_1 = m_2$ nullhipotézis esetén:

<table>
<tr><td></td><td colspan="2">a két minta
független</td><td>a két minta
nem független</td></tr>
<tr><td>σ_1 és σ_2 ismert</td><td colspan="2">b.) kétmintás u-próba</td><td>egymintás u-próba
a különbségekre</td></tr>
<tr><td rowspan="3">σ_1 és σ_2 ismeretlen</td><td colspan="2">előzetes F-próba</td><td rowspan="3">egymintás t-próba
a különbségekre</td></tr>
<tr><td>$\sigma_1 = \sigma_2$</td><td>$\sigma_1 \neq \sigma_2$</td></tr>
<tr><td>c.) kétmintás t-próba</td><td>d.) Welch-próba</td></tr>
</table>

3.) χ^2-**próbák**

a.) Diszkrét illeszkedésvizsgálat

Feladat: adott egy $\mathbf{X} = (X_1, \ldots, X_n)$ n elemű minta, és azt akarjuk eldönteni, hogy a minta egy általunk "remélt" eloszlásból származik-e. *Diszkrét* illeszkedésvizsgálatnál feltesszük, hogy a mintaelemek r különböző értéket vehetnek fel: $\mathrm{P}(X_i = x_j) = p_j \quad j = 1, \ldots, r$. Jelöljük N_j-vel a gyakoriságokat, azaz azt, hogy az n elemű mintában hány darab x_j szerepel.

Osztályok	1	2	...	r	Összesen
Valószínűségek	p_1	p_2	...	p_r	1
Gyakoriságok	N_1	N_2	...	N_r	n

H_0 : a valószínűségek: $\mathbf{p}=(p_1, \ldots, p_r)$
H_1: nem ezek a valószínűségek

A próbastatisztika: $T_n = \sum\limits_{i=1}^{r} \frac{(N_i - np_i)^2}{np_i} \xrightarrow{H_0 \text{ esetén}} \chi^2_{r-1}$ eloszlásban, ha $n \to \infty$

A kritikus tartomány: $\mathcal{X}_k = \{\mathbf{x} : T_n(\mathbf{x}) > \chi^2_{r-1,1-\alpha}\}$

b.) Függetlenségvizsgálat

Feladat: van egy minta, két szempont szerint csoportosítva. Azt kell eldönteni, hogy a két szempont független-e egymástól.
$\mathrm{p}_{i,j}$=P(egy megfigyelés az (i,j) osztályba kerül)
$\mathrm{N}_{i,j}$=ennyi megfigyelés kerül az (i,j) osztályba

A mintavétel eredménye:

<table>
<tr><td></td><td></td><td colspan="5">2. szempont</td><td></td></tr>
<tr><td></td><td></td><td>1</td><td>...</td><td>j</td><td>...</td><td>s</td><td>Összesen</td></tr>
<tr><td rowspan="5">1. szempont</td><td>1</td><td>N_{11}</td><td>...</td><td>N_{1j}</td><td>...</td><td>N_{1s}</td><td>$N_{1\bullet}$</td></tr>
<tr><td>$\vdots$</td><td>$\vdots$</td><td></td><td>$\vdots$</td><td></td><td>$\vdots$</td><td>$\vdots$</td></tr>
<tr><td>i</td><td>N_{i1}</td><td>...</td><td>N_{ij}</td><td>...</td><td>N_{is}</td><td>$N_{i\bullet}$</td></tr>
<tr><td>$\vdots$</td><td>$\vdots$</td><td></td><td>$\vdots$</td><td></td><td>$\vdots$</td><td>$\vdots$</td></tr>
<tr><td>r</td><td>N_{r1}</td><td>...</td><td>N_{rj}</td><td>...</td><td>N_{rs}</td><td>$N_{r\bullet}$</td></tr>
<tr><td colspan="2">Összesen</td><td>$N_{\bullet 1}$</td><td>...</td><td>$N_{\bullet j}$</td><td>...</td><td>$N_{\bullet s}$</td><td>n</td></tr>
</table>

$$N_{i\bullet} = \sum_{j=1}^{s} N_{i,j}$$
$$N_{\bullet j} = \sum_{i=1}^{r} N_{i,j}$$

H_0 : a szempontok függetlenek, azaz $p_{i,j} = p_{i\bullet} \cdot p_{\bullet j}$ $\forall i, j$-re
H_1: nem azok

A próbastatisztika: $T_n = n \left(\sum\limits_{i=1}^{r} \sum\limits_{j=1}^{s} \frac{N_{i,j}^2}{N_{i\bullet} N_{\bullet j}} - 1 \right) \xrightarrow{H_0 \text{ esetén}} \chi^2_{(r-1)(s-1)}$ eloszlásban, ha $n \to \infty$

A kritikus tartomány: $\mathcal{X}_k = \{\mathbf{x} : T_n(\underline{x}) > \chi^2_{(r-1)(s-1),1-\alpha}\}$

Bizonyítások, Definíciók:

1. Szimetrikus bolyongás

a. visszatéri idejének értéke végtelen:

Megoldás:
Legyen $n>0$ egész szám és jelölje ξ a lépések számát n-ből 0-ba! Jelöljük továbbá A-val azt az eseményt, hogy az első lépés jobbra történik, B-vel pedig azt, hogy az első lépés balra történik. A és B teljes eseményrendszert alkot. **(1 pont)**
Ekkor mivel A és B is ½ valószínűségű, ezért a teljes várható érték tétel **(1 pont)** szerint

$$v(n) = E\xi = E(\xi|A)P(A) + E(\xi|B)P(B) = (1+v(n+1))\frac{1}{2} + (1+v(n-1))\frac{1}{2}.$$ **(1 pont)**

Ebből

$$v(n+1) - v(n) = v(n) - v(n-1) - 2 = \ldots = v(1) - v(0) - 2n.$$ **(1 pont)**

Mivel ξ nem-negatív, ezért várható értéke is az, így **(1 pont)**

$$0 \le v(n) = v(n) - v(n-1) + v(n-1) - \ldots + v(1) - v(0) + v(0) =$$
$$= nv(1) - 2\frac{n(n-1)}{2} = n\left(v(1) - (n-1)\right).$$ **(1 pont)**

Így

$$v(1) \ge n-1 \Rightarrow v(1) = +\infty$$ **(1 pont)**

Jelöljük X-el a 0-ból 0-ba történő visszatérés idejét. C-vel azt az eseményt, hogy az első lépés jobbra történik, D-vel pedig azt, hogy az első lépés balra történik. C és D teljes eseményrendszert alkot.
Mivel $v(1) = +\infty$, ezért

$$EX = E(X|C)P(C) + E(X|D)P(D) = (1+v(1))\frac{1}{2} + (1+v(1))\frac{1}{2} = \infty$$ **(1 pont)**

b. egy valószínűséggel visszatér nullába:

Legyen D_1, hogy az első lépésben jobbra, ill. D_2, hogy az első lépésben balra megyünk.
$P(C) = P(C|D_1)\cdot\frac{1}{2} + P(C|D_2)\cdot\frac{1}{2}$
$P(C|D_1)$: annak valószínűsége, hogy 1-ből eljutunk 0-ba
$P(C|D_1) \ge P(\text{1-ből előbb jutunk el 0-ba, mint } n\text{-be}) = 1 - \frac{1}{n}$ minden n-re $\Rightarrow$
$P(C|D_1) \ge 1 - \frac{1}{n}$ minden n-re $\Rightarrow$
$P(C|D_1) = 1$, ugyanígy $P(C|D_2) = 1$, azaz $P(C) = 1$.

2. Statisztika alaptétele:

Tétel: $\xi_1,\ldots,\xi_n$ független, azonos F eloszlásfüggvényűek (1 pont). Ekkor
$\sup_z |F_n(z) - F(z)| \xrightarrow[n\to\infty]{} 0$ (1 pont) majdnem mindenütt (1 vszgel) (1 pont).
Biz.: Csak folytonos F eloszlásfüggvényekre látjuk be (1 pont). Ebből következik, hogy tetszőleges pozitív egész N-hez léteznek olyan valós $z_1,\ldots,z_N$ számok, hogy

$$F(z_0) = 0, F(z_1) = \frac{1}{N},\ldots, F(z_i) = \frac{i}{N},\ldots, F(z_{N-1}) = \frac{N-1}{N}, F(z_N) = 1,$$

$z_0 = -\infty, z_N = \infty$. (1 pont)
Ekkor, ha $z \in [z_k, z_{k+1})$, akkor

$$F_n(z) - F(z) \le F_n(z_{k+1}) - F(z_k) = F_n(z_{k+1}) - F(z_{k+1}) + \frac{1}{N},$$

$$F_n(z) - F(z) \ge F_n(z_k) - F(z_{k+1}) = F_n(z_k) - F(z_k) - \frac{1}{N}.$$

Ebből következik, hogy

$$\sup_z |F_n(z) - F(z)| \le \max_{0\le k\le N} |F_n(z_k) - F(z_k)| + \frac{1}{N}.$$ (1 pont)

Tudjuk, hogy rögzített x – re

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \chi\{\xi_i < x\},$$ (1 pont)

ahol $\chi\{\xi_i < x\}$ független, azonos eloszlású indikátor valószínűségi változók, melyek várható értéke

$E\chi\{\xi_i < x\} = P(\xi_i < x) = F(x)$.(1 pont)

Így a nagy számok erős törvénye szerint

$F_n(x) = \frac{1}{n}\sum_{i=1}^{n}\chi\{\xi_i < x\} \xrightarrow[n\to\infty]{} E\chi\{\xi_i < x\} = F(x)$ mm. (1 pont)

Legyen $A_{k,N} = \left\{\omega : \frac{1}{n}\sum_{i=1}^{n}\chi\{\xi_i(\omega) < z_k\} \xrightarrow[n\to\infty]{} F(z_k)\right\}$, ekkor

$P(A_{k,N}) = 1$ és $B_N = \left\{\omega : \max_{0\le k\le N}\left|F_n(z_k) - F(z_k)\right| \xrightarrow[n\to\infty]{} 0\right\} = \bigcap_{k=1}^{N-1} A_{k,N}$.

$B_N -$en $\limsup_{n\to\infty}\left|F_n(z) - F(z)\right| \le \frac{1}{N}$. Ebből következik, hogy $\bigcap_{N=1}^{\infty} B_N$-en $\limsup_{n\to\infty}\left|F_n(z) - F(z)\right| = 0$. 1 valószínűségű események metszete is 1 valószínűségű, így $\bigcap_{N=1}^{\infty} B_N = \bigcap_{N=1}^{\infty}\bigcap_{k=1}^{N-1} A_{k,N}$ is 1 valószínűségű.(1 pont)

3. Bayes –tétel:

4.3. Tétel. (**Bayes tétele**) Legyen $A \in \mathcal{A}$ esemény, és $B_1, B_2, \ldots$ teljes eseményrendszer. Ekkor

$$P(B_k \big| A) = \frac{P(A \big| B_k)\cdot P(B_k)}{\sum_{i=1}^{\infty} P(A \big| B_i)\cdot P(B_i)}.$$

Bizonyítás. A jobboldal számlálója $\frac{P(A\cap B_k)}{P(B_k)}\cdot P(B_k) = P(A\cap B_k)$, a jobboldal nevezője pedig éppen $P(A)$ a teljes valószínűség tétele szerint. ■

4. Teljes valószínűség tétele:

4.2. Tétel. (**Teljes valószínűség tétele**) Legyen $A \in \mathcal{A}$ tetszőleges esemény, és $B_1, B_2, \ldots$ TER. Ekkor

$$P(A) = \sum_{i=1}^{\infty} P(A|B_i)\cdot P(B_i).$$

Bizonyítás. $P(A) = P(A\cap\Omega) = P(A\cap(\cup B_i)) = P(\bigcup(B_i\cap A)) \overset{\text{diszj}}{=} \sum_{i=1}^{\infty} P(B_i\cap A) =$

$$\sum_{i=1}^{\infty}\frac{P(B_i\cap A)}{P(B_i)}\cdot P(B_i) = \sum_{i=1}^{\infty} P(A|B_i)\cdot P(B_i). \quad ■$$

5. Markov és Csebisev egyenlőtlenségek:

13.1. Tétel. Markov egyenlőtlenség:

Legyen $X \geqslant 0$ valószínűségi változó ; $E(X)$ létezik. Ekkor $P(X \geqslant K) \leqslant \frac{E(X)}{K}$.

Biz.: $\tilde{X} = \begin{cases} K & \text{ha } X \geqslant K \\ 0 & \text{egyébként} \end{cases}$

$\tilde{X} \leqslant X$; $E(\tilde{X}) \leqslant E(X)$

$E(X) \geqslant E(\tilde{X}) = K\cdot P(\tilde{X} = K) + 0\cdot P(\tilde{X} = 0) = K\cdot P(X \geqslant K)$

13.2. Tétel. Csebisev-egyenlőtlenség:

Legyen X tetsz. valószínűségi változó ; $E(X), D(X)$ létezik. Ekkor $P(|X - E(X)| \geqslant K) \leqslant \frac{D^2(X)}{K^2}$.

Biz.: $P(\ (X-E(X))^2 \geqslant K^2\ ) \overset{\text{Markov}}{\leqslant} \dfrac{E[\ (X-E(X))^2\ ]}{K^2} = \dfrac{D^2(X)}{K^2}$

Példa: pénzérme; jelöljük n dobásból a fejek számát: S_n

$P(S_n \geqslant 0,6n)$-t szeretnénk becsülni.

a) Markov: $P(S_n \geqslant 0,6n) \leqslant \dfrac{E(S_n)}{0,6n} = \dfrac{0,5n}{0,6n} = 0,83.$

b) Csebisev: $P(S_n \geqslant 0,6n) = P(S_n - 0,5n \geqslant 0,1n) \leqslant \dfrac{1}{2} P(|S_n - 0,5n| \geqslant 0,1n) \leqslant \dfrac{1}{2} \cdot \dfrac{n \cdot \frac{1}{2} \cdot \frac{1}{2}}{(0,1n)^2} = \dfrac{100}{8n}$

itt $\dfrac{1}{2}$-es szorzó: szimmetria, 1000 dobásból 520-at vagy 480-at u.a. valószínűséggel dobhatok

c) Még jobb: $P(S_n \geqslant 0,6n) = P(\ (1,5)^{S_n} \geqslant (1,5)^{0,6n}\ ) \leqslant \dfrac{E(1,5^{S_n})}{1,5^{0,6n}} = \dfrac{\left(\frac{5}{4}\right)^n}{(1,5^{0,6})^n} = 0,98^n$

ahol $E(1,5^{S_n}) = E(1,5^{X_1+\cdots+X_n}) = E(1,5^{X_1} \cdots 1,5^{X_n}) \overset{\text{fgtln}}{=} E(1,5^{X_1}) \cdots E(1,5^{X_n}) = \left(\dfrac{5}{4}\right)^n$

$X_i = \begin{cases} 1 & \text{ha az i. fej} \\ 0 & \text{egyébként} \end{cases}$

$E(1,5^{X_i}) = 1,5^0 \cdot \dfrac{1}{2} + 1,5^1 \cdot \dfrac{1}{2} = \dfrac{5}{4}$

$n = 1000$: $P(S_{1000} \geqslant 600)$

Markov: $\leqslant 0,83$

Cseb.: $\leqslant \dfrac{25}{2n} = \dfrac{25}{2000} = 0,0125$

1.5-ös: $1,68 \cdot 10^{-9}$

6. Nagy számok erős törvény:

13.8. Tétel. Nagy számok erős törvénye. Legyen $X_1, X_2, \ldots$ független, azonos eloszlású valószínűségi változók, melyekre $E(X) = m$ létezik. Ekkor $S_n/n \to m$ 1 valószínűséggel.

7. Paraméterbecslések:

- **Maximum likelihood módszer (ML-módszer):** Azt a paraméterértéket keressük, ahol a likelihood függvény a legnagyobb értéket veszi fel: $\max\limits_{\theta} L(\theta, \mathbf{x})$

- **Momentum módszer:** A mintából számítható tapasztalati momentumokat ($m_i := \frac{\sum_j x_j^i}{n}$) egyenlővé tesszük az elméleti momentumokkal ($M_i := E_\theta X^i$), az elsőtől kezdve, mégpedig annyit, amennyi paraméter van. Tehát p darab ismeretlen paraméter esetén a következő p ismeretlenes egyenletrendszert oldjuk meg:

$$M_1 = m_1$$
$$\vdots$$
$$M_p = m_p$$

Megjegyzés: $m_1 = \overline{x}$