**Uppsala Universitet**
Matematiska Institutionen
Marcus Vaktnäs

# Tentamen i Matematik

Sannolikhetsteori 1, 1MS034
2022-10-28

*Skrivtid: 5 timmar. Tillåtna hjälpmedel: miniräknare och formelsamlingen. Varje uppgift ger som mest 5 poäng. Maxpoäng är 40, plus 2 möjliga bonuspoäng från inlämningsuppgifter. 18-24 ger betyg 3, 25-31 ger betyg 4, och 32-42 ger betyg 5. Lycka till!*

**1.** Låt $A$, $B$ och $C$ vara händelser i ett sannolikhetsrum $(\Omega, \mathcal{F}, P)$ med $P(B \cap C) > 0$.

a) Visa att $P(B) > 0$.

*Lösning.* $B \supseteq B \cap C$, så $P(B) \geq P(B \cap C) > 0$. □

b) Visa att $P(A^c \mid B) = 1 - P(A \mid B)$.

*Lösning.* Vi har $B = (B \cap A) \cup (B \cap A^c)$, och $B \cap A$ och $B \cap A^c$ är disjunkta mängder, så vi får $P(B) = P(B \cap A) + P(B \cap A^c)$, och därmed har vi

$$P(A^c \mid B) = \frac{P(A^c \cap B)}{P(B)} = \frac{P(B) - P(A \cap B)}{P(B)} = 1 - \frac{P(A \cap B)}{P(B)} = 1 - P(A \mid B).$$

□

c) För sannolikhetsmåttet $Q(A) = P(A \mid B)$, visa att $Q(A \mid C) = P(A \mid B \cap C)$.

*Lösning.*

$$Q(A \mid C) = \frac{Q(A \cap C)}{Q(C)} = \frac{P(A \cap C \mid B)}{P(C \mid B)} = \frac{\frac{P(A \cap B \cap C)}{P(B)}}{\frac{P(B \cap C)}{P(B)}} = \frac{P(A \cap B \cap C)}{P(B \cap C)} = P(A \mid B \cap C).$$

□

**2.** Den tvådimensionella slumpvariabeln $(X, Y)$ har simultan täthetsfunktion $f_{X,Y}(x, y) = cx$ för $x, y \in [0, 1]$ och $f_{X,Y}(x, y) = 0$ annars.

b) Bestäm de marginella täthetsfunktionerna till $X$ och $Y$.

*Lösning.* Om $x \notin [0, 1]$ så är $f_{X,Y}(x, y) = 0$ för alla $y$, så $f_X(x) = 0$ i detta fall. För $x \in [0, 1]$ får vi

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy = \int_0^1 cx \, \mathrm{d}y = cx.$$

På samma sätt får vi $f_Y(y) = 0$ om $y \notin [0, 1]$ och annars

$$f_Y(y) = \int_0^1 cx \, \mathrm{d}x = c/2.$$

□

a) Bestäm $c$.

*Lösning.* Vi får $c = 2$, eftersom

$$1 = \int_{-\infty}^{\infty} f_X(x) \, \mathrm{d}x = \int_0^1 cx \, \mathrm{d}x = c/2.$$

□

c) Bestäm kovariansen $\mathrm{Cov}(X, Y)$.

*Lösning.* Eftersom $f_{X,Y}(x, y) = f_X(x) f_Y(y)$ så är $X$ och $Y$ oberoende, så $\mathrm{Cov}(X, Y) = 0$. □

**3.** En apa spelar en rad på stryktipset varje vecka. En tipsrad består av 13 fotbollsmatcher där apan på varje match fyller i 1, $X$ eller 2.

a) Hur länge kommer det ta i genomsnitt (medelvärde) för apan att få minst 12 rätt på en rad?

*Lösning.* Antalet rätt apan får under en given vecka är en slumpvariabel $X \sim \text{Bin}(13, 1/3)$, så sannolikheten att apan får minst 12 rätt på en given vecka är

$$P(X \geq 12) = P(X = 12) + P(X = 13) = 13\left(\frac{1}{3}\right)^{12} \cdot \frac{2}{3} + \left(\frac{1}{3}\right)^{13} = \frac{1}{3^{10}}.$$

Antalet veckor apan måste spela för att få minst 12 rätt på en rad är alltså $Y \sim \text{ffg}(1/3^{10})$, så vi får $EY = 3^{10} = 59049$ veckor $= 59049 \cdot 7 = 413343$ dagar $\approx 413343/365$ år $\approx 1100$ år. □

b) Hur länge måste apan leva för att den ska ha minst 90% chans att få minst 12 rätt under sin livstid?

*Lösning.* Om $Y \sim \text{ffg}(p)$ så är $P(X > k) = (1-p)^k$ för $k = 0, 1, 2, \ldots$, och i vårat fall är $p = 1/3^{10}$, så vi söker alltså det minsta heltalet $k$ så att

$$P(X > k) = (1 - \frac{1}{3^{10}})^k \leq 0.1.$$

Vi får alltså

$$k = \left\lceil \frac{\log(0.1)}{\log(1 - 1/3^{10})} \right\rceil = 135964 \text{ veckor} \approx 2600 \text{ år}.$$

□

**4.** Låt $X \sim \text{N}(0, 1)$ och $Y \sim \text{N}(0, 1)$ vara oberoende slumpvariabler.

a) Använd faltningsformlerna för att hitta fördelningen till $X + Y$.

*Lösning.*

$$\begin{aligned}
f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,\mathrm{d}x &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\sqrt{2\pi}} e^{-(z-x)^2/2}\,\mathrm{d}x \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-x^2+xz-z^2/2}\,\mathrm{d}x \\
&= \frac{e^{-z^2/4}}{2\pi} \int_{-\infty}^{\infty} e^{-x^2+xz-z^2/4}\,\mathrm{d}x \\
&= \frac{e^{-z^2/4}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(x-z/2)^2}\,\mathrm{d}x \\
&= \frac{e^{-z^2/(2\cdot 2)}}{\sqrt{2\pi \cdot 2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi \cdot 1/2}} e^{-(x-z/2)^2/(2\cdot 1/2)}\,\mathrm{d}x.
\end{aligned}$$

Eftersom integranden i sista raden är täthetsfunktionen till $\text{N}(z/2, 1/2)$ så får vi

$$f_{X+Y}(z) = \frac{e^{-z^2/(2\cdot 2)}}{\sqrt{2\pi \cdot 2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi \cdot 1/2}} e^{-(x-z/2)^2/(2\cdot 1/2)}\,\mathrm{d}x = \frac{e^{-z^2/(2\cdot 2)}}{\sqrt{2\pi \cdot 2}}.$$

Detta är täthetsfunktionen till $\text{N}(0, 2)$, så vi får $X + Y \sim \text{N}(0, 2)$. □

b) Använd genererande funktioner för att hitta fördelningen till $X + Y$.

*Lösning.* Den momentgenererande funktionen till $\text{N}(\mu, \sigma^2)$ är $e^{\mu t + \sigma^2 t^2/2}$, så vi får

$$\psi_{X+Y}(t) = \psi_X(t)\psi_Y(t) = e^{t^2/2} e^{t^2/2} = e^{t^2} = e^{0t + 2t^2/2}.$$

Vi ser alltså att mgf till $X + Y$ är samma som mgf till $\text{N}(0, 2)$, så $X + Y \sim \text{N}(0, 2)$. □

**5.** Kasta ett mynt 50 gånger och skriv $p$ för sannolikheten att myntet landade på krona minst 30 ggr.

a) Använd Chebyshevs olikhet för att visa att $p \leq 1/4$.

*Lösning.* Antalet ggr myntet landade på krona är en slumpvariabel $X \sim \text{Bin}(50, 1/2)$ och vi vill uppskatta $P(X \geq 30)$. Genom symmetrin för binomialfördelning har vi att $P(X \geq 30) = P(X \leq 20)$, så vi får

$$P(X \geq 30) = \frac{1}{2}P(\{X \geq 30\} \cup \{X \leq 20\}) = \frac{1}{2}P(|X - 25| \geq 5).$$

Vi har $EX = 25$ och $\text{Var}(X) = 25/2$, så Chebyshevs olikhet ger

$$P(X \geq 30) = \frac{1}{2}P(|X - 25| \geq 5) \leq \frac{1}{2} \cdot \frac{\text{Var}(X)}{5^2} = \frac{25}{2 \cdot 2 \cdot 25} = \frac{1}{4}.$$

□

b) Använd en noggrant utvald tabell för att approximera $p$.

*Lösning.* Eftersom $50 \cdot 1/2 \cdot 1/2 > 5$ så använder vi normalapproximation, med halvkorrektion då variabeln är diskret. Vi får då

$$\begin{aligned} P(X \geq 30) = P(X \geq 30.5) = P\left(\frac{X - 25}{5/\sqrt{2}} \geq \frac{30.5 - 25}{5/\sqrt{2}}\right) = P\left(\frac{X - 25}{5/\sqrt{2}} \geq 1.1\sqrt{2}\right) &\approx 1 - \Phi(1.56) \\ &\approx 1 - 0.9406 \\ &= 0.0594 \approx 0.06. \end{aligned}$$

□

**6.** Kasta en symmetrisk tärning med $n$ sidor till någon sida dykt upp två gånger. Låt $X_n$ vara antalet kast som behövs. Bestäm fördelningsfunktionen till $X_n$ och visa att gränsfördelningen till $X_n/\sqrt{n}$ är $\sqrt{X}$ där $X \sim \text{Exp}(1/2)$.

**Ledning:** För att hitta gränsfördelningen, beräkna $\lim_{n\to\infty} \log P(X_n/\sqrt{n} > t)$ genom att använda Taylorutvecklingen $\log(1 - x) = -x + \mathcal{O}(x^2)$ när $x \to 0$, och använd er av att

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2} \quad , \quad \sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}.$$

*Lösning.* Notera att $P(X_n > 1) = 1$, eftersom vi behöver minst 2 kast. Notera också att vi har $P(X_n > k + 1 \mid X_n > k) = (n - k)/n$ för $k = 1, \dots, n - 1$, eftersom om de första $k$ kasten ger olika sidor så har vi $n - k$ möjliga sidor kvar som gör att kast $k + 1$ ger en ny sida. För $k = 2, \dots, n - 1$ får vi då

$$\begin{aligned} P(X_n > k) &= P(X_n > k, X_n > k - 1) \\ &= P(X_n > k \mid X_n > k - 1)P(X_n > k - 1) \\ &= P(X_n > k \mid X_n > k - 1)P(X_n > k - 1 \mid X_n > k - 2)P(X_n > k - 2) \\ &= \dots \\ &= P(X_n > k \mid X_n > k - 1) \dots P(X_n > 2 \mid X_n > 1)P(X_n > 1) \\ &= \frac{n - k - 1}{n} \dots \frac{n - 1}{n} \cdot 1 \\ &= \prod_{j=1}^{k-1} \frac{n - j}{n}. \end{aligned}$$

Fördelningsfunktionen blir alltså

$$F_{X_n}(t) = P(X_n \leq t) = 1 - P(X_n > t) = 1 - \prod_{j=1}^{k-1} \frac{n - j}{n} = 1 - \prod_{j=1}^{k-1} \left(1 - \frac{j}{n}\right)$$

för $k = 2, \dots, n - 1$. Mer generellt har vi $P(X_n \leq t) = 0$ för $t < 2$, $P(X_n \leq t) = 1$ om $t \geq n$, och om $2 \leq t < n$ får vi

$$F_{X_n}(t) = 1 - \prod_{j=1}^{\lfloor t \rfloor - 1} \left(1 - \frac{j}{n}\right).$$

Vi vill nu hitta $\lim_{n\to\infty} P(X_n/\sqrt{n} \le t)$. Om $t \le 0$ är $P(X_n/\sqrt{n} \le t) = 0$ för alla $n$, så gränsvärdet blir 0 i detta fall. Ta nu ett $t > 0$. Om $n > \max\{t^2, 2/t^2\}$ så är $2 < t\sqrt{n} < n$, så att

$$P(X_n/\sqrt{n} > t) = P(X_n > t\sqrt{n}) = \prod_{j=1}^{\lfloor t\sqrt{n}\rfloor - 1} \left(1 - \frac{j}{n}\right),$$

och vi har

$$\log \prod_{j=1}^{\lfloor t\sqrt{n}\rfloor - 1} \left(1 - \frac{j}{n}\right) = \sum_{j=1}^{\lfloor t\sqrt{n}\rfloor - 1} \log\left(1 - \frac{j}{n}\right) = -\sum_{j=1}^{\lfloor t\sqrt{n}\rfloor - 1} \frac{j}{n} + \sum_{j=1}^{\lfloor t\sqrt{n}\rfloor - 1} \left(\log\left(1 - \frac{j}{n}\right) + \frac{j}{n}\right).$$

För den första summan har vi

$$\sum_{j=1}^{\lfloor t\sqrt{n}\rfloor - 1} \frac{j}{n} = \frac{(\lfloor t\sqrt{n}\rfloor - 1)\lfloor t\sqrt{n}\rfloor}{2n} = s_n.$$

Eftersom $t\sqrt{n} - 1 < \lfloor t\sqrt{n}\rfloor \le t\sqrt{n}$ har vi

$$\frac{(t\sqrt{n} - 2)(t\sqrt{n} - 1)}{2n} < s_n \le \frac{(t\sqrt{n} - 1)t\sqrt{n}}{2n},$$

så vi ser att $s_n \to t^2/2$. För den andra termen använder vi att $|\log(1 - x) + x| \le Cx^2$ för $x \in (-\delta, \delta)$ för något tillräckligt litet $\delta > 0$ och något tillräckligt stort $C \ge 0$. Eftersom $\lfloor t\sqrt{n}\rfloor / n \to 0$ så är $0 < j/n < \delta$ för alla $j = 1, \dots, \lfloor t\sqrt{n}\rfloor - 1$ om $n$ är stort nog, så att vi har, för stora nog $n$,

$$\begin{aligned}\left| \sum_{j=1}^{\lfloor t\sqrt{n}\rfloor - 1} \left(\log\left(1 - \frac{j}{n}\right) + \frac{j}{n}\right) \right| \le \sum_{j=1}^{\lfloor t\sqrt{n}\rfloor - 1} C\frac{j^2}{n^2} &= \frac{C(\lfloor t\sqrt{n}\rfloor - 1)\lfloor t\sqrt{n}\rfloor(2\lfloor t\sqrt{n}\rfloor - 1)}{6n^2} \\ &\le \frac{C \cdot t\sqrt{n} \cdot t\sqrt{n} \cdot 2t\sqrt{n}}{6n^2} = \frac{Ct^3}{3\sqrt{n}} \xrightarrow{n\to\infty} 0.\end{aligned}$$

Vi får alltså $\lim_{n\to\infty} \log P(X_n/\sqrt{n} > t) = -t^2/2$, så $P(X_n/\sqrt{n} > t) \to e^{-t^2/2}$ och då har vi

$$P(X_n/\sqrt{n} \le t) = 1 - P(X_n/\sqrt{n} > t) \xrightarrow{n\to\infty} 1 - e^{-t^2/2}.$$

Om $X \sim \text{Exp}(1/2)$ så är $P(X \le t) = 1 - e^{-t/2}$ för $t > 0$ och $P(X \le t) = 0$ om $t \le 0$, så $P(X^2 \le t) = 0$ om $t \le 0$ och om $t > 0$ så är $P(\sqrt{X} \le t) = P(X \le t^2) = 1 - e^{-t^2/2}$. Vi drar alltså slutsatsen att $X_n/\sqrt{n}$ konvergerar i fördelning mot $\sqrt{X}$. $\square$

**7.** Vi skriver $X \sim Y$ om $X$ och $Y$ har samma fördelning. Avgör om följande påståenden om diskreta slumpvariabler är sanna eller falska. Om ett givet påstående är sant, förklara varför, och om falskt, ge ett motexempel.

a) Om $P(X > 0) = 1$ så är $X > 0$.

*Lösning.* FALSKT. Ta till exempel $\Omega = \{a, b\}$ med $P(\{a\}) = 1, P(\{b\}) = 0$, och sätt $X(a) = 1, X(b) = 0$. Då är $X(b) \le 0$, så vi har inte $X > 0$, men $P(X > 0) = P(\{a\}) = 1$. $\square$

b) Om $X > 0$ så är $EX > 0$.

*Lösning.* SANT. X är diskret så är alla observationer med positiv sannolikhet finns i mängden $A = \{x_1, x_2, \dots\}$. $X > 0$ så vi kan välja $x_n > 0$ för alla $n$, och $P(X = x_m) > 0$ för något $m$, så vi får

$$EX = \sum_{n=1}^{\infty} x_n P(X = x_n) \ge x_m P(X = x_m) > 0.$$

$\square$

c) Om $X_1 \sim X_2$ och $Y_1 \sim Y_2$ så är $X_1 + Y_1 \sim X_2 + Y_2$.

*Lösning.* FALSKT. Ta till exempel $X_1 = X_2 = Y_1 \sim \mathrm{Be}(1/2)$ och låt $Y_2 \sim \mathrm{Be}(1/2)$ så att $X_2$ och $Y_2$ är oberoende. Då är $X_1 + Y_1 = 2X_1$ så att fördelningen till $X_1 + Y_1$ är $p_{X_1+Y_1}(0) = p_{X_1+Y_1}(2) = 1/2$, medan $X_2 + Y_2 \sim \mathrm{Bin}(2, 1/2)$. Så till exempel har vi $0 = P(X_1 + Y_1 = 1) \neq P(X_2 + Y_2 = 1) = 1/2$.

Ett exempel på sådana slumpvariabler skulle kunna vara att låta $\Omega = \{a, b\}^2$ med likformig sannolikhet, och $X_1(\omega) = X_2(\omega) = Y_1(\omega) = 1$ om $\omega_1 = a$, $X_1(\omega) = X_2(\omega) = Y_1(\omega) = 0$ om $\omega_1 = b$, $Y_2(\omega) = 1$ om $\omega_2 = a$, och $Y_2(\omega) = 0$ om $\omega_2 = b$. Då är $X_1 = X_2 = Y_1 \sim \mathrm{Be}(1/2)$ och $Y_2 \sim \mathrm{Be}(1/2)$ oberoende. □

d) Om $X \sim Y$ så är $cX \sim cY$ för alla $c \in \mathbb{R}$.

*Lösning.* SANT. Om $c = 0$ så är $cX = cY$ så $cX \sim cY$. Annars, om $P(X = t) = P(Y = t)$ för varje $t \in \mathbb{R}$, så är $P(cX = t) = P(X = t/c) = P(Y = t/c) = P(cY = t)$, så $cX \sim cY$. □

e) Om $X \sim Y$ så är $EX = EY$.

*Lösning.* SANT. $X \sim Y$, så $P(X = t) = P(Y = t)$ för alla $t \in \mathbb{R}$. Låt $A = \{t_1, t_2, \dots\}$ vara en mängd som innehåller alla observationer med positiv sannolikhet (för både $X$ och $Y$). Vi får

$$EX = \sum_{n=1}^{\infty} t_n P(X = t_n) = \sum_{n=1}^{\infty} t_n P(Y = t_n) = EY.$$

□

**8.** Hitta diskreta slumpvariabler $X_1, X_2, \dots$ som bara antar värden i $\mathbb{N} = \{0, 1, 2, \dots\}$ så att följande påståenden stämmer.

a) $EX_1 = 1$ och $\mathrm{Var}(X_1) = 32$.

*Lösning.* Pröva till exempel $P(X_1 = n) = 1/n$ och $P(X_1 = 0) = 1 - 1/n$. Då är $EX_1 = 0 \cdot P(X_1 = 0) + n \cdot P(X_1 = n) = 0 + n/n = 1$. Eftersom $EX_1^2 = n^2 P(X_1 = n) = n$ så är $\mathrm{Var}(X_1) = EX_1^2 - (EX_1)^2 = n - 1$. Om $\mathrm{Var}(X) = 32$ måste alltså $n = 33$ och vi får då en slumpvariabel som uppfyller egenskaperna. □

b) $EX_2 < \infty$ och $\mathrm{Var}(X_2) = \infty$.

*Lösning.* Vi vet att $\sum 1/n^3$ konvergerar, så om vi har $P(X_2 = n) = c/n^3$ för $n = 1, 2, \dots$ där $c$ är vald så att $\sum c/n^3 = 1$ så får vi en sannolikhetsfördelning, och

$$EX_2 = \sum_{n=1}^{\infty} n \cdot \frac{c}{n^3} = \sum_{n=1}^{\infty} \frac{c}{n^2} < \infty,$$

men vi får också

$$\mathrm{Var}(X_2) = \sum_{n=1}^{\infty} (n - EX_2)^2 \cdot \frac{c}{n^3} = \sum_{n=1}^{\infty} \left( \frac{c}{n} - \frac{2cEX_2}{n^2} + \frac{c(EX_2)^2}{n^3} \right) = \infty$$

eftersom $\sum c/n$ divergerar medan $\sum 2cEX_2/n$ och $\sum c(EX_2)^2/n$ konvergerar. □

c) $\mathrm{Cov}(X_3, X_4) = 0$ men $X_3$ och $X_4$ är beroende.

*Lösning.* Om vi först låter $X \in \{0, 1, 2\}$ vara så att $P(X = 1) = P(X = -1) = P(X = 0) = 1/3$, så är även $P(X^3 = 1) = P(X^3 = -1) = P(X^3 = 0) = 1/3$, så vi får $EX = EX^3 = 0$. Då är $\mathrm{Cov}(X, X^2) = EX^3 - EXEX^2 = 0$, men $X$ och $X^2$ är tydligt beroende, eftersom till exempel $P(X = 0, X^2 = 0) = P(X = 0) = 1/3$, medan $P(X = 0)P(X^2 = 0) = P(X = 1)^2 = 1/9 \neq 1/3$. Låt nu $X_3 = X + 1$ och $X_4 = X^2$. Då är $X_3 \in \mathbb{N}$ och $X_4 \in \mathbb{N}$, och

$$\mathrm{Cov}(X_3, X_4) = \mathrm{Cov}(X + 1, X^2) = \mathrm{Cov}(X, X^2) + \mathrm{Cov}(1, X^2) = 0,$$

men $X_3$ och $X_4$ är beroende eftersom till exempel $P(X_3 = 1, X_4 = 0) \neq P(X_3 = 1)P(X_4 = 0)$. □

d) $P(X_5 > m + n \mid X_5 > m) = P(X_5 > n)$ för alla $m \in \mathbb{N}$ och $n \in \mathbb{N}$.

*Lösning.* Om $X_5 \sim \text{ffg}(p)$ för något $p$ så är $P(X_5 > k) = (1-p)^k$ för $k \in \mathbb{N}$, så vi får

$$\begin{aligned} P(X_5 > m+n \mid X_5 > m) = \frac{P(X_5 > m+n, X_5 > m)}{P(X_5 > m)} = \frac{P(X_5 > m+n)}{P(X_5 > m)} &= \frac{(1-p)^{m+n}}{(1-p)^m} \\ &= (1-p)^n \\ &= P(X_5 > n). \end{aligned}$$

□

e) $\lim_{n\to\infty} EX_n = \infty$ men $\lim_{n\to\infty} P(X_n \geq \epsilon) = 0$ för varje $\epsilon > 0$.

*Lösning.* Vi sätter $P(X_n = n^2) = 1/n$ och $P(X_n = 0) = 1 - 1/n$ för $n \geq 6$. Då är $EX_n = 0 \cdot P(X_n = 0) + n^2 \cdot P(X_n = n^2) = 0 + n^2/n \xrightarrow{n\to\infty} \infty$, men $P(X_n \geq \epsilon) \leq 1/n \xrightarrow{n\to\infty} 0$. □