UPPSALA UNIVERSITET
Matematiska institutionen
Örjan Stenflo

TENTAMEN I MATEMATIK
Sannolikhetsteori I, 1MS034
2017-10-25

Skrivtid: 8–13. Varje problem ger max 5 poäng. För betygen 3,4 resp. 5 krävs 18, 25 resp. 32 poäng (inklusive eventuella bonuspoäng). Tillåtna hjälpmedel: miniräknare samt formelsamling för kursen Sannolikhetsteori I, 1MS034. Lösningarna skall vara väl motiverade och försedda med förklarande text.

1. Louise ska sälja sitt hus i skärgården och hoppas få minst 7 miljoner för huset. Hon bedömer att hon kommer att lyckas med detta med sannolikhet 0.5 om det är fler än två budgivare, sannolikhet 0.1 om det är exakt två budgivare och sannolikhet 0.02 om det är färre än två budgivare. Antag att sannolikheten att det är färre än två budgivare är 0.4, sannolikheten att det är exakt två budgivare är 0.2 och att sannolikheten att det är fler än två budgivare är 0.4.

   (a) Beräkna sannolikheten att Louise lyckas.

   (b) Antag att Louise lyckas. Beräkna sannolikheten att det var fler än två budgivare.

   Lösning: Låt E_1, E_2, och E_3 beteckna händelerna att det är färre än, lika med, respektive fler än två budgivare. Låt L vara händelsen att Louise lyckas sälja sitt hus för minst 7 miljoner. Då är $P(E_1) = 0.4$, $P(E_2) = 0.2$, $P(E_3) = 0.4$, $P(L|E_1) = 0.02$, $P(L|E_2) = 0.1$, $P(L|E_3) = 0.5$.

   (a) Lagen om total sannolikhet ger

$$\begin{aligned} P(L) &= P(L|E_1)P(E_1) + P(L|E_2)P(E_2) + P(L|E_3)P(E_3) \\ &= 0.02 \cdot 0.4 + 0.1 \cdot 0.2 + 0.5 \cdot 0.4 = 0.228 \end{aligned}$$

   (b) Enligt definitionen av betingad sannolikhet är

$$P(E_3|L) = \frac{P(E_3 \cap L)}{P(L)} = \frac{P(L|E_3) \cdot P(E_3)}{P(L)} = \frac{0.5 \cdot 0.4}{0.228} = 0.877.$$

2. I en klass med 40 elever fördelar sig födelsedagarna enligt följande tabell:

Födelsemånad	Jan	Feb	Mar	Apr	Maj	Jun	Jul	Aug	Sep	Okt	Nov	Dec
Antal	3	1	6	6	5	5	0	1	3	4	3	3

   Beräkna sannolikheten att det i en slumpmässigt vald grupp om 5 av dessa elever finns minst 2 stycken som är födda i april.

Börja med att identifiera vilken sannolikhetsfördelning som är relevant här.

Lösning: Vi kan identifiera detta problem med dragning utan återläggning av 5 bollar ur en urna bestående av 6 vita (april-bollar) och 34 svarta (andra bollar). Om X är antalet personer bland de 5 utvalda som är födda i april, så gäller $X \sim Hyp(40, 5, 6/40)$. Alltså är

$$P(X \geq 2) = 1 - P(X \leq 1) = 1 - P(X = 0) - P(X = 1) = 1 - \underbrace{\frac{\binom{6}{0}\binom{34}{5}}{\binom{40}{5}}}_{\approx 0.42288} + \underbrace{\frac{\binom{6}{1}\binom{34}{4}}{\binom{40}{5}}}_{\approx 0.42288} \approx 0.154.$$

3. Låt A, B och C vara händelser med $P(A) = 0.5$, $P(B) = 0.7$ och $P(C) = 0.9$.
   (a) Bestäm det kortaste intervallet $[a, b]$ så att
   $$a \leq P(A \cap B \cap C) \leq b,$$
   alltid gäller.
   (b) Bestäm $P(A \cup B \cup C)$ under antagandet att händelserna A, B och C är oberoende.

Lösning:

(a) Eftersom $A \cap B \cap C \subseteq A$, $A \cap B \cap C \subseteq B$, och $A \cap B \cap C \subseteq C$, är

$$P(A \cap B \cap C) \leq \min(P(A), P(B), P(C)) = 0.5.$$

Vidare är

$$\begin{aligned} P(A \cap B \cap C) &= 1 - P((A \cap B \cap C)^c) \underbrace{=}_{De\ Morgan} 1 - P(A^c \cup B^c \cup C^c) \\ &\geq 1 - (P(A^c) + P(B^c) + P(C^c)) \\ &= 1 - ((1 - 0.5) + (1 - 0.7) + (1 - 0.9)) = 0.1. \end{aligned}$$

Alltså är

$$0.1 \leq P(A \cap B \cap C) \leq 0.5.$$

Den övre gränsen uppnås exempelvis om $A \subseteq B \subseteq C$, och den undre gränsen uppnås om A^c, B^c och C^c är disjunkta.

(b) Om A, B och C är oberoende, är

$$P(A \cup B \cup C) = 1 - P((A \cup B \cup C)^c) \underbrace{=}_{De\ Morgan} 1 - P(A^c \cap B^c \cap C^c)$$
$$\underbrace{=}_{ober.} 1 - P(A^c)P(B^c)P(C^c) = 1 - (1 - 0.5) \cdot (1 - 0.7) \cdot (1 - 0.9) = 0.985.$$

4. Låt (X, Y) vara en 2-dimensionell diskret stokastisk variabel med simultan sannolikhetsfunktion

$$P(X=i, Y=j) = \begin{cases} c & \text{om } i=j=1 \\ 2c & \text{om } i=j=2 \\ 0 & \text{annars} \end{cases},$$

där c är en konstant.

(a) Bestäm den marginella sannolikhetsfunktionen för X.

(b) Bestäm den betingade sannolikhetsfunktionen $q(x) := P(X = x|Y = 1)$ och avgör om X och Y är oberoende.

Lösning:

(a)

$$\begin{aligned} P(X=1) &= \textstyle\sum_j P(X=1, Y=j) = P(X=1, Y=1) = c \\ P(X=2) &= \textstyle\sum_j P(X=2, Y=j) = P(X=2, Y=2) = 2c \end{aligned}$$

$$P(X=1) + P(X=2) = 1 \Rightarrow c = \frac{1}{3}.$$

Alltså är

$$P(X=1) = \frac{1}{3}, \quad P(X=2) = \frac{2}{3}.$$

(b)

$$P(Y=1) = \sum_i P(X=i, Y=1) = P(X=1, Y=1) = \frac{1}{3}.$$

$$q(1) \quad = \quad P(X=1|Y=1) = \frac{P(X=1, Y=1)}{P(Y=1)} = \frac{\frac{1}{3}}{\frac{1}{3}} = 1.$$

$q(x)$ är en betingad sannolikhet och eftersom $q(1) = 1$ så måste $q(x) = 0$ för $x \neq 1$. $q(x) \neq P(X = x)$ och alltså är X och Y inte oberoende. Detta kan man även inse direkt från definitionen av den simultana fördelningen, eftersom man från den får att $X = Y$.

5. Låt

$$X \sim \text{Exp}(2), \text{ och låt } \quad Z = \frac{\sum_{i=1}^{200} X_i - 100}{100},$$

där X_i har samma fördelning som X för alla $1 \leq i \leq 200$.

(a) Antag att $X_i = X$ för alla i. Bestäm täthetsfunktionen för Z.

(b) Antag att X_i är oberoende. Bestäm approximativt täthetsfunktionen för Z.

Lösning: Då $X \sim \text{Exp}(2)$ är

$$E(X) = \frac{1}{2}, \quad V(X) = \frac{1}{4},$$

och X har täthetsfunktion

$$f_X(t) = \begin{cases} 2e^{-2t} & \text{om } t \geq 0 \\ 0 & \text{annars} \end{cases},$$

och fördelningsfunktion

$$F_X(t) = P(X \leq t) = \begin{cases} 1 - e^{-2t} & \text{om } t \geq 0 \\ 0 & \text{annars} \end{cases}.$$

(a) Om $X_i = X$ för alla i är $Z = \frac{\sum_{i=1}^{200} X_i - 100}{100} = \frac{200X - 100}{100} = 2X - 1$. Denna stokastiska variabel har fördelningsfunktion

$$\begin{aligned} F_Z(t) &= P(Z \leq t) = P(2X - 1 \leq t) = P(X \leq (t+1)/2) \\ &= F_X((t+1)/2) = \begin{cases} 1 - e^{-(t+1)} & \text{om } t \geq -1 \\ 0 & \text{annars} \end{cases}. \end{aligned}$$

Täthetsfunktionen för Z som ges av derivatan av dess fördelningsfunktion, blir därför $f_Z(t) = \frac{1}{2} f_X((t+1)/2) = \begin{cases} e^{-(t+1)} & \text{om } t \geq -1 \\ 0 & \text{annars} \end{cases}$.

(b) Om X_i är oberoende med samma fördelning som X för alla i är

$$E(\sum_{i=1}^{200} X_i) = 200 \cdot \frac{1}{2} = 100,$$

$$V(\sum_{i=1}^{200} X_i) = 200 \cdot \frac{1}{4} = 50,$$

och

$$N := \frac{\sum_{i=1}^{200} X_i - E(\sum_{i=1}^{200} X_i)}{D(\sum_{i=1}^{200} X_i)} = \frac{\sum_{i=1}^{200} X_i - 100}{\sqrt{50}},$$

approximativt $N(0,1)$-fördelad enligt CGS. Alltså är $Z = \frac{\sqrt{50}}{100} N$ ungefär $N(0, \sigma^2)$-fördelad där $\sigma = \frac{\sqrt{50}}{100}$. Täthetsfunktionen för Z blir därför ungefär

$$f_Z(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-t^2/2\sigma^2}, \; t \in \mathbb{R}.$$

där $\sigma = \frac{\sqrt{50}}{100}$.

6. Låt $X \sim \text{Re}(-1, 1)$.

   (a) Visa att $E[f(X)] = f(E[X])$ om $f(x) = x^3 + 1$.

   (b) Visa med ett motexempel att det inte alltid gäller att $E[f(X)] = f(E[X])$ för en godtycklig funktion $f : \mathbb{R} \to \mathbb{R}$.

Lösning:

(a) Den stokastiska variabeln X har täthetsfunktion $f_X(x) = \frac{1}{2}$ för $-1 \leq x \leq 1$ och $f_X(x) = 0$ för övriga värden på x. Då

$$E[f(X)] = \int_{-1}^{1} \frac{1}{2}(x^3 + 1)dx = \frac{1}{2}\left[\frac{x^4}{4} + x\right]_{-1}^{1} = 1,$$

och

$$f(E[X]) = f(0) = 1,$$

ser vi att $f(E[X]) = f(E[X])$, då $f(x) = x^3 + 1$.

(b) Ett enkelt exempel är $f(x) = x^2$, som ger $f(E[X]) = f(0) = 0$ och $E[f(X)] = E[X^2] = V(X) > 0$.

7. En nybörjar-roddare uppskattar att längden han kommer på ett godtyckligt årtag (mätt i enheten meter) är en viss stokastisk variabel med väntevärde 9 och varians 9. Bestäm sannolikheten att han klarar att ro 3000 meter på 325 årtag under antagandet att det inte finns några samband mellan längden på olika årtag. (Redogör som vanligt för de beteckningar, antaganden och approximationer du gör.)

Lösning: Låt X_i var längden av det i:te årtaget, och låt $S_n = \sum_{i=1}^{n} X_i$. Vi söker $P(S_{325} > 3000)$. $E(S_n) = n \cdot E(X_1) = 9n$ och tack vare oberoendet mellan de olika årtagen blir $V(S_n) = nV(X_1) = 9n$. Särskilt så får vi $E(S_{325}) = V(S_{325}) = 2925$. På grund av Centrala Gränsvärdessatsen är S_n approximativt normalfördelad då n är stort.

$$\begin{aligned}
P(S_{325} > 3000) &= 1 - P(S_{325} \leq 3000) \\
&= 1 - P\left(\frac{S_{325} - 2925}{\sqrt{2925}} \leq \underbrace{\frac{3000 - 2925}{\sqrt{2925}}}_{\doteq 1.39}\right) \\
\{\text{enligt CGS}\} &\approx 1 - \Phi(1.39) \\
\{\text{enligt tabell}\} &\doteq 1 - 0.9177 = 0.0823.
\end{aligned}$$

8. Låt X vara en Binomialfördelad stokastiska variabel med $E(X) = 4$ och $V(X) = 3$.

   (a) Bestäm $P(X \leq 2)$.

   (b) Bestäm $E(X^3)$.

Lösning:

(a) Om $X \sim Bin(n, p)$ är $E(X) = np$ och $V(X) = np(1-p)$. Då $E(X) = 4$ och $V(X) = 3$ är därför $X \sim Bin(16, 1/4)$. Enligt formelsamlingen (fördelningsfunktionen för en Binomialfördelad stokastisk variabel, $n = 16, k = 2, p = 0.25$) blir därför $P(X \leq 2) = 0.1971$.

(b) Den stokastiska variabeln X har (se formelsamlingen) momentgenererande funktion

$$\Psi(t) = (0.75 + 0.25e^t)^{16}.$$

Derivering ger

$$\Psi'(t) = 4e^t(0.75 + 0.25e^t)^{15},$$

$$\Psi''(t) = 4e^t(0.75 + 0.25e^t)^{15} + 15e^{2t}(0.75 + 0.25e^t)^{14},$$

$$\Psi'''(t) = \Psi''(t) + 30e^{2t}(0.75 + 0.25e^t)^{14} + \frac{210}{4}e^{3t}(0.75 + 0.25e^t)^{13}.$$

Det följer att

$$E(X^3) = \Psi'''(0) = \Psi''(0) + 30 + \frac{210}{4} = 19 + 30 + \frac{210}{4} = \frac{98 + 105}{2} = \frac{203}{2} = 101.5.$$

Alternativ lösning: $X = \sum_{i=1}^{16} X_i$, där $X_i \sim Be(1/4)$ är oberoende, ger

$$\begin{aligned} E(X^3) &= E((X_1 + \ldots + X_{16})^3) = 16 \cdot 15 \cdot 14 \cdot E(X_1X_2X_3) + 3 \cdot 16 \cdot 15 \cdot E(X_1X_2^2) + 16 \cdot E(X_1^3) \\ &= 16 \cdot 15 \cdot 14 \cdot (1/4)^3 + 3 \cdot 16 \cdot 15 \cdot (1/4)^2 + 16 \cdot (1/4) = 101.5, \end{aligned}$$

.

där vi använt att väntevärdet av en produkt av de oberoende och likafördelade X_i variablerna endast beror av antal distinkta variabler, ty $E(X_i^n) = 1/4$, för alla $n \geq 1$.