Tillämpad matematisk statistik LMA521
Lösningar tentamen 2022-04-13

Redovisa lösningarna i detalj. Räkna exakt så långt som möjligt. Svaret kan ges numeriskt/approximativt.
OBS: Uppgiftstext på flera sidor.

---

1. (2+2+3 poäng) Antag vi har två händelser A och B och antag

$$P(A) = 0.3, \quad P(B) = 0.4, \quad P(A \cup B) = 0.5.$$

   (a) Är händelserna A och B oberoende? Motivera ditt svar.
   (b) Beräkna $P(B \mid A)$.
   (c) Beräkna $P(A \mid B^c)$.

   Lösning:

   (a) Additionssatsen:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B),$$

   då är $P(A \cap B) = 0.3 + 0.4 - 0.5 = 0.2$. Men

$$P(A) \cdot P(B) = 0.12 \neq P(A \cap B),$$

   så A och B är *inte* oberoende.

   (b) Betingad sannolikhet:

$$\begin{aligned} P(B \mid A) &= \frac{P(B \cap A)}{P(A)} \\ &= \frac{0.2}{0.3} = 2/3 \approx 0.67. \end{aligned}$$

   (c) Betingad sannolikhet:

$$\begin{aligned} P(A \mid B^C) = \frac{P(A \cap B^C)}{P(B^C)} &= \frac{P(A) - P(A \cap B)}{1 - P(B)} \\ &= \frac{0.3 - 0.2}{0.6} = 1/6 \approx 0.17. \end{aligned}$$

2. (2+2+3 poäng) I en fabrik finns två olika maskiner. Antalet gånger maskin A går sönder under ett givet år kan antas vara Poissonfördelat med väntevärde 4, och antalet gånger maskin B går sönder under ett givet år kan antas vara Poissonfördelat med väntevärde 1. De två maskinerna går sönder oberoende av varandra.

   (a) Beräkna sannolikheten att maskin A går sönder fler än eller lika med två gånger under ett givet år.
   (b) Beräkna sannolikheten att det totala antalet maskinfel från de två maskinerna under ett givet år är högst tre.
   (c) Anta att varje gång maskin A går sönder kostar det 900 kronor för att repareras, och varje gång maskin B går sönder kostar det 1 500 kronor. Beräkna väntevärdet och standardavvikelsen för den totala kostnaden under ett år.

Lösning: Låt $X_1 \in \text{Po}(4)$ vara antal gånger maskin A går sönder, och $X_2 \in \text{Po}(1)$ antal gånger maskin B går sönder.

(a) Vi har

$$\begin{aligned} P(X_1 \geq 2) &= 1 - (P(X_1 = 0) + P(X_1 = 1)) \\ &= 1 - \frac{e^{-4}4^0}{0!} - \frac{e^{-4}4^1}{1!} \approx 0.91. \end{aligned}$$

(b) Låt $Y = X_1 + X_2$ vara det totala antalet maskinfel under året. Vi vet att $Y \in \text{Po}(4+1) = \text{Po}(5)$, så

$$\begin{aligned} P(Y \leq 3) &= \frac{e^{-5}5^0}{0!} + \frac{e^{-5}5^1}{1!} + \frac{e^{-5}5^2}{2!} + \frac{e^{-5}5^3}{3!} \\ &\approx 0.265. \end{aligned}$$

(Man kunde istället räkna ut alla separata situationer och summera dem, men det är mycket mer arbetskrävande!)

(c) Låt K vara den totala kostnaden under året, så att $K = 900X_1 + 1500X_2$. Vi har

$$\text{E}(K) = 900 \cdot \text{E}(X_1) + 1500 \cdot \text{E}(X_2) = 3600 + 1500 = 5\,100 \text{ kronor.}$$

Eftersom X_1 och X_2 är oberoende har vi

$$\text{Var}(K) = 900^2 \cdot \text{Var}(X_1) + 1500^2 \cdot \text{Var}(X_2) = 900^2 \cdot 4 + 1500^2 \cdot 1 = 5\,490\,000$$

och då

$$\text{S}(K) = \sqrt{\text{Var}(K)} = \sqrt{5\,490\,000} = 2\,343 \text{ kronor.}$$

3. (2+3 poäng) Tio oberoende mätningar av en normalfördelad stokastisk variabel med oänt väntevärde μ och okänd standardavvikelse σ gav värdena

$$82,\ 84,\ 78,\ 73,\ 79,\ 81,\ 76,\ 84,\ 91,\ 83.$$

Medelvärdet för stickprovet är $\overline{x} = 81.1$ och stickprovsstandardavvikelsen beräknades till $s = 5.0$.

(a) Ge ett tvåsidigt symmetriskt 95%:s konfidensintervall för μ.

(b) Ge ett tvåsidigt 95%:s konfidensintervall för σ.

Lösning:

(a) Tvåsidigt 95%:s konfidensintervall för μ. Vi använder t-fördelningen eftersom σ är okänd och $n < 30$.

$$\overline{x} = 81.1, \quad s = 5.0, \quad t_{1-\alpha/2}(\nu) = t_{0.975}(9) = 2.26 \quad (\text{från } \alpha = 0.05 \text{ kolumn}).$$

Intervallet är

$$\left(81.1 - 2.26 \cdot \frac{5.0}{\sqrt{10}},\ 81.1 + 2.26 \cdot \frac{5.0}{\sqrt{10}}\right) = (77.5,\ 84.7)\,.$$

(b) Tvåsidigt 95%:s konfidensintervall för σ. Vi använder χ^2-fördelningen.

$$\chi^2_{\alpha/2}(\nu) = \chi^2_{0.025}(9) = 19.023, \quad \chi^2_{1-\alpha/2}(\nu) = \chi^2_{0.975}(9) = 2.700.$$

Intervallet är

$$\left(\sqrt{\frac{9 \cdot 5.0^2}{19.023}}, \sqrt{\frac{9 \cdot 5.0^2}{2.700}} \right) = (3.4,\, 9.1)\,.$$

4. (5 poäng) En klass med 40 studenter ska delas upp i projektgrupper om fem personer. På hur många sätt kan grupperna bildas? (Svaret kan ges på exakt form och får lov att innehålla fakulteter)

Lösning: Först väljer vi de fem studenterna i den första gruppen. Det finns $\binom{40}{5}$ sätt att göra så. Sedan har vi 35 studenter kvar och väljer fem av dem för den andra gruppen, o.s.v. Vi får därför (enligt multiplikationsprincipen)

$$\binom{40}{5} \cdot \binom{35}{5} \cdot \binom{30}{5} \cdot \binom{25}{5} \cdot \binom{20}{5} \cdot \binom{15}{5} \cdot \binom{10}{5} \cdot \binom{5}{5} = \frac{40!}{(5!)^8}.$$

Men ordningen av grupperna är inte viktig, så nu dividerar vi med 8! för att inte ta hänsyn till den, och får

$$\frac{40!}{(5!)^8 \cdot 8!} = 4.7 \cdot 10^{26}$$

olika sätt att bilda projektgrupperna.

5. (2+4 poäng) Anta att livslängden i timmar för en viss typ av glödlampa är en kontinuerlig stokastisk variabel med följande frekvensfunktion:

$$f(x) = \begin{cases} \frac{1}{450} e^{-x/450}, & x \geq 0 \\ 0, & x < 0. \end{cases}$$

(a) Vad är sannolikheten att en given glödlampa fungerar längre än 500 timmar?

(b) Anta att vi har 1 000 sådana glödlampor som fungerar oberoende av varandra. Vad är approximativt sannolikheten att den genomsnittliga livslängden av glödlamporna blir mellan 440 och 460 timmar?

Lösning:

(a) Observera att livslängden, X, är exponentialfördelad med parameter $\lambda = 1/450$. Då är

$$P(X > 500) = 1 - P(X \leq 500) = 1 - F(X) = 1 - (1 - e^{-500/450}) \approx 0.33.$$

Det går också att lösa uppgiften från början:

$$\int_{500}^{\infty} \frac{1}{450} e^{-x/450}\, dx = \Big[-e^{-x/450} \Big]_{500}^{\infty} = 0 - (-e^{-500/450}) \approx 0.33.$$

(b) Låt Y vara den genomsnittliga livslängden av 1 000 glödlampor,

$$Y = \frac{\sum_{i=1}^{1000} X_i}{1000}.$$

Enligt centrala gränsvärdessatsen är Y approximativt normalfördelad med

$$\mu = E(X) \text{ och } \sigma = \frac{S(X)}{\sqrt{n}}.$$

Vi har $E(X) = 1/\lambda = 450$ och $S(X) = 1/\lambda = 450$, vilket betyder att

$$Y \approx N(450, 450/\sqrt{1000}) = N(450, 14.23).$$

Därför har vi

$$\begin{aligned}
P(440 < Y < 460) &= P(Y < 460) - P(Y < 440) \\
&\approx P(Z < \frac{460-450}{14.23}) - P(Z < \frac{440-450}{14.23}) \\
&= 2\Phi(0.70) - 1 = 2 \cdot 0.7580 - 1 \\
&= 0.516.
\end{aligned}$$

6. (3+3 poäng) Ett företag har utfört ett fullständigt faktorförsök för att undersöka hur faktorerna A, B och C påverkar deras produkt. Resultaten syns i tabell 1.

Försök	A	B	C	resultat
1	+	+	+	2.01
2	-	+	+	1.82
3	+	-	+	1.94
4	-	-	+	1.75
5	+	+	-	2.04
6	-	+	-	1.85
7	+	-	-	2.05
8	-	-	-	1.73

Tabell 1: Fullständigt 2^3 faktorförsök.

(a) Beräkna huvudeffekten l_A och samspelseffekterna l_{AB} och l_{ABC}.

Lösning:

$$l_A = \frac{2.01 + 1.94 + 2.04 + 2.05 - 1.82 - 1.75 - 1.85 - 1.73}{4} = 0.2225$$

$$l_{AB} = \frac{2.01 + 1.75 + 2.04 + 1.73 - 1.82 - 1.94 - 1.85 - 2.05}{4} = -0.0325$$

$$l_{ABC} = \frac{2.01 + 1.75 + 1.85 + 2.05 - 1.82 - 1.94 - 2.04 - 1.73}{4} = 0.0325$$

(b) Antag man har ytterligare två faktorer D och E, men man kan bara göra åtta försök så man gör ett reducerat faktorförsök. Om man har generatorerna $D = AB$ och $E = BC$, vilka alias får A och vad blir upplösningen?

Lösning: Orden blir $I_1 = ABD$, $I_2 = EBC$ och $I_3 = I_1 I_2 = ABDEBC = ACDE$. Längden på kortaste ordet är tre, så upplösningen blir III. Alias till A blir $AI_1 = BD$, $AI_2 = ABCE$ och $AI_3 = CDE$.

7. (1+2+2 poäng)

En viss typ av fluortablett skall innehålla $0.5mg$ fluor. För att kontrollera att tillverkningsprocessen är i statistisk kontroll tar man med jämna mellanrum ut en provgrupp på 3 tabletter och kontrollerar mängden fluor. Vid 10 mättillfällen får man följande resultat:

Mättillfälle	1	2	3	4	5	6	7	8	9	10
	0.49	0.48	0.51	0.51	0.52	0.50	0.51	0.50	0.49	0.48
	0.50	0.51	0.49	0.49	0.51	0.51	0.51	0.49	0.48	0.46
	0.51	0.51	0.53	0.50	0.50	0.52	0.51	0.48	0.47	0.50
$\bar{x}_i$	0.50	0.50	0.51	0.50	0.51	0.51	0.51	0.49	0.48	0.48
R_i	0.02	0.03	0.04	0.02	0.02	0.02	0	0.02	0.02	0.04
s_i	0.01	0.017	0.02	0.01	0.01	0.01	0	0.01	0.01	?

I tabellen står $\bar{x}_i$ för provgruppsmedelvärde, R_i för provgruppens variationsbredd, och s_i för provgruppens standardavvikelse.

(a) Beräkna det saknade s_i värdet för den tionde och sista provgruppen.
Lösning: $s_{10} = \sqrt{((0.48-0.48)^2 + (0.46-0.48)^2 + (0.5-0.48)^2)/2} = 0.02$

(b) Använd x_i och R_i värdena för att beräkna styrgränser (kontrollgränser) för lämpliga diagram för att avgöra ifall processen är i statistisk kontroll. Är processen i statistik kontroll? Räknehjälp: $\bar{\bar{x}} = 0.499$ och $\bar{R} = 0.023$.
Lösning: Provgruppsstorleken är 3 så gränserna blir $K_u = 0.499 - 1.023 \times 0.023 = 0.475$ och $K_{ö} = 0.499 + 1.023 \times 0.023 = 0.523$ för $\bar{x}$ diagram, och $K_u = 0$ och $K_{ö} = 2.575 \times 0.023 = 0.059$ för R-diagram. Processen är i statistisk kontroll eftersom alla $\bar{x}_i$ och R_i värden är inom gränserna.

(c) Gör samma sak som i deluppgift b, fast använd nu x_i och s_i värdena i stället. Blir det samma slutsats med de olika metoderna? Om du ej gjort deluppgift a så får du lösa uppgiften baserat på de 9 första provgrupperna istället.
Lösning: Vi har $\bar{s} = \sum_{i=1}^{10} s_i/10 = 0.0117$. Gränserna blir $K_u = 0.499 - 1.954 \times 0.0117 = 0.476$ och $K_{ö} = 0.499 + 1.954 \times 0.0117 = 0.522$ för $\bar{x}$-diagrammet, och $K_u = 0$ och $K_{ö} = 2.568 \times 0.0117 = 0.030$ för s-diagrammet. Även denna metod ger att processen är i statistisk kontroll.

8. (5+2+2 poäng) Ett företag skall köpa in ett parti på 6000 reklampennor. De använder en dubbel provtagningsplan för att avgöra ifall de skall acceptera partiet eller inte. I ett första urval kontrolleras 50 pennor, och partiet accepteras ifall 2 eller färre pennor är defekta. Om 5 eller fler är defekta avvisas det. I övriga fall går man vidare till ett andra urval. I det andra urvalet kontrolleras 50 ytterligare pennor. Om totala antalet defekta i urval 1 och 2 är fler än eller lika med 5 avvisas partiet, och om totala antalet defekta i urval 1 och 2 är mindre än eller lika med 4 så accepteras partiet. Antag nu att felkvoten i partiet är $p = 0.06$.

(a) Vad är sannolikheten att partiet accepteras?

Lösning: Binomialapproximation ger

$$\begin{aligned}
L(0.06) &= P(d_1 = 0) + P(d_1 = 1) + P(d_1 = 2) \\
&\quad + P(d_1 = 3)P(d_2 = 0) + P(d_1 = 3)P(d_2 = 1) + P(d_1 = 4)P(d_2 = 0) \\
&= \binom{50}{0} 0.06^0 0.94^{50} + \binom{50}{1} 0.06^1 0.94^{49} + \binom{50}{2} 0.06^2 0.94^{48} \\
&\quad + \binom{50}{3} 0.06^3 0.94^{47} \binom{50}{0} 0.06^0 0.94^{50} + \binom{50}{3} 0.06^3 0.94^{47} \binom{50}{1} 0.06^1 0.94^{49} \\
&\quad + \binom{50}{4} 0.06^4 0.94^{46} \binom{50}{0} 0.06^0 0.94^{50} \approx 0.468.
\end{aligned}$$

(b) Vad blir ASN och ATI?

Lösning: Sannolikheten att man går vidare till urval 2 ges av

$$P(d_1 = 3) + P(d_1 = 4) = \binom{50}{3} 0.06^3 0.94^{47} + \binom{50}{4} 0.06^4 0.94^{46} \approx 0.40435.$$

Därför blir $ASN = 50 + 50 \times 0.40435 = 70.2175$.
Vi har $A_1 = P(d_1 = 0) + P(d_1 = 1) + P(d_1 = 2) \approx 0.416$ och $A_2 = L(0.06) - A_1 = 0.052$. Så

$$\begin{aligned}
ATI &= n_1 A_1 + (n_1 + n_2) A_2 + N(1 - A_1 - A_2) \\
&= 50 \times 0.416 + 100 \times 0.052 + 6000 \times (1 - 0.416 - 0.052) = 3218.
\end{aligned}$$

(c) Vi ser att det är tre defekta pennor i urval 1. Givet detta, vad är den betingade sannolikheten att partiet avvisas?

Lösning: Om $d_1 = 3$, så avvisas partiet ifall $d_2 \geq 2$. Så den sökta betingade sannolikheten blir

$$1 - P(d_2 = 0) - P(d_2 = 1) = 1 - \binom{50}{0} 0.06^0 0.94^{50} - \binom{50}{1} 0.06^1 0.94^{49} \approx 0.810.$$

Motivera eventuella approximationer.

Lycka till!