Tillämpad matematisk statistik LMA521
Lösningar Tentamen 2022-01-11

1. (2+2 poäng) Annika spelar ett bordsspel och kastar två symmetriska 20-sidiga tärningar (numrerade från 1 till 20). Tärningarna antas oberoende. Låt A vara händelsen att hon får mer än eller lika med 16 på den första tärningen, och B att hon får mer än eller lika med 16 på den andra tärningen. Beräkna

   (a) $P(A \cup B)$,

   (b) $P(A \cap B \mid A \cup B)$.

   Lösning:

   (a)

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \quad \text{(additionssatsen)} \\ &= P(A) + P(B) - P(A) \cdot P(B) \quad \text{(oberoende)} \\ &= \frac{1}{4} + \frac{1}{4} - \frac{1}{4} \cdot \frac{1}{4} = \frac{7}{16}. \end{aligned}$$

   (b)

$$\begin{aligned} P(A \cap B \mid A \cup B) &= \frac{P((A \cap B) \cap (A \cup B))}{P(A \cup B)} \quad \text{(betingad)} \\ &= \frac{P(A \cap B)}{P(A \cup B)} \\ &= \frac{\frac{1}{16}}{\frac{7}{16}} = \frac{1}{7} \quad \text{(se a-uppgiften).} \end{aligned}$$

2. (2+4 poäng)

   (a) I en urna finns två gröna och åtta röda bollar. Man drar slumpmässigt tre bollar från urnan (utan återläggning). Vad är sannolikheten att man får exakt en grön boll?

   (b) I en stor låda finns 200 gröna och 800 röda bollar. Man drar slumpmässigt 15 bollar från lådan (utan återläggning). Vad är sannolikheten att man får högst två gröna bollar?

   Lämpliga approximationer kan användas.

   Lösning:

   (a) Låt X vara antalet gröna bollar i urvalet från urnan. $X \in \text{Hyp}(10, 3, 0.2)$.

$$P(X = 1) = \frac{\binom{2}{1}\binom{8}{2}}{\binom{10}{3}} = \frac{56}{120} = \frac{7}{15} = 0.47$$

   (b) Låt Y vara antalet gröna bollar i urvalet från lådan. $Y \in \text{Hyp}(1000, 15, 0.2)$. Eftersom $n/N = 15/1000 < 0.1$ kan den hypergeometriska fördelningen approximeras med binomialfördelningen. Y är approximativt $\text{Bin}(15, 0.2)$.

$$\begin{aligned} P(Y \leq 2) &= P(Y = 0) + P(Y = 1) + P(Y = 2) \\ &= \binom{15}{0} \cdot 0.2^0 \cdot 0.8^{15} + \binom{15}{1} \cdot 0.2^1 \cdot 0.8^{14} + \binom{15}{2} \cdot 0.2^2 \cdot 0.8^{13} \\ &= 0.035 + 0.132 + 0.231 = 0.40. \end{aligned}$$

3. (1+2+3 poäng) Man har gjort 10 oberoende mätningar av en normalfördelad stokastisk variabel med okänt väntevärde μ och okänd standardavvikelse σ. Från stickprovet har man beräknat följande summor:

$$\sum_{i=1}^{10} x_i = 332 \quad \text{och} \quad \sum_{i=1}^{10} x_i^2 = 11098.$$

(a) Beräkna stickprovsmedelvärdet $\overline{x}$ och stickprovsstandardavvikelsen s.
Om du ej löst deluppgift a, sätt in egna rimliga värden i deluppgifter b och c.

(b) Beräkna ett tvåsidigt symmetriskt 99%:s konfidensintervall för μ.

(c) Beräkna ett ensidigt uppåt begränsat 99%:s konfidensintervall för σ.

Lösning:

(a) Stickprovsmedelvärdet och -standardavvikelsen.

$$\overline{x} = \frac{332}{10} = 33.2, \quad s = \sqrt{\frac{11098 - \frac{332^2}{10}}{10-1}} = \sqrt{8.4} \quad (= 2.9).$$

(b) Tvåsidigt 99%:s konfidensintervall för μ. Vi använder t-fördelningen eftersom σ är okänt och $n < 30$.

$$\overline{x} = 33.2, \quad s = \sqrt{8.4}, \quad t_{1-\alpha/2}(\nu) = t_{0.995}(9) = 3.25.$$

Intervallet är

$$\left(33.2 - 3.25 \cdot \sqrt{\frac{8.4}{10}}, 33.2 + 3.25 \cdot \sqrt{\frac{8.4}{10}}\right) = (30.2, 36.2)\,.$$

(c) Ensidigt uppåt begränsat 99%:s konfidensintervall för σ.

$$\chi^2_{1-\alpha}(\nu) = \chi^2_{0.99}(9) = 2.088.$$

Intervallet är

$$\left[0, \sqrt{\frac{(10-1)\cdot 8.4}{2.088}}\right) = [0, 6.0).$$

4. (2+2+3 poäng) Antalet jordbävningar i ett visst område under ett år kan antas följa en Poissonprocess med parameter $\lambda = 0.5$. Det innebär bland annat att antalet jordbävningar under ett år är Poissonfördelat med parameter $\lambda = 0.5$, och att antalet jordbävningar under olika år är oberoende. Dessutom innebär det att vid en given tidpunkt är tiden till nästa jordbävning exponentialfördelad med väntevärde 2 år, och tiderna mellan jordbävningarna är oberoende exponentialfördelade stokastiska variabler med väntevärde 2 år.

   (a) Beräkna sannolikheten att högst två jordbävningar inträffar i området under ett givet år.

   (b) Om en jordbävning inträffar idag, vad är sannolikheten att väntetiden till nästa jordbävning blir kortare än sex månader?

   (c) Forskare vill studera jordbävningarna i området under lång tid. Beräkna approximativt sannolikheten att nästa 50 jordbävningar inträffar inom 70 år.

   Lösning:

   (a) Låt X vara antalet jordbävningar i området under ett år. $X \in \text{Po}(0.5)$.

$$\begin{aligned} P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= \frac{e^{-0.5} \cdot 0.5^0}{0!} + \frac{e^{-0.5} \cdot 0.5^1}{1!} + \frac{e^{-0.5} \cdot 0.5^2}{2!} \\ &= 0.607 + 0.303 + 0.076 = 0.986. \end{aligned}$$

   (b) Låt Y vara väntetiden mellan jordbävningar. Vi har $Y \in \text{Exp}(0.5)$.

$$P(Y < 0.5) = F(Y) = 1 - e^{-0.5 \cdot 0.5} = 0.221.$$

   (c) Man kan använda normalapproximation av Poissonfördelning. Eller CGS approximation enligt nedanstående (metoderna ger olika svar, men båda är OK)
   Eftersom $n = 50 > 30$ kan en normalapproximation användas enligt CGS. Summan av väntetiderna är approximativt normalfördelad med väntevärdet $n \cdot E(Y) = 50 \cdot \frac{1}{\lambda} = 100$ och standardavvikelsen $\sqrt{n} \cdot S(Y) = \sqrt{50} \cdot \frac{1}{\lambda} = 10\sqrt{2}$.

$$\begin{aligned} P(\sum_{i=1}^{50} Y_i \leq 70) &\approx P(Z \leq \frac{70 - 100}{10\sqrt{2}}) \\ &= 1 - P(Z \leq \frac{3}{\sqrt{2}}) = 1 - P(Z \leq 2.12) \\ &= 1 - 0.9830 = 0.017 \quad \text{(från tabellen).} \end{aligned}$$

5. (1+2+3 poäng)

   (a) Hur många ”ord” kan bildas från bokstäverna i ordet KANSAS? (Alla sex bokstäverna skall användas i ett ”ord”.)

   (b) Man väljer slumpmässigt ett ”ord” från a-uppgiften. Vad är sannolikheten att de två A:na står intill varandra?

   (c) Man väljer slumpmässigt ett ”ord” från a-uppgiften. Vad är den betingade sannolikheten att de två A:na står intill varandra givet att ”ordet” börjar med K?

Lösning:

(a) Antalet "ord":

$$\frac{6!}{2! \cdot 2!} = \frac{720}{2 \cdot 2} = 180.$$

(b) Om de två A:na står intill varandra kan de uppfattas som en enda enhet. Antalet gynnsamma fall är därför

$$\frac{5!}{2!} = \frac{120}{2} = 60.$$

Antalet möjliga fall är 180 (se a-uppgiften). Sannolikheten är då lika med

$$\frac{60}{180} = \frac{1}{3}.$$

(c) Antalet möjliga fall att få en ordning som börjar med K är

$$\frac{5!}{2! \cdot 2!} = \frac{120}{2 \cdot 2} = 30.$$

Antalet fall att få en ordning som börjar med K samt A:na står intill varandra är

$$\frac{4!}{2!} = \frac{24}{2} = 12.$$

Därför är den betingade sannolikheten

$$P(\text{AA intill varandra} \mid \text{börjar med K}) = \frac{12}{30} = \frac{2}{5}.$$

6. (2 + 3 +3 poäng) Medlemmarna i motorsportsklubben Mycket Fart brukar ofta ställa upp i dragracing. På grund av många förluster i rad vill de nu undersöka vilka faktorer som påverkar tiden det tar tills de kommer in i mål. De faktorer som de finner intressanta att undersöka är:

	faktor	$-$	$+$
A	*längd på bilen*	$3m$	$5m$
B	*tjocklek på hjulen*	$0.5dm$	$1dm$
C	*målade flammor på bilen*	*nej*	*ja*

För att undersöka hur mycket varje faktor påverkar har man utgått från ett fullständigt faktorförsök och mät upp körtiden enligt följande tabell:

Nr.	A	B	C	y
1	$-$	$-$	$-$	14.01
2	$+$	$-$	$-$	12.27
3	$-$	$+$	$-$	14.56
4	$+$	$+$	$-$	8.24
5	$-$	$-$	$+$	12.18
6	$+$	$-$	$+$	14.42
7	$-$	$+$	$+$	12.7
8	$+$	$+$	$+$	10.25

(a) Skatta huvudeffekten l_A och samspelseffekten l_{AC}.

Lösning:

$$I_A = \frac{-14.01 + 12.27 - 14.56 + 8.24 - 12.18 + 14.42 - 12.7 + 10.25}{4} = -2.0675.$$

$$I_{AC} = \frac{14.01 - 12.27 + 14.56 - 8.24 - 12.18 + 14.42 - 12.7 + 10.25}{4} = 1.9625$$

(b) Skattar man övriga effekter får man $l_B = -1.78$, $l_C = -0.12$, $l_{AB} = -1.78$, $l_{BC} = -0.04$, $l_{ABC} = -0.03$. Dessutom blir medelvärdet av de åtta mätningarna $l_M = 12.33$. Man vet att mätfelen vid mätningarna är normalfördelade med standardavvikelse 0.21, vilket gör att ett 95% referensintervall ges av $[-0.29, 0.29]$. Med hjälp av detta, bestäm vilka effekter som är signifikanta. Ange sedan en lämplig matematisk modell för körtiden baserat på de signifikanta effekterna.

Lösning: De effekter som blir signifikanta är de med ett absolutbelopp större än 0.29, d.v.s M, A,B,AB och AC. Modellen blir

$$y = 12.33 - \frac{-2.0676}{2}x_A - \frac{1.78}{2}x_B - \frac{1.78}{2}x_Ax_B + \frac{1.9625}{2}x_Ax_C + \xi$$
$$= 12.33 + 1.0338x_A - 0.89x_B - 0.89x_Ax_B + 0.98125x_Ax_C + \xi;\ \ \xi \in N(0{,}0.21).$$

där $x_A = 1$ om A på $+$ nivå och $x_A = -1$ om A på $-$ nivå osv...

(c) Efter att man har gjort mätningarna kom man på att resultatet kan ha påverkats av att man använde två olika förare, person 1 och person 2. Person 2 körde de korta bilarna utan flammor samt de långa bilarna med flammor. Person 1 körde de övriga. Vi har alltså ytterligare en faktor $D =$ *bilförare*. Vi har alltså egentligen gjort ett reducerat faktorförsök. Antag vi säger att nivå $+$ för D svarar mot att person 2 kör och nivå $-$ svarar mot att person 1 kör. Vad är generatorn, och vad blir alias till A, B och C?

Lösning: Vi har nu ytterligare en faktor $D =$ bilförare. Från texten får vi veta att D är på $+$ i försök 1,3,6,8 och på $-$ i försök 2,4,5,7. Generatorn blir alltså $D = AC$ och ordet blir $I = ACD$. Genom att multplicera båda sidorna av ordet fås aliasen $A = CD$, $B = ABCD$, och $C = AD$.

7. (6 poäng) En företagare köper och säljer klarinetter. Då hen är ny på marknaden och anser att ett gott omdöme är av yttersta vikt använder hen sig av en dubbel provplan för att avgöra om ett parti är bra nog. Av ett parti på 100 klarinetter plockar hen först ut 4 stycken och provspelar dessa. Om inga av dessa är defekta accepteras partiet. Är 3 eller fler dåliga avvisar hen partiet. Annars går hen vidare till urval 2. I urval 2 kontrollerar företagaren ytterligare 4 klarinetter. Om totala antalet defekta i urval 1 och urval 2 är större än eller lika med 3 avvisas partiet. Annars godkänns det. Om partiet blir avvisat kontrolleras hela partiet. Antag 15 defekta klarinetter i partiet, dvs felkvot 0.15. Beräkna sannolikheten att partiet accepteras, och beräkna väntevärdet av antalet klarinetter företagaren måste kontrollera (dvs ATI). Motivera eventuella approximationer du gör.

Lösning: Eftersom $8/100 < 0.1$ kan vi använda binomialapproximation. Acceptanssannolikheten $L(0.15)$ blir approximativt

$$\binom{4}{0}0.15^0 0.85^4 + \binom{4}{1}0.15^1 0.85^3\binom{4}{0}0.15^0 0.85^4$$
$$+\binom{4}{1}0.15^1 0.85^3\binom{4}{1}0.15^1 0.85^3 + \binom{4}{2}0.15^2 0.85^2\binom{4}{0}0.15^0 0.85^4 \approx \ldots \approx 0.90.$$

För ATI har vi att $A_1 = \binom{4}{0}0.15^0 0.85^4 \approx 0.52$ (sannolikheten att acceptera efter urval 1), och $A_2 \approx 0.9 - 0.52 = 0.38$ (sannolikheten att acceptera efter urval 2). ATI blir

$$4 \times A_1 + 8 \times A_2 + 100 \times (1 - A_1 - A_2) = 15.12.$$

8. (3+2+2 poäng) En glasstillverkare måste regelbundet se till så att mängden socker i glassen inte får oväntade förändringar under tillverkningsprocessen. Därför tar hen regelbundet ut 10 burkar med glass och mäter mängden socker per 100 ml i dessa. Resultatet för 10 mät-tillfällen syns i tabellen nedan, där $\bar{x}$ står för provgruppsmedelvärde och s står för provgrupps-standardavvikelse.

tillfälle	1	2	3	4	5	6	7	8	9	10
$\bar{x}$	22.28	22.78	26.53	20.55	23.35	18.62	21.22	21.71	15.83	26.29
s	2.89	2.11	2.51	4.8	4.91	4.86	5.14	7.6	6.26	7.38

(a) Ta fram styrgränser (kontrollgränser) för lämpliga diagram och avgör om processen är under statistisk kontroll. Du behöver ej rita diagrammen.

Lösningar: I och med att vi studerar medelvärden samt standardavvikelser är det lämpligt att använda sig av X-S diagram. Vi börjar med ett X-diagram Centrallinjen för vårt X-diagram ges av

$$CL = \bar{\bar{X}} = \frac{22.28 + 22.78 + 26.53 + 20.55 + 23.35 + 18.62 + 21.22 + 21.71 + 15.83 + 26.29}{10} = 21.9$$

Den övre och undre styrgränsen ges av

$$CL \pm A_3\bar{s} = 21.916 \pm 0.975 \cdot 4.845 \approx [17.191{,}26.641]$$

där $\bar{s}$ ges av

$$\bar{s} = \frac{2.89 + 2.11 + 2.51 + 4.8 + 4.91 + 4.86 + 5.14 + 7.6 + 6.26 + 7.38}{10} = 4.845.$$

Styrgränserna blir

- $CL = 21.916$
- $S_u = 17.191$
- $S_ö = 26.641$.

Vi beräknar vårt $S - diagram$. Styrgränserna här ges av:

- $CL = \bar{s} = 4.846$
- $S_u = B_3\bar{s} = 0.284 \cdot 4.846 \approx 1.376$

- $S_{ö} = B_4\bar{s} = 1.716 \cdot 4.846 \approx 8.316$.

Utifrån detta ser vi att processen är ej under statistisk kontroll eftersom $\bar{x}_9$ är utanför våra styrgränser.

(b) Senast glasstillverkaren räknade ut sina styrdiagram beräknade hen också ut kapabilitetsindex till $C_p = 2$. Dock har hen glömt bort vilka kravgränser som hen utgick ifrån. Hjälp glasstillverkaren att ta reda på bredden på styrgränsintervallet, dvs bestäm $T_{ö} - T_u$. Kom ihåg att vi kan skatta standardavvikelsen som $\sigma = \frac{\bar{s}}{c_4}$ där $c_4 = 0.9727$ då $n = 10$.

(c) Hen får också reda på att det korrigerade kapabilitetsindex blev $C_{pk} = 1.7$. Ta fram värden på $T_{ö}$ och T_u som uppfyller detta. Vid lösandet får du skatta väntevärdet μ på lämpligt vis med hjälp av mätvärdena ovan. Vi antar också att processens målvärde ligger precis i mitten av intervallet mellan $T_{ö}$ och T_u.

Lösning av b och c:

Vi har fått reda på att $C_p = 2$ samt att $C_{pk} = 1.7$. För att finna $T_{ö}$ och T_u utgår vi från deffinitionerna för C_p och C_{pk}

(a) $C_p = \frac{T_{ö}-T_u}{6\sigma}$

(b) $C_{kp} = C_p(1 - CM)$

(c) $CM = 2\frac{|M-\mu|}{T_{ö}-T_u}$.

Vi börjar med att beräkna vad CM måste vara. Utifrån samband 2 får vi att

$$1.7 = 2(1 - CM) \Rightarrow CM = 1 - \frac{1.7}{2} = 0.15.$$

Vi börjar nu med att beräkna bredden på vårt kravintervall. D.v.s $T_{ö} - T_u$ Vi känner inte till σ men vi kan göra en skattning av sigma. Detta gör vi genom formeln

$$s = \frac{\bar{s}}{c_4} = \frac{4.845}{0.9727} = 4.982.$$

Från samband 1 får vi att $T_{ö} - T_u = 2 \cdot 6 \cdot 4.982 = 59.784$. Detta innebär att bredden på kravintervallet blir 59.784.

Vi tar nu fram mittpunkten för intervallet. Detta gör vi genom vårt samband för CM. Stoppar vi in $T_{ö} - T_u$ samt att μ skattas med 21.916 (centrumlinjen i diagrammet) får vi

$$2\frac{|M - 21.916|}{59.784} = 0.15 \Rightarrow |M - 21.916| = 0.15 \cdot 59.784/2 = 4.4838.$$

Denna ekvation har nu två lösningar $M = 21.916 \pm 4.4838$. Båda dessa fungerar bra för att få rätt på uppgiften. Låt oss utgå från att $M = 21.916 + 4.4838 = 26.3998$. Vi vet att M ligger mitt i mellan $T_{ö}$ samt T_u samt att $T_{ö} - T_u = 59.784$. Detta gör att vi kan skriva styrgränserna som $T_u = 26.3998 - \frac{59.784}{2} = -3.4922$, samt $T_{ö} = 26.3998 + \frac{59.784}{2} = 56.2918$. Det verkar som om de använda toleransgränserna är ganska grova, i synnerhet den undre är alltid uppfylld. Företagaren borde nog tänka över sina parametrar.