LINKÖPINGS UNIVERSITET

Matematiska institutionen

08 juni 2021, kl. 08.00-12.00
Examinator: Jörg-Uwe Löbus

Alla hjälpmedel är tillåtna.

Betygsgränser 3: minst 8 poäng, 4: minst 11.5 poäng, 5: minst 15 poäng (av 18)

(1) Låt A och B vara händelser så att $P(A) = 0.4$ och $P(B \cap A^*) = 0.3$, där A^* betecknar komplementet till A.

(a) Bestäm $P(A \cup B)$ och $P(A^*)$. (1p)

(b) Om händelserna är oberoende, vad är $P(B)$? (1p)

(c) Om händelserna inte är oberoende, men man vet att $P(A \mid B) = 0.5$, vad är $P(B)$? (1p)

(2) Vi har en rättvis tärning med 6 sidor som vi kastar 4 gånger och räknar antalet 6:or vi får, låt X vara en stokastisk variabel som innehåller resultatet (antalet 6:or). I ett spel så satsar man 10 kronor och vinner enligt följande. Om vi får två eller färre 6:or vinner vi inget. Om vi får tre 6:or vinner vi 500 kronor, och om vi lyckas slå fyra 6:or vinner vi 2000 kronor.

(a) Vad är $P(X \geq 3)$ och $E(X)$? (2p)

(b) Om man spelar en gång, vad är väntevärdet för vinsten? Tycker du spelet är rättvist? (1p)

(3) Filmsfantasten Sture samlar på italienska filmer. Sture brukar beställa dessa från en importör där han riskerar att få betala tullavgifter om värdet är för högt, så han beställer endast en film åt gången och inväntar leveransen innan han genast gör en ny beställning. Denna leverantör har 32 filmer som Sture inte äger, och leveranstiden från och med att beställningen görs uppskattar Sture är exponentialfördelad med väntevärde 8 dagar. Vad är sannolikheten att Sture måste vänta i högst 230 dagar innan han stolt kan visa upp sin bokhylla med alla dessa filmer? Ett approximativt svar duger om det motiveras. Vi antar också att leveranstiderna för olika filmer är tillräckligt oberoende av varandra. (3p)

(4) Antag att vi vid följande x-värden har mätt upp motsvarande y-värden.

x	1	2	4	5	8
y	3.20	4.72	8.00	9.88	13.81

Modellen är att y_j är en observation av en stokastisk variabel $Y_j = b_0 + b_1 x_j + \epsilon_j$, där $\epsilon_j \sim N(0, \sigma)$ är parvis oberoende, $j = 1, 2, 3, 4, 5$.

(a) Hitta den linjära regressionslinjen för y med avseende på x. (1p)

(b) Visa $E[Y_j] = b_0 + b_1 x_j$. (0.5p)

(c) Visa att skattningarna för parametrarna b_0 och b_1 är väntevärdesriktiga. Ledning: Betrakta de motsvarande stokastiska variablerna,

$$B_1 = \frac{\sum x_j Y_j - n\bar{x}\bar{Y}}{\sum x_j^2 - n\bar{x}^2}$$

och

$$B_0 = \bar{Y} - B_1 \bar{x}$$

och använd (b)-delen. (1.5p)

(5) Man korsar två typer av bönor med runda gula resp. skrynkliga gröna frön. Enligt Mendels ärftlighetslära skall man få fyra typer av frön med följande sannolikheter:

$$P(\text{runda gula}) = \frac{9}{16}, \quad P(\text{skrynkliga gula}) = \frac{3}{16},$$

$$P(\text{runda gröna}) = \frac{3}{16}, \quad P(\text{skrynkliga gröna}) = \frac{1}{16}.$$

När man genomförde ett försök med 560 korsningar av den runda gula med den skrynkliga gröna typen fick man 330 runda gula, 100 skrynkliga gula, 112 runda gröna, 18 skrynkliga gröna. Testa på den approximativa risknivån 5 % om Mendels teori verkar vara förenlig med ovanstående resultat. (3p)

(6) Låt $X_1, \ldots, X_n$ vara oberoende poissonfördelade stokastiska variabler med parameter λ.

(a) Härled ML-skattningen av λ baserat på de observationerna $x_1, \ldots, x_n$ av $X_1, \ldots, X_n$. (2p)

(b) Beräkna skattningens väntevärde och varians. (1p)

Lösningar

(1) (a) Eftersom händelserna A och $B \cap A^*$ är oförenliga och $A \cup B = A \cup (B \cap A^*)$ (rita Venndiagram) så får vi $P(A \cup B) = P(A) + P(B \cap A^*) = 0.7$. Vidare så är $P(A^*) = 1 - P(A) = 0.6$.

(b) Om A och B är oberoende så gäller att $P(A \cap B) = P(A)P(B)$. Vi söker $P(B)$, och vet att

$$0.7 = P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B)(1 - P(A))$$

där vi utnyttjat att A och B är oberoende. Vi löser ut $P(B)$ och erhåller

$$P(B) = \frac{0.7 - P(A)}{1 - P(A)} = \frac{0.3}{0.6} = 0.5$$

(c) Ifrån uppgiften vet vi att $P(A \mid B) = 0.5$. Vidare så gäller fortfarande

$$0.7 = P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

så

$$0.5 = \frac{P(A \cap B)}{P(B)} = \frac{P(B) - 0.3}{P(B)}$$

Vi löser ut $P(B) = 0.3/0.5 = 0.6$.

(2) (a) Situationen ger att X är binomialfördelad: $X \sim \mathrm{Bin}(4, 1/6)$. Vidare så låter vi funktionen g vara definierad enligt $g(k) = 500$ för $k = 3$ och $g(k) = 2000$ för $k = 4$. Annars är $g(k) = 0$. Vi söker nu följande: $P(X \geq 3), E(X)$ samt $E(g(X))$. Vi börjar med sannolikheten:

$$\begin{aligned} P(X \geq 3) &= P(X = 3) + P(X = 4) \\ &= \binom{4}{3}\left(\frac{1}{6}\right)^3 \frac{5}{6} + \binom{4}{4}\left(\frac{1}{6}\right)^4 = \frac{20+1}{6^4} \approx 0.0162 \end{aligned}$$

(b) Tydligt att det är ganska låg sannolikhet för vinst. Eftersom X är binomialfördelad blir

$$E(X) = 4 \cdot 1/6 = 2/3 \approx 0.6667$$

Vi utnyttjar nu formel för väntevärde av en funktion av en stokastisk variabel för att beräkna $E(g(X))$:

$$\begin{aligned} E(g(X)) &= \sum_{k=0}^{4} g(k) \binom{4}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{4-k} \\ &= 500 \cdot \binom{4}{3}\left(\frac{1}{6}\right)^3 \frac{5}{6} + 2000 \cdot \binom{4}{4}\left(\frac{1}{6}\right)^4 \\ &= \frac{500 \cdot 20 + 2000}{6^4} \approx 9.2593 \end{aligned}$$

Det lönar sig alltså inte i längden att spela spelet då insatsen är högre än det förväntade värdet av vinsten i varje omgång (en följd av de stora talens lag).

(3) Vi låter X_i beteckna leveranstiden för film $i, i = 1, 2, \ldots, 32$. De stokastiska variablerna X_i antas vara parvis oberoende och likafördelade: $X_i \sim \text{Exp}(8)$. Vi betraktar summan

$$Y = X_1 + X_2 + \cdots + X_{32}.$$

Fördelningen för Y är bökig att räkna ut direkt (i själva verket blir Y gammafördelad med parametrarna 32 och 8), så vi försöker med en normalapproximation istället. Väntevärde och varians beräknas enkelt till $E(Y) = 32 \cdot 8 = 256$ och $V(Y) = 32 \cdot V(X_1) = 32 \cdot 8^2 = 2048$. Enligt centrala gränsvärdessatsen så blir summan Y approximativt normalfördelad, så vi får

$$P(Y \leq 230) \approx \Phi((230 - 256)/\sqrt{2048}) = 1 - \Phi(26/\sqrt{2048}) = 0.2828.$$

(4) (a) Ifrån siffrorna kan vi räkna ut

$$\bar{x} = 4, \quad \bar{y} = 7.922, \quad \sum_{j=1}^{5} x_j y_j = 204.52, \quad \sum_{j=1}^{5} x_j^2 = 110.$$

Vi använder dessa siffror för att beräkna koefficienterna:

$$b_1 = \frac{\sum_{j=1}^{5} x_j y_j - n\bar{x}\bar{y}}{\sum_{j=1}^{5} x_j^2 - n\bar{x}^2} = \frac{204.52 - 5 \cdot 4 \cdot 7.922}{110 - 5 \cdot 4^2} = 1.536$$

och

$$b_0 = \bar{y} - b_1\bar{x} = 7.922 - 1.536 \cdot 4 = 1.778.$$

Regressionslinjen ges nu av $y = b_0 + b_1 x = 1.778 + 1.536x$.

(b,c) För att visa att b_0 och b_1 är väntevärdesriktiga skattningar betraktar vi motsvarande stokastiska variabler:

$$B_1 = \frac{\sum x_j Y_j - n\bar{x}\bar{Y}}{\sum x_j^2 - n\bar{x}^2}$$

och

$$B_0 = \bar{Y} - B_1\bar{x}.$$

Nämnaren i B_1 är icke-slumpmässig (inget som varierar när x-värderna är fixa), så vi beräknar väntevärdet av täljaren:

$$E\left(\sum x_j Y_j - n\bar{x}\bar{Y}\right) = \sum x_j E(Y_j) - n\bar{x}E(\bar{Y})$$

Då $Y_j = b_0 + b_1 x_j + \epsilon_j$ följer att $E(Y_j) = b_0 + b_1 x_j$ (ϵ_j har väntevärde noll), så högerledet ovan kan skrivas om som $\sum x_j (b_0 + b_1 x_j) - n\bar{x}\frac{1}{n}\sum (b_0 + b_1 x_j) = b_0 n\bar{x} + b_1 \sum x_j^2 - (n\bar{x}b_0 + b_1 n\bar{x}^2) = b_1 \left(\sum x_j^2 - n\bar{x}^2\right)$där vi känner igen den andra faktorn som nämnaren i ekvation ovan. Följdaktligen blir

$$E(B_1) = \frac{1}{\sum x_j^2 - n\bar{x}^2} E\left(\sum x_j Y_j - n\bar{x}\bar{Y}\right) = b_1$$

så B_1 är en väntevärdesriktig skattning. För B_0 utnyttjar vi att B_1 är väntevärdesriktig:

$$E(B_0) = b_0 + b_1\bar{x} - E(B_1)\bar{x} = b_0.$$

(5) Vi vill testa H_0 given i uppgiften, dvs.

$$H_0 : P(\text{runda gula}) = \frac{1}{16}, \quad \ldots$$

Teststorheten är

$$Q = \frac{\left(330 - 560 \cdot \frac{9}{16}\right)^2}{560 \cdot \frac{9}{16}} + \frac{\left(100 - 560 \cdot \frac{3}{16}\right)^2}{560 \cdot \frac{3}{16}} + \frac{\left(112 - 560 \cdot \frac{3}{16}\right)^2}{560 \cdot \frac{3}{16}} + \frac{\left(18 - 560 \cdot \frac{1}{16}\right)^2}{560 \cdot \frac{1}{16}}$$

$$= 9.68 > \chi^2_{0.05}(3) = 7.81 .$$

Detta innebär att vi har svaret: H_0 förkastas på $\alpha = 0.05$.

(6) (a)

$$\begin{aligned} L(\lambda) &= P(X_1 = x_1 \ldots X_n = x_n; \lambda) \\ &= P(X_1 = x_1) \cdot \ldots \cdot P(X_n = x_n) \\ &= \prod_{i=1}^{n} e^{-\lambda} \cdot \frac{\lambda^{x_i}}{x_i!} \\ &= e^{-n\lambda} \cdot \prod_{i=1}^{n} \frac{1}{x_i!} \cdot \lambda^{\sum_{i=1}^{n} x_i} , \end{aligned}$$

$$\ln L(\lambda) = -n\lambda - \sum_{i=1}^{n} \ln(x_i!) + \ln \lambda \cdot \sum_{i=1}^{n} x_i ,$$

$$\frac{d \ln L(\lambda)}{d\lambda} = -n + \frac{1}{\lambda} \cdot \sum_{i=1}^{n} x_i .$$

Av $\dfrac{d \ln L(\lambda)}{d\lambda} = 0$ fås

$$\lambda^*_{\text{obs}} = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x} .$$

Eftersom $\lim \ln L(\lambda) = -\infty$ då $\lambda \to 0$ eller $\lambda \to \infty$ så följer att extremvärdet är ett maximum och att $\lambda^*_{\text{obs}} = \bar{x}$ är ML-skattningen av λ.

(b)

$$E[\lambda^*] = E[\bar{X}] = E[X_1] = \lambda ,$$

$$Var(\lambda^*) = \frac{1}{n^2} Var(X_1 + \ldots + X_n) = \frac{1}{n} Var(X_1) = \frac{\lambda}{n} .$$