Kurs: MVE650
Lärare: T. Vilkas

14 mars 2022
kl. 14.00-18.00

CHALMERS
UNIVERSITY OF TECHNOLOGY

Tentamen

i kursen

Statistik och sannolikhetsteori

Hjälpmedel: *Godkänd räknedosa, egen* **handskriven** *formelsamling (fyra A4-sidor) samt med skrivningen utdelade tabellsidor.*

Examinator: Timo Vilkas, 031 772-5316

OBS! Markera tydligt på varje inlämnat blad, vilken uppgift det gäller. Var noga med att förklara vad du gör, hur och varför. Det är i hög grad lösningen som ger poäng, inte själva svaret.

För betyg 3 (godkänt) krävs 16 poäng, 24 poäng ger betyg 4 och 32 poäng ger betyg 5 .

Uppgift 1. Vi kastar en gul och en blå symmetrisk sexsidig tärning samtidigt. Låt X vara minimum av antal prickar som de båda tärningar visar.

(a) Vilka utfall är möjliga för X? Bestäm sannolikhetsfunktionen p_X. (1p)

(b) Beräkna väntevärde och varians för X. (2p)

(c) Beräkna den betingade sannolikheten att båda tärningar visar 4 prickar, givet att $X = 4$. (2p)

Uppgift 2. En kurs med $n = 65$ deltagare examineras genom en skriftlig tentamen med 7 uppgifter och 40 poäng totalt. Antag att antalet poäng hos en godtycklig kursdeltagare har en fördelning med väntevärde $\mu = 24$ och standardavvikelse på $\sigma = 10$ samt att studenternas individuella resultat $X_1, \ldots, X_n$ är oberoende.

(a) Vilken approximativ fördelning har då kursens genomsnittliga resultat $\overline{X}_n$ och vilken sats är denna approximation baserat på? Beräkna (approximativt) sannolikheten att kursgenomsnittet landar mellan 22 och 25 poäng. (3p)

(b) Låt sannolikheten för att en godtycklig student löste uppgift 7 korrekt vara $p = 0.04$ (samma för varje, oberoende för olika studenter). Låt vidare beteckna Y antalet korrekta lösningar av uppgift 7 i hela kursen. Vilken exakt fördelning har slumpvariabeln Y och vilken approximation lämpar sig? Beräkna (approximativt) sannolikheten att exakt tre studenter löste uppgift 7 korrekt. (2p)

Uppgift 3. Låt (X,Y) vara en kontinuerligt fördelad slumpvektor med täthetsfunktion

$$f(x,y) = 2x \cdot \mathrm{e}^{-y}, \quad \text{för alla} \quad 0 \leq x \leq 1,\ 0 \leq y < \infty$$

och lika med 0 annars.

(a) Beräkna de marginella täthetsfunktionerna f_X (för slumpvariabeln X) och f_Y (för slumpvariabeln Y) samt väntevärdet $\mathbb{E}X$. (3p)

(b) Är X och Y oberoende? (1p)

(c) Beräkna sannolikheten $\mathbb{P}(X \geq Y)$. (2p)

Uppgift 4. Karl-Gustav åker skidor och noterar varje gång han åker upp med linbanan numret på gondolen han sitter i. Om det totalt finns $\theta \in \mathbb{N}$ gondoler får han på så sätt ett stickprov ur den likformiga fördelningen $\mathrm{unif}(\{1,2,\ldots,\theta\})$. Antalet θ har Karl-Gustav dock ingen kunskap om och vill därför skatta det baserat på sina n observationer $X_1, X_2, \ldots, X_n$.

(a) Bestäm skattaren $\hat{\theta}_{\mathrm{MM}}$ baserat på momentmetoden. (2p)

(b) Vad innebär det att en skattare är väntevärdesriktigt? Vad innebär konsistens? Undersök om $\hat{\theta}_{\mathrm{MM}}$ är väntevärdesriktig resp. konsistent. (2p)

(c) Bestäm skattaren $\hat{\theta}_{\mathrm{ML}}$ baserat på maximum-likelihood-metoden. Är denna väntevärdesriktigt? (2p)

Ledning: För att kunna besvara frågan om konsistens behöver man inte beräkna det exakta värdet på variansen för en $\mathrm{unif}(\{1,2,\ldots,\theta\})$-fördelning, utan det räcker med att denna är konstant (med avseende på stickprovets storlek n).

Uppgift 5. När man ringer till SJs kundservice hamnar man nästan alltid i kön. Det är rimligt att anta att väntetiden i kön vid en slumpvist vald tidspunkt (som faller inom deras öppetider) är normalfördelad. Dock är varken väntevärdet μ eller variansen σ^2 för väntetiden känd. I en statistisk undersökning ringde man till kundservicen ett flertal gånger (vi kan anta att resulterande värden är oberoende) och observerade följande väntetider (i minuter):

24.2	13.7	17.0	7.5	5.3
8.8	14.4	35.0	14.7	9.4

(a) Beräkna stickprovsmedelvärdet och stickprovsvariansen. (2p)

(b) Beräkna ett 95% konfidensintervall för väntevärdet μ baserat på stickprovet. (2p)

(c) Beräkna ett 95% konfidensintervall för variansen σ^2 baserat på stickprovet. (2p)

(d) Genomför en ensidig hypotesprövning med signifikansnivå $\alpha = 0.05$ för att kunna ta ställning till påståendet att man inte behöver vänta i kön mer än 11 minuter i genomsnitt. (2p)

Uppgift 6. I en studie undersöktes skärmtiden bland unga kvinnor och män. Studien omfattade 25 unga kvinnor och 22 unga män som rapporterade sin genomsnittliga skärmtid per dag. Stickprov 1 (kvinnor) hade ett medelvärde på 5.8 timmar, stickprov 2 (män) ett medelvärde på 4.9 timmar. Tidigare undersökningar har visat att skärmtiden kan antas vara normalfördelad, med en varians $\sigma_1^2 = 3$ för kvinnor och $\sigma_2^2 = 2.5$ för män. Man har tidigare utgått ifrån att skärmtiden i genomsnitt är samma för båda kön, dvs. $\mu_1 = \mu_2$.

(a) Genomför en hypotesprövning (med signifikansnivå $\alpha = 0.1$) baserat på datan ovan och ta ställning till frågan om stickproven ger stöd till likhets-hypotesen eller om det finns en signifikant skillnad. (2p)

(b) Beräkna motsvarande P-värde för testet. (1p)

(c) Beräkna testets styrka givet att $\mu_1 - \mu_2 = 0.5$, dvs. att medelvärden faktiskt skiljer sig om en halvtimme. (2p)

Ledning: Observera att förkastningsområdet består av två intervall.

Uppgift 7. När det gäller besökare på Matematiska institutionens webbsida utgår man ifrån att det (under dagstid) kan modelleras med en Poisson process med intensitet 1 (per tidsenhet minut), dvs. i genomsnitt en besökare per minut och antal besökare i disjunkta tidsintervall är oberoende slumpvariabler.
Antalet besökare X under en tidsperiod av två minuter har då en Poisson fördelning med parameter $\lambda = 2$, alltså

$$\mathbb{P}(X = k) = \frac{2^k}{k!}\,\mathrm{e}^{-2} \quad \text{för alla } k \in \mathbb{N}_0.$$

Vi vill hypotespröva detta fördelningsantagande med hjälp av χ^2-testet. Antalet besökare i 100 disjunkta två minuters tidsintervall registreras och leder till följande statistik:

antal besökare	0	1	2	3	4	≥ 5
antal tidsintervall	16	20	26	24	5	9

(a) Beräkna sannolikheten utifrån modellen för varje kategori i indelningen ovan, alltså $p_0 = \mathbb{P}(X = 0), \ldots, p_4 = \mathbb{P}(X = 4)$, $p_5 = \mathbb{P}(X \geq 5)$. (1p)

(b) Formulera motsvarande hypoteser och pröva fördelningsantagandet $X \sim \text{Poiss}(2)$, med hjälp av χ^2-testet baserat på stickprovet/indelningen ovan, på signifikansnivå $\alpha = 0.05$. Vilken slutsats kan du dra utifrån din beräkning? (4p)