Avd. Matematisk statistik

KTH Matematik

EXEMPELTENTAMEN I SANNOLIKHETSTEORI OCH STATISTIK,
Tillåtna hjälpmedel: Formel- och tabellsamling i Matematisk statistik (utdelas vid tentamen).

Tentamen består av två delar, benämnda del I och del II. Del I består av 12 uppgifter. På denna del skall endast svar anges, antingen i form av ett numeriskt värde med tre värdesiffrors noggrannhet eller i form av val av ett av de möjliga svarsalternativen. Studenter som är godkända på kontrollskrivningen behöver ej besvara uppgift 1-3, utan får tillgodoräkna sig dessa tre uppgifter. Gränsen för godkänt är preliminärt 9 poäng. Möjlighet att komplettera ges för tentander med, preliminärt, 8 poäng. Tid och plats för komplettering kommer att anges på kursens hemsida.
Del II består av 4 uppgifter och varje korrekt lösning ger 10 poäng. Del II rättas bara för studenter som är godkända på del I och poäng på del II krävs för högre betyg än E. På denna del skall resonemang och uträkningar skall vara så utförliga och väl motiverade att de är lätta att följa. Införda beteckningar skall förklaras och definieras och numeriska svar skall anges med minst två värdesiffrors noggrannhet. Studenter som är godkända på datorlaboration II får 3 bonuspoäng på del II och uppg 12 på del I tillgodoräknad.

Tentamen kommer att vara rättad inom tre arbetsveckor från skrivningstillfället och kommer att finnas tillgänglig på studentexpeditionen minst sju veckor efter skrivningstillfället.

Del I

Uppgift 1

Låt A och B vara två oberoende händelser. Det gäller att $P(A) = 0.4$ och att $P(B) = 0.3$.
Bestäm $P(B|A^*)$.
Svar:...........

Uppgift 2

Låt a vara en positiv konstant och låt X vara en kontinuerlig stokastisk variabel med täthetsfunktion

$$f_X(x) = 3x^2/a^3,$$

för $x \in [0, a]$ och $f_X(x) = 0$ annars. Bestäm $E(X^2)$.

A: $3a/4$

B: $9a/7$

C: $3a^2/5$

D: $3a^2/80$

Uppgift 3

Vi har två urnor, där den ena innehåller 2 vita och 4 svarta kulor och den andra innehåller 1 vit och 5 svarta kulor. Vi drar en kula ur vardera urnan. Bestäm sannolikheten att vi får en vit och en svart kula.

Svar:...........

Uppgift 4

Betrakta en kundmottagning där tiden X_i som det tar att betjäna kund i är Exp(λ)-fördelad med $\lambda = 10h^{-1}$. Antag att de s.v. X_i är oberoende och att kund $i + 1$ börjar betjänas omedelbart efter att kund i betjänats färdigt. Bestäm den approximativa fördelningen för den totala tiden som det tar att betjäna de 100 första kunderna.

A: $\mathrm{N}(10, 1/10)$

B: $\mathrm{N}(10, 1)$

C: $\mathrm{N}(10, 10)$

D: $\mathrm{N}(100, 10)$

Uppgift 5

Ett varuparti om 100 enheter innehåller 6 defekta enheter. En köpare väljer på måfå och utan återläggning ut 5 enheter och undersöker dessa. Låt X beteckna antalet defekta enheter bland de 5 utvalda. Bestäm fördelningen för X.

A: $\mathrm{Bin}(5, 6/100)$

B: $\mathrm{Bin}(6, 5/100)$

C: $\mathrm{Hyp}(100, 5, 6/100)$

D: $\mathrm{Hyp}(100, 6, 5/100)$

E: $\mathrm{Po}(3/10)$

Uppgift 6

Antalet tåg per dag som ankommer med mer än fem minuters försening till en viss station antas vara Po(2)-fördelat. Bestäm sannolikheten att minst tre tåg är mer än fem minuter försenade under en given dag.

A: 0.143

B: 0.180

C: 0.323

D: 0.857

Uppgift 7

Låt datamängden $\{1, 3, 8, 7, 5\}$ vara given. Varje data antas vara ett utfall av en stokastisk variabel X_i, där X_i:na antas vara oberoende. Vi skattar $E(X_i) = \mu$ med medelvärdet $\bar{x}$ och $D(X_i) = \sigma$ med roten ur stickprovsvariansen. Ange medelfelet för skattningen av μ.

A: 1.28

B: 1.43

C: 2.86

D: 5.72

Uppgift 8

Låt mätdatat $x_1 = 0.7$ och $x_2 = 0.4$ vara givet och antag att varje mätdata är utfall av oberoende stokastiska variabler som alla har täthetsfunktion

$$f_X(x) = \theta x^{\theta-1}, \quad \text{för } 0 < x < 1,$$

för något fixt värde på parametern $\theta > 1$. Ange det numeriska värdet av Maximum Likelihood-skattningen av θ.

Svar:...........

Uppgift 9

Låt $\bar{x} = 137.0$, $\bar{y} = 208.0$, $s_x^2 = 92.5$, $s_y^2 = 163.0$, $n_x = 5$ och $n_y = 5$ vara givet och antag att $X_i \in N(\mu_x, \sigma)$ och $Y_i \in N(\mu_y, \sigma)$ samt att alla dessa stokastiska variabler är oberoende. Ange undre gränsen för det 99%-iga tvåsidiga konfidensintervallet $I_{\mu_x - \mu_y}$.

A: -85.0

B: -89.4

C: -95.0

D: -98.8

Uppgift 10

ntag att $X \in Bin(200, p)$. Vi gör ett försök och får $x = 23$. Ange den övre gränsen för det ensidigt uppåt begränsade konfidensintervallet I_p för p. Använd approximativ konfidensgrad 95%.

A: 0.152

B: 0.159

C: 0.162

D: 0.776

Uppgift 11

Antag att $X \in \text{ffg}(p)$ och låt $H_0 : p = \frac{1}{13}$. Vi förkastar H_0 om vi lyckas redan i första eller andra försöket (dvs. om $x = 1$ eller $x = 2$). Bestäm signifikansnivån för detta hypotestest.

Svar: ....

Uppgift 12

Vid test av given fördelning förkastas H_0 om

A: $Q < \chi^2_\alpha(f)$

B: $Q > \chi^2_\alpha(f)$

C: $Q < \chi^2_{\frac{\alpha}{2}}(f)$

D: $Q > \chi^2_{\frac{\alpha}{2}}(f)$

Del II

Uppgift 1

I faderskapsmål undersöks ofta blodgrupper hos moder, barn och den utpekade fadern. För enkelhets skull betraktar vi bara det s.k. $AB0$-systemet (i verkligheten undersöks ett flertal sådana blodgruppssystem). För barn vars moder har blodgrupp 0 ges den av faderns blodgrupp betingade sannolikheten för att barnet skall få blodgrupp $A0$ av följande tabell:

Faderns blodgrupp	Sannolikheten för att barnet skall få blodgrupp $A0$
$A0$	1/2
AA	1
AB	1/2
Övriga	0

Sannolikheten att en på måfå vald man skall ha blodgrupp $A0, AA, AB$ respektive ”Övriga”är 0.36, 0.08, 0.02 respektive 0.54. Beräkna sannolikheten att fadern har AA när barnet har $A0$ och modern 0. (10 p)

Uppgift 2

Den danske missionären Hans Egede Saabye som var stationerad på Grönland åren 1770-1778 nedtecknade i sin dagbok följande (i fri översättning): På Grönland är alla vintrar stränga, men de är ändå inte likadana. Danskarna har noterat att när vintern i Danmark är sträng, med vårt mått mätt, så är vintern på Grönland mild, i grönländska mått mätt”. För att testa påståendet att det finns ett beroende mellan grönländska och danska vintertemperaturer kan vi undersöka nedanstående tabell. Tabellen är baserad på mätningar av medeltemperaturen i Januari i Nuuk på Grönland och i den danska huvudstaden Köpenhamn under åren 1866-2013. På vardera platsen är januarimånaderna indelade i tre kategorier: sträng (om medeltemperaturen är minst 0.8 standardavvikelser lägre än normalvärdet), mild (om medeltemperaturen är minst 0.8 standardavvikelser högre än normalvärdet) och normal (om medeltemperaturen avviker mindre än 0.8 standardavvikelser från normalvärdet).

	Sträng vinter Nuuk	Normal vinter Nuuk	Mild vinter Nuuk
Sträng vinter Köpenhamn	2	19	9
Normal vinter Köpenhamn	13	51	22
Mild vinter Köpenhamn	12	18	2

Formulera ett hypotestest på signifikansnivån 1% och undersök om det finns ett beroende mellan grönländska och danska vintertemperaturer. (10 p)

Uppgift 3

SCB:s partisympatiundersökning görs två gånger om året. Lite förenklat får ett riksomfattande slumpmässigt urval av i riksdagsval röstberättigade personer besvara frågan “Vilket parti skulle du rösta på om det vore riksdagsval någon av de närmaste dagarna?”. Nedan finner du resultaten från SCB:s mätningar i maj 2018 och november 2017 för alliansen, de rödgröna partierna respektive övriga partier.

	Alliansen	Rödgröna	Övriga	Totalt
Maj 2018	1885	1945	802	4632
November 2017	1910	2126	679	4715

Vi vill undersöka utvecklingen för de rödgröna partierna.

(a) Skatta förändringen i stödet för de rödgröna partierna samt medelfelet för denna skattning. (3 p)

(b) Bestäm ett konfidensintervall med approximativ konfidensgrad 95% för förändringen i stödet för de rödgröna partierna mellan de båda undersökningstillfällena. (4 p)

(c) Utför ett test på signifikansnivå 5% om det varit någon statistiskt säkerställd förändring i stödet för de rödgröna partierna. Ange tydligt de uppställda hypoteserna och motivera tydligt vilken slutsats som dras från testet. (3 p)

Uppgift 4

I en processor för akustiska signaler observerar man en stokastisk variabel Y som ges av absolutbeloppet av $X \in N(0, \sigma)$, $\sigma > 0$, dvs. $Y = |X|$. Vi har inte tillgång till direkta observationer av X. Standardavvikelsen σ är inte känd och bör skattas på basis av n oberoende observationer $y_1, \ldots, y_n$ av Y.

a) En intuitivt tilltalande skattning av σ^2 ges av

$$s^* = \frac{1}{n}\sum_{i=1}^{n} y_i^2,$$

där $y_1, \ldots, y_n$ är oberoende observationer av Y. Avgör om s^* är en väntevärdesriktig skattning. (4 p)

b) Härled täthetsfunktionen $f_Y(y)$ för Y. (6 p)

Lycka till!

Avd. Matematisk statistik

KTH Matematik

LÖSNINGSFÖRSLAG EXEMPELTENTAMEN I SANNOLIKHETSTEORI OCH STATISTIK

Del I

1. 0.300
2. C
3. 0.389
4. B
5. C
6. C
7. A
8. 1.57
9. C
10. A
11. 0.148
12. B

Del I

Uppgift 1

Vi väljer händelsen att modern har blodgrupp 0 som utfallsrum Ω, eftersom detta är givet som förutsättning för de redovisade sannolikheterna. Vi inför händelserna

$$\begin{aligned} A &= \{\text{barnet har blodgrupp A0}\}, \\ H_1 &= \{\text{fadern har blodgrupp A0}\}, \\ H_2 &= \{\text{fadern har blodgrupp AA}\}, \\ H_3 &= \{\text{fadern har blodgrupp AB}\}, \\ H_4 &= \{\text{fadern har en annan blodgrupp}\}, \end{aligned}$$

och vi söker $P(H_2|A)$. Enligt texten är

$$P(H_1) = 0.36, \quad P(H_2) = 0.08, \quad P(H_3) = 0.02, \quad P(H_4) = 0.54,$$

och

$$P(A \mid H_1) = 1/2, \quad P(A \mid H_2) = 1, \quad P(A \mid H_3) = 1/2 \text{ och } P(A \mid H_4) = 0.$$

Från Bayes sats och satsen om total sannolikhet får vi

$$\begin{aligned} P(H_2|A) &= \frac{P(A|H_2)P(H_2)}{P(A)} = \frac{P(A|H_2)P(H_2)}{\sum_{i=1}^4 P(A|H_i)P(H_i)} = \\ &= \frac{1 \cdot 0.08}{1/2 \cdot 0.36 + 1 \cdot 0.08 + 1/2 \cdot 0.02 + 0 \cdot 0.54} = \frac{0.08}{0.27} = \frac{8}{27} \approx 0.296. \end{aligned}$$

Svar: Sannolikheten att fadern har AA när barnet har $A0$ och modern 0 är 0.296.

Uppgift 2

Vi utför här ett test av oberoende. Som nollhypotes H_0 väljer vi att medeltemperaturerna i januari i Köpenhamn och Nuuk är oberoende, medan mothypotesen H_1 är att dessa temperaturer är beroende. Vi gör här en tabell med observerade antal enligt

	Sträng Nuuk	Normal Nuuk	Mild Nuuk	Totalt
Sträng Köpenhamn	2	19	9	30
Normal Köpenhamn	13	51	22	86
Mild Köpenhamn	12	18	2	32
Totalt	27	88	33	148

Låt x_{ij} beteckna antalet vintrar som faller inom kategori i i Köpenhamn och inom kategori j i Nuuk. Låt vidare n_i beteckna det totala antalet vintrar som faller inom kategori i i Köpenhamn och låt m_j beteckna det totala antalet vintrar som faller inom kategori j i Nuuk. Om vi låter N beteckna det totala antalet undersökta vintrar, så blir teststorheten

$$\begin{aligned} Q = \sum_{i=1}^{3}\sum_{j=1}^{3} \frac{(x_{ij} - \frac{n_i m_j}{N})^2}{\frac{n_i m_j}{N}} &= \frac{(2 - \frac{30\cdot 27}{148})^2}{\frac{30\cdot 27}{148}} + \frac{(19 - \frac{30\cdot 88}{148})^2}{\frac{30\cdot 88}{148}} + \frac{(9 - \frac{30\cdot 33}{148})^2}{\frac{30\cdot 33}{148}} \\ &+ \frac{(13 - \frac{86\cdot 27}{148})^2}{\frac{86\cdot 27}{148}} + \frac{(51 - \frac{86\cdot 88}{148})^2}{\frac{86\cdot 88}{148}} + \frac{(22 - \frac{86\cdot 33}{148})^2}{\frac{86\cdot 33}{148}} \\ &+ \frac{(12 - \frac{32\cdot 27}{148})^2}{\frac{32\cdot 27}{148}} + \frac{(18 - \frac{32\cdot 88}{148})^2}{\frac{32\cdot 88}{148}} + \frac{(2 - \frac{32\cdot 33}{148})^2}{\frac{32\cdot 33}{148}} = 14.21. \end{aligned}$$

Om H_0 är sann så är 14.21 ett utfall av en stokastisk variabel som är approximativt χ^2-fördelad med $(3-1)(3-1) = 4$ frihetsgrader. Approximationen är applicerbar eftersom $n_i m_j/N \geq 30 \cdot 27/148 = 5.47 > 5$. Eftersom $\chi^2_{0.01}(4) = 13.3 < 14.21$, så kan H_0 förkastas på signifikansnivån 1%. Vi drar följande slutsats.

Svar: På signifikansnivån 1% finns ett beroende mellan grönländska och danska vintertemperaturer.

Uppgift 3

itemize

Antalet personer som sades sig vilja rösta på något av de rödgröna partierna i de två undersökningarna är utfall av $X \in \mathrm{Bin}(4632, p_1)$ respektive $Y \in \mathrm{Bin}(4715, p_2)$, där p_1 och p_2 är dessa partiers verkliga andelar i maj 2018 respektive november 2017. Vi skattar dessa andelar med $(p_1)^*_{obs} = 1945/4632 \approx 0.4199$ respektive $(p_2)^*_{obs} = 2126/4715 \approx 0.4509$ och skattningen av förandringen är

$$(p_1)^*_{obs} - (p_2)^*_{obs} = 0.4199 - 0.4509 = -0.031.$$

Medelfelet är en skattning av standardavvikelsen för skattningen. Vi skattar $p_1 - p_2$ med $X/4632 - Y/4715$ som har variansen

$$\frac{V(X)}{4632^2} + \frac{V(Y)}{4715^2} = \frac{p_1(1-p_1)}{4632} + \frac{p_2(1-p_2)}{4715}.$$

Medelfelet får vi om vi ersätter de okända p_1 och p_2 med skattningarna $(p_1)^*_{obs} = 0.4199$ respektive $(p_2)^*_{obs} = 0.4509$. Vi erhåller då medelfelet

$$\begin{aligned} &\sqrt{\frac{(p_1)^*_{obs}(1-(p_1)^*_{obs})}{4632} + \frac{(p_2)^*_{obs}(1-(p_2)^*_{obs})}{4715}} \\ \approx\ &\sqrt{\frac{0.4199 \cdot (1-0.4199)}{4632} + \frac{0.4509 \cdot (1-0.4509)}{4715}} \approx 0.0103. \end{aligned}$$

Svar: Förändringen i stödet för de rödgröna partierna skattas till -0.031 och medelfelet för skattningen är 0.0103.

$X/4632$ och $Y/4715$ är approximativt normalfördelade $N(p_1, \sqrt{p_1(1-p_1)/4632})$ respektive $N(p_2, \sqrt{p_2(1-p_2$ där det är tillåtet att göra normalapproximation ty (t ex) $4632 p_1(1-p_1) \approx 4632 \cdot 0.4199(1 - 0.4199) \geq 10$ med mycket bred marginal. Vi vill beräkna ett (approximativt) 95%-igt konfidensintervall för $p_1 - p_2$ och får med approximativa metoden enligt FS 11.3 intervallet

$$\begin{aligned} &(p_1)^*_{obs} - (p_2)^*_{obs} \pm \lambda_{0.025}\sqrt{\frac{(p_1)^*_{obs}(1-(p_1)^*_{obs})}{4632} + \frac{(p_2)^*_{obs}(1-(p_2)^*_{obs})}{4715}} \\ \approx\ & 0.4199 - 0.4509 \pm 1.9600 \cdot 0.0103 = -0.031 \pm 0.0202 = (-0.0512, -0.0108). \end{aligned}$$

Svar: Ett konfidensintervall med approximativ konfidensgrad 95% för förändringen i stödet för de rödgröna partierna ges av $(-0.051, -0.011)$.

Vi testar nollhypotesen $H_0 : p_1 = p_2$ mot mothypotesen $H_1 : p_1 \neq p_2$ på signifikansnivån 5%. Eftersom konfidensintervallet i deluppgift (b) inte innehåller 0, så kan vi förkasta nollhypotesen. 1808-sos-exempeltenta.tex

Svar: Förändringen i stödet för de rödgröna partierna är statistiskt säkerställd.

Uppgift 4

a) Låt $X_1, X_2, \ldots$ vara oberoende $N(0, \sigma)$-fördelade stokastiska variabler och låt $Y_i = |X_i|$, $i = 1, 2, \ldots$. Beteckna med S^* den stickprovsvariabel som svarar mot s^*, dvs

$$S^* = \frac{1}{n}\sum_{i=1}^{n} Y_i^2.$$

Linjäritet av väntevärden ger att

$$E[S^*] = \frac{1}{n}\sum_{i=1}^{n} E[Y_i^2].$$

Vidare har vi

$$E[Y_i^2] = E[|X_i|^2] = E[X_i^2] = \sigma^2.$$

Det följer att

$$E[S^*] = \frac{1}{n}\sum_{i=1}^{n} \sigma^2 = \sigma^2,$$

så eftersom $E[S^*] = \sigma^2$, så är skattningen väntevärdesriktig.

Svar: Skattningen s^* är väntevärdesriktig.

b) Vi inleder med att härleda fördelningsfunktionen $F_Y(y) = P(Y \leq y)$. Från definitionen av Y_i står det klart att $F_Y(y) = 0$ för $y < 0$. För $y \geq 0$ fås

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(|X| \leq y) = P(-y \leq X \leq y) = P(\frac{-y}{\sigma} \leq \frac{X}{\sigma} \leq \frac{y}{\sigma}) \\ &= \Phi\left(\frac{y}{\sigma}\right) - \Phi\left(\frac{-y}{\sigma}\right). \end{aligned}$$

I det sista steget har vi utnyttjat att $X/\sigma \in N(0,1)$. Symmetri ger sedan $\Phi\left(\frac{-y}{\sigma}\right) = 1 - \Phi\left(\frac{y}{\sigma}\right)$, vilket insatt i ovanstående uttryck ger

$$F_Y(y) = 2\Phi\left(\frac{y}{\sigma}\right) - 1.$$

Derivering med avseende på y ger

$$f_Y(y) = \frac{d}{dy}F_Y(y) = 2\left[\Phi'(x)\right]_{x=\frac{y}{\sigma}} \times \frac{1}{\sigma} = \frac{2}{\sigma}\frac{1}{\sqrt{2\pi}}e^{-y^2/2\sigma^2}.$$

Svar: Täthetsfunktionen för Y ges av $f_Y(y) = \begin{cases} \sqrt{\frac{2}{\pi\sigma^2}}e^{-y^2/2\sigma^2}, & \text{då } y \geq 0, \\ 0, & \text{då } y < 0. \end{cases}$