Göteborgs Universitetet
GU Lärarprogrammet

Tentamen L9MA30, LGMA30

Matematik 3 för lärare, åk 7-9, Sannolikhetslära och statistik,
Matematik 3 för gymnasielärare, Sannolikhetslära och statistik

8 januari 2019, kl 08.30-12.30

Examinator: Serik Sagitov, tel. 031 772 5351
Jour: Anton Johansson, tel. 031 772 3558

Hjälpmedel: valfri miniräknare med tomt minne och ingen möjlighet att koppla upp sig till nätet, dubbelsidigt A4 ark med egna anteckningar samt med skrivningen utdelade tabellsidor och formelsamling.

Maxpoäng: 30. För godkänd krävs minst 14 p, för Väl Godkänd minst 22 p. Lösningar skall redovisas till varje uppgift. Uppgifterna är ej ordnade efter svårighetsgrad. Rättningsprotokoll anslås senast tre veckor efter tentamen.

1. (5 poäng) I långväga digitala dataöverföringar (t.ex. från en satellit) är risken för överföringsfel relativt hög. Antag i en specifik application att i överföringen av en bit, som är antingen en nolla eller en etta, felrisken är 2%. Vi vet att man sänder ungefär 40% nollor och 60% ettor.

(a) Låt X vara antal fel vid mottagning av 10 bit. Vad är sannolikheter av händelser

$$A_0 = \{X = 0\}, \quad A_1 = \{X = 1\}, \quad A_2 = \{X = 2\}?$$

Vilka antaganden utgår du ifrån?

(b) Låt X vara antal fel vid mottagning av 100 bit. Beräkna sannolikheten $\mathrm{P}(X \leq 2)$ med hjälp av en utdelad tabell.

(c) Hur stor är sannolikheten att en etta tagits emot om en okänd bit sändes?

(d) Givit att en etta tagits emot, hur stor är sannolikheten att det var en etta som sändes?

2. (5 poäng) I en urna finns 5 kulor: 3 vita och 2 svarta. Elin drar på måfå två kulor utan återläggning. För varje svart kula får Elin tre kronor och för varje vit kula förlorar hon två kronor. Låt X_1 (respektive X_2) vara Elins vinst/förlust i kronor för den första (respektive andra) dragna kulan. Elins totala vinst/förlust blir $Y = X_1 + X_2$.

(a) Rita en Venndiagram med 4 händelser

- V_1 första kulan är vit,
- S_1 första kulan är svart,
- V_2 andra kulan är vit,
- S_2 andra kulan är svart,

på så sätt att arean är nogurlunda proportionell sannolikheten.

(b) Vad är simultan fördelning av (X_1, X_2)?

(c) Vad är fördelningen av Y? Beräkna väntevärde av Elins totala vinst.

(d) En skolelev frågar sin lärare: "Varför är det så viktigt att räkna sannolikheter med urnor och kulor?" Hur skulle du svara på frågan?

3. (5 poäng) För en viss försäkring betalar man 0.6 tkr per år och får 10 tkr vid ett skadetillfälle. Låt det årliga antalet skador per kund, X, vara Poisson fördelat med parameter λ. Låt Y vara vinsten/förlusten som försäkringen gör från en viss kund per år.

(a) Vad är väntevärdet och standardavvikelsen av Y? Låt $\lambda = 0.01$. Hur stor är standardavvikelsen av Y jämfört med väntevärdet? Vilka slutsatser om riskerna för försäkrinsbolaget drar du från denna jämföresle?

(b) Hur stor kan parametern av Poisson fördelningen av de årliga skadetillfällena vara, så att det med 99% sannolikhet inte finns något skadetillfälle för en viss kund?

(c) Låt Y_n vara en årsvinst om försäkringsbolaget har n kunder. Vad är väntevärde och standardavvikelse av Y_n? Vilka antaganden utgår du ifrån?

(d) Om antal kunder är $n = 10000$ och det sker 0.01 skadetillfälle per kund och per år, vad är sannolikheten att försäkringen går i förlust under ett visst år? Ledning: hänvisa till centrala gränsvärdesatsen och tre-sigma regeln.

4. (5 poäng) Ett stickprov av storlek 16 har följande stickprovs medelvärde och varians:

$$\bar{x} = 3.8, \quad s^2 = 2.25.$$

(a) Vad exakt betyder påståendet att 3.8 är en väntevärdesriktig punktskattning av μ? Vad står här μ för?

(b) Vad exakt betyder påståendet att 2.25 är en väntevärdesriktig punktskattning av σ^2? Vilka antagande gäller här?

(c) Är $\sqrt{2.25} = 1.5$ en väntevärdesriktig punktskattning av σ? Förklara.

(d) Bilda ett motsvarande exakt 95% konfidensintervall för μ. Vilka antagande gäller här?

5. (5 poäng) År 2011 förbrukade svenskarna i genomsnitt 36.1 kg griskött per person och år. I dagens samhälle verkar det vara trendigt att äta vegetariskt, så vi är intresserade av att undersöka om detta ändrats. Därför observerade vi grisköttförbrukningen hos $n = 40$ personer under 2018 och observerade data som kan sammanfattas på följande vis:

$$x_1 + \ldots + x_n = 1350, \quad x_1^2 + \ldots + x_n^2 = 48523.$$

(a) Bilda ett 99% konfidensintervall för μ, den genomsnittliga grisköttförbrukningen per person och år för 2018 baserat på data.

(b) Antag att vi visste att det sanna värdet på standardavvikelsen $\sigma = 6$. Hur många hade vi minst behövt tillfråga för att få ett 99% konfidensintervall av mindre längd än intervallet vi fått fram i (a)?

(c) Baserad på data skulle du förkasta $H_0 : \mu = 36.1$? Vilken signifikansnivå har ditt statistiska test? Varför denna nollhypotes är intressant?

6. (5 poäng) Under den sista rubriken i formelsamlingen:

Exakta konfidensintervaller för väntevärdesskillnad

finns tre olika formler för konfidensinterval $I_{\mu_1-\mu_2}$. Förklara principiella skillnader mellan de tre formlerna.

Lycka till!

Lösningar

1a. Anta att felen uppstår oberoende vid överföring av olika bit. I så fall X kan betraktas som en slumpvariabel med $\mathrm{Bin}(n, p)$ fördelning där $n = 10$ och $p = 0.02$. Vi får

$$\begin{aligned}\mathrm{P}(X = 0) &= (0.98)^{10} = 0.817,\\ \mathrm{P}(X = 1) &= 10 \cdot 0.02 \cdot (0.98)^9 = 0.167,\\ \mathrm{P}(X = 2) &= \binom{10}{2}(0.02)^2(0.98)^8 = 0.015.\end{aligned}$$

1b. Vi kan använda Poisson approximation för binomial fördelning $X \approx \mathrm{Po}(\lambda)$ med $\lambda = 100 \cdot 0.02 = 2$. Enligt Tabell 2

$$\mathrm{P}(X \leq 2) = 0.677.$$

1c. Vi vill beräkna $\mathrm{P}(A|B)$, där

$$\begin{aligned}A &= \{1 \text{ sändes}\},\\ B &= \{1 \text{ tagits emot}\}.\end{aligned}$$

Vi vet att

$$\mathrm{P}(B|A) = 0.98, \quad \mathrm{P}(A) = 0.6, \quad \mathrm{P}(B|A^c) = 0.02, \quad \mathrm{P}(A^c) = 0.4,$$

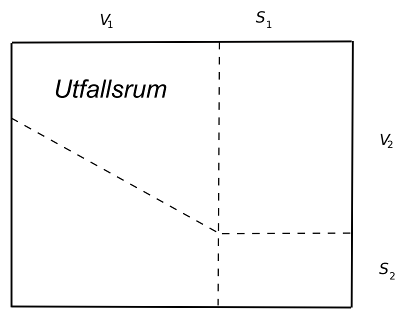
och därför

$$\mathrm{P}(B) = \mathrm{P}(B|A)\mathrm{P}(A) + \mathrm{P}(B|A^c)\mathrm{P}(A^c) = 0.98 \cdot 0.6 + 0.02 \cdot 0.4 = 0.596.$$

1d.

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(B|A)\mathrm{P}(A)}{\mathrm{P}(B)} = 0.987.$$

2a.



Venndiagram problem 2a

2b.

	$X_1 = -2$	$X_1 = 3$
$X_2 = -2$	0.3	0.3
$X_2 = 3$	0.3	0.1

2c. Slumpvariabeln $Y = X_1 + X_2$ tar en av tre värdena

$$y_1 = -4, \quad y_2 = 1, \quad y_3 = 6,$$

med sannolikheter

$$p_1 = 0.3, \quad p_1 = 0.6, \quad p_1 = 0.1.$$

Det innebär

$$\mathrm{E}(Y) = (-4) \cdot 0.3 + 1 \cdot 0.6 + 6 \cdot 0.1 = 0.$$

Samma svar får man med hjälp av additionsregel

$$\mathrm{E}(Y) = \mathrm{E}(X_1) + \mathrm{E}(X_2) = 0 + 0 = 0.$$

3a. Eftersom $Y = 0.6 - 10X$ och

$$\mathrm{E}(X) = \mathrm{Var}(X) = \lambda,$$

får vi

$$\mathrm{E}(Y) = 0.6 - 10\lambda, \quad \mathrm{Var}(Y) = 100\lambda, \quad \sigma_Y = 10\sqrt{\lambda}.$$

Om $\lambda = 0.01$, då

$$\mathrm{E}(Y) = 0.6 - 10 \cdot 0.01 = 0.5, \quad \sigma_Y = 10\sqrt{0.01} = 1.$$

Även om väntevärdet är positiv: $\mathrm{E}(Y) = 0.5$, risken är stor: $\sigma_Y = 1$, att hamna i minus om det är bara en kund.

3b. Enligt $\mathrm{Po}(\lambda)$ fordelnings fromel

$$\mathrm{P}(X = 0) = e^{-\lambda}.$$

Genom att lösa ekvationen

$$e^{-\lambda} = 0.99$$

får vi svaret till (b)

$$\lambda = -\ln 0.99 = 0.01.$$

3c. Anta att alla n kunder är oberoende av varandra och har samma risker att hamna i olyckshändelse. Då

$$\mathrm{E}(Y_n) = n\mathrm{E}(Y) = n(0.6 - 10\lambda), \quad \sigma_{Y_n} = 10\sqrt{n\lambda}.$$

3d. Om $\lambda = 0.01$ och $n = 10000$, då

$$\mathrm{E}(Y_n) = 10000 \cdot 0.5 = 5000, \quad \sigma_{Y_n} = 10\sqrt{100} = 100.$$

Enligt centrala gränsvärdesatsen Y_n är approximativt normalfördelad med $\mu = 5000$ och $\sigma = 100$. Det vill säga normalfördelningen har väntevärde som är 50 gånger större än standardavvikelse. Därför enligt tre-sigma regeln sannolikheten att hamna i minus är försvinnande liten.

4c. Stickprovs varians s^2 är väntevärdesriktig skattning av σ^2 eftersom

$$\mathrm{E}(S^2) = \sigma^2.$$

Stickprovs standardavvikelse s används som punktskattning av σ. Variansen

$$\mathrm{Var}(S) = \mathrm{E}(S^2) - (\mathrm{E}S)^2$$

är positiv, därför

$$(\mathrm{E}S)^2 < \mathrm{E}(S^2) = \sigma^2$$

och vi observerar att

$$\mathrm{E}S < \sigma.$$

Det betyder att s är inte väntevärdesriktig skattning av σ, eftersom s systematiskt underskattar σ.

4d. Om vi antar att populationsfördelning är normal, då 95% konfidensintervall beräknas som

$$I_\mu = \bar{x} \pm t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}} = 3.8 \pm 2.131\frac{1.5}{4} = 3.8 \pm 0.8.$$

5a. Vi räknar medelvärde

$$\bar{x} = \frac{1350}{40} = 33.75,$$

och stickprovs varians

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}(x_i^2) - n\bar{x}^2\right) = \frac{1}{39}\left(48523 - 40(33.75)^2\right) = 75.91.$$

Under antagandet om normalfördelning för grisköttsförbrukning beräknar vi 99% konfidensintervall som

$$I_\mu = \bar{x} \pm t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}} = 33.75 \pm 2.704\frac{8.71}{6.32} = 33.75 \pm 3.72.$$

5b. En motsvarande formel med känd $\sigma = 6$ är

$$I_\mu = \bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 33.75 \pm 2.58\frac{6}{\sqrt{n}}.$$

Svaret kommer från ekvationen

$$2.58\frac{6}{\sqrt{n}} = 3.72.$$

Vi får

$$n = (2.58\tfrac{6}{3.72})^2 = 17.32,$$

och det visar sig att minsta n är 18.

5c. Eftersom 36.1 ligger inom 33.75 ± 3.72 kan vi inte förkasta nollhypotesen $H_0 : \mu = 36.1$. Signifikansnivån av vårt test är 1%. Denna nollhypotes är relevant eftersom vår huvudfråga var om köttförbrukning i 2018 förändrades jämfört med 2011.