

Tentamen L9MA30, LGMA30

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

2017-10-20 kl. 08:30-12:30

Examinator: Claes Andersson, tel. 0734 031540 alt. 031 772 8294

Jour: Malin Palö Forsström, tel 031 772 8294

Hjälpmedel: Valfri miniräknare (tomt minne och ingen möjlighet att koppla upp sig till nätet; kontrolleras av examinatoren), (dubbelsidigt) A4 ark med egna anteckningar samt med skrivningen utdelade tabellsidor och formelsamling.

Maxpoäng: 30. För godkänd krävs minst 15 p, för Väl Godkänd 24 p. Eventuell bonuspoäng från 2017 adderas. Lösningar skall redovisas till varje uppgift. Uppgifterna är ej ordnade efter svårighetsgrad. Rättningsprotokoll anslås senast tre veckor efter tentamen.

1. Antag att B_1 och B_2 är två disjunkta händelser, och A en händelse som är oberoende av både B_1 och B_2 . Visa att A är oberoende av $B_1 \cup B_2$.

Ledning: Använd att $A \cap (B_1 \cup B_2) = (A \cap B_1) \cup (A \cap B_2)$.

2p

Lösning: A är oberoende av $B_1 \cup B_2$ om $\mathbb{P}(A \cap (B_1 \cup B_2)) = \mathbb{P}(A)\mathbb{P}(B_1 \cup B_2)$ (0.5 p).

$$\begin{aligned}\mathbb{P}(A \cap (B_1 \cup B_2)) &\stackrel{\text{ledn.}}{=} \mathbb{P}((A \cap B_1) \cup (A \cap B_2)) \stackrel{\text{disjunkta}(0.5p)}{=} \mathbb{P}(A \cap B_1) + \mathbb{P}(A \cap B_2) \stackrel{\text{oberoende}(0.5p)}{=} \\ &\mathbb{P}(A)\mathbb{P}(B_1) + \mathbb{P}(A)\mathbb{P}(B_2) = \mathbb{P}(A)(\mathbb{P}(B_1) + \mathbb{P}(B_2)) \stackrel{\text{disjunkta}(0.5p)}{=} \mathbb{P}(A)\mathbb{P}(B_1 \cup B_2)\end{aligned}$$

2. Far föder upp får. Han har två baggar, en svart och en vit. Vi vet att lamm till en vit bagge blir svarta med sannolikhet 0.4, medan lamm till en svart bagge blir svarta med sannolikhet 0.8, oberoende av varandra inom en kull. En av tackorna blir dräktig utan fars vetskap, och far bedömer att det är lika sannolikt att den svarta baggen är far till kullen som den vita.

a) Vad är sannolikheten att 3 av totalt 5 lamm till en svart bagge är svarta?

1p

b) När kullen föds är just 3 av totalt 5 lamm svarta. Vad är sannolikheten att den svarta baggen är far till lammen?

2p

Lösning: Om X = "antal svarta får i en kull" så gäller $X \sim \text{Bin}(n, p)$ (0.5 p till a)), där n är storleken på kullen och p beror på vilken bagge som är far till lammen.

- a) Om den svarta baggen är far är $p = 0.8$, så

$$\mathbb{P}(X = 3) = \binom{5}{3} 0.8^3 0.2^2 = 0.2048 \quad (0.5p)$$

- a) Om vi låter S = "den svarta baggen är far till kullen" har vi fått att $\mathbb{P}(S) = 1/2$ och vi har att

$$\mathbb{P}(S|X = 3) \stackrel{\text{Bayes}(1p)}{=} \frac{\mathbb{P}(X = 3|S)\mathbb{P}(S)}{\mathbb{P}(X = 3|S)\mathbb{P}(S) + \mathbb{P}(X = 3|S^c)\mathbb{P}(S^c)},$$

där $\mathbb{P}(X = 3|S)$ är beräknad i a), $\mathbb{P}(X = 3|S^c) = \binom{5}{3} 0.4^3 0.6^2 = 0.2304$, vilket ger att

$$\mathbb{P}(S|X = 3) = \frac{0.5 \cdot 0.2048}{0.5 \cdot 0.2048 + 0.5 \cdot 0.2304} = 0.47 \quad (1p),$$

dvs trots att fler lamm var svarta än vita är det mer sannolikt att den vita baggen är far.

3. I en påse har vi 26 brickor med bokstäverna A-Z på (21 vokaler, 6 konsonanter). Vi drar fem brickor utan återläggning. Beräkna sannolikheterna för följande händelser:

- a) A är inte med bland de dragna brickorna.

1 p

- b) Med bokstäverna vi dragit kan vi stava till "LEGO" (och få en bricka över).

2 p

Lösning:

- a) $\frac{25}{26} \frac{24}{25} \frac{23}{24} \frac{22}{23} \frac{21}{22} = \frac{21}{26} = 0.80769$. (1p)

- b) X = "Antalet bokstäver av L, E, G och O vi får upp" $\sim \text{Hyp}(26, 5, 4/26)$ (1p). För att kunna bilda LEGO måste $X = 4$.

$$\mathbb{P}(X = 4) = \frac{\binom{4}{4} \binom{22}{1}}{\binom{26}{5}} \approx 3.34 \cdot 10^{-4} \quad (1p)$$

4. Claes sadlar om och blir fiskare. Varje morgon åker han ut och lägger nät med vilka han fångar torsk och sill. Mängden torsk han fångar är $\mathcal{N}(50, 10)$ -fördelad och mängden sill är $\mathcal{N}(100, 15)$ -fördelad (i kg). Vid dagens slut säljer han fisken han fångat, och får 50 kr/kg för torsk och 20 kr/kg för sill. Dessutom måste han betala en daglig båthyra på 4000 kr.

- a) Vad är den förväntade vikten av Claes totala fångst på en dag?

1p

- b) Vad Claes förväntade vinst per dag?

2p

- c) Vad är sannolikheten att Claes går med vinst en given dag?

2p

Lösning:

- a) Låt T = "mängden fångad torsk" $\sim \mathcal{N}(50, 10)$ och S = "mängden fångad sill" $\sim \mathcal{N}(100, 15)$, samt F = "totala fångsten" $= T + S$. Då är

$$\mathbb{E}[F] = \mathbb{E}[T + S] = \mathbb{E}[T] + \mathbb{E}[S] = 50 + 100 = 150 \quad (1p)$$

- b) Claes *inkomst* är $50T + 20S$, och vinsten skrivs som $V = 50T + 20S - 4000$ (1p). Den förväntade vinsten är således

$$\mathbb{E}[V] = 50\mathbb{E}[T] + 20\mathbb{E}[S] - 4000 = 500 \quad (1p)$$

- c) Eftersom vinsten är en linjärkombination av $\mathcal{N}$ -fördelade slumpvariabler gäller att

$$V \sim \mathcal{N}(\mu_V, \sigma_V),$$

där $\mu_V = 500$ och $\sigma_V = \sqrt{50^2 \cdot 10^2 + 20^2 \cdot 15^2} = 583.1$ (1p). Vi söker

$$\mathbb{P}(V > 0) = \mathbb{P}\left(\frac{V - 500}{583.1} > \frac{-500}{583.1}\right) = \mathbb{P}(Z > -0.86) = \mathbb{P}(Z \leq 0.86) = 0.8051 \quad (1p)$$

där $Z \sim \mathcal{N}(0, 1)$ och värdet fås ur tabell.

5. Visa att för godtyckliga slumpvariabler X och Y gäller

$$V(X + Y) = V(X) + V(Y) + 2C(X, Y)$$

Ledning: Använd att $V(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$ samt $C(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$.

3p

Lösning:

$$\begin{aligned} V(X + Y) &= \mathbb{E}[(X + Y)^2] - (\mathbb{E}[X + Y])^2 \stackrel{(1p)}{=} \\ &\quad \mathbb{E}[X^2] + \mathbb{E}[Y^2] + 2\mathbb{E}[XY] - \mu_X^2 - \mu_Y^2 - 2\mu_X\mu_Y \stackrel{(1p)}{=} \\ &\quad (\mathbb{E}[X^2] - \mu_X^2) + (\mathbb{E}[Y^2] - \mu_Y^2) + 2(\mathbb{E}[XY] - \mu_X\mu_Y) \stackrel{(1p)}{=} \\ &\quad V(X) + V(Y) + 2C(X, Y) \end{aligned}$$

6. En tvådimensionell diskret slumpvariabel har simultan (gemensam) sannolikhetsfunktion

$$p_{XY}(j, k) = \begin{cases} C(kj + j) & 0 \leq j \leq 2, \quad 0 \leq k \leq 3 \\ 0 & \text{annars} \end{cases}$$

a) Bestäm konstanten C så att p_{XY} verkligen är en sannolikhetsfunktion.

1p

b) Beräkna $p_X(j)$, $p_Y(k)$ och $\mathbb{E}[X]$.

2p

c) Är X och Y oberoende?

1p

Lösning:

a) För att det ska vara en sannolikhetsfunktion måste den summera till 1 (0.5p). Vi sätter

$$1 = \sum_{j=0}^2 \sum_{k=0}^3 C(kj + j) = C \sum_{j=0}^2 j \sum_{k=0}^3 (k+1) = C \sum_{j=0}^2 10j = 30C \Rightarrow C = 1/30 \quad (0.5p)$$

b) För att beräkna de marginella sannolikhetsfunktionerna summerar vi enligt

$$p_X(j) \stackrel{0.5p}{=} \sum_k p_{XY}(j, k) = \frac{1}{30} \sum_{k=0}^3 j(k+1) \stackrel{0.5p}{=} \frac{j}{3}$$

$$p_Y(k) = \sum_j p_{XY}(j, k) = \frac{1}{30} \sum_{j=0}^2 j(k+1) = \frac{k+1}{10}$$

Väntevärdet av X ges av

$$E[X] = \sum_j j p_X(j) = \frac{1}{3} \sum_{j=0}^2 j^2 = \frac{5}{3} \quad (1p)$$

c) Ja, ty

$$p_{XY}(j, k) = \frac{kj + j}{30} = \frac{j(k+1)}{30} = \frac{j}{3} \frac{k+1}{10} = p_X(j) p_Y(k) \quad (1p)$$

för $0 \leq j \leq 2$ och $0 \leq k \leq 3$. Både marginalerna och den gemensamma är 0 annars. Så för alla j, k gäller $p_{XY}(j, k) = p_X(j) p_Y(k)$.

7. År 2011 förbrukade svenskarna i genomsnitt 36,1 kg griskött per person och år¹. I dagens samhälle verkar det vara trendigt att äta vegetariskt², så vi är intresserade av att undersöka om detta ändrats. Därför observerade vi grisköttförbrukningen hos 41 personer under 2016 och observerade data som kan sammanfattas på följande vis:

$$\sum_i x_i = 1415, \quad \sum_i x_i^2 = 49947$$

- a) Bilda ett 90% konfidensintervall för den genomsnittliga grisköttförbrukningen per person och år för 2016 baserat på data.

2p

- b) Antag att vi visste att det sanna värdet på standardavvikelsen, σ , var 6. hur många hade vi minst behövt tillfråga för att få ett 99% konfidensintervall av mindre längd än intervallet vi fått fram i a)?

2p

Lösning:

- a) Det som efterfrågas är ett 90% KI för μ . Från datan vi fått kan vi beräkna

$$\bar{x} = 1415/41 = 34.5122, \quad \text{och } s^2 = \frac{1}{40} \left(\sum x_i^2 - n(\bar{x})^2 \right) = 27.8061, \quad (0.5p)$$

där $n = 41$. Vi vet inte den sanna variansen, utan skattar den med s^2 . Därför ska vi använda

$$I_\mu = \bar{x} \pm t(n-1)_{\alpha/2} \frac{s}{\sqrt{n}} \quad (0.5p),$$

där $t(n-1)_{\alpha/2} = t(40)_{0.05} = 1.684$ vilket ger

$$I_\mu = 34.51 \pm 1.683 \cdot \frac{\sqrt{27.81}}{\sqrt{41}} = 34.51 \pm 1.387 = [33.13, 35.90] \quad (1p)$$

- a) b) Med känt σ ges konfidensintervallet av

$$I_\mu = \bar{x} \pm \lambda_{\alpha/2} \frac{\sigma}{\sqrt{n^*}} \quad (1p),$$

där $\lambda_{\alpha/2} = \lambda_{0.005} = 2.576$, eftersom vi nu var intresserade av ett 99% KI, och n^* är antalet vi ska tillfråga i denna tänkta undersökning. För att detta ska bli kortare än det vi fann i a) så måste

$$\lambda_{0.005} \frac{6}{\sqrt{n^*}} \leq t(40)_{0.05} \frac{s}{\sqrt{n}} \Leftrightarrow n^* \geq \left(\frac{6\lambda_{0.005}}{t(40)_{0.05} \frac{s}{\sqrt{n}}} \right)^2 = \left(\frac{6 \cdot 2.576}{1.387} \right)^2 = 124.2 \quad (1p),$$

dvs vi hade behövt fråga minst 125 personer.

¹Källa: <http://www.jordbruksverket.se>

²Källa: Majorna.

8. I Sverige genomförs varje månad opinionsundersökningar av olika företag och institut för att undersöka hur människor skulle rösta i riksdagsvalet om det var val idag. I Figur 1 ses ett litet utdrag av olika resultat³.

År	Månad	Företag	M	L (Ff)	C	KD	S	V	MP	SD	Övriga	Intervjuer
2017	September	Aftonbladet	16,5	3,6	12,3	4,7	28,8	6,5	3,5	19,9	4,2	2 066
2017	September	Demoskop	17,3	5,2	11,2	2,8	30,3	8	4,6	18,1	2,5	1 253
2017	September	Novus	16,3	5,4	12	3,9	29,7	6,8	4,8	18,1	3	4 001
2017	Augusti	Ipsos	17	6	11	3	29	8	5	19	2	2 004
2017	Augusti	Sifo	16,2	6,2	12,4	3,6	30	6,4	4,1	18,3	3	10 625
2017	Augusti	Sentio	15,1	4,1	10,5	3,2	23,5	8,6	4,3	27,4	3,3	1 000
2017	Augusti	Yougov	16,1	4,6	11,5	2,9	25,4	8,2	4,3	23,6	3,5	1 526
2017	Augusti	Aftonbladet	16	4,1	12,9	4,8	28,7	6,8	2,6	20,1	3,9	2 037
2017	Augusti	Novus	15,2	5,5	12,8	4,3	29,3	7,6	3,9	18,7	2,7	2 000
2017	Augusti	Demoskop	17,2	6,4	12,9	4	28,4	7,4	4,4	16,6	2,7	1 250

Figure 1: Redovisning av olika väljarbarometrar Augusti-September 2017. Tabellen är beskuren.

Som man kan utläsa i tabellen ger olika undersökningar lite olika resultat, och om skillnaderna är alltför stora kan man fundera över om företagen verkligen gör sitt intervjuurval på likvärdiga sätt. Man kan notera att resultatet för vänsterpartiet (V) skiljer sig ganska markant mellan Aftonbladets och Demoskops undersökningar i September 2017 (6,5% respektive 8%) där man tillfrågat 2066 respektive 1253 personer.

- a) Bilda ett 95% konfidensintervall för skillnaden i andelen som skulle rösta på V i Aftonbladets urvalsgrupp och Demoskops urvalsgrupp. (Vi kan betrakta båda grupperna som oändliga).

2p

- b) Kan vi vid signifikansnivå $\alpha = .05$ fastslå att andelen som skulle rösta på V i de två urvalsgrupperna är olika?

1p

Lösning:

- a) Låt p_1 vara andelen som röstar på V i Aftonbladets urvalsgrupp och p_2 motsvarande i Demoskops urvalsgrupp⁴. Det som söks är ett 95% KI för $p_1 - p_2$. Detta bildas genom

$$I_{p_1-p_2} = \hat{p}_1 - \hat{p}_2 \pm \lambda_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \quad (1p),$$

där $\lambda_{\alpha/2} = \lambda_{0.025} = 1.96$. Från datan vi fått har vi att

$$\begin{aligned} \hat{p}_1 &= 0.065 & \hat{p}_2 &= 0.08 \\ n_1 &= 2066 & n_2 &= 1253 \end{aligned}$$

Detta ger oss

$$I_{p_1-p_2} = -0.015 \pm 1.96 \sqrt{\frac{0.065 \cdot 0.935}{2066} + \frac{0.08 \cdot 0.92}{1253}} = [-0.0334, 0.0034] \quad (1p)$$

- b) Vi vill testa

$$H_0 : p_1 = p_2 \text{ mot } H_1 : p_1 \neq p_2,$$

och ska då förkasta H_0 om 0 är utanför vårt framräknade KI. Vi ser från lösningen i a) att 0 är i vårt KI, så vi kan inte förkasta H_0 . (1p)

³Källa: www.novus.se

⁴Givetvis bör dessa urvalsgrupperna vara samma, dvs rösberättigade svenska medborgare, men om urvalmetoderna är olika kan vi tänka på det som att grupperna är olika.

9. Egon och Lisa har två stickprov, $X_1, \dots, X_n$ respektive $Y_1, \dots, Y_n$, av samma storlek från en fördelning med väntevärde μ och varians σ^2 , och är intresserade av att skatta μ . Egon föreslår att de bara ska ta hans stickprov och använda $\hat{\mu} = \bar{X}$ som skattare. Lisa, däremot, föreslår att dom ska mötas på hälften och istället använda $\hat{\mu} = \frac{\bar{X} + \bar{Y}}{2}$.

a) Visa att Lisas förslag också ger en väntevärdesriktig skattare för μ .

1 p

b) Vilken skattare skulle du säga är bäst? Motivera.

2 p

Lösning:

a) Om Lisas skattare, som vi kan beteckna $\hat{\mu}_L$, är väntevärdesriktig ska vi ha $\mathbb{E}[\hat{\mu}_L] = \mu$.

$$\mathbb{E}[\hat{\mu}_L] = \mathbb{E}\left[\frac{\bar{X} + \bar{Y}}{2}\right] \stackrel{0.5p}{=} \frac{\mathbb{E}[\bar{X}] + \mathbb{E}[\bar{Y}]}{2} \stackrel{0.5p}{=} \frac{\mu + \mu}{2} = \mu,$$

eftersom vi vet att $\mathbb{E}[\bar{X}] = \mu$.

b) Vi vet nu att båda skattarna är väntevärdesriktiga, så för att avgöra vilken som är bäst får vi titta på vilken som är effektivast, dvs vilken som har minst varians (0.5 p). Låt oss betrakta variansen för Lisas skattare:

$$V(\hat{\mu}_L) = V\left(\frac{\bar{X} + \bar{Y}}{2}\right) \stackrel{regler}{=} \frac{1}{4}V(\bar{X} + \bar{Y}) \stackrel{oberoende}{=} \frac{1}{4}(V(\bar{X}) + V(\bar{Y})) = \frac{1}{2}V(\bar{X}) = \frac{1}{2}V(\hat{\mu}_E) \quad (1p),$$

eftersom $V(\bar{X}) = V(\bar{Y}) = \frac{\sigma^2}{n}$ och där $\hat{\mu}_E$ är Egons skattare. Lisas skattare har hälften så stor varians som Egons, och är därmed bättre (0.5p). Egon borde lära sig att man inte ska kasta bort data på det där viset.

Lycka till!