Matematisk statistik TMS063
Tentamen 2018-08-21

Tid: 14:00-18:00 **Tentamensplats:** SB
Hjälpmedel: Bifogad formelsamling och tabell samt Chalmersgodkänd räknare.
Kursansvarig: Olof Elias
Telefonvakt/jour: Olof Elias, 0762026293. Till salen ca 15.00 och 17.00
Betygsgränser: För betyg **3**, **4** resp. **5** krävs minst **12**, **18** resp. **24** poäng.
För att bli godkänd på kursen krävs godkänt på tentan **och** godkänt resultat på båda flervariabel-duggorna.
Till varje uppgift skall fullständig lösning lämnas!

---

1. I en urna ligger 13 bollar varav 7 st är röda och 6 st är blå. Antag att vi drar 5 bollar utan återläggning. Beräkna:
   (a) Antalet sätt kan man dra 5 bollar? **2 poäng**
   (b) Sannolikheten att vi drar 0 röda bollar samt sannolikheten att vi drar 2 röda bollar. **1+1 poäng**
   (c) Sannolikheten att vi drar k röda bollar, där k är något positivt heltal. **2 poäng**

   Lösning:
   Antalet sätt man kan dra 5 bollar på ges av binomialkoefficienten:
   $$N = \binom{13}{5}.$$
   De sökta sannolikheten ges av
   $$\frac{\text{Antalet sätt man kan dra 0 röda av 13}}{\text{Antalet sätt man kan dra 5 bollar av 13}} = \frac{m_0}{N}$$
   och
   $$\frac{\text{Antalet sätt man kan dra 2 röda av 13}}{\text{Antalet sätt man kan dra 5 bollar av 13}} = \frac{m_2}{N}.$$
   så vi behöver helt enkelt bara beräkna m_0 och m_2. Om vi drar 0 röda bollar så innebär det helt att vi drar 5 blå bollar vilket ger att
   $$m_0 = \binom{6}{5}.$$
   Den andra kvantiteten kan beräknas av följande argument. Antalet sätt att dra 2 röda bollar av 7 möjliga ges av $\binom{7}{2}$. De resterande 3 bollarna skall vara blå vilket innebär att dessa kan dras på $\binom{6}{3}$ sätt. Totala antalet sätt man kan dra 2 blå ges av att multiplicera ihop dessa kvantiteter:
   $$m_2 = \binom{7}{2}\binom{6}{3}.$$
   Det generlla uttrycket ges av
   $$\binom{7}{k}\binom{6}{5-k}/N,\ \ k = 0, 1, 2, 3, 4, 5$$

2. Lasse Kongo har bestämt sig för att bli basketproffs. För att göra detta börjar med han med att öva på straffar. Om man antar varje försök är oberoende och lyckas (dvs. gör mål) med en sannolikhet $p = 1/20$.

   (a) Vad är sannolikheten att hans första mål kommer först efter 5 kast? **1 poäng**

   (b) Hur många kast förväntar man sig att Lasse måste göra tills att han lyckas för första gången? **1 poäng**

   (c) Givet att Lasse har kastat mer än 10 kast, vad är sannolikheten att han måste kasta mer än 10 kast *till* för att träffa korgen? **2 poäng**

   Lösning: Låt X vara antalet försök som krävs för att Lasse Kongo ska göra mål för första gången. Då är $X \sim \text{Geo}(p)$ med $p = 1/20$ vilket ger täthet och fördelningsfunktion:

$$P(X = x) = f(x) = (1-p)^{x-1}p, x = 1, 2, ... \tag{1}$$
$$P(X \leq x) = F(x) = 1 - (1-p)^x, x = 1, 2, ... \tag{2}$$

   vilket ger att den sökta sannolikheten ges av

$$P(X \geq 5) = 1 - P(X \leq 4) = 1 - (1 - (1-p)^4) = (1-p)^4 = (19/20)^4.$$

   Väntevärdet ges av
$$E[X] = 1/p = 20.$$

   Slutligen så får vi att

$$P(X > 20|X > 10) = \frac{P(X > 20, X > 10)}{P(X \geq 10)} = \frac{P(X > 20)}{P(X > 10)}$$

   vilket ger

$$\frac{P(X > 20)}{P(X > 10)} = \frac{1 - (1 - (1-p)^{20})}{1 - (1 - (1-p)^{10})} = (1-p)^{10} = P(X > 10).$$

3. Låt
$$f(x, y) = cxy,\ 0 \leq x \leq y \leq 1,$$
   där $c > 0$ är en positiv konstant.

   (a) Bestäm c sådan att f är en täthet för en stokastisk vektor $Z = (X, Y)$. **2 poäng**

   (b) Beräkna marginaltätheterna och motsvarande fördelningsfunktioner för Z. **2 poäng**

   (c) Beräkna kovariansen mellan X och Y. **2 poäng**

   Lösning:
   Konstanten bestäms av

$$\int_0^1 \int_0^y cxy\mathrm{d}x\mathrm{d}y = 1 \Leftrightarrow \frac{1}{c} = \int_0^1 \int_0^y xy\mathrm{d}x\mathrm{d}y,$$

vilket ger

$$\int_0^1 \int_0^y xy\mathrm{d}x\mathrm{d}y = \int_0^1 y\left[\frac{x^2}{2}\right]_0^y \mathrm{d}y = \int_0^1 \frac{y^3}{2}\mathrm{d}y = \left[\frac{y^4}{8}\right]_0^1 = \frac{1}{8} \Rightarrow c = 8.$$

Marginaltätheterna ges av

$$f_X(x) = \int f(x,y)\mathrm{d}y = 8x\int_x^1 y\mathrm{d}y = 8x\left[\frac{y^2}{2}\right]_x^1 = 4x\left(1-x^2\right),\ 0 \le x \le 1$$
$$f_Y(y) = \int f(x,y)\mathrm{d}x = 8y\int_0^x x\mathrm{d}x = 8\frac{y^3}{2} = 4y^3, 0 \le y \le 1$$

vilket ger fördelningsfunktioner

$$F_X(x) = \int_0^x 4t\left(1-t^2\right)\mathrm{d}t = x^2(2-x^2)$$
$$F_Y(y) = \int_0^y 4t^3\mathrm{d}t = t^4.$$

Kovariansen ges av

$$\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y].$$

Vi börjar med att beräkna

$$E[XY] = \int_0^1 \int_0^y 8x^2y^2\mathrm{d}x\mathrm{d}y = 8\int_0^1 y^2\left[\frac{x^3}{3}\right]_0^y \mathrm{d}y = \frac{8}{3}\int_0^1 y^5\mathrm{d}5 = \frac{2}{9}.$$

Vidare så får vi

$$E[X] = \int_0^1 x4x\left(1-x^2\right)\mathrm{d}x = \frac{8}{15}$$

och

$$E[Y] = \int_0^1 y4y^3\mathrm{d}y = \frac{4}{5}$$

vilket ger kovarians

$$\mathrm{Cov}(X,Y) = \frac{2}{9} - \frac{8}{15}\frac{4}{5} = \frac{-46}{225}.$$

4. Låt $x = (x_1, ..., x_n)$ vara ett observerat stickprov från $N(1,\sigma^2)$. Beräkna Maximum Likelihood-skattaren av σ^2. **4 poäng**
Lösning :
Tätheten ges av

$$f(x|\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(x-1)^2\right\}$$

vilket ger likelihood funktion

$$L(\sigma^2|x) = \prod_{i=1}^n f(x_i|\sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-1)^2\right]$$

och log-likelihood funktionen ges därav

$$l(\sigma^2|x) = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-1)^2 - \frac{n}{2}\log(2\pi\sigma^2).$$

För enkelhetens skull sätt $t = \sigma^2$ och derivera med avseende på t.
(Man kan derivera med avseende på σ och sen kvadrera men det är enklare att derivara m.a.p. σ^2).

$$\begin{aligned} l'(t|x) = \frac{\mathrm{d}}{\mathrm{d}t}l(t|x) &= \frac{\mathrm{d}}{\mathrm{d}t}\left(-\frac{1}{2t}\sum_{i=1}^{n}(x_i-1)^2 - \frac{n}{2}\log(2\pi t)\right) \\ &= \frac{1}{2t^2}\sum_{i=1}^{n}(x_i-1)^2 - \frac{n}{2t} = \frac{1}{2t}\left(\frac{1}{t}\sum_{i=1}^{n}(x_i-1)^2 - n\right). \end{aligned}$$

Derivatan byter tecken vid

$$t = \frac{1}{n}\sum_{i=1}^{n}(x_i-1)^2$$

vilket även är maximat för $l(t|x)$ vilket ger att ML-skattaren för σ^2 ges av

$$(\sigma^*)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-1)^2.$$

5. Låt $x = (x_1, x_2, ..., x_n)$ vara ett observerat stickprov från $N(\mu, 4)$, med $\bar{x} = 1.37,\ n = 20$. Testa hypotesen

$$\begin{aligned} H_0 &: \mu \leq 1.1 \\ H_1 &: \mu > 1.1 \end{aligned}$$

med en signifikansnivå $\alpha = 0.05$. **(4 poäng)**
Lösning: Låt $\alpha = 0.05$ och låt $X = (X_1, ..., X_n)$ vara ett stick prov från $N(\mu, 4)$, då gäller att

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

Vidare, låt

$$z_{\text{obs}} = \frac{\bar{x}-1.1}{2/\sqrt{20}} = 0.6$$

vara den motsvarande observerade kvantiteten. Eftersom $H_0 : \mu \leq 1.1$ så ges p-värdet av

$$p = P(Z > z_{\text{obs}}) = P(Z > 0.6) = 1 - 0.73 = 0.27.$$

Då $p > \alpha$ så förkastar vi inte H_0.

6. Bill Skarsgård vill undersöka om han har blivit mer poppis bland allmänheten efter filmen It. För att testa detta väljer Bill att fråga 100 slumpmässigt utvalda personer på stan före och 100 personer efter premiären av filmen. Antalet personer som visste vem han var före premiären fick han till 45 personer. Antalet personer som visste vem han var efter premiären uppgick till 50. Bill antar att hans båda stickprov är oberoende. Hjälp Bill att undersöka om han faktiskt blivit mer populär genom att:

(a) Ställa upp lämplig hypotes och test-statistika **3 poäng**

(b) Beräkna det approximativa p-värdet med hjälp av centrala gränsvärdessatsen och förkasta med en signifikansnivå $\alpha = 0.01$. **3 poäng**

Lösning:

Lämplig hypotes ges av

$$\begin{aligned} H_0 &: p_{\text{efter}} \leq p_{\text{före}} \\ H_1 &: p_{\text{efter}} > p_{\text{före}} \end{aligned}$$

och test-statistikan ges av

$$Z = \frac{\hat{p}_f - \hat{p}_e}{\sqrt{\frac{1}{100}\left(\hat{p}_f(1-\hat{p}_f) + \hat{p}_e(1-\hat{p}_e)\right)}}$$

Om $\hat{p}_f$ och $\hat{p}_e$ är skattarna för våra två proportioner så säger centrala gränsvärdessatsen att

$$Z = \frac{\hat{p}_f - \hat{p}_e}{\sqrt{\frac{1}{100}\left(\hat{p}_f(1-\hat{p}_f) + \hat{p}_e(1-\hat{p}_e)\right)}},$$

är approximativt normalfördelad med väntevärde 0 och varians 1, d.v.s. $Z \sim N(0,1)$. Därmed kan vi göra samma test för differensen $\hat{p}_f - \hat{p}_e$ som vi gör för ett normalfördelat stickprov.

Låt

$$z_{obs} = \frac{0.46 - 0.50}{\sqrt{\frac{1}{100}\left(0.46 \cdot 0.54 + 0.5 \cdot 0.5\right)}} \approx -10.05$$

vara den observerade kvantiteten. Våran hypotes är ensidig och vi förkastar H_0 och vi ser ett stort och negativt värde, dvs. vi förkastar H_0 om sannolikheten

$$P(Z < z_{obs})$$

är liten. Går vi till Tabel III ser vi att det minsta värdet vi kan få tag i är

$$P(Z < -3.99) = 0.000033.$$

Eftersom

$$P(Z < -10) < P(Z < -3.99) = 0.00033 < 0.01$$

så förkastar vi H_0 och accepterar hypotesen att Bill har blivit mer poppis efter premiären av It.