Matematisk statistik TMS063/TMS064
Tentamen 2019-10-11

Tid: 8:30-12:30 **Tentamensplats:** M
Hjälpmedel: Bifogad formelsamling och tabell samt Chalmersgodkänd räknare.
Kursansvarig: Olof Elias
Telefonvakt/jour: Olof Zetterqvist, 031-7725325. Till salen ca 9.30 och 11.30
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 flervariabelduggorna.
Till varje uppgift skall fullständig lösning lämnas!

---

1. Antag att A, B, C är händelser som uppfyller

$$P(A) = P(B) = P(C) = 0.2$$

   (a) Antag att A, B, C är disjunkta. Beräkna sannolikheten att någon av händelserna inträffar. **2 poäng**

   (b) Antag nu istället att A, B, C är oberoende. Beräkna sannolikheten att exakt 1 av händelserna inträffar. **2 poäng**

   Lösning (a)
   För (a) så observerar man att

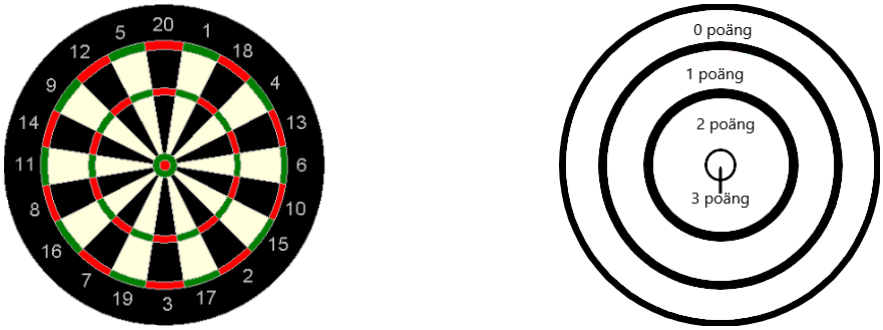
   $P($ Någon av händelserna inträffar $) = P(A \cup B \cup C) = P(A) + P(B) + P(C) = 3 \cdot 0.2 = 0.6$

   (b)
   För (b) så observerar man att

   $P($ Exakt 1 av händelserna inträffar $) = P(A \cap B^c \cap C^c) + P(A^c \cap B \cap C^c) + P(A^c \cap B^c \cap C) = 3 \cdot (0.2 \cdot 0.8 \cdot 0.8)$

2. Benjamin spelar en förenklad form av dart där poängen han får beror enbart på hur nära mitten han kommer. Antalet poäng per kast ges av följande:

   - 0 poäng om han träffar utanför den yttre röd-gröna cirkelskivan.
   - 1 poäng om han träffar den yttre cirkelskivan, d.v.s, området mellan de två yttre röd-gröna ringarna.
   - 2 poäng om han träffar den inre cirkel skivan
   - 3 poäng om han träffar den innersta cirkeln



Figur 1: En dart tavla och motsvarande poängområden.

För enkelhetens skull antar vi att darttavlan kan beskrivas med enhetscirkeln in planet:

$$C = \left\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\right\}$$

och att de fyra områdena kan beskrivas enligt

$$C_0 = \left\{(x, y) \in \mathbb{R}^2 : \frac{2}{3} \leq x^2 + y^2 \leq 1\right\},\ C_1 = \left\{(x, y) \in \mathbb{R}^2 : \frac{1}{3} \leq x^2 + y^2 < \frac{2}{3}\right\},$$
$$C_2 = \left\{(x, y) \in \mathbb{R}^2 : \frac{1}{4} \leq x^2 + y^2 < \frac{1}{3}\right\},\ C_3 = \left\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 < \frac{1}{4}\right\}.$$

Antag nu att Benjamin är tillräckligt duktig så att han alltid träffar darttavlan och att platsen där pilen landar på darttavlan är likformigt fördelat på darttavlan, d.v.s. platsen där pilen landar har täthetsfunktion

$$f(x, y) = \frac{1}{\pi},\ (x, y) \in C.$$

(a) Låt T vara antalet poäng Benjamin får vid ett kast. Härled tätheten för T. **3 poäng**

(b) Beräkna väntevärdet samt variansen av T. **1+1 poäng**

(c) Antag nu att Benjamin kastar 100 gånger och att varje kast kan antas vara oberoende. Beräkna (approximativt) sannolikheten att den totala poängen efter 100 kast överstiger 75. (Om du inte lyckas räkna ut väntevärdet eller variansen i del (b) så kan du anta att väntevärdet är 0.7 och variansen är 1). **3 poäng**

Lösning :
(a)
Om vi låter $Z = (X, Y)$ vara den stokastiska vektor som beskriver platsen där pilen landar så observerar vi att

$$\{T = i\} = \{Z \in C_i\}, i = 0, 1, 2, 3.$$

Med andra ord så gäller att följande sannolikheter är lika

$$P(T = i) = P(Z \in C_i) = \frac{1}{\pi} \int_{C_i} \mathrm{d}x\mathrm{d}y = \frac{\text{Area}(C_i)}{\pi},$$

där Area(C_i) är arean av området C_i. Allt som återstår att beräkna är därmed Area(C_i):

$$\text{Area}(C_0) = \pi - \left(\frac{2}{3}\right)^2 \pi = \frac{5\pi}{9}, \ \text{Area}(C_1) = \left(\frac{2}{3}\right)^2 \pi - \left(\frac{1}{3}\right)^2 \pi = \frac{\pi}{3},$$
$$\text{Area}(C_2) = \left(\frac{1}{3}\right)^2 \pi - \left(\frac{1}{4}\right)^2 \pi = \frac{7\pi}{144}, \ \text{Area}(C_3) = \left(\frac{1}{4}\right)^2 \pi = \frac{\pi}{16}.$$

Vilket ger sannolikheter :

$$P(T=0) = \frac{5}{9}, \ P(T=1) = \frac{1}{3},$$
$$P(T=2) = \frac{7}{144}, \ P(T=3) = \frac{1}{16}.$$

(b)
Väntevärde och varians fås av en enkel beräkning:

$$E[T] = \frac{1}{3} + 2\frac{7}{144} + 3\frac{1}{16} = \frac{89}{144} \approx 0.62,$$
$$E[T^2] = \frac{1}{3} + 2^2\frac{7}{144} + 3^2\frac{1}{16} = \frac{157}{144} \approx 1.1,$$
$$\text{Var}(T) = \frac{157}{144} - \left(\frac{89}{144}\right)^2 = \frac{14687}{20736} \approx 0.71.$$

(c)
Om vi låter S vara den totala poängen efter 100 kast så ger centrala gränsvärdessatsen att

$$S \overset{approx}{\sim} N\left(100E[T], 100\,\text{Var}(T)\right)$$

vilket ger sannolikheten

$$P(S \geq 75) \approx P\left(Z \geq \frac{75 - 100E[T]}{\sqrt{100\,\text{Var}(T)}}\right) = 1 - \Phi\left(\frac{75 - 100\frac{89}{144}}{\sqrt{100\frac{14687}{20736}}}\right) = 1 - \Phi\left(1.57\right) = 1 - 0.941792 = 0.058208.$$

3. Antalet bilar som passerar en viss korsning kan beskrivas med hjälp av en Poisson process där det förväntade antalet bilar som passerar korsningen under en timma är 6.
Beräkna sannolikheten under de första 30 minuterna passerar minst en bil. **2 poäng**
Lösning:
Låt $N(t)$ vara Poisson processen. Enligt uppgiftsformuleringen så har vi

$$\lambda = E[N(1)] = 6,$$

vilket ger

$$P(N(0.5) \geq 1) = 1 - \exp\{-0.5\lambda\} = 1 - \mathrm{e}^{-3}.$$

4. Låt $x = (x_1, x_2, ..., x_n)$ vara ett observerat stickprov från $N(\mu, \sigma^2)$.
   (a) Antag att σ är känd. Härled maximum-likelihoodskattaren för μ. **3 poäng**
   (b) Antag nu att σ är okänd. Härled momentmetodsskattaren för μ och σ. **2 poäng**

(c) Visa att både maximum-likelihoodskattaren och momentmetodsskattaren för μ är väntevärdesriktiga. **1 poäng**

Lösning: (a)
Log-likelihood funktionen ges av

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

och derivatan ges av

$$l'(\mu|x) = \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(x_i-\mu\right) = \frac{n}{\sigma^2}\left(\bar{x}-\mu\right).$$

Vilket ger ML skattare

$$\mu_{ML} = \bar{x}.$$

(b)
Momentmetoden ges av

$$\frac{1}{n}\sum_{1}^{n} x_i = E[X]$$
$$\frac{1}{n}\sum_{1}^{n} x_i^2 = E[X^2]$$

där $X \sim N(\mu, \sigma^2)$. Eftersom $\mathrm{Var}(X) = \sigma^2 = E[X^2] - E[X]^2 = E[X^2] - \mu^2$ så ser vi att ekvationssystemet ovan ges av

$$\frac{1}{n}\sum_{1}^{n} x_i = \mu$$
$$\frac{1}{n}\sum_{1}^{n} x_i^2 = \sigma^2 + \mu^2$$

vilket ger momentmetods skattare:

$$\mu_{MM} = \bar{x}, \sigma^2_{MM} = \frac{1}{n}\sum_{1}^{n} x_i^2 - \bar{x}^2.$$

(c)
Att båda skattare är väntevärdesriktiga fås genom att låta $X_1, ..., X_n$ vara ett stickprov av $N(\mu, \sigma^2)$ och observera att

$$\mathrm{E}\left[\mu_{MM}\right] = \mathrm{E}\left[\mu_{ML}\right] = \mathrm{E}\left[\frac{1}{n}\sum_{1}^{n} X_i\right] = \frac{1}{n} n\,\mathrm{E}[X_1] = \mu.$$

5. En persons IQ brukar antas vara normalfördelat med väntevärde 100 och standardavvikelse 15. För att undersöka om en viss populations genomsnitts-IQ skiljer sig från detta har ett stickprov på 50 personer dragits och IQ-testats. Deras medel-IQ beräknades till $\bar{x} = 104.3$.

   (a) Konstruera ett 95%-igt konfidensintervall för populationens medel-IQ. **2 poäng**

(b) Hur stort stickprov behövs för att konstruera ett 99%-igt av samma längd som i (a)? **2 poäng**

Lösning :
(a)
Då σ är känd ges konfidensintervallet av

$$\bar{x} \pm z_{0.025}\frac{15}{\sqrt{50}} \approx [100.1, 108.5]$$

där $z_{0.025} = 1.96$.
(b)
Låt n vara storleken på det nya stickprovet. Om det skall vara samma längd som konfidensintervallet krävs det att

$$z_{0.025}\frac{15}{\sqrt{50}} = z_{0.005}\frac{15}{\sqrt{n}} \Leftrightarrow n = 50 \cdot \frac{z_{0.005}^2}{z_{0.025}^2} = 50 \cdot \frac{2.576^2}{1.96^2} = 86.367... \approx 87.$$

6. Låt $x_1, ..., x_{100}$ vara ett observerat stickprov av den stokastiska variabeln X som har fördelning

$$P(X = 0) = 1 - p, P(X = 1) = p$$

och låt $y_1, ..., y_{120}$ vara ett observerat stickprov av stokastiska variabeln Y som har fördelning

$$P(Y = 0) = 1 - q, P(Y = 1) = q$$

med samt:

$$\sum_{i=1}^{100} x_i = 53, \ \sum_{i=1}^{120} y_i = 44.$$

Välj signifikansnivå och testa hypotesen att $p = q$ med två olika metoder. **3 poäng per metod, totalt 6 poäng** .
Lösning
Låt $\alpha = 0.05$ vara signifikansnivån. Jag väljer att testa med konfidensintervall samt p-värde. Först, låt $\hat{p} = 53/100, \hat{q} = 44/120$ vilket ger konfidensintervall

$$\hat{p} - \hat{q} \pm z_{0.025}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{\hat{q}(1-\hat{q})}{m}} \approx [0.0329, 0.2937] = I$$

och eftersom $0 \notin I$ så förkastar vi hypotesen att $p = q$. För att beräkna p-värdet så observerar vi att

$$\hat{p} - \hat{q} \overset{approx}{\sim} N\left(p - q, \frac{\hat{p}(1-\hat{p})}{n} + \frac{\hat{q}(1-\hat{q})}{m}\right)$$

eller ekvivalent

$$Z_0 = \frac{\hat{p} - \hat{q} - (p - q)}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{\hat{q}(1-\hat{q})}{m}}} \overset{approx}{\sim} N(0, 1).$$

Under hypotesen $p = q$ ges det observerade värdet på Z_o av

$$z_0 = \frac{\hat{p} - \hat{q}}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{\hat{q}(1-\hat{q})}{m}}} = 2.455.$$

P-värdet för hypotesen ges av

$$P(|Z| \geq z_o) = 2(1 - \Phi(z_o)) = 0.0141 < 0.05$$

och vi förkastar därmed hypotesen att $p = q$.