LINKÖPINGS UNIVERSITET
MAI
Johan Thim

Tentamen i matematisk statistik (92MA31, STN2)
2015-08-19 kl 8–12

Hjälpmedel är: miniräknare med tömda minnen och formelbladet bifogat.
Varje uppgift är värd 6 poäng. För godkänd tentamen räcker 16 poäng. Noggrann motivering krävs där alla viktiga detaljer skall motiveras.

För lösningsskisser, se kurshemsidan efter skrivningens slut. Lycka till!

---

1. Låt A, B och C vara händelser så att $P(A) = 0.4$, $P(B \cap A^*) = 0.2$ och $P(B \cap C) = 0.05$. Vidare så är $P(A \cap B) = P(A \cap C) = 0.10$ och $A \cap B$ samt $A \cap C$ är disjunkta (de har inget snitt). Händelserna B och C är dessutom oberoende.

   (a) Bestäm $P(B)$. (2p)

   (b) Är A och B oberoende? Oförenliga? Motivera ditt svar! (2p)

   (c) Beräkna $P(A \mid B \cup C)$. (2p)

2. Vid en serie oberoende mätningar av en process erhöll man följande värden:

   $$2.49 \quad 3.03 \quad 2.73 \quad 1.70 \quad 2.29 \quad 1.21 \quad 2.89 \quad 0.85\ .$$

   Vi antar att dessa värden utgör ett stickprov av en normalfördelad variabel $X \sim N(\mu_1, \sigma)$. Efter en tid misstänker man att något gått snett då produkterna som produceras inte längre håller samma kvalitet. Man gör nya mätningar på processen och erhåller då värderna

   $$1.32 \quad 1.31 \quad 0.14 \quad 0.97 \quad 0.84\ .$$

   Antag att detta är ett stickprov från en normalfördelad variabel $Y \sim N(\mu_2, \sigma)$.

   Beräkna 95% konfidensintervall (dubbelsidiga) för μ_1, μ_2, samt $\mu_1 - \mu_2$. Kan man dra någon statistiskt säker slutsats med 5% felrisk från dessa intervall för om det är någon skillnad mellan μ_1 och μ_2? Motivera ditt svar.

3. Svampplockaren Svea brukar spendera sina helger med att plocka små svampar i skogen vid en närliggande sjö. När hon kommer hem brukar hon fotografera och väga svamparna. På senare tid har hon misstänkt att hennes våg inte fungerar felfritt längre, så hon bestämmer sig för att mäta om en serie med svampar och jämnföra de nya vikterna med de vikter hon skrivit ned vid förra mätningen. Antag att $X_j \sim N(\mu_j, \sigma_1)$ är de gamla vikterna och att $Y_j \sim N(\mu_j + \Delta, \sigma_2)$ är de nya. Låt $Z_j = Y_j - X_j$. Stinas observerade värden blev:

x_j	56.68	53.22	55.20	53.91	55.61	53.80	55.98	56.48	58.42	54.61
y_j	61.82	57.65	58.76	53.88	55.06	55.46	58.20	55.44	57.40	51.90
z_j	5.14	4.43	3.56	−0.03	−0.54	1.66	2.22	−1.03	−1.02	−2.72

   Beräkna ett 99% konfidensintervall för Δ. Kan hypotesen att vågens funktion inte förändrats förkastas på 1% nivån? (6p)

Vänd!

4. Låt oss betrakta en 33 cl burk med läsk. Antag att vikten för innehållet (d v s vätskan) är normalfördelat med väntevärde 376.2 g och standardavvikelse 1.63 g, samt att vikten hos den tomma burken är normalfördelad med väntevärde 13.5 g och standardavvikelse 0.21 g. Variablerna anses vara oberoende.

   (a) Vad är sannolikheten att en slumpmässigt vald burk (med läsk) väger mer än 390 g? (3p)

   (b) Vad är väntevärdet för antalet plattor (en platta innehåller 24 burkar) vi behöver väga innan vi hittar en platta som väger över 9.35 kg om vi bortser vikten som tillkommer hos pappförpackningarna. (3p)

5. Låt (X, Y) vara en diskret tvådimensionell s.v. där X och Y är oberoende och $p_{X,Y}(j, k) = 0$ om $j < 0$, $j > 2$, $k < 0$ eller $k > 2$. För övriga värden på j och k ges sannolikhetsfunktionen av

j \ k	$k = 0$	$k = 1$	$k = 2$
$j = 0$	0.06	0.12	0.12
$j = 1$	0.08	?	?
$j = 2$	?	?	?

   (a) Räkna ut vad det måste stå där det finns frågetecken. Motivera! (3p)

   (b) Beräkna sannolikheten $P(X + Y < 2)$. (2p)

   (c) Beräkna kovariansen $C(X, Y)$. (1p)

6. Låt $X \sim \text{Exp}(\mu)$ där $E(X) = \mu$. Definiera $Y = \ln(1 + |X|)$ och bestäm vad Y har för täthetsfunktion.

Lösningsskisser för matematisk statistik 2015-08-19

1. (a) Vi vet att $0.20 = P(B \cap A^*) = P(B) - P(A \cap B)$, och då $P(A \cap B) = 0.10$ följer det att $P(B) = 0.30$.

   (b) Om A och B är oberoende så gäller att $P(A \cap B) = P(A)P(B)$. Vi vet att $P(A \cap B) = 0.10$, men $P(A)P(B) = 0.40 \cdot 0.30 = 0.12$. Händelserna är alltså inte oberoende. Då $P(A \cap B) \neq 0$ kan händelserna inte heller vara oförenliga (att vara oförenliga skulle innebära att $A \cap B = \emptyset$, så $P(A \cap B) = 0$ i det fallet).

   (c) Enligt definitionen av betingad sannolikhet så gäller att

$$P(A \mid B \cup C) = \frac{P(A \cap (B \cup C))}{P(B \cup C)}.$$

   Eftersom B och C är oberoende så är $P(B \cap C) = P(B)P(C)$. Detta ger att

$$P(C) = P(B \cap C)/P(B) = 0.05/0.3 = 1/6.$$

   Vi erhåller nu att

$$P(B \cup C) = P(B) + P(C) - P(B \cap C) = 0.30 + 0.1667 - 0.05 = 0.4167$$

   och att

$$P(A \cap (B \cup C)) = P((A \cap B) \cup (A \cap C)) = P(A \cap B) + P(A \cap C) = 0.20.$$

   Vi kan nu beräkna att

$$P(A \mid B \cup C) \approx \frac{0.20}{0.4167} \approx 0.48.$$

   Svar: (a) $P(B) = 0.30$. (b) Händelserna är varken oberoende eller oförenliga. (c) ≈ 0.48.

2. Vi har två olika stickprov, ett med 8 mätningar och ett med 5. Modellen förutsätter att de kommer från två källor som har samma varians (är detta rimligt?). Vi punktskattar med medelvärdet: $\mu_1^* = \bar{X} \sim N(\mu_1, \sigma_1^2/8)$ som vanligt och skattar σ med s_1, där

$$s_1^2 = \frac{1}{7}\sum_{i=1}^{8}(x_i - \bar{x})^2 \approx 0.6528125$$

   är stickprovsvariansen. Vi skapar testvariabeln

$$T = \frac{\mu_1^* - \mu_1}{s_1/\sqrt{8}} \sim t(7).$$

   Det följer att

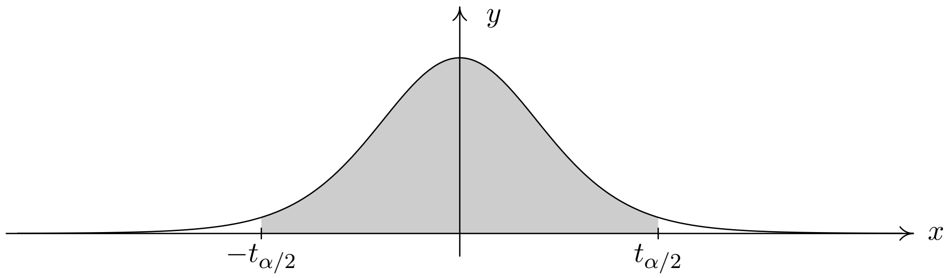
$$P\big(-t_{\alpha/2}(7) < T < t_{\alpha/2}(7)\big) = 1 - \alpha,$$

   där $t_\alpha(7)$ är α-kvantilen till $t(7)$-fördelningen.

   Ur tabell finner vi $t_{0.025}(7) = 2.365$. Genom att lösa ut μ_1 ur olikheten i sannolikhetsmåttet får vi uttrycket

$$\mu_1^* - \frac{2.365 \cdot 0.8080}{\sqrt{8}} < \mu_1 < \mu_1^* + \frac{2.365 \cdot 0.8080}{\sqrt{8}}.$$

   Om vi ersätter μ_1^* med den observerade punktskattningen $(\mu_1)^*_{\text{obs}} = 2.149$ (medelvärdet av observationerna) så får vi ett konfidensintervall $I_{\mu_1} = [1.47, 2.82]$ med konfidensgrad 95%.



Figur 1: $t(7)$-fördelningen med 95% av arean markerad.

På samma sätt kan vi finna ett konfidensintervall för μ_2. Värderna vi använder är $s_2^2 = 0.2323$ och $(\mu_2)^*_{\text{obs}} = 0.9160$, och fördelningen för testvariabeln är $t(4)$. Motsvarande kvantil kan hittas i tabell och ges av $t_{0.025}(4) = 2.776$. Vi erhåller $I_{\mu_2} = [0.32, 1.51]$.

Ur dessa intervall kan vi inte säga något om skillnaden $\mu_1 - \mu_2$ med konfidensgrad 95%. Istället punktskattar vi $\mu_1 - \mu_2$ med $\mu^* = \bar{X} - \bar{Y}$. Vi betraktar sedan variabeln

$$T = \frac{\mu^* - (\mu_1 - \mu_2)}{s\sqrt{1/8 + 1/5}},$$

där

$$s^2 = \frac{7s_1^2 + 4s_2^2}{11}$$

är den sammanvägda skattningen av σ^2. Det följer att $T \sim t(11)$. Nu fullföljer vi som vanligt, med $s^2 = 0.4999$, $\mu^*_{\text{obs}} = 1.2328$, $s\sqrt{1/8 + 1/5} = 0.4031$, och $t_{0.025}(11) = 2.201$, och får konfidensintervallet $I_{\mu_1 - \mu_2} = [0.35, 2.12]$. Eftersom nollan INTE ingår i intervallet kan vi med 5% felrisk påstå att $\mu_1 > \mu_2$.

3. Modellen är stickprov i par. Vi betraktar skillnaden som ett enda stickprov och skattar väntevärdet med

$$\bar{\Delta} = 1.167$$

och variansen med

$$s = \left(\frac{1}{10 - 1} \sum_{i=1}^{10} (x_i - y_i - \bar{\Delta})^2 \right)^{1/2} = 2.63.$$

Vi bildar testvariabeln

$$T = \frac{\Delta^* - \Delta}{s/\sqrt{n}} = \frac{\Delta^* - \Delta}{0.8338} \sim t(9)$$

och ser att

$$P(-t_{\alpha/2}(9) < T < t_{\alpha/2}(9)) = 1 - \alpha.$$

Vi löser ut Δ ur intervallet i sannolikhetsmåttet och får att

$$\Delta^* - 0.8338 t_{\alpha/2}(9) < \Delta < \Delta^* + 0.8338 t_{\alpha/2}(9).$$

Vi ersätter Δ^* med den observerade punktskattningen $\Delta^*_{\text{obs}} = 1.167$, d v s med medelvärdet av våra z_j, och erhåller ett konfidensintervall av konfidensgrad 99%:

$$I_\Delta = [-1.54, 3.88],$$

där vi använt $\alpha = 0.01$ och $t_{0.005}(9) = 3.25$. Eftersom $0 \in I_\Delta$ kan vi **inte** förkasta hypotesen att vågens funktion inte förändrats.

Svar: Hypotesen att vågens funktion inte förändrats kan inte förkastas på 1%-nivån.

4. Låt $B \sim N(13.5, 0.21^2)$ och $L \sim N(376.2, 1.63^2)$ vara stokastiska variabler som representerar en burk respektive innehållet (läsken). Vikten X för en full burk med läsk blir då normalfördelad (eftersom variablerna förutsattes vara oberoende): $X \sim N(389.7, 1.643^2)$.

   (a) För en burk erhåller vi

$$P(X > 390) = P\left(\frac{X - 389.7}{1.64} > \frac{390 - 389.7}{1.64}\right) = 1 - \Phi(0.183) \approx 1 - 0.5714 = 0.4286.$$

   (b) En hel platta får vikten $Y \sim N(24 \cdot 389.7, 1.643^2 \cdot 24) = N(9352.8, 8.03^2)$. Låt Z vara antalet plattor man behöver välja ut innan man för första gången får en som väger mer än 9.35kg. Den s.v. Z är Ffg(p)-fördelad, där $p = P(Y > 9350) = 1 - \Phi(0.349) \approx 0.36$. Enligt formelsamling erhåller vi $E(Z) = 1/p \approx 2.75$.

   Svar: a) 0.429. b) 2.75 så 3 st. behöver undersökas.

5. (a) Eftersom X och Y är oberoende så kommer sannolikhetsfunktionen att ges av produkten av de marginella sannolikhetsfunktionerna: $p_{X,Y}(j,k) = p_X(j)p_Y(k)$. Vi kan direkt läsa ut att

$$p_X(0) = \sum_{k=0}^{2} p_{X,Y}(0,k) = 0.06 + 0.12 + 0.12 = 0.30.$$

   Från detta följer att $p_{X,Y}(0,k) = p_X(0)p_Y(1) = 0.30 \cdot p_Y(k)$, och om vi utnyttjar informationen given i tabellen kan vi nu lösa ut att

$$p_Y(0) = 0.06/0.30 = 0.2,$$
$$p_Y(1) = p_Y(2) = 0.12/0.30 = 0.40.$$

   Vi kan även lösa ut att $p_X(1) = 0.08/0.40 = 0.20$, vilket vi sedan kan använda för att räkna ut att $p_{X,Y}(2,0) = 0.30 \cdot 0.20 = 0.06$.
   Med samma teknik som ovan kan vi räkna ut att $p_Y(1) = p_Y(2) = 0.40$, och därefter resterande fält i tabellen. Vi erhåller sålunda

j \ k	$k=0$	$k=1$	$k=2$	$p_X(j)$
$j=0$	0.06	0.12	0.12	0.30
$j=1$	0.08	0.16	0.16	0.40
$j=2$	0.06	0.12	0.12	0.30
$p_Y(k)$	0.20	0.40	0.40	1.00

   (b) Vi söker sannolikheten att $X + Y < 2$. Om $X = j$ och $Y = k$ händer detta endast vid $(j,k) = (0,0)$, $(0,1)$ samt $(1,0)$ med nollskild sannolikhet. Vi får alltså

$$P(X + Y < 2) = 0.06 + 0.12 + 0.08 = 0.26.$$

   (c) Kovariansen kan, t ex, beräknas med formeln $C(X,Y) = E(XY) - E(X)E(Y)$. Enklare är att utnyttja att vi vet att variablerna är oberoende, så $C(X,Y) = 0$ är nödvändigt.

   Svar: (a) Se tabell ovan. (b) 0.26. (c) $C(X,Y) = 0$.

6. Eftersom $1 + |X| \geq 1$ så är $Y \geq 0$ med sannolikhet ett. Därför är $f_Y(y) = 0$ för $y \leq 0$. Låt $y > 0$. Då gäller att

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(\ln(1 + |X|) \leq y) \\ &= P(1 + |X| \leq e^y) = P(|X| \leq e^y - 1) \\ &= P(1 - e^y \leq X \leq e^y - 1) = P(X \leq e^y - 1) \\ &= F_X(e^y - 1), \end{aligned}$$

eftersom $1 - e^y < 0$ för $y > 0$. Eftersom $f_Y(y) = F'_Y(y)$ (snälla funktioner) så erhåller vi nu

$$f_Y(y) = \frac{d}{dy}\left(F_X(e^y - 1)\right) = e^y f_X(e^y - 1) = \frac{e^y}{\mu}\exp\left(\frac{e^y - 1}{\mu}\right).$$

Svar: $f_Y(y) = \dfrac{e^y}{\mu}\exp\left(\dfrac{e^y - 1}{\mu}\right)$ för $y > 0$.