LINKÖPINGS UNIVERSITET
MAI
Johan Thim

Tentamen i matematisk statistik (9MA241/9MA341, STN2) 2012-08-24 kl 08-12

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 `www.mai.liu.se/~jothi/kurser/9MA241-stat/` efter skrivningens slut. Lycka till!

---

1. Låt A och B vara händelser så att $P(A \cup B) = 0.6$ och $P(A \cap B) = 0.3$.

   (a) Beräkna $P(A) + P(B)$. (2p)

   (b) Antag vidare att $P(A) = 0.4$. Beräkna $P(A \,|\, B^*)$. (2p)

   (c) Antag att $P(B^*) \neq 0$ och att $P(A \,|\, B^*) = P(A)$. Är A och B oberoende i detta fall? Bevisa eller finn motexempel. (2p)

2. (a) Vid 16 upprepade mätningar av en process fann man att medelvärdet blev $\bar{x} = 21$ och stickprovsvariansen $s_x^2 = 7.2$. Antag att mätningarna är observationer av en stokastisk variabel $X \sim \mathrm{N}(\mu, \sigma)$. Konstruera ett 99% konfidensintervall för väntevärdet μ. (3p)

   (b) På grund av hur processen ser ut vet man att det verkliga värdet för σ är $\sigma = 7.0$. Konstruera ett 99% konfidensintervall för μ där vi utnyttjar vad vi vet om σ. (3p)

3. 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)

Var god vänd!

4. (a) Låt X vara en s.v. med täthetsfunktion

$$f_X(x) = \begin{cases} 3x(2-x)/4, & 0 \leq x \leq 2, \\ 0, & \text{för övrigt.} \end{cases}$$

Visa att $f_X(x)$ är en täthetsfunktion och beräkna $E(X)$ samt $V(X)$. (4p)

(b) Låt Y vara en summa av 100 stycken oberoende variabler med samma täthetsfunktioner som $f_X(x)$ i deluppgift (a). Bestäm approximativt sannolikheten att $Y < 105$. (2p)

5. Kalle Kanon har varit och köpt budget-ammunition som tillverkaren hävdar har 1% felrisk (risken att få en blindgångare). Kalle planerar att testa tillverkarens påstående genom att skjuta 1000 stycken slumpmässigt utvalda kulor. Vi antar att Kalles vapen är felfritt och fungerar perfekt samt att Kalle inte har några problem med koncentrationen vid upprepande sysselsättning

(a) Antag att tillverkaren har rätt och felrisken är $p = 0.01$. Beräkna approximativt sannolikheten för precis 10 blindgångare vid Kalles test. Vad blir approximativt sannolikheten att Kalle finner färre än 20 stycken blindgångare? (3p)

(b) Vid Kalles test fann han 22 stycken blindgångare. Använd detta för att finna ett approximativt 95% konfidensintervall för den verkliga felrisken p. Vad kan du dra för slutsats angående företagets påstående om 1% felrisk? (3p)

6. Låt X vara likformigt fördelad på intervallet $[0, 1]$, d.v.s. $X \sim \text{Re}(0, 1)$. Definiera den s.v. Y genom $Y = -\dfrac{1}{\lambda}\ln(1 - X)$ för något $\lambda > 0$. Visa att $Y \sim \text{Exp}(\lambda)$ där Y har väntevärde $1/\lambda$. (6p)

Lösningsskisser för tentamen i matematisk statistik, 9MA241, 2012-08-24

1. (a) Eftersom $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, följer det att $P(A) + P(B) = 0.6 + 0.3 = 0.9$.
   (b) Om det är givet att $P(A) = 0.4$ söker vi $P(A|B^*)$. Eftersom $P(A) + P(B) = 0.9$ så måste $P(B) = 0.5$. Vi vet också enligt definitionen att

$$P(A \,|\, B^*) = \frac{P(A \cap B^*)}{P(B^*)} = \frac{P(A) - P(A \cap B)}{1 - P(B)} = \frac{0.4 - 0.3}{1 - 0.5} = 0.2.$$

   (c) Låt $P(A \,|\, B^*) = P(A)$. Eftersom $P(B^*) \neq 0$ gäller att

$$\begin{aligned} \frac{P(A \cap B^*)}{P(B^*)} = P(A) \quad &\Leftrightarrow \quad P(A \cap B^*) = P(B^*)P(A) \\ &\Leftrightarrow \quad P(A) - P(A \cap B) = P(A)(1 - P(B)) \\ &\Leftrightarrow \quad P(A \cap B) = P(A)P(B). \end{aligned}$$

   Således är A och B oberoende.

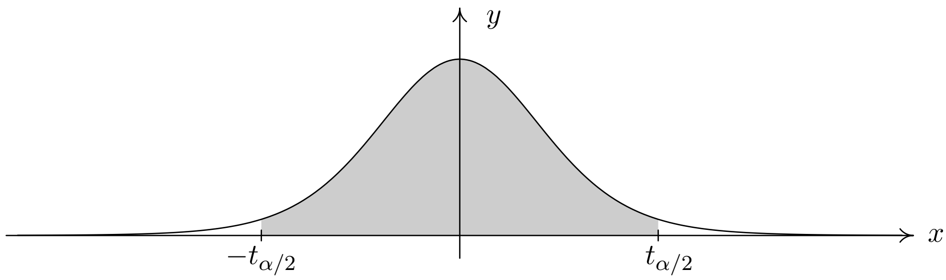
   Svar: (a) 0.9. (b) 0.2. (c) Oberoende. Se ovan.

2. (a) Vi har $n = 16$ och bildar testvariabeln

$$T = \frac{\overline{X} - \mu}{S_x/\sqrt{n}} \sim t(15)$$

   och ser att

$$P(-t_{\alpha/2}(15) \leq T \leq t_{\alpha/2}(15)) = 1 - \alpha.$$



Figur 1: Den markerade arean innehåller 99% av sannolikheten för $t(15)$-fördelningen.

   Vi löser ut μ ur intervallet i sannolikhetsmåttet och får att

$$\overline{X} - \frac{S_x}{\sqrt{n}} t_{\alpha/2}(15) \leq \mu \leq \overline{X} + \frac{S_x}{\sqrt{n}} t_{\alpha/2}(15).$$

   Vi ersätter $\overline{X}$ med den observerade punktskattningen $\overline{X}^*_{\text{obs}} = \overline{x} = 21$ och S_x med $s_x = \sqrt{7.2}$. Då erhåller vi ett konfidensintervall med konfidensgrad 99%:

$$I_\mu = [19.02, 22.98],$$

   där vi använt $\alpha = 0.01$ och $t_{0.005}(15) = 2.947$.

(b) Vi använder oss av testvariabeln

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

Vi vet att $\sigma = \sqrt{7.0}$ och det följer att

$$P(-\lambda_{\alpha/2} \leq Z \leq \lambda_{\alpha/2}) = 1 - \alpha.$$

På samma sätt som ovan löser vi ut μ ur intervallet i sannolikhetsmåttet och får att

$$\overline{X} - \lambda_{\alpha/2}\sigma/\sqrt{n} \leq \mu \leq \overline{X} + \lambda_{\alpha/2}\sigma/\sqrt{n}.$$

Vi ersätter $\overline{X}$ med den observerade punktskattningen $\overline{x} = 21$ och $\sigma = \sqrt{7.0}$. Då erhåller vi ett konfidensintervall av konfidensgrad 99%:

$$I_\mu = [19.30, 22.70],$$

där vi använt $\alpha = 0.01$ och $\lambda_{0.005} = 2.575$.

Svar: (a) $I_\mu = [19.02, 22.98]$. (b) $I_\mu = [19.30, 22.70]$.

3. (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

$j \backslash k$	$k=0$	$k=1$	$k=2$	$p_X(j)$
$j=0$	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) $C(X,Y) = 0$.

4. (a) Det är klart att $f_X(x) \geq 0$ för alla x, och vi ser att

$$\int_{-\infty}^{\infty} f_X(x)\,dx = \frac{3}{4}\int_0^2 (2x - x^2)dx = \frac{3}{4}\left[x^2 - \frac{x^3}{3}\right]_0^2 = \frac{3}{4}\frac{4}{3} = 1,$$

så f_X är en täthetsfunktion. För att beräkna väntevärdet använder vi definitionen:

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx = \frac{3}{4}\int_0^2 (2x^2 - x^3)dx = \frac{3}{4}\left[\frac{2x^3}{3} - \frac{x^4}{4}\right]_0^2 = 1.$$

Eftersom $V(X) = E(X^2) - E(X)^2$, och

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x)\,dx = \frac{3}{4}\int_0^2 (2x^3 - x^4)dx = \frac{3}{4}\left[\frac{x^4}{2} - \frac{x^5}{5}\right]_0^2 = \frac{6}{5},$$

erhåller vi $V(X) = 6/5 - 1^2 = 1/5$.

(b) Låt X_i, $i = 1, 2, \ldots, 100$, vara oberoende s.v. med samma täthetsfunktion $f_X(x)$ som i uppgift (a). Vi definierar $Y = X_1 + X_2 + \cdots + X_{100}$. Enligt CGS kommer Y att vara approximativt normalfördelad med väntevärde och varians enligt

$$E(Y) = 100 \cdot E(X) = 100,$$
$$V(Y) = 100 \cdot V(X) = 100/5.$$

Vi beräknar approximativt

$$P(Y < 105) \approx \Phi\left(\frac{105 - 100}{\sqrt{100/5}}\right) = \Phi\left(\frac{5\sqrt{5}}{10}\right) = \Phi(\sqrt{5}/2) \approx \Phi(1.12) = 0.87.$$

Att det är ok att använda CGS följer av att vi summerar så många variabler, och att fördelningen för varje X_i redan är ganska symmetrisk (skissa upp $f_X(x)$).

Svar: (a) $E(X) = 1$ och $V(X) = 1/5$. (b) ≈ 0.87.

5. (a) Låt X vara antalet blindgångare i Kalles test. Då blir X binomialfördelad: $X \sim \text{Bin}(n, p)$ med $n = 1000$ och $p = 0.01$ (vi antar att tillverkaren talat sanning). Vi antar även att varje skott Kalle tar är oberoende av övriga (kanske inte helt sant då det finns gemensamma faktorer så som Kalles vapen och Kalle själv). Enklast nu är att Poissonapproximera X: $X \overset{\text{appr.}}{\sim} \text{Po}(10)$, vilket är OK då $p \leq 0.1$ och $n \geq 10$. Vi finner sedan att

$$P(X = 10) = P(X \leq 10) - P(X \leq 9) \approx 0.5830 - 0.4579 = 0.1251$$

och att

$$P(X < 20) = P(X \leq 19) \approx 0.9965$$

ur tabell för Po(10).

Anmärkning: Redan här ser vi att $p = 0.01$ inte kommer att förefalla rimligt med resultatet av Kalles försök i deluppgift (b). Om p verkligen har det värdet är resultatet $X = 22$ orimligt. Detta är ett alternativt sett att testa hypoteser (istället för med konfidensintervall). Läs mer i boken om s.k. hypotestest!

(b) Vi söker konfidensintervall för felrisken vid Kalles test. Kalle har testat $n = 1000$ stycken kulor, och fått $X = 22$ stycken blindgångare. En naturlig skattning på den verkliga andelen ges av

$$\widehat{P} = \frac{X}{n},$$

och vårt observerade värde är $\widehat{p} = 22/1000 = 0.022$. Vi vet att variabeln X är binomialfördelad: $X \sim \text{Bin}(n, p)$. Vidare ser vi att

$$np(1-p) \approx 1000 \cdot 0.022 \cdot (1 - 0.022) = 21.516,$$

så vi borde kunna göra en normalapproximation: $X \overset{\text{appr.}}{\sim} N(np, D)$ där $D = \sqrt{np(1-p)}$. Alltså blir $\widehat{P} \overset{\text{appr.}}{\sim} N(p, d)$, där

$$d = \sqrt{\widehat{p}(1-\widehat{p})/n} = \sqrt{21.516/1000^2} = 0.00464.$$

Vår testvariabel blir alltså

$$Z = \frac{\widehat{P} - p}{d} \overset{\text{appr.}}{\sim} N(0, 1).$$

Vi stänger in Z:

$$P(-\lambda_{\alpha/2} \leq Z \leq \lambda_{\alpha/2}) \approx 1 - \alpha,$$

där λ_α är normalfördelningens α-kvantil. Vi har $\alpha = 0.05$ och $\lambda_{\alpha/2} = \lambda_{0.025} = 1.96$. Vi löser ut p ur olikheten inuti sannolikhetsmåttet, och erhåller då

$$\widehat{P} - \lambda_{0.025} d \leq p \leq \widehat{P} + \lambda_{0.025} d.$$

Detta ger oss konfidensintervallet $I_p = [0.017, 0.027]$ om vi ersätter $\widehat{P}$ med det observerade värdet $\widehat{p} = 0.022$. Vi ser att $p = 0.01 \notin I_p$, så tillverkarens påstående verkar inte stämma.

Svar: (a) $P(X = 10) \approx 0.13$ och $P(X < 20) \approx 0.997$.

(b) $I_p = [0.017, 0.027]$. Det verkar inte som företaget varit helt ärliga.

6. Låt $y \geq 0$. Fördelningsfunktionen $F_Y(y)$ för den s.v. Y fås enligt

$$\begin{aligned} F_Y(y) = P(Y \leq y) &= P\left(-\frac{1}{\lambda}\ln(1 - X) \leq y\right) \\ &= P(\ln(1 - X) \geq -\lambda y) = P(1 - X \geq \exp(-\lambda y)) \\ &= P(X \leq 1 - \exp(-\lambda y)) \end{aligned}$$

Eftersom X är likformigt fördelad på $[0, 1]$ så gäller att $P(X \leq a) = a$ för varje $0 \leq a \leq 1$. Vidare vet vi att $0 \leq 1 - \exp(-\lambda y) \leq 1$ och således erhåller vi

$$F_Y(y) = P(X \leq 1 - \exp(-\lambda y)) = 1 - \exp(-\lambda y).$$

Vi kan nu derivera fram täthetsfunktionen för Y:

$$f_Y(y) = F_Y'(y) = \lambda \exp(-\lambda y),$$

vilket mycket riktigt är täthetsfunktionen för en exponentialfördelad variabel Y med väntevärde $1/\lambda$.

Svar: Se ovan.