

Petter Mostad
Tillämpad matematik och statistik
Matematiska Vetenskaper, Chalmers

**Lösningsförslag för
MVE051/MSG810 Matematisk statistik och diskret matematik**

Omtenta 12 April 2017, 8:30 - 12:30

1. (a) Minvärdet är 0.8 och maxvärdet 1.0: Vi har $\Pr(A \cup B) \geq \Pr(A) = 0.8$, och detta realiserar om $A = B$. Samtidigt skulle B kunna innehålla alla utfall som inte är i A , ty dessa har sannolikhet 0.4, och då får vi $\Pr(A \cup B) = 1$.
- (b) Minvärdet är 0.6 och maxvärdet 0.8: Om A och B har minimalt överlapp, så är storleken på detta överlapp 0.6, ty det totala värdet på sannolikheterna är $0.8 + 0.8 = 1.6$. Överlappet är maximalt om $A = B$, och då är $\Pr(A \cap B) = \Pr(A) = 0.8$.

2. (a)

$$\binom{6}{4} 0.69^4 (1 - 0.69)^2 = \frac{6 \cdot 5}{1 \cdot 2} 0.69^4 (1 - 0.69)^2 = 0.3267$$

- (b) Negativ Binomialfördelning.

- (c) Vi approximerar med 1 minus värdet i 149.5 av fördelningsfunktionen för en normalfördelning med väntevärde $200 \cdot 0.69 = 138$ och varians $200 \cdot 0.69 \cdot (1 - 0.69) = 42.78$. Vi använder en tabell för att hitta värdet av fördelningsfunktionen för en standard normalfördelning i

$$\frac{149.5 - 138}{\sqrt{42.78}} = 1.76.$$

Tabellen ger sannolikheten 0.9608, och vi får därmed som resultat $1 - 0.9608 = 0.0392$.

3. (a) Vi beräknar först c :

$$1 = \int_0^1 [(x-1)^2 + c] dx + \int_1^2 c dx = \left[\frac{1}{3}x^3 - x^2 + x \right]_0^1 + c + c = \frac{1}{3} + 2c$$

som ger $c = \frac{1}{3}$. Därmed får vi

$$\Pr(X > 1) = \int_1^2 \frac{1}{3} dx = \frac{1}{3}.$$

- (b) Vi får

$$\mathbf{E}(X) = \int_0^1 x \left[(x-1)^2 + \frac{1}{3} \right] dx + \int_1^2 x \frac{1}{3} dx = \left[\frac{1}{4}x^4 - \frac{2}{3}x^3 + \frac{2}{3}x^2 \right]_0^1 + \left[\frac{1}{6}x^2 \right]_1^2 = \frac{3}{4}$$

(c) Vi får

$$\mathbf{E}(X^2) = \int_0^1 x^2 \left[(x-1)^2 + \frac{1}{3} \right] dx + \int_1^2 x^2 \frac{1}{3} dx = \left[\frac{1}{5}x^5 - \frac{2}{4}x^4 + \frac{4}{9}x^3 \right]_0^1 + \left[\frac{1}{9}x^3 \right]_1^2 = \frac{83}{90}$$

$$\text{och d\u00e4rmed } \mathbf{Var}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2 = \frac{83}{90} - \frac{9}{16} = 0.3597.$$

4. (a) Vi f\u00e5r estimatet $\hat{p} = \frac{51}{87} = 0.5862$ och 90% konfidensintervall

$$\hat{p} \pm z_{0.05} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.5862 \pm 1.65 \cdot \sqrt{\frac{0.5862 \cdot 0.4138}{87}} = 0.5862 \pm 0.0977$$

- (b) Vi ser fr\u00e5n formeln f\u00f6r konfidensintervallet att det beror p\u00e5 antalet observationer, n , via faktorn $1/\sqrt{n}$. Om vi vill att denna faktorn skall \u00e4ndras till att bli en tredel s\u00e5 stor, s\u00e5 m\u00e5ste n bli 9 g\u00e5nger s\u00e5 stor, allts\u00e5 totalt $9 \cdot 87 = 783$ observationer. Eftersom vi redan har 87 observationer beh\u00f6ver vi $783 - 87 = 696$ nya observationer.

5. (a) $m_X(t) = \mathbf{E}(e^{tX})$.

$$(b) m_X(t) = \mathbf{E}(e^{tX}) = \sum_{x=0}^n \binom{n}{x} p^x (1-p)^{n-x} e^{tx} = \sum_{x=0}^n \binom{n}{x} (pe^t)^x (1-p)^{n-x} = (pe^t + 1-p)^n.$$

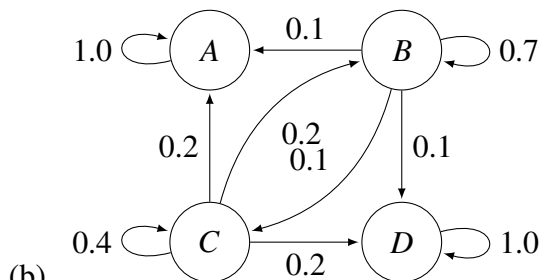
- (c) Derivering ger $m'_X(t) = n(pe^t + 1-p)^{n-1} pe^t$ och $\mathbf{E}(X) = m'_X(0) = np$. Vidare f\u00e5r vi

$$m''_X(t) = n(n-1)(pe^t + 1-p)^{n-2} p^2 e^{2t} + n(pe^t + 1-p)^{n-1} pe^t.$$

och d\u00e4rmed $m''_X(0) = n(n-1)p^2 + np$. Detta ger

$$\mathbf{Var}(X) = \mathbf{E}(X^2) - \mathbf{E}(X)^2 = m''_X(0) - m'_X(0)^2 = n(n-1)p^2 + np - n^2 p^2 = np(1-p).$$

6. (a) En Markovk\u00e4dja \u00e4r station\u00e4r om f\u00f6rdelningen f\u00f6r X_n givet X_{n-1} \u00e4r oberoende av n .



(c)

$$\begin{bmatrix} 1.0 & 0 & 0 & 0 \\ 0.1 & 0.7 & 0.1 & 0.1 \\ 0.2 & 0.2 & 0.4 & 0.2 \\ 0 & 0 & 0 & 1.0 \end{bmatrix}$$

- (d) Tillst\u00e5nden A och D \u00e4r absorberande, medan B och C \u00e4r transienta. K\u00e4djan \u00e4r absorberande.

- (e) Om vi skriver övergångsmatrisen på standard form, genom att använda ordningen B, C, A, D av tillstånden, så får vi

$$\begin{bmatrix} Q & R \\ 0 & I \end{bmatrix} = \begin{bmatrix} 0.7 & 0.1 & 0.1 & 0.1 \\ 0.2 & 0.4 & 0.2 & 0.2 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Den fundamentala matrisen blir därmed

$$N = (I - Q)^{-1} = \begin{bmatrix} 0.3 & -0.1 \\ -0.2 & 0.6 \end{bmatrix}^{-1} = \frac{1}{0.3 \cdot 0.6 - 0.1 \cdot 0.2} \begin{bmatrix} 0.6 & 0.1 \\ 0.2 & 0.3 \end{bmatrix} = \begin{bmatrix} 3.75 & 0.625 \\ 1.25 & 1.875 \end{bmatrix}$$

Om man startar i tillståndet C så är förväntad antal steg till absorption $1.25 + 1.875 = 3.125$.

7. (a)

$$1 + x + \dots + x^6 + x^7 = \frac{1 - x^8}{1 - x}$$

(b)

$$x^5 + x^6 + \dots = \frac{x^5}{1 - x}$$

(c) Produkten av de genererande funktionerna blir

$$\frac{1 - x^8}{1 - x} \cdot \frac{x^5}{1 - x} \cdot \frac{1}{1 - x} \cdot \frac{1}{1 - x} = \frac{x^5 - x^{13}}{(1 - x)^4} = (x^5 - x^{13}) \sum_{n=0}^{\infty} \binom{n+3}{3} x^n$$

Vi får att koefficienten framför x^{17} i denna serien blir

$$\binom{12+3}{3} - \binom{4+3}{3} = \frac{15!}{12!3!} - \frac{7!}{4!3!} = 420$$

som blir lika med antalet heltalslösningar av de ursprungliga ekvationerna.

8. Vi behöver anta att approximationen där väntevärdet av styrkan i fiberna är normalfördelad med väntevärde μ är tillräckligt bra. Nollhypotesen blir

$$H_0 : \mu = 3.4$$

medan den alternativa hypotesen blir

$$H_1 : \mu < 3.4$$

Teststatistikan blir

$$T = \frac{3.24 - 3.4}{\sqrt{0.12/16}} = -1.8475$$

och den skall jämföras med en T-fördelning med 15 frihetsgrader. Tabellen ger att sannolikheten för att vara mindre än detta värde är någonstans mellan 0.05 och 0.025. Därmed blir p-värdet någonstans mellan 0.05 och 0.025.

9. (a) Residualerna är

$$e_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$$

för $i = 1, \dots, 14$. Summan av residualerna i en regression är alltid lika med noll.

- (b) Om x och y är oberoende så betyder det att väntevärdet för lutningskoefficienten för regressionslinjen är noll. Men det betyder inte att estimatet $\hat{\beta}_1$ är noll; detta värde kan gott vara skild från noll på grund av stokastisk variation i data.
- (c) Om $\hat{\beta}_1 = 0$ så är estimatet för det *linjära* beroende mellan x och y lika med noll. Dock skulle det kunna vara så att det finns något litet linjärt beroende mellan variablerna, medan $\hat{\beta}_1$ blir noll av slumpen. Det är också så att generellt kan x och y vara beroende utan att de har ett *linjärt* beroende.
- (d) Längden av konfidensintervall kommer bero på x : Om x är nära medelvärdet av alla observerade x så kommer längden vara minst, medan ju längre från detta medelvärdet x , ju större kommer konfidensintervallet att vara. Orsaken är att osäkerheten beror på osäkerheten i regressionslinjen, och den ökar ju längre x är från medelvärdet av observerade x .