Lösningar till tentamen i Matematisk statistik, TMS155, 18 april 2006

1. a) $E[X] = 3 \cdot 0.5 + 2 \cdot 0.3 + 1 \cdot 0.1 = 2.2$; $Var(X) = (3-2.2)^2 0.5 + ... + (0-2.2)^2 .1 = 0.96$

   b) Centrala gränsvärdessatesen ger att summan är approximativt normalfördelad med väntevärde $E[\sum_{i=1}^{100} X_i] = 100E[X_1] = 220$, och varians $Var(\sum_{i=1}^{100} X_i) = 100Var(X_1) = 96$.

2. Låt A = Produkt från maskin A, B = Produkt från maskin B och C = Produkt från maskin C. Vi har att $P(A) = 0.5, P(B) = 0.4$, och $P(C) = 0.1$. Låt D = Produkt är defekt. Vi har att $P(D|A) = 0.01, P(D|B) = 0.02$, och $P(D|C) = 0.05$. Sannolikheten för att en produkt är defekt blir då

$$P(D) = P(D|A)P(A) + P(D|B)P(B) + P(D|C)P(C) = 0.018.$$

   (b) $P(A|D) = P(D|A)P(A)/P(D) \approx 0.28$.

3. Medelvärde: $\overline{x} = 21.81$
   Stickprovsstandardavvikelese: $s^2 \approx 1.65$
   t-tab, 9 frihetsgrader: $t_{0.025} = 2.262$.
   95%igt konfidensintervall: $[\overline{x} - t_{0.025}s/\sqrt{n}, \overline{x} + t_{0.025}s/\sqrt{n}] = [20.89, 22.73]$.

4. $H_0 : p = 0.3$
   $H_1 : p < 0.3$

   a) Låt X = Antal försök bland 5 som lyckas. X är binomial (5,p), där $p = 0.3$ under H_0. Vi räknar ut vårt p-värde: $P(X \leq 1) = 0.7^5 + 5 \cdot 0.7^4 0.3 \approx 0.53$. Vi kan ej förkasta H_0.

   (b) Låt Y= antalet försök tills det första lyckade. Y är geometriskt fördelad med parameter p , där $p = 0.3$ under H_0.
   p-värde=$P(Y \geq 6) = P(f$örsta 5 försöken misslyckas$) = 0.7^5 = 0.17$. Vi kan ej förkasta H_0.

5. Låt $Z = \max\{X, Y\}$.
   $F_Z(z) = P(max\{X,Y\} \leq z) = P(X \leq z)P(Y \leq z) = z^2$, för $0 < z < 1$.
   $f_Z(z) = F_Z'(z) = 2z$, för $0 < z < 1$.

6. a)

$$Var(X) = E[(X-\mu)^2] = E[X^2 - 2X\mu + \mu^2] = E[X^2] - 2\mu E[X] + \mu^2 = E[X^2] - \mu^2$$

   där $\mu = E[X]$.
   b) Om $korr(X, Y) = 1$ så $Y = a + bX$ där b är positiv, och om $korr(X, Y) = -1$ så $Y = a - bX$ där b är positiv.
   Om X och Y är oberoende så är korrelationen 0.

7. a) $f(15) = f(-30) = 1/8$, $f(0) = f(-15) = 3/8$
   b)

$$F(x) = \begin{array}{rl} 0, & -\infty < x < -30 \\ 1/8, & -30 \leq x < -15 \\ 1/2, & -15 \leq x < 0 \\ 7/8, & 0 \leq x < 15 \\ 1, & 15 \leq x < \infty \end{array}$$

8.

$$A \int_0^a 4x - 2x^2 dx = A2a^2(1 - a/3)$$

a) $a = 2$. Om A=3/8 så utgör $f(x)$ en täthetsfunktion ty $f(x) \geq 0$ för alla x och $\int f(x)dx) = 1$.
b) $a = 3$. $f(x)$ kan aldrig vara en täthetsfunktion ty $A \int_0^a 4x - 2x^2 dx = 0$.
c) $a = 4$. Om $A > 0$ så $f(x) < 0$ för $x > 2$ och om $A < 0$ så $f(x) < 0$ för $x < 2$, alltså kan aldrig A väljas så att $f(x)$ blir en täthetsfunktion.

9. a) p-värde= P(att man får ett minst lika extremt utfall som det observerade $|H_0$ är sann). Det används för att besluta om man ska förkasta H_0 eller ej. Ju mindre p-värde desto större större skäl har man att förkasta H_0. Vanliga gränser då man förkastar är 0.01 och 0.05, men det beror på flera saker, tex vilka konsekvenser det kan få om man förkastar felaktigt.

   b) Ett $100(1 - \alpha)\%$ konfidensintervall för en paramter θ är ett intervall $[L_1, L_2]$ baserat på ett stickprov sådant att $P(L_1 \leq \theta \leq Ł_2) = 1 - \alpha$.

   Det är alltså ett osäkerhetsintervall för en skattning. Om man tar nya stickprov 100 gånger och varje gång konstruerar tex ett 95% konfidensintervall kommer ca 95 st av de 100 intervallen att innehålla det sanna värdet på parametern θ.

10. a) X = antal bilar under 2 minuter, $X \sim Po(2 \cdot 4)$.

$$P(X = 7) = e^{-8} 8^7 / 7! \approx 0.140$$

   b) X_i = antal bilar minut i $\sim Po(4)$, oberoende. CGS ger att $(\sum_{i=1}^{90} X_i - 360)/\sqrt{360}$ approx N(0,1).

$$P(\sum_{i=1}^{90} X_i \geq 300) \approx 1 - \Phi(-3.11) = 0.9991.$$