Matematisk statistik
Matematikcentrum
Lunds tekniska högskola
Lunds universitet

Lösningsförslag dugga 2009–10–16 kl 8^{00}–10^{00}
FMS 012 — Matematisk statistik för CDI, 9 hp

1. Om vi inför händelserna M = "Ett äpple har mask" och B = "Ett äpple har skorv" så har vi enligt uppgift följande sannolikheter

$$P(M) = 0.3, \qquad P(S) = 0.4, \qquad P(M|S) = 0.6.$$

(a) Sannolikheten att valt äpple är friskt fås med, i tur och ordning, komplement, additionssatsen och definitionen av betingad sannolikhet till

$$P(\text{friskt}) = 1 - P(\text{sjukt}) = 1 - P(M \cup S) = 1 - [P(M) + P(S) - P(M \cap S)] =$$
$$= 1 - [P(M) + P(S) - P(M \mid S)P(S)] = 1 - [0.3 + 0.4 - 0.6 \cdot 0.4] = \underline{0.54}$$

(b) Om vi låter X vara antalet maskätna äpplen av de sex utvalda (bland mååånga) är det rimligt att anta att $X \in Bin(n, p)$ där $n = 6$ och $p = 0.3$. Den sökta sannolikheten blir då

$$P(X \geq 3) = 1 - P(X \leq 2) = \sum_{k=0}^{2} p_X(k) = 1 - \sum_{k=0}^{2} \binom{6}{k} 0.3^k \cdot 0.7^{6-k} = \underline{0.2557}.$$

Alternativt kan man använda fördelningsfunktionen i tabell 6 ($n = 6$, $p = 0.30$, $x = 2$) och får $1 - P(X \leq 2) = 1 - F_X(2) = 1 - 0.74431 = \underline{0.2557}$.

2. (a) Sannolikheten att X är minst 2 fås med en partiell integration till

$$P(X \geq 2) = \int_2^\infty x\, e^{-x} dx = \left[-xe^{-x}\right]_2^\infty + \int_2^\infty e^{-x} dx = 2e^{-2} + \left[-e^{-x}\right]_2^\infty =$$
$$= 2e^{-2} + e^{-2} = 3e^{-2} \approx \underline{0.406}.$$

(b) Eftersom den största av de tre oberoende och likafördelade variablerna är mindre än 2 bara då alla tre är mindre än två fås

$$P(\max(X_1, X_2, X_3) < 2) = P(X_1 < 2, X_2 < 2, X_3 < 2) = [\text{oberoende}] =$$
$$= P(X < 2)^3 = (1 - 0.406)^3 \approx \underline{0.2096}.$$

3. (a) Den marginella sannolikhetsfunktionen för X resp. Y fås ur

$$p_X(j) = \sum_k p_{X,Y}(j,k), \qquad p_Y(k) = \sum_j p_{X,Y}(j,k)$$

dvs summor radvis respektive kolonnvis i den simultana sannolikhetsfunktionen. Vi kan sätta de endimensionella fördelningarna i marginalerna till den tvådimensionella.

$j \backslash k$	0	1	2	$p_X(j)$
0	0.03	0.18	0.09	0.30
1	0.02	0.12	0.06	0.20
2	0.05	0.30	0.15	0.50
$p_Y(k)$	0.10	0.60	0.30	

(b) X väntevärde och varians blir

$$E(X) = \sum_j j\, p_X(j) = 0 \cdot 0.30 + 1 \cdot 0.20 + 2 \cdot 0.50 = \underline{1.20}$$

$$E(X^2) = \sum_j j^2\, p_X(j) = 0^2 \cdot 0.30 + 1^2 \cdot 0.20 + 2^2 \cdot 0.50 = 2.20$$

$$V(X) = E(X^2) - E(X)^2 = 2.20 - 1.20^2 = \underline{0.76}$$

Om man studerar tabellen med den simultana och de marginella sannolikhetsfunktionerna ser vi att elementen i tabellen hela tiden är produkten av motsvarande marginalelement, dvs X och Y är oberoende av varandra. Därför är $\underline{C(X,Y) = 0}$. Missade man det kan man naturligtvis räkna ut kovariansen. Vi behöver komplettera med $E(Y)$ och $E(XY)$.

$$E(Y) = \sum_k k\, p_Y(k) = 0 \cdot 0.10 + 1 \cdot 0.60 + 2 \cdot 0.30 = 1.20$$

$$E(XY) = \sum_{j,k} j\,k\, p_{X,Y}(j,k) = \sum_{j=0}^{2}\sum_{k=0}^{2} j\,k\, p_{X,Y}(j,k) = [\text{termerna blir 0 då } j \text{ eller } k \text{ är 0}] =$$

$$= \sum_{j=1}^{2}\sum_{k=1}^{2} j\,k\, p_{X,Y}(j,k) = 1 \cdot 1 \cdot 0.12 + 1 \cdot 2 \cdot 0.06 + 2 \cdot 1 \cdot 0.30 + 2 \cdot 2 \cdot 0.15 = 1.44$$

$$C(X,Y) = E(XY) - E(X)E(Y) = 1.44 - 1.20 \cdot 1.20 = \underline{0}$$

4. (a) Av formelsamligen framgår att, för en exponentialfördelning, så gäller det att variansen är lika med kvadraten på väntevärdet ($E = 1/\lambda$ och $V = 1/\lambda^2$), dvs

$$V(X_1) = 2^2 = 4,\ V(X_2) = 3^2 = 9 \text{ och } V(X_3) = 6^2 = 36.$$

Sätt $U = X_1 + X_2 + X_3 =$ ”sammanlagda tiden för en kund hos lager A. Då har vi att

$$E(U) = E(X_1 + X_2 + X_3) = E(X_1) + E(X_2) + E(X_3) = 2 + 3 + 6 = \underline{11 \text{ minuter}}.$$

Vi får också att

$$V(U) = V(X_1 + X_2 + X_3) = V(X_1) + V(X_2) + V(X_3) = 4 + 9 + 36 = 49$$

och standardavvikelsen

$$D(U) = \sqrt{V(U)} = \sqrt{49} = \underline{7 \text{ minuter}}.$$

(b) Eftersom $n = 100$ är stort gäller, enligt CGS, att $V_a = \sum_{i=1}^{100} U_i =$ ”sammanlagda tiden för 100 kunder hos lager A” $\underset{\sim}{\in} N\left(\sum_{i=1}^{100} E(U_i), \sqrt{\sum_{i=1}^{100} V(U_i)}\right) = N(100 \cdot 11, \sqrt{100 \cdot 49}) =$ $N(1100, 70)$.

På samma sätt får vi att $V_b = \sum_{i=1}^{100} W_i =$ ”sammanlagda tiden för 100 kunder hos lager B” $\underset{\sim}{\in} N\left(\sum_{i=1}^{100} E(W_i), \sqrt{\sum_{i=1}^{100} V(W_i)}\right) = N(100 \cdot 10, \sqrt{100 \cdot 6^2}) = N(1000, 60)$.

Vi vill beräkna $P(V_a < V_b) = P(V_a - V_b < 0)$ men eftersom V_a och V_b är oberoende och approximativt normalfördelade får vi också att
$V_a - V_b \underset{\sim}{\in} N(E(V_a) - E(V_b), \sqrt{V(V_a) + (-1)^2 \cdot V(V_b)}) =$
$= N(1100 - 1000, \sqrt{70^2 + 60^2}) = N(100, \sqrt{8500})$ så att $P(V_a - V_b < 0) = \Phi(\frac{0-100}{\sqrt{8500}}) =$
$1 - \Phi(\frac{100}{\sqrt{8500}}) = 1 - \Phi(1.08) = 1 - 0.8610 = \underline{0.139}$.