Matematisk statistik
Matematikcentrum
Lunds tekniska högskola, Lunds universitet

Kortfattade lösningar till tentamen: 2012–12–18 kl 8^{00}–13^{00}
FMS 012 — Matematisk statistik för PiE, F, CDI, 9 hp
MAS B03 — Matematisk statistik för fysiker, 9 hp

1. (a) $E(3X + 2Y) = 3E(X) + 2E(Y) = 3 \cdot 1 + 2 \cdot 2 = 7$.
$V(3X+2Y) = 3^2V(X)+2^2V(Y)+2\cdot 3\cdot 2C(X, Y)$. Men eftersom $C(X, Y) = \rho(X, Y)D(X)D(Y) = -0.5\sqrt{5}\sqrt{4} = -\sqrt{5}$ blir $V(3X + 2Y) = 9 \cdot 5 + 4 \cdot 4 - 12\sqrt{5} = 61 - 12\sqrt{5} \approx 34.167$.

   (b) X= mängden aktiv substans $\in N(2, \sigma)$. Bestäm σ så att $0.05 \geq P(X \leq 1.9)$, vilket ger $P(X \leq 1.9) = \Phi(\frac{1.9-2}{\sigma}) \leq 0.05$ och $\frac{1.9-2}{\sigma} \leq -\lambda_{0.05} = -1.645$ med $\sigma \leq \frac{0.1}{1.645} = 0.061$.

   (c) $p_Y(k) = P(Y = k) = P([X] = k) = P(k \leq X < k + 1) = \int_k^{k+1}(x + 1)^{-2}dx = [-(x + 1)^{-1}]_k^{k+1} = \frac{1}{(k+1)(k+2)},\ k = 0, 1, 2, \ldots$.

2. Bilda differenser före-efter för de 8 olika vinsorterna:

   z_i: 0.27 0.38 0.26 0.04 0.02 0.23 0.16 0.19

   Modell: $z_1, \ldots, z_8$ observationer av Z=skillnad i vinsyra (före-efter behandling); $Z \in N(\Delta, \sigma)$.

   Intressanta hypoteser: $H_0 : \Delta = 0$ (ingen förändring i vinsyra); $H_1 : \Delta > 0$ (behandlingen sänker vinsyran).

   Från data fås: $\Delta^* = \bar{z} = 0.1937$ och $\sigma^* = s = \sqrt{\frac{1}{8-1}\sum_{i=1}^{8}(z_i - \bar{z})^2} = 0.1202$.

   Beräkna ett ensidigt, nedåt begränsat intervall för $\Delta : I_\Delta = (\bar{z} - t_{0.05}(8 - 1)\frac{s}{\sqrt{8}},\ \infty) = (0.11,\ \infty)$.

   Eftersom intervallet ej täcker över 0, förkastas H_0. Ja, det tycks som om behandlingen sänker vinsyran.

3. $P(X > Y) = \int\int_{x>y, x\geq 0, y\geq 0} 2e^{-x}e^{-2y}dxdy =$ [rita figur!] $= 2\int_0^\infty e^{-x}(\int_0^x e^{-2y}dy)dx = 2\int_0^\infty e^{-x}[-\frac{1}{2}e^{-2y}]_0^x dx =$

   $\int_0^\infty e^{-x}(1 - e^{-2x})dx = [-e^{-x} + \frac{1}{3}e^{-3x}]_0^\infty = 1 - \frac{1}{3} = \frac{2}{3}$.

4. (a) $X =$ "antalet glassar som säljs en viss dag" med $(X \mid \text{mulet}) = (X \mid E_2) \in Po(10)$.
$$\mathsf{P}(X = 0 \mid E_2) = \mathrm{e}^{-10}\,\frac{10^0}{0!} = \mathrm{e}^{-10} \approx 4.54 \cdot 10^{-5}.$$

   (b) Vi har dessutom att $(X \mid E_1) \in Po(30)$ med $\mathsf{P}(X = 0 \mid E_1) = \mathrm{e}^{-30} \approx 9.4 \cdot 10^{-14}$ och $(X \mid E_3) \in Po(2)$ med $\mathsf{P}(X = 0 \mid E_3) = \mathrm{e}^{-2} = 0.1353$.
Vi har att $\mathsf{P}(E_2 \to E_1) = 0.2$, $\mathsf{P}(E_2 \to E_2) = 0.4$ och $\mathsf{P}(E_2 \to E_3) = 0.4$ så att
$$\mathsf{P}(X = 0 \text{ i morgon} \mid \text{mulet idag}) = \sum_{i=1}^{3}\mathsf{P}(X = 0 \mid E_i) \cdot \mathsf{P}(E_1 \to E_i) =$$
$$= 9.4 \cdot 10^{-14} \cdot 0.2 + 4.54 \cdot 10^{-5} \cdot 0.4 + 0.1353 \cdot 0.4 \approx 0.0542.$$

   (c) Lös $\pi = \pi P$ där $\sum_{i=1}^{3}\pi_i = 1$, dvs
$$\begin{cases} \pi_1 = 0.7\pi_1 + 0.2\pi_2 \\ \pi_2 = 0.3\pi_1 + 0.4\pi_2 + 0.5\pi_3 \\ \pi_3 = 0.4\pi_2 + 0.5\pi_3 \\ 1 = \pi_1 + \pi_2 + \pi_3 \end{cases} \Rightarrow \pi = \left(\frac{10}{37}, \frac{15}{37}, \frac{12}{37}\right)$$

   (d) Med "mot slutet av sommaren" menas att den asymptotiska fördelningen gäller, och
$$\mathsf{E}(X) = \sum_{i=1}^{3}\mathsf{E}(X \mid E_i) \cdot \pi_i = 30 \cdot \frac{10}{37} + 10 \cdot \frac{15}{37} + 2 \cdot \frac{12}{37} \approx 12.81 \text{ glassar.}$$

5. (a) Hypoteserna $H_0 : \beta_i = 0; H_1 : \beta_i \neq 0$ kan testas på nivå 0.05 genom att göra ett 95% konfidensintervall för β_i: $I_{\beta_i} = (\beta_i^* \pm t_{0.025}(28 - 4)d(\beta_i^*))$. Om I_{β_i} ej täcker 0 förkastas H_0 och den förklarande variabeln x_i påverkar y och bör därmed vara med i modellen.
För de aktuella parametrarna fås:
$I_{\beta_1} = (-0.0613 \pm 2.06 \cdot 0.2644) = (-0.6060,\ 0.4834)$
$I_{\beta_2} = (0.9390 \pm 2.06 \cdot 0.1504) = (0.6292,\ 1.2488)$
$I_{\beta_3} = (-0.0328 \pm 2.06 \cdot 0.1908) = (-0.4258,\ 0.3602)$
Det är enbart parameter β_2 som är signifikant skilt från 0.

(b) Skatta σ_2^2 med $s^2 = \frac{Q_0}{28-2} = \frac{253.09}{26} = 9.7342$, d.v.s. $s = \sigma_2^* = 3.120$. Antalet frihetsgrader för denna skattning är $28 - 2 = 26$.

Vi söker ett intervall för $\mu_0 = \beta_0 + 3\beta_2$.

$\mu_0^* = \beta_0^* + 3\beta_2^* = -0.4258 + 3 \cdot 0.9359 = 2.3819$.

$V(\mu_0^*) = V(\beta_0^* + 3\beta_2^*) = V(\beta_0^*) + 3^2 V(\beta_2^*) + 2 \cdot 3 \cdot C(\beta_0^*, \beta_2^*) =$

$0.0364 \cdot \sigma_2^2 + 9 \cdot 0.0021 \cdot \sigma_2^2 + 2 \cdot 3 \cdot (-0.0012)\sigma_2^2 = 0.0481\sigma_2^2$, där varianser och kovarians är hämtad från matrisen $(X^T X)^{-1}$.

$d(\mu_0^*) = \sigma_2^* \sqrt{0.0481} = 3.120\sqrt{0.0481} = 0.6843$.

$I_{\mu_0} = (\mu_0^* \pm t_{0.025}(26) d(\mu_0^*)) = (2.3819 \pm 2.06 \cdot 0.6843) = (0.972,\ 3.791)$.

Alternativt arbetar vi direkt med matrisformuleringen: Låt $\mathbf{x_0} = (1\ 3)$ och $\boldsymbol{\beta}^T = (\beta_0\ \beta_2)$. Då är $\mu_0 = \mathbf{x_0}\boldsymbol{\beta}$ och intervallet fås genom

$I_{\mu_0} = (\mathbf{x_0}\boldsymbol{\beta}^* \pm t_{0.025}(26)\sigma_2^* \sqrt{\mathbf{x_0}(X^T X)^{-1}\mathbf{x_0}^T}) = (2.3819 \pm 2.06 \cdot 3.120\sqrt{0.0481}) = (0.972,\ 3.791)$.

(c) Vi vill skatta $Y_{x_2=4} - Y_{x_2=2} = (\beta_0 + \beta_2 \cdot 4) - (\beta_0 + \beta_2 \cdot 2) = 2\beta_2$. Från tabellen fås $\beta_2^* = 0.9359$ så sökt punktskattning är $2 \cdot 0.9359 = 1.8718$.

6. (a) $F(x) = P(X \leq x) = 1 - e^{-\frac{1}{a}x^c}; f(x) = F'(x) = \frac{c}{a}x^{c-1}e^{-\frac{1}{a}x^c}, x \geq 0$.

(b) Då $c = 2$ blir $f(x) = \frac{2}{a}xe^{-\frac{1}{a}x^2}, x \geq 0$.

Antag att n observationer av X görs (i uppgiften är $n = 5$).

$L(a) = \Pi_{i=1}^n \frac{2x_i}{a}e^{-\frac{x_i^2}{a}} = \frac{2^n}{a^n}e^{-\frac{1}{a}\sum_{i=1}^n x_i^2} \cdot \Pi_{i=1}^n x_i$.

$\ln L(a) = -n \cdot \ln a - \frac{1}{a}\sum_{i=1}^n x_i^2 + \sum_{i=1}^n \ln x_i$.

$\frac{d \ln L(a)}{da} = -\frac{n}{a} + \frac{1}{a^2}\sum_{i=1}^n x_i^2 = \frac{1}{a}(-n + \frac{1}{a}\sum_{i=1}^n x_i^2)$ där $a > 0$.

$\frac{d \ln L(a)}{da} = 0 \Rightarrow \frac{1}{a}\sum_{i=1}^n x_i^2 = n \Leftrightarrow a = \frac{1}{n}\sum_{i=1}^n x_i^2$.

Eftersom a maximerar $L(a)$ (behöver ej visas) väljer vi ML-skattningen $a_{ML}^* = \frac{1}{n}\sum_{i=1}^n x_i^2 = \frac{1}{5} \cdot 1.7233 = 0.345$.

(c) Bestäm percentilen $L_{0.1}$ så att $0.1 = P(X \leq L_{0.1}) = F(L_{0.1}) = 1 - e^{-\frac{1}{a}L_{0.1}^2}$. Detta ger $L_{0.1} = \sqrt{-a \cdot \ln 0.9}$ som med skattningen $a_{ML}^* = 0.345$ ger att $L_{0.1}$ uppskattas till $\sqrt{-0.345 \cdot \ln 0.9} = 0.191$.

(d) I uppgiften behövs $E(X^2) = V(X) + E(X)^2 = \frac{(4-\pi)a}{4} + \frac{\pi a}{4} = a$.

$E(a^*) = E(\frac{1}{n}\sum_{i=1}^n X_i^2) = \frac{1}{n}\sum_{i=1}^n E(X_i^2) = \frac{1}{n}\sum_{i=1}^n a = \frac{a \cdot n}{n} = a$.

Ja, skattningen är väntevärdesriktig.