UPPSALA UNIVERSITET
Matematiska institutionen
G. Berg, J. Björnberg, J. Rydén

TENTAMEN I MATEMATISK STATISTIK
Sannolikhet och statistik 1MS005
2012-01-10

LÖSNINGSFÖRSLAG

1. Inför händelserna H_1 =”Möjligt resa A till B” samt H_2 =”Omöjligt ta sig mellan A och C”. Sökt sannolikhet:

$$p_0 = \mathsf{P}(H_1 \,|\, H_2) = \frac{\mathsf{P}(H_1 \cap H_2)}{\mathsf{P}(H_2)}.$$

Inför händelserna S_1 =”Samtliga vägar AB igensnöade” samt S_2 =”Samtliga vägar BC igensnöade”. Då följer

$$\mathsf{P}(H_2) = \mathsf{P}(S_1 \cup S_2) = \mathsf{P}(S_1) + \mathsf{P}(S_2) - \mathsf{P}(S_1 \cap S_2) = p^2 + p^2 - (p^2)^2 = 2p^2 - p^4.$$

Händelsen $H_1 \cap H_2$ kan formuleras i ord som ”Omöjligt att ta sig mellan A och C, *samtidigt som* minst en av vägarna AB är snöfri”. Oberoende ger

$$\mathsf{P}(H_1 \cap H_2) = p^2(1 - p^2)$$

där den första faktorn följer av att bägge vägarna BC måste vara igensnöade, den andra med komplementresonemang. Sammanfattningsvis:

$$p_0 = \frac{p^2(1-p^2)}{2p^2 - p^4} = \frac{1-p^2}{2-p^2}.$$

2. (a) För täthetsfunktionen skall gälla

$$\int_0^2 f_(x)\,\mathrm{d}x = 1.$$

Vi finner här

$$1 = \int_0^2 a(2x - x^2)\,\mathrm{d}x = a\left[x^2 - \frac{x^3}{3}\right]_0^2 = \frac{4}{3}a,$$

dvs. $a = 3/4$.

(b) Väntevärde:

$$\mathsf{E}[X] = \int_0^2 x f(x)\,\mathrm{d}x = a\int_0^2 2x^2 - x^3\,\mathrm{d}x = 1.$$

Eftersom

$$\mathsf{E}[X^2] = \int_0^2 x^2 f(x)\,\mathrm{d}x = a\int_0^2 2x^3 - x^4\,\mathrm{d}x = \frac{8a}{5} = \frac{6}{5}$$

följer variansen som

$$\mathsf{V}[X] = \mathsf{E}[X^2] - (\mathsf{E}[X])^2 = \frac{6}{5} - 1^2 = \frac{1}{5}.$$

(c) Den sökta sannolikheten ges av

$$\mathsf{P}(X > 0.5) = 1 - \mathsf{P}(X \le 0.5) = 1 - a\int_0^{0.5} 2x - x^2\,\mathrm{d}x = \frac{27}{32} = 0.84375.$$

3. Vi har att $X+Y+Z \sim N(-0.2, 1.1)$, så $Z' \sim N(0,1)$ där
$$Z' = \frac{X+Y+Z+0.2}{\sqrt{1.1}}.$$
Så
$$0.025 = P(Z' > 1.96) = P((X+Y+Z)/3 > (\sqrt{1.1}\cdot 1.96 - 0.2)/3).$$
Svaret är alltså $(\sqrt{1.1}\cdot 1.96 - 0.2)/3$.

4. Inför slumpvariabeln X =”Tid för putsning av ett par skor”, med $\mathsf{E}[X] = 10$, $\mathsf{V}[X] = 3^2 = 9$ (min). Sökt sannolikhet: $\mathsf{P}(X_1 + \cdots + X_{30} \leq 290)$. Inför
$$Y = X_1 + \cdots + X_{30}.$$
Enligt centrala gränsvärdessatsen gäller approximativt (30 termer anses vara ett stort antal) att $Y \sim \mathrm{N}(\mu_Y, \sigma_Y^2)$ där
$$\mu_Y = 30\cdot 10 = 300, \quad \sigma_Y^2 = 30\cdot 9 = 270.$$
(Här antas oberoende tider.) Den sökta sannolikheten ges av
$$\mathsf{P}(Y \leq 290) = \mathsf{P}((Y-300)/\sqrt{270} \leq (290-300)/\sqrt{270}) \doteq 1 - \Phi(0.61) \doteq 0.27.$$

5. Ett intervall med approximativ konfidensgrad kan beräknas om stickprovet är tillräckligt stort (motiveras av centrala gränsvärdessatsen). Detta ges då av
$$I_\mu = [\bar{x} \pm \lambda_{\alpha/2} s/\sqrt{n}].$$
Vi anser här att 200 observationer är ett stort stickprov och finner från den givna informationen $\bar{x} = 2.49$, $s = 1.08$. Ett konfidensintervall med approximativ konfidensgrad 95% följer som
$$I_\mu = [2.49 \pm 1.96\cdot 1.08/\sqrt{200}] \doteq [2.34,\ 2.64].$$

6. Från det sedvanliga intervallet med approximativ konfidensgrad $(1-\alpha)$ för andelen p
$$p^* \pm \lambda_{\alpha/2}\sqrt{p^*(1-p^*)/n}$$
finner man
$$n = \frac{\lambda_{\alpha/2}^2 p^*(1-p^*)}{d^2}$$
där $2d$ är intervallets längd.

   I detta problem har vi $p^* = 0.89$, $d = 0.01$ samt $\lambda_{0.025} = 1.96$. Antalet observationer:
$$n = \frac{(1.96)^2\cdot 0.89\cdot 0.11}{(0.01)^2} = 3760.$$

7. (a) Modell: X = ”Uppmätt överhöjning hos element från fabrik A”, Y =”Uppmätt överhöjning hos element från fabrik A”, enligt uppgift gäller $X \sim \mathrm{N}(\mu_A, \sigma^2)$, $Y \sim \mathrm{N}(\mu_B, \sigma^2)$ (lika standardavvikelser, okända). Ett 99%-konfidensintervall för $\mu_A - \mu_B$ ges av
$$\bar{x} - \bar{y} \pm t_{0.005}(9+16-2) s_P \sqrt{\frac{1}{9} + \frac{1}{16}}$$
Med $\bar{x}$, $\bar{y}$, s_1, s_2 givna i texten återstår att beräkna den poolade standardavvikelsen
$$s_P = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = \sqrt{41.572} = 6.4476$$
och med $t_{0.005}(23) = 2.8073$ följer intervallet $[-4.04,\ 11.0]$.

(b) Intervallet täcker noll, vi kan inte på nivån 0.99 uttala oss om att det finns någon skillnad mellan överhöjningarna fabrikerna emellan.

8. (a) $\beta^* = S_{xy}/S_{xx} = -0.4833$, $\alpha^* = \bar{y} - \bar{x}\beta^* = 1.7889$, där vi utnyttjat $\bar{x} = 0$, $\bar{y} = 1.7889$.

(b) En skattning av σ^2 ges av

$$s^2 = \frac{1}{7}(S_{yy} - S_{xy}^2/S_{xx}) = 0.03532$$

och medelfelet för β ges av $\mathsf{d}[\beta] = s/\sqrt{S_{xx}} = 0.07672$. Ett konfidensintervall följer som

$$I_\beta = [\beta^* \pm t_{0.025}(7)\mathsf{d}[\beta]] = [-0.66,\ -0.30].$$