1. (a) Vi har $\mathsf{P}(A \cap B) = \mathsf{P}(A) + \mathsf{P}(B) - \mathsf{P}(A \cup B) = 0.2 + 0.3 - 0.4 = 0.1$ och

$$\mathsf{P}(A|B) = \frac{\mathsf{P}(A \cap B)}{\mathsf{P}(B)} = \frac{0.1}{0.3} = \frac{1}{3}$$

(b) Låt S beteckna händelsen att en telefon har en sprucken skärm och låt T_A, T_B respektive T_C beteckna händelserna att en telefon är av märke A, B respektive C. Vi vill beräkna $\mathsf{P}(T_A|S)$. Satsen om total sannolikhet ger

$$\begin{aligned}\mathsf{P}(S) &= \mathsf{P}(S|T_A)\mathsf{P}(T_A) + \mathsf{P}(S|T_B)\mathsf{P}(T_B) + \mathsf{P}(S|T_C)\mathsf{P}(T_C) \\ &= 0.2 \cdot 0.1 + 0.3 \cdot 0.02 + 0.5 \cdot 0.04 = 0.046\end{aligned}$$

Bayes sats ger

$$\mathsf{P}(T_A|S) = \frac{\mathsf{P}(S|T_A)\mathsf{P}(T_A)}{\mathsf{P}(S)} = \frac{0.2 \cdot 0.1}{0.046} \approx 0.435$$

2. Låt X_i och Y_i vara reaktionstiderna för patient i före respektive efter alkoholintag. Bilda differensen $D_i = Y_i - X_i$

Person i	1	2	3	4	5	6	7	8
D_i	0.07	0.03	0.03	0.04	0.05	0.04	0.03	0.06

och ansätt modellen $D_i \sim \mathsf{N}(\Delta, \sigma^2)$ där vi antar att alla D_i är oberoende. Skattningen av skillnad i reaktionstid ges nu av $\bar{D} = 0.35/8 = 0.0437$ och vi skattar standardavvikelsen σ med $s = 0.0151$. Vi vill testa $H_0 : \Delta = 0$ mot $H_1 : \Delta \neq 0$. Vi skriver upp teststorheten $T_{obs} = \bar{D}/s = 8.21$. Vi ser från tabell att $T_{obs} > t_{\alpha/2}(7)$ för $\alpha = 0.05$ och kan därför förkasta H_0 på nivå $\alpha = 0.05$. Alternativt bildar vi först konfidensintervallet

$$I_\Delta = \left[\bar{\Delta} \pm t_{\alpha/2}(7)\frac{s}{\sqrt{n}}\right] = [0.0437 \pm 2.3646 \cdot 0.0053] = [0.0312, 0.0563],$$

och ser att vi kan förkasta H_0 eftersom intervallet inte täcker 0.

3. (a) Vi har

$$f_X(x) = \int_0^1 \left(\frac{5}{4} - xy\right) dy = \left[\frac{5y}{4} - \frac{xy^2}{2}\right]_0^1 = \frac{5}{4} - \frac{x}{2}$$

$$f_Y(y) = \int_0^1 \left(\frac{5}{4} - xy\right) dx = \left[\frac{5x}{4} - \frac{x^2y}{2}\right]_0^1 = \frac{5}{4} - \frac{y}{2}$$

(b) Vi vill beräkna $C(X,Y) = \mathsf{E}(XY) - \mathsf{E}(X)\mathsf{E}(Y)$. Vi har

$$\mathsf{E}(X) = \int_0^1 x\left(\frac{5}{4} - \frac{x}{2}\right) dx = \left[\frac{5}{8}x^2 - \frac{x^3}{6}\right]_0^1 = \frac{5}{8} - \frac{1}{6} = \frac{11}{24}$$

På grund av symmetri är $\mathsf{E}(Y) = \mathsf{E}(X)$ och vi har

$$\begin{aligned}\mathsf{E}(XY) &= \int_0^1\int_0^1 xy\left(\frac{5}{4} - xy\right) dydx = \int_0^1\int_0^1 x\left(\frac{5}{8} - \frac{x}{3}\right) dx \\ &= \left[\frac{5}{16}x^2 - \frac{x^3}{9}\right]_0^1 = \frac{5}{16} - \frac{1}{9} = \frac{29}{144}\end{aligned}$$

och vi får alltså

$$C(X,Y) = \frac{29}{144} - \frac{11}{24}\frac{11}{24} = -\frac{5}{576}$$

(c) Eftersom $C(X, Y) \neq 0$ kan variablerna inte vara oberoende.

4. Antalet kunder under två timmar följer en Poissonfördelning, $\mathrm{Po}(2\lambda)$, och vi vill testa

$$H_0 : \lambda = 2 \quad \text{mot} \quad H_1 : \lambda > 2$$

Vi väljer $\alpha = 0.05$ för testet. Direktmetoden ger p-värdet

$$\begin{aligned}\mathsf{P}(\text{att få det vi fick eller värre}|H_0) &= \mathsf{P}(X \geq 7|\lambda = 2)\\ &= 1 - \mathsf{P}(x \leq 6|\lambda = 2) = 1 - \sum_{i=0}^{6} e^{-4}\frac{4^k}{k!} = 0.11\end{aligned}$$

Eftersom p-värdet är större än α kan vi inte förkasta, och inte ens om vi gjorde testet på nivå $\alpha = 0.1$.

5. (a) Låt x_i vara avståndet till jordbruket för observation i och låt y_i vara den motsvarande nitrathalten. Vi ansätter modellen $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, där $\varepsilon_i \sim \mathsf{N}(0, \sigma^2)$ är oberoende. Vi har $S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 = 108976, S_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = 64.4, S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = -2426$ och skattningarna ges av $\beta_1^* = \frac{S_{xy}}{S_{xx}} = -0.022, \beta_0^* = \bar{y} - \beta_1^*\bar{x} = 48.641, s^2 = (S_{yy} - \frac{S_{xy}^2}{S_{xx}})/(n-2) = 12.99$.

(b) Vi söker ett konfidensintervall för $100\beta_1$, vilket ges av

$$I_{100\beta_1} = \left(100\beta_1^* \pm t_{\alpha/2}(n-2)\frac{100s}{\sqrt{S_{xx}}}\right) = (-3.02, -1.43)$$

där vi valt $\alpha = 0.05$

(c) Vi söker prediktionsintervallet för $Y(700)$, vilket ges av

$$I_{Y(700)} = \left(\beta_0^* + \beta_1^* 700 \pm t_{\alpha/2}(n-2)s\sqrt{1 + \frac{1}{n} + \frac{(700 - \bar{x})^2}{S_{xx}}}\right) = (24.16, 41.95)$$

6. (a) Log-likelihood funktionen är

$$l(\theta) = \log L(\theta) = n \log \theta + \sum_{i=1}^{n}(x_i - 1)\log(1 - \theta)$$

För att maximera $l(\theta)$ sätter vi dess derivata till noll:

$$\begin{aligned} 0 &= \frac{\partial\, l(\theta)}{\partial\, \theta} = n\frac{1}{\theta} - \sum_{i=1}^{n}(x_i - 1)\frac{1}{1-\theta} \Rightarrow\\ 0 &= (1-\theta)n - \theta\sum_{i=1}^{n}(x_i - 1) \Rightarrow\\ \theta &= \frac{n}{\sum_{i=1}^{n} x_i}\end{aligned}$$

Alltså är ML-skattaren $\hat{\theta} = \frac{n}{\sum_{i=1}^{n} x_i}$.

(b) Skattaren är väntevärdesriktig eftersom väntevärdet är

$$\mathsf{E}\left(\frac{1}{\hat{\theta}}\right) = \mathsf{E}\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathsf{E}(x_i) = \frac{1}{n}n\frac{1}{\theta} = \frac{1}{\theta}$$

Variansen ges av

$$\mathsf{V}\left(\frac{1}{\hat{\theta}}\right) = \mathsf{V}\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathsf{V}(x_i) = \frac{1}{n}\frac{1-\theta}{\theta^2}$$

(d) Enligt CGS är $1/\hat{\theta}$ approximativt normalfördelad, $\mathsf{N}(\frac{1}{\theta}, \frac{1}{n}\frac{1-\theta}{\theta^2})$, vi använder detta och får ett konfidensintervall som

$$I_{\frac{1}{\theta}} = \left[\frac{1}{\hat{\theta}} \pm z_{\alpha/2}\sqrt{\mathsf{V}\left(\frac{1}{\hat{\theta}}\right)}\right] \approx \left[\frac{1}{\hat{\theta}} \pm z_{\alpha/2}\frac{\sqrt{1-\hat{\theta}}}{\hat{\theta}\sqrt{n}}\right]$$