Solutions Exercise 1

1(a): Compute the log likelihood:

$$\begin{aligned} l_X(\theta) &= \log\left(\prod_{i=1}^{n}\binom{x_i+r-1}{x_i}\cdot(1-\theta)^r\cdot\theta^{x_i}\right) \\ &= \log\left((\prod_{i=1}^{n}\binom{x_i+r-1}{x_i})\cdot(1-\theta)^{nr}\cdot\theta^{\sum_{i=1}^n x_i}\right) \\ &= \log\left(\prod_{i=1}^{n}\binom{x_i+r-1}{x_i}\right)+nr\log(1-\theta)+(\sum_{i=1}^n x_i)\log(\theta) \end{aligned}$$

Take the derivative w.r.t. θ and set it to 0:

$$\frac{-nr}{1-\theta}+\frac{\sum_{i=1}^n x_i}{\theta}=0 \;\Leftrightarrow\; -nr\theta+(\sum_{i=1}^n x_i)(1-\theta)=0 \Leftrightarrow -(nr+\sum_{i=1}^n x_i)\theta+\sum_{i=1}^n x_i=0$$

$$\Leftrightarrow\; \theta=\frac{\sum_{i=1}^n x_i}{nr+\sum_{i=1}^n x_i}\Rightarrow\hat{\theta}_{ML}=\frac{\bar{X}}{r+\bar{X}}$$

1(b): For $n=1$ we have: $\frac{d^2}{d\theta^2}l_{X_1}(\theta)=\frac{-r}{(1-\theta)^2}-\frac{X_1}{\theta^2}$, and the Fisher information is:

$$\begin{aligned} I(\theta) &= -E_\theta\left[\frac{d^2}{d\theta^2}l_{X_1}(\theta)\right]=E_\theta\left[\frac{r}{(1-\theta)^2}+\frac{X_1}{\theta^2}\right]=\frac{r}{(1-\theta)^2}+\frac{E[X_1]}{\theta^2} \\ &= \frac{r}{(1-\theta)^2}+\frac{\frac{r\theta}{(1-\theta)}}{\theta^2}=\frac{r}{(1-\theta)^2}+\frac{r\theta}{(1-\theta)\theta^2}=\frac{r\theta+r(1-\theta)}{(1-\theta)^2\theta}=\frac{r}{\theta(1-\theta)^2} \end{aligned}$$

1(c): Asymptotically $\sqrt{I(\theta)}\sqrt{n}\cdot(\hat{\theta}_{ML}-\theta)\sim\mathcal{N}(0,1)$, hence:

$$P(q_{0.05}\le\sqrt{I(\theta)}\sqrt{n}\cdot(\hat{\theta}_{ML}-\theta))=0.95\Leftrightarrow P(\hat{\theta}_{ML}-\frac{q_{0.05}}{\sqrt{I(\theta)}\cdot\sqrt{n}}\ge\theta)=0.95$$

With $q_{0.05}=-1.6$ the one-sided 95% CI for θ is: $(-\infty,\hat{\theta}_{ML}+\frac{1.6}{\sqrt{I(\hat{\theta}_{ML})\cdot\sqrt{n}}}]$

Here we have $\hat{\theta}_{ML}=0.8$ and $\frac{1.6}{\sqrt{I(\hat{\theta}_{ML})}\sqrt{n}}=\frac{1.6}{\sqrt{\frac{2}{0.8\cdot0.2^2}}\sqrt{20}}\approx0.045$.

So the one-sided CI is: $(-\infty,0.845]$.

1(e): Asymptotically: $\frac{\frac{d}{d\theta}l_X(\theta)}{\sqrt{n\cdot I(\theta)}}\sim N(0,1)$ where $\frac{d}{d\theta}l_X(\theta)=\frac{-nr}{1-\theta}+\frac{\sum_{i=1}^n x_i}{\theta}$.

Given $r=2$, $\bar{X}=8$ and $n=20$ and $\theta_0=0.9$ we get:
$\frac{-nr}{1-\theta}+\frac{\sum_{i=1}^n x_i}{\theta}=\frac{-40}{1-0.9}+\frac{20\cdot8}{0.9}\approx-222$ and $\sqrt{n\cdot I(\theta)}=\sqrt{20\cdot\frac{2}{0.9\cdot0.1^2}}\approx66.67$

Therefore the score test statistic takes the value: $\frac{\frac{d}{d\theta}l_X(\theta)}{\sqrt{n\cdot I(\theta)}}=\frac{-222}{66.67}\approx-3.33$. As the value is lower than the $q_{0.01}$ quantile -2.3 of the $N(0,1)$, **the score test would reject the null hypothesis** to the level 0.02.

Solutions Exercise 2

2(a) The log-likelihood is given by

$$\ell_Y(\theta) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(y_i - \theta x_i)^2$$

Set the derivative of the log-likelihood equal to zero

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\ell_Y(\theta) = \sum_{i=1}^{n}(y_i - \theta x_i)\,x_i = 0.$$

We find $\hat{\theta}_{\mathrm{ML}} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$. and we don't need to check whether this is really a maximum.

2(b) We have

$$\mathbb{E}\left(\hat{\theta}_{\mathrm{ML}}\right) = \mathbb{E}\left(\frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}\right) = \frac{\sum_{i=1}^n x_i \mathbb{E}(Y_i)}{\sum_{i=1}^n x_i^2} = \frac{\sum_{i=1}^n x_i \theta x_i}{\sum_{i=1}^n x_i^2} = \theta.$$

Therefore $\hat{\theta}_{\mathrm{ML}}$ is unbiased, and:

$$\mathrm{Var}\left(\hat{\theta}_{\mathrm{ML}}\right) = \sum_{i=1}^{n}\left(\frac{x_i}{\sum_{i=1}^n x_i^2}\right)^2 \mathrm{Var}(Y_i) = \frac{\sum_{i=1}^n x_i^2}{\left(\sum_{i=1}^n x_i^2\right)^2} = \frac{1}{\sum_{i=1}^n x_i^2}.$$

2(c) Under the null-hypothesis ($\theta = \theta_0$) we have:

$$\frac{\hat{\theta}_{\mathrm{ML}} - \theta_0}{\sqrt{\frac{1}{\sum_{i=1}^n x_i^2}}} \sim \mathcal{N}(0,1).$$

Hence, we have that

$$\mathbb{P}\left(\left|\frac{\hat{\theta}_{\mathrm{ML}} - \theta_0}{\sqrt{\frac{1}{\sum_{i=1}^n x_i^2}}}\right| > q_{1-\frac{\alpha}{2}}\right) = \alpha.$$

This implies the rejection region:

$$C_\alpha = \left\{ y \in \mathbb{R}^n : \left| \left(\frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} - \theta_0\right)\sqrt{\sum_{i=1}^{n} x_i^2}\right| > q_{1-\frac{\alpha}{2}} \right\}$$

We reject the null-hypothesis if $y \in C_\alpha$.

2(d) For $\alpha = 0.05$ we have $q_{1-\frac{\alpha}{2}} \approx 2$ (see Table 1), and the 95% confidence interval is:

$$\left[\frac{27}{81} - 2\cdot\sqrt{\frac{1}{81}}\ ;\ \frac{27}{81} + 2\cdot\sqrt{\frac{1}{81}}\right] = [1/9\ ;\ 5/9]$$

Solutions Exercise 3

(3a) We have the joint density

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} \frac{1}{2} \cdot \theta^3 \cdot x_i^2 \cdot \exp\{-x_i\theta\} = \frac{1}{2^n} \cdot \theta^{3n} \cdot \left(\prod_{i=1}^{n} x_i\right)^{2n} \cdot \exp\{-\theta \sum_{i=1}^{n} x_i\}$$

$\sum_{i=1}^{n} X_i$ is a sufficient statistic, because we can factorize:

$$L(\theta) = g(x_1, \ldots, x_n) \cdot h(\sum_{i=1}^{n} x_i, \theta)$$

where $g(x_1, \ldots, x_n) := \frac{1}{2^n} \cdot \left(\prod_{i=1}^{n} x_i\right)^{2n}$ and $h(\sum_{i=1}^{n} x_i, \theta) := \theta^{3n} \cdot \{-\theta \sum_{i=1}^{n} x_i\}$

3(b) The UMP test rejects H_0 if the likelihood ratio $W(X)$ is smaller than a constant k:

$$W(X) = \frac{L(4)}{L(2)} = \frac{\frac{1}{2^n} \cdot 4^{3n} \cdot \left(\prod_{i=1}^{n} X_i\right)^{2n} \cdot \exp\{-4 \sum_{i=1}^{n} X_i\}}{\frac{1}{2^n} \cdot 2^{3n} \cdot \left(\prod_{i=1}^{n} X_i\right)^{2n} \cdot \exp\{-2 \sum_{i=1}^{n} X_i\}} = 2^{3n} \cdot \exp\{-2 \sum_{i=1}^{n} X_i\}$$

$W(X)$ is a monotone decreasing function in $\sum_{i=1}^{n} X_i$, so that: $W(X) < k \Leftrightarrow \sum_{i=1}^{n} X_i > k_0$

This shows that a test who rejects H_0 if $\sum_{i=1}^{n} X_i > k_0$ is UMP.

Solutions Exercise 4

4(a): Under H_0 we have $\sqrt{n} \cdot \frac{(\bar{X}_n - 0)}{2} \sim \mathcal{N}(0, 1)$
For $n = 100$ this means: $5 \cdot \bar{X}_{100} \sim \mathcal{N}(0, 1)$ and

$$\bar{X}_{100} < -0.4 \Leftrightarrow 5 \cdot \bar{X}_{100} < -2$$

From Table 1 we see $q_{0.025} = -2$, so that the test is (at least) to the level 0.025.

4(b): Under H_1 we have $5 \cdot (\bar{X}_{100} + 0.54) \sim \mathcal{N}(0, 1)$, and

$$\bar{X}_{100} < -0.4 \Leftrightarrow \bar{X}_{100} + 0.54 < 0.14 \Leftrightarrow 5(\bar{X}_{100} + 0.54) < 0.7$$

From Table 1 we see that 0.7 corresponds to the $q_{0.75}$ quantile, so that the power is 0.75.

4(c): Under H_1 we have $\sqrt{n} \cdot \frac{(\bar{x}_n + 0.54)}{2} \sim \mathcal{N}(0, 1)$ and the relationship:

$$\bar{x}_n < -0.4 \Leftrightarrow \sqrt{n} \cdot \frac{(\bar{x}_n + 0.54)}{2} < 0.07 \cdot \sqrt{n}$$

From Table 1 we see that the 0.9 quantile corresponds to $q_{0.9} = 1.3$. We thus need:

$$0.07 \cdot \sqrt{n} = q_{0.9} = 1.3 \Leftrightarrow n = \left(\frac{1.3}{0.07}\right)^2 \approx 344.9$$

That is, the sample size must be at least $n = 345$.