Statistical Data Analysis
Problem sheet #9 solutions

**1(a)** The likelihood function is given by the binomial distribution evaluated with the single observed value $n$ and regarded as a function of the unknown parameter $\theta$:

$$L(\theta) = \frac{N!}{n!(N-n)!}\theta^n(1-\theta)^{N-n} \ .$$

The log-likelihood function is therefore

$$\ln L(\theta) = n \ln \theta + (N-n)\ln(1-\theta) + C \ ,$$

where $C$ represents terms not depending on $\theta$. Setting the derivative of $\ln L$ equal to zero,

$$\frac{\partial \ln L}{\partial \theta} = \frac{n}{\theta} - \frac{N-n}{1-\theta} = 0 \ ,$$

we find the ML estimator to be

$$\hat{\theta} = \frac{n}{N} \ .$$

**1(b)** We are given the expectation and variance of a binomial distributed variable as $E[n] = N\theta$ and $V[n] = N\theta(1-\theta)$. Using these results we find the expectation value of $\hat{\theta}$ to be

$$E[\hat{\theta}] = E\left[\frac{n}{N}\right] = \frac{E[n]}{N} = \frac{N\theta}{N} = \theta \ ,$$

and therefore the bias is $b = E[\hat{\theta}] - \theta = 0$. Similarly we find the variance to be

$$V[\hat{\theta}] = V\left[\frac{n}{N}\right] = \frac{1}{N^2}V[n] = \frac{N\theta(1-\theta)}{N^2} = \frac{\theta(1-\theta)}{N} \ .$$

**1(c)** Suppose we observe $n = 0$ for $N = 10$ trials. The upper limit on $\theta$ at a confidence level of $\mathrm{CL} = 1 - \alpha$ is the value of $\theta$ for which there is a probability $\alpha$ to find as few events as we found or fewer, i.e.,

$$\alpha = P(n \leq 0; N, \theta) = \frac{N!}{0!(N-0)!}\theta^0(1-\theta)^{N-0} \ .$$

Solving for $\theta$ gives the 95% CL upper limit

$$\theta_{\mathrm{up}} = 1 - \alpha^{1/N} = 1 - 0.05^{1/10} = 0.26 \ .$$

**1(d)** To find the Jeffreys prior we need the second derivative of $\ln L$,

$$\frac{\partial^2 \ln L}{\partial \theta^2} = -\frac{n}{\theta^2} - \frac{N-n}{(1-\theta)^2} \; .$$

The expected Fisher information is therefore

$$I(\theta) = -E\left[\frac{\partial^2 \ln L}{\partial \theta^2}\right] = \frac{N\theta}{\theta^2} + \frac{N(1-\theta)}{(1-\theta)^2} = \frac{N}{\theta} + \frac{N}{1-\theta} = \frac{N}{\theta(1-\theta)} \; .$$

The Jeffreys prior is therefore

$$\pi(\theta) \propto \frac{1}{\sqrt{\theta(1-\theta)}} \; .$$

Using this in Bayes theorem to find the posterior pdf gives

$$p(\theta|n) \propto L(n|\theta)\pi(\theta) \propto \frac{\theta^n (1-\theta)^{N-n}}{\sqrt{\theta(1-\theta)}} = \theta^{n-1/2}(1-\theta)^{N-n-1/2} \; .$$

**1(e)** To find a Bayesian upper limit on $\theta$ one simply integrates the posterior pdf so that a specified probability $1-\alpha$ is contained below $\theta_{\rm up}$, i.e.,

$$1-\alpha = \int_0^{\theta_{\rm up}} p(\theta|n)\, d\theta \; ,$$

solving for $\theta_{\rm up}$ gives the upper limit.

A frequentist upper limit as found in (c) is a function of the data designed to be greater than the true value of the parameter with a fixed probability (the confidence level) regardless of the parameter's actual value. A Bayesian interval can be regarded as reflecting a range for the parameter where it is believed to lie with a fixed probability (the credibility level). Note that with the Jeffreys prior, one may not necessary use the degree of belief interpretation of the interval, but rather take it to have a certain probability to cover the true $\theta$ (which in general will depend on $\theta$).

**2(a)** The variables $x$ and $y$ are independent, so the likelihood function is given by the product of the two pdfs, i.e.,

$$L(\theta_1, \theta_2) = f(x|\theta_1, \theta_2) g(y|\theta_2) = \frac{1}{2\pi\sigma^2} \exp\left[ -\frac{(x - \theta_1 - \theta_2)^2}{2\sigma^2} - \frac{(y - \theta_2)^2}{2\sigma^2} \right] .$$

The log-likelihood function is therefore

$$\ln L(\theta_1, \theta_2) = -\frac{1}{2} \frac{(x - \theta_1 - \theta_2)^2}{\sigma^2} - \frac{1}{2} \frac{(y - \theta_2)^2}{\sigma^2} + C ,$$

where $C = -\ln 2\pi\sigma^2$ is a constant (i.e., does not depend on $\theta_1$ or $\theta_2$) and thus can be dropped.

**2(b)** To find the ML estimators we set the derivatives of $\ln L$ with respect to the parameters equal to zero:

$$\frac{\partial \ln L}{\partial \theta_1} = -\frac{1}{2} \frac{2(x - \theta_1 - \theta_2)(-1)}{\sigma^2} = 0 , \quad (1)$$

$$\frac{\partial \ln L}{\partial \theta_2} = -\frac{1}{2} \frac{2(x - \theta_1 - \theta_2)(-1)}{\sigma^2} - \frac{1}{2} \frac{(y - \theta_2)(-1)}{\sigma^2} = 0 . \quad (2)$$

From Eq. (2) we get $\theta_1 + \theta_2 = x$, and from this the first term in Eq. (2) is zero. We therefore find the ML estimators

$$\begin{aligned} \hat{\theta}_1 &= x - y , \\ \hat{\theta}_2 &= y . \end{aligned}$$

**2(c)** From the pdfs of $x$ and $y$ given we can see the expectation values and variances are

$$\begin{aligned} E[x] &= \theta_1 + \theta_2 , \\ E[y] &= \theta_2 , \\ V[x] &= V[y] = \sigma^2 . \end{aligned}$$

The expectation values of $\hat{\theta}_1$ and $\hat{\theta}_2$ are

$$\begin{aligned} E[\hat{\theta}_1] &= E[x - y] = E[x] - E[y] = \theta_1 + \theta_2 - \theta_2 = \theta_1 , \\ E[\hat{\theta}_2] &= E[y] = \theta_2 , \end{aligned}$$

and therefore we see that both estimators are unbiased. The variances of $\hat{\theta}_1$ and $\hat{\theta}_2$ are

$$\begin{aligned} V[\hat{\theta}_1] &= V[x-y] = V[x] + V[y] = 2\sigma^2 \,, \\ V[\hat{\theta}_2] &= V[y] = \sigma^2 \,, \end{aligned}$$

and the covariance of $\hat{\theta}_1$ and $\hat{\theta}_2$ is

$$\mathrm{cov}[\hat{\theta}_1, \hat{\theta}_2] = \mathrm{cov}[x-y, y] = \mathrm{cov}[x,y] - \mathrm{cov}[y,y] = 0 - V[y] = -\sigma^2 \,.$$

Combining the ingredients above gives the correlation coefficient

$$\rho = \frac{\mathrm{cov}[\hat{\theta}_1, \hat{\theta}_2]}{\sqrt{V[\hat{\theta}_1]V[\hat{\theta}_2]}} = \frac{-\sigma^2}{\sqrt{\sigma^2 \times 2\sigma^2}} = -\frac{1}{\sqrt{2}} \,.$$

**2(d)** Figure 1 shows a contour of the log-likelihood $\ln L(\theta, \theta_2) = \ln L_{\max} - 1/2$, which is centred about the ML estimators $(\hat{\theta}_1, \hat{\theta}_2)$. The standard deviations are determined from the distance from the ML estimators to the tangent lines to the contour. (In the large sample limit the contour is symmetric so the distance to either tangent line can be used.) The negative correlation is indicated by the tilt of the contour from upper left to lower right.

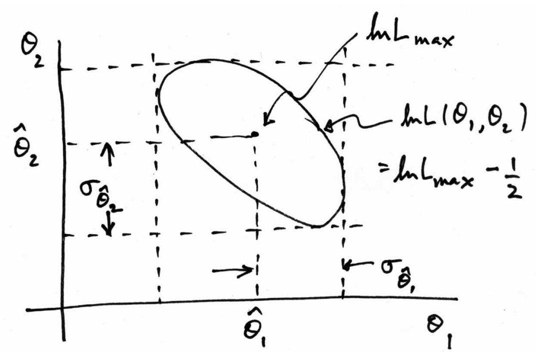


Figure 1: The standard deviations $\sigma_{\hat{\theta}_1}$ and $\sigma_{\hat{\theta}_2}$ are determined from the tangent lines to the contour of $\ln L(\theta_1, \theta_2) = \ln L_{\max} - 1/2$.

**2(e)** From Eq. (2) above setting the derivative of $\ln L$ with respect to $\theta_2$ equal to zero we have

$$\frac{x - \theta - \theta_2}{\sigma^2} + \frac{y - \theta_2}{\sigma^2} = 0 \,.$$

Treating $\theta_1$ as fixed and solving for $\theta_2$ gives the profiled value

$$\hat{\hat{\theta}}_2(\theta_1) = \frac{x + y - \theta_1}{2} \,.$$

The profile likelihood $L_{\mathrm{p}}(\theta_1)$ is defined by evaluating $L(\theta_1, \theta_2)$ with $\hat{\hat{\theta}}_2(\theta_1)$ as found above. After dropping constant terms we find

$$
\begin{aligned}
\ln L_{\mathrm{p}}(\theta_1) &= -\frac{1}{2\sigma^2}\left[\left(x-\theta_1-\frac{x+y-\theta_1}{2}\right)^2+\left(y-\frac{x+y-\theta_1}{2}\right)^2\right] \\
&= -\frac{1}{2\sigma^2}\left[\left(\frac{x}{2}-\frac{y}{2}-\frac{\theta_1}{2}\right)^2+\left(\frac{y}{2}-\frac{x}{2}+\frac{\theta_1}{2}\right)^2\right] \\
&= -\frac{1}{4}\frac{x-y-\theta_1)^2}{\sigma^2}\ .
\end{aligned}
$$

The derivatives of $\ln L_{\mathrm{p}}$ are

$$
\begin{aligned}
\frac{\partial \ln L_{\mathrm{p}}}{\partial \theta_1} &= \frac{1}{2}\frac{x-y-\theta_1}{\sigma^2}\ , \\
\frac{\partial^2 \ln L_{\mathrm{p}}}{\partial \theta_1^2} &= -\frac{1}{2\sigma^2}\ .
\end{aligned}
$$

Using the profile likelihood to find the Fisher information therefore gives

$$
I(\theta) = -E\left[\frac{\partial^2 \ln L_{\mathrm{p}}}{\partial \theta_1^2}\right] = \frac{1}{2\sigma^2}
$$

Using this to determine (approximately) the variance thus gives the same as the exact result found above

$$
V[\hat{\theta}_1] \approx I^{-1}(\theta) = 2\sigma^2\ .
$$