Exam STK 4021/9021 December 2015:
Notes, by Nils Lid Hjort

Exercise 1

(a) First, $y_i \,|\, \theta$ is a binomial $(1,\theta)$, with mean θ and variance $\theta(1-\theta)$. Secondly, $z = \sum_{i=1}^n y_i$ given θ is binomial (n,θ), hence with mean and variance $n\theta$ and $n\theta(1-\theta)$.

(b) Now θ is uniform on $(0,1)$.

(i) Its mean and variance are $\frac{1}{2}$ and $\frac{1}{12}$.

(ii) Then the marginal distribution of y_i:

$$P(Y_i = 1) = \int_0^1 P(Y_i = 1 \,|\, \theta) p(\theta)\,\mathrm{d}\theta = \int_0^1 \theta\,\mathrm{d}\theta = \tfrac{1}{2}.$$

Hence Y_i marginally is binomial $(1, \frac{1}{2})$, with mean and variance $\frac{1}{2}$ and $\frac{1}{4}$.

(iii) We have $P(Y_i = 1, Y_j = 1 \,|\, \theta) = \theta^2$, so

$$\mathrm{cov}(Y_i, Y_j) = \mathrm{E}\, Y_i Y_j - \tfrac{1}{4} = \mathrm{E}\,\theta^2 - \tfrac{1}{4} = \mathrm{Var}\,\theta = \tfrac{1}{12},$$

with ensuing correlation $\frac{1}{3}$.

(iv) From rules of double expectation, $\mathrm{E}\, z = \mathrm{E}\mathrm{E}(z \,|\, \theta) = \mathrm{E}\, n\theta = \frac{1}{2}n$, and

$$\mathrm{Var}\, z = \mathrm{E}\, n\theta(1-\theta) + \mathrm{Var}\,(n\theta) = n/6 + n^2/12.$$

(c) The marginal distribution of data is

$$f(y_1, \ldots, y_n) = \int \theta^z (1-\theta)^{n-z}\,\mathrm{d}\theta = \frac{z!(n-z)!}{(n+1)!},$$

with $z = \sum_{i=1}^n y_i$.

(d) We have

$$P(M_0 \,|\, \mathrm{data}) = \frac{\frac{1}{2} f_0}{\frac{1}{2} f_0 + \frac{1}{2} f_1} \quad \text{and} \quad P(M_1 \,|\, \mathrm{data}) = \frac{\frac{1}{2} f_1}{\frac{1}{2} f_0 + \frac{1}{2} f_1},$$

where

$$f_0 = \int p(y_1, \ldots, y_{20} \,|\, \theta) p(\theta)\,\mathrm{d}\theta = \frac{11!\,9!}{21!}$$

and

$$f_1 = \int p(y_1, \ldots, y_{10} \,|\, \theta_A) p(y_{11}, \ldots, y_{20} \,|\, \theta_B) p(\theta_A) p(\theta_B)\,\mathrm{d}\theta_A\,\mathrm{d}\theta_B = \frac{7!\,3!}{11!}\frac{4!\,6!}{11!}.$$

I find

$$f_0 = \frac{1}{3527160} \quad \text{and} \quad f_1 = \frac{1}{1320}\frac{1}{2310}$$

leading to probabilities 0.464 and 0.536 for Model Zero and Model One.

Exercise 2

(a) The posterior for θ given y is proportional to $\exp(-\theta)\theta^y\theta^{a-1}\exp(-b\theta)$, and hence a Gamma $(a+y, b+1)$. The Bayes estimate under quadratic loss is the posterior mean,

$$\widehat{\theta} = \frac{a+y}{b+1} = (1-w)\theta_0 + wy,$$

where $\theta_0 = a/b$ is the prior mean and $w = 1/(b+1)$.

(b) The risk function for some $\widetilde{\theta}$ estimator under quadratic loss is $\mathrm{E}_\theta(\widetilde{\theta}-\theta)^2$. For the ML we have $R(\theta^*, \theta) = \theta$. For the Bayes estimator we find

$$R(\widehat{\theta}, \theta) = w^2\theta + (1-w)^2(\theta-\theta_0)^2.$$

This is indeed smaller than θ when $|\theta-\theta_0|$ is small. The Bayes is in fact better than the ML as long as

$$w^2\theta + (1-w)^2(\theta-\theta_0)^2 \le \theta$$

or

$$(\theta-\theta_0)^2 \le \frac{1+w}{1-w}\theta = (b+2)\theta.$$

(c) One finds $\mathrm{E}\, y = a/b$ and $\mathrm{Var}\, y = \theta_0(1+1/b)$. So $\bar{y}$ is a good estimator for $\theta_0 = a/b$, and s^2 estimates $\theta_0(1+1/b)$. This means $\bar{y}/s^2$ estimates $b/(b+1)$, and we may solve for b to find $\widehat{b} = \bar{y}/(s^2-\bar{y})$. In case the variance is smaller than $\bar{y}$ (which may happen, but typically with small probability), we may put $a/b = \theta_0 = \bar{y}$ and $\widehat{b} = \infty$.

(d) The marginal distribution of y_i is

$$\begin{aligned} f(y) &= \int_0^\infty \exp(-\theta)\frac{\theta^y}{y!}\frac{b^a}{\Gamma(a)}\theta^{a-1}\exp(-b\theta)\,\mathrm{d}\theta \\ &= \frac{b^a}{\Gamma(a)}\frac{1}{y!}\frac{\Gamma(a+y)}{(b+1)^{a+y}}, \end{aligned}$$

for $y = 0, 1, 2, \ldots$. The log-likelihood function becomes

$$\ell_n(a,b) = \sum_{i=1}^n \{a\log b - \log\Gamma(a) + \log\Gamma(a+y_i) - (a+y_i)\log(b+1)\},$$

and the marginal ML are the $(\widehat{a}, \widehat{b})$ maximising this expression. Taking the derivative with respect to b and setting it to zero leads to the insight that $\bar{y} = \widehat{a}/\widehat{b}$.

(e) For estimating $w = 1/(b+1)$ one may use moment estimates or ML, both agreeing that $\bar{y} = \widehat{a}/\widehat{b}$, but with different schemes for determining the b. In any case the suggestion is

$$\widetilde{\theta}_i = (1-\widehat{w})\bar{y} + \widehat{w}y_i,$$

for example

$$\widetilde{\theta}_i = \frac{\bar{y}}{s^2}\bar{y} + \Big(1 - \frac{\bar{y}}{s^2}\Big)y_i.$$

Using this 'lending strength' empirical Bayes method will give a risk function better than for the ML method, for a wide range of $(\theta_1, \ldots, \theta_n)$ in the parameter space, those for which the spread is not too big. The ML might be better for cases where the parameters have a big spread.

Exercise 3

(a) This is Bayes' formula. The marginal distribution of y is $p(y) = \sum_{k=1}^{10} \pi_k f_k(y)$.

(b) The expected loss, given data y, associated with the decision $\widehat{c}$, is

$$\mathrm{E}\,\{L(c, \widehat{c}) \,|\, y\} = 1 \cdot P(c \neq \widehat{c} \,|\, y) = 1 - P(c = \widehat{c} \,|\, y).$$

Minimising this posterior expected loss is hence the same as finding the $\widehat{c}$ that makes $P(c = \widehat{c} \,|\, y)$ the biggest. I classify y to come from the most probable class. Its risk function is a function of the parameter $c = 1, \ldots, 10$, and is equal to

$$r(c) = \mathrm{E}_c L(c, \widehat{c}) = 1 \cdot \mathrm{pmc}(c) = 1 - \mathrm{pcc}(c),$$

which is the error rate or probability of misclassification $\mathrm{pmc}(c)$ at c, or one minus the hit rate or probability of correct classification $\mathrm{pcc}(c)$ at c. The Bayes method succeeds in minimising the overall error rate, or maximising the overall rate of correct classification:

$$\text{error rate} = \sum_{k=1}^{10} \pi_k \mathrm{pmc}(k) = 1 - \sum_{k=1}^{10} \pi_k \mathrm{pcc}(k).$$

(c) The posterior expected loss is still

$$\mathrm{E}\,\{L(c, \widehat{c}) \,|\, y\} = 1 - P(c = \widehat{c} \,|\, y)$$

if $\widehat{c}$ is among $1, \ldots, 10$, and is equal to k if $\widehat{c}$ is set equal to D. The Bayes solution is to allocate y to the most probable class $\widehat{c}$, as long as this probability $P(c = \widehat{c} \,|\, y)$ is bigger than $1 - k$; if all class probabilities are smaller than $1 - k$, then use Doubt.

Exercise 4

(a) The likelihood for these data is $1/\theta^n$ for $y \geq y_{\max}$. Here $n = 10$ and $y_{\max} = 6.931$.

(b) With prior $1/\theta$, the posterior takes the form

$$p(\theta \,|\, \text{data}) = \frac{c}{\theta^{n+1}} \quad \text{for } \theta \geq y_{\max},$$

with c the appropriate integration constant. One finds that this corresponds to cumulative posterior

$$F_n(\theta) = 1 - (y_{\max}/\theta)^n \quad \text{for } \theta \geq y_{\max},$$

with posterior density $p_n(\theta) = n y_{\max}^n / \theta^{n+1}$ for this range of θ. The Bayes estimate is the posterior median, which is found to be

$$\widehat{\theta} = \frac{y_{\max}}{(\frac{1}{2})^{1/n}}$$

which here is equal to $1.072\, y_{\max}$. The ML estimator is $y_{\max}$, and the best unbiased estimator is $1.1\, y_{\max}$.

The Project, Exercise 1

(a) It is positive and integrates to 1. Also, the cdf is $1-\exp(-\theta\sqrt{y})$, and the median solves $\theta\sqrt{\mu}=\log 2$, giving $(\log 2)^2/\theta^2$. The mean is found to be $2/\theta^2$, by integration, or by noting that $x=\theta\sqrt{y}$ is a unit exponential, so $y=x^2/\theta^2$ etc.

(b) The log-likelihood is

$$\ell_n(\theta)=\sum_{i=1}^n(\log\theta-\theta y_i^{1/2}),$$

and the ML is $\widehat\theta=1/W_n$, with $W_n=(1/n)\sum_{i=1}^n y_i^{1/2}$. Via $y_i=x_i^2/\theta^2$, for iid unit exponentials, we have

$$\widehat\theta=\frac{\theta}{\bar x}=\theta\frac{2n}{\sum_{i=1}^n 2x_i}=\theta\frac{2n}{\chi^2_{2n}},$$

from which the exact density etc. may be found. Also, $J(\theta)=1/\theta^2$, so

$$\widehat\theta\approx \mathrm{N}(\theta,\theta^2/n)\approx \mathrm{N}(\theta,\widehat\theta^2/n)$$

from curriculum.

(c) The posterior is proportional to $\theta^{a-1}\exp(-b\theta)\theta^n\exp(-\theta n w_n)$, and is hence a Gamma $(a+n,b+nw_n)$.

(d) The marginal density for y becomes

$$f(y)=\int_0^\infty\frac{\theta}{2\sqrt{y}}\exp(-\theta\sqrt{y})\frac{b^a}{\Gamma(a)}\theta^{a-1}\exp(-b\theta)\,\mathrm{d}\theta=\frac{b^a}{\Gamma(a)}\frac{1}{2\sqrt{y}}\frac{\Gamma(a+1)}{(b+\sqrt{y})^{a+1}}.$$

For the uniform case $(a,b)=(1,1)$,

$$f(y)=\frac{1}{2\sqrt{y}(1+\sqrt{y})^2}\quad\text{og}\quad F(y)=1-\frac{1}{1+\sqrt{y}}=\frac{\sqrt{y}}{1+\sqrt{y}}.$$

(e) The cdf for the mean $2/\theta^2$ is

$$P\{2/\theta^2\le x\}=P\{(2/x)^{1/2}\le\theta\}=1-G((2/x)^{1/2},a,b).$$

We must choose (a,b) to have

$$1-G((2/0.20)^2,a,b)=0.10\quad\text{og}\quad 1-G((2/2.00)^2,a,b)=0.90.$$

I find $(a_0,b_0)=(5.315,2.656)$.

(f) The ML is 0.856, the posterior mean is

$$\widehat\theta_B=\frac{a_0+n}{b_0+\sum_{i=1}^n y_i^{1/2}}=1.038.$$

(g) In the same graph, can display the Gamma with parameters $(a_0+nn, b_0+\sum_{i=1}^n y_i^{1/2}) = (17.315, 16.673)$ and the normal approximation $\mathrm{N}(0.856, 0.856^2/12)$. They are close, but not very much so (but will become closer for higher n). There are slightly more informative versions of 'lazy Bayes' that take the prior into account.

(h) The predictive density is as for the marginal density above, but with (a,b) replaced by the updated $(a',b') = (a_0+n, b_0+\sum_{i=1}^n \sqrt{y_i})$. I simulate 10^6 values of θ from the Gamma posterior, and for each of these I simulate $y_{\rm new}$ from the model density, which is the same as using $y_{\rm new} = x_{\rm new}^2/\theta^2$, with $x_{\rm new}$ from the unit exponential. I read off quantiles via `quantile(ysim,c(0.1,0.5,0.9))` and find $(0.010, 0.463, 5.623)$.

The Project, Exercise 2

(a) The posterior is proportional to

$$\theta^{a-1}(1-\theta)^{b-1}\theta(1-\theta)^{y-1}$$

and hence a Beta $(a+1, b+y-1)$.

(b) For the uniform case $(a,b)=(1,1)$, the posterior is a Beta $(2,y)$, and the marginal distribution for y is

$$f(y) = \int_0^1 (1-\theta)^{y-1}\theta\,{\rm d}\theta = \frac{(y-1)!}{(y+1)!} = \frac{1}{y(y+1)} \quad \text{for } y = 1,2,3,\ldots.$$

Its mean is infinity, which also may be seen via $\mathrm{E}(y\,|\,\theta) = 1/\theta$, which yields $\mathrm{E}\,y = \mathrm{E}\,1/\theta = \infty$.

(c) We have $\log f = \log\theta + (y-1)\log(1-\theta)$ with derivative $u = 1/\theta - (y-1)/(1-\theta)$, hence

$$J(\theta) = \mathrm{Var}_\theta\, u(Y,\theta) = \frac{1}{\theta^2(1-\theta)}.$$

The square-root is the Jeffreys prior, which may be seen as a Beta $(0, \frac{1}{2})$. It is improper, but yields a proper posterior as soon as there is one or more y.

(d) Risk functions are computed as

$$r(\theta) = \mathrm{E}_\theta(\widetilde\theta - \theta)^2 = \sum_{y\ge1}\{\widetilde\theta(y) - \theta\}^2(1-\theta)^{y-1}\theta.$$

Doing this with $1/y$ and $2/(2+y)$ gives nice risk functions $r_1(\theta)$ and $r_2(\theta)$, for the ML and the Bayes. The Bayes is better if $\theta \le 0.836$ (and partly much better); only for $\theta > 0.836$ is ML better.

(e) The posterior means are

$$\widehat\theta_i = \frac{a+1}{a+b+y_i} \quad \text{for } i = 1,\ldots,n.$$

(f) The marginal distribution for y_i is

$$f(y_i) = \int (1-\theta)^{y_i-1}\theta p(\theta\,|\,a,b)\,\mathrm{d}\theta = a\frac{\Gamma(a+b)}{\Gamma(b)}\frac{\Gamma(b+y_i-1)}{\Gamma(a+b+y_i)}$$

for $y_i = 1,2,3,\ldots$. Can then form the marginal log-likelihood

$$\ell_n(a,b) = \sum_{i=1}^n \{\log a + \log\Gamma(a+b) - \log\Gamma(b) + \log\Gamma(b+y_i-1) - \log\Gamma(a+b+y_i)\}.$$

Numerical maximisation, using `nlm` and starting at e.g. $(10,10)$, gives ML estimates $(1.714, 6.355)$.

(g) MCMC for (a,b) can be implemented in the usual fashion. I have used normal proposals, of the type $a_{\rm new} \sim \mathrm{N}(a_{\rm old}, \frac{1}{2}\mathrm{se}_a)$ and $b_{\rm new} \sim \mathrm{N}(b_{\rm old}, \frac{1}{2}\mathrm{se}_b)$, with $(\mathrm{se}_a, \mathrm{se}_b) = (0.973, 5.113)$ the approximate standard errors from likelihood analysis. With a million MCMC steps I have acceptance rate about 0.80, and also mean probability of acceptance 0.80. The a is better estimated than b, the former having $(3.836, 2.792)$ as mean and standard deviation, the latter $(59.725, 27.147)$. The correlation is 0.468. A plot of $a/(a+b)$ is also useful, the prior mean in the distribution of geometric probabilities.

(h) With a bit of work I can find and display `cbind(1:nn,yy,1/yy,eB,low,mid,up)`, involving the full million of MCMC steps:

```
 [1,]  1  1 1.0000 0.2992 0.0160 0.0705 0.2258
 [2,]  2  1 1.0000 0.2992 0.0159 0.0704 0.2263
 [3,]  3  2 0.5000 0.2695 0.0157 0.0691 0.2139
 [4,]  4  2 0.5000 0.2695 0.0157 0.0690 0.2138
 [5,]  5  2 0.5000 0.2695 0.0157 0.0691 0.2146
 [6,]  6  2 0.5000 0.2695 0.0157 0.0691 0.2144
 [7,]  7  2 0.5000 0.2695 0.0157 0.0691 0.2142
 [8,]  8  3 0.3333 0.2452 0.0154 0.0678 0.2044
 [9,]  9  3 0.3333 0.2452 0.0155 0.0678 0.2041
[10,] 10  3 0.3333 0.2452 0.0154 0.0678 0.2043
[11,] 11  3 0.3333 0.2452 0.0155 0.0678 0.2043
[12,] 12  3 0.3333 0.2452 0.0154 0.0678 0.2042
[13,] 13  5 0.2000 0.2076 0.0149 0.0654 0.1885
[14,] 14  6 0.1667 0.1929 0.0146 0.0642 0.1817
[15,] 15  6 0.1667 0.1929 0.0146 0.0643 0.1819
[16,] 16 14 0.0714 0.1230 0.0128 0.0564 0.1485
[17,] 17 15 0.0667 0.1176 0.0126 0.0555 0.1454
[18,] 18 15 0.0667 0.1176 0.0126 0.0555 0.1453
[19,] 19 20 0.0500 0.0967 0.0118 0.0517 0.1332
[20,] 20 75 0.0133 0.0327 0.0065 0.0298 0.0766
```

The Project, Exercise 3

(a) The posterior for τ becomes

$$\pi(\tau \mid \text{data}) \propto f_L(y_1)\cdots f_L(y_\tau) f_R(y_{\tau+1})\cdots f_R(y_n).$$

The right hand side can be computed for each τ (typically using the logarithm and summing) and then normalised.

(b) I simulate $\tau = 66$ data points from $\text{N}(1.1, 1)$ and 34 data points from $\text{N}(2.2, 1)$. Then I compute

$$Q(\tau) = -\tfrac{1}{2}\sum_{i\le\tau}(y_i - \xi_L)^2 - \tfrac{1}{2}\sum_{i\ge\tau+1}(y_i - \xi_R)^2$$

for each τ. The posterior for τ is then proportional to $\exp(Q(\tau))$ and can be displayed. It will have most of its mass close to the true τ.

(c) The log-likelihood function for the Poisson case is

$$\begin{aligned}\ell(\tau, \theta_L, \theta_R) &= \sum_{i\le\tau}(-\theta_L + y_i \log\theta_L) + \sum_{i\ge\tau+1}(-\theta_R + y_i\log\theta_R)\\ &= -\tau\{\theta_L + \bar y_L \log\theta_L\} - (n-\tau)\{\theta_R + \bar y_R \log\theta_R\},\end{aligned}$$

in terms of left and right averages $\bar y_L$ and $\bar y_R$. This is for given τ maximised by setting θ_L and θ_R equal to $\bar y_L$ and $\bar y_R$. Hence it remains to maximise

$$\ell(\tau, \widehat\theta_L(\tau), \widehat\theta_R(\tau)) = -\tau\{\widehat\theta_L(\tau) + \widehat\theta_L(\tau)\log\widehat\theta_L(\tau)\} - (n-\tau)\{\widehat\theta_R(\tau) + \widehat\theta_R(\tau)\log\widehat\theta_R(\tau)\}$$

over the possible values of τ. I find $\widehat\tau = 22$. Then, given this change-point value, I find ML for θ_L and θ_R in the usual fashion, based on 22 Poisson data to the left and 29 Poisson data to the right. Values are 3.045 and 0.896.

(d) The full distribution for $(\tau, \theta_L, \theta_R, y_1, \ldots, y_n)$ can be written

$$\pi(\tau)p(\theta_L)p(\theta_R)\prod_{i\le\tau}\exp(-\theta_L)\theta_L^{y_i}/y_i!\prod_{i\ge\tau+1}\exp(-\theta_R)\theta_R^{y_i}/y_i!.,$$

with $\pi(\tau)$ the prior for τ. But this may be rewritten as proportional to

$$\begin{aligned}&\pi(\tau)\theta_L^{a-1}\exp(-b\theta_L)\theta_L^{\tau\bar y_l(\tau)}\exp(-\tau\theta)\theta_R^{a-1}\exp(-b\theta_R)\theta_R^{(n-\tau)\bar y_R(\tau)}\exp(-(n-\tau)\theta_R)\\ &\pi(\tau)\theta_L^{a+\tau\bar y_l(\tau)-1}\exp\{-(b+\tau)\theta_L\}\theta_R^{a+(n-\tau)\bar y_R(\tau)-1}\exp\{-(b+n-\tau)\theta_R(\tau)\}.\end{aligned}$$

Integrating over (θ_L, θ_R) gives

$$\pi(\tau \mid \text{data}) \propto \pi(\tau)\frac{\Gamma(a+\tau\bar y_L(\tau))}{(b+\tau)^{a+\tau\bar y_L(\tau)}}\frac{\Gamma(a+(n-\tau)\bar y_R(\tau))}{(b+n-\tau)^{a+(n-\tau)\bar y_R}}$$

which may be computed and displayed. Also, given τ, the posteriors for θ_L and θ_R are independent, with

$$\begin{aligned}\theta_L \mid \text{data}, \tau &\sim \text{Gam}(a+\tau\bar y_L(\tau), b+\tau),\\ \theta_R \mid \text{data}, \tau &\sim \text{Gam}(a+(n-\tau)\bar y_R(\tau), b+n-\tau).\end{aligned}$$

(e) To simulate from the posterior of $(\tau, \theta_L, \theta_R)$ is therefore easy: we simulate τ from $\pi(\tau \,|\, \text{data})$ using `sample`, and then conditional on such draws we simulate θ_L and θ_R from Gamma distributions (involving $\bar{y}_L$ and $\bar{y}_R$ computed with different τ each time). Doing this a million times, I find $(17, 21, 27)$ for the 0.05, 0.50, 0.95 posterior quantiles for τ. For $\gamma = \theta_R/\theta_L$ I find $(0.212, 0.314, 0.453)$.