The four-hour exam

Exercise 1

(a) First, $y_i \mid \theta$ 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 $\theta \sim \mathrm{Beta}(2, 2)$.

(i) The mean is $\frac{1}{2}$, the variance is $\frac{1}{2}\frac{1}{2}/5 = 1/20$.

(ii) $\mathrm{E}\, y_i = \mathrm{E}\,\mathrm{E}\,(y_i \mid \theta) = \mathrm{E}\,\theta = \frac{1}{2}$, and

$$\begin{aligned}\mathrm{Var}\, y_i &= \mathrm{E}\,\mathrm{Var}\,(y_i \mid \theta) + \mathrm{Var}\,\mathrm{E}\,(y_i \mid \theta) \\ &= \mathrm{E}\,\theta(1-\theta) + \mathrm{Var}\,\theta \\ &= \tfrac{1}{2} - \mathrm{E}\,\theta^2 + \mathrm{E}\,\theta^2 - \tfrac{1}{4} = \tfrac{1}{4}.\end{aligned}$$

Also, y_i is a 0-1 variable, with $P(y_i = 1) = \frac{1}{2}$; hence its variance must be $\frac{1}{2}\frac{1}{2} = \frac{1}{4}$.

(iii) We have $\mathrm{E}\,(y_i y_j \mid \theta) = \theta^2$ and hence $\mathrm{cov}(y_i, y_j) = \mathrm{E}\,\theta^2 - (\frac{1}{2})^2 = \mathrm{Var}\,\theta$, i.e. 1/20. Hence the correlation is

$$\rho = \frac{1/20}{1/4} = \frac{4}{20} = 1/5 = 0.20.$$

(iv) We have

$$\mathrm{E}\, z = \mathrm{E}\, n\theta = \tfrac{1}{2} n$$

and

$$\mathrm{Var}\, z = \mathrm{E}\, n\theta(1-\theta) + \mathrm{Var}\, n\theta = n(\tfrac{1}{2} - \tfrac{1}{4} - 1/20) + n^2/20 = 0.20\, n + 0.05\, n^2.$$

(c) The distribution of z is

$$\begin{aligned}f(z) &= \int_0^1 \binom{n}{z} \theta^z (1-\theta)^{n-z} 6\theta(1-\theta)\, \mathrm{d}\theta \\ &= \frac{n!}{z!\,(n-z)!} \frac{(z+1)!\,(n-z+1)!}{(n+2)!} \\ &= 6\frac{(z+1)(n-z+1)}{n(n+1)}\end{aligned}$$

for $z = 0, 1, \ldots, n$.

(d) The posterior is proportional to

$$\theta^z(1-\theta)^{n-z}\theta(1-\theta) = \theta^{z+1}(1-\theta)^{n-z+1},$$

which is a $\mathrm{Beta}(z+2, n-z+2)$. The conditional mean is

$$\widehat{\theta}_B = \mathrm{E}\,(\theta \mid z) = \frac{z+2}{n+4}.$$

(e) The standard estimator is $\widetilde\theta = z/n$, which is unbiased. Hence the risk, the mean of $(\widetilde\theta-\theta)^2$, is $\theta(1-\theta)/n$. It starts and ends at zero, and is biggest for $\theta=\frac12$.

(f) The Bayes estimator has risk function

$$\begin{aligned} r(\theta) &= \mathrm{Var}\,\widehat\theta_B + (\mathrm{E}\,\widehat\theta_B-\theta)^2 \\ &= \frac{n\theta(1-\theta)}{(n+4)^2} + \Big(\frac{n\theta+2}{n+4}-\frac{n\theta+4\theta}{n+4}\Big)^2 \\ &= \frac{n}{(n+4)^2}\theta(1-\theta) + \frac{(4\theta-2)^2}{(n+4)^2}. \end{aligned}$$

It is better than the usual ML estimator for those θ where

$$\frac{16(\theta-\frac12)^2}{(n+4)^2} \le \Big\{\frac1n - \frac{n}{(n+4)^2}\Big\}\theta(1-\theta),$$

or

$$16(\theta-\tfrac12)^2 \le \frac{(n+4)^2-n^2}{n}\theta(1-\theta) = \frac{16+8n}{n}\theta(1-\theta),$$

or $(\theta-\frac12)^2\le(\frac12+1/n)\theta(1-\theta)$. This means a certain interval around $\frac12$, where Bayes is better. For large n, the inequality is close to $(\theta-\frac12)^2\le\frac12\theta(1-\theta)$, and this holds for θ inside $\frac12\pm\sqrt3/6$, which is $[0.211, 0.789]$, i.e. a pretty wide interval.

Exercise 2

(a) The likelihood function becomes

$$\begin{aligned} L &= (p(1-q))^{212}\,((1-p)q)^{103}\,(pq)^{39}\,((1-p)(1-q))^{148} \\ &= p^{212+39}(1-p)^{103+148}q^{103+39}(1-q)^{212+148} \\ &= p^{251}(1-p)^{251}q^{142}(1-q)^{360}. \end{aligned}$$

(b) With independent uniforms for p and q, the posteriors are also independent, with

$$p\,|\,\mathrm{data}\sim\mathrm{Beta}(252,252),\quad q\,|\,\mathrm{data}\sim\mathrm{Beta}(143,361).$$

(c) The expected number of AB cases, if this theory is correct, is

$$e_{AB} = \mathrm{E}\,(npq\,|\,\mathrm{data}) = n\frac{252}{502}\frac{143}{502} = 71.5.$$

But this is far off from the observed 39. So the theory looks very suspicious, indeed. As Landsteiner and others found out, about a hundred years ago, the two-loci theory stinks and sucks; the one-locus theory, however, is splendid. – One may put in more detail here, including computing the probability that one should get a number as far off as 39 (or more), as measured through the lens of the posterior distribution for npq. This will be a microscopic probability. The essence is simply to compare the observed far too small 39 with the mean of 71.5 – and, of course, similar calculations for the other three cells.

Exercise 3

(a) The posterior density $p(\theta)$ is the derivative of the cumulative $P(\theta)$, and one finds $\theta\exp(-\theta)$. This is also a gamma $(2,1)$. Its mean is 2. The density is zero at zero, climbs to $\exp(-1)=0.368$ at the value 1, and then decreases slowly to zero.

(b) The Bayes decision is to pick among A, B, C the action that has the smallest expected posterior loss. These three expected posterior losses are

$$\begin{aligned} 2\,\{1-P(1.1)\} &= 2\cdot 0.6990 = 1.3981,\\ 3\,\{P(1.1)+1-P(3.3)\} &= 3\cdot 0.4596 = 1.3787,\\ 4\,P(3.3) &= 4\cdot 9.8414 = 3.3656, \end{aligned}$$

for respectively A, B, C. So we take action B.

(c) The risk function, for a credibility interval $[\widehat a,\widehat b]$ constructed from the data, becomes

$$r(\theta) = \mathrm{E}_\theta\, L(\theta,[\widehat a,\widehat b]) = 0.10\, E_\theta\,(\widehat b-\widehat a) + 1 - P_\theta\{\theta\in[\widehat a,\widehat b]\}.$$

A good method is one with short expected length and with high probability of containing the right parameter.

(d) Again we ought to minimise posterior expected loss. This means minimising

$$\mathrm{E}\,\{L(\theta,[a,b])\,|\,\mathrm{data}\} = 0.10\,(b-a) + P(a) + 1 - P(b)$$

over (a,b). Taking derivatives leads to the equations $p(a)=0.10$, $p(b)=0.10$. This again means finding a to the left of 1 and b to the right of 1, as solutions to $p(x)=0.10$. I find $[a,b]=[0.112, 3.576]$.

Exercise 4

(a) The likelihood for the data becomes

$$\prod_{i=1}^n \{(3\theta y_i^2)\exp(-\theta y_i^3)\} \propto \theta^n \exp(-n\theta W_n),$$

where $W_n=(1/n)\sum_{i=1}^n y_i^3$. Its logarithm is

$$\ell_n(\theta) = n\log\theta - n\theta W_n,$$

with first derivative $n/\theta - nW_n$ and second derivative $-n/\theta^2$. So the maximum likelihood (ML) estimator is

$$\widehat\theta = 1/W_n.$$

(b) The prior times the likelihood is proportional to

$$\theta^{a-1}\exp(-b\theta)\theta^n\exp(-n\theta W_n) = \theta^{a+n-1}\exp\{-(b+nW_n)\theta\},$$

which means the posterior is a Gamma with parameters $(a+n, b+nW_n)$. Its mean, by the way, is the Bayes estimator

$$\widehat\theta_B = \frac{a+n}{b+nW_n},$$

which is close to the ML estimator.

(c) First, by frequentist ML theory, and assuming the model is actually correct,

$$\widehat\theta \approx_d \mathrm{N}(\theta_0, \theta_0^2/n),$$

with θ_0 signalling the true parameter value. Secondly, by Bayes theory for larger sample sizes,

$$\theta \,|\, \mathrm{data} \approx_d \mathrm{N}(\widehat\theta, \widehat\theta^2/n).$$

So there's a mirror situation, and the Bayesian and the frequentist will have the same inferences, for large n.