Petter Mostad
Matematisk Statistik
Chalmers

Suggested solution for exam in MSA100/MVE185 Computer Intensive Statistical Methods, October 2008

1. (a) We get

$$
\begin{aligned}
\pi(p \mid y) &\propto \pi(y \mid p)\pi(p) \\
&\propto p^r(1-p)^y p^{\alpha-1}(1-p)^{\beta-1} \\
&\propto p^{r+\alpha-1}(1-p)^{y+\beta-1}
\end{aligned}
$$

which means that the posterior is a Beta distribution with parameters $r+\alpha$ and $y+\beta$.

(b) We get, for example,

$$
\begin{aligned}
\pi(y) &= \frac{\pi(y \mid p)\pi(p)}{\pi(p \mid y)} \\
&= \frac{\frac{\Gamma(y+r)}{\Gamma(y+1)\Gamma(r)} p^r(1-p)^y \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}}{\frac{\Gamma(r+\alpha+y+\beta)}{\Gamma(r+\alpha)\Gamma(y+\beta)} p^{r+\alpha-1}(1-p)^{y+\beta-1}} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha+r)}{\Gamma(r)} \cdot \frac{\Gamma(y+\beta)}{\Gamma(y+1)} \cdot \frac{\Gamma(y+r)}{\Gamma(y+r+\alpha+\beta)}
\end{aligned}
$$

(c) The number of unsuccessful trials until 10 successful trials is obtained is Negative Binomially distributed with parameters p and $r = 10$, and Karl has observed $y = 36 - 10 = 26$ in such an experiment. The prior mentioned corresponds to an improper Beta distribution with $\alpha = 0$ and $\beta = 0$. Thus we can get the result from part (a): The posterior is a Beta distribution with parameters 10 and 26. The expectation of such a Beta distribution is $\frac{10}{10+26} = \frac{10}{36} = 0.278$.

(d) Karl can now use the posterior from (c) as his prior. The number of unsuccessful experiments y^* necessary to get n successful ones is then given by the distribution found in (b), with $\alpha = 10, \beta = 26, r = n$ and $y = y^*$:

$$
\pi(y^*) = \frac{\Gamma(36)}{\Gamma(10)\Gamma(26)} \cdot \frac{\Gamma(10+n)}{\Gamma(n)} \cdot \frac{\Gamma(y^*+26)}{\Gamma(27)} \cdot \frac{\Gamma(n+y^*)}{\Gamma(n+y^*+36)}.
$$

2. (a) Using the Holm method we get adjusted p-values

$$
\begin{aligned}
H_{01}: &\quad \max(0.071 \cdot 1, 0.128) = 0.128 \\
H_{02}: &\quad 0.002 \cdot 4 = 0.008 \\
H_{03}: &\quad 0.064 \cdot 2 = 0.128 \\
H_{04}: &\quad 0.027 \cdot 3 = 0.081.
\end{aligned}
$$

This shows that we can reject H_{02} and H_{04}, while still guaranteeing a FWER<10%.

(b) An alternative is now to use Sidak adjusted p-values, which would become

$$
\begin{aligned}
H_{01}: &\quad 1-(1-0.071)^4 = 0.255 \\
H_{02}: &\quad 1-(1-0.002)^4 = 0.008 \\
H_{03}: &\quad 1-(1-0.064)^4 = 0.232 \\
H_{04}: &\quad 1-(1-0.027)^4 = 0.104.
\end{aligned}
$$

Using this method, we can only reject H_{02}. However, as the assumptions in (a) are actually weaker, we can still reject both H_{02} and H_{04}, guaranteeing a FWER<10%.

3. (a) We get

$$\begin{aligned}\frac{d}{dx}\log f(x) &= -2\frac{2x+\frac{2x}{1+x^2}}{1+x^2+\log(1+x^2)} \\ &= -4x\frac{1+\frac{1}{1+x^2}}{1+x^2+\log(1+x^2)}.\end{aligned}$$

Setting $\frac{d}{dx}\log f(x) = 0$ gives $x = 0$, and as this is clearly a maximum, the mode is at $x = 0$. Computing the second derivative, keeping in mind that we only need to know its value when $x = 0$, we get

$$\frac{d^2}{dx^2}\log f(x) = -4\frac{1+\frac{1}{1+x^2}}{1+x^2+\log(1+x^2)} - 4x\cdot g(x)$$

for some continuous function $g(x)$, showing that the value of the second derivative at $x = 0$ is -8.

The normal distribution with expectation 0 and with the second derivative at 0 of the logarithm of its density equal to -8 is the one with density

$$h(x) \propto \exp\left(-\frac{8}{2}x^2\right)$$

so that the approximative normal probaiblity is the one with expectation 0 and precision 8, i.e., variance $\frac{1}{8} = 0.125$.

(b) We get

$$f(0) \approx h(0) = \sqrt{\frac{8}{2\pi}}\exp\left(-\frac{8}{2}\cdot 0^2\right) = \frac{2}{\sqrt{\pi}}.$$

As $f(0) = C$, this gives us $C \approx \frac{2}{\sqrt{\pi}}$.

4. (a) We have

$$\begin{aligned}\pi(\lambda_i \mid y_i, \alpha, \beta) &\propto \pi(y_i \mid \lambda_i, \alpha, \beta)\pi(\lambda_i \mid \alpha, \beta) \\ &\propto \lambda_i \exp(-\lambda_i y_i)\lambda_i^{\alpha-1}\exp(-\beta\lambda_i) \\ &\propto \lambda_i^{\alpha}\exp(-(\beta+y_i)\lambda_i),\end{aligned}$$

so that the posterior is a Gamma distribution with parameters $\alpha + 1$ and $\beta + y_i$.

(b) We have

$$\begin{aligned}\pi(y_i \mid \alpha, \beta) &= \frac{\pi(y_i \mid \lambda_i, \alpha, \beta)\pi(\lambda_i \mid \alpha, \beta)}{\pi(\lambda_i \mid y_i, \alpha, \beta)} \\ &= \frac{\lambda_i \exp(-\lambda_i y_i)\frac{\beta^\alpha}{\Gamma(\alpha)}\lambda_i^{\alpha-1}\exp(-\beta\lambda_i)}{\frac{(\beta+y_i)^{\alpha+1}}{\Gamma(\alpha+1)}\lambda_i^{\alpha}\exp(-(\beta+y_i)\lambda_i)} \\ &= \frac{\alpha\beta^\alpha}{(\beta+y_i)^{\alpha+1}}.\end{aligned}$$

(c) We have

$$\begin{aligned}
\pi(\alpha,\beta \mid y_1,\dots,y_n) &\propto \prod_{i=1}^n \pi(y_i \mid \alpha,\beta)\pi(\alpha,\beta) \\
&\propto \prod_{i=1}^n \frac{\alpha\beta^\alpha}{(\beta+y_i)^{\alpha+1}} \cdot \frac{1}{\beta(\alpha+1)^2} \\
&\propto \frac{\alpha^n\beta^{n\alpha-1}}{(\alpha+1)^2}\prod_{i=1}^n(\beta+y_i)^{-(\alpha+1)},
\end{aligned}$$

so that we can write

$$\begin{aligned}
&\log(\pi(\alpha,\beta \mid y_1,\dots,y_n)) + C \\
= \; & n\log\alpha - 2\log(\alpha+1) + (n\alpha-1)\log\beta - (\alpha+1)\sum_{i=1}^n(\beta+y_i)
\end{aligned}$$

(d) Writing

$$\begin{aligned}
&\pi(\alpha,\beta,\lambda_1,\dots,\lambda_n \mid y_1,\dots,y_n) \\
= \; & \pi(\alpha,\beta \mid y_1,\dots,y_n)\prod_{i=1}^n \pi(\lambda_i \mid, y_i,\alpha,\beta),
\end{aligned}$$

we can simulate from this distribution by first simulating from $\pi(\alpha,\beta \mid y_1,\dots,y_n)$ and then simulating from each $\pi(\lambda_i \mid y_i,\alpha,\beta)$, using (a). To simulate from $\pi(\alpha,\beta \mid y_1,\dots,y_n)$ we use the results from (c). One of several possible methods is the following: First, transform from the variables (α,β) to new variables (a,b), setting $\alpha = e^a$ and $\beta = e^b$; this gives us a probability distribution defined on all of $\mathbb{R}^2$, where the logaritm of its density is given, up to a constant, by

$$(n+1)a + b - 2\log(e^a+1) + (ne^a-1)e^b - (e^a+1)\sum_{i=1}^n(e^b+y_i).$$

Numerical optimization of this function can give you a bivariate normal distribution approximating it. This can again be used to find parameters for an MCMC simulation method, or possibly a rejection sampling algorithm. A brute-force simulation algorithm could also be employed, using computed values of the function above on a grid.

(e) If the toys that lasted longest and shortest were numbered i and j, respectively, then one could count the number of rows in A where $\frac{1}{\lambda_i} > \frac{2}{\lambda_j}$ and divide by the total number of rows in A; this would give an approximation to the probability in question.

5. (a) A possible simulation method would be Gibbs sampling: Alternatively simulating from $\pi(\theta_1 \mid \theta_2)$ and $\pi(\theta_2 \mid \theta_1)$.

(b) When θ_2 is fixed, the function is an exponential of a second-degree polynomial in θ_1. Completing the square, we get

$$\begin{aligned}
\pi(\theta_1 \mid \theta_2) &\propto \exp\left(-\theta_1^2\theta_2 + \theta_1\log\theta_2\right) \\
&\propto \exp\left(-\theta_2\left(\theta_1^2 - \frac{\log\theta_2}{2\theta_2}\theta_1\right)\right) \\
&\propto \exp\left(-\frac{2\theta_2}{2}\left(\theta_1 - \frac{\log\theta_2}{2\theta_2}\right)^2\right),
\end{aligned}$$

which means that $\pi(\theta_1 \mid \theta_2)$ is a Normal distribution with expectation $\frac{\log \theta_2}{2\theta_2}$ and precision $2\theta_2$.

(c) We can write

$$\pi(\theta_2 \mid \theta_1) \propto \exp(-\theta_1^2\theta_2 + \theta_1 \log \theta_2) \propto \theta_2^{\theta_1} \exp(-\theta_1^2\theta_2),$$

which shows that $\pi(\theta_2 \mid \theta_1)$ is a Gamma distribution with parameters $\theta_1 + 1$ and θ_1^2.