Introduction to Bayesian Econometrics, Second Edition

Answers to Exercises, Chapter 2

1. First, since AB^c, A^cB, and AB are disjoint,

$$P(A) = P(AB^c \cup AB) = P(AB^c) + P(AB) \Rightarrow P(AB^c) = P(A) - P(AB) \tag{1}$$

$$P(B) = P(AB \cup A^cB) = P(AB) + P(A^cB) \Rightarrow P(A^cB) = P(B) - P(AB) \tag{2}$$

$$P(A \cup B) = P(AB^c \cup A^cB \cup AB) = P(AB^c) + P(A^cB) + P(AB).$$

Therefore,

$$\begin{aligned} P(A \cup B) &= P(A) - P(AB) + P(B) - P(AB) + P(AB) \quad \text{by substituting (1) and (2)} \\ &= P(A) + P(B) - P(AB). \end{aligned}$$

Second,

Event	Gain
AB^c	$(1-p_1)S_1 - p_2S_2 - p_3S_3 + (1-p_4)S_4$
A^cB	$-p_1S_1 + (1-p_2)Ss_2 - p_3S3 + (1-p_4)S_4$
AB	$(1-p_1)S_1 + (1-p_2)S2 + (1-p_3)S_3 + (1-p_4)S_4$
$(A \cup B)^c$	$-p_1S_1 - p_2S_2 - p_3S_3 - p_4S_4$

By coherence, p_1, p_2, p_3, and p_4 must satisfy

$$\begin{vmatrix} 1-p_1 & -p_2 & -p_3 & 1-p_4 \\ -p_1 & 1-p_2 & -p_3 & 1-p_4 \\ 1-p_1 & 1-p_2 & 1-p_3 & 1-p_4 \\ -p_1 & -p_2 & -p_3 & -p_4 \end{vmatrix} = 0.$$

Therefore,

$$p_4 = p_1 + p_2 - p_3 \Leftrightarrow P(A \cup B) = P(A) + P(B) - P(AB).$$

2. When the random variable is the number of heads, $y = \sum y_i$, y has the binomial distribution,

$$p(y|n, \theta) = \binom{n}{y} \theta^y \, (1 - \theta)^{n-y}, \quad y = 0, \ldots, n.$$

By Bayes theorem,

$$\pi(\theta|n, y) \propto \theta^y \, (1 - \theta)^{n-y} \, \pi(\theta).$$

When the random variable is the number of trials n necessary to obtain y heads, we have the negative binomial distribution,

$$p(n|y, \theta) = \binom{n-1}{y-1} \theta^y \, (1 - \theta)^{n-y}, \quad n = y, y + 1, \ldots, .$$

And, by Bayes theorem,

$$\pi(\theta|n, y) \propto \theta^y \, (1 - \theta)^{n-y} \, \pi(\theta).$$

Accordingly, if we have the same prior distribution over θ for the two experiments, Bayesian statistical inference about θ does not depend on whether n or y is fixed: given that an experiment yielded y successes in n trials, inferences from the Bayesian viewpoint do not depend on the details of the experiment. This is not true for the frequentist viewpoint. For example, in the binomial experiment, y/n is an unbiased estimator for θ, but it is a biased estimator in the negative binomial case.

Bayesian inference is consistent with the *likelihood principle*, which states that all relevant experimental information about a parameter is contained in the likelihood function for the observed random variable and that two likelihood functions contain the same information about the parameter if they are proportional to each other as functions of the parameter. See Berger (1985, secs. 1.6.4, 1.7) for an excellent discussion of the likelihood principle and its relation to the sufficiency and weak conditionality principles.

3. (a) The likelihood function is

$$\begin{aligned} p(y_1, \ldots, y_n|\theta) &= \frac{\theta^{y_1}\, e^{-\theta}}{y_1!} \cdots \frac{\theta^{y_n} e^{-\theta}}{y_n!} \\ &= \frac{1}{\prod y_i!}\, \theta^{\sum y_i}\, e^{-n\theta}. \end{aligned}$$

The prior distribution is

$$\pi(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta}.$$

Therefore, the posterior distribution is

$$\begin{aligned} \pi(\theta|y_1, \ldots, y_n) &\propto \frac{1}{\prod y_i!}\, \theta^{\sum y_i}\, e^{-n\theta}\, \frac{\beta^\alpha}{\Gamma(\alpha)}\, \theta^{\alpha-1}\, e^{-\beta\theta} \\ &\propto \theta^{\alpha+\sum y_i - 1}\, e^{-(\beta+n)\theta}, \end{aligned}$$

which is a gamma distribution.

(b) To find θ that maximizes the likelihood function, take the derivative of the likelihood function with respect to θ and set it equal to 0. The MLE for θ

satisfies this first order condition,

$$
\begin{aligned}
p'(y_1,\ldots,y_n|\theta) &= \sum y_i \theta^{\sum y_i - 1}\, e^{-n\theta} - n\theta^{\sum y_i}\, e^{-n\theta} = 0 \\
&\Rightarrow \sum y_i - n\theta = 0 \\
\theta^{MLE} &= \frac{\sum y_i}{n} = \bar{y}.
\end{aligned}
$$

(c)

$$
\begin{aligned}
E(\theta|y_1,\ldots,y_n) &= \frac{\alpha + \sum y_i}{\beta + n} \\
&= \frac{\alpha}{\beta + n} + \frac{\sum y_i}{\beta + n} \\
&= \left(\frac{\beta}{\beta + n}\right) \frac{\alpha}{\beta} + \left(\frac{n}{\beta + n}\right) \bar{y}.
\end{aligned}
$$

(d)

$$
\lim_{n \to \infty} \left(\frac{\beta}{\beta + n}\right) = 0.
$$

4. (a) First, the likelihood function is

$$
\begin{aligned}
f(y_1,\ldots,y_n|\theta) &= \theta\, e^{-\theta y_i} \cdots \theta\, e^{-\theta y_n} \\
&= \theta^n\, e^{-\theta \sum y_i}.
\end{aligned}
$$

Second, the prior distribution is

$$
\pi(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\, \theta^{\alpha - 1}\, e^{-\beta\theta}.
$$

Therefore, the posterior distribution is

$$\begin{aligned}\pi(\theta|y_1,\ldots,y_n) &\propto \theta^n\, e^{-\theta\sum y_i}\,\frac{\beta^\alpha}{\Gamma(\alpha)}\,\theta^{\alpha-1}e^{-\beta\theta}\\ &\propto \theta^{\alpha+n-1}\,e^{-(\beta+\sum y_i)\theta},\end{aligned}$$

which is a gamma distribution.

(b) To find the θ that maximizes the likelihood function, take the derivative of the likelihood function with respect to θ and set it equal to 0. The MLE for θ should satisfy the first order condition

$$\begin{aligned}f'(y_1,\ldots,y_n|\theta) &= n\,\theta^{n-1}\,e^{-\theta\sum y_i} - \sum y_i\,\theta^n\,e^{-\theta\sum y_i} = 0\\ &\Rightarrow n - \theta\sum y_i = 0\\ \theta^{MLE} &= \frac{n}{\sum y_i} = \frac{1}{\bar{y}}.\end{aligned}$$

(c)

$$\begin{aligned}E(\theta|y_1,\ldots,y_n) &= \frac{\alpha_0+n}{\beta_0+\sum y_i}\\ &= \left(\frac{\beta_0}{\beta_0+\sum y_i}\right)\frac{\alpha_0}{\beta_0} + \left(\frac{\sum y_i}{\beta_0+\sum y_i}\right)\frac{1}{\bar{y}}.\end{aligned}$$

(d)

$$\lim_{n\to\infty}\left(\frac{\beta_0}{\beta_0+\sum y_i}\right) = 0 \text{ because } y_i > 0.$$

5. (a) First, the likelihood function is

$$\begin{aligned}f(y_1,\ldots,y_n|\theta) &= \frac{1}{\theta}\cdots\frac{1}{\theta}\\ &= \theta^{-n} \quad \text{for} \quad 0 \le y_i \le \theta.\end{aligned}$$

Second, the prior distribution is

$$\pi(\theta) = \begin{cases} ak^a\theta^{-(a+1)}, & \theta \geq k, \quad a > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Set $M = \max(k, y_1, \ldots, y_n)$. Then, for $\theta \geq M$, $a > 0$, the posterior distribution is proportional to

$$\pi(\theta|y_1, \ldots, y_n) \propto \theta^{-n}\, \theta^{-(a+1)} = \theta^{-(a+n+1)},$$

which implies a Pareto distribution with parameters $(a + n, M)$.

(b) To find the θ that maximizes the likelihood, note that the likelihood θ^{-n} is strictly decreasing in θ. The MLE is therefore the smallest possible value of θ, and, since $0 \leq y_i \leq \theta$, we set $\theta^{MLE} = \max(y_1, \ldots, y_n)$.

(c) From (a), we have that the posterior distribution is Pareto with parameters $(a + n, M)$. Accordingly,

$$E(\theta|y_1, \ldots, y_n) = \frac{(a+n)\, M}{a+n-1}.$$

Thanks to John P. Burkett of the University of Rhode Island for finding an error in the original answer.