Solutions to the Exercises of Section 1.8.

1.8.1. $\mathrm{E}(Z-b)^2 = \mathrm{Var}(Z) + (\mathrm{E}Z-b)^2$ obviously takes on its minimum value of $\mathrm{Var}(Z)$ when $b = \mathrm{E}Z$.

1.8.2. We say b_0 is a median of a random variable Z if $\mathrm{P}(Z \le b_0) \ge 1/2$ and $\mathrm{P}(Z \ge b_0) \ge 1/2$. Let b_0 be any median of Z. First suppose $b > b_0$. Then,

$$|z-b_0| - |z-b| = \begin{cases} -(b-b_0) & \text{if } z \le b_0 \\ 2(z-b_0)-(b-b_0) & \text{if } b_0 < z \le b \\ b-b_0 & \text{if } z > b \end{cases} \qquad \le \begin{cases} -(b-b_0) & \text{if } b \le b_0 \\ b-b_0 & \text{if } b > b_0 \end{cases}.$$

Then, provided Z has a finite first moment, $\mathrm{E}|Z-b_0| - \mathrm{E}|Z-b| \le (b-b_0)\mathrm{P}(Z > b_0) - (b-b_0)\mathrm{P}(Z \le b_0) = (b-b_0)(1-2\mathrm{P}(Z \le b_0)) \le 0$. Similarly, for $b < b_0$, we also have $\mathrm{E}|Z-b_0| - \mathrm{E}|Z-b| \le 0$. This shows that $f(b) = \mathrm{E}|Z-b|$ is minimized at $b - b_0$.

1.8.3. **Rule:** For the decision problem with $\Theta = \mathcal{A} = \Re$ and loss function,

$$L(\theta,a) = \begin{cases} k_1|\theta-a|, & \text{if } a \le \theta, \\ k_2|\theta-a|, & \text{if } a > \theta, \end{cases}$$

where k_1 and k_2 are known positive numbers, a Bayes rule with respect to a given prior distribution is to estimate θ as the pth quantile of the posterior distribution of θ given the observations, where $p = k_1/(k_1+k_2)$.

A number b is said to be a *pth quantile* of a distribution of a random variable θ if $\mathrm{P}(\theta \le b) \ge p$ and $\mathrm{P}(\theta \ge b) \ge 1-p$. Suppose $a > b$. We are to show that $\mathrm{E}(L(\theta,a)) - \mathrm{E}(L(\theta,b)) \ge 0$, where b is a pth quantile. Since for all θ,

$$L(\theta,a) - L(\theta,b) \ge (a-b)[k_2 I(\theta \le b) - k_1 I(\theta > b)],$$

where $I(S)$ is the indicator function of the set S, we have,

$$\begin{aligned} \mathrm{E}(L(\theta,a) - L(\theta,b)) &\ge (a-b)[k_2\mathrm{P}(\theta \le b) - k_1\mathrm{P}(\theta > b)] \\ &\ge (a-b)[k_2 p - k_1(1-p)] = 0, \end{aligned}$$

as was to be shown. A similar method works to show that for $a < b$, $\mathrm{E}L(\theta,a) \ge \mathrm{E}L(\theta,b)$.

1.8.4. In the example with prior distribution (1.27) and distribution of X given θ (1.26), the Bayes estimate of θ for absolute error loss is the median of the posterior distribution of θ, given as $g(\theta|x)$ at the bottom of page 45. To find the median of this distribution, we solve for the median, m:

$$1/2 = \int_m^\infty e^{-(\theta-x)}\,d\theta = e^{(x-m)}.$$

Solving for m gives $m = x + \log(2) = x + .693\cdots$ as the Bayes estimate of θ using absolute error loss, which is to be compared to the estimate $d(x) = x+1$, the Bayes estimate using squared error loss.

1.8.5. An interval of length $2c$, say $(b-c, b+c)$, is said to be a *modal interval of length $2c$* for the distribution of a random variable θ, if $\mathrm{P}(b-c \le \theta \le b+c)$ takes on its maximum value out of all such intervals. For the loss function

$$L(\theta,a) = \begin{cases} 0 & \text{if } |\theta-a| \le c \\ 1 & \text{if } |\theta-a| > c, \end{cases}$$

$\mathrm{E}L(\theta,a) = \mathrm{P}(|\theta-a| > c) = 1 - \mathrm{P}(a-c \le \theta \le a+c)$ is minimized if a is chosen to be the midpoint of the modal interval of length $2c$. **Rule:** In the problem of estimating a real parameter θ with the above loss function, a Bayes decision rule with respect to a given prior is to estimate θ as the midpoint of the modal interval of length $2c$ of the posterior distribution of θ given the observations.

1.8.6. If τ is a prior distribution for θ with density $g(\theta)$, and if $c = \mathrm{E}w(\theta) = \int w(\theta)g(\theta)\,d\theta < \infty$, then $g^*(x) = (1/c)w(\theta)g(\theta)$ is a density of a prior distribution, call it τ^*, for θ. If d is Bayes with respect to τ

for loss $L(\theta,a)=(\theta-a)^2w(\theta)$, then d is Bayes with respect to τ^* for loss $L^*(\theta,a)=(\theta-a)^2$, because in either case d minimizes $\int(\theta-d)^2f(x|\theta)w(\theta)g(\theta)\,d\theta$. Hence, d cannot be unbiased unless $r^*(\tau^*,d)$, which is equal to $r(\tau,d)=\int\int(\theta-d(x))^2f(x|\theta)w(\theta)g(\theta)\,d\theta\,dx$, is zero.

1.8.7. (a) The joint density of θ and X is

$$\begin{aligned} h(\theta,x) &= f_X(x|\theta)g(\theta) \\ &= e^{-\theta}\theta^x/x!\cdot(\Gamma(\alpha)\beta^\alpha)^{-1}e^{-\theta/\beta}\theta^{\alpha-1} \end{aligned}$$

for $x=0,1,\ldots$ and $\theta>0$. Hence $g(\theta|x)$ is proportional to

$$e^{-((\beta+1)/\beta)\theta}\theta^{\alpha+x-1}$$

for $\theta>0$, which makes $g(\theta|x)$ the gamma distribution $\mathcal{G}(\alpha+x,\beta/(\beta+1))$.

(b) Since the loss is squared error, we have $d_{\alpha,\beta}(x)=\mathrm{E}(\theta|x)=(\alpha+x)\beta/(\beta+1)$.

(c) If $d(x)$ is Bayes with respect to τ, then $r(\tau,d)=0$. On the other hand, $r(\tau,d)=\mathrm{E}(\theta-X)^2=\mathrm{E}(\mathrm{E}\{(\theta-X)^2|\theta\})=\mathrm{E}(\theta)$. If $\Theta=(0,\infty)$, then $\mathrm{E}(\theta)>0$, since $\theta>0$. But if $\Theta=[0,\infty)$, then $\mathrm{E}(\theta)$ can be zero, and in fact, d, or any rule d such that $d(0)=0$, is Bayes with respect to the distribution degenerate at 0. (For $\theta=0$, $\mathcal{P}(\theta)$ is defined to be degenerate at 0.)

(d) $d_{\alpha,\beta}(x)=(\alpha+x)\beta/(\beta+1)\to d(x)=x$ as $\alpha\to0$ and $\beta\to\infty$.

(e) We want to find d to minimize

$$\int_0^\infty(\theta-d)^2e^{-\theta}\theta^x(1/\theta)\,d\theta.$$

If $x=0$, then this integral is $+\infty$ unless $d=0$. Hence, $d(0)=0$. If $x>0$, then minimizing this integral is equivalent to finding d to minimize $\mathrm{E}(\theta-d)^2$ when θ has the distribution $\mathcal{G}(x,1)$, and so $d(x)=x$.

(f) Given $\epsilon>0$, let τ_ϵ be the gamma distribution $\mathcal{G}(1,\epsilon)$. Then,

$$\begin{aligned} r(\tau_\epsilon,d) &= \mathrm{E}(\theta-d(X))^2=\mathrm{E}(\mathrm{E}\{(\theta-X)^2|\theta\}) \\ &= \mathrm{E}(\mathrm{Var}(X|\theta))=\mathrm{E}(\theta)=\epsilon. \end{aligned}$$

Since the minimum Bayes risk cannot be negative, d certainly comes within ϵ of minimizing the Bayes risk. Since ϵ is arbitrary, d is extended Bayes. (The same must be true of any rule d such that $d(0)=0$.)

1.8.8. (a) Writing the joint density $h(\theta,x)=g(\theta)f(x|\theta)$, and neglecting factors that do not involve θ, we have

$$g(\theta|x)\propto\theta^{\alpha+x-1}\cdot(1-\theta)^{\beta+n-x-1},$$

which is the beta distribution, $\mathcal{B}e(\alpha+x,\beta+n-x)$.

(b) Since the loss is squared error, we have $d_{\alpha,\beta}(x)=\mathrm{E}(\theta|x)=(\alpha+x)/(\alpha+\beta+n)$.

(c) $d(x)=x/n$ is an unbiased estimate of θ; so if d were also a Bayes rule with respect to τ, we would have $r(\tau,d)=0$. But $r(\tau,d)=\mathrm{E}(\theta-X/n)^2=\mathrm{E}(\mathrm{E}\{(\theta-X)^2|\theta\})=\mathrm{E}(\theta(1-\theta))/n$, which is greater than zero because $\theta(1-\theta)>0$ for $\theta\in\Theta$. This shows that d is not Bayes.

(d) $d_{\alpha,\beta}(x)=(\alpha+x)/(\alpha+\beta+n)\to d(x)$ as $\alpha\to0$ and $\beta\to0$.

(e) Consider the measure τ on (0,1) such that $d\tau(\theta)=d\theta/(\theta(1-\theta))$. We want to minimize

$$I=\int_0^1(\theta-a)^2\theta^{x-1}(1-\theta)^{n-x-1}\,d\theta.$$

If $x=0$, then $I=\infty$ unless $a=0$. Hence, $d(0)=0$. If $x=n$, then $I=\infty$ unless $a=1$. Hence, $d(n)=1$. If $0<x<n$, then I is minimized by the mean of the $\mathcal{B}e(x,n-x)$ distribution, $d(x)=x/n$. Thus $d(x)=a=x/n$ minimizes I.

(f) Let $\epsilon>0$, and consider the prior distribution $\tau=\mathcal{B}e(\epsilon,1)$.

$$r(\tau,d)=\mathrm{E}(\theta(1-\theta))/n\le\mathrm{E}(\theta)/n=(\epsilon/(\epsilon+1))/n\le\epsilon.$$

This must come within ϵ of the minimum Bayes risk for this τ.

1.8.9. For loss $L(\theta, a) = (\theta - a)^2/(\theta(1-\theta))$ and uniform prior distribution $g(\theta) = 1$, the Bayes rule will minimize the integral I of 1.8.8(e). Hence the Bayes rule is as found there: $d(x) = x/n$. Even though d is an unbiased estimate of θ, this does not contradict Exercise 1.8.6 because $\mathrm{E}(w(\theta)) = \mathrm{E}(1/(\theta(1-\theta))) = \infty$.

1.8.10. Since the prior distribution of p_1, p_2 is $g(p_1, p_2) = 1$ on the unit square, the joint density of p_1, p_2, X, Y is proportional to $h(p_1, p_2, x, y) = f_{X,Y}(x, y|p_1, p_2)$. Hence,

$$g(p_1, p_2) \propto p_1^x(1-p_1)^{n-x}p_2^y(1-p_2)^{n-y},$$

so that the posterior distribution of p_1 and p_2 are as independent random variables with $p_1 \in \mathcal{B}e(x+1, n-x+1)$ and $p_2 \in \mathcal{B}e(y+1, n-y+1)$. The Bayes estimate of $p_1 - p_2$ is therefore

$$d(x, y) = \mathrm{E}(p_1 - p_2|x, y) = (x+1)/(n+2) - (y+1)/(n+2).$$