Solutions to the Exercises of Section 3.4.

3.4.1. We are to show $\mathrm{E}(\mathrm{median}(X_i)|T) = \overline{X}$ when $X_1, \ldots, X_n$ is a sample from $\mathcal{N}(\theta, 1)$ and $T = X_1 + \cdots + X_n$. Since $\overline{X}$ is a one-to-one function of T, $\mathrm{E}(\mathrm{median}(X_i)|T) = \mathrm{E}(\mathrm{median}(X_i)|\overline{X})$. Note that $\mathrm{median}(X_i - \overline{X}) = \mathrm{median}(X_i) - \overline{X}$ a.s. But $\overline{X}$ and the differences, $(X_1 - \overline{X}, \ldots, X_n - \overline{X})$, are stochastically independent, so $\mathrm{E}(\mathrm{median}(X_i)|\overline{X}) = \overline{X} + \mathrm{E}(\mathrm{median}(X_i - \overline{X})|\overline{X}) = \overline{X} + \mathrm{E}(\mathrm{median}(X_i - \overline{X}))$. But $E(\mathrm{median}(X_i - \overline{X})) = \mathrm{E}_\theta(\mathrm{median}(X_i)) - \mathrm{E}_\theta(\overline{X}) = \theta - \theta = 0$. Hence, $\mathrm{E}(\mathrm{median}(X_i)|T) = \overline{X}$.

3.4.2. If $X_1, \ldots, X_n$ is a sample from the uniform distribution, $\mathcal{U}(\alpha, \beta)$, then $T = (\min X_j, \max X_j)$ is a sufficient statistic for (α, β) (see bottom of page 118). Hence, since the loss function, $((\alpha + \beta)/2 - a)^2$, is a convex function of a for all (α, β), the decision rule $d_0 = \mathrm{E}(\overline{X}_n|T)$ is as good as $\overline{X}_n$. Using Exercise 3.4.6, since the distribution of $\overline{X}_n$ given $T = t$ is nondegenerate for almost all t when $n \geq 3$, and the loss is strictly convex, d_0 is an improvement over $\overline{X}_n$ when $n \geq 3$. When $n = 2$, $d_0 = \overline{X}_n$ so it is not an improvement in this case.

To find d_0, we use the following symmetry argument. The conditional distribution of the order statistics, $X_{(1)}, \ldots, X_{(n)}$, given T, that is given $X_{(1)}$ and $X_{(n)}$, has $X_{(2)}, \ldots, X_{(n-1)}$ as the order statistics of a sample of size $n - 2$ from the uniform distribution on the interval $(X_{(1)}, X_{(n)})$. Thus the conditional distribution of $\overline{X}_n$ given T is symmetric about the midpoint of the interval, namely, about the midrange, $M = (\min X_j + \max X_j)/2$. Hence, $d_0 = \mathrm{E}(\overline{X}_n|T) = M$.

3.4.3. (The conclusion should have read "..., then the maximum likelihood estimate can be taken to be a function of T." Note that if the parameter space consists of two points, $\Theta = \{1/3, 2/3\}$, and if X_1 and X_2 are independent Bernoulli, $\mathcal{B}(1, \theta)$, then $\hat{\theta}(X_1, X_2) = (X_1 + 1)/3$ is a maximum likelihood estimate of θ that is not a function of the sufficient statistic $X_1 + X_2$. This occurs because the maximum of $f(x_1, x_2|\theta)$ over θ is not achieved at a unique value of θ when $x_1 + x_2 = 1$. But we can still find a maximum likelihood estimate that is a function of T.)

If $T = t(X)$ is sufficient for θ and the factorization theorem holds, then $f_X(x|\theta) = g(t(x), \theta)h(x)$. If for a given x there is a value of θ that achieves the maximum in $f(x|\theta)$, then the same value of θ achieves the maximum in $g(t, \theta)$ where $t = t(x)$. Then if the maximum likelihood estimate exists, we can choose for each t in the range of $t(x)$ a value $\hat{\theta}(t)$ that maximizes $g(t, \theta)$. The the estimate $\hat{\theta}(t(x))$ is a maximum likelihood estimate of θ.

3.4.4. (a) Let $f_j = n_j \hat{p}_j (1 - \hat{p}_j)$. Then, setting the derivatives of logit χ^2 with respect to α and β to zero gives

$$\alpha \sum f_j + \beta \sum x_j f_j = \sum f_j \operatorname{logit} \hat{p}_j$$
$$\alpha \sum x_j f_j + \beta \sum x_j^2 f_j = \sum x_j f_j \operatorname{logit} \hat{p}_j$$

The determinant, $(\sum f_j)(\sum x_j^2 f_j) - (\sum x_j f_j)^2$, is nonnegative by Schwarz inequality. We may assume without loss of generality that the x_j are distinct (since the Y_j corresponding to equal x_j could be combined). Then the determinant is zero if and only if at most one f_j is positive. In this case, the minimum logit χ^2 estimates are not uniquely determined.

(b) For $N = 3$, $n_1 = n_2 = n_3 = 10$, $x_1 = -1$, $x_2 = 0$, $x_3 = 1$, and $Y_1 = 0$, $Y_2 = 4$, $Y_3 = 9$, the equations become,

$$33\alpha + 9\beta = 24 \log(2/3) + 9 \log(9)$$
$$9\alpha + 9\beta = 9 \log(9)$$

Hence, the minimum logit χ^2 estimates are

$$\hat{\alpha}(\mathbf{y}) = \log 2 - \log 3 = -.405\cdots$$
$$\hat{\beta}(\mathbf{y}) = -\log 2 + 3 \log 3 = 2.603\cdots.$$

To find the Rao-Blackwellized version, we need the conditional distribution of Y_1, Y_2, Y_3 given $Y_1 + Y_2 + Y_3 = 13$ and $Y_3 - Y_1 = 9$. The only vectors (y_1, y_2, y_3) of integers y_j with $0 \leq y_j \leq 10$, such that $y_1 + y_2 + y_3 = 13$ and $y_3 - y_1 = 9$ are

$$\mathbf{y} = (0, 4, 9) \quad \text{and} \quad \mathbf{y}' = (1, 2, 10).$$

When $\mathbf{Y}=\mathbf{y}$, the minimum logit χ^2 estimates are found as above. When $\mathbf{Y}=\mathbf{y}'$, the minimum logit χ^2 estimates are

$$\tilde{\alpha}(\mathbf{y}') = -2\log 2 = -1.386\cdots$$
$$\tilde{\beta}(\mathbf{y}') = 2(\log 3 - \log 2) = .811\cdots.$$

We compute

$$\mathrm{P}(\mathbf{Y}=\mathbf{y}|\mathbf{Y}=\mathbf{y} \quad \text{or} \quad \mathbf{Y}=\mathbf{y}') = \frac{\binom{10}{0}\binom{10}{4}\binom{10}{9}}{\binom{10}{0}\binom{10}{4}\binom{10}{9}+\binom{10}{1}\binom{10}{2}\binom{10}{10}} = .824\cdots$$

from which we have $\mathrm{P}(\mathbf{Y}=\mathbf{y}'|\mathbf{Y}=\mathbf{y} \quad \text{or} \quad \mathbf{Y}=\mathbf{y}') = 1-.824\cdots = .176\cdots$. The Rao-Blackwellized estimates are therefore

$$\alpha^*(13,9) = \mathrm{E}\{\alpha(\mathbf{Y})|\mathbf{T}=(13,9)\} = .824\hat{\alpha}(\mathbf{y}) + .176\tilde{\alpha}(\mathbf{y}') = -.579\cdots$$

and

$$\beta^*(13,9) = \mathrm{E}\{\hat{\beta}(\mathbf{Y})|\mathbf{T}=(13,9)\} = .824\hat{\beta}(\mathbf{y}) + .176\tilde{\beta}(\mathbf{y}') = 2.286\cdots.$$

3.4.5. (Again as in Exercise 3, the conclusion should state that there exists a nonrandomized Bayes rule that is a function of T.)

If $T=t(X)$ is sufficient for θ and the factorization theorem holds, then $f_X(x|\theta) = g(t(x),\theta)h(x)$. For a prior τ, the Bayes rule minimizes the conditional Bayes risk given $X=x$ which is proportional to

$$\int_\Theta L(\theta,d)g(t(x),\theta)\,d\tau(\theta).$$

This depends on x only through the value of $t(x)$. Hence any Bayes rule may be taken to be a function of $t(x)$, as in Exercise 3.

3.4.6. Suppose $L(\theta,a)$ is strictly convex in a for all $\theta\in\Theta$, and that the conditional distribution of $d(X)$ given $T=t$ is nondegenerate. Then from Exercise 2.8.9 we may conclude that $\mathrm{E}(L(\theta,d(X)|T=t) > L(\theta,\mathrm{E}(d(X)|T=t)) = L(\theta,\hat{\theta}(t))$, the inequality being strict. If the distribution of $d(X)$ given $T=t$ is nondegenerate for a set of t with positive probability under some θ, then in the proof of the Rao-Blackwell Theorem we may conclude that $R(\theta,d) = \mathrm{E}_\theta[\mathrm{E}(L(\theta,d(X)|T)] > \mathrm{E}_\theta[L(\theta,\hat{\theta}(T))] = R(\theta,\hat{\theta})$.