Solution to Exercises 1.

I.3.3.1 (We do this for the case when the prior is discrete so all integrals are sums.)

For the inference base $(\pi, \{f_\theta : \theta \in \Theta\}, x)$ the posterior for θ is $\pi(\theta \mid x) = \pi(\theta) f_\theta(x) / m(x)$ where $m(x) = \sum_{\theta \in \Theta} \pi(\theta) f_\theta(x)$. Then the marginal posterior for $\psi = \Psi(\theta)$ is $\pi_\Psi(\psi \mid x) = \sum_{\theta \in \Psi^{-1}\{\psi\}} \pi(\theta \mid x)$

$$= \sum_{\theta \in \Psi^{-1}\{\psi\}} \frac{\pi(\theta) f_\theta(x)}{m(x)} = \frac{\pi_\Psi(\psi)}{m(x)} \sum_{\theta \in \Psi^{-1}\{\psi\}} \frac{\pi(\theta)}{\pi_\Psi(\psi)} f_\theta(x)$$

$$= \frac{\pi_\Psi(\psi) f_\psi(x)}{m(x)}$$ where $\pi_\Psi(\psi) = \sum_{\theta \in \Psi^{-1}\{\psi\}} \pi(\theta)$ is the marginal prior of ψ, so $\pi(\theta)/\pi_\Psi(\psi) = \pi(\theta \mid \psi)$ is the conditional prior of θ given $\psi = \Psi(\theta)$, and $f_\psi(x) = \sum_{\theta \in \Psi^{-1}\{\psi\}} \pi(\theta \mid \psi) f_\theta(x)$. Now

$$\sum_{\psi \in \Psi(\Theta)} \pi_\Psi(\psi) f_\psi(x) = \sum_{\psi \in \Psi(\Theta)} \sum_{\theta \in \Psi^{-1}\{\psi\}} \pi(\theta) f_\theta(x)$$

$= \sum_{\theta \in \Theta} \pi(\theta) f_\theta(x) = m(x)$. Therefore the posterior $\pi_\Psi(\cdot \mid x)$ arises from $(\pi, \{f_\theta : \theta \in \Theta\}, x)$ by marginalizing or from the inference base $(\pi_\Psi, \{f_\psi : \psi \in \Psi(\Theta)\}, x)$.

$\boxed{\text{Ex III.2.1}}$ We have the joint likelihood for (σ, y) is

$$L(\sigma^2, y \mid x) = (\sigma^2)^{-\frac{n}{2}} \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2\} (\sigma^2)^{-\frac{1}{2}} \exp\{-\frac{y^2}{2\sigma^2}\}$$

$$= \sigma^{-(n+1)} \exp\{-\frac{1}{2\sigma^2}(\sum_{i=1}^{n} x_i^2 + y^2)\},$$

$$\log L(\sigma^2, y \mid x) = -\frac{(n+1)}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(\sum_{i=1}^{n} x_i^2 + y^2)$$

$$\frac{\partial \log L(\sigma^2, y \mid x)}{\partial \sigma^2} = -\frac{(n+1)}{2\sigma^2} + \frac{1}{2\sigma^4}(\sum_{i=1}^{n} x_i^2 + y^2) = 0$$

and so $\hat{\sigma}^2 = \frac{1}{n+1}(\sum_{i=1}^{n} x_i^2 + y^2)$ maximizes $L(\sigma^2, y \mid x)$ for each y ($L(\sigma^2, y \mid x) \to 0$ as $\sigma^2 \to \infty$).

Therefore, $L^{y}(\sigma^2 \mid x) = (\hat{\sigma}^2)^{-\frac{n+1}{2}} \exp\{\frac{-1}{2\hat{\sigma}^2}(\sum_{i=1}^{n} x_i^2 + y^2)\}$.

Now $\frac{\partial \log L(\sigma^2, y \mid x)}{\partial y} = -\frac{1}{2\sigma^2} 2y = -\frac{y}{\sigma^2} = 0$

implies $\hat{y} = 0$. Therefore $L^{\sigma^2}(\sigma^2 \mid x)$

$$= (\sigma^2)^{-\frac{n+1}{2}} \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2\},$$

$\log L^{\sigma^2}(\sigma^2 \mid x) = -\frac{(n+1)}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2$ which leads to profile MLE of σ^2 equaling $\frac{1}{n+1}\sum_{i=1}^{n} x_i^2$ but based on the usual likelihood $L(\sigma^2 \mid x)$

$= (\sigma^2)^{-\frac{n}{2}} \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2\}$ the MLE is $\frac{1}{n}\sum_{i=1}^{n} x_i^2$.

II.5.1 So now the jailer generates $U_1 \sim U(0,1)$ puts $p = U_1$ and then generates U as described in Example II.5.1. So the conditional probability given {"II is reported", "$U_1 = p$"} is

$P($"I lives"$\mid \{\cdot\}) = \frac{p}{1+p}$ and so now

$P($"I lives"$\mid$ "II is reported"$) = \int_0^1 \frac{p}{1+p}\,dp$

$= 1 - \int_0^1 (1+p)^{-1}\,dp = 1 - \log(1+p)\Big|_0^1 = 1 - \log 2.$

II.5.2 See next page from Evans (2015)

Example 2.2.2 *The Monte Hall Problem.*

In the game show Let's Make a Deal a contestant is shown three closed doors I, II and III and there is a desirable prize behind one and a goat behind each of the other doors. The contestant is asked to pick a door. By the principle of insufficient reason the contestant concludes that the probability they will win the prize by selecting any door is $1/3$ and chooses I. After the contestant picks door I, the host of the show, Monte Hall, opens door II to reveal a goat. The contestant is then offered the opportunity to switch the door they have selected to door III. Should the contestant switch or not?

The set of possible outcomes is $\Omega = \{(I,II,III),(I,III,II),(II,III,I)\}$ where the first two coordinates of $\omega \in \Omega$ indicate the doors concealing goats. Then $A = \{(II,III,I)\}$ is the event that the contestant will win the prize by not switching and $B = \{(I,II,III),(II,III,I)\}$ is the event that door II conceals a goat. Using the principle of insufficient reason $P(A) = 1/3, P(B) = 2/3$ and then, by the principle of conditional probability, $P(A\,|\,B) = P(A \cap B)/P(B) = (1/3)/(2/3) = 1/2$ and so there doesn't appear to be any reason to switch.

It is necessary to consider, however, how the information was generated. Again it is clear that $\Xi(I,II,III) = II, \Xi(I,III,II) = III$ and the value of $\Xi(II,III,I)$ is not determined. If $\Xi(II,III,I) = II$, then the relevant conditional probability of winning by not switching is $1/2$ but, if $\Xi(II,III,I) = III$, then the relevant conditional probability is 0 since $\Xi^{-1}\{II\} = \{(I,II,III)\}$. More generally, it may seem plausible that the host used a random system to generate what is reported to the contestant. So suppose there is an auxiliary variable U such that $P(U = II) = p, P(U = III) = 1 - p$ where $p \in [0,1]$ is unknown to the contestant. Then consider the information generator $\Xi^*(\omega, U)$ for the enlarged response (ω, U), which is again given by (2.1). With this information generator the unconditional probability that II will be opened by the host is $1(1/3) + 0(1/3) + p(1/3) = (1 + p)/3$ and the unconditional probability that the contestant will win by not switching and II will be opened is $p(1/3)$. Therefore, the conditional probability that the contestant will win by not switching is $p/(1+p)$ and this can be any number in $[0, 1/2]$. But this implies that the conditional probability that the contestant will win by switching is always greater than $1/2$ and so the contestant should always switch doors. Note that when $p = 1/2$, then the conditional probability the contestant will win by switching is $2/3$. Similarly, when the principle of insufficient reason is applied to p, the probability of winning by switching is $1 - 0.307 = 0.693$. So overall, based on this analysis, the contestant should always switch.

Sometimes the conclusion that the contestant should switch is reached by the following argument: the probability is $2/3$ that the initial choice conceals a goat so the probability is $2/3$ that switching will result in the contestant winning. But this ignores the way the information is generated and is only appropriate when the host randomly chooses a door based on the value of an auxiliary variable U with $P(U = II) = 1/2$. If the contestant is told how the information was generated, then the probability of winning by switching is not ambiguous but otherwise it is. It can still be concluded, however, that switching is appropriate.

This example has engendered a considerable amount of discussion among both statisticians and non-statisticians. The paper by Morgan, Chaganty, Dahiya and Dovial (1991) presents a correct analysis. ■

III.1.1 The likelihood regions are as follow

$$C_k(1) = \begin{cases} \phi & k > 1/3 \\ \{\text{hearts}\} & 1/4 < k \le 1/3 \\ \{\text{hearts, spades}\} & 1/8 < k \le 1/4 \\ \Theta & 0 \le k \le 1/8 \end{cases}$$

$C_k(2) = C_k(1)$ since $L(\theta|1) = L(\theta|2)$

$$C_k(3) = \begin{cases} \phi & k > 1/3 \\ \{\text{hearts}\} & 1/4 < k \le 1/3 \\ \Theta & 0 \le k \le 1/4 \end{cases}$$

$$C_k(4) = \begin{cases} \phi & k > 1/2 \\ \{\text{diamond}\} & 1/4 < k \le 1/2 \\ \{\text{diamond, hearts}\} & 0 < k \le 1/4 \\ \Theta & k = 0 \end{cases}$$

III.3.1 Suppose T is a mss for $\{f_\theta : \theta \in \Theta\}$. Then if U is a sufficient statistic for $\{f_\theta : \theta \in \Theta\}$ then there is a function h_U s.t. $T(x) = h_U(U(x))$ for all x. Now suppose $T'(x) = h(T(x))$ for all x where h is 1-1. Now the conditional distribution of $x \mid T'(x)$ is

$$f_\theta(x \mid T'(x)) = \frac{f_\theta(x)}{\sum_{z \in \{z : T'(z) = T'(x)\}} f_\theta(z)} = \frac{f_\theta(x)}{\sum_{z \in \{z : h(T(z)) = h(T(x))\}} f_\theta(z)}$$

since h is 1-1 $$= \frac{f_\theta(x)}{\sum_{z \in \{z : T(z) = T(x)\}} f_\theta(z)} = f_\theta(x \mid T(x))$$

Since T is sufficient Theorem III.3.2 implies that $f_\theta(x \mid T'(x))$ is the same for all θ where the conditional distribution is defined and this is defined whenever $f_\theta(x \mid T(x))$ is defined. Therefore T' is sufficient. For any other sufficient U we have $T'(x) = h(T(x)) = h(h_u(U(x)))$ so $T' = h \circ h_u(U(x))$ which implies T' is a mss.

$\boxed{\text{III.3.2}}$ We have that $L(\cdot \mid 1)$ is not proportional to $L(\cdot \mid 2), L(\cdot \mid 3), L(\cdot \mid 4)$ so $[1] = \{1\}$. Now $L(\cdot \mid 2)$ is not proportional to $L(\cdot \mid 3)$ or $L(\cdot \mid 4)$ so $[2] = \{2\}$. However, $L(\cdot \mid 3) = \frac{1}{2} L(\cdot \mid 4)$ and so $[3] = [4] = \{3, 4\}$. This defines the mss $[x]$.

III.3.3 Consider the mss $[x]$. If $z \in [x]$ then $L(\cdot \mid z) = c\,L(\cdot \mid x)$ for some $c > 0$. Therefore all data in $[x]$ give rise to the same likelihood function. Also if z and x have proportional likelihoods (so the <u>same</u> likelihoods) we have $[z] = [x]$. Therefore the value $[x]$ determines the likelihood and the likelihood $L(\cdot \mid x)$ determines $[x]$.