Solutions to the Exercises of Section 3.3.

3.3.1. (a) Independent $X_j \in \mathcal{B}(n_j, p)$ for $j = 1, \ldots, n$.

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n|p) = \prod_{j=1}^n \binom{n_j}{x_j} p^{x_j}(1-p)^{n_j-x_j} = h(\mathbf{x})p^{\sum x_j}(1-p)^{\sum n_j - \sum x_j}.$$

By the Factorization Theorem, $S = \sum X_j$ is sufficient for p. Since S is the total number of successes in $\sum n_j$ independent trials with probability p of success on each trial, $S \in \mathcal{B}(\sum n_j, p)$.

(b) Independent $X_j \in \mathcal{NB}(r_j, p)$ for $j = 1, \ldots, n$.

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n|p) = \prod_{j=1}^n \binom{r_j + x_j - 1}{x_j}(1-p)^{r_j}p^{x_j} = h(\mathbf{x})(1-p)^{\sum r_j}p^{\sum x_j},$$

shows that $S = \sum x_j$ is sufficient for p. Since S represents the number of trials required to obtain $\sum r_j$ successes in a sequence of independent trials with probability p of success on each trial, we have $S \in \mathcal{NB}(\sum r_j, p)$.

(c) Independent $X_j \in \mathcal{P}(\lambda)$ for $j = 1, \ldots, n$.

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n|\lambda) = \prod_{j=1}^n e^{-\lambda}\lambda^{x_j}/x_j! = h(\mathbf{x})e^{-n\lambda}\lambda^{\sum x_j}.$$

So $S = \sum X_j$ is sufficient for λ. The Moment Generating Function of $X \in \mathcal{P}(\lambda)$ is $M_X(t) = \exp\{\lambda(e^t - 1)\}$. Thus, the MGF of S is $M_S(t) = \prod M_{X-j}(t) = \exp\{n\lambda(e^t - 1\}$, the MGF of $\mathcal{P}(n\lambda)$.

(d) Independent $X_j \in \mathcal{G}(\alpha, \beta)$ for $j = 1, \ldots, n$.

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n|\alpha,\beta) = \prod_{j=1}^n \frac{1}{\Gamma(\alpha)\beta^\alpha} e^{-x_j/\beta}x_j^{\alpha-1} = h(\alpha,\beta)e^{\sum x_j/\beta}(\prod x_j)^{\alpha-1}.$$

Thus, $(\sum X_j, \prod X_j)$ is sufficient for (α, β); $\sum X_j$ is sufficient for β if α is known; and $\prod X_j$ is sufficient for α if β is known. The MGF of $\mathcal{G}(\alpha, \beta)$ is $(1-\beta t)^\alpha$, so the MGF of $\sum X_j$ is $(1-\beta t)^{n\alpha}$, which is the MGF of $\mathcal{G}(n\alpha, \beta)$.

(e) Independent $X_j \in \mathcal{B}e(\alpha, \beta)$ for $j = 1, \ldots, n$.

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n|\alpha,\beta) = \prod_{j=1}^n h(\alpha,\beta)x_j^{\alpha-1}(1-x_j)^{\beta-1} = h(\alpha,\beta)^n(\prod x_j)^{\alpha-1}(\prod(1-x_j))^{\beta-1}.$$

Thus, $(\prod X_j, \prod(1-X_j))$ is sufficient for (α, β); $\prod X_j$ is sufficient for α if β is known; and $\prod(1-X_j)$ is sufficient for β if α is known.

3.3.2. The joint probability mass function of $X_1, \ldots, X_n$ is

$$\begin{aligned} f_{X_1,\ldots,X_n}(x_1,\ldots,x_n|\theta,p) &= \prod_{j=1}^n (1-p)p^{x_j-\theta}\mathrm{I}(x_j \in \{\theta, \theta+1, \ldots\}) \\ &= (1-p)^n p^{-n\theta}p^{\sum x_j}\mathrm{I}(\min x_j \in \{\theta, \theta+1, \ldots\}) \\ &\qquad \cdot \prod_{i=1}^n \mathrm{I}(x_i \in \{\min x_j, \min x_j + 1, \ldots\}). \end{aligned}$$

This is of the form $h(\mathbf{x})g_1(\sum x_j, p)g_2(\min x_j, \theta)$, so by the Factorization Theorem, $(\min X_j, \sum X_j)$ is sufficient for (θ, p). Moreover, if p is known, then $\min X_j$ is sufficient for θ, and if θ is known, then $\sum X_j$ is sufficient for p.

3.3.3. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be i.i.d from a distribution $F(x|\theta)$. Then for every permutation π of $(x_1, \ldots, x_n)$, the distribution of $\pi\mathbf{X}$ is the same as the distribution of $\mathbf{X}$. Let $\mathbf{T} = (X_{(1)}, \ldots, X_{(n)})$ be the vector of order statistics. Then for any measurable set A in E_n,

$$\mathrm{P}(\mathbf{X} \in A) = \frac{1}{n!}\sum_{\pi} \mathrm{P}(\pi\mathbf{X} \in A) = \frac{1}{n!}\sum_{\pi} \mathrm{P}(\pi\mathbf{T} \in A)$$

where the sumation is over the set of all $n!$ permutations. This shows that the conditional distribution of $(X_1, \ldots, X_n)$ given the order statistics, $(X_{(1)}, \ldots, X_{(n)})$, is uniform on the set of all permutations of $(X_{(1)}, \ldots, X_{(n)})$. Since this conditional distribution does not depend on θ, it follows that $\mathbf{T}$ is sufficient for θ.

In fact, the above proof works for a family of exchangeable distributions. (The distribution of a sequence, $\mathbf{X} = (X_1, \ldots, X_n)$ of random variables is said to be exchangeable if for every permutation π, of $(x_1, \ldots, x_n)$, the distribution of $\pi\mathbf{X}$ is the same as the distribution of $\mathbf{X}$.)

3.3.4. First, suppose that the parameter space is $\Theta = \{(\mu, \Sigma) : \Sigma \text{ nonsingular}\}$. Then the joint density of $\mathbf{X}_1, \ldots, \mathbf{X}_n$ exists and is of the form

$$f(\mathbf{x}_1, \ldots, \mathbf{x}_n|\mu, \Sigma) = (2\pi\det(\Sigma))^{-n/2}\exp\{-U/2\},$$

where

$$\begin{aligned} U &= \sum_{i=1}^{n}(\mathbf{x}_i - \mu)^T\Sigma^{-1}(\mathbf{x}_i - \mu) \\ &= \sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}} + \overline{\mathbf{x}} - \mu)^T\Sigma^{-1}(\mathbf{x}_i - \overline{\mathbf{x}} + \overline{\mathbf{x}} - \mu) \\ &= \sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})^T\Sigma^{-1}(\mathbf{x}_i - \overline{\mathbf{x}}) + n(\overline{\mathbf{x}} - \mu)^T\Sigma^{-1}(\overline{\mathbf{x}} - \mu). \end{aligned}$$

The last term depends on $\mathbf{x}_1, \ldots, \mathbf{x}_n$ only through $\overline{\mathbf{x}}$. The factorization theorem will imply that $\overline{\mathbf{X}}, \mathbf{S}^2$ is sufficient for μ, Σ when we show the first term depends on $\mathbf{x}_1, \ldots, \mathbf{x}_n$ only through $\mathbf{S}^2$. Let σ^{hj} denote the (h, j)-element of Σ^{-1}. Then

$$\begin{aligned} \sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})^T\Sigma^{-1}(\mathbf{x}_i - \overline{\mathbf{x}}) &= \sum_{i=1}^{n}\sum_{h}\sum_{j}(x_{ih} - \overline{x}_{\cdot h})\sigma^{hj}(x_{ij} - \overline{x}_{\cdot j}) \\ &= \sum_{h}\sum_{j}\sigma^{hj}\sum_{i=1}^{n}(x_{ih} - \overline{x}_{\cdot h})(x_{ij} - \overline{x}_{\cdot j}). \end{aligned}$$

The inside sum is n times the (h, j)-element of $\mathbf{S}^2$.

Now suppose that Σ is allowed to be singular in Θ. First note that Σ and $\mathbf{S}^2$ have the same null space almost surely; more precisely, $\mathbf{a}^T\Sigma\mathbf{a} = 0$ if and only if $\mathbf{a}^T\mathbf{S}^2\mathbf{a} = 0$ w.p. 1. This is because $\mathrm{Var}(\mathbf{a}^T\mathbf{X}) = \mathbf{a}^T\Sigma\mathbf{a}$ and this is zero if and only if $\mathbf{a}^T\mathbf{X}_i = \mathbf{a}^T\mu$ w.p. 1 for all i, so that $\mathbf{a}^T\mathbf{S}^2\mathbf{a} = (1/n)\sum_{i=1}^{n}\mathbf{a}^T(\mathbf{X}_i - \mu)(\mathbf{X}_i - \mu)^T\mathbf{a} = 0$ w.p. 1. Hence when $\mathbf{S}^2$ is nonsingular, the distribution of $\mathbf{X}_1, \ldots, \mathbf{X}_n$ given $\overline{\mathbf{X}}, \mathbf{S}^2$ is the same as in the nonsingular Σ case and does not depend on μ, Σ. If $\mathbf{S}^2$ is singular, say of rank $r < k$, then the distribution of $\mathbf{X}_1, \ldots, \mathbf{X}_n$ given $\overline{\mathbf{X}}, \mathbf{S}^2$ is found as follows. There exists an r-dimensional subvector $\mathbf{X}^*$ of $\mathbf{X}$ with nonsingular subcovariance matrix $\mathbf{S}^{*2}$ of $\mathbf{S}^2$. The conditional distribution of these components of $\mathbf{X}_1, \ldots, \mathbf{X}_n$ given $\overline{\mathbf{X}}, \mathbf{S}^2$ is the same as in the nonsingular Σ case but with dimension r. The remaining components of the vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n$ may be determined exactly from these and $\overline{\mathbf{X}}, \mathbf{S}^2$. This distribution is described idependently of μ, Σ.