Appendix F
Some Results from Classical Statistics

In this section we provide some definitions and state some theorems (without proof) from classical statistics. More details can be found in Schervish (1995). Let $\boldsymbol{y} = [y_1, \ldots, y_n]^{\mathrm{T}}$ be a random sample from $p(y \mid \theta)$.

Definition. The statistic $T(\boldsymbol{Y})$ is *sufficient* for θ within a family of probability distributions $p(y \mid \theta)$ if $p(\boldsymbol{y} \mid T(\boldsymbol{y}))$ does not depend upon θ.

Theorem. *The Fisher–Neyman factorization theorem states that $T(\boldsymbol{Y})$ is sufficient for θ if and only if*

$$p(\boldsymbol{y} \mid \theta) = g[T(\boldsymbol{y}) \mid \theta] \times h(\boldsymbol{y}).$$

Intuitively, all of the information in the sample with respect to θ is contained in $T(\boldsymbol{Y})$.

Definition. The statistic $T(\boldsymbol{Y})$ is *minimal sufficient* for θ within a family of probability distributions $p(y \mid \theta)$ if no further reduction from T is possible while retaining sufficiency.

Theorem. *The Lehmann–Scheffé theorem states that if $T(\boldsymbol{Y})$ satisfies the following property: for every pair of sample points $\boldsymbol{y}, \boldsymbol{z}$ the ratio $p(\boldsymbol{y} \mid \theta)/p(\boldsymbol{z} \mid \theta)$ is free of θ if and only if $T(\boldsymbol{y}) = T(\boldsymbol{z})$, then T is minimal sufficient.*

Example. Let $Y_1, \ldots, Y_n$ be independent and identically distributed from the one-parameter exponential family of distributions:

$$p(y \mid \theta) = \exp[\theta T(y) - b(\theta) + c(y)]$$

for functions $b(\cdot)$ and $c(\cdot)$. Then $\sum_{i=1}^{n} T(Y_i)$ is sufficient for θ by the factorization theorem and minimal sufficient by the Lehmann–Sheffé theorem.

Definition. A statistic $V = V(\boldsymbol{y})$ is *ancilliary* for θ within a family of probability distributions $p(y \mid \theta)$ if its distribution does not depend on θ.

J. Wakefield, *Bayesian and Frequentist Regression Methods*, Springer Series in Statistics, DOI 10.1007/978-1-4419-0925-1_18,

Definition. If a minimal sufficient statistic is $\boldsymbol{T} = [T_1, T_2]$ and T_2 is ancillary then T_1 is called *conditionally sufficient given* T_2.

Example. In a linear normal linear regression with covariate x, suppose that x has distribution $p(x)$ and

$$Y_i \mid X_i = x_i \sim \mathrm{N}(\beta_0 + \beta_1 x_i, \sigma^2), \qquad i = 1, \ldots, n.$$

Then, letting $\boldsymbol{x} = [x_1, \ldots, x_n]^{\mathrm{T}}$, $\boldsymbol{y} = [y_1, \ldots, y_n]^{\mathrm{T}}$ and $\boldsymbol{\beta} = [\beta_0, \beta_1]^{\mathrm{T}}$ the distribution for the data is

$$p(\boldsymbol{x}, \boldsymbol{y} \mid \boldsymbol{\beta}, \sigma^2) = p(\boldsymbol{x})(2\pi\sigma)^{-n/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2\right].$$

The sufficient statistic for $[\boldsymbol{\beta}, \sigma^2]$ is

$$\boldsymbol{S} = \left[\widehat{\boldsymbol{\beta}}, \widehat{\sigma}^2, \sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2\right],$$

with the last two components being an ancillary statistic.

Definition. A statistic T is *complete* if for every real-valued function $g(\cdot)$, $\mathrm{E}[g(T)] = 0$ for every θ implies $g(T) = 0$.

Definition. Suppose we wish to estimate $\phi = \phi(\theta)$ based on $Y \mid \theta \sim p(\cdot \mid \theta)$. An unbiased estimator $\widehat{\phi}$ of ϕ is a *uniformly minimum-variance unbiased estimator (UMVUE)* if, for all other unbiased estimators $\widetilde{\phi}$,

$$\mathrm{var}(\widehat{\phi}) \leq \mathrm{var}(\widetilde{\phi})$$

for all θ.

Lemma. *If T is complete then $\phi(\theta)$ admits at most one unbiased estimator $\widehat{\phi}(T)$ depending on T.*

Theorem (Rao–Blackwell–Lehmann–Scheffé). *Let $T = T(\boldsymbol{Y})$ be complete and sufficient for θ. If there exists at least one unbiased estimator $\widetilde{\phi} = \widetilde{\phi}(\boldsymbol{Y})$ for $\phi(\theta)$ then there exists a unique UMVUE $\widehat{\phi} = \widehat{\phi}(T)$ for $\phi(\theta)$, namely,*

$$\phi(\boldsymbol{Y}) = \mathrm{E}[\widetilde{\phi}(\boldsymbol{Y} \mid T].$$

Corollary. *Let $T = T(\boldsymbol{Y})$ be complete and sufficient for θ. Then any function $g(T)$ is the UMVUE of its expectation $\mathrm{E}[g(T)] = \phi(\theta)$.*

Theorem. *The Cramér–Rao lower bound for any unbiased estimator $\widehat{\phi}$ of a scalar function of interest $\phi = \phi(\theta)$ is*

$$\mathrm{var}(\widehat{\phi}) \geq -\frac{\left[\phi'(\theta)\right]^2}{\mathrm{E}\left[\frac{\partial^2 l}{\partial \theta^2}\right]},$$

where $l(\theta) = \sum_{i=1}^{n} \log p(y_i \mid \theta)$*, is the log of the joint distribution, viewed as a function of* θ*. Equality holds if and only if* $p(y \mid \theta)$ *is a one-parameter exponential family.*