9 Parametric Inference I: Parameters and Models

Suppose we have reason to believe that the distribution from which the data were drawn has a density $f(x;\ \theta)$ where θ is an unknown parameter or a vector of unknown parameters. The problem of inference then reduces to the problem of estimating the parameter θ. In some cases, we might be interested in some function $T(\theta)$. In the next few chapters we discuss how to infer θ and $T(\theta)$.

9.1 Parametric Models

Suppose we have iid data $X_1, \ldots, X_n$ whose pdf (or mass function) is contained in the set

$$\mathcal{M} = \{f(x;\theta);\ \theta \in \Theta\}$$

where θ is a real number or a vector of real numbers. We call θ a parameter and Θ is called the parameter space. The set of pdf's $\mathcal{M}$ is called a *parametric statistical model*. Our goal is to estimate θ or some function of θ.

EXAMPLE 9.1 *Let* $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$. *The parameter is* p *and the parameter space is* $[0, 1]$.

EXAMPLE 9.2 *Let* $X_1, \ldots, X_n \sim \text{Normal}(\mu, \sigma^2)$. *The parameter is* $\theta = (\mu, \sigma)$ *is* $\Theta = \{(\mu, \sigma) :\ \mu \in \mathcal{R},\ \sigma > 0\}$. *Suppose we are interested in estimating the mean of the distribution. We can write the quantity of interest as* $T(\mu, \sigma) = \mu$. *On the other hand, suppose that* X_i *is the outcome of a blood test and suppose we are interested in* τ, *defined as the fraction of the population whose test score is larger than 1. How do we express this? Let* Z *denote a standard Normal random variable. Then*

$$\begin{aligned}
\tau &= P(X > 1) \\
&= 1 - P(X < 1) \\
&= 1 - P\left(\frac{X-\mu}{\sigma} < \frac{1-\mu}{\sigma}\right) \\
&= 1 - P\left(Z < \frac{1-\mu}{\sigma}\right) \\
&= 1 - \Phi\left(Z < \frac{1-\mu}{\sigma}\right).
\end{aligned}$$

The parameter of interest is $\tau = T(\mu, \sigma)$ where $T(\mu, \sigma) = 1 - \Phi((1-\mu)/\sigma)$. The important point is that the quantity of interest can be thought of as a function of the parameter θ.

EXAMPLE 9.3 *Recall that X has a* Gamma(α, β) *distribution if*

$$f(x;\, \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, \quad x > 0$$

where $\alpha, \beta > 0$ and

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$$

is the Gamma function. The Gamma distribution is sometimes used to model lifetimes of people, animals, and electronic equipment. Suppose we want to estimate the average lifetime. The mean of X_i is $T(\alpha, \beta) = \alpha/\beta$; see chapter 3 for a proof.

9.2 Review of Mean Squared Error and Consistency

Let $\widehat{\theta}_n$ be an estimator of θ. Recall that $\widehat{\theta}_n$ is consistent if $\widehat{\theta}_n \xrightarrow{p} \theta$. Let $\overline{\theta}_n = E_\theta(\widehat{\theta}_n)$ be the expectation of $\widehat{\theta}_n$. If $\overline{\theta}_n = \theta$ we say the estimator is *unbiased* otherwise it is biased and we define $b_n \equiv \overline{\theta}_n - \theta$ to be the bias. Bias is not necessarily a bad thing but we would like the bias to get small as the sample size increases. Recall that the MSE (mean squared error) is defined by $MSE = E_\theta(\widehat{\theta}_n - \theta)^2$. Earlier we showed that

$$MSE = \mathrm{Var}_\theta(\widehat{\theta_n}) + \mathrm{bias}^2.$$

If $\mathrm{Var}_\theta(\widehat{\theta_n})$ and bias both tend to 0, then $\widehat{\theta}_n \xrightarrow{q.m.} \theta$ and hence $\widehat{\theta}_n \xrightarrow{p} \theta$. So one way to show that $\widehat{\theta}_n$ is consistent is to show that the bias and variance go to 0.