2 Point Estimation for Parametric Families of Probability Distributions

2.1 Parametric Families

We now shift gears to discuss the statistical idea of point estimation. We will be consider the problem of *parametric* point estimation, so we will first need to understand what is a parametric family of probability distributions.

Here a *parameter space* Θ will be a subset of $\mathbb{R}^k$ for some k.

Definition 2. A *parametric* family of probability distributions is a collection of probability density functions (or probability mass functions) on $\mathbb{R}^n$ indexed by the parameter space Θ, that is, a collection of densities of the form $\{f(x;\theta) : \theta \in \Theta\}$.

Given a parametric family, each $\theta \in \Theta$ uniquely specifies a probability density function $f(x;\theta)$.

Example 1 (Normal family). The family of normal probability densities has parameter space $\Theta = \mathbb{R} \times (0, \infty)$. In this case, the parameter is the ordered pair $\theta = (\mu, \sigma^2)$, and the density specified by θ is (in the case of an i.i.d. sample $(X_1, \ldots, X_n)$ of size n)

$$f(x;\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2} .$$

Suppose that the distribution of the random vector $\boldsymbol{X}$ has a density belonging to a parametric family, that is, $\boldsymbol{X} \sim f(x;\theta)$ for some $\theta \in \Theta$. Given a function $g : \mathbb{R}^n \to \mathbb{R}$, we write $E_\theta(g(\boldsymbol{X}))$ to indicate that we are taking an expectation with respect to the density $f(\boldsymbol{x};\theta)$. Similarly, we write $P_\theta(\boldsymbol{X} \in A)$ to indicate we are computing a probability using the density $f(\boldsymbol{x};\theta)$. To be precise,

$$P_\theta(\boldsymbol{X} \in A) = \int\limits_A f(\boldsymbol{x};\theta) d\boldsymbol{x}$$

$$E_\theta(g(\boldsymbol{X})) = \int\limits_{\mathbb{R}^n} g(\boldsymbol{x}) f(\boldsymbol{x};\theta) \boldsymbol{x} .$$

A parametric family can have more than one parameterization. For example, we can parameterize the exponential family by

$$\mu \mapsto \frac{1}{\mu} e^{-x/\mu} \mathbf{1}\{x \geq 0\}, \quad \mu > 0.$$

Alternatively, it is sometimes parameterized by

$$\lambda \mapsto \lambda e^{-\lambda x} \mathbf{1}\{x \geq 0\}, \quad \lambda > 0.$$

When we talk about a parametric family of probability distributions, we should be sure to specify explicitly which parameterization we are using.

2.2 Statistical Inference

The problem of parametric statistical inference is the following: We observe data $\boldsymbol{X} = (X_1, \ldots, X_n)$ which has a joint density belonging to some parametric family $\{f(x;\theta) : \theta \in \Theta\}$. More exactly, the joint density of $\boldsymbol{X}$ is $f(x;\theta_0)$, where $\theta_0 \in \Theta$. We often say that θ_0 is the "true" parameter value, in that the observed random variables in fact came from a distribution having density $f(x;\theta_0)$. The situation is that *we do not know the value of* θ_0, but we would like to *infer* information about θ_0, based on $\boldsymbol{X}$.

It is intuitively clear that the data $\boldsymbol{X}$ contains information about the value of θ_0. An illustrative example is the following:

Example 2. Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be an i.i.d. sample from a uniform random variable on the interval $[\theta_0, \theta_0 + 1]$. Thus, we have the family of densities

$$\begin{aligned} f(\boldsymbol{x};\theta) &= \prod_{i=1}^{n} \mathbf{1}\{\theta \leq x_i \leq \theta + 1\} \\ &= \mathbf{1}\left\{\theta \leq \min_{1 \leq i \leq n} x_i \leq \max_{1 \leq i \leq n} x_i \leq \theta + 1\right\}, \end{aligned}$$

where $\theta \in \mathbb{R}$. If we observe $\min_{1 \leq i \leq n} X_i = a$ and $\max_{1 \leq i \leq n} X_i = b$, then it must be that $\theta_0 < a < b < \theta_0 + 1$. In particular, $b - 1 < \theta_0 < a$, and we have narrowed down the possible values of θ_0 to those in an interval of length at most 1.

2.3 Point Estimation Set-Up

In the context of point estimation, we may be interested in a function $\tau : \Theta \to \mathbb{R}^p$, and we wish to know $\tau(\theta_0)$. Again, we want to make an *estimate* of $\tau(\theta_0)$ based on $\boldsymbol{X}$.

The notion of a *statistic* is elementary, but must be stated:

Definition 3. A *statistic* is a random variable T so that $T = t(X_1, X_2, \ldots, X_n)$ for some function $t : \mathbb{R}^n \to \mathbb{R}^m$.

Thus a statistic is any random variable which is a function of the sample $(X_1, \ldots, X_n)$.

Definition 4. An *estimator* is any statistic T which is used to estimate the fixed vector $\tau(\theta_0)$. If we observe $(X_1, \ldots, X_n) = (x_1, \ldots, x_n)$ for a (non-random) vector $\boldsymbol{x} = (x_1, \ldots, x_n)$, then the value $t(\boldsymbol{x})$ is an *estimate* of $\tau(\theta_0)$ when $\boldsymbol{X} = \boldsymbol{x}$.

Notice that there is not much to the definition of an estimator, as any stupid statistic can qualify. For example, the statistic $T \stackrel{\text{def}}{=} 1$ which is always equal to 1 can be claimed to be an estimator of μ in Example 1. It is, however, a stupid estimator and others are better.

2.4 Some criterion

We briefly mention here some criterion to decide if an estimator is good. We will not say much here, but will return to this topic at a later point.

We would like an estimator T to be, "on average" centered around the number we are trying to estimate, $\tau(\theta_0)$.

Definition 5. An estimator T of $\tau(\theta_0)$ is called *unbiased* for $\tau(\theta_0)$ if

$$E_{\theta_0}(T) = \tau(\theta_0) \quad \forall \theta_0 \in \Theta \,. \tag{1}$$

We now suppose that for each n, we have a parametric family of densities $\{f(x_1, \ldots, x_n; \theta) : \theta \in \Theta\}$, where the parameter space Θ does not depend on n. This is the case when $(X_1, \ldots, X_n)$ is an i.i.d. sample from a parametric family of distributions on $\mathbb{R}$. The following definition applies in this situation.

Definition 6. Suppose that we have a sequence of estimators $\{T_n\}_{n=1}^{\infty}$, where T_n is an estimator based on a sample of size n. Then the sequence $\{T_n\}$ is *consistent* of $\tau(\theta_0)$ if $T_n \xrightarrow{\text{Pr}} \tau(\theta_0)$ for all $\theta_0 \in \Theta$.

Example 3. Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from a parametric family on $\mathbb{R}$. Suppose that for each θ, the density $f(x;\theta)$ has finite variance. Consider the function of θ given by

$$\tau(\theta) = E_\theta(X_1) \ .$$

We note that this is indeed a function of θ. Since the distribution of X_1 has density $f(x;\theta)$, the expectation of X_1 depends on θ. By the Weak Law of Large Numbers, we have that

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\text{Pr}} \tau(\theta) \, ,$$

and so the sequence of estimators $T_n \stackrel{\text{def}}{=} n^{-1}\sum_{i=1}^{n} X_i$ is consistent for $\tau(\theta)$.

We will return to these properties later.