Lecture 6 Empirical Measures and Method of Moments

Point estimation is an area in statistical inference, which concerns the question: given a sample X generated from a model $f(x|\theta)$ with an unknown parameter $\theta \in \Theta$, how should we construct an estimator $T(X)$ such that $T(X)$ is "close" to a quantity $q(\theta)$ of interest, where q is a known function? There are a number of approaches. One of them, based on *empirical frequencies*, serves as a basis for nonparametric estimation. We will introduce empirical frequencies and various versions in some special cases.

(i) (empirical frequencies) Let $X_1, ..., X_n$ be iid random variables with a common unknown distribution P (not necessarily a parametric family $\{P_\theta\}$). Assign an equal weight $1/n$ to each X_i, and define an *empirical* (*probability*) *measure* $\bar{P}_n$ based on X^n, i.e.

$$\bar{P}_n(A) = \frac{1}{n} \sum_{i=1}^{n} I_{\{X_i \in A\}}, \quad A \subset \mathcal{X}, \tag{6.1}$$

which is just the relative frequency of the n observations falling into set A. We use $\bar{P}_n$ as an estimate of P.

(ii) The empirical cumulative distribution function (cdf) $\bar{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} I_{\{X_i \le x\}}$ is an estimate of cdf $F(x) = P(X_1 \le x), \ x \in I\!R$.

(iii) To estimate a mass function $p_\xi = P(X_1 = \xi)$, $\xi \in D$ — a discrete set in $\mathcal{X}$, we use $\bar{P}_n(\xi) = S_n^{(\xi)}/n$.

(iv) To estimate a density function $f(x)$, $x \in I\!R$, consider a naive estimator (a histogram or step function)

$$\bar{f}_n(x) = \frac{1}{2h} \frac{\# \text{ of } X_i's \text{ falling in } (x-h, \ x+h]}{n},$$

which is motivated by the approximation of infinitesimal probability $f(x)\ 2h \approx P(X_1 \in (x-h, x+h])$ with a small $h > 0$. Note that the choice of h (called a *bandwidth* or *smoothing factor*) is a central issue in non-parametric estimation. Many alternative methods can be proposed to replace $\bar{f}_n$, such as kernel, spline, local polynomial, wavelet, etc.

(v) For $j = 1, ..., k$, we first estimate the jth moment $m^{(j)} = EX_1^j$ by $\bar{m}_n^{(j)} = \frac{1}{n} \sum_{i=1}^{n} X_i^j$, then estimate $q(m^{(1)}, ..., m^{(k)})$ by $q(\bar{m}_n^{(1)}, ..., \bar{m}_n^{(k)})$ with a known continuous function q. This is called *method of moments*. Estimating $VarX_1$ by $\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$ is an example following this approach. (why?)

(vi) For $\alpha \in (0,1)$, the α-th quantile q_α is defined such that $P(X_1 \leq q_\alpha) \geq \alpha$ and $P(X_1 \geq q_\alpha) \geq 1-\alpha$. Write $q_\alpha = q_\alpha(P)$. Then the α-th sample quantile, as an estimate of q_α, is defined by

$$q_\alpha(\bar{P}_n) = \begin{cases} X_{([\alpha n]+1)}, & \text{if } \alpha n \text{ is not an integer} \\ \text{any } x \in [X_{(\alpha n)},\ X_{(\alpha n+1)}], & \text{if } \alpha n \text{ is an integer} \end{cases}$$

In particular, when $\alpha = 1/2$, we have a sample median

$$m(\bar{P}_n) \triangleq q_{\frac{1}{2}}(\bar{P}_n) = \begin{cases} X_{(\frac{n+1}{2})}, & \text{if } n \text{ is odd} \\ \frac{X_{(\frac{n}{2})}+X_{(\frac{n}{2}+1)}}{2}, & \text{if } n \text{ is even} \end{cases}$$

Note:

- (v) and (vi) illustrate a general strategy: If we want to estimate a numerical characteristic of P, defined by $q(P)$ as a continuous functional of P, then $q(\bar{P}_n)$ is a natural "plug-in" estimator. The motivation is that for large n, $\bar{P}_n$ is close to P, so is $q(\bar{P}_n)$ to $q(P)$.
- Estimators (i) – (vi) are non-parametric, i.e. model-free, because they do not depend on particular forms of P.
  - ∘ Virtues: easy to construct, robust, a useful first-step when little is known about P;
  - ∘ Flaws: too flexible thus not efficient when some information about P is available, also not easy to perform simulation when Monte Carlo studies are relevant.

Example 6.1 Let $X_1, ..., X_n$ be iid random samples from a Poisson distribution with parameter θ. Both $\bar{X}$ and $\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$ are estimators by method of moments. Which one is preferable?

Example 6.2 Let $X_1, ..., X_n$ be iid random samples from $\mathcal{U}(0,\theta)$. $T_1(X^n) = 2\ \bar{X}$ is a method-of-moments estimator. When comparing T_1 with $T_2(X^n) = \frac{n+1}{n}X_{(n)}$, note that both T_1 and T_2 are *unbiased* in the sense $E_\theta T_1 = E_\theta T_2 = \theta\ \forall\ \theta > 0$. However, $Var_\theta T_1 = \frac{\theta^2}{3n} >> \frac{\theta^2}{n(n+2)} = Var_\theta T_2$ for large n. In fact, this phenomenon is generic, i.e. with large samples, method-of-moments estimators are often inferior to so-called maximum likelihood estimators (MLEs). T_2 is an unbiased modification of MLE for θ in this example. See Lecture 7 for an introduction to MLE.