Stat 400, section 6.2 Methods of Point Estimation

notes by Tim Pilachowski

"A point estimate of a parameter θ is a single number that can be regarded as a sensible value for θ… The selected statistic is called the point estimator of θ." The symbol $\hat{\theta}$ is used for both the random variable and the calculated value of the point estimate.

Ideally, the point estimator $\hat{\theta}$ is unbiased, i.e. $E(\hat{\theta}) = \theta$. In words, the sampling distribution based on the statistic has an expected value equal to the actual (but unknown) population parameter.

When we have a choice between point estimators which are all unbiased, how do we pick the one we should use? We already have one criterion. When considering the population parameter "mean = μ", both random variable X and random variable $\overline{X}$ are unbiased. However, since the variance of X $\left[V(X) = \sigma_X^2\right]$ is larger than the variance of $\overline{X}$ $\left[V(\overline{X}) = \frac{\sigma_X^2}{n}\right]$, the sample statistic x has a lower probability of being representative of the population than the sample statistic $\bar{x}$.

More importantly, how do we find a possibility in the first place? Section 6.2 formalizes this choice process by looking at two methods of picking and estimator: a) the method of moments and b) the method of maximum likelihood.

a) the method of moments

Definition 1. Given a random variables $X_1, X_2, \dots X_n$ and positive integer k, the k^{th} population moment of X is

$$m_k(X) = E\left(X^k\right), \quad k \geq 1.$$

Thus, $m_1(X) = E(X) = \mu$ and $m_2(X) = E\left(X^2\right) = \sigma^2 + \mu^2$. [For the derivation of the second assertion, see Lecture 6.1c.]

Definition 2. Given random sample values $x_1, x_2, \dots x_n$, from the sample space of a random variable X and non-negative integer k, the k^{th} sample moment S_k is

$$S_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k, \quad k \geq 1.$$

Thus, $S_1 = \bar{x}$.

It is important to note that the the k^{th} population moments of X are functions of parameter θ while the k^{th} sample moment S_k is not. Rather S_k is an average sum of powers of the sample values x_i.

Here's how to implement the method of moments: a) Set the k^{th} population moment of X, $m_k(X)$, equal to the k^{th} sample moment S_k, then b) solve the resulting equation for θ. That is, solve

$$m_k(X) = S_k, \quad 1 \leq k \leq \infty.$$

Lecture 6.1c Example F revisited. You toss a coin n times. Define a random variable $W = 0$ if a toss is tails, and $W = 1$ if a toss is heads. Use the method of moments to determine an estimator $\hat{\theta}$ for the population parameter "proportion of successes". [Although this scenario is described in terms of flipping a coin, the mathematics would be the same for any Bernoulli/binomial distribution.]

For each toss, we have a binomial probability density function.

w	0	1
$P(W = w)$	$1 - p$	p

Lecture 6.1c Example F. (continued)

Example A (see Lecture 6.1b). Use the method of moments to determine estimators for parameters mean, μ, and variance, σ^2, for random variable X for which we know a probability distribution.

A better estimator would be sample variance $s^2 = \frac{1}{n-1}\left[\sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2\right]$ because (as demonstrated in Lecture 6.1b) it is unbiased.

Dr. Millson develops a moment estimator for a uniform distribution (Lecture 23 pp. 9-15). The text covers exponential, gamma and negative binomial distributions (Examples 6.12-6.14).

b) the method of maximum likelihood.

Random variables X_1, X_2, … , X_n form a (simple) random sample of size n if they meet two (important) requirements:

1. The X_i's are independent random variables.
2. Every X_i has the same probability distribution.

Given random sample values x_1, x_2, … x_n, from the sample space of a discrete random variable X with a probability mass function $p_X(x;\theta)$ [the pmf is a function of unknown parameter θ], what is

$$P = P(X_1 = x_1, X_2 = x_2, \ldots X_n = x_n)?$$

In other words, what is the probability of getting (by random chance) the sample we actually got?
Since the X_i's are independent,

$$P = P(X_1 = x_1, X_2 = x_2, \ldots X_n = x_n) = P(X_1 = x_1) * P(X_2 = x_2) * \ldots * P(X_n = x_n).$$

Since every X_i has the same probability distribution,

$$P = P(X_1 = x_1, X_2 = x_2, \ldots X_n = x_n) = p_X(x_1;\theta) * p_X(x_2;\theta) * \ldots * p_X(x_n;\theta).$$

Because the sample values x_1, x_2, … x_n are numbers, P is a function only of parameter θ, the likelihood function, which we will designate $L(\theta)$. We want to find the value of θ that maximizes $L(\theta)$. The maximum likelihood estimator $\hat{\theta}$ will be a critical value, such that $L'(\hat{\theta}) = 0$.

short side trip: logarithmic differentiation
Let $h(\theta) = \ln[L(\theta)]$. (Domain is not a problem because $0 \le p_X(x_i;\theta) \le 1$, and we aren't interested in occasions when probability equals 0.) Then by the chain rule,

$$\frac{d}{dx}[h(\theta)] = \frac{d}{dx}(\ln[L(\theta)])$$

$$h'(\theta) = \frac{1}{L(\theta)} * L'(\theta)$$

Here's the important thing: h and L share the same critical values. Also, since $a < b \iff \ln(a) < \ln(b)$, order is preserved and both h and L will have a maximum at the same critical value!

To use the method of maximum likelihood,

Let $h(\theta) = \ln[L(\theta)]$.
Find $h'(\theta)$.
Set $h'(\theta) = 0$ and solve for θ in terms of x_1, x_2, … , x_n.

Lecture 6.1c Example F revisited. You toss a coin n times. Define a random variable Y = proportion of heads. Use the method of maximum likelihood to determine an estimator $\hat{\theta}$ for the population parameter. [Although this scenario is described in terms of flipping a coin, the mathematics would be the same for any Bernoulli/binomial distribution.]

We're looking for an estimator of the population parameter p (population proportion, or probability of success).

Recall from Lecture 6.1c, we have $Y = \frac{\text{number of heads}}{\text{number of tosses}} = \frac{X}{n} = \frac{W_1 + W_2 + \ldots + W_n}{n} = \overline{W} = p$.

That is, for each toss, we have a binomial probability density function.

w	0	1
$P(W = w)$	$1 - p$	p

$$\Rightarrow \quad p_W(w;\, p) = \begin{cases} p^w (1-p)^{1-w} & w = 0, 1 \\ 0 & \text{otherwise} \end{cases}$$

First step: Write out and simplify the likelihood function using our generic parameter θ.

notes on the proof:

$$L(\theta) = p_W(w_1;\theta) * p_W(w_2;\theta) * \ldots * p_W(w_n;\theta)$$

$$= \theta^{w_1}(1-\theta)^{1-w_1} * \theta^{w_2}(1-\theta)^{1-w_2} * \ldots * \theta^{w_n}(1-\theta)^{1-w_n}$$

$$= \theta^{w_1}\theta^{w_2}\ldots\theta^{w_n} * (1-\theta)^{1-w_1}(1-\theta)^{1-w_2}\ldots(1-\theta)^{1-w_n}$$

$$= \theta^{w_1+w_2+\ldots+w_n} * (1-\theta)^{n-(w_1+w_2+\ldots+w_n)}$$

Second step: Create $h(\theta)$ and simplify.

notes on the proof:

$$h(\theta) = \ln[L(\theta)]$$

$$= \ln\left[\theta^{w_1+w_2+\ldots+w_n} * (1-\theta)^{n-(w_1+w_2+\ldots+w_n)}\right]$$

$$= \ln\left[\theta^{w_1+w_2+\ldots+w_n}\right] + \ln\left[(1-\theta)^{n-(w_1+w_2+\ldots+w_n)}\right]$$

$$= (w_1+w_2+\ldots+w_n)\ln[\theta] + (n-(w_1+w_2+\ldots+w_n))\ln[(1-\theta)]$$

Final step: Differentiate $h(\theta)$, set equal to 0, and solve for θ.

notes on the proof:

$$h'(\theta) = (w_1+w_2+\ldots+w_n) * \frac{1}{\theta} + (n-(w_1+w_2+\ldots+w_n)) * \frac{-1}{1-\theta}$$

$$0 = \frac{w_1+w_2+\ldots+w_n}{\theta} - \frac{n-(w_1+w_2+\ldots+w_n)}{1-\theta}$$

$$\frac{n-(w_1+w_2+\ldots+w_n)}{1-\theta} = \frac{w_1+w_2+\ldots+w_n}{\theta}$$

$$\theta[n-(w_1+w_2+\ldots+w_n)] = (1-\theta)(w_1+w_2+\ldots+w_n)$$

$$\theta * n - \theta(w_1+w_2+\ldots+w_n) = w_1+w_2+\ldots+w_n - \theta(w_1+w_2+\ldots+w_n)$$

$$\theta * n = w_1+w_2+\ldots+w_n$$

$$\theta = \frac{w_1+w_2+\ldots+w_n}{n}$$

$$\hat{\theta} = \frac{\text{number of heads}}{n} = \hat{p}$$

Example B. A random variable X has an exponential probability distribution. Use the method of maximum likelihood to determine an estimator for parameter λ.

Example B extended. A random variable X has an exponential probability distribution. Determine an estimator for parameter λ^2.

We'll need the Invariance Principle:

Let $\hat{\theta}_1, \hat{\theta}_2, \ldots \hat{\theta}_n$ be the maximum likelihood estimators for parameters $\theta_1, \theta_2, \ldots \theta_n$. Then the maximum likelihood estimator of any function $h(\theta_1, \theta_2, \ldots \theta_n)$ of these parameters is the function $h(\hat{\theta}_1, \hat{\theta}_2, \ldots \hat{\theta}_n)$ of the maximum likelihood estimators.