Lecture 12
Methods of Point Estimation

Chao Song

College of Ecology

Lanzhou University

October 24, 2024

Essential terminologies

In statistics, we often consider random variables for which the functional form of the PMF or PDF is known, but the distribution depends on unknown parameters.. We take a random sample $X_1, X_2, \ldots, X_n$ from the distribution to elicit some information about the unknown parameters, say θ.

- All potential values of the parameters is called the **parameter space**;
- The statistic $u(X_1, X_2, \ldots, X_n)$ used to estimate θ is called an **estimator**;
- The computed value of the estimator is called an **estimate**.
- Since we are estimating one single value for each parameter, this is also referred to as **point estimation**

While intuition can lead us to good estimators, we need a more methodical way of estimating parameters for complex problems we encounter in practice.

Method of moments

Empirical distribution of a sample should "converge" to the probability distribution. Hence, the corresponding moments should be about equal.

Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a distribution with PMF or PDF $f(x|\theta_1, \theta_2, \ldots, \theta_n)$. **Method of moments** estimator can be found by equating the moments of the distribution to the moments of the sample as

$$E(X^k) = \frac{1}{n}\sum_{i=1}^{n} X^k$$

until all parameters can be solved. Here, $E(X^k)$ is the moments of distribution and $\sum_{i=1}^{n} X_i^k / n$ is the moments of the sample.

Method of moments

Example: Suppose $X_1, X_2, \ldots, X_n$ are a random sample from $N(\mu, \sigma^2)$. Find the method of moment estimators of μ and σ^2.

The moment generating function of X is $m(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$. Thus, the moments of the distribution is

$$m'(0) = (\mu + \sigma^2 t)e^{\mu t + \frac{1}{2}\sigma^2 t^2}\Big|_{t=0} = \mu$$

$$m''(0) = \left(\sigma^2 e^{\mu t + \frac{1}{2}\sigma^2 t^2} + (\mu + \sigma^2 t)^2 e^{\mu t + \frac{1}{2}\sigma^2 t^2}\right)\Big|_{t=0} = \mu^2 + \sigma^2$$

Thus, by equating moments of the distribution and moments of the sample

$$\mu = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\mu^2 + \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

Method of moments

We obtain the method of moments estimator for parameters of a normal distribution

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

The method of moments estimator coincides with our intuition and perhaps gives some credence to both. The method is somewhat more helpful when no obvious estimator suggests itself.

Methods of moments

Example: Let $X_1, X_2, \ldots, X_n$ be a random sample from a binomial distribution with unknown parameters k and p. Find the method of moment estimators.

Equating the first two sample moments to those of the distribution yields the system of equations

$$kp = \overline{X}$$
$$kp(1-p) + k^2p^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

Solving the equations yields the estimates

$$\hat{k} = \frac{\overline{X}^2}{\overline{X} - (1/n)\sum_{i=1}^{n}(X_i - \overline{X}^2)}$$
$$\hat{p} = \frac{\overline{X}}{\hat{k}}$$

Maximum likelihood estimators

Suppose we flipped a coin 3 times and observed heads, heads, and tail. What is the probability of observing such a result if $p = 0.5$ or $p = 0.6$?

The result of a coin flipping follows a Bernoulli distribution. Thus, the probability of observing heads, heads, and tail is

$$P(HHT) = p \times p \times (1 - p)$$

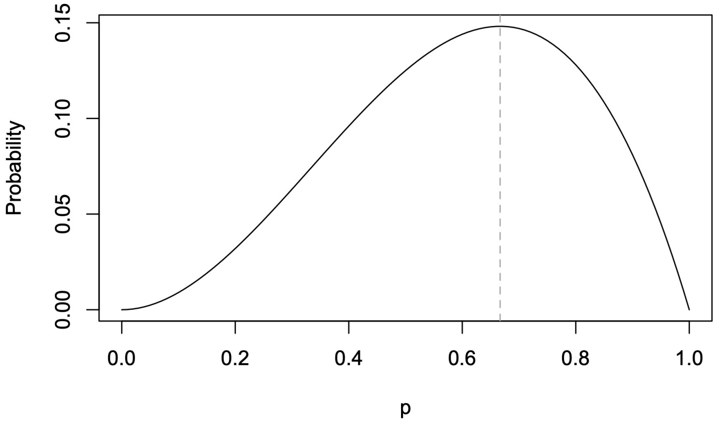
Thus, we have

$$\begin{aligned} P(HHT) &= 0.5 \times 0.5 \times (1 - 0.5) = 0.125 \quad \textit{if } p = 0.5 \\ P(HHT) &= 0.6 \times 0.6 \times (1 - 0.6) = 0.144 \quad \textit{if } p = 0.5 \end{aligned}$$

In this case, if we do not know the probability of success and want to estimate it from observation, what would be the best estimates?

Maximum likelihood estimates

The probability of getting heads, heads, and tail depends on what the success probability is. Thus, the value of p that makes the observation most likely seems to be a reasonable estimator.



Maximum likelihood estimators

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with PDF or PMF $f(x \big| \theta_1, \theta_2, \ldots, \theta_k)$, the joint PDF or PMF regarded as a function the $\theta_1, \theta_2, \ldots, \theta_k$ is called the **likelihood** function

$$
\begin{aligned}
L(\theta|X) &= L(\theta_1, \theta_2, \ldots, \theta_k \big| X_1, X_2, \ldots, X_n) \\
&= f(X_1, X_2, \ldots, X_n | \theta_1, \theta_2, \ldots, \theta_k)
\end{aligned}
$$

Most commonly, the random samples we draw are independent and are from the same distribution. We refer to as **independent and identically distributed (iid)** data. In this case, the likelihood function is

$$
L(\theta|X) = \prod_{i=1}^{n} f(x_i | \theta_1, \theta_2, \ldots, \theta_k)
$$

Maximum likelihood estimators

Definition: For a partcular sample, let $\hat{\theta}$ be the parameter value at which $L(\theta|X)$ attains its maximum as a function of θ, with X held fixed. $\hat{\theta}$ is called the **maximum likelihood estimator (MLE)** of the parameter θ based on the sample X.

Invariance property of MLEs: If $\hat{\theta}$ is the maximum likelihood estimator of θ, then for any function $\tau(\theta)$, the maximum likelihood estimator for $\tau(\theta)$ is $\tau(\hat{\theta})$.

Maximum likelihood estimators

Let $X_1, X_2, \ldots, X_n$ be the results a n independent Bernoulli trials. We know the PDF of X is $f(x) = p^x(1-p)^{1-x}$. Find the maximum likelihood estimate of p.

The likelihood function is

$$L(p|X) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum x_i}(1-p)^{n-\sum x_i}$$

The derivative of $L(p|X)$ is

$$L'(p|X) = (\sum x_i)p^{\sum x_i - 1}(1-p)^{n-\sum x_i} + (n - \sum x_i)p^{\sum x_i}(1-p)^{n-\sum x_i - 1}$$

Setting $L'(p|X) = 0$, we have

$$p^{\sum x_i}(1-p)^{n-\sum x_i}\left(\frac{\sum x_i}{p} - \frac{n-\sum x_i}{1-p}\right) = 0$$

We solve for p and get

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

Maximum likelihood estimators

We finding maximum likelihood estimators, it is often easier to deal with the natural logarithm of the likelihood function. Because logarithm is a monotonically increasing function, parameter values that maximize likelihood function also maximize the logarithm of likelihood.

In the example above, the log likelihood is

$$\log L(p) = (\sum_{i=1}^{n} x_i) \log p + (n - \sum_{i=1}^{n} x_i) \log(1 - p)$$

Thus, the derivative is

$$\frac{d \log L(p)}{dp} = \frac{\sum x_i}{p} - \frac{n - \sum x_i}{1 - p}$$

Setting the derivative to zero and we obtain

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Maximum Likelihood estimators

Let Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$. Find the maximum likelihood estimator of μ and σ^2.

The likelihood function is

$$\begin{aligned} L(\mu, \sigma) &= \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x_i - \mu)^2}{2\sigma^2} \right] \\ &= \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n \exp\left[-\frac{\sum_{i=1}^{n} (x_i - \mu)^2}{2\sigma^2} \right] \end{aligned}$$

The log likelihood function is

$$\log L(\mu, \sigma) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{2\sigma^2}$$

Setting the derivative with respect to μ and σ to 0, we get

$$\hat{\mu} = \overline{X}$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2$$