Math/Stat 310: Intro to Mathematical Statistics February 12, 2018

Lectures 6: I.I.D. Sampling and Method of Moments

Instructor: Hyunseung Kang *Scribe: Meenmo Kang*

1 IID Sampling

In previous lectures, we discussed SRS sampling were we sampled n individuals from a population of N people at random. In this lecture, we discuss another form of sampling, known as i.i.d. sampling where n individuals are sampled from a population in independent and identically distributed fashion from some population density.

2 IID Sampling vs SRS Sampling

When estimating parameter(s), there are differences between SRS and $i.i.d.$ since characteristics of their properties vary. The most distinguishable thing is their variables. In the case of SRS, we do not make any assumption on individual ξ_i because it can be either sampled or not sampled. Meanwhile, parameter X_i of $i.i.d.$ represents the actual value of measurement. Therefore, different methods should be approached by different sampling.

Table 1: Example of Normal Distribution

	SRS	**I.I.D.**
Population	$(\xi_1, ..., \xi_N)$ No assumption was made on individual ξ_i e.g. ξ_i is height of a ith person	pdf_θ that describes population parameter θ maps θ(domain) $\rightarrow$ pdf(range) e.g. Population of height is normally distributed with mean μ and variance σ^2
Sample	$(X_1, ..., X_N)$ Each variable is dependent on the other	$X_i \overset{iid}{\sim} pdf_\theta$
Parameters	$\mu = \frac{1}{N}\sum_{l=1}^{N} \xi_i$ $\sigma^2 = \frac{1}{N}\sum_{l=1}^{N}(\xi_i - \mu)^2$	$\mu = E[X_i]$ $\sigma^2 = Var[X_i]$

Example	ξ_i: Height of ith person X_i: Sampled(1); Otherwise(0) X_i and X_j are dependent	X_i of $i.i.d.$ is ith measurement X_i and X_j are independent

3 Estimating Parameters from IID Sampling

To estimate population parameter(s), we are going to use the method of moments. The major reason why the method of moments is used for estimation is that, by law of large numbers, as $n \to \infty$, $\frac{1}{n}\sum_{i=1}^{n} X$ becomes E[X]. Consequently, $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n}(X_i^k)$ is a good approximation for $E[X^k]$. So, firstly, we should define the moments of X: $(\mu_1 = E[X^1], ..., \ \mu_k = E[X^k])$. Then, a mapping between $(\mu_1, \mu_2, ..., \mu_k) \to \theta$ should be found.

Example: Normal Distribution
Suppose that we sample the height of **n** students on the campus again and the samples come from a population of heights which are assumed to be Normally distributed with mean μ and variance σ^2. Formally, $X_1, ..., X_n \sim N(\mu, \sigma^2)$. Since the parameters of normal distribution are $\theta = (\mu, \sigma^2)$, the first and second moments need to be estimated.

$\mu_1 = E[X] = \mu$
$\mu_2 = E[X^2] = \mu^2 + \sigma^2 \Rightarrow \sigma^2 = \mu_2 - \mu_1^2 \ (\because \sigma^2 = E[X^2] - E[x]^2 \Rightarrow E[X^2] = E[X]^2 + \sigma^2)$

Therefore, we can deduce that

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

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

Example: Binomial Distribution
Suppose that we flip a coin infinite number of times to check whether it is fair. When the variable X is the total number of heads, the distribution is expressed as $X \sim Bin(n, p)$. However, unlike normal distribution, there is only one parameter. Because only a single parameter exists, only the first moment is needed. In the meantime, since what we are observing is just one thing, the number of heads, we assume that $E[X]$ can be well approximated by X as like Bernoulli distribution.

$$E[X] = np \Rightarrow p = \frac{E[x]}{n} \Rightarrow \widehat{p} = \frac{X}{n}$$

Example: Gamma Distribution
As for normal distribution, there are two different parameters for Gamma distribution $\theta(\alpha, \lambda)$. Hence, the first and second moments are required for this estimation. μ_1 and μ_2 are defined by $\mu_1 = \frac{\alpha}{\lambda}$, $\quad \mu_2 = \frac{\alpha(\alpha+1)}{\lambda^2}$ (See Textbook p.157)

Using one of the properties we proved above,

$$\mu_2 = \mu^2 + \sigma^2 = \mu_1^2 + \frac{\mu_1}{\lambda} \Rightarrow \lambda = \frac{\mu_1}{\mu_2 - \mu_1^2}$$

$$\alpha = \lambda\mu_1 = \frac{\mu_1^2}{\mu_2 - \mu_1^2}$$

can be derived. Therefore, estimations for the method of moments are

$$\hat{\lambda} = \frac{\bar{X}}{\hat{\sigma^2}} \qquad \hat{\alpha} = \frac{\bar{X}^2}{\hat{\sigma^2}}$$