MAT2013: Probability and Statistics [1] [2]	**Spring 2017**

Lecture 7: Statistics: Sampling

Lecturer: Jae Guk, Kim *Scribes: Jisung Lim*

Scribes: Jisung Lim, *B.S. Candidate of Industrial Engineering in Yonsei University, South Korea.*

Chapter Objectives: How one characterizes the variability inherent **in samples**, $\{x_i\}_{i=1}^n$, as distinct from, but obviously related to, characterizing the variability inherent **in individual observations of a random variable**, X.

Inferential statistics is primarily concerned with drawing inference about a population from sample information. We therefore must begin with **sampling**—a formal study of samples from a population. But a sample is a finite collection of individual observations being itself susceptible to random variation (i.e., different samples from the same population under indentical conditions will be differ). Therefore, for samples to be useful for statistical inference (concerning the populations that produced them), the variability inherent in them must be characterized mathematically.

7.1 Introductory Concept

We've seen that the role of the sample space V_X of X in the probability theory is analogous to that of the population in statistics. this sense of analogy can be extended to the relationship between *individual observation* x from *the sample space* V_X and finite-sized *sample* from *the population.* In the probability theory, the variability inherent to individual observations x of X from the V_X is characterized by the pdf $f(x)$. It is then used to carry out theoretical probabilistic analysis for the elements of V_X.

There is analogous problem in statistics: in order to characterize the population appropriately, we must first figure out "how to characterize the variability intrinsic to the finite-sized sample".

Hence the ***"sampling theory"*** is

1. characterizing and analyzing samples from a population, and
2. employing such results to make statistical inference statements about the population.

There are three central concepts to sampling:

1. The random sample
2. The statistic
3. The distribution of statistic (sampling distribution)

7.1.1 The Random Sample

Definition 7.1.1. Random Sample
Let $X_1, X_2, X_3, \ldots, X_n$ denotes n mutually independent random variables, each of which has the same, but possibly unknown, probability distribution P_X. Then the random variable

$$X_1, X_2, X_3, \ldots, X_n \tag{7.1}$$

constitute a *random sample* from a distribution. Alternatively, $X_1, X_2, X_3, \ldots, X_n$ are called independent and identically distributed (iid) random variables with pdf $f(x)$, which is given by

$$f(x_1, x_2, \ldots, x_n|\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i|\boldsymbol{\theta}) \tag{7.2}$$

□

The motivation of the definition of the random sample is as follows:

- **Finite observations representative of population:**
  Consider a set of observations (data) $\{x_1, x_2, \ldots, x_n\}$ drawn from a population of size N (where N is possibly infinite), then the sample given to us is representative of the population.
- **Equiprobable choice of n-sized subset:**
  If all possible n-sized subsets of the N elements of the population have equal probabilities of being chosen (i.e., no particular subset will preferentially favor any particular aspect of the population), then the observations constitute a *random sample* from the population.

Example 7.1.1. *exponential random sample*
Let $X_1, X_2, \ldots, X_n \quad \sim \quad$ iid *exponential*(β). Then,

$$\begin{aligned} f(x_1, x_2, \ldots, x_n|\beta) &= \prod_{i=1}^{n} f(x_i|\beta) \\ &= \prod_{i=1}^{n} \frac{1}{\beta} \exp\left(-\frac{x_i}{\beta}\right) \\ &= \frac{1}{\beta} \exp\left(\frac{1}{\beta}\sum_{i=1}^{n} x_i\right) \end{aligned}$$

◊

7.1.2 Statistic and Its Distribution

Definition 7.1.2. Statistics
Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample of size n from a population and let $T(x_1, x_2, \ldots, x_n)$ be a real-valued or vector-valued function. Then, the random variable or random vector

$$Y = T(X_1, X_2, X_3, \ldots, X_n) \tag{7.3}$$

is called a statistic. The probability distribution of a statistic Y is called the sampling distribution of Y. $\square$

It turns out that most of the unknown parameters of population distribution are contained in the mean μ and the variance σ^2 of the distribution i.e., once the μ and σ^2 are known, the naturally occuring parameters can then be deduced. It is therefore customary for *sampling theory* to concentrate on the sampling distributions of the mean and of the variance. The following statistics represent the mean and variance of a population.

- Sample mean $\bar{X}$:

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \cdots + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{7.4}$$

- Sample variance S^2:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n} \{X_i - \bar{X}\}^2 \tag{7.5}$$

- Sample standard deviation S:

$$S = \sqrt{S^2} \tag{7.6}$$

Properties 7.1.2. Useful facts about $\bar{X}, S^2, S$:

1. $\min_a \sum_{i=1}^{n} (x_i - a)^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2$
2. $(n-1)s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$
3. $\mathbb{E}[\bar{X}] = \mathbb{E}[X_1] = \mu, Var[X_1] = \sigma^2$
4. $Var[\bar{X}] = \frac{\sigma^2}{n}$
5. $\mathbb{E}[S^2] = \sigma^2$

Proof:

1. At first, isolate a variable x_i and a by manipulating the equation.

$$\begin{aligned}\sum_{i=1}^{n}(x_i-a)^2 &= \sum_{i=1}^{n}(x_i-\bar{x}+\bar{x}-a)^2 \\ &= \sum_{i=1}^{n}(x_i-\bar{x})^2+2\sum_{i=1}^{n}(x_i-\bar{x})(\bar{x}-a)+\sum_{i=1}^{n}(\bar{x}-a)^2 \\ &= \sum_{i=1}^{n}(x_i-\bar{x})^2+\sum_{i=1}^{n}(\bar{x}-a)^2\end{aligned}$$

Now, let

$$\begin{aligned}U(x) &\coloneqq \sum_{i=1}^{n}(x_i-\bar{x})^2 \\ V(a) &\coloneqq \sum_{i=1}^{n}(\bar{x}-a)^2\end{aligned}$$

where $x=(x_1,x_2,\ldots,x_n)$. Clearly, we can see that the value of $U(x)$ change only depends on the data $\{x_i\}_{i=1}^n$. Since we want to find minimum value with respect to a, the first term $U(x)$ is out of concern, we therefore can take $a=\bar{x}$ to vanish the second term V (i.e., $V(0)=0$). Hence the minimum value is given by

$$\min_a \sum_{i=1}^{n}(x_i-a)^2=\sum_{i=1}^{n}(x_i-\bar{x})^2+V(0)=\sum_{i=1}^{n}(x_i-\bar{x})^2$$

2. We may see that

$$\begin{aligned}E[S^2] &= \mathbb{E}\left[\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2\right]=\mathbb{E}\left[\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i^2-2X_i\bar{X}+\bar{X}^2\right)\right] \\ &= \mathbb{E}\left[\frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^2-n\bar{X}^2\right)\right]=\frac{1}{n-1}\left(\sum_{i=1}^{n}\mathbb{E}[X_i^2]-n\mathbb{E}[\bar{X}^2]\right)\end{aligned}$$

Since $\forall i=1,2,\ldots,n,\quad \sigma^2=\mathbb{E}[X_i^2]-\mu^2$ and $\sigma_{\bar{X}}^2=\mathbb{E}[\bar{X}^2]+\mu_{\bar{X}}^2$ where $\sigma_{\bar{X}}^2=\sigma^2/n$ and $\mu_{\bar{X}}^2=\mu$, then

$$\begin{aligned}E[S^2] &= \frac{1}{n-1}\left(\sum_{i=1}^{n}\mathbb{E}[X_i^2]-n\mathbb{E}[\bar{X}^2]\right) \\ &= \frac{1}{n-1}\left(n(\sigma^2+\mu^2)-n(\frac{\sigma^2}{n}+\mu^2)\right) \\ &= \sigma^2\end{aligned}$$

■

Before we go on to examine the *sampling distribution*, let us consider the following statement of the problem of statistical inference:

- **Random Variable** $X \leftarrow$ **complete** $P_X \leftarrow$ **entire population**
  The distribution P_X, or equivalently pdf $f(x)$ with its parameter $\boldsymbol{\theta}$, characterizes the random variable X. Therfore, were it possible to observe the complete population in its entirety, we would be able to *constuct the complete* P_X, or $f(x|\boldsymbol{\theta})$.
- **no entire population, just finite sample** $\rightarrow$ **inference about** P_X **or** θ
  Unfortunately, since only a finite sample from the population is available via experiment or observation, we should determine the distribution P_X from that finite sample $\{x_i\}_{i=1}^n$. In the case where we know the form of the pdf $f(x)$, we are left with the issue of determining the unknown parameter $\boldsymbol{\theta}$ *making inference about the population parameter* $\boldsymbol{\theta}$ from sample data.
- **inference** $\leftarrow$ **statistics** $\leftarrow$ **random sample**
  We make this inferences by investigating *random samples*, using appropriate *statistics* (quatities calculated from the random samples) that will provide information about the parameters.
- **statistics** $\rightarrow$ **sample distribution** $\rightarrow$ **statement about** P_X **or** θ
  This statistics, which enable us to determine the unknown parameters, are themselves *random variables*. Hence, *the distribution of such statistics* then enable us to make probability statements about these statistics and hence the unknown parameters.

Hence the primary utility of the statistics and its distribution is in determining unknown population parameters from samples and quantifying the inherent variability.

7.1.3 The Sampling Distribution

Because a statistic is an observable function of random variables, determining sampling distributions requires techniques for obtaining distributions of functions of random variables. The general problem of interest may be started as follows:

Given the joint pdf for n random variables $(X_1, X_2, \ldots, X_n)$, let Y be a random vector given by $Y = (Y_1, Y_2, \ldots, Y_m)$ and find the pdf $f_Y(y)$ for the random vector Y defined as

$$Y = \Phi(X)$$

where Φ is a transformation given by

$$\begin{aligned} \Phi : \mathbb{R}^n &\rightarrow \mathbb{R}^m \\ X &\mapsto Y \end{aligned}$$

We may be alternatively denote $Y = \Phi(X)$ as follows

$$\begin{aligned} Y_1 &= \phi_1(X) \\ Y_2 &= \phi_2(X) \\ Y_3 &= \phi_3(X) \\ &\cdots \\ Y_m &= \phi_m(X) \end{aligned}$$

1. **From Φ and find Ψ**
   From this, we can find inverse transformation Ψ that $X = \Psi(Y)$, i.e.,

$$\begin{aligned} X_1 &= \psi_1(Y) \\ X_2 &= \psi_2(Y) \\ X_3 &= \psi_3(Y) \\ &\cdots \\ X_n &= \psi_n(Y) \end{aligned}$$

2. **Evaluate Jacobian $J_\Psi(X)$ and it's determinant $||J||$**

$$J_\Psi(X) = \frac{\partial \Psi}{\partial y} = \begin{bmatrix} \frac{\partial \psi_1}{\partial y_1} & \cdots & \frac{\partial \psi_1}{\partial y_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial \psi_n}{\partial y_1} & \cdots & \frac{\partial \psi_n}{\partial y_m} \end{bmatrix}$$

3. **Find $f_Y(y)$**
   From definition of pdf, $f_Y(y)$ is defined by

$$P[Y \in A] = F_Y(y) = \int_{\infty}^{y} f_Y(y)\, dy$$

From our existing knowledge about $f_X(x)$, we can evalute any probability $P_X[X \in B]$ for all $B \in \mathcal{B}(\mathbb{R}^n)$. We then can evaluate $P[Y \in A]$ for any given subset $A \in \mathcal{B}(\mathbb{R}^m)$ also as follows

$$\begin{aligned} P[Y \in A] &= P[X \in \Psi(A)] \\ &= \int_{\Psi(A)} f_X(x)\, dx \\ &= \int_A f_X(\Psi(y))\, ||J|| dy \end{aligned}$$

where $||J|| = \frac{\partial \Psi}{\partial y}$. Therefore, by definition,

$$f_Y(y) = f_X(\Psi(y))||J||$$

7.1.3.1 Some Important Sampling Distribution Results

1. Linear combination

References

[1] Youngstown State University G. Jay Kerns. *Introduction to Probability and Statistics Using R*. `http://ipsur.org/index.html`. G. Jay Kerns, 2010.

[2] Babatunde A. Ogunnaike. *Random Phenomena: Fundamentals of Probability and Statistics for Engineers*. CRC Press, 2009.