208711 Statistical Theory 1 **Fall 2019**

Lecture 15: October 8

Lecturer: Donlapark Pornnopparath

15.1 Random sample

In an experiment, we usually collect several observations from the same underlying probability distribution. We will study them in details.

Definition 15.1. n random variables $X_1, X_2, \ldots, X_n$ are called a *random sample* if they are independent of each other and the marginal pdf is the same function $f(x)$ for each X_i. We say that these variables are *independent and identically distributed* (iid).

For the indepence we can easily compute the joint pdf of a random sample:

$$f(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i).$$

If the pdf is of the form $f(X|\theta)$ i.e. it is described by a paramater θ, then we have $f(x_1, x_2, \ldots, x_n|\theta) = \prod_i f(x_i|\theta)$. In the next few sections, we will study how the distribution of the random sample behaves under different values of θ.

Example 15.2. Consider $X_i \sim \text{Exp}(\beta)$. In this case X_i could be the times (in hours) before someone call in each of the n identical telephones. Then the sample pdf is

$$f(x_1, x_2, \ldots, x_n|\beta) = \prod_{i=1}^{n} \frac{1}{\beta} e^{-x_i/\beta} = \frac{1}{\beta^n} e^{-(x_1+x_2+\ldots+x_n)/\beta}$$

Now the probability that no one has called in the first hour can be calculated as follows:

$$\begin{aligned}
\mathbb{P}(X_1 > 1, X_2 > 1, \ldots, X_n > 1) &= \frac{1}{\beta^n} \int_1^\infty \cdots \int_1^\infty e^{-(x_1+x_2+\ldots+x_n)/\beta} \, dx_1 dx_2 \ldots dx_n \\
&= \frac{1}{\beta^n} \prod_{i=1}^{n} \int_1^\infty e^{-x_i/\beta} \, dx_i \\
&= e^{-n/\beta}.
\end{aligned}$$

An easier way to calculate the above probability is to rely on the independence of the sample:

$$\begin{aligned}
\mathbb{P}(X_1 > 1, X_2 > 1, \ldots, X_n > 1) &= \mathbb{P}(X_1 > 1)\mathbb{P}(X_2 > 1) \ldots \mathbb{P}(X_n > 1) \\
&= \Big[\frac{1}{\beta} \int_1^\infty e^{-x/\beta} \, dx\Big]^n \\
&= e^{-n/\beta}.
\end{aligned}$$

◇

In the case of finite population, the result above only apply when we have a sampling *with replacement*, where each observation is returned back to the population before picking the next one. The situation if different when we have a sampling *without replacement*, since the random variables obtained from the process are not independent of each other. To see this, consider the following example:

Example 15.3. Consider the set $S = \{1, 2, \ldots, N\}$ with discrete uniform distribution. Then after picking $X_1 = x_1 \in S$ then $\mathbb{P}(X_2 = x_1 | X_1 = x_1) = 0$ and $\mathbb{P}(X_2 = y | X_1 = x_1) = \frac{1}{N-1}$ for all $y \neq x_1$. Since the conditional probability of X_2 depends on the observed value of X_1, X_1 and X_2 are not independent. ⋄

We can see from the above example that for sufficiently large N and relatively small i compared to N, the sample becomes *almost independent* in the sense that

$$\mathbb{P}(X_i = x_i | X_1 = x_1, \ldots, X_{i-1} = x_{i-1}) = \frac{1}{N - i + 1} \to 0 \quad \text{as} \quad N \to \infty.$$

Hence, in such case, some events dependent only on $X_1, \ldots X_i$ can be approximated by assuming independence.

15.2 Sample statistics

We can gain some insight on the underlying distribution by computing a summary of the observed values, such as the "location" or the "scale" of the distribution. If it is parametrized, then the goal is usually to find the parameter itself. This is going to be the main focus of the second half of the course.

Definition 15.4. Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution and $T(x_1, x_2, \ldots, x_n)$ be a real-valued or vector-valued function of observed values. Then the random variable

$$Y = T(X_1, X_2, \ldots, X_n)$$

is a *statistic* of this random sample. The probability distribution of Y is called the *sampling distribution* of Y.

The following statistics are popular choices of summarizing the sample.

Definition 15.5. The *sample mean* is given by

$$\overline{X} = \frac{X_1, X_2, \ldots, X_n}{n}$$

The *sample variance* is given by

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

and the *sample standard deviation* is given by

$$S = \sqrt{S^2}.$$

Intuitively, $\overline{X}$ and S tells us the overall location and variability of the observations, respectively. We now consider basic properties of these statistics.

Theorem 15.6. Let $X_1, X_2, \ldots, X_n$ be a random sample from a probability distribution with mean μ and variance σ^2. Then

1. $\mathbb{E}[\bar{X}] = \mu$
2. $\text{Var}[\bar{X}] = \frac{\sigma^2}{n}$
3. $\mathbb{E}[S^2] = \sigma^2$

Proof: The first two identities follow from straightforward computations:

$$\mathbb{E}[\bar{X}] = \frac{\mathbb{E}[X_1] + \mathbb{E}[X_2] + \ldots + \mathbb{E}[X_n]}{n} = \frac{\mu + \mu + \ldots + \mu}{n} = \mu$$

$$\begin{aligned}\text{Var}[\bar{X}] &= \frac{1}{n^2}(\text{Var}[X_1] + \text{Var}[X_2] + \ldots + \text{Var}[X_n]) \\ &= \frac{1}{n^2}(n\text{Var}[X_1]) \\ &= \frac{\sigma^2}{n}\end{aligned}$$

We then use these two identities to prove the last one.

$$\begin{aligned}\mathbb{E}[S^2] &= \frac{1}{n-1}\mathbb{E}[\sum_i (X - \bar{X})^2] \\ &= \frac{1}{n-1}\big[\sum_i \mathbb{E}[X^2] - n\mathbb{E}[(\bar{X})^2]\big] \\ &= \frac{1}{n-1}\big[n\mathbb{E}[X_1^2] - n\mathbb{E}[(\bar{X})^2]\big]\end{aligned}$$

Since $\sigma^2 = \mathbb{E}[X_1^2] - n\mu^2$ and $\text{Var}[\bar{X}] = \mathbb{E}[(\bar{X})^2] - n\mu^2$, the line above is equal to

$$\begin{aligned}&= \frac{1}{n-1}\big[n(\sigma^2 + n\mu^2) - n(\frac{\sigma^2}{n} + n\mu^2)\big] \\ &= \sigma^2.\end{aligned}$$

■

Such strong relations between sample statistics and population parameters, connected by expected value, give rise to a notion of *unbiased estimator* which we will see later in the course. In this case, we say that $\bar{X}$ and S^2 is an unbiased estimator of μ and σ^2, respectively.

15.3 Moment generating function

We now take a little detour and focus on a powerful tool that allows us to find the probability distribution of sample statistics.

Definition 15.7. Let X be a random variable. The moment generating function (mgf) of X is

$$M_X(t) = \mathbb{E}[e^{tX}].$$

The most useful property of mgf is that mgf determines the distribution.

Theorem 15.8. Let X and Y be two probability distributions with cdf F_X and F_Y, respectively. If there exists $\delta > 0$ such that $M_X(t) = M_Y(t)$ for all $t \in (-\delta, \delta)$, then

$$F_X(t) = F_Y(t) \quad \text{for all} \quad t \in \mathbb{R}.$$

Since the proof is out of scope for this class, we will omit it and take the theorem as granted.

Example 15.9. We will compute the mgf of $Z \sim N(0,1)$, starting off with the definition:

$$\begin{aligned} M_Z(t) &= \int \frac{1}{\sqrt{2\pi}} e^{zt} e^{-\frac{1}{2}z^2} \; dz \\ &= \int \frac{1}{\sqrt{2\pi}} e^{zt-\frac{1}{2}z^2} \; dz \end{aligned}$$

By completing the square, $zt - \frac{1}{2}z^2 = -\frac{1}{2}(z-t)^2 + \frac{1}{2}t^2$, we obtain

$$\begin{aligned} M_Z(t) &= e^{\frac{1}{2}t^2} \int \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-t)^2} \; dz \\ &= e^{\frac{1}{2}t^2}, \end{aligned}$$

where the last integral is equal to one because the integrand is the pdf of a random variable distributed as $N(t,1)$. ◇

Example 15.10. Now we consider a general case of normal distribution: $X \sim N(\mu, \sigma^2)$. The mgf in this case is

$$M_X(t) = \int \frac{1}{\sqrt{2\pi\sigma^2}} e^{xt} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2} \; dx$$

By changing the variable $z = \frac{x-\mu}{\sigma}$, which is equivalent to $x = z\sigma + \mu$, we get

$$\begin{aligned} M_X(t) &= e^{\mu t} \int \frac{1}{\sqrt{2\pi\sigma^2}} e^{z\sigma t} e^{-\frac{1}{2}z^2} \left| \frac{dx}{dz} \right| \; dz \\ &= e^{\mu t} \int \frac{1}{\sqrt{2\pi}} e^{z\sigma t} e^{-\frac{1}{2}z^2} \; dz \\ &= e^{\mu t} e^{\frac{1}{2}\sigma^2 t^2}. \end{aligned}$$

◇