EE2012/ST2334 Discussion 5*

Liu Zichen

April 8, 2019

1. **[Population, sample]** A ***population*** is a set of similar items or events which is of interest for some question or experiment, and a ***sample*** is any subset of population. Every outcome or observation can be recorded as a *numerical* or a *categorical* value. Population may be finite or infinite.

2. **[Random sampling]** ***Simple random sample*** of n observations is a sample such that every subset of n observations of the population has the same probability of being selected.

   (a) When we sample from a finite population, we can sample with/without replacement. This corresponds to the counting problems.

   (b) When we sample from an infinite population, if we assume that all random variables have the ***same*** distribution and are ***independent***, we say that the sample is random.

3. **[Sampling distribution]** The main purpose of sampling is to estimate some ***unknown population parameters***, so that we can make some ***inference*** regarding the true population. A value computed from a sample is called a ***statistic***, and it varies (why?). Hence a statistic should be a random variable. The *probability distribution of a statistic* is called a ***sampling distribution***.

   Sample mean defined by the statistic:

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

   Theorem (see an example): For random samples of size n taken from an *infinite* population or from a *finite population with replacement* having population mean μ and population standard deviation σ, the **sampling distribution of the sample mean** has:

$$\mu_{\overline{X}} = \mu_X \quad \text{and} \quad \sigma^2_{\overline{X}} = \frac{\sigma^2_X}{n} \tag{2}$$

   Law of large number (LLN): Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from a population having any distribution with mean μ and *finite* population variance σ^2. Then for any $\epsilon \in \mathcal{R}$

$$P(|\overline{X} - \mu| > \epsilon) \to 0 \text{ as } n \to \infty \tag{3}$$

   Central limit theorem (CLT): Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from a population having any distribution with mean μ and *finite* population variance σ^2. If n is sufficiently large, the sampling distribution of the sample mean $\overline{X}$ is approximately normal with mean μ and variance $\frac{\sigma^2}{n}$:

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \text{ approx } \sim N(0,1) \tag{4}$$

   Theorem: if $X_i, i = 1, 2, \cdots, n$ are $N\left(\mu, \sigma^2\right)$, the sample mean $\overline{X}$ is $N\left(\mu, \frac{\sigma^2}{n}\right)$ regardless of the sample size n.

   What about the **sampling distribution of the difference of two sample means**? If independent samples of size $n_1 (\geq 30)$ and $n_2 (\geq 30)$ are drawn from two large or infinite populations, discrete or continuous, with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, respectively. The sampling distribution of the difference of means, $\overline{X}_1$ and $\overline{X}_2$, is approximately normally distributed with mean and standard deviation given by $\mu_{\overline{X}_1 - \overline{X}_2} = \mu_1 - \mu_2$ and $\sigma_{\overline{X}_1 - \overline{X}_2} = \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$:

$$\frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \text{ approx } \sim N(0,1) \tag{5}$$

4. **[Chi-square distribution]** Y is a chi-square distribution with n *degrees of freedom* if

$$f_Y(y) = \frac{1}{2^{n/2}\Gamma(n/2)} y^{n/2-1} e^{-y/2}, \quad \text{for } y > 0, \text{and 0 otherwise} \tag{6}$$

   It is denoted as $\chi^2(n)$, and n is a positive integer and $\Gamma(\cdot)$ is the gamma function: $\Gamma(n) = \int_0^\infty x^{n-1} e^{-x} dx = (n-1)!$.

   $E(Y) = n$, $V(Y) = 2n$; if $X \sim N(0,1)$, then $X^2 \sim \chi^2(1)$; let $X_1, X_2, \cdots, X_n$ be a random sample from a normal population with mean μ and variance σ^2, $Y = \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2} \sim \chi^2(n)$.

*Some examples from bookmarked pages of the lecture notes