Online Review Course of
Undergraduate Probability and Statistics

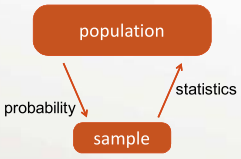
Review Lecture 10
Sampling Distribution

Chris A. Mack
Adjunct Associate Professor

Course Website: www.lithoguru.com/scientist/statistics/review.html

Probability vs. Statistics

- Probability predicts the behavior of a sample given knowledge of the population
- Statistics infers properties of the population given knowledge of a sample
- The two are tied together by the sampling distribution



Sampling

- Statistic: any quantity computed from values in a sample
  - Sampling variability – variation of a statistic for different samples
  - Sampling distribution – the distribution of the statistic across all possible samples of the given size
- Example: sample mean
  - Sample = $\{X_1, X_2, \ldots, X_n\}$
  - Sample mean = $\bar{X}$ (a random variable)
  - Assume each X_i is *iid*

independent and identically distributed

Sampling Distribution of the Mean

- What is the expectation of the sample mean?

$$\bar{X} = \frac{1}{n}\sum_{i=1,n} X_i$$

population mean

$$E(\bar{X}) = \frac{1}{n}\sum_{i=1,n} E(X_i) = E(X_i) = \mu$$

- When the expectation value of the sample statistic is equal to the population statistic, we say that the sample statistic is an unbiased estimator

Sampling Distribution of the Mean

- What is the variance of the sample mean?

$$\bar{X} = \frac{1}{n}\sum_{i=1,n} X_i$$

population variance

$$var(\bar{X}) = \frac{1}{n^2} var\left(\sum_{i=1,n} X_i\right) = \frac{\sigma^2}{n}$$

Assumes an infinite population

Standard deviation of the sample mean $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

Law of Large Numbers

- Let $X_1, X_2, \ldots, X_n$ be iid random variables with mean μ and variance σ^2.
- Then,

$$\mathbb{P}(|\bar{X} - \mu| > \varepsilon) = 0 \quad as \quad n \to \infty$$

- The average of the results obtained from a large number of trials will tend to become closer to the expected value as more trials are performed

Central Limit Theorem

- Let $X_1, X_2, \ldots, X_n$ be iid random variables with mean μ and finite variance σ^2.
- Then,

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

is a RV with a distribution that approaches $N(0,1)$ as $n \to \infty$

- This is true regardless of the distribution for X_i

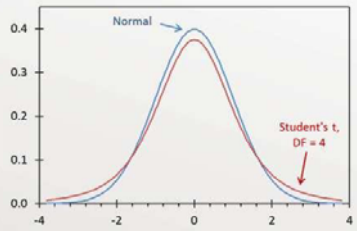
Student's t Distribution

- Alas, we rarely know the population variance
  - We estimate σ with *s*, the sample standard deviation
- If each X_i is $N(\mu, \sigma^2)$, let's define the Student's *t*-statistic:

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$

- This *t* is a RV with the Student's *t* distribution and parameter DF = *n* -1

Student's t Distribution



$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

- As DF = *n* -1 → ∞, the Student's *t* becomes $N(0,1)$
- For n > 30 or so, $N(0,1)$ is a good approximation

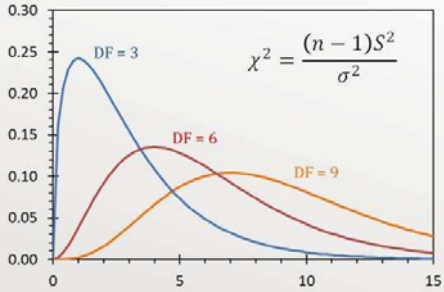
Sampling Distribution of the Variance

- Let $X_1, X_2, \ldots, X_n$ be iid random variables with $X_i \sim N(\mu, \sigma^2)$
- For a sample variance S^2,

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

is a RV with a chi square distribution and parameter DF = *n* -1

Chi Square Distribution



Review #10: What have we learned?

- What is a sampling distribution, and why do we care about it?
- What is the Central Limit Theorem?
- What is the sampling distribution of the Mean?
- What is the Student's t distribution?
- What is the chi square distribution?

