BIO5312 Biostatistics
Lecture 04: Central Limit Theorem and Three Distributions Derived from the Normal Distribution

Dr. Junchao Xia
Center of Biophysics and Computational Biology

Fall 2016
Introduction

In this lecture we will talk about main concepts as listed below,

- Expected value and variance of functions of random variables
- Linear combination of random variables
- Moments and moment generating function
- Theoretical derivation of the central limit theorem
- Three distributions derived from the normal distribution
Measures of location and spread can be developed for a random variable in much the same way as for samples.

The analog of the arithmetic mean is called the expected value of a random variable, or population mean, and is denoted by E(X) or \( \mu \) and represents the “average” value of the random variable.

\[
E(X) = \mu = \sum_{i} x_i p(x_i) = \int_{-\infty}^{+\infty} x f(x) dx
\]

Expected value of a binomial distribution \((n, p) = np\)

Expected value of a Poisson distribution with \(\mu = \lambda t\)

\[
E(X) = \sum_{k=0}^{\infty} k \frac{\mu^k}{k!} e^{-\mu} = \mu e^{-\mu} \sum_{k=1}^{\infty} \frac{\mu^{k-1}}{(k-1)!} = \mu e^{-\mu} \sum_{k=0}^{\infty} \frac{\mu^k}{k!} = \mu e^{-\mu} e^\mu = \mu
\]

Expected value of a normal distribution with \(N(\mu, \sigma^2)\)

\[
E(X) = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{\frac{-1}{2} \frac{(x-\mu)^2}{\sigma^2}} dx = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{+\infty} ze^{\frac{-z^2}{2\sigma^2}} dz + \frac{\mu}{\sigma \sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{\frac{-z^2}{2\sigma^2}} dz = 0 + \mu = \mu
\]
Variance of a Random Variable

The analog of the sample variance \( s^2 \) for a random variable is called the variance of a random variable, or population variance, and is denoted by \( \text{Var}(X) \) or \( \sigma^2 \). The standard deviation of a random variable \( X \), denoted by \( \text{sd}(X) \) or \( \sigma \), is defined by the square root of its variance.

\[
\text{Var}(X) = E \{ [X - E(X)]^2 \} = E \{ (X - \mu)^2 \} = E(X^2 - 2\mu X + \mu^2) = E(X^2) - \mu^2
\]

\[
\sigma^2 = \sum_i (x_i - \mu)^2 p(x_i) = \int (x - \mu)^2 f(x) dx
\]

- Variance of a binomial distribution \((n,p) = npq\)
- Variance of a Poisson distribution \( \sigma^2 = \mu = \lambda t \)
- Variance of a normal distribution \( N(\mu, \sigma^2) \)

\[
\text{Var}(X) = E \{ (X - \mu)^2 \} = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{+\infty} (x - \mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx
\]

\[
= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z^2 e^{-\frac{z^2}{2}} dz = \frac{\sigma^2}{\sqrt{2\pi}} \int_{0}^{+\infty} \sqrt{2u} e^{-u} du = \frac{\sigma^2}{\sqrt{\pi}} \Gamma(\alpha = \frac{1}{2} + 1) = \sigma^2
\]
Expected Value and Variance of Functions of Random Variables

- **Expected value of functions of random variables**

\[ Y = g(X), \quad E(Y) = \sum_{i} g(x_i) p(x_i) = \int_{-\infty}^{+\infty} g(x) f(x) dx \]

\[ E(X + Y) = E(X) + E(Y), \quad \text{(for any random variables independent or not)} \]

if \( X \) and \( Y \) are independent random variables,

\[ E(XY) = E(X)E(Y) \]

\[ E[g(X)h(Y)] = \sum_{i} g(x_i)h(y_i)p(x_i)p(y_i) = E[g(X)]E[h(Y)] \]

\( a \) and \( b \) are constants

\[ E(a + bX) = a + bE(X) \]

- **Variance of functions of random variables**

\[ Y = g(X), \quad Var(Y) = \sum_{i} [g(x_i) - E[g(X)]]^2 p(x_i) = \int_{-\infty}^{+\infty} [g(x) - E[g(X)]]^2 f(x) dx \]

\( X \) and \( Y \) are independent random variables,

\[ Var(X + Y) = Var(X) + Var(Y) \]

\( a \) and \( b \) are constants

\[ Var(a + bX) = E\{[a + bX - a - bE(X)]^2\} = E\{b^2[X - E(X)]^2\} = b^2Var(X) \]
Linear Combinations of Random Variables

Sums or difference or more complicated linear functions of random variables (either continuous or discrete) are often used.

A linear combination $L$ of the random variables $X_1,...,X_n$ is defined as any function of the form $L = c_1X_1 + ...+c_nX_n$. A linear combination also called a linear contrast.

To compute the expected value and variance of linear combinations of random variables, we use the principle that the expected value of the sum of $n$ random variables is the sum of the $n$ respective expected values.

No matter $X_i$ independent to each or not,

$$L = \sum_{i=1}^{n} c_iX_i, \quad E(L) = \sum_{i=1}^{n} c_iE(X_i) = \sum_{i=1}^{n} c_i\mu_i$$

Variance of $L$ where $X_1,...,X_n$ are independent is

$$L = \sum_{i=1}^{n} c_iX_i, \quad \text{Var}(L) = \sum_{i=1}^{n} c_i^2\text{Var}(X_i) = \sum_{i=1}^{n} c_i^2\sigma_i^2$$

The corresponding case for dependent random variables will be shown later.
Covariance of Dependent Random Variables

➢ The covariance between two random variables $X$ and $Y$ is denoted by $\text{Cov}(X,Y)$ and is defined by

$$\text{Cov}(X,Y) = E[(X - E(X))(Y - E(Y))] = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - \mu_X E(Y) + \mu_Y E(X) + \mu_X \mu_Y$$

where $\mu_X$ is the average value of $X$, $\mu_Y$ is the average value of $Y$, and $E(XY)$ is the average value of the product of $X$ and $Y$.

$$\text{Cov}(a + X, Y) = \text{Cov}(X, Y), \text{Cov}(aX, bY) = ab \text{Cov}(X, Y), \text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z)$$

➢ If $X$ and $Y$ are independent, then the covariance between them is 0. If large values of $X$ and $Y$ occur among the same subjects (as well as small values of $X$ and $Y$), then the covariance is positive. If large values of $X$ and small values of $Y$ (or conversely, small values of $X$ and large values of $Y$) tend to occur among the same subjects, then the covariance is negative.

➢ To obtain a measure of relatedness or association between two random variables $X$ and $Y$, we consider the **correlation coefficient**, denoted by $\text{Corr}(X,Y)$ or $\rho$ and is defined by

$$\rho = \text{Corr}(X,Y) = \text{Cov}(X,Y)/(\sigma_X \sigma_Y)$$

where $\sigma_X$ and $\sigma_Y$ are the standard deviations of $X$ and $Y$, resp.

It is a dimensionless quantity that is independent of the units of $X$ and $Y$ and ranges between -1 and 1.
If $X$ and $Y$ are approx. linearly related, a correlation coefficient of 0 implies independence. Correlation coefficient close to 1 implies nearly perfect positive dependence with large values of $X$ corresponding to large values of $Y$ and small values of $X$ corresponding to small values of $Y$. Corr close to -1 implies $\approx$ perfect negative dependence, with large values of $X$ corresponding to small values of $Y$ and vice versa.
To compute the variance of a linear contrast involving two dependent random variables $X_1$ and $X_2$, we can use the general equation:

$$Var(c_1 X_1 + c_2 X_2) = c_1^2 Var(X_1) + c_2^2 Var(X_2) + 2c_1 c_2 Cov(X_1, X_2)$$

$$= c_1^2 Var(X_1) + c_2^2 Var(X_2) + 2c_1 c_2 \sigma x \sigma y Corr(X_1, X_2)$$

Variance of linear combination of random variables (general case)

The variance of the linear contrast $L = \sum_{i=1}^{n} c_i X_i$ can be calculated as:

$$Var(L) = \sum_{i=1}^{n} c_i^2 Var(X_i) + 2 \sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j Cov(X_i X_j)$$

$$= \sum_{i=1}^{n} c_i^2 Var(X_i) + 2 \sum_{i=1}^{n} \sum_{i<j}^{n} c_i c_j \sigma_i \sigma_j Corr(X_i, X_j)$$
In statistics, higher-order statistics involves using the third or higher power of a sample such as the third or higher moments as defined below:

\[ \mu_m = E(X^m) = \sum_{i} x_i^m p(x_i) = \int_{-\infty}^{+\infty} x^m f(x)dx \]

\[ \mu_1 = E(X) = \mu, \mu_2 = E(X^2) = \sigma^2 + E(X)^2 = \sigma^2 + \mu^2 \]

\[ m \text{-th central moments} = E[(X-\mu)^m] \]

The normalised \(m\)-th central moment or standardized moment is the \(m\)-th central moments divided by \(\sigma^m\).

\[ m \text{-th normalised central moments} = E[(X-\mu)^m] / \sigma^m \]
The normalised third central moment is called the skewness, often denoted by $\gamma$.

$$\gamma = E[(X-\mu)^3]/\sigma^3$$
$$\gamma = 0 \text{ for normal distribution}$$

The normalised four central moment or standardized moment is called Kurtosis, often denoted by $\kappa$.

$$\kappa = E[(X-\mu)^4]/\sigma^4$$

The *excess kurtosis* is defined as kurtosis minus 3.
The *excess kurtosis* = 0, for a normal distribution.
The *excess kurtosis* > 0, fatter tails than a normal distribution.
The *excess kurtosis* < 0, thinner tails than a normal distribution.
Sampling Distribution

How is a specific random sample $X_1, \ldots, X_n$ used to estimate $\mu$ and $\sigma^2$, the mean and variance of the underlying distribution?

A natural estimator to select $n$ number for the sequence and take the average,

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

- $\bar{x}$ is a single realization of a random variable $\bar{X}$ over all possible samples of size $n$ that could have been selected from the population.

- $X$ denotes a random variable, and $x$ denotes a specific realization of the random variable $X$ in a sample.

- The sampling distribution of $\bar{X}$ is the distribution of values of $\bar{x}$ over all possible samples of size $n$ that could have been selected from the reference population.
Example of Sampling Distribution

Figure 6.1  Sampling distribution of $\bar{X}$ over 200 samples of size 10 selected from the population of 1000 birthweights given in Table 6.2 (100 = 100.0–100.9, etc.)

Frequency distribution of the sample mean from 200 randomly selected samples of size 10.
The expected value of $\bar{X}$ over its sampling distribution is equal to $\mu$. 
Let $X_1, ..., X_n$ be a random sample from some population with mean $\mu$ and variance $\sigma^2$. Then for large $n$, $X \sim N(\mu, \sigma^2/n)$ even if the underlying distribution of individual observations in the population is not normal. (The symbol $\sim$ is used to represent “approximately distributed.”)

This theorem allows us to perform statistical inference based on the approximate normality of the sample mean despite the nonnormality of the distribution of individual observations.

The skewness of the distribution can be reduced by transformation data using log scale. The central-limit theorem can then be applicable for smaller sizes than if the data are retained in the original scale.
Examples for Central Limit Theorem

Figure 6.4 Illustration of the central-limit theorem: 100 = 100–101.9

Figure 6.5 Distribution of single serum-triglyceride measurements and of means of such measurements over samples of size $n$
The moment-generating function (mgf) of a random variable $X$ is defined as

$$M(t) = E(e^{tX}) = \sum_{i} e^{tx_i} p(x_i) = \int_{-\infty}^{+\infty} e^{tx} f(x)dx$$

- **Property A**: If the moment-generating function exists for $t$ in an open interval containing zero, it uniquely determine the probability distribution.
- **Property B**: if the moment-generating function exists in an open interval containing zero, then the $n$-th derivative at $t=0$, $M^{(n)}(0)=E(X^n)$
- **Property C**: If $X$ has the mgf $M_X(t)$ and $Y=a+bX$, then $Y$ has the mgf $M_Y(t)=e^{at} M_X(bt)$.
- **Property D**: if $X$ and $Y$ are independent random variables with mgf’s $M_X$ and $M_Y$ and $Z=X+Y$, then $M_Z(t)=M_X(t)M_Y(t)$ on the common interval where both mgf’s exist.
The moment-generating function (mgf) of the standard normal distribution is defined as

$$M(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{tx} e^{-x^2/2} \, dx$$

Since

$$\frac{x^2}{2} - tx = \frac{1}{2} (x^2 - 2tx + t^2) - \frac{t^2}{2} = \frac{1}{2} (x - t)^2 - \frac{t^2}{2}$$

therefore,

$$M(t) = \frac{e^{t^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-(x-t)^2/2} \, dx = \frac{e^{t^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-u^2/2} \, du = e^{t^2/2}$$

$$E(X) = M'(t = 0) = te^{t^2/2} \bigg|_{t=0} = 0$$

$$E(X^2) = M''(t = 0) = e^{t^2/2} + te^{t^2/2} \bigg|_{t=0} = 1$$

$$Var(X) = E(X^2) - E(X)^2 = 1$$
Theoretical Derivation of Central Limit Theorem

\[ S_n = n\overline{X}_n = \sum_{i=1}^{n} X_i, \] where \( X_i \) are independent and identically distributed from \((0, \sigma^2)\)

Let \( Z_n = (S_n)/(\sigma\sqrt{n}) \).

We need to show the mgf \( Z_n \) tends to the mgf of standard normal distribution.

Since \( S_n \) is a sum of independent random variables,

\[ M_{S_n}(t) = [M(t)]^n \text{ and } M_{Z_n}(t) = [M\left(\frac{t}{\sigma\sqrt{n}}\right)]^n \]

\( M(s) \) has a Taylor series expansion about zero:

\[ M(s) = M(0) + sM'(0) + \frac{1}{2}s^2M''(0) + \varepsilon_s \frac{s^2}{s^2} \rightarrow 0 \text{ as } s \rightarrow 0 \]

Since \( E(X) = 0, M(0) = 1, M'(0) = 0, M''(0) = \sigma^2 \), as \( n \rightarrow \infty, t/(\sigma\sqrt{n}) \rightarrow 0 \), and

\[ M\left(\frac{t}{\sigma\sqrt{n}}\right) = 1 + \frac{1}{2}\sigma^2\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + \varepsilon_n \text{ where } \varepsilon_n/(t^2/(n\sigma^2)) \rightarrow 0 \text{ as } n \rightarrow \infty \]

We have, \( M_{Z_n}(t) = (1 + \frac{t^2}{2n} + \varepsilon_n)^n \)

It can be shown that if \( a_n \rightarrow a \), then, \( \lim_{n \rightarrow \infty} (1 + \frac{a_n}{n})^n = e^a \)

From this result, we get \( M_{Z_n}(t) \rightarrow e^{t^2/2} \) as \( n \rightarrow \infty \),

where \( e^{t^2/2} \) is the mgf of the standard normal distribution.
To obtain an interval estimate for $\sigma^2$, a new family of distributions, called chi-square ($\chi^2$) distributions, must be introduced to enable us to find the sampling distribution of $S^2$ from sample to sample.

If $Z$ is a standard normal random variable, the distribution of $U=Z^2$ called the chi-square distribution with 1 degree of freedom, denoted by $\chi_1^2$. If $X \sim N(\mu, \sigma^2)$, the $(X - \mu)/ \sigma \sim N(0, 1)$, and therefore $[(X - \mu)/ \sigma]^2 \sim \chi_1^2$.

If $U_1, U_2, \ldots, U_n$ are independent chi-square random variables with 1 degree of freedom, the distribution of $V = U_1 + U_2 + \ldots + U_n$ is called chi-square distribution with $n$ degrees of freedom and is denoted by $\chi_n^2$, with the pdf as below,

$$f(v) = g_{\alpha, \lambda} \frac{1}{2^{n/2} \Gamma(n/2)} v^{(n/2) - 1} e^{-v/2}, \ v \geq 0$$

Gamma density function $g_{\alpha, \lambda}(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{(\alpha - 1)} e^{-\lambda t}, \ t \geq 0$

Gamma function $\Gamma(\alpha) = \int_0^\infty x^{\alpha - 1} e^{-x} dx, \ x > 0$

$M(t) = (1 - 2t)^{-n/2}, \ E(V) = n, \ Var(V) = 2n$
Example of Chi-Square Distribution

If \( G = \sum_{i=1}^{n} X_i^2 \) where \( X_1, \ldots, X_n \sim \text{N}(0,1) \) and the \( X_i \)'s are independent, then \( G \) is said to follow a **chi-square distribution** with \( n \) **degrees of freedom (df)**. The distribution is often dented by \( \chi_n^2 \). The chi-square distribution only takes on positive values and is always skewed to the right.

For \( n \geq 3 \), the distribution has a mode greater than 0 and is skewed to the right. The skewness diminishes as \( n \) increases.

The \( u \)th percentile of a \( \chi_d^2 \) distribution (that is, a chi-square distribution with \( d \) \( df \)) is denoted by \( \chi_d^2 \) where \( \text{Pr}(\chi_d^2 < \chi_d^2) = u \).
If \( Z \sim N(0,1) \) and \( U \sim \chi^2_n \), and \( Z \) and \( U \) are independent, then the distribution of \( W = Z/\sqrt{U/n} \) is called a \( t \) distribution with \( n \) degrees of freedom, solved by a statistician named William Gossett (“Student”). The \( t \) distribution is not unique but is a family of distributions indexed by a parameter, the degrees of freedom (df) of the distribution.

The density function with \( n \) degrees of freedom is

\[
f(t) = \frac{\Gamma[(n+1)/2]}{\sqrt{n\pi\Gamma(n/2)}} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}
\]

If \( X_1, \ldots, X_n \sim N(\mu, \sigma^2) \) and are independent random variables, then \((\bar{X} - \mu)/(s/\sqrt{n})\) is distributed as a \( t \) distribution with \((n - 1)\) df. The \( t \) distribution with \( d \) degrees of freedom is referred to as the \( t_d \) distribution.

The **100 \times u\text{th} \) percentile of a \( t \) distribution with \( d \) degrees of freedom is denoted by \( t_{d,u} \), that is \( \Pr(t_d < t_{d,u}) \equiv u \)

---

**Figure 6.6** Comparison of Student’s \( t \) distribution with 5 degrees of freedom with an \( N(0, 1) \) distribution

![Comparison of Student's t distribution with 5 degrees of freedom with an N(0, 1) distribution](image)

\[
f(t) = f(-t)
\]

\( n \to \infty, f(t) \to \) normal distribution
Let U and V be independent chi-square random variables with m and n degrees of freedom respectively. The distribution of a new random variable \( W = (U/m)/(V/n) \) is called the F distribution with m and n degrees of freedom and is denoted by \( F_{m,n} \), with the density function as

\[
f(w) = \frac{\Gamma[(m+n)/2]}{\Gamma(m/2)\Gamma(n/2)} \left(\frac{m}{n}\right)^{m/2} w^{m/2-1} (1 + \frac{m}{n} w)^{-(m+n)/2}, \quad w \geq 0
\]
Summary

In this chapter, we discussed

- Expected and variance of functions of random variables
- Moment generating function
- Central-limit theorem
- Three distributions from the normal distribution: t, chi-square, and F distributions
The End