Probability and Distribution Theory

Ming-Ching Luoh

2005.9.13

Random Variables

Some Specific Probability Distributions
- Normal Distribution
- Chi-squared, t and F Distributions

Joint Distributions

Multivariate Distributions
- Multivariate Normal Distribution

Samples and Sampling Distributions

Large Sample Distribution Theory

Random Variable

Probability Distributions

- For a **discrete** random variable,

$$f(x) = \text{Prob}(X = x)$$

The axioms of probability requires that

1. $0 \leq \text{Prob}(X = x) \leq 1$.
2. $\sum_x f(x) = 1$.

- For a **continuous** random variable, the probability density function (pdf) is defined so that $f(x) \geq 0$ and
  1. $\text{Prob}(a \leq x \leq b) = \int_a^b f(x)dx \geq 0.$
  2. $\int_{-\infty}^{+\infty} f(x)d(x) = 1.$

Cumulative Distribution Function

- For discrete random variable,
  $F(x) = \sum_{X \leq x} f(x) = \text{Prob}(X \leq x)$, and
  $f(x_i) = F(x_i) - F(x_{i-1})$.
- For a continuous variable,
  $F(x) = \int_{-\infty}^{x} f(t)dt$ and $f(x) = \frac{dF(x)}{dx}$.

Expectations of a Random Variable

Discuss the continuous random variable only.

Mean of a random variable is,

$$\begin{aligned} \mathrm{E}(x) &= \int_x x f(x) dx \\ \mathrm{E}(g(x)) &= \int_x g(x) f(x) dx \end{aligned}$$

If $g(x) = a + bx$ for constants a and b, then $\mathrm{E}(a + bx) = a + b\mathrm{E}(x)$.

Then the variance of a random variable is

$$\begin{aligned} \mathrm{Var}(x) &= \mathrm{E}((x-\mu)^2) = \int_x (x-\mu)^2 f(x)dx \\ &= \mathrm{E}(x^2) - \mu^2 \\ \mathrm{E}(x^2) &= \sigma^2 + \mu^2 \\ \mathrm{Var}(a+bx) &= b^2 \mathrm{Var}(x) \end{aligned}$$

where $\mu = \mathrm{E}(x)$.

For any two functions $g_1(x)$ and $g_2(x)$,

$$E(g_1(x) + g_2(x)) = E(g_1(x)) + E(g_2(x)).$$

For the general case of a possibly nonlinear $g(x)$,

$$\begin{aligned} E(g(x)) &= \int_x g(x) f(x) dx \\ Var(g(x)) &= \int_x (g(x) - E(g(x)))^2 f(x) dx \end{aligned}$$

By a linear Taylor expansion around the mean of x, μ, we have

$$\begin{aligned} g(x) &\simeq g(\mu) + g'(\mu)(x-\mu) \\ \text{then } \mathrm{E}(g(x)) &\simeq g(\mu) \\ \mathrm{Var}(g(x)) &\simeq (g'(\mu)^2)\mathrm{Var}(x) \end{aligned}$$

Probability Distributions

The Normal Distribution

$$f(x|\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} \sim N(\mu,\sigma^2)$$

Normal distributions are preserved under linear transformation.

If $x \sim N(\mu,\sigma^2)$, then $(a+bx) \sim N(a+b\mu, b^2\sigma^2)$

One particular important transformation is $z = \frac{x-\mu}{\sigma}$, so that $a = -\frac{\mu}{\sigma}$, $b = \frac{1}{\sigma}$, then $z \sim N(0,1)$.

The density function of the standard normal distribution is

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

and

$$f(x) = \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right).$$

We also denote the c.d.f. of ϕ as Φ.
Since normal distribution is symmetric on both sides of zero, $\Phi(-z) = 1 - \Phi(z)$.

The Chi-Squared, t, and F Distributions

These distributions are derived from the normal distribution and arise from as sums of n or n_1 and n_2 other variables. The results are

- If $z \sim N(0, 1)$, then

$$x = z^2 \sim \chi^2(1).$$

  That is chi-squared with one degree of freedom. It can be shown that $\mathrm{E}(x) = 1$ and $\mathrm{Var}(x) = 2$.

- If $x_1, \cdots, x_n$ are n independent $\chi^2(1)$ variables, then

$$\sum_{i=1}^{n} x_i \sim \chi^2(n).$$

  The mean and variance are
  $E(\sum_{i=1}^n x_i) = \sum_{i=1}^n E(x_i) = n$ and
  $\mathrm{Var}(\sum_{i=1}^n x_i) = \sum_{i=1}^n \mathrm{Var}(x_i) = 2n$.

- If $z_i, i = 1, \cdots, n$, are independent $N(0, 1)$ variables, then

$$\sum_{i=1}^{n} z_i^2 \sim \chi^2(n).$$

- If z_i, $i = 1, \cdots, n$, are independent $N(0, \sigma^2)$ variables, then

$$\sum_{i=1}^{n} \left(\frac{z_i}{\sigma}\right)^2 \sim \chi^2(n).$$

- If x_1 and x_2 are independent chi-squared variables with n_1 and n_2 degrees of freedom, then

$$x_1 + x_2 \sim \chi^2(n_1 + n_2).$$

- If x_1 and x_2 are two independent chi-squared variables with n_1 and n_2 degrees of freedom, then the ratio

$$F(n_1, n_2) = \frac{x_1/n_1}{x_2/n_2}$$

has the F distribution with n_1 and n_2 degrees of freedom.

- If $z \sim N(0, 1)$ and $x \sim \chi^2(n)$ and is independent of z, then

$$t(n) = \frac{z}{\sqrt{x/n}}$$

  has the t distribution with n degrees of freedom. t distribution has the same shape as the normal distribution but has thicker tails. When the degrees of freedom is beyond 100, t is equivlent to the standard normal distribution.
- If $t \sim t(n)$, then $t^2 = \frac{z^2/1}{x/n} = F(1, n)$.

Joint Distributions

A joint density function for two random variables X and Y denoted $f(x, y)$ is defined so that

$$\begin{aligned}
\text{Prob}(a \leq x \leq b, c \leq y \leq d) &= \int_a^b \int_c^d f(x, y) dy dx \\
f(x, y) &\geq 0 \\
\int_x \int_y f(x, y) dy dx &= 1 \\
F(x, y) = \text{Prob}(X \leq x, Y \leq Y) &= \int_{-\infty}^x \int_{-\infty}^y f(t, s) ds dt
\end{aligned}$$

The marginal probability density is

$$f_x(x) = \int_y f(x, s)ds,$$

and similar for $f_y(y)$.

$$f(x, y) = f_x(x) f_y(y) \Leftrightarrow x \text{ and } y \text{ are independent.}$$

If x and y are independent, then from the definition of cdf, $F(x, y) = F_x(x)F_y(y)$.

Expectations in a joint distribution is defined similar to expectation is a sigle random variable,

$$
\begin{aligned}
\mathrm{E}(x) &= \int_x x f_x(x) dx = \int_x \int_y x f(x, y) dy dx \\
\mathrm{Var}(x) &= \int_x (x - \mathrm{E}(x))^2 f_x(x) dx \\
&= \int_x \int_y (x - \mathrm{E}(x))^2 f(x, y) dy dx
\end{aligned}
$$

Covariance and Correlation

$$\begin{aligned}\mathrm{Cov}(x, y) &= \mathrm{E}((x-\mu_x)(y-\mu_y)) \\ &= \mathrm{E}(xy) - \mu_x\mu_y = \sigma_{xy}\end{aligned}$$

If x and y are independent, then $f(x, y) = f_x(x) f_y(y)$. Therefore,

$$\begin{aligned}\sigma_{xy} &= \int_x \int_y (x-\mu_x)(y-\mu_y) f(x, y) dy\, dx \\ &= \int_x \int_y (x-\mu_x)(y-\mu_y) f_x(x) f_y(y) dy\, dx \\ &= \int_x (x-\mu_x) f_x(x) dx \int_y (y-\mu_y) f_y(y) dy \\ &= \mathrm{E}(x-\mu_x)\mathrm{E}(y-\mu_y) = 0\end{aligned}$$

The size of σ_{xy} depends on the own variance, it can be normalize to

$$r_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \in (-1, 1).$$

which is called **correlation coefficient**.

Some general results regarding expectations in a joint distribution.

$$\begin{aligned} \mathrm{E}(ax+by+c) &= a\mathrm{E}(x)+b\mathrm{E}(y)+c \\ \mathrm{Var}(ax+by+c) &= \mathrm{Var}(ax+by) \\ &= a^2\mathrm{Var}(x)+b^2\mathrm{Var}(y) \\ &\quad +2ab\mathrm{Cov}(x,y) \end{aligned}$$

Conditioning in a Bivariate Distribution

Conditional distributions are

$$f(y|x) = \frac{f(x,y)}{f_x(x)},\ f(x|y) = \frac{f(x,y)}{f_y(y)}$$

If x and y are independent, then $f(y|x) = f_y(y)$ and $f(x|y) = f_x(x)$.

Bivariate Normal Distribution

$$
\begin{aligned}
f(x, y) &= \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-1/2\left((\epsilon_x^2\epsilon_y^2 - 2\rho\epsilon_x\epsilon_y)/(1-\rho^2)\right)} \\
\epsilon_x &= \frac{x-\mu_x}{\sigma_x}, \epsilon_y = \frac{y-\mu_y}{\sigma_y}, \rho = \frac{\sigma_{xy}}{\sigma_x\sigma_y} \\
(x, y) &\sim N(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)
\end{aligned}
$$

- The marginal distributions are normal.
  $f_x(x) = N(\mu_x, \sigma_x^2)$, $f_y(y) = N(\mu_y, \sigma_y^2)$.

- The conditional distributions are normal:

$$
\begin{aligned}
f(y|x) &= N(\alpha + \beta x, \sigma_y^2(1-\rho^2)) \\
\alpha &= \mu_y - \beta \mu_x \\
\beta &= \frac{\sigma_{xy}}{\sigma_x^2}
\end{aligned}
$$

- x and y are independent if and only if $\rho = 0$.

Multivariate Distributions

x is a random vectors with n dimensions, then

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} = \begin{bmatrix} \mathrm{E}(x_1) \\ \mathrm{E}(x_2) \\ \vdots \\ \mathrm{E}(x_n) \end{bmatrix} = \mathrm{E}(x)$$

$$(x-\mu)(x-\mu)'$$

$$= \begin{bmatrix} (x_1-\mu_1)(x_1-\mu_1) & (x_1-\mu_1)(x_2-\mu_2) & \cdots & (x_1-\mu_1)(x_n-\mu_n) \\ (x_2-\mu_2)(x_1-\mu_1) & (x_2-\mu_2)(x_2-\mu_2) & \cdots & (x_2-\mu_2)(x_n-\mu_n) \\ \vdots & \vdots & \vdots & \vdots \\ (x_n-\mu_n)(x_1-\mu_1) & (x_n-\mu_n)(x_2-\mu_2) & \cdots & (x_n-\mu_n)(x_n-\mu_n) \end{bmatrix}$$

$$
\begin{aligned}
\mathrm{E}((x-\mu)(x-\mu)') &= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix} \\
&= \mathrm{E}(xx') - \mu \mathrm{E}(x') - \mathrm{E}(x)\mu' + \mu\mu' \\
&= \mathrm{E}(xx') - \mu\mu' = \Sigma
\end{aligned}
$$

Σ is called the **covariance matrix.**

In a set of linear functions,

$$\begin{aligned}
\mathrm{E}(a'x) &= a'\mu \\
\mathrm{Var}(a'x) &= \mathrm{E}((a'x - a'\mu)(a'x - a'\mu)') \\
&= \mathrm{E}(a'(x-\mu)(x-\mu)'a) = a'\mathrm{E}((x-\mu)(x-\mu)')a \\
&= a'\Sigma a
\end{aligned}$$

If $y = Ax$, then

$$\begin{aligned}
\mathrm{E}(y) &= \mathrm{E}(Ax) = A\mathrm{E}(x) = A\mu \\
\mathrm{Var}(y) &= \mathrm{E}((Ax - A\mu)(Ax - A\mu)') = \mathrm{E}(A(x-\mu)(x\mu)A') \\
&= A\mathrm{E}((x-\mu)(x\mu))A' = A\Sigma A'
\end{aligned}$$

Multivariate Normal Distribution

$$f(x) = (2\pi)^{-n/2}|\Sigma|^{-1/2}e^{(-1/2)(x-\mu)'\Sigma^{-1}(x-\mu)}$$

Any linear function of a vector of joint normally distributed variabels is also normally distributed. If $A \sim N(\mu, \Sigma)$, then

$$Ax + b \sim N(A\mu + b, A\Sigma A').$$

Consider a quadratic form in a standard normal vector x, $q = x'Ax$. Then

$$q = x'C\Lambda C'x = y'\Lambda y = \sum_{i=1}^{n} \lambda_i y_i^2, \;\; y = C'x$$

Since x is normally distributed, y is also normally distributed.
If A is idempotent, λ's are either 1 or 0.
Therefore, if A is idempotent then q is a χ^2 distribution with degree of freedom being the number of non-zero eigenvalues.

Theorem

Distribution of a Standardized Normal Vector. *If* $x \sim N(\mu, \Sigma)$, *then* $\Sigma^{-1/2}(x - \mu) \sim N(0, I)$.

From the above theorem,

$$(\Sigma^{-1/2}(x-\mu))'\Sigma^{-1/2}(x-\mu) \sim \chi^2(n),$$

then

$$(x-\mu)'\Sigma^{-1/2}\Sigma^{-1/2}(x-\mu) = (x-\mu)'\Sigma^{-1}(x-\mu) \sim \chi^2(n).$$

Samples and Sampling Distributions

A sample of n observations denoted $x_1, x_2, \cdots, x_n$, is a **random sample** if the n observations are drawn **independently** from the **same population**, or probability distribution, $f(x_i, \theta)$.

The sample of observations, denoted $\{x_1, x_2, \cdots, x_n\}$ or $\{x_i\}_{i=1,\cdots,n}$ is said to be independently, identically distributed, or $i.i.d$.

Some often-used descriptive statistics are

- mean: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.
- median:
- standard deviation: $s_x = \left[\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}\right]^{1/2}$.
- covariance: $s_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}$, covariance matrix: $S = [s_{ij}]$.
- correlation: $r_{xy} = \frac{s_{xy}}{s_x s_y}$, correlation matrix: $R = [r_{ij}]$.

Definition

Statistic. A statistic is any function computed from the data in a sample.

Statistic is a random variable with a probability distribution called a **sampling distribution**.

Point Estimation of Parameters

Estimator is a rule for using the data to estimate the parameter.

Estimation in a Finite Sample

Definition

Unbiased Estimator.

An estimator of a parameter θ is unbiased if the mean of its sampling distribution is θ. Formally, $\mathrm{E}(\hat{\theta}) = \theta$.

There are so many unbiased estimator. For example, the first observation is an unbiased estimator of the mean. we need more criteria.

Definition

Efficient Unbiased Estimator.
An unbiased estimator $\hat{\theta}_1$ is more efficient than another unbiased estimator $\hat{\theta}_2$ if $\text{Var}(\hat{\theta}_2) - \text{Var}(\hat{\theta}_1)$ is a **positive definite** matrix.

Large Sample Distribution Theory

Definition

Convergence in Probability The random variable x_n converges in probability to a constant c if $\lim_{n\to\infty} \text{Prob}(|x_n - c| > \epsilon) = 0$ for any positive ϵ.
For example, x_n takes two values, 0 and n with probabilities $1 - \frac{1}{n}$ and $\frac{1}{n}$. As n increases, the probability of taking the second point becomes less and less. It will converges to 0 in probability.
This is denoted as plim $x_n = c$.

Theorem

Convergence in Mean Square. *If x_n has mean μ_n and variance σ_n^2 such that the ordinary limits of μ_n and σ_n^2 are c and 0 respectively. Then x_n converges in mean square to c and* $\operatorname{plim} x_n = c$.

Theorem

Chebychev's Inequality. *If* x_n *is a random variable and* c *and* ϵ *are constants, then* $\text{Prob}(|x_n - c| > \epsilon) \leq \frac{\text{E}[(x_n - c)^2]}{\epsilon^2}$.

To establish Chebychev's inequality, we use another result.

Theorem

Markov's Inequality. *If y_n is nonnegative random variable and δ is a positive constant, then* $\text{Prob}(y_n \geq \delta) \leq \frac{E(y_n)}{\delta}$.

pf: $E(y_n) = \text{Prob}(y_n < \delta)E(y_n | y_n < \delta) + \text{Prob}(y_n \geq \delta)E(y_n | y_n \geq \delta)$.

Since y_n is nonnegative, both terms must be nonnegative, so $E(y_n) \geq \text{Prob}(y_n \geq \delta)E(y_n | y_n \geq \delta)$.

Since $E(y_n | y_n \geq \delta) \geq \delta$, $E(y_n) \geq \text{Prob}(y_n \geq \delta)\delta$, so $\text{Prob}(y_n \geq \delta) \leq \frac{E(y_n)}{\delta}$.

To prove Chebychev's inequality, let y_n be $(x_n - c)^2$ is a nonnegative random variable, and δ be ϵ^2. Since $y_n \geq \delta$ implies $|x_n - c| \geq \epsilon$, then $\text{Prob}(|x_n - c| \geq \epsilon) = \text{Prob}(|x_n - c| > \epsilon) \leq \frac{\text{E}[(x_n - c)^2]}{\epsilon^2}$.

Take a sepcial case of $c = \mu_n$, we have

$$\text{Prob}(|x_n - \mu_n| > \epsilon) \leq \frac{\sigma_n^2}{\epsilon^2}.$$

If $\lim_{n \to \infty} \text{E}(x_n) = c$ and $\lim_{n \to \infty} \text{Var}(x_n) = 0$, then

$$\lim_{n \to \infty} \text{Prob}(|x_n - c| > \epsilon) \leq \lim_{n \to \infty} \frac{\sigma_n^2}{\epsilon^2} = 0.$$

Therefore,

$$\text{plim}\, x_n = c.$$

In other word, **convergence in mean square implies convergence in probability**.

Definition

Consistent Estimator. An estimator $\hat{\theta}_n$ of a parameter θ is a consistent estimator of θ if and only if

$$\text{plim}\ \hat{\theta}_n = \theta.$$

Theorem

Consistency of the Sample Mean. *The mean of a random sample from any population with finite mean μ and finite variance σ^2 is a consistent estimator of μ.*

pf: Since $E(\bar{x}_n) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \mu$ and $\text{Var}(\bar{x}_n) = \frac{\sigma^2}{n}$, then $\bar{x}_n$ converges in mean square to μ which implies $\text{plim}\, \bar{x}_n = \mu$.

Corollary to the above theorem is the **Consistency of a Mean of Functions:** In random sampling, for any function $g(x)$, if $\mathrm{E}(g(x))$ and $\mathrm{Var}(g(x))$ are finite constants, then

$$\operatorname{plim} \frac{1}{n} \sum_{i=1}^{n} g(x_i) = \mathrm{E}(g(x)).$$

pf: Define $y_i = g(x_i)$, then $\mathrm{E}(y_i) = \mathrm{E}(g(x))$ and $\mathrm{Var}(y_i) = \mathrm{Var}(g(x))$ are finite constants. Apply the theorem above will prove the result.

Theorem

Slutsky Theorem. *For a continuous function $g(x_n)$ that is not a function of n,*

$$\text{plim}\, g(x_n) = g(\text{plim}\, x_n).$$

Theorem

Jensen's Inequality. *If $g(x_n)$ is a concave function of x_n, then $g(\mathrm{E}(x_n)) \geq \mathrm{E}(g(x_n))$.*

Theorem

Rules for Probability Limits. *If* x_n *and* y_n *are random variables with* $\text{plim } x_n = c$ *and* $\text{plim } y_c = d$*, then*

$$\begin{aligned} \text{plim } (x_n + y_n) &= c + d \\ \text{plim } x_n y_n &= cd \\ \text{plim } x_n / y_n &= c/d \text{ if } d \neq 0 \end{aligned}$$

If W_n is a matrix whose elements are random variables and if plim $W_n = \Omega$, then

$$\text{plim } W_n^{-1} = \Omega^{-1}.$$

If X_n and Y_n are random matrices with plim $X_n = A$ and plim $Y_n = B$, then

$$\text{plim } X_n Y_n = AB.$$

Definition

Convergence in Distribution.

x_n converges in distribution to a random variable x with cdf $F(x)$ if $\lim_{n\to\infty} |F_n(x) - F(x)| = 0$ at all continuous points of $F(x)$.

$F(x)$ is the limiting distribution of x, denoted as

$$x_n \to x.$$

This does not require x_n to converge at all. For example, $\text{Prob}(x_n = 1) = \frac{1}{2} + \frac{1}{n+1}$ and $\text{Prob}(x_n = 2) = \frac{1}{2} - \frac{1}{n+1}$. The distribution converges to a distribtution with probability $\frac{1}{2}$ in the two points, but x_n never converges to any constants.

Theorem

Rules for Limiting Distributions.

- *If* $x_n \rightarrow x$ *and* $\text{plim } y_n = c$*, then*

$$x_n y_n \rightarrow cx,$$

which means that the limiting distribution of $x_n y_n$ *is the distribution of* cx*. Also,*

$$\begin{aligned} x_n + y_n &\rightarrow x + c \\ \frac{x_n}{y_n} &\rightarrow \frac{x}{c}, c \neq 0 \end{aligned}$$

- If $x_n \to x$ and $g(x)$ is a continuous function, then $g(x_n) \to g(x)$.
- If y_n has a limiting distribution and plim $(x_n - y_n) = 0$, then x_n has the same limiting distribution as y_n.

Theorem

Central Limit Theorem (Univariate). *If* $x_1, x_2, \cdots, x_n$ *are a randome sample from a probability distribution with finite mean* μ *and finite variance* σ^2 *and* $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$*, then*

$$\sqrt{n}(\bar{x}_n - \mu) \to N(0, \sigma^2),$$

Theorem

Central Limit Theorem with Unequal Variances. *Suppose that* $\{x_i\}, i = 1, \cdots, n$, *is a set of random variables with finite means* μ_i *and finite positive variance* σ_i^2. *Let* $\bar{\mu}_n = \frac{1}{n}\sum_{i=1}^{n}\mu_i$ *and* $\bar{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\sigma_i^2$. *If no sigle term dominates this average variance, which we could state as* $\lim_{n\to\infty} \frac{\max(\sigma_i)}{n\bar{\sigma}_n^2} = 0$, *and if the average variance converges to a finite constant,* $\bar{\sigma}^2 = \lim_{n\to\infty} \bar{\sigma}_n^2$, *then*

$$\sqrt{n}(\bar{x}_n - \bar{\mu}_n) \to N(0, \bar{\sigma}^2).$$

$$\sqrt{n}(\bar{x}_n - \bar{\mu}_n) \to N(0, \bar{\sigma}^2).$$

This version of CLT does not require the variables come from the same underlying distribution. It requires only that the mean be a mixture of many random variables, and none of which is large compared with their sum.

For multivariate cases, we have

Theorem

Lindberg-Levy Central Limit Theorem. *If* $x_1, \cdots, x_n$ *are a random sample from a multivariate distribution with finite mean vector* μ *and finite positive definite matrix* Q*, then*

$$\sqrt{n}(\bar{x}_n - \mu) \rightarrow N(0, Q),$$

where $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Theorem

Lindberg-Feller Central Limit Theorem. *If* $x_1, \cdots, x_n$ *are a sample of random vectors such that* $\mathrm{E}(x_i) = \mu_i$, $\mathrm{Var}(x_i) = Q_i$, *and all mixed third moments of the multivariate distribution are finite. Assume that* $\lim_{n\to\infty} \bar{Q}_n = Q$, *where* Q *is a finite, positive definite matrix and that for every* i,

$$\lim_{n\to\infty} (n\bar{Q}_n)^{-1} Q_i = \lim_{n\to\infty} \Big(\sum_{i=1}^{n} Q_i\Big)^{-1} Q_i = 0.$$

then

$$\sqrt{n}(\bar{x}_n - \bar{\mu}_n) \to N(0, Q),$$

Delta method of getting the limiting distribution of a function. Having the fact that $g(z_n) \simeq g(\mu) + g'(\mu)(z_n - \mu)$ for univariate case, and $c(z_n) = c(\mu) + \frac{\partial c(\mu)}{\partial \mu'}$ for multivariate case.

Theorem

Limiting Normal Distribution of a Function. *If $\sqrt{n}(z_n - \mu) \to N(0, \sigma^2)$ and $g(z_n)$ is a continuous function not involving n, then*

$$\sqrt{n}(g(z_n) - g(\mu)) \to N(0, g'(\mu)^2\sigma^2).$$

Theorem

Limiting Normal Distribution of a Set of Functions. *If* $\sqrt{n}(z_n - \mu) \rightarrow N(0, \Sigma)$ *and* $c(z_n)$ *is a set of* J *continuous functions not involving* n*, then*

$$\sqrt{n}(c(z_n) - c(\mu)) \rightarrow N(0, C(\mu)\Sigma C(\mu)').$$

where $C(\mu) = \frac{\partial c(\mu)}{\partial \mu'}$.