Large sample tools[1]

STA2053 Fall 2022

[1] See last slide for copyright information.

Overview

Sample Space Ω, $\omega \in \Omega$

- Ω is a set, the underlying sample space.
- $\mathcal{F}$ is a class of subsets of Ω.
- There is a probability measure $\mathcal{P}$ defined on the elements of $\mathcal{F}$.

$$\text{Probability space } (\Omega, \mathcal{F}, \mathcal{P})$$

Random variables are functions from Ω into the set of real numbers

$$Pr\{X \in B\} = Pr(\{\omega \in \Omega : X(\omega) \in B\})$$

Random Sample $X_1(\omega), \ldots, X_n(\omega)$

- $T = T(X_1, \ldots, X_n)$
- $T = T_n(\omega)$
- Let $n \to \infty$ to see what happens for large samples.

Modes of Convergence

- Almost Sure Convergence
- Convergence in Probability
- Convergence in Distribution

Almost Sure Convergence

We say that T_n converges *almost surely* to T, and write $T_n \stackrel{a.s.}{\rightarrow} T$ if

$$Pr\{\omega : \lim_{n \rightarrow \infty} T_n(\omega) = T(\omega)\} = 1.$$

- Acts like an ordinary limit, except possibly on a set of probability zero.
- All the usual rules apply.
- Called convergence with probability one or sometimes strong convergence.
- In this course, convergence will usually be to a constant.

$$Pr\{\omega : \lim_{n \rightarrow \infty} T_n(\omega) = c\} = 1.$$

Strong Law of Large Numbers

Let $X_1, \ldots, X_n$ be independent with common expected value μ.

$$\overline{X}_n \stackrel{a.s.}{\rightarrow} E(X_i) = \mu$$

The only condition required for this to hold is the existence of the expected value.

Probability is long run relative frequency

- Statistical experiment: Probability of "success" is θ.
- Carry out the experiment many times independently.
- Code the results $X_i = 1$ if success, $X_i = 0$ for failure, $i = 1, 2, \ldots$

Sample proportion of successes converges to the probability of success

Recall $X_i = 0$ or 1.

$$\begin{aligned} E(X_i) &= \sum_{x=0}^{1} x \, Pr\{X_i = x\} \\ &= 0 \cdot (1-\theta) + 1 \cdot \theta \\ &= \theta \end{aligned}$$

Relative frequency is

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

Simulation

Using pseudo-random number generation by computer

- Estimate almost any probability that's hard to figure out.
- Statistical power
- Weather model
- Performance of statistical methods

- Tests or confidence intervals for estimated probabilities.

Back to the Law of Large Numbers

Recall the Change of Variables formula: Let $Y = g(X)$

$$E(Y) = \int_{-\infty}^{\infty} y\, f_Y(y)\, dy = \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx$$

Or, for discrete random variables

$$E(Y) = \sum_y y\, p_Y(y) = \sum_x g(x)\, p_X(x)$$

This is actually a big theorem, not a definition.

Applying the change of variables formula

To approximate $E[g(X)]$

Simulate $X_1, \ldots, X_n$ from the distribution of X. Calculate

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} g(X_i) &= \frac{1}{n}\sum_{i=1}^{n} Y_i \stackrel{a.s.}{\rightarrow} E(Y) \\
&= E(g(X))
\end{aligned}
$$

So for example

$$\frac{1}{n}\sum_{i=1}^{n} X_i^k \stackrel{a.s.}{\rightarrow} E(X^k)$$

$$\frac{1}{n}\sum_{i=1}^{n} U_i^2 V_i W_i^3 \stackrel{a.s.}{\rightarrow} E(U^2VW^3)$$

That is, sample moments converge almost surely to population moments.

Approximate an integral: $\int_{-\infty}^{\infty} h(x)\,dx$

Where $h(x)$ is a nasty function.

Let $f(x)$ be a density with $f(x) > 0$ wherever $h(x) \neq 0$.

$$\begin{aligned}\int_{-\infty}^{\infty} h(x)\,dx &= \int_{-\infty}^{\infty} \frac{h(x)}{f(x)} f(x)\,dx \\ &= E\left[\frac{h(X)}{f(X)}\right] \\ &= E[g(X)],\end{aligned}$$

So

- Sample $X_1, \ldots, X_n$ from the distribution with density $f(x)$
- Calculate $Y_i = g(X_i) = \frac{h(X_i)}{f(X_i)}$ for $i = 1, \ldots, n$
- Calculate $\overline{Y}_n \stackrel{a.s.}{\rightarrow} E[Y] = E[g(X)]$
- Confidence interval for $\mu = E[Y]$ is routine.

Convergence in Probability

We say that T_n converges *in probability* to T, and write $T_n \stackrel{P}{\rightarrow} T$

if for all $\epsilon > 0$,

$$\lim_{n\to\infty} P\{\omega : |T_n(\omega) - T(\omega)| < \epsilon\} = 1$$

For us, convergence will usually be to a constant:

$$\lim_{n\to\infty} P\{|T_n - c| < \epsilon\} = 1$$

Convergence in probability (say to c) means no matter how small the interval around c, for large enough n (that is, for all $n > N_1$) the probability of getting that close to c is as close to one as you like.
We will seldom use the definition in this class.

Weak Law of Large Numbers

$$\overline{X}_n \xrightarrow{p} \mu$$

- Almost Sure Convergence implies Convergence in Probability
- Strong Law of Large Numbers implies Weak Law of Large Numbers

Consistency

$T = T(X_1, \ldots, X_n)$ is a statistic estimating a parameter θ

The statistic T_n is said to be *consistent* for θ if $T_n \stackrel{P}{\rightarrow} \theta$ for all θ in the parameter space.

$$\lim_{n \to \infty} P\{|T_n - \theta| < \epsilon\} = 1$$

The statistic T_n is said to be *strongly consistent* for θ if $T_n \stackrel{a.s.}{\rightarrow} \theta$.

Strong consistency implies ordinary consistency.

Consistency is great but it's not enough.

$$T_n \stackrel{a.s.}{\to} \theta \Rightarrow U_n = T_n + \frac{100,000,000}{n} \stackrel{a.s.}{\to} \theta$$

Consistency of the Sample Variance

$$
\begin{aligned}
\widehat{\sigma}_n^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2
\end{aligned}
$$

By SLLN, $\overline{X}_n \stackrel{a.s.}{\rightarrow} \mu$ and $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \stackrel{a.s.}{\rightarrow} E(X^2) = \sigma^2 + \mu^2$.

Because the function $g(x,y) = x - y^2$ is continuous,

$$
\widehat{\sigma}_n^2 = g\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2, \overline{X}_n\right) \stackrel{a.s.}{\rightarrow} g(\sigma^2+\mu^2, \mu) = \sigma^2 + \mu^2 - \mu^2 = \sigma^2
$$

Convergence in Distribution

Sometimes called *Weak Convergence*, or *Convergence in Law*

Denote the cumulative distribution functions of $T_1, T_2, \ldots$ by $F_1(t), F_2(t), \ldots$ respectively, and denote the cumulative distribution function of T by $F(t)$.

We say that T_n converges *in distribution* to T, and write $T_n \stackrel{d}{\rightarrow} T$ if for every point t at which F is continuous,

$$\lim_{n\rightarrow\infty} F_n(t) = F(t)$$

Again, we will seldom use this definition directly.

Univariate Central Limit Theorem

Let $X_1, \ldots, X_n$ be a random sample from a distribution with expected value μ and variance σ^2. Then

$$Z_n = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \stackrel{d}{\rightarrow} Z \sim N(0,1)$$

Connections among the Modes of Convergence

- $T_n \xrightarrow{a.s.} T \Rightarrow T_n \xrightarrow{p} T \Rightarrow T_n \xrightarrow{d} T$.
- If a is a constant, $T_n \xrightarrow{d} a \Rightarrow T_n \xrightarrow{p} a$.

Sometimes we say the distribution of the sample mean is approximately normal, or asymptotically normal.

- This is justified by the Central Limit Theorem.
- But it does *not* mean that $\overline{X}_n$ converges in distribution to a normal random variable.
- The Law of Large Numbers says that $\overline{X}_n$ converges almost surely (and in probability) to a constant, μ.
- So $\overline{X}_n$ converges to μ in distribution as well.

Why would we say that for large n, the sample mean is approximately $N(\mu, \frac{\sigma^2}{n})$?

Have $Z_n = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \stackrel{d}{\rightarrow} Z \sim N(0,1)$.

$$\begin{aligned} Pr\{\overline{X}_n \leq x\} &= Pr\left\{\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \leq \frac{\sqrt{n}(x-\mu)}{\sigma}\right\} \\ &= Pr\left\{Z_n \leq \frac{\sqrt{n}(x-\mu)}{\sigma}\right\} \approx \Phi\left(\frac{\sqrt{n}(x-\mu)}{\sigma}\right) \end{aligned}$$

Suppose Y is *exactly* $N(\mu, \frac{\sigma^2}{n})$:

$$\begin{aligned} Pr\{Y \leq x\} &= Pr\left\{\frac{\sqrt{n}(Y - \mu)}{\sigma} \leq \frac{\sqrt{n}(x-\mu)}{\sigma}\right\} \\ &= Pr\left\{Z_n \leq \frac{\sqrt{n}(x-\mu)}{\sigma}\right\} = \Phi\left(\frac{\sqrt{n}(x-\mu)}{\sigma}\right) \end{aligned}$$

Convergence of random vectors I

1. Definitions (All quantities in boldface are vectors in $\mathbb{R}^m$ unless otherwise stated)
   - ⋆ $\mathbf{T}_n \stackrel{a.s.}{\rightarrow} \mathbf{T}$ means $P\{\omega : \lim_{n\to\infty} \mathbf{T}_n(\omega) = \mathbf{T}(\omega)\} = 1$.
   - ⋆ $\mathbf{T}_n \stackrel{P}{\rightarrow} \mathbf{T}$ means $\forall \epsilon > 0$, $\lim_{n\to\infty} P\{||\mathbf{T}_n - \mathbf{T}|| < \epsilon\} = 1$.
   - ⋆ $\mathbf{T}_n \stackrel{d}{\rightarrow} \mathbf{T}$ means for every continuity point $\mathbf{t}$ of $F_{\mathbf{T}}$, $\lim_{n\to\infty} F_{\mathbf{T}_n}(\mathbf{t}) = F_{\mathbf{T}}(\mathbf{t})$.
2. $\mathbf{T}_n \stackrel{a.s.}{\rightarrow} \mathbf{T} \Rightarrow \mathbf{T}_n \stackrel{P}{\rightarrow} \mathbf{T} \Rightarrow \mathbf{T}_n \stackrel{d}{\rightarrow} \mathbf{T}$.
3. If $\mathbf{a}$ is a vector of constants, $\mathbf{T}_n \stackrel{d}{\rightarrow} \mathbf{a} \Rightarrow \mathbf{T}_n \stackrel{P}{\rightarrow} \mathbf{a}$.
4. Strong Law of Large Numbers (SLLN): Let $\mathbf{X}_1, \ldots \mathbf{X}_n$ be independent and identically distributed random vectors with finite first moment, and let $\mathbf{X}$ be a general random vector from the same distribution. Then $\overline{\mathbf{X}}_n \stackrel{a.s.}{\rightarrow} E(\mathbf{X})$.
5. Central Limit Theorem: Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be i.i.d. random vectors with expected value vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Then $\sqrt{n}(\overline{\mathbf{X}}_n - \boldsymbol{\mu})$ converges in distribution to a multivariate normal with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$.

Convergence of random vectors II

6. Slutsky Theorems for Convergence in Distribution:
    1. If $\mathbf{T}_n \in \mathbb{R}^m$, $\mathbf{T}_n \xrightarrow{d} \mathbf{T}$ and if $f : \mathbb{R}^m \to \mathbb{R}^q$ (where $q \leq m$) is continuous except possibly on a set C with $P(\mathbf{T} \in C) = 0$, then $f(\mathbf{T}_n) \xrightarrow{d} f(\mathbf{T})$.
    2. If $\mathbf{T}_n \xrightarrow{d} \mathbf{T}$ and $(\mathbf{T}_n - \mathbf{Y}_n) \xrightarrow{P} 0$, then $\mathbf{Y}_n \xrightarrow{d} \mathbf{T}$.
    3. If $\mathbf{T}_n \in \mathbb{R}^d$, $\mathbf{Y}_n \in \mathbb{R}^k$, $\mathbf{T}_n \xrightarrow{d} \mathbf{T}$ and $\mathbf{Y}_n \xrightarrow{P} \mathbf{c}$, then

$$\begin{pmatrix} \mathbf{T}_n \\ \mathbf{Y}_n \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \mathbf{T} \\ \mathbf{c} \end{pmatrix}$$

Convergence of random vectors III

7. Slutsky Theorems for Convergence in Probability:
   1. If $\mathbf{T}_n \in \mathbb{R}^m$, $\mathbf{T}_n \xrightarrow{P} \mathbf{T}$ and if $f : \mathbb{R}^m \to \mathbb{R}^q$ (where $q \leq m$) is continuous except possibly on a set C with $P(\mathbf{T} \in C) = 0$, then $f(\mathbf{T}_n) \xrightarrow{P} f(\mathbf{T})$.
   2. If $\mathbf{T}_n \xrightarrow{P} \mathbf{T}$ and $(\mathbf{T}_n - \mathbf{Y}_n) \xrightarrow{P} 0$, then $\mathbf{Y}_n \xrightarrow{P} \mathbf{T}$.
   3. If $\mathbf{T}_n \in \mathbb{R}^d$, $\mathbf{Y}_n \in \mathbb{R}^k$, $\mathbf{T}_n \xrightarrow{P} \mathbf{T}$ and $\mathbf{Y}_n \xrightarrow{P} \mathbf{Y}$, then

$$\begin{pmatrix} \mathbf{T}_n \\ \mathbf{Y}_n \end{pmatrix} \xrightarrow{P} \begin{pmatrix} \mathbf{T} \\ \mathbf{Y} \end{pmatrix}$$

Convergence of random vectors IV

8. Delta Method (Theorem of Cramér, Ferguson p. 45): Let $g : \mathbb{R}^d \to \mathbb{R}^k$ be such that the elements of $\dot{g}(\mathbf{x}) = \left[\frac{\partial g_i}{\partial x_j}\right]_{k \times d}$ are continuous in a neighborhood of $\boldsymbol{\theta} \in \mathbb{R}^d$. If $\mathbf{T}_n$ is a sequence of d-dimensional random vectors such that $\sqrt{n}(\mathbf{T}_n - \boldsymbol{\theta}) \stackrel{d}{\to} \mathbf{T}$, then $\sqrt{n}(g(\mathbf{T}_n) - g(\boldsymbol{\theta})) \stackrel{d}{\to} \dot{g}(\boldsymbol{\theta})\mathbf{T}$. In particular, if $\sqrt{n}(\mathbf{T}_n - \boldsymbol{\theta}) \stackrel{d}{\to} \mathbf{T} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, then $\sqrt{n}(g(\mathbf{T}_n) - g(\boldsymbol{\theta})) \stackrel{d}{\to} \mathbf{Y} \sim N(\mathbf{0}, \dot{g}(\boldsymbol{\theta})\boldsymbol{\Sigma}\dot{g}(\boldsymbol{\theta})')$.

An application of the Slutsky Theorems

- Let $X_1, \ldots, X_n \stackrel{i.i.d.}{\sim} ?(\mu, \sigma^2)$
- By CLT, $Y_n = \sqrt{n}(\overline{X}_n - \mu) \stackrel{d}{\rightarrow} Y \sim N(0, \sigma^2)$
- Let $\widehat{\sigma}_n$ be *any* consistent estimator of σ.
- Then by 6.3, $\mathbf{T}_n = \begin{pmatrix} Y_n \\ \widehat{\sigma}_n \end{pmatrix} \stackrel{d}{\rightarrow} \begin{pmatrix} Y \\ \sigma \end{pmatrix} = \mathbf{T}$
- The function $f(x, y) = x/y$ is continuous except if $y = 0$ so by 6.1,

$$f(\mathbf{T}_n) = \frac{\sqrt{n}(\overline{X}_n - \mu)}{\widehat{\sigma}_n} \stackrel{d}{\rightarrow} f(\mathbf{T}) = \frac{Y}{\sigma} \sim N(0, 1)$$

Another application: Asymptotic normality of the sample variance

- Let $X_1, \ldots, X_n \stackrel{i.i.d.}{\sim} ?(\mu, \sigma^2)$, and $\widehat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}_n)^2$.
- Want to show $\sqrt{n}\left(\widehat{\sigma}_n^2 - \sigma^2\right)$ converges to a normal.
- Substitute μ for $\overline{X}_n$? Look at $\frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2$?
- If so, it's easy.
    - Let $Y_i = (X_i - \mu)^2$
    - $E(Y_i) = \sigma^2$
    - $Var(Y_i) = E(Y_i^2) - \left(E(Y_i)\right)^2 = E(X_i - \mu)^4 - \sigma^4 = \sigma_y^2$.
    - $\overline{Y}_n = \frac{1}{n}\sum_{i=1}^n (X_i - \mu)^2$
    - By CLT, $\sqrt{n}\left(\overline{Y}_n - \sigma^2\right) \stackrel{d}{\rightarrow} Y \sim N(0, \sigma_y^2)$.

Show $\sqrt{n}\left(\widehat{\sigma}_n^2 - \sigma^2\right) - \sqrt{n}\left(\overline{Y}_n - \sigma^2\right) \stackrel{p}{\to} 0$

See 6.2

$$
\begin{aligned}
\widehat{\sigma}_n^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}(X_i {\color{red}-\mu + \mu} - \overline{X}_n)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[(X_i-\mu)^2 + 2(X_i-\mu)(\mu-\overline{X}_n) + (\mu-\overline{X}_n)^2\right] \\
&= \ldots \\
&= \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 - (\overline{X}_n-\mu)^2 \\
&= \overline{Y}_n - (\overline{X}_n-\mu)^2
\end{aligned}
$$

Using $\widehat{\sigma}_n^2 = \overline{Y}_n - (\overline{X}_n - \mu)^2$

$$
\begin{aligned}
\sqrt{n}\left(\overline{Y}_n - \sigma^2\right) - \sqrt{n}\left(\widehat{\sigma}_n^2 - \sigma^2\right) &= \sqrt{n}\left(\overline{Y}_n - \widehat{\sigma}_n^2\right) \\
&= \sqrt{n}\left(\overline{Y}_n - (\overline{Y}_n - (\overline{X}_n - \mu)^2)\right) \\
&= \sqrt{n}\left(\overline{X}_n - \mu\right)^2 \\
&= \sqrt{n}(\overline{X}_n - \mu) \cdot (\overline{X}_n - \mu)
\end{aligned}
$$

- First term goes in distribution to $X \sim N(0, \sigma^2)$ by CLT.
- Second term goes to zero in probability by LLN.
- $\begin{pmatrix} \sqrt{n}(\overline{X}_n - \mu) \\ \overline{X}_n - \mu \end{pmatrix} \stackrel{d}{\rightarrow} \begin{pmatrix} X \\ 0 \end{pmatrix}$ by 6.3.
- By continuous mapping 6.1, $\sqrt{n}(\overline{X}_n - \mu) \cdot (\overline{X}_n - \mu) \stackrel{d}{\rightarrow} X \cdot 0 = 0$
- Convergence in distribution to a constant implies convergence in probability (Rule 3), so the difference converges in probability to zero, and the result follows by 6.2 ■

The Result

- Because the difference between $\sqrt{n}\left(\widehat{\sigma}_n^2 - \sigma^2\right)$ and $\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 - \sigma^2\right)$ goes to zero in probability, they converge in distribution to the same target.
- Target is $N(0, \sigma_y^2)$
- $\sigma_y^2 = E(X_i - \mu)^4 - \sigma^4$.

Univariate delta method

In the multivariate Delta Method 8, the matrix $\dot{\text{g}}(\boldsymbol{\theta})$ is a Jacobian. The univariate version of the delta method says that

If $\sqrt{n}\,(T_n - \theta) \stackrel{d}{\rightarrow} T$ and $g''(x)$ is continuous in a neighbourhood of θ, then

$$\sqrt{n}\,(g(T_n) - g(\theta)) \stackrel{d}{\rightarrow} g'(\theta)\,T.$$

When using the Central Limit Theorem, *especially* if there is a $\theta \neq \mu$ in the model, it's safer to write

$$\sqrt{n}\left(g(\overline{X}_n) - g(\mu)\right) \stackrel{d}{\rightarrow} g'(\mu)\,T.$$

and then substitute for μ in terms of θ.

Delta Method Example

$X_1, \ldots, X_n \overset{i.i.d.}{\sim} \text{Poisson}(\lambda)$

$E(X_i) = Var(X_i) = \lambda$

$\frac{\sqrt{n}(\overline{X}_n - \lambda)}{\sqrt{\overline{X}_n}} \xrightarrow{d} Z_1 \sim N(0,1)$

Confidence interval $\left(\overline{X}_n - z_{\alpha/2}\sqrt{\frac{\overline{X}_n}{n}} \, , \, \overline{X}_n + z_{\alpha/2}\sqrt{\frac{\overline{X}_n}{n}} \right)$

Maybe we can do better.

Delta Method says $\sqrt{n}\,(g(T_n) - g(\theta)) \xrightarrow{d} g'(\theta)\,T$

$$\sqrt{n}(\overline{X}_n - \lambda) \xrightarrow{d} X \sim N(0, \lambda).$$

$$\sqrt{n}\left(g(\overline{X}_n) - g(\lambda)\right) \xrightarrow{d} g'(\lambda)\,X \sim N(0, g'(\lambda)^2\lambda)$$

- Choose g to make the variance not depend on λ.
- How about $g(\lambda) = 2\sqrt{\lambda}$

  $g'(\lambda) = 2\,\frac{1}{2}\lambda^{-1/2} = \frac{1}{\sqrt{\lambda}}$.
- Variance of the target is $g'(\lambda)^2\lambda = 1$.

  So,

$$\sqrt{n}\left(2\sqrt{\overline{X}_n} - 2\sqrt{\lambda}\right) \xrightarrow{d} Z_2 \sim N(0, 1).$$

$\sqrt{n}\left(2\sqrt{\overline{X}_n} - 2\sqrt{\lambda}\right) \stackrel{d}{\rightarrow} Z_2 \sim N(0,1)$

$$
\begin{aligned}
0.95 &\approx P\left\{-z_{\alpha/2} < \sqrt{n}\left(2\sqrt{\overline{X}_n} - 2\sqrt{\lambda}\right) < z_{\alpha/2}\right\} \\
&= P\left\{-\frac{z_{\alpha/2}}{2\sqrt{n}} < \sqrt{\overline{X}_n} - \sqrt{\lambda} < \frac{z_{\alpha/2}}{2\sqrt{n}}\right\} \\
&= P\left\{\sqrt{\overline{X}_n} - \frac{z_{\alpha/2}}{2\sqrt{n}} < \sqrt{\lambda} < \sqrt{\overline{X}_n} + \frac{z_{\alpha/2}}{2\sqrt{n}}\right\} \\
&= P\left\{\left(\sqrt{\overline{X}_n} - \frac{z_{\alpha/2}}{2\sqrt{n}}\right)^2 < \lambda < \left(\sqrt{\overline{X}_n} + \frac{z_{\alpha/2}}{2\sqrt{n}}\right)^2\right\}.
\end{aligned}
$$

Compare $P\left\{\overline{X}_n - z_{\alpha/2}\sqrt{\frac{\overline{X}_n}{n}} < \lambda < \overline{X}_n + z_{\alpha/2}\sqrt{\frac{\overline{X}_n}{n}}\right\}$

The delta method comes from Taylor's Theorem

Taylor's Theorem: Let the nth derivative $f^{(n)}$ be continuous in $[a, b]$ and differentiable in (a, b), with x and x_0 in (a, b). Then there exists a point ξ between x and x_0 such that

$$\begin{aligned} f(x) &= f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)(x-x_0)^2}{2!} + \ldots \\ &+ \frac{f^{(n)}(x_0)(x-x_0)^n}{n!} + \frac{f^{(n+1)}(\xi)(x-x_0)^{n+1}}{(n+1)!} \end{aligned}$$

where $R_n = \frac{f^{(n+1)}(\xi)(x-x_0)^{n+1}}{(n+1)!}$ is called the *remainder term*.
If $R_n \to 0$ as $n \to \infty$, the resulting infinite series is called the *Taylor Series* for $f(x)$.

Taylor's Theorem with two terms plus remainder

Very common in applications

Let $g(x)$ be a function for which $g''(x)$ is continuous in an open interval containing $x = \theta$. Then

$$g(x) = g(\theta) + g'(\theta)(x - \theta) + \frac{g''(\theta^*)(x - \theta)^2}{2!}$$

where θ^* is between x and θ.

Delta method

Using $g(x) = g(\theta) + g'(\theta)(x-\theta) + \frac{1}{2}g''(\theta^*)(x-\theta)^2$

Let $\sqrt{n}(T_n - \theta) \xrightarrow{d} T$ so that $T_n \xrightarrow{p} \theta$.

$$
\begin{aligned}
\sqrt{n}\left(g(T_n) - g(\theta)\right) &= \sqrt{n}\left(g(\theta) + g'(\theta)(T_n - \theta) + \frac{1}{2}g''(\theta_n^*)(T_n - \theta)^2 - g(\theta)\right) \\
&= \sqrt{n}\left(g'(\theta)(T_n - \theta) + \frac{1}{2}g''(\theta_n^*)(T_n - \theta)^2\right) \\
&= g'(\theta)\sqrt{n}(T_n - \theta) \\
&\quad + \frac{1}{2}g''(\theta_n^*) \cdot \sqrt{n}(T_n - \theta) \cdot (T_n - \theta) \\
&\xrightarrow{d} g'(\theta)T + 0
\end{aligned}
$$

That was fun, but it was all univariate.

- The multivariate CLT establishes convergence to a multivariate normal.
- Vectors of MLEs are approximately multivariate normal for large samples.
- The multivariate delta method can yield the asymptotic distribution of useful functions of the MLE vector.

We need to look at random vectors and the multivariate normal distribution.

Copyright Information