STSCI 5080
Probability Models and Inference
Lecture 16: CLT, χ^2 and t Distributions

October 23, 2018

Example 16.1

Example

If X_n and X have pdfs f_n and f, respectively, and $X_n \xrightarrow{d} X$, then does $f_n(x) \to f(x)$ pointwise?

Example 16.1

Example

If X_n and X have pdfs f_n and f, respectively, and $X_n \xrightarrow{d} X$, then does $f_n(x) \to f(x)$ pointwise?

Answer: No (but the opposite is true).

Counterexample

Consider a pdf

$$f_n(x) = \begin{cases} 1 - \cos(2\pi n x) & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}.$$

For $x < 0$, $F_n(x) = 0$, for $0 \le x \le 1$,

$$F_n(x) = \int_0^x f_n(y)dy = x - \frac{1}{2\pi n}\sin(2\pi n x) \to x,$$

and for $x > 1$, $F_n(x) = 1$. Hence,

$$X_n \xrightarrow{d} U[0,1].$$

But $f_n(x)$ does not converge pointwise.

CLT

Denote by $\Phi(x)$ the cdf of $N(0,1)$.

Theorem

Let $X_1, \ldots, X_n$ be a random sample from a cdf F, where F has mean μ and variance $\sigma^2 > 0$. Then

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0,1),$$

or equivalently

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2).$$

Verify that

$$X_n \xrightarrow{d} X \Rightarrow \sigma X_n \xrightarrow{d} \sigma X.$$

We also note that

$$X \sim N(\mu, \sigma^2) \Rightarrow aX + b \sim N(a\mu + b, a^2\sigma^2).$$

The CLT implies that

$$\left.\begin{array}{l} P(a < \sqrt{n}(\overline{X}_n - \mu)/\sigma < b) \\ P(a \leq \sqrt{n}(\overline{X}_n - \mu)/\sigma < b) \\ P(a < \sqrt{n}(\overline{X}_n - \mu)/\sigma \leq b) \\ P(a \leq \sqrt{n}(\overline{X}_n - \mu)/\sigma \leq b) \end{array}\right\} \to \Phi(b) - \Phi(a)$$

for any $a < b$.
We note that

$$\Phi(-x) = 1 - \Phi(x)$$

for any $x > 0$.

Example 16.2

Example

Let $X_1, \ldots, X_{12} \sim U[0,1]$ i.i.d. Use CLT to approximate $P(|\overline{X}_{12} - 1/2| < 0.1)$.

Example 16.2

Example

Let $X_1, \ldots, X_{12} \sim U[0,1]$ i.i.d. Use CLT to approximate $P(|\overline{X}_{12} - 1/2| < 0.1)$.

We have

$$\mu = \frac{1}{2} \quad \text{and} \quad \sigma^2 = \frac{1}{12},$$

so that

$$P(12|\overline{X}_{12} - 1/2| < x) \approx \Phi(x) - \Phi(-x) = 2\Phi(x) - 1.$$

Hence,

$$\begin{aligned} P(|\overline{X}_{12} - 1/2| < 0.1) &= P(12|\overline{X}_{12} - 1/2| < 1.2) \\ &\approx 2\Phi(1.2) - 1 \\ &\approx 0.729. \end{aligned}$$

Theorem

If $Y_n \sim Bin(n, p)$*, then*

$$\frac{Y_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0, 1).$$

Proof

By definition, $Y_n = X_1 + \cdots + X_n$ for independent Bernoulli trials $X_1, \ldots, X_n$ with success probability p. We note that $E(X_1) = p$ and $\mathrm{Var}(X_1) = p(1-p)$, and so

$$\frac{Y_n - np}{\sqrt{np(1-p)}} = \frac{\sqrt{n}(\overline{X}_n - p)}{\sqrt{p(1-p)}} \xrightarrow{d} N(0,1)$$

by CLT.

Example 16.3

Example

If $Y \sim Bin(100, 1/2)$, then use CLT to approximate $P(Y > 60)$.

Example 16.3

Example

If $Y \sim Bin(100, 1/2)$, then use CLT to approximate $P(Y > 60)$.

In this case, $np = 50$ and $np(1-p) = 25$. Hence,

$$P(Y > 60) = P\left(\frac{Y-50}{5} > \frac{60-50}{5}\right) \approx 1 - \Phi(2) \approx 0.023.$$

Proof of CLT

Theorem (Continuity theorem for mgfs)

Let X_n *and* X *have mgfs* ψ_n *and* ψ*, respectively. If* $\psi_n(\theta) \to \psi(\theta)$ *for any* θ *in an open interval containing the origin, then* $X_n \xrightarrow{d} X$.

Proof of CLT

Theorem (CLT)

Let $X_1, \ldots, X_n$ be a random sample from a cdf F, where F has mean μ and variance $\sigma^2 > 0$. Then

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0, 1).$$

Proof of CLT using mgfs

Let

$$Y_i = \frac{X_i - \mu}{\sigma}, \; i = 1, \dots, n.$$

We note that $Y_1, \dots, Y_n$ are i.i.d. with mean zero and unit variance, and

$$\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} = \sqrt{n}\overline{Y}_n.$$

Denote by $\psi(\theta)$ the mgf of Y_1: $\psi(\theta) = E(e^{\theta Y_1})$. The mgf of $\sqrt{n}\overline{Y}_n$ is

$$\begin{aligned}\psi_n(\theta) &= E(e^{\theta \sum_{i=1}^n Y_i/\sqrt{n}}) = E(e^{\theta Y_1/\sqrt{n}} \cdots e^{\theta Y_n/\sqrt{n}}) \\ &= E(e^{\theta Y_1/\sqrt{n}}) \cdots E(e^{\theta Y_n/\sqrt{n}}) = \{\psi(\theta/\sqrt{n})\}^n.\end{aligned}$$

Now, since $\psi'(0) = E(Y_1) = 0$ and $\psi''(0) = E(Y_1^2) = 1$, we can expand $\psi(\theta)$ as

$$\begin{aligned}\psi(\theta) &= \psi(0) + \psi'(0)\theta + \frac{\theta^2}{2}\psi''(0) + \theta^2 R(\theta) \\ &= 1 + \frac{\theta^2}{2} + \theta^2 R(\theta)\end{aligned}$$

by Taylor's theorem, where $\lim_{\theta\to 0} R(\theta) = 0$. Substituting this expansion, we have

$$\psi_n(\theta) = \{\psi(\theta/\sqrt{n})\}^n = \left(1 + \frac{\theta^2}{2n} + \frac{\theta^2}{n}R(\theta/\sqrt{n})\right)^n \to e^{\theta^2/2},$$

which is the mgf of $N(0,1)$. By the continuity theorem, we have $\sqrt{n}\overline{Y}_n \xrightarrow{d} N(0,1)$.

Functions of sample mean

Theorem (Continuous mapping theorem)

If $Y_n \xrightarrow{P} \mu$ (constant) and if $g(x)$ is continuous at $x = \mu$, then $g(Y_n) \xrightarrow{P} g(\mu)$.

Proof

Since $g(x)$ is continuous at $x = \mu$, for any $\varepsilon > 0$, there exists $\delta > 0$ such that

$$|x - \mu| < \delta \Rightarrow |g(x) - g(\mu)| < \varepsilon.$$

This implies that

$$\{|Y_n - \mu| < \delta\} \subset \{|g(Y_n) - g(\mu)| < \varepsilon\},$$

so that

$$P\{|g(Y_n) - g(\mu)| < \varepsilon\} \geq P(|Y_n - \mu| < \delta),$$

but $\lim_{n\to\infty} P(|Y_n - \mu| < \delta) = 1$. Hence, we have

$$\lim_{n\to\infty} P\{|g(Y_n) - g(\mu)| < \varepsilon\} = 1.$$

Since $\varepsilon > 0$ is arbitrary, we conclude that $g(Y_n) \xrightarrow{P} g(\mu)$.

Delta method

Example

If $\sqrt{n}(Y_n - \mu) \xrightarrow{d} N(0, \sigma^2)$ and $g(x)$ is differentiable at $x = \mu$, then

$$\sqrt{n}\{g(Y_n) - g(\mu)\} \xrightarrow{d} N(0, \{g'(\mu)\}^2\sigma^2).$$

Proof (heuristic)

By differentiability, we have

$$g(x) \approx g(\mu) + g'(\mu)(x - \mu).$$

Plugging in $x = Y_n$, we have

$$\sqrt{n}\{g(Y_n) - g(\mu)\} \approx g'(\mu)\sqrt{n}(Y_n - \mu) \xrightarrow{d} N(0, \{g'(\mu)\}^2\sigma^2).$$

Example 16.4

Example

Let $X_1, \ldots, X_n \sim Ex(\lambda)$ i.i.d. Since $E(X_1) = 1/\lambda$, i.e., $\lambda = 1/E(X_1)$, it is reasonable to "estimate" λ by $\widehat{\lambda} = 1/\overline{X}_n$ (which is indeed the MLE). By the continuous mapping theorem, $\widehat{\lambda}$ is consistent, i.e., $\widehat{\lambda} \xrightarrow{P} \lambda$. What is the limiting distribution of $\sqrt{n}(\widehat{\lambda} - \lambda)$?

Example 16.4

Example

Let $X_1, \ldots, X_n \sim Ex(\lambda)$ i.i.d. Since $E(X_1) = 1/\lambda$, i.e., $\lambda = 1/E(X_1)$, it is reasonable to "estimate" λ by $\widehat{\lambda} = 1/\overline{X}_n$ (which is indeed the MLE). By the continuous mapping theorem, $\widehat{\lambda}$ is consistent, i.e., $\widehat{\lambda} \xrightarrow{P} \lambda$. What is the limiting distribution of $\sqrt{n}(\widehat{\lambda} - \lambda)$?

We first note that

$$\sqrt{n}(\overline{X}_n - 1/\lambda) \xrightarrow{d} N(0, 1/\lambda^2).$$

Since the derivative of $g(x) = 1/x$ at $x = 1/\lambda$ is

$$g'(1/\lambda) = -\frac{1}{1/\lambda^2} = -\lambda^2$$

we have

$$\sqrt{n}(\widehat{\lambda} - \lambda) \xrightarrow{d} N(0, \lambda^2).$$

Example 16.5

Example

Let $X_1, \ldots, X_n \sim Po(\lambda)$ i.i.d. We want to estimate $\theta = g(\lambda) = P(X_1 = 0) = e^{-\lambda}$. Find the limiting distribution of $\sqrt{n}(\widehat{\theta} - \theta)$ where $\widehat{\theta} = g(\overline{X}_n) = e^{-\overline{X}_n}$.

Example 16.5

Example

Let $X_1, \ldots, X_n \sim Po(\lambda)$ i.i.d. We want to estimate $\theta = g(\lambda) = P(X_1 = 0) = e^{-\lambda}$. Find the limiting distribution of $\sqrt{n}(\widehat{\theta} - \theta)$ where $\widehat{\theta} = g(\overline{X}_n) = e^{-\overline{X}_n}$.

Since the $Po(\lambda)$ distribution has mean λ and variance λ, we have

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

by CLT. Applying the delta method with $g(\lambda) = e^{-\lambda}$, we have

$$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} N(0, e^{-2\lambda}\lambda).$$

Example 16.5 (cont)

Example

Another natutral estimator for $\theta = e^{-\lambda}$ is $\widetilde{\theta} = \overline{Y}_n$ where

$$Y_i = \begin{cases} 1 & \text{if } X_i = 0 \\ 0 & \text{otherwise} \end{cases},$$

because $E(Y_i) = P(X_i = 0) = e^{-\lambda} = \theta$. Find the limiting distribution of $\sqrt{n}(\widetilde{\theta} - \theta)$? Which estimator do you think is better?

Since Y_i has mean θ and variance $\theta(1-\theta) = e^{-\lambda}(1-e^{-\lambda})$, we have

$$\sqrt{n}(\widetilde{\theta} - \theta) \xrightarrow{d} N(0, e^{-\lambda}(1-e^{-\lambda}))$$

by CLT.
Next, both $\widehat{\theta}$ and $\widetilde{\theta}$ are approximately unbiased, but

$$\frac{e^{-\lambda}(1-e^{-\lambda})}{e^{-2\lambda}\lambda} = \frac{e^{\lambda}-1}{\lambda} = \frac{\lambda + \frac{\lambda^2}{2} + \cdots}{\lambda} > 1,$$

we can argue that $\widehat{\theta}$ is better than $\widetilde{\theta}$ (we will discuss optimality of estimators in more detail in the future).

Chapter 6 Distributions Derived from the Normal Distribution

χ^2 distribution

Definition

Let $Z_1, \ldots, Z_n \sim N(0,1)$ i.i.d. Then $V = Z_1^2 + \cdots + Z_n^2$ is said to follow the χ^2 distribution with n degrees of freedom, $V \sim \chi^2(n)$ in short.

Recall that $Y = Z_1^2$ has pdf

$$g(y) = \begin{cases} \frac{1}{\sqrt{2\pi}} y^{-1/2} e^{-y/2} & \text{if } y > 0 \\ 0 & \text{otherwise} \end{cases},$$

which coincides with the pdf of $Ga(1/2, 2)$. By the regeneration property of the gamma distribution, we have:

Theorem

$\chi^2(n) = Ga(n/2, 2)$. *Hence, the pdf of* $V \sim \chi^2(n)$ *is*

$$f(v) = \frac{1}{2^{n/2}\Gamma(n/2)} v^{n/2-1} e^{-v/2} \quad \textit{for } v > 0,$$

and the mgf of V *is*

$$\psi(\theta) = (1 - 2\theta)^{-n/2} \quad \textit{for } \theta < 1/2.$$

t distribution

Definition

If $Z \sim N(0,1)$ and $V \sim \chi^2(n)$, and Z and V are independent, then

$$T = \frac{Z}{\sqrt{V/n}}$$

is said to follow the t distribution with n degrees of freedom, $T \sim t(n)$ in short.

Theorem

The pdf of $T \sim t(n)$ *is*

$$f_T(t) = \frac{\Gamma\{(n+1)/2\}}{\sqrt{n\pi}\,\Gamma(n/2)} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}, \quad -\infty < t < \infty.$$

Proof (outline)

The cdf of $U = \sqrt{V/n}$ is

$$P(U \leq u) = P(\sqrt{V/n} \leq u) = P(V \leq nu^2) = F_V(nu^2),$$

so that the pdf of U is

$$f_U(u) = 2nuf_V(nu^2).$$

The pdf of $T = Z/U$ is given by

$$f_T(t) = \int_0^\infty f_U(u)f_Z(ut)du.$$

Properties of t distribution

Denote by $f_n(t)$ the pdf of $t(n)$.

- If $n = 1$, then the pdf is

$$f_1(t) = \frac{1}{\pi(1+t^2)},$$

  which coincides with the Cauchy density.

- If $Y \sim \chi^2(n)$, then for any positive integer k,

$$E(|Y|^k) \begin{cases} < \infty & \text{if } k < n \\ = \infty & \text{if } k \geq n \end{cases}.$$

- If $n \to \infty$, then $f_n(t) \to e^{-t^2/2}/\sqrt{2\pi}$ (pdf of $N(0,1)$) pointwise.