STSCI 5080
Probability Models and Inference
Lecture 18: Sample Mean/Variance from Normal Population and Estimation

November 1, 2018

Some common mistakes in Exam 2

- $\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$ does NOT mean that

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

  You have to distinguish between two. CLT means that

$$\lim_{n\to\infty} P\{\sqrt{n}(\overline{X}_n - \mu) \le x\} = P(\sigma Z \le x),\ Z \sim N(0, 1)$$

  for any x, but whenever n is finite, the cdf of $\sqrt{n}(\overline{X}_n - \mu)$ is different from that of σZ.

- $Y_n - n \xrightarrow{d} N(0, 2n)$ as $n \to \infty$ for $Y_n \sim \chi^2(n)$ does NOT make sense. Since we are taking the limit $n \to \infty$, n should not appear in the limit.

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.

Theorem

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

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

and the mgf of V *is*

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

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

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

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 $Y \sim t(n)$, 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.

Sample mean and variance

- Random sample from F:

$$X_1, \ldots, X_n \sim F \text{ i.i.d.}$$

  where F has mean μ (population mean) and variance σ^2 (population variance).

- The sample mean is

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

  We know that

$$E(\overline{X}) = \mu \quad \text{and} \quad \mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}.$$

- The sample variance is

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2.$$

Theorem

We have

$$E(S^2) = \sigma^2.$$

Proof

Let

$$Y_i = X_i - \mu,$$

so that $E(Y_i) = 0$ and $E(Y_i^2) = \mathrm{Var}(X_i)$. Then

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\{X_i - \mu - (\overline{X} - \mu)\}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2.$$

In addition,

$$\begin{aligned}\sum_{i=1}^{n}(Y_i - \overline{Y})^2 &= \sum_{i=1}^{n}(Y_i - 2\overline{Y}Y_i + \overline{Y}^2) \\ &= \sum_{i=1} Y_i^2 - 2\overline{Y}\underbrace{\sum_{i=1}^{n} Y_i}_{=n\overline{Y}} + n(\overline{Y})^2 \\ &= \sum_{i=1}^{n} Y_i^2 - n(\overline{Y})^2.\end{aligned}$$

Hence,

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n} Y_i^2 - \frac{n}{n-1}(\overline{Y})^2.$$

Since $E(Y_i) = 0$ and $E(Y_i^2) = \mathrm{Var}(X_i) = \sigma^2$, we have

$$\begin{aligned} E(S^2) &= \frac{n}{n-1}\sigma^2 - \frac{n}{n-1}\underbrace{E\{(\overline{Y})^2\}}_{=\mathrm{Var}(\overline{Y})=\sigma^2/n} \\ &= \frac{n}{n-1}\sigma^2 - \frac{n}{n-1}\frac{\sigma^2}{n} = \sigma^2. \end{aligned}$$

Corollary

We have

$$\sum_{i=1}^{n}(X_i - \overline{X})^2 = \sum_{i=1}^{n} X_i^2 - n(\overline{X})^2.$$

Sample from normal population

- Consider now
$$X_1, \ldots, X_n \sim N(\mu, \sigma^2) \text{ i.i.d.}$$
- We know that
$$\overline{X} \sim N(\mu, \sigma^2/n).$$
- What is the distribution of S^2?

Theorem

(a) $\overline{X}$ *and* S^2 *are independent.*

(b) $\overline{X} \sim N(\mu, \sigma^2/n)$.

(c) $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$.

Caution:

- Any of these properties does not hold if the common distribution is not normal.
- $X_1 - \overline{X}, \ldots, X_n - \overline{X}$ are not independent.

Proof of Part (c)

We note that

$$U = (n-1)S^2/\sigma^2 = \sum_{i=1}^{n} (X_i/\sigma - \overline{X}/\sigma)^2.$$

Let

$$Z_i = (X_i - \mu)/\sigma,$$

so that $Z_i \sim N(0,1)$. Then

$$U = \sum_{i=1}^{n} (Z_i - \overline{Z})^2 = \underbrace{\sum_{i=1}^{n} Z_i^2}_{=V} - \underbrace{(\sqrt{n}\overline{Z})^2}_{=W},$$

namely,

$$V = U + W.$$

By Part (a),

$$\sum_{i=1}^{n}(Z_i - \overline{Z})^2 \quad \text{and} \quad \overline{Z}$$

are independent, so that the mgf of V coincides with the product of mgfs of U and W:

$$\psi_V(\theta) = \psi_U(\theta)\psi_W(\theta).$$

Now, by definition, $V = \sum_{i=1}^{n} Z_i^2 \sim \chi^2(n)$, and $W = (\sqrt{n}\overline{Z})^2 \sim \chi^2(1)$ (why?), so that

$$\psi_V(\theta) = (1 - 2\theta)^{-n/2} \quad \text{and} \quad \psi_W(\theta) = (1 - 2\theta)^{-1/2}$$

for $\theta < 1/2$. Hence,

$$\psi_U(\theta) = \frac{\psi_V(\theta)}{\psi_W(\theta)} = (1 - 2\theta)^{-(n-1)/2}$$

for $\theta < 1/2$, which coincides with the mgf of $\chi^2(n-1)$.

Corollary

We have

$$\frac{\sqrt{n}(\overline{X} - \mu)}{\sqrt{S^2}} \sim t(n-1).$$

Chapter 8 Estimation of Parameters and Fitting of Probability Distributions

Setting

- Random sample
$$X_1, \ldots, X_n \sim F \text{ i.i.d.}$$

- We fit a class of pmfs/pdfs $\{f_\theta \mid \theta \in \Theta\}$ to F, where $\Theta \subset \mathbb{R}^k$, and assume that F (or more precisely its pmf/pdf) is among the class:
$$X_1, \ldots, X_n \sim f_\theta \text{ i.i.d.}$$
The class $\{f_\theta : \theta \in \Theta\}$ is called a (statistical) model, and θ is called a parameter, and Θ is called a parameter space.

- Estimation tries to find a "guess" at the value of θ based on the sample.

Definition

An estimator $\widehat{\theta} = \widehat{\theta}(X_1, \ldots, X_n)$ for θ is a function (statistic) of $X_1, \ldots, X_n$ that takes values in $\mathbb{R}^k$. If the estimator $\widehat{\theta}$ is evaluated at some specific values of $X_1, \ldots, X_n$, i.e., $X_1 = x_1, \ldots, X_n = x_n$, then $\widehat{\theta}(x_1, \ldots, x_n)$ is called an estimate.

An estimator is a random variable (vector), but an estimate is a non-random number.

Poisson distribution

The class of Poisson distributions

$$\{Po(\lambda) \mid \lambda > 0\}$$

has parameter λ, and the parameter space is the positive real line:

$$(0, \infty).$$

If $X_1, \ldots, X_n \sim Po(\lambda)$ i.i.d., a natural estimator for λ is the sample mean:

$$\widehat{\lambda} = \widehat{\lambda}(X_1, \ldots, X_n) = \overline{X}.$$

If $n = 3$ and $X_1 = 3, X_2 = 0, X_3 = 1$, then the value

$$\widehat{\lambda}(3, 0, 1) = 4/3$$

is an estimate.

Goal

Construct an estimator $\widehat{\theta}$ that is "close" to θ *whatever the value of* θ *is*.

“whatever the value of θ is”?

- We want to exclude “betting on a specific value”.
- Consider to estimate λ in $Po(\lambda)$, but you believe that λ is 1:

$$\widehat{\lambda}(X_1, \ldots, X_n) \equiv 1.$$

  If the true value λ is 1, then this estimator has no error.
- But the estimator is disastrous if $\lambda = 200$.

Other classes of distributions

- The class of exponential distributions:

$$\{Ex(\lambda) \mid \lambda > 0\} \quad \theta = \lambda,\ \Theta = (0, \infty).$$

- The class of normal distributions with unit variance:

$$\{N(\mu, 1) \mid \infty < \mu < \infty\} \quad \theta = \mu,\ \Theta = \mathbb{R}.$$

- The class of normal distributions with mean zero:

$$\{N(0, \sigma^2) \mid \sigma^2 > 0\} \quad \theta = \sigma^2,\ \Theta = (0, \infty).$$

- The class of normal distributions with unknown mean and variance:

$$\{N(\mu, \sigma^2) \mid -\infty < \mu < \infty, \sigma^2 > 0\}, \quad \theta = (\mu, \sigma^2),\ \Theta = \mathbb{R} \times (0, \infty).$$

Let f be a pdf on $\mathbb{R}$ (e.g. the Cauchy density).

- The location family with base pdf f is

$$\{f(x-\mu) \mid -\infty < \mu < \infty\}.$$

  μ may not be the mean of $f(x-\mu)$.

- The scale family with base pdf f is

$$\left\{\frac{1}{\sigma} f(x/\sigma) \mid \sigma > 0\right\}.$$

  σ^2 may not be the variance of $\sigma^{-1} f(x/\sigma)$.

- The local-scale family with base pdf f:

$$\left\{\frac{1}{\sigma} f((x-\mu)/\sigma) \mid -\infty < \mu < \infty, \sigma > 0\right\}.$$

μ and σ are called location and scale parameters, resp.

Maximal Likelihood Estimation (MLE)

Example

Suppose that $X \in \{0, 1, 2\}$ associated with a model (a class of pmfs) with parameter $\theta \in \{\theta_0, \theta_1\}$ described by the table

θ	$x = 0$	$x = 1$	$x = 2$
θ_0	0.8	0.1	0.1
θ_1	0.1	0.5	0.4

- If $X = 0$ is observed, then it is more likely that $\theta = \theta_0$.
- If $X = 1$ or 2 is observed, then it is more likely that $\theta = \theta_1$.

This suggests the following estimator:

$$\widehat{\theta} = \widehat{\theta}(X) = \begin{cases} \theta_0 & \text{if } X = 0 \\ \theta_1 & \text{if } X \in \{1, 2\} \end{cases}.$$

Example (Cont.)

Denote by f_θ the pmf with parameter θ; then

$$\widehat{\theta}(X) \text{ maximizes the function } \quad \underbrace{\theta \mapsto f_\theta(X)}_{\text{likelihood function}} \quad ,$$

and $\widehat{\theta}$ is called the Maximum Likelihood Estimator (MLE).