Math 362: Mathematical Statistics II

Le Chen
le.chen@emory.edu
chenle02@gmail.com

Emory University
Atlanta, GA

Last updated on Spring 2021
Last compiled on January 15, 2023

2021 Spring

Chapter 7. Inference Based on The Normal Distribution

§ 7.1 Introduction

§ 7.2 Comparing $\frac{\overline{Y}-\mu}{\sigma/\sqrt{n}}$ and $\frac{\overline{Y}-\mu}{S/\sqrt{n}}$

§ 7.3 Deriving the Distribution of $\frac{\overline{Y}-\mu}{S/\sqrt{n}}$

§ 7.4 Drawing Inferences About μ

§ 7.5 Drawing Inferences About σ^2

§ 7.1 Introduction

§ 7.2 Comparing $\frac{\overline{Y}-\mu}{\sigma/\sqrt{n}}$ and $\frac{\overline{Y}-\mu}{S/\sqrt{n}}$

§ 7.3 Deriving the Distribution of $\frac{\overline{Y}-\mu}{S/\sqrt{n}}$

§ 7.4 Drawing Inferences About μ

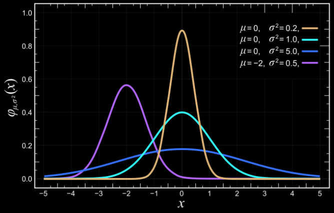
§ 7.5 Drawing Inferences About σ^2

Carl Friedrich Gauss discovered the normal distribution in 1809 as a way to rationalize the method of least squares.

(1777-1855)

Marquis de Laplace proved the central limit theorem in 1810, consolidating the importance of the normal distribution in statistics.

(1749-1827)



Notation	$\mathcal{N}(\mu, \sigma^2)$
Parameters	$\mu \in \mathbb{R}$ = mean (location) $\sigma^2 > 0$ = variance (squared scale)
Support	$x \in \mathbb{R}$
PDF	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
CDF	$\frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$
Quantile	$\mu + \sigma\sqrt{2}\operatorname{erf}^{-1}(2p-1)$
Mean	μ
Median	μ
Mode	μ
Variance	σ^2
MAD	$\sigma\sqrt{2/\pi}$
Skewness	0
Ex. kurtosis	0
Entropy	$\frac{1}{2}\log(2\pi e\sigma^2)$
MGF	$\exp(\mu t + \sigma^2 t^2/2)$
CF	$\exp(i\mu t - \sigma^2 t^2/2)$
Fisher information	$\mathcal{I}(\mu,\sigma) = \begin{pmatrix} 1/\sigma^2 & 0 \\ 0 & 2/\sigma^2 \end{pmatrix}$ $\mathcal{I}(\mu,\sigma^2) = \begin{pmatrix} 1/\sigma^2 & 0 \\ 0 & 1/(2\sigma^4) \end{pmatrix}$
Kullback-Leibler divergence	$D_{\text{KL}}(\mathcal{N}_0 \parallel \mathcal{N}_1) = \frac{1}{2}\left\{\left(\frac{\sigma_0}{\sigma_1}\right)^2 + \frac{(\mu_1-\mu_0)^2}{\sigma_1^2} - 1 + 2\ln\frac{\sigma_1}{\sigma_0}\right\}$

https://en.wikipedia.org/wiki/Normal_distribution

Test for normal parameters (one sample test)

Let $Y_1, \cdots, Y_n$ be a random sample from $N(\mu, \sigma^2)$.

Prob. 1 Find a test statistic Λ in order to test $H_0 : \mu = \mu_0$ v.s. $H_1 : \mu \neq \mu_0$.

When σ^2 is known: $$\Lambda = \frac{\overline{Y} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$$

When σ^2 is unknown: $\Lambda = ?$ $$\Lambda \stackrel{?}{=} \frac{\overline{Y} - \mu_0}{s/\sqrt{n}} \sim ?$$

Prob. 2 Find a test statistic Λ in order to test $H_0 : \sigma^2 = \sigma_0^2$ v.s. $H_1 : \sigma^2 \neq \sigma_0^2$.

Prob. 1 Find a test statistic for $H_0 : \mu = \mu_0$ v.s. $H_1 : \mu \neq \mu_0$, with σ^2 unknown

Sol. Composite-vs-composite test with:

$$\omega = \left\{(\mu, \sigma^2) : \mu = \mu_0,\ \sigma^2 > 0\right\}$$

$$\Omega = \left\{(\mu, \sigma^2) : \mu \in \mathbb{R},\ \sigma^2 > 0\right\}$$

The MLE under the two spaces are:

$$\omega_e = (\mu_e, \sigma_e^2) : \qquad \mu_e = \mu_0 \quad \text{and} \quad \sigma_e^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \mu_0)^2 \quad (\text{Under } \omega)$$

$$\Omega_e = (\mu_e, \sigma_e^2) : \qquad \mu_e = \bar{y} \quad \text{and} \quad \sigma_e^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 \quad (\text{Under } \Omega)$$

$$L(\mu, \sigma^2) = (2\pi\sigma^2)^{-n} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i - \mu}{\sigma}\right)^2\right)$$

$$L(\omega_e) = \cdots = \left[\frac{ne^{-1}}{2\pi \sum_{i=1}^{n}(y_i - \mu_0)^2}\right]^{n/2}$$

$$L(\Omega_e) = \cdots = \left[\frac{ne^{-1}}{2\pi \sum_{i=1}^{n}(y_i - \bar{y})^2}\right]^{n/2}$$

Hence,

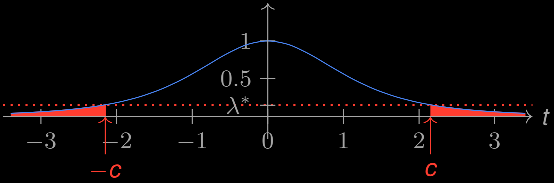
$$\lambda = \frac{L(\omega_e)}{L(\Omega_e)} = \left[\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \mu_0)^2}\right]^{n/2} = \cdots = \left[1 + \frac{n(\bar{y} - \mu_0)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}\right]^{-n/2}$$

$$= \left[1 + \frac{1}{n-1}\left(\frac{\bar{y} - \mu_0}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2} \Big/ \sqrt{n}}\right)^2\right]^{-n/2}$$

$$= \left[1 + \frac{1}{n-1}\left(\frac{\bar{y} - \mu_0}{s / \sqrt{n}}\right)^2\right]^{-n/2}$$

$$= \left[1 + \frac{t^2}{n-1}\right]^{-n/2}, \quad t = \frac{\bar{y} - \mu_0}{s / \sqrt{n}}$$

$$\lambda(t) = \left(1 + \frac{t^2}{n-1}\right)^{-\frac{n}{2}}$$



$$\lambda \in (0, \lambda^*] \qquad \Leftrightarrow \qquad |t| \geq c.$$

Finally, the test statistic is

$$\boxed{T = \frac{\overline{Y} - \mu_0}{S/\sqrt{n}}}$$

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

The critical region takes the form: $|t| \geq c$.

Question: Find the exact distribution of T.

Prob. 2 Find a test statistic for $H_0 : \sigma^2 = \sigma_0^2$ v.s. $H_1 : \sigma^2 \neq \sigma_0^2$, with μ unknown

Sol. Composite-vs-composite test with:

$$\omega = \left\{(\mu, \sigma^2) : \mu \in \mathbb{R},\ \sigma^2 = \sigma_0^2\right\}$$
$$\Omega = \left\{(\mu, \sigma^2) : \mu \in \mathbb{R},\ \sigma^2 > 0\right\}$$

The MLE under the two spaces are:

$$\omega_e = (\mu_e, \sigma_e^2) : \qquad \mu_e = \bar{y} \quad \text{and} \quad \sigma_e^2 = \sigma_0^2 \qquad \text{(Under } \omega)$$

$$\Omega_e = (\mu_e, \sigma_e^2) : \qquad \mu_e = \bar{y} \quad \text{and} \quad \sigma_e^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 \qquad \text{(Under } \Omega)$$

$$L(\mu, \sigma^2) = (2\pi\sigma^2)^{-n} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i - \mu}{\sigma}\right)^2\right)$$

$$L(\omega_e) = (2\pi\sigma^2)^{-n} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i - \bar{y}}{\sigma_0}\right)^2\right)$$

$$L(\Omega_e) = \cdots = \left[\frac{ne^{-1}}{2\pi \sum_{i=1}^{n}(y_i - \bar{y})^2}\right]^{n/2}$$

Hence,

$$\lambda = \frac{L(\omega_e)}{L(\Omega_e)} = \left[\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n\sigma_0^2}\right]^{n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{y_i - \bar{y}}{\sigma_0}\right)^2 + \frac{n}{2}\right)$$

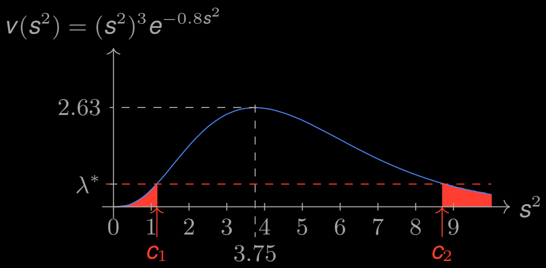
$$= \left[\frac{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}{\frac{n}{n-1}\sigma_0^2}\right]^{n/2} \exp\left(-\frac{n-1}{2\sigma_0^2}\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 + \frac{n}{2}\right)$$

$$= \left[\frac{s^2}{\frac{n}{n-1}\sigma_0^2}\right]^{n/2} \exp\left(-\frac{n-1}{2\sigma_0^2}s^2 + \frac{n}{2}\right)$$

$$\Downarrow$$

$$\lambda(s^2) = \left[\frac{s^2}{\frac{n}{n-1}\sigma_0^2}\right]^{n/2} \exp\left(-\frac{n-1}{2\sigma_0^2}s^2 + \frac{n}{2}\right) \quad \Longleftrightarrow \quad v(s^2) = (s^2)^{\frac{n}{2}} e^{-\lambda s^2}$$

By setting $n = 6$ and $\lambda = 0.8$, we see ...



This suggests that the critical region should be of the form in terms of s^2:

$$(0, c_1) \cup (c_2, \infty)$$

For convenience, we put $\alpha/2$ mass on each tails of S^2:

Find c_1 and c_2 such that

$$\int_0^{c_1} f_{S^2}(z)dz = \int_{c_2}^{\infty} f_{S^2}(z)dz = \frac{\alpha}{2}.$$

Finally, the test statistic is

$$\boxed{S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2} \quad \text{with} \quad \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

Question: Find the exact distribution of S^2.