Introduction to mathematical Statistics Lecture7

Inference on two population variances

-- An introduction and preview

* Both pop's are normal, two independent samples

Sample 1 : $X_1, X_2, \cdots, X_{n_1} \overset{i.i.d.}{\sim} N(\mu_1, \sigma_1^2) \quad \Rightarrow \dfrac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi^2_{n_1-1}$

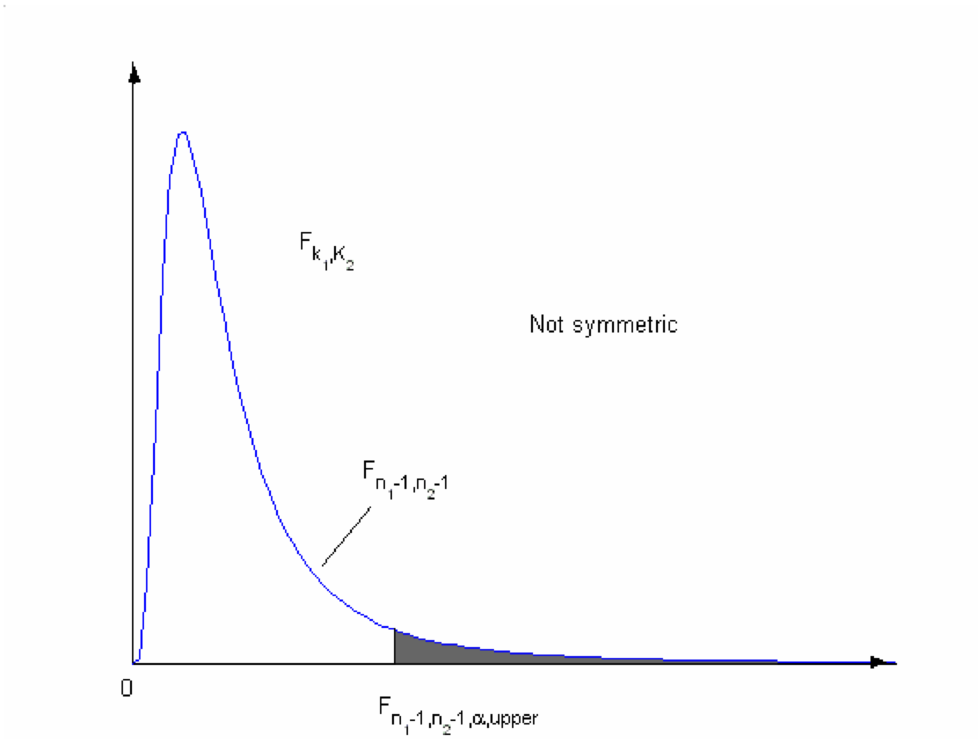
Sample 2 : $Y_1, Y_2, \cdots, Y_{n_2} \overset{i.i.d.}{\sim} N(\mu_2, \sigma_2^2) \quad \Rightarrow \dfrac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi^2_{n_2-1}$

1. point estimator : $\hat{\sigma}_1^2 = S_1^2$ (parameter of interest : $\dfrac{\sigma_1^2}{\sigma_2^2}$)

$\hat{\sigma}_2^2 = S_2^2$

Def. F-distribution Let $W_1 \sim \chi^2_{k_1}$, $W_2 \sim \chi^2_{k_2}$, W_1, W_2 are independent.

Then, $F = \dfrac{W_1/k_1}{W_2/k_2} \sim F_{k_1,k_2}$



$\alpha = P(F \le F_{k_1,k_2,\alpha,lower})$

$$= P\left(\frac{1}{F} \geq \frac{1}{F_{k_1,k_2,\alpha,lower}}\right)$$

$$\boxed{\frac{1}{F} \sim F_{k_2,k_1}}$$

$$\boxed{F_{k_1,k_2,\alpha,lower} = \frac{1}{F_{k_2,k_1,\alpha,upper}}}$$

$$\boxed{F = \frac{\dfrac{(n_1-1)S_1^2}{\sigma_1^2} \Big/ (n_1-1)}{\dfrac{(n_2-1)S_2^2}{\sigma_2^2} \Big/ (n_2-1)} = \frac{\dfrac{S_1^2}{S_2^2}}{\dfrac{\sigma_1^2}{\sigma_2^2}} \sim F_{n_1-1,n_2-1}}$$

The F distribution was tabulated - and the letter introduced - by G. W. Snedecor Calculation and Interpretation of Analysis of Variance and Covariance (1934). For this reason it is sometimes called "Snedecor's F distribution".

George Waddel Snedecor (October 20, 1881 – February 15, 1974) was an American mathematician and statistician. He contributed to the foundations of analysis of variance, data analysis, experimental design, and statistical methodology. Snedecor's F distribution and the George W. Snedecor Award of the American Statistical Association are named after him.

Snedecor founded the first academic department of statistics in the United States, at Iowa State University. He also created the first statistics laboratory in the U.S. at Iowa State, and was a pioneer of modern applied statistics in the U.S. His 1938 textbook *Statistical Methods* became an essential resource: "In the 1970s, a review of citations in published scientific articles from all areas of science showed that Snedecor's *Statistical Methods* was the most frequently cited book."[1]

2. $100(1-\alpha)\%$ CI for $\frac{\sigma_1^2}{\sigma_2^2}$

$$1-\alpha = P\left(F_{\alpha/2,lower} \leq \frac{\dfrac{S_1^2}{S_2^2}}{\dfrac{\sigma_1^2}{\sigma_2^2}} \leq F_{\alpha/2,upper}\right)$$

$$= P\left(\frac{1}{F_{\alpha/2, lower}} \geq \frac{\frac{\sigma_1^2}{\sigma_2^2}}{\frac{S_1^2}{S_2^2}} \geq \frac{1}{F_{\alpha/2, upper}}\right)$$

$$= P\left(\frac{\frac{S_1^2}{S_2^2}}{F_{\alpha/2, upper}} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{\frac{S_1^2}{S_2^2}}{F_{\alpha/2, lower}}\right)$$

3. Test

$$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$$

$$H_a : \frac{\sigma_1^2}{\sigma_2^2} \neq 1$$

Test Statistic $F_0 = \frac{S_1^2}{S_2^2} \overset{H_0}{\sim} F_{n_1-1, n_2-1}$

At the significance level α, we reject H_0 if F_0 is too large or too small.

$F_0 \geq c_1$, $F_0 \leq c_2$

* conventional boundaries / thresholds

$c_1 = F_{n_1-1, n_2-1, \alpha/2, upper}$

$c_2 = F_{n_1-1, n_2-1, \alpha/2, lower}$

Homework #2. Due Monday, March 7th, before class

4.1.5, 4.1.8, 4.1.9, 4.2.3, 4.2.5, 4.2.11, 4.2.12, 4.2.15, 4.2.18, 4.2.20, 4.2.23, 4.2.25, 4.3.1, 4.3.2, 4.3.6, 4.4.2, 4.4.7, 4.4.8, 4.4.12

Typos found in our text book:

1. **Page 201, Example 4.2.7, "t" should be changed to "Z" as what we have there is a standard normal random variable – and not t with 19 degrees of freedom -- since the population standard deviation σ was given. The numerical results would still be about 0.**
2. **Page 210, Example 4.3.4, since 1/λ = 2, so λ = ½. The example incorrectly uses λ = 2.**

Review. The exponential distribution

Exercise:

A post office has two clerks, Lucy and Ricky. It is known that their service times are two independent exponential random variables with the same parameter λ, and it is known that Lucy spends on the average 10 minutes with each customer. Suppose Lucy and Ricky are each serving one customer at the moment,

(a) What is the distribution of the waiting time for the next customer in line?
(b) What is the probability that the next customer in line will be the last customer (among the two customers being served and him/herself) to leave the post office?

Solution:

(a) Let X_1 and X_2 denote the service time for Lucy and Ricky, respectively. Then the waiting time for the next customer in line is:
$Y=\min(X_1,X_2)$, where $X_i \sim$ iid $\exp(\lambda)$, $i=1,2$.
Furthermore, $E(X_1)=1/\lambda=10$, $\lambda=1/10$. From the definition of exponential R.V.,

$f_{X_i}(x_i)=\lambda e^{-\lambda x_i}$, $x_i>0$, $i=1,2$ and $P(X_i>x)=\int_x^\infty f_{x_i}(u)du = e^{-\lambda x_i}$, $x_i>0$, $i=1,2$.

$F_Y(y)=P(Y\le y)=1-P(Y>y)=1-P(X_1>y,X_2>y)=1-P(X_1>y)\cdot P(X_2>y)=1-(e^{-\lambda y})^2=1-e^{-2\lambda y}$

Therefore $f_Y(y)=\frac{d}{dy}F_Y(y)=(2\lambda)e^{-(2\lambda)y}$, $y>0$. *Thus* $Y\sim \exp(2\lambda)$, *where* $\lambda=1/10$.

Note: Because of the memoryless property of the exponential distribution, it does not matter how long the current customers has been served by Lucy or Ricky.

(b)One of the two customers being served right now will leave first. Now between the customer who is still being served and the next customer in line, their service time would follow the same exponential distribution because of the memoryless property of the exponential random variable. Therefore, each would have the same chance to finish first or last. That is, the next customer in line will be the last customer to leave with probability 0.5.

Note: The memoryless property of an exponential random variable X is described as follows:

$$P(X>a+b\mid X>a)=P(X>a)$$

where a and b are any positive constants.

Chapter 5: Point Estimation

Point Estimators

Example. Let $X_1, X_2, \cdots, X_n$ be a random sample from $N(\mu, \sigma^2)$.

Please find a good point estimator for $1. \mu \quad 2. \sigma^2$

Solutions. $1. \hat{\mu} = \bar{X}$

$2. \hat{\sigma}^2 = S^2$

There are the typical estimators for μ and S^2. Both are unbiased estimators.

Property of Point Estimators

Unbiased. $\hat{\theta}$ is said to be an unbiased estimator for θ if $E(\hat{\theta}) = \theta$.

$$E(\bar{X}) = E\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right)$$

$$= \frac{E(X_1) + E(X_2) + \cdots + E(X_n)}{n} = \frac{\mu + \mu + \cdots + \mu}{n}$$

$$= \mu$$

$E(S^2) = \sigma^2$ (Proof: Homework #2 extra problem)

Unbiased estimator may not be unique.

Example. $E[\sum a_i X_i] = \left(\sum a_i\right)\mu$

$$\tilde{\mu} = \frac{\sum_{i=1}^{n} a_i X_i}{\sum_{i=1}^{n} a_i} \text{ since } E(\tilde{\mu}) = \mu$$

Variance of the unbiased estimators

$$\left.\begin{aligned} Var(\bar{X}) &= \frac{\sigma^2}{n} \\ Var(X_i) &= \sigma^2 \end{aligned}\right\} Var(\bar{X}) \prec Var(X_i) \text{ when } n \succ 1$$

Methods for deriving point estimators

1. Maximum Likelihood Estimator (MLE)

2. Method of Moment Estimator (MOME)

Example (continued). $X_i \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$, $i = 1, \cdots, n$

1. Derive the MLE for μ and σ^2.

2. Derive the MOME for μ and σ^2.

Solution.

1. MLE

[i] pdf: $f(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right]$, $x_i \in R, i = 1, \cdots, n$

[ii] likelihood function $= L = f(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} f(x_i)$

$$= \prod_{i=1}^{n} \left\{ (2\pi\sigma^2)^{-\frac{1}{2}} \exp[-\frac{(x_i-\mu)^2}{2\sigma^2}] \right\}$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2} \right]$$

[iii] log likelihood function

$$l = \ln L = (-\frac{n}{2})\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}$$

[iv] Find the MLE's

$$\begin{cases} \frac{\partial l}{\partial \mu} = \frac{2\sum_{i=1}^{n}(x_i-\mu)}{2\sigma^2} = 0 \\ \frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^4} = 0 \end{cases}$$

$$\Rightarrow \begin{cases} \hat{\mu} = \bar{X} \\ \hat{\sigma}^2 = \frac{\sum (X_i - \bar{X})^2}{n} \end{cases}$$

2. MOME

Order	Population Moment	Sample Moment
1st	$E(X)$	$= \frac{X_1 + X_2 + \cdots + X_n}{n}$
2nd	$E(X^2)$	$= \frac{X_1^2 + X_2^2 + \cdots + X_n^2}{n}$
⋮	⋮	⋮
kth	$E(X^k)$	$= \frac{X_1^k + X_2^k + \cdots + X_n^k}{n}$

⋮	⋮	⋮

$$E(X) = \mu = \bar{X}$$

$$E(X^2) = \mu^2 + \sigma^2 = \frac{\sum_{i=1}^{n} X_i^2}{n}$$

$$\Rightarrow \begin{cases} \hat{\mu} = \bar{X} \\ \hat{\sigma}^2 = \dfrac{\sum_{i=1}^{n} X_i^2}{n} - (\bar{X})^2 \end{cases}$$

Proof.

$$\frac{\sum_{i=1}^{n} X_i^2}{n} - (\bar{X})^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X} + \bar{X})^2}{n} - (\bar{X})^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2 + \sum_{i=1}^{n} 2\bar{X}(X_i - \bar{X}) + \sum_{i=1}^{n} (\bar{X})^2}{n} - (\bar{X})^2$$

$$= \frac{\sum (X_i - \bar{X})^2}{n}$$

Therefore, the MLE and MOME for σ^2 are the same for the normal population.

$$E(\hat{\sigma}^2) = E\left(\frac{\sum (X_i - \bar{X})^2}{n}\right) = E\left[\frac{n-1}{n}\frac{\sum (X_i - \bar{X})^2}{n-1}\right]$$

$$= \frac{n-1}{n} E(S^2) = \frac{n-1}{n}\sigma^2 \overset{n\to\infty}{\Longrightarrow} \sigma^2 \text{ (asymptotically unbiased)}$$

Example. Let $X_i \overset{i.i.d.}{\sim} Bernoulli(p)$, $i = 1, \cdots, n$.

Please derive 1. The MLE of p   2. The MOME of p.

Solution.

1. MLE

   [i] $f(x_i) = p^{x_i}(1-p)^{1-x_i}$, $i = 1, \cdots, n$

   [ii] $L = \prod f(x_i) = p^{\sum x_i}(1-p)^{n - \sum x_i}$

   [iii] $l = \ln L = (\sum x_i)\ln p + (n - \sum x_i)\ln(1-p)$

   [iv] $\dfrac{dl}{dp} = \dfrac{\sum x_i}{p} - \dfrac{n - \sum x_i}{1-p} = 0$

   $\Rightarrow \hat{p} = \dfrac{\sum x_i}{n}$

2. MOME

   $E(X) = p = \dfrac{X_1 + X_2 + \cdots + X_n}{n} \qquad \Rightarrow \hat{p} = \dfrac{\sum x_i}{n}$