HUL315: Econometric Methods

Harshit Goyal: 2021MT10143

January 28, 2024

Some Properties

- **Linearity of Expectation**: If X and Y are random variables, $a \in \mathbb{R}$ be any constant:

$$E[X + a] = E[X] + a \tag{1}$$
$$E[aX] = aE[X] \tag{2}$$
$$E[X + Y] = E[X] + E[Y] \tag{3}$$

- Variance: If X and Y are random variables, $a \in \mathbb{R}$ be any constant:

$$Var(X + a) = Var(X) \tag{4}$$
$$Var(aX) = a^2 * Var(X) \tag{5}$$
$$Var(X + Y) = Var(X) + Var(Y) + 2 * Cov(X, Y) \tag{6}$$

- X and Y are **independent** $\implies$ X and Y are **uncorrelated**, $Cov(X, Y) = 0$ and:

$$Var(X + Y) = Var(X) + Var(Y) \tag{7}$$

- For any constant $a \in \mathbb{R}$:

$$X \sim N(\mu, \sigma^2) \implies X - a \sim N(\mu - a, \sigma^2) \tag{8}$$

- For any constant $a \in \mathbb{R}, a \neq 0$:

$$X \sim N(\mu, \sigma^2) \implies aX \sim N(a\mu, a^2\sigma^2) \tag{9}$$

- If X and Y are **independent** Random Variables, $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$ then:

$$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2) \tag{10}$$

- If $X_1, X_2, X_3....X_n$ are Random Samples from a population with mean μ and variance σ^2 then $\forall i \in \{1, 2, ...n\}$

$$E[X_i] = \mu \tag{11}$$
$$Var(X_i) = \sigma^2 \tag{12}$$

Problem 1

The t-distribution with r degrees of freedom can be defined as the ratio of two independent random variables. The numerator being a $N(0,1)$ random variable and the denominator being the square-root of a χ^2 random variable divided by its degrees of freedom. The t-distribution is a symmetric distribution like the Normal distribution but with fatter tails. As $r \to \infty$ the t-distribution approaches the Normal distribution.

1. Verify that if $X_1, ..., X_n$ are a random samples drawn from a $N(\mu, \sigma^2)$ distribution, then $z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ is $N(0,1)$.

$$\bar{X} = \frac{X_1 + X_2 .....X_n}{n}$$

Hence,

$$\begin{aligned} E[\bar{X}] &= E[\frac{X_1 + X_2 .....X_n}{n}] \\ &= \frac{E[X_1] + E[X_2].....E[X_n]}{n} \\ &= \frac{\mu + \mu.....\mu}{n} = \frac{n\mu}{n} = \mu \end{aligned}$$

So we get,

$$E[\bar{X}] = \mu \tag{13}$$

Since, X_i's are sampled independently, $Cov(X_i, X_j) = 0 \ \ \forall \ \ i \neq j$

$$\begin{aligned} \implies Var[\bar{X}] &= \frac{1}{n^2} Var[X_1 + X_2....X_n] \\ &= \frac{Var[X_1] + Var[X_2].....Var[X_n]}{n^2} \\ &= \frac{\sigma^2 + \sigma^2.....\sigma^2}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n} \end{aligned}$$

So we get,

$$Var[\bar{X}] = \frac{\sigma^2}{n} \tag{14}$$

Using (10), (13) and (14), we can say

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \tag{15}$$

Using (8),

$$\bar{X} - \mu \sim N(0, \frac{\sigma^2}{n})$$

Using (9),

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0, \frac{1}{(\sigma/\sqrt{n})^2}\frac{\sigma^2}{n})$$
$$\implies \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

2. Use the fact that $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$ to show that $t = \frac{z}{\sqrt{S^2/\sigma^2}} = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ has a t-distribution with $(n-1)$ degrees of freedom.

$$\begin{aligned}\frac{\bar{X}-\mu}{S/\sqrt{n}} &= \frac{\bar{X}-\mu}{(S*\sigma)/(\sqrt{n}*\sigma)} \\ &= \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\frac{1}{S/\sigma} \\ &= \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\frac{1}{\sqrt{S^2/\sigma^2}} \\ &= \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\frac{1}{\frac{\sqrt{(n-1)S^2/\sigma^2}}{\sqrt{(n-1)}}}\end{aligned}$$

$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ and $\frac{\sqrt{(n-1)S^2/\sigma^2}}{\sqrt{(n-1)}}$ is $\chi^2_{(n-1)}$ divided by square root of it's degree of freedom $\sqrt{n-1}$. Hence, $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ is t-distribution, with $(n-1)$ degrees of freedom.

3. For $n = 16, \bar{x} = 20$ and $s^2 = 4$, construct a 95% confidence interval for μ.

   For 95% Confidence Interval, we need c

$$P(-c < \frac{\bar{X}-\mu}{S/\sqrt{n}} < c) = 0.95$$

   From t-table, (15 degrees of freedom, 95% confidence) $c = 2.131$

$$\begin{aligned}\text{Confidence Interval} &= [\bar{x} - \frac{c*s}{\sqrt{n}}, \bar{x} + \frac{c*s}{\sqrt{n}}] \\ &= [20 - \frac{2.131*2}{\sqrt{16}}, 20 + \frac{2.131*2}{\sqrt{16}}] \\ &= [18.9345, 21.0655]\end{aligned}$$

   Hence the 95% Confidence Interval is $[18.9345, 21.0655]$

Problem 2

Let $\bar{Y}$ denote the sample average from a random sample with mean μ and variance σ^2. Consider two alternative estimators of μ: $W_1 = \frac{n-1}{n}\bar{Y}$ and $W_2 = \frac{\bar{Y}}{2}$.
We know,

$$E[\bar{Y}] = \mu$$

Bias of an estimator (of parameter θ) W is $E[W - \theta] = E[W] - \theta$, denoted as

$$Bias(W, \theta)$$

1. Show that W_1 and W_2 are both biased estimators of μ and find the biases. What happens to the biases as $n \to \infty$.

$$E[W_1] = E[\frac{n-1}{n}\bar{Y}] = \frac{n-1}{n}E[\bar{Y}] = \frac{n-1}{n}\mu$$

$$\boxed{E[W_1] = \frac{n-1}{n}\mu}$$

Now,

$$Bias(W_1, \mu) = E[W_1] - \mu = \frac{n-1}{n}\mu - \mu = \frac{-\mu}{n}$$

Since $Bias(W_1, \mu) \neq 0 \;\; (\mu \neq 0)$, $\;W_1$ is a biased estimator of μ. Now.

$$\lim_{n\to\infty} Bias(W_1, \mu) = \lim_{n\to\infty} \frac{-\mu}{n} = 0$$

Similarly,

$$E[W_2] = E[\frac{\bar{Y}}{2}] = \frac{E[\bar{Y}]}{2} = \frac{\mu}{2}$$

$$\boxed{E[W_2] = \frac{\mu}{2}}$$

Now,

$$Bias(W_2, \mu) = E[W_2] - \mu = \frac{\mu}{2} - \mu = \frac{-\mu}{2}$$

Since $Bias(W_2, \mu) \neq 0(\mu \neq 0)$, $\;W_2$ is a biased estimator of μ. Now,

$$\lim_{n\to\infty} Bias(W_2, \mu) = \lim_{n\to\infty} \frac{-\mu}{2} = \frac{-\mu}{2}$$

2. Find the probability limits of W_1 and W_2. Which estimator is consistent?

Denote W_1 constructed using n samples as W_1^n. Similarly define W_2^n.

$$\boxed{plim(W_1^n) = \mu}$$

$plim(W_1) = \mu$ shows W_1 is a consistent estimator of μ.

$$\boxed{plim(W_2^n) = \frac{\mu}{2}}$$

$plim(W_2) \neq \mu$ shows W_2 not a consistent estimator of μ.

Proof: For every $\epsilon > 0$

$$\begin{aligned}
P(|W_1^n - \mu| < \epsilon) &= P(|\frac{n-1}{n}\bar{Y} - \mu| < \epsilon) \\
&= P(\mu - \epsilon < \frac{n-1}{n}\bar{Y} < \mu + \epsilon) \\
&= P(\frac{n}{n-1}(\mu - \epsilon) < \bar{Y} < \frac{n}{n-1}(\mu + \epsilon)) \\
&= P(\frac{\mu}{n-1} - \frac{n}{n-1}\epsilon) < \bar{Y} - \mu < \frac{\mu}{n-1} + \frac{n}{n-1}\epsilon)) \\
&= P(\frac{\mu}{\sigma}\frac{\sqrt{n}}{n-1} - \frac{\epsilon}{\sigma}\frac{n\sqrt{n}}{n-1} < \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} < \frac{\mu}{\sigma}\frac{\sqrt{n}}{n-1} + \frac{\epsilon}{\sigma}\frac{n\sqrt{n}}{n-1})
\end{aligned}$$

Let $l_n = \frac{\mu}{\sigma}\frac{\sqrt{n}}{n-1} - \frac{\epsilon}{\sigma}\frac{n\sqrt{n}}{n-1}$ and $r_n = \frac{\mu}{\sigma}\frac{\sqrt{n}}{n-1} + \frac{\epsilon}{\sigma}\frac{n\sqrt{n}}{n-1}$
Notice:

$$\begin{aligned}
\lim_{n\to\infty} \frac{\mu}{\sigma}\frac{\sqrt{n}}{n-1} &= 0 \\
\lim_{n\to\infty} \frac{\epsilon}{\sigma}\frac{n\sqrt{n}}{n-1} &= +\infty
\end{aligned}$$

So,

$$\begin{aligned}
\lim_{n\to\infty} l_n &= -\infty \\
\lim_{n\to\infty} r_n &= +\infty
\end{aligned}$$

Using the Central Limit Theorem, in the limit of large n,

$$\frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$$\begin{aligned}
&P(|W_1^n - \mu| < \epsilon) = \Phi(r_n) - \Phi(l_n) \\
\implies &\lim_{n\to\infty} P(|W_1^n - \mu|) = \Phi(+\infty) - \Phi(-\infty) = 1 - 0 = 1 \\
\implies &\lim_{n\to\infty} P(|W_1^n - \mu| > \epsilon) = 0
\end{aligned}$$

Here Φ is the CDF of $N(0,1)$

$$
\begin{aligned}
P(|W_2^n - \frac{\mu}{2}| < \epsilon) &= P(|\frac{\bar{Y}}{2} - \frac{\mu}{2}| < \epsilon) \\
&= P(-2\epsilon < \bar{Y} - \mu < 2\epsilon) \\
&= P(\frac{-2\sqrt{n}\epsilon}{\sigma} < \frac{\bar{Y} - \mu}{\sigma/\sqrt{n}} < \frac{-2\sqrt{n}\epsilon}{\sigma})
\end{aligned}
$$

Let $l_n = \frac{-2\sqrt{n}\epsilon}{\sigma}$ and $r_n = \frac{2\sqrt{n}\epsilon}{\sigma}$
Notice:

$$
\begin{aligned}
\lim_{n\to\infty} l_n &= -\infty \\
\lim_{n\to\infty} r_n &= +\infty
\end{aligned}
$$

$$
\begin{aligned}
&P(|W_2^n - \mu| < \epsilon) = \Phi(r_n) - \Phi(l_n) \\
&\implies \lim_{n\to\infty} P(|W_2^n - \mu|) = \Phi(+\infty) - \Phi(-\infty) = 1 - 0 = 1 \\
&\implies \lim_{n\to\infty} P(|W_2^n - \mu| > \epsilon) = 0
\end{aligned}
$$

Problem 3

Suppose that you have two independent unbiased estimators of the same parameter a, say $\hat{a_1}$ and $\hat{a_2}$ with different standard deviations σ_1 and σ_2 . What linear combination of $\hat{a_1}$ and $\hat{a_2}$ is the minimum variance unbiased estimator of a?

Let $\hat{a}$ be the desired estimator. For some $\lambda, \mu \in \mathbb{R}$ we can write it as:

$$\hat{a} = \lambda\hat{a_1} + \mu\hat{a_2}$$

Since $\hat{a}, \hat{a_1}, \hat{a_2}$ are unbiased, $E[\hat{a}] = E[\hat{a_1}] = E[\hat{a_2}] = a$, we get:

$$E[\hat{a}] = E[\lambda\hat{a_1} + \mu\hat{a_2}] = 0$$

Using **Linearity of Expectation**,

$$\lambda E[\hat{a_1}] + \mu E[\hat{a_1}] = a \implies \lambda a + \mu a = a$$

Assuming $a \neq 0$, we get

$$\lambda + \mu = 1 \implies \mu = 1 - \lambda$$

Now, $\hat{a}$ can be written as:

$$\hat{a} = \lambda\hat{a_1} + (1 - \lambda)\hat{a_2}$$

Variance of $\hat{a}$, σ^2 is given as:

$$\begin{aligned}
\sigma^2 &= E[(\hat{a} - E[\hat{a}])^2] \\
&= E[(\hat{a} - a)^2] \\
&= E[(\lambda\hat{a_1} + (1-\lambda)\hat{a_2} - a)^2] \\
&= E[(\lambda(\hat{a_1} - a) + (1-\lambda)(\hat{a_2} - a))^2] \\
&= \lambda^2 E[(\hat{a_1} - a)^2] + (1-\lambda)^2 E[(\hat{a_2} - a)^2] + 2 * \lambda(1-\lambda)E[(\hat{a_1} - a)(\hat{a_2} - a)]
\end{aligned}$$

Now, $E[(\hat{a_1} - a)^2] = \sigma_1$ and $E[(\hat{a_2} - a)^2] = \sigma_2$.
Since $\hat{a_1}$ and $\hat{a_2}$ are **independent**, they are **uncorrelated**, hence $E[(\hat{a_1} - a)(\hat{a_2} - a)] = Cov(\hat{a_1}, \hat{a_2}) = 0$. This gives:

$$\sigma^2 = \lambda^2\sigma_1^2 + (1-\lambda)^2\sigma_2^2$$

Since σ^2 is a continuous and differentiable function of λ, to get the minima, we set:

$$\begin{aligned}
&\frac{d\sigma^2}{d\lambda} = 0 \\
\implies &\sigma^2 = \lambda^2\sigma_1^2 + (1-\lambda)^2\sigma_2^2 \\
\implies &2\lambda\sigma_1^2 - 2(1-\lambda)\sigma_2^2 = 0 \\
\implies &\lambda = \frac{\sigma_2^2}{\sigma_2^2 + \sigma_1^2}
\end{aligned}$$

So unbiased estimator $\hat{a}$ minimising variance is,

$$\boxed{\hat{a} = \frac{\sigma_2^2}{\sigma_2^2 + \sigma_1^2}\hat{a_1} + \frac{\sigma_1^2}{\sigma_2^2 + \sigma_1^2}\hat{a_2}}$$

Above is a minima, as the expression for σ^2 is a quadratic function of λ with positive leading coefficient ($= \sigma_1^2 + \sigma_2^2$).