STAB57: An Introduction to Statistics

Shahriar Shams

Week 3 (Sampling distribution of S^2 and some related distributions)

Winter 2023

Recap of Week 2

- Learned two formal ways of defining an estimator
  - Method of moments estimator.
  - Maximum Likelihood Estimator (MLE).

- Sampling distribution
  - Sampling distribution of ($\bar{X}$)

- Considering T as an estimator of θ,

$$MSE(T) = var(T) + (Bias(T))^2$$

- $Bias(T) = 0 \implies$ The estimator is Unbiased.

NOTE:

- "$\bar{X}$ is an unbiased estimator" is an *incomplete sentence*!
- We have to say "$\bar{X}$ is an unbiased estimator of μ".
- This same $\bar{X}$ can be biased for some other parameters.

Learning goals for this week

- Reviewing MSE (slides that we couldn't cover in week-2)
- Which formula to use for sample variance (S^2)?
  - Should we divide $\sum_{i=1}^{n}(X_i - \bar{X})^2$ by n or $n-1$?
- † Sampling distribution of S^2 (under Normal distribution)
- Some relationships among distributions (for future use)

† Evans and Rosenthal: theorem 4.6.6 (using theorem 4.6.2) and John A. Rice: Chap 6.3

Section 1

$MSE = Var + Bias^2$; topics from week 2 that we couldn't finish

Measuring quality of an estimator [E&R-page 322]

- Let $\psi(\theta)$ be any real valued function of θ
- Suppose, T is an estimator of $\psi(\theta)$
- The most commonly used measurement of accuracy of an estimator is *Mean Squared Error (MSE)*
- $MSE_\theta(T) = E_\theta[(T - \psi(\theta))^2]$
- The smaller the value of $MSE_\theta(T)$, the more concentrated the sampling distribution of T is about the value $\psi(\theta)$
- Since the true value of θ is unknown, often we evaluate the $MSE_\theta(T)$ at $\theta = \hat{\theta}$

More on MSE

$$MSE_\theta(T) = var_\theta(T) + (E_\theta(T) - \psi(\theta))^2$$

Proof...

Unbiasedness

- **Bias:** The bias of an estimator T of $\psi(\theta)$ is given by $E_\theta(T) - \psi(\theta)$

- **Unbiased estimator:** When the bias of an estimator is zero, it's called unbiased.
  - So T is unbiased estimator of $\psi(\theta)$ when $E_\theta(T) = \psi(\theta)$
  - In other words, T is unbiased if $\psi(\theta)$ is the mean of the sampling distribution of T.
  - Example: On slide 21(week-2), we have shown $E[\bar{X}] = \mu$. Therefore, sample mean is an unbiased estimator of the population mean.

Comments on MSE and Unbiasedness

- $MSE(T) = var(T) + (Bias(T))^2$
- For unbiased estimators, $MSE(T) = var(T)$
- If all the other properties (we haven't studied them yet) are similar, then an unbaised estimator is preferred over a biased estimator.
- In practice, often an biased estimator with lower variance is preferred over an unbiased estimator with really high variance $\implies$ we minimize MSE.

Section 2

Which formula to use for sample variance (S^2)

Let's start with Population variance (σ^2)

- **Definition of σ^2:**
  - $\sigma^2 = E[(X - \mu)^2]$ where $\mu = E[X]$
  - if we have equally likely N *data points* in our Population this is equivalent of

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)^2$$

- **In words:** it's the AVERAGE squared difference of each of the data points (X_i) from the mean (μ)
- We are not estimating anything here. We are calculating σ^2 based on the population.

Let's estimate σ^2 based on a sample of size=n

- When we are estimating based on the sample of size $= n$,
  - we replace μ by $\bar{X}$
  - So the numerator is $\sum_{i=1}^{n}(X_i - \bar{X})^2$
  - To get an estimator, should we divide it by n or $n-1$?
- The ans is: we can do both!
- They both can be used as an estimator of σ^2
- **Difference:** one of them is an unbiased and the other one is a biased estimator of σ^2.

Identity needed to check unbiasedness

- An identity that we need here (and will need in future)

$$\sum_i (X_i - \mu)^2 = \sum_i (X_i - \bar{X})^2 + n(\bar{X} - \mu)^2 \tag{1}$$

Proof...

- Re-writing it

$$\sum_i (X_i - \bar{X})^2 = \sum_i (X_i - \mu)^2 - n(\bar{X} - \mu)^2$$

Unbiased estimator of σ^2

- Taking Expectation on both sides,

$$
\begin{aligned}
E[\sum_i (X_i - \bar{X})^2] &= E[\sum_i (X_i - \mu)^2] - E[n(\bar{X} - \mu)^2] \\
&= ... \\
&= ... \\
&= (n-1)\sigma^2 \qquad (2)
\end{aligned}
$$

Unbiased estimator of σ^2 (cont...)

- Dividing both sides of eq-2 [slide ~~8~~ 13] by n $\implies$

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

So, $\frac{1}{n}\sum_i (X_i - \bar{X})^2$ is a *biased estimator* of σ^2.

- Dividing both sides of eq-2 [slide ~~8~~ 13] by $n-1$ $\implies$

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

So, $\frac{1}{n-1}\sum_i (X_i - \bar{X})^2$ is an *unbiased estimator* of σ^2.

Few comments on the choice of estimator for σ^2

- For Normal distribution, both Method of moments and Maximum likelihood estimation gives $\frac{1}{n}\sum_i (X_i - \bar{X})^2$ as an estimator of σ^2 (we did this last week)
- The fraction, $\frac{n-1}{n} \to 1$ as $n \to \infty$
- For large n, both estimators will produce similar estimate.
- In statistical literature, whenever we say *sample variance* we refer to the *unbiased* one.
- Hence, from now on (at least for this course),

sample variance,

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

Section 3

Sampling distribution of S^2 (under Normal distribution)

A well known theorem

- Suppose $X_1, X_2, ..., X_n \overset{iid}{\sim} N(\mu, \sigma^2)$
- $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

Then

- $\bar{X}$ and S^2 are independent. [slide ~~13-16~~ 18-21]
- $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(df=n-1)}$ [slide ~~17-18~~ 22-23]

Proving $\bar{X}$ and S^2 are independent

- The rigorous proof uses the geometric properties of a multivariate Normal distribution (which is beyond the scope of this course)
- John A. Rice gave a proof using the moment generating function (page 195-197).
- We will try a different way using theorem 4.6.2 of Evans and Rosenthal.

E&R theorem 4.6.2 (page-235)

- $X_1, X_2, ..., X_n \stackrel{iid}{\sim} N(\mu, \sigma^2)$
- U and V are two different linear combinations of the X_i's
- $cov[U, V] = 0$ if and only if U and V are independent.

NOTE:

- In general, zero covariance doesn't imply independence (we will give an example later).
- But for **bi-variate Normal** distribution,

$$\text{zero covariance} \implies \text{independence}$$

proof of this theorem is available on page 248 (uses two dimensional change of variables)

$\bar{X}$ is independent of $X_i - \bar{X}$

- Say $i = 1$
- $\bar{X} = \frac{1}{n}(X_1 + X_2 + ... + X_n)$
    - It's a linear combination of X_i's
- $X_1 - \bar{X} = (1 - \frac{1}{n})X_1 - \frac{1}{n}X_2 - ... - \frac{1}{n}X_n$
    - It's also a linear combination of X_i's
- $cov[\bar{X}, X_1 - \bar{X}] = 0$ (proof)
- Hence using E&R theorem 4.6.2, $\quad \bar{X} \perp\!\!\!\perp X_1 - \bar{X}$
- we can show this for all the values of $i = 1, 2, ...n$
- Hence,

$$\bar{X} \perp\!\!\!\perp X_i - \bar{X} \text{ for } i = 1, 2, ...n$$

$\bar{X}$ is independent of S^2

- From last slide, $\bar{X}$ is independent of all the $(X_i - \bar{X})$'s
- $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$ which is a function of all the $(X_i - \bar{X})$'s
- Hence, $\bar{X}$ is independent of S^2

Proving $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(df=n-1)}$

- Dividing eq-1 [slide-~~7~~ 12] by σ^2 we get

$$\frac{\sum_i (X_i - \mu)^2}{\sigma^2} = \frac{\sum_i (X_i - \bar{X})^2}{\sigma^2} + \frac{n(\bar{X} - \mu)^2}{\sigma^2}$$
$$\implies \sum_i \left(\frac{X_i - \mu}{\sigma}\right)^2 = \frac{(n-1)S^2}{\sigma^2} + \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right)^2$$

- $\sum_i \left(\frac{X_i - \mu}{\sigma}\right)^2 \sim ?$
- $\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right)^2 \sim ?$

Proving $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(df=n-1)}$ (cont...)

- Using moment generating function(MGF) and independence of $\bar{X}$ and S^2

$$\text{MGF of } \chi^2_{(df=n)} = [\text{MGF of } \frac{(n-1)S^2}{\sigma^2}] * [\text{MGF of } \chi^2_{(df=1)}]$$

$$\implies (1-2t)^{-n/2} = [\text{MGF of } \frac{(n-1)S^2}{\sigma^2}] * (1-2t)^{-1/2}$$

$$\implies [\text{MGF of } \frac{(n-1)S^2}{\sigma^2}] = (1-2t)^{-(n-1)/2}$$

which is the MGF of a $\chi^2_{df=(n-1)}$

- Hence,

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(df=n-1)}$$

End of theorem...

Unbiasedness of S^2 using the Chi-sq distribution

- Recall: The mean of a Chi-sq distribution is it's degrees of freedom, *df* (in other words it's parameter).

Then,

$$E\left[\frac{(n-1)S^2}{\sigma^2}\right] = (n-1)$$
$$\implies E[S^2] = \sigma^2$$

Note:

- This proves S^2 is an unbiased estimator for σ^2 under Normal distribution.
- Slide [~~7-9~~ 12-14] proves it under any arbitrary distribution with the assumption that X_i's are $i.i.d.$ and μ, σ^2 exists.

An example of [cov=0 $\nRightarrow$ independence]

- Say we have $X \sim N(0,1)$
- $Y = X^2$
- Clearly X and Y are dependent.
- But their covariance is zero!

$$\begin{aligned} cov[X,Y] &= E[XY] - E[X]E[Y] \\ &= E[X.X^2] - 0.E[X^2] \\ &= E[X^3] \\ &= 0 \end{aligned}$$

Note: Odd moments (e.g. $E[X]$, $E[X^3]$, $E[X^5]$...) of any distribution which is symmetric around zero $= 0$

Section 4

Some relationships among distributions (for future use)

$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{(df=n-1)}$

Recall from week 1: Z=standard Normal, U=chi-sq with df=m and $Z \perp\!\!\!\perp U$. Then $\frac{Z}{\sqrt{U/m}} \sim t_{(df=m)}$

- (Week-2) Sampling distribution of $\bar{X}$ under Normal distribution $\implies \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$
- This week we proved $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(df=n-1)}$
- Also $\bar{X} \perp\!\!\!\perp S^2$
- Then,

$$\implies \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{(df=n-1)}$$

Note: we will use this when we do the interval estimation.

Few comments on $\chi^2_{(m)}$ distribution

- $\chi^2_{(m)}$ is a special case of *Gamma* dist. $[G(m/2, 1/2)]$
- $\chi^2_{(m)}/m = \frac{1}{m}(Z_1^2 + Z_2^2 + ...Z_m^2)$ where $Z_1, Z_2, ...Z_m$ are independent $N(0,1)$ variables.
  Then by LLN,
  $$\frac{1}{m}(Z_1^2 + Z_2^2 + ...Z_m^2) \xrightarrow{P} E[Z_i^2] = 1$$
  Therefore,
  $$\chi^2_{(m)}/m \xrightarrow{P} 1$$

Homework (Non-credit)

Assuming $X_1, X_2, ...X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ and using the properties of χ^2 distribution, calculate the MSE of S^2 as an estimator of σ^2.

Assuming $X_1, X_2, ...X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ and using the properties of χ^2 distribution, calculate the MSE of $\hat{\sigma}^2$ as an estimator of σ^2, where $\hat{\sigma}^2 = \frac{(n-1)S^2}{n}$