1/13/22

Saturday, January 15, 2022 2:55 AM

Sampling Distributions : $[\chi^2, T, F]$

① If $Z_1, Z_2, \cdots, Z_n$ are i.i.d. $\sim N(0,1)$ then $W = Z_1^2 + Z_2^2 + \cdots + Z_n^2$ is χ^2 with n (df) degrees of freedom

$\boxed{\text{HW-Hint: } (\text{qchisquare}(df, 'x'))}$

$T = \frac{Z}{\frac{\sqrt{W}}{\sqrt{n}}} = \frac{Z\sqrt{n}}{\sqrt{W}}$

② If $Z \sim N(0,1)$, $W \sim \chi^2_n$, Then $T = \frac{Z}{\sqrt{W/n}}$, given Z & W are independent.

We call this the "student's t-distribution" ← given 'n df'!

③ If $W_1 \sim \chi^2_{n_1}$ & $W_2 \sim \chi^2_{n_2}$; $F = \frac{W_1/n_1}{W_2/n_2}$ or $\frac{W_1}{n_1} \cdot \frac{n_2}{W_2}$ is F_{n_1, n_2}

Think of $Z_1, Z_2, \cdots Z_n$ as a random sample drawn from a standard normal, $N(0,1)$, population ...

HW Hint $P\left(a < \frac{Z_1^2 + Z_2^2}{Z_3^2 + Z_4^2 + Z_5^2} < b\right) = .90$ Can be transformed to an F RV Ex. $\frac{\frac{W_2}{2}}{\frac{W_3}{3}}$

Statistic := A number derived from data.

data being ~ Randomly Generated Values. (Sampled RVs).

Thus A statistic is a number that itself is a function of data.

Thus A statistic is also Randomly Generated (RV).

Ex. Suppose $X_1, X_2, \cdots X_{100}$ are iid & $N(\mu, \sigma^2)$, with both μ & σ^2 unknown

What we may want to do is solve, or guess μ.

here $\mu \approx \bar{X} = \dfrac{X_1 + X_2 + \cdots + X_{100}}{100}$ * $\bar{X}$ is a statistic, the sample mean.

μ is a parameter, the population mean.

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

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Assuming $X_1, X_2, \cdots X_n$ are i.i.d. R.V.'s

Sample mean : $\bar{X} = \dfrac{\sum_{i=1}^{n} x_i}{n}$ (usuall used as an estimator of μ).

Sample Variance : $S^2 = \dfrac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}$ (usuall used as as estimator of σ^2)

* if we KNOW μ, but not σ^2, we can use $\dfrac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}$ to estimate σ^2.

Note here: $\mathbb{E}\left[\frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2\right] = \mathbb{E}\left[(X_1 - \mu)^2\right] = \sigma^2$

Otherwise : $\mathbb{E}\left[\sum_{i=1}^{n} (X_i - \bar{X})^2\right]$ is usually smaller

Proof: $\sum_{i=1}^{n} (X_i - \bar{X}) = \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \bar{X} = \sum_{i=1}^{n} X_i - n\bar{X} = 0$

$$\boxed{\sum_{i=1}^{n} (X_i - \mu)^2} = \sum_{i=1}^{n} (X_i - \bar{X} + \bar{X} - \mu)^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 + \overbrace{2 \underbrace{\sum_{i=1}^{n} (X_i - \bar{X})}_{0} \underbrace{(\bar{X} - \mu)}_{\text{constant}}}^{0} + \sum_{i=1}^{n} (\bar{X} - \mu)^2$$

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

$$= \sum_{i=1} (x_i - \bar{x})^2 \quad \cdots \sum_{i=1} \quad \cdots \quad \sum_{i=1}$$

So $\sum_{i=1}^{n} (x_i - \mu)^2 \geq \sum_{i=1}^{n} (x_i - \bar{x})^2$ as $n(\bar{x} - \mu)^2 \geq 0$

Following, $\mathbb{E}\left[\sum_{i=1}^{n} (x_i - \mu)^2\right] = \mathbb{E}\left[\sum_{i=1}^{n} (x_i - \bar{x})^2\right] - n(\bar{x} - \mu)^2$

$$n\sigma^2 = \mathbb{E}\left[\sum_{i=1}^{n} (x_i - \bar{x})^2\right] - \frac{\cancel{n}\,\sigma^2}{\cancel{n}}$$

$$\therefore \mathbb{E}\left[\sum_{i=1}^{n} (x_i - \bar{x})^2\right] = \sigma^2 (n-1)$$

Facts: That will help with last hw problem *

- Suppose $X_1, \ldots, X_n$ are iid $N(\mu, \sigma^2)$

Then ① $\bar{X} = \dfrac{X_1 + \ldots + X_n}{n} \sim N\left(\mu = \mu, \sigma^2 = \dfrac{\sigma^2}{n}\right)$

② $\underbrace{\dfrac{(n-1)S^2}{\sigma^2}}_{} = \dfrac{(X_1 - \bar{X})^2}{\sigma^2} + \cdots + \dfrac{(X_n - \bar{X})^2}{\sigma^2} = \dfrac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{\sigma^2}$

↳ Turns out to be χ^2 with $n-1$ 'df'

③ $\bar{X}$ & S, are independant.

Now Recall $T = \dfrac{Z}{\sqrt{W/n}}$ ← where Z is standard Normal & W is χ^2 with n 'df'

④ $\dfrac{\bar{X} - \mu}{S/\sqrt{n}} \sim T$ with $n-1$ 'df' $\left[\dfrac{Z}{\chi^2}\right]$