Introduction to mathematical Statistics **Lecture 4**

Sampling from the Normal Population

*__Example__: We wish to estimate the distribution of heights of adult US male. It is believed that the height of adult US male follows a normal distribution $N(\mu, \sigma^2)$

Def. Simple random sample: A sample in which every subject in the population has the same chance to be selected.

X : The Random Variable denote the height of a adult male we will choose randomly from the population

So $X \sim N(\mu, \sigma^2)$: the distribution of a randomly selected subject is the population distribution.

Theorem 1 Sampling from the normal population

Let $X_1, X_2, ..., X_n \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$, where i.i.d stands for independent and identically distributed

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

2. $\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(\mathrm{X}_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}$ (Chi Square distribution with $(n-1)$ degrees of freedom),

*Reminder: The Sample variance S^2 is defined as: $S^2 = \frac{\sum_{i=1}^{n}(\mathrm{X}_i - \bar{X})^2}{n-1}$

*Def 1: The Chi-square distribution is a special gamma distribution (*** Please find out which one it is.)

*Def 2: Let $Z_1, Z_2, ..., Z_k \overset{\text{i.i.d.}}{\sim} N(0,1)$,

Then $W = \sum_{i=1}^{k} Z_i^2 \sim \chi^2_k$

3. $Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ （Z score of $\bar{X}$ ） $\sim N(\mu, \sigma^2)$

4. $\bar{X}$ & S^2 are independent.

5. $T = \dfrac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \sim t_{n-1}$ (t distribution with (n-1) degrees of freedom)

***Def. of t-distribution** (student's t-distribution, first introduced by William Sealy Gosset)

Let $Z \sim N(0,1)$, $W \sim \chi_k^2$, where Z & W are independent.

Then, $T = \dfrac{Z}{\sqrt{W/k}} \sim t_k$

***Proof of 5**

We know that $Z = \dfrac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$ and $W = \dfrac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$

Furthermore, since $\bar{X}$ & S^2 are independent, thus Z & W are independent.

Therefore by the definition of t-distribution, $\dfrac{Z}{\sqrt{W/n-1}} \sim t_{n-1}$

***We will prove #2 & #4 for the special case of n=2. In that case, we have:**

Independent $\begin{cases} X_1 \sim N(\mu,\sigma^2) \\ X_2 \sim N(\mu,\sigma^2) \end{cases}$

$$S^2 = \frac{(X_1-\bar{X})^2+(X_2-\bar{X})^2}{1} = \frac{1}{2}(X_1-X_2)^2$$

*#2

$$W = \frac{(n-1)S^2}{\sigma^2} = \frac{S^2}{\sigma^2} = \frac{(X_1-X_2)^2}{2\sigma^2}$$

$Z = \dfrac{X_1-X_2}{\sqrt{2}\sigma} \sim \mathrm{N}(0,1)$ ***You can prove this using the pivotal quantity method**

$\therefore W = Z^2 \sim \chi_1^2$ (using the 2nd Definition of Chi Square Distribution.)

***#4**

If we can show X_1+X_2 and X_1-X_2 are independent then we have proven that $\bar{X}$ and S^2 are independent.

Approach 1: p.d.f. $f_{X_1+X_2,X_1-X_2}(\mu,\nu)=f_{X_1+X_2}(\mu)f_{X_1-X_2}(\nu)$

Approach2: m.g.f. $M_{X_1+X_2,X_1-X_2}(t_1,t_2)=M_{X_1+X_2}(t_1)M_{X_1-X_2}(t_2)$

The joint m.g.f. of X and Y is defined as:

$M_{X,Y}(t_1,t_2)=E(e^{t_1X+t_2Y})$

Note: We have done this already.

Additional Questions and Solutions

Question 1. Prove $E(S^2)=\sigma^2$ for any distribution/population.

Solution

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

$$=\sum_{i=1}^{n}\left(X_i-\mu+\mu-\bar{X}\right)^2=\sum_{i=1}^{n}\left[\left(X_i-\mu\right)^2-2\left(X_i-\mu\right)\left(\bar{X}-\mu\right)+\left(\bar{X}-\mu\right)^2\right]$$

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

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

$$(n-1)E\left(S^2\right)=E\left[\sum_{i=1}^{n}\left(X_i-\mu\right)^2\right]-nE\left[\left(\bar{X}-\mu\right)^2\right]$$

$$=n\sigma^2-nVar(\bar{X})$$

$$\rightarrow(n-1)\sigma^2$$

$\therefore E(S^2) = \sigma^2$

Question 2.

- Please point out a chi-square random variable with k degrees of freedom corresponds to which particular gamma distribution.
- Please write down the pdf, mgf, mean, and variance of a general gamma distribution and of a chi-square random variable with k degrees of freedom.

Solution

Let $W \sim \chi_k^2$. W is indeed a special random variable.

Gamma distribution

$X \sim gamma(\alpha, \beta)$ (Some books use $\lambda = \frac{1}{\beta}$)

if $f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}},\ 0 \prec x \prec \infty$

or

$f(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x},\ 0 \prec x \prec \infty$ if r is a non-negative integer, then $\Gamma(r) = (r-1)!$

$$1 = \int_{-\infty}^{\infty} f(x)dx = \int_0^\infty \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} dx$$

$$\Gamma(r) = \lambda^r \int_0^\infty x^{r-1} e^{-\lambda x} dx$$

$$M_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-r} = \left(\frac{\lambda - t}{\lambda}\right)^{-r}$$

$$\Rightarrow M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^r$$

$$E(X) = \frac{r}{\lambda}$$

$$Var(X) = \frac{r}{\lambda^2}$$

Special case : when $r = \frac{k}{2},\ \lambda = \frac{1}{2} \quad \Rightarrow \quad X \sim \chi_k^2$

Special case : when $r = 1 \quad \Rightarrow \quad X \sim \exp(\lambda)$

Review

$X \sim \exp(\lambda)$

p.d.f. $f(x) = \lambda e^{-\lambda x},\ x \succ 0$

m.g.f. $M_X(t) = \frac{\lambda}{\lambda - t}$

e.g. Let $X_i \overset{i.i.d.}{\sim} \exp(\lambda)$, $i = 1, \cdots, n$. What is the distribution of $\sum_{i=1}^{n} X_i$?

Solution

$$M_{\sum X_i}(t) = \prod_{i=1}^{n} M_{X_i}(t) = \left(\frac{\lambda}{\lambda - t} \right)^n$$

$$\therefore \sum X_i \sim gamma(r = n, \lambda)$$

e.g. Let $W \sim \chi_k^2$. What is the mgf of W ?

Solution

$$M_W(t) = \left(\frac{\frac{1}{2}}{\frac{1}{2} - t} \right)^{\frac{k}{2}}$$

$$\Rightarrow M_W(t) = \left(\frac{1}{1 - 2t} \right)^{\frac{k}{2}}$$

e.g. Let $W_1 \sim \chi_{k_1}^2$, $W_2 \sim \chi_{k_2}^2$, and W_1 and W_2 are independent. What is the distribution of $W_1 + W_2$?

Solution

$$M_{W_1+W_2}(t) = M_{W_1}(t)\cdot M_{W_2}(t) = \left(\frac{1}{1-2t}\right)^{\frac{k_1}{2}}\left(\frac{1}{1-2t}\right)^{\frac{k_2}{2}} \quad = \left(\frac{1}{1-2t}\right)^{\frac{k_1+k_2}{2}}$$

$$\Rightarrow W_1 + W_2 \sim \chi^2_{k_1+k_2}$$

Theorem. If the joint domain of W and V are rectangular and free of any parameter, then X_1 and X_2 are independent iff

1. $f(w,v) = f(w)\cdot f(v)$
2. $M_{(W,V)}(t_1,t_2) = M_W(t_1)\cdot M_V(t_2)$

[**Definition:** $M_{(W,V)}(t_1,t_2) = E(e^{t_1W+t_2V}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t_1w+t_2v} f(w,v)dwdv$]

Question 3. Let $X_1 \sim \exp(\lambda)$, $X_2 \sim \exp(\lambda)$. X_1 and X_2 are independent.

- What is the m.g.f. of $X_1 + X_2$?
- What is the distribution of $X_1 + X_2$?

Solution $gamma(r = 2, \lambda)$