Sampling from the Normal Distribution.

Theorem 5.3.1:

Let $X_1, X_2, \ldots X_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution. and let $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ and $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$ then

(a) $\bar{X}$ and S^2 are independent random variables

(b) $\bar{X}$ has a $N(\mu, \frac{\sigma^2}{n})$ distribution

(c) $(n-1) \frac{S^2}{\sigma^2}$ has a chi-square

Proof:

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

$$= \frac{1}{n-1} \left((x_1 - \bar{x}) + \sum_{i=2}^{n} (x_i - \bar{x})^2 \right)$$

$$= \frac{1}{n-1} \left(x_1 - n\bar{x} + (n-1)\bar{x} + \sum_{i=2}^{n} (x_i - \bar{x})^2 \right)$$

$$= \frac{1}{n-1} \left(x_2 + x_3 + \dots + x_n + (n-1)\bar{x} + \sum_{i=2}^{n} (x_i - \bar{x})^2 \right)$$

$$= \frac{1}{n-1} \left(\sum_{i=2}^{n} (x_i - \bar{x}) + \sum_{i=2}^{n} (x_i - \bar{x})^2 \right)$$

$$S^2 (x_2 - \bar{x}, x_3 - \bar{x}, \dots, x_n - \bar{x}).$$

we need show S^2 is independent of $\bar{x}$

$$\Rightarrow \quad f_{X_1, X_2, \ldots X_n}(x_1, x_2, \ldots x_n) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\sum_{i=1}^{n} x_i^2}$$

$-\infty < x_i < \infty$

<u>Transformation:</u>

$$\begin{aligned} y_1 &= \bar{x} \\ y_2 &= x_2 - \bar{x} \\ y_3 &= x_3 - \bar{x} \\ &\vdots \\ y_n &= x_n - \bar{x} \end{aligned} \quad \Longrightarrow \quad \begin{aligned} \bar{x} &= y_1 \\ x_2 &= y_2 + y_1 \\ x_3 &= y_3 + y_1 \\ &\vdots \\ x_n &= y_n + y_1 \end{aligned}$$

$$\Downarrow$$

$$x_1 = n(\bar{x}) - x_2 - x_3 \ldots - x_n$$

$$x_1 = -(x_2 - \bar{x}) - (x_3 - \bar{x}) \ldots - (x_n - \bar{x}) + \bar{x}$$

$$x_1 = y_1 - y_2 - y_3 \ldots - y_n$$

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & & & \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix} = \begin{vmatrix} 1 & -1 & -1 & \cdots & -1 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & & & & \\ 1 & 0 & 0 & \cdots & 1 \end{vmatrix}$$

$$= n$$

$$f(y_1, y_2, \dots y_n) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}(x_1^2 + x_2^2 + \dots + x_n^2)} \cdot n$$

$$= \frac{n}{(2\pi)^{n/2}} e^{-\frac{1}{2}\left(\left(y_1 - \sum_{i=2}^{n} y_i\right)^2 + (y_2 + y_1)^2 + (y_3 + y_1)^2 + \dots + (y_n + y_1)^2\right)}$$

$$(y_1 - y_2 - y_3 \dots - y_n)^2 = y_1^2 + \cancel{y_2^2} + \cancel{y_3^2} + \dots + \cancel{y_n^2}$$

$$- 2 \sum_{1 \le i \le j \le n} y_i y_j$$

$$+ \qquad\qquad +$$

$$(y_2 + y_1)^2 + (y_3 + y_1)^2 + \dots + (y_n + y_1)^2 = (n-1) y_1^2 + \cancel{y_2^2} + \cancel{y_3^2} + \dots + \cancel{y_n^2}$$

$$+ 2 \sum_{i=2}^{n} y_i y_1$$

$$\Rightarrow \quad n y_1^2 + 2 \sum_{2 \le i < j < n} y_i y_j$$

$$\Rightarrow f(y_1, y_2, \dots y_n) = \left(\frac{n}{2\pi}\right)^{1/2} \cdot e^{-\frac{n y_1^2}{2}}$$

$$\times \frac{n^{1/2}}{(2\pi)^{(n-1)/2}} \cdot e^{-\frac{1}{2}\left(\sum_{i=2}^{n} y_i^2 + \left(\sum_{i=2}^{n} y_i\right)^2\right)} \Big]$$

$\Rightarrow$ X_1 is indendent of $Y_1, Y_2 \dots Y_n$

S^2 has the distribution of Chi-square

$$f(x) = \frac{1}{\Gamma(p/2)\, 2^{p/2}} x^{\frac{p}{2}-1} \cdot e^{-\frac{x}{p}}$$

$$X_p^2 \sim \text{Chi square}(p)$$

$$X_p^2 \sim \text{gamma}\left(\frac{p}{2}, 2\right)$$

Lemma 5.3.2:

FActs about Chi squared random variables.

(a) if $Z \sim n(0,1)$, then $Z^2 \sim X_1^2$;
that is, the square of a standard normal r.v is a chi squared r.v

(b) if $X_1, X_2, \ldots X_n$ are independent and $X_i \sim \chi^2_{p}$, then

$$X_1 + \cdots + X_n \sim \chi^2_{p_1+p_2+\cdots+p_n}$$

That is, independent Chi squared variable add to a chi squared variable, and the d.o.f also add

Proof:

(a) $$Z \sim n(0,1)$$

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

$$Y = Z^2$$

$A_1 = (-\infty, 0)$ $\quad Y = Z^2 \quad Z = -\sqrt{y}$

$A_2 = (0, \infty)$ $\quad Y = Z^2 \quad Z = \sqrt{y}$

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y}{2}} \cdot \frac{1}{2\sqrt{y}} \cdot 2$$

$$= \frac{1}{\sqrt{2\pi}} y^{-1/2} \cdot e^{-\frac{y}{2}}$$

$$X_1^2 \sim \text{Chisquer}(1)$$

$$f_{X_1^2}(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{x}} \cdot e^{-\frac{x}{2}} \quad 0 < y < \infty$$

(b)

MGf of $X \sim \text{gamma}(\alpha, \beta)$

$$M_x(t) = \left(\frac{1}{1-\beta t}\right)^{\alpha}$$

using MGF of sum of r.v

Assume $X_1, X_2, \ldots X_n$ independent r.v

with $X_i \sim \text{gamma}(\alpha, \beta)$

$$Z = X_1 + X_2 + \ldots + X_n$$

$$M_Z(t) = \prod_{i=1}^{n} M_{X_i}(t)$$

$$= \prod_{i=1}^{n} \left(\frac{1}{1-\beta t} \right)^{\alpha_i}$$

$$= \left(\frac{1}{1-\beta t} \right)^{\alpha_1+\alpha_2+\cdots+\alpha_n}$$

$Z \sim \text{gamma}(\alpha_1+\alpha_2+\cdots+\alpha_n, \beta)$

$\Rightarrow$ if $X_i \sim \text{gamma}(\alpha_i, \beta)$

then $X_1+X_2+\cdots+X_n \sim \text{gamma}(\alpha_1+\alpha_2+\cdots+\alpha_n, \beta)$

Using this, we have

$X_1, X_2, \ldots X_n \sim$ Chisquare independent

when $X_i \sim \chi^2_{p_i} \sim \text{gamma}(p_i/2, 2)$

$\Rightarrow X_1+X_2+\cdots X_n = \text{gamma}\left(\frac{p_1+p_2+\cdots+p_n}{2}, 2\right)$

$\Rightarrow X_1+X_2+\cdots+X_n \sim \chi^2_{(p_1+p_2+\cdots p_n)}$

Proof of Theorem 5.3.1 (c)

$$(n-1)S_n^2 = (n-2)S_{n-1}^2 + \left(\frac{n-1}{n}\right)(X_n - \overline{X}_{n-1})^2$$

Proof of the Identity

$$(n-1)S_n^2 = \sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

$$\overline{X}_n = \frac{X_1 + X_2 + \cdots + X_{n-1} + X_n}{n}$$

$$\frac{n\overline{X}_n - X_n}{n-1} = \frac{X_1 + X_2 + \cdots + X_{n-1}}{n-1}$$

$$\Rightarrow \frac{n}{n-1}\overline{X}_n - \frac{X_n}{n-1} = \overline{X}_{n-1}$$

$$\Rightarrow \overline{X}_n = \frac{n-1}{n}\overline{X}_{n-1} + \frac{X_n}{n}$$

$$\Rightarrow (n-1) S_n^2 = \sum_{i=1}^{n} \left(x_i - \overline{x}_n\right)^2$$

$$= \sum_{i=1}^{n} \left(x_i - \frac{n-1}{n} \overline{x}_{n-1} - \frac{x_n}{n} \right)^2$$

$$= \sum_{i=1}^{n} \left(x_i - \frac{n-1}{n} \overline{x}_{n-1} - \frac{x_n}{n} \right)^2$$

$$= \sum_{i=1}^{n} \left(x_i - \overline{x}_{n-1} + \frac{\overline{x}_{n-1} - x_n}{n} \right)^2$$

$$= \sum_{i=1}^{n-1} \left(x_i - \overline{x}_{n-1} + \frac{\overline{x}_{n-1} - x_n}{n} \right)^2$$

$$+ \left(x_n - \overline{x}_{n-1} + \frac{\overline{x}_{n-1} - x_n}{n} \right)^2$$

$$= \sum_{i=1}^{n-1} \left[\left(x_i - \overline{x}_{n-1}\right)^2 + \left(\frac{\overline{x}_{n-1} - x_n}{n} \right)^2 + 2\left(x_i - \overline{x}_{n-1}\right) \left(\frac{\overline{x}_{n-1} - x_n}{n} \right) \right]$$

$$+ \left(1 - \frac{1}{n}\right)^2 \left(x_n - \overline{x}_{n-1}\right)^2$$

$$= \sum_{i=1}^{n-1} \left(x_i - \bar{x}_{n-1}\right) + \sum_{i=1}^{n-1} \left(\frac{\bar{x}_{n-1} - x_n}{n} \right)^2$$

$$+ \; 0 \; + \left(1 - \frac{1}{n}\right)^2 \left(x_n - \bar{x}_{n-1}\right)^2$$

$$= (n-2)\, S_{n-1}^2 + \left(\frac{n-1}{n^2}\right)\left(x_n - \bar{x}_{n-1}\right)^2 + \left(1 - \frac{1}{n}\right)^2 \left(x_n - \bar{x}_{n-1}\right)^2$$

$$= (n-2)\, S_{n-1}^2 + \left(x_n - \bar{x}_{n-1}\right)^2 \left[\frac{1}{n} - \frac{\cancel{1}}{n^2} + 1 + \frac{\cancel{1}}{n^2} - \frac{2}{n} \right]$$

$$= (n-2)\, S_{n-1}^2 + \left(\frac{n-1}{n}\right)\left(x_n - \bar{x}_{n-1}\right)^2$$

Therefore

$$(n-1)\, S_n^2 = (n-2)\, S_{n-1}^2 + \left(\frac{n-1}{n}\right)\left(x_n - \bar{x}_{n-1}\right)^2$$

Proof By Induction

for $n=2$

$$S_2^2 = \frac{1}{2}(X_2 - \bar{X}_1)^2$$

$$\Rightarrow S_2^2 = \frac{1}{2}(X_n - X_1)^2$$

$$\Rightarrow S_2^2 \sim \text{chi square}(1)$$

for $n=k$

if $(k-1)S_k^2 \sim \chi^2_{k-1}$

then

$$k S_{k+1}^2 = (k-1)S_k^2 + \frac{k}{k+1}(X_{k+1} - \bar{X}_k)^2$$

if we can show that

$$\frac{k}{k+1}(X_{k+1} - \bar{X}_k)^2 \sim \chi_1^2$$

independent of S_k^2

it will be easy

we have proved $\overline{X}_k$, S_k^2 are independent

then $(X_{k+1}, \overline{X}_k)$, S_k^2 are independent.

$\Rightarrow$ $(X_{k+1} - \overline{X}_k)$, S_k^2 are independent.

$$X_{k+1} - \overline{X}_k \sim N(0,1) - N\left(0, \frac{1}{k}\right)$$

$$\operatorname{var}\left((X_{k+1} - \overline{X}_k)^2\right) = \frac{k+1}{k}$$

$$\Rightarrow \left(\frac{k}{k+1}\right)(X_{k+1} - \overline{X}_k)^2 \sim \chi_1^2$$

therefore
$$\left.\begin{array}{l} (k-1) S_k^2 \sim \chi^2_{k-1} \\ \left(\frac{k}{k+1}\right)(X_{k+1} - \overline{X}_k)^2 \sim \chi^2_1 \end{array}\right\} \text{independ.}$$

$\Rightarrow$ summation of these two results to χ^2_k