556: MATHEMATICAL STATISTICS I

SAMPLING DISTRIBUTION FOR NORMAL SAMPLES

THEOREM
If $X_1, ..., X_n$ is a random sample from a normal distribution, say $X_i \sim N(\mu, \sigma^2)$, then

(a) $\overline{X} \sim N(\mu, \sigma^2/n)$

(b) $\overline{X}$ is independent of $\{X_i - \overline{X}, i = 1, ..., n\}$, and $\overline{X}$ and s^2 are independent random variables

(c) The random variable

$$\frac{(n-1)s^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

has a **chi-squared distribution** with $n-1$ degrees of freedom.

Proof. (a) See earlier proof using mgfs.

(b) The joint pdf $X_1, ..., X_n$ is the normal density

$$f_{X_1,...,X_n}(x_1, ..., x_n) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right\}$$

Consider the multivariate transformation to $Y_1, ..., Y_n$ where

$$\left.\begin{aligned} Y_1 &= \overline{X} \\ Y_i &= X_i - \overline{X},\ i = 2, ..., n \end{aligned}\right\} \Longleftrightarrow \left\{\begin{aligned} X_1 &= Y_1 - \sum_{i=2}^{n} Y_i \\ X_i &= Y_i + Y_1,\ \ i = 2, ..., n \end{aligned}\right.$$

Thus $\underset{\sim}{Y} = A\underset{\sim}{X}$, or equivalently $\underset{\sim}{X} = A^{-1}\underset{\sim}{Y}$, where A is the $n \times n$ matrix with (i,j)th element

$$[A]_{ij} = \begin{cases} 1 - 1/n & i = j \text{ and } i \neq 1, \\ 1/n & i = 1 \\ -1/n & \text{otherwise} \end{cases}$$

that is, we have a linear transformation, and the Jacobian of the transformation does not depend on any y. Note that

$$\sum_{i=1}^{n}(x_i - \mu)^2 = \sum_{i=1}^{n}(x_i - \overline{x} + \overline{x} - \mu)^2 = \sum_{i=1}^{n}(x_i - \overline{x})^2 + n(\overline{x} - \mu)^2$$

where $\overline{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$. Thus the joint pdf of $X_1, ..., X_n$ is, in scalar form

$$f_{X_1,..,X_n}(x_1, .., x_n) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(x_i - \overline{x})^2 + n(\overline{x} - \mu)^2\right]\right\}.$$

Now

$$x_1 - \overline{x} = -\sum_{i=2}^{n}(x_i - \overline{x}) = -\sum_{i=2}^{n} y_i$$

and so

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

The Jacobian of the transformation is n, so the joint density of $Y_1,...,Y_n$ is given by

$$\begin{aligned}
f_{Y_1,..,Y_n}(y_1,..,y_n) &= n\left(\frac{1}{2\pi\sigma^2}\right)^{n/2}\exp\left\{-\frac{1}{2\sigma^2}\left[\left(-\sum_{i=2}^{n}y_i\right)^2+\sum_{i=2}^{n}y_i^2+n\left(y_1-\mu\right)^2\right]\right\} \\
&= n\left(\frac{1}{2\pi\sigma^2}\right)^{n/2}\exp\left\{-\frac{1}{2\sigma^2}\left[\left(-\sum_{i=2}^{n}y_i\right)^2+\sum_{i=2}^{n}y_i^2\right]\right\}\times\exp\left\{-\frac{n}{2\sigma^2}\left(y_1-\mu\right)^2\right\}
\end{aligned}$$

Hence

$$f_{Y_1,..,Y_n}(y_1,..,y_n) = f_{Y_2,..,Y_n}(y_2,..,y_n)f_{Y_1}(y_1)$$

and therefore Y_1 is independent of $Y_2,...,Y_n$. Hence $\overline{X}$ is **independent** of the random variables terms $\left\{Y_i=X_i-\overline{X}, i=2,...,n\right\}$. Finally, $\overline{X}$ is also independent of $X_1-\overline{X}$ as

$$X_1-\overline{X}=-\sum_{i=2}^{n}\left(X_i-\overline{X}\right)$$

and s^2 is a function only of $\left\{X_i-\overline{X},\ i=1,...,n\right\}$. As $\overline{X}$ is independent of these variables, $\overline{X}$ and s^2 are also independent.

(c) The random variables that appear as sums of squares terms that joint pdf are

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

or $V_1=V_2+V_3$, say. Now, $X_i\sim N(\mu,\sigma^2)$, so therefore

$$\frac{(X_i-\mu)^2}{\sigma^2}\sim N(0,1) \quad\Longrightarrow\quad \frac{(X_i-\mu)^2}{\sigma^2}\sim\chi_1^2\equiv Ga\left(\frac{1}{2},\frac{1}{2}\right) \quad\Longrightarrow\quad \frac{\sum_{i=1}^{n}\left(X_i-\mu\right)^2}{\sigma^2}=V_1\sim\chi_n^2$$

as the X_is are independent, and the sum of n independent $Ga(1/2,1/2)$ variables has a $Ga(n/2,1/2)$ distribution. Similarly, as $\overline{X}\sim N(\mu,\sigma^2/n)$, $V_3\sim\chi_1^2$ By part (b), V_2 and V_3 are independent, and so the mgfs of V_1, V_2 and V_3 are related by

$$M_{V_1}(t)=M_{V_2}(t)M_{V_3}(t)\Longrightarrow M_{V_2}(t)=\frac{M_{V_1}(t)}{M_{V_3}(t)}$$

As V_1 and V_3 are Gamma random variables, M_{V_1} and M_{V_3} are given by

$$M_{V_1}(t)=\left(\frac{1/2}{1/2-t}\right)^{n/2}, M_{V_3}(t)=\left(\frac{1/2}{1/2-t}\right)^{1/2}\Longrightarrow M_{V_2}(t)=\left(\frac{1/2}{1/2-t}\right)^{(n-1)/2}$$

which is also the mgf of a Gamma random variable, and hence

$$V_2=\frac{(n-1)s^2}{\sigma^2}\sim\chi_{n-1}^2$$

and the result follows. ∎