Math 562 Mathematical Statistics

Lecture Notes on $\overline{X}$ and S^2

<u>Theorem</u> Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from $N(\mu, \sigma^2)$, $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$. Then

1) $\overline{X}$ and S^2 are independent;

2) $\frac{(n-1)}{\sigma^2} S^2 = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{\sigma^2} \sim \chi^2(n-1)$

<u>Proof</u>. 1) We first show that $\overline{X}$ and $X_i - \overline{X}$ are independent, $1 \le i \le n$.

Observe that
$$\sum_{i=1}^{n} \frac{(x_i - \mu)^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(x_i - \bar{x} + \bar{x} - \mu)^2}{\sigma^2}$$
$$= \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{\sigma^2} + \sum_{i=1}^{n} \frac{(\bar{x} - \mu)^2}{\sigma^2} + 2\sum_{i=1}^{n} \frac{(x_i - \bar{x})(\bar{x} - \mu)}{\sigma^2}$$
$$= \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{\sigma^2} + \sum_{i=1}^{n} \frac{(\bar{x} - \mu)^2}{\sigma^2} + 2\,\frac{\bar{x} - \mu}{\sigma^2}\left[\sum_{i=1}^{n} x_i - n\bar{x}\right]$$
$$= \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{\sigma^2} + \frac{n(\bar{x} - \mu)^2}{\sigma^2}$$

The joint density function of $(X_1, X_2, \cdots, X_n)$ can be written as
$$f(x_1, x_2, \cdots, x_n) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left[-\frac{1}{2}\sum_{i=1}^{n} \frac{(x_i - \mu)^2}{\sigma^2}\right]$$
$$= \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left[-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i - \bar{x})^2 + n(\bar{x} - \mu)^2\right)\right]$$

Let $y_1 = \bar{x}$, $y_i = x_i - \bar{x}$, $2 \le i \le n$. Then $x_1 - \bar{x} = -\sum_{i=2}^{n}(x_i - \bar{x}) = -\sum_{i=2}^{n} y_i$, and $\sum_{i=1}^{n}(x_i - \bar{x})^2 = \left(-\sum_{i=2}^{n} y_i\right)^2 + \sum_{i=2}^{n} y_i^2$. The joint density function of $Y_1 = \bar{X}$, $Y_i = X_i - \bar{X}$, $2 \le i \le n$, is

$$g(y_1, y_2, \cdots, y_n) = \frac{K}{(2\pi)^{n/2}\sigma^n} \exp\left[-\frac{1}{2\sigma^2}\left(\left(-\sum_{i=2}^{n} y_i\right)^2 + \sum_{i=2}^{n} y_i^2 + n(y_1-\mu)^2\right)\right]$$

where K is a constant (Jacobian of the transformation). Now $Y_1 = \bar{X} \sim N(\mu, \frac{\sigma^2}{n})$, so Y_1 is independent of Y_i, $2 \le i \le n$. It follows from $X_1 - \bar{X} = -\sum_{i=2}^{n}(X_i - \bar{X}) = -\sum_{i=2}^{n} Y_i$ that $\bar{X}$ and $X_1 - \bar{X}$ are independent. By a similar arguement, this is true for any i, $1 \le i \le n$.

Now that $\bar{X}$ is independent of $X_i - \bar{X}$, $1 \le i \le n$, $\bar{X}$ is of course independent of $\sum_{i=1}^{n}(X_i - \bar{X})^2$, and hence $\bar{X}$ and S^2 are independent.

2. To show $\frac{(n-1)}{\sigma^2} S^2 \sim \chi^2(n-1)$, observe that

$$V_1 = \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \bar{X})^2 + \frac{n(\bar{X}-\mu)^2}{\sigma^2} = \frac{(n-1)}{\sigma^2} S^2 + \frac{n(\bar{X}-\mu)^2}{\sigma^2}.$$

If we let $V_2 = \frac{(n-1)}{\sigma^2} S^2$ and $V_3 = \frac{n(\bar{X}-\mu)^2}{\sigma^2}$, then V_2 and V_3 are independent. Therefore, $M_{V_1}(t) = M_{V_2}(t) M_{V_3}(t)$,

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

$\therefore V_2 \sim \chi^2(n-1)$. //

Remark. $\frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1)$, $E\left[\frac{n-1}{\sigma^2} S^2\right] = n-1$ so

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

This is the reason we use $n-1$ instead of n.