5.3.1 Properties of the sample mean and variance

Lemma 5.3.2 (Facts about chi-squared random variables)

We use the notation χ_p^2 to denote a chi-squared random variable with p degrees of freedom.

(a) If Z is a $N(0,1)$ random variable, then $Z^2 \sim \chi_1^2$; that is, the square of a standard normal random variable is a chi-squared random variable.

(b) If $X_1, \ldots, X_n$ are independent and $X_i \sim \chi_p^2$, then $X_1 + \cdots + X_n \sim \chi^2_{p_1+\cdots+p_n}$; that is, independent chi-squared variables add to a chi-squared variables, and the degrees of freedom also add.

Proof: . Part (a) can be established based on the density formula for variable transformations. Part (b) can be established with the moment generating function. □

Theorem 5.3.1 Let $X_1, \ldots, X_n$ be a random sample from a $N(\mu, \sigma^2)$ distribution, and let $\bar{X} = (1/n)\sum_{i=1}^n X_i$ and $S^2 = [1/(n-1)]\sum_{i=1}^n (X_i - \bar{X})^2$. Then

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

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

(c) $(n-1)S^2/\sigma^2$ has a chi-squared distribution with $n-1$ degrees of freedom.

Proof: Without loss of generality, we assume that $\mu = 0$ and $\sigma = 1$. Parts (a) and (c) are proved as follows.

$$
\begin{aligned}
S^2 &= \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1}[(X_1 - \bar{X})^2 + \sum_{i=2}^n (X_i - \bar{X})^2)] \\
&= \frac{1}{n-1}[\big(\sum_{i=2}^n (X_i - \bar{X})\big)^2 + \sum_{i=2}^n (X_i - \bar{X})^2)]
\end{aligned}
$$

The last equality follows from the fact $\sum_{i=1}^n (X_i - \bar{X}) = 0$. Thus, S^2 can be written as a function only of $(X_1 - \bar{X}, \ldots, X_n - \bar{X})$. We will now show that these random variables are independent of $\bar{X}$. The joint pdf of the sample $X_1, \ldots, X_n$ is given by

$$
f(x_1, \ldots, x_n) = \frac{1}{(2\pi)^{n/2}} e^{-(1/2)\sum_{i=1}^n x_i^2}, \quad -\infty < x_i < \infty.
$$

Make the transformation

$$y_1 = \bar{x},$$
$$y_2 = x_2 - \bar{x},$$
$$\vdots$$
$$y_n = x_n - \bar{x}.$$

This is a linear transformation with a Jacobian equal to $1/n$. We have

$$\begin{aligned}
f(y_1, \ldots, y_n) &= \frac{n}{(2\pi)^{n/2}} e^{-(1/2)(y_1 - \sum_{i=2}^n y_i)^2} e^{-(1/2)\sum_{i=2}^n (y_i + y_1)^2}, \quad -\infty < y_i < \infty \\
&= \big[(\frac{n}{2\pi})^{1/2} e^{(-ny_1^2)/2}\big]\big[\frac{n^{1/2}}{(2\pi)^{(n-1)/2}} e^{-(1/2)[\sum_{i=2}^n y_i^2 + (\sum_{i=2}^n y_i)^2]}\big], \quad -\infty < y_i < \infty.
\end{aligned}$$

Hence, Y_1 is independent of $Y_2, \ldots, Y_n$, and $\bar{X}$ is independent of S^2.

Since

$$\bar{x}_{n+1} = \frac{\sum_{i=1}^{n+1} x_i}{n+1} = \frac{x_{n+1} + n\bar{x}_n}{n+1} = \bar{x}_n + \frac{1}{n+1}(x_{n+1} - \bar{x}_n),$$

we have

$$\begin{aligned}
nS_{n+1}^2 &= \sum_{i=1}^{n+1} (x_i - \bar{x}_{n+1})^2 = \sum_{i=1}^{n+1} [(x_i - \bar{x}_n) - \frac{1}{n+1}(x_{n+1} - \bar{x}_n)]^2 \\
&= \sum_{i=1}^{n+1} [(x_i - \bar{x}_n)^2 - 2(x_i - \bar{x}_n)(\frac{x_{n+1} - \bar{x}_n}{n+1}) + \frac{1}{(n+1)^2}(x_{n+1} - \bar{x}_n)^2] \\
&= \sum_{i=1}^{n} (x_i - \bar{x}_n)^2 + (x_{n+1} - \bar{x}_n)^2 - 2\frac{(x_{n+1} - \bar{x}_n)^2}{n+1} + \frac{(n+1)}{(n+1)^2}(x_{n+1} - \bar{x}_n)^2 \\
&= (n-1)S^2 + \frac{n}{n+1}(x_{n+1} - \bar{x}_n)^2.
\end{aligned}$$

Now consider $n = 2$, $S_2^2 = \frac{1}{2}(X_2 - X_1)^2$. Since $(X_2 - X_1)/\sqrt{2} \sim N(0, 1)$, part (a) of Lemma 5.3.2 shows that $S_2^2 \sim \chi_1^2$. Proceeding with the induction, we assume that for $n = k$, $(k-1)S_k^2 \sim \chi_{k-1}^2$. For $n = k + 1$, we have

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

Since S_k^2 is independent of X_{k+1} and $\bar{X}_k$, and $X_{k+1} - \bar{X}_k \sim N(0, \frac{k+1}{k})$, $kS_{k+1}^2 \sim \chi_k^2$. $\square$

Lemma 5.3.3

Let $X_j \sim N(\mu_j, \sigma_j^2)$, $j = 1, \dots, n$, independent. For constants a_{ij} and b_{rj} ($j = 1, \dots, n; i = 1, \dots, k; r = 1, \dots, m$), where $k + m \leq n$, define

$$U_i = \sum_{j=1}^{n} a_{ij} X_j, \quad i = 1, \dots, k,$$
$$V_r = \sum_{j=1}^{n} b_{rj} X_j, \quad r = 1, \dots, m.$$

(a) The random variables U_i and V_r are independent if and only if $\text{Cov}(U_i, V_r) = 0$. Furthermore, $\text{Cov}(U_i, V_r) = \sum_{j=1}^{n} a_{ij} b_{rj} \sigma_j^2$.

(b) The random vectors $(U_1, \dots, U_k)$ and $(V_1, \dots, V_m)$ are independent if and only if U_i is independent of V_r for all pairs i, r ($i = 1, \dots, k; r = 1, \dots, m$).