Difference between special sums of squares of IID $N(0,1)$ random variables

- Covariance. For two random variables X and Y, the covariance between X and Y is

$$E\left((X-E(X))(Y-E(Y))\right)=E(XY)-E(X)E(Y),$$

  denoted by $Cov(X,Y)$.

  - Note. If X and Y are independent, then $E(XY)=E(X)E(Y)$.

- Fact 1 Suppose that Z_1, …, Z_ℓ are IID $N(0,1)$ and U_1, …, U_ℓ satisfy (i) and (ii):

  (i) For $1\le i\le \ell$, $U_i=a_{i,1}Z_1+\cdots+a_{i,\ell}Z_\ell$ for some constants $a_{i,1}$, …, $a_{i,\ell}$.

  (ii) For $1\le i\le \ell$, $Var(U_i)=1$; if $\ell\ge 2$, then $Cov(U_i,U_j)=0$ for $i\ne j$, $1\le i,j\le \ell$.

  Then U_1, …, U_ℓ are IID $N(0,1)$ and

$$U_1^2+\cdots+U_\ell^2=Z_1^2+\cdots+Z_\ell^2.$$

- Fact 2 Suppose that Z_1, …, Z_ℓ are IID $N(0,1)$, $m<\ell$, and U_1, …, U_m satisfy (i) and (ii):

  (i) For $1\le i\le m$, $U_i=a_{i,1}Z_1+\cdots+a_{i,\ell}Z_\ell$ for some constants $a_{i,1}$, …, $a_{i,\ell}$.

  (ii) For $1\le i\le m$, $Var(U_i)=1$; if $m\ge 2$, then $Cov(U_i,U_j)=0$ for $i\ne j$, $1\le i,j\le m$.

  Then, $Z_1^2+\cdots+Z_\ell^2-(U_1^2+\cdots+U_m^2)$ is a $\chi^2(\ell-m)$ random variable that is independent of U_1, …, U_m.

  Proof of Fact 2. We can construct U_{m+1}, …, U_ℓ such that (i) and (ii) in Fact 1 hold. By Fact 1, U_1, …, U_ℓ are IID $N(0,1)$ randoma variables such that

$$\begin{aligned}&Z_1^2+\cdots+Z_\ell^2-(U_1^2+\cdots+U_m^2)\\&=U_{m+1}^2+\cdots+U_\ell^2,\end{aligned}$$

  which implies that $Z_1^2+\cdots+Z_\ell^2-(U_1^2+\cdots+U_m^2)$ is a $\chi^2(\ell-m)$ random variable that is independent of U_1, …, U_m.

- Example 1. Suppose that $(X_1,\ldots,X_n)$ is a random sample from $N(\mu,\sigma^2)$. Let $\bar{X}$ be the sample mean and

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

be the sample standard deviation. Then

$$\frac{(n-1)S^2}{\sigma^2} \text{ is a } \chi^2(n-1) \text{ random variable that is independent of } \bar{X}. \tag{1}$$

Proof of (1). We will prove the result assuming $\mu = 0$ and $\sigma = 1$ and the proof for the general case is left as an exercise. When $\mu = 0$ and $\sigma = 1$, we have that $X_1, \ldots, X_n$ are IID $N(0,1)$ random variables, $Var(\sqrt{n}\bar{X}) = 1$, and

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

Apply Fact 2 with $(Z_1, \ldots, Z_\ell) = (X_1, \ldots, X_n)$, $m = 1$, and $U_1 = \sqrt{n}\bar{X}$, then we have (1).

- Example 2. Suppose that $X_{i,j} : 1 \le i \le k$, $1 \le j \le n_i$ are independent and $X_{i,j} \sim N(\mu, \sigma^2)$. Let $n = \sum_{i=1}^{k} n_i$. For $1 \le i \le k$, let

$$\bar{X}_i = \frac{1}{n_i}\left(X_{i,1} + \cdots + X_{i,n_i}\right)$$

and

$$\bar{X}_G = \frac{\sum_{i=1}^{k} n_i \bar{X}_i}{n} = \frac{n_1\bar{X}_1 + \cdots + n_k\bar{X}_k}{n_1 + \cdots + n_k}.$$

Let $SST = \sum_{i=1}^{k} n_i(\bar{X}_i - \bar{X}_G)^2$ and

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_1}(X_{i,j} - \bar{X}_i)^2.$$

Show that $SSE/\sigma^2 \sim \chi^2(n-k)$, $SST/\sigma^2 \sim \chi^2(k-1)$ and SSE is independent of SST.

Sol. We will prove the result assuming $\mu = 0$ and $\sigma = 1$ and the proof for the general case is left as an exercise. When $\mu = 0$ and $\sigma = 1$, $X_{i,j}$s are IID $N(0,1)$. Let $Z_i = \sqrt{n_i}\bar{X}_i$ for $i = 1, \ldots, k$, and

$$U_1 = \sqrt{n}\bar{X}_G = \sqrt{n}\sum_{i=1}^{k} n_i\bar{X}_i/n = \sum_{i=1}^{k}\frac{\sqrt{n_i}Z_i}{\sqrt{n}},$$

then $Z_1, \ldots, Z_k$ are IID $N(0,1)$ random variables and $Var(U_1) = 1$, so (i) and (ii) in Fact 2 hold with $\ell = k$, $m = 1$, and $U_1 = U_1$. Moreover,

$$SST = \sum_{i=1}^{k} n_i(\bar{X}_i - \bar{X}_G)^2$$

$$
\begin{aligned}
&= \left(\sum_{i=1}^{k} (\sqrt{n_i}\bar{X}_i)^2 \right) - (\sqrt{n}\bar{X}_G)^2 \qquad (2) \\
&= \left(\sum_{i=1}^{k} Z_i^2 \right) - U_1^2,
\end{aligned}
$$

so by Fact 2, SST is a $\chi^2(k-1)$ random variable that is independent of $U_1 = \sqrt{n}\bar{X}_G$.

To find the distribution of SSE, note that

$$
\begin{aligned}
SSE &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{i,j} - \bar{X}_G)^2 - SST \\
&\overset{(2)}{=} \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{i,j} - \bar{X}_G)^2 - \left(\left(\sum_{i=1}^{k} (\sqrt{n_i}\bar{X}_i)^2 \right) - (\sqrt{n}\bar{X}_G)^2 \right) \\
&= \left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{i,j}^2 \right) - \left(\sum_{i=1}^{k} (\sqrt{n_i}\bar{X}_i)^2 \right),
\end{aligned}
$$

which is a $\chi^2(n-k)$ random variable that is independent of $(\bar{X}_1, \ldots, \bar{X}_k)$ by Fact 2. Since SST is a function of $(\bar{X}_1, \ldots, \bar{X}_k)$ and SSE is independent of $(\bar{X}_1, \ldots, \bar{X}_k)$, SST and SSE are independent. We have shown that $SSE/\sigma^2 \sim \chi^2(n-k)$, $SST/\sigma^2 \sim \chi^2(k-1)$ and SSE is independent of SST assuming $\mu = 0$ and $\sigma = 1$.