Sample Mean and Variance for I.I.D. Normals

Suppose that $X_i, i = 1, \ldots, n$ are i.i.d. $N(0,1)$ random variables, $n \geq 2$. The sample mean and variance are then given by

$$\begin{aligned}\bar{X} &= \frac{1}{n}\sum_{i=1}^{n} X_i \\ s^2 &= \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\end{aligned}$$

Theorem 1 *Under the condition above, that is $X_i, i = 1, \ldots, n$ are i.i.d. $N(0,1)$ random variables, then*

1. *the random variables $\bar{X}$ and $\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$ are independent*
2. $$\sqrt{n}\bar{X} \sim N(0,1)$$
3. $$(n-1)s^2 = \sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \sim \chi^2_{(n-1)}$$

The rest of this handout gives a proof of this statement using a matrix or linear algebra approach. This is different from the method used in the Rice text, chapter 6. However this method is more useful in other settings such as regression and other linear models.

First we review some linear algebra results needed. Let $\mathbf{x}$ and $\mathbf{y}$ be column vectors of length n. For an $n \times n$ matrix $\mathbf{A}$ consider

$$\mathbf{y} = \mathbf{A}\mathbf{x} \tag{1}$$

Suppose that the i-th row of $\mathbf{A}$ is $\mathbf{a}_i$, which is a row vector of length n. The transpose of a vector or matrix is denoted by a superscript t, for example $\mathbf{a}^t$ is the transpose of a vector $\mathbf{a}$. Recall that the inner product of these vectors is given by

$$\langle \mathbf{a}_i, \mathbf{a}_k \rangle = \mathbf{a}_i^t \mathbf{a}_k = \sum_{j=1}^{n} a_{i,j}, a_{k,j}$$

where

$$\mathbf{a}_i = (a_{i,1}, a_{1,2}, \ldots, a_{i,n})$$

that is $a_{i,j}$ is the j-th component of the vector.

The matrix $\mathbf{A}$ is said to be orthogonal if the rows are orthogonal, that is

$$\langle \mathbf{a}_i, \mathbf{a}_k \rangle = 0 \text{ for } i \neq k$$

The matrix is said to orthonormal if it is orthogonal and each row has length 1, that is

$$\langle \mathbf{a}_i, \mathbf{a}_i \rangle = 1 \ .$$

Aside : Sometimes orthogonal matrices is the terminology of what is called orthonormal matrices. This is not universal. In general vectors can be orthogonal but not have length 1. Non square matrices can also be orthogonal matrices.

A property of a a square matrix $\mathbf{A}$ (that is $n \times n$) that is also an orthonormal matrix is that

$$\mathbf{A}^t\mathbf{A} = \mathbf{A}\mathbf{A}^t = \mathbf{I}_n$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix. This follows from the uniqueness of the inverse for a square matrix and the fact that $\mathbf{A}^t$ is a right inverse of $\mathbf{A}$.

In (1), suppose that $\mathbf{A}$ is orthonormal. Then

$$\mathbf{y}^t\mathbf{y} = \mathbf{x}^t\mathbf{A}^t\mathbf{A}\mathbf{x} = \mathbf{x}^t\mathbf{x} \ . \tag{2}$$

Proposition 1 *Suppose that $X_1, \ldots, X_n$ are i.i.d. $N(0,1)$ r.v.'s. Let $\mathbf{X}$ be the column vector such that $\mathbf{X}^t = (X_1, \ldots, X_n)$, that is $\mathbf{X}$ is the column vector with i-th component X_i. Consider the transformation*

$$\mathbf{Y} = \mathbf{A}\mathbf{X}$$

where $\mathbf{A}$ is orthonormal. Then $\mathbf{Y} = (Y_1, \ldots, Y_n)$ has the distribution such that the r.v.'s are i.i.d. $N(0,1)$.

Proof

The pdf of $\mathbf{Y}$ is

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(\mathbf{A}^t\mathbf{y})|\det(\mathbf{A}^t)| = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2}\mathbf{y}^t\mathbf{y}\right\} = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2}\left(y_1^2 + \ldots y_n^2\right)\right\} \ .$$

From this we see that the joint pdf factors into the product of n standard normal pdf's. From this we can then calculate the marginal pdf of Y_i, which we find is the standard normal pdf. Thus we find $Y_i, i = 1, \ldots, n$ are iid $N(0,1)$. This completes the proof of Proposition 1

We next construct an orthonormal matrix $\mathbf{A}$ that will be a preliminary step in the proof of Theorem 1. Consider row vectors (of length n)

$$\begin{aligned}
\alpha_1 &= (1, 1, \ldots, 1) \\
\alpha_2 &= (-1, 1, 0, \ldots, 0) \\
\alpha_3 &= \left(-\frac{1}{2}, -\frac{1}{2}, 1, 0, \ldots, 0\right) \\
&\vdots \\
\alpha_i &= \left(\underbrace{-\frac{1}{i-1}, \ldots, -\frac{1}{i-1}}_{i-1 \text{ terms}}, 1, \underbrace{0 \ldots, 0}_{n-i \text{ terms}}\right) \\
&\vdots \\
\alpha_n &= \left(\underbrace{-\frac{1}{n-1}, \ldots, -\frac{1}{n-1}}_{n-1 \text{ terms}}, 1\right)
\end{aligned}$$

These vectors α_i are orthogonal. We find they have length

$$\begin{aligned}\langle \alpha_1, \alpha_1 \rangle &= n \\ \langle \alpha_i, \alpha_i \rangle &= \frac{i}{i-1} \text{ for } 2 \le i \le n \ .\end{aligned}$$

Define the row vector

$$\mathbf{a}_i = \frac{\alpha_i}{\sqrt{\langle \alpha_i, \alpha_i \rangle}} \ .$$

The first few of these are

$$\begin{aligned}\mathbf{a}_1 &= \frac{1}{\sqrt{n}}(1, 1, \ldots, 1) \\ \mathbf{a}_2 &= \sqrt{\frac{1}{2}}(-1, 1, 0, \ldots, 0) \\ \mathbf{a}_3 &= \sqrt{\frac{2}{3}}\left(-\frac{1}{2}, -\frac{1}{2}, 1, 0, \ldots, 0\right) \\ \mathbf{a}_i &= \left(\underbrace{-\frac{1}{\sqrt{i(i-1)}}, \ldots, -\frac{1}{\sqrt{i(i-1)}}}_{i-1 \text{ terms}}, \sqrt{\frac{i-1}{i}}, \underbrace{0 \ldots, 0}_{n-i \text{ terms}}\right)\end{aligned}$$

Construct the matrix $\mathbf{A}$ to have i-th row $\mathbf{a}_i$. This matrix is orthonormal.

For this matrix, and $\mathbf{y} = \mathbf{A}\mathbf{x}$ the following properties hold.

1. $y_1 = \sqrt{n}\bar{x}$
2. from (2)

$$y_2^2 + \ldots y_n^2 = \sum_{i=1}^{n} x_i^2 - (\sqrt{n}\bar{x})^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Proof of Theorem 1

For the transformation $\mathbf{Y} = \mathbf{A}\mathbf{X}$, we have Y_i are i.i.d. $N(0, 1)$ by Proposition 1. Thus $\bar{X} = Y_1/\sqrt{n} \sim N(0, 1/n)$ and $Y_2^2 + \ldots Y_n^2 \sim \chi^2_{(n-1)}$ are independent. From the second property above we thus have

$$\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2 \sim \chi^2_{(n-1)}$$

and that it is independent of $\bar{X}$. This concludes the proof of the Theorem.

Theorem 1 has some very useful and immediate corollaries.

However it does not say that X_1 is independent of $Y_2, \ldots, Y_n$. In particular X_1 and $\sum_{i=1}^{n}(X_i - \bar{X})^2$ are dependent.

The Rice text obtains these results in a different manner, by using moment generating functions. This method is very special, while the method above is useful provided one can construct appropriate orthonormal matrices, such as in linear regression. We now consider the method in Rice Chapter 6. For this we need some preliminary results.

R.v.s (X, Y) have joint moment generating function

$$M_{X,Y}(s,t) = \mathrm{E}(e^{sX+tY})$$

provided this expectation exists for all (s,t) in an open neighbourhood of $(0,0)$. This means there is a $\delta > 0$ such that $M_{X,Y}(s,t)$ is finite for all $||(s,t)|| = \sqrt{s^2+t^2} < \delta$. $||(s,t)||$ is the Euclidean norm or length for the vector (s,t). This will imply that $M_{X,Y}$ will have a power series with radius of convergence r which is at least as big as δ. The marginal mgf of X is then easily obtained as

$$M_X(s) = \mathrm{E}\left(e^{sX}\right) = \mathrm{E}\left(e^{sX+0Y}\right) = M_{X,Y}(s,0)\ .$$

The moment generating function generalizes to (joint) mgf of n r.v.s $(X_1,\ldots,X_n)$ by

$$M_{X_1,\ldots,X_n}(s_1,\ldots,s_n) = \mathrm{E}\left(e^{s_1X_1+\ldots+s_nX_n}\right)$$

provided this expectation is finite for all $(s_1,\ldots,s_n)$ in an open neighbourhood of $\mathbf{0} \in R^n$. Marginal moment generating functions can be obtained by setting appropriate arguments of $M_{X_1,\ldots,X_n}$ to 0, just as in the bivariate mgf case above.

A r.v. X_1 is independent of r.v.'s $X_2,\ldots,X_n$ if the joint mgf factors into the product of the marginal mgf's. That is if

$$M_{X_1,\ldots,X_n}(t_1,\ldots,t_n) = M_{X_1}(t_1)M_{X_2,\ldots,X_n}(t_2,\ldots,t_n) \tag{3}$$

then X_1 is independent of $X_2,\ldots,X_n$. It is this property that Rice uses in the proof of Theorem 6.3A.

Before proceeding with the proof let us look more closely at Theorem A. It deals with a transformation that maps n r.v.s $X_1,\ldots,X_n$ into $n+1$ r.v.s $\bar{X}, X_1-\bar{X}, X_2-\bar{X},\ldots,X_n-\bar{X}$. This transformation or mapping is from $R^n \mapsto R^{n+1}$ and cannot be a 1 to 1 mapping and so our direct method of change of variables and Jacobians will not apply.

Theorem 2 (Rice Theorem 6.3A) *Suppose $X_1,\ldots,X_n$ are iid $N(\mu,\sigma^2)$. Then the $\bar{X}$ is independent of the random vector $(X_1-\bar{X},\ldots,X_n-\bar{X})$.*

Proof : Let M be the mgf of the $n+1$ random variables $\bar{X}, X_1-\bar{X}, X_2-\bar{X},\ldots,X_n-\bar{X}$, that is

$$M(s,t_1,\ldots,t_n) = \mathrm{E}\left(\exp\left\{s\bar{X}+t_1(X_1-\bar{X})+t_2(X_2-\bar{X})+\ldots+t_n(X_n-\bar{X})\right\}\right)$$

With some careful algebra

$$\begin{aligned} & s\bar{X}+t_1(X_1-\bar{X})+t_2(X_2-\bar{X})+\ldots+t_n(X_n-\bar{X}) \\ = & \sum_{i=1}^{n}\left[\frac{s}{n}+t_i-\bar{t}\right]X_i \\ = & \sum_{i=1}^{n} a_iX_i \end{aligned}$$

where

$$a_i = \frac{s}{n}+t_i-\bar{t}\ .$$

Then

$$\sum_{i=1}^{n} a_i = s$$

and

$$\sum_{i=1}^{n} a_i^2 \quad = \quad \frac{s^2}{n} + \sum_{i=1}^{n} (t_i - \bar{t})^2 \ .$$

Therefore

$$\begin{aligned} M(s, t_1, \ldots, t_n) &= \prod_{i=1}^{n} \mathrm{E}\left(e^{a_i X_i}\right) \\ &= \exp\left\{\sum_{i=1}^{n} a_i \mu + \frac{1}{2}\sum_{i=1}^{n} a_i^2 \sigma^2\right\} \\ &= \exp\left\{s\mu + \frac{\sigma^2}{2n}s^2\right\} \exp\left\{\frac{\sigma^2}{2}\sum_{i=1}^{n}(t_i - \bar{t})^2\right\} \end{aligned}$$

This is the product of two functions, but are they both mgfs? Recall $M(s, 0, \ldots, 0)$ is the mgf of $\bar{X}$, and is $\exp\left\{s\mu + \frac{\sigma^2}{2n}s^2\right\}$. The other term is therefore $M(0, t_1, \ldots, t_n)$ which therefore the mgf of $(X_1 - \bar{X}, X_2 - \bar{X}, \ldots, X_n - \bar{X})$. Thus by the property (3) we conclude that Rice Theorem 6.3A is proved.

End of Proof

From the proof we might not easily recognize what is the distribution of $(X_1 - \bar{X}, X_2 - \bar{X}, \ldots, X_n - \bar{X})$. We can however conclude that $\bar{X}$ and S^2 are independent. Rice Theorem 6.3B is an indirect way of finding the distribution of S^2. The student should read this Rice. Here we examine the properties that he is using.

Note that

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

Then

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

The LHS is a $\chi^2_{(n)}$ r.v..

The RHS is the sum of two independent r.v.s (say U , V) and its mgf is thus the product of these two mgfs. The distribution of V is $\chi^2_{(1)}$. Thus

$$\left(\frac{1}{1-2t}\right)^{\frac{n}{2}} = M_U(t)\left(\frac{1}{1-2t}\right)^{\frac{1}{2}} \ .$$

We can solve for M_U giving

$$M_U(t) = (1-2t)^{-\frac{n-1}{2}}$$

and hence $U \sim \chi^2_{(n-1)}$.