10 Distributions of the sample mean and variance

Statisticians do it. After all, it's only normal.

10.1 Simpson's paradox

Example 10.1 *These are some Cambridge admissions statistics for 1996.*

	Women			Men		
	applied	accepted	%	applied	accepted	%
Computer Science	26	7	27	228	58	25
Economics	240	63	26	512	112	22
Engineering	164	52	32	972	252	26
Medicine	416	99	24	578	140	24
Veterinary medicine	338	53	16	180	22	12
Total	1184	274	23	2470	584	24

In all five subjects women have an equal or better success rate in applications than do men. However, taken overall, 24% of men are successful but only 23% of women are successful! This is called **Simpson's paradox** (though it was actually discovered by Yule 50 years earlier). It can often be found in real data. Of course it is not a paradox. The explanation here is that women are more successful in each subject, but tend to apply more for subjects that are hardest to get into (e.g., Veterinary medicine). This example should be taken as a warning that pooling contingency tables can produce spurious associations. The correct interpretation of this data is that, for these five subjects, women are significantly more successful in gaining entry than are men.

In order to produce an example of Simpson's paradox I carefully selected five subjects from tables of 1996 admissions statistics. Such 'data snooping' is cheating; a similar table that reversed the roles of men and women could probably be constructed by picking different subjects.

10.2 Transformation of variables

The rest of this lecture is aimed at proving some important facts about distribution of the statistics $\bar{X}$ and $S_{XX} = \sum_i (X_i - \bar{X})^2$, when $X_1, \dots, X_n$ are IID $N(\mu, \sigma^2)$. We begin by reviewing some ideas about transforming random variables.

Suppose the joint density of $X_1, \dots, X_n$ is f_X, and there is a 1–1 mapping between $X_1, \dots, X_n$ and $Y_1, \dots, Y_n$ such that $X_i = x_i(Y_1, \dots, Y_n)$. Then the joint density of

$Y_1, \dots, Y_n$ is

$$f_Y(y_1, \dots, y_n) = f_X(x_1(y), \dots, x_n(y)) \begin{vmatrix} \frac{\partial x_1(y)}{\partial y_1} & \cdots & \frac{\partial x_1(y)}{\partial y_n} \\ \vdots & & \vdots \\ \frac{\partial x_n(y)}{\partial y_1} & \cdots & \frac{\partial x_n(y)}{\partial y_n} \end{vmatrix}$$

where the **Jacobian** $:= J(y_1, \dots, y_n)$ is the absolute value of the determinant of the matrix $(\partial x_i(y)/\partial y_j)$.

The following example is an important one, which also tells us more about the beta distribution.

Example 10.2 Let $X_1 \sim \text{gamma}(n_1, \lambda)$ and $X_2 \sim \text{gamma}(n_2, \lambda)$, independently. Let $Y_1 = X_1/(X_1 + X_2)$, $Y_2 = X_1 + X_2$. Since X_1 and X_2 are independent we multiply their pdfs to get

$$f_X(x) = \frac{\lambda^{n_1} x_1^{n_1 - 1}}{(n_1 - 1)!} e^{-\lambda x_1} \times \frac{\lambda^{n_2} x_2^{n_2 - 1}}{(n_2 - 1)!} e^{-\lambda x_2}.$$

Then $x_1 = y_1 y_2$, $x_2 = y_2 - y_1 y_2$, so

$$J(y_1, y_2) = \begin{vmatrix} \frac{\partial x_1(y)}{\partial y_1} & \frac{\partial x_1(y)}{\partial y_2} \\ \frac{\partial x_2(y)}{\partial y_1} & \frac{\partial x_2(y)}{\partial y_2} \end{vmatrix} = \begin{vmatrix} y_2 & y_1 \\ -y_2 & 1 - y_1 \end{vmatrix} = y_2$$

Hence making the appropriate substitutions and arranging terms we get

$$f_Y(y) = \frac{(n_1 + n_2 - 1)!}{(n_1 - 1)!(n_2 - 1)!} y_1^{n_1 - 1}(1 - y_1)^{n_2 - 1} \times \frac{\lambda^{n_1 + n_2} y_2^{n_1 + n_2 - 1}}{(n_1 + n_2 - 1)!} e^{-\lambda y_2}$$

from which it follows that Y_1 and Y_2 are independent RVs (since their joint density function factors into marginal density functions) and $Y_1 \sim \text{beta}(n_1, n_2)$, $Y_2 \sim \text{gamma}(n_1 + n_2, \lambda)$.

10.3 Orthogonal transformations of normal variates

Lemma 10.3 *Let $X_1, \dots, X_n$, be independently distributed with distributions $N(\mu_i, \sigma^2)$ respectively. Let $A = (a_{ij})$ be an orthogonal matrix, so that $A^\top A = AA^\top = I$. Then the elements of $\boldsymbol{Y} = A\boldsymbol{X}$ are independently distributed, and $Y_i \sim N\big((A\mu)_i, \sigma^2\big)$, where $\mu = (\mu_1, \dots, \mu_n)^\top$.*

Proof. The joint density of $X_1, \dots, X_n$ is

$$\begin{aligned} f_X(x_1, \dots, x_n \mid \mu, \sigma^2) &= \prod_i f_{X_i}(x_i \mid \mu_i, \sigma^2) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\sum_i (x_i - \mu_i)^2/2\sigma^2} \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-(\mathbf{x} - \mu)^\top (\mathbf{x} - \mu)/2\sigma^2} \end{aligned}$$

Since $\mathbf{x} = A^\top \mathbf{y}$, we have $\partial x_i/\partial y_j = a_{ji}$ and hence $J(y_1, \dots, y_n) = |\det(A^\top)| = 1$. Thus

$$\begin{aligned}
f_Y(y_1, \dots, y_n \mid \mu, \sigma^2) &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-(A^\top\mathbf{y} - \mu)^\top(A^\top\mathbf{y} - \mu)/2\sigma^2\right] \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-(A^\top\mathbf{y} - A^\top A\mu)^\top(A^\top\mathbf{y} - A^\top A\mu)/2\sigma^2\right] \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-(\mathbf{y} - A\mu)^\top AA^\top(\mathbf{y} - A\mu)/2\sigma^2\right] \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-(\mathbf{y} - A\mu)^\top(\mathbf{y} - A\mu)/2\sigma^2\right]
\end{aligned}$$

■

Remark. An alternative proof can be given using moment generating functions. For $\theta \in \mathbb{R}^n$, the mgf of the joint distribution is

$$\begin{aligned}
\mathbb{E}\exp\left[\theta^\top \mathbf{Y}\right] &= \mathbb{E}\exp\left[\theta^\top A\mathbf{X}\right] \\
&= \mathbb{E}\exp\left[(A^\top\theta)^\top \mathbf{X}\right] \\
&= \exp\left[(A^\top\theta)^\top\mu + \tfrac{1}{2}\sigma^2(A^\top\theta)^\top(A^\top\theta)\right] \\
&= \exp\left[\theta^\top A\mu + \tfrac{1}{2}\sigma^2\theta^\top\theta\right]
\end{aligned}$$

which we recognise as the mgf of independent RVs with distributions $N\big((A\mu)_i, \sigma^2\big)$.

10.4 The distributions of $\bar{X}$ and S_{XX}

Lemma 10.4 *Let $X_1, \dots, X_n$ be IID $N(\mu, \sigma^2)$ and let $\bar{X} = n^{-1}\sum_{i=1}^n X_i$, $S_{XX} = \sum_{i=1}^n (X_i - \bar{X})^2$. Then:*

(i) $\bar{X} \sim N(\mu, \sigma^2/n)$ and $n(\bar{X} - \mu)^2 \sim \sigma^2\chi_1^2$.

(ii) $X_i - \mu \sim N(0, \sigma^2)$, so $\sum_{i=1}^n (X_i - \mu)^2 \sim \sigma^2\chi_n^2$.

(iii) $\sum_{i=1}^n (X_i - \mu)^2 = S_{XX} + n(\bar{X} - \mu)^2$.

(iv) $S_{XX}/(n-1)$ is an unbiased estimator of σ^2.

(v) $\bar{X}$ and S_{XX} are independent.

(vi) $S_{XX} \sim \sigma^2\chi_{n-1}^2$.

Proof.

(i) and (ii) are immediate from the fact that linear combinations of normal RVs are normally distributed and the definition of χ_n^2. To prove (iii) and (iv) we note

that

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

Let A be an orthogonal matrix such that

$$\mathbf{Y} = A(\mathbf{X} - \mu 1) = \big(\sqrt{n}(\bar{X} - \mu), Y_2, \dots, Y_n\big).$$

I.e., we take

$$A = \begin{pmatrix} 1/\sqrt{n} & 1/\sqrt{n} & \cdots & 1/\sqrt{n} \\ \vdots & \vdots & & \vdots \\ . & . & \dots & . \end{pmatrix}$$

where the rows below the first are chosen to make the matrix orthogonal. Then $Y_1 = \sqrt{n}(\bar{X} - \mu) \sim N(0, \sigma^2)$ and Y_1 is independent of $Y_2, \dots, Y_n$. Since $\sum_{i=1}^n Y_i^2 = \sum_i (X_i - \mu)^2$, we must have

$$\sum_{i=2}^n Y_i^2 = \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2 = S_{XX}.$$

Hence S_{XX} and Y_1 (and equivalently S_{XX} and $\bar{X}$) are independent. This gives (v).

Finally, (vi) follows from $S_{XX} = \sum_{i=2}^n Y_i^2$ and the fact that $Y_2, \dots, Y_n$ are IID $N(0, \sigma^2)$. ■

10.5 Student's t-distribution

If $X \sim N(0, 1)$, $Y \sim \chi_n^2$, independently of X, then

$$Z = X/(Y/n)^{\frac{1}{2}} \sim t_n,$$

where t_n is the Student's **t-distribution** with (or on) n degrees of freedom. Like the normal distribution, this distribution is symmetric about 0, and bell-shaped, but has more probability in its tails, i.e., for all $t > 0$, $\mathbb{P}(Z > t) > \mathbb{P}(X > t)$.

From Lemma 10.4 we have $\sqrt{n}(\bar{X} - \mu) \sim \sigma N(0, 1)$ and $S_{XX} \sim \sigma^2\chi_{n-1}^2$, independently. So from these and the definition of the t-distribution follows the important fact that if $X_1, \dots, X_n$ are IID $N(\mu, \sigma^2)$, then

$$\frac{\sqrt{n}(\bar{X} - \mu)}{\sqrt{S_{XX}/(n-1)}} \sim t_{n-1}.$$