Wilks' Λ and Hotelling's T^2.

Steffen Lauritzen, University of Oxford

BS2 Statistical Inference, Lecture 13, Hilary Term 2008

March 2, 2008

If X and Y are independent , $X \sim \Gamma(\alpha_x, \gamma)$, and $Y \sim \Gamma(\alpha_y, \gamma)$, then the ratio $X/(X+Y)$ follows a Beta distribution:

$$B = \frac{X}{X+Y} \sim \mathcal{B}(\alpha_x, \alpha_y).$$

A multivariate analogue of this result involves the Wishart distribution and asserts.

If $W_1 \sim \mathcal{W}_d(f_1, \Sigma)$ and $W_2 \sim \mathcal{W}_d(f_2, \Sigma)$ with $f_1 \geq d$, then the distribution of

$$\Lambda = \frac{\det(W_1)}{\det(W_1 + W_2)}$$

does not depend on Σ and is denoted by $\Lambda(d, f_1, f_2)$. The distribution is known as *Wilks' distribution*.

To see that the distribution of Λ does not depend on Σ, we choose a matrix A such that $A\Sigma A^\top = I_d$. Then

$$\tilde{W}_i = AW_iA^\top \sim \mathcal{W}_d(f_i, I_d)$$

and

$$\tilde{\Lambda} = \frac{\det(\tilde{W}_1)}{\det(\tilde{W}_1 + \tilde{W}_2)} = \frac{\det(A)\det(W_1)\det(A^\top)}{\det(A)\det(W_1 + W_2)\det(A^\top)} = \Lambda.$$

Clearly, the distribution of $\tilde{\Lambda}$ does not depend on Σ and as $\tilde{\Lambda} = \Lambda$ this also holds for the latter.

Wilks' distribution is closely related to the Beta distribution. *It holds that*

$$\Lambda \stackrel{\mathcal{D}}{=} \prod_{i=1}^{d} B_i$$

where B_i are independent and follow Beta distributions with

$$B_i \sim \mathcal{B}\{(f_1 + 1 - i)/2, f_2/2)\}.$$

Indeed the distribution of

$$(W_1 + W_2)^{-1} W_1$$

is also known as *the multivariate Beta distribution*.

We first need a useful result about determinants of block matrices.

If A is a $d \times d$ symmetric matrix partitioned into blocks of dimension $r \times r$, $r \times s$, and $s \times s$ as

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},$$

it holds that

$$\det A = \det(A_{11} - A_{12}A_{22}^{-1}A_{21}) \det(A_{22}). \quad (1)$$

Here the entire expression should be considered equal to 0 if A_{22} is not invertible and $\det(A_{22}) = 0$.

This follows from a simple calculation

$$\begin{aligned}
\det(A) &= \det\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \det\begin{pmatrix} I_{r\times r} & 0_{r\times s} \\ -A_{22}^{-1}A_{21} & I_{s\times s} \end{pmatrix} \\
&= \det\begin{pmatrix} A_{11} - A_{12}A_{22}^{-1}A_{21} & A_{12} \\ 0_{s\times r} & A_{22} \end{pmatrix} \\
&= \det(A_{11} - A_{12}A_{22}^{-1}A_{21})\det(A_{22}).
\end{aligned}$$

Consider a partitioning of W and Σ into blocks as

$$W = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

where Σ_{11} is an $r \times r$ matrix, Σ_{22} is $s \times s$, etc.

If $W \sim \mathcal{W}_d(f, \Sigma)$ and $\Sigma_{12} = \Sigma_{21} = 0$ then

$$\frac{\det(W)}{\det(W_{11})\det(W_{22})} \sim \Lambda(r, f-s, s) = \Lambda(s, f-r, r).$$

To see this is true we first use the matrix identity (1) to write

$$\frac{\det(W)}{\det(W_{11})\det(W_{22})} = \frac{\det(W_{1|2})}{\det(W_{11})} = \frac{\det(W_{1|2})}{\det(W_{1|2} + W_{12}W_{22}^{-1}W_{21})},$$

where $W_{1|2} = W_{11} - W_{12}W_{22}^{-1}W_{21}$.

Next we need to use that if $\Sigma_{12} = 0$ and thus $\Sigma_{1|2} = \Sigma_{11}$, it further holds that $W_{1|2}$ and $W_{12}W_{22}^{-1}W_{21}$ are independent and both Wishart distributed as

$$W_{1|2} \sim \mathcal{W}_r(f - s, \Sigma_{11}), \quad W_{12}W_{22}^{-1}W_{21} \sim \mathcal{W}_r(s, \Sigma_{11}).$$

We abstain from giving further details.

Wilks' distribution occurs as the likelihood ratio test for independence. Consider $X_1, \ldots, X_n \sim \mathcal{N}_d(0, \Sigma)$. The likelihood function is

$$L(K) = (\det K)^{n/2} e^{-\operatorname{tr}(KW)/2}.$$

As this is maximized by

$$\hat{K} = nW^{-1}$$

we have

$$L(\hat{K}) = (\det W)^{-n/2} e^{-nd/2}.$$

If $\Sigma_{12} = 0$ we similarly have

$$L(\hat{K}_{11}, \hat{K}_{22}) = (\det W_{11})^{-n/2} e^{-nr/2} (\det W_{22})^{-n/2} e^{-ns/2}.$$

Hence the likelihood ratio statistic is

$$\frac{L(\hat{K}_{11}, \hat{K}_{22})}{L(\hat{K})} = \left\{ \frac{\det(W)}{\det(W_{11}) \det(W_{22})} \right\}^{n/2} = \Lambda^{n/2}.$$

Let $Y \sim \mathcal{N}_d(\mu, c\Sigma)$ and $W \sim \mathcal{W}_d(f, \Sigma)$ with $f \geq d$, and $Y \perp\!\!\!\perp W$. Then

$$T^2 = f(Y - \mu)^\top W^{-1}(Y - \mu)/c$$

is known as *Hotelling's T^2*. This is the multivariate analogue of Student's t (or rather t^2).

It is equivalent to the likelihood ratio statistic for testing $\mu = 0$ from a sample $X_1, \ldots, X_n$ where then $Y = \bar{X}$, $W = \sum_i (X_i - \bar{X})$, $f = n - 1$, and $c = 1/n$.

It holds that

$$\frac{1}{1 + T^2/f} \sim \Lambda(d, f, 1) = \Lambda(1, f - d + 1, d).$$

To see this we exploit the matrix identity (1) and calculate a determinant in two different ways. We may without loss of generality let $\mu = 0$. We have

$$\det \begin{pmatrix} W & -Y/\sqrt{c} \\ Y/\sqrt{c} & 1 \end{pmatrix} = \det(W + YY^\top/c) \cdot 1,$$

But we also have

$$\begin{aligned} \det \begin{pmatrix} W & -Y/\sqrt{c} \\ Y/\sqrt{c} & 1 \end{pmatrix} &= \det(1 + Y^\top W^{-1} Y/c) \det W \\ &= (1 + Y^\top W^{-1} Y/c) \det W \\ &= (1 + T^2/f) \det W. \end{aligned}$$

Hence

$$\frac{1}{1+T^2/f} = \frac{1}{1+Y^\top W^{-1}Y/c} = \frac{\det W}{\det(W+YY^\top/c)}.$$

The result now follows by noting that $Y \sim \mathcal{N}_d(0, c\Sigma)$ implies $YY^\top/c \sim \mathcal{W}_d(1, \Sigma)$. Since

$$\Lambda(d, f, 1) = \Lambda(1, f-d+1, d)$$

and the latter is a Beta distribution, *it also holds that*

$$\frac{f-d+1}{fd} T^2 \sim F(d, f+1-d)$$

where F denotes Fisher's F-distribution.