Wishart and Inverse Wishart Distributions

Steffen Lauritzen, University of Oxford

BS2 Statistical Inference, Lecture 9, Hilary Term 2009

February 23, 2009

The Wishart distribution is the sampling distribution of the matrix of sums of squares and products. More precisely:

A random $d \times d$ matrix W has a *d-dimensional Wishart distribution* with parameter Σ and n *degrees of freedom* if

$$W \stackrel{\mathcal{D}}{=} \sum_{i=1}^{n} X_\nu X_\nu^\top$$

where $X_\nu \sim \mathcal{N}_d(0, \Sigma)$. We then write

$$W \sim \mathcal{W}_d(n, \Sigma).$$

The Wishart is the multivariate analogue to the χ^2:

$$\mathcal{W}_1(n, \sigma^2) = \sigma^2 \chi^2(n).$$

If $W \sim \mathcal{W}_d(n, \Sigma)$ its mean is $\mathbf{E}(W) = n\Sigma$.

If W_1 and W_2 are independent with $W_i \sim \mathcal{W}_d(n_i, \Sigma)$, then

$$W_1 + W_2 \sim \mathcal{W}_d(n_1 + n_2, \Sigma).$$

If A is an $r \times d$ matrix and $W \sim \mathcal{W}_d(n, \Sigma)$, then

$$AWA^\top \sim \mathcal{W}_r(n, A\Sigma A^\top).$$

For $r = 1$ we get that when $W \sim \mathcal{W}_d(n, \Sigma)$ and $\lambda \in R^d$,

$$\lambda^\top W \lambda \sim \sigma_\lambda^2 \chi^2(n),$$

where $\sigma_\lambda^2 = \lambda^\top \Sigma \lambda$.

If $W \sim \mathcal{W}_d(n, \Sigma)$, where Σ is regular, then *W is regular with probability one if and only if $n \geq d$*.

When $n \geq d$ the Wishart distribution has density

$$
\begin{aligned}
& f_d(w \mid n, \Sigma) \\
&= \quad c(d, n)^{-1} (\det \Sigma)^{-n/2} (\det w)^{(n-d-1)/2} e^{-\operatorname{tr}(\Sigma^{-1} w)/2}
\end{aligned}
$$

for w positive definite, and 0 otherwise.

The *Wishart constant* $c(d, n)$ is

$$
c(d, n) = 2^{nd/2} (2\pi)^{d(d-1)/4} \prod_{i=1}^{d} \Gamma\{(n + 1 - i)/2\}.
$$

Let $W \sim \mathcal{W}_d(n, \Sigma)$ with Σ regular and $n > d$. Then W_{22} is regular with probability one and

(i) $W_{1|2}$ is independent of (W_{12}, W_{22});

(ii) $W_{1|2} \sim \mathcal{W}_r(n - s, \Sigma_{1|2})$;

(iii) $W_{22} \sim \mathcal{W}_s(n, \Sigma_{22})$;

(iv) The conditional distribution of W_{12} given $W_{22} = w_{22}$ is multivariate Gaussian $\mathcal{N}_{r\times s}(\Sigma_{12}\Sigma_{22}^{-1}w_{22}, \Lambda)$ where

$$\Lambda_{ij,kl} = \mathrm{Cov}(W_{ij}, W_{kl} \mid W_{22} = w_{22}) = \sigma_{ik}^{1|2} w_{jl}.$$

In the special case with $\Sigma_{12} = 0$ this can be simplified to $W_{1|2} \sim \mathcal{W}_r(n - s, \Sigma_{11})$ and

$$W_{12} \,|\, W_{22} = w_{22} \sim \mathcal{N}_{r \times s}(0, \Lambda)$$

with $\Lambda_{ij,kl} = \sigma_{ik} w_{jl}$.

It follows that in this case, i.e. when $\Sigma_{12} = 0$, it holds that

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

Consider $\mathcal{N}_3(0, \Sigma)$ with covariance matrix

$$\Sigma = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}.$$

The conditional distribution of (X_1, X_2) given X_3 has covariance matrix

$$\Sigma_{12|3} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 3 \end{pmatrix}.$$

Suppose we have $W \sim \mathcal{W}(n, \Sigma)$ with Σ as specified. Then

$$W_{12|3} = \begin{pmatrix} W_{11} - W_{33}^{-1} W_{13}^2 & W_{12} - W_{33}^{-1} W_{13} W_{23} \\ W_{21} - W_{33}^{-1} W_{21} W_{23} & W_{22} - W_{33}^{-1} W_{23}^2 \end{pmatrix} \sim \mathcal{W}(n-1, \Sigma_{12|3})$$

and independent of (W_{13}, W_{23}, W_{33}).

The conditional distribution of $(W_{13}, W_{23})^\top$ given $W_{33} = w_{33}$ is bivariate Gaussian, with mean

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix} \sigma_{33}^{-1} w_{33} = \begin{pmatrix} w_{33}/2 \\ w_{33}/2 \end{pmatrix}$$

and covariance matrix

$$w_{33} \Sigma_{12|3} = \frac{w_{33}}{2} \begin{pmatrix} 1 & 1 \\ 1 & 3 \end{pmatrix}.$$

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

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

is Wilks' distribution and denoted by $\Lambda(d, f_1, f_2)$. It holds that

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

where B_i are independent and follow Beta distributions with

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

Wilks' distribution occurs as the likelihood ratio test for independence. Consider $W \sim \mathcal{W}_d(f, \Sigma)$ and the hypothesis that $\Sigma_{12} = 0$ for a fixed block partitioning of Σ into $r \times r$, $r \times s$ and $s \times s$ matrices. The likelihood ratio statistic then becomes

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

where

$$U \sim \Lambda(r, f - s, s) = \Lambda(s, f - r, r).$$

It follows that

$$\Lambda(d, f_1, f_2) = \Lambda(f_2, f_1 + f_2 - d, d).$$

Example: the bivariate case

Consider $Z = (X, Y)^\top$ and assume $Z \sim \mathcal{N}(0, \Sigma)$ with

$$\Sigma = \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_X^2 \end{pmatrix}.$$

From data $Z_1, \dots, Z_n$, form the Wishart matrix

$$W = \begin{pmatrix} \sum_i X_i^2 & \sum_i X_i Y_i \\ \sum_i X_i Y_i & \sum_i Y_i^2 \end{pmatrix}.$$

Wilks' Λ for independence then becomes

$$\Lambda = \mathrm{LR}^{2/n} = \frac{\sum_i X_i^2 \sum_i Y_i^2 - (\sum_i X_i Y_i)^2}{\sum_i X_i^2 \sum_i Y_i^2} = 1 - R^2.$$

This is $\Lambda(1, n-1, 1)$ so $(n-1)R^2/(1-R^2) \sim F(n-1, 1)$.

Hotelling's T^2 is the equivalent of Student's t-distribution. Let $Y \sim \mathcal{N}_d(\mu, c\Sigma)$, $W \sim \mathcal{W}_d(f, \Sigma)$ with $f \geq d$, and $Y \perp\!\!\!\perp W$.

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

is known as Hotelling's T^2.

It holds that

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

and

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

where F denotes Fisher's F-distribution.

Recall that the Wishart density has the form

$$f_d(w \mid n, \Sigma) \propto (\det w)^{(n-d-1)/2} e^{-\operatorname{tr}(\Sigma^{-1}w)/2}.$$

Since the likelihood function for Σ is

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

a conjugate family of distributions for K is given by

$$\pi(K; a, \Psi) \propto (\det K)^{a/2-1} e^{-\operatorname{tr}(K\Psi)/2},$$

which thus specifies a Wishart distribution for the concentration matrix.

We then say that Σ *follows an inverse Wishart distribution* if $K = \Sigma^{-1}$ follows a Wishart distribution, formally expressed as

$$\Sigma \sim \mathcal{IW}_d(\delta, \Psi) \iff K = \Sigma^{-1} \sim \mathcal{W}_d(\delta + d - 1, \Psi^{-1}),$$

i.e. if the density of K has the form

$$f(K \mid \delta, \Psi) \propto (\det K)^{\delta/2-1} e^{-\operatorname{tr}(\Psi K)/2}.$$

We repeat the expression for the standard Wishart density:

$$f_d(w \mid n, \Sigma) \propto (\det w)^{(n-d-1)/2} e^{-\operatorname{tr}(\Sigma^{-1} w)/2}.$$

It follows that the family of inverse Wishart distributions is a conjugate family for Σ.

If the prior distribution of Σ is $\mathcal{IW}_d(\delta, \Psi)$ and $W \mid \Sigma \sim \mathcal{W}_d(n, \Sigma)$, we get for the posterior density of K that

$$
\begin{aligned}
f(K \mid \delta, \Psi, W) &\propto (\det K)^{n/2} e^{-\operatorname{tr}(KW)/2} \\
&\quad \times (\det K)^{\delta/2-1} e^{-\operatorname{tr}(\Psi K)/2} \\
&= (\det K)^{(n+\delta)/2-1} e^{-\operatorname{tr}\{(\Psi+W)K\}/2},
\end{aligned}
$$

and hence the posterior distribution is simply $\mathcal{IW}_d(\delta + n, \Psi + W) = \mathcal{IW}_d(\delta^*, \Psi^*)$.

We can thus interpret the parameter δ as a prior equivalent sample size and Ψ as the value of a matrix of sums and squares and products from a previous sample.