Inverse Wishart Distribution and Conjugate Bayesian Analysis

Steffen Lauritzen, University of Oxford

BS2 Statistical Inference, Lecture 14, Hilary Term 2008

March 2, 2008

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

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

This 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

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

Since the likelihood function for Σ is

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

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:

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

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 $W \mid \Sigma \sim \mathcal{W}_d(f, \Sigma)$, we get for the posterior density of K that

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

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

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.

We need the full form of the Wishart density for K, as constants may become important and recall that

$$\begin{aligned} f_d(K \mid \delta, \Psi) \\ &= q(d,\delta)^{-1}(\det \Psi)^{(\delta+d-1)/2}(\det K)^{\delta/2-1}e^{-\operatorname{tr}(\Psi K)/2} \end{aligned}$$

The constant $q(d,\delta)$ is

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

Consider now alternative models M_1 with Σ arbitrary and M_2 with Σ of block diagonal form:

$$\Sigma = \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}.$$

If the associated prior distributions are for M_1 that $\Sigma \sim \mathcal{IW}_d(\delta, I_d)$ and for M_2 that $\Sigma_{11} \sim \mathcal{IW}_r(\delta, I_r)$, $\Sigma_{22} \sim \mathcal{IW}_s(\delta, I_s)$, we can now calculate the Bayes factor.