The Wishart and Inverse Wishart Distributions

Steven W. Nydick

May 25, 2012

Contents

1 The Wishart Distribution

1.1 Intuitive Understanding

The Wishart distribution is the multivariate extension of the gamma distribution, although most statisticians use the Wishart distribution in the special case of integer degrees of freedom, in which case it simplifies to a multivariate generalization of the χ^2 distribution. As the χ^2 distribution describes the sums of squares of n draws from a univariate normal distribution, the Wishart distribution represents the sums of squares (and cross-products) of n draws from a multivariate normal distribution.

1.2 Mathematical Understanding

PDF Let $\mathbf{S} \sim \text{Wish}_p(\mathbf{\Sigma}, \nu)$, where $\mathbf{\Sigma}$ denotes a positive definite scale matrix (which can be thought of as a variance/covariance matrix from a multivariate normal distribution), ν is the parameter that denotes the degrees of freedom, and p indicates the dimensions of $\mathbf{S}$ (i.e., $\mathbf{S} \in \mathbb{R}^{p \times p}$). Then $\mathbf{S}$ is positive definite with probability density function (pdf)

$$f(\mathbf{S}) = \frac{|\mathbf{S}|^{\frac{\nu-p-1}{2}}}{2^{\frac{\nu p}{2}} |\mathbf{\Sigma}|^{\frac{\nu}{2}} \mathbf{\Gamma}_p\left(\frac{\nu}{2}\right)} \exp\left[-\frac{1}{2}\operatorname{tr}(\mathbf{\Sigma}^{-1}\mathbf{S})\right] \tag{1}$$

where $|\mathbf{A}|$ represents the determinant of square matrix $\mathbf{A}$, $\operatorname{tr}(\mathbf{A})$ is the trace of square matrix $\mathbf{A}$ (i.e., the sum of the diagonal elements of $\mathbf{A}$), and

$$\mathbf{\Gamma}_p(x) = \pi^{\frac{1}{2}\binom{p}{2}} \prod_{j=1}^{p} \mathbf{\Gamma}[x + (1-j)/2] \tag{2}$$

is the multivariate generalization of the gamma function, $\mathbf{\Gamma}$. Note that we must have $\nu > p - 1$ to ensure that $\mathbf{S}$ is invertible. If $\nu > p - 1$ does not hold, then $\text{Wish}_p(\mathbf{\Sigma}, \nu)$ is called a Singular Wishart distribution due to $\mathbf{\Sigma}$ being a singular matrix.

Individual Variates Similar to the χ^2 distribution, draws from a Wishart distribution represent sums of squares and not variances. Because the scale matrix ($\mathbf{\Sigma}$) can be thought of as a population variance/covariance matrix, individual draws from the Wishart distribution will often be several times the magnitude of the variance/covariance matrix. Note that if $X \sim \chi^2(n)$, then $X = \sum_{i=1}^n Z_i^2$ where $Z_1, \ldots, Z_n \overset{\text{iid}}{\sim} N(0, 1)$, so the sums of squares extension to the Wishart distribution is necessary.

Sum of Individual Variates A Wishart distribution acts like a distribution for the sums of squares and cross-products. To see the relationship between a Wishart distribution and sums of squares, let $\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_i, \ldots, \mathbf{S}_k$ be independent from k Wishart distributions where $\mathbf{S}_i \sim \text{Wish}(\mathbf{\Sigma}, \nu_i)$. Then

$$\mathbf{SS} = \sum_{i=1}^{k} \mathbf{S}_i \sim \text{Wish}\left(\mathbf{\Sigma}, \sum_{i=1}^{k} \nu_i\right) \tag{3}$$

In other words, $\mathbf{S}_i$ can be thought of as the sums of squares matrix for a sub-sample of ν_i scores from a multivariate normal distribution. And $\mathbf{SS} = \sum_{i=1}^k \mathbf{S}_i$ is the sums of squares matrix across the k sub-samples. However, when the k sub-samples are independent (all with population variance/covariance matrix $\mathbf{\Sigma}$) then the sums of squares across the k sub-samples is equivalent to the sums of squares across the $N = \sum_{i=1}^k \nu_i$ total scores.

Expected Value The expected value of $\mathbf{S}$ is

$$\text{E}(\mathbf{S}) = \nu\mathbf{\Sigma} \tag{4}$$

Also in parallel with the χ^2 distribution, the expected value of a Wishart distribution depends on number of draws one makes from the multivariate normal distribution. In comparison, the expected value of a $\chi^2(\nu)$ distribution is ν, so that the only differences between a Wishart expectation and a χ^2 expectation are the underlying dimensionality of the data and a scale component.

Variance We can find the individual variances of the elements of $\mathbf{S}$. For instance, the variance of the ij^{th} element of $\mathbf{S}$ is:

$$\text{Var}(\mathbf{S}_{ij}) = \nu(\sigma_{ij}^2 + \sigma_{ii}\sigma_{jj}) \tag{5}$$

where σ_{ij} is the ij^{th} element of the $\mathbf{\Sigma}$ matrix and can be thought of as the population covariance between variable i and variable j. Note that if $X \sim \chi^2(\nu)$, then $p = 1$, so that the only element of the variance/covariance matrix is $\sigma_{11} = \sigma_{11}^2 = 1$. Therefore, we get $\text{Var}(X) = \nu(1 + 1 \times 1) = 2\nu$, which is the familiar variance of a $\chi^2(\nu)$ variable.

Equation (5) is a set of variances rather than depicting the variance/covariance matrix because every observation of the Wishart distribution is a matrix. Therefore, describing all combinations of variances and covariance of $\mathbf{S}$ requires either an array of higher order or an outer/tensor/Kronecker operation to represent that higher order array as a matrix. The covariance matrix of $\mathbf{S}$ can be represented as[1]

$$\begin{aligned}
\text{Cov}(\mathbf{S}) &= \text{Cov}\left(\sum_{i=1}^{\nu} \mathbf{x}_i \mathbf{x}_i^T\right) \\
&= \sum_{i=1}^{\nu} \text{Cov}(\mathbf{x}_i \mathbf{x}_i^T) && \text{(by independence)} \\
&= \nu \text{Cov}(\mathbf{C}\mathbf{z}_i\mathbf{z}_i^T\mathbf{C}^T) && (6)
\end{aligned}$$

where $\mathbf{\Sigma} = \mathbf{C}\mathbf{C}^T$ is the Cholesky decomposition of the (square, symmetric) matrix $\mathbf{\Sigma}$, and $\text{E}(\mathbf{z}_i\mathbf{z}_i^T) = \mathbf{I}_p$. Next, apply the vec operator to $\mathbf{S}$, which forms a long vector by stacking the columns of $\mathbf{S}$, so that $\text{Cov}[\text{vec}(\mathbf{S})]$ is a matrix rather than an array, and letting $\mathbf{z} = \mathbf{z}_i$ (because all of the $\mathbf{z}$ vectors have the same distribution), we have

$$\begin{aligned}
\text{Cov}[\text{vec}(\mathbf{S})] &= \nu \text{Cov}[\text{vec}(\mathbf{C}\mathbf{z}\mathbf{z}^T\mathbf{C}^T)] \\
&= \nu \text{Cov}[(\mathbf{C} \otimes \mathbf{C})\text{vec}(\mathbf{z}\mathbf{z}^T)] && \text{(by the vec to Kronecker property)} \\
&= \nu (\mathbf{C} \otimes \mathbf{C})\text{Cov}[\text{vec}(\mathbf{z}\mathbf{z}^T)](\mathbf{C} \otimes \mathbf{C})^T \\
&= \nu (\mathbf{C} \otimes \mathbf{C})\text{Cov}[\mathbf{z} \otimes \mathbf{z}](\mathbf{C}^T \otimes \mathbf{C}^T) && \text{(by vec and Kronecker properties)}
\end{aligned}$$

To determine $\text{Cov}[\text{vec}(\mathbf{S})]$ (as a proxy for $\text{Cov}(\mathbf{S})$), one would only need to know $\text{Cov}[\mathbf{z} \otimes \mathbf{z}]$. Because $\mathbf{z}$ is a random vector of z-scores from a normal population, there are only five, unique terms in $\text{Cov}[\mathbf{z} \otimes \mathbf{z}]$: (1) The variance of Z_k^2 (where k is any element of $\mathbf{z}$); (2) The variance of $Z_k Z_l$ (where $k \neq l$ are any two elements of $\mathbf{z}$); (3) The covariance between $Z_k Z_l$ and $Z_l Z_k$; (4) The covariance between Z_k^2 and Z_l^2; and (5) The covariance between $Z_i Z_j$ and $Z_k Z_l$ (where at most two of i, j, k, or l are the same).

1. Z_k is standard normally distributed, so Z_k^2 follows a $\chi^2(1)$ distribution with variance equal to $2(1) = 2$. Therefore $\text{Var}(Z_k^2) = 2$ for all k.

[1]The initial idea for this derivation was presented in Eaton (2007), although in a nearly incomprehensible form

2. Z_k and Z_l are uncorrelated standard normal random variables, which implies that they are also independent. Therefore, $\text{Var}(Z_k Z_l) = \text{E}[(Z_k Z_l)^2] - \text{E}[Z_k]\text{E}[Z_l] = \text{E}(Z_k^2 Z_l^2) - 0 = \text{E}(Z_k^2)\text{E}(Z_l^2)$ due to independence, so that $\text{Var}(Z_k Z_l) = E(Z_k^2)E(Z_l^2) = 1 \times 1 = 1$ because Z_k^2 and Z_l^2 both follow a $\chi^2(1)$ distribution with expected value equal to $1(1) = 1$.

3. Z_k and Z_l are uncorrelated standard normal random variables, so $\text{Cov}(Z_k Z_l, Z_l Z_k) = \text{E}(Z_k Z_l Z_l Z_k) - \text{E}(Z_k Z_l)\text{E}(Z_l Z_k) = \text{E}(Z_k^2 Z_l^2) - \text{E}(Z_k)^2 \text{E}(Z_l)^2 = 1 - 0 = 1$ due to independence and $\chi^2(1)$ properties.

4. Z_k and Z_l are uncorrelated standard normal random variables, so that $\text{Cov}(Z_k^2, Z_l^2) = \text{E}(Z_k^2 Z_l^2) - \text{E}(Z_k^2)\text{E}(Z_l^2) = 1 - 1 = 0$ due to independence and $\chi^2(1)$ properties.

5. Because Z_i, Z_j, Z_k, and Z_l are all independent, $\text{Cov}(Z_i Z_j, Z_k Z_l) = 0$ (even if two of i, j, k, or l are the same).

Therefore, the $[p(k-1)+k, p(k-1)+k]$ elements of $\text{Cov}[\mathbf{z} \otimes \mathbf{z}]$ will all be 2 because $\text{Var}(Z_k^2) = 2$ for all k, and the remaining diagonal elements of $\text{Cov}[\mathbf{z} \otimes \mathbf{z}]$ will all be 1 because $\text{Var}(Z_k Z_l) = 1$ for all $k \neq l$. The off-diagonal elements of $\text{Cov}[\mathbf{z} \otimes \mathbf{z}]$ must be 0 except for those elements symbolizing the covariance between $Z_i Z_j$ and $Z_j Z_i$, which will be 1. Ultimately, $\text{Cov}[\mathbf{z} \otimes \mathbf{z}]$ can be written as $(\mathbf{I}_p \otimes \mathbf{I}_p + \mathbf{M}_p)$, where $\mathbf{M}_p$ is a $p^2 \times p^2$ matrix of 1s and 0s. The elements of $\text{vec}(\mathbf{M}_p)$ can be described as follows:

- A single 1 always surrounding a sequence of 0s.
- The length of a 0 sequence is one of the following two forms:
  - Define Big $= p^2 + (p-1)$.
  - Define Small $= p$.
- A set can be defined as
  - A repetition of $p-1$ Big sequences followed by
  - one Small sequence, where
  - each sequence is bracketed by 1s.

Using the above terminology, $\text{vec}(\mathbf{M}_p)$ is composed of $p-1$ sets followed by a repetition of $p-1$ Big sequences. For instance, if $p = 3$, then

$$\text{vec}(\mathbf{M}_p) = [1, 11\,0s, 1, 11\,0s, 1, 3\,0s, 1, 11\,0s, 1, 11\,0s, 1, 3\,0s, 1, 11\,0s, 1, 11\,0s, 1]$$

Therefore:

$$\begin{aligned}
\text{Cov}[\text{vec}(\mathbf{S})] &= \nu(\mathbf{C} \otimes \mathbf{C})\text{Cov}[\mathbf{z} \otimes \mathbf{z}](\mathbf{C}^T \otimes \mathbf{C}^T) \\
&= \nu(\mathbf{C} \otimes \mathbf{C})(\mathbf{I}_p \otimes \mathbf{I}_p + \mathbf{M}_p)(\mathbf{C}^T \otimes \mathbf{C}^T) \\
&= \nu[(\mathbf{C} \otimes \mathbf{C})(\mathbf{C}^T \otimes \mathbf{C}^T) + (\mathbf{C} \otimes \mathbf{C})(\mathbf{M}_p)(\mathbf{C}^T \otimes \mathbf{C}^T)] \\
&= \nu[\boldsymbol{\Sigma} \otimes \boldsymbol{\Sigma} + (\mathbf{C} \otimes \mathbf{C})(\mathbf{M}_p)(\mathbf{C}^T \otimes \mathbf{C}^T)] \qquad (7)
\end{aligned}$$

And we can check the derivation by simulating draws from a Wishart distribution and comparing the simulated covariance with the empirical covariance matrix calculated using Equation (7).

```
> # A test variance covariance matrix:
> Sig <- matrix(c(1, .7, .6,
+                 .7,  1, .4,
+                 .6, .4,  1), nrow = 3)
> p   <- dim(Sig)[1]
> df  <- 4
> ## EMPIRICAL (USING SIMULATION) ##
>
> set.seed(901254)   # for replication
> reps    <- 100000 # number of obs in our sampling dist
> W.empir <- matrix( nrow = reps, ncol = length( c(Sig) ) )
> for(i in 1:reps)
+   W.empir[i, ] <- c(rwish(v = df, S = Sig))
> ## THEORETICAL (USING EQUATION) ##
>
> # The Cholesky decomposition of Sig:
> C  <- t( chol(Sig) )
> # The strange M matrix:
> M <- matrix(c(rep(c(rep(c(1, rep(0, times = p*p+(p-1))), times = p-1),
+                     1, rep(0, times = p)),
+                   times = p-1),
+               rep(c( 1, rep(0, times = p*p+(p-1)) ), times = p-1),
+               1),
+             nrow = p^2)
> # Which can also be written as follows (when p = 3):
> # M <- matrix( c(1, rep(0, 11), 1, rep(0, 11), 1, rep(0, 3),
> #                1, rep(0, 11), 1, rep(0, 11), 1, rep(0, 3),
> #                1, rep(0, 11), 1, rep(0, 11), 1),
> #              nrow = p^2 )
>
> # And looks like the following:
> M
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
 [1,]    1    0    0    0    0    0    0    0    0
 [2,]    0    0    0    1    0    0    0    0    0
 [3,]    0    0    0    0    0    0    1    0    0
 [4,]    0    1    0    0    0    0    0    0    0
 [5,]    0    0    0    0    1    0    0    0    0
 [6,]    0    0    0    0    0    0    0    1    0
 [7,]    0    0    1    0    0    0    0    0    0
 [8,]    0    0    0    0    0    1    0    0    0
 [9,]    0    0    0    0    0    0    0    0    1
> # The function derived above (with four degrees of freedom):
```

```
> W.theor <- {df * ( kronecker(Sig, Sig) +
+                    kronecker(C, C) %*% M %*% kronecker(t(C), t(C)) ) }
> ## COMPARING VAR/COV MATRICES ##
>
> ## 1. EMPIRICAL ##
> round(var(W.empir), digits = 2)
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
 [1,] 7.99 5.62 4.76 5.62 3.95 3.34 4.76 3.34 2.86
 [2,] 5.62 6.00 3.27 6.00 5.63 3.51 3.27 3.51 1.93
 [3,] 4.76 3.27 5.41 3.27 2.24 3.74 5.41 3.74 4.76
 [4,] 5.62 6.00 3.27 6.00 5.63 3.51 3.27 3.51 1.93
 [5,] 3.95 5.63 2.24 5.63 8.02 3.19 2.24 3.19 1.30
 [6,] 3.34 3.51 3.74 3.51 3.19 4.63 3.74 4.63 3.18
 [7,] 4.76 3.27 5.41 3.27 2.24 3.74 5.41 3.74 4.76
 [8,] 3.34 3.51 3.74 3.51 3.19 4.63 3.74 4.63 3.18
 [9,] 2.86 1.93 4.76 1.93 1.30 3.18 4.76 3.18 7.95
> ## 2. THEORETICAL ##
> round(W.theor, digits = 3)
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
 [1,] 8.00 5.60 4.80 5.60 3.92 3.36 4.80 3.36 2.88
 [2,] 5.60 5.96 3.28 5.96 5.60 3.52 3.28 3.52 1.92
 [3,] 4.80 3.28 5.44 3.28 2.24 3.76 5.44 3.76 4.80
 [4,] 5.60 5.96 3.28 5.96 5.60 3.52 3.28 3.52 1.92
 [5,] 3.92 5.60 2.24 5.60 8.00 3.20 2.24 3.20 1.28
 [6,] 3.36 3.52 3.76 3.52 3.20 4.64 3.76 4.64 3.20
 [7,] 4.80 3.28 5.44 3.28 2.24 3.76 5.44 3.76 4.80
 [8,] 3.36 3.52 3.76 3.52 3.20 4.64 3.76 4.64 3.20
 [9,] 2.88 1.92 4.80 1.92 1.28 3.20 4.80 3.20 8.00
```

Notice that `reps` must be very large to be for the empirical covariance matrix to be close to the theoretical covariance matrix.

1.3 Relationship to the Normal Distribution

It is useful to diagram the explicit relationship between a multivariate normally distributed variable and the Wishart distribution. As we described earlier, the Wishart distribution can be thought of drawing sums of squares from a multivariate normal distribution with variance/covariance parameter $\mathbf{\Sigma}$. Specifically, let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ be independent and identically distributed draws from a p-dimensional multivariate normal distribution, or

$$\mathbf{x}_i \overset{\text{iid}}{\sim} N(\mathbf{0}, \mathbf{\Sigma}) \qquad\qquad i = 1, 2, \ldots, n \tag{8}$$

Then we can stack the random vectors as rows of an $n \times p$ matrix $\mathbf{X}$,

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_n^T \end{bmatrix}$$

so that the Wishart distribution describes the probability density function of the $p \times p$ matrix

$$\mathbf{S} = \mathbf{X}^T\mathbf{X} \sim \text{Wish}_p(\boldsymbol{\Sigma}, n). \tag{9}$$

Note that we used ν to indicate the general degrees of freedom parameter of the Wishart distribution, but we changed the degrees of freedom parameter to n when referring to sample size of the normal distribution (for the sake of convention).

1.4 Relationship to the χ^2 Distribution

It is also useful to diagram the relationship between the Wishart distribution and the χ^2 distribution (assuming, of course, an integer degrees of freedom, ν). Let $\mathbf{S} \sim \text{Wish}_p(\boldsymbol{\Sigma}, \nu)$. Then given any $\boldsymbol{\lambda} \in \mathbb{R}^p$, the quadratic form $\boldsymbol{\lambda}^T\mathbf{S}\boldsymbol{\lambda}$ is a scaled $\chi^2(\nu)$ variable, such that

$$\boldsymbol{\lambda}^T\mathbf{S}\boldsymbol{\lambda} \sim \boldsymbol{\lambda}^T\boldsymbol{\Sigma}\boldsymbol{\lambda} \times \chi^2(\nu). \tag{10}$$

Many people refer to $\sigma_{\boldsymbol{\lambda}} = \boldsymbol{\lambda}^T\boldsymbol{\Sigma}\boldsymbol{\lambda}$ as the scale parameter for a $\chi^2(\nu)$ distribution. And if $\boldsymbol{\Sigma}$ is an identity matrix and $\boldsymbol{\lambda}$ is a vector of 1s, then $\boldsymbol{\lambda}\boldsymbol{\Sigma}\boldsymbol{\lambda} = p$.

1.5 Application in Bayesian Statistics

The Wishart distribution is frequently used as the prior on the precision matrix parameter ($\mathbf{T} = \boldsymbol{\Sigma}^{-1}$) of a multivariate normal distribution. Because the gamma distribution is the conjugate prior for the precision parameter ($\tau = 1/\sigma^2$) of a univariate normal distribution, the Wishart distribution (as its multivariate generalization) extends conjugacy to the multivariate normal distribution. Importantly, when using the Wishart distribution as the posterior distribution for $\mathbf{T}$, then the degrees of freedom parameter (ν) represents the tuning parameter. The larger the ν, the more prior observations we are supposed to have collected, and the more the prior distribution influences the posterior distribution for $\mathbf{T}$.

Specifically, let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ be independent and identically distributed draws from a p-dimensional multivariate normal distribution with mean vector $\mathbf{0}$ and precision matrix $\mathbf{T} = \boldsymbol{\Sigma}^{-1}$, where $\boldsymbol{\Sigma}$ is the variance/covariance matrix of the multivariate normal distribution. If we put a Wishart prior distribution on the parameter $\mathbf{T}$ such that $\mathbf{T} \sim \text{Wish}_p(\boldsymbol{\Lambda}, \nu)$, then the posterior distribution of $\mathbf{T}$ (i.e., after the data have been collected) will also be Wishart distributed with $\mathbf{T}|\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]^T \sim \text{Wish}_p\big((\boldsymbol{\Lambda}^{-1} + \mathbf{S})^{-1}, \nu + n\big)$ where $\mathbf{S}$ is the sample sums of squares matrix. This derivation follows naturally from the conjugacy of the Inverse-Wishart distribution and the variance/covariance parameter of the multivariate normal distribution, as derived in Section 2.4.1. Intuitively, $\boldsymbol{\Lambda}$ is the **inverse** of the prior sums of squares matrix based off of ν observations. Thererfore, $\boldsymbol{\Lambda}^{-1}$ is the prior sums of squares matrix, so that $\boldsymbol{\Lambda}^{-1} + \mathbf{S}$ is the posterior sums of squares matrix based off of $\nu + n$ observations. But the Wishart distribution is modeling the precision matrix (and not the variance/covariance matrix), so we must re-invert the posterior sums of squares matrix to model the precision: $(\boldsymbol{\Lambda}^{-1} + \mathbf{S})^{-1}$.

1.6 An Example Using R

The following example highlights properties of the Wishart distribution.

```
> # A function to calculate the Wishart-y things:
> go.wish <- function(reps, Sig, df){
+
+ # A fancy-schmancy way of generating Wishart draws:
+   SSs <- sapply(1:reps,
+                 function(x, df, sig) rwish(df, sig)[!upper.tri(sig)],
+                 df = df, sig = Sig)
+
+ # --> The mean of the Wishart draws (as a matrix):
+   SS <- matrix( 0, nrow = nrow(Sig), ncol = ncol(Sig) ) # set matrix
+   SS[!upper.tri(SS)] <- colMeans( t(SSs) )              # fill lower-diag
+   SS <- SS + t(SS) - diag( diag(SS) )                   # fill upper-diag
+
+ # --> The variance/covariance matrix from the Wishart draws:
+   simSig <- SS / df
+
+ # --> Returning a bunch of stuffy-stuff:
+   list(simSS = SS,       simSig = simSig,
+        expSS = df * Sig, expSig = Sig)
+
+ } # END go.wish FUNCTION
> # Population-y stuff (for every run)
> nreps <- 1000                           # number of random draws
> Sigma <- matrix(c(10, 5,
+                   5,  10), nrow = 2)  # population covariance mat
> #####
> # 1 # (LOW DEGREES OF FREEDOM)
> #####
> df1 <- 3
> # Running the Wishart Function:
> set.seed(125)
> SS1 <- go.wish(reps = nreps, Sig = Sigma, df = df1)
> # --> The mean of the Wishart draws:
> SS1$simSS
         [,1]     [,2]
[1,] 29.01173 14.25635
[2,] 14.25635 27.81542
> # --> The expected value of the Wishart distribution:
> SS1$expSS
     [,1] [,2]
[1,]   30   15
[2,]   15   30
```

```
> # --> The variance/covariance matrix from the Wishart draws:
> SS1$simSig
         [,1]     [,2]
[1,] 9.670578 4.752117
[2,] 4.752117 9.271806
> # --> The original variance/covariance matrix:
> SS1$expSig
     [,1] [,2]
[1,]   10    5
[2,]    5   10
> # The estimated variance/covariance matrix is fairly similar to
> # the actual variance/covariance matrix.
>
> # What happens when we increase the degrees of freedom?
>
> #####
> # 2 # (HIGH DEGREES OF FREEDOM)
> #####
>
> df2 <- 100
> # Running the Wishart Function:
> set.seed(126)
> SS2 <- go.wish(reps = nreps, Sig = Sigma, df = df2)
> # --> The mean of the Wishart draws:
> SS2$simSS
         [,1]     [,2]
[1,] 1002.058 501.3810
[2,]  501.381 998.8175
> # --> The expected value of the Wishart distribution:
> SS2$expSS
     [,1] [,2]
[1,] 1000  500
[2,]  500 1000
> # --> The variance/covariance matrix from the Wishart draws:
> SS2$simSig
         [,1]     [,2]
[1,] 10.02058 5.013810
[2,]  5.01381 9.988175
> # --> The original variance/covariance matrix:
> SS2$expSig
     [,1] [,2]
[1,]   10    5
[2,]    5   10
```

```
>
> # The estimated sums of squares matrix is further away from
> # the actual sums of squares matrix, but the estimated
> # variance/covariance matrix is closer to the actual
> # variance/covariance matrix.
```

Notice in the above code that when $\nu = 3$ (a small degrees of freedom), the average sums of squares matrix was slightly closer to the population sums of squares matrix than when $\nu = 100$. A sum is more variable than individual scores. But the average of the sums is much higher as the degrees of freedom increase. By putting the sums of squares onto the same metric (e.g., turning the sums of squares into variances), increasing the degrees of freedom results in a tighter distribution.

2 The Inverse-Wishart Distribution

2.1 Intuitive Understanding

The Inverse-Wishart distribution is the multivariate extension of the inverse-gamma distribution (or, similar to the Wishart distribution, the inverse-χ^2 distribution in the case of integer degrees of freedom). Oddly, even though the Wishart distribution generates sums of squares matrices, one can think of the Inverse-Wishart distribution as generating random covariance matrices. However, those covariance matrices would be **inverses** of the covariance matrices generated under Wishart distribution. Therefore, the covariance matrices generated in either case (as well as the scale parameter matrix) can be thought of as (1) a covariance matrix, or (2) a precision matrix. The interpretation of the random variable depends on the research context.

2.2 Mathematical Understanding

PDF Let $\mathbf{T} \sim \text{InvWish}_p(\mathbf{\Psi}, m)$, where $\mathbf{\Psi}$ denotes a positve definite scale matrix (which can be thought of as a sums of squares matrix from a multivariate normal distribution), m is the parameter that denotes the degrees of freedom, and p indicates the dimensions of $\mathbf{T}$ (i.e., $\mathbf{T} \in \mathbb{R}^{p\times p}$). Then $\mathbf{T}$ is positive definite with probability density function (pdf)

$$f(\mathbf{T}) = \frac{|\mathbf{\Psi}|^{m/2}}{|\mathbf{T}|^{\frac{m+p+1}{2}} 2^{\frac{mp}{2}} \mathbf{\Gamma}_p\left(\frac{m}{2}\right)} \exp\left[-\frac{1}{2}\,\text{tr}(\mathbf{\Psi}\mathbf{T}^{-1})\right] \tag{11}$$

with $|\mathbf{A}|$, $\text{tr}(\mathbf{A})$, and $\mathbf{\Gamma}_p(x)$ defined as in Equation (1). Note that we must have $m > p - 1$ to ensure that $\mathbf{S}$ is invertible. If $m > p - 1$ does not hold, then we must use the Moore-Penrose inverse as our random variable and $\text{InvWish}_p(\mathbf{\Psi}, m)$ is then called a Generalized Inverse Wishart distribution. Cook and Forzani (2011) discuss the moments of the Generalized Inverse Wishart distribution (and, by extension, the Inverse Wishart distribution).

Individual Variates One can think of individual draws from the Inverse-Wishart distribution as the exact opposite as those from the Wishart distribution. In other words, comparing the pdf of both distributions, it appears as though $\mathbf{S}$ (the *observation* of the Wishart distribution) is similar to $\mathbf{\Psi}$ (the *parameter* of the Inverse-Wishart distribution) and $\mathbf{\Sigma}$ (the *parameter* of the Wishart distribution) is similar to $\mathbf{T}$ (the *observation* of the Inverse-Wishart distribution). Therefore, it is helpful to view the Inverse-Wishart distribution as taking a sums-of-squares matrix as its parameter and generating random covariance matrices, although technically, the Inverse-Wishart distribution is taking an inverse covariance matrix as its parameter and generating inverse sums-of-squares matrices. But the scales are essentially the same.

Expected Value The expected value of $\mathbf{T}$ is

$$\mathrm{E}(\mathbf{T}) = \frac{\mathbf{\Psi}}{m-p-1} \tag{12}$$

And as we described with respect to the χ^2 distribution, the only differences between the Inverse-Wishart expectation and the inverse-χ^2 expectation are the underlying dimensionality of the data and a scale component. Whereas the χ^2 distribution always has expectation, the inverse-χ^2 distribution only has calculable expectation if $m > 2$. Similarly, the Inverse-Wishart distribution has finite expectation only when $m > p - 1$.

Variance We can find the individual variances of the elements of $\mathbf{T}$. For instance, the variance of the ij^{th} element of $\mathbf{T}$ is:

$$\mathrm{Var}(\mathbf{T}_{ij}) = \frac{(m-p+1)\psi_{ij}^2 + (m-p-1)\psi_{ii}\psi_{jj}}{(m-p)(m-p-1)^2(m-p-3)} \tag{13}$$

where ψ_{ij} is the ij^{th} element of the $\mathbf{\Psi}$ matrix and can be thought of as the inverse-population covariance between variable i and variable j (or alternatively, the sum of cross-products between variable i and variable j). And if $X \sim \text{Inv-}\chi^2(m)$, then $p = 1$, so that the only element of the variance/covariance matrix is $\psi_{11} = \psi_{11}^2 = 1$. Therefore, we get $\mathrm{Var}(X) = \frac{(m)\times 1+(m-2)\times 1\times 1}{(m-1)(m-2)^2(m-4)} = \frac{2(m-1)}{(m-1)(m-2)^2(m-4)} = \frac{2}{(m-2)^2(m-4)}$, which is the variance of an inverse-$\chi^2(m)$ variable.

Similar to the Wishart distribution, we would need to use complicated linear algebra properties (including tensor/Kronecker products) to represent all of the variances and covariances of the individual Inverse-Wishart elements in a nice form.

2.3 Relationship to the Wishart Distribution

The relationship between the Inverse-Wishart distribution and other distributions can be viewed through its relationship to the Wishart distribution. As the Wishart distribution is related to the normal distribution, χ^2 distribution, and gamma distribution, the Inverse-Wishart distribution is related to those distributions in a similar way. Let $\mathbf{S} \sim \mathrm{Wish}_p(\mathbf{\Sigma}, \nu)$, as defined in Equation (1) with all of the stipulations therein. Then $\mathbf{S}^{-1} \sim \mathrm{InvWish}_p(\mathbf{\Sigma}^{-1}, m)$ where $m = \nu$ is the degrees of freedom. Note that many sources diverged on the degrees of freedom relationship between the Wishart and Inverse-Wishart distributions. Wikipedia claims that $m = \nu$, Yu (n.d.) claimed $m = \nu + p + 1$, and Lauritzen (2009) claimed $m = \nu + p - 1$. Unfortunately, the Inverse-Wishart distribution can be parameterized differently, and each parameterization results in a different degrees of freedom. The most common parameterization sets $m = \nu$, so that the Inverse-Wishart pdf is directly comparible to that of the Wishart distribution.

2.4 Application in Bayesian Statistics

The Inverse-Wishart distribution is frequently used as the prior on the variance/covariance matrix parameter ($\mathbf{\Sigma}$) of a multivariate normal distribution. Note that the inverse-gamma distribution is the conjugate prior for the variance parameter (σ^2) of a univariate normal distribution, and the Inverse-Wishart distribution (as its multivariate generalization) extends conjugacy to the multivariate normal distribution.

Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ be independent and identically distributed draws from a p-dimensional multivariate normal distribution with mean vector $\mathbf{0}$ and variance/covariance matrix $\mathbf{\Sigma}$. In other words, $\mathbf{x}_i \sim N(\mathbf{0}, \mathbf{\Sigma})$ where $i = 1, 2, \ldots, n$. If we put an Inverse-Wishart prior distribution on the parameter $\mathbf{\Sigma}$ such that $\mathbf{\Sigma} \sim \text{InvWish}_p(\mathbf{\Psi}, m)$, then the posterior distribution of $\mathbf{\Sigma}$ (i.e., after the data have been collected) will also be Inverse-Wishart distributed with $\mathbf{\Sigma}|\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]^T \sim \text{InvWish}_p(\mathbf{\Psi} + \mathbf{S}, m + n)$ where $\mathbf{S}$ is the sample sums of squares matrix. Therefore, m essentially acts as the number of observations we had observed *prior* to collecting the data, or, alternatively, the number of observations on which our prior sums-of-squares matrix $\mathbf{\Psi}$ is based.

2.4.1 Demonstration of Conjugacy

It should be fairly straightforward to demonstrate conjugacy. As was set up in the above section, let $\mathbf{x}_i$ $(i = 1, \ldots, n)$ be i.i.d. draws from a p-dimensional multivariate normal distribution with mean vector $\mathbf{0}$ and variance/covariance matrix $\mathbf{\Sigma}$. The mean vector, $\boldsymbol{\mu}$, can actually be anything, as long as it is known, and we can re-center the scores by subtracting the mean vector from each of them. Next, assuming that $\mathbf{\Sigma} \sim \text{InvWish}_p(\mathbf{\Psi}, m)$, we can write the posterior distribution as proportional to the likelihood times the prior (where f designates the data pdf and g indicates the prior pdf):

$$
\begin{aligned}
\text{Posterior} \propto\ & \text{Likelihood} \times \text{Prior} \\
&= f(\mathbf{X}|\mathbf{\Sigma}) \times g(\mathbf{\Sigma}|\mathbf{\Psi}, m) \\
&= \left[\prod_{i=1}^{n} f(\mathbf{x}_i|\mathbf{\Sigma})\right] \times g(\mathbf{\Sigma}|\mathbf{\Psi}, m) \\
&= \left[\prod_{i=1}^{n} (2\pi)^{-\frac{k}{2}} |\mathbf{\Sigma}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\mathbf{x}_i^T \mathbf{\Sigma}^{-1}\mathbf{x}_i\right]\right] \times \frac{|\mathbf{\Psi}|^{m/2}}{|\mathbf{\Sigma}|^{\frac{m+p+1}{2}} 2^{\frac{mp}{2}} \mathbf{\Gamma}_p\left(\frac{m}{2}\right)} \exp\left[-\frac{1}{2}\operatorname{tr}(\mathbf{\Psi}\mathbf{\Sigma}^{-1})\right] \\
&= (2\pi)^{-\frac{nk}{2}} |\mathbf{\Sigma}|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i^T \mathbf{\Sigma}^{-1}\mathbf{x}_i)\right] \times \frac{|\mathbf{\Psi}|^{m/2}}{|\mathbf{\Sigma}|^{\frac{m+p+1}{2}} 2^{\frac{mp}{2}} \mathbf{\Gamma}_p\left(\frac{m}{2}\right)} \exp\left[-\frac{1}{2}\operatorname{tr}(\mathbf{\Psi}\mathbf{\Sigma}^{-1})\right]
\end{aligned}
\tag{14}
$$

Note that many terms in Equation (14) are multiplicative constants and can be safely removed without affecting the *shape* of the function. Those multiplicative constants are: $(2\pi)^{-\frac{nk}{2}}$, $|\mathbf{\Psi}|^{m/2}$, $2^{\frac{mp}{2}}$, and $\mathbf{\Gamma}_p\left(\frac{m}{2}\right)$. Removing those constants and factoring, we find that

$$\text{Posterior} \propto (2\pi)^{-\frac{nk}{2}}|\mathbf{\Sigma}|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i^T\mathbf{\Sigma}^{-1}\mathbf{x}_i)\right] \times \frac{|\mathbf{\Psi}|^{m/2}}{|\mathbf{\Sigma}|^{\frac{m+p+1}{2}}2^{\frac{mp}{2}}\mathbf{\Gamma}_p\left(\frac{m}{2}\right)} \exp\left[-\frac{1}{2}\operatorname{tr}(\mathbf{\Psi}\mathbf{\Sigma}^{-1})\right]$$
$$\propto |\mathbf{\Sigma}|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i^T\mathbf{\Sigma}^{-1}\mathbf{x}_i)\right] \times |\mathbf{\Sigma}|^{-\frac{m+p+1}{2}} \exp\left[-\frac{1}{2}\operatorname{tr}(\mathbf{\Psi}\mathbf{\Sigma}^{-1})\right].$$

To finish the demonstration, we need to combine like terms and consider determinant and trace properties. First, the determinant of a product is the product of the determinants, or $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$, which allows us to combine the $|\mathbf{\Sigma}|^{-\frac{n}{2}}$ from the multivariate normal density with the $|\mathbf{\Sigma}|^{-\frac{m+p+1}{2}}$ from the Inverse-Wishart density. Second, a scalar is equal to the trace of itself, or $a = \operatorname{tr}(a)$ (where a is a scalar), which allows us to write $\sum_{i=1}^{n}(\mathbf{x}_i^T\mathbf{\Sigma}^{-1}\mathbf{x}_i) = \operatorname{tr}\left[\sum_{i=1}^{n}(\mathbf{x}_i^T\mathbf{\Sigma}^{-1}\mathbf{x}_i)\right]$. Third, the trace operator is invariant under cyclic permutation, or $\operatorname{tr}(\mathbf{AB}) = \operatorname{tr}(\mathbf{BA})$, which allows us to rotate our newly formed trace so that $\operatorname{tr}\left[\sum_{i=1}^{n}(\mathbf{x}_i^T\mathbf{\Sigma}^{-1}\mathbf{x}_i)\right] = \operatorname{tr}\left[\sum_{i=1}^{n}(\mathbf{x}_i\mathbf{x}_i^T\mathbf{\Sigma}^{-1})\right] = \operatorname{tr}\left[\left(\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^T\right)\mathbf{\Sigma}^{-1}\right] = \operatorname{tr}(\mathbf{S}\mathbf{\Sigma}^{-1})$ (where $\mathbf{S}$ is short hand for the sums of squares matrix). Finally, the sum of traces is equal to the trace of a sum, or $\operatorname{tr}(\mathbf{X}) + \operatorname{tr}(\mathbf{Y}) = \operatorname{tr}(\mathbf{X}+\mathbf{Y})$ (as long as $\mathbf{X}$ and $\mathbf{Y}$ are square matrices of the same dimension), which allows us to combine the $\operatorname{tr}(\mathbf{S}\mathbf{\Sigma}^{-1})$ (from the multivariate normal density) with the $\operatorname{tr}(\mathbf{\Psi}\mathbf{\Sigma}^{-1})$ (from the Inverse-Wishart density).

Taking into consideration all of the properties of traces and determinants, we ultimately have

$$\begin{aligned}
\text{Posterior} &\propto \text{Likelihood} \times \text{Prior} \\
&\propto |\mathbf{\Sigma}|^{-\frac{n}{2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i^T\mathbf{\Sigma}^{-1}\mathbf{x}_i)\right] \times |\mathbf{\Sigma}|^{-\frac{m+p+1}{2}} \exp\left[-\frac{1}{2}\operatorname{tr}(\mathbf{\Psi}\mathbf{\Sigma}^{-1})\right] \\
&= |\mathbf{\Sigma}|^{-\frac{n}{2}}|\mathbf{\Sigma}|^{-\frac{m+p+1}{2}} \exp\left[-\frac{1}{2}\operatorname{tr}(\mathbf{S}\mathbf{\Sigma}^{-1})\right] \exp\left[-\frac{1}{2}\operatorname{tr}(\mathbf{\Psi}\mathbf{\Sigma}^{-1})\right] \\
&= |\mathbf{\Sigma}|^{-\frac{n+m+p+1}{2}} \exp\left[-\frac{1}{2}\operatorname{tr}(\mathbf{S}\mathbf{\Sigma}^{-1}) - \frac{1}{2}\operatorname{tr}(\mathbf{\Psi}\mathbf{\Sigma}^{-1})\right] \\
&= |\mathbf{\Sigma}|^{-\frac{(n+m)+p+1}{2}} \exp\left[-\frac{1}{2}\operatorname{tr}(\mathbf{S}\mathbf{\Sigma}^{-1} + \mathbf{\Psi}\mathbf{\Sigma}^{-1})\right] \\
&= |\mathbf{\Sigma}|^{-\frac{(n+m)+p+1}{2}} \exp\left[-\frac{1}{2}\operatorname{tr}\left((\mathbf{S}+\mathbf{\Psi})\mathbf{\Sigma}^{-1}\right)\right].
\end{aligned} \tag{15}$$

Note that Equation (15) is the kernal of (i.e., proportional to) an Inverse-Gamma distribution with new parameters $\mathbf{\Psi}' = \mathbf{\Psi} + \mathbf{S}$ and $m' = n + m$. Therefore, we have shown that $\mathbf{\Sigma}|\mathbf{X} \sim \text{InvWish}_p(\mathbf{\Psi}+\mathbf{S}, m+n)$, and our demonstration is complete. ■

2.5 An Example Using R

The following example highlights properties of the Inverse Wishart distribution.

```
> # A function to calculate the Inverse-Wishart-y things:
> go.iwish <- function(reps, Psi, df){
+
```

```
+ # A fancy-schmancy way of generating Wishart draws:
+   Sigs <- sapply(1:reps,
+                  function(x, df, ss) riwish(df, ss)[!upper.tri(ss)],
+                  df = df, ss = Psi)
+
+ # --> The mean of the Inverse-Wishart draws (as a matrix):
+   Sig <- matrix( 0, nrow = nrow(Psi), ncol = ncol(Psi) ) # set matrix
+   Sig[!upper.tri(Sig)] <- colMeans( t(Sigs) )            # fill lower-diag
+   Sig <- Sig + t(Sig) - diag( diag(Sig) )                # fill upper-diag
+
+ # --> The sums-of-squares matrix from the Inverse-Wishart draws:
+   simSS <- Sig * (df - dim(Psi)[1] - 1)
+
+ # --> Returning a bunch of stuffy-stuff:
+   list(simSig = Sig,       simSS = simSS,
+        expSig = Psi / (df - dim(Psi)[1] - 1),
+        expSS  = Psi)
+
+ } # END go.iwish FUNCTION
> # Population-y stuff (for every run)
> nreps <- 1000                          # number of random draws
> Psi <- matrix(c(100, 90,
+                 90,  100), nrow = 2)  # population SS mat
> #####
> # 1 # (LOW DEGREES OF FREEDOM)
> #####
> df1 <- 6
> # Running the Inverse-Wishart Function:
> set.seed(124)
> Sig1 <- go.iwish(reps = nreps, Psi = Psi, df = df1)
> # --> The mean of the Inverse-Wishart draws:
> Sig1$simSig
         [,1]     [,2]
[1,] 32.72979 29.89009
[2,] 29.89009 33.89546
> # --> The expected value of the Inverse-Wishart distribution:
> Sig1$expSig
         [,1]     [,2]
[1,] 33.33333 30.00000
[2,] 30.00000 33.33333
> # --> The sums-of-squares matrix from the Inverse-Wishart draws:
> Sig1$simSS
         [,1]      [,2]
[1,] 98.18936  89.67026
[2,] 89.67026 101.68638
```

```
> # --> The original sums-of-squares matrix:
> Sig1$expSS
     [,1] [,2]
[1,]  100   90
[2,]   90  100
> # What happens when we increase the degrees of freedom?
>
> #####
> # 2 # (HIGH DEGREES OF FREEDOM)
> #####
>
> df2 <- 100
> # Running the Wishart Function:
> set.seed(125)
> Sig2 <- go.iwish(reps = nreps, Psi = Psi, df = df2)
> # --> The mean of the Inverse-Wishart draws:
> Sig2$simSig
          [,1]      [,2]
[1,] 1.0281028 0.9232746
[2,] 0.9232746 1.0259488
> # --> The expected value of the Inverse-Wishart distribution:
> Sig2$expSig
          [,1]      [,2]
[1,] 1.0309278 0.9278351
[2,] 0.9278351 1.0309278
> # --> The sums-of-squares matrix from the Inverse-Wishart draws:
> Sig2$simSS
         [,1]     [,2]
[1,] 99.72598 89.55764
[2,] 89.55764 99.51703
> # --> The original sums-of-squares matrix:
> Sig2$expSS
     [,1] [,2]
[1,]  100   90
[2,]   90  100
>
> # Notice how the resulting variance/covariance matrix is smaller due
> # to the extra number of observations resulting in the SAME sums-of-squares.
```

Notice in the above code that when $m = 5$ (a small degrees of freedom), the average sums of squares matrix was further away from the population sums of squares matrix than when $m = 100$. But increasing the degrees of freedom results in *not just* a tighter distribution of the variances (as was the case with the Wishart distribution) but also a smaller overall variance.

3 Graphs of the Wishart/Inverse Wishart

Because the Wishart/Inverse Wishart are distributions describing variance/covariance matrices, they are impossible to visualize in multidimensional form. However, as a simplification, we can view both distributions in univariate form via the χ^2 and inverse-χ^2 distributions.

3.1 The Wishart (err... χ^2) Distribution

Figure 1 displays the χ^2 distribution for various degrees of freedom as a simplification of a Wishart distribution. Note that as $df \to \infty$, the density shifts to the right and becomes more variable. However, if one were to put the density into the variance metric rather than the sums-of-squares metric, s/he would find that the variances were actually less variable than the sums-of-squares.

```
> x <- seq(0, 10, by = .01)
> plot(x = x, y = dchisq(x, df = 1),
+      ylim = c(0, .5),
+      xlab = expression(chi^2), ylab = "Density",
+      main = "",
+      type = "l", lwd = 2, col = "purple", axes = FALSE)
> axis(1, col = "grey")
> axis(2, col = "grey")
> points(x = x, y = dchisq(x, df = 3),
+        type = "l", lwd = 2, col = "blue")
> points(x = x, y = dchisq(x, df = 6),
+        type = "l", lwd = 2, col = "green")
> legend(x = "topright", inset = .05, bty = "n",
+        legend = c(expression(paste(chi^2, "(df = 1)", sep = "")),
+                   expression(paste(chi^2, "(df = 3)", sep = "")),
+                   expression(paste(chi^2, "(df = 6)", sep = ""))),
+        lwd = 2, col = c("purple", "blue", "green"))
```

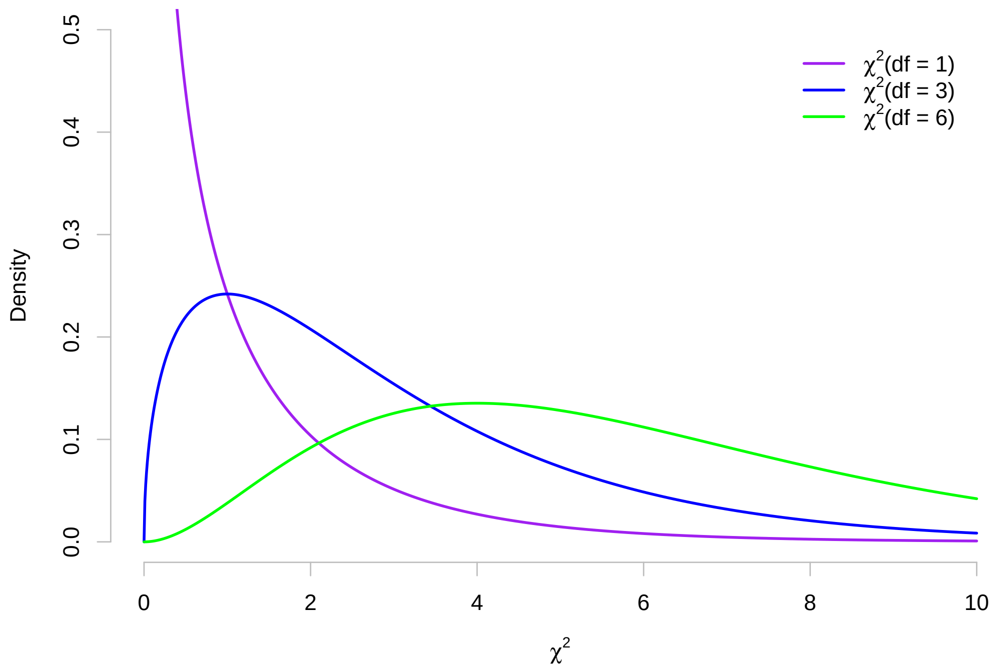

Figure 1: Several different χ^2 densities. The purple curve corresponds to χ^2 with $df = 1$; the blue curve corresponds to χ^2 with $df = 3$; and the green curve corresponds to χ^2 with $df = 6$. Note that the χ^2 distribution has a non-zero mode for $df > 2$.

3.2 The Inverse-Wishart (err... inverse-χ^2) Distribution

Figure 2 displays the inverse-χ^2 distribution for various degrees of freedom as a simplification of a Inverse-Wishart distribution. Note that as $df \to \infty$, the distribution becomes more peaked around inverse-$\chi^2 = 1$. The inverse-χ^2 distribution is in the variance metric, so by increasing the degrees of freedom (i.e., the sample size), the variance estimates become concentrated around the population variance of 1.0.

```
> x <- seq(0, 3, by = .01)
> plot(x = x, y = dinvchisq(x, df = 1, scale = 1),
+      ylim = c(0, 1),
+      xlab = expression(paste("Inverse-", chi^2, sep = "")), ylab = "Density",
+      main = "",
+      type = "l", lwd = 2, col = "darkmagenta", axes = FALSE)
> axis(1, col = "grey")
> axis(2, col = "grey")
> points(x = x, y = dinvchisq(x, df = 3, scale = 1),
+         type = "l", lwd = 2, col = "navy")
> points(x = x, y = dinvchisq(x, df = 6, scale = 1),
+         type = "l", lwd = 2, col = "green4")
```

```
> legend(x = "topright", inset = .05, bty = "n",
+        legend = c(expression(paste("Inv-", chi^2, "(df = 1)", sep = "")),
+                   expression(paste("Inv-", chi^2, "(df = 3)", sep = "")),
+                   expression(paste("Inf-", chi^2, "(df = 6)", sep = ""))),
+        lwd = 2, col = c("darkmagenta", "navy", "green4"))
```

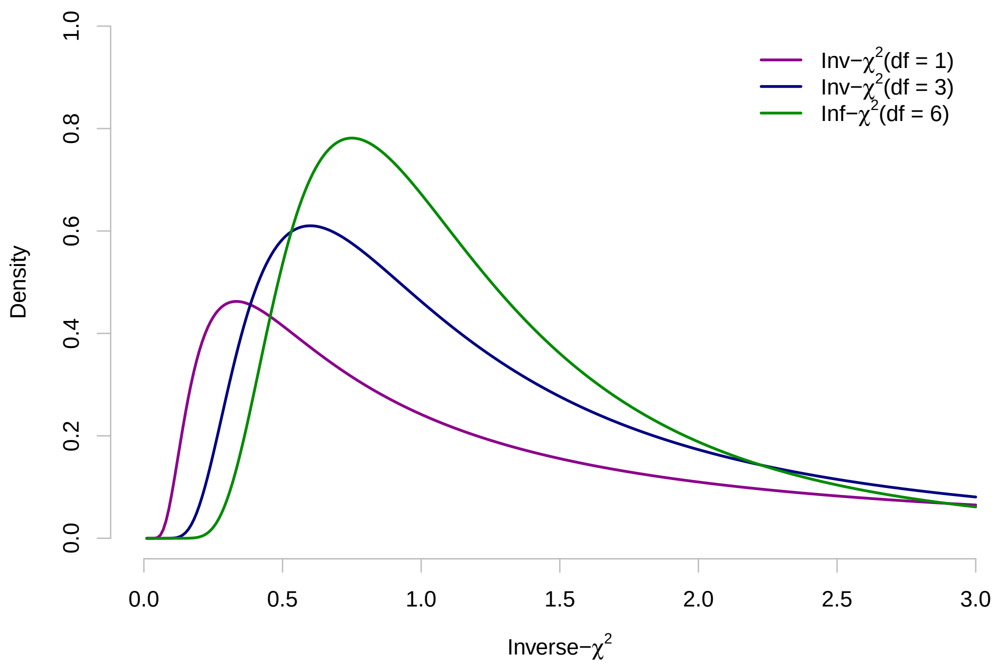

Figure 2: Several different inverse-χ^2 densities. The magenta curve corresponds to inverse-χ^2 with $df = 1$; the navy curve corresponds to inverse-χ^2 with $df = 3$; and the dark green curve corresponds to inverse-χ^2 with $df = 6$.

References

[1] Cook, R. D. (2011). On the mean and variance of the generalized inverse of a singular Wishart matrix (2011). *Electronic Journal of Statistic*, *5*, 146–158.

[2] Eaton, M. L. (2007) Chapter 8: The Wishart distribution. In M. L. Eaton, *Multivariate Statistics: A Vector Space Approach* (pp. 302–333). OH: Institute of Mathematical Statistics.

[3] Lauritzen, S. (2009). *Wishart and inverse Wishart distributions.* [LaTeX Slides] Retrieved from: http://www.stats.ox.ac.uk/~steffen/teaching/bs2HT9/inverse.pdf

[4] Yu, K. (n.d.). A note on the inverted Wishart distribution. Retrieved from: http://www.dbs.informatik.uni-muenchen.de/~yu_k/wishart.pdf