

# The Matrix-variate F prior

Matt Simpson

February 11, 2014

## 1 Mixing Wisharts and inverse Wisharts

The main goal of this section is to make precise and prove the statement “an inverse Wishart mixture of Wisharts is the same as a Wishart mixture of inverse Wisharts” as well as describe some of the properties of the resulting class of distributions. We will work with a slightly more general class of distribution in order to somewhat simplify the mathematics – the (inverse) matrix gamma distribution. Let  $X$  be a  $p \times p$  nonnegative definite random matrix with a matrix gamma distribution given shape and scale parameters  $\alpha > (p-1)/2$  and  $\beta > 0$ , and scale matrix parameter  $\Sigma$ , a  $p \times p$  positive definite matrix, i.e.  $X \sim MG_p(\alpha, \beta, \Sigma)$ . Then the density of  $X$  is

$$p(X) = \frac{|\Sigma|^{-\alpha}}{\beta^{p\alpha} \Gamma_p(\alpha)} |X|^{\alpha-(p+1)/2} \exp \left[ \text{tr} \left( -\frac{1}{\beta} \Sigma^{-1} X \right) \right]$$

Where  $|\cdot|$  denotes the determinant operator,  $\text{tr}(\cdot)$  the trace operator, and  $\Gamma_p(\cdot)$  is the multivariate gamma function.  $Y = X^{-1}$  has an inverse matrix gamma distribution,  $Y \sim IMG_p(\alpha, \beta, \Psi)$  with  $\Psi = \Sigma^{-1}$ , and the density of  $Y$  is

$$p(Y) = \frac{|\Psi|^\alpha}{\beta^{p\alpha} \Gamma_p(\alpha)} |Y|^{-\alpha-(p+1)/2} \exp \left[ \text{tr} \left( -\frac{1}{\beta} \Psi Y^{-1} \right) \right].$$

The (inverse) Wishart distribution is a special case with  $\alpha = n/2$  and  $\beta = 0$ , where  $n$  is the degrees of freedom parameter of the (inverse) Wishart distribution. Now we can state the two main theorems of this section.

**Theorem 1.0.1.** *Suppose  $X|Y \sim MG_p(\alpha_1, \beta_1, Y)$  and  $Y \sim IMG_p(\alpha_2, \alpha_2, \Sigma)$  Then the marginal distribution of  $X$  is*

$$p(X) = \frac{\left| \frac{\beta_1}{\beta_2} \Sigma \right|^{\alpha_2}}{B_p(\alpha_1, \alpha_2)} |X|^{\alpha_1-(p+1)/2} \left| X + \frac{\beta_1}{\beta_2} \Sigma \right|^{-(\alpha_1+\alpha_2)}$$

**Theorem 1.0.2.** *Suppose  $X|Y \sim IMG_p(\alpha_2, \beta_2, Y)$  and  $Y \sim MG_p(\alpha_1, \beta_1, \Sigma)$  Then the marginal distribution of  $X$  is*

$$p(X) = \frac{\left| \frac{\beta_1}{\beta_2} \Sigma \right|^{\alpha_2}}{B_p(\alpha_1, \alpha_2)} |X|^{\alpha_1-(p+1)/2} \left| X + \frac{\beta_1}{\beta_2} \Sigma \right|^{-(\alpha_1+\alpha_2)}$$

where  $B_p(\alpha_1, \alpha_2)$  is the multivariate beta function with the property  $B_p(\alpha_1, \alpha_2) = \Gamma_p(\alpha_1) \Gamma_p(\alpha_2) / \Gamma_p(\alpha_1 + \alpha_2)$ .

The proofs of these theorems are actually quite simple. We'll start with Theorem 1.0.1. The joint distribution of  $(X, Y)$  can be written as

$$p(X, Y) = \frac{|\Sigma|^{\alpha_2} \Gamma_p(\alpha_1 + \alpha_2)}{(\beta_1^{\alpha_1} \beta_2^{\alpha_2})^p \Gamma_p(\alpha_1) \Gamma_p(\alpha_2)} \left| \frac{1}{\beta_1} X + \frac{1}{\beta_2} \Sigma \right|^{-(\alpha_1 + \alpha_2)} |X|^{\alpha_1 - (p+1)/2} \\ \times \frac{\left| \frac{1}{\beta_1} X + \frac{1}{\beta_2} \Sigma \right|^{\alpha_1 + \alpha_2}}{\Gamma_p(\alpha_1 + \alpha_2)} |Y|^{-(\alpha_1 + \alpha_2) - (p+1)/2} \exp \left[ \text{tr} \left( - \left( \frac{1}{\beta_1} X + \frac{1}{\beta_2} \Sigma \right) Y^{-1} \right) \right].$$

The second line is the density of an inverse matrix gamma distribution, so marginalizing out  $Y$  and rearranging a bit we get

$$p(X) = \frac{\left| \frac{\beta_1}{\beta_2} \Sigma \right|^{\alpha_1}}{B_p(\alpha_1, \alpha_2)} |X|^{\alpha_1 - (p+1)/2} \left| X + \frac{\beta_1}{\beta_2} \Sigma \right|^{-(\alpha_1 + \alpha_2)}$$

which completes the proof.

To prove Theorem 1.0.2 we'll do something similar but ignore the normalizing constant. In this case the joint distribution of  $X$  and  $Y$  can be written as

$$P(X, Y) \propto |X|^{-\alpha_2 - (p+1)/2} |Y|^{\alpha_1 - (p+1)/2} \exp \left[ \text{tr} \left( - \left( \frac{1}{\beta_2} X^{-1} + \frac{1}{\beta_1} \Sigma^{-1} \right) Y \right) \right].$$

Now we marginalize out  $Y$  to get the desired result:

$$p(X) \propto |X|^{-\alpha_2 - (p+1)/2} \left| \frac{1}{\beta_2} X^{-1} + \frac{1}{\beta_1} \Sigma^{-1} \right|^{-(\alpha_1 + \alpha_2)} \\ \propto |X|^{\alpha_1 - (p+1)/2} \left| \frac{\beta_1}{\beta_2} \Sigma + X \right|^{-(\alpha_1 + \alpha_2)}$$

where the last line comes from  $|X \beta_1 \Sigma|^{\alpha_1 + \alpha_2} |X \beta_1 \Sigma|^{-(\alpha_1 + \alpha_2)} = 1$ .

When  $\alpha_1 = 2n_1$ ,  $\alpha_2 = 2n_2$  and  $\beta_1 = \beta_2$  (or both  $\beta$ 's are absorbed into  $\Sigma$ ),  $X$  has the multivariate (or matrix-variate) F distribution ( $X \sim F_p(n_1, n_2, \Sigma)$ ) with density

$$p(X) \propto |X|^{\frac{n_1 - (p+1)}{2}} |I_p + \Sigma^{-1} X|^{-(n_1 + n_2)/2} p$$

where  $I_p$  is the  $p \times p$  identity matrix. We'll focus on the multivariate F distribution.

## 2 Properties of the multivariate F distribution

It is commonly observed that the multivariate F distribution can be seen as a sort of generalization of the Wishart distribution. The following corollary illustrates this.

**Corollary 2.0.3.** *Suppose for  $i = 1, 2, \dots, n_1$ ,  $x_i$  is a  $p$ -dimensional random vector with*

$$x_i \stackrel{iid}{\sim} T_p(n_2 - p + 1, 0, \Sigma / (n_2 - p + 1))$$

*where  $T_p(\nu, \mu, \Omega)$  is the multivariate  $T$  distribution with degrees of freedom  $\nu$ ,  $p \times 1$  location parameter  $\mu$  and  $p \times p$  scale matrix  $\Omega$ . Suppose  $n_1, n_2 > p - 1$  and let  $S = \sum_{i=1}^{n_1} x_i x_i'$ . Then  $S \sim F_p(n_1, n_2, \Sigma)$*

To see this, recall that if  $x|Y \sim N_p(0, Y)$  and  $Y \sim IW_p(\nu, \Sigma)$  then marginally  $x \sim T_p(\nu - p + 1, 0, \Sigma / (\nu - p + 1))$ . So we can write the distribution of the  $x_i$ 's as independent inverse Wishart mixtures of independent normals, i.e.  $x_i|Y_1, Y_2, \dots, Y_{n_1} \stackrel{ind}{\sim} N_p(0, Y_i)$  and  $Y_i \stackrel{iid}{\sim} IW_p(n_2, \Sigma)$ . But conditional on  $Y$ , we know  $S \sim W_p(n_1, Y)$ . So by Theorem 1.0.1  $S \sim F_p(n_1, n_2, \Sigma)$ .

**Proposition 2.0.4.** *If  $X \sim F_p(n_1, n_2, \Sigma)$  then the marginal distribution of  $s_{ii} = \sqrt{x_{ii}}$  is the “scale mixed Nakagami” distribution with density*

$$p(s) \propto s^{n_2} \left( 1 + \frac{1}{\nu} \frac{s}{\sigma/\nu} \right)^{-(\nu+1)/2}$$

where  $\nu = n_1 + n_2 - p$

First, write the distribution of  $X$  as  $X|Y \sim IW_p(n_1, Y)$  and  $Y \sim W_p(1, \Sigma)$ . Recall that conditional on  $Y$ , the marginal distribution of  $x_{ii}$  is  $IG((n_1 - p + 1)/2, y_{ii}/2)$  and that the marginal distribution of  $y_{ii}$  is  $G(n_2/2, 1/(2\sigma_{ii}))$ . Put together (and omitting the subscripts of  $x_{ii}$ ,  $y_{ii}$  and  $\sigma_{ii}$ ) we have

$$p(x, y) \propto x^{-(n_1-p+1)/2-1} y^{(n_1+n_2-p+1)/2-1} \exp \left[ -y \frac{1}{2} \left( \frac{1}{x} + \frac{1}{\sigma} \right) \right]$$

so that

$$\begin{aligned} p(x) &\propto x^{-(n_1-p+1)/2-1} \left( \frac{1}{\sigma} + \frac{1}{x} \right)^{-(n_1+n_2-p-1)/2-1} \\ &\propto x^{n_2/2-1} (1 + x/\sigma)^{-(n_1+n_2-p-1)/2-1} \end{aligned}$$

Now let  $s = \sqrt{x}$ . Then the Jacobian is  $2s$  and the density of  $s$  can be written as

$$\begin{aligned} p(s) &\propto s^{n_2} \left( 1 + \frac{s^2}{\sigma} \right)^{(n_1+n_2-p+1)/2} \\ &\propto s^{n_2} \left( 1 + \frac{1}{\nu} \frac{s}{\sigma/\nu} \right)^{-(\nu+1)/2} \end{aligned}$$

where  $\nu = n_1 + n_2 - p$ . I’m calling this the scale mixed Nakagami distribution because it is what results when you put an inverse gamma mixing distribution on the scale parameter of a Nakagami distribution. Similarly, the Nakagami distribution is a Gaussian kernel of the variable  $s$  multiplied by  $s^\alpha$  where  $\alpha > 0$ . Here we use a  $T$  kernel multiplied by  $s^\alpha$ .

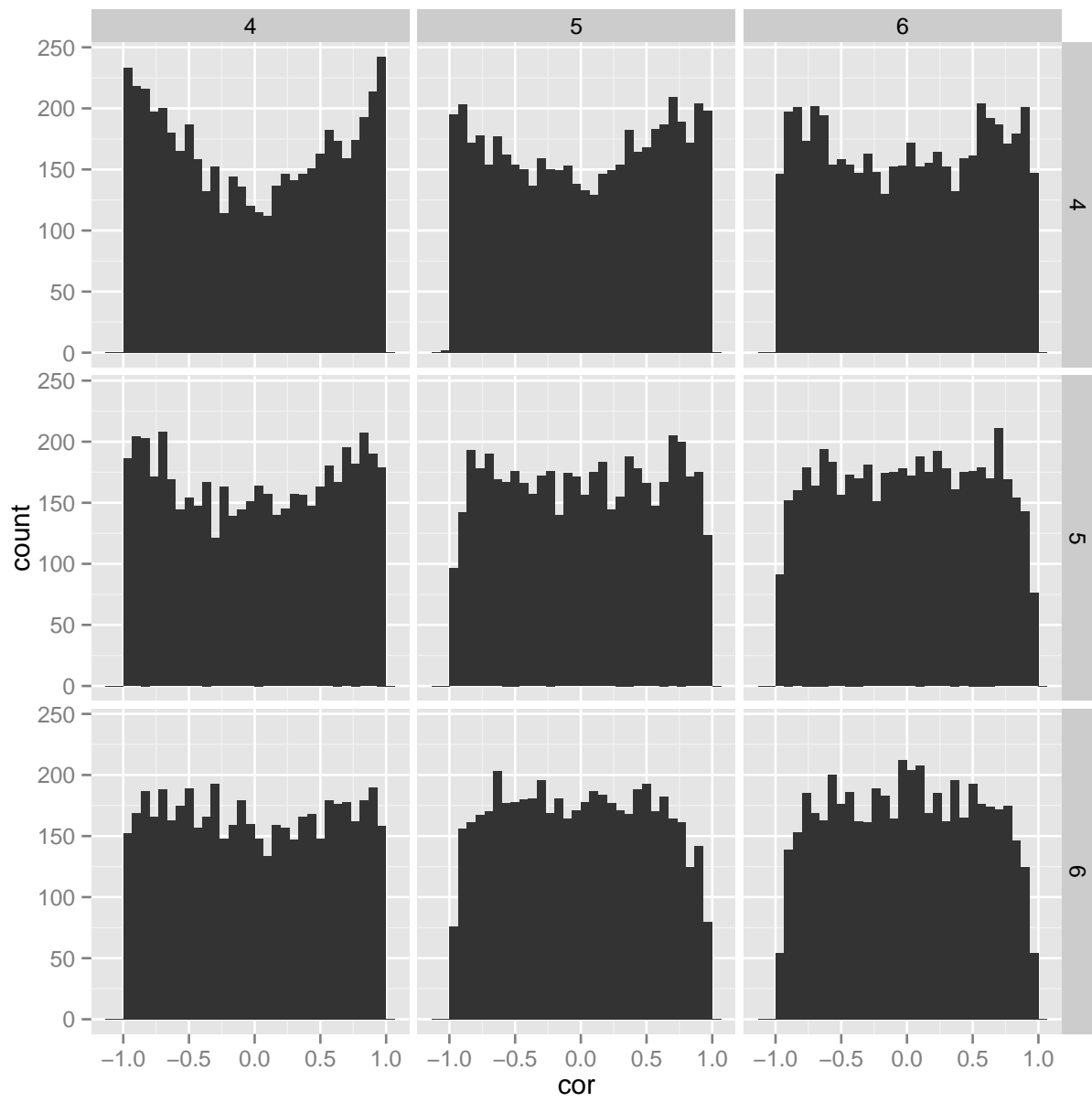
Because the inverse Wishart distribution is restricted to  $n_1 > (p - 1)$ , in general we can’t enforce half- $t$  distributions on the standard deviations by setting  $n_1 = 1$ . However, this distribution should probably satisfy ?’s main complaint about sensitivity to hyperparameter choice.

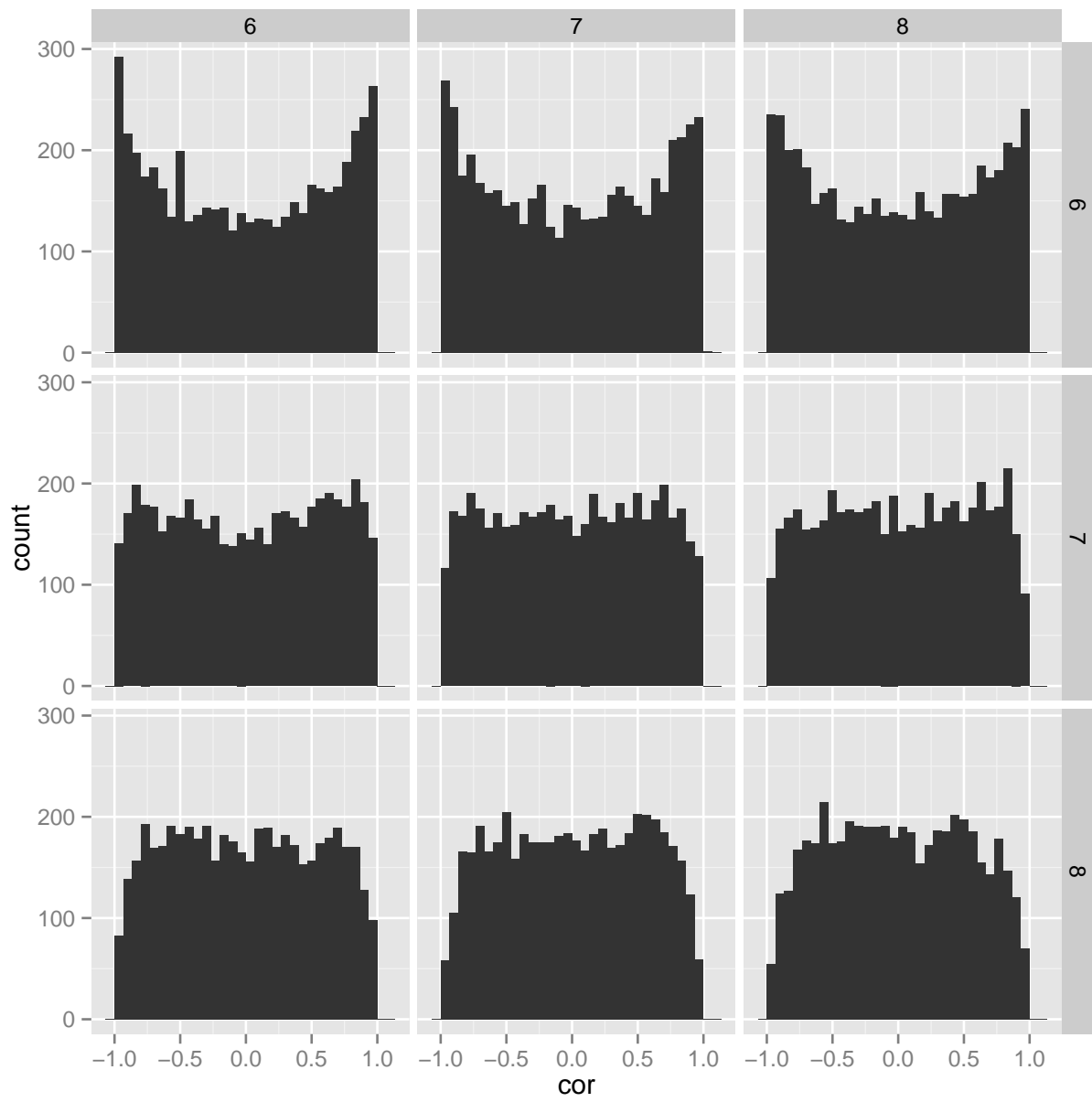
*LOOK AT WHAT HAPPENS WHEN EITHER  $n_1$  OR  $n_2$  GO TO  $\infty$ . WHEN  $n_2 \rightarrow \infty$  WE GET THE WISHART. SUSPECT WHEN  $n_2 \rightarrow \infty$  WE GET THE INVERSE WISHART*

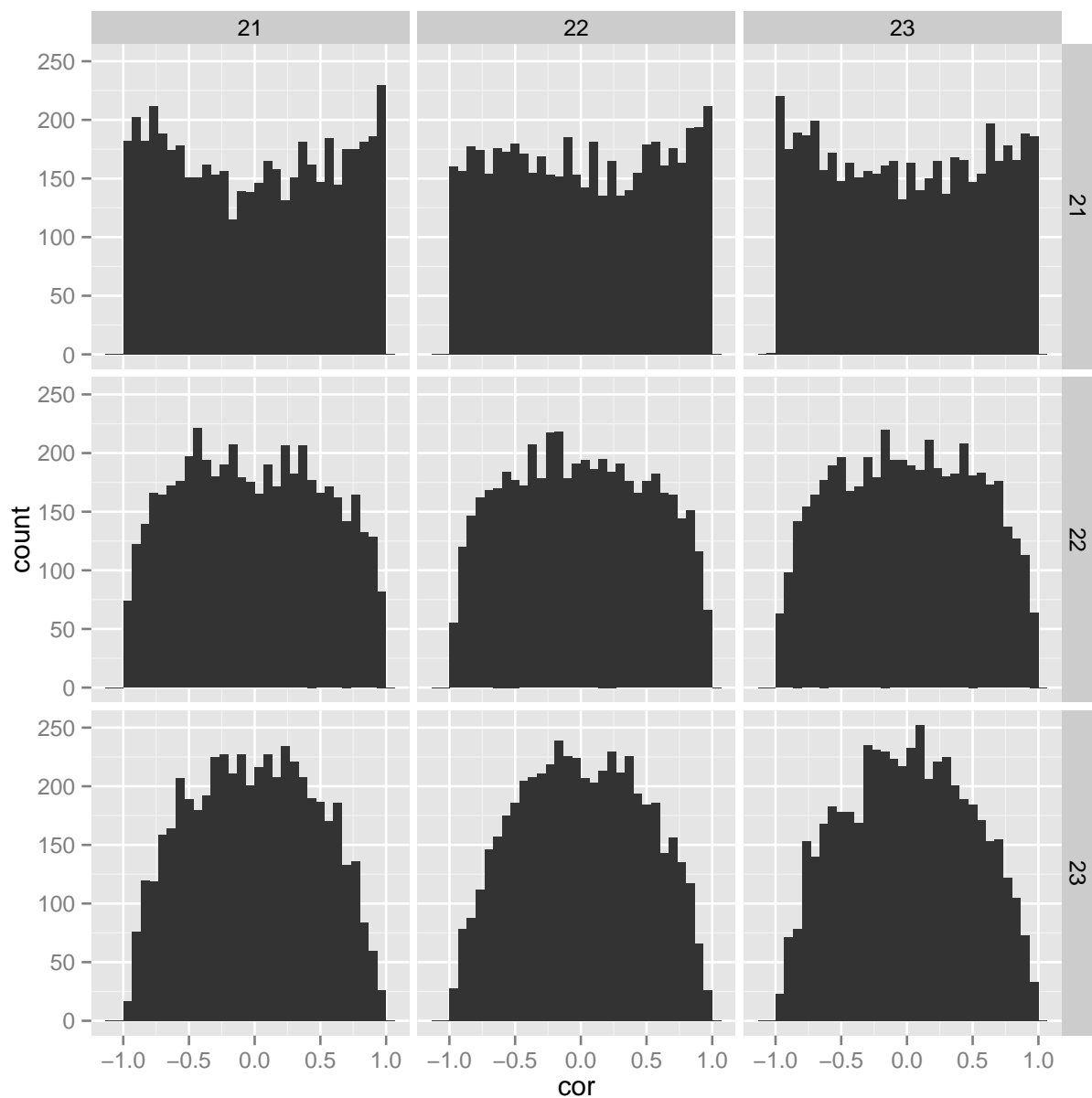
### 3 Examining the correlations

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```

```
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
## stat_bin: binwidth defaulted to range/30. Use 'binwidth = x' to adjust this.
```







## 4 Discussion

The significance of this distribution and its mixture representations is hopefully clear — the matrixvariate F distribution provides a much more general class of prior distributions for both covariance and precision matrices than typically used while still allowing computation to be cheap and easily implemented. The main cost here is an additional Gibbs step where we draw from an (inverse) Wishart distribution for the scale matrix in the prior for the covariance (precision) matrix. A similar idea has been put forth by ?. Their suggestion is to use an inverse Wishart prior for the covariance matrix, but force scale matrix parameter to be diagonal. Then put independent gamma hyperpriors on the diagonal elements of the scale matrix. They note that this allows for a more flexible covariance matrix prior, including marginal half-t distributions for standard deviations and, if desired, marginal uniform distributions for the correlations.