

A note on priors

Matt Simpson

June 24, 2014

1 Introduction

This note shows that depending on how you parameterize a DLM, a different natural conjugate prior results for each of the covariance matrices. However, there is a unifying framework from which to view each of these priors.

2 The Model

Consider the standard dynamic linear model with time varying variances. Let y_t denote a $p \times 1$ vector of data for $t = 1, 2, \dots, T$ and θ_t denote a $k \times 1$ vector of unobserved latent states for $t = 0, 1, \dots, T$. Then the DLM can be written as:

$$y_t = F_t \theta_t + v_t \quad (1)$$

$$\theta_t = G_t \theta_{t-1} + w_t \quad (2)$$

where F_t is a $k \times p$ matrix and G_t is a $k \times k$ matrix for $t = 1, 2, \dots, T$ and for $t = 1, 2, \dots, T$

$$\begin{bmatrix} v_t \\ w_t \end{bmatrix} \stackrel{\text{ind}}{\sim} N \left(\begin{bmatrix} \mathbf{0}_{p \times 1} \\ \mathbf{0}_{k \times 1} \end{bmatrix}, \begin{bmatrix} V_t & \mathbf{0}_{p \times k} \\ \mathbf{0}_{k \times p} & W_t \end{bmatrix} \right).$$

The covariance matrices, $V_{1:T}$ and $W_{1:T}$, are the unknown parameters in addition to any unknown parameters contained in $F_{1:T}$ or $G_{1:T}$. For simplicity, we'll assume that these are matrices of constants. Equation (??) is called the observation equation and equation (??) is called the system equation. Similarly v_t 's are called the observation errors and the V_t 's are called the observation covariances while the w_t 's are called the system disturbances and the W_t 's are called the system covariances. The latent states, $\theta_{0:T}$, are a data augmentation (DA) that are used in the MCMC sampling algorithm in order to draw from the posterior distribution of $(V_{1:T}, W_{1:T})$ though, often, we're interested in the posterior distribution of $\theta_{0:T}$ as well.

To complete the model, we need priors on θ_0 , $V_{1:T}$ and $W_{1:T}$. The standard priors are the conditionally conjugate priors, i.e.

$$p(\theta_0, V_{1:T}, W_{1:T}) = p(\theta_0) \prod_{t=1}^T p(V_t) p(W_t)$$

with

$$\begin{aligned} \theta_0 &\sim N(m_0, C_0) \\ V_t &\sim IW(\Psi_t, \eta_t) \text{ for } t = 1, 2, \dots, T \\ W_t &\sim IW(\Omega_t, \delta_t) \text{ for } t = 1, 2, \dots, T \end{aligned}$$

for some fixed hyperparameters m_0 & C_0 and Ψ_t , η_t , Ω_t , & δ_t for $t = 1, 2, \dots, T$. Using these priors, we can use FFBS in order to sample the latent states conditional on $V_{1:T}$ and $W_{1:T}$, and each V_t and W_t has independent inverse Wishart full conditional distributions. ?, ?, and ? contain the details of FFBS and the rest of this algorithm.

To complete the model we need a prior for (θ_0, V, W) . Sylvia's suggestion is to use what amounts to a $G\left(\frac{1}{2}, \frac{1}{2Q}\right)$ prior for each of V and W with Q chosen appropriately, but to do this by specifying normal priors on $\pm\sqrt{V}$ and $\pm\sqrt{W}$.

Let's ignore the dynamic setting for a moment and essentially impose that $W = 0$ so that the model is just

$$y_t \stackrel{iid}{\sim} N(\theta, V).$$

So now we need a prior for (θ, V) . We'll suppose θ is known to simplify even more, then suppose that V has the $G(1/2, 1/(2Q))$ prior. This means that the prior density of V is

$$p(V) \propto V^{-1/2} e^{-V/(2Q)}.$$

Now let $\tau = \pm\sqrt{V}$ and suppose that a priori $\tau \sim N(0, Q)$. Then

$$p(\tau) \propto e^{-\tau^2/(2Q)}.$$

If we transform from τ to V , $V = \tau^2$ and the jacobian is the same for both the positive and negative sides of τ :

$$|J| = V^{-1/2}/2$$

which gives

$$p(V) \propto V^{-1/2} e^{-V/(2Q)}.$$

So a $N(0, Q)$ prior on τ and a $G(1/2, 1/(2Q))$ prior on V are equivalent.

Given this prior we can write the posterior distribution of τ as

$$\begin{aligned} p(\tau|y_{1:T}) &\propto \tau^{-T} e^{\sum_{t=1}^T (y_t - \theta)^2 / (2\tau^2)} e^{-\tau^2 / (2Q)} \\ &\propto \tau^{-T} e^{-\frac{1}{2}(\sum_{t=1}^T (y_t - \theta)^2 / \tau^2 + \frac{1}{Q}\tau^2)} \end{aligned}$$

or the posterior distribution of V as

$$\begin{aligned} p(V|y_{1:T}) &\propto V^{-(T+1)/2} e^{\sum_{t=1}^T (y_t - \theta)^2 / (2V)} e^{-V / (2Q)} \\ &\propto V^{-(T+1)/2} e^{-\frac{1}{2}(\sum_{t=1}^T (y_t - \theta)^2 / V + \frac{1}{Q}V)}. \end{aligned}$$

The distribution of V is an instance of the generalized inverse gaussian distribution, or GIG. A random variable $x > 0$ has a $GIG(a, b, p)$ distribution if the density of x is given by

$$p(x) \propto x^{p-1} e^{-(ax+b/x)/2}$$

for $a \geq 0$, $b \geq 0$, and p real valued. The gamma and inverse gamma distributions (among others) are special cases.

If we treat y_t as a vector and thus V as a matrix, let $CC' = V$ i.e. C is the lower triangular cholesky decomposition of V . Then in a similar fashion you can put a normal prior on each of the components of C (treating it as $\pm chol(V)$). Things get complicated here, but in a different setting this ends up being a pretty convenient prior for the variance of a hierarchical distribution for regression parameters, and in the Gibbs sampler C ends up being drawn from a normal distribution. We'll ignore that for now and go back to the LLM.

Suppose now that $V = \tau^2$ and $W = \sigma^2$ with $(\tau, \sigma) \sim N(0, Q)N(0, R)$, i.e. independent mean zero normal distributions with potentially different variances. We'll assume that θ_0 is independent of (τ, σ) in the prior with $\theta_0 \sim N(m_0, C_0)$. Now we can write the conditional posteriors of τ and σ as

$$\begin{aligned} p(\tau|\theta_{0:T}, \dots) &\propto \tau^{-T} e^{-(\sum_{t=1}^T (y_t - \theta_t)^2 / \tau^2 + \frac{1}{Q} \tau^2) / 2} \\ p(\sigma|\theta_{0:T}, \dots) &\propto \sigma^{-T} e^{-(\sum_{t=1}^T (\theta_{t-1} - \theta_t)^2 / \sigma^2 + \frac{1}{R} \sigma^2) / 2}. \end{aligned}$$

Suppose now we reparameterize in terms of $\gamma_t = (\theta_t - \theta_{t-1}) / \sqrt{W} = (\theta_t - \theta_{t-1}) / \sigma$ for $t = 1, \dots, T$ and $\gamma_0 = \theta_0$. The determinant of the jacobian of this transformation ends up being $W^{T/2}$. Then we have the conditional posteriors given $\gamma_{0:T}$ instead of $\theta_{0:T}$:

$$\begin{aligned} p(\tau|\gamma_{0:T}, \dots) &\propto \tau^{-T} e^{-(\sum_{t=1}^T (y_t - \sigma \sum_{s=1}^t \gamma_s - \gamma_0)^2 / \tau^2 + \frac{1}{Q} \tau^2) / 2} \\ p(\sigma|\gamma_{0:T}, \dots) &\propto e^{-(\sum_{t=1}^T (y_t - \sigma \sum_{s=1}^t \gamma_s - \gamma_0)^2 / \tau^2 + \frac{1}{R} \sigma^2) / 2}. \end{aligned}$$

So τ still has a GIG distribution, but notice that σ now has a normal distribution with variance $\hat{\Sigma} = \left(\frac{1}{\tau^2} \sum_{t=1}^T \left(\sum_{s=1}^t \gamma_s \right)^2 + \frac{1}{R} \right)^{-1}$ and mean $\hat{\mu} = \hat{\Sigma} \frac{1}{\tau^2} \sum_{t=1}^T (y_t - \gamma_0) \sum_{s=1}^t \gamma_s$. This means no awful rejection sampling in order to draw $W|\gamma_{0:T}, \dots$ or $V|\psi_{0:T}, \dots$.

3 Attempts to generalize

Computationally cheap MCMC is a good reason to use this prior provided that it's flexible enough for the purpose of inference. However, note that the shape parameter of the implied gamma distribution on the variance is held fixed in this family of probability distributions. In other words, we're assuming that $V \sim G(\alpha, \beta)$ with $\beta = 1/(2Q)$ and α fixed at $\alpha = 1/2$. The question is, then, is there a prior on $\tau = \pm\sqrt{V}$ that implies the $G(\alpha, 1/(2Q))$ prior on V ? There is, and it looks like this:

$$p(\tau) \propto (\tau^2)^{\alpha-1/2} \exp[-\beta\tau^2].$$

With $\beta = 1/(2Q)$. To see that this results in a $G(\alpha, 1/(2Q))$ prior on $V = \tau^2$, consider the transformation. The jacobian is the same as before, giving

$$p(V) \propto V^{\alpha-1} \exp[-\beta V].$$

I'm calling the distribution of τ the "signed Nakagami distribution" for reasons that will be clear shortly. The normalizing constant of the signed Nakagami distribution is

$$\begin{aligned} \int_{-\infty}^{\infty} (\tau^2)^{\alpha-1/2} \exp[-\beta\tau^2] d\tau &= \int_0^{\infty} y^{\alpha-1} \exp[-\beta V] V \\ &= \Gamma(\alpha) / \beta^\alpha \end{aligned}$$

so

$$\begin{aligned} p(\tau) &= \frac{\beta^\alpha}{\Gamma(\alpha)} (\tau^2)^{\alpha-1/2} \exp[-\beta\tau^2]. \\ &= \frac{1}{\Gamma(\alpha)(2Q)^\alpha} (\tau^2)^{\alpha-1/2} \exp\left[-\frac{1}{2Q}\tau^2\right]. \end{aligned}$$

Then the CDF of the signed Nakagami distribution is:

$$P(\tau \leq x) = \frac{1 + \text{sgn}(x)F(x^2; \alpha, \beta)}{2}$$

where $F(x; \alpha, \beta)$ is the CDF of the gamma distribution with shape parameter α and rate parameter β . To see this, note that if $x < 0$ then

$$\begin{aligned} P(\tau \leq x) &= \int_{-\infty}^x \frac{\beta^\alpha}{\Gamma(\alpha)} (\tau^2)^{\alpha-1/2} \exp[-\beta\tau^2] d\tau \\ &= -\frac{1}{2} \int_{\infty}^{x^2} \frac{\beta^\alpha}{\Gamma(\alpha)} (\tau^2)^{\alpha-1/2} \exp[-\beta\tau] d\tau \\ &= \frac{1 - F(x^2; \alpha, \beta)}{2}. \end{aligned}$$

Now since the signed Nakagami distribution is clearly symmetric around 0, $P(\tau \leq 0) = 1/2$ and thus if $x > 0$

$$\begin{aligned} P(\tau \leq x) &= \int_{-\infty}^x \frac{\beta^\alpha}{\Gamma(\alpha)} (\tau^2)^{\alpha-1/2} \exp[-\beta\tau^2] d\tau \\ &= \frac{1}{2} + \int_0^x \frac{\beta^\alpha}{\Gamma(\alpha)} (\tau^2)^{\alpha-1/2} \exp[-\beta\tau^2] d\tau \\ &= \frac{1}{2} + \frac{1}{2} \int_0^{x^2} \frac{\beta^\alpha}{\Gamma(\alpha)} (\tau^2)^{\alpha-1/2} \exp[-\beta\tau] d\tau \\ &= \frac{1 + F(x^2; \alpha, \beta)}{2}. \end{aligned}$$

Now the Nakagami distribution has pdf

$$p(x) = \frac{2\beta^\alpha}{\Gamma(\alpha)} x^{2\alpha-1} \exp[-\beta x^2]$$

for $x > 0$, and cdf $P(x \leq t) = F(t^2; \alpha, \beta)$ for $t > 0$ where again F is the cdf of the gamma distribution with shape parameter α and rate parameter β . This is not the usual parameterization of the Nakagami distribution, but it's more convenient for our purposes. Now suppose we sign this distribution by flipping a fair coin. If it comes up heads, we'll call the drawn value of x positive, while if it comes tails we'll call it negative. Let $w = 1$ when the coin comes up heads and $w = -1$ when it comes up tails. Then if $t > 0$

$$\begin{aligned} P(xw \leq t) &= P(x \leq t|w = 1)P(w = 1) + P(x \geq t|w = -1)P(w = -1) \\ &= F(t^2; \alpha, \beta) \frac{1}{2} + \frac{1}{2} \\ &= \frac{1 + F(t^2; \alpha, \beta)}{2} \end{aligned}$$

while if $t < 0$

$$\begin{aligned} P(xw \leq t) &= P(x \leq t|w = 1)P(w = 1) + P(x \geq t|w = -1)P(w = -1) \\ &= 0 + [1 - F(t^2; \alpha, \beta)] \frac{1}{2} \end{aligned}$$

and finally $P(xw \leq 0) = P(w = -1) = 1/2$. So by flipping a fair coin in order to choose the sign of a random variate drawn from the Nakagami distribution, we obtain a draw from the signed Nakagami distribution. This is useful since in order to draw from the *Nakagami*(α, β) distribution, we can simply draw from the

$G(\alpha, \beta)$ distribution and take the positive square root, thus drawing from the signed Nakagami distribution is simple and cheap.

In the sources I've seen, the Nakagami distribution is only defined for $\alpha > 1/2$ and as $\alpha \rightarrow 1/2$ it converges to a half-normal distribution. However, the normalizing constant is finite for any $\alpha > 0$ since the Nakagami distribution is simply the positive square root of the gamma distribution.

Suppose we have the basic local level model from above again:

$$\begin{aligned} y_t &= \theta_t + v_t \\ \theta_t &= \theta_{t-1} + w_t \end{aligned}$$

with $v_t \stackrel{iid}{\sim} N(0, \tau^2)$ and $w_t \stackrel{iid}{\sim} N(0, \sigma^2)$. For priors we'll assume $p(\theta_0, \tau, \sigma) = p(\theta_0)p(\tau)p(\sigma)$ with $\theta_0 \sim N(\theta_0, C_0)$ and signed Nakagami priors on τ and σ , i.e. $\tau \sim sNaka(a, 1/(2Q))$ and $\sigma \sim sNaka(b, 1/(2R))$. Then we can write the full conditional posteriors of τ and σ as

$$\begin{aligned} p(\tau|\theta_{0:T}, \dots) &\propto (\tau^2)^{a-(T+1)/2} \exp \left[-\frac{1}{2} \left(\frac{1}{Q} \tau^2 + \sum_{t=1}^T (y_t - \theta_t)^2 \frac{1}{\tau^2} \right) \right] \\ p(\sigma|\theta_{0:T}, \dots) &\propto (\sigma^2)^{b-(T+1)/2} \exp \left[-\frac{1}{2} \left(\frac{1}{R} \sigma^2 + \sum_{t=1}^T (\theta_t - \theta_{t-1})^2 \frac{1}{\sigma^2} \right) \right]. \end{aligned}$$

This results in GIG distributions for both τ^2 and σ^2 - again, this is something we can easily sample from, but consider conditioning on $\gamma_{0:T}$ instead of $\theta_{0:T}$ where again $\gamma_0 = \theta_0$ and $\gamma_t = (\theta_t - \theta_{t-1})/\sigma$. This results in the conditional posteriors

$$\begin{aligned} p(\tau|\gamma_{0:T}, \dots) &\propto (\tau^2)^{a-(T+1)/2} \exp \left[-\frac{1}{2} \left(\frac{1}{Q} \tau^2 + \sum_{t=1}^T \left(y_t - \gamma_0 - \sigma \sum_{s=1}^t \gamma_s \right)^2 \frac{1}{\tau^2} \right) \right] \\ p(\sigma|\gamma_{0:T}, \dots) &\propto (\sigma^2)^{b-1/2} \exp \left[-\frac{1}{2} \left(\frac{1}{R} \sigma^2 + \sum_{t=1}^T \left(y_t - \gamma_0 - \sigma \sum_{s=1}^t \gamma_s \right)^2 \frac{1}{\tau^2} \right) \right] \\ &\propto (\sigma^2)^{b-1/2} \exp \left[-\frac{1}{2} \left(\frac{1}{R} + \sum_{t=1}^T \left(\sum_{s=1}^t \frac{\gamma_s}{\tau} \right)^2 \right) \left(\sigma - \left(\frac{1}{R} + \sum_{t=1}^T \left(\sum_{s=1}^t \frac{\gamma_s}{\tau} \right)^2 \right)^{-1} \sum_{t=1}^T \frac{y_t - \gamma_0}{\tau} \sum_{s=1}^t \frac{\gamma_s}{\tau} \right)^2 \right] \end{aligned}$$

Again, τ has a GIG full conditional distribution. However, σ has something a bit different — it's related to the signed Nakagami distribution, but the density now has the form:

$$p(x) \propto (x^2)^{\alpha-1/2} \exp \left[-\frac{1}{2\beta} (x - \mu)^2 \right].$$

Using the known formulas for the absolute noncentral moments of the normal distribution, we can obtain the normalizing constant. If $x \sim N(\mu, \beta)$ and $p > 0$ then

$$E[|x|^p] = \frac{(2\beta)^{p/2}}{\sqrt{\pi}} \Gamma \left(\frac{1+p}{2} \right) M \left(-\frac{1}{2}p, \frac{1}{2}, -\frac{1}{2} \frac{\mu^2}{\beta} \right)$$

where $M(a, b, z)$ is the confluent hypergeometric function of the first kind, i.e.

$$M(a, b, z) = \sum_{n=0}^{\infty} \frac{a_{(n)} z^n}{b_{(n)} n!}$$

where the notation $a_{(n)}$ means $a_{(0)} = 1$ and $a_{(n)} = a(a+1)(a+2)\cdots(a+n-1)$. Thus we can write the distribution in question as

$$\begin{aligned} p(x) &= \frac{1}{E[|z|^{2\alpha-1}]} \frac{1}{\sqrt{2\pi\beta}} (x^2)^{\alpha-1/2} \exp\left[-\frac{1}{2\beta}(x-\mu)^2\right] \\ &= \frac{1}{(2\beta)^\alpha \Gamma(\alpha) M\left(\frac{1}{2} - \alpha, \frac{1}{2}, -\frac{1}{2}\frac{\mu^2}{\beta}\right)} (x^2)^{\alpha-1/2} \exp\left[-\frac{1}{2\beta}(x-\mu)^2\right] \\ &= \frac{1}{(2\beta)^\alpha e^{-\frac{1}{2}\frac{\mu^2}{\beta}} \Gamma(\alpha) M\left(\alpha, \frac{1}{2}, \frac{1}{2}\frac{\mu^2}{\beta}\right)} (x^2)^{\alpha-1/2} \exp\left[-\frac{1}{2\beta}(x-\mu)^2\right] \end{aligned}$$

where z is a $N(\mu, \beta)$ random variable and the third line follows from Kummer's transformation: $M(a, b, z) = e^z M(b-a, b, -z)$. I'll call this the noncentral signed Nakagami distribution with noncentrality parameter μ , or $nsNaka(\alpha, \beta, \mu)$ — though note that this is not simply a location transformation of the original signed Nakagami distribution. Also note that while the signed Nakagami distribution only requires $\alpha > 0$, the noncentral signed Nakagami distribution requires $\alpha \geq 1/2$ whenever $\mu \neq 0$. This actually restricts our use of the signed Nakagami distribution as a prior to $\alpha \geq 1/2$ because if we put a $sNaka(\alpha, \beta)$ prior on σ , its full conditional posterior under the scaled disturbance parameterization (the $\gamma_{0:T}$'s) is $nsNaka(\alpha, \delta, \mu)$ where $\mu > 0$ and δ are functions of the data, the γ 's and τ . So if we set $\alpha < 1/2$ the full conditional isn't actually a probability distribution.

4 Properties of the noncentral signed Nakagami distribution

For reference, the density of the noncentral signed Nakagami distribution is

$$p(x) = \frac{1}{(2\beta)^\alpha \Gamma(\alpha) M\left(\frac{1}{2} - \alpha, \frac{1}{2}, -\frac{1}{2}\frac{\mu^2}{\beta}\right)} (x^2)^{\alpha-1/2} \exp\left[-\frac{1}{2\beta}(x-\mu)^2\right]. \quad (3)$$

with $\alpha \geq 1/2$, $\beta > 0$ and $\mu \in \Re$ where $\Gamma(\cdot)$ is the gamma function and $M(\cdot, \cdot, \cdot)$ is the confluent hypergeometric function of the first kind, i.e.

$$M(a, b, z) = \sum_{n=0}^{\infty} \frac{\Gamma(a+n)\Gamma(b)z^n}{\Gamma(a)\Gamma(b+n)n!}$$

. The main thing I would like to do is draw from this distribution cheaply. The cdf inversion method doesn't look too promising since ultimately I want to generalize this to multiple dimensions (for covariance matrices), though it might work fine in this case as long as the approximation routines for the confluent hypergeometric function work well. Rejection sampling is an option, but some care is involved because the distribution is often bimodal. If $\alpha = 1/2$ clearly this is just a normal distribution so the mode is μ . If $\alpha > 1/2$, however, consider the log density

$$\log(p(x)) = \log(C) + (\alpha - 1/2) \log(x^2) - \frac{1}{2\beta}(x-\mu)^2.$$

The FOC is

$$\begin{aligned} (\alpha - 1/2) \frac{2}{x} - \frac{1}{\beta}(x - \mu) &= 0 \\ \implies \frac{1}{\beta}x^2 - \frac{\mu}{\beta}x + 1 - 2\alpha &= 0 \end{aligned}$$

so that

$$\begin{aligned} x^{mode} &= \frac{\mu/2 \pm \sqrt{\mu^2/\beta^2 + 4(2\alpha - 1)/\beta}}{2/\beta} \\ &= \mu/2 \pm \sqrt{\mu^2/4 + (2\alpha - 1)\beta}. \end{aligned}$$

The SOSC requires

$$\begin{aligned} -(2\alpha - 1)(x^{mode})^2 - \frac{1}{\beta} &< 0 \\ \iff (x^{mode})^2 &> -\beta(2\alpha - 1) \end{aligned}$$

but

$$(x^{mode})^2 = \mu^2/2 + (2\alpha - 1)\beta \pm \mu\sqrt{\mu^2/4 + (2\alpha - 1)\beta}.$$

So $\mu/2 + \sqrt{\mu^2/4 + (2\alpha - 1)\beta}$ is always a mode while $\mu/2 - \sqrt{\mu^2/4 + (2\alpha - 1)\beta}$ is a mode if

$$\begin{aligned} \mu^2/2 + 2\beta(2\alpha - 1) - \mu\sqrt{\mu^2/4 + (2\alpha - 1)\beta} &> 0 \\ \iff 2(\mu^2/4 + (2\alpha - 1)\beta) &> \mu\sqrt{\mu^2/4 + (2\alpha - 1)\beta} \\ \iff \mu^2/4 + (2\alpha - 1)\beta &> \mu^2/4. \end{aligned}$$

So if $\alpha > 1/2$ then both solutions are maximums and thus there are two modes.

The first moment of this distribution is tricky to compute, but the second moment is pretty easy (up to approximating the confluent hypergeometric function). Let $C(\alpha, \beta, \mu)$ be the normalizing constant of the $nsNaka(\alpha, \beta, \mu)$ distribution, i.e.

$$\int_{-\infty}^{\infty} C(\alpha, \beta, \mu)(x^2)^{\alpha-1/2} \exp\left[-\frac{1}{2\beta}(x - \mu)^2\right] dx = 1.$$

Then

$$\begin{aligned} E[x^2] &= \int_{-\infty}^{\infty} C(\alpha, \beta, \mu)(x^2)^{\alpha} \exp\left[-\frac{1}{2\beta}(x - \mu)^2\right] \\ &= \frac{C(\alpha, \beta, \mu)}{C(\alpha + 1, \beta, \mu)} \\ &= \frac{(2\beta)^{\alpha+1}\Gamma(\alpha + 1)M\left(-\frac{1}{2} - \alpha, \frac{1}{2}, -\frac{1}{2}\frac{\mu^2}{\beta}\right)}{(2\beta)^{\alpha}\Gamma(\alpha)M\left(\frac{1}{2} - \alpha, \frac{1}{2}, -\frac{1}{2}\frac{\mu^2}{\beta}\right)} \\ &= 2\alpha\beta \frac{M\left(\alpha + 1, \frac{1}{2}, \frac{1}{2}\frac{\mu^2}{\beta}\right)}{M\left(\alpha, \frac{1}{2}, \frac{1}{2}\frac{\mu^2}{\beta}\right)} \end{aligned}$$

5 Some notes on other priors

I started with the $N(0, Q)$ prior on $\pm\sqrt{V}$ where V is a variance parameter. This implies a $Gamma(1/2, 1/(2Q))$ prior on V . I tried to generalize this by allowing the shape parameter of the implied Gamma distribution to be something other than $1/2$. However, Another option is to pick a distribution for $\pm\sqrt{V}$ which is more general than the normal distribution, e.g. the non-standardized t distribution, $t(m, Q, v)$ where m is a location parameter, Q is a scale parameter, and v is the degrees of freedom. Using the fact that the t distribution can be written as a scale mixture of normals (if $x \sim t(m, Q, v)$ then $x|\omega \sim N(m, \omega)$ and $\omega \sim IG(v/2, vQ/2)$),

we can get nice full conditional distributions for either V or $\pm\sqrt{V}$ and either W or $\pm\sqrt{W}$. This adds another Gibbs step to the algorithm but doesn't change much other than that, i.e. we still have GIG distributions for $V|\theta, \dots$ and $W|\theta, \dots$, but now they're also conditional on scale parameters ω_V and ω_W which now each have to be drawn in separate steps, but these will each be inverse gamma draws. The draw for $W|\gamma, \dots$ is still a normal draw, but again that's conditional on ω_W which also has to be drawn in a separate inverse gamma step. It would be interesting to generalize this even more - e.g. a scale mixture of signed Nakagami distributions, though this is equivalent to just putting a hyperprior on the scale parameter of the signed Nakagami distribution. In that sense, then, we're back to the problems of the previous section.

Another thing that we can easily do and should mention is allow for dependence in the prior - $p(V, W) \neq p(V)p(W)$. The only sampling algorithm that draws (V, W) jointly is the state sampler (and interweaving algorithms that depend on it), so it shouldn't change much elsewhere - provided that $p(V|W)$ and $p(W|V)$ are of sufficiently nice form.