

# 1 Efficiently drawing from $p(W|V, \gamma, y)$ and $p(V|W, \psi, y)$

Both of these two densities are of the form

$$\log p(x) = lp(x) - ax + b\sqrt{x} - (\alpha + 1) \log x - \beta/x + C$$

for  $x > 0$  where  $C$  is some constant,  $\alpha$  and  $\beta$  are the hyperparameters for  $x$ , and  $a > 0$  &  $b \in \Re$  are parameters that depend on the data,  $y$ , the relevant data augmentation ( $\psi$  or  $\gamma$ ), and the other variable ( $W$  or  $V$ ). This density is not a known form and is difficult to sample from. We provide three different rejection sampling strategies below that work well under different circumstances, and combine them into a single strategy.

## 1.1 Adaptive rejection sampling

One nice strategy is to use adaptive rejection sampling, e.g. ?. This requires  $lp(x)$  to be concave, which is easy enough to check. The second derivative of  $lp(x)$  is:

$$\frac{\partial^2 lp(x)}{\partial x^2} = -\frac{1}{4}bx^{-3/2} + (\alpha + 1)x^{-2} - 2\beta x^{-3}.$$

Then we have

$$\begin{aligned} \frac{\partial^2 lp(x)}{\partial x^2} < 0 &\iff \\ -\frac{b}{4}x^{3/2} + (\alpha + 1)x - 2\beta &< 0 \end{aligned}$$

which would imply that  $lp(x)$  is concave. We can maximize the left hand side of the last equation very easily. When  $b \leq 0$  the max occurs at  $x = \infty$  such that  $LHS > 0$ , but when  $b > 0$ :

$$\begin{aligned} \frac{\partial LHS}{\partial x} &= -\frac{3}{8}bx^{1/2} + \alpha + 1 = 0 \\ \implies x^{max} &= \frac{(\alpha + 1)^2}{b^2} \frac{64}{9}. \end{aligned}$$

Then we have

$$LHS \leq LHS|_{x=x^{max}} = \frac{(\alpha + 1)^3}{b^2} \frac{64}{27} - 2\beta$$

so that

$$\begin{aligned} LHS|_{x=x^{max}} < 0 &\iff \frac{(\alpha + 1)^3}{b^2} \frac{64}{27} < 2\beta \\ &\iff b > \left( \frac{(\alpha + 1)^3}{\beta} \right)^{1/2} \frac{4\sqrt{2}}{3\sqrt{3}}. \end{aligned}$$

This last condition is necessary and sufficient for  $lp(x)$  to be globally (for  $x > 0$ ) concave since  $b < 0$  forces  $LHS > 0$  for some  $x$ . When the condition is satisfied, we can use adaptive rejection sampling — which is already implemented in the R package **ars**. We input the initial evaluations of  $lp(x)$  at the mode  $x^{mode}$  and at  $2x^{mode}$  and  $0.5x^{mode}$  in order to get the algorithm going.

## 1.2 Rejection sampling on the log scale

When  $b \leq \left(\frac{(\alpha+1)^3}{\beta}\right)^{1/2} \frac{4\sqrt{2}}{3\sqrt{3}}$ , which happens often — especially for small  $T$  — we need to rely on a different method to sample from  $p(x)$ . A naive approach would be to construct a normal or  $t$  approximation to  $p(x)$  and use that as a proposal in a rejection sampler. It turns out that this is often very inefficient, but for  $y = \log(x)$  the approach works well. Note that

$$p_y(y) = p_x(e^y)e^y$$

so that we can write the log density of  $y$  as (dropping the subscripts):

$$lp(y) = -ae^y + be^{y/2} - \alpha y - \beta e^{-y}.$$

The mode of this density  $y^{mode}$  can be easily found numerically, and the second derivative is:

$$\frac{\partial^2 lp(y)}{\partial y^2} = -ae^y + \frac{b}{4}e^{y/2} - \beta e^{-y}.$$

The  $t$  approximation then uses the proposal distribution  $t_{df}\left(y^{mode}, \left[\frac{\partial^2 lp(y)}{\partial y^2}\Big|_{y=y^{mode}}\right]^{-1}\right)$ . In practice choosing  $df = 1$  works very well over the region of the parameter space where adaptive rejection sampling cannot be used. We can easily use this method when adaptive rejection sampling does not work, then transform  $y$  back to  $x$ .

## 1.3 Adaptive rejection sampling on the log scale

We can also try adaptive rejection sampling on for  $y = \log(x)$ . First we need to check the concavity of  $lp(y)$ . Using the second derivative of  $lp(y)$  from above

$$\frac{\partial^2 lp(y)}{\partial y^2} < 0 \iff -ae^{2y} + \frac{b}{4}e^{3y/2} - \beta < 0.$$

So for  $b \leq 0$  we have concavity of  $lp(y)$  and can use adaptive rejection sampling. When  $b > 0$ , we can maximize the left hand side as before:

$$\begin{aligned} \frac{\partial LHS}{\partial y} &= -2ae^{2y} + \frac{3}{8}be^{3y/2} = 0 \\ \implies y^{max} &= 2 \log\left(\frac{3b}{16a}\right) \end{aligned}$$

so that

$$\begin{aligned} LHS &< LHS|_{y=y^{max}} = -a\left(\frac{3b}{16a}\right)^4 + \frac{b}{4}\left(\frac{3b}{16a}\right)^3 - \beta < 0 \\ \iff \frac{1}{16}b\left(\frac{3b}{16a}\right)^3 &< \beta \\ \iff b &\leq 16\left(\frac{\beta a^3}{3^3}\right)^{1/4}. \end{aligned}$$

This last condition is necessary and sufficient for  $lp(y)$  to be concave so that we can use adaptive rejection sampling for  $y$  then transform back to  $x$ . This is convenient because  $lp_x(x)$  and  $lp_y(y)$  are concave in roughly opposite situations — for  $b$  large and  $b$  small respectively.

## 1.4 Intelligently choosing a rejection sampler

In practice, adaptive rejection sampling is relatively efficient for  $p_x(x)$  but inefficient for  $p_y(y)$  — so much so that rejection sampling with the  $t$  approximation for  $p_y(y)$  is more efficient. To minimize computation time, it is best to use adaptive rejection sampling for  $p_x(x)$  when the concavity condition is satisfied. When it is not, the  $t$  approximation works well.

## 2 Efficiently drawing from $p(W|V, \tilde{\gamma}, y)$ and $p(V|W, \tilde{\psi}, y)$

$$p(V, W|\tilde{\gamma}_{0:T}, y_{1:T}) \propto |V|^{-(\lambda_V + k + 2)/2} \exp \left[ -\frac{1}{2} \text{tr} (\Lambda_V V^{-1}) \right] \exp \left[ -\frac{1}{2} \sum_{t=1}^T (y_t - F_t \theta_t(\tilde{\gamma}_{0:T}))' V^{-1} (y_t - F_t \theta_t(\tilde{\gamma}_{0:T})) \right] \\ \times |W|^{-(\lambda_W + k + T + 2)/2} \exp \left[ -\frac{1}{2} \text{tr} (\Lambda_W W^{-1}) \right] \exp \left[ -\frac{1}{2} \sum_{t=1}^T \tilde{\gamma}_t' L_V' W^{-1} L_V \tilde{\gamma}_t \right].$$

Since  $\theta(\tilde{\gamma}_{0:T})$  does not depend on  $V$ , we have  $V|W, \tilde{\gamma}_{0:T}, y_{1:T} \sim IW \left( \Lambda_W + L_V \sum_{t=1}^T \tilde{\gamma}_t \tilde{\gamma}_t' L_V', \lambda_W + T \right)$ . The distribution of  $W|V, \tilde{\gamma}_{0:T}, y_{1:T}$  is more complicated. In section REFERENCE TO LLM SECTION we will show an example of how to sample from this distribution in the local level model.

We can write the joint conditional posterior of  $V$  and  $W$  in this case as

$$p(V, W|\tilde{\psi}_{0:T}, y_{1:T}) \propto |V|^{-(\lambda_V + k + T + 2)/2} \exp \left[ -\frac{1}{2} \text{tr} (\Lambda_V V^{-1}) \right] \exp \left[ -\frac{1}{2} \sum_{t=1}^T \tilde{\psi}_t' L_W' V^{-1} L_W \tilde{\psi}_t \right] \\ \times |W|^{-(\lambda_W + k + 2)/2} \exp \left[ -\frac{1}{2} \text{tr} (\Lambda_W W^{-1}) \right] \exp \left[ -\frac{1}{2} \sum_{t=1}^T (y_t - \tilde{\mu}_t)' (F_t W F_t')^{-1} (y_t - \tilde{\mu}_t) \right].$$

Now since  $\tilde{\mu}_t$  doesn't depend on  $V$  we have  $V|W, \tilde{\gamma}_{0:T}, y_{1:T} \sim IW(\Lambda_V + L_W \sum_{t=1}^T \tilde{\psi}_t \tilde{\psi}_t' L_W', \lambda_V + T)$ . The distribution of  $W|V, \tilde{\psi}_{0:T}$  is once again complicated, but is analogous to that of  $V|W, \tilde{\gamma}_{0:T}$  — in section REFERENCE we give an example using the local level model.