Supplementary Material

Christian A. Naesseth[†‡] **Francisco J. R. Ruiz**[‡§] **Scott W. Linderman**[‡] **David M. Blei**[‡]
[†]Linköping University [‡]Columbia University [§]University of Cambridge

1 Distribution of ε

Here we formalize the claim in the main manuscript regarding the distribution of the accepted variable ε in the rejection sampler. Recall that $z = h(\varepsilon, \theta)$, $\varepsilon \sim s(\varepsilon)$ is equivalent to $z \sim r(z\,;\theta)$, and that $q(z\,;\theta) \leq M_\theta r(z\,;\theta)$. For simplicity we consider the univariate continuous case in the exposition below, but the result also holds for the discrete and multivariate settings. The cumulative distribution function for the accepted ε is given by

$$
\begin{aligned}
\mathbb{P}(E \leq \varepsilon) &= \sum_{i=1}^{\infty} \mathbb{P}(E \leq \varepsilon, E = E_i) \\
&= \sum_{i=1}^{\infty} \Bigg[\mathbb{P}\left(E_i \leq \varepsilon, U_i < \frac{q(h(E_i,\theta)\,;\theta)}{M_\theta r(h(E_i,\theta)\,;\theta)} \right) \\
&\quad \prod_{j=1}^{i-1} \mathbb{P}\left(U_j \geq \frac{q(h(E_j,\theta)\,;\theta)}{M_\theta r(h(E_j,\theta)\,;\theta)} \right) \Bigg] \\
&= \sum_{i=1}^{\infty} \int_{-\infty}^{\varepsilon} s(e) \frac{q(h(e,\theta)\,;\theta)}{M_\theta r(h(e,\theta)\,;\theta)} \mathrm{d}e \prod_{j=1}^{i-1} \left(1 - \frac{1}{M_\theta} \right) \\
&= \int_{-\infty}^{\varepsilon} s(e) \frac{q(h(e,\theta)\,;\theta)}{r(h(e,\theta)\,;\theta)} \mathrm{d}e \cdot \frac{1}{M_\theta} \cdot \sum_{i=1}^{\infty} \left(1 - \frac{1}{M_\theta} \right)^{i-1} \\
&= \int_{-\infty}^{\varepsilon} s(e) \frac{q(h(e,\theta)\,;\theta)}{r(h(e,\theta)\,;\theta)} \mathrm{d}e.
\end{aligned}
$$

Here, we have applied that $z = h(\varepsilon, \theta)$, $\varepsilon \sim s(\varepsilon)$ is a reparameterization of $z \sim r(z\,;\theta)$, and thus

$$
\begin{aligned}
&\mathbb{P}\left(U_j \geq \frac{q(h(E_j,\theta)\,;\theta)}{M_\theta r(h(E_j,\theta)\,;\theta)} \right) \\
&= \int_{-\infty}^{\infty} s(e) \left(1 - \frac{q(h(e,\theta)\,;\theta)}{M_\theta r(h(e,\theta)\,;\theta)} \right) \mathrm{d}e \\
&= 1 - \frac{1}{M_\theta} \mathbb{E}_{s(e)} \left[\frac{q(h(e,\theta)\,;\theta)}{r(h(e,\theta)\,;\theta)} \right] \\
&= 1 - \frac{1}{M_\theta} \mathbb{E}_{r(z\,;\theta)} \left[\frac{q(z\,;\theta)}{r(z\,;\theta)} \right] = 1 - \frac{1}{M_\theta}.
\end{aligned}
$$

The density is obtained by taking the derivative of the cumulative distribution function with respect to ε,

$$
\frac{\mathrm{d}}{\mathrm{d}\varepsilon} \mathbb{P}(E \leq \varepsilon) = s(\varepsilon) \frac{q(h(\varepsilon,\theta)\,;\theta)}{r(h(\varepsilon,\theta)\,;\theta)},
$$

which is the expression from the main manuscript.

The motivation from the main manuscript is basically a standard "area-under-the-curve" or geometric argument for rejection sampling [Robert and Casella, 2004], but for ε instead of z.

2 Derivation of the Gradient

We provide below details for the derivation of the gradient. We assume that h is differentiable (almost everywhere) with respect to θ, and that $f(h(\varepsilon,\theta)) \frac{q(h(\varepsilon,\theta)\,;\theta)}{r(h(\varepsilon,\theta)\,;\theta)}$ is continuous in θ for all ε. Then, we have

$$
\begin{aligned}
&\nabla_\theta \mathbb{E}_{q(z\,;\theta)}[f(z)] = \nabla_\theta \mathbb{E}_{\pi(\varepsilon\,;\theta)}[f(h(\varepsilon,\theta))] \\
&= \int s(\varepsilon) \nabla_\theta \left(f\left(h(\varepsilon,\theta)\right) \frac{q\left(h(\varepsilon,\theta)\,;\theta\right)}{r\left(h(\varepsilon,\theta)\,;\theta\right)} \right) \mathrm{d}\varepsilon \\
&= \int s(\varepsilon) \frac{q\left(h(\varepsilon,\theta)\,;\theta\right)}{r\left(h(\varepsilon,\theta)\,;\theta\right)} \nabla_\theta f\left(h(\varepsilon,\theta)\right) \mathrm{d}\varepsilon \\
&+ \int s(\varepsilon) f\left(h(\varepsilon,\theta)\right) \nabla_\theta \left(\frac{q\left(h(\varepsilon,\theta)\,;\theta\right)}{r\left(h(\varepsilon,\theta)\,;\theta\right)} \right) \mathrm{d}\varepsilon \\
&= \underbrace{\mathbb{E}_{\pi(\varepsilon\,;\theta)}[\nabla_\theta f(h(\varepsilon,\theta))]}_{=:g_{\mathrm{rep}}} + \\
&\quad \underbrace{+ \mathbb{E}_{\pi(\varepsilon\,;\theta)} \left[f(h(\varepsilon,\theta)) \nabla_\theta \log \frac{q(h(\varepsilon,\theta)\,;\theta)}{r(h(\varepsilon,\theta)\,;\theta)} \right]}_{=:g_{\mathrm{cor}}},
\end{aligned}
$$

where in the last step we have identified $\pi(\varepsilon\,;\theta)$ and made use of the log-derivative trick

$$
\nabla_\theta \frac{q\left(h(\varepsilon,\theta)\,;\theta\right)}{r\left(h(\varepsilon,\theta)\,;\theta\right)} = \frac{q\left(h(\varepsilon,\theta)\,;\theta\right)}{r\left(h(\varepsilon,\theta)\,;\theta\right)} \nabla_\theta \log \frac{q\left(h(\varepsilon,\theta)\,;\theta\right)}{r\left(h(\varepsilon,\theta)\,;\theta\right)}.
$$

Gradient of Log-Ratio in g_{cor} For invertible reparameterizations we can simplify the evaluation of the gradient of the log-ratio in g_{cor} as follows using standard results on transformation of a random variable

$$\begin{aligned}
&\nabla_\theta \log \frac{q(h(\varepsilon,\theta)\,;\theta)}{r(h(\varepsilon,\theta)\,;\theta)} = \nabla_\theta \log q(h(\varepsilon,\theta)\,;\theta) + \\
&+ \nabla_\theta \log \left| \frac{dh}{d\varepsilon}(\varepsilon,\theta) \right| - \nabla_\theta \log \underbrace{s(h^{-1}(h(\varepsilon,\theta),\theta))}_{=\, s(\varepsilon)} \\
&= \nabla_\theta \log q(h(\varepsilon,\theta)\,;\theta) + \nabla_\theta \log \left| \frac{dh}{d\varepsilon}(\varepsilon,\theta) \right|.
\end{aligned}$$

3 Examples of Reparameterizable Rejection Samplers

We show in Table 1 some examples of reparameterizable rejection samplers for three distributions, namely, the gamma, the truncated normal, and the von Misses distributions (for more examples, see Devroye [1986]). We show the distribution $q(z\,;\theta)$, the transformation $h(\varepsilon,\theta)$, and the proposal $s(\varepsilon)$ used in the rejection sampler.

We show in Table 2 six examples of distributions that can be reparameterized in terms of auxiliary gamma-distributed random variables. We show the distribution $q(z\,;\theta)$, the distribution of the auxiliary gamma random variables $p(\tilde{z}\,;\theta)$, and the mapping $z = g(\tilde{z},\theta)$.

4 Reparameterizing the Gamma Distribution

We provide details on reparameterization of the gamma distribution. In the following we consider rate $\beta = 1$. Note that this is not a restriction, we can always reparameterize the rate. The density of the gamma random variable is given by

$$q(z\,;\alpha) = \frac{z^{\alpha-1}e^{-z}}{\Gamma(\alpha)},$$

where $\Gamma(\alpha)$ is the gamma function. We make use of the reparameterization defined by

$$\begin{aligned}
&z = h(\varepsilon,\alpha) = \left(\alpha - \frac{1}{3}\right)\left(1 + \frac{\varepsilon}{\sqrt{9\alpha-3}}\right)^3, \\
&\varepsilon \sim \mathcal{N}(0,1).
\end{aligned}$$

Because h is invertible we can make use of the simplified gradient of the log-ratio derived in Section 2 above. The gradients of $\log q$ and $-\log r$ are given by

$$\begin{aligned}
&\nabla_\alpha \log q(h(\varepsilon,\alpha)\,;\alpha) \\
&= \log(h(\varepsilon,\alpha)) + (\alpha-1)\frac{\frac{dh(\varepsilon,\alpha)}{d\alpha}}{h(\varepsilon,\alpha)} - \frac{dh(\varepsilon,\alpha)}{d\alpha} - \psi(\alpha), \\
&\nabla_\alpha - \log r(h(\varepsilon,\alpha)\,;\alpha) = \nabla_\alpha \log \left| \frac{dh}{d\varepsilon}(\varepsilon,\alpha) \right| \\
&= \frac{1}{2\left(\alpha - \frac{1}{3}\right)} - \frac{9\varepsilon}{\left(1 + \frac{\varepsilon}{\sqrt{9\alpha-3}}\right)(9\alpha-3)^{\frac{3}{2}}},
\end{aligned}$$

where $\psi(\alpha)$ is the digamma function and

$$\begin{aligned}
&\frac{dh(\varepsilon,\alpha)}{d\alpha} \\
&= \left(1 + \frac{\varepsilon}{\sqrt{9\alpha-3}}\right)^3 - \frac{27\varepsilon}{2(9\alpha-3)^{\frac{3}{2}}}\left(1 + \frac{\varepsilon}{\sqrt{9\alpha-3}}\right)^2.
\end{aligned}$$

References

L. Devroye. *Non-Uniform Random Variate Generation.* Springer-Verlag, 1986.

C. Robert and G. Casella. *Monte Carlo statistical methods.* Springer Science & Business Media, 2004.

$q(z\,;\theta)$	$h(\varepsilon,\theta)$	$s(\varepsilon)$
Gamma$(\alpha, 1)$	$\left(\alpha - \frac{1}{3}\right)\left(1 + \frac{\varepsilon}{\sqrt{9\alpha-3}}\right)^3$	$\varepsilon \sim \mathcal{N}(0,1)$
Truncated $\mathcal{N}(0,1,a,\infty)$	$\sqrt{a^2 - 2\log\varepsilon}$	$\varepsilon \sim \mathcal{U}[0,1]$
vonMises(κ)	$\frac{\mathrm{sign}(\varepsilon)}{\cos\left(\frac{1+c\cos(\pi\varepsilon)}{c+\cos(\pi\varepsilon)}\right)}$, $c = \frac{1+\rho^2}{2\rho}$, $\rho = \frac{r-\sqrt{2r}}{2\kappa}$, $r = 1+\sqrt{1+4\kappa^2}$	$\varepsilon \sim \mathcal{U}[-1,1]$

Table 1: Examples of reparameterizable rejection samplers; many more can be found in Devroye [1986]. The first column is the distribution, the second column is the transformation $h(\varepsilon,\theta)$, and the last column is the proposal $s(\varepsilon)$.

$q(z\,;\theta)$	$g(\tilde{z},\theta)$	$p(\tilde{z}\,;\theta)$
Beta(α,β)	$\frac{\tilde{z}_1}{\tilde{z}_1+\tilde{z}_2}$	$\tilde{z}_1 \sim$ Gamma$(\alpha,1)$, $\tilde{z}_2 \sim$ Gamma$(\beta,1)$
Dirichlet$(\alpha_{1:K})$	$\frac{1}{\sum_\ell \tilde{z}_\ell}\left(\tilde{z}_1,\ldots,\tilde{z}_K\right)^\top$	$\tilde{z}_k \sim$ Gamma$(\alpha_k,1)$, $k=1,\ldots,K$
St(ν)	$\sqrt{\frac{\nu}{2\tilde{z}_1}}\tilde{z}_2$	$\tilde{z}_1 \sim$ Gamma$(\nu/2,1)$, $\tilde{z}_2 \sim \mathcal{N}(0,1)$
$\chi^2(k)$	$2\tilde{z}$	$\tilde{z} \sim$ Gamma$(k/2,1)$
F(d_1,d_2)	$\frac{d_2\tilde{z}_1}{d_1\tilde{z}_2}$	$\tilde{z}_1 \sim$ Gamma$(d_1/2,1)$, $\tilde{z}_2 \sim$ Gamma$(d_2/2,1)$
Nakagami(m,Ω)	$\sqrt{\frac{\Omega\tilde{z}}{m}}$	$\tilde{z} \sim$ Gamma$(m,1)$

Table 2: Examples of random variables as functions of auxiliary random variables with reparameterizable distributions. The first column is the distribution, the second column is a function $g(\tilde{z},\theta)$ mapping from the auxiliary variables to the desired variable, and the last column is the distribution of the auxiliary variables $\tilde{z}$.