1 Distribution of $\varepsilon$

Here we formalize the claim in the main manuscript regarding the distribution of the accepted variable $\varepsilon$ 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 $\varepsilon$ is given by

$$
P(E \leq \varepsilon) = \sum_{i=1}^{\infty} P(E \leq \varepsilon, E = E_i)
= \sum_{i=1}^{\infty} \left[ P\left( E_i \leq \varepsilon, U_i < \frac{q(h(E_i, \theta); \theta)}{M_\theta r(h(E_i, \theta); \theta)} \right) \prod_{j=1}^{i-1} P\left( U_j \geq \frac{q(h(E_j, \theta); \theta)}{M_\theta r(h(E_j, \theta); \theta)} \right) \right]
= \sum_{i=1}^{\infty} \int_{-\infty}^{\varepsilon} s(e) \frac{q(h(e, \theta); \theta)}{M_\theta r(h(e, \theta); \theta)} \, de \prod_{j=1}^{i-1} \left( 1 - \frac{1}{M_\theta} \right)
= \int_{-\infty}^{\varepsilon} s(e) \frac{q(h(e, \theta); \theta)}{r(h(e, \theta); \theta)} \, de.
$$

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

$$
P\left( U_j \geq \frac{q(h(E_j, \theta); \theta)}{M_\theta r(h(E_j, \theta); \theta)} \right)
= \int_{-\infty}^{\infty} s(e) \left( 1 - \frac{q(h(e, \theta); \theta)}{M_\theta r(h(e, \theta); \theta)} \right) \, de
= 1 - \frac{1}{M_\theta} E_{\pi_\theta}[q(h(e, \theta); \theta) r(h(e, \theta); \theta)]
= 1 - \frac{1}{M_\theta} E_{\pi_\theta}[q(z; \theta) r(z; \theta)] = 1 - \frac{1}{M_\theta}.
$$

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

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

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 $\varepsilon$ 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 $\theta$, and that $f(h(\varepsilon, \theta)) r(h(\varepsilon, \theta); \theta)$ is continuous in $\theta$ for all $\varepsilon$. Then, we have

$$
\nabla_\theta \mathbb{E}_{\pi_\theta}[f(h(\varepsilon, \theta))] = \nabla_\theta \mathbb{E}_{\pi_\theta}[f(h(\varepsilon, \theta))]
= \int s(\varepsilon) \nabla_\theta \left( f(h(\varepsilon, \theta)) \frac{q(h(\varepsilon, \theta); \theta)}{r(h(\varepsilon, \theta); \theta)} \right) \, d\varepsilon
= \int s(\varepsilon) q(h(\varepsilon, \theta); \theta) r(h(\varepsilon, \theta); \theta) \nabla_\theta f(h(\varepsilon, \theta)) \, d\varepsilon
+ \int s(\varepsilon) f(h(\varepsilon, \theta)) \nabla_\theta \left( \frac{q(h(\varepsilon, \theta); \theta)}{r(h(\varepsilon, \theta); \theta)} \right) \, d\varepsilon
= \mathbb{E}_{\pi_\theta}[\nabla_\theta f(h(\varepsilon, \theta))]
+ \mathbb{E}_{\pi_\theta}[f(h(\varepsilon, \theta)) \nabla_\theta \log \frac{q(h(\varepsilon, \theta); \theta)}{r(h(\varepsilon, \theta); \theta)}],
$$

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

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

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 stan-
dard results on transformation of a random variable
\[ \nabla_{\theta} \log q(h(\varepsilon, \theta); \theta) = \nabla_{\theta} \log q(h(\varepsilon, \theta); \theta) + \nabla_{\theta} \log \left| \frac{dh}{d\varepsilon}(\varepsilon, \theta) \right| = \nabla_{\theta} \log q(h(\varepsilon, \theta); \theta) + \sum_{\varepsilon} \log \left| \frac{dh}{d\varepsilon}(\varepsilon, \theta) \right| = \sum_{\varepsilon} \nabla_{\theta} \log q(h(\varepsilon, \theta); \theta) + \nabla_{\theta} \log \left| \frac{dh}{d\varepsilon}(\varepsilon, \theta) \right|. \]

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
\[ z = h(\varepsilon, \alpha) = \left( \alpha - \frac{1}{3} \right) \left( 1 + \frac{\varepsilon}{\sqrt{9\alpha - 3}} \right)^3, \]
\[ \varepsilon \sim \mathcal{N}(0, 1). \]

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
\[ \nabla_{\alpha} \log q(h(\varepsilon, \alpha); \alpha) = \log(h(\varepsilon, \alpha)) + (\alpha - 1) \frac{dh(\varepsilon, \alpha)}{d\alpha} - \frac{dh(\varepsilon, \alpha)}{d\alpha} - \psi(\alpha), \]
\[ \nabla_{\alpha} - \log r(h(\varepsilon, \alpha); \alpha) = \nabla_{\alpha} \log \left| \frac{dh}{d\varepsilon}(\varepsilon, \alpha) \right| = \frac{1}{2} \left( \alpha - \frac{1}{3} \right) \left( 1 + \frac{\varepsilon}{\sqrt{9\alpha - 3}} \right)^3 \frac{9\varepsilon}{(9\alpha - 3)^2}. \]

References


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 \( s(\varepsilon) \).

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Transformation ( h(\varepsilon, \theta) )</th>
<th>Proposal ( s(\varepsilon) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gamma((\alpha, 1))</td>
<td>((\alpha - \frac{1}{3}) \left(1 + \frac{\varepsilon}{\sqrt{3\alpha - 3}}\right)^3)</td>
<td>(\varepsilon \sim \mathcal{N}(0, 1))</td>
</tr>
<tr>
<td>Truncated (\mathcal{N}(0, 1, a, \infty))</td>
<td>(\sqrt{a^2 - 2 \log \varepsilon})</td>
<td>(\varepsilon \sim \mathcal{U}[0, 1])</td>
</tr>
<tr>
<td>vonMises((\kappa))</td>
<td>(\frac{\text{sign}(\varepsilon)}{\cos\left(\frac{\pi \text{sign}(\varepsilon)}{2\kappa}\right)}), (c = \frac{1 + \rho^2}{2\rho}), (\rho = \frac{r - \sqrt{2r}}{2\kappa}), (r = 1 + \sqrt{1 + 4\kappa^2})</td>
<td>(\varepsilon \sim \mathcal{U}[-1, 1])</td>
</tr>
</tbody>
</table>

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 \( g(\bar{z}, \theta) \) mapping from the auxiliary variables to the desired variable, and the last column is the distribution of the auxiliary variables \(\bar{z}\).

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Function ( g(\bar{z}, \theta) )</th>
<th>Distribution of Auxiliary Variables (\bar{z})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beta((\alpha, \beta))</td>
<td>(\frac{\bar{z}_1}{\bar{z}_1 + \bar{z}_2})</td>
<td>(\bar{z}_1 \sim \text{Gamma}(\alpha, 1)), (\bar{z}_2 \sim \text{Gamma}(\beta, 1))</td>
</tr>
<tr>
<td>Dirichlet((\alpha_1:K))</td>
<td>(\frac{1}{\sum \bar{z}_\ell} (\bar{z}_1, \ldots, \bar{z}_K)^\top)</td>
<td>(\bar{z}_k \sim \text{Gamma}(\alpha_k, 1), k = 1, \ldots, K)</td>
</tr>
<tr>
<td>St((\nu))</td>
<td>(\sqrt{\frac{\nu}{2\bar{z}_2}})</td>
<td>(\bar{z}_1 \sim \text{Gamma}(\nu/2, 1)), (\bar{z}_2 \sim \mathcal{N}(0, 1))</td>
</tr>
<tr>
<td>(\chi^2(k))</td>
<td>(2\bar{z})</td>
<td>(\bar{z} \sim \text{Gamma}(k/2, 1))</td>
</tr>
<tr>
<td>(F(d_1, d_2))</td>
<td>(\frac{d_2 \bar{z}_1}{d_1 \bar{z}_2})</td>
<td>(\bar{z}_1 \sim \text{Gamma}(d_1/2, 1)), (\bar{z}_2 \sim \text{Gamma}(d_2/2, 1))</td>
</tr>
<tr>
<td>Nakagami((m, \Omega))</td>
<td>(\sqrt{\frac{\Omega \bar{z}}{m}})</td>
<td>(\bar{z} \sim \text{Gamma}(m, 1))</td>
</tr>
</tbody>
</table>