Electronic Journal of Statistics
Vol. 0 (0000)
ISSN: 1935-7524
DOI: 10.1214/154957804100000000

Rejection Sampling for an Extended Gamma Distribution

Ying Liu, Michael J. Wichura and Mathias Drton

Department of Statistics
The University of Chicago
Chicago, IL 60637, U.S.A.
e-mail: summeryingl@gmail.com; wichura@galton.uchicago.edu; drton@uchicago.edu

Abstract: We propose and analyze rejection samplers for simulating from an extended Gamma distribution. This distribution is supported on $(0, \infty)$ and has density proportional to $t^{\alpha-1}\exp(-t - 2\sqrt{t}\gamma)$, where α and γ are two real parameters with $\alpha > 0$. The proposed samplers use normal and Gamma instrumental distributions. We show that when $\alpha \geq 1/2$, a suitable combination of the samplers has acceptance rate at least 0.8 for all γ and near 1 when $|\gamma|$ is small or large. When $\alpha < 1/2$, our optimal acceptance rates are uniformly close to 1 for $\gamma \geq 0$.

Keywords and phrases: Alternative t-distribution, Gamma distribution, rejection sampling.

1. Introduction

This paper treats the problem of drawing random samples from the probability distribution with density

$$f(t \,|\, \alpha, \gamma) := \frac{1}{2Z(\alpha, \gamma)}\, t^{\alpha-1} \exp\left(-t - 2\sqrt{t}\gamma\right), \qquad t > 0. \tag{1.1}$$

Here, $\alpha > 0$ and $\gamma \in \mathbb{R}$ are two parameters, and $Z(\alpha, \gamma)$ is the normalizing constant; the factor '2' in the denominator is kept for later convenience. This two-parameter family provides an extension of Gamma distributions in that $f(\,\cdot\,|\,\alpha, 0)$ defines the Gamma distribution with shape parameter α and scale parameter 1. Some examples of graphs of $f(\,\cdot\,|\,\alpha, \gamma)$ are shown in Figure 1.

We begin this introduction by motivating our interest in the extended Gamma distributions (Section 1.1). Next, we describe the contribution of the paper, which consists of a rejection sampling algorithm. For $\alpha \geq 1/2$, as is the case in our motivating application, our sampler has an acceptance rate that never drops below 0.8 and is close to 1 when $|\gamma|$ is near zero or sufficiently large. The description of the algorithm is split into two subsections. We first state three different types of rejection samplers (Section 1.2) and then explain how they can be combined in an optimal fashion (Section 1.3). This part of the introduction includes pseudo-code and provides all information necessary for implementation of our sampler. We conclude the introduction with an outline of the remainder of the paper (Section 1.4), in which we prove correctness and analyze the acceptance rates of the rejection samplers we propose and combine.

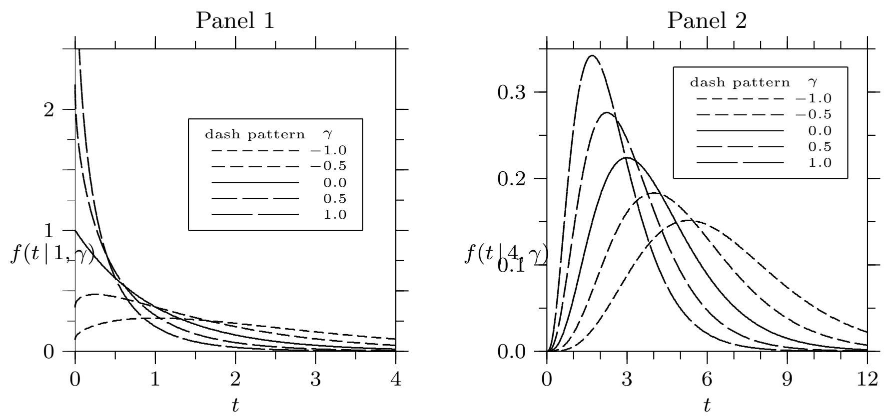

Fig 1. *Graphs of the target density $f(t \,|\, \alpha, \gamma)$ versus t for $\gamma = -1$, $-1/2$, 0, 1/2, and 1, and $\alpha = 1$ (Panel 1) and $\alpha = 4$ (Panel 2).*

The remainder also includes a discussion of the case $\alpha < 1/2$, which our methods address satisfactorily only for $\gamma \geq 0$.

1.1. Motivation

Our interest in the density from (1.1) is due to its relevance for statistical computation for certain multivariate t-distributions. Let Z be a Gamma random variable, with $Z \sim \Gamma(\nu/2, \nu/2)$ for a degrees of freedom parameter $\nu > 0$; we parametrize the density $g(x)$ of $\Gamma(\alpha, \beta)$ to be proportional to $x^{\alpha-1}e^{-\beta x}$. Suppose that given Z, the random vector $\mathbf{Y}$ is p-variate normal with

$$(\mathbf{Y} \,|\, Z) \;\sim\; \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}/Z) \tag{1.2}$$

for a mean vector $\boldsymbol{\mu}$ and a positive definite matrix $\boldsymbol{\Sigma}$. Then the (marginal) distribution of $\mathbf{Y}$ is the classical multivariate t-distribution with ν degrees of freedom. Statistical modelling with t-distributions, multivariate or not, is a common approach when data are heavy-tailed or may contain outliers; compare, e.g., Besag and Higdon (1999) or Gottardo et al. (2006). Solutions to computational problems in such models often exploit the above scale mixture representation of the t-distribution. In particular, it is straightforward to set up a Gibbs sampler because the conditional distribution of Z given Y is a Gamma distribution. This fact is also important for EM algorithms (Liu and Rubin, 1995).

The scale mixture that yields the classical multivariate t-distribution involves a single divisor Z for the entire vector $\mathbf{Y}$. This can be unsuitable when modelling highly multivariate data that contain outliers in small parts of many observations. Instead, Finegold and Drton (2011) propose an alternative t-distribution that is obtained by associating independent $\Gamma(\nu/2, \nu/2)$ divisors $Z_1, \ldots, Z_p$ with the different coordinates of $\mathbf{Y} = (Y_1, \ldots, Y_p)^\top$. Defining the diagonal matrix

$\mathbf{D}(\mathbf{Z}) = \mathrm{diag}(Z_1, \dots, Z_p)$, the alternative t-distribution is the (marginal) distribution of $\mathbf{Y}$ when

$$(\mathbf{Y} \mid \mathbf{Z}) \sim \mathcal{N}_p\big(\boldsymbol{\mu}, \mathbf{D}(\mathbf{Z})^{-1/2}\, \boldsymbol{\Sigma}\, \mathbf{D}(\mathbf{Z})^{-1/2}\big). \tag{1.3}$$

For statistical computation, as in Gibbs sampling for Bayesian inference or in Monte Carlo EM algorithms, we now need to be able to sample from the conditional distributions of Z_i given $\mathbf{Y}$ and the remaining divisors Z_j, $j \neq i$. Up to rescaling Z_i by a constant, this conditional distribution is of the extended Gamma type as defined in (1.1). Similarly, the extended Gamma distribution arises as a conditional distribution in Finegold and Drton (2012), where a clustering model allows for sharing of Gamma divisors across different coordinates.

Example 1.1. Form the diagonal matrix $\mathbf{D}(\mathbf{Z}) = \mathrm{diag}(Z_1, Z_2, Z_2)$ from two independent $\Gamma(\nu/2, \nu/2)$ random variables Z_1, Z_2. Let the conditional distribution of $\mathbf{Y} = (Y_1, Y_2, Y_3)$ given (Z_1, Z_2) be trivariate normal, following (1.3) with mean vector $\boldsymbol{\mu} = \mathbf{0}$ and inverse covariance matrix

$$\mathbf{D}(\mathbf{Z})^{1/2}\boldsymbol{\Sigma}^{-1}\mathbf{D}(\mathbf{Z})^{1/2} = \begin{pmatrix} 2Z_1 & \sqrt{Z_1 Z_2} & 0 \\ \sqrt{Z_1 Z_2} & 2Z_2 & 0 \\ 0 & 0 & 2Z_2 \end{pmatrix}.$$

Then the conditional density of Z_1 given $(\mathbf{Y}, Z_2) = (y_1, y_2, z_2)$ is proportional to the function

$$z_1 \mapsto z_1^{(\nu-1)/2} \exp\left[-z_1 \cdot \left(\frac{\nu}{2} + y_1^2\right) - \sqrt{z_1} \cdot (y_1 y_2 \sqrt{z_2})\right],$$

and the conditional density of Z_2 given $(\mathbf{Y}, Z_1) = (y_1, y_2, z_1)$ is proportional to

$$z_2 \mapsto z_2^{\nu/2} \exp\left[-z_2 \cdot \left(\frac{\nu}{2} + y_2^2 + y_3^2\right) - \sqrt{z_2} \cdot (y_1 y_2 \sqrt{z_1})\right].$$

Up to rescaling of z_1 and z_2, respectively, the two densities have the extended Gamma form with shape parameter α that is larger than $1/2$ and depends on the degrees of freedom ν as well as the number of coordinates of $\mathbf{Y}$ that Z_1 or Z_2 were associated to. The second parameter γ can be any real number.

1.2. Three rejection samplers

The contribution of this paper are rejection samplers for the extended Gamma distributions from (1.1). One advantage of rejection sampling is that it avoids the need to compute the normalizing constant $Z(\alpha, \gamma)$ in (1.1). While software such as `Mathematica` readily expresses $Z(\alpha, \gamma)$ in terms of special functions, evaluating those numerically becomes unreliable when γ or α are of larger magnitude. Similarly, quadrature methods for numerical integration fail to converge in more extreme cases when $Z(\alpha, \gamma)$ can be very large or small. These numerical issues aside, it is also natural to turn to rejection sampling for our problem because the state-of-the-art methods for fast sampling from Gamma distributions are rejection samplers. Throughout the paper we assume that an efficient routine for drawing from Gamma distributions is available; see Marsaglia and Tsang (2000) and the older references therein for this (non-trivial) problem.

For a brief review of the principle underlying rejection sampling, consider two probability densities f and g. If there is a finite constant M such that $f(t) \leq Mg(t)$ for all t in the domain of f, then we can draw from f by rejection sampling with g as instrumental distribution. In other words, we generate draws t from g until one such draw is accepted, where the probability of accepting draw t is $f(t)/[Mg(t)]$. The acceptance rate of the sampler, that is, the unconditional probability of accepting a draw from g, is $1/M$; compare Liu (2008) or Robert and Casella (2004). Clearly, the smaller M is the larger is the acceptance rate; the smallest allowable M is

$$M_{\text{opt}} := \sup\{\, f(t)/g(t) : f(t) > 0 \,\}. \tag{1.4}$$

We now give three rejection samplers for extended Gamma distributions. Depending on the parameters α and γ, each sampler selects an optimal instrumental distribution from a particular parametric class. Here, optimality refers to maximizing the acceptance rate $1/M_{\text{opt}}$ resulting from (1.4).

a) Gamma distributions on original scale. The density $f(\,\cdot \mid \alpha, 0)$ defines the standard Gamma distribution $\Gamma(\alpha, 1)$. The family $\{\Gamma(r, \delta) : r, \delta > 0\}$ thus provides natural candidates for instrumental distributions in rejection sampling. For $\gamma \leq 0$, numerical evaluation of the acceptance rate $1/M_{\text{opt}}(r, \delta)$ for a wide range of choices of (r, δ) suggests taking the shape parameter to be $r = \alpha$, for which the optimal choice of the scale parameter is

$$\delta_0(\alpha, \gamma) := \frac{\gamma^2 + 2\alpha - \sqrt{\gamma^4 + 4\alpha\gamma^2}}{2\alpha} = \frac{4\alpha}{\left(\sqrt{\gamma^2 + 4\alpha} + |\gamma|\,\right)^2}. \tag{1.5}$$

Similarly, for $\gamma > 0$, we take $\delta = 1$ and altough the acceptance rate is maximized by $r_0 = \theta\alpha$, where θ is root of function $s(\theta) = \frac{\Gamma'(\theta\alpha)}{\Gamma(\theta\alpha)} - 2log(\frac{(1-\theta)\alpha)}{\gamma})$. But there is additional computational cost to get this optimal r and computing $\frac{f(t)}{Mg(t)}$ in the sampling process. So for simplicity and computational speed, we use $\Gamma(\alpha, 1)$ for this case. Pseudo-code for the rejection sampler based on these choices is given in Algorithm 1. We denote the sampler by $\mathcal{S}_0$. Figure 2 graphs the sampler's acceptance rate $A_0(\alpha, \gamma)$ for a selection of fixed values of α and for γ of the form $C\sqrt{\alpha}$ for $|C| \leq 4$. The sampler works for any $\alpha > 0$ and is perfect for $\gamma = 0$, but Figure 2 shows that the acceptance rates grow ever smaller as $|\gamma|$ increases.

Remark. The pseudo-code in Algorithm 1, and the later Algorithms 2 and 3, is not meant to be an efficient implementation. For instance, it will typically be beneficial to take logarithms in the acceptance rule.

In order to remedy the problem of decreasing acceptance rates, we consider

---

Algorithm 1 (Gamma rejection sampler on original scale: $\mathcal{S}_0$)

Input: Parameters $\alpha > 0$ and $\gamma \in \mathbb{R}$
if $\gamma \le 0$ **then**
  $\delta_0 \leftarrow \frac{4\alpha}{\left(\sqrt{\gamma^2+4\alpha}+|\gamma|\right)^2}$.
  repeat
    Generate $T \sim \Gamma(r, \delta_0)$ and $U \sim$ Uniform$(0, 1)$.
  until $U < \exp(-T(1-\delta_0) - 2\sqrt{T}\gamma - \frac{\alpha}{\delta})$.
else
  repeat
    Generate $T \sim \Gamma(\alpha, 1)$ and $U \sim$ Uniform $(0, 1)$.
  until $U < \exp(-2\alpha\sqrt{T})$.
end if
Output: Accepted draw T.

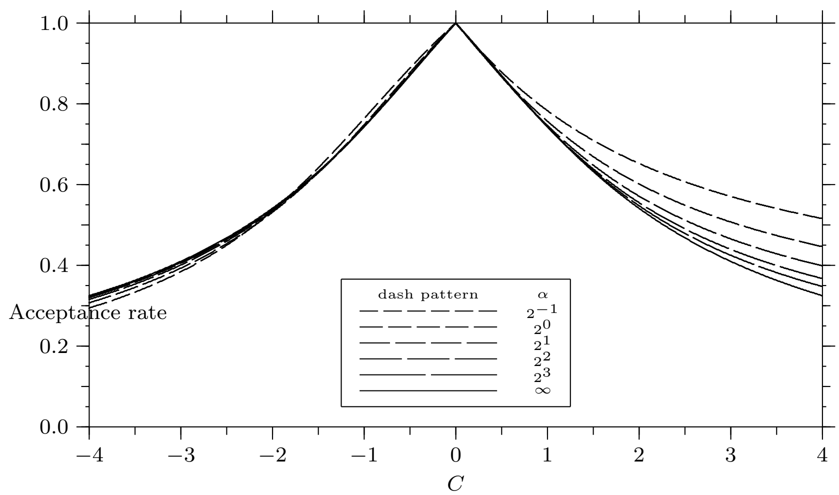

FIG 2. *Graphs of the acceptance rate $A_0(\alpha, C\sqrt{\alpha})$ versus C for $\alpha = 1/2$, 1, 2, 4, 8, and in the limit as $\alpha \to \infty$, when using Gamma instrumental distributions on the original scale. See (2.9) for the case $\alpha < \infty$ and (5.22) and (5.29) for the limiting case.*

an alternative approach in which we transform $f(t \mid \alpha, \gamma)$ to

$$h(x \mid \alpha, \gamma) := \frac{1}{Z(\alpha, \gamma)} x^{2\alpha-1} \exp(-x^2 - 2\gamma x), \qquad x > 0, \tag{1.6}$$

by the substitution $t = x^2$. A squared draw from $h(\cdot \mid \alpha, \gamma)$ follows $f(\cdot \mid \alpha, \gamma)$.

b) Normal distributions on square-root scale. For $\gamma \to -\infty$, we can approximate $h(\cdot \mid \alpha, \gamma)$ by a normal density. Therefore, we consider normal distributions as instruments for drawing from $h(\cdot \mid \alpha, \gamma)$. This approach works only for $\alpha \geq 1/2$; the supremum M_{opt} is never finite for $\alpha < 1/2$. For an optimal choice of the normal instrument, we may set the variance to $\sigma_0^2 = 1/2$ and the mean to

$$m(\alpha, \gamma) = \begin{cases} \frac{-\gamma + \sqrt{\gamma^2 + 4\alpha - 2}}{2} = \frac{2\alpha - 1}{(\gamma + \sqrt{\gamma^2 + 4\alpha - 2})}, & \text{if } \alpha > 1/2, \\ \max(-\gamma, 0) & \text{if } \alpha = 1/2. \end{cases} \tag{1.7}$$

Pseudo-code for the resulting sampler $\mathcal{S}_{\mathcal{N}}$ is given in Algorithm 2. The sampler's acceptance rate $A_{\mathcal{N}}(\alpha, \gamma)$ is strictly decreasing in γ; compare Figure 3 which plots $A_{\mathcal{N}}$ together with the acceptance rates of the next sampler we consider.

Algorithm 2 (Normal rejection sampler on square-root scale: $\mathcal{S}_{\mathcal{N}}$)

Input: Parameters $\alpha \geq 1/2$ and $\gamma \in \mathbb{R}$
if $\alpha = 1/2$ **then**
  $m \leftarrow \max(-\gamma, 0)$
else
  $m \leftarrow \frac{2\alpha - 1}{\gamma + \sqrt{\gamma^2 + 4\alpha - 2}}$
end if
repeat
  Generate $X \sim \mathcal{N}(m, 1/2)$ and $U \sim \text{Uniform}(0, 1)$
until $X > 0$ and $m^{2\alpha-1} \cdot U < X^{2\alpha-1} \exp\{-2(m + \gamma)(X - m)\}$
Output: Squared accepted draw X^2.

c) Gamma distributions on square-root scale. For $\gamma \to \infty$, the density $h(\cdot \mid \alpha, \gamma)$ can be approximated by a Gamma distribution. In an approach that works for all $\alpha > 0$, we use Gamma instruments $\Gamma(r, \delta)$ to sample from $h(\cdot \mid \alpha, \gamma)$. A computer study involving a broad range of α's and γ's suggests that in this setting the best choice of the shape parameter is $r = 2\alpha$, regardless of the value of γ. The optimal choice of the scale parameter is then

$$\delta_1(\alpha, \gamma) := \gamma + \sqrt{\gamma^2 + 4\alpha} = \frac{4\alpha}{\sqrt{\gamma^2 + 4\alpha} - \gamma}. \tag{1.8}$$

Pseudo-code for this sampler $\mathcal{S}_{\mathcal{G}}$ is given in Algorithm 3. The associated acceptance rate $A_{\mathcal{G}}(\alpha, \gamma)$ is strictly increasing in γ and again graphed in Figure 3.

Algorithm 3 (Gamma rejection sampler on square-root scale: $\mathcal{S}_{\mathcal{G}}$)

Input: Parameters $\alpha > 0$ and $\gamma \in \mathbb{R}$
$\delta_1 \leftarrow \gamma + \sqrt{\gamma^2 + 4\alpha}$
$r \leftarrow 2\alpha$
repeat
  Generate $X \sim \Gamma(r, \delta_1)$ and $U \sim$Uniform(0,1)
until $U < \exp(-(X - (\delta_1/2 - \gamma))^2)$
Output: Accepted draw X^2.

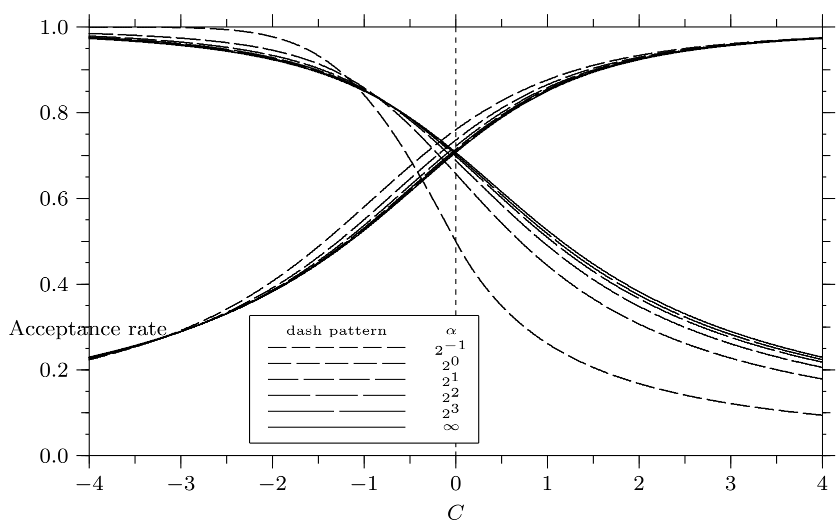

FIG 3. *Graphs of the acceptance rates $A_{\mathcal{N}}(\alpha, C\sqrt{\alpha})$ and $A_{\mathcal{G}}(\alpha, C\sqrt{\alpha})$ versus C for $\alpha = 1/2$, 1, 2, 4, 8, and in the limit as $\alpha \to \infty$, when using normal and Gamma instrumental distributions on the square-root scale. The $A_{\mathcal{N}}$ curves slope downward, the $A_{\mathcal{G}}$ curves upward. See (3.12) and (3.13) for the case $\alpha < \infty$ and (5.5) and (5.12) for the limiting cases.*

1.3. Optimal combination

The Gamma instrumental distributions on the original scale yield high acceptance rates when $C = \gamma/\sqrt{\alpha}$ is of small magnitude (Figure 2). Using the square-root scale works best for C of large magnitude with the Gamma and the normal instruments yielding large acceptance rates for C positive and C negative, respectively (Figure 3). Recall, however, that normal instruments are applicable only for $\alpha \geq 1/2$. Clearly, it is beneficial to combine all three approaches.

Due to the monotonicity of the acceptance rates as a function of γ, optimal combination requires determination of two thresholds for γ, one negative and one positive. Equivalently, we may specify the thresholds in terms of C. Then sampler $\mathcal{S}_0$ is applied when $C \in (\mathfrak{C}_{\mathcal{N},0-}(\alpha), \mathfrak{C}_{0+,\mathcal{G}}(\alpha))$, sampler $\mathcal{S}_\mathcal{N}$ when $C \leq \mathfrak{C}_{\mathcal{N},0-}(\alpha)$, and sampler $\mathcal{S}_\mathcal{G}$ when $C \geq \mathfrak{C}_{0+,\mathcal{G}}(\alpha)$; compare Panel 1 of Figure 4 that shows graphs of all three acceptance rates for the smaller range of $|C| \leq 2$. The thresholds $\mathfrak{C}_{\mathcal{N},0-}(\alpha)$ and $\mathfrak{C}_{0+,\mathcal{G}}(\alpha)$ that optimize the acceptance rate can be precomputed as functions of $1/\alpha$ on a fine grid from $(0, 2]$; the graphs of the functions are shown in Panel 2 of Figure 4. In particular, the thresholds for C vary only modestly with α. The combination of $\mathcal{S}_0$, $\mathcal{S}_\mathcal{N}$ and $\mathcal{S}_\mathcal{G}$ achieves high acceptance rates uniformly across all values of (α, γ) with $\alpha \geq 1/2$. As can be seen by studying the top two curves in Panel 3 of Figure 4, the acceptance rate for our algorithm never drops below 0.8. For C close to zero or of moderately large value, the rate will be close to one; see again Panel 1 of Figure 4.

1.4. Outline of the paper

In the remainder of the paper we first prove correctness of the three rejection samplers from Section 1.2. The sampler $\mathcal{S}_0$ with Gamma instrumental distributions on the original scale is treated in Section 2. The analysis of this sampler separates the cases $\gamma > 0$ and $\gamma < 0$, and we indicate this in our notation by referring to two samplers $\mathcal{S}_{0+}$ and $\mathcal{S}_{0-}$. The two samplers $\mathcal{S}_\mathcal{N}$ and $\mathcal{S}_\mathcal{G}$ that use normal and Gamma instruments on the square-root scale are treated in Section 3. Next, in Section 4, we give some more details on the optimal combination of the three samplers $\mathcal{S}_0$, $\mathcal{S}_\mathcal{N}$ and $\mathcal{S}_\mathcal{G}$ that was described in Section 1.3.

The optimal combination relies on monotonocity properties of the acceptance rates of $\mathcal{S}_0$, $\mathcal{S}_\mathcal{N}$ and $\mathcal{S}_\mathcal{G}$ that are suggested by Figures 2 and 3. Section 5 establishes those properties rigorously and quantifies them. Section 6 carries the analysis further by developing asymptotic expansions for the samplers' acceptance rates $A(\alpha, \gamma)$. An expansion is given for $A(\alpha, C\sqrt{\alpha}\,)$ as α tends to infinity with C held fixed, and also as C tends to infinity with α held fixed.

Section 7 addresses the case $\alpha < 1/2$. When $\gamma > 0$, the acceptance rate of $\mathcal{S}_\mathcal{G}$ tends to 1 uniformly in γ as $\alpha \downarrow 0$. Unfortunately, when $\gamma < 0$, the acceptance rate of our best sampler falls off quickly as γ increases in magnitude. These properties are illustrated in Figure 6. Section 8 gives some concluding remarks.

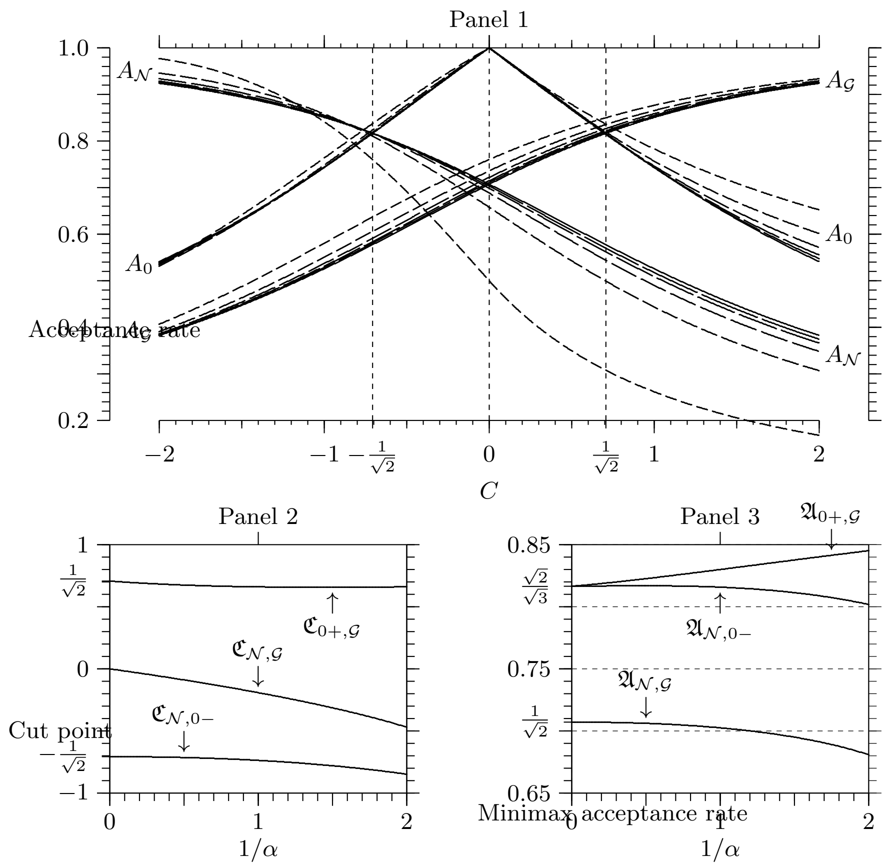

FIG 4. *Panel 1 graphs the acceptance rates $A_\mathcal{N}(\alpha, C\sqrt{\alpha})$, $A_\mathcal{G}(\alpha, C\sqrt{\alpha})$, and $A_0(\alpha, C\sqrt{\alpha})$ versus C for $\alpha = 1/2$, 1, 2, 4, 8, and in the limit as $\alpha \to \infty$, using the same dash patterns as in Figures 2 and 3. For $\alpha \geq 1/2$ and $(\mathcal{S}_1, \mathcal{S}_2) = (\mathcal{S}_\mathcal{N}, \mathcal{S}_\mathcal{G})$, $(\mathcal{S}_\mathcal{N}, \mathcal{S}_{0-})$, and $(\mathcal{S}_{0+}, \mathcal{S}_\mathcal{G})$, Panel 2 graphs versus $1/\alpha$ the number $C = \mathfrak{C}_{\mathcal{S}_1,\mathcal{S}_2}(\alpha)$ which minimizes the maximum of $A_{\mathcal{S}_1}(\alpha, C\sqrt{\alpha})$ and $A_{\mathcal{S}_2}(\alpha, C\sqrt{\alpha})$. Similarly, Panel 3 plots the minimax acceptance rate $\mathfrak{A}_{\mathcal{S}_1,\mathcal{S}_2}(\alpha) := A_{\mathcal{S}_1}(\alpha, \mathfrak{C}_{\mathcal{S}_1,\mathcal{S}_2}(\alpha)\sqrt{\alpha}) = A_{\mathcal{S}_2}(\alpha, \mathfrak{C}_{\mathcal{S}_1,\mathcal{S}_2}(\alpha)\sqrt{\alpha})$ versus $1/\alpha$.*

2. Rejection samplers on the original scale

Let $\mathcal{G}(r, \delta)$ be the Gamma distribution with shape parameter r and mean r/δ, that is, $\mathcal{G}(r, \delta)$ has density

$$g(t \,|\, r, \delta) := \frac{\delta^r}{\Gamma(r)} t^{r-1} \exp(-\delta t), \qquad t > 0. \tag{2.1}$$

Here we consider rejection samplers for $f(\,\cdot\,|\,\alpha, \gamma)$ with Gamma instrumental distributions. A computer study involving a broad range of α's and γ's suggests that

$$M := \sup_{t>0}\Big\{\frac{f(t \,|\, \alpha, \gamma)}{g(t \,|\, r, \delta)}\Big\} = \sup_{t>0}\Big(\frac{\Gamma(r)}{2Z(\alpha, \gamma)} \frac{t^{\alpha - r}}{\delta^r} \exp\big((\delta - 1)t - 2\gamma\sqrt{t}\,\big)\Big) \tag{2.2}$$

is minimized by taking $r = \alpha$ when $\gamma \le 0$, whereas when $\gamma \ge 0$, M is minimized by taking $\delta = 1$. Accordingly, in what follows we will only consider those situations; we will choose the remaining free parameter optimally.

Theorem 2.1. *Let $\alpha > 0$ and $\gamma \in \mathbb{R}$ be given. Define f by (1.1) and g by (2.1). If $\gamma < 0$, then it holds for $\delta \in (0, \infty)$ that*

$$M_{0,-}(\delta) := \sup_{t>0}\Big\{\frac{f(t \,|\, \alpha, \gamma)}{g(t \,|\, \alpha, \delta)}\Big\} = \begin{cases} \dfrac{\Gamma(\alpha)\exp(\gamma^2/(1-\delta))}{2Z(\alpha, \gamma)\delta^\alpha}, & \text{if } 0 < \delta < 1, \\ \infty, & \text{if } \delta \ge 1; \end{cases} \tag{2.3}$$

$M_{0,-}(\delta)$ is uniquely minimized by taking δ to be

$$\delta_0(\alpha, \gamma) := \frac{\gamma^2 + 2\alpha - \sqrt{\gamma^4 + 4\alpha\gamma^2}}{2\alpha} = \frac{4\alpha}{\big(\sqrt{\gamma^2 + 4\alpha} + |\gamma|\,\big)^2}; \tag{2.4}$$

the minimum is

$$M_{0,-}(\alpha, \gamma) := \frac{\Gamma(\alpha)}{2Z(\alpha, \gamma)e^\alpha}\Big(\frac{\exp(1/\delta)}{\delta}\Big)^\alpha \tag{2.5}$$

for $\delta = \delta_0(\alpha, \delta)$. If $\gamma > 0$, then it holds for $r \in (0, \infty)$ that

$$M_{0,+}(r) := \sup_{t>0}\Big\{\frac{f(t \,|\, \alpha, \gamma)}{g(t \,|\, r, 1)}\Big\} = \begin{cases} \dfrac{\Gamma(r)}{2Z(\alpha, \gamma)}\Big(\dfrac{\alpha - r}{\gamma e}\Big)^{2(\alpha - r)}, & \text{if } 0 < r \le \alpha, \\ \infty, & \text{if } r > \alpha, \end{cases} \tag{2.6}$$

with the convention that $0^0 = 1$ if $r = \alpha$; for $r \in (0, \alpha)$, the function

$$s(\theta) := \frac{1}{\alpha}\frac{d}{d\theta}\log\big(M_{0,+}(\theta\alpha)\big) = \Big(\frac{d}{dr}\log\Gamma(r)\Big)\Big|_{r=\theta\alpha} - 2\log\Big(\frac{(1-\theta)\alpha}{\gamma}\Big) \tag{2.7}$$

is strictly increasing in $\theta \in (0, 1)$ with $s(0+) = -\infty$ and $s(1-) = \infty$; $M_{0,+}(r)$ is uniquely minimized by taking $r = r(\alpha, \gamma) = \theta\alpha$, where $\theta = \theta(\alpha, \gamma)$ is the root of the equation $s(\theta) = 0$.

Proof. Suppose first that $\gamma < 0$. Ignoring additive constants in t, the logarithm of $f(t\,|\,\alpha,\gamma)/g(t\,|\,\alpha,\delta)$ is

$$L(t) = (\delta - 1)t - 2\sqrt{t}\gamma.$$

For $\delta \geq 1$, the function $L(t)$ is unbounded above over $(0, \infty)$. For $\delta \in (0, 1)$, the derivative

$$\frac{dL(t)}{dt} = (\delta - 1) - \frac{\gamma}{\sqrt{t}}$$

is decreasing in $t \in (0, \infty)$ and equals 0 at $t_0 = \gamma^2/(1-\delta)^2$. Thus $f(t\,|\,\alpha,\gamma)/g(t\,|\,\alpha,\delta)$ is a maximum at $t = t_0$ and (2.3) holds.

For $\delta \in (0, 1)$, $\Lambda(\delta) := \log\big(M_{0,-}(\delta)\big)$ has derivative

$$\Lambda'(\delta) = \gamma^2/(1 - \delta)^2 - \alpha/\delta,$$

$\Lambda'(\delta)$ is strictly increasing in $\delta \in (0, 1)$, and the quadratic equation $\Lambda'(\delta) = 0$ has the root $\delta_0 = \delta_0(\alpha, \gamma)$ given by (2.4). Therefore, $M_{0,-}(\delta)$ has a unique minimum at δ_0. Since $\Lambda'(\delta_0) = 0$, we have $\gamma^2/(1 - \delta_0) = \alpha(1 - \delta_0)/\delta_0$; thus (2.5) holds.

Next suppose that $\gamma > 0$. Here

$$L(t) := \log\big(f(t\,|\,\alpha,\gamma)/g(t\,|\,r,1)\big) = C + (\alpha - r)\log(t) - 2\gamma\sqrt{t}$$

for $C = \log\big(\Gamma(r)/(2Z(\alpha,\gamma))\big)$. For $r > \alpha$, $L(t)$ is unbounded above over $(0, \infty)$. For $r = \alpha$, $\sup_{t>0} L(t) = L(0-) = C$. For $r \in (0, \alpha)$,

$$L'(t) = \frac{\alpha - r}{t} - \frac{\gamma}{\sqrt{t}} \quad \text{and} \quad L''(t) = -\frac{\alpha - r}{t^2} + \frac{1}{2}\frac{\gamma}{t^{3/2}}.$$

The equation $L'(t) = 0$ has the unique root $t_0 = (\alpha - r)^2/\gamma^2$ and $L''(t_0) < 0$, so $L(t)$ is a maximum at $t = t_0$. Thus (2.6) holds, with the usual convention that $0^0 = 1$.

For $\theta \in (0, 1]$,

$$\Lambda(\theta) := \log\big(M_{0,+}(\theta\alpha)\big) = -\log(2Z(\alpha,\gamma)) + (\log\Gamma)(\theta\alpha) + 2\alpha(1-\theta)\log\big((1-\theta)c/e\big)$$

for $c = \alpha/\gamma$. For $\theta \in (0, 1]$, the function $\Lambda(\theta)$ is continuous in θ. For $\theta \in (0, 1)$,

$$s(\theta) := \frac{1}{\alpha}\Lambda'(\theta) = (\log\Gamma)'(\theta\alpha) - 2\log\big((1-\theta)c\big),$$
$$s'(\theta) = \alpha(\log\Gamma)''(\theta\alpha) + \frac{2}{1-\theta}.$$

Since $(\log\Gamma)''(r) > 0$, see (2.89) in Andrews (1985), s is strictly increasing over $(0, 1)$. Since $s(0+) = -\infty$ and $s(1-) = \infty$, $\Lambda(\theta)$ has a unique minimum at the root $\theta = \theta(\alpha, \gamma)$ to the equation $s(\theta) = 0$. □

In what follows we let $A_0(\alpha,\gamma)$ be the acceptance rate for the rejection sampler with target density $f(\cdot\,|\,\alpha,\gamma)$ and instrumental distribution $\mathcal{G}(r,\delta)$ with

$$\begin{cases} r=\alpha,\ \delta=\delta_0(\alpha,\gamma), & \text{if } \gamma<0,\\ r=\alpha,\ \delta=1, & \text{if } \gamma=0,\\ r=\alpha\theta(\alpha,\gamma),\ \delta=1, & \text{if } \gamma>0, \end{cases} \tag{2.8}$$

where $\delta_0(\alpha,\gamma)$ and $\theta(\alpha,\gamma)$ are as specified in Theorem 2.1. Thus

$$A_0(\alpha,\gamma)=\begin{cases} A_{0-}(\alpha,\gamma):=\dfrac{2Z(\alpha,\gamma)e^{\alpha}}{\Gamma(\alpha)}\Big(\dfrac{\delta}{\exp(1/\delta)}\Big)^{\alpha}, & \text{if } \gamma<0,\\ 1, & \text{if } \gamma=0,\\ A_{0+}(\alpha,\gamma):=\dfrac{2Z(\alpha,\gamma)}{\Gamma(r)}\Big(\dfrac{\gamma e}{\alpha-r}\Big)^{2(\alpha-r)}, & \text{if } \gamma>0, \end{cases} \tag{2.9}$$

where $\delta=\delta_0(\alpha,\gamma)$ and $r=\theta(\alpha,\gamma)\alpha$. Note that

$$\lim_{\gamma\uparrow 0}\delta_0(\alpha,\gamma)=1=\lim_{\gamma\downarrow 0}\theta(\alpha,\gamma)\ \text{ and }\ \lim_{\gamma\uparrow 0}A_{0-}(\alpha,\gamma)=A_0(\alpha,0)=\lim_{\gamma\downarrow 0}A_{0+}(\alpha,\gamma);$$

thus the case $\gamma=0$ can be regarded as a limiting form of the case $\gamma<0$, and also of the case $\gamma>0$. Recall that $A_0(\alpha,C\sqrt{\alpha}\,)$ is plotted versus C for selected α's in Figure 2. The plots suggest various properties of A_0 — for example, decay of $A_0(\alpha,C\sqrt{\alpha}\,)$ as C increases in magnitude — which are developed in Sections 5 and 6. For future reference, we'll denote the rejection sampler for $\gamma\le 0$ by $\mathcal{S}_{0-}=\mathcal{S}_{0-}(\alpha,\gamma)$, and the one for $\gamma\ge 0$ by $\mathcal{S}_{0+}=\mathcal{S}_{0+}(\alpha,\gamma)$. Sampler $\mathcal{S}_{0-}$ has acceptance rate A_{0-}, and $\mathcal{S}_{0+}$ has acceptance rate A_{0+}. Panel 1 in Figure 5 plots $\theta(\alpha,C\sqrt{\alpha}\,)$ versus C for selected increasing α's. The message here is that for moderate to large α, $\theta(\alpha,C\sqrt{\alpha}\,)$ is small, except when C is near 0. Panel 2 in the figure plots

$$\tau(\alpha,\gamma):=\alpha\exp\big(1/(2\alpha)-\Gamma'(1)/2\big)\times\big(1-\theta(\alpha,\gamma)\big)$$

versus γ for selected α's tending to 0. The message here is that for small α, $\tau(\alpha,\gamma)$ is about γ, whence $\theta(\alpha,\gamma)$ must be *extremely* close to 1 and for practical purposes can be taken to be 1.

3. Rejection samplers on the square-root scale

Recall from (1.6) that

$$h(x\,|\,\alpha,\gamma)=\frac{1}{Z(\alpha,\gamma)}\,x^{2\alpha-1}\exp\big(-x^2-2\gamma x\big)$$

for $x>0$ (and for $x=0$ when $\alpha\ge 1/2$). Let

$$\varphi(x\,|\,\mu,\sigma^2):=\frac{1}{\sqrt{2\pi\sigma^2}}\,\exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big) \tag{3.1}$$

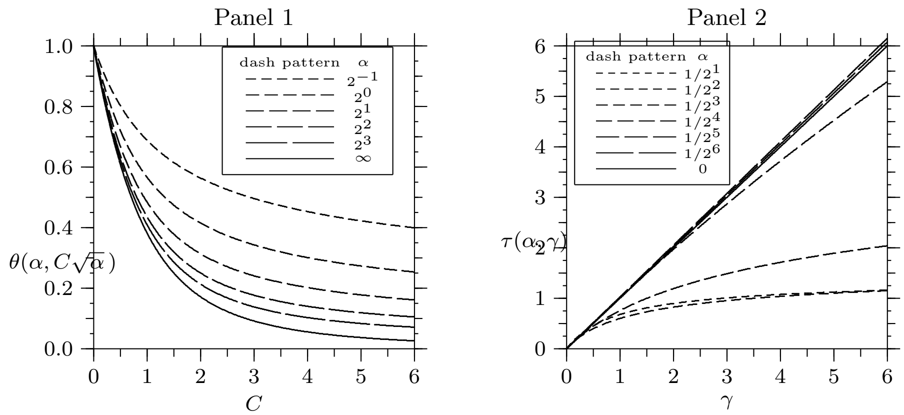

FIG 5. *The rejection sampler $\mathcal{S}_{0+}(\alpha,\gamma)$ uses a Gamma instrumental distribution with shape parameter $r=\alpha\theta(\alpha,\gamma)$; see (2.8). Panel 1 graphs $\theta(\alpha, C\sqrt{\alpha})$ versus C for $\alpha=1/2$, 1, 2, 4, 8, and in the limit as $\alpha\to\infty$; see (5.28). Panel 2 graphs $\tau(\alpha,\gamma):=\alpha\exp\big(1/(2\alpha)-\Gamma'(1)/2\big)\times\big(1-\theta(\alpha,\gamma)\big)$ versus γ for $\alpha=1/2^1, 1/2^2, \dots, 1/2^6$, and 0, with $\tau(0,\gamma):=\lim_{\alpha\downarrow 0}\tau(\alpha,\gamma)=\gamma$.*

be the density of the normal distribution $\mathcal{N}(\mu,\sigma^2)$ with mean μ and variance σ^2. The next theorem considers rejection samplers with target $h(\,\cdot\,|\,\alpha,\gamma)$ and normal instrumental distributions. Since the ratio

$$R(x)=R(x\,|\,\alpha,\gamma,\mu,\sigma^2):=\frac{h(x\,|\,\alpha,\gamma)}{\varphi(x\,|\,\mu,\sigma^2)} \tag{3.2}$$

is unbounded above over $(0,\infty)$ for $\alpha<1/2$, this approach requires $\alpha\ge 1/2$. For such α's, the following theorem says that the optimal choice of the parameters of the instrumental normal distribution are $\mu=m$ and $\sigma^2=1/2$, where m is the mode of $h(\,\cdot\,|\,\alpha,\gamma)$. In particular, the target $h(\,\cdot\,|\,\alpha,\gamma)$ and the matching normal have the same mode.

For $\alpha>1/2$, the mode of $h(\,\cdot\,|\,\alpha,\gamma)$ is

$$m=m(\alpha,\gamma)=\frac{-\gamma+\sqrt{\gamma^2+4\alpha-2}}{2}=\frac{2\alpha-1}{\big(\gamma+\sqrt{\gamma^2+4\alpha-2}\,\big)}, \tag{3.3}$$

which is the positive root of the quadratic equation

$$q(x)=0 \quad \text{where} \quad q(x):=-\frac{1}{2}\frac{d}{dx}\log\big(h(x\,|\,\alpha,\gamma)\big)=x+\gamma-\frac{2\alpha-1}{2x}. \tag{3.4}$$

For $\alpha=1/2$, the mode is

$$m=m(1/2,\gamma)=\gamma^-:=\max(-\gamma,0). \tag{3.5}$$

For $\alpha<1/2$, the mode is at 0. Note that $m(\alpha,\gamma)$ is a decreasing function of γ.

Theorem 3.1. *Let $\alpha \geq 1/2$ and $\gamma \in \mathbb{R}$ be given. Define h by (1.6), φ by (3.1), m by (3.3) and (3.5), put*

$$M_{\mathcal{N}}(\mu, \sigma^2) := \sup_{x \geq 0} \frac{h(x \,|\, \alpha, \gamma)}{\varphi(x \,|\, \mu, \sigma^2)}, \tag{3.6}$$

and set $\sigma_0^2 := 1/2$. Then $M_{\mathcal{N}}(\mu, \sigma^2)$ is uniquely minimized by taking $\mu = m$ and $\sigma = \sigma_0$. Moreover,

$$M_{\mathcal{N}}(m, \sigma_0^2) = \frac{h(m \,|\, \alpha, \gamma)}{\varphi(m \,|\, m, \sigma_0^2)} = \frac{\sqrt{\pi}\, m^{2\alpha-1} \exp(-m^2 - 2\gamma m)}{Z(\alpha, \gamma)} \tag{3.7}$$

with the convention that $m^{2\alpha-1} = 1$ when $\alpha = 1/2$ and $m = 0$.

Proof. First consider $M_{\mathcal{N}}(m, \sigma_0^2)$. Suppose $\alpha > 1/2$. The ratio $R(x)$ in (3.2) is then proportional to

$$x^{2\alpha-1} \exp(-x^2 - 2\gamma x + (x-m)^2).$$

Hence, the maximum of $R(x)$ is attained at the value of x that maximizes

$$L(x) = (2\alpha - 1)\log(x) - x^2 - 2\gamma x + (x-m)^2.$$

Since $\alpha > 1/2$, the derivative

$$L'(x) = \frac{2\alpha - 1}{x} - 2x - 2\gamma + 2(x - m) = \frac{2\alpha - 1}{x} - 2\gamma - 2m$$

is decreasing in $x \in (0, \infty)$ and zero at $x = m$; recall (3.4). Therefore, $M_{\mathcal{N}}(m, \sigma_0^2) = R(m)$.

The same conclusion holds when $\alpha = 1/2$, since then

$$L'(x) = -2(\gamma + m) = \begin{cases} 0, & \text{if } \gamma \leq 0, \\ -2\gamma, & \text{if } \gamma > 0. \end{cases}$$

Thus when $\gamma \leq 0$, $L(x)$ is constant and has a maximum at $x = -\gamma = m$, whereas when $\gamma > 0$, $L(x)$ is decreasing in x and has a maximum at $x = 0 = m$.

Now consider general μ and σ^2. If $\sigma^2 < 1/2$, then $M_{\mathcal{N}}(\mu, \sigma^2) = \infty$. If $\sigma^2 \geq 1/2$,

$$\begin{aligned} M_{\mathcal{N}}(\mu, \sigma^2) &\geq \frac{h(m \,|\, \alpha, \gamma)}{\varphi(m \,|\, \mu, \sigma^2)} = \frac{h(m \,|\, \alpha, \gamma)}{\varphi(m \,|\, m, \sigma_0^2)} \times \frac{\varphi(m \,|\, m, \sigma_0^2)}{\varphi(m \,|\, \mu, \sigma^2)} \\ &= M_{\mathcal{N}}(m, \sigma_0^2) \times \Big(\frac{\sigma}{\sigma_0}\Big) \exp\Big(\frac{(m-\mu)^2}{2\sigma^2}\Big) \geq M_{\mathcal{N}}(m, \sigma_0^2). \end{aligned} \tag{3.8}$$

The final inequality is strict unless $\sigma = \sigma_0$ and $\mu = m$. □

The next theorem considers rejection samplers for $h(\,\cdot\,|\,\alpha,\gamma)$ with Gamma instrumental distributions. A computer study involving a broad range of α's and δ's suggests that in this setting the best choice of the shape parameter r for the Gamma distribution is 2α, regardless of the value of γ. Accordingly, the following theorem just deals with the case $r = 2\alpha$. Here we don't need to require $\alpha \geq 1/2$.

Theorem 3.2. *Let $\alpha > 0$ and $\gamma \in \mathbb{R}$ be given. Define h by (1.6) and g by (2.1). Then*

$$M_{\mathcal{G}}(\delta) := \sup_{x>0} \frac{h(x\,|\,\alpha,\gamma)}{g(x\,|\,2\alpha,\delta)} = \frac{\Gamma(2\alpha)}{Z(\alpha,\gamma)}\frac{1}{\delta^{2\alpha}}\left(\begin{cases}\exp\big((\delta/2-\gamma)^2\big), & \text{if } \delta/2 \geq \gamma\\ 1, & \text{otherwise}\end{cases}\right) \tag{3.9}$$

is uniquely maximized by taking δ to be

$$\delta_1(\alpha,\gamma) := \gamma + \sqrt{\gamma^2+4\alpha} = \frac{4\alpha}{\sqrt{\gamma^2+4\alpha}-\gamma}. \tag{3.10}$$

Note that it always the case that $\delta_1(\alpha,\gamma)/2 > \gamma$. Also, when $\alpha > 1/2$, the instrumental distribution $G\big(2\alpha,\delta_1(\alpha,\gamma)\big)$ has mode $(2\alpha-1)/\delta_1(\alpha,\gamma)$, which is smaller than the mode of the target $h(\,\cdot\,|\,\alpha,\gamma)$, recall (3.3), but only slightly so when α is large.

Proof. The ratio

$$\frac{h(x\,|\,\alpha,\gamma)}{g(x\,|\,2\alpha,\delta)} = \frac{\Gamma(2\alpha)}{Z(\alpha,\gamma)\delta^{2\alpha}}\,\exp\big((\delta/2-\gamma)^2\big)\exp\Big[-\big(x-(\delta/2-\gamma)\big)^2\Big]$$

is obviously maximized over $x \in [0,\infty)$ by taking x to be $x_0 := \delta/2-\gamma$ if $x_0 \geq 0$, and otherwise 0. Therefore (3.9) holds:

$$L(\delta) := \log\big(M_{\mathcal{G}}(\delta)\big) = C - 2\alpha\log(\delta) + \big[(\delta/2-\gamma)\vee 0\,\big]^2$$

for $C = \log\big(\Gamma(2\alpha)/Z(\alpha,\gamma)\big)$. To make the best choice of $\delta \in (0,\infty)$, put

$$L_\bullet(\delta) := C - 2\alpha\log(\delta) + (\delta/2-\gamma)^2.$$

Since

$$L'_\bullet(\delta) = -2\alpha/\delta + \delta/2 - \gamma, \tag{3.11}$$

$L_\bullet(\delta)$ is uniquely minimized over $(0,\infty)$ by taking δ to be the root $\delta_1 := \delta_1(\alpha,\gamma)$ of the equation $L'_\bullet(\delta) = 0$. We claim that $\delta = \delta_1$ also uniquely minimizes $L(\delta)$. That's obvious when $\gamma \leq 0$, since then $L = L_\bullet$. When $\gamma > 0$ it holds because $L(\delta) = L_\bullet(\delta)$ for $\delta \geq 2\gamma$ while for $\delta < 2\gamma$ we have $L(\delta) > L(2\gamma) = L_\bullet(2\gamma) > L_\bullet(\delta_1)$. □

In the light of Theorems 3.1 and 3.2, we have two rejection samplers for $h(\,\cdot\,|\,\alpha,\gamma)$ with instrumental distributions $\mathcal{N}(m,1/2)$ and $\mathcal{G}(2\alpha,\delta)$, respectively, where $m = m(\alpha,\gamma)$ is given by (3.3) and (3.5) and $\delta = \delta_1(\alpha,\gamma)$ is given by (3.10).

For future reference, we'll denote these rejection samplers by $\mathcal{S}_{\mathcal{N}} = \mathcal{S}_{\mathcal{N}}(\alpha,\gamma)$ and $\mathcal{S}_{\mathcal{G}} = \mathcal{S}_{\mathcal{G}}(\alpha,\gamma)$, respectively. The corresponding acceptance rates are

$$A_{\mathcal{N}}(\alpha,\gamma) = \frac{Z(\alpha,\gamma)}{\sqrt{\pi}m^{2\alpha-1}\exp(-m^2-2\gamma m)} \tag{3.12}$$

and

$$A_{\mathcal{G}}(\alpha,\gamma) = \frac{Z(\alpha,\gamma)}{\Gamma(2\alpha)}\frac{\delta^{2\alpha}}{\exp\big((\delta/2-\gamma)^2\big)} = \frac{Z(\alpha,\gamma)}{\Gamma(2\alpha)}\frac{\delta^{2\alpha}}{\exp\big((2\alpha/\delta)^2\big)}, \tag{3.13}$$

the second equality in (3.13) holding because $L'_{\bullet}(\delta) = 0$; recall (3.11). The Gamma sampler $\mathcal{S}_{\mathcal{G}}$ is available for all $\alpha > 0$, but the normal sampler $\mathcal{S}_{\mathcal{N}}$ only for $\alpha \ge 1/2$. Figure 3 graphs $A_{\mathcal{N}}$ and $A_{\mathcal{G}}$ in the same manner that A_0 is graphed in Figure 2. We observe that for all $\alpha \ge 1/2$, $A_{\mathcal{N}}(\alpha,\gamma)$ is uniformly of practically useful size when $\gamma \le 0$ and the same is true for $A_{\mathcal{G}}(\alpha,\gamma)$ when $\gamma \ge 0$.

4. Combining the rejection samplers

Consider the two rejection samplers $\mathcal{S}_{\mathcal{N}}$ and $\mathcal{S}_{\mathcal{G}}$ for making a draw from h. We may combine these two samplers, using in any particular instance the one with the larger acceptance rate. As Figure 3 suggests, and the analysis in the following section confirms, for any fixed $\alpha \ge 1/2$, as C runs from $-\infty$ to ∞ the acceptance rate for $\mathcal{S}_{\mathcal{N}}(\alpha, C\sqrt{\alpha})$ decreases from 1 to 0, while the rate for $\mathcal{S}_{\mathcal{G}}(\alpha, C\sqrt{\alpha})$ increases from 0 to 1. Therefore there is a cut-point $\mathfrak{C}_{\mathcal{N},\mathcal{G}}(\alpha)$ such that $\mathcal{S}_{\mathcal{N}}(\alpha, C\sqrt{\alpha})$ is more efficient than $\mathcal{S}_{\mathcal{G}}(\alpha, C\sqrt{\alpha})$ for $C < \mathfrak{C}_{\mathcal{N},\mathcal{G}}(\alpha)$, while the opposite holds for $C > \mathfrak{C}_{\mathcal{N},\mathcal{G}}(\alpha)$. The cut-point $\mathfrak{C}_{\mathcal{N},\mathcal{G}}(\alpha)$ is the unique solution to the equation

$$1 = \frac{A_{\mathcal{N}}(\alpha, C\sqrt{\alpha})}{A_{\mathcal{G}}(\alpha, C\sqrt{\alpha})} = \frac{\Gamma(2\alpha)\exp\big((2\alpha/\delta)^2\big)}{\sqrt{\pi}m^{2\alpha-1}\exp\big(-m^2-2mC\sqrt{\alpha}\big)\,\delta^{2\alpha}} \tag{4.1}$$

where $m = m(\alpha,\gamma)$ and $\delta = \delta_1(\alpha,\gamma)$; recall (3.3) and (3.10). This equation does not have a closed form solution, but it is easily solved numerically. The values of $\mathfrak{C}_{\mathcal{N},\mathcal{G}}(\alpha)$ are plotted versus $1/\alpha$ for $\alpha \ge 1/2$ in Panel 2 of Figure 4, which shows that as α runs from 1/2 to ∞, $\mathfrak{C}_{\mathcal{N},\mathcal{G}}(\alpha)$ increases from about -0.47 to 0. The combined sampler, call it $\mathcal{S}_{\mathcal{N}\vee\mathcal{G}}$, has acceptance rate

$$A_{\mathcal{N}\vee\mathcal{G}}(\alpha, C\sqrt{\alpha}) = \begin{cases} A_{\mathcal{N}}(\alpha, C\sqrt{\alpha}), & \text{if } C \le \mathfrak{C}_{\mathcal{N},\mathcal{G}}(\alpha), \\ A_{\mathcal{G}}(\alpha, C\sqrt{\alpha}), & \text{if } C \ge \mathfrak{C}_{\mathcal{N},\mathcal{G}}(\alpha). \end{cases} \tag{4.2}$$

$A_{\mathcal{N}\vee\mathcal{G}}(\alpha, C\sqrt{\alpha})$ is smallest at $C = \mathfrak{C}_{\mathcal{N},\mathcal{G}}(\alpha)$. Panel 3 in Figure 4 graphs the worst case acceptance rate $\mathfrak{A}_{\mathcal{N},\mathcal{G}}(\alpha) := A_{\mathcal{N}\vee\mathcal{G}}(\alpha, \mathfrak{C}_{\mathcal{N},\mathcal{G}}(\alpha)\sqrt{\alpha})$ versus $1/\alpha$ for $\alpha \ge 1/2$. The plot shows that as α runs from 1/2 to ∞, $\mathfrak{A}_{\mathcal{N},\mathcal{G}}(\alpha)$ increases from a minimum of about 0.68 at $\alpha = 1/2$ to about 0.71 at $\alpha = \infty$. In fact, it follows from (5.5) and (5.12) below that $\mathfrak{C}_{\mathcal{N},\mathcal{G}}(\infty-) = 0$ and $\mathfrak{A}_{\mathcal{N},\mathcal{G}}(\infty-) = 1/\sqrt{2}$. As

Figure 3 shows, $\mathcal{S}_{\mathcal{N}\vee\mathcal{G}}$ is very efficient for moderate and large C's; in particular, $A_{\mathcal{N}\vee\mathcal{G}}(\alpha, C\sqrt{\alpha}\,) \geq 0.957$ for all $|C| \geq 3$ and all $\alpha \geq 1/2$.

The rejection samplers $\mathcal{S}_{0-}$ and $\mathcal{S}_{0+}$ for f on the original scale perform well for the parameter values for which $\mathcal{S}_{\mathcal{N}\vee\mathcal{G}}$ performs least well. It is natural then to combine these three samplers (squaring a draw from $\mathcal{S}_{\mathcal{N}\vee\mathcal{G}}$ to get back to the original scale), using whichever is the more efficient. As Panel 1 in Figure 4 suggests, and the analysis in the following section confirms, for any fixed $\alpha \geq 1/2$, the most efficient combination of our various rejection samplers uses

$$\begin{Bmatrix} \mathcal{S}_{\mathcal{N}}(\alpha, C\sqrt{\alpha}\,) \\ \mathcal{S}_{0-}(\alpha, C\sqrt{\alpha}\,) \\ \mathcal{S}_{0+}(\alpha, C\sqrt{\alpha}\,) \\ \mathcal{S}_{\mathcal{G}}(\alpha, C\sqrt{\alpha}\,) \end{Bmatrix} \text{ for } C \text{ in } \begin{Bmatrix} (-\infty, \mathfrak{C}_{\mathcal{N},0-}(\alpha)] \\ [\mathfrak{C}_{\mathcal{N},0-}(\alpha), 0] \\ [0, \mathfrak{C}_{0+,\mathcal{G}}(\alpha)] \\ [\mathfrak{C}_{0+,\mathcal{G}}(\alpha), \infty) \end{Bmatrix}, \tag{4.3}$$

where $\mathfrak{C}_{\mathcal{N},0-}(\alpha)$ is the unique solution C to the equation

$$1 = \frac{A_{\mathcal{N}}(\alpha, C\sqrt{\alpha}\,)}{A_{0-}(\alpha, C\sqrt{\alpha}\,)} \tag{4.4}$$

for $C < 0$, and similarly $\mathfrak{C}_{0+,\mathcal{G}}(\alpha)$ is the unique solution to the equation

$$1 = \frac{A_{0+}(\alpha, C\sqrt{\alpha}\,)}{A_{\mathcal{G}}(\alpha, C\sqrt{\alpha}\,)} \tag{4.5}$$

for $C > 0$. As with (4.1), these equations are easily solved numerically, without having to evaluate $Z(\alpha, C\sqrt{\alpha}\,)$. Panel 2 in Figure 4 plots $\mathfrak{C}_{\mathcal{N},0-}(\alpha)$ and $\mathfrak{C}_{0+,\mathcal{G}}(\alpha)$ versus $1/\alpha$ for $\alpha \geq 1/2$; the plot shows that $\mathfrak{C}_{\mathcal{N},0-}(\alpha)$ is close to -0.7 and $\mathfrak{C}_{0+,\mathcal{G}}(\alpha)$ is close to 0.7, for all α. Denote the combined rejection sampler (4.3) by $\mathcal{S}_{\mathcal{N}\vee\mathcal{G}\vee 0}$. Its acceptance rate is

$$A_{\mathcal{N}\vee\mathcal{G}\vee 0}(\alpha, C\sqrt{\alpha}\,) = \begin{cases} A_{\mathcal{N}}(\alpha, C\sqrt{\alpha}\,) \vee A_{0-}(\alpha, C\sqrt{\alpha}\,), & \text{for } C \leq 0, \\ A_{\mathcal{G}}(\alpha, C\sqrt{\alpha}\,) \vee A_{0+}(\alpha, C\sqrt{\alpha}\,), & \text{for } C \geq 0. \end{cases} \tag{4.6}$$

Panel 3 in Figure 4 plots the worst case acceptance rates

$$\mathfrak{A}_{\mathcal{N},0-}(\alpha) := \min\nolimits_{C<0} A_{\mathcal{N}\vee\mathcal{G}\vee 0}(\alpha, C\sqrt{\alpha}\,) = A_{\mathcal{N}\vee\mathcal{G}\vee 0}\big(\alpha, \mathfrak{C}_{\mathcal{N},0-}(\alpha)\sqrt{\alpha}\,\big), \tag{4.7}$$

$$\mathfrak{A}_{0+,\mathcal{G}}(\alpha) := \min\nolimits_{C>0} A_{\mathcal{N}\vee\mathcal{G}\vee 0}(\alpha, C\sqrt{\alpha}\,) = A_{\mathcal{N}\vee\mathcal{G}\vee 0}\big(\alpha, \mathfrak{C}_{0+,\mathcal{G}}(\alpha)\sqrt{\alpha}\,\big) \tag{4.8}$$

versus $1/\alpha$ for $\alpha \geq 1/2$. We see that $\mathcal{S}_{\mathcal{N}\vee\mathcal{G}\vee 0}$ is always at least 80% efficient, and at least 81.7% efficient for large α. That $\mathfrak{C}_{\mathcal{N},0-}(\infty-) = -1/\sqrt{2} = -\mathfrak{C}_{0+,\mathcal{G}}(\infty-)$ and $\mathfrak{A}_{\mathcal{N},0-}(\infty-) = \sqrt{2/3} = \mathfrak{A}_{0+,\mathcal{G}}(\infty-)$ follow from (5.5), (5.12), and (5.36) below. As Panel 1 in Figure 4 shows, $\mathcal{S}_{\mathcal{N}\vee\mathcal{G}\vee 0}$ is very efficient for $|C|$ small or moderate; in particular $A_{\mathcal{N}\vee\mathcal{G}\vee 0}(\alpha, C\sqrt{\alpha}\,) \geq 0.95$ for $|C| \leq 0.2$ and for $|C| \geq 3$, for all $\alpha \geq 1/2$.

Remark 4.1. Suppose $\mathcal{S}_1$ and $\mathcal{S}_2$ are two rejection samplers for the same target density t. For $i = 1$ and 2, let r_i be the instrumental distribution for $\mathcal{S}_i$, let p_i be the probability with which $\mathcal{S}_i$ accepts a draw from r_i, and let ν_i be the average time necessary to make such a draw. Assuming that draws are independent and identically distributed, the average time it takes $\mathcal{S}_i$ to produce a draw from t is then $\mu_i := \nu_i/p_i$. The sampler $\mathcal{S}_1$ is more efficient than $\mathcal{S}_2$ when $\mu_1 < \mu_2$, that is, when the p-ratio p_1/p_2 exceeds the ν-ratio ν_1/ν_2. The preceding discussion tacitly assumed that for our various rejection samplers the ν-ratios were unity. If speed tests on a computer show that this assumption is seriously wrong, it could be worthwhile adjusting the cut-points $\mathfrak{C}_{\mathcal{N},\mathcal{G}}$, $\mathfrak{C}_{\mathcal{N},0-}$, and $\mathfrak{C}_{0+,\mathcal{G}}$ to make the p-ratios coincide with the corresponding ν-ratios, thereby minimizing the expected time for the combined rejection sampler to produce a draw.

5. Analysis of the acceptance rates

We now study the acceptance rates $A_0(\alpha, \gamma)$, $A_{\mathcal{N}}(\alpha, \gamma)$, and $A_{\mathcal{G}}(\alpha, \gamma)$, in reverse order. For real numbers C, put

$$b = b(C) := \frac{C + \sqrt{C^2+4}}{2} \quad \text{and} \quad B = B(C) := \frac{1}{b^2+1}. \tag{5.1}$$

Since $b'(C) = \big(1 + C/\sqrt{C^2+4}\,\big)/2 > 0$, as C runs from $-\infty$ to ∞, b increases from 0 to ∞, B decreases from 1 to 0, and the quantity

$$L(C) := b\sqrt{B} = b/\sqrt{b^2+1} = \sqrt{1-B}$$

in the limiting results (5.5), (5.12), and (5.22) below increases from 0 to 1. Since $b^2 - bC - 1 = 0$, we have $C = b - 1/b$. In addition,

$$b(-C) = \frac{1}{b(C)}, \;\; B(-C) = 1 - B(C), \;\; \text{and} \;\; L(-C) = \sqrt{B(C)} = \sqrt{1 - L^2(C)}. \tag{5.2}$$

5.1. Gamma rejection sampler on the square-root scale

Theorem 5.1. *For $\alpha > 0$ and $\gamma \in \mathbb{R}$, let $A_{\mathcal{G}}(\alpha, \gamma)$ be the acceptance rate defined by* (3.13) *with $\delta = \delta_1(\alpha, \gamma)$ given by* (3.10)*. Let $b = b(C)$ and $B = B(C)$ be as in* (5.1)*.*

(i) For $C \in \mathbb{R}$,

$$A_{\mathcal{G}}(\alpha, C\sqrt{\alpha}\,) = E\Big(\exp\Big(-\frac{W_\alpha^2}{2b^2}\Big)\Big) \tag{5.3}$$

for

$$W_\alpha := \frac{V_{2\alpha} - 2\alpha}{\sqrt{2\alpha}} \tag{5.4}$$

where $V_{2\alpha}$ is a random variable with a $\mathcal{G}(2\alpha, 1)$ distribution.

(ii) For each fixed α, $A_{\mathcal{G}}(\alpha, C\sqrt{\alpha}\,)$ is strictly increasing in $C \in \mathbb{R}$, tends to 0 as $C \downarrow -\infty$, and tends to 1 as $C \uparrow \infty$.
(iii) For each fixed C,

$$\lim_{\alpha\to\infty} A_{\mathcal{G}}(\alpha, C\sqrt{\alpha}\,) = b\sqrt{B}\,. \tag{5.5}$$

(iv) Suppose $\alpha = 1/2$. For $C \in \mathbb{R}$,

$$A_{\mathcal{G}}(\alpha, C\sqrt{\alpha}\,) = \sqrt{2\pi}\, b\, e^{b^2/2-1} P[\, Z \geq C\,]\,, \tag{5.6}$$

where Z is a standard normal random variable.

Proof. (i) Recall that $\delta = \delta_1(\alpha, \gamma) = \gamma + \sqrt{\gamma^2 + 4\alpha}$. By (3.13),

$$\begin{aligned}
A_{\mathcal{G}}(\alpha, \gamma) &= \frac{Z(\alpha,\gamma)}{\Gamma(2\alpha)} \delta^{2\alpha} \exp\bigl(-(2\alpha)^2/\delta^2\bigr) \\
&= \frac{1}{\Gamma(2\alpha)} \int_{x=0}^{\infty} x^{2\alpha-1} \exp(-x^2 - 2\gamma x)\; \delta^{2\alpha} \exp\bigl(-(2\alpha)^2/\delta^2\bigr)\, dx \\
&=_{(x=v/\delta)}\; \frac{1}{\Gamma(2\alpha)} \int_{v=0}^{\infty} v^{2\alpha-1} \exp\Bigl(-\frac{2v(2\alpha+\delta\gamma)}{\delta^2}\Bigr) \exp\Bigl(-\frac{(v-2\alpha)^2}{\delta^2}\Bigr)\, dv \\
&=_{(\text{by } \delta^2/2-\gamma\delta-2\alpha=0)}\; \frac{1}{\Gamma(2\alpha)} \int_{v=0}^{\infty} v^{2\alpha-1} \exp(-v)\; \exp\Bigl(-\frac{(v-2\alpha)^2}{\delta^2}\Bigr)\, dv \\
&= E\Bigl(\exp\Bigl(-\frac{(V_{2\alpha}-2\alpha)^2}{\delta^2}\Bigr)\Bigr).
\end{aligned}$$

Now use the fact that for $\gamma = C\sqrt{\alpha}$,

$$\delta = (C + \sqrt{C^2+4})\sqrt{\alpha} = 2b\sqrt{\alpha}. \tag{5.7}$$

(ii) This follows from (5.3), the way $b(C)$ varies with C, and the monotone convergence theorem.

(iii) Recall that here α tends to ∞ while C (and so also b and B) remains fixed. By the central limit theorem, W_α converges in distribution to $Z \sim \mathcal{N}(0,1)$ as $\alpha \to \infty$. Since $Z^2 \sim \mathcal{G}(1/2, 1/2)$ and since $G \sim \mathcal{G}(r, \lambda)$ has moment generating function $E(e^{tG}) = 1/(1-t/\lambda)^r$ for $t < \lambda$,

$$\lim_{\alpha\to\infty} A_{\mathcal{G}}(\alpha, C\sqrt{\alpha}\,) = E\bigl[\exp\bigl(-Z^2/(2b^2)\bigr)\bigr] = \frac{1}{(1+2/(2b^2))^{1/2}} = \frac{b}{\sqrt{1+b^2}}. \tag{5.8}$$

(iv) Suppose $\alpha = 1/2$. Then $V_{2\alpha}$ has a standard exponential distribution, so by (5.3),

$$\begin{aligned}
A_{\mathcal{G}}(\alpha, C\sqrt{\alpha}\,) &= \int_0^\infty \exp(-v)\; \exp\Bigl(-\frac{(v-1)^2}{2b^2}\Bigr)\, dv \\
&= e^{b^2/2-1} \int_0^\infty \exp\Bigl(-\frac{\bigl(v-(1-b^2)\bigr)^2}{2b^2}\Bigr)\, dy = \sqrt{2\pi}\, b\, e^{b^2/2-1} P[\, Z \geq C\,];
\end{aligned}$$

at the last step we used $(b^2-1)/b = b - 1/b = C$. □

5.2. Normal rejection sampler on the square-root scale

The next theorem gives properties of $A_{\mathcal{N}}(\alpha,\gamma)$ analogous to those for $A_{\mathcal{G}}(\alpha,\gamma)$ in the preceding one. We get a nice correspondence between properties by taking γ to be of the form $-C\sqrt{\alpha-1/2}$ below, in place of $C\sqrt{\alpha}$ in Theorem 5.1; this correspondence also links Theorems 6.1 and 6.2 below.

Theorem 5.2. *For $\alpha \ge 1/2$ and $\gamma \in \mathbb{R}$, let $A_{\mathcal{N}}(\alpha,\gamma)$ be the acceptance rate defined by* (3.12) *with $m = m(\alpha,\delta)$ given by* (3.3) and (3.5)*. Set $\alpha_\bullet = \alpha_\bullet(\alpha) := \alpha - 1/2$, and let Z be a standard normal random variable. Let $b = b(C)$ and $B = B(C)$ be as in* (5.1)*.*

(i) For $\alpha > 1/2$ and $C \in \mathbb{R}$,

$$A_{\mathcal{N}}(\alpha, -C\sqrt{\alpha_\bullet}\,) = E\big(v(\sigma Z)\big) \tag{5.9}$$

where

$$\sigma^2 := \frac{1}{2m^2(\alpha,-C\sqrt{\alpha_\bullet}\,)} = \frac{1}{2\alpha_\bullet b^2} \tag{5.10}$$

and

$$v(z) = \big(u(z)\big)^{2\alpha_\bullet} \quad \text{with } u(z) := \big((1+z)\exp(-z)\big)\mathbf{1}_{(-1,\infty)}(z). \tag{5.11}$$

(ii) For each fixed $\alpha > 1/2$, $A_{\mathcal{N}}(\alpha,-C\sqrt{\alpha_\bullet}\,)$ is strictly increasing in $C \in \mathbb{R}$, tends to 0 as $C \downarrow -\infty$, and tends to 1 as $C \uparrow \infty$.
(iii) For each fixed $C \in \mathbb{R}$,

$$\lim_{\alpha\to\infty} A_{\mathcal{N}}(\alpha,-C\sqrt{\alpha}\,) = \lim_{\alpha\to\infty} A_{\mathcal{N}}(\alpha,-C\sqrt{\alpha_\bullet}\,) = b\sqrt{B}\,. \tag{5.12}$$

(iv) Suppose $\alpha = 1/2$. For $C \in \mathbb{R}$,

$$A_{\mathcal{N}}(\alpha, C\sqrt{\alpha}\,) = \exp\big((C^+)^2/2\big)\,P[\,Z \ge C\,] \tag{5.13}$$

where $C^+ := \max(C,0)$. The rate $A_{\mathcal{N}}(\alpha,C\sqrt{\alpha}\,)$ is strictly decreasing in C, tends to 1 as $C \to -\infty$ and tends to 0 as $C \to \infty$.

Proof. (i) Writing $Z(\alpha,\gamma)$ as an integral, we have

$$A_{\mathcal{N}}(\alpha,\gamma) = \int_0^\infty \frac{1}{\sqrt{\pi}}\left(\frac{x}{m}\right)^{2\alpha-1}\exp(-x^2-2\gamma x+m^2+2\gamma m)\,dx. \tag{5.14}$$

Making the substitution $x = mz+m$ and then using that $2m^2+2\gamma m = 2\alpha-1$ by (3.4), we obtain that

$$\begin{aligned} A_{\mathcal{N}}(\alpha,\gamma) &= \int_{-1}^\infty \frac{m}{\sqrt{\pi}}\,(1+z)^{2\alpha-1}\exp(-m^2z^2-2m^2z-2\gamma mz)\,dz \\ &= \int_{-1}^\infty \big[(1+z)e^{-z}\big]^{2\alpha-1}\,\frac{m}{\sqrt{\pi}}\exp(-m^2z^2)\,dz = E\big(v(\sigma Z)\big) \end{aligned} \tag{5.15}$$

for $\sigma^2 = 1/(2m^2)$. Now use the fact that for $\gamma = -C\sqrt{\alpha_\bullet}\,$,

$$m = \frac{-\gamma+\sqrt{\gamma^2+4\alpha-2}}{2} = \sqrt{\alpha_\bullet}\,\frac{C+\sqrt{C^2+4}}{2} = \sqrt{\alpha_\bullet}\,b(C). \tag{5.16}$$

(ii) The function $v(z)$ is increasing in z for $z \in (-1, 0]$ and decreasing in z for $z \in [0, \infty)$; moreover $v(0) = 1$ and $\lim_{|z|\to\infty} v(z) = 0$. As C runs from $-\infty$ to ∞, $b = b(C)$ increases from 0 to ∞ so σ decreases from ∞ to 0; by the monotone convergence theorem, $E\big(v(\sigma Z)\big)$ increases from 0 to 1.

(iii) Recall that here $\alpha \to \infty$ while C (and so also b and B) remains fixed. Note that

$$\log\big(u(z)\big) = -z^2/2 + z^3/3 - z^4/4 + \cdots \tag{5.17}$$

for $|z| < 1$. Thus, as $\alpha \to \infty$,

$$\log\big(v(\sigma Z)\big) = 2\alpha_\bullet \log\big(u(\sigma Z)\big) = -\alpha_\bullet \sigma^2 Z^2\big(1 + o(1)\big) \to -Z^2/(2b^2);$$

by the dominated convergence theorem, $E\big(v(\sigma Z)\big) \to E\big(\exp(-Z^2/b^2)\big) = b\sqrt{B}$, as in (5.8).

(iv) Suppose $\alpha = 1/2$ and $\gamma \in \mathbb{R}$. The technique of completing the square gives

$$Z(\alpha, \gamma) = \int_0^\infty \exp(-x^2 - 2\gamma x)\, dx = \sqrt{\pi} e^{\gamma^2} P[\, Z \geq \sqrt{2}\, \gamma\,].$$

Since $m = m(\alpha, \gamma) = \gamma^-$,

$$A_{\mathcal{N}}(\alpha, \gamma) = \frac{Z(\alpha, \gamma)}{\sqrt{\pi}\exp(-m^2 - 2\gamma m)} = \begin{cases} P[\, Z \geq \sqrt{2}\, \gamma\,], & \text{if } \gamma < 0, \\ Z(\alpha, \gamma)/\sqrt{\pi}, & \text{if } \gamma \geq 0. \end{cases}$$

The right-hand side above is clearly strictly decreasing in γ with limits of 1 and 0 at $\mp\infty$. To get (iv) as stated, take $\gamma = C\sqrt{\alpha} = C/\sqrt{2}\,$. □

5.3. Gamma rejection sampler on the original scale: $\gamma < 0$

Now we turn to A_{0-}. The following theorem allows properties of A_{0-} to be deduced from those of $A_{\mathcal{G}}$. Recall that A_{0-} applies only to negative γ's.

Theorem 5.3. *For $\alpha > 0$ and $\gamma < 0$, let $A_{0-}(\alpha, \gamma)$ and $A_{\mathcal{G}}(\alpha, \gamma)$ be defined by* (2.9) *and* (3.13)*, respectively. For each fixed α, one has*

$$\frac{A_{0-}(\alpha, \gamma)}{A_{\mathcal{G}}(\alpha, \gamma)} = \rho(\alpha) := 2\frac{\Gamma(2\alpha)}{\Gamma(\alpha)}\Big(\frac{e}{4\alpha}\Big)^\alpha \tag{5.18}$$

for all $\gamma < 0$. Moreover, $\rho(\alpha)$ is a strictly increasing function of α such that

$$\lim\nolimits_{\alpha\downarrow 0} \rho(\alpha) = 1 \quad \textit{and} \quad \lim\nolimits_{\alpha\uparrow\infty} \rho(\alpha) = \sqrt{2}. \tag{5.19}$$

Proof. In the following argument, suppose $\gamma = C\sqrt{\alpha}$ for $C < 0$. By (2.9) and (2.4),

$$A_{0-}(\alpha, \gamma\,) = \frac{2Z(\alpha, \gamma)e^\alpha}{\Gamma(\alpha)}\Big(\frac{\delta_0}{\exp(1/\delta_0)}\Big)^\alpha$$

where

$$\delta_0 = \frac{4\alpha}{\left(\sqrt{\gamma^2+4\alpha}+|\gamma|\right)^2} = \frac{4}{\left(\sqrt{C^2+4}+|C|\right)^2} = b^2$$

for $b = b(C)$ as in (5.1); at the last step, we used the fact that $b(C) = 1/b(|C|)$. By (3.13) with $\gamma = C\sqrt{\alpha}$ and formula (5.7) for $\delta_1(\alpha, C\sqrt{\alpha}\,)$,

$$A_{\mathcal{G}}(\alpha,\gamma) = \frac{Z(\alpha,\gamma)}{\Gamma(2\alpha)} 4^\alpha \alpha^\alpha \Big(\frac{b^2}{\exp(1/b^2)}\Big)^\alpha. \tag{5.20}$$

If follows that (5.18) holds. To show that $\rho(\alpha)$ is strictly increasing in α, we use the fact that

$$\Psi(r) := \frac{d}{dr}\log\big(\Gamma(r)\big) = \log(r) - \frac{1}{r}E\Big(p\Big(\frac{V}{r}\Big)\Big) \tag{5.21a}$$

where

$$V \sim \mathcal{G}(1,1) \quad \text{and} \quad p(t) := 1 - \Big(\frac{1}{t} - \frac{1}{e^t-1}\Big) \text{ is strictly increasing in } t; \tag{5.21b}$$

confer formulas (2.67) and (2.68) in Andrews (1985). This implies that

$$\frac{d}{d\alpha}\log\big(\rho(\alpha)\big) = 2\big[\Psi(2\alpha)-\log(2\alpha)\big]-\big[\Psi(\alpha)-\log(\alpha)\big] = \frac{1}{\alpha}E\Big[p\Big(\frac{V}{\alpha}\Big)-p\Big(\frac{V}{2\alpha}\Big)\Big] > 0.$$

That $\rho(\alpha) \to 1$ as $\alpha \downarrow 0$ follows from the identity $r\Gamma(r) = \Gamma(r+1)$, and that $\rho(\alpha) \to \sqrt{2}$ as $\alpha \uparrow \infty$ follows from Stirling's formula. □

Here are some consequences of the theorem. First, since $\rho(\alpha) > 1$ for all $\alpha > 0$, for $\gamma < 0$ the Gamma rejection sampler on the original scale always outperforms the one on the square-root scale. Second, as α increases to ∞, $\rho(\alpha)$ increases to $\sqrt{2}$, so $A_{\mathcal{G}}(\alpha, 0) = 1/\rho(\alpha)$ decreases to $1/\sqrt{2}$, as one sees in Figure 3; moreover by (5.5),

$$\lim_{\alpha\to\infty} A_{0-}(\alpha, C\sqrt{\alpha}\,) = \sqrt{2}\, b(C)\sqrt{B(C)} \tag{5.22}$$

for all $C \le 0$. Third, for each fixed α, $A_{0-}(\alpha,\gamma)$ is strictly increasing in $\gamma \in (-\infty, 0)$, with limits 0 and 1 at $-\infty$ and 0, respectively.

5.4. Gamma rejection sampler on the original scale: $\gamma > 0$

Finally we take up A_{0+}. Recall that A_{0+} only applies to positive γ's. By (2.9),

$$A_{0+}(\alpha, C\sqrt{\alpha}\,) = \frac{2Z(\alpha, C\sqrt{\alpha}\,)}{\Gamma(\theta_\alpha(C)\alpha)}\Big(\frac{Ce}{\sqrt{\alpha}\,\big(1-\theta_\alpha(C)\big)}\Big)^{2\alpha(1-\theta_\alpha(C))} \tag{5.23}$$

where $\theta_\alpha(C) := \theta(\alpha,\gamma)$ for $\gamma = C\sqrt{\alpha}$ is the unique solution to the equation

$$s_\alpha(\theta) = -2\log(C) \tag{5.24}$$

with

$$s_\alpha(\theta) := (\log\Gamma)'(\theta\alpha) - \log(\alpha) - 2\log(1-\theta) \tag{5.25}$$

for $\theta \in (0,1)$. Also for $\theta \in (0,1)$, set

$$s_\infty(\theta) := \log(\theta) - 2\log(1-\theta). \tag{5.26}$$

Theorem 5.4. *For $\alpha > 0$ and $C > 0$, let $A_{0+}(\alpha, C\sqrt{\alpha}\,)$ be the acceptance rate defined in* (5.23) *and let $\theta_\alpha(C)$ be the root of the equation* (5.24) *with $s_\alpha(\theta)$ given by* (5.25)*. Let $s_\infty(\theta)$ be defined by* (5.26)*. Let $b = b(C)$ and $B = B(C)$ be as in* (5.1)*.*

(i) For each fixed α, $\theta_\alpha(C)$ is strictly decreasing in $C \in (0,\infty)$ with $\lim_{C\downarrow 0}\theta_\alpha(C) = 1$ and $\lim_{C\uparrow\infty}\theta_\alpha(C) = 0$.
(ii) For each fixed C, $\theta_\alpha(C)$ is strictly decreasing in α.
(iii) The solution $\theta_\infty(C)$ to the equation $s_\infty(\theta) = -2\log(C)$ is

$$\theta_\infty(C) = 1/b^2 = B/(1-B) = 1 - C/b. \tag{5.27}$$

(iv) If α tends to ∞, then $\theta_\alpha(C)$ converges to $\theta_\infty(C)$ at rate $\big(1-\theta_\infty(C)\big)\times O(1/\alpha)$ uniformly for C bounded away from ∞, in the sense that for each fixed $C_0 \in (0,\infty)$,

$$\sup\nolimits_{0<C\le C_0}\left|\frac{\theta_\alpha(C)-\theta_\infty(C)}{1-\theta_\infty(C)}\right| = O(1/\alpha). \tag{5.28}$$

(v) For each fixed C,

$$\lim\nolimits_{\alpha\to\infty} A_{0+}(\alpha, C\sqrt{\alpha}\,) = \sqrt{2B}\,. \tag{5.29}$$

Proof. (i) In terms of the function $s(\theta)$ in (2.7), $s_\alpha(\theta) = s(\theta) - 2\log(C)$. By Theorem 2.1, $s_\alpha(\theta)$ is strictly increasing in $\theta \in (0,1)$, with $\lim_{\theta\downarrow 0} s_\alpha(\theta) = -\infty$ and $\lim_{\theta\uparrow\infty} s_\alpha(\theta) = \infty$. The claims in (i) follow from this.

(ii) Write $s_\alpha(\theta)$ as $\Psi(r) - \log(r) + s_\infty(\theta)$ where $\Psi(r)$ is the digamma function $(\log\Gamma)'(r)$ and $r = \theta\alpha$. Since $\Psi(r) - \log(r)$ is strictly increasing in r by (5.21), $s_\alpha(\theta)$ is strictly increasing in α. The claim in (ii) follows from this.

(iii) $\theta_\infty(C) \in (0,1)$ is the unique root $\theta \in (0,1)$ of the equation $\theta/(1-\theta)^2 = 1/C^2$. That root is $1/b^2$ because $1 - 1/b^2 = C/b$.

(iv) From analysis, $\Psi(r)$ has the asymptotic expansion

$$\Psi(r) = \log(r) - 1/(2r) - 1/(12r^2) + O(1/r^4) \tag{5.30}$$

as $r \to \infty$; confer (2.74) in Andrews (1985). Consequently as $\alpha \to \infty$,

$$s_\alpha(\theta) = s_\infty(\theta) + O(1/\alpha)$$

uniformly for $\theta \in (0,1)$ bounded away from 0. Equation (5.28) follows from this, the identity

$$\begin{aligned}
0 &= -2\log(C) - \big(-2\log(C)\big) = s_\alpha\big(\theta_\alpha(C)\big) - s_\infty\big(\theta_\infty(C)\big)\\
&= \big[s_\alpha\big(\theta_\alpha(C)\big) - s_\infty\big(\theta_\alpha(C)\big)\big] + \big[s_\infty\big(\theta_\alpha(C)\big) - s_\infty\big(\theta_\infty(C)\big)\big]\\
&= \big[s_\alpha\big(\theta_\alpha(C)\big) - s_\infty\big(\theta_\alpha(C)\big)\big] + \big(\theta_\alpha(C) - \theta_\infty(C)\big)s_\infty'(\theta^*)
\end{aligned}$$

where θ^* is some point between $\theta_\alpha(C)$ and $\theta_\infty(C)$, and the inequality

$$s'_\infty(\theta^*) = \frac{1}{\theta^*} + \frac{2}{1-\theta^*} \geq \frac{2}{1-\theta^*} \geq \frac{2}{1-\theta_\infty(C)}.$$

(v) We will just sketch the argument; the algebraic manipulations were done by computer. By (5.23),

$$\log\bigl(A_{0+}(\alpha, C\sqrt{\alpha}\,)\bigr) = K - J\bigl(\theta_\alpha(C)\bigr) \tag{5.31}$$

where

$$K := \log\bigl(2Z(\alpha, C\sqrt{\alpha}\,)\bigr),$$
$$J(\theta) := \log\Bigl[\Gamma(\alpha\theta)\Bigl(\frac{\sqrt{\alpha}\,(1-\theta)}{Ce}\Bigr)^{2\alpha(1-\theta)}\Bigr],$$

and $\theta = \theta_\alpha(C)$ is the root of the equation $J'(\theta) = 0$. We do not have a closed form expression for $\theta_\alpha(C)$, but from (5.28) we know that $\theta_\alpha(C) = \theta_\infty(C) + O(1/\alpha)$ as $\alpha \to \infty$. By Taylor's theorem,

$$J\bigl(\theta_\infty(C)\bigr) = J\bigl(\theta_\alpha(C)\bigr) + \tfrac{1}{2}J''(\theta^*)\bigl(\theta_\infty(C) - \theta_\alpha(C)\bigr)^2$$

for some point θ^* between $\theta_\alpha(C)$ and $\theta_\infty(C)$. It turns out that $J''(\theta^*)$ is of order α, so

$$J\bigl(\theta_\alpha(C)\bigr) = J\bigl(\theta_\infty(C)\bigr) + O(1/\alpha) \tag{5.32}$$

as $\alpha \to \infty$. Now write $J\bigl(\theta_\infty(C)\bigr)$ as a linear combination of logarithms, approximate $\log\bigl(\Gamma(\alpha\theta_\infty(C))\bigr)$ using Stirling's formula

$$\log\bigl(\Gamma(r)\bigr) = \tfrac{1}{2}\log(2\pi) + r\log(r) - r - \tfrac{1}{2}\log(r) + O(1/r) \text{ as } r \to \infty, \tag{5.33}$$

use the fact that $\theta = \theta_\infty(C)$ is the root of the equation $\log(\theta) - 2\log(\theta) = -2\log(C)$, and replace $\log\bigl(1 - \theta_\infty(C)\bigr)$ by $\log(C) - \log(b)$. The upshot is that as $\alpha \to \infty$,

$$J\bigl(\theta_\infty(C)\bigr) = \log(D) + O(1/\alpha) \text{ for } D := \sqrt{2\pi}\,\alpha^{\alpha-1/2}\,e^{-2\alpha}\,(e^{1/b^2}/b^2)^a\,b\,. \tag{5.34}$$

By (5.20) and by (5.33) with $r = 2\alpha$

$$K = \log\bigl(\sqrt{2B}\,\bigr) + \log(D) + o(1) \tag{5.35}$$

as $\alpha \to \infty$. (The $o(1)$ in (5.35) is actually an $O(1/\alpha)$; see (6.14) below.) Together with (5.31), (5.32), and (5.34), this gives (5.29). □

Panel 1 in Figure 5 illustrates parts (i)-(iv) of the theorem. Together with (5.22) and (5.2), the identity (5.29) in part (v) implies that

$$\lim_{\alpha\to\infty} A_{0-}(\alpha, -C\sqrt{\alpha}\,) = \sqrt{2}b(-C)\sqrt{B(-C)} = \sqrt{2B(C)} = \lim_{\alpha\to\infty} A_{0+}(\alpha, C\sqrt{\alpha}\,) \tag{5.36}$$

for $C > 0$; Figure 2 illustrates this limiting symmetry about $C = 0$.

6. Asymptotic expansions of the acceptance rates

This section gives asymptotic expansions for the acceptance rates $A_{\mathcal{G}}(\alpha, C\sqrt{\alpha}\,)$, $A_{\mathcal{N}}(\alpha, C\sqrt{\alpha}\,)$, and $A_{0+}(\alpha, C\sqrt{\alpha}\,)$ as $\alpha \to \infty$ with C held fixed, and for the rates $A_{\mathcal{G}}(\alpha, C\sqrt{\alpha}\,)$ and $A_{\mathcal{N}}(\alpha, -C\sqrt{\alpha}\,)$ as $C \to \infty$ with α held fixed. The expansions sharpen the results from Section 5, and as we explain in more detail below, they can yield trivial to evaluate and only slightly suboptimal bounds on the ratio between target and instrumental distribution; recall the opening paragraph of the introduction for the role of such bounds in rejection sampling.

6.1. Gamma rejection sampler on the square-root scale

Theorem 6.1. *Let $A_{\mathcal{G}}$ be as in Theorem* 5.1*, and let $b = b(C)$ and $B = B(C)$ be as in* (5.1)*.*

(i) The last assertion in part (ii) of Theorem 5.1 *admits the following sharpening. Fix $\alpha > 0$, put $r = 2\alpha$, and for $j = 0, 1, 2, 3$ and $C \geq 0$ set*

$$A^{\circ}_{\mathcal{G},j}(\alpha, C) := \sum\nolimits_{i=0}^{j} a^{\circ}_{\mathcal{G},i}(\alpha)\frac{1}{b^{2i}} \tag{6.1}$$

where

$$\begin{aligned} a^{\circ}_{\mathcal{G},0}(\alpha) &:= 1, & a^{\circ}_{\mathcal{G},2}(\alpha) &:= \frac{1}{8}\Big(3+\frac{6}{r}\Big), \\ a^{\circ}_{\mathcal{G},1}(\alpha) &:= -\frac{1}{2}, & a^{\circ}_{\mathcal{G},3}(\alpha) &:= -\frac{1}{48}\Big(15+\frac{130}{r}+\frac{120}{r^2}\Big). \end{aligned} \tag{6.2}$$

Then for each j

$$R^{\circ}_{\mathcal{G},j}(\alpha, C) := A^{\circ}_{\mathcal{G},j}(\alpha, C) - A_{\mathcal{G}}(\alpha, C\sqrt{\alpha}\,) = O\Big(\frac{1}{b^{2j+2}}\Big) \text{ as } C \to \infty, \tag{6.3}$$

$$R^{\circ}_{\mathcal{G},j}(\alpha, C) \text{ is positive or negative according as } j \text{ is even or odd.} \tag{6.4}$$

Moreover, for all $\alpha \geq 10$,

$$\sup\nolimits_{C\geq 0}\, b^{2j+2}\big|R^{\circ}_{\mathcal{G},j}(\alpha, C)\big| \leq r^{\circ}_{\mathcal{G},j} \text{ with } \begin{array}{ll} r^{\circ}_{\mathcal{G},0} = .50, & r^{\circ}_{\mathcal{G},2} = 0.45, \\ r^{\circ}_{\mathcal{G},1} = .41, & r^{\circ}_{\mathcal{G},3} = 0.63. \end{array} \tag{6.5}$$

(ii) Part (iii) of Theorem 5.1 *admits the following sharpening. Fix $C \in \mathbb{R}$. For $j = 0, 1, 2$, set*

$$A^{\bullet}_{\mathcal{G},j}(\alpha, C) := b\sqrt{B}\,\sum\nolimits_{i=0}^{j} a^{\bullet}_{\mathcal{G},i}(C)/(2\alpha)^i \tag{6.6}$$

with

$$a^{\bullet}_{\mathcal{G},0}(C) := 1, \tag{6.7a}$$

$$a^{\bullet}_{\mathcal{G},1}(C) := \frac{B^2(9-10B)}{12}, \tag{6.7b}$$

$$a^{\bullet}_{\mathcal{G},2}(C) := \frac{B^3(1540B^3-3780B^2+2961B-720)}{288}. \tag{6.7c}$$

For all $\alpha \geq 1/2$, one has

$$\sup_{C\in\mathbb{R}}\big|A^{\bullet}_{\mathcal{G},j}(\alpha,C) - A_{\mathcal{G}}(\alpha, C\sqrt{\alpha}\,)\big| \leq \begin{cases} 0.06/(2\alpha), & \textit{if } j=0,\\ 0.019/(2\alpha)^2, & \textit{if } j=1,\\ 0.014/(2\alpha)^3, & \textit{if } j=2.\end{cases} \tag{6.8}$$

Proof. (i) Recall that here $C \to \infty$ while α remains fixed. (6.3) and (6.4) follow from (5.3) and Lemma 6.1 below with $Y = W_\alpha^2$, $\mathcal{Y} = [0,\infty)$, $F(y) = e^{-y}$, $q = n = 4$, and $\delta = 1/(2b^2)$; since $W_\alpha = (V_r - r)/\sqrt{r}$ with $r = 2\alpha$ and $V_r \sim \mathcal{G}(r,1)$, one has $E(Y) = 1$, $E(Y^2) = (3r+6)/r$, $E(Y^3) = (15r^2+130r+120)/r^2$, and $E(Y^4) < \infty$. A computer study gave the specific bounds in (6.5).

(ii) Recall that here α tends to ∞ while C remains fixed. The density f_{W_α} of W_α has an Edgeworth expansion Kolassa (2006)

$$f_{W_\alpha}(x) \doteq E_r(x) := \varphi(x)\Big[1 + \frac{1}{\sqrt{r}}\frac{H_3(x)}{3} + \frac{1}{r}\Big(\frac{H_4(x)}{4} + \frac{H_6(x)}{18}\Big) + \tag{6.9}$$

$$\frac{1}{r^{3/2}}\begin{pmatrix}\text{linear combination of}\\ H_j(x) \text{ for } j=5,7,9\end{pmatrix} + \frac{1}{r^2}\Big(\frac{H_6(x)}{6} + \frac{47H_8(x)}{480} + \frac{H_{10}(x)}{72} + \frac{H_{12}(x)}{1944}\Big) + \cdots\Big],$$

where $r = 2\alpha$, φ is the density of $Z \sim \mathcal{N}(0,1)$, and H_k is the kth Hermite polynomial, defined by $\varphi^{(k)} = (-1)^k\varphi H_k$. Since

$$E\big[H_\ell(Z)\exp\big(-Z^2/(2b^2)\big)\big] = \begin{cases}(-1)^k(2k-1)!!\, b\sqrt{B}\,B^k, & \text{if } \ell = 2k \text{ is even},\\ 0, & \text{otherwise},\end{cases}$$

where $(2k-1)!! = (2k-1)\cdot(2k-3)\cdots 3\cdot 1$, we get the asymptotic expansion

$$A_{\mathcal{G}}(\alpha, C\sqrt{\alpha}\,) = E\big(-W_\alpha^2/(2b^2)\big) = \int_{-\infty}^{\infty} f_{W_\alpha}(x)\exp\big(-x^2/(2b^2)\big)\,dx$$

$$\doteq \int_{-\infty}^{\infty} E_r(x)\exp\big(-x^2/(2b^2)\big)\,dx = b\sqrt{B}\Big[a^{\bullet}_{\mathcal{G},0}(C) + a^{\bullet}_{\mathcal{G},1}(C)/r + a^{\bullet}_{\mathcal{G},2}(C)/r^2 + \cdots\Big]. \tag{6.10}$$

One can use the error bounds for the Edgeworth expansion to show that the difference between $A_{\mathcal{G}}(\alpha, C\sqrt{\alpha}\,)$ and $A^{\bullet}_{\mathcal{G},j}(\alpha, C)$ is order $O(1/r^{j+1})$ as $r \to \infty$, uniformly in C. A computer study gave the specific bounds in (6.8). □

We used the following lemma in the proof of Theorem 6.1 and will call on it again.

Lemma 6.1. *Suppose that n is a positive integer, q is a finite number at least n, Y is a random variable taking values in $\mathcal{Y} = [0,\infty)$ or $\mathbb{R}$ such that $E(|Y|^q) < \infty$, and $F\colon \mathcal{Y} \to \mathbb{R}$ is a function which has a finite nth derivative at 0, is bounded on bounded intervals, and satisfies $F(y) = O(|y|^q)$ as $|y| \to \infty$. Then*

$$E\big(F(\delta Y)\big) = \sum_{i=0}^{n-1}\frac{F^{(i)}(0)E(Y^i)}{i!}\delta^i + O(\delta^n) \tag{6.11}$$

as $\delta \downarrow 0$. *Moreover, if* $F^{(n)}(y)$ *exists for all* $y \in \mathcal{Y}$ *and has a constant sign* s *throughout* $\mathcal{Y}$, *and if either* $\mathcal{Y} = [0, \infty)$ *or* $\mathcal{Y} = \mathbb{R}$ *and* n *is even, then the* O *term in* (6.11) *has sign* s.

Proof. Define $R\colon \mathcal{Y} \to \mathbb{R}$ by the equation

$$F(y) = \sum\nolimits_{i=0}^{n-1} \frac{F^{(i)}(0)}{i!} y^i + R(y).$$

The assumptions on F imply that there exists a finite constant c such that $|R(y)| \le c(|y|^n + |y|^q)$ for all $y \in \mathcal{Y}$, whence

$$\begin{aligned}\Big|E\big(F(\delta Y)\big) - \sum\nolimits_{i=0}^{n-1} \frac{F^{(i)}(0)E(Y^i)}{i!}\delta^i\Big| &= \big|E\big(R(\delta Y)\big)\big| \le E\big(\big|R(\delta Y)\big|\big) \\ &\le c\big(\delta^n E(|Y|^n) + \delta^q E(|Y|^q)\big) = O(\delta^n)\end{aligned}$$

as $\delta \downarrow 0$. That establishes (6.11). For the final assertion of the lemma, use Taylor's theorem to represent $R(y)$ as $F^{(n)}(\eta) y^n/n!$ for some number η between 0 and y. □

Here are some remarks on part (i) of Theorem 6.1. First, for C large and α bounded away from 0, part (i) implies that $A_{\mathcal{G}}(\alpha, C\sqrt{\alpha}\,)$ depends only slightly on α; this is what one sees in Figure 3 even for C as small as 2. Second, the asymptotic expansion for $A_{\mathcal{G}}(\alpha, C\sqrt{\alpha}\,)$ as $C \to \infty$ in (6.3) gives rise to an asymptotic expansion for $Z(\alpha, \gamma)$ as $\gamma \to \infty$, as follows. Fix α. Given γ, set $C = \gamma/\sqrt{\alpha}$, write $b(C)$ as $\gamma^\star/\sqrt{\alpha}$ with $\gamma^\star := (\gamma + \sqrt{\gamma^2 + 4\alpha}\,)/2$, write the δ in (3.10) as $2\gamma^\star$, and then use (3.13) and (6.3) to get

$$\frac{Z(\alpha,\gamma)}{\Gamma(2\alpha)/(2\gamma^\star)^{2\alpha}}\, e^{-(\alpha/\gamma^\star)^2} = A_{\mathcal{G}}(\alpha,\gamma) = \sum\nolimits_{i=0}^{j-1} \frac{a^\circ_{\mathcal{G},i}(\alpha)\alpha^i}{\gamma^{\star 2i}} + O\Big(\frac{1}{\gamma^{2j}}\Big) \qquad (6.12)$$

as $\gamma \to \infty$, for $j = 1, 2, 3, 4$. The classical asymptotic development of $Z(\alpha, \gamma)$ as $\gamma \to \infty$ uses

$$\begin{aligned}\frac{Z(\alpha,\gamma)}{\Gamma(2\alpha)/(2\gamma)^{2\alpha}} &= E\Big[\exp\Big(-\frac{V_{2\alpha}^2}{(2\gamma)^2}\Big)\Big] \\ &= \sum\nolimits_{i=0}^{j-1} (-1)^i \frac{(2\alpha)^{[\text{rise } 2i]}}{2^{2i} i!} \frac{1}{\gamma^{2i}} + O\Big(\frac{1}{\gamma^{2j}}\Big),\end{aligned} \qquad (6.13)$$

where $x^{[\text{rise } k]} := x(x+1)\cdots(x+k-1)$. It turns out that the expansion in (6.12) is more accurate — typically considerably so — than the expansion in (6.13). Indeed, numerical computations show that for given α and j, the ratio of the O term in (6.12) to the one in (6.13) increases with γ. Thus the supremum of those ratios equals their limit as $\gamma \to \infty$, which is

$$a^\circ_{\mathcal{G},j}(\alpha)\alpha^j \Big/ \Big(\frac{(2\alpha)^{[\text{rise } 2j]}}{2^{2j} j!}\Big).$$

This quantity is less than 1 for $j = 1, 2, 3, 4$ and $\alpha > 0$; as $\alpha \to \infty$ it asymptotes to $\big(\binom{2j}{j}/2^{2j}\big)/(\alpha^j/j!)$, which of course is small when α is large.

Here are some remarks bearing on part (ii) of Theorem 6.1. First, the linear factor $9 - 10B$ in (6.7b) has a root at $B = 9/10$, corresponding to $b = 1/3$ and $C = -2\,2/3$. So to first order in $1/\alpha$, as α increases to ∞, $A_{\mathcal{G}}(\alpha, C\sqrt{\alpha}\,)$ decreases down to $b\sqrt{B}$ for $C > -2\,2/3$, and otherwise increases up to $b\sqrt{B}$; this is what one sees in Figure 3. Second, in order to implement the rejection sampler $\mathcal{S}_{\mathcal{G}}$, one needs to compute the quantity $M = 1/A := 1/A_{\mathcal{G}}(\alpha, C\sqrt{\alpha}\,)$. Since, by (6.8), $A > A^\bullet := A^\bullet_{\mathcal{G},2}(\alpha, C) - 0.014/(2\alpha)^3$, one has $M^\bullet := 1/A^\bullet > M$. By the opening paragraph of the introduction, one can use $M^\bullet$ in place of M; doing so reduces the acceptance rate from A to $A^\bullet$, but that's a negligible decrease. The advantage is that it is trivial to compute $M^\bullet$. A similar remark applies to the asymptotic expansions we give for the other acceptance rates. Third, by (5.20), for each fixed C one has

$$\frac{Z(\alpha, C\sqrt{\alpha}\,)}{\dfrac{\Gamma(2\alpha)}{4^\alpha \alpha^\alpha}\Big(\dfrac{\exp(1/b^2)}{b^2}\Big)^\alpha} = A_{\mathcal{G}}(\alpha, C\sqrt{\alpha}\,) = A^\bullet_{\mathcal{G},2}(\alpha, C) + O\big(1/\alpha^3\big) \tag{6.14}$$

as $\alpha \to \infty$. We will use this in the proof of Theorem 6.3.

6.2. Normal rejection sampler on the square-root scale

Theorem 6.2. *Let α, $\alpha_\bullet$, and $A_{\mathcal{N}}$ be as in Theorem 5.2. For $C \in \mathbb{R}$, let $b = b(C)$ and $B = B(C)$ be as in (5.1).*

(i) The last assertion in part (ii) of Theorem 5.2 admits the following sharpening. Fix $\alpha > 1/2$, put $r = 2\alpha_\bullet$, and for $j = 0, 1, 2, 3$ and $C \geq 0$ set

$$A^\circ_{\mathcal{N},j}(\alpha, C) = \sum\nolimits_{i=0}^{j} a^\circ_{\mathcal{N},i}(\alpha)\frac{1}{b^{2i}}, \tag{6.15}$$

where, for $i = 0, 1, 2, 3$,

$$a^\circ_{\mathcal{N},i}(\alpha) := a^\circ_{\mathcal{G},i}(-\alpha_\bullet) \tag{6.16}$$

with the $a^\circ_{\mathcal{G},i}$'s as in (6.2). Then for each j

$$R^\circ_{\mathcal{N},j}(\alpha, C) := A^\circ_{\mathcal{N},j}(\alpha, C) - A_{\mathcal{N}}(\alpha, -C\sqrt{\alpha_\bullet}\,) = O\Big(\frac{1}{b^{2j+2}}\Big) \tag{6.17}$$

as $C \to \infty$. Moreover,

$$\sup\nolimits_{C\geq 0} b^{2j+2}\big|R^\circ_{\mathcal{N},j}(\alpha, C)\big| \leq r^\circ_{\mathcal{N},j} \ \textit{ with } \begin{array}{ll} r^\circ_{\mathcal{N},0} = .50, & r^\circ_{\mathcal{N},2} = 0.33, \\ r^\circ_{\mathcal{N},1} = .38, & r^\circ_{\mathcal{N},3} = 0.28. \end{array} \tag{6.18}$$

for all $\alpha \geq 5$.

(ii) Part (iii) of Theorem 5.2 admits the following sharpening. Fix $C \in \mathbb{R}$. For $j = 0, 1, 2$, set

$$A^\bullet_{\mathcal{N},j}(\alpha, C) := b\sqrt{B}\,\sum\nolimits_{i=0}^{j} a^\bullet_{\mathcal{N},i}(C)/(2\alpha_\bullet)^i, \tag{6.19}$$

where

$$a^{\bullet}_{\mathcal{N},0}(C) := a^{\bullet}_{\mathcal{G},0}(C), \quad a^{\bullet}_{\mathcal{N},1}(C) := -a^{\bullet}_{\mathcal{G},1}(C), \quad \textit{and} \quad a^{\bullet}_{\mathcal{N},2}(C) := a^{\bullet}_{\mathcal{G},2}(C), \tag{6.20}$$

with the $a^{\bullet}_{\mathcal{G},i}$*'s as in* (6.7)*. For all* $\alpha \geq 2$*, one has*

$$\sup_{C\in\mathbb{R}}\left|A^{\bullet}_{\mathcal{N},j}(\alpha, C) - A_{\mathcal{N}}(\alpha, -C\sqrt{\alpha_\bullet}\,)\right| \leq \begin{cases} 0.06/(2\alpha_\bullet), & \textit{if } j = 0, \\ 0.019/(2\alpha_\bullet)^2, & \textit{if } j = 1, \\ 0.014/(2\alpha_\bullet)^3, & \textit{if } j = 2. \end{cases} \tag{6.21}$$

Proof. (i) Recall that here $C \to \infty$ while α remains fixed. The claim in (6.17) follows from (5.9) and (6.11) with $Y = Z$, $\mathcal{Y} = \mathbb{R}$, $F(y) = v(y)$, $q = n = 8$, and $\delta = \sigma = 1/(\sqrt{2\alpha_\bullet}\, b)$. We got the derivatives of F at 0 via computer algebra and then used $E(Z^{2j-1}) = 0$ and $E(Z^{2j}) = (2j-1)!!$. A computer study gave the specific bounds in (6.21).

(ii) Recall that here $\alpha \to \infty$ while C remains fixed. By (5.15) and (5.16),

$$A_{\mathcal{N}}(\alpha, -C\sqrt{\alpha_\bullet}\,) = \frac{\sqrt{\alpha_\bullet}\, b}{\sqrt{\pi}} \int_{-1}^{\infty} \exp\bigl(2\alpha_\bullet w(z)\bigr)\, dz \tag{6.22}$$

where

$$w(z) = \log\bigl(u(z)\bigr) - b^2 z^2/2$$

for $u(z)$ as in (5.11). $w(z)$ has the Taylor expansion

$$w(z) = -z^2/(2B) + z^3/3 - z^4/4 + \cdots$$

about $z = 0$, and $w(z)$ is strictly increasing in z for $z \in [-1, 0]$ and strictly decreasing in z for $z \in [0, \infty)$. Laplace's method (see, e.g., Section II.1 in Wong (2001)) gives $A_{\mathcal{N}}(\alpha, -C\sqrt{\alpha_\bullet}\,) = A^{\bullet}_{\mathcal{N},2}(\alpha, C) + O(1/(2\alpha_\bullet)^3)$ as $\alpha \to \infty$. A computer study gave the explicit error bounds in (6.21). □

Part (i) of Theorem 6.2 has implications similar to those of part (i) of Theorem 6.1. First, for large C and α bounded away from 1/2, the acceptance rate $A_{\mathcal{N}}(\alpha, -C\sqrt{\alpha_\bullet}\,)$ depends only slightly on α; the same is true, but somewhat less so, for $A_{\mathcal{N}}(\alpha, -C\sqrt{\alpha}\,)$. Second, the asymptotic expansion for $A_{\mathcal{N}}(\alpha, -C\sqrt{\alpha_\bullet}\,)$ as $C \to \infty$ in (6.17) gives rise to an asymptotic expansion for $Z(\alpha, -\gamma)$ as $\gamma \to \infty$, as follows. Fix $\alpha > 1/2$. Given γ, set $C = \gamma/\sqrt{\alpha_\bullet}$, write $b(C)$ as $\gamma^\star/\sqrt{\alpha_\bullet}$ here with $\gamma^\star := (\gamma + \sqrt{\gamma^2 + 4\alpha_\bullet}\,)/2$, write $m(\alpha, -\gamma)$ as $\gamma^\star$ (see (3.3)), note that $\gamma^\star - \gamma = \alpha_\bullet/\gamma^\star$, and then use (3.12) and (6.17) to get

$$\frac{Z(\alpha, -\gamma)}{\sqrt{\pi}\,(\gamma^\star)^{2\alpha_\bullet} e^{\gamma^2} e^{-(\alpha_\bullet/\gamma^\star)^2}} = A_{\mathcal{N}}(\alpha, -\gamma) = \sum\nolimits_{i=0}^{j-1} \frac{a^{\bullet}_{\mathcal{G},i}(-\alpha_\bullet)\alpha_\bullet^i}{\gamma^{\star 2i}} + O\Bigl(\frac{1}{\gamma^{2j}}\Bigr) \tag{6.23}$$

as $\gamma \to \infty$, for $j = 1, 2, 3, 4$. The classical asymptotic development of $Z(\alpha, -\gamma)$ as $\gamma \to \infty$ uses

$$\frac{Z(\alpha,-\gamma)}{\sqrt{\pi}\,\gamma^{2\alpha_\bullet}e^{\gamma^2}} = E\Big[\Big(1+\frac{Z}{2\sqrt{\gamma}}\Big)^{2\alpha_\bullet}\mathbf{1}_{(-2\sqrt{\gamma},\infty)}(Z)\Big]$$
$$= \sum\nolimits_{i=0}^{j-1}\frac{(2\alpha_\bullet)^{[\text{fall } 2i]}}{2^{2i}i!}\frac{1}{\gamma^{2i}} + O\Big(\frac{1}{\gamma^{2j}}\Big), \tag{6.24}$$

where $x^{[\text{fall } k]} := x(x-1)\cdots(x-k+1)$. It turns out that for $\alpha \geq 12$, the expansion in (6.23) is more accurate — typically considerably so — than the expansion in (6.24). Indeed, numerical computations show that for given $j = 1, 2, 3, 4$ and $\alpha \geq 12$, the magnitude of the ratio of the O term in (6.23) to the one in (6.24) increases with γ. Thus the supremum of those ratios equals their limit as $\gamma \to \infty$, which is

$$a^{\circ}_{\mathcal{G},j}(-\alpha_\bullet)\alpha_\bullet^j\Big/\Big(\frac{(2\alpha_\bullet)^{[\text{fall } 2j]}}{2^{2j}j!}\Big).$$

This quantity is less than 1 for $j = 1, 2, 3, 4$ and $\alpha \geq 5.12$; as $\alpha \to \infty$ it asymptotes to $\big(\binom{2j}{j}/2^{2j}\big)/(\alpha^j/j!)$. The story is more complicated for $\alpha < 12$, because in that situation there are pairs (α, j) such that $a^{\bullet}_{\mathcal{G},j}(-\alpha_\bullet)$ and/or $(2\alpha_\bullet)^{[\text{fall } 2j]}$ equals 0; the expansion in (6.23) can be considerably less accurate than the one in (6.24), as well as considerably more accurate.

Here are some remarks on part (ii) of Theorem 6.2. Suppose $\alpha > 1/2$. Given $C_\diamond \in \mathbb{R}$, determine $C = C(\alpha, C_\diamond)$ so that $C_\diamond\sqrt{\alpha} = -C\sqrt{\alpha_\bullet}$; thus $C = -\sqrt{\alpha/\alpha_\bullet}\,C_\diamond$. Then

$$A_{\mathcal{N}}(\alpha, C_\diamond\sqrt{\alpha}\,) = A_{\mathcal{N}}(\alpha, -C\sqrt{\alpha_\bullet}\,) \tag{6.25}$$
$$\doteq b(C)\sqrt{B(C)}\,\Big[1 + a^{\bullet}_{\mathcal{N},1}(C)/(2\alpha_\bullet) + a^{\bullet}_{\mathcal{N},2}(C)/(2\alpha_\bullet)^2 + \cdots\Big]. \tag{6.26}$$

The uniform error bounds in (6.21) apply to the asymptotic expansion (6.26) because $C_\diamond$ and C are one-to-one functions of each other. However, one needs to bear in mind that in (6.26), C depends on α as well as $C_\diamond$. Expanding the right-hand side of (6.26) in powers of $1/\alpha$ gives an asymptotic expansion analogous to the one in (6.6) for $A_{\mathcal{G}}(\alpha, C\sqrt{\alpha}\,)$:

$$A_{\mathcal{N}}(\alpha, C_\diamond\sqrt{\alpha}\,) \doteq b_\diamond\sqrt{B_\diamond}\,\big[\,a^{\diamond}_{\mathcal{N},0}(C_\diamond) + a^{\diamond}_{\mathcal{N},1}(C_\diamond)/(2\alpha) + a^{\diamond}_{\mathcal{N},2}(C_\diamond)/(2\alpha)^2 + \cdots\big] \tag{6.27}$$

where

$$a^{\diamond}_{\mathcal{N},0}(C_\diamond) := 1, \tag{6.28a}$$
$$a^{\diamond}_{\mathcal{N},1}(C_\diamond) := \frac{B_\diamond(10B_\diamond^2 - 21B_\diamond + 6)}{12}, \tag{6.28b}$$
$$a^{\diamond}_{\mathcal{N},2}(C_\diamond) := \frac{B_\diamond^2(1540B_\diamond^4 - 5460B_\diamond^3 + 7041B_\diamond^2 - 3756B_\diamond + 612)}{288} \tag{6.28c}$$

with $b_\diamond = b_\diamond(C_\diamond) := b(-C_\diamond)$ and $B_\diamond = B_\diamond(C_\diamond) := B(-C_\diamond)$. The expansion (6.27) is not quite as accurate as (6.26): for $\alpha \geq 2$, one has

$$\sup_{C_\diamond \in \mathbb{R}} \left| b_\diamond \sqrt{B_\diamond} \sum_{i=0}^{j} \frac{a^\diamond_{\mathcal{N},i}(C_\diamond)}{(2\alpha)^i} - A_\mathcal{N}(\alpha, C_\diamond\sqrt{\alpha}\,) \right| \leq \begin{cases} 0.146/(2\alpha), & \text{if } j=0, \\ 0.078/(2\alpha)^2, & \text{if } j=1, \\ 0.044/(2\alpha)^3, & \text{if } j=2. \end{cases} \tag{6.29}$$

Nonetheless, (6.27) confirms some features one sees in Figure 3. First of all, there is symmetry about $C = 0$ in the limit: $\lim_{\alpha\to\infty} A_\mathcal{N}(\alpha, C\sqrt{\alpha}\,) = b(-C)\sqrt{B(-C)} = \lim_{\alpha\to\infty} A_\mathcal{G}(\alpha, -C\sqrt{\alpha}\,)$. Second, the function $a^\diamond_{\mathcal{N},1}(C_\diamond)$ is positive for $C_\diamond$ to the left of -0.67 and negative to the right of that point; thus at least to first order, as $\alpha \uparrow \infty$, $A_\mathcal{N}(\alpha, C_\diamond\sqrt{\alpha}\,)$ decreases down to $b_\diamond\sqrt{B_\diamond}$ for $C_\diamond < -0.67$ and increases up to that limit for $C_\diamond > -0.67$.

6.3. Gamma rejection sampler on the original scale: $\gamma > 0$

Theorem 6.3. *Let $\theta_\alpha(C)$, $A_{0+}(\alpha, C\sqrt{\alpha}\,)$ $s_\alpha(\theta)$, and $s_\infty(\theta)$ be as in Theorem 5.4. Let $b = b(C)$ and $B = B(C)$ be as in (5.1).*

(i) Part (iv) of Theorem 5.4 admits the following sharpening. For $j = 0, 1, 2$, set

$$T_j(\alpha, C) := \sum\nolimits_{i=0}^{j} \tau_i(C)/(2\alpha)^i. \tag{6.30}$$

with

$$\tau_0(C) := 1/b^2, \tag{6.31a}$$
$$\tau_1(C) := 1 - 2B, \tag{6.31b}$$
$$\tau_2(C) := -(1-B)(1-2B)(1-6B^2)/(6B). \tag{6.31c}$$

For $j = 0, 1, 2$ and for each fixed $C_0 \in (0, \infty)$, one has

$$\sup_{0<C\leq C_0} \left| \frac{\theta_\alpha(C) - T_j(\alpha, C)}{1 - \theta_\infty(C)} \right| = O(1/\alpha^{j+1}) \tag{6.32}$$

as $\alpha \to \infty$. In particular, for all $\alpha \geq 1/2$,

$$\sup_{0<C\leq 1} \left| \frac{\theta_\alpha(C) - T_j(\alpha, C)}{1 - \theta_\infty(C)} \right| \leq \begin{cases} 0.73/(2\alpha), & \textit{if } j=0, \\ 0.26/(2\alpha)^2, & \textit{if } j=1, \\ 0.53/(2\alpha)^3, & \textit{if } j=2. \end{cases} \tag{6.33}$$

(ii) Part (v) of Theorem 5.4 admits the following sharpening. For $j = 0, 1, 2$, set

$$A^\bullet_{0+,j}(\alpha, C) := \sqrt{2B} \sum\nolimits_{i=0}^{j} a^\bullet_{0+,i}(C)/(2\alpha)^i \tag{6.34}$$

with

$$a^{\bullet}_{0+,0}(C) := 1, \tag{6.35a}$$

$$a^{\bullet}_{0+,1}(C) := -(B-1)(-1+2B)(5B^2+3B-1)/(12B), \tag{6.35b}$$

$$a^{\bullet}_{0+,2}(C) := (-1+2B)(15B^4-B-1)(B-1)^2/(3B) + \big(a^{\bullet}_{0+,1}(C)\big)^2. \tag{6.35c}$$

For $j = 0, 1, 2$ *and for each fixed* $C_0 \in (0, \infty)$*, one has*

$$\sup\nolimits_{0<C\le C_0}\left|\frac{A^{\bullet}_{0+,j}(\alpha, C) - A_{0+}(\alpha, C\sqrt{\alpha}\,)}{1-\theta_\infty(C)}\right| = O(1/\alpha^{j+1}) \tag{6.36}$$

as $\alpha \to \infty$*. In particular, for all* $\alpha \ge 1/2$*,*

$$\sup\nolimits_{0<C\le 1}\left|\frac{A^{\bullet}_{0+,j}(\alpha, C) - A_{0+}(\alpha, C\sqrt{\alpha}\,)}{1-\theta_\infty(C)}\right| \le \begin{cases} 0.074/(2\alpha), & \text{if } j = 0,\\ 0.205/(2\alpha)^2, & \text{if } j = 1,\\ 0.132/(2\alpha)^3, & \text{if } j = 2.\end{cases} \tag{6.37}$$

Proof. (i) This follows by an elaboration of the argument for part (iv) of Theorem 5.4. (5.30) implies that as $a := 1/(2\alpha)$ tends to 0,

$$s_\alpha(\theta) = s_\infty(\theta) - a/\theta - a^2/(3\theta^2) + O(a^3) \tag{6.38}$$

uniformly for θ bounded away from 0. Formally, (6.32) follows from (6.38) via the usual procedure for inverting an asymptotic expansion. The unboundedness of the first, second, and third derivatives of $s_\infty(\theta)$ for θ near 1 turns out not to cause any problems because in the error analysis, each time one of those derivatives arises, it gets multiplied by a term which is on the order of the reciprocal of the derivative. A computer study gave the specific bounds in (6.33).

(ii) This follows by an elaboration of the argument for part (v) of Theorem 5.4, now using

$$\log\big(\Gamma(r)\big) = \tfrac{1}{2}\log(2\pi) + r\log(r) - r - \tfrac{1}{2}\log(r) + \frac{1}{12r} + O(1/r^3) \text{ as } r \to \infty. \tag{6.39}$$

One first argues that $J(\theta_\alpha) = \hat{J} + O(a^3)$ where $\hat{J} := J\big(\theta_\infty(C) + a\tau_1(C) + a^2\tau_2(C)\big)$ with $a = 1/(2\alpha)$ and $\tau_1(C)$ and $\tau_2(C)$ as in (6.31b) and (6.31c). Then one expands $\hat{J}$ in a Taylor series about $\theta_\infty(C)$; since the nth derivative of J is of order α for $n \ge 2$, one needs to take the third order term in the Taylor expansion into account to get the second order term for the expansion of $\hat{J}$. Finally, one combines the expansion of $\hat{J}$ with that for K coming from (6.14). A computer study gave the specific numerical bounds in (6.37). □

In Panel 1 of Figure 5, $\theta_\alpha(C)$ appears to tend down to $\theta_\infty(C)$ at rate $1/\alpha$; that is what part (i) of the theorem predicts, since the function $\tau_1(C)$ in (6.31b) is strictly positive. Part (ii) underlies something one sees in Figure 2. The function $a^{\bullet}_{0+,1}(C)$ is negative for $C \in (0, 2/3)$ but with a minimum value of only about -0.03, positive and increasing for $C > 2/3$, and asymptotic to $C^2/12$ as

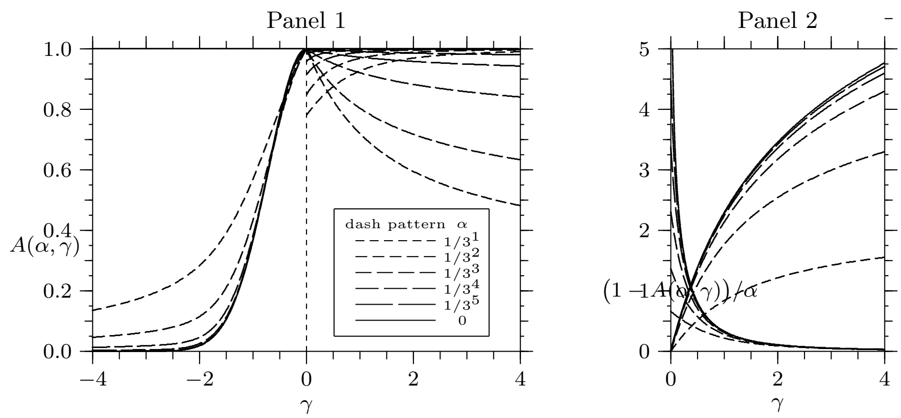

FIG 6. *The left half of Panel 1 graphs* $A_{0-}(\alpha, \gamma)$ *versus* γ *for* $\alpha = 1/3,\ 1/3^2,\ \ldots,\ 1/3^5$ *and in the limit as* $\alpha \to 0$*; see (2.9) for the case* $\alpha > 0$ *and* (7.2) *for the limiting case. Similarly, the right half of Panel 1 graphs* $A_{0+}(\alpha, \gamma)$ *(the downward sloping curves) and* $A_{\mathcal{G}}(\alpha, \gamma)$ *(the upward sloping curves) versus* γ*; see* (2.9) *and* (3.13) *for the case* $\alpha > 0$*. Here the limits as* $\alpha \to 0$ *are identically 1. Panel 2 graphs* $\big(1 - A(\alpha, \gamma)\big)/\alpha$ *versus* γ*, for the same acceptance curves as in the right half of Panel 1, using the same dash patterns; see* (7.3) *and* (7.4) *for the limits as* $\alpha \to 0$*.*

$C \to \infty$. Thus to first order, as α increases, $A_{0+}(\alpha, C\sqrt{\alpha}\,)$ effectively decreases at rate $1/\alpha$, but ever more slowly the larger is C. It is important to bear in mind that the asymptotic expansions (6.30) for $\theta_\alpha(C)$ and (6.34) for $A_{0+}(\alpha, C\sqrt{\alpha}\,)$ are *not* uniform in $C \in (0, \infty)$. These expansions are based on Stirling's approximation to $\Gamma(r)$ for $r = \theta\alpha$ where for large C, $\theta = \theta_\alpha(C)$ tends to be small, on the order $1/C^2$; therefore α needs to be at least a moderate multiple of C^2 in order for the expansion to "kick in".

7. Behavior of the rejection samplers when $\alpha < 1/2$

Here we augment the analysis in the preceding two sections by exploring how our various rejection samplers behave as $\alpha \downarrow 0$. Throughout suppose that $\alpha \in (0, 1/2)$. The sampler $\mathcal{S}_{\mathcal{N}}$ is not available in this situation, confer (3.2), but $\mathcal{S}_{0-}$, $\mathcal{S}_{0+}$, and $\mathcal{S}_{\mathcal{G}}$ are. Panel 1 in Figure 6 graphs the corresponding acceptance rates A_{0-} (for $\gamma < 0$) and A_{0+} and $A_{\mathcal{G}}$ (for $\gamma > 0$).

Throughout this paragraph, suppose $\gamma < 0$. By Theorem 5.3, $A_{0-}(\alpha, \gamma)/A_{\mathcal{G}}(\alpha, \gamma) = \rho(\alpha) > 1$. Accordingly, we have not displayed $A_{\mathcal{G}}$ in the left half of Panel 1. Since $\lim_{\alpha\downarrow 0} \rho(\alpha) = 1$, the rates $A_{0-}(\alpha, \gamma)$ and $A_{\mathcal{G}}(\alpha, \gamma)$ behave the same in the limit; according to (7.2) below, $\lim_{\alpha\downarrow 0} A_{0-}(\alpha, \gamma) = \exp(-\gamma^2)$. The situation for $\gamma < 0$ is thus disappointing: $\mathcal{S}_{0-}(\alpha, \gamma)$ behaves adequately well only for γ's between (say) -1 and 0.

Throughout this paragraph, suppose $\gamma > 0$. The situation in this case is much rosier. Indeed, the inequality $A_{\mathcal{G}}(\alpha, \gamma) \geq A_{\mathcal{G}}(\alpha, 0) = 1/\rho(\alpha)$ implies that as $\alpha \downarrow 0$, $A_{\mathcal{G}}(\alpha, \gamma)$ tends to 1 uniformly in γ. Moreover, as Panel 2 in Figure 6 shows, for all α, $A_{\mathcal{G}}(\alpha, \gamma)$ exceeds $A_{0+}(\alpha, \gamma)$ except when γ is less than (about) 3/8.

One could use $\mathcal{S}_{0+}(\alpha,\gamma)$ for $\gamma < 3/8$ and $\mathcal{S}_{\mathcal{G}}(\alpha,\gamma)$ otherwise. Recall from the discussion following (2.9) that for small α, say $\alpha < 1/32$, the quantity $M := 1/A_{0+}(\alpha,\gamma)$ is effectively $M_{0,+}(\alpha) = \Gamma(\alpha)/\big(2Z(\alpha,\gamma)\big)$.

Theorem 7.1. *(i) Let $2Z(\alpha,\gamma)$ be the normalizing constant for the target density* (1.1). *For each $\gamma \in \mathbb{R}$, $2Z(\alpha,\gamma) \sim \Gamma(\alpha)$ as $\alpha \downarrow 0$; more precisely,*

$$\lim_{\alpha\downarrow 0} \frac{1}{\alpha}\Big(1 - \frac{2Z(\alpha,\gamma)}{\Gamma(\alpha)}\Big) = J(\gamma) := \int_0^\infty \exp(-t)\, \frac{1-\exp\big(-2\gamma\sqrt{t}\,\big)}{t}\, dt. \tag{7.1}$$

(ii) For $\gamma < 0$, let $A_{0-}(\alpha,\gamma)$ be defined by (2.9)*. One has*

$$\lim\nolimits_{\alpha\downarrow 0} A_{0-}(\alpha,\gamma) = \exp(-\gamma^2). \tag{7.2}$$

(iii) For $\gamma > 0$, let $A_{\mathcal{G}}(\alpha,\gamma)$ and $A_{0+}(\alpha,\gamma)$ be defined by (3.13) *and* (2.9)*, respectively. One has*

$$\lim\nolimits_{\alpha\to 0}\Big(\frac{1 - A_{\mathcal{G}}(\alpha,\gamma)}{\alpha}\Big) = J(\gamma) + \Gamma'(1) - \log(4\gamma^2), \tag{7.3}$$

$$\lim\nolimits_{\alpha\downarrow 0}\Big(\frac{1 - A_{0+}(\alpha,\gamma)}{\alpha}\Big) = J(\gamma). \tag{7.4}$$

Proof. (i) Since

$$\frac{1}{\alpha}\Big(1 - \frac{2Z(\alpha,\gamma)}{\Gamma(\alpha)}\Big) = \frac{1}{\Gamma(\alpha+1)}\int_0^\infty t^\alpha y(t)\, dt \ \text{ for } \ y(t) := \exp(-t)\, \frac{1-\exp\big(-2\gamma\sqrt{t}\,\big)}{t}$$

and since $\int_0^\infty (1 \vee t^{1/4})\, |y(t)|\, dt < \infty$, (7.1) follows from the dominated convergence theorem.

(ii) According to (2.9) and (2.4),

$$A_{0-}(\alpha,\gamma) = \frac{2Z(\alpha,\gamma)}{\Gamma(\alpha)}\, e^\alpha \Big(\frac{\delta}{\exp(1/\delta)}\Big)^\alpha \quad \text{with} \quad \delta = \frac{4\alpha}{\big(\sqrt{\gamma^2+4\alpha} + |\gamma|\big)^2}.$$

Since $\delta \sim \alpha/\gamma^2$ as $\alpha \downarrow 0$, (7.2) follows from (7.1).

(iii) First note that for numbers ξ and x_α, one has

$$\lim\nolimits_{\alpha\downarrow 0} \frac{1 - x_\alpha}{\alpha} = \xi \iff \lim\nolimits_{\alpha\downarrow 0} \frac{\log(x_\alpha)}{\alpha} = -\xi. \tag{7.5}$$

Now use (3.13) and (3.10) to write

$$A_{\mathcal{G}}(\alpha,\gamma) = \frac{Z(\alpha,\gamma)}{\Gamma(2\alpha)}\frac{\delta^{2\alpha}}{\exp\big((2\alpha/\delta)^2\big)} = \frac{2Z(\alpha,\gamma)}{\Gamma(\alpha)} \times \frac{\Gamma(1+\alpha)}{\Gamma(1+2\alpha)} \times \frac{\delta^{2\alpha}}{\exp\big((2\alpha/\delta)^2\big)}$$

with $\delta = \gamma + \sqrt{\gamma^2 + 4\alpha} = 2\gamma + O(\alpha)$. By (7.1) and (7.5),

$$\log\big(A_{\mathcal{G}}(\alpha,\gamma)\big) = -\alpha J(\gamma) + (\log\Gamma)'(1)(\alpha - 2\alpha) + 2\alpha\log(2\gamma) + o(\alpha);$$

(7.3) follows from this and (7.5). Next, according to (2.9) and Theorem 2.1,

$$A_{0+}(\alpha,\gamma) = \frac{2Z(\alpha,\gamma)}{\Gamma(\alpha)} \times \frac{\Gamma(\alpha)}{\Gamma(\alpha\theta)} \times \Big(\frac{\gamma e}{\alpha(1-\theta)}\Big)^{2\alpha(1-\theta)}$$

where $\theta = \theta(\alpha,\gamma)$ is the root to the equation $s(\theta) = 0$. As $\alpha \to 0$,

$$(1-\theta) \sim \frac{\gamma}{\alpha \exp(1/(2\alpha)) - \Gamma'(1)/2},$$

see Figure 5, so $1-\theta \to 0$ extremely fast. Therefore,

$$\log\big(A_{0+}(\alpha,\gamma)\big) = \log\big(2Z(\alpha,\gamma)/\Gamma(\alpha)\big) + o(\alpha).$$

Now use (7.1) and (7.5). □

8. Conclusion

We proposed rejection samplers that generate draws from the distribution with density $f(\cdot \,|\, \alpha,\gamma)$, defined in (1.1). The first type of samplers, introduced in Section 2, use Gamma instrumental distributions to target $f(\cdot \,|\, \alpha,\gamma)$ directly. However, this approach leads to practically useful acceptance rates only for small to moderate values of γ. We showed that this problem can be remedied by using alternative samplers that operate on a square-root scale, targeting the distribution with density $h(\cdot \,|\, \alpha,\gamma)$, defined in (1.6). These alternative samplers use normal and Gamma distributions as instrumental distributions; recall Section 3.

Ultimately, we suggest combining all of the proposed samplers, as described in detail in Section 4. To this end, we derived, for any fixed $\alpha \geq 1/2$, a partition of the real line, and the rejection sampler to be used is chosen based upon which of the intervals contains the parameter γ. As pointed out in Remark 4.1, the partition is obtained under the not entirely realistic assumption that generating draws from the different instrumental distributions comes at equal cost. If the assumption is seriously wrong in the programming environment used, then experimenting with adjusted partitions could be beneficial.

Finally, as mentioned in the introduction, the case $\alpha < 1/2$ is not directly relevant for the problem motivating our work. Nevertheless, the samplers with Gamma instrumental distributions apply to this case and were analyzed in this regime in Section 7.

References

Andrews, L. C. (1985). *Special functions for engineers and applied mathematicians.* Macmillan Co., New York. MR779819 (86g:33001)

Besag, J. and Higdon, D. (1999). Bayesian analysis of agricultural field experiments. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **61** 691–746. With discussion and a reply by the authors. MR1722238

Finegold, M. and Drton, M. (2011). Robust graphical modeling of gene networks using classical and alternative t-distributions. *Ann. Appl. Stat.* **5** 1057–1080. MR2840186

Finegold, M. and Drton, M. (2012). Robust Bayesian inference of networks using Dirichlet t-distributions. arXiv:1207.1221.

Gottardo, R., Raftery, A. E., Yeung, K. Y. and Bumgarner, R. E. (2006). Bayesian robust inference for differential gene expression in microarrays with multiple samples. *Biometrics* **62** 10–18, 313. MR2226551

Kolassa, J. E. (2006). *Series approximation methods in statistics*, third ed. *Lecture Notes in Statistics* **88**. Springer-Verlag, New York.

Liu, J. S. (2008). *Monte Carlo strategies in scientific computing. Springer Series in Statistics*. Springer, New York. MR2401592 (2010b:65013)

Liu, C. and Rubin, D. B. (1995). ML estimation of the t distribution using EM and its extensions, ECM and ECME. *Statistica Sinica* **5** 19–39. MR1329287 (96b:62039)

Marsaglia, G. and Tsang, W. W. (2000). A simple method for generating gamma variables. *ACM Trans. Math. Software* **26** 363–372. MR1809948 (2001k:65015)

Robert, C. P. and Casella, G. (2004). *Monte Carlo statistical methods*, second ed. *Springer Texts in Statistics*. Springer-Verlag, New York. MR2080278 (2005d:62006)

Wong, R. (2001). *Asymptotic approximations of integrals. Classics in Applied Mathematics* **34**. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. Corrected reprint of the 1989 original. MR1851050 (2002f:41023)