

# Sampling From a 1-D Truncated Normal Distribution: Theory, Algorithms, and Applications

Michael Hughes  
Computer Science  
Brown University  
Providence, RI  
mhughes@cs.brown.edu

August 3, 2011

## Abstract

The 1-D *truncated* Normal distribution is a one-dimensional probability density function proportional to the 1-D Normal (Gaussian) distribution, except with support truncated to some interval  $a < x < b$  rather than the entire real line. Sampling from this distribution is straight-forward, except when the truncated interval occurs far from the mean. In this report, we discuss efficient and exact sampling techniques for the one-sided truncation  $x > a$  even as  $a \rightarrow \infty$ . We develop a sampling algorithm appropriate for all possible truncation points  $a$ , and discuss its practical utility in variety of applications such as censored data tasks, probit regression, and multinomial regression.

## 1 Introduction

### 1.1 Normal Distribution

The traditional one-dimensional Normal (Gaussian) distribution, parameterized by mean  $\mu$  and variance  $\sigma^2$ , is defined by the probability density function

$$\begin{aligned} p(x) &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\ &= \frac{1}{Z} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \end{aligned} \tag{1}$$

For brevity and notational clarity, we will often use the PDF in the second line, where we have collapsed the normalization constant terms with no dependence on  $x$  into  $Z = \sqrt{2\pi}\sigma$ .

Random variables sampled from this distribution are denoted by  $x \sim \mathcal{N}(\mu, \sigma^2)$ . Samples  $x$  are real numbers and can take values over the entire real line  $(-\infty, +\infty)$ . However, with extremely high probability any sample  $x$  will fall within just a few standard deviations of the mean. For example, the probability that  $x \in (\mu - 4\sigma, \mu + 4\sigma)$  is at least 0.9999.

#### 1.1.1 Standard parameterization

Note that we can always directly translate between a random variable  $X$  drawn from the normal with arbitrary parameters  $\mu, \sigma^2$ , and a similar variable  $\bar{X}$  drawn from the *standard* Normal distribution with fixed mean and variance  $\mu = 0$  and  $\sigma^2 = 1$ .

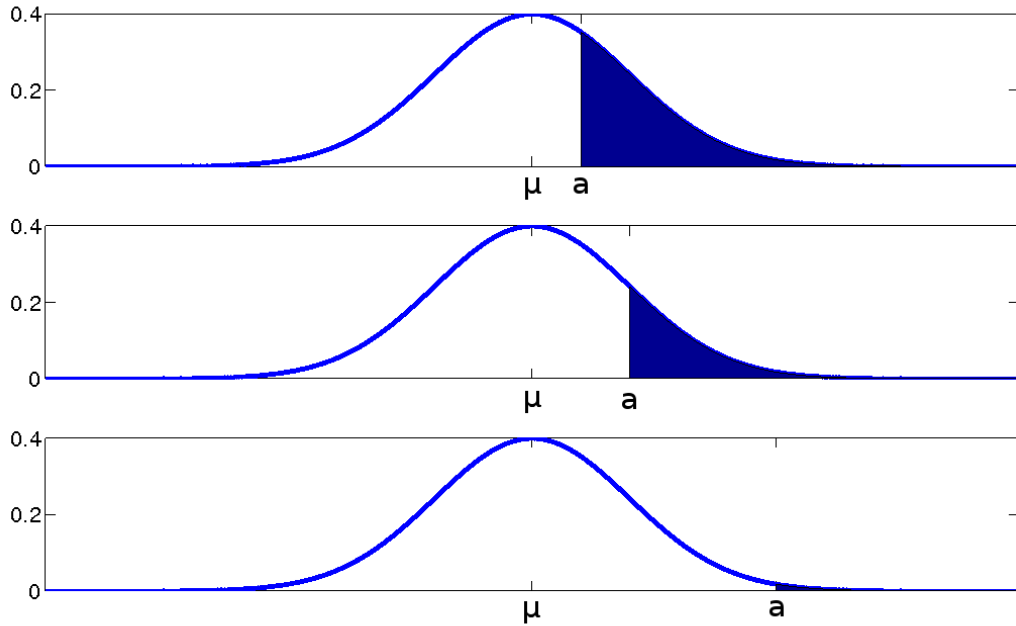


Figure 1: Normal PDF (line) vs. Truncated Normal PDF (shaded) for various values of truncation parameter  $a$ . As  $a$  grows far from the mean, sampling strategies based on evaluations of the normal PDF/CDF will become undesirable, and alternative sampling strategies are necessary.

$$\bar{X} = \frac{X - \mu}{\sigma} \qquad X = \sigma \bar{X} + \mu \qquad (2)$$

This transformation is useful, as it allows us to sample any normal random variable by first sampling from the standard normal, and then scaling and shifting appropriately. Thus, the parameters  $\mu$  and  $\sigma$  are of central importance when considering sampling algorithms for the normal distribution and its relatives.

## 1.2 Truncated Normal Distribution

We are interested in the *truncated* Normal distribution. Random variables  $x$  described by this distribution are drawn from the normal distribution *conditioned* on the fact that  $x \geq a$  for some  $a$ . We have the probability distribution function of a truncated normal random variable  $x$  with parameters mean  $\mu$ , variance  $\sigma^2$ , and lower bound  $a$  as follows:

$$p(x) = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{Z_a} \quad \text{for } x \geq a \qquad (3)$$

$$\text{where } Z_a = \int_a^\infty e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \quad \text{is a normalization constant} \qquad (4)$$

Random variables sampled from this distribution are denoted by  $X \sim \mathcal{N}_{x \geq a}(\mu, \sigma^2)$ , where the subscript denotes the truncation interval. Samples  $x$  are real numbers and can take values over the interval  $(a, +\infty)$  where the parameter  $a$  is any real number.

### 1.2.1 Interpreting the parameter $a$

We observe that  $a$ 's relative value to the mean  $\mu$  controls how close this distribution is to the non-truncated normal.

Consider three regimes:

1.  $a \ll \mu - 3\sigma$  Here, the truncation is effectively irrelevant and this distribution is nearly indistinguishable from the normal distribution
2.  $a$  is near the mean ( $\pm 3\sigma$  or so) Here, a significant fraction of the probability mass of the normal distribution remains after the truncation.
3.  $a \gg \mu + 3\sigma$  Here almost all of the probability mass of the bell curve has been discarded. This is a very different distribution.

### 1.2.2 Standard parameterization

Just as with the normal, the truncated normal has a nice standard parameterization that makes studying sampling much easier. We define the special case where  $\mu = 0, \sigma = 1$  to be the *standard* truncated normal distribution with parameter  $\bar{a}$ . We write  $\bar{x} \sim \mathcal{N}_{\bar{x} \geq a}(\mu = 0, \sigma^2 = 1)$ . Given some arbitrary truncated normal random variable  $X$  with parameters  $\mu, \sigma, a$ , we can always define a translation to  $\bar{X}$  as we did in equation 2, and similarly a translation between  $a$  and  $\bar{a}$  as shown below

$$\bar{a} = \frac{a - \mu}{\sigma} \qquad a = \sigma \bar{a} + \mu \qquad (5)$$

$$(6)$$

This transformation is useful, as it allows us to sample any truncated normal random variable by first sampling from the standard truncated normal with appropriate  $\bar{a}$ , and then scaling and shifting appropriately. Thus, the parameters  $\mu$  and  $\sigma$  are not of central importance here, although *bara* will be.

### 1.2.3 Cumulative distribution function

The cumulative distribution function (CDF) of the standard truncated normal can be seen as a direct scaling and shifting of the standard normal CDF  $\Phi(x)$ .

$$\begin{aligned} p(X \leq x) &= \int_a^x \frac{1}{Z_a} \tilde{p}(x) dx \\ &= \int_a^x \frac{1}{Z_a} \frac{Z}{Z} \tilde{p}(x) dx \\ &= \frac{Z}{Z_a} \int_a^x \frac{1}{Z} \tilde{p}(x) dx \qquad \text{inside of integral is the pdf of the standard normal} \\ &= \frac{Z}{Z_a} (\Phi(x) - \Phi(a)) \end{aligned} \qquad (7)$$

## 2 Sampling from the Standard Truncated Normal: An Overview

Here we consider how to develop a sampling scheme for the standard truncated normal distribution. Our discussion above indicates that the parameter  $\bar{a}$  controls similarity to the standard normal distribution. If

$\bar{a} \ll -3$  is very small, we can essentially just sample from the standard normal distribution. However, if  $\bar{a} \gg 3$ , we will need very different methods. Thus, we are probably best off considering a piece-wise sampling scheme rather than one to handle all possible values of  $\bar{a}$ . We consider regimes (1) and (2) first, where  $a$  is relatively close to  $\mu$  or far below it. Later we develop advanced schemes for cases where  $a$  is far above the mean, where we sample from the extreme tail of the bell curve.

## 2.1 Regime (1): $a \ll \mu$

When  $a \ll 3\sigma$ , the truncation leaves the normal density nearly unchanged, so we can just sample from the standard normal distribution in a straightforward fashion. As  $a$  approaches the mean, we can still use the standard normal as a proposal distribution, although we must adopt a rejection scheme to handle the truncation (see below).

## 2.2 Regime (2): $a \approx \mu$

The most naive approach here is a simple rejection sampler that considers proposals  $x^*$  from the non-truncated standard normal distribution, and accepts them if they are larger than the lower bound  $\bar{a}$ . This sampler is exact, although its efficiency is less than desirable. For  $\bar{a} > 0$ , the acceptance ratio is necessarily less than  $1/2$ , and decays towards zero as  $\bar{a}$  increases.

A better approach is to consider an inversion sampler based on a numerical approximation to the standard truncated normal CDF. We assume the user has access to an efficient evaluator of the standard normal CDF, which we denote  $\Phi(x)$ , and its inverse,  $\Phi^{-1}(p)$ .

The basic idea stems from the fact that one way to sample from the plain vanilla standard normal is to draw  $u \sim \text{Unif}(0, 1)$ , and then let  $\bar{x} = \Phi^{-1}(u)$ . In the truncated case, we simply confine the uniform draw  $u$  to the interval  $(\Phi(\bar{a}), 1)$ , and then assign  $\bar{x} = \Phi^{-1}(u)$ . This ensures the output of the sample lies within  $(a, +\infty)$ . The algorithm is defined fully in 1. See Appendix D for details about its derivation.

## 2.3 Algorithm

---

**Algorithm 1** Sampling inversion of numerical approximation to CDF

---

**Input:** Mean  $\mu$ , variance  $\sigma^2$ , truncation bound  $a$

Functions  $\Phi(\cdot)$  and  $\Phi^{-1}(\cdot)$  for evaluating CDF of standard normal distribution

**Output:**  $x$  s.t.  $x \sim \mathcal{N}_{x \geq a}(\mu, \sigma^2)$

$\bar{a} \leftarrow \frac{a-\mu}{\sigma}$   
 $u \sim \text{Unif}(\Phi(\bar{a}), 1)$   
 $\bar{x} \leftarrow \Phi^{-1}(u)$   
 $x \leftarrow \sigma\bar{x} + \mu$

---

This method turns out to be very efficient and reasonably exact when  $a \approx \mu$ . However, we must remember that the normal CDF  $\Phi(x)$  and its inverse are *numerical approximations*, and so the quality of this scheme will degrade as  $a$  becomes much larger than the mean, and at some large enough  $a$  the numerical methods used to compute  $\Phi(x)$  will not be precise enough to yield correct results. As a case in point, my matlab implementation sometimes gives samples equal to  $+\text{Inf}$  when  $\bar{a} \geq 7$ . In practice it may be difficult to tell what value of  $\bar{a}$  will cause breakdown on a specific platform and algorithm implementation, so I tend to recommend a conservative approach that limits the use of the direct inversion sampler to  $\bar{a} \leq 6$  or so. We'll see that the sampling scheme for regime (3) is reasonably efficient in this regime, so we don't lose too much in this case.

## 2.4 Regime (3): $a \gg \mu$

In this regime, we turn to rejection sampling. The essential idea here is that we can find a tractable distribution  $q(x)$  that tightly bounds  $p(x)$ , and use the bounding relation between the two to convert a sample from  $q(x)$  into a sample from  $p(x)$ . Any function  $q(x)$  is a valid candidate so long as it satisfies two criteria: (1) it is everywhere non-negative and (2) it integrates to one over the domain of  $p(x)$ , which in our case is  $(a, +\infty)$ .

In the next few pages, we'll outline in detail two distinct options for the proposal distribution  $q(x)$ . Our discussion follows from the classic text on sampling from the tail of the normal distribution by Devroye (Ch. 9, pp. 379-382).

## 3 Rejection Sampling based on Rayleigh proposal distribution

Here we consider a rejection sampling scheme with a proposal distribution  $q(x)$  equal to the Rayleigh distribution, truncated such that  $x \geq a$ . We will provide a derivation for a tight bound  $\tilde{p}(x) \leq kq(x)$ , and show that even as  $a \gg \mu$  this rejection sampling scheme remains efficient (and in fact gets more efficient the farther  $a$  increases from the mean).

In the interest of summary, we first provide the entire algorithm. The following subsections provide the necessary derivations and commentary, which more thorough mathematics found in the appendix.

### 3.1 Algorithm

---

**Algorithm 2** Rejection sampler for tail of normal distribution via Rayleigh proposal

---

**Input:** Mean  $\mu$ , variance  $\sigma^2$ , truncation bound  $a$

**Output:**  $x$  s.t.  $x \sim \mathcal{N}_{x \geq a}(\mu, \sigma^2)$

$\bar{a} \leftarrow \frac{a-\mu}{\sigma}$

**repeat**

$u \sim \text{Unif}(0, 1)$

$\bar{x} \leftarrow \sqrt{\bar{a}^2 - 2 \log(1 - u)}$

$v \sim \text{Unif}(0, 1)$

**until**  $v \leq \frac{\bar{x}}{\bar{a}}$

$x \leftarrow \sigma \bar{x} + \mu$

---

### 3.2 Proposal Distribution

We consider the standard Rayleigh distribution, [http://en.wikipedia.org/wiki/Rayleigh\\_distribution](http://en.wikipedia.org/wiki/Rayleigh_distribution), truncated such that  $x \geq a$ . It's PDF and CDF are given below.

$$q(x) = \frac{x\tilde{p}(x)}{C} = \frac{xe^{-\frac{x^2}{2}}}{e^{-\frac{a^2}{2}}} = xe^{\frac{a^2-x^2}{2}} \quad x \geq a \tag{8}$$

$$Q(x) = 1 - e^{\frac{a^2-x^2}{2}} \tag{9}$$

We can confirm that  $q(x)$  meets the necessary criteria for a proposal distribution (see Appendix A for details).

### 3.2.1 Efficient Sampling from $q(x)$

The analytical form of the CDF here makes sampling via the inversion method tractable. If  $u \sim \text{Unif}(0, 1)$ , then we can transform the random variable  $u$  into  $x \sim \text{Rayleigh}$  as follows:

$$x = \sqrt{a^2 - 2 \log(1 - u)} \quad (10)$$

### 3.3 Tight Bound between $q(x)$ and $\tilde{p}(x)$

We need to find some constant  $k$  such that  $\tilde{p}(y) \leq kq(y)$  for all  $y$  in  $(0, \infty)$ . We proceed step-by-step as follows, turning a simple bound into the desired bound one careful algebraic manipulation at a time. Note that some of following requires  $a > 0$ , which is perfectly fine in Regime 3 where  $a \gg \mu = 0$ .

$$\begin{aligned} \tilde{p}(x) &\leq \frac{x}{a} \tilde{p}(x) && \frac{x}{a} \geq 1 \text{ when } a > 0 \\ e^{\frac{a^2}{2}} \tilde{p}(x) &\leq e^{\frac{a^2}{2}} \frac{x}{a} \tilde{p}(x) && \text{positive constant on both sides doesn't change inequality} \\ e^{\frac{a^2}{2}} \tilde{p}(x) &\leq \frac{1}{a} x e^{\frac{a^2 - x^2}{2}} && \text{substitute definition of } \tilde{p}(x) \\ e^{\frac{a^2}{2}} \tilde{p}(x) &\leq \frac{1}{a} q(x) && \text{collapse to definition of } q(x) \\ \tilde{p}(x) &\leq \frac{1}{a} e^{-\frac{a^2}{2}} q(x) \end{aligned} \quad (11)$$

$$(12)$$

We thus have  $k = \frac{1}{a} e^{-\frac{a^2}{2}}$ . We can then proceed to compute the probability of accepting a draw  $x^*$  from  $q(x)$  as a valid sample from  $\tilde{p}(x)$ .

$$\begin{aligned} p(\text{accept}|x^*, a) &= \frac{\tilde{p}(x^*)}{kq(x^*)} \\ &= \frac{\tilde{p}(x^*)}{kx^* e^{\frac{a^2}{2}} \tilde{p}(x^*)} \\ &= \frac{a}{x^*} \end{aligned} \quad (13)$$

### 3.4 Acceptance Rate Analysis

We wish to understand how efficient this rejection scheme is as a function of the lower bound parameter  $a$ . We assume the key quantity of interest is the *expected efficiency* of this rejection sampling scheme. We define this as the mean acceptance rate across all possible values  $x^*$  that could be proposed by  $q(x)$ , weighted by the probability of observing that sample from  $q(x)$ .

$$\begin{aligned}
\mathbb{E}_q[p(\text{accept}|x^*, a)] &= \int_a^\infty \frac{a}{x^*} q(x) dx \\
&= a \int_a^\infty \exp\left[\frac{a^2 - x^2}{2}\right] dx \\
&= ae^{\frac{a^2}{2}} \int_a^\infty e^{-\frac{x^2}{2}} dx \\
&= ae^{\frac{a^2}{2}} \cdot \sqrt{2\pi} [1 - \Phi(a)] \quad \text{where } \sqrt{2\pi} = \int_{-\infty}^\infty \tilde{p}(x) dx
\end{aligned} \tag{14}$$

$$(15)$$

The key question to ask is, what happens as  $a \rightarrow \infty$ ? Taking the limit of the expression in eq. 14 and applying some algebra and L'Hopital's rule yields the following (see Appendix E for details).

$$\begin{aligned}
\lim_{a \rightarrow \infty} \mathbb{E}_q[p(\text{accept}|x^*, a)] &= \lim_{a \rightarrow \infty} ae^{\frac{a^2}{2}} \cdot \sqrt{2\pi} [1 - \Phi(a)] \\
&= \lim_{a \rightarrow \infty} \frac{\sqrt{2\pi} [1 - \Phi(a)]}{\frac{1}{a} e^{-\frac{a^2}{2}}} \\
&= \lim_{a \rightarrow \infty} \frac{(-1)e^{-\frac{a^2}{2}}}{(-1) \left[1 + \frac{1}{a^2}\right] e^{-\frac{a^2}{2}}} \quad \text{l'Hopital's rule applied here} \\
&= \lim_{a \rightarrow \infty} \frac{1}{1 + \frac{1}{a^2}} = 1
\end{aligned} \tag{16}$$

### 3.5 Optimization

We note that this algorithm requires a computation of a square root inside the tight loop, which could be costly if many rejections occur (as noted by Devroye). To postpone computation of this square root until *after* acceptance, we rework the scheme so that it uses transformed variables rather than the originals.

First, we define  $b = \frac{a^2}{2}$ . Next, we draw a candidate value  $w^*$  via the inversion method given  $u \sim \text{Unif}(0, 1)$  as follows:

$$w = b - \log(1 - u) \tag{17}$$

We can verify that  $w^*$  has the same distribution as  $\frac{1}{2}x^{*2}$  when  $x^*$  is drawn as in the original scheme via eq. 10. This implies that the appropriate inverse transform  $x^* = \sqrt{2w^*}$  can yield a valid draw from the proposal  $q(x)$ .

Next, we alter our acceptance procedure to handle these transformed  $w^*$  variables. Formerly, given an candidate proposal  $x^*$ , we computed its acceptance probability by drawing a uniform random variable  $V$  and accepting if  $V \leq \frac{a}{x^*}$ . Here, we use the fact that for any uniformly distributed random variable  $U$ ,  $p(U \leq z) = p(U^2 \leq z^2)$  for any constant  $0 \leq z \leq 1$ . This implies that

$$p(V \leq \frac{a}{x^*} = p(V^2 \leq \frac{a^2}{(x^*)^2}) = p(V^2 \leq \frac{b}{w^*}) \quad \text{since } b = \frac{a^2}{2}, w = \frac{(x^*)^2}{2} \tag{18}$$

We can thus evaluate the acceptance probability in terms of  $b$  and  $w^*$  just by squaring a uniform random variate  $V$ , which allows us to postpone any square root evaluations until we're sure we've drawn an acceptable sample from the target distribution.

---

**Algorithm 3** Optimized rejection sampler for tail of normal distribution via Rayleigh proposal

---

**Input:** Mean  $\mu$ , variance  $\sigma^2$ , truncation bound  $a$ **Output:**  $x$  s.t.  $x \sim \mathcal{N}_{x \geq a}(\mu, \sigma^2)$ 

$$b \leftarrow \frac{1}{2} \left( \frac{a - \mu}{\sigma} \right)^2$$

**repeat**

$$u \sim \text{Unif}(0, 1)$$

$$w \leftarrow C - \log(1 - u)$$

$$v \sim \text{Unif}(0, 1)$$

**until**  $v \leq \frac{w}{b}$ 

$$x \leftarrow \sigma \sqrt{2} * w + \mu$$

---

This scheme leads to the following algorithm, which will probably be slightly faster than 2.

Note also that  $1 - u$  has the same exact distribution as  $u$  when  $u \sim \text{Unif}(0, 1)$ , so we can omit the subtraction step  $1 - u$  in any implementation.



## 4 Rejection Sampling based on Exponential proposal distribution

Here, we consider an alternative proposal based on the Exponential distribution. We first institute a change of variables. Let  $y = x - a$ . Our non-normalized target pdf then becomes:

$$\tilde{p}(y) = e^{-\frac{(y+a)^2}{2}} = e^{-\frac{1}{2}[y^2+2ay+a^2]} \quad y \geq 0 \quad (19)$$

As with the Rayleigh proposal, we first provide the general algorithm for rejection sampling, followed by detailed discussion and mathematical derivation.

### 4.1 Algorithm

---

**Algorithm 4** Rejection sampler for tail of normal distribution via Exponential proposal

---

**Input:** Mean  $\mu$ , variance  $\sigma^2$ , truncation bound  $a$

**Output:**  $x$  s.t.  $x \sim \mathcal{N}_{x \geq a}(\mu, \sigma^2)$

$$\bar{a} \leftarrow \frac{a-\mu}{\sigma}$$

**repeat**

$$u \sim \text{Unif}(0, 1)$$

$$\bar{y} \leftarrow \frac{-1}{\bar{a}} \log(1 - u)$$

$$v \sim \text{Unif}(0, 1)$$

**until**  $v \leq e^{-\frac{\bar{y}^2}{2}}$

$$x \leftarrow \sigma(\bar{y} + \bar{a}) + \mu$$


---

### 4.2 Proposal Distribution

The standard exponential distribution with parameter  $a$  is given by:

$$q(y) = ae^{-ay} \quad y \geq 0 \quad Q(y) = 1 - e^{-ay} \quad (20)$$

We can confirm that  $q(y)$  meets the necessary criteria for a proposal distribution (see Appendix B for details).

#### 4.2.1 Efficient Sampling from $q(x)$

The analytical form of the CDF here makes sampling via the inversion method tractable. If  $u \sim \text{Unif}(0, 1)$ , then we can transform the random variable  $u$  into  $x \sim \text{Exp}(a)$  as follows:

$$y = -\frac{1}{a} \log(1 - u) \quad (21)$$

### 4.3 Tight Bound between $q(x)$ and $\tilde{p}(x)$

We need to find some constant  $k$  such that  $\tilde{p}(y) \leq kq(y)$  for all  $y$  in  $(0, \infty)$ . We proceed step-by-step as follows, turning a simple bound into the desired bound one careful algebraic manipulation at a time.

$$\begin{aligned}
\tilde{p}(y) &\leq e^{\frac{y^2}{2}} \tilde{p}(y) && e^{\frac{y^2}{2}} \geq 1 \text{ since } y^2/2 \geq 0 \text{ for all } y \\
\tilde{p}(y) &\leq e^{\frac{y^2}{2} - \frac{1}{2}[y^2 + 2ay + a^2]} \\
\tilde{p}(y) &\leq e^{-\frac{a^2}{2}} e^{-ay} \\
a\tilde{p}(y) &\leq a e^{-\frac{a^2}{2}} e^{-ay} && \text{multiply both sides by positive constant } a \\
a\tilde{p}(y) &\leq e^{-\frac{a^2}{2}} q(y) \\
\tilde{p}(y) &\leq \frac{1}{a} e^{-\frac{a^2}{2}} q(y)
\end{aligned} \tag{22}$$

We thus have  $k = \frac{1}{a} e^{-\frac{a^2}{2}}$ .

We can then proceed to compute the probability of accepting a draw  $y^*$  from  $q(y)$  as a valid sample from  $\tilde{p}(y)$ .

$$\begin{aligned}
p(\text{accept}|y^*, a) &= \frac{\tilde{p}(y)}{kq(y)} \\
&= \frac{e^{-\frac{1}{2}[y^2 + 2ay + a^2]}}{\frac{1}{a} e^{-\frac{a^2}{2}} a e^{-ay}} = e^{-\frac{(y^*)^2}{2}}
\end{aligned} \tag{23}$$

#### 4.4 Acceptance Rate Analysis

We wish to understand how efficient this rejection scheme is as a function of the lower bound parameter  $a$ . We assume the key quantity of interest is the *expected efficiency* of this rejection sampling scheme given the proposal distribution  $y^*$ . A brief summary of the algebra required here is given below, for full derivation see the Appendix C.

$$\begin{aligned}
\mathbb{E}_q[p(\text{accept}|y^*, a)] &= \int_0^\infty e^{-\frac{1}{2}y^2} a e^{-ay} dy \\
&= a e^{\frac{1}{2}a^2} \int_0^\infty e^{-\frac{(y+a)^2}{2}} dy && \text{complete the square} \\
&= a e^{\frac{1}{2}a^2} \int_a^\infty e^{-\frac{(x)^2}{2}} dx && \text{substitution } x = y + a \\
&= a e^{\frac{1}{2}a^2} \sqrt{2\pi} [1 - \Phi(a)]
\end{aligned} \tag{24}$$

Surprisingly, we notice that our resulting expression here is *exactly* the same as that found in eq 14 for the Rayleigh distribution. Thus, when we let  $a \rightarrow \infty$  we expect that acceptance efficiency approaches 1. This sampler is just as theoretically efficient as the Rayleigh.

#### 4.5 Alternative Algorithm

The original algorithm 4 is straight-forward to understand, but it can be instructive to consider a slight modification, which is the algorithm presented in Devroye (1986).

Recall that the probability of accepting a sample  $y^*$  from  $q(y)$  here is  $e^{-\frac{1}{2}(y^*)^2}$ . We can readily decide whether to accept a given  $y^*$  by drawing a uniform random variate  $V$ , and checking whether it is less than or equal to  $e^{-\frac{1}{2}(y^*)^2}$ , because  $p(V \leq c) = c$  for any constant  $c$ . However, an alternate way to decide is to draw an exponential random variate  $Z \sim \text{Exp}(1)$ , and then accept  $y^*$  if  $Z > \frac{1}{2}(y^*)^2$ .

This is valid because for any constant  $c$  and  $Z \sim \text{Exp}(1)$ , we have the following fact for the random variable  $Z$ :

$$\begin{aligned} p(Z > c) &= 1 - p(Z \leq c) \\ &= 1 - [1 - e^{-c}] \\ &= e^{-c} \end{aligned} \tag{25}$$

We build our alternative algorithm by realizing that if  $w^* \sim \text{Exp}(1)$ , then  $\frac{w^*}{a}$  has an equivalent distribution to  $y^*$ . So the new sampling algorithm simply consists of a two step tight-loop: first, we draw samples  $w^*$  and  $z$  from an exponential distribution with parameter 1, and then decide to accept  $w^*$  with probability  $z > \frac{(w^*)^2}{2a^2}$ . Once we have an accepted  $w^*$ , we transform it to  $y^* = \frac{w^*}{a}$  to have a valid sample from  $q(y)$ . This is summarized in Algorithm 5. This algorithm is ideally *slightly* more efficient because we do not have to multiply by  $1/a$  inside the proposal loop as we did in Alg. 4, which might save a very very small fraction of time (although this might depend on the architecture).

Note that a simple, efficient, and exact numerical method to draw a sample from  $z \sim \text{Exp}(1)$  is to first draw  $u \sim \text{Unif}(0, 1)$  and compute the inverse transform  $z = -\log(u)$ .

---

**Algorithm 5** Alternative rejection sampler for tail of normal distribution via Exponential proposal

---

**Input:** Mean  $\mu$ , variance  $\sigma^2$ , truncation bound  $a$

**Output:**  $x$  s.t.  $x \sim \mathcal{N}_{x \geq a}(\mu, \sigma^2)$

$$\bar{a} \leftarrow \frac{a - \mu}{\sigma}$$

**repeat**

$$w \sim \text{Exp}(1)$$

$$z \sim \text{Exp}(1)$$

**until**  $2\bar{a}^2 \cdot z > w^2$

$$x \leftarrow \sigma \left( \frac{w}{\bar{a}} + \bar{a} \right) + \mu$$


---

## 5 Theoretical and Experimental Comparison of different sampling schemes

Given the above sampling schemes, it is of interest to compare qualitative and quantitative performance in the regime  $a \gg \mu$ . Here, for clarity and brevity we give nicknames to each sampling scheme: **CDF** refers to the numerical inversion method, **Ray** refers to the Rayleigh proposal rejection sampler, and **Exp** is the exponential proposal rejection sampler.

### 5.1 Qualitative Checks on Proposal Bounds

First, a qualitative investigation compares how tight the bounds on  $\tilde{p}(x)$  are for the different proposals. Figure 5.1 shows the non-normalized log pdf of the standard truncated normal (aka  $\tilde{p}(x)$ ) alongside the Rayleigh and Exponential proposal distributions for various  $\bar{a}$ . We can observe that in general, the Rayleigh proposal captures the shape of the target pdf much better than the Exponential. However, this trend is much better

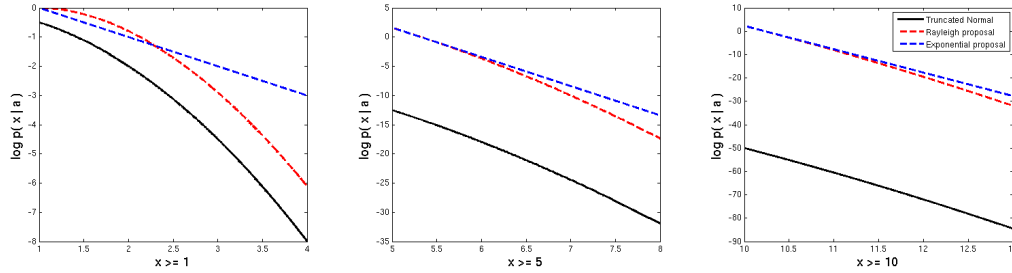


Figure 2: Comparison of  $\log q(x)$  vs.  $\log \tilde{p}(x)$  for different proposal distributions and different values of  $\bar{a}$  for the standard truncated normal.

for low  $\bar{a}$ , and gets worse as  $\bar{a}$  increases. Furthermore, we remember that the bulk of the probability mass is distributed at  $x$  values very near  $\bar{a}$ , so a good fit in the extreme tails probably doesn't mean much for sampling performance.

## 5.2 Assurance of Implementation Correctness

After implementing each of the three schemes, we can collect samples from each and compare against each other as well as theoretical results to ensure correctness. One of the most straightforward ways to check that the correct distribution is achieved is to create a quantile-quantile (QQ) plot<sup>1</sup>. We compare the quantiles of one distribution (either empirical or theoretical) directly to those of another, and check how well they align (draws from the same distribution should be perfectly aligned). In a QQ-plot, computed quantiles for distributions  $X$  and  $Y$  are plotted as points  $(x, y)$  in a scatterplot, and the resulting plot should have strong alignment along the line  $y = x$  if the distributions  $X, Y$  are the same.

We can see QQ-plots for all possible pairwise comparisons in figure 5.2. All comparisons show strong correlations along the line  $y = x$ , which indicates that all three implementations seem to be drawing correctly from the target distribution.

## 5.3 Empirical Acceptance Rates Comparison

For the rejection samplers, we can also validate that we achieve the empirical acceptance rates predicted by our theoretical analysis. Figure 5.3 shows a plot of the measured/theoretical expected efficiency of each scheme as a function of the truncation bound  $\bar{a}$ . We can see that all sampler implementation seem to meet theoretical expectations, with empirical efficiency rapidly approaching 1 as  $\bar{a}$  increases. Furthermore, the samplers appear to be practically efficient: For moderate values like  $\bar{a} = 1$  we achieve acceptance rates above 60%, while for  $\bar{a} = 3$  the rate is above 90% and for  $\bar{a} = 7$  (roughly the boundary at which the CDF method becomes unstable and rejection schemes are necessary) we achieve slightly above 98% efficiency.

## 5.4 Execution Time Comparison

Since both Exp and Ray schemes provide satisfactory and comparable performance in terms of correctness and rejection efficiency, there remains only to test the executime time of each algorithm to determine if one should be preferred. Devroye hints that the Exp scheme might be more efficient, but

Here, we report the results comparing Matlab implementations of the different schemes discussed. The actual Matlab code used to generate samples as well as test harness code for comparison is provided alongside this report. As an experiment, we fixed the value of  $\bar{a} = 7$  and generated 10 million samples using four schemes: plain (Alg. 2) as well as and fast (Alg. 3) versions of the Rayleigh proposal sampler, as well as plain (Alg.

<sup>1</sup><http://itl.nist.gov/div898/handbook/eda/section3/qqplot.htm>

derivation

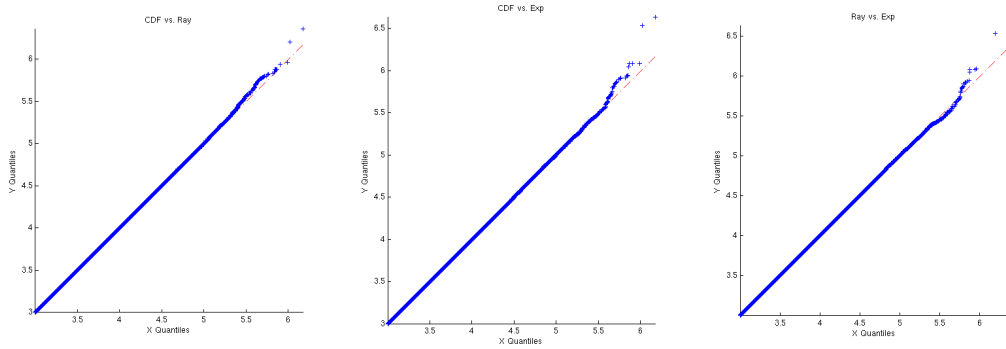


Figure 3: QQ-plot pair-wise comparison of empirical distributions achieved from 5 million samples from each sampling scheme. All comparisons show strong correlations along the line  $y = x$ , which indicates that all three implementations seem to be drawing correctly from the target distribution.

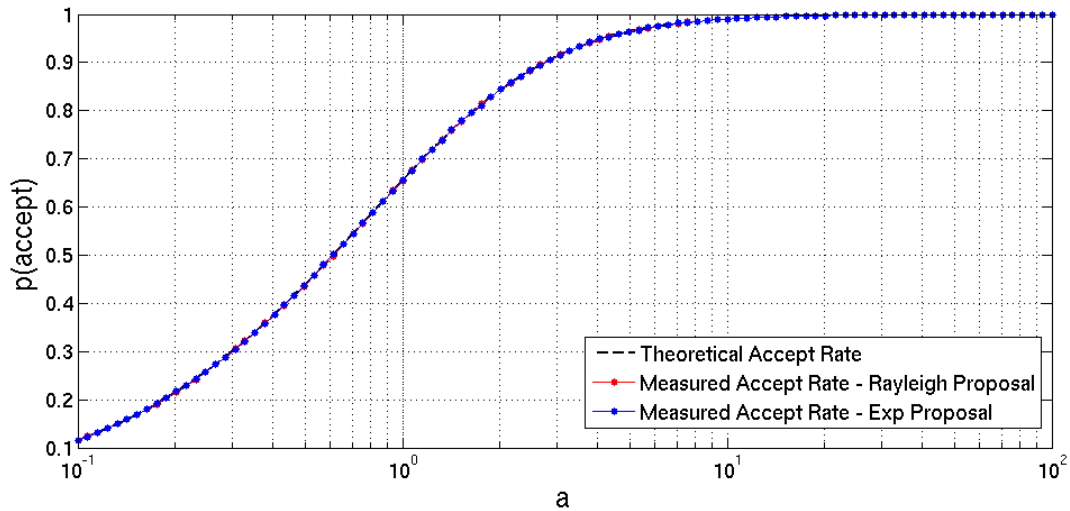


Figure 4: Mean acceptance rates as a function of truncation bound  $\bar{a}$  for different rejection sampling schemes. Appears to be negligible qualitative difference between theoretical expectations and numerical implementation.

Table 1: Execution Time to Draw 10 million samples from standard truncated normal where  $\bar{a} = 7$ .

	Ray	RayFast	Exp	Exp2	CDF
Absolute Time (s)	2.65	2.35	2.70	2.70	1.38
PercDiff from ExpFast	+13%		+15%	+15%	

4) and fast (see Devroye) versions of the Exp sampler. Coding styles were kept as similar as possible (see attached code for details).

We report results in absolute time required as well as percent difference from the leading algorithm (ExpFast) in Table 1.

We observe two important results. First, it *does* seem that the Rayleigh proposal is about 15% faster than the Exponential proposal, even in high acceptance regimes ( $\bar{a} \approx 7$ ). Even without optimization, the Rayleigh proposal appears to be slightly better, but the difference is close enough to probably not matter. Second, our attempt to optimize the Rayleigh proposal appears successful, since postponing the square root allows a  $\approx 10\%$  improvement in speed even in high-acceptance regimes. Thus, we recommend the optimized Rayleigh proposal rejection sampler for its moderate gains in execution time as well as its slightly easier implementation.

Finally, we note that the direct CDF method is significantly faster (1.7x) than any rejection sampler, even in this high-acceptance regime. Thus, we should be somewhat careful to tune a cross-regime algorithm to find the right trade-off between execution time and numerical instability. Yielding to the rejection sampler too early could require unnecessary computation time, but erring on the wrong side of numerical stability could be disastrous.

## A Rayleigh Distribution for $x \geq a$

Here we provide proof that the truncated Rayleigh distribution (defined below and in eq. 8 ) used as a proposal for a rejection sampler meets the necessary criteria.

$$q(x) = \frac{x\tilde{p}(x)}{C} = \frac{xe^{-\frac{x^2}{2}}}{e^{-\frac{a^2}{2}}} = xe^{\frac{a^2-x^2}{2}} \quad x \geq a$$

$$Q(x) = 1 - e^{\frac{a^2-x^2}{2}} \tag{27}$$

$$\tag{28}$$

These criteria are: (1)  $q(x)$  is non-negative and (2) integrates to unity. Both must be satisfied across the entire support of  $\tilde{p}(x)$ , which here is the interval  $(a, \infty)$ .

**Criteria 1:** Restricted to the case  $x \geq a \geq 0$ , we can see by inspection that  $xe^{\frac{a^2-x^2}{2}}$  is non-negative across  $x \in (a, \infty)$ , since  $x \geq 0$  in this interval and  $\exp(y) \geq 0$  for any  $y$ .

**Criteria 2:** We show that  $q(x)$  integrates to one via a few steps of calculus and algebra.

$$\begin{aligned} \int_a^\infty q(x)dx &= \int_a^\infty \frac{xe^{-\frac{x^2}{2}}}{e^{-\frac{a^2}{2}}} dx \\ &= e^{\frac{a^2}{2}} \int_a^\infty xe^{-\frac{x^2}{2}} dx \\ &= e^{\frac{a^2}{2}} (-1) \int_{-a^2/2}^{-\infty} e^w dw \quad \text{substitution } w = -\frac{x^2}{2} \end{aligned} \tag{29}$$

$$= e^{\frac{a^2}{2}} (-1) e^w \Big|_{-a^2/2}^{-\infty} \tag{30}$$

$$= e^{\frac{a^2}{2}} \left[ e^{-a^2/2} - e^{-\infty} \right] \tag{31}$$

$$= \exp \left[ \frac{a^2 - a^2}{2} \right] = e^0 = 1 \tag{32}$$

**Valid CDF:** derivation We can verify also that the CDF is valid by ensuring that at the lowest value of  $x$  ( $x = a$ ), the CDF evaluates to zero, while at the largest  $x = \infty$  we have the CDF equal to one.

$$Q(a) = 1 - e^{\frac{a^2-a^2}{2}} = 1 - e^0 = 0 \quad \lim_{x \rightarrow \infty} Q(x) = 1 - e^{\frac{a^2-\infty^2}{2}} = 1 - e^{-\infty} = 1 \tag{33}$$

## B Exponential Distribution for $x \geq a$

Here we provide proof that the Exponential distribution used as a proposal for a rejection sampler meets the necessary criteria.

$$q(y) = ae^{-ay} \quad y \geq 0 \quad Q(y) = 1 - e^{-ay} \tag{34}$$

**Criteria 1:** Restricted to the case  $y \geq 0$  and  $a \geq 0$ , we can see by inspection that  $ae^{-ay}$  is non-negative across  $y \in (0, \infty)$ , since  $a \geq 0$  in this interval and  $\exp(y) \geq 0$  for any  $y$ , positive or negative.

**Criteria 2:** The properties of the Exponential distribution for  $0 \leq y \leq \infty$  are well-known, so we omit this proof as unnecessary.

## C Expected Efficiency of Rejection Sampler with Exponential Proposal Distribution

$$\begin{aligned}
\mathbb{E}_q[p(\text{accept}|y^*, a)] &= \int_0^\infty e^{-\frac{1}{2}y^2} a e^{-ay} dy \\
&= a \int_0^\infty e^{-\frac{1}{2}y^2 - ay} dy \\
&= a e^{\frac{1}{2}a^2} \int_0^\infty e^{-\frac{1}{2}[y^2 + 2ay + a^2]} dy && \text{complete the square} \\
&= a e^{\frac{1}{2}a^2} \int_0^\infty e^{-\frac{(y+a)^2}{2}} dy && \text{complete the square} \\
&= a e^{\frac{1}{2}a^2} \int_a^\infty e^{-\frac{(x)^2}{2}} dx && \text{substitution } x = y + a \\
&= a e^{\frac{1}{2}a^2} \sqrt{2\pi} [1 - \Phi(a)] && (35)
\end{aligned}$$

## D Inverting the CDF of the Truncated Normal Distribution

Suppose we had a valid numerical CDF for the truncated normal distribution, which we denote  $F(x)$  and its inverse  $F^{-1}(p)$ . Then by standard arguments, we can always take a uniform random variable  $U$  over the interval  $(0, 1)$  and compute a sample from our target distribution by transforming  $U$  into  $X$  via the inverse CDF:  $X \rightarrow F^{-1}(U)$ .

However, we do not have a way to evaluate  $F$  or its inverse directly. Instead, we have the relation defined in eq. (7):

$$F(x) = \frac{Z}{Z_a} (\Phi(x) - \Phi(a)) \quad (36)$$

Inverting the right-hand side of this expression, we can find a sample  $x$  from the truncated normal given  $u$  uniformly distributed in  $(0, 1)$ .

$$\text{derivation } x = \Phi^{-1}\left(\frac{Z_a}{Z}u + \Phi(a)\right) \quad (37)$$

Unfortunately, this expression involves the normalization constant  $Z_a$  that we do not have in closed form, so we can't evaluate this directly. Instead, we next recognize that the quantity  $V_a = \frac{Z_a}{Z}u + \Phi(a)$  inside the inverse CDF here can be thought of as a random variable with a particular distribution.

Given a fixed value  $a$ , we have  $\frac{Z_a}{Z}$  and  $\Phi(a)$  as constants. It is straightforward to show that for any uniform  $(0, 1)$  random variable  $U$ , the quantity  $V = cU + b$  has a uniform distribution over the range  $(b, b + c)$ . In our case, we have that  $V_a$  is uniform over the interval  $(\Phi(a), \Phi(a) + \frac{Z_a}{Z})$ .



We also recognize the following identity:

$$\begin{aligned}
\Phi(a) &= \int_{-\infty}^a \frac{1}{Z} \tilde{p}(x) dx \\
&= \frac{1}{Z} \int_{-\infty}^a \tilde{p}(x) dx \\
&= \frac{1}{Z} \left[ \int_{-\infty}^{\infty} \tilde{p}(x) dx - \int_a^{\infty} \tilde{p}(x) dx \right] \\
&= \frac{1}{Z} [Z - Z_a] \\
&= \frac{Z - Z_a}{Z}
\end{aligned} \tag{38}$$

$$\tag{39}$$

Rearranging this, we find that  $1 = \Phi(a) + \frac{Z_a}{Z}$ , which implies that  $V_a$  is uniform over the interval  $(\Phi(a), 1)$ . Thus, as long as we have  $\Phi$  and  $\Phi^{-1}$  functions available to compute the standard normal CDF and its inverse, we can draw a sample  $X$  from the truncated standard normal via the two-step inversion process:

1. Draw  $V_a \sim \text{Unif}(\Phi(a), 1)$
2.  $X \rightarrow \Phi^{-1}(V_a)$ .

## E Computing Expected Efficiency as $a \rightarrow \infty$ via L'Hopital's rule

Recall that we study rejection samplers because the naive approach for sampling from a truncated normal becomes horribly inefficient as  $a$  increases, and the numerical inversion method becomes inexact and unstable for  $a \gg \mu$ . We thus wish to show that the mean acceptance probability of the rejection samplers remains high as  $a$  grows to infinity.

We have from the main text that the expected acceptance probability of both the Rayleigh and Exponential sampling schemes is given by

$$\mathbb{E}_q(\text{accept}|a) = ae^{\frac{1}{2}a^2} \sqrt{2\pi} [1 - \Phi(a)] \tag{40}$$

We find it convenient to expression this as the ratio of two functions,  $f(a)$  and  $g(a)$ , defined as follows:

$$\begin{aligned}
f(a) &= \sqrt{2\pi} [1 - \Phi(a)] \\
g(a) &= \frac{1}{a \exp\left[\frac{a^2}{2}\right]} = \frac{1}{a} \exp\left[-\frac{1}{2}a^2\right]
\end{aligned} \tag{41}$$

Now we have

$$\begin{aligned}
\lim_{a \rightarrow \infty} \mathbb{E}_q(\text{accept}|a) &= \lim_{a \rightarrow \infty} \frac{f(a)}{g(a)} \\
&= \frac{0}{0}
\end{aligned} \tag{42}$$

This limit has the indeterminate form  $\frac{0}{0}$ , since

$$\begin{aligned}
\lim_{a \rightarrow \infty} f(a) &= \sqrt{2\pi} \lim_{a \rightarrow \infty} [1 - \Phi(a)] = 0 \\
\lim_{a \rightarrow \infty} g(a) &= \lim_{a \rightarrow \infty} \frac{1}{a} \exp\left[-\frac{a^2}{2}\right] = 0
\end{aligned} \tag{43}$$

However, both of these separate terms are individually differentiable, so we can apply l'Hopital's rule. For a quick reference, see [http://en.wikipedia.org/wiki/L'H%C3%B4pital's\\_rule](http://en.wikipedia.org/wiki/L'H%C3%B4pital's_rule), and compute this limit by computing the limit of  $\frac{f'(a)}{g'(a)}$ .

The first-order derivatives with respect to variable  $a$  are defined below:

$$\begin{aligned}
 \frac{d}{da}f(a) &= \sqrt{2\pi} \frac{d}{da} [1 - \Phi(a)] \\
 &= \sqrt{2\pi} \left[ (-1) \frac{1}{\sqrt{2\pi}} e^{-a^2/2} \right] && \text{derivative of the CDF is the PDF} \\
 &= (-1)e^{-\frac{a^2}{2}}
 \end{aligned} \tag{44}$$

$$\begin{aligned}
 \frac{d}{da}g(a) &= \frac{d}{da} \left[ \frac{1}{a} \exp\left(-\frac{1}{2}a^2\right) \right] \\
 &= \frac{d}{da} \left[ \frac{1}{a} \right] \exp\left(-\frac{1}{2}a^2\right) + \frac{1}{a} \frac{d}{da} \left[ \exp\left(-\frac{1}{2}a^2\right) \right] \\
 &= \frac{-1}{a^2} e^{-\frac{a^2}{2}} + \frac{1}{a} \left[ e^{-\frac{a^2}{2}} \frac{d}{da} \left[ -\frac{a^2}{2} \right] \right] \\
 &= \frac{-1}{a^2} e^{-\frac{a^2}{2}} + \frac{1}{a} \left[ e^{-\frac{a^2}{2}} (-a) \right] \\
 &= e^{-\frac{a^2}{2}} \left[ \frac{-1}{a^2} - 1 \right] \\
 &= (-1)e^{-\frac{a^2}{2}} \left[ \frac{1}{a^2} + 1 \right]
 \end{aligned} \tag{45}$$

We can now return to computing our limit

$$\begin{aligned}
 \lim_{a \rightarrow \infty} \mathbb{E}_q(\text{accept}|a) &= \lim_{a \rightarrow \infty} \frac{f'(a)}{g'(a)} \\
 &= \lim_{a \rightarrow \infty} \frac{(-1)e^{-\frac{a^2}{2}}}{(-1)e^{-\frac{a^2}{2}} \left[ \frac{1}{a^2} + 1 \right]} \\
 &= \lim_{a \rightarrow \infty} \frac{1}{\frac{1}{a^2} + 1} = 1
 \end{aligned} \tag{46}$$