CHAPTER 7

GENERATING STANDARD NORMALS

Here we discuss several techniques for transforming Uniform $(0, 1)$ random numbers into standard normal random numbers. Specifically, we discuss the inverse transform, Box-Muller, and polar methods. We have already seen how to generate standard normals via accept/reject; the methods presented here are much more efficient.

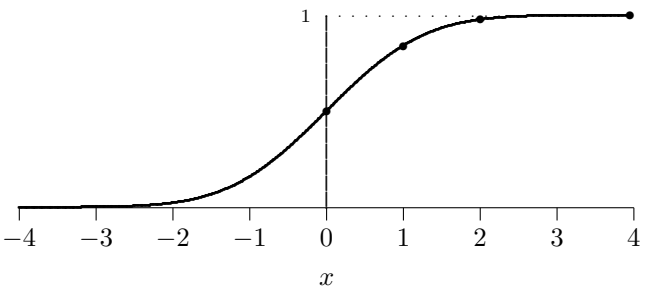
§1. Inverse Transform

Let $\Psi(x)$ denote the distribution function for a standard normal,

$$\Psi(x) \;=\; \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt,$$

as shown in Figure 1. We have seen in Chapter 6 that the random variable $X = \Psi^{-1}(U)$ has distribution function $\Psi(x)$ and is therefore a standard normal. Unfortunately, neither the distribution function $\Psi(x)$ nor its inverse $\Psi^{-1}(u)$ may be expressed in terms of familiar functions. However, the inverse transform method may be used to construct random variables that are *approximately* standard normal.

Figure 1. The standard normal distribution function $u = \Psi(x)$.



Our objective, then, is to construct an approximate inverse function. The function $\Psi^{-1}(u)$ qualitatively resembles a rational function (a ratio of two polynomials): it has vertical asymptotes at $u = 0$ and $u = 1$ and is 0 at $u = 1/2$ (look ahead to Figure 5, where Ψ^{-1} is plotted). Indeed, the rational function $\frac{u-(1/2)}{u(1-u)}$ exhibits this same behavior, although it is not a good approximation of

Ψ^{-1}. The problem with this approach is that $\Psi(x)$ approaches 0 (respectively 1) very very rapidly as x goes to $-\infty$ (respectively $+\infty$). As a consequence, rational functions are not a good approximation to $\Psi^{-1}(u)$ for values of u near the asymptotes at 0 and 1.

In fact, it can be shown that $\Psi(x) \approx 1 - \frac{1}{\sqrt{2\pi}x} e^{-x^2/2}$ as $x \to \infty$ (see §4). Now the expression $\frac{1}{x} e^{-x^2/2}$ is not analytically invertible, whereas $e^{-x^2/2}$ is. Suppose $u = \Psi(x)$ were precisely equal to $1 - e^{-x^2/2}$ for all $x \geq 0$. Then we would have $x = \Psi^{-1}(u) = \sqrt{-2\ln(1-u)}$ for $u \geq \frac{1}{2}$. Since this is not an equality, the strategy is to approximate $\Psi^{-1}(u)$ by a rational function of the quantity $w(u) = \sqrt{-2\ln(1-u)}$, i.e., take

$$\Psi^{-1}(u) \;\approx\; R(w(u)) \tag{1}$$

for some rational function $R(w)$.

The table below shows the pair $(w(u), \Psi^{-1}(u))$ corresponding to the four points (x, u) that are highlighted in Figure 1. (See the §4 for numerical approaches for computing $u = \Psi(x)$.)

Figure 2. The four points highlighted in Figure 1.

x	$u = \Psi(x)$	$(w(u),\ \Psi^{-1}(u))$
0	0.50000	(1.17741, 0)
1	0.84134	(1.91887, 1)
2	0.97725	(2.75070, 2)
4	0.99997	(4.55195, 4)

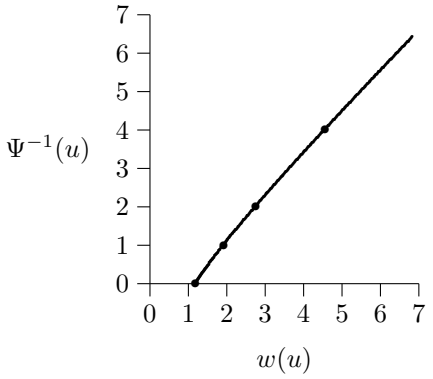
Figure 3 shows a graph of points of the form $(w(u), \Psi^{-1}(u))$ for values of u ranging from $\frac{1}{2}$ to almost 1 with the four points in this table highlighted. This graph is nearly linear and is therefore nicely modeled by rational functions where the degree of the numerator is one more than the degree of the denominator. To this end we take

$$R(w) \;=\; \frac{Aw^2 + Bw + C}{Dw + 1}, \tag{2}$$

which has 4 free parameters. (The constant 1 in the denominator does not reduce generality as both top and bottom may be scaled as necessary to produce 1 in that location.) Here we choose the parameters to explain the points on the graph indicated by the $\bullet$s, i.e., so that:

$$\begin{gathered} R(1.17741) = 0; \quad R(1.91887) = 1; \quad R(2.75070) = 2; \\ \text{and} \quad R(4.55195) = 4. \end{gathered} \tag{3}$$

Figure 3. A plot of points $(w(u), \Psi^{-1}(u))$ *for* $\frac{1}{2} \le u < 1$.



Rearranging (2) gives

$$w^2 \cdot A + w \cdot B + C - wR(w) \cdot D = R(w).$$

Each item of data in (3) generates a linear equation producing the system

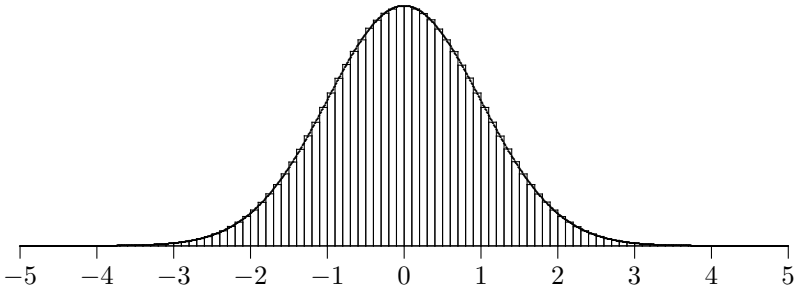
$$\begin{aligned} 1.38629A + 1.17741B + C &= 0 \\ 3.68204A + 1.91887B + C - 1.91887D &= 1 \\ 7.56637A + 2.75070B + C - 5.50141D &= 2 \\ 20.72029A + 4.55195B + C - 18.20782D &= 4, \end{aligned}$$

whose solution is $A = 1.20447$, $B = 0.69322$, $C = -2.48595$, and $D = 1.18776$. With these parameter values, the right side of (1) provides a good approximation of $\Psi^{-1}(u)$ for $\frac{1}{2} \le u < 1$. Reflecting the symmetry $\Psi^{-1}(u) = -\Psi^{-1}(1-u)$, we take

$$\Psi^{-1}(u) \approx \begin{cases} R\big(\sqrt{-2\ln(1-u)}\,\big) & \text{if } \frac{1}{2} \le u \\ -R\big(\sqrt{-2\ln u}\,\big) & \text{if } u < \frac{1}{2}. \end{cases} \tag{4}$$

This is implemented in `CDHNormals.cpp`, where approximate standard normals are generated at the rate of 11 million per second. A histogram of the data is shown in Figure 4 with a scaled normal density function for comparison. Of course, a serious implementation of this approach would involve constructing an R that fits more than just 4 points on the graph of Figure 3.

Figure 4. Histogram of `CDHNormals.cpp`.



A More Efficient Variation. The above algorithm is computationally intensive due to the logarithm and square root functions that are invoked with each application. A more efficient and widely-used algorithm due to Peter Acklam capitalizes on the fact that $\Psi^{-1}(u)$ may be well approximated in a central region $p \le u \le 1-p$ by a rational function of u, rather than of $w(u) = \sqrt{-2\log(1-u)}$, thus largely bypassing those expensive calculations.

Figure 5. $\Psi^{-1}(u)$ and an approximating rational function.

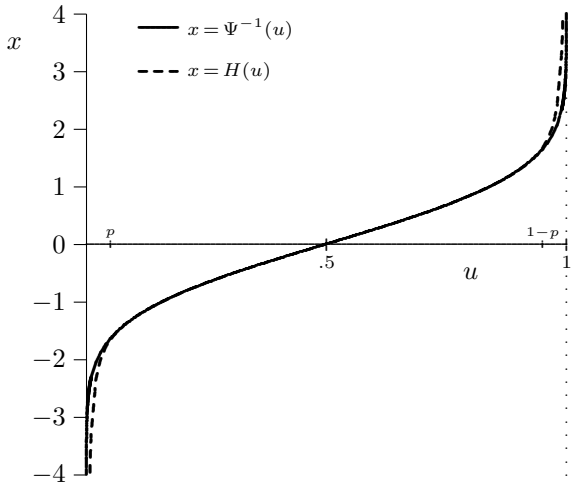


Figure 5 Shows a rational function $H(u)$ that closely approximates $\Psi^{-1}(u)$ on an interval $[p, 1-p]$. Outside of this central region, the graph of H runs off to ∞ ($-\infty$ to the left) much faster than Ψ^{-1}. As previously discussed, this is because the normal distribution has very slim tails. This shows up clearly in Figure 5 where we see that the graph of Ψ^{-1} is much closer to the asymptotes than the graph of H. Any approximation of Ψ^{-1} by a rational function will exhibit this phenomenon. Consequently, this algorithm uses the approximation $\Psi^{-1}(u) \approx H(u)$ only in the central range $u \in [p, 1-p]$. Outside of this range it

uses the previous methodology (4), yielding

$$\Psi^{-1}(u) \;\approx\; \begin{cases} R\big(\sqrt{-2\ln(1-u)}\,\big) & \text{if } u > 1-p \\ H(u) & \text{if } u \in [p, 1-p] \\ -R\big(\sqrt{-2\ln u}\,\big) & \text{if } u < p. \end{cases} \tag{5}$$

In Acklam's implementation $p = 0.02425$ and the rational functions H and R are of much higher degree. There are 21 free parameters allowing for a near perfect approximation of Ψ^{-1}. This algorithm is implemented in the program `PANormals`. On my laptop, it generates high quality standard normals at the rate of 17 million per second.

§2. Box-Muller

Box-Muller is a technique based on a transformation to polar coordinates. We need the follow fact.

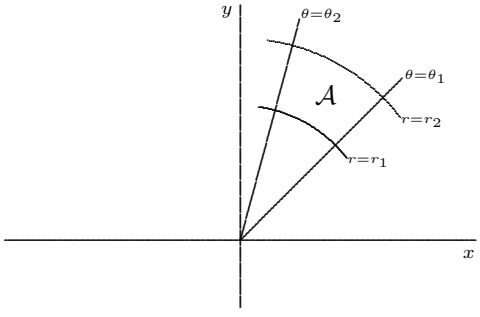
Theorem 1. Suppose X and Y are random variables and let Θ and R denote the polar coordinates corresponding to the Cartesian coordinates (X, Y), so $X = R\cos(\Theta)$ and $Y = R\sin(\Theta)$ with $0 \le \Theta < 2\pi$ and $R \ge 0$. Then the following are equivalent:

(i) X and Y have joint density function $f(x, y)$ where, for any $r > 0$, $f(x, y)$ is constant on the circle $x^2 + y^2 = r^2$.

(ii) Θ and R are independent where Θ is uniformly distributed on the interval $[0, 2\pi]$ and R is a continuous non-negative random variable.

Proof. For any point with Cartesian coordinates (x, y), let (θ, r) denote its corresponding polar coordinates. For $0 \le \theta_1 < \theta_2 \le 2\pi$ and $0 \le r_1 < r_2$, let $\mathcal{A}$ denote the region of the (x, y) plane shown below in Figure 6.

Figure 6. The region $\mathcal{A} \;=\; \{(x, y) : \theta_1 < \theta < \theta_2,\ r_1 < r < r_2\}$.



(i) $\implies$ *(ii):* Since $f(x, y)$ is constant on circles centered at the origin, we may write $f(x, y) = h(\sqrt{x^2 + y^2}) = h(r)$ for some function $h(\cdot)$. Then

$$\{\theta_1 < \Theta < \theta_2\} \cap \{r_1 < R < r_2\} \;=\; \{(X, Y) \in \mathcal{A}\},$$

so

$$\begin{aligned} P[\{\theta_1 < \Theta < \theta_2\} \cap \{r_1 < R < r_2\}] &= P\{(X,Y) \in \mathcal{A}\} \\ &= \iint_{\mathcal{A}} f(x,y)\,dxdy \\ &= \int_{\theta_1}^{\theta_2} \int_{r_1}^{r_2} f(r\cos\theta, r\sin\theta)\, r dr d\theta \\ &= \int_{\theta_1}^{\theta_2} \int_{r_1}^{r_2} h(r)\, r dr d\theta \\ &= (\theta_2 - \theta_1) \int_{r_1}^{r_2} h(r)\, r dr \\ &= \frac{\theta_2 - \theta_1}{2\pi} \cdot \int_{r_1}^{r_2} 2\pi r h(r)\, dr. \end{aligned} \tag{6}$$

Applying this to $\theta_1 = 0$, $\theta_2 = 2\pi$, $r_1 = 0$, $r_2 = \infty$, we get that

$$1 = P[\{0 < \Theta < 2\pi\} \cap \{0 < R < \infty\}] = \int_0^{\infty} 2\pi r h(r)\, dr. \tag{7}$$

Applying (6) now to the case $r_1 = 0$, $r_2 = \infty$, and then using (7) gives

$$P[\theta_1 < \Theta < \theta_2] = P[\{\theta_1 < \Theta < \theta_2\} \cap \{0 < R < \infty\}] = \frac{\theta_2 - \theta_1}{2\pi}, \tag{8}$$

which shows that Θ is uniformly distributed on $[0, 2\pi]$. An additional application of (6) then shows that

$$P[r_1 < R < r_2] = P[\{0 < \Theta < 2\pi\} \cap \{r_1 < R < r_2\}] = \int_{r_1}^{r_2} 2\pi r h(r)\, dr. \tag{9}$$

Combining (6), (8) and (9), we have that

$$P[\{\theta_1 < \Theta < \theta_2\} \cap \{r_1 < R < r_2\}] = P[\theta_1 < \Theta < \theta_2] \cdot P[r_1 < R < r_2],$$

establishing that Θ and R are independent. From (9) we see that R is continuous with density function $2\pi r h(r)$, $r > 0$.

(ii) $\Longrightarrow$ *(i):* This argument basically works in reverse. Suppose $\Theta \sim$ Uniform $(0, 2\pi)$, R is continuous with density function $g(r)$ supported on $r > 0$, and Θ and R are independent. For $r > 0$, put $h(r) = \frac{g(r)}{2\pi r}$ so $g(r) = 2\pi r h(r)$. Then

$$\begin{aligned} P\{(X,Y) \in \mathcal{A}\} &= P[\{\theta_1 < \Theta < \theta_2\} \cap \{r_1 < R < r_2\}] \\ &= \frac{\theta_2 - \theta_1}{2\pi} \cdot \int_{r_1}^{r_2} 2\pi r h(r)\, dr \\ &= \iint_{\mathcal{A}} f(x,y)\, dxdy, \end{aligned} \tag{10}$$

where $f(x,y) = h(\sqrt{x^2+y^2})$. The third equality in (10) follows from a reversal of the steps in (6). Since $\mathcal{A}$ is an arbitrary region of the form shown in Figure 6, it must be that $f(x,y)$ is the joint density function of (X,Y). Clearly $f(x,y)$ is constant on circles about the origin. □

We use this presently to generate independent standard normals (X,Y). In this application we have $f(x,y) = \frac{1}{2\pi}e^{-(x^2+y^2)/2}$, which is clearly constant on circles about the origin. Here, $h(r) = \frac{1}{2\pi}e^{-r^2/2}$ and the proof of our theorem shows that the density function for R is $g(r) = 2\pi r h(r) = re^{-r^2/2}$. We will generate X and Y by first generating $\Theta \sim \text{Uniform}\,(0, 2\pi)$ and then independent $R \sim g$ and putting

$$X \;=\; R\cos(\Theta) \quad \text{and} \quad Y \;=\; R\sin(\Theta). \tag{11}$$

We generate R via the inverse transform method. R has distribution function

$$G(r) \;=\; \int_0^r g(s)\,ds \;=\; 1 - e^{-r^2/2},$$

so $G^{-1}(u) = \sqrt{-2\ln(1-u)}$. Hence, if $U \sim \text{Uniform}\,(0,1)$, $\sqrt{-2\ln(1-U)} \sim g$. As noted previously, $1-U$ is also uniformly distributed on the interval $[0,1]$ so we will take $R = \sqrt{-2\ln(U)}$, where $U \sim \text{Uniform}\,(0,1)$. This straightforward approach, known as the Box–Muller transformation, is summarized as: (i) generate fresh uniforms U, and V; (ii) put $R = \sqrt{-2\ln(U)}$ and $\Theta = 2\pi V$; (iii) compute X and Y as in (11). The computational cost of this procedure involves a logarithm, square root, sine, and cosine — all of which are time-intensive. It is implemented in the program `BoxMullerNormals.cpp`, which generates about 7 million standard normals per second.

§3. The Polar Method

The polar method is less direct but more efficient. Here we first generate a pair (A,B) uniformly distributed on the unit disk $\mathcal{D} = \{(x,y) : x^2+y^2 < 1\}$ by generating a pair (A,B) uniformly distributed on the 2×2 square $\mathcal{S} = \{(x,y) : -1 < x < 1,\ -1 < y < 1\}$ and rejecting the pair if it falls outside $\mathcal{D}$: (i) generate fresh uniforms U and V; (ii) put $A = -1 + 2U$ and $B = -1 + 2V$; (iii) if $S = A^2 + B^2 \geq 1$, go back to (i), otherwise keep A, B, and S. (We retain S as well for later use.)

Since the resulting (A,B) is uniformly distributed on $\mathcal{D}$, the pair's joint density function is the constant $\frac{1}{\pi}$ on $\mathcal{D}$ (using that $\mathcal{D}$ has area π) and zero off $\mathcal{D}$. This density function is clearly constant on circles about the origin (the constant is either $\frac{1}{\pi}$ or 0, according to whether the radius of the circle is less than 1 or not). Letting $(\Theta, \sqrt{S})$ denote (A,B)'s polar coordinates, we find that

Θ and $\sqrt{S}$ are independent with $\Theta \sim \text{Uniform}\,(0, 2\pi)$. For any region $\mathcal{R} \subset \mathcal{D}$, $P[(A, B) \in \mathcal{R}] = \text{Area of}\,(\mathcal{R})/\text{Area of}\,(\mathcal{D})$, so, for $0 < s < 1$,

$$P[S < s] \;=\; P\big[\sqrt{A^2 + B^2} < \sqrt{s}\,\big] \;=\; \frac{\pi(\sqrt{s}\,)^2}{\pi} \;=\; s,$$

i.e., $S \sim \text{Uniform}\,(0, 1)$. To summarize, Θ and S are independent with Θ uniformly distributed on $(0, 2\pi)$ and $S \sim \text{Uniform}\,(0, 1)$. If we take $R = \sqrt{-2\ln(S)}$ and define X and Y as in (11), the pair (X, Y) will be independent standard normals. By construction, $\cos(\Theta) = A/\sqrt{S}$ and $\sin(\Theta) = B/\sqrt{S}$ (draw a picture!), so we take:

$$X \;=\; A\sqrt{\frac{-2\ln(S)}{S}} \quad \text{and} \quad Y \;=\; B\sqrt{\frac{-2\ln(S)}{S}}. \tag{12}$$

The logarithm and square roots are still computed in the polar method, but the trigonometric functions are avoided. The resulting histograms look just like Figure 4. The program `PolarNormals.cpp` implements this algorithm, generating about 15 million standard normals per second.

§4. The Computation of $\Psi(x)$

Here we describe the implementation of the function `Psi`, found in the function library, which approximates the standard normal distribution function $\Psi(x)$. Since $\Psi(0) = \frac{1}{2}$,

$$\begin{aligned}
\Psi(x) \;&=\; \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2}\,dt \\
&=\; \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^x \sum_{n=0}^{\infty} \frac{1}{n!} \left(-\frac{t^2}{2}\right)^n dt \\
&=\; \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n}{n!2^n} \int_0^x t^{2n}\,dt \\
&=\; \frac{1}{2} + \sum_{n=0}^{\infty} \frac{(-1)^n}{n!2^n(2n+1)\sqrt{2\pi}}\, x^{2n+1}.
\end{aligned} \tag{13}$$

For $|x| > 1$, the general term in this summation, call it t_n, is the product of a very small factor (the fraction) and a very large factor (x^{2n+1}) which will lead to computational issues even with double precision. Thus we compute

$$r_n \;=\; \frac{t_n}{t_{n-1}} \;=\; -\frac{2n-1}{2n(2n+1)} x^2,$$

which is a number that is reasonable in size. Then, with $t_0 = x/\sqrt{2\pi}$, we compute $t_n = t_{n-1} \cdot r_n$ for $n = 1, 2, \ldots$, summing as we go. We break out of

the summation when $|t_n| \le 10^{-8}$ and $n > \frac{1}{2}x^2$. The condition $n > \frac{1}{2}x^2$ ensures that $|r_n| < 1$, since $\frac{2n-1}{2n+1} < 1$, and hence $|t_n| < |t_{n-1}|$. Since the series in (13) is alternating the error at break-out is at most $|t_n|$, which is at most 10^{-8}.

This procedure works well for $|x| \le 5$. For extreme values of x we use that, for large positive x, $\Psi(-x) = 1 - \Psi(x) \approx \frac{1}{\sqrt{2\pi}x} e^{-x^2/2}$, yielding

$$\Psi(x) \approx \texttt{Psi}(x) = \begin{cases} \frac{0.967}{\sqrt{2\pi}|x|} e^{-x^2/2} & \text{if } x < -5 \\ (13), \text{ summed to break-out} & \text{if } |x| \le 5 \\ 1 - \frac{0.967}{\sqrt{2\pi}x} e^{-x^2/2} & \text{if } x > 5. \end{cases} \tag{14}$$

The quantity 0.967 is $c(5)$ (see (15) below). This makes the two different case expressions for $\texttt{Psi}(x)$ in (14) equal at $x = -5$ and at $x = 5$. This approximation is accurate to within 10^{-8} across all values of x.

An Asymptotic Estimate. Suppose $X \sim \text{Normal}\,(0,1)$. We close by proving that

$$P[X \ge x] = c(x) \cdot \frac{1}{\sqrt{2\pi}x} e^{-x^2/2}, \tag{15}$$

where $c(x) \to 1$ as $x \to \infty$. The quantity $c(x)$ is a 'plug' whose value is defined by this equality. We have informally stated this as $P[X \ge x] \approx \frac{1}{\sqrt{2\pi}x} e^{-x^2/2}$ and have used this fact as motivation for algorithms approximating Ψ^{-1} and Ψ. First note that for $x > 0$,

$$\begin{aligned} P[X \ge x] &= \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2}\, dt \\ &\le \frac{1}{\sqrt{2\pi}x} \int_x^\infty t e^{-t^2/2}\, dt = \frac{1}{\sqrt{2\pi}x} e^{-x^2/2}. \end{aligned} \tag{16}$$

The inequality holds because $x \le t$ on the range of integration $[x, \infty)$. On the other hand,

$$\begin{aligned} P[X \ge x] &\ge \frac{1}{\sqrt{2\pi}} \int_x^{x+1} e^{-t^2/2}\, dt \\ &\ge \frac{1}{\sqrt{2\pi}} \frac{1}{x+1} \int_x^{x+1} t e^{-t^2/2}\, dt \\ &= \frac{1}{\sqrt{2\pi}x} \frac{x}{x+1} \left[e^{-x^2/2} - e^{-(x+1)^2/2} \right] \\ &= \frac{x}{x+1} \left(1 - e^{-(2x+1)/2} \right) \cdot \frac{1}{\sqrt{2\pi}x} e^{-x^2/2}, \end{aligned} \tag{17}$$

where the second inequality holds since $x+1 \geq t$ on the range of integration $[x, x+1]$. Combining (15), (16) and (17) we see that

$$\begin{aligned} \boxed{\frac{x}{x+1}\left(1-e^{-x-\frac{1}{2}}\right)} \cdot \frac{1}{\sqrt{2\pi x}} e^{-x^2/2} &\leq \boxed{c(x)} \cdot \frac{1}{\sqrt{2\pi x}} e^{-x^2/2} \\ &\leq \boxed{1} \cdot \frac{1}{\sqrt{2\pi x}} e^{-x^2/2}. \end{aligned}$$

The quantity in the box at left, which is less than one, approaches one as $x \to \infty$, proving the estimate since $c(x)$ lies between this quantity and one.

Accompanying Code

`CDHNormals.cpp` implements the four parameter approximation to Ψ^{-1} given in (4). It generates standard normals via inverse transform using this approximation. `PANormals.cpp` implements the Acklam 21 parameter approximation to Ψ^{-1} given in (5). The algorithm is contained in the function `PsiInv`, which is found in the function header file. `BoxMullerNormals.cpp` uses the Box-Muller method for generating standard normals and `PolarNormals.cpp` generates standard normals via the polar method. All of these programs generate a histogram of the standard normals and time the execution.

`CDHCoefficients.cpp` calculates the coefficients A, B, C, and D in (2).