Sampling Random Variables

Bo Li, Math, UCSD, Spring 2019

Let \((\Omega, \mathcal{A}, \mathbb{P})\) be a probability space. Let \(X: \mathbb{R} \to \mathbb{R}\) be a random variable (RV), with probability density function (PDF) \(f_X: \mathbb{R} \to [0, \infty)\). Sampling \(X\) with the given density \(f_X\) means to produce \(X_1, X_2, \ldots\) i.i.d. (independent identically distributed) RVs with the common density/distribution \(f_X\).

Often, a starting point is to generate \(U \sim U[0,1]\) i.e., RV \(U\) that is uniformly distributed on \([0,1]\). This is a random number \(U \sim U[0,1]\). Then, using some method to generate/produce the needed RVs \(X_1, X_2, \ldots\) i.i.d. with the given/target density \(f_X\). \(U \sim U[0,1] \Rightarrow X \sim f_X\).

Notation. \(X \sim f\) means \(X\) is a RV distributed according to \(f\), or with the PDF \(f\). E.g., \(U \sim U[0,1]\), \(U\) is uniformly distributed on \([0,1]\). \(Z \sim N(0,1)\): \(Z\) has the standard normal or Gaussian distribution.

If \(F = F_X(x)\) is the distribution (or cumulative distribution function) of a RV, then \(X \sim F\) means the RV \(X\) has the CDF \(F\), or \(X\) is \(F\)-distributed.
Example 1: Sample a RV $X \sim \text{Bernoulli}(\frac{1}{3})$

Define $h : [0, 1] \rightarrow \{0, 1\}$ by

$$h(u) = \begin{cases} 
0 & \text{if } 0 \leq u \leq \frac{1}{3} \\
1 & \text{if } \frac{2}{3} < u \leq 1 
\end{cases}$$

If $U(1) \sim \text{Uniform}[0, 1]$ then $h(U) \sim \text{Bernoulli}(\frac{1}{3})$. 

**Proof.** \( \forall w \in [0, 1) \)

$$P(U(w) \leq \frac{1}{3}) = \frac{1}{3}$$

So we can define

$$h(U(w)) = \begin{cases} 
0 & \text{if } U(w) \in [0, \frac{1}{3}], \\
1 & \text{if } U(w) \in (\frac{2}{3}, 1]. 
\end{cases}$$

So

$$P(h(U) = 0) = P(0 \leq U < \frac{2}{3}) = F_U(\frac{2}{3}) - F_U(0) = \frac{2}{3}$$

$$P(h(U) = 1) = P(\frac{2}{3} < U \leq 1) = F_U(1) - F_U(\frac{2}{3}) = 1 - \frac{2}{3} = \frac{1}{3}$$

So

$$P(h(U) = 0) = 0$$

Hence, $h(U) \sim \text{Bernoulli}(\frac{1}{3})$.

Example 2: Sample $X \sim \text{Exp}(7)$. Recall the PDF for $	ext{Exp}(7)$ [exponential distribution with $\lambda = 7$] is given by

$$f(x) = \begin{cases} 
0 & \text{if } x < 0, \\
7e^{-7x} & \text{if } x \geq 0. 
\end{cases}$$

Define $h : [0, 1] \rightarrow [0, \infty)$ by $h(u) = -\frac{1}{7} \log u$ (log: natural logarithm).

Let $U(1) \sim \text{Uniform}[0, 1]$. We show that $h(U) \sim \text{Exp}(7)$.

We compute the PDF (prob. distribution function) of the RV $Y = h(U)$.

Since $h(U) \equiv 0 \forall u \in [0, 1]$, we have

$$P(h(U) \leq 0) = 0$$

i.e. $F_{h(U)}(0) = 0$.

Let $x > 0$ we have

$$F_{h(U)}(x) = F_{h(U)}(0) = 0$$

So $F_{h(U)}(x) = 0$ if $x \leq 0$.

Note: We should define

$$x = \begin{cases} 
h(U) & \text{if } U \in [0, 1), \\
0 & \text{if } U \not\in [0, 1]. 
\end{cases}$$

So $F_x(x) = 0$ if $x \leq 0$. 

<Refer to the image>
Let \( x > 0 \). We have
\[
P(h(U) \leq x) = P \left( \log(U) \geq -7x \right)
= P(U \geq e^{-7x})
= P(U \geq e^{-7x}) + P(U < e^{-7x}) - P(U < e^{-7x})
= 1 - P(U < e^{-7x})
= 1 - F_U(e^{-7x})
= 1 - e^{-7x} \text{ (since } e^{-7x} \text{ is strictly increasing)}
\]

i.e.,
\[
F_U(U)(x) = \begin{cases} 
1 - e^{-7x} & \text{if } x \leq 0, \\
-x^{-7} & \text{if } x > 0.
\end{cases}
\]
Together
\[
F_U(U)(x) = \begin{cases} 
0 & \text{if } x \leq 0, \\
-x^{-7} & \text{if } x > 0.
\end{cases}
\]

The Inversion Method

**Lemma** Let \( F : \mathbb{R} \rightarrow [0,1] \) be the distribution of an RV \( X \rightarrow \mathbb{R} \). If \( U \sim U(0,1) \), then
\[
X = F^{-1}(U) \sim F.
\]

Note: We should really define
\[
X = \begin{cases} 
F^{-1}(U) & \text{if } F \text{ strictly increasing}, \\
\text{constant} & \text{otherwise}.
\end{cases}
\]

Here \( F^{-1} \) is the generalized inverse of \( F \), defined as
\[
F^{-1}(u) = \inf \{ z \in \mathbb{R} : F(z) \geq u \} \text{ for } 0 \leq u \leq 1.
\]

If \( F \) is strictly increasing, then this is the same as the inverse function of \( F \).

Prove this fact. You need the properties of the CDF \( F \).

**Proof of Lemma** Let \( z \in \mathbb{R} \).
\[
P(X = \infty) = P(U < 0)
= P(U < 0) \quad \text{Use the right-continuity of } F
= P(U > 1)
= P(U > 1) \quad \text{by definition of } F
= 0.
\]

\[
P(U < 0) = P(F^{-1}(U) \leq z) \quad \text{by definition of } F^{-1}
\]

Since \( F(z) \in [0,1] \).

**Proof of (x)** Since \( P(U < 0) = 0 \), \( P(U > 1) = 0 \). We consider only \( 0 < U < 1 \) and show
\[
F^{-1}(U) \leq z \iff U \leq F(z)
\]
for any \( z \in \mathbb{R} \). This follows from the def. \( F^{-1} \).
Fix $z \in \mathbb{R}$ and assume $F^{-1}(U) \leq z$. By the def. of $F^{-1}(U)$, let $z_n \downarrow z$ s.t. $F^{-1}(U) = \lim_{n \to \infty} z_n$, and $F(z_n) \geq U$. Note that $z_n$ is bounded below, for otherwise $z_n \to -\infty$ and $F(z_n) \to 0$ contradiction $F(z_n) \geq U > 0$ ($n=1,2,\ldots$). So, $z_n \downarrow z'$ for some $z' \in \mathbb{R}$. But all $z_n \leq F^{-1}(U)$. So, $z \leq F^{-1}(U) \leq z$. Moreover, the right-continuity implies $F(z') = \lim_{n \to \infty} F(z_n) \geq U$. Hence, the monotonicity of $F$ implies that $U \leq F(z') \leq F(z)$.

Example Sample a RV that is $\text{Exp}(\lambda)$ distributed for some $\lambda > 0$. The CDF for $\text{Exp}(\lambda)$ is $F(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1 - e^{-\lambda x} & \text{if } x \geq 0. \end{cases}$ Let $U \sim U(0,1)$. Since $P(U \in (0,1)) = 0$, let's just assume $U \in (0,1)$. Let $X = F^{-1}(U)$.

$$x = F^{-1}(U), \quad U \in (0,1) \implies F(x) = U \in (0,1).$$

$$1 - e^{-\lambda x} = U \implies x = -\frac{1}{\lambda} \log(1-U).$$

So, $X = -\frac{1}{\lambda} \log(1-U)$. Check that $F_X = F$.

So $X \sim \text{Exp}(\lambda)$.

Example Sample a RV with density $f(r) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-\frac{r^2}{2}} & \text{if } r \geq 0, \\ 0 & \text{if } r < 0. \end{cases}$ The distribution is $F(r) = \int_0^r f(u) du = 1 - e^{-\frac{r^2}{2}}$ if $r \geq 0$, and $F(r) = 0$ if $r < 0$. So, we have $F(r) = U \in (0,1)$.

$$\Rightarrow r = \sqrt{-2 \log(1-U)}.$$ 

So, if $U \sim U(0,1)$. Then $R = \sqrt{-2 \log(1-U)}$ is a RV with the density $f$.

[See next page for the discrete inversion method.]

**Numerical Approximation** — if an explicit formula of $F^{-1}$ is not available.

Sample a RV with CDF $F$.

**Step 1** Find $a,b \in \mathbb{R}$, $a < b$, and set $x_n = 2^k < x_1 < \ldots < x_n = b$.

So that $F(x_i) \leq 1 - F(b) < x_i - F(x_{i+1}) < F(x_{i+1}) - F(x_i) < 1$ for all $i$. 
Step 2. Generate $U \sim U[0,1].$

Step 3. Find $i$ s.t. $F(2i) \leq U \leq F(2i+1)

(Need to use some search algorithm)

Step 4. Define

\[
X = 2i + \frac{U - F(2i)}{F(2i+1) - F(2i)} \cdot F(2i+1) - F(2i).
\]

Then, approximately, $X$ has the distribution function $F.$

Why? By the inversion method, $F(X) = U.$ But, we don't have a formula for $F^{-1}.$ So, we solve $F(2i) = U$ numerically to find $z,$ then use this $z$ as our $X.$

\[
U = F(2i + (z-i)(2-2i)) \quad \Rightarrow \quad z = 2i + \frac{U - F(2i)}{F'(2i)}.
\]

But $F'(2i) = \frac{F(2i+1) - F(2i)}{2i+1 - 2i}.$

Hence

\[
z = 2i + \frac{U - F(2i)}{F(2i+1) - F(2i)}.
\]

Replace $z$ by $X$ and $i$ by $x.$
The Discrete Inversion Method (Or the Inversion Method for Discrete RVs)

Let \( X : \mathbb{N} \rightarrow \mathbb{R} \) be a discrete RV taking values \( x_1 < x_2 < \ldots \) with probabilities \( p_1, p_2, \ldots \), where \( p_k = P(X = x_k) > 0 \) and \( \sum_{k=1}^{\infty} p_k = 1 \).

The PMF (probability mass function), or the discrete distribution (function) \( F = F_X : \mathbb{R} \rightarrow [0,1] \) is a step function:

\[
F(x) = \begin{cases} 
0 & \text{if } x < x_1 \\
p_1 & \text{if } x_1 \leq x < x_2 \\
p_1 + p_2 & \text{if } x_2 \leq x < x_3 \\
p_1 + p_2 + \ldots + p_k & \text{if } x_k \leq x < x_{k+1} \\
1 & \text{if } x \geq x_k 
\end{cases}
\]

If \( U \sim U[0,1] \), then, for each \( x_k \), we find the smallest integer \( k \geq 1 \) such that \( F(x_k) > U \), and return \( X = x_k \).

The \( X \) is a discrete RV with \( X \sim F \).

This algorithm involves searching, which is in general \( O(M) \) complexity if it is a \( M \)-state RV. With a binary search, it is \( O(\log M) \).
The Transformation Method

More general than the inversion method.

Note: \( x = \alpha(x) \)

may not itself be a distribution.

Let \( X: \mathbb{R} \to \mathbb{R} \) be a RV with the distribution \( F_X: \mathbb{R} \to [0, 1] \) and the density \( f_X: \mathbb{R} \to [0, \infty) \).

Let \( \alpha: \mathbb{R} \to \mathbb{R} \) be a \( C^1 \) function. Assume \( \alpha'(x) > 0 \) \( \forall x \in \mathbb{R} \), i.e., \( \alpha \) is increasing.

Define \( Y = \alpha(X) \). Then \( Y: \mathbb{R} \to \mathbb{R} \) is also a RV. We would like to find the distribution and/or density for \( Y \).

Let \( y = \alpha(x) \) (\( x \in \mathbb{R} \)). Then

\[
F_Y(y) = P(Y \leq y) = P(\alpha(X) \leq \alpha(x)) = P(X \leq x) = F_X(x)
\]

\[
F_Y'(y) = F_Y'(y) = F_X'(x)
\]

So,

\[
\frac{d}{dx}[F_Y(y(x))] = F_X'(x) = f_X(x)
\]

\[
F_Y'(y) \cdot y'(x) = f_X(x)
\]

\[
F_Y(y) \frac{dx}{dy} = f_X(x)
\]

\[
f_Y(y) = f_X(x) \left( \frac{dx}{dy} \right)^{-1}, \quad y = x = \alpha'(y)
\]

Let \( \beta \in C^1(\mathbb{R}) \) be \( \beta \) new \( \beta(x_1) < 0 \) \( \forall x_1 \in \mathbb{R} \).

\[\text{strictly} \]

i.e., \[\beta'(x_1) < 0 \] \( \forall x_1 \in \mathbb{R} \).
Consider now \( Y = \beta(X) \).

What is \( f_Y \)?

Let \( y = \beta(x) \). \( F_Y(y) = P(Y \leq y) = P(\beta(X) \leq y) \)

\[ = P(X \geq x) = 1 - P(X < x) \]

\[ = 1 - F_X(x) \]

\[ \frac{d}{dx}: \quad f_Y(y) \frac{d\beta}{dx} = -f_X(x) \]

\[ f_Y(y) = f_X(x) \left| \frac{d\beta}{dx} \right|^{-1}, \quad y = \beta^{-1}(x) \]

For both cases \( |f_X(x)\,dx| = |f_Y(y)\,dy| \).

**Example** \( X \sim N(0,1) \) \( f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \)

Let \( d(x) = 0x + \mu \), for some \( \sigma > 0 \) and \( \mu \in \mathbb{R} \)

Let \( Y = d(X) \). \( f_Y(y) = f_X(x) \left| \frac{dy}{dx} \right|^{-1} = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \cdot \frac{1}{\sigma} \cdot y = \beta(x) \)

But \( y = d(x) = 0x + \mu \). So \( x = \frac{y - \mu}{\sigma} \). Hence,

\[ f_Y(y) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(y - \mu)^2}{2\sigma^2}} \]

And, \( Y \sim N(\mu, \sigma^2) \).

**Example** \( X \sim U[0,1] \) \( d(x) = -\log x, \quad x \in (0,1] \)

Let \( Y = d(X) = -\log X \). For \( y = d(x), \ x \in (0,1] \)

\[ f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right|^{-1} = \left| -\frac{1}{x} f_X(x) \right| = x f_X(x) = xf_X(x) \]
If \( x \in (0, 1) \) then \( y = x \log x \in (0, \infty) \)

Also, \( X \sim U[0,1] \). So \( F_X(x) = 1 \) for \( x \in (0, 1) \)

Hence \( f_Y(y) = y = e^{-y} \) for \( y \in (0, \infty) \)

Note that \( P(x \leq 0) = F_X(0) = 0 \)

\( P(x \geq 1) = 0 \) (since \( F_X(1) = 1 = 0 \))

So, on the subset \( \{ X \leq 0 \} \in \sqrt{y} \), we can define \( Y = 0 \). So.

\[
Y = \begin{cases} 
0 & \text{on } \{ X \leq 0 \} \\
-\log X & \text{on } \{ X > 0 \} \end{cases}
\]

Thus, \( f_Y(y) = P(Y \leq y) = 0 \) if \( y \leq 0 \).

For \( y > 0 \), \( f_Y(y) = \frac{d}{dy} f_Y(y) = e^{-y} \)

Hence, \( F_Y(y) = \int_{-\infty}^{y} f_Y(u) \, du 
= \int_{0}^{y} e^{-u} \, du = 1 - e^{-y} \)

Summary. \( F_Y(y) = \begin{cases} 
0 & \text{if } y < 0 \\
1 - e^{-y} & \text{if } y \geq 0 \end{cases} \)

This is the exponential distribution with \( \lambda = 1 \).

In what follows, if \( U \sim U[0,1] \), i.e., \( U \sim U \mathbb{R} \)

is a RV \( U[0,1] \)-distributed. Then we can assume \( 0 \leq U \leq 1 \). Since \( P(U \in [0,1]) = 1 \).

In computer, a RND (a random number)
generated in \([0,1] \) is always in \([0,1] \).
The transformation method for sampling multi-variate (multiple RVs).

Let $X_1, \ldots, X_n : \mathcal{X} \to \mathbb{R}$ be $n$ RVs. Then, $X = (X_1, \ldots, X_n) : \mathcal{X} \to \mathbb{R}^n$ is a vector-valued RV.

The joint distribution function is $F_X : \mathbb{R}^n \to [0,1]$. 

$$F_X(x) = P \left( \bigcap_{i=1}^{n} (X_i \leq x_i) \right) \quad \forall x = (x_1, \ldots, x_n) \in \mathbb{R}^n$$

The joint density (assumed to exist) is denoted 

$$f_X(x) = f_{X_1, \ldots, X_n} (x_1, \ldots, x_n) : \mathbb{R}^n \to \mathbb{R}_+$$

It is related to $F_X$ by

$$F_X(x) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} f(u_1, \ldots, u_n) \, du_1 \ldots du_n$$

and invertible

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a C-map with the Jacobian for $T^{-1}$: \text{Range}(T) \to \mathbb{R}^n$

$$J(y) = \det \left( \frac{\partial x_i}{\partial y_j} \right)$$

non-singular at any $y \in \text{Range}(T)$, where

$T(x_1, \ldots, x_n) = (y_1, \ldots, y_n)$,

$(X_1, \ldots, X_n) = T^{-1}(y_1, \ldots, y_n)$,

$x_i = x_i(y_1, \ldots, y_n)$ (i = 1, \ldots, n).

Define $Y = T(X_1, \ldots, X_n) : \mathcal{X} \to \mathbb{R}^n$. What is the joint density for $Y$?
Theorem. Under the above assumptions, we have for $Y = T(X_1, \ldots, X_n) = (Y_1, \ldots, Y_n)$:

$$f_Y(y) = \begin{cases} f_X(T^{-1}(y)) \left| \frac{\partial T}{\partial y} \right| & \text{if } y \in \text{Range}(T), \\ 0 & \text{otherwise}. \end{cases}$$

**PF.** Assume Range$(T) = \mathbb{R}^n$. Write $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n)$. $\forall B \subseteq \mathbb{R}^n$ a Borel set. [It is enough to assume $B$ is a box, $B = \prod_{i=1}^n [a_i, b_i]$.]

We have

$$P(X \in B) = \int_{T^{-1}(B)} f_X(x) \, dx = \int_{T(B)} f_X(T(y)) \left| \frac{\partial T}{\partial y} \right| \, dy$$

Since $Y = T(X)$, $P(Y \in T(B)) = \int_{T(B)} f_Y(y) \, dy$

Hence $f_Y(y) = f_X(T^{-1}(y)) \left| \frac{\partial T}{\partial y} \right|$. \[ □ \]

The result is $f_Y(y) \, dy = f_X(x) \, dx. \quad \frac{dx}{dy} = \left| \frac{\partial T}{\partial y} \right|$

The Box-Muller Method for Sampling Two Independent Gaussian RVs

and Standard Indep.

Let $X, Y : \mathbb{R} \to \mathbb{R}$ be two standard Gaussian variables.

1. Independent: $f_{X,Y}(x,y) = f_X(x) f_Y(y)$
2. $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$

\[ \text{i.e. } X \sim N(0,1), ~ Y \sim N(0,1) \]

We would like to sample $X, Y$ by generating $z, \eta \sim U([0,1])$ and use the transformation method.
We implement this using two steps.

**Step 1. Change variables** $\mathbf{(x, y)}$ to $\mathbf{(r, \theta)}$

\[
\begin{align*}
x &= r \cos \theta & x &= R \cos \theta \\
y &= r \sin \theta & y &= R \sin \theta \\
\end{align*}
\]

Generate RVs $R$ and $\theta$ that correspondingly to $X, Y$.

Find $f_R, f_\theta$.

**Step 2** Generate $R, \theta \sim U[0, 1]$. And then using the transformation method again to generate $R$ and $\theta$ or the inversion method.

Let's begin with

\[
\int f_{x, y}(x, y) \, dx \, dy = f_{R, \theta}(r, \theta) \, dr \, d\theta.
\]

Independence:

\[
\int f_x(x) f_y(y) \, dx \, dy = f_R(r) f_\theta(\theta) \, dr \, d\theta.
\]

But, the left-hand side is

\[
\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \, dx \, dy
\]

\[
x = r \cos \theta = \frac{1}{\sqrt{2\pi}} e^{-\frac{r^2}{2}} (x^2 + y^2) \, dx \, dy
\]

\[
y = r \sin \theta = \frac{1}{\sqrt{2\pi}} e^{-\frac{r^2}{2}} r \, dx \, dy
\]

Hence, we choose

\[
\begin{align*}
f_R(r) &= r e^{-\frac{r^2}{2}} \quad (r \geq 0) \\
\frac{1}{2\pi} \quad &\text{so that} \quad x = R \cos \theta \\
\frac{1}{2\pi} \quad &\text{so that} \quad y = R \sin \theta \\
\end{align*}
\]

We can define $f_R(r) = 0$ if $r < 0$ and $f_\theta(\theta) = 0$ if $\theta \notin [0, 2\pi)$.

Now, (step 2) we generate $R, \theta$ with the densities $f_R$ and $f_\theta$, respectively, by the inversion method.
By the second example on page 4 of this set of notes, we have
\[ R = \sqrt{-2 \log(1-\xi)} \quad \text{if } \xi \sim U[0,1] \]

For sampling \( \Theta \) note first from \( f_\Theta(\theta) = \frac{1}{2\pi} \) only for \( \theta \in (0, 2\pi) \), we have
\[ F_\Theta(\theta) = \left\{ \begin{array}{ll}
0 & \text{if } \theta \leq 0 \\
\frac{\theta}{2\pi} & \text{if } 0 < \theta \leq 2\pi \\
1 & \text{if } \theta > 2\pi.
\end{array} \right. \]

Let \( \eta \sim U[0,1] \).

Solve: \( F_\Theta(\theta) = \frac{\theta}{2\pi} \text{ in } 0 < \theta < 2\pi \).

Hence, \( \Theta = 2\pi \eta \), \( \theta = 2\pi \eta \in (0, 2\pi) \)

Now, we have the Box-Muller method
\[ X = \sqrt{-2 \log(1-\xi)} \cos(2\pi \eta) \quad \xi, \eta \sim U[0,1] \]
\[ Y = \sqrt{-2 \log(1-\xi)} \sin(2\pi \eta) \quad \text{independent} \]

Let us verify that \( X, Y \sim N(0,1) \), independent.

Since \( \xi, \eta \sim U[0,1] \) independent,
\[ R = \sqrt{-2 \log(1-\xi)} \]
\[ \Theta = 2\pi \eta \]
are independent RVs. So, \( f_{X,Y} = f_X \cdot f_Y \)

\[ F_\Theta(\theta) = \frac{\theta}{2\pi} \quad \text{if } 0 \leq \theta \leq 2\pi \\
0 \quad \text{elsewhere if } \theta > 2\pi. \]

So, \( f_{\Theta}(\theta) = \left\{ \begin{array}{ll}
\frac{1}{2\pi} & \text{if } \theta \in [0,2\pi] \\
0 & \text{elsewhere.}
\end{array} \right. \)
Similarly,
\[ F_R(r) = P(R \leq r) = P(\xi \geq e^{-\frac{r^2}{2}}) \]
\[ = 1 - P(\xi < e^{-\frac{r^2}{2}}) \]
\[ = 1 - F_\xi(e^{-\frac{r^2}{2}}) \]
\[ = 1 - e^{-\frac{r^2}{2}} \quad \text{since } \xi \sim \mathcal{N}(0,1) \]

\[ F_R(r) = P(R \leq r) = 0. \]

So,
\[ f_R(r) = \begin{cases} 0 & \text{if } r < 0 \\ F_R'(r) & \text{if } r \geq 0 \end{cases} = \begin{cases} 0 & \text{if } r < 0 \\ re^{-\frac{r^2}{2}} & \text{if } r \geq 0. \end{cases} \]

Combine:
\[ f_{R@}(r,\theta) = f_R(r) f_\theta(\theta) = \begin{cases} \frac{1}{2\pi} re^{-\frac{r^2}{2}} & \text{if } r \geq 0, \theta \in [0,2\pi] \\ 0 & \text{elsewhere}. \end{cases} \]

Now, by the theorem on the transformation method for multivariate, we have for the new RVs \( X = R \cos \Theta \), \[ \begin{align*}
X = R \cos \Theta, & \quad \text{following the transform.} \\
y = R \sin \Theta & \quad \text{with } x = r \cos \alpha, y = r \sin \alpha.
\end{align*} \]

that
\[ f_{x,y}(x,y) = f_{R@}(r,\theta) \left| J(x,y) \right| \]
\[ \left| J(x,y) \right| = \left| \frac{\partial (r,\theta)}{\partial (x,y)} \right| = \left| \det \begin{pmatrix} \frac{\partial r}{\partial x} & \frac{\partial r}{\partial y} \\ \frac{\partial \theta}{\partial x} & \frac{\partial \theta}{\partial y} \end{pmatrix} \right| = \frac{1}{r} \]

for \( r \geq 0 \). Hence
\[ f_{x,y}(x,y) = f_{R@}(r,\theta) \frac{1}{r} \]
\[ = \frac{1}{2\pi} e^{-\frac{r^2}{2}} \cdot \frac{1}{r} \]
\[ = \frac{1}{2\pi} e^{-\frac{r^2}{2}} (x^2 + y^2) \]

The marginal distribution is
\[ f_X(x) = \int f_{x,y}(x,y) \, dy = \int_0^{2\pi} \frac{1}{2\pi} e^{-\frac{r^2}{2}} \, dy \]
\[ = \frac{1}{2\pi} \int_0^{2\pi} e^{-\frac{x^2}{2}} \, dy = \left[ \frac{1}{2\pi} \int_0^{2\pi} e^{-\frac{x^2}{2}} \, dy \right] = 1. \]
Hence \( f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \)

Similarly \( f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} \)

Hence, \( X, Y \sim N(0, 1) \). Moreover
\[
f_{X,Y}(x, y) = f_X(x) f_Y(y)
\]
So, \( X \) and \( Y \) are independent.

Note. To generate a single RV \( X \sim N(0, 1) \), one can use the CLT (Central Limit Theorem)
(1) Generate \( \delta_1, \delta_2, \ldots, \delta_N \) \( \text{i.i.d.} \ U[0,1] \)-distributed
(2) Set \( X_N = \sqrt{\frac{N}{2}} \left( \sum_{k=1}^{N} \delta_k - \frac{N}{2} \right) \)
Then, Approximately \( X_N \sim N(0, 1) \). If \( N \gg 1 \).

**Accept-Reject (or Acceptance-Rejection)**

The Acceptance-Rejection Method
(Due to John von Neumann, early 1950s.) necessary

This is a direct sampling method, without using some other distributions (e.g., \( U[0,1] \)). So, it is different from the inversion, or more generally, the transformation method.

**Basic idea.** Let \( f : \mathbb{R} \to [0, \infty) \) be the pdf of a RV from \( X \) to \( \mathbb{R} \). Assume \( f = 0 \) outside \([a, b]\), a finite interval \( (a, b \in \mathbb{R}, a < b) \). We would like to sample a RV with this pdf. (The assumption \( f = 0 \) outside \([a, b]\) can be relaxed, to get approximate results.)
Let $c \in \mathbb{R}$ be an upper bound of $f$, e.g.,

\[ c = \sup \{ f(x) : x \in [a, b] \}. \]

The algorithm for generating $Z \sim f$ is:

**Step 1** Generate $X \sim U(a, b)$

**Step 2** e.g., generate $U \sim U(0, 1)$.

$X \leftarrow (b-a) U + a$

**Step 2** Generate $Y \sim U(0, c)$, independent of $X$.

**Step 3** If $Y \leq f(X)$, return $Z = X$.

Otherwise, reject $X$ and return to Step 1.

Note. Here, we still use a reference distribution $U([0, 1])$.

Why the algorithm generates $Z \sim f$? The reason is as follows. $(X, Y)$ is uniformly distributed, and the accepted $X$ satisfies $Y \leq f(X)$. Thus, the marginal distribution for the accepted $X$ is $f$. Thus, the marginal $f_X(x) = \int f(x, y) dy = f(x)$. 

A more general acceptance-rejection method is based on the following idea:

Let \( f : \mathbb{R}^n \rightarrow (0, \infty) \) be the PDF of a RV in \( \mathbb{R}^n \). Let \( g : \mathbb{R}^n \rightarrow (0, \infty) \) be the PDF of a RV in \( \mathbb{R}^n \), and \( \alpha \geq 1 \) be such that

1. We know how to sample RVs with \( g \) the PDF.
2. \( \alpha g(x) \geq f(x) \) for all \( x \).

But the region between the surfaces \( y = \alpha g(x) \) and \( y = f(x) \) is as small as possible. So, \( \alpha = \sup_x \frac{f(x)}{g(x)} \) is a good candidate.

Call \( \alpha g(x) \) a majorizing function of \( f(x) \), and \( g(x) \) a proposal PDF.

**Algorithm: Acceptance-Rejection Method**

1. **Step 1** Generate \( X \sim g \).
2. **Step 2** Generate \( Y \sim U[0, \alpha g(X)] \).
3. **Step 3** If \( Y \leq f(X) \) then accept \( Z = X \), otherwise reject. Return to step 1.

**Theorem** The algorithm generates \( Z \sim f \).
Proof. Let \( A = \{ (x, y) : 0 \leq y \leq g(x) \} \subseteq \mathbb{R}^n \times \mathbb{R}^m \),
\( B = \{ (x, y) : 0 \leq y \leq f(x) \} \subseteq \mathbb{R}^n \times \mathbb{R}^m \).

Note that the volume of \( A \) is \( |A| = \alpha \), since \( g \) is a PDF, and the volume of \( B \) is \( |B| = 1 \), since \( f \) is a PDF.

From Steps 1 and 2, \((X, Y)\) is uniformly distributed on \( A \). To see this, let \( \varrho(x, y) \) denote the joint PDF of \((X, Y)\), and let \( \vartheta(y|x) \) denote the conditional PDF of \( Y \) given \( X = x \). Then,
\[
\varrho(x, y) = \begin{cases} g(x) \vartheta(y|x) & \text{if } (x, y) \in A, \\ 0 & \text{otherwise}. \end{cases}
\]

By step 2 implies that \( \vartheta(y|x) = \frac{1}{\alpha g(x)} \) for \( y \in [0, g(x)] \), and \( \vartheta(y|x) = 0 \) elsewhere. Thus,
\[
\varrho(x, y) = \frac{1}{\alpha} \quad \text{for every } (x, y) \in A.
\]

Let \((X^*, Y^*)\) be the first accepted point, i.e., the first point in \( B \). Since \((X, Y)\) is uniformly distributed on \( A \), \((X^*, Y^*)\) is uniformly distributed on \( B \). But, the volume \( |B| = 1 \). So, this joint PDF of \((X^*, Y^*)\) is \( 1 \) Thus, the marginal PDF of \( Z = X^* \) is
\[
\int_{0}^{\infty} \frac{1}{\alpha} \, dy = f(x). \quad \square
\]

Note that the efficiency of this algorithm is defined as
\[
P\left((X, Y)\text{'s accepted}\right) = \frac{\alpha \text{vol}(B)}{\text{vol}(A)} = \frac{\alpha}{\alpha} = 1.
\]

Also, in generating many RVs \( X \) we introduce a similar concept:
acceptance rate = \# of \( X \) accepted \# of \( X \) generated (= efficiency)
Often a slightly modified version of the above algorithm is used.

\[ X \sim U[0, \alpha g(x)] \text{ is same as setting } \]
\[ Y = U | \alpha g(x), \text{ where } U \sim U[0,1]. \]

Then, \( Y \leq f(x) \) is equivalent to \( U \leq \frac{f(x)}{\alpha g(x)}. \)

Hence, the modification is:

Generate \( X \) from \( g(x) \) and accept it with probability \( \frac{f(x)}{\alpha g(x)} \).

**Algorithm: Modified Accept-Reject Method**

1. Sample a RV \( Z \sim f \).
2. Let \( f \) be a given PDF of RV \( \sim \mathbb{R}^n \).
3. Let \( g \) be a PDF of RV \( \sim \mathbb{R}^n \) and \( \alpha \geq 1 \).
4. Suppose \( f(x) \leq \alpha g(x) \forall x \in \mathbb{R}^n \) (Common R^n cube replaced by a finite region, the support of \( f \), if \( f \) is finitely supported).

**Steps**

1. **Step 1** Generate \( X \) from \( g \).
2. **Step 2** Generate \( U \sim U[0,1] \) independent of \( X \).
3. **Step 3** If \( U \leq \frac{f(x)}{\alpha g(x)} \) then \( Z \leftarrow X \).
   Otherwise, reject \( X \), and go to Step 1.

**Example**

Let \( f(x) = \frac{1}{2\pi R} \sqrt{R^2 - x^2}, -12 \leq x \leq 12 \), for some \( R > 0 \). \( f \geq 0 \) and \( \int_{-12}^{12} f(x) \, dx = 1 \). \( f \) is a PDF.

2. Take the proposed distribution \( g(x) = \frac{1}{2R} \) for \( x \in [-R, R] \). Choose \( \alpha = \text{const. small}, \text{ s.t. } \alpha g(x) \geq f(x), x \in [-12, 12] \).

The smallest such \( \alpha \) is \( \alpha = \frac{1}{16R}. \)
We sample a RV \( Z \) with distribution \( f \) as follows:

1. Generate independent \( U_1, U_2 \sim U[0, 1] \).
2. Use \( U_2 \) to generate \( X \) via the inversion method: 
   \[ X = (2U_2 - 1)R \]
3. Calculate 
   \[ f(X) = \sqrt{1 - (2U_2 - 1)^2} \]
4. If \( U_1 \leq f(X) / \sqrt{g(X)} \), i.e., 
   \[ (2U_2 - 1)^2 \leq 1 - U_1^2 \]
   or 
   \[ U_1^2 + 4U_2^2 - 4U_1 \leq 0 \]
   then accept \( X \) and return \( Z = X = (2U_2 - 1)R \).
   Otherwise, reject \( X \) and go to 1.

The expected number of trials is \( \alpha = \frac{4}{\pi} \) and the efficiency is \( \frac{1}{\alpha} = \frac{\pi}{4} \approx 0.785 \).

Another variation of the above algorithm is the so-called "squeeze accept-reject method"

Let \( f \) be a PDF.

Let \( g, h \) also be PDFs that are relatively easy to sample. (e.g., \( h(x) \) can be a piecewise linear function)

Assume 
\[ h(x) \leq f(x) \leq \alpha g(x) \quad \forall x. \]

**Algorithm**

1. **Step 1** Generate \( X \sim g(x), U \sim U[0, 1] \).
2. **Step 2** Accept \( X \) if \( U \leq \frac{f(x)}{\alpha g(x)} \).
3. **Step 3** Otherwise, accept \( X \) if \( U \leq f(x) / \alpha g(x) \).
Example. Sample $X \sim N(0, 1)$ by the accept-reject method. The PDF is $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \ (x \in \mathbb{R})$.

First, we choose a proposal PDF:

$$g(x) = \frac{1}{2} e^{-\frac{|x|}{2}} \ (x \in \mathbb{R}).$$

Observe that we can describe $g$ as the distribution of an exponential $RV \ (\text{Exp}(λ))$, with $λ = 1$ multiplied randomly by $±1$. Thus, we can generalize a $RV \ V \sim g$ as follows:

1. Generate $U_1, U_2 \sim \text{U}[0, 1]$, independent.
2. If $U_1 < \frac{1}{2}$ then set $V = -\log U_2$.
   Otherwise, set $V = \log U_2$.

Now, set

$$\lambda = \sup_{x \in \mathbb{R}} \frac{f(x)}{g(x)} = \sup_{x \geq 0} \frac{\frac{2}{\sqrt{\pi}} e^{-\frac{x^2}{2}}}{\frac{1}{2} e^{-\frac{x}{2}}} = \sqrt{\frac{2}{\pi}} \approx 1.33.$$ 

So, $\lambda g(x)$ is a majorizing function of $f(x)$, i.e.,

$$0 \leq f(x) \leq \lambda g(x) \ (x \in \mathbb{R}).$$

The accept-reject method for generating $X \sim N(0, 1)$

Step 1. Generate $V \sim g$ using $U_1, U_2 \sim \text{U}[0, 1]$ independent, as in (1), (2) above.

Step 2. General $X \sim \text{U}[\log(V), 1]$ by generating $U_3 \sim \text{U}[0, 1]$ and setting $Y = \lambda g(V) U_3 = \frac{1}{2} U_3 e^{-V} \ (\iff U_3 \leq \frac{f(V)}{\lambda g(V)}).$

Step 3. If $Y < f(V)$ then accept and set $X \sim V$. Otherwise, reject and go to Step 1.

Efficiency. Require an average of $\frac{1}{\lambda} = 0.76$ proposals since acceptance rate

$$\frac{1}{\lambda} \approx \frac{1}{1.33} = 0.75.$$ 

So, need to draw an average of $3 \times 0.75 = 2.25$ uniform random numbers to generate $X \sim N(0, 1)$. 

\[21\]
Discrete Accept-Reject Method

Generate a discrete $\mathbf{RV} \mathbf{X}$ from a known target PMF $p: \mathcal{X} \rightarrow [0, 1]$. $p_i = P(X = x_i), \ i \in \mathcal{X}$. $\mathcal{X}$ is finite or countably infinite. $x_i \in \mathcal{X}$, distinct.

\[ \sum_{i \in \mathcal{X}} p_i = 1. \]

Suppose we know how to generate a RV $V$ from a proposal PMF $q: \mathcal{X} \rightarrow [0, 1]$.

$\hat{q}_i = (i \in \mathcal{X}), \sum_{i \in \mathcal{X}} \hat{q}_i = 1$. Suppose also there exists $\alpha \in \mathbb{R}$, $\alpha > 0$ such that

\[ p_i \leq \alpha q_i, \ \forall i \in \mathcal{X}. \]

Then, we can proceed as follows:

1. Generate $V \sim q$.
2. Generate $Y \sim U[0, \alpha q_V]$.
3. If $Y \leq \frac{1}{\alpha} p_V$ then accept. Set $X \leftarrow V$ and stop.
   Otherwise, reject. Go to (1).

Example: $\mathcal{X} = \{1, 2, \ldots\}$. The target PMF $p$ is given by $p_i = \frac{1}{i^2}$, $i = 1, 2, \ldots$. ($p_i \geq \frac{1}{\hat{q}} \forall i, \sum_{i \in \mathcal{X}} q_i = 1$).

Take the proposal PMF to be $\hat{q}$ with

\[ \hat{q}_i = \frac{1}{(i+1)^2}, \ i = 1, 2, \ldots. \]

We can check that:

\[ p_i \leq \alpha \hat{q}_i = \frac{1}{i} - \frac{1}{i+1} \Rightarrow \alpha = 2. \]

(\( \lfloor \cdot \rfloor \) = greatest integer $\leq \cdot$.) So, easy to generate $V \sim q$. Also,\[ \max_{i \in \mathcal{X}} \frac{p_i}{q_i} = \frac{p_1}{q_1} = \frac{1}{2}. \]

Now, the algorithm generating $X \sim p$ is:

1. Generate $U \sim U[0, 1]$ and set $V \leftarrow -\ln U$.
2. Generate $(i \in \mathbb{N}_0, 1)$ and set $Y \leftarrow U \left( \frac{i}{\alpha} + 1 \right)$.\[ q_1 \leq Y \leq q_i \]
3. If $Y \leq \frac{1}{\alpha} p_V$ then accept. Set $X \leftarrow V$ and stop.
   Otherwise, goto Step 1.

Let \( X_1, \ldots, X_M \) be \( M \) distinct real numbers. Denote \( V = \{ X_1, \ldots, X_M \} \) and \( S = \{ 1, 2, \ldots, M \} \). Let \( X : S \to V \) be a discrete RV such that
\[
P_i = \mathbb{P}(X = X_i) > 0, \quad \forall i \in S
\]
and
\[
\sum_{i=1}^{M} P_i = 1.
\]

So, the discrete density for \( X \) is given by
\[
f(x) = \begin{cases} 
P_i & \text{if } x = X_i \text{ for some } i \in S; \\ 0 & \text{otherwise}. \end{cases}
\]

We wish to sample \( Z \sim f \).

The discrete inversion method: Generate \( U \sim U[0,1] \), search \( m \) s.t. \( p_1 + \cdots + p_{m-1} < U \leq p_1 + \cdots + p_{m-1} + p_m \), and set \( Z \leftarrow X_m \). This involves the search of \( m \) with the complexity \( O(M) \) for linear search or \( O(\log M) \) for binary search.

The alias method consists of two parts: set up and sampling. Set up means the construction of \( M \) evenly weighted 2-point densities, equivalent to the original density, requiring \( \mathcal{O}(M) \) work with the complexity same as that of the discrete inversion method. Then, sampling is to sample first a \( M \) even values uniformly, followed by sampling a two-point density, i.e., generating a RV to be some \( x \) with some probability \( p \leftarrow p \) and some \( y \) with probability \( 1 - p \).
Setup: Original distribution/density

\[ \begin{array}{cccc}
T_1 & T_2 & \cdots & T_M \\
1 & 2 & \cdots & M \\
\end{array} \]

\[ \beta > 0, \sum \beta = 1 \]

Construct 2xM table

\[ \begin{array}{cccc}
a_1 & a_2 & \cdots & a_M \\
c_1 & c_2 & \cdots & c_M \\
b_1 & b_2 & \cdots & b_M \\
j_1 & j_2 & \cdots & j_M \\
\end{array} \]

Rules:
1. \( \alpha_i \geq 0, \beta_i \geq 0 \)
2. \( \alpha_k \in \{1, \ldots, M\} \) (labels).
3. \( \alpha_k + \beta_k = 1 \)

Example:
\[ \begin{array}{cccccc}
\beta_1 = 0.41, \beta_2 = 0.27, \beta_3 = 0.07 \\
\beta_4 = 0.14, \beta_5 = 0.11 \end{array} \]

\[ \begin{array}{ccccc}
0.41 & 0.27 & 0.07 & 0.14 & 0.11 \\
1 & 2 & 3 & 4 & 5 \\
\end{array} \]

Setup:

- Poor guys = \( \alpha_k \) with \( \beta_k \leq \frac{1}{M} \) stand in 1st row.
- Rich guys = \( \alpha_k \) with \( \beta_k > \frac{1}{M} \).

First column then becomes poor. \( \alpha_k = \frac{1}{M} \), call it a middle class number (\( \frac{1}{M} \)).

Note: From \( M \) to \( 2M \), sum uniform in \( \alpha \) and columns. \( \alpha_k + \beta_k = \frac{1}{M} \) (\( k = 1, 2, \ldots, M \)).

Note: It is possible that a rich \( \alpha_k \) becomes 0 after donating its value to poorer \( \beta_k \) numbers/guys. In this case, other rich numbers need to be two or more at values, e.g., 0.25, 0.06, 0.29, 0.3, 0.3.
**Alias Algorithm**

*Set up* (Construct $M$ evenly weighted two-point densities that are equivalent to the given density.)

**Step 1** Set $S = \{1, 2, ..., M\}$.
- Set $t = 1$.
- Set $v_k = \frac{p_k}{M}$ for each $k$.

**Step 2** Set $i_t$ to value of $k$ such that $v_{i_t}$ minimizes $v_k$.
- Set $j_t$ to value of $k$ such that $v_{j_t}$ maximizes $v_k$.
- Set $a_t = v_{i_t}$.
- Set $b_t = \frac{1}{M} - a_t$.
- Set $v_{i_t}$.
- Remove $i_t$ from $S$.

**Step 3** Set $t = t + 1$.
- If $t = M$, go to step 2; otherwise, stop.

**Alias Algorithm**

**Sampling**

1. Generate $U_i \sim U[0, 1]$ and set $K = \lceil MU_i \rceil$.
   ([x] = smallest integer $\geq x$.)

2. Generate $U_i \sim U[0, 1]$.
   - Set $Z \leftarrow X_{i_K}$ if $U_i \leq a_{i_K}$.
   - Otherwise set $Z \leftarrow X_{j_K}$.

**Theorem** The algorithm generates $Z \sim f$.

**Key Equivalence:** $P_k = \frac{1}{M} \sum_{m=1}^{M} \frac{a_{m} \cdot I_{[m, m+1)} + b_{m} \cdot I_{[m+1, \infty[}}{v_{k \in S}} \left[ I_{[m, m+1)} + I_{[m+1, \infty[} \right]$.

Prove this by induction.
The composition method

Let \( f \) be the PDF of a RV \( X : \mathbb{R} \to \mathbb{R} \). Let \( f \geq 0 \) and \( \int f(x) \, dx = 1 \). Divide the region between the graph of \( y = f(x) \) and the x-axis into finitely many regions, say \( S_1, S_2, ..., S_m \), with areas \( x_1, x_2, ..., x_m \), respectively. So \( \sum_{k=1}^{m} x_k = \int f(x) \, dx = 1 \).

(See the above figure.)

The composition method to generate a RV \( X \sim f \) is as follows:

1. Generate a RV \( \mathcal{I} \in \{1, 2, ..., m\} \) with the discrete density \( (x_1, x_2, ..., x_m) \).
2. Generate \( (V, W) \) uniformly in \( S_\mathcal{I} \).
3. Set \( X \leftarrow V \).

If \( S_\mathcal{I} \) are mostly regular (e.g., rectangles), then step 2 is more efficient.

Generalization: Write \( f = \sum_{i=1}^{M} x_i f_i \) as mixture of known densities \( f_i \).