Sampling Random Variables

Bo Li, Math, UCSD. Spring 2019

Let $(\Omega, \mathcal{A}, P)$ be a probability space. Let $X: \Omega \to \mathbb{R}$ be a random variable (RV) with probability density function (PDF), or simply the density $f_X : \mathbb{R} \to [0, \infty)$. Sampling X with the given density f_X means to produce $X_1, X_2, \ldots$ i.i.d (independent identitically 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 $\in [0,1]$. 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. $U \sim U[0,1] \Longrightarrow X \sim f$.

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 $[0,1]$.

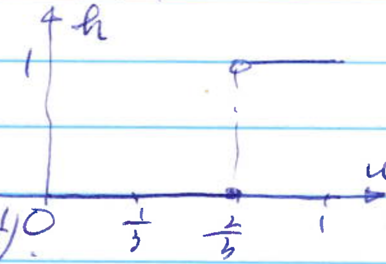
$\xi \sim N(0,1)$: ξ has the standard normal or Gaussian distribution.

If $F = F(x)$ is the distribution (or cumulative distribution function — CDF) of a RV, then $X \sim F$ means the RV X has the CDF F, or X is F-distributed.

<u>Example 1</u> Sample a RV $X \sim \text{Bernoulli}(\frac{1}{3})$.

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

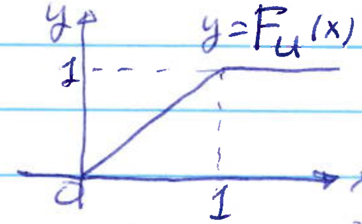
$$h(u) = \begin{cases} 0 & \text{if } 0 \le u \le 2/3 \\ 1 & \text{if } 2/3 < u \le 1 \end{cases}$$



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

<u>Proof.</u> $\forall \omega \in \Omega$.

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



Note that $P(U \notin [0,1]) = 0$. So, we can define $h(U(\omega)) = -\infty$ if $U(\omega) \notin [0,1]$. So $P(h(U) \ne 0,1) = 0$

So, $P(h(U) = 0) = P(0 \le U \le \frac{2}{3}) = F_U(\frac{2}{3}) - F_U(0) = \frac{2}{3}$

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

Hence, $h(U) \sim$ Bernoulli$(1/3)$.

<u>Example 2</u> Sample $X \sim \text{Exp}(7)$. Recall the ~~PDF~~ ~~PDF~~ CDF (or just the distribution) for $\text{Exp}(7)$ [exponential distribution with $\lambda = 7$] is given by

$$F(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1 - e^{-7x} & \text{if } x \ge 0. \end{cases}$$

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

(log: natural logarithm).

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

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

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

~~$P(h(U) \le 0) = 0$ i.e. $F_{h(U)}(x) = 0$ if $x < 0$~~

~~Let $t > 0$ we have~~

Note: We should define

$$X = \begin{cases} h(U) & \text{if } U \in [0,1) \\ 0 & \text{if } U \notin [0,1]. \end{cases}$$

But $P(U < 0 \text{ or } U > 1) = 0$. So, $F_X(x) = 0$ if $x \le 0$

Let $t>0$. We have

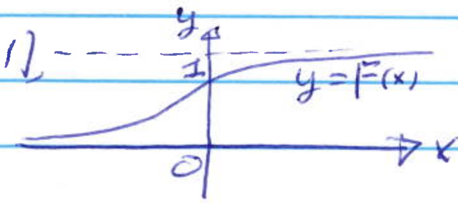
$$P(h(U)\le t) = P(\log U \ge -7t)$$
$$= P(U \ge e^{-7t})$$
$$= P(U\ge e^{-7t}) + P(U<e^{-7t}) - P(U<e^{-7t})$$
$$= 1 - P(U<e^{-7t})$$
$$= 1 - F_U(e^{-7t})$$
$$= 1 - e^{-7t} \quad (\text{since } e^{-7t}\in(0,1))$$

i.e., $F_{h(U)}(x) = 1-e^{-7x}$ if $x>0$.

Together $F_X(x) = \begin{cases} 0 & \text{if } x\le 0 \\ 1-e^{-7x} & \text{if } x>0. \end{cases}$ $\quad X = \begin{cases} h(U) & \text{if } 0\le U\le 1 \\ 0 & \text{otherwise} \end{cases}$

The Inversion Method

Lemma Let $F:\mathbb{R}\to[0,1]$ be the distribution of a RV $:\Omega\to\mathbb{R}$. If $U\sim\mathcal{U}[0,1]$ then $X := F^{-1}(U)\sim F$.



Note: we should really define $X = \begin{cases} F^{-1}(U) & \text{if } 0\le U\le 1 \\ -\infty & \text{otherwise} \end{cases}$ Then $P(X=-\infty) = P(U<0 \text{ or } U>1) = P(U<0) + P(U>1) = 0$.

Here F^{-1} is the generalized inverse ~~function~~ of F defined by

$$F^{-1}(u) = \inf\{z\in\mathbb{R} : F(z)\ge u\} \quad \forall u\in(0,1).$$

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 $\forall z\in\mathbb{R}$

$$P(F^{-1}(U)\le z) \overset{(*)}{=} P(U\le F(z))$$

Use the right-continuous of F

$= F_U(F(z)) = F(z)$ since $F(z)\in[0,1]$. $\square$

Pf of $(*)$. Since $P(U\le 0)=0$, $P(U\ge 1)=0$. We consider only $0<U<1$ and show $F^{-1}(U)\le z \iff U\le F(z)$ for any $z\in\mathbb{R}$. "$\Leftarrow$" Follows from the def. F^{-1}.

"$\Longrightarrow$" Fix $z \in \mathbb{R}$ and assume $F^{-1}(U) \le z$. By the def. of $F^{-1}(U)$, $\exists\, z_n' \downarrow$ s.t. $F^{-1}(U) = \lim_{n\to\infty} z_n'$, $z_n' \le F^{-1}(U)$, and $F(z_n') \ge U$. Note that z_n' is bounded below, for otherwise $z_n' \to -\infty$ and $F(z_n') \to 0$ contradicting $F(z_n') \ge U > 0$ $(n = 1, 2, \cdots)$. So, $z_n' \downarrow z'$ for some $z' \in \mathbb{R}$. But all $z_n' \le F^{-1}(U)$. So $z' \le F^{-1}(U) \le z$. Moreover, the right-continuity implies $F(z') = \lim_{n\to\infty} F(z_n') \ge U$. Hence, the monotonicity of F implies that $U \le F(z') \le F(z)$. $\square$

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

$$x = F^{-1}(u),\ u \in (0,1). \Leftrightarrow F(x) = u \in (0,1).$$

$$1 - e^{-\lambda x} = u \in (0,1). \Longleftrightarrow x = -\frac{1}{\lambda}\log(1-u).$$

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

Example Sample a RV with density $f(r) = \begin{cases} 0 & \text{if } r < 0 \\ r e^{-\frac{1}{2}r^2} & \text{if } r \ge 0. \end{cases}$ The distribution is $F(r) = \int_{-\infty}^{r} f(u)\,du = 1 - e^{-\frac{1}{2}r^2}$, if $r \ge 0$, and $F(r) = 0$ if $r \le 0$. Solve $F(r) = u \in (0,1)$

$$\Longrightarrow r = \sqrt{-2\log(1-u)}.$$

So, if $U \sim \mathcal{U}[0,1]$. 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.

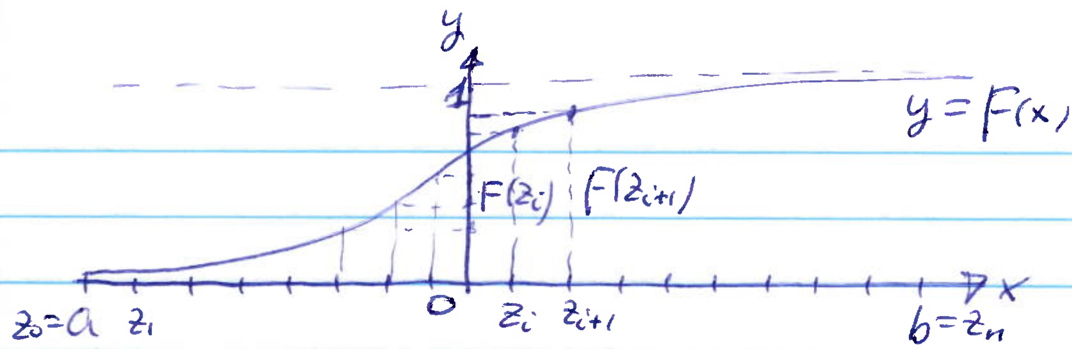
Sample a RV with CDF F.

Step 1. Find $a, b \in \mathbb{R}$, $a < b$, and set

$$a = z_0 < z_1 < \cdots z_n = b.$$

So that $F(a) \ll 1$, $1 - F(b) \ll 1$, $|F(z_{i+1}) - F(z_i)| \ll 1$ for all i.

Cont. from end of page 4.



<u>step 2</u> Generate $U \sim \mathcal{U}[0,1]$.

<u>step 3</u> Find i s.t. $F(z_i) \le U \le F(z_{i+1})$

(Need to use some search algorithm)

<u>step 4.</u> Define

$$X = z_i + (z_{i+1} - z_i)\frac{U - F(z_i)}{F(z_{i+1}) - F(z_i)}.$$

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 $F(z) = U$ numerically to find z, then use this z as our X.

$$U = F(z) \approx F(z_i) + F'(z_i)(z - z_i)$$

$$\Rightarrow \quad z \approx z_i + \frac{U - F(z_i)}{F'(z_i)}. \qquad \text{But } F'(z_i) = \frac{F(z_{i+1}) - F(z_i)}{z_{i+1} - z_i}$$

Hence. $$z \approx z_i + (z_{i+1} - z_i)\frac{U - F(z_i)}{F(z_{i+1}) - F(z_i)}.$$

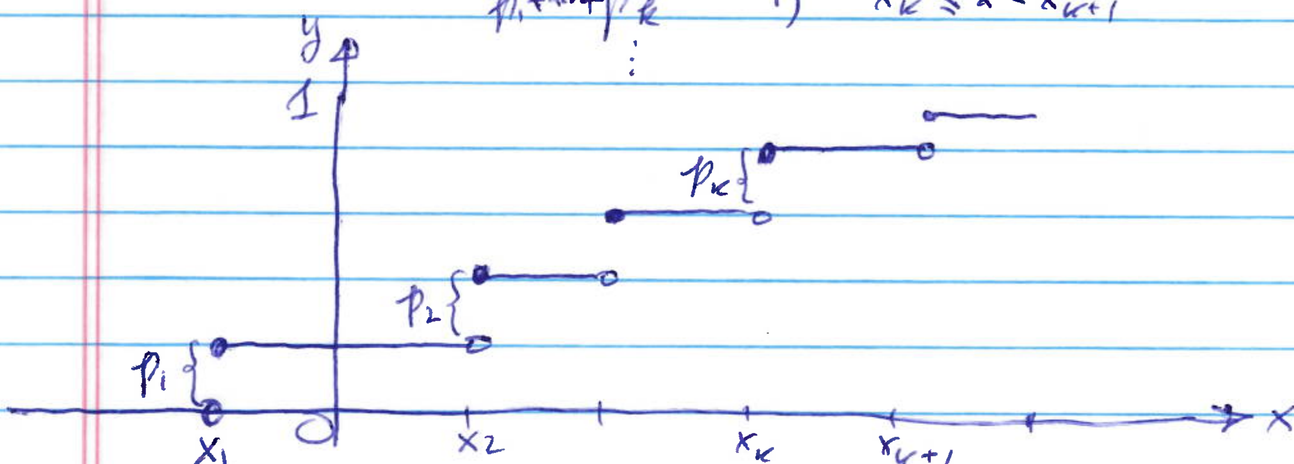
Replace z by X and "$\approx$" by "$=$".

The Discrete Inversion Method (Or the Inversion Method for Discrete RVs)

Let $X: \Omega \to \mathbb{R}$ be a discrete RV taking values $x_1 < x_2 < \cdots$ with probabilities $p_1, p_2, \ldots$, where $p_k = P(X = x_k) > 0$ and $\sum_{k \geq 1} p_k = 1$

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

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



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

The X is a discrete RV with $X \sim F$.

This alogrithm 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: $\alpha = \alpha(x)$ may not itself be a distrib. function. $\alpha(x)$ can be negative

Let $X : \Omega \to \mathbb{R}$ 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. α is increasing.

Define $Y = \alpha(X)$. Then $Y : \Omega \to \mathbb{R}$ 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

$$F_Y(y) = P(Y \le y) = P(\alpha(X) \le \alpha(x))$$
$$= P(X \le x) = F_X(x)$$

$F_Y' = f_Y$
$F_X' = f_X$

So, $$\frac{d}{dx}\left[F_Y(y(x))\right] = F_X'(x) = f_X(x)$$
$$\overset{\parallel}{F_Y'(y)\, y'(x)}$$

$$f_Y(y)\frac{d\alpha}{dx} = f_X(x)$$

$$\boxed{f_Y(y) = f_X(x)\left(\frac{d\alpha}{dx}\right)^{-1}, \quad x = \alpha^{-1}(y)}$$

Let $\beta \in C^1(\mathbb{R})$ $\beta \downarrow$ strictly i.e., $\beta'(x) < 0$ $\forall x \in \mathbb{R}$.

Consider now $Y = \beta(X)$.
What is f_Y?

Let $y = \beta(x)$. $F_Y(y) = P(Y \le y) = P(\beta(X) \le y(x))$
$$= P(X \geq x) = 1 - P(X < x)$$
$$= 1 - F_X(x)$$
$\frac{d}{dx}$: $$f_Y(y)\frac{d\beta}{dx} = -f_X(x)$$
$$\boxed{f_Y(y) = f_X(x)\left|\frac{d\beta}{dx}\right|^{-1} \qquad y = \beta^{-1}(x)}$$

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

<u>Example</u> $X \sim N(0,1)$ $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$
let $\alpha(x) = \sigma x + \mu$, for some $\sigma > 0$ and $\mu \in \mathbb{R}$.
Let $Y = \alpha(X)$. $f_Y(y) = f_X(x)\left(\frac{dy}{dx}\right)^{-1} = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \cdot \frac{1}{\sigma}$, $y = \alpha(x)$
But $y = \alpha(x) = \sigma x + \mu$. 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 $Y \sim N(\mu, \sigma^2)$.

<u>Example</u> $X \sim U[0,1]$ $\alpha(x) = -\log x$, $x \in (0,1]$.
let $Y = \alpha(X) = -\log X$ ~~and Y = 0 if X = 0~~. For $y = \alpha(x)$, $x \in (0,1]$,
$$f_Y(y) = f_X(x)\left|\frac{d\alpha}{dx}\right|^{-1} = \left|\frac{-1}{x}\right|^{-1} f_X(x) = x f_X(x)$$

If $x \in (0,1)$ then $y = \alpha(x) = -\log x \in (0,\infty)$

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

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

Note that $P(X \le 0) = F_U(0) = 0$

$P(X > 1) = 0$ $(= 1 - P(X<1) = 1-1 = 0)$

So, on the subset $\{X \le 0\} \subseteq \Omega$, we can define $Y = 0$. So,

$$Y = \begin{cases} 0 & \text{on } \{X \le 0\} \\ -\log X & \text{on } \{0 < X \le 1\} \end{cases}$$

Thus, $F_Y(y) = P(Y \le y) = 0$ if $y \le 0$.

$\boxed{\text{Hence } f_Y(y) = F_Y'(y) = 0 \text{ if } y < 0}$

Now, for $y > 0$ $F_Y'(y) = f_Y(y) = e^{-y}$.

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

$$= \int_0^y e^{-u}\,du = 1 - e^{-y}.$$

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

This is the exponential distribution with $\lambda = 1$.

In what follows, if $U \sim U[0,1]$, i.e., $U : \Omega \to \mathbb{R}$ is a RV $U[0,1]$-distributed. Then we can assume $0 \le U \le 1$. Since $P(U \notin [0,1]) = 0$.

In computer, a RND (a random number) generated in $[0,1]$ is always in $[0,1)$.

The transformation method for sampling multi-variates (multiple RVs)

Let $X_1, \cdots, X_n : \Omega \to \mathbb{R}$ be n RVs. Then,
$X = (X_1, \cdots, X_n) : \Omega \to \mathbb{R}^n$ is a vector-valued RV.
The joint distribution function is $F_X : \mathbb{R}^n \to [0,1]$.

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

The joint density (assumed to exist) is denoted

$$f_X(x) = f_{X_1 \cdots X_n}(x_1, \cdots, x_n) : \mathbb{R}^n \to [0, \infty)$$

It is related to F_X by

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

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

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

nonsingular at any $y \in \text{Range}(T)$, where

$$T(x_1, \cdots, x_n) = (y_1, \cdots, y_n),$$
$$(x_1, \cdots, x_n) = T^{-1}(y_1, \cdots, y_n).$$
$$x_i = x_i(y_1, \cdots, y_n) \quad (i = 1, \cdots, n).$$

Define $Y = T(X_1, \cdots, X_n) : \Omega \to \mathbb{R}^n$. What is the joint density for Y?

Theorem Under the above assumptions, we have for $Y = T(X_1, \dots, X_n) = (Y_1, \dots, Y_n)$

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

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

We have

$$P(X \in B) = \int_B f_X(x)\,dx \overset{x = T^{-1}(y)}{=} \int_{T(B)} f_X(T^{-1}(y))\,|J(y)|\,dy$$

$$\|$$

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

Hence $f_Y(y) = f_X(T^{-1}(y))\,|J(y)|$. $\square$

The result is formally: $f_Y(y)\,dy = f_X(x)\,dx$ $\qquad \frac{dx}{dy} = |J(y)|$.

The Box-Muller Method for Sampling Two Independent Gaussian RVs

~~and Standard~~

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

(1) Independent: $f_{X,Y}(x,y) = f_X(x)\, f_Y(y)$

(2) $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $\quad f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$

~~(3)~~ i.e. $X \sim N(0,1)$, $Y \sim N(0,1)$.

We would like to sample X, Y by generating $\xi, \eta \sim U[0,1]$ and use the transformation method.

We implement this using two steps.

Step 1. Change variables (x, y) to (r, θ)

$$x = r\cos\theta \qquad X = R\cos\Theta$$
$$y = r\sin\theta. \qquad Y = R\sin\Theta$$

Generate RVs R and Θ that corresponding to X, Y.

Find f_R, f_Θ

Step 2 Generate $\xi, \eta \sim U([0,1])$. And then using the transformation method again (or the inversion method) to generate R and Θ.

Let's begin with

$$f_{X,Y}(x,y)\,dx\,dy = f_{R,\Theta}(r,\theta)\,dr\,d\theta$$

Independence:

$$f_X(x) f_Y(y)\,dx\,dy = f_R(r) f_\Theta(\theta)\,dr\,d\theta$$

But, the left-hand side is

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

$$= \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}\,dx\,dy$$

$$\begin{cases} x = r\cos\theta \\ y = r\sin\theta \end{cases} \Downarrow \quad = \frac{1}{2\pi} e^{-\frac{1}{2}r^2}\, r\,dr\,d\theta$$

So, define R, Θ so that $X = R\cos\Theta$, $Y = R\sin\Theta$.

Hence we choose $f_R(r) = r e^{-\frac{1}{2}r^2}$ $(r \geq 0)$

$$f_\Theta(\theta) = \frac{1}{2\pi} \qquad \theta \in [0, 2\pi)$$

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, Θ with the densities f_R and f_Θ respectively, by the inversion method.

By the second example on page [4] of this set of notes, we have

$$R = \sqrt{-2\log(1-\xi)} \qquad \text{if } \xi \sim U[0,1]$$

For sampling Θ, note first from $f_\Theta(\theta) = \frac{1}{2\pi}$ only for $\theta \in [0, 2\pi)$, we have

$$F_\Theta(\theta) = \begin{cases} 0 & \text{if } \theta \le 0 \\ \frac{\theta}{2\pi} & \text{if } 0 < \theta \le 2\pi \\ 1 & \text{if } \theta > 2\pi. \end{cases}$$

Let $\eta \sim U[0,1]$.

Solve: $F_\Theta(\theta) = \eta$ on $0 < \theta \le 2\pi$.

$$\frac{\theta}{2\pi} = \eta, \quad \theta = 2\pi\eta \in [0, 2\pi)$$

Hence

$$R \quad \Theta = 2\pi\eta$$

Now, we have $\boxed{\text{the Box-Muller method}}$

$$X = \sqrt{-2\log(1-\xi)}\cos(2\pi\eta)$$
$$Y = \sqrt{-2\log(1-\xi)}\sin(2\pi\eta)$$

$\xi, \eta \sim U[0,1]$ 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, $f_{R,\Theta} = f_R \cdot f_\Theta$.

$$F_\Theta(\theta) = P(\Theta \le \theta) = P\left(\eta \le \frac{\theta}{2\pi}\right) = F_\eta\left(\frac{\theta}{2\pi}\right) \quad \cancel{\text{since } \eta \sim U[0,1]}$$

Since $\eta \sim U[0,1]$ $\rightarrow$

$$= \begin{cases} 0 & \text{if } \theta < 0 \\ \frac{\theta}{2\pi} & \text{if } 0 \le \theta \le 2\pi \\ 1 & \cancel{\text{elsewhere}}\ \text{if } \theta > 2\pi. \end{cases}$$

So, $f_\Theta(\theta) = \begin{cases} \frac{1}{2\pi} & \text{if } \theta \in [0, 2\pi] \\ 0 & \text{elsewhere.} \end{cases}$

Similarly,

$$\forall r \geq 0 \quad F_R(r) = P(R \leq r) = P(\xi \geq e^{-\frac{r^2}{2}})$$
$$= 1 - P(\xi < e^{-\frac{r^2}{2}})$$
$$= 1 - F_\xi(e^{-\frac{r^2}{2}})$$
$$= 1 - e^{-\frac{r^2}{2}} \quad \text{since } \xi \sim U[0,1]$$

$$\forall r < 0 \quad F_R(r) = P(R \leq r) = 0.$$

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

Now, by the theorem on the transformation method for multi-variates, we have for the new RVs $X = R\cos\Theta$, $Y = R\sin\Theta$. [following the transform $T(r,\theta) = (x,y)$, $x = r\cos\theta$, $y = r\sin\theta$.] that $f_{X,Y}(x,y) = f_{R,\Theta}(r,\theta)\,|J(x,y)|$

$$|J(x,y)| = \left|\frac{\partial(r,\theta)}{\partial(x,y)}\right| = \left|\det\begin{pmatrix} \partial_x r & \partial_y r \\ \partial_x \theta & \partial_y \theta \end{pmatrix}\right| = \frac{1}{r}$$

for $r \geq 0$. Hence

$$f_{X,Y}(x,y) = f_{R,\Theta}(r,\theta)\frac{1}{r}$$
$$= \frac{1}{2\pi} e^{-\frac{r^2}{2}}$$
$$= \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}$$

$$F_{X,Y}(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f_{X,Y}(u,v)\,du\,dv$$

The marginal distribution is
$$F_X(x) = F_{X,Y}(x,\infty) = \int_{-\infty}^{\infty} dy \int_{-\infty}^{x} du\, f_{X,Y}(u,v)$$
$$= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}\,du. \quad \left[\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{v^2}{2}}\,dv = 1.\right]$$

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, $X, Y \sim \mathcal{N}(0,1)$. Moreover

$$f_{X,Y}(x,y) = f_X(x) f_Y(y)$$

So, X and Y are independent.

Note. To generate a single RV $X \sim \mathcal{N}(0,1)$, one can use the CLT (Central Limit Theorem)

(1) Generate $\xi_1, \xi_2, \dots, \xi_N$, i.i.d. $\mathcal{U}[0,1]$-distributed.

(2) Set $X_N = \sqrt{12N}\left(\sum_{k=1}^{N} \xi_k - \frac{1}{2}N\right)$

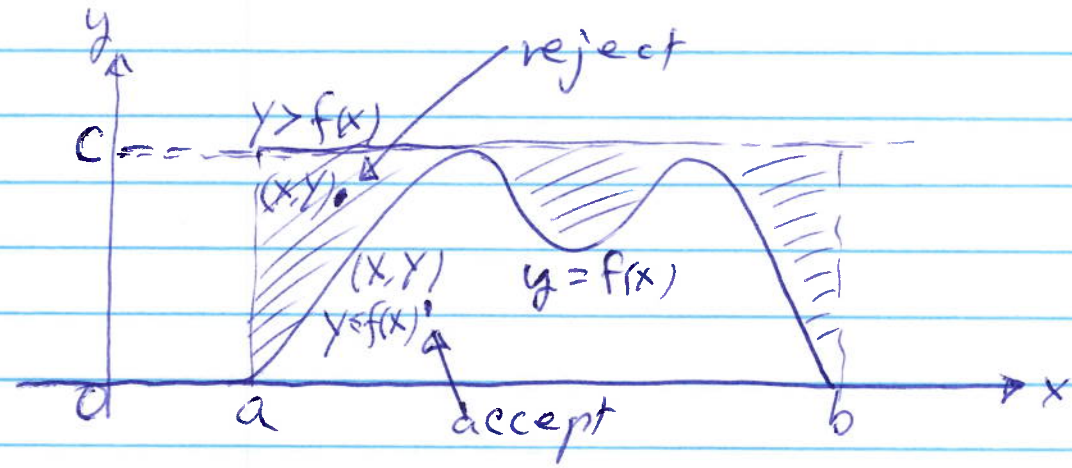
Then, Approximately $X_N \sim \mathcal{N}(0,1)$. if $N \gg 1$.

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~~ transforming from some other distributions (e.g., $\mathcal{U}[0,1]$). 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 Ω 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.,

(Simple acceptance-rejection) $c = \sup\{ f(x) : x \in [a,b] \}$.

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

<u>step 1</u> Generate $X \sim U[a,b]$

~~step 1~~ e.g., generate $U \sim U[0,1]$.

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

<u>step 2</u> Generate $Y \sim U[0,c]$, independent of X

<u>step 3</u> If $Y \le f(X)$, accept X and 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 on $[a,b]\times[0,c]$, and the accepted X satisfies $Y \le f(X)$. Thus the magnal distribution for the accepted X is f.

~~and~~ X and Y are independent.

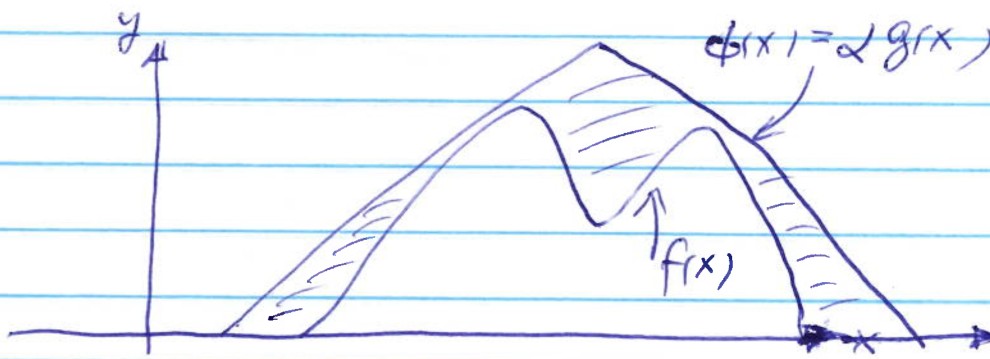
magnal: $f_X(x) = \int_0^{f(x)} 1\,dy = f(x)$.

A more general (accept-reject or ~~acceptance-rejection~~) method is based on the following idea:

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

1. We know how to sample RVs with g the PDF.
2. $\alpha g(x) \ge f(x)$ 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 <u>majorizing function</u> of $f(x)$, and $g(x)$ a proposal PDF.

<u>Algorithm: Accept~~ance~~ Rejection Method</u> for generating $Z \sim f$

<u>step 1</u> Generate $X \sim g$.
<u>step 2</u> Generate $Y \sim U[0, \alpha g(X)]$.
<u>step 3</u> If $Y \le f(X)$ then accept. $Z \leftarrow X$.
Otherwise, reject. Return to step 1.

<u>Theorem</u> The algorithm generates $Z \sim f$.

Proof Let $A = \{(x,y) : 0 \le y \le \alpha g(x)\} \subseteq \mathbb{R}^n \times \mathbb{R}^1$,
$B = \{(x,y) : 0 \le y \le f(x)\} \subseteq \mathbb{R}^n \times \mathbb{R}^1$.

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 $q(x,y)$ denote the joint PDF of (X,Y), and let $q(y|x)$ denote the conditional PDF of Y given $X = x$. Then

$$q(x,y) = \begin{cases} g(x)\, q(y|x) & \text{if } (x,y) \in A, \\ 0 & \text{otherwise}. \end{cases}$$

By step 2 implies that $q(y|x) = \frac{1}{\alpha g(x)}$ for $y \in [0, \alpha g(x)]$, and $q(y|x) = 0$ elsewhere. Thus,

$$q(x,y) = \frac{1}{\alpha} \quad \text{for every } (x,y) \in A.$$

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

$$\int_0^{f(x)} 1 \, dy = f(x). \qquad \square$$

Note that the efficiency of this algorithm is defined as

$$P((X,Y) \text{ is accepted}) = \frac{\cancel{\alpha}\, \text{vol}(B)}{\text{vol}(A)} = \frac{1}{\alpha}.$$

Also, in generating many RVs $\sim f$, we introduce a similar concept:

$$\text{acceptance rate} = \frac{\#\text{ of } X \text{ accepted}}{\#\text{ of } X \text{ generated}} \ (= \text{efficiency})$$

Often a slightly modified version of the above algorithm is used.

$Y \sim U[0, \alpha g(X)]$ is same as setting

$Y = U \alpha g(X)$, where $U \sim U[0,1]$.

Then, $Y \le f(X)$ is equivalent to $U \le \frac{f(X)}{\alpha g(X)}$.

Hence, the modification is:

Generate X from $g(x)$ and accept it with probability $f(X)/[\alpha g(X)]$.

Algorithm: Modified Accept-Reject Method to Sample a RV $Z \sim f$.

Let f be a given PDF of RV in $\mathbb{R}^n$.

Let g be a PDF of RV in $\mathbb{R}^n$ and $\alpha \ge 1$

Suppose $f(x) \le \alpha g(x)$ $\forall x \in \mathbb{R}^n$ ~~(comm~~

$\mathbb{R}^n$ can be replaced by a finite region, the support of f, if f is finitely supported

Step 1 Generate X from g

Step 2 Generate $U \sim U[0,1]$ independent of X.

Step 3 If $U \le \frac{f(X)}{\alpha g(X)}$, then accept X, and $Z \leftarrow X$.

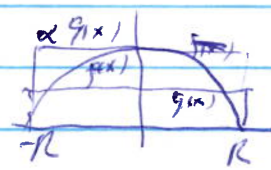
Otherwise, reject X, go to Step 1.

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

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

$\alpha g(x) \ge f(x)$, $x \in [-R, R)$

The smallest such α is $\alpha = \frac{4}{\pi}$.



Z

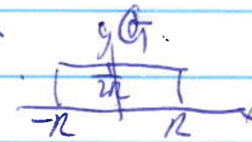
We sample a RV with distribution f as follows.

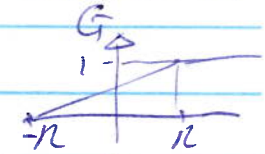
① Generate independent $U_1, U_2 \sim U[0,1]$.

② Use U_2 to generate $X \sim g$ via the inversion method $X = (2U_2 - 1)R$.

↖ Need first to calculate G

③ Calculate
$$\frac{f(X)}{\alpha g(X)} = \sqrt{1 - (2U_2 - 1)^2}.$$





④ If $U_1 \le f(X)/[\alpha g(X)]$, i.e; $(2U_2-1)^2 \le 1 - U_1^2$ or $U_1^2 + 4U_2^2 - 4U_2 \le 0$, 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 $\frac{1}{\alpha} = \frac{\pi}{4} \approx 0.785$.

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

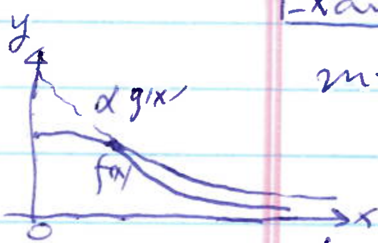
Let f be a PDF.
Let g, h also be PDFs that are relatively easy to sample. (e.g., $h(x)$ can be a piecewise linear function). Assume
$$h(x) \le f(x) \le \alpha g(x) \quad \forall x.$$

Algorithm

step 1 Generate $X \sim g(x)$, $U \sim U[0,1]$.

step 2 Accept X if $U \le \frac{h(x)}{\alpha g(x)}$.

step 3 Otherwise, accept X if $U \le f(x)/[\alpha g(x)]$.



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

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

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

Now, set
$$\alpha = \sup_{x \in \mathbb{R}} \frac{f(x)}{g(x)} = \sup_{x \ge 0} \sqrt{\frac{2}{\pi}} e^{x - \frac{x^2}{2}} = \sqrt{\frac{2e}{\pi}} \doteq 1.32$$

So, $\alpha g(x)$ is a majorizing function of $f(x)$, i.e.,
$0 \le f(x) \le \alpha g(x)$ $(x \in \mathbb{R})$

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

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

Step 2 General $Y \sim U[0, \alpha g(V)]$ by generating $U_3 \sim U[0,1]$ and setting $Y = \alpha g(V) U_3$
$= \frac{\alpha}{2} U_3 e^{-|V|}$ $\left(\Leftrightarrow U_3 \le \frac{f(V)}{\alpha g(V)}\right)$

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

Efficiency: Require an average of $\alpha = 1.32$ proposals/trials (since acceptance rate $= 1/\alpha$). So, need ~~to~~ an average of $3 \times 1.32 = 3.94$ uniform random numbers to generate $X \sim N(0,1)$.

Discrete Accept-Reject Method

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

$p_i > 0$, $\sum_{i \in S} p_i = 1$.

Suppose we know how to generate a RV V from a proposal PMF $q: S \to [0,1]$. $q_i > 0$ $(i \in S)$, $\sum_{i \in S} q_i = 1$. Suppose also there exists $\alpha \in \mathbb{R}$, $\alpha > 0$ such

$$p_i \le \alpha q_i \quad \forall i \in S.$$

Then, we can proceed as follows:

(1) Generate $V \sim q$.

(2) Generate $Y \sim U[0, \alpha q_V]$.

(3) If $Y \le p_V$ then accept. Set $X \leftarrow V$ and stop. Otherwise, reject. Go to (1).

Example $S = \{1, 2, \cdots\}$ The target PMF p is given by $p_i = \frac{6}{\pi^2 i^2}$ $(i = 1, 2, \cdots)$. ($p_i > 0$ $\forall i$, $\sum_i^{\infty} p_i = 1$). Take the proposal PMF to be q with $q_i = \frac{1}{i(i+1)}$ $(i = 1, 2, \cdots)$. We can check that if $U \sim U[0,1]$ then $P(\lfloor U^{-1} \rfloor = i) = \frac{1}{i} - \frac{1}{i+1} = q_i$. ($\lfloor x \rfloor$ = greatest integer $\le x$.) So, easy to generate $V \sim q$. Also, $\max_{i \ge 1} \frac{p_i}{q_i} = \frac{p_1}{q_1} = \frac{12}{\pi^2}$.

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

step 1 Generate $U_1 \sim U[0,1]$ and set $V \leftarrow \lfloor U_1^{-1} \rfloor$.

step 2 Generate $U_2 \sim U[0,1]$ and set $Y \leftarrow U_2 \left(\frac{12}{\pi^2} \frac{1}{V(V+1)} \right)$

step 3 If $Y \le \frac{6}{\pi^2} V^{-2}$ then accept, set $X \leftarrow V$ and stop. Otherwise. Go to step 1.

The Alias Method for Generate Discre RVs with Finite States. (Walker, 1974, 1977)

Let $x_1, \dots, x_M$ be M distinct real numbers. Denote $V = \{x_1, \dots, x_M\}$ and $S = \{1, 2, \dots, M\}$. Let $X: \Omega \to V$ be a discrete RV such that

$$p_i := P(X = x_i) > 0, \quad \forall i \in S$$

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 $\underbrace{Z}_{\text{a RV}} \sim f$.

The discrete inversion method: Generate $U \sim U[0,1]$. search m s.t. $p_1 + \dots + p_{m-1} < U \le p_1 + \dots + p_{m-1} + p_m$ 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 work with the complexity same as that of the discrete inversion method. Then, sampling is to sample first M even values uniformly, followed by sampling a two-point density, i.e., generating a RV to be some x with ~~the~~ some propability $p \in (0,1)$ and some y with probability $1-p$.

Set up: original distribution/density

p_1	p_2	...	p_M
1	2	...	M

$p_i > 0$, $\sum_1^M p_i = 1$

Construct $2 \times M$ table

a_1	a_2	...	a_M
i_1	i_2	...	i_M
b_1	b_2	...	b_M
j_1	j_2	...	j_M

Rules

① $a_i \geq 0$, $b_i \geq 0$

② $i_k \in j_k \in \{1, \cdots, M\}$ (labels).

③ $a_k + b_k = \frac{1}{M}$, $k = 1, \cdots, M$.

④ $\sum_{k: i_k = r} a_k + \sum_{k: j_k = r} b_k = p_r$ $\forall r \in \{1 \cdots M\}$

5

If $p_k \leq \frac{1}{M}$, then p_k is one of a_j's.

If $p_k > \frac{1}{M}$ then p_k is splitted in to a few b_j's and a possible a_j

Example $p_1 = 0.41$, $p_2 = 0.27$, $p_3 = 0.07$, $p_4 = 0.14$, $p_5 = 0.11$

$M = 5$

$\frac{1}{M} = 0.2$

0.41	0.27	0.07	0.14	0.11
1	2	3	4	5

Set up:

"Poor guys" = a_k's with $a_k < \frac{1}{M}$, stand in 1st row.

"Rich guys" = a_k's with $a_k > \frac{1}{M}$. First donates, then becomes "poor" ($< \frac{1}{M}$) then stand in Row 1.

0.07	0.11	0.14	0.19	0.20
3	5	4	1	2
0.13	0.09	0.06	0.01	0
1	1	2	2	2

$p_k = \frac{1}{M}$: call it a "middle class" number

Note: From M to $2M$. But, uniform in columns. $a_k + b_k = \frac{1}{M}$. $(k = 1, 2, \cdots, M)$

Note: It is possible that a rich p_k becomes 0 after donating its values to poor numbers/guys. In this case, other rich numbers need to be two or more a_i 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 $S \leftarrow \{1, 2, \dots, M\}$.

Set $t \leftarrow 1$.

set $r_k \leftarrow p_k$ $\cancel{(k=1,2,\dots,M)}$ for each $k \in S$.

Step 2 Set $i_t \leftarrow$ value of $k \in S$ that minimizes r_k.

Set $j_t \leftarrow$ value of $k \in S$ that maximizes r_k.

Set $a_t = r_{i_t}$.

Set $b_t = 1/M - a_t$.

Set $r_{j_t} \leftarrow r_{j_t} - b_t$.

Remove i_t from S.

Step 3 Set $t \leftarrow t+1$

If $t \le M$, go to step 2; otherwise stop.

Alias Algorithm

Sampling

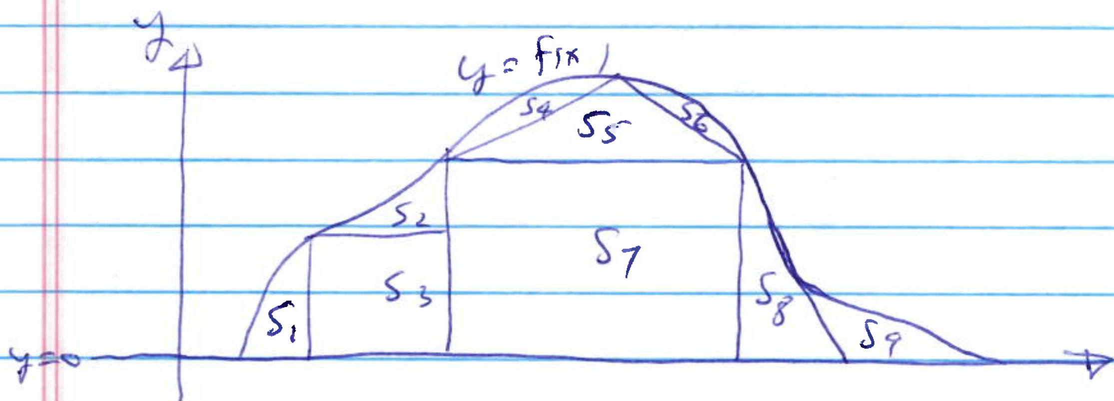
step 1 Generate $U_1 \sim U[0,1]$ and set $K = \lceil M U_1 \rceil$.
($\lceil x \rceil$ = smallest integer $\ge x$.)

Step 2 Generate $U_2 \sim U[0,1]$. Set $Z \leftarrow x_{i_K}$ if $U_2 \le a_{i_K}$. Otherwise set $Z \leftarrow x_{j_K}$.

Theorem The algorithm generates $Z \sim f$.

key. Equivalence: $$p_k = \frac{1}{M} \sum_{m=1}^{M} \left(a_m I_{[i_m = k]} + b_m I_{[j_m = k]} \right) \quad \forall k \in S.$$
Pf this by induction.

The composition method



Let f be the PDF of a RV : $\Omega \to \mathbb{R}$. So $f \geq 0$, $\int_{\mathbb{R}} f\,dx = 1$. Divide the region between the graph $y = f(x)$ and the x-axis into finitely many regions, say, $S_1, S_2, \cdots, S_M$, with areas $\alpha_1, \alpha_2, \cdots, \alpha_M$, respectively. So $\sum_{k=1}^{M} \alpha_k = \int_{\mathbb{R}} f(x)\,dx = 1$. (See the above figure.)

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

step 1 Generate a RV $I \in \{1, \cdots, M\}$ with the discrete density $(\alpha_1, \alpha_2, \cdots, \alpha_M)$.

step 2 Generate (V, W) uniformly in S_I.

step 3 Set $X \leftarrow V$.

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

Generalization: write $f = \sum_{i=1}^{M} \alpha_i f_i$ as mixture of known densities f_i.