The Acceptance Rejection Method

The Inverse-Transform Method is often time consuming in generating random numbers which have positive mass at **many** points, unless for some special distributions; and yet there are **not** many continuous distributions whose cdf has closed from.

Theorem: Let X (to be generated) have pdf $f_X(x)$ on support I. (i.e. $f_X(x) > 0$ if $x \in I$; and $f_X(x) = 0$, otherwise.) Let Y have pdf $f_Y(y)$ **on** I, $U \sim U(0,1)$ and U **and** Y **be independent**. Let $c \geq 1$ and $g(x) = \frac{f_X(x)}{cf_Y(x)}$ such that $0 < g(x) \leq 1, \forall x \in I$. Then

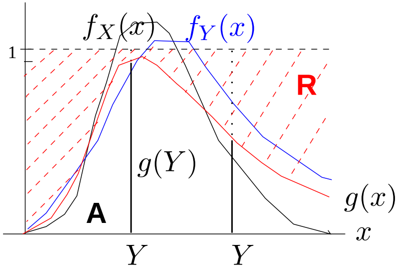
$$f_Y(x|U \leq g(Y)) = f_X(x), \ \forall x \in I.$$

Proof: Casella and Berger (2002).

Theorem: Let X (to be generated) have pdf $f_X(x)$ on support I. (i.e. $f_X(x) > 0$ if $x \in I$; and $f_X(x) = 0$, otherwise.) Let Y have pdf $f_Y(y)$ **on** I, $U \sim U(0,1)$ and U **and** Y **be independent**. Let $c \geq 1$ and $g(x) = \frac{f_X(x)}{cf_Y(x)}$ such that $0 < g(x) \leq 1$, $\forall x \in I$. Then

$$f_Y(x | U \leq g(Y)) = f_X(x), \ \forall x \in I.$$

Proof: Casella and Berger (2002).


$f_X(x)$
$f_Y(x)$
1
R
$g(Y)$
$g(x)$
A
x
Y
Y

Note: 1. It says that the distribution of X is same as the conditional distribution of Y **given** $U \leq g(Y)$.

2. One can generate U and Y independently and see if $U \leq g(Y)$.

3. $U \leq g(Y) \Longleftrightarrow U \leq \frac{f_X(Y)}{cf_Y(Y)}$.

4. The density f_Y is called the *majorizing density*, *trial density*, or *proposal density*.

5. $cf_Y(x)$ is called the majorizing function, *envelope* **function** or *hat function*.

6. The larger c the larger rejection region, so want **small** c.

Note: 1. It says that the distribution of X is same as the conditional distribution of Y **given** $U \leq g(Y)$.

2. One can generate U and Y independently and see if $U \leq g(Y)$.

3. $U \leq g(Y) \Longleftrightarrow U \leq \frac{f_X(Y)}{cf_Y(Y)}$.

4. The density f_Y is called the *majorizing density*, *trial density*, or *proposal density*.

5. $cf_Y(x)$ is called the majorizing function, *envelope* **function** or *hat function*.

6. The larger c the larger rejection region, so want **small** c.

Note: 1. It says that the distribution of X is same as the conditional distribution of Y **given** $U \leq g(Y)$.

2. One can generate U and Y independently and see if $U \leq g(Y)$.

3. $U \leq g(Y) \Longleftrightarrow U \leq \frac{f_X(Y)}{cf_Y(Y)}$.

4. The density f_Y is called the *majorizing density*, *trial density*, or *proposal density*.

5. $cf_Y(x)$ is called the majorizing function, *envelope* **function** or *hat function*.

6. The larger c the larger rejection region, so want **small** c.

Note: 1. It says that the distribution of X is same as the conditional distribution of Y **given** $U \leq g(Y)$.

2. One can generate U and Y independently and see if $U \leq g(Y)$.

3. $U \leq g(Y) \Longleftrightarrow U \leq \frac{f_X(Y)}{cf_Y(Y)}$.

4. The density f_Y is called the *majorizing density*, *trial density*, or *proposal density*.

5. $cf_Y(x)$ is called the majorizing function, *envelope* **function** or *hat function*.

6. The larger c the larger rejection region, so want **small** c.

Note: 1. It says that the distribution of X is same as the conditional distribution of Y **given** $U \leq g(Y)$.

2. One can generate U and Y independently and see if $U \leq g(Y)$.

3. $U \leq g(Y) \Longleftrightarrow U \leq \frac{f_X(Y)}{cf_Y(Y)}$.

4. The density f_Y is called the *majorizing density*, *trial density*, or *proposal density*.

5. $cf_Y(x)$ is called the majorizing function, *envelope* **function** or *hat function*.

6. The larger c the larger rejection region, so want **small** c.

Discrete case

Discrete Case: If the (discrete) r.v. Y with $P(Y = x) = q_x$, $x \in I$, countable, can be generated easily, then a r.v. X with $P(X = x) = p_x$, $x \in I$ can be generated based on Y and an independent $U(0,1)$ r.v..

A-R Algorithm: 1. Generate $U \sim U(0,1)$.

2. Generate $Y \sim P(Y = y) = q_y$, $y \in I$.

3. Set $X = Y$, if $U \leq \frac{p_Y}{cq_Y}$; return to Step 1 otherwise.

Discrete case

Discrete Case: If the (discrete) r.v. Y with $P(Y = x) = q_x$, $x \in I$, countable, can be generated easily, then a r.v. X with $P(X = x) = p_x$, $x \in I$ can be generated based on Y and an independent $U(0, 1)$ r.v..

A-R Algorithm: 1. Generate $U \sim U(0, 1)$.

2. Generate $Y \sim P(Y = y) = q_y$, $y \in I$.

3. Set $X = Y$, if $U \leq \frac{p_Y}{cq_Y}$; return to Step 1 otherwise.

Note: 1. The number of trials N that a successful pair (Y, U) is found has a **geometric** distribution with parameter

$$p = P(U < \frac{p_Y}{cq_Y}) = \sum_{y \in I} P(U < \frac{p_Y}{cq_Y} | Y = y) P(Y = y) = \sum_{y \in I} \frac{p_y}{cq_y} q_y = 1/c.$$

2. The **mean** number of trials to generate an X is $EN = c$.

   So want c as **small** as possible.

3. Take $\mathbf{c = \max_x \frac{f_X(x)}{f_Y(x)} \geq 1}$.

4. More efficient if $\mathbf{c \approx 1} \iff f_Y(\cdot)$ and $f_X(\cdot)$ are **similar**.

Note: 1. The number of trials N that a successful pair (Y, U) is found has a **geometric** distribution with parameter

$$p = P(U < \frac{p_Y}{cq_Y}) = \sum_{y \in I} P(U < \frac{p_Y}{cq_Y} | Y = y) P(Y = y)$$

$$= \sum_{y \in I} \frac{p_y}{cq_y} q_y = 1/c.$$

2. The **mean** number of trials to generate an X is $EN = c$.

So want c as **small** as possible.

3. Take $\mathbf{c} = \max_{\mathbf{x}} \frac{\mathbf{f_X(x)}}{\mathbf{f_Y(x)}} \geq \mathbf{1}$.

4. More efficient if $\mathbf{c} \approx \mathbf{1} \iff f_Y(\cdot)$ and $f_X(\cdot)$ are **similar**.

Note: 1. The number of trials N that a successful pair (Y, U) is found has a **geometric** distribution with parameter

$$\begin{aligned} p &= P\left(U < \frac{p_Y}{cq_Y}\right) = \sum_{y \in I} P\left(U < \frac{p_Y}{cq_Y} \middle| Y = y\right) P(Y = y) \\ &= \sum_{y \in I} \frac{p_y}{cq_y} q_y = 1/c. \end{aligned}$$

2. The **mean** number of trials to generate an X is $EN = c$.

   So want c as **small** as possible.

3. Take $\mathbf{c = \max_x \frac{f_X(x)}{f_Y(x)} \geq 1}$.

4. More efficient if $\mathbf{c \approx 1} \Longleftrightarrow f_Y(\cdot)$ and $f_X(\cdot)$ are **similar**.

Note: 1. The number of trials N that a successful pair (Y, U) is found has a **geometric** distribution with parameter

$$
\begin{aligned}
p &= P\left(U < \frac{p_Y}{cq_Y}\right) = \sum_{y \in I} P\left(U < \frac{p_Y}{cq_Y} \middle| Y = y\right) P(Y = y) \\
&= \sum_{y \in I} \frac{p_y}{cq_y} q_y = 1/c.
\end{aligned}
$$

2. The **mean** number of trials to generate an X is $EN = c$.

So want c as **small** as possible.

3. Take $\mathbf{c = \max_x \frac{f_X(x)}{f_Y(x)} \geq 1}$.

4. More efficient if $\mathbf{c \approx 1} \Longleftrightarrow f_Y(\cdot)$ and $f_X(\cdot)$ are **similar**.

Note: 1. The number of trials N that a successful pair (Y, U) is found has a **geometric** distribution with parameter

$$p = P(U < \frac{p_Y}{cq_Y}) = \sum_{y \in I} P(U < \frac{p_Y}{cq_Y} | Y = y) P(Y = y)$$
$$= \sum_{y \in I} \frac{p_y}{cq_y} q_y = 1/c.$$

2. The **mean** number of trials to generate an X is $EN = c$.

So want c as **small** as possible.

3. Take $\mathbf{c = \max_x \frac{f_X(x)}{f_Y(x)} \geq 1}$.

4. More efficient if $\mathbf{c \approx 1} \iff f_Y(\cdot)$ and $f_X(\cdot)$ are **similar**.

Note: 1. The number of trials N that a successful pair (Y, U) is found has a **geometric** distribution with parameter

$$\begin{aligned} p &= P(U < \frac{p_Y}{cq_Y}) = \sum_{y \in I} P(U < \frac{p_Y}{cq_Y} | Y = y) P(Y = y) \\ &= \sum_{y \in I} \frac{p_y}{cq_y} q_y = 1/c. \end{aligned}$$

2. The **mean** number of trials to generate an X is $EN = c$.

   So want c as **small** as possible.

3. Take $\mathbf{c} = \max_{\mathbf{x}} \frac{\mathbf{f_X(x)}}{\mathbf{f_Y(x)}} \geq \mathbf{1}$.

4. More efficient if $\mathbf{c} \approx \mathbf{1} \Longleftrightarrow f_Y(\cdot)$ and $f_X(\cdot)$ are **similar**.

Note: 1. The number of trials N that a successful pair (Y, U) is found has a **geometric** distribution with parameter

$$\begin{aligned} p &= P(U < \frac{p_Y}{cq_Y}) = \sum_{y \in I} P(U < \frac{p_Y}{cq_Y} | Y = y) P(Y = y) \\ &= \sum_{y \in I} \frac{p_y}{cq_y} q_y = 1/c. \end{aligned}$$

2. The **mean** number of trials to generate an X is $EN = c$.

   So want c as **small** as possible.

3. Take $\mathbf{c} = \max_{\mathbf{x}} \frac{\mathbf{f_X(x)}}{\mathbf{f_Y(x)}} \geq \mathbf{1}$.

4. More efficient if $\mathbf{c} \approx \mathbf{1} \iff f_Y(\cdot)$ and $f_X(\cdot)$ are **similar**.

Note: 1. The number of trials N that a successful pair (Y, U) is found has a **geometric** distribution with parameter

$$
\begin{aligned}
p &= P(U < \frac{p_Y}{cq_Y}) = \sum_{y \in I} P(U < \frac{p_Y}{cq_Y} | Y = y) P(Y = y) \\
&= \sum_{y \in I} \frac{p_y}{cq_y} q_y = 1/c.
\end{aligned}
$$

2. The **mean** number of trials to generate an X is $EN = c$.

   So want c as **small** as possible.

3. Take $\mathbf{c} = \max_{\mathbf{x}} \frac{\mathbf{f_X(x)}}{\mathbf{f_Y(x)}} \geq \mathbf{1}$.
4. More efficient if $\mathbf{c} \approx \mathbf{1} \iff f_Y(\cdot)$ and $f_X(\cdot)$ are **similar**.

Note: 1. The number of trials N that a successful pair (Y, U) is found has a **geometric** distribution with parameter

$$\begin{aligned} p &= P(U < \frac{p_Y}{cq_Y}) = \sum_{y \in I} P(U < \frac{p_Y}{cq_Y} | Y = y) P(Y = y) \\ &= \sum_{y \in I} \frac{p_y}{cq_y} q_y = 1/c. \end{aligned}$$

2. The **mean** number of trials to generate an X is $EN = c$.

   So want c as **small** as possible.

3. Take $\mathbf{c} = \max_{\mathbf{x}} \frac{\mathbf{f_X(x)}}{\mathbf{f_Y(x)}} \geq \mathbf{1}$.
4. More efficient if $\mathbf{c} \approx \mathbf{1} \iff f_Y(\cdot)$ and $f_X(\cdot)$ are **similar**.

Continuous case

Continuous Case: For X with bounded support $[a, b]$ and density f_X, we may consider $Y \sim U(a, b)$, i.e. $f_Y(y) = \frac{1}{b-a}\mathbf{1}_{[a,b]}(y)$ and let $\mathbf{c} = \max_{a \le x \le b} \frac{f_X(x)}{f_Y(x)} = \mathbf{(b-a)M}$, where M = the density at the **mode of** X.

Then $g(x) = f_X(x)/M$, $a \le x \le b$.

Continuous case

Continuous Case: For X with bounded support $[a,b]$ and density f_X, we may consider $Y \sim U(a,b)$, i.e. $f_Y(y) = \frac{1}{b-a}\mathbf{1}_{[a,b]}(y)$ and let $\mathbf{c} = \max_{a\le x\le b} \dfrac{f_X(x)}{f_Y(x)} = (\mathbf{b}-\mathbf{a})\mathbf{M}$, where M = the density at the **mode of** X.

Then $g(x) = f_X(x)/M$, $a \le x \le b$.

Continuous case

<u>**Continuous Case**</u>: For X with bounded support $[a, b]$ and density f_X, we may consider $Y \sim U(a, b)$, i.e. $f_Y(y) = \frac{1}{b-a}\mathbf{1}_{[a,b]}(y)$ and let $\mathbf{c} = \max_{a \le x \le b} \frac{f_X(x)}{f_Y(x)} = \mathbf{(b-a)M}$, where M = the density at the **mode of** X.

Then $g(x) = f_X(x)/M$, $a \le x \le b$.

A-R Algorithm: 1. Generate $U_1, U_2 \sim U(0,1)$.

2. Set $Y = a + U_1(b-a)$.

3. If $U_2 \leq f_X(Y)/M$, set $X = Y$; otherwise, return to Step 1.

Note: $EN = (b-a)M$, so if f_X has high maximum density, $U(a,b)$ is *not* a good choice for the proposal density.

Ex. 1. $X \sim \textbf{Gamma}(\alpha, \beta)$.

i.e. $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, x > 0$, where $\alpha, \beta > 0$.

Recall that $EX = \alpha/\beta$ and the support of X is $(0, \infty)$.

Consider $Y \sim \textbf{Exp}(\lambda)$, $\lambda = ???$

Ex. 1. $X \sim \mathbf{Gamma}(\alpha, \beta)$.

i.e. $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, x > 0$, where $\alpha, \beta > 0$.

Recall that $EX = \alpha/\beta$ and the support of X is $(0, \infty)$.

Consider $Y \sim \mathbf{Exp}(\lambda)$, $\lambda =$???

Ex. 1. $X \sim \mathbf{Gamma}(\alpha, \beta)$.

i.e. $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, x > 0$, where $\alpha, \beta > 0$.

Recall that $EX = \alpha/\beta$ and the support of X is $(0, \infty)$.

Consider $Y \sim \mathbf{Exp}(\lambda)$, $\lambda =$???

Ex. 1. $X \sim \textbf{Gamma}(\alpha, \beta)$ (Cont'd).

For given λ, we need to find

$$c_\lambda = \max_x \frac{f_X(x)}{f_Y(x|\lambda)} = \max_x \frac{K x^{\alpha-1} e^{-\beta x}}{\lambda e^{-\lambda x}}.$$

Hence, we want to maximize

$$h_\lambda(x) = x^{\alpha-1} e^{-(\beta-\lambda)x}, \text{ for given } \lambda.$$

Note that when $0 < \alpha < 1$, $\lim_{x \to 0} h_\lambda(x) = \infty$, so Y can *not* be used.

Ex. 1. $X \sim \mathbf{Gamma}(\alpha, \beta)$ (Cont'd).

For given λ, we need to find

$$c_\lambda = \max_x \frac{f_X(x)}{f_Y(x|\lambda)} = \max_x \frac{K x^{\alpha-1} e^{-\beta x}}{\lambda e^{-\lambda x}}.$$

Hence, we want to maximize

$$h_\lambda(x) = x^{\alpha-1} e^{-(\beta-\lambda)x}, \text{ for given } \lambda.$$

Note that when $0 < \alpha < 1$, $\lim_{x\to 0} h_\lambda(x) = \infty$, so Y can *not* be used.

Ex. 1. $X \sim \textbf{Gamma}(\alpha, \beta)$ (Cont'd).

For given λ, we need to find

$$c_\lambda = \max_x \frac{f_X(x)}{f_Y(x|\lambda)} = \max_x \frac{Kx^{\alpha-1}e^{-\beta x}}{\lambda e^{-\lambda x}}.$$

Hence, we want to maximize

$$h_\lambda(x) = x^{\alpha-1}e^{-(\beta-\lambda)x}, \text{ for given } \lambda.$$

Note that when $0 < \alpha < 1$, $\lim_{x\to 0} h_\lambda(x) = \infty$, so Y can *not* be used.

Ex. 1. $X \sim \mathbf{Gamma}(\alpha, \beta)$ (Cont'd).

Now for $\alpha > 1$,

$\frac{d}{dx} h_\lambda(x) = (\alpha - 1)x^{\alpha-2} e^{-(\beta-\lambda)x} - (\beta - \lambda)x^{\alpha-1} e^{-(\beta-\lambda)x} = 0$ occurs at

$$x = \frac{\alpha - 1}{\beta - \lambda}, \text{ when } \lambda < \beta;$$

and $h_\lambda(x) \uparrow \infty$ when $\lambda \geq \beta$. So

$$c_\lambda = \frac{f_X\left(\frac{\alpha-1}{\beta-\lambda}\right)}{f_Y\left(\frac{\alpha-1}{\beta-\lambda} \middle| \lambda\right)} = K\lambda^{-1}\left(\frac{\alpha - 1}{\beta - \lambda}\right)^{\alpha-1} e^{1-\alpha}, \quad \text{for given } \lambda < \beta.$$

Ex. 1. $X \sim \textbf{Gamma}(\alpha, \beta)$ (Cont'd).

Now for $\alpha > 1$,

$\frac{d}{dx} h_\lambda(x) = (\alpha - 1)x^{\alpha-2} e^{-(\beta-\lambda)x} - (\beta - \lambda)x^{\alpha-1} e^{-(\beta-\lambda)x} = 0$ occurs at

$$x = \frac{\alpha - 1}{\beta - \lambda}, \text{ when } \lambda < \beta;$$

and $h_\lambda(x) \uparrow \infty$ when $\lambda \geq \beta$. So

$$c_\lambda = \frac{f_X(\frac{\alpha-1}{\beta-\lambda})}{f_Y(\frac{\alpha-1}{\beta-\lambda}|\lambda)} = K\lambda^{-1} \left(\frac{\alpha - 1}{\beta - \lambda}\right)^{\alpha-1} e^{1-\alpha}, \quad \text{for given } \lambda < \beta.$$

<u>Ex</u>. 1. $X \sim \mathbf{Gamma}(\alpha, \beta)$ (Cont'd).

Then one should choose λ such that c_λ is as **small** as possible. That is $\lambda(\beta-\lambda)^{\alpha-1}$ to be maximized. Solving

$$(\beta-\lambda)^{\alpha-1} - (\alpha-1)\lambda(\beta-\lambda)^{\alpha-2} = 0$$

yields $\lambda = \beta/\alpha$, $c = (\alpha/\beta)^\alpha K e^{-(\alpha-1)}$. Therefore,

$$g(y) = \left(\frac{e\beta}{\alpha}\right)^{\alpha-1} y^{\alpha-1} e^{-\frac{\beta}{\alpha}(\alpha-1)y}.$$

Ex. 1. $X \sim \mathbf{Gamma}(\alpha, \beta)$ (Cont'd).

Algorithm: 1. Generate $U_1 \sim U(0,1)$, set $Y = -\frac{\alpha}{\beta} \log U_1$.

2. Generate $U \sim U(0,1)$.

3. If $U \leq (\frac{e\beta}{\alpha})^{\alpha-1} Y^{\alpha-1} e^{-\frac{\beta}{\alpha}(\alpha-1)Y}$, set $X = Y$;

otherwise, return to 1. □

Ex. 1. $X \sim \textbf{Gamma}(\alpha, \beta)$ (Cont'd).

Algorithm: 1. Generate $U_1 \sim U(0,1)$, set $Y = -\frac{\alpha}{\beta} \log U_1$.

2. Generate $U \sim U(0,1)$.

3. If $U \leq (\frac{e\beta}{\alpha})^{\alpha-1} Y^{\alpha-1} e^{-\frac{\beta}{\alpha}(\alpha-1)Y}$, set $X = Y$;

otherwise, return to 1. $\square$

Ex. 1. $X \sim \mathbf{Gamma}(\alpha, \beta)$ (Cont'd).

Note: 1. $EX = \alpha/\beta$ and $EY = 1/\lambda$. So try λ such that $EX = EY$.

We get $\lambda = \beta/\alpha$ ($\alpha > 1$) $\Longrightarrow$ **optimal**!

2. $Gamma(n, \beta)$ can be obtained by summing up n i.i.d.

$Exp(\beta)$ r.v.'s, if n is integer.

3. If $Y \sim f_Y(y|\lambda)$, a given parametric family, take

$c = \min_{\lambda} \max_{x} \dfrac{f_X(x)}{f_Y(x|\lambda)}$, in general.

Ex. 1. $X \sim \mathbf{Gamma}(\alpha, \beta)$ (Cont'd).

Note: 1. $EX = \alpha/\beta$ and $EY = 1/\lambda$. So try λ such that $EX = EY$.

We get $\lambda = \beta/\alpha$ ($\alpha > 1$) $\Longrightarrow$ **optimal**!

2. $Gamma(n, \beta)$ can be obtained by summing up n i.i.d. $Exp(\beta)$ r.v.'s, if n is integer.

3. If $Y \sim f_Y(y|\lambda)$, a given parametric family, take

$c = \min_{\lambda} \max_{x} \frac{f_X(x)}{f_Y(x|\lambda)}$, in general.

Ex. 1. $X \sim \mathbf{Gamma}(\alpha, \beta)$ (Cont'd).

Note: 1. $EX = \alpha/\beta$ and $EY = 1/\lambda$. So try λ such that $EX = EY$.

We get $\lambda = \beta/\alpha$ ($\alpha > 1$) $\Longrightarrow$ **optimal**!

2. $Gamma(n, \beta)$ can be obtained by summing up n i.i.d. $Exp(\beta)$ r.v.'s, if n is integer.

3. If $Y \sim f_Y(y|\lambda)$, a given parametric family, take $c = \min_{\lambda} \max_{x} \frac{f_X(x)}{f_Y(x|\lambda)}$, in general.

Ex. 1. $X \sim \mathbf{Gamma}(\alpha, \beta)$ (Cont'd).

Note: 1. $EX = \alpha/\beta$ and $EY = 1/\lambda$. So try λ such that $EX = EY$.

We get $\lambda = \beta/\alpha$ $(\alpha > 1) \Longrightarrow$ **optimal**!

2. $Gamma(n, \beta)$ can be obtained by summing up n i.i.d.

$Exp(\beta)$ r.v.'s, if n is integer.

3. If $Y \sim f_Y(y|\lambda)$, a given parametric family, take

$c = \min_{\lambda} \max_{\mathbf{x}} \frac{f_X(x)}{f_Y(x|\lambda)}$, in general.

Ex. 2. $X \sim \mathbf{Beta}(\alpha, \beta)$.

i.e. $f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, 0 < x < 1$ and $\alpha, \beta > 0$.

Method I: ($\alpha, \beta \geq 1$.) **Acceptance-Rejection**: Use

$f_Y(y) = \alpha y^{\alpha-1}, 0 < y < 1,$ or $f_Y(y) = \beta y^{\beta-1}, 0 < y < 1.$

Ex. 2. $X \sim \mathbf{Beta}(\alpha, \beta)$.

i.e. $f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, 0 < x < 1$ and $\alpha, \beta > 0$.

Method I: $(\alpha, \beta \geq 1.)$ **Acceptance-Rejection**: Use

$$f_Y(y) = \alpha y^{\alpha-1}, 0 < y < 1, \text{ or } f_Y(y) = \beta y^{\beta-1}, 0 < y < 1.$$

Method II: If $X \sim Gamma(\alpha, 1)$ independent of $Y \sim Gamma(\beta, 1)$, then $X/(X+Y) \sim Beta(\alpha, \beta)$.(Exercise.)

Method III: $(\alpha, \beta < 1.)$ Let $Y_1 = U_1^{1/\alpha}$ and $Y_2 = U_2^{1/\beta}$, where $U_1, U_2 \overset{i.i.d.}{\sim} U(0,1)$. Then

if $Y_1 + Y_2 \leq 1, \implies Y_1/(Y_1+Y_2) \sim Beta(\alpha, \beta)$. (Exercise).

Note: The efficiency is $P(Y_1 + Y_2 \leq 1) = \frac{\alpha\beta}{\alpha+\beta}\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.

(Exercise.) □

Method II: If $X \sim Gamma(\alpha, 1)$ independent of $Y \sim Gamma(\beta, 1)$, then $X/(X+Y) \sim Beta(\alpha, \beta)$.(Exercise.)

Method III: $(\alpha, \beta < 1.)$ Let $Y_1 = U_1^{1/\alpha}$ and $Y_2 = U_2^{1/\beta}$, where $U_1, U_2 \overset{i.i.d.}{\sim} U(0,1)$. Then

if $Y_1 + Y_2 \leq 1, \implies Y_1/(Y_1+Y_2) \sim Beta(\alpha, \beta)$. (Exercise).

Note: The efficiency is $P(Y_1 + Y_2 \leq 1) = \frac{\alpha\beta}{\alpha+\beta} \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.

(Exercise.) □

Method II: If $X \sim Gamma(\alpha, 1)$ independent of $Y \sim Gamma(\beta, 1)$, then $X/(X+Y) \sim Beta(\alpha, \beta)$.(Exercise.)

Method III: $(\alpha, \beta < 1.)$ Let $Y_1 = U_1^{1/\alpha}$ and $Y_2 = U_2^{1/\beta}$, where $U_1, U_2 \overset{i.i.d.}{\sim} U(0,1)$. Then

if $Y_1 + Y_2 \leq 1, \implies Y_1/(Y_1+Y_2) \sim Beta(\alpha, \beta)$. (Exercise).

<u>**Note**</u>: The efficiency is $P(Y_1 + Y_2 \leq 1) = \frac{\alpha\beta}{\alpha+\beta}\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.

(Exercise.) □

Ex. 3. $Z \sim \mathbf{N(0,1)}$.

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, -\infty < z < \infty.$$

Need 1) $Y \sim f_Y(y)$, such that $f_Y(y) > 0, \forall y \in (0, \infty)$.

2) $c = \max_x \frac{f_X(x)}{f_Y(y)}$ can be easily found.

Ex. 3. $Z \sim \mathbf{N(0,1)}$.

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, -\infty < z < \infty.$$

Need 1) $Y \sim f_Y(y)$, such that $f_Y(y) > 0, \forall y \in (0, \infty)$.

2) $c = \max_x \frac{f_X(x)}{f_Y(y)}$ can be easily found.

Ex. 3. $Z \sim \mathbf{N(0,1)}$(Cont'd).

Consider $X = |Z| > 0$ with density $f_X(x) = \sqrt{\frac{2}{\pi}} e^{-x^2/2}, x > 0.$

Z is symmetric so $Z = \pm X$ with probability **1/2** of each, if X is determined.

Method I: Take $Y \sim \mathbf{Exp(1)}$ $(EY = 1 \approx EX = \sqrt{2/\pi} \approx 0.8.)$

Then $c = \max_x \sqrt{2/\pi} \frac{e^{-x^2/2}}{e^{-x}} = \sqrt{2/\pi} e^{1/2}$. Thus,

$$g(y) = \frac{f_X(y)}{cf_Y(y)} = e^{-\frac{1}{2}(y-1)^2}.$$

Ex. 3. $Z \sim \mathbf{N(0,1)}$(Cont'd).

Consider $X = |Z| > 0$ with density $f_X(x) = \sqrt{\frac{2}{\pi}} e^{-x^2/2}, x > 0$.

Z is symmetric so $Z = \pm X$ with probability **1/2** of each, if X is determined.

Method I: Take $Y \sim \mathbf{Exp(1)}$ $(EY = 1 \approx EX = \sqrt{2/\pi} \approx 0.8.)$

Then $c = \max_x \sqrt{2/\pi} \frac{e^{-x^2/2}}{e^{-x}} = \sqrt{2/\pi} e^{1/2}$. Thus,

$$g(y) = \frac{f_X(y)}{c f_Y(y)} = e^{-\frac{1}{2}(y-1)^2}.$$

Algorithm: 1. Generate $U \sim U(0,1)$, and $Y \sim Exp(1)$.

2. If $U \leq e^{-\frac{1}{2}(Y-1)^2}$, set $X = Y$; otherwise, return to 1.

3. Generate $U_1 \sim U(0,1)$.

4. **Set $Z = X$ if $U_1 \leq 1/2$; otherwise, $Z = -X$.**

Note: $U \leq e^{-\frac{1}{2}(Y-1)^2} \iff -\log U \geq (Y-1)^2/2$

$\iff Y_1 \geq (Y-1)^2/2$, where $Y_1 \sim Exp(1)$.

Algorithm: 1. Generate $U \sim U(0,1)$, and $Y \sim Exp(1)$.

2. If $U \leq e^{-\frac{1}{2}(Y-1)^2}$, set $X = Y$; otherwise, return to 1.

3. Generate $U_1 \sim U(0,1)$.

4. **Set $Z = X$ if $U_1 \leq 1/2$; otherwise, $Z = -X$.**

Note: $U \leq e^{-\frac{1}{2}(Y-1)^2} \iff -\log U \geq (Y-1)^2/2$

$\iff Y_1 \geq (Y-1)^2/2$, where $Y_1 \sim Exp(1)$.

Method II: Take $Y \sim$ logistic;

$f_Y(y|\theta) = \frac{e^{-y/\theta}}{\theta(1+e^{-y/\theta})^2}, -\infty < y < \infty, \theta > 0.$

Find optimal θ, i.e. $\min_\theta c_\theta = \min_\theta \sup_x f_Z(x)/f_Y(x|\theta)$. $\theta = .626657$.

(Exercise.) $\square$

The Composition Method

Def: f is a **mixture** of g_i's if $f_X(x) = \sum_i p_i g_i(x)$, where g_i are densities and $\sum_i p_i = 1$, $0 \le p_i \le 1$,

i.e. $X \sim g_i$ **with probability** p_i.

Note: One can sample X from g_i with probability p_i if $X \sim g_i$ is inexpensive.

Algorithm: 1. Generate $U \sim U(0,1)$.

2. Set $I = i$, if $\sum_{j=1}^{i-1} p_j \le U < \sum_{j=1}^{i} p_j$, $i = 1, 2, \ldots$.

3. Generate $X \sim g_I$.

The Composition Method

Def: f is a **mixture** of g_i's if $f_X(x) = \sum_i p_i g_i(x)$, where g_i are densities and $\sum_i p_i = 1$, $0 \le p_i \le 1$,

i.e. $X \sim g_i$ **with probability** p_i.

Note: One can sample X from g_i with probability p_i if $X \sim g_i$ is inexpensive.

Algorithm: 1. Generate $U \sim U(0,1)$.

2. Set $I = i$, if $\sum_{j=1}^{i-1} p_j \le U < \sum_{j=1}^{i} p_j$, $i = 1, 2, \ldots$.

3. Generate $X \sim g_I$.

The Composition Method

Def: f is a **mixture** of g_i's if $f_X(x) = \sum_i p_i g_i(x)$, where g_i are densities and $\sum_i p_i = 1$, $0 \le p_i \le 1$,

i.e. $X \sim g_i$ **with probability** p_i.

Note: One can sample X from g_i with probability p_i if $X \sim g_i$ is inexpensive.

Algorithm: 1. Generate $U \sim U(0,1)$.

2. Set $I = i$, if $\sum_{j=1}^{i-1} p_j \le U < \sum_{j=1}^{i} p_j$, $i = 1, 2, \ldots$.

3. Generate $X \sim g_I$.

<u>Ex</u>. $f(x) = \frac{5}{12}[1 + (x-1)^4]$, $0 \le x \le 2$.

Let $g_1(x) = c_1 \cdot 1$, $0 \le x \le 2$ and $g_2 = c_2(x-1)^4$, for $0 \le x \le 2$, be densities of the two components. So $c_1 = 1/2$, $c_2 = 5/2$; and

$$f(x) = \frac{5}{12}[2 \cdot \frac{1}{2} + \frac{2}{5} \cdot \frac{5}{2}(x-1)^4] = \frac{5}{6} g_1(x) + \frac{1}{6} g_2(x).$$

Thus $G_1^{-1}(x) = 2x$ and $G_2^{-1}(x) = 1 + \sqrt[5]{2x-1}$.

Algorithm: 1. Generate $U_1, U_2 \sim U(0,1)$.

2. If $U_1 \le 5/6$, set $X = 2U_2$ (i.e. $X \sim g_1$); otherwise, set $X = 1 + \sqrt[5]{2U_2 - 1}$ (i.e. $X \sim g_2$). □

Ex. $f(x) = \frac{5}{12}[1 + (x-1)^4]$, $0 \le x \le 2$.

Let $g_1(x) = c_1 \cdot 1$, $0 \le x \le 2$ and $g_2 = c_2(x-1)^4$, for $0 \le x \le 2$, be densities of the two components. So $c_1 = 1/2$, $c_2 = 5/2$; and

$$f(x) = \frac{5}{12}[2 \cdot \frac{1}{2} + \frac{2}{5} \cdot \frac{5}{2}(x-1)^4] = \frac{5}{6} g_1(x) + \frac{1}{6} g_2(x).$$

Thus $G_1^{-1}(x) = 2x$ and $G_2^{-1}(x) = 1 + \sqrt[5]{2x-1}$.

Algorithm: 1. Generate $U_1, U_2 \sim U(0,1)$.

2. If $U_1 \le \mathbf{5/6}$, set $X = 2U_2$ (i.e. $X \sim g_1$); otherwise,

set $X = 1 + \sqrt[5]{2U_2 - 1}$ (i.e. $X \sim g_2$). □

Ex. $f(x) = \frac{5}{12}[1 + (x-1)^4]$, $0 \le x \le 2$.

Let $g_1(x) = c_1 \cdot 1$, $0 \le x \le 2$ and $g_2 = c_2(x-1)^4$, for $0 \le x \le 2$, be densities of the two components. So $c_1 = 1/2$, $c_2 = 5/2$; and

$$f(x) = \frac{5}{12}[2 \cdot \frac{1}{2} + \frac{2}{5} \cdot \frac{5}{2}(x-1)^4] = \mathbf{\frac{5}{6}} g_1(x) + \mathbf{\frac{1}{6}} g_2(x).$$

Thus $G_1^{-1}(x) = 2x$ and $G_2^{-1}(x) = 1 + \sqrt[5]{2x-1}$.

Algorithm: 1. Generate $U_1, U_2 \sim U(0,1)$.

2. If $U_1 \le \mathbf{5/6}$, set $X = 2U_2$ (i.e. $X \sim g_1$); otherwise,

set $X = 1 + \sqrt[5]{2U_2 - 1}$ (i.e. $X \sim g_2$).

Ex. $f(x) = \frac{5}{12}[1 + (x-1)^4]$, $0 \le x \le 2$.

Let $g_1(x) = c_1 \cdot 1$, $0 \le x \le 2$ and $g_2 = c_2(x-1)^4$, for $0 \le x \le 2$, be densities of the two components. So $c_1 = 1/2$, $c_2 = 5/2$; and

$$f(x) = \frac{5}{12}[2 \cdot \frac{1}{2} + \frac{2}{5} \cdot \frac{5}{2}(x-1)^4] = \frac{\mathbf{5}}{\mathbf{6}} g_1(x) + \frac{\mathbf{1}}{\mathbf{6}} g_2(x).$$

Thus $G_1^{-1}(x) = 2x$ and $G_2^{-1}(x) = 1 + \sqrt[5]{2x-1}$.

Algorithm: 1. Generate $U_1, U_2 \sim U(0,1)$.

2. If $U_1 \le \mathbf{5/6}$, set $X = 2U_2$ (i.e. $X \sim g_1$); otherwise, set $X = 1 + \sqrt[5]{2U_2 - 1}$ (i.e. $X \sim g_2$). □

For $X \sim f(x)$ difficult to simulate, if we know another r.v. Y with $P(Y = y_i) = p_i$ and $\sum_i p_i = 1$ such that $f(x|y_i)$ can be easily generated for each i, then by

$$f(x) = \sum_i f(x|y_i)p(Y = y_i) = \sum_i f(x|y_i)p_i,$$

we can **generate Y first, then generate $X \sim f(x|Y)$**.

Continuous Version:

$$f_X(x) = \int g(x|y)h_Y(y)dy,$$

where $h_Y(y)$ is a density and $g(x|y)$ is the conditional density of X given $Y = y$.

Hence, **generate $Y \sim h_Y(y)$ first, then generate $X \sim g(x|Y)$**.

Ex. $f_X(x) = n\int_1^\infty y^{-n}\mathbf{e}^{-\mathbf{xy}}dy, x > 0 \overset{?}{=} \int \mathbf{g}(\mathbf{x}|\mathbf{y})h_Y(y)dy.$

Here $y > 1$ and $x > 0$.

Let $\mathbf{g}(\mathbf{x}|\mathbf{y}) = ye^{-xy} \sim \mathbf{Exp}(\mathbf{y})$ and $h(y) = ny^{-n-1}, y > 1$.

(Check $\int_1^\infty h_Y(y)dy = 1$, $\therefore$ pdf.) Hence,

$$f_X(x) = \int_1^\infty (ye^{-xy})(ny^{-(n+1)})dy = \int_1^\infty g(x|y)h_Y(y)dy.$$

Ex. $f_X(x) = n\int_1^\infty y^{-n}\mathbf{e^{-xy}}dy, x > 0 \overset{?}{=} \int \mathbf{g(x|y)}h_Y(y)dy.$

Here $y > 1$ and $x > 0$.

Let $\mathbf{g(x|y)} = ye^{-xy} \sim \mathbf{Exp(y)}$ and $h(y) = ny^{-n-1}, y > 1$.

(Check $\int_1^\infty h_Y(y)dy = 1$, $\therefore$ pdf.) Hence,

$$f_X(x) = \int_1^\infty (ye^{-xy})(ny^{-(n+1)})dy = \int_1^\infty g(x|y)h_Y(y)dy.$$

Ex. $f_X(x) = n\int_1^\infty y^{-n}\mathbf{e^{-xy}}dy, x > 0 \overset{?}{=} \int \mathbf{g(x|y)}h_Y(y)dy.$

Here $y > 1$ and $x > 0$.

Let $\mathbf{g(x|y)} = ye^{-xy} \sim \mathbf{Exp(y)}$ and $h(y) = ny^{-n-1}, y > 1.$

(Check $\int_1^\infty h_Y(y)dy = 1$, $\therefore$ pdf.) Hence,

$$f_X(x) = \int_1^\infty (ye^{-xy})(ny^{-(n+1)})dy = \int_1^\infty g(x|y)h_Y(y)dy.$$

Ex. $f_X(x) = n\int_1^\infty y^{-n}\mathbf{e^{-xy}}dy, x > 0 \overset{?}{=} \int \mathbf{g(x|y)}h_Y(y)dy.$

Here $y > 1$ and $x > 0$.

Let $\mathbf{g(x|y)} = ye^{-xy} \sim \mathbf{Exp(y)}$ and $h(y) = ny^{-n-1}, y > 1$.

(Check $\int_1^\infty h_Y(y)dy = 1$, $\therefore$ pdf.) Hence,

$$f_X(x) = \int_1^\infty (ye^{-xy})(ny^{-(n+1)})dy = \int_1^\infty g(x|y)h_Y(y)dy.$$

Ex. $f_X(x) = n\int_1^\infty y^{-n}\mathbf{e^{-xy}}dy, x > 0 \overset{?}{=} \int \mathbf{g(x|y)}h_Y(y)dy.$

Here $y > 1$ and $x > 0$.

Let $\mathbf{g(x|y)} = ye^{-xy} \sim \mathbf{Exp(y)}$ and $h(y) = ny^{-n-1}, y > 1$.

(Check $\int_1^\infty h_Y(y)dy = 1$, $\therefore$ pdf.) Hence,

$$f_X(x) = \int_1^\infty (ye^{-xy})(ny^{-(n+1)})dy = \int_1^\infty g(x|y)h_Y(y)dy.$$

Ex. $f_X(x) = n\int_1^\infty y^{-n}\mathbf{e}^{-\mathbf{xy}}dy, x > 0 \overset{?}{=} \int \mathbf{g}(\mathbf{x}|\mathbf{y})h_Y(y)dy.$

Here $y > 1$ and $x > 0$.

Let $\mathbf{g}(\mathbf{x}|\mathbf{y}) = ye^{-xy} \sim \mathbf{Exp}(\mathbf{y})$ and $h(y) = ny^{-n-1}, y > 1$.

(Check $\int_1^\infty h_Y(y)dy = 1$, $\therefore$ pdf.) Hence,

$$f_X(x) = \int_1^\infty (ye^{-xy})(ny^{-(n+1)})dy = \int_1^\infty g(x|y)h_Y(y)dy.$$

Algorithm: 1. Generate $U_1, U_2 \sim U(0,1)$.

2. Set $Y = U_1^{-1/n}$ and $X = -\frac{1}{Y} \ln U_2$. $\square$

Ex. **T-distribution**. $X \sim T_1(p; 0, 1)$,

$$f_X(x) = \frac{\Gamma(\frac{p+1}{2})}{(p\pi)^{1/2}\Gamma(p/2)} \frac{1}{(1 + \frac{x^2}{p})^{(p+1)/2}}.$$

Fact: $X \sim T_1(p; \theta, \sigma^2)$ if and only if

$$\mathbf{X|z \sim N(\theta, z\sigma^2)} \quad \text{and} \quad \mathbf{\frac{1}{z} \sim Gamma(\frac{p}{2}, \frac{p}{2})}.$$

i.e.

$$f_X(x) = \int_0^\infty f(x|z) f_Z(z) dz.$$

Ex. **T-distribution**. $X \sim T_1(p; 0, 1)$,

$$f_X(x) = \frac{\Gamma(\frac{p+1}{2})}{(p\pi)^{1/2}\Gamma(p/2)} \frac{1}{(1 + \frac{x^2}{p})^{(p+1)/2}}.$$

Fact: $X \sim T_1(p; \theta, \sigma^2)$ if and only if

$$\mathbf{X|z \sim N(\theta, z\sigma^2)} \quad \text{and} \quad \mathbf{\frac{1}{z} \sim Gamma(\frac{p}{2}, \frac{p}{2})}.$$

i.e.

$$f_X(x) = \int_0^\infty f(x|z) f_Z(z) dz.$$

Generation of Random Vectors

Case 1: $X_1, \ldots, X_n$ independent if and only if

$$f(x_1, \ldots, x_n) = f_1(x_1) \ldots f_n(x_n).$$

STEP 1: Generate $X_i \sim f_i$, $i = 1, \ldots, n$ and set $X_i = F_i^{-1}(U_i)$

STEP 2: Set $\boldsymbol{X} = (X_1, \ldots, X_n)$.

Generation of Random Vectors

Case 1: $X_1, \ldots, X_n$ independent if and only if

$$f(x_1, \ldots, x_n) = f_1(x_1) \ldots f_n(x_n).$$

STEP 1: Generate $X_i \sim f_i$, $i = 1, \ldots, n$ and set $X_i = F_i^{-1}(U_i)$

STEP 2: Set $\boldsymbol{X} = (X_1, \ldots, X_n)$.

Case 2: $X_1, \ldots, X_n$ dependent, then

$$f(x_1, \ldots, x_n) = f_1(x_1) f_2(x_2|x_1) \ldots f_n(x_n|x_1, \ldots, x_{n-1}).$$

STEP 1: Generate $X_1 \sim f_1$.

STEP 3: Generate $X_2 \sim f_2(x_2|X_1)$.

$\vdots$

STEP n: Generate $X_n \sim f_n(x_n|X_1, \ldots, X_{n-1})$.

Case 2: $X_1, \dots, X_n$ dependent, then

$$f(x_1, \dots, x_n) = f_1(x_1) f_2(x_2|x_1) \dots f_n(x_n|x_1, \dots, x_{n-1}).$$

STEP 1: Generate $X_1 \sim f_1$.

STEP 3: Generate $X_2 \sim f_2(x_2|X_1)$.

$\vdots$

STEP n: Generate $X_n \sim f_n(x_n|X_1, \dots, X_{n-1})$.

Case 2: $X_1, \ldots, X_n$ dependent, then

$$f(x_1, \ldots, x_n) = f_1(x_1) f_2(x_2|x_1) \ldots f_n(x_n|x_1, \ldots, x_{n-1}).$$

STEP 1: Generate $X_1 \sim f_1$.

STEP 3: Generate $X_2 \sim f_2(x_2|X_1)$.

$\vdots$

STEP n: Generate $X_n \sim f_n(x_n|X_1, \ldots, X_{n-1})$.

I

. **Inverse Transform Method**: Let

$$\begin{cases} U_1 &= F_1(X_1) \\ U_2 &= F_2(X_2|X_1) \\ &\vdots \\ U_n &= F_n(X_n|X_1,\ldots,X_{n-1}) \end{cases}$$

Sovle for $\boldsymbol{X} = (X_1,\ldots,X_n)$ in terms of $U_1,\ldots,U_n$.

Ex. $f_{X_1,X_2}(x_1,x_2) = 6x_1$ for $x_1, x_2 \geq 0$ and $x_1 + x_2 \leq 1$; and zero, otherwise.

1). Find

$$\left\{\begin{array}{lcl} f_1(x_1) & = & \int_0^{1-x_1} f_{X_1,X_2}(x_1,x_2)dx_2 \\ f_2(x_2|x_1) & = & f_{X_1,X_2}(x_1,x_2)/f_1(x_1) \end{array}\right.$$

$$\Longrightarrow \left\{\begin{array}{lcl} F_1(x_1) & = & \int_0^{x_1} f_1(t)dt \\ F_2(x_2|x_1) & = & \int_0^{x_1} f_2(t|x_1)dt. \end{array}\right.$$

Ex. $f_{X_1,X_2}(x_1,x_2) = 6x_1$ for $x_1, x_2 \geq 0$ and $x_1 + x_2 \leq 1$; and zero, otherwise.

1). Find

$$\begin{cases} f_1(x_1) & = & \int_0^{1-x_1} f_{X_1,X_2}(x_1,x_2)dx_2 \\ f_2(x_2|x_1) & = & f_{X_1,X_2}(x_1,x_2)/f_1(x_1) \end{cases}$$

$$\Longrightarrow \begin{cases} F_1(x_1) & = & \int_0^{x_1} f_1(t)dt \\ F_2(x_2|x_1) & = & \int_0^{x_1} f_2(t|x_1)dt. \end{cases}$$

2). Find $f_2(x_2)$ and $f_1(x_1|x_2) \Longrightarrow F_2(x_2)$ and $F_1(x_1|x_2)$.

$$X_1 = \sqrt{U_2}U_1^{1/3}, \quad X_2 = 1 - U_1^{1/3}, \quad \text{where } U_1, U_2 \overset{i.i.d}{\sim} U(0,1).$$

Easier! $\square$

Note: **No** general rule for the optimal order!

2). Find $f_2(x_2)$ and $f_1(x_1|x_2) \Longrightarrow F_2(x_2)$ and $F_1(x_1|x_2)$.

$$X_1 = \sqrt{U_2} U_1^{1/3}, \quad X_2 = 1 - U_1^{1/3}, \quad \text{where } U_1, U_2 \overset{i.i.d}{\sim} U(0,1).$$

Easier! $\square$

Note: **No** general rule for the optimal order!

II. Acceptance-Rejection Method.

<u>Case 1</u>: $\boldsymbol{X} \sim (X_1, \ldots, X_n) \sim f_{\boldsymbol{X}}(\boldsymbol{x})$, $\boldsymbol{Y} \sim (Y_1, \ldots, Y_n) \sim h_{\boldsymbol{Y}}(\boldsymbol{y})$.

If $f_{\boldsymbol{X}}(\boldsymbol{x}) = c \cdot g(\boldsymbol{x}) h_{\boldsymbol{Y}}(\boldsymbol{x})$, for some $c \geq 1$, $0 < g(\boldsymbol{x}) < 1$, $\forall \boldsymbol{x}$, then

$$f_{\boldsymbol{Y}}(\boldsymbol{x} | U \leq g(\boldsymbol{Y})) = f_{\boldsymbol{X}}(\boldsymbol{x}), \text{ where } U \sim U(0,1), \text{ independent of } \boldsymbol{Y}.$$

Case 2: $\boldsymbol{X} = (X_1, \ldots, X_n) \in G$ uniformly. (i.e. $\boldsymbol{X}$ is uniformly distributed over a region G.)

STEP 1: Generate $\boldsymbol{Y}$ uniformly in Ω, where Ω is an n-dimensional **rectangle** containing G.

STEP 2: If $\boldsymbol{Y} \in G$, set $\boldsymbol{X} = \boldsymbol{Y}$; otherwise, return to Step 1.

Note: Good if $|G|/|\Omega|$ is **large**.

Case 2: $\boldsymbol{X} = (X_1, \ldots, X_n) \in G$ uniformly. (i.e. $\boldsymbol{X}$ is uniformly distributed over a region G.)

STEP 1: Generate $\boldsymbol{Y}$ uniformly in Ω, where Ω is an n-dimensional **rectangle** containing G.

STEP 2: If $\boldsymbol{Y} \in G$, set $\boldsymbol{X} = \boldsymbol{Y}$; otherwise, return to Step 1.

Note: Good if $|G|/|\Omega|$ is **large**.

Case 2: $\boldsymbol{X} = (X_1, \ldots, X_n) \in G$ uniformly. (i.e. $\boldsymbol{X}$ is uniformly distributed over a region G.)

STEP 1: Generate $\boldsymbol{Y}$ uniformly in Ω, where Ω is an n-dimensional **rectangle** containing G.

STEP 2: If $\boldsymbol{Y} \in G$, set $\boldsymbol{X} = \boldsymbol{Y}$; otherwise, return to Step 1.

Note: Good if $|G|/|\Omega|$ is **large**.

III. **Multivariate Transformation Method**.

Idea: If the desired random vector $\boldsymbol{Y}$ is a transformation of other r.v.'s, i.e. $\boldsymbol{Y} = g(X_1, \ldots, X_k)$, where $X_1, \ldots, X_k$ are 'easy' to be generated. Then

STEP 1: Generate X_i, $i = 1, \ldots, k$.

STEP 2: Plug into g to receive $\boldsymbol{Y}$.

Ex. **Normal random variates.**

(1) **Box-Muller Transformation**: If $X, Y \overset{i.i.d.}{\sim} N(0,1)$, then $R = X^2 + Y^2 \sim Exp(1/2)$, $\Theta = \arctan \frac{Y}{X} \sim U(0, 2\pi)$ and R and Θ are independent.

Conversely, if $R \sim Exp(1/2)$ and $\Theta \sim U(0, 2\pi)$ are independent, then $X = \sqrt{R}\cos\Theta$ and $Y = \sqrt{R}\sin\Theta$ are i.i.d. $N(0,1)$.

Proof: Exercise. $\square$

Note: 1. If $U_1, U_2 \overset{i.i.d.}{\sim} U(0,1)$, then $X = \sqrt{-2\log U_1}\cos(2\pi U_2)$ and $Y = \sqrt{-2\log U_1}\sin(2\pi U_2)$ are i.i.d. $N(0,1)$.

2. Time-comsuming to compute sin, cos.

Note: 1. If $U_1, U_2 \overset{i.i.d.}{\sim} U(0,1)$, then $X = \sqrt{-2\log U_1}\cos(2\pi U_2)$ and $Y = \sqrt{-2\log U_1}\sin(2\pi U_2)$ are i.i.d. $N(0,1)$.

2. Time-comsuming to compute sin, cos.

(2) **Polar Method**: If (X, Y) is uniformly distributed over the unit circle, then $R = X^2 + Y^2$ and $\Theta = \arctan\frac{Y}{X}$ are independent with

$$f_{R,\theta}(r, \theta) = \frac{1}{2\pi}, \ 0 < r < 1, 0 < \theta < 2\pi.$$

Proof: Exercise. $\square$

Note: $R \sim U(0, 1)$ and $\Theta \sim U(0, 2\pi)$, **independent**!

(2) **Polar Method**: If (X, Y) is uniformly distributed over the unit circle, then $R = X^2 + Y^2$ and $\Theta = \arctan \frac{Y}{X}$ are independent with

$$f_{R,\theta}(r, \theta) = \frac{1}{2\pi}, \ 0 < r < 1, 0 < \theta < 2\pi.$$

Proof: Exercise. $\square$

Note: $R \sim U(0, 1)$ and $\Theta \sim U(0, 2\pi)$, **independent**!

Thus, if (X, Y) is a point randomly selected within the unit circle, then $R \sim U(0,1)$ and $\Theta \sim U(0, 2\pi)$ independently; moreover, $\cos\Theta = X/\sqrt{R}$ and $\sin\Theta = Y/\sqrt{R}$ both are independent of R.

Algorithm: 1. Generate $U_1, U_2 \sim U(0,1)$.

2. Set $X = 2U_1 - 1$ and $Y = 2U_2 - 1$ ($\sim U(-1,1)$).

3. If $X^2 + Y^2 \leq 1$, set $\mathbf{R} = X^2 + Y^2$ ($\sim \mathbf{U(0,1)}$); otherwise, return to 1.

4. Set $Z_1 = \sqrt{-2\log \mathbf{R}}\frac{\mathbf{X}}{\sqrt{\mathbf{R}}}$ and $Z_2 = \sqrt{-2\log R}\frac{Y}{\sqrt{R}}$.

Thus, if (X, Y) is a point randomly selected within the unit circle, then $R \sim U(0,1)$ and $\Theta \sim U(0, 2\pi)$ independently; moreover, $\cos\Theta = X/\sqrt{R}$ and $\sin\Theta = Y/\sqrt{R}$ both are independent of R.

Algorithm: 1. Generate $U_1, U_2 \sim U(0,1)$.

2. Set $X = 2U_1 - 1$ and $Y = 2U_2 - 1$ ($\sim U(-1,1)$).

3. If $X^2 + Y^2 \leq 1$, set $\mathbf{R} = X^2 + Y^2$ ($\sim \mathbf{U(0,1)}$); otherwise, return to 1.

4. Set $Z_1 = \sqrt{-2\log \mathbf{R}}\dfrac{\mathbf{X}}{\sqrt{\mathbf{R}}}$ and $Z_2 = \sqrt{-2\log R}\dfrac{Y}{\sqrt{R}}$.

Thus, if (X, Y) is a point randomly selected within the unit circle, then $R \sim U(0,1)$ and $\Theta \sim U(0, 2\pi)$ independently; moreover, $\cos\Theta = X/\sqrt{R}$ and $\sin\Theta = Y/\sqrt{R}$ both are independent of R.

Algorithm: 1. Generate $U_1, U_2 \sim U(0,1)$.

2. Set $X = 2U_1 - 1$ and $Y = 2U_2 - 1$ ($\sim U(-1,1)$).

3. If $X^2 + Y^2 \leq 1$, set $\mathbf{R} = X^2 + Y^2$ ($\sim \mathbf{U}(\mathbf{0}, \mathbf{1})$); otherwise, return to 1.

4. Set $Z_1 = \sqrt{-2\log \mathbf{R}}\frac{\mathbf{X}}{\sqrt{\mathbf{R}}}$ and $Z_2 = \sqrt{-2\log R}\frac{Y}{\sqrt{R}}$.

Thus, if (X, Y) is a point randomly selected within the unit circle, then $R \sim U(0,1)$ and $\Theta \sim U(0, 2\pi)$ independently; moreover, $\cos\Theta = X/\sqrt{R}$ and $\sin\Theta = Y/\sqrt{R}$ both are independent of R.

Algorithm: 1. Generate $U_1, U_2 \sim U(0,1)$.

2. Set $X = 2U_1 - 1$ and $Y = 2U_2 - 1$ ($\sim U(-1,1)$).

3. If $X^2 + Y^2 \leq 1$, set $\mathbf{R} = X^2 + Y^2$ ($\sim \mathbf{U(0,1)}$); otherwise, return to 1.

4. Set $Z_1 = \sqrt{-2\log \mathbf{R}}\frac{\mathbf{X}}{\sqrt{\mathbf{R}}}$ and $Z_2 = \sqrt{-2\log R}\frac{Y}{\sqrt{R}}$.

Note: 1. The mean number of iterations

$$= [P(X^2 + Y^2 \leq 1)]^{-1} = 4/\pi \approx 1.273.$$

2. $\mathbf{X} = \sigma\mathbf{Z} + \mu \sim N(\mu, \sigma^2)$ if and only if $Z \sim N(0, 1)$.

3. $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$, where $\Sigma_{p\times p}$ is symmetric, p.s.d., then $\exists\, C_{p\times p}$ such that $\Sigma = CC'$ and $\boldsymbol{X} = C\boldsymbol{Z} + \boldsymbol{\mu}$ if and only if $\boldsymbol{Z} \sim N_p(\mathbf{0}, I_p)$. $\square$

Note: 1. The mean number of iterations

$$= [P(X^2 + Y^2 \leq 1)]^{-1} = 4/\pi \approx 1.273.$$

2. $\mathbf{X} = \sigma \mathbf{Z} + \mu \sim N(\mu, \sigma^2)$ if and only if $Z \sim N(0, 1)$.

3. $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$, where $\Sigma_{p \times p}$ is symmetric, p.s.d., then $\exists\, C_{p \times p}$ such that $\Sigma = CC'$ and $\boldsymbol{X} = C\boldsymbol{Z} + \boldsymbol{\mu}$ if and only if $\boldsymbol{Z} \sim N_p(\mathbf{0}, I_p)$. $\square$

<u>Ex</u>. χ^2**-distribution**.

(1) $Z_1, \ldots, Z_k \overset{i.i.d.}{\sim} N(0,1) \Longrightarrow X = \sum_{i=1}^k Z_i^2 \sim \chi^2_{(k)}$.

(2) $\chi^2_{(k)} \equiv Gamma(\frac{k}{2}, \frac{1}{2})$.

(i) k even, $\chi^2_{(k)} = X_1 + \cdots + X_{k/2} = -2\sum_{i=1}^{k/2} \log U_i$,

$\mathbf{X_i \sim Exp(1/2)}$, $U_i \sim U(0,1)$.

(ii) k odd, $\chi^2_{(k)} = X_1 + \cdots + X_{[k/2]} + Z^2$, $Z \sim N(0,1)$.

Ex. χ^2**-distribution**.

(1) $Z_1, \ldots, Z_k \overset{i.i.d.}{\sim} N(0,1) \Longrightarrow X = \sum_{i=1}^{k} Z_i^2 \sim \chi^2_{(k)}$.

(2) $\chi^2_{(k)} \equiv Gamma(\frac{k}{2}, \frac{1}{2})$.

(i) k even, $\chi^2_{(k)} = X_1 + \cdots + X_{k/2} = -2\sum_{i=1}^{k/2} \log U_i$,

$\mathbf{X_i \sim Exp(1/2)}$, $U_i \sim U(0,1)$.

(ii) k odd, $\chi^2_{(k)} = X_1 + \cdots + X_{[k/2]} + Z^2$, $Z \sim N(0,1)$.

Ex. χ^2-**distribution**.

(1) $Z_1, \ldots, Z_k \overset{i.i.d.}{\sim} N(0,1) \Longrightarrow X = \sum_{i=1}^{k} Z_i^2 \sim \chi^2_{(k)}$.

(2) $\chi^2_{(k)} \equiv Gamma(\frac{k}{2}, \frac{1}{2})$.

(i) k even, $\chi^2_{(k)} = X_1 + \cdots + X_{k/2} = -2\sum_{i=1}^{k/2} \log U_i$,

$\mathbf{X_i \sim Exp(1/2)}$, $U_i \sim U(0,1)$.

(ii) k odd, $\chi^2_{(k)} = X_1 + \cdots + X_{[k/2]} + Z^2$, $Z \sim N(0,1)$.

Ex. χ^2-**distribution**.

(1) $Z_1, \ldots, Z_k \overset{i.i.d.}{\sim} N(0,1) \Longrightarrow X = \sum_{i=1}^k Z_i^2 \sim \chi^2_{(k)}$.

(2) $\chi^2_{(k)} \equiv Gamma(\frac{k}{2}, \frac{1}{2})$.

(i) k even, $\chi^2_{(k)} = X_1 + \cdots + X_{k/2} = -2\sum_{i=1}^{k/2} \log U_i$,

$\mathbf{X_i \sim Exp(1/2)}$, $U_i \sim U(0,1)$.

(ii) k odd, $\chi^2_{(k)} = X_1 + \cdots + X_{[k/2]} + Z^2$, $Z \sim N(0,1)$.

$\square$

Ex. **Multinomial distribution.**

$\boldsymbol{X} = (X_1, \cdots, X_d) \sim Mult(n, d; \pi_1, \cdots, \pi_d)$ with

$$P(X_1 = x_1, \cdots, X_d = x_d) = \frac{n!}{\prod_{j=1}^{d} x_j!} \prod_{j=1}^{d} \pi_j^{x_j}, \ \pi_j, \ x_j \geq 0,$$

$\sum x_j = n$, and $\sum \pi_j = 1$.

Recall: The marginals are **binomials** and the conditional marginals are also binomials.

Idea: Use successive **conditional** marginals and begin with the one with the largest probability.

Ex. **Multinomial distribution**.

$\boldsymbol{X} = (X_1, \cdots, X_d) \sim Mult(n, d; \pi_1, \cdots, \pi_d)$ with

$$P(X_1 = x_1, \cdots, X_d = x_d) = \frac{n!}{\prod_{j=1}^{d} x_j!} \prod_{j=1}^{d} \pi_j^{x_j}, \ \pi_j, \ x_j \geq 0,$$

$\sum x_j = n$, and $\sum \pi_j = 1$.

Recall: The marginals are **binomials** and the conditional marginals are also binomials.

Idea: Use successive **conditional** marginals and begin with the one with the largest probability.

Ex. **Multinomial distribution**.

$\boldsymbol{X} = (X_1, \cdots, X_d) \sim Mult(n, d; \pi_1, \cdots, \pi_d)$ with

$$P(X_1 = x_1, \cdots, X_d = x_d) = \frac{n!}{\prod_{j=1}^{d} x_j!} \prod_{j=1}^{d} \pi_j^{x_j}, \ \pi_j, \ x_j \geq 0,$$

$\sum x_j = n$, and $\sum \pi_j = 1$.

Recall: The marginals are **binomials** and the conditional marginals are also binomials.

Idea: Use successive **conditional** marginals and begin with the one with the largest probability.

Without loss of generality, assume $\pi_1 \geq \pi_2 \geq \cdots \geq \pi_d$.

Algorithm: 1. Generate $X_1 \sim bin(n, \pi_1)$, say $X_1 = x_1$.

2. Generate

$X_2 \sim X_2|X_1 = x_1 \sim bin(\mathbf{n - x_1}, \frac{\pi_2}{1-\pi_1})$, say $X_2 = x_2$.

3. Generate $X_3 \sim X_3|X_1 = x_1, X_2 = x_2$

$\sim bin(\mathbf{n - x_1 - x_2}, \frac{\pi_2}{1-\pi_1-\pi_2})$.

$\vdots$

n. $X_d = n - (x_1 + \cdots + x_{d-1})$. $\square$

Without loss of generality, assume $\pi_1 \geq \pi_2 \geq \cdots \geq \pi_d$.

Algorithm: 1. Generate $X_1 \sim bin(n, \pi_1)$, say $X_1 = x_1$.

2. Generate

$$X_2 \sim X_2 | X_1 = x_1 \sim bin(\mathbf{n - x_1}, \frac{\boldsymbol{\pi_2}}{\mathbf{1-\boldsymbol{\pi_1}}}), \text{ say } X_2 = x_2.$$

3. Generate $X_3 \sim X_3 | X_1 = x_1, X_2 = x_2$

$$\sim bin(\mathbf{n - x_1 - x_2}, \frac{\boldsymbol{\pi_2}}{\mathbf{1-\boldsymbol{\pi_1}-\boldsymbol{\pi_2}}}).$$

$\vdots$

n. $X_d = n - (x_1 + \cdots + x_{d-1})$.

Without loss of generality, assume $\pi_1 \geq \pi_2 \geq \cdots \geq \pi_d$.

Algorithm: 1. Generate $X_1 \sim bin(n, \pi_1)$, say $X_1 = x_1$.

2. Generate

$$X_2 \sim X_2 | X_1 = x_1 \sim bin(\mathbf{n - x_1}, \tfrac{\pi_2}{1-\pi_1}),$$ say $X_2 = x_2$.

3. Generate $X_3 \sim X_3 | X_1 = x_1, X_2 = x_2$

$$\sim bin(\mathbf{n - x_1 - x_2}, \tfrac{\pi_2}{1-\pi_1-\pi_2}).$$

$$\vdots$$

n. $X_d = n - (x_1 + \cdots + x_{d-1})$. $\square$

Ex. **Dirichlet distribution**.

— a multivariate extension of a *Beta* distribution.

$\boldsymbol{X} = (X_1, \cdots, X_{d+1}) \sim Dir(d+1; \alpha_1, \cdots, \alpha_{d+1})$ with

$$f(\boldsymbol{x}) = \frac{\Gamma(\sum_{j=1}^{d+1} \alpha_j)}{\prod_{j=1}^{d+1} \Gamma(\alpha_j)} x_1^{\alpha_1 - 1} x_2^{\alpha_2 - 1} \cdots x_d^{\alpha_d - 1} (1 - x_1 - \cdots - x_d)^{\alpha_{d+1} - 1},$$

$0 \leq x_j \leq 1.$

Ex. **Dirichlet distribution**.

— a multivariate extension of a *Beta* distribution.

$\boldsymbol{X} = (X_1, \cdots, X_{d+1}) \sim Dir(d+1; \alpha_1, \cdots, \alpha_{d+1})$ with

$$f(\boldsymbol{x}) = \frac{\Gamma(\sum_{j=1}^{d+1} \alpha_j)}{\prod_{j=1}^{d+1} \Gamma(\alpha_j)} x_1^{\alpha_1 - 1} x_2^{\alpha_2 - 1} \cdots x_d^{\alpha_d - 1} (1 - x_1 - \cdots - x_d)^{\alpha_{d+1} - 1},$$

$0 \le x_j \le 1.$

Fact: If $Y_1, Y_2, \cdots, Y_{d+1}$ are independent, each of $\mathbf{Gamma}(\alpha_\mathbf{i}, \beta)$, $i = 1, \ldots, d+1$, respectively, then $\boldsymbol{X} = (X_1, \cdots, X_{d+1})$ with

$$\mathbf{X_j} = \frac{\mathbf{Y_j}}{\sum_{\mathbf{k=1}}^{\mathbf{d+1}} \mathbf{Y_k}}, \; j = 1, \ldots, d+1,$$

has a $Dir(d+1; \alpha_1, \cdots, \alpha_{d+1})$ distribution. $\square$