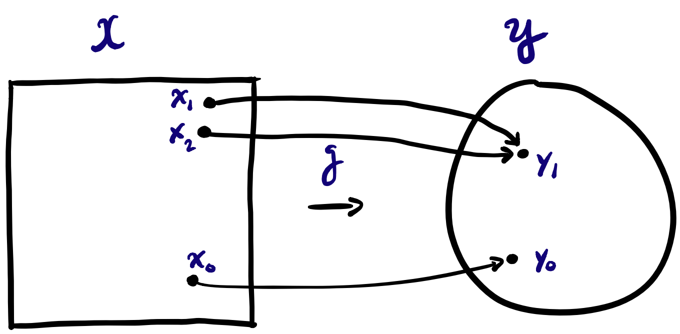
TRANSFORMATIONS OF A RANDOM VARIABLE

Let X be a rv with support

$$\mathcal{X} = \begin{cases} \{x : p_X(x) > 0\} & \text{if } X \text{ is discrete with pmf } p_X \\ \{x : f_X(x) > 0\} & \text{if } X \text{ is continuous with pdf } f_X \end{cases}$$

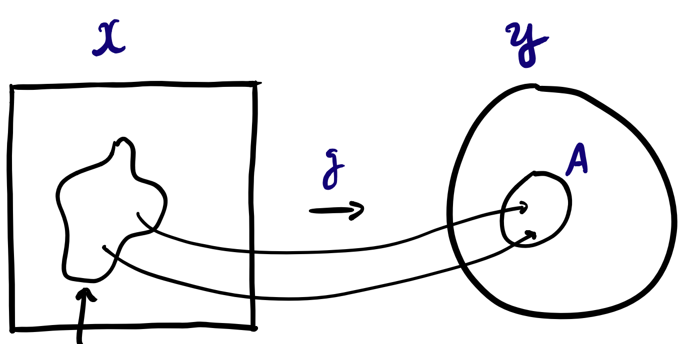
Consider a function $g : \mathcal{X} \to \mathcal{Y}$ taking values in $\mathcal{X}$ and returning values in some set $\mathcal{Y}$, and define the rv Y to be $Y = g(X)$.

Illustration:



Each point in $\mathcal{X}$ maps to a single point in $\mathcal{Y}$.

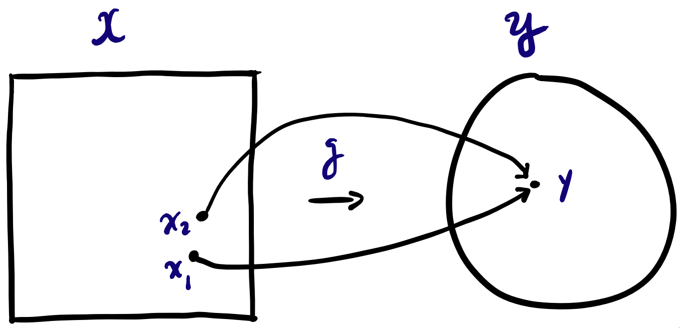
Different points in $\mathcal{X}$ may map to the same point in $\mathcal{Y}$.



For any set $A \subset \mathcal{Y}$, define the inverse mapping g^{-1}, which takes subsets of $\mathcal{Y}$ and returns subsets of $\mathcal{X}$, as

$$g^{-1}(A) = \{x \in \mathcal{X} : g(x) \in A\}$$

$$g^{-1}(A) = \underbrace{\{x \in \mathcal{X} : g(x) \in A\}}_{\text{"all the points in } \mathcal{X} \text{ which } g \text{ takes into } A\text{"}}$$



For a single point $\{y\} \subset \mathcal{Y}$, we write

$$g^{-1}(y) = \{x \in \mathcal{X} : g(x) = y\}$$
$$\text{"} \; g^{-1}(\{y\})$$

$$g^{-1}(y) = g^{-1}(\{y\}) = \{x \in \mathcal{X} : g(x) = y\} = \{x_1, x_2\}$$

The points in $\mathcal{X}$ for which g returns the value y.

We find we may make a probability statement about Y by making the appropriate probability statement about X:

We have

$$P(Y \in A) = P(g(X) \in A) = P\left(X \in \{x \in \mathcal{X} : g(x) \in A\}\right) = P(X \in g^{-1}(A))$$

This defines the probability distribution of Y.

$\boxed{X \text{ DISCRETE}}$

(NOT ~~GO~~ DISCREET)

If X is discrete, $\mathcal{X}$ is finite or countable, so the set $\mathcal{Y}$ of values $y = g(x)$, $x \in \mathcal{X}$ is also finite or countable, so $Y = g(X)$ is discrete.

The pmf of Y is obtained as

$$p_Y(y) = P(Y = y) = P\left(X \in \underbrace{g^{-1}(y)}\right) = \sum_{x \in g^{-1}(y)} P(X = x) = \sum_{x \in g^{-1}(y)} p_X(x),$$

may contain a single value or multiple values

so $p_Y(y)$ is the sum of $p_X(x)$ over x such that $g(x) = y$.

Example: Roll a die twice and let X be the first roll minus the second roll. Then the support of X is

$$\mathcal{X} = \{-5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, 6\},$$

over which the pmf of X takes the values

x	-5	-4	-3	-2	-1	0	1	2	3	4	5
$p_X(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$3/36$	$2/36$	$1/36$

Now let $Y = g(X)$, where $g(x) = |x|$. Then

$$\mathcal{Y} = \{0, 1, 2, 3, 4, 5\},$$

over which the pmf of Y takes the values

y	0	1	2	3	4	5
$g^{-1}(y)$	$\{0\}$	$\{-1,1\}$	$\{-2,2\}$	$\{-3,3\}$	$\{-4,4\}$	$\{-5,5\}$
$p_Y(y)$	$6/36$	$10/36$	$8/36$	$6/36$	$4/36$	$2/36$
	$= p_X(0)$	$= p_X(-1)+p_X(1)$	$\cdots$			

Example: Let $X \sim \text{Binomial}(n, p)$ and let $Y = g(X)$, where $g(x) = n - x$. Recall that

$$p_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{if } x \in \{0, 1, \ldots, n\} \\ 0 & \text{otherwise.} \end{cases}$$

The transformation $g(x) = n - x$ gives $\mathcal{Y} = \{0, 1, \ldots, n\}$, and its inverse is defined by

$$g^{-1}(y) = \{x \in \mathcal{X} : \underbrace{n - x = y}_{\text{solve for } x, \text{ getting } x = n-y}\} = n - y$$

So the pmf of Y for $y \in \mathcal{Y}$ is given by

$$
\begin{aligned}
p_Y(y) &= \sum_{x \in g^{-1}(y)} p_X(x) \\
&= p_X(n-y) \\
&= \binom{n}{n-y} p^{n-y} (1-p)^{n-(n-y)} \\
&= \binom{n}{y} (1-p)^y p^{n-y},
\end{aligned}
$$

$$\binom{n}{n-y} = \frac{n!}{(n-y)!(n-(n-y))!} = \frac{n!}{(n-y)!\,y!} = \binom{n}{y}$$

so that $Y \sim \text{Binomial}(n, 1-p)$.

Indeed, if $X = \#$ successes, then $Y = n - X = \#$ failures in n independent Bernoulli trials with success probability p.

$\boxed{X \text{ CONTINUOUS}}$

Let X be continuous with pdf $f_X(x)$. We first consider the cdf of $Y = g(X)$, which is given by

$$
\begin{aligned}
F_Y(y) &= P_Y(Y \le y) \\
&= P_X(g(X) \le y) \\
&= P_X(X \in \{x \in \mathcal{I} : g(x) \le y\}) \\
&= \int_{\{x \in \mathcal{I} : g(x) \le y\}} f_X(x)\, dx .
\end{aligned}
$$

Evaluating the integral involves identifying the integration bounds corresponding to the set $\{x \in \mathcal{I} : g(x) \le y\}$.

Then, if $\frac{d}{dy} F_Y(y)$ is continuous over $y \in \mathcal{Y}$, the pdf f_Y of Y is given by

$$f_Y(y) = \frac{d}{dy} F_Y(y) \quad \text{for} \quad y \in \mathcal{Y}.$$

Obtaining the distribution of Y by carrying out the above integration to get the cdf of Y is sometimes called the "METHOD OF DISTRIBUTION FUNCTIONS."

<u>Example</u>: Let $X \sim f_X(x) = \frac{3}{2}x^2 \mathbb{1}(-1 \leq x \leq 1)$ and let $Y = |X|$. Find the pdf f_Y of Y.

We have $Y = |X|$, so $g(x) = |x|$, giving $\mathcal{Y} = [0,1]$.

First find the cdf $F_Y(y)$ for $y \in [0,1]$:

$$
\begin{aligned}
F_Y(y) &= P_Y(Y \leq y) \\
&= P_X(|X| \leq y) \\
&= \int_{\{x \in [-1,1] : \underbrace{|x| \leq y}_{-y < x < y}\}} \tfrac{3}{2} x^2 \, dx \\
&= \int_{-y}^{y} \tfrac{3}{2} x^2 \, dx \\
&= \tfrac{1}{2} x^3 \Big|_{-y}^{y} \\
&= \frac{y^3 - (-y)^3}{2} \\
&= y^3,
\end{aligned}
$$

so that

$$
F_Y(y) = \begin{cases} 1, & y > 1 \\ y^3, & 0 \leq y \leq 1 \\ 0, & y < 0 \end{cases} \qquad \begin{array}{l} (Y \text{ is always } \leq 1) \\ \\ (Y \text{ is never } < 0). \end{array}
$$

Then, taking $\frac{d}{dy} F_Y(y)$ on $y \in [0,1]$, we get

$$
f_Y(y) = 3y^2 \mathbb{1}(0 \leq y \leq 1).
$$

Example: Let $X \sim \text{Uniform}(0,1)$ and let $Y = -\log X$.

Find the pdf f_Y of Y.

We have $Y = -\log X$, so $g(x) = -\log x$, giving $\mathcal{Y} = (0,\infty)$.

First find the cdf $F_Y(y)$ for $y \in (0,\infty)$:

$$F_Y(y) = \int_{\{x \in (0,1):\, \underbrace{-\log x \le y}_{x \ge e^{-y}}\}} 1 \cdot dx = \int_{e^{-y}}^{1} 1 \cdot dx = 1 - e^{-y}$$

so

$$F_Y(y) = \begin{cases} 1 - e^{-y}, & y > 0 \\ 0, & y \le 0. \end{cases}$$

Taking $\frac{d}{dy} F_Y(y)$ over $y > 0$, we get

$$f_Y(y) = e^{-y}\, \mathbb{1}(y > 0).$$

In some "nice" situations, we have an explicit formula for f_Y.

"Nice" meaning:
(i) the function g is monotone
(ii) g^{-1} has a continuous derivative
(iii) the pdf of X is continuous.

The next theorem gives us the formula.

Theorem: Let X be a continuous rv with support $\mathcal{X}$ and pdf f_X which is continuous on $\mathcal{X}$. Then let $Y = g(X)$, such that $g : \mathcal{X} \to \mathcal{Y}$ is monotone and the derivative of g^{-1} is continuous on $\mathcal{Y}$. Then the pdf f_Y of Y is given by

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|, & y \in \mathcal{Y} \\ 0, & \text{otherwise.} \end{cases} \qquad (◆)$$

Applying this theorem to get the density of Y is sometimes called the "METHOD OF TRANSFORMATIONS".

Proof: Since $g: \mathcal{X} \to \mathcal{Y}$ is monotone over $\mathcal{X}$, for any $u, v \in \mathcal{X}$

$u > v \iff g(u) > g(v)$ or $u > v \iff g(u) < g(v)$.

In this case the transformation is one-to-one, so that the inverse g^{-1} is single-valued, satisfying

$$g^{-1}(y) = x \iff y = g(x).$$

We consider the cdf of Y in the cases

(i) g is monotone increasing:

$$F_Y(y) = P_Y(Y \le y) = P_X(g(X) \le y) = P_X\left(X \le g^{-1}(y)\right) = F_X(g^{-1}(y))$$

(ii) g is monotone decreasing:

$$F_Y(y) = P_Y(Y \le y) = P_X(g(X) \le y) = P_X\left(X \ge g^{-1}(y)\right) = 1 - F_X(g^{-1}(y))$$

We have used here the continuity of X:

$$P_X(X \ge g^{-1}(y)) = 1 - P_X(X < g^{-1}(y)) = 1 - P_X(X \le g^{-1}(y)) = 1 - F_X(g^{-1}(y))$$

Then

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \begin{cases} \dfrac{d}{dy} F_X(g^{-1}(y)), & g \text{ increasing} \\ \dfrac{d}{dy}\left[1 - F_X(g^{-1}(y))\right], & g \text{ decreasing}, \end{cases}$$

where by the chain rule we have

$$\frac{d}{dy} F_X(g^{-1}(y)) = f_X(g^{-1}(y)) \underbrace{\frac{d}{dy} g^{-1}(y)}_{\ge 0, \text{ because } g \text{ increasing} \Rightarrow g^{-1} \text{ increasing}} = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$$

and

$$\frac{d}{dy}\left[1 - F_X(g^{-1}(y))\right] = -f_X(g^{-1}(y)) \underbrace{\frac{d}{dy} g^{-1}(y)}_{\le 0, \text{ because } g \text{ decreasing} \Rightarrow g^{-1} \text{ decreasing}} = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|.$$

Combining the two cases gives the result. $\square$

Example: Let $X \sim f_X(x) = \frac{1}{\lambda} e^{-x/\lambda} \mathbb{1}(x>0)$. Find the pdf of $Y = \sqrt{X}$.

We have $g(x) = \sqrt{x}$, which is monotone and gives $\mathcal{Y} = (0,\infty)$.

Also $g^{-1}(y) = y^2$ (solve $y = \sqrt{x}$ for x), for which

$$\frac{d}{dy} g^{-1}(y) = \frac{d}{dy} y^2 = 2y,$$

which is continuous over $\mathcal{Y} = (0,\infty)$.

So applying ($\blacklozenge$) gives

$$f_Y(y) = \begin{cases} \frac{1}{\lambda} e^{-\frac{y^2}{\lambda}} |2y|, & y>0 \\ 0, & y \le 0 \end{cases}$$

$$= (2/\lambda)\, y\, e^{-y^2/\lambda} \mathbb{1}(y>0).$$

Example: Let $X \sim f_X(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} \mathbb{1}(x>0)$. Find the pdf of $Y = X^{-1}$.

We have $g(x) = 1/x$ is monotone on $\mathcal{X} = (0,\infty)$ and gives $\mathcal{Y} = (0,\infty)$.

Also $g^{-1}(y) = 1/y$ (solve $y = 1/x$ for x), for which

$$\frac{d}{dy} g^{-1}(y) = -\frac{1}{y^2} = \frac{1}{y^2},$$

which is continuous over $\mathcal{Y} = (0,\infty)$.

So applying ($\blacklozenge$) gives

$$f_Y(y) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} \left(\frac{1}{y}\right)^{\alpha-1} e^{-\frac{1}{y\beta}} \left|\frac{1}{y^2}\right|, & y>0 \\ 0, & y \le 0 \end{cases}$$

$$= \frac{1}{\Gamma(\alpha)\beta^\alpha} \left(\frac{1}{y}\right)^{\alpha+1} e^{-\frac{1}{y\beta}} \mathbb{1}(y>0).$$

Example: Let $X \sim f_X(x) = e^{-x}\,\mathbb{1}(x>0)$ and let $Y = -\log X$. Find the pdf and cdf of Y.

We have $y = g(x) = -\log x$ monotone on $\mathcal{X} = (0,\infty)$ and gives $\mathcal{Y} = (-\infty,\infty)$.

Also $x = g^{-1}(y) = e^{-y}$, for which

$$\frac{d}{dy}\, g^{-1}(y) = -e^{-y},$$

which is continuous over $\mathcal{Y} = (-\infty,\infty)$.

So applying (◆) gives

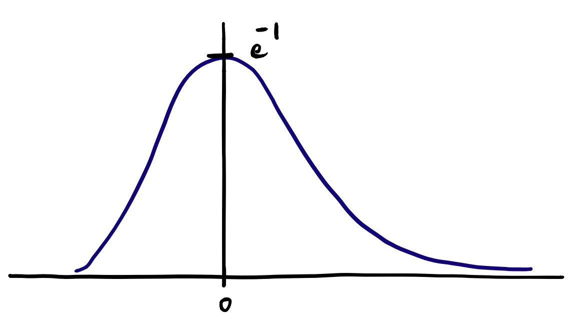
$$f_Y(y) = e^{-e^{-y}}\left|-e^{-y}\right| = e^{-(y+e^{-y})} \quad \text{for all } y \in \mathbb{R}.$$

The cdf of Y is given by

$$F_Y(y) = \int_{-\infty}^{y} e^{-(t+e^{-t})}\,dt = e^{-e^{-t}}\Big|_{-\infty}^{y} = e^{-e^{-y}} \quad \text{for all } y \in \mathbb{R}.$$

The distribution with this cdf is called the Gumbel distribution.

The Gumbel pdf looks like this:

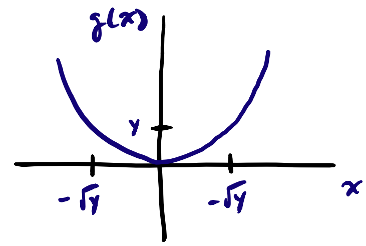


SQUARE TRANSFORMATION OF CONTINUOUS X

Let X be continuous with pdf f_X with support over positive and negative numbers. Find the pdf of $Y = X^2$.

We have $g(x) = x^2$, which is <u>not</u> <u>monotone</u> over $\mathcal{X}$.

For $y > 0$, we have



$$
\begin{aligned}
F_Y(y) &= P_Y(Y \le y) \\
&= P_X(X^2 \le y) \\
&= P_X(-\sqrt{y} \le X \le \sqrt{y}) \\
&= P_X(X \le \sqrt{y}) - P_X(X \le -\sqrt{y}) \\
&= F_X(\sqrt{y}) - F_X(-\sqrt{y}),
\end{aligned}
$$

so that, by the chain rule,

$$
\begin{aligned}
f_Y(y) &= \frac{d}{dy}\left[F_X(\sqrt{y}) - F_X(-\sqrt{y})\right] \\
&= f_X(\sqrt{y})\left(\frac{1}{2}\right)\frac{1}{\sqrt{y}} - f_X(-\sqrt{y})\left(-\frac{1}{2}\right)\frac{1}{\sqrt{y}}.
\end{aligned}
$$

Simplify to get

$$
f_Y(y) = \frac{1}{2}\frac{1}{\sqrt{y}}\left[f_X(\sqrt{y}) + f_X(-\sqrt{y})\right]. \qquad (\spadesuit)
$$

This formula can be used to get the following very important result...

Application: Square of Normal(0,1) is Chi-square(1):

Let $X \sim \text{Normal}(0,1)$. Find the pdf of $Y = X^2$.

We have $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, so for $y > 0$, by $(\spadesuit)$,

$$f_Y(y) = \frac{1}{2} \frac{1}{\sqrt{y}} \left[\frac{1}{\sqrt{2\pi}} e^{-\frac{(\sqrt{y})^2}{2}} + \frac{1}{\sqrt{2\pi}} e^{-\frac{(-\sqrt{y})^2}{2}} \right]$$

$$= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{y}} e^{-\frac{y}{2}}.$$

We find we may write $\quad$ ($\Gamma(\frac{1}{2}) = \sqrt{\pi}$, see lec 8 from STAT 511)

$$f_Y(y) = \frac{1}{\Gamma(\frac{1}{2})\, 2^{1/2}} y^{(1/2 - 1)} e^{-\frac{y}{2}} \mathbb{1}(y > 0),$$

which we recognize as the pdf of the Gamma$(\alpha = 1/2, \beta = 2)$ distribution, which is also called the Chi-square distribution with 1 degree of freedom.

THE PROBABILITY INTEGRAL TRANSFORM

The probability integral transform turns any continuous rv into a Uniform(0,1) rv by passing it through its own cdf — the cdf is the "probability integral."

We present it in the following theorem:

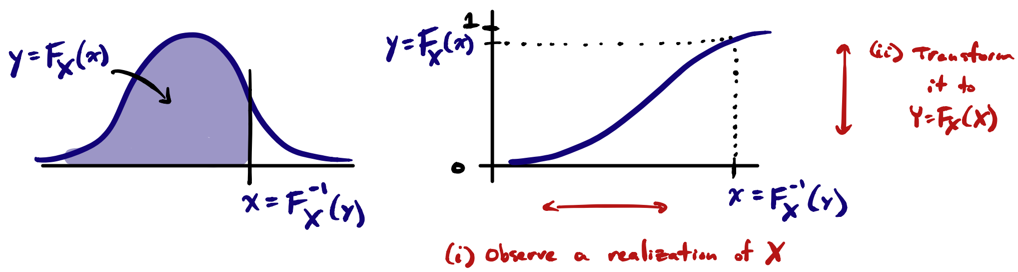
Theorem: Let X be a rv with a continuous cdf F_X and define $Y = F_X(X)$. Then Y has the uniform distribution on $(0,1)$.

Proof: * Suppose F_X is monotone (we don't need this assumption, but it simplifies the proof).

Then $y = F_X(x) \iff F^{-1}(y) = x$.

Also, $\mathcal{Y} = (0,1)$, since $F_X(x) \in (0,1)$ for $x \in \mathcal{X}$.

Illustration:



The cdf of Y for $y \in (0,1)$ is

$$F_Y(y) = P_Y(Y \le y) = P_X(F_X(X) \le y) = P_X(X \le \underbrace{F_X^{-1}(y)}_{y^{th}\text{ quantile of } X}) = y,$$

So

$$F_Y(y) = \begin{cases} 1, & 1 \le y \\ y, & 0 < y < 1 \\ 0, & y \le 0, \end{cases}$$

which is the Uniform $(0,1)$ cdf. $\square$

RANDOM NUMBER GENERATION

We can apply the probability integral transform, in reverse, to generate random numbers.

To generate $X \sim F_X$, do the following:

(i) Generate $U \sim \text{Uniform}(0,1)$

(ii) Set $X = F_X^{-1}(U)$, where $F_X^{-1}(u) = \inf\{x : F_X(x) \geq u\}$

This is the probability integral transform in reverse.

We see that the cdf of $X = F_X^{-1}(U)$ is F_X:

$$P_U\left(F^{-1}(U) \leq x\right) = P_U\left(U \leq F_X(x)\right) = F_X(x).$$

__So:__ We need only to figure out how to generate numbers from the Uniform $(0,1)$ distribution.

It is not easy to make a computer generate random numbers; it is in fact impossible, so we must content ourselves with pseudorandom numbers.

The best, as of now, generator of pseudorandom numbers is the "Mersenne Twister," introduced by Makoto Matsumoto and Takuji Nishimura in 1997. It generates pseudorandom Uniform $(0,1)$ values:

Throughout this paper, bold letters, such as $\mathbf{x}$ and $\mathbf{a}$, denote *word vectors*, which are w-dimensional row vectors over the two-element field $\mathbb{F}_2 = \{0, 1\}$, identified with machine words of size w (with the least significant bit at the right).

The MT algorithm generates a sequence of word vectors, which are considered to be uniform pseudorandom integers between 0 and $2^w - 1$. Dividing by $2^w - 1$, we regard each word vector as a real number in [0,1]. w is some large number

The algorithm is based on the following linear recurrence

$$\mathbf{x}_{k+n} := \mathbf{x}_{k+m} \oplus (\mathbf{x}_k^u | \mathbf{x}_{k+1}^l)A, \quad (k = 0, 1, \cdots). \tag{2.1}$$

Makoto Matsumoto and Takuji Nishimura (1998). Mersenne Twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Trans. Model. Comput. Simul. 8, 1, 3-30.

Example: Generate a realization of $X \sim \text{Exponential}(\lambda)$.

We have

$$F_X(x) = \begin{cases} 1 - e^{-x/\lambda}, & x > 0 \\ 0, & x \leq 0, \end{cases}$$

which has inverse (quantile function)

$$F_X^{-1}(u) = -\lambda \log(1-u) \quad \left(\text{solve } u = 1 - e^{-x/\lambda} \text{ for } x\right)$$

for $u \in (0,1)$.

So generate $U \sim \text{Uniform}(0,1)$ and set $X = -\lambda \log(1-U)$.

TRANSFORMATIONS WITH MOMENT GENERATING FUNCTIONS

If the moment generating function $M_Y(t)$ of a rv Y exists, it tells us what distribution Y has.

In some situations, the easiest way to find the distribution of $Y = g(X)$ is to compute

$$M_Y(t) = \mathbb{E}e^{tY} = \mathbb{E}e^{tg(X)},$$

provided the expectation exists for all t in a neighborhood of zero. This strategy is most advantageous when g is a "shift-and-scale" transformation, that is, when $g(x) = ax + b$ for some $a, b \in \mathbb{R}$.

This is due to the following result (In STAT 511).

Theorem: For any constants a and b, the mgf of $aX + b$ is

$$M_{aX+b}(t) = e^{tb} M_X(at)$$

Proof: $M_{aX+b}(t) = \mathbb{E}\exp[t(aX+b)] = \mathbb{E}\exp[(ta)X]\exp[tb] = e^{tb}M_X(at)$

Finding the distribution of $Y = g(X)$ by finding the mgf of Y is sometimes called the METHOD OF MOMENT GENERATING FUNCTIONS.

__Example:__ Let $X \sim \text{Gamma}(\alpha, \beta)$ and let $Y = X/\beta$.

We know that $M_X(t) = (1-\beta t)^{-\alpha}$, so

$$M_Y(t) = M_{X/\beta}(t) = M_X(t/\beta) = (1-\beta(t/\beta))^{-\alpha} = (1-t)^{-\alpha}.$$

And $(1-t)^{-\alpha}$ is the mgf of the $\text{Gamma}(\alpha, 1)$ dist.

__Example:__ Let $X \sim \text{Normal}(\mu, \sigma^2)$ and let $Z = \frac{X-\mu}{\sigma}$.

We have $M_X(t) = e^{\mu t + \sigma^2 t^2/2}$, so

$$\begin{aligned} M_Z(t) &= M_{\frac{X-\mu}{\sigma}}(t) \\ &= e^{-\frac{\mu}{\sigma}t} M_X(t/\sigma) \\ &= e^{-\frac{\mu}{\sigma}t} e^{\mu\left(\frac{t}{\sigma}\right) + \sigma^2\left(\frac{t}{\sigma}\right)^2/2} \\ &= e^{t^2/2}, \end{aligned}$$

which we recognize as the mgf of the $\text{Normal}(0,1)$ distribution.