MATH 5010 –Foundations of Statistical Theory and Probability

❖ Transformations

Instructor: He Wang
Department of Mathematics
Northeastern University

❖ Outline:

1. **Function of a random variable**

2. **Function of two or more random variables**

❖ Function of a random variable

Let X be a continuous random variable with cdf $F_X(x)$.

Given a function $g(x)$, we can define a random variable $Y := g(X)$.

Goal: Find the distribution for Y.

For any subset A,

$$P(Y \in A) = P(\ g(X) \in A\ )$$

$$= P(X \in g^{-1}(A))$$

Here the inverse mapping g^{-1} is from subset to subset.

If X is a discrete random variable, the pmf of Y is

$$f_Y(y) = P(Y = y) = P(X \in g^{-1}(y))$$

$$= \sum_{x \in g^{-1}(y)} P(X = x)$$

$$= \sum_{x \in g^{-1}(y)} f_X(x)$$

If X is a continuous random variable, the cdf of Y is

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = P(x|\ g(x) \leq y)$$

$$= \int_{\{x|\ g(x) \leq y\}} f_X(x)\, dx$$

Linear Transformation

Example 1:

$X \sim Normal(\mu, \sigma^2)$ $\quad\quad$ $g(x) = ax + b$

$Y = aX + b$ is a Normal distribution $E[Y] = a\mu + b$ and $Var[Y] = a^2\sigma^2$.

Example 2:

$X \sim U[0,1]$ $\quad\quad$ $g(x) = ax + b$

$Y = aX + b$ is a Uniform distribution $U[b, a + b]$

Example(non-linear):

Let X be a discrete random variable with $P_X(x) = \frac{1}{5}$ for x=2,−1,0,1,2.

Let $Y = g(X) = 2|X|$.

Find the range and pmf of Y.

x					
$P(x)$					

$$Range(Y) = \{g(x) \mid x \in Range(X)\}$$

$$P_Y(y) = P(Y = y) = P(g(X) = y) = \sum_{\{x \mid g(x)=y\}} P(X = x)$$

y					
$P(y)$					

Theorem: Suppose $X \in [a, b]$, $g(x)$ is differentiable and $g'(x) > 0$ over $[a, b]$, then the pdf of $Y = g(X)$ is

$$p_Y(y) = \begin{cases} \dfrac{p_X\left(g^{-1}(y)\right)}{g'(g^{-1}(y))} & g(a) \leq y \leq g(b) \\ 0 & \text{otherwise} \end{cases}$$

Proof: The CDF of Y is

$$P(Y \leq y) = P(g(X) \leq y) = P\left(X \leq g^{-1}(y)\right)$$

The pdf is the derivative of the cdf

$$p_Y(y) = \frac{dP\left(X \leq g^{-1}(y)\right)}{dy} = p_X\left(g^{-1}(y)\right) \frac{d}{dy}\, g^{-1}(y) = \frac{p_X\left(g^{-1}(y)\right)}{g'(g^{-1}(y))}$$

Example (linear): Let X be a random variable with pdf $p_X(x)$.

$$g(x) = ax + b$$

The pdf of $Y = aX + b$ is

$$p_Y(y) = \frac{1}{|a|} p_X\left(\frac{y - b}{z}\right)$$

In particular,

$$E[Y] = aE[X] + b$$

$$Var[y] = a^2 Var[X]$$

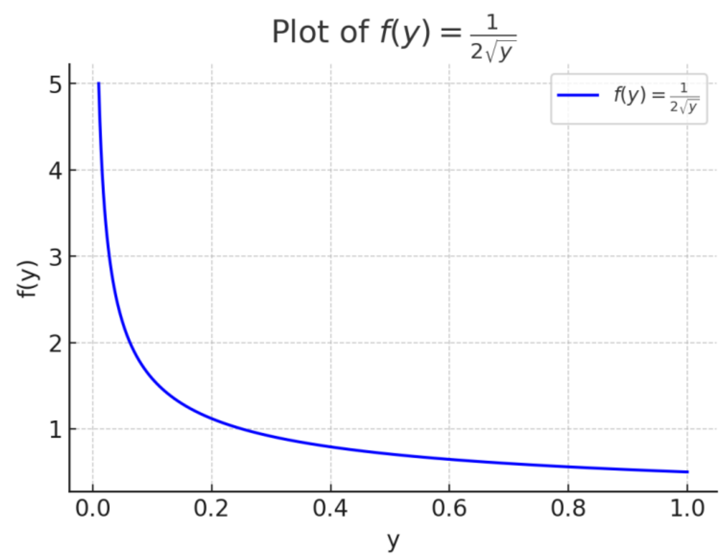
Example (Quadratic)

$$X \sim U[0,1] \qquad g(x) = x^2$$

$$g'(x) = 2x \qquad g^{-1}(y) = \sqrt{y}$$

By theorem, the pdf of $Y = X^2$ is

$$p_Y(y) = \frac{1}{2\sqrt{y}} \mathbb{I}(0 \le y \le 1)$$



Example (Logarithmic)

$X \sim U[0,1]$ $\qquad$ $g(x) = -2\log X$

$g'(x) = -\dfrac{2}{x}$ $\qquad$ $g^{-1}(y) = e^{-\frac{1}{2}y}$

Since $g'(x) \leq 0$, we can not use the theorem directly, but we can modify by the absolute value. The pdf function

$$p_Y(y) = \frac{1}{2} e^{\frac{1}{2}y} \mathbb{I}(0 \leq y)$$

Y is the Exponential distribution with parameter $\lambda = \frac{1}{2}$.

Example (chi-squared distribution)

$$X \sim Normal(0,1) \qquad\qquad g(x) = x^2$$

The CDF of $Y = X^2$ is

$$F_Y(y) = P(Y \leq y) = P(X^2 \leq y) = P(-\sqrt{y} \leq x \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

By chain rule and $\frac{dF_X(x)}{dx} = p_X(x)$, the pdf of Y is

$$p_Y(y) = \frac{dF_Y(y)}{dy} = \frac{1}{2\sqrt{y}}\left(p_X(\sqrt{y}) + p_X(-\sqrt{y})\right) = \frac{1}{\sqrt{y}}\left(p_X(\sqrt{y})\right)$$

By symmetric of normal

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

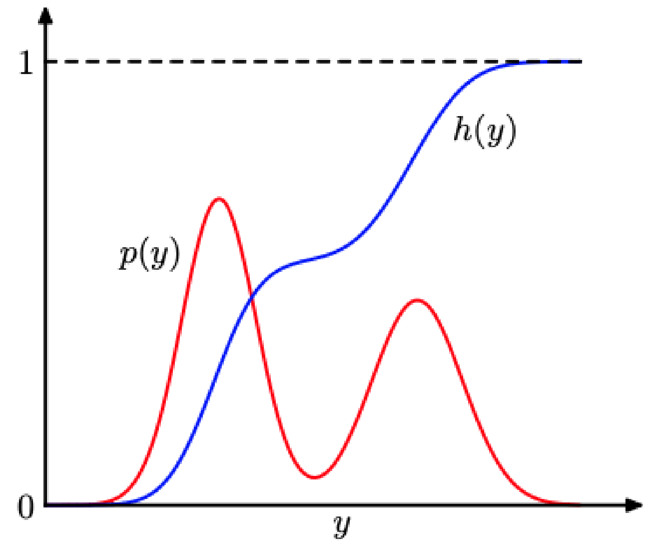
$Y \sim Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$ and also χ_1^2 the chi-squared distribution with degree of freedom 1.

Example: Inverse-Transform

Suppose $X \sim U[0,1]$, So, the pdf is $p(x) = 1$

Suppose the distribution Y has cdf $h(y)$.

Then, $Y = h^{-1}(X)$.



➢ Function of two or more random variables

Let $X_1, X_2, \ldots, X_n$ be n random variables.

Find the PDF of random variable: $Y = g(X_1, \ldots, X_n)$

As before, we investigate the underlying CDF and take the derivative to obtain the corresponding pdf.

Example:

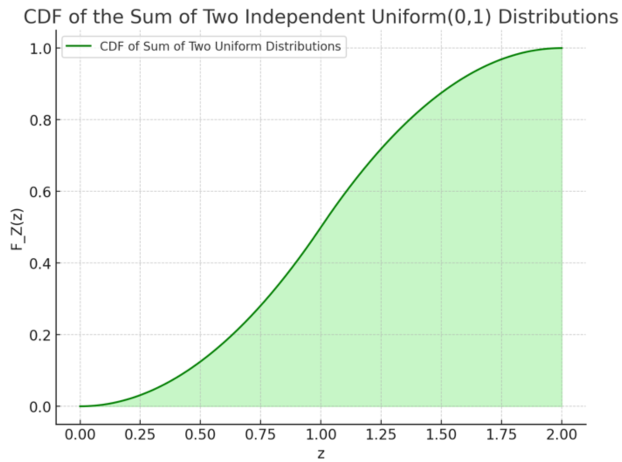
$X_i \sim Bernouli(p)$ are independent

$$X = X_1 + X_2 + \cdots + X_n \sim Binomial(n, p)$$

Example.
Let (X, Y) be a uniform distribution over $[0,1] \times [0,1]$. Then X, Y are independent.

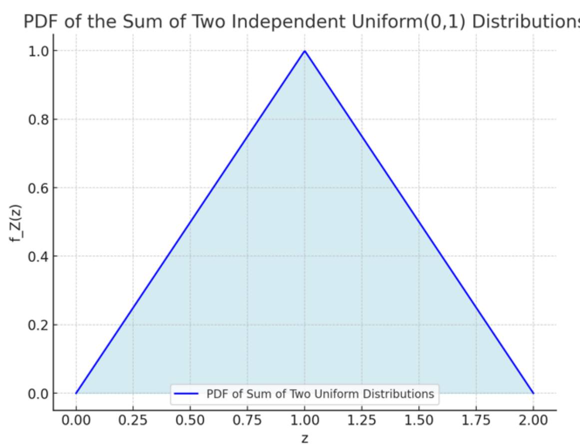
(1) (**Sum.**) The **CDF** of $U = X + Y$

$$F_U(u) = P(U \leq u) = \begin{cases} 0 & u \leq 0 \\ \dfrac{u^2}{2} & 0 \leq u \leq 1 \\ 1 - \dfrac{(2-u)^2}{2} & 1 \leq u \leq 2 \\ 1 & u \geq 2 \end{cases}$$



The **pdf** of $U = X + Y$

$$f(u) = F_U{}'(u) = \begin{cases} u & 0 \leq u \leq 1 \\ 2 - u & 1 \leq u \leq 2 \\ 0 & otherwise \end{cases}$$



(2) **(Max)** The CDF of $V = \max\{X, Y\}$

$$F_V(v) = P(V \leq v) = P(\max\{X, Y\} \leq v) = P(X \leq v, Y \leq v) = P(X \leq v)P(Y \leq v)$$
$$= v^2 \text{ for } v \in [0,1]$$

The pdf of $V = \max\{X, Y\}$

$$p_V(v) = 2v \text{ for } v \in [0,1]$$

(3) **(Min)** The CDF of $W = \min\{X, Y\}$. Consider a reverse event

$$1 - F_W(w) = 1 - P(W \leq w) = P(\min\{X, Y\} \geq w)$$

$$= P(X > w, Y > w) = P(X > w)P(Y > w) = (1 - w)^2 \text{ for } w \in [0,1]$$

The pdf of $V = \min\{X, Y\}$

$$p_V(v) = 2(1 - w) \text{ for } w \in [0,1]$$

Example (Minimum of several uniform distributions).

Let $X_1, \dots, X_n$ be IID uniform distributions on $[0,1]$.

$W = n \cdot \min\{X_1, \dots, X_n\}$

For $w \in [0,1]$, the compliment of the CDF is

$$1 - F_W(w) = 1 - P(W \le w) = P\left(\min\{X_1, \dots, X_n\} \ge \frac{w}{n}\right)$$

$$= P\left(X_1 > \frac{w}{n}, \dots, X_n > \frac{w}{n}\right) = P\left(X_1 > \frac{w}{n}\right) \cdots P\left(X_n > \frac{w}{n}\right)$$

$$= \left(1 - \frac{w}{n}\right)^n$$

$\to e^{-w}$ when $n \to \infty$

So, the pdf is $p_W(w) \to e^{-w}$ when $n \to \infty$

Random Variable W behaves like from an Exponential distribution when n is large

Example: Normal Distributions

(1. **Sum**) Suppose $X \sim Normal\ (\mu_1, \sigma_1^2)$ and $Y \sim Normal\ (\mu_2, \sigma_2^2)$ are independent,

$$X + Y \sim Normal\ (\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

(2. **Sample Mean**) Suppose $X_i \sim Normal\ (\mu\,, \sigma^2)$ are independent for $i = 1, \dots, n$

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \sim Normal\ \left(\mu\,, \frac{\sigma^2}{n}\right)$$

(3. **Chi Square**) Suppose $X_i \sim Normal\ (0,1)$ are independent for $i = 1, \dots, n$

$Z_n = X_1^2 + \cdots + X_n^2$ follows χ^2 distribution with degree of freedom n.

Example (**Exponential distributions**). Consider X, Y are IID from exponential distribution with parameter 1, i.e., $f(x) = e^{-x}$ for $x \geq 0$.

The cdf of $U = X + Y$

$$F_U(u) = P(U \leq u) = P(X + Y \leq u) = \int_{x+y\leq u} e^{-x-y}\, dxdy$$

$$= \int_{x=0}^{u} \int_{y=0}^{u-y} e^{-x-y}\, dydx$$

$$= \int_{x=0}^{u} e^{-x}(1 - e^{x-u})\, dx$$

$$= 1 - e^{u} - ue^{-u}$$

The pdf of $U = X + Y$

$$f_U(u) = ue^{-u} \text{ for } u \geq 0.$$

- Minimum of two exponentials $V = \min(X, Y) \sim Exp(2)$

- Difference of two exponentials $W = |X - Y| = \max(X, Y) - \min(X, Y)$ is

  $Exp(1)$

- Ratio $U = \frac{X}{X+Y}$ is Uniform[0,1]

Transformations of two or more random variables

Example. Let X_1, X_2, X_3, X_4 be random variables with joint pdf.

$$f_{\vec{X}}(\vec{x}) = 24e^{-x_1-x_2-x_3-x_4} \text{ for } 0 < x_1 < x_2 < x_3 < x_4 < \infty$$

Consider the transformation

$$\begin{aligned} Y_1 &= X_1 \\ Y_2 &= X_2 - X_1 \\ Y_3 &= X_3 - X_2 \\ Y_4 &= X_4 - X_3 \end{aligned}$$

Find the joint pdf for Y_1, Y_2, Y_3, Y_4.

Transformations of two or more random variables

Let $X_1, X_2, \dots, X_n$ be n random variables with joint pdf $f_{\vec{X}}(\vec{x})$.

Suppose $Y_i = g_i(\vec{X})$ for $i = 1, \dots, n$

Suppose the **inverse** functions are

$$X_i = h_i(\vec{Y}) \text{ for } i = 1, \dots, n$$

Problem: Find the joint pdf for $Y_1, Y_2, \dots, Y_n$.

The **Jacobian** of the inverse functions is the determinant

$$J = \left|\frac{\partial(x_1, \dots, x_n)}{\partial(y_1, \dots, y_n)}\right| = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \vdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix} = \begin{vmatrix} \frac{\partial h_1}{\partial y_1} & \frac{\partial h_1}{\partial y_2} & \cdots & \frac{\partial h_1}{\partial y_n} \\ \frac{\partial h_2}{\partial y_1} & \frac{\partial h_2}{\partial y_2} & \cdots & \frac{\partial h_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial h_n}{\partial y_1} & \frac{\partial h_n}{\partial y_2} & \vdots & \frac{\partial h}{\partial y_n} \end{vmatrix}$$

Theorem: The joint pdf for $Y_1, Y_2, \dots, Y_n$ is given by

$$p(y_1, \dots, y_n) = p(x_1, \dots, x_n)\left|\frac{\partial(x_1, \dots, x_n)}{\partial(y_1, \dots, y_n)}\right|$$

Remark: The theorem can be more general.

Solution of the example:

The inverse is given by
$$\begin{aligned} X_1 &= Y_1 \\ X_2 &= Y_1 + Y_2 \\ X_3 &= Y_1 + Y_2 + Y_3 \\ X_4 &= Y_1 + Y_2 + Y_3 + Y_4 \end{aligned}$$

The Jacobian of the inverse is

$$J = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{vmatrix} = 1.$$

So, by theorem,

$$f(y_1, \dots, y_n) = f(x_1, \dots, x_n) \left| \frac{\partial(x_1, \dots, x_n)}{\partial(y_1, \dots, y_n)} \right|$$

$$= 24e^{-y_1-(y_1+y_2)-(y_1+y_2+y_3)\ -(y_1+y_2+y_3+y_4)}$$

$$= 24e^{-4y_1-3y_2-2y_3-y_4)}$$

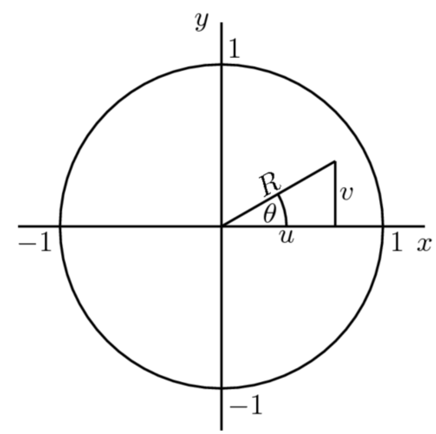
Example: The joint pdf of two independent normal variables (Z_1, Z_2) is

$$f(z_1, z_2) = \frac{1}{2\pi} e^{-\frac{z_1^2 + z_2^2}{2}}$$

Transform the problem into **polar coordinates**:

$$Z_1 = R\cos\Theta$$

$$Z_2 = R\sin\Theta$$



$$R^2 = u^2 + v^2$$
$$\cos\theta = \frac{u}{R}$$
$$\sin\theta = \frac{v}{R}$$

where R is the **radius** and Θ is the **angle** in a 2D plane.

Then $R^2 = Z_1^2 + Z_2^2$, then

$$f(R, \Theta) = \frac{1}{2\pi} e^{-\frac{R^2}{2}} R$$

- Θ is **uniformly** distributed between 0 and 2π.

- R follows the Rayleigh distribution **pdf**:

$$f(R) = Re^{-\frac{R^2}{2}}$$

with cdf $F(R) = 1 - e^{-R^2/2}$ and inverse transform

- $R = \sqrt{-2 \ln U_1}$, where $U_1 \sim \text{Uniform}[0, 1]$

- $\Theta = 2\pi U_2$, where $U_2 \sim \text{Uniform}[0, 1]$

Example: Multi-Normal distribution (Box-Muller method)

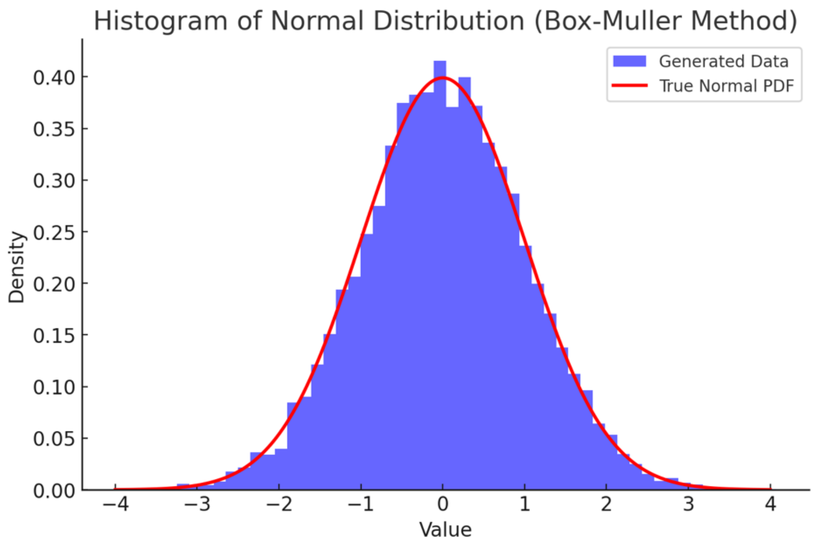
The Box-Muller method converts two independent uniform random variables into two **independent** standard normal random variables.

$U_1 \sim \mathrm{Uniform}[0,1]$ and $U_2 \sim \mathrm{Uniform}[0,1]$ are independent.

Let $$Z_1 = R\cos\Theta = \sqrt{-2\ln U_1}\cos(2\pi U_2)$$

$$Z_2 = R\sin\Theta = \sqrt{-2\ln U_1}\sin(2\pi U_2)$$

Then, Z_1 *and* Z_2 are independent standard normal random variables.



Theorem (Sum): Suppose that X and Y are independent random variables. Let $W = X + Y.$

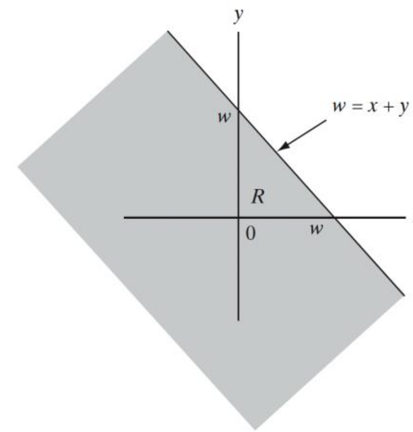
(1) If X *and* Y are *discrete* random variables with pdfs $p_X(x)$ *and* $p_Y(y)$, then

$$p_W(w) = \sum_{\text{all } x} p_X(x) p_Y(w - x)$$

Proof:

$$\begin{aligned} p_W(w) &= P(W = w) = P(X + Y = w) \\ &= P\Big(\bigcup_{\text{all } x} (X = x, Y = w - x)\Big) = \sum_{\text{all } x} P(X = x, Y = w - x) \\ &= \sum_{\text{all } x} P(X = x) P(Y = w - x) \\ &= \sum_{\text{all } x} p_X(x) p_Y(w - x) \end{aligned}$$

(2) If X *and* Y are *continuous* random variables with pdfs $f_X(x)$ *and* $f_Y(y)$, then



$$f_W(w) = \int_{-\infty}^{\infty} f_X(x) f_Y(w-x)\, dx$$

Proof:

CDF:

$$F_w(w) = \int_{-\infty}^{\infty}\int_{-\infty}^{w-x} f_X(x) f_Y(y)\, dy\, dx = \int_{-\infty}^{\infty} f_X(x)\left[\int_{-\infty}^{w-x} f_Y(y)\, dy\right] dx$$

$$= \int_{-\infty}^{\infty} f_X(x) F_Y(w-x)\, dx$$

Pdf:

$$f_W(w) = \frac{d}{dw} F_W(w) = \frac{d}{dw}\int_{-\infty}^{\infty} f_X(x) F_Y(w-x)\, dx = \int_{-\infty}^{\infty} f_X(x)\left[\frac{d}{dw} F_Y(w-x)\right] dx$$

$$= \int_{-\infty}^{\infty} f_X(x) f_Y(w-x)\, dx$$

Theorem (Quotient): Suppose that X *and* Y are independent *continuous* random variables with pdfs $f_X(x)$ *and* $f_Y(y)$. Let $W = Y/X$. Then

$$f_W(w) = \int_{-\infty}^{\infty} |x| \, f_X(x) \, f_Y(wx) \, dx$$

Proof:

$$\begin{aligned}
F_W(w) &= P(Y/X \le w) \\
&= P(Y/X \le w \quad \text{and} \quad X \ge 0) + P(Y/X \le w \quad \text{and} \quad X < 0) \\
&= P(Y \le wX \quad \text{and} \quad X \ge 0) + P(Y \ge wX \quad \text{and} \quad X < 0) \\
&= P(Y \le wX \quad \text{and} \quad X \ge 0) + 1 - P(Y \le wX \quad \text{and} \quad X < 0) \\
&= \int_0^{\infty} \int_{-\infty}^{wx} f_X(x) f_Y(y) \, dy \, dx + 1 - \int_{-\infty}^{0} \int_{-\infty}^{wx} f_X(x) f_Y(y) \, dy \, dx
\end{aligned}$$

$$
\begin{aligned}
f_W(w) = \frac{d}{dw} F_W(w) &= \frac{d}{dw} \int_0^{\infty} \int_{-\infty}^{wx} f_X(x) f_Y(y)\, dy\, dx - \frac{d}{dw} \int_{-\infty}^{0} \int_{-\infty}^{wx} f_X(x) f_Y(y)\, dy\, dx \\
&= \int_0^{\infty} f_X(x) \left(\frac{d}{dw} \int_{-\infty}^{wx} f_Y(y)\, dy \right) dx - \int_{-\infty}^{0} f_X(x) \left(\frac{d}{dw} \int_{-\infty}^{wx} f_Y(y)\, dy \right) dx
\end{aligned}
$$

By the Fundamental Theorem of Calculus and the chain rule

$$
\begin{aligned}
f_W(w) &= \int_0^{\infty} x\, f_X(x) f_Y(wx)\, dx - \int_{-\infty}^{0} x\, f_X(x) f_Y(wx)\, dx \\
&= \int_0^{\infty} x\, f_X(x) f_Y(wx)\, dx + \int_{-\infty}^{0} (-x) f_X(x) f_Y(wx)\, dx \\
&= \int_0^{\infty} |x|\, f_X(x) f_Y(wx)\, dx + \int_{-\infty}^{0} |x|\, f_X(x) f_Y(wx)\, dx \\
&= \int_{-\infty}^{\infty} |x|\, f_X(x) f_Y(wx)\, dx
\end{aligned}
$$

Theorem (Product): Suppose that X *and* Y are independent *continuous* random variables with pdfs $f_X(x)$ *and* $f_Y(y)$. Let $W = XY$.

$$f_W(w) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_X(x) f_Y(w/x)\, dx = \int_{-\infty}^{\infty} \frac{1}{|x|} f_X(w/x) f_Y(x)\, dx$$

The above results can be adapted to situations where X *and* Y are not independent by replacing the product of the marginal pdfs with the joint pdf.

Theorem (Max, Min): Suppose that $Y_1, \dots, Y_n$ are *continuous IID* random variables with pdfs $f_Y(y)$ and cdfs $F_Y(y)$.

- The pdf of $Y_{max} = \max\{Y_1, \dots, Y_n\}$ is

$$f_{Y_{max}}(y) = n[F_Y(y)]^{n-1} f_Y(y)$$

- The pdf of $Y_{min} = \min\{Y_1, \dots, Y_n\}$ is

$$f_{Y_{min}}(y) = n[1 - F_Y(y)]^{n-1} f_Y(y)$$

Proof:

CDF of Y_{max} :

$$F_{Y_{\max}}(y) = P(Y_{\max} \leq y)$$

$$= P(Y_1 \leq y, Y_2 \leq y, \cdots, Y_n \leq y)$$

$$= P(Y_1 \leq y) \cdot P(Y_2 \leq y) \cdots P(Y_n \leq y)$$

$$= [F_Y(y)]^n$$

PDF of Y_{max} :

$$f_{\max}(y) = d/dy[[F_Y(y)]^n] = n[F_Y(y)]^{n-1} f_Y(y)$$

CDF of Y_{max} :

$$F_{Y_{\min}}(y) = P(Y_{\min} \leq y)$$

$$= 1 - P(Y_{\min} > y) = 1 - P(Y_1 > y) \cdot P(Y_2 > y) \cdots P(Y_n > y)$$

$$= 1 - [1 - F_Y(y)]^n$$

PDF of Y_{max} :

$$f_{\min}(y) = d/dy[1 - [1 - F_Y(y)]^n] = n[1 - F_Y(y)]^{n-1} f_Y(y)$$

References:

- **Book 1. [CB] Statistical Inference**, by Casella, George, Berger, Roger L, 2nd edition
- **Book 2. [W]: All of Statistics: Larry Wasserman**
- **Book 3. Introduction to Probability**. C.M. Grinstead and J.L. Snell. American Mathematical Society, 2012
- **Book 4. Introduction to Probability Models**, S. Ross, 12th edition (published by Academic Press).

https://online.stat.psu.edu/stat414/lesson/23/23.1