Lecture 18

(10 October 2024)

Functions of Random Variables

$X : \Omega \to \mathbb{R}$, $\quad Y = g(X)$

For the discrete case,

$$p_Y(y) = \sum_{x : g(x) = y} p_X(x).$$

For the continuous case we follow the two-step procedure outlined below.

1) Calculate the CDF F_Y of Y using

$$F_Y(y) = P(Y \le y) = P(g(X) \le y)$$

in terms of F_X

2) Differentiate the CDF to obtain the PDF of Y:

$$f_Y(y) = \frac{d}{dy} F_Y(y).$$

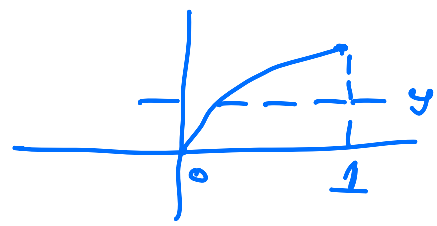
<u>Example</u> (proved earlier).

$$Y = aX + b \quad a \neq 0$$

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right).$$

<u>Example</u>. Let X be a uniformly distributed random variable on $[0, 1]$ and let $Y = \sqrt{X}$.

$$\text{Range}(Y) = [0, 1].$$



For $y < 0$ $\quad F_Y(y) = P(Y \leq y)$

$$= P(\sqrt{X} \leq 0)$$

$$= 0.$$

For $y \geq 1$ $\quad F_Y(y) = P(\sqrt{X} \leq y) = P(X \leq y^2)$

$$= 1$$

$$\therefore f_Y(y) = 0 \quad y \notin \text{Range}(Y).$$

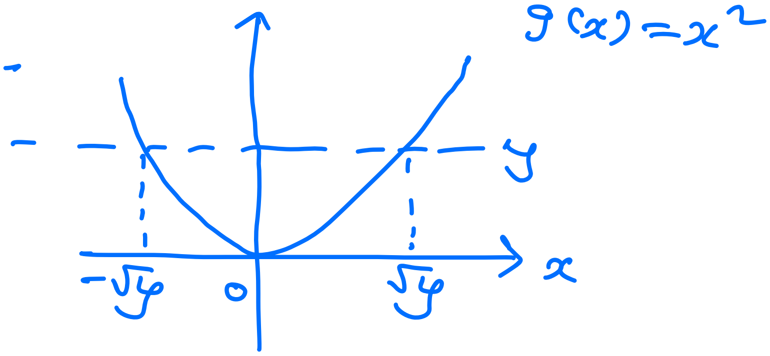
For $0 \leq y < 1$

$$F_Y(y) = P(\sqrt{X} \leq y) = P(X \leq y^2) = y^2$$

$$\Rightarrow f_Y(y) = 2y \quad 0 \leq y < 1$$

Example. $y = x^2 \in [0, \infty)$

$f_Y(y) = 0, \quad y < 0.$



For $y \geq 0$

$$F_Y(y) = P(X^2 \leq y)$$

$$= P(-\sqrt{y} \leq X \leq \sqrt{y})$$

$$= F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

$$\Longrightarrow f_Y(y) = \frac{f_X(\sqrt{y})}{2\sqrt{y}} + \frac{f_X(-\sqrt{y})}{2\sqrt{y}}$$

$x_1 = a(y) = -\sqrt{y} \qquad x_2 = b(y) = \sqrt{y}$

$$F_Y(y) = F_X(x_2) - F_X(x_1)$$

$$= F_X(b(y)) - F_X(a(y))$$

$$\Longrightarrow f_Y(y) = f_X(b(y)) \cdot b'(y) - f_X(a(y)) a'(y)$$

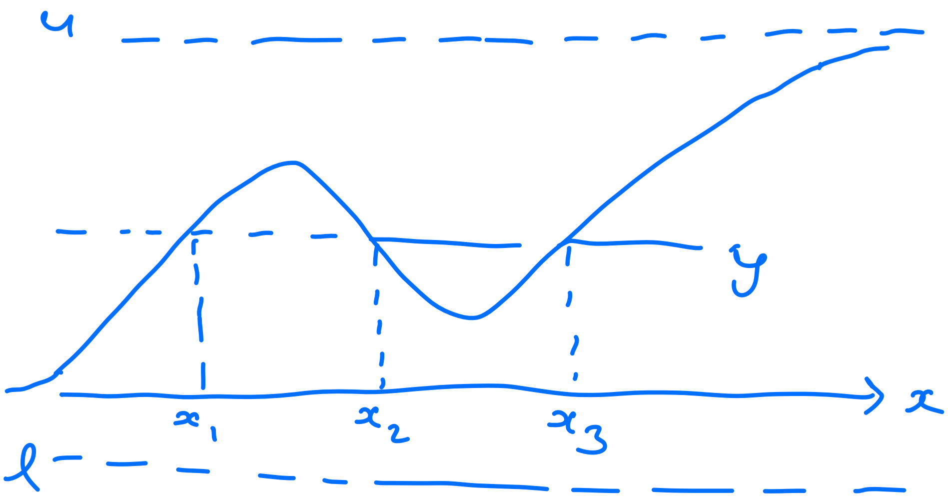
$$= \frac{f_X(b(y))}{|g'(x_2)|} + \frac{f_X(a(y))}{|g'(x_1)|}.$$

We have used $\frac{dx}{dy} \cdot \frac{dy}{dx} = 1$ in the above.

$x_1 = a(y) \Longrightarrow \frac{dx_1}{dy} = a'(y)$

$y = g(x_1)$

$\frac{dy}{dx_1} = g'(x_1)$



$y = g(x)$

$x_1 = a_1(y)$

$x_2 = a_2(y)$

$x_3 = a_3(y)$

$$P(Y \le y) = P(X \le x_1) + P(x_2 < X \le x_3)$$

$$= F_X(x_1) - F_X(x_2) + F_X(x_3)$$

$$F_Y(y) = F_X(a_1(y)) - F_X(a_2(y)) + F_X(a_3(y))$$

$$\Longrightarrow f_Y(y) = f_X(a_1(y))\, a_1'(y) - f_X(a_2(y))\, a_2'(y) + f_X(a_3(y))\, a_3'(y)$$

$$= \frac{f_X(x_1)}{|g'(x_1)|} + \frac{f_X(x_2)}{|g'(x_2)|} + \frac{f_X(x_3)}{|g'(x_3)|},$$

$x_i = a_i(y)$.

$\left(h_i'(y) \cdot g'(x_i) = 1 \text{ as } \frac{dx}{dy} \cdot \frac{dy}{dx} = 1 \right)$

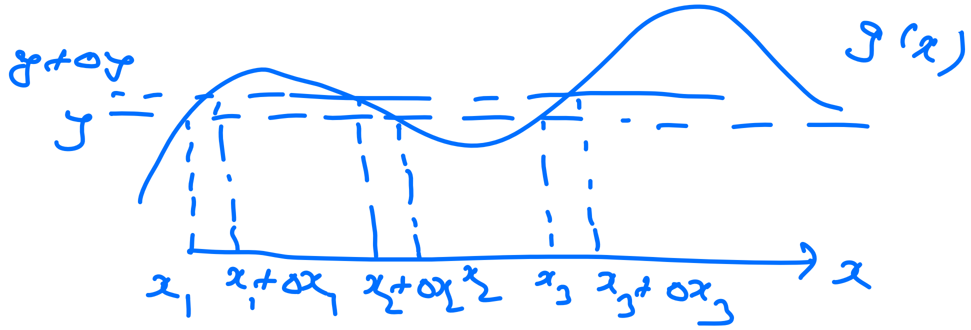
<u>Theorem</u>. Let X and $Y = g(X)$ be continuous random variables. Suppose we can partition $\mathbb{R}$ into intervals $I_1, I_2, \ldots, I_n$ such that $g(x)$ is strictly monotone and differentiable on each I_i, $\forall i \in [1:n]$. Then the PDF of Y is given by

$$f_Y(y) = \sum_{i=1}^{n} \frac{f_X(x_i)}{|g'(x_i)|},$$

where $x_1, x_2, \ldots, x_n$ are real roots to $g(x) = y$ in the respective intervals. In other words let $x_i = h_i(y)$ be the root in interval I_i.

$$f_Y(y) = \sum_{i=1}^{n} \frac{f_X(h_i(y))}{|g'(h_i(y))|}.$$

Proof.



$$P(y < Y \le y+\Delta y) \simeq f_Y(y)\,\Delta y \qquad \boxed{x_i = a_i(y)}$$

To compute the LHS it suffices to find the set of values x such that $y < g(x) \le y+\Delta y$ and the probability that x is in this set.

$$f_Y(y)\,\Delta y = \sum_{i=1}^{n} f_X(x_i)\,|\Delta x_i|$$

$$\Rightarrow f_Y(y) = \sum_{i=1}^{n} f_X(x_i)\left|\frac{\Delta x_i}{\Delta y}\right|$$

As $\Delta y \to 0$, we have

$$f_Y(y) = \sum_{i=1}^{n} f_X(x_i)\underbrace{\frac{dx_i}{dy}}_{a_i'(y)}$$

$$= \sum_{i=1}^{n} \frac{f_X(x_i)}{|g'(a_i(y))|}.$$

<u>Example</u>. $y = g(x) = \dfrac{a}{1+x^2}$.

$g(x) = y \implies x^2+1 = \dfrac{a}{y} \implies x = \pm\sqrt{\dfrac{a}{y}-1}$.

$x_1 = \sqrt{\dfrac{a}{y}-1}$, $x_2 = -\sqrt{\dfrac{a}{y}-1}$.

$$g'(x) = \frac{-a}{(1+x^2)^2}\,2x = \frac{-2ax}{(1+x^2)^2}$$

$$f_Y(y) = \frac{f_X(x_1)}{|g'(x_1)|} + \frac{f_X(x_2)}{|g'(x_2)|}$$

$g'(x_1) = \dfrac{-2y^2}{a}\sqrt{\dfrac{a}{y}-1}$, $g'(x_2) = \dfrac{2y^2}{a}\sqrt{\dfrac{a}{y}-1}$

$$\therefore f_Y(y) = \frac{a}{2y^2\sqrt{\frac{a}{y}-1}}\left(f_X\left(-\sqrt{\frac{a}{y}-1}\right) + f_X\left(\sqrt{\frac{a}{y}-1}\right)\right),$$

for $0 \le y \le a$,

0, otherwise.

Functions of Two Random Variables

$Z = g(X, Y)$, i.e.,

$$Z(\omega) = g(X(\omega), Y(\omega)), \quad \forall \omega \in \Omega.$$

Sum of Independent Random Variables:

$Z = X + Y$, X and Y are independent

$$F_Z(t) = P(Z \leq t)$$

$$= P(X + Y \leq t)$$

$$= \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{t-x} f_{XY}(x, y)\, dy\, dx$$

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

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

$$\Rightarrow F_Z'(t) = \frac{d}{dt} \int_{x=-\infty}^{\infty} f_X(x) F_Y(t-x)\, dx$$

$$= \int_{x=-\infty}^{\infty} f_X(x) \frac{d}{dt} F_Y(t-x)\, dx$$

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

$$\therefore f_Z(z) = \int_{x=-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx .$$