STAT 315: Functions of Random Variables I: CDF Method

Luc Rey-Bellet

University of Massachusetts Amherst

luc@math.umass.edu

April 16, 2025

Function of Random Variables

If we have a function of random variables, say

$$Z = g(Y)$$

or

$$Z = g(Y_1, Y_2)$$

or maybe a sample average

$$Z = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{Y_1 + \cdots Y_n}{n}$$

We know how to compute expectations $E[g(Y)]$ or $E[g(Y_1, Y_2)]$ but often we need more

How do we compute the pdf or cdf of Z?

The CDF method

The CDF method

For $Z = g(Y_1, Y_2, \cdots Y_n)$ compute the CDF of Z by

- Identify the region $Z = g(Y_1, \cdots, Y_n) = z$ in the $y_1, \cdots, y_n$ space.
- Identify the region $Z = g(Y_1, \cdots, Y_n) \leq z$ in the $y_1, \cdots, y_n$ space.
- Compute the integral

$$F(z) = P(Z \leq z) = \int \cdots \int_{g(y_1, \cdots, y_n) \leq z} f(y_1, y_2, \cdots, y_n)\, dy_1 dy_2 \cdots dy_n$$

- Compute the pdf of Z by

$$f(z) = F'(z)\,.$$

Example: Linear transformations

Suppose that the random variable X has PDF $f_X(x)$ and CDF $F_X(x)$. Find the density of $\mathrm{Y} = \mathrm{aX} + \mathrm{b}$?

First take $a > 0$:

$$\text{CDF of } Y: \quad \begin{aligned} F_Y(y) &= P(Y \le y) = P(aX + b \le y) \\ &= P(aX \le y - b) \underbrace{=}_{\text{use } a>0} P\left(X \le \frac{y-b}{a}\right) \\ &= F_X\left(\frac{y-b}{a}\right). \end{aligned}$$

Differentiating we find

$$f_Y(y) = F_Y'(y) = \frac{d}{dy} F_X\left(\frac{y-b}{a}\right) = F_X'\left(\frac{y-b}{a}\right)\frac{1}{a} = \frac{1}{a} f_X\left(\frac{y-b}{a}\right)$$

Next assume take $a < 0$:

$$\text{CDF of } Y: \quad F_Y(y) = P(Y \le y) = P(aX + b \le y) = P(aX \le y - b) \underbrace{=}_{\text{use } a<0} P\left(X \ge \frac{y-b}{a}\right) = 1 - F_X\left(\frac{y-b}{a}\right).$$

Then

$$f_Y(y) = \frac{d}{dy}\left(1 - F_X\left(\frac{y-b}{a}\right)\right) = -F_X'\left(\frac{y-b}{a}\right)\frac{1}{a} = \frac{1}{-a} f_X\left(\frac{y-b}{a}\right)$$

PDF of $Y = aX + b$

$$X \text{ has PDF } f_X(x) \Longrightarrow Y = aX + b \text{ has PDF } f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

Example 1: X is a normal RV with mean μ and variance σ^2

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Then $Y = aX + b$ has density

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right) = \frac{1}{|a|}\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\left(\frac{y-b}{a}-\mu\right)^2}{2\sigma^2}} = \frac{1}{|a|\sigma\sqrt{2\pi}} e^{-\frac{(y-(a\mu+b))^2}{2a^2\sigma^2}}$$

so Y is normal with mean $a\mu + b$ and variance $a^2\sigma^2$.

Example 2: Suppose X is an exponential random variable with parameter β. So $f(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}}$
Then $Y = aX$ (with $a > 0$) has density

$$f_Y(y) = \frac{1}{a} f_X(y/a) = \frac{1}{a}\frac{1}{\beta} e^{-\frac{y/a}{\beta}} = \frac{1}{a\beta} e^{-\frac{y}{a\beta}}$$

so Y is exponential with parameter $a\beta$.

The function $Y = X^2$

Given the PDF $f(x)$ of X we want the pdf of $Y = X^2$. We have

$$\{Y \le y\} \Longleftrightarrow \{X^2 \le y\} \Longleftrightarrow \{-\sqrt{y} \le X \le \sqrt{y}\}$$

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

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

PDF of $Y = X^2$

X has PDF $f_X(x) \Rightarrow Y = X^2$ has PDF $f_Y(y) = \frac{1}{2\sqrt{y}}(f_X(\sqrt{y}) + f_X(-\sqrt{y}))$

The relation between normal RV and χ^2 RV (a.k.a Gamma with $\alpha = \frac{1}{2}$ and $\beta = 2$)

If Z is a standard normal RV, (PDF $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$) then $Y = Z^2$ has pdf

$$\begin{aligned} f_Y(y) &= f_Z(\sqrt{y})\frac{1}{2\sqrt{y}} + f_Z(-\sqrt{y})\frac{1}{2\sqrt{y}} \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{(\sqrt{y})^2}{2}} \frac{1}{2\sqrt{y}} + \frac{1}{\sqrt{2\pi}} e^{-\frac{(-\sqrt{y})^2}{2}} \frac{1}{2\sqrt{y}} = \frac{y^{-1/2} e^{-\frac{y}{2}}}{\sqrt{\pi}} \end{aligned}$$

Compare with the PDF of Gamma RV ($f(y) = \frac{z^{\alpha-1} e^{-\frac{z}{\beta}}}{2^{1/2}\beta^{\alpha}\Gamma(\alpha)}$) gives $\alpha = 1/2$, $\beta = 2$, $\Gamma(1/2) = \sqrt{\pi}$

Normal vs χ^2

Z standard normal RV $\Longrightarrow$ Z^2 is Gamma RV with $\alpha = \frac{1}{2}, \beta = 2$ (=χ^2 RV)

Simulation of Random Variable on a computer

The only source of randomness on a computer is a (pseudo-) number generator which gives a uniform random variable U on $[0, 1]$. How do we generate other random variables?

The inverse CDF method

Suppose Y is a continuous random variable with CDF $F(y)$ and U is uniform on $[0, 1]$. Then we have

$$Y = F^{-1}(U)$$

Proof We check that $Y = F^{-1}(U)$ has the correct CDF, namely F itself.

$$\begin{aligned} P(Y \leq y) &= P\left(F^{-1}(U) \leq y\right) \\ &= P\left(U \leq F(y)\right) \quad (F \text{ invertible}) \\ &= F(y) \qquad\quad (\text{since } P(U \leq a) = a) \end{aligned}$$

Example An exponential random variable with parameter β has CDF $F(y) = 1 - e^{-y/\beta}$. The inverse function F^{-1} is

$$z = 1 - e^{-y/\beta} \Leftrightarrow e^{-y/\beta} = 1 - z \Leftrightarrow y = -\beta \ln(1-z)$$

So the inverse function of $F(y) = 1 - e^{-y/\beta}$ is $F^{-1}(z) = -\beta \ln(1-z)$ and so

$$Y = -\beta \ln(1 - U)$$

has an exponential distribution with parameter β.

Pseudocode to generate exponentials random variables

- Generate a random number U.
- Set $Y = -\beta \ln(1 - U)$

The method of transformation

This is special case of the CDF method which avoids the computation of integral

The transformation method

- The RV Y has pdf $f_Y(y)$.
- For the RV $Z = h(Y)$ we assume that h is an increasing (or decreasing function)

$$\text{if } y_1 < y_2 \text{ then } h(y_1) < h(y_2) \quad (\text{or } h(y_2) < h(y_1))\,.$$

 and so the inverse function $h^{-1}(z)$ exists.
- Then the RV Z has pdf

$$f_Z(z) = f_Y(h^{-1}(z)) \left| \frac{dh^{-1}(z)}{dz} \right|$$

Example for the transformation method

- Suppose Y is an exponential random variable, we want to find the distribution of $Z = \sqrt{Y}$.
- The transformation $h(y) = \sqrt{y}$ is invertible with inverse

$$h(y) = \sqrt{y} = z \iff y = z^2 = h^{-1}(z)$$

- The derivative is

$$\frac{d}{dz} h^{-1}(z) = \frac{d}{dz} z^2 = 2z$$

- The density of Z is

$$f_Z(z) = f_Y(h^{-1}(z)) \left| \frac{dh^{-1}(z)}{dz} \right| = z f_Y(z^2) = \frac{z}{\beta} e^{-z^2/\beta}$$

Z is called a Weibull random variable.

More examples: functions of 2 uniform random variables

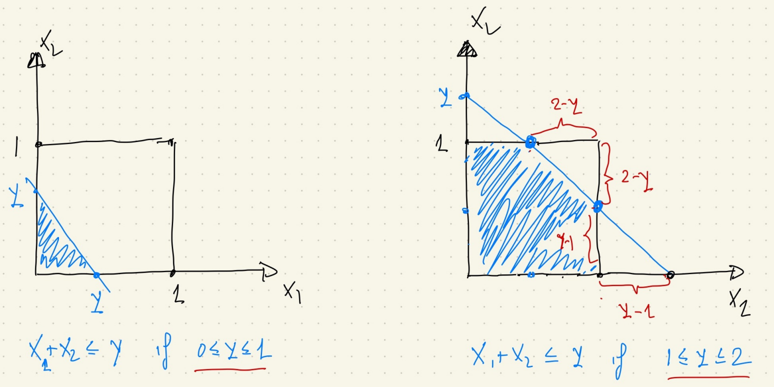
X_1 and X_2 independent and uniform random variables on $[0, 1]$ so the joint PDF is

$$f(x_1, x_2) = 1 \qquad \begin{array}{l} 0 \le x_1 \le 1 \\ 0 \le x_2 \le 1 \end{array}$$

Since the density is constant we can compute

$$P((X_1, X_2) \in A) = \text{area of } A \qquad \text{(maybe avoid integrals)}$$

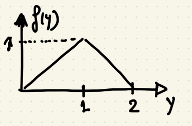
Example 1: Find the PDF of $Y = X_1 + X_2$.
We have $0 \le Y = X_1 + X_2 \le 2$ and we need to compute the CDF $F_Y(y) = P(Y \le y)$ for $0 \le y \le 2$. There are 2 cases.



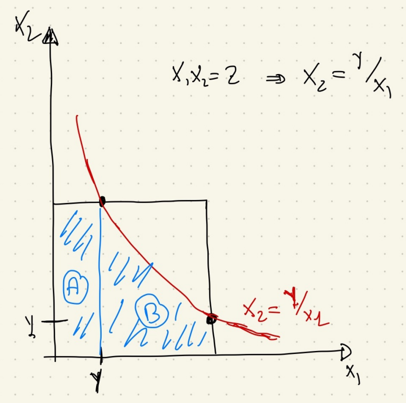
$$P(Y \leq y) = \frac{y^2}{2} \qquad P(Y \leq y) = 1 - \frac{(2-y)^2}{2}$$

So we have

$$\text{PDF} \quad f(y) = F'(y) = \begin{cases} y & \text{if } 0 \leq y \leq 1 \\ 2 - y & \text{if } 1 \leq y \leq 2 \end{cases}$$



Example 2: Find the PDF of $Y = X_1 X_2$. We have $0 \le Y \le 1$ and we need to draw the region $\{x_2 \le \frac{y}{x_1}\}$.



$$
\begin{aligned}
P(Y \le y) &= \text{area A} + \text{area B} \\
\text{area A} &= y \\
\text{area B} &= \int_y^1 \int_0^{y/x_1} 1 dx_2 dx_1 \\
&= \int_y^1 \frac{y}{x_1} dx_1 \\
&= y \ln(x_1)\big|_{x_1=y}^{x_1=1} \\
&= -y\ln(y)
\end{aligned}
$$

$$
\text{So } F(y) = y - y\ln(y)
$$

PDF $\quad f(y) = F'(y) = -\ln(y) \quad 0 \le y \le 1$