CDF method for distribution of products and powers of normal random variables

1 Notation

For a random variable, X, we denote the Probability Density Function (PDF) of X by $f_X(x)$ and the corresponding Cumulative Density Function (CDF) by $F_X(x)$. By definition we have

$$\frac{dF_X(x)}{dx} = f_X(X). \tag{1}$$

Given a pair of random variables, X and Y, we will denote their joint PDF by $f_{X,Y}(x, y)$.

Let X be distributed normally with mean μ and standard deviation σ:

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

$$F_X(x) = \frac{1}{2}\left[1 + erf\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)\right]. \tag{3}$$

2 PDF of square of a normal random variable

Now consider $Z = X^2$. To find an explicit formula for the PDF of Z, we first construct the CDF and then differentiate it to get the PDF.

$$\begin{aligned} F_Z(z) &= \mathbb{P}(Z \le z) \\ &= \mathbb{P}(X^2 \le z) \\ &= \mathbb{P}(-\sqrt{z} \le X \le \sqrt{z}) \\ &= \mathbb{P}(X \le \sqrt{z}) - \mathbb{P}(X \le -\sqrt{z}) \\ &= F_X(\sqrt{z}) - F_X(-\sqrt{z}). \end{aligned}$$

We can now differentiate with respect to z with the chain rule to get the PDF:

$$\begin{aligned} f_Z(z) &= \frac{dF_Z(z)}{dz} \\ &= F'_X(\sqrt{z})\,\frac{d}{dz}\sqrt{z} - F'_X(-\sqrt{z})\,\frac{d}{dz}(-\sqrt{z}) \\ &= \frac{1}{2\,\sqrt{z}}\,\left(f_X(\sqrt{z}) + f_X(-\sqrt{z})\right) \end{aligned}$$

The calculation so far works for any distribution, $f_X(x)$, for the random variable X. For the specific case of a normal distribution we have:

$$f_Z(z) = \frac{1}{2\,\sqrt{2\,\pi\,\sigma^2\,z}}\left(e^{-\frac{(\sqrt{z}-\mu)^2}{2\,\sigma^2}} + e^{-\frac{(\sqrt{z}+\mu)^2}{2\,\sigma^2}}\right). \tag{4}$$

This formula is checked against some empirical data in Fig. 1.

3 PDF of cube of a normal random variable

A similar calculation works for the cube. Let $Z = X^3$. Then

$$\begin{aligned} F_Z(z) &= \mathbb{P}(Z \le z) \\ &= \mathbb{P}(X^3 \le z) \\ &= \mathbb{P}(X \le \mathrm{sgn}(z)\,|z|^{\frac{1}{3}}) \\ &= F_X(\mathrm{sgn}(z)\,|z|^{\frac{1}{3}}). \end{aligned}$$

We can again differentiate with respect to z with the chain rule to get the PDF:

$$\begin{aligned} f_Z(z) &= \frac{dF_Z(z)}{dz} \\ &= F'_X(\mathrm{sgn}(z)\,|z|^{\frac{1}{3}})\,\frac{d}{dz}\left(\mathrm{sgn}(z)\,|z|^{\frac{1}{3}}\right) \\ &= \frac{1}{3\,|z|^{\frac{2}{3}}}\,f_X(\mathrm{sgn}(z)\,|z|^{\frac{1}{3}}) \end{aligned}$$

Specialising to the case of a normal distribution we get

$$f_Z(z) = \frac{1}{3\,\sqrt{2\,\pi\,\sigma^2}\,|z|^{\frac{2}{3}}}\,e^{-\frac{1}{2\,\sigma^2}\left[\mathrm{sgn}(z)|z|^{\frac{1}{3}}-\mu\right]^2}. \tag{5}$$

This formula is checked against some empirical data in Fig. 2.

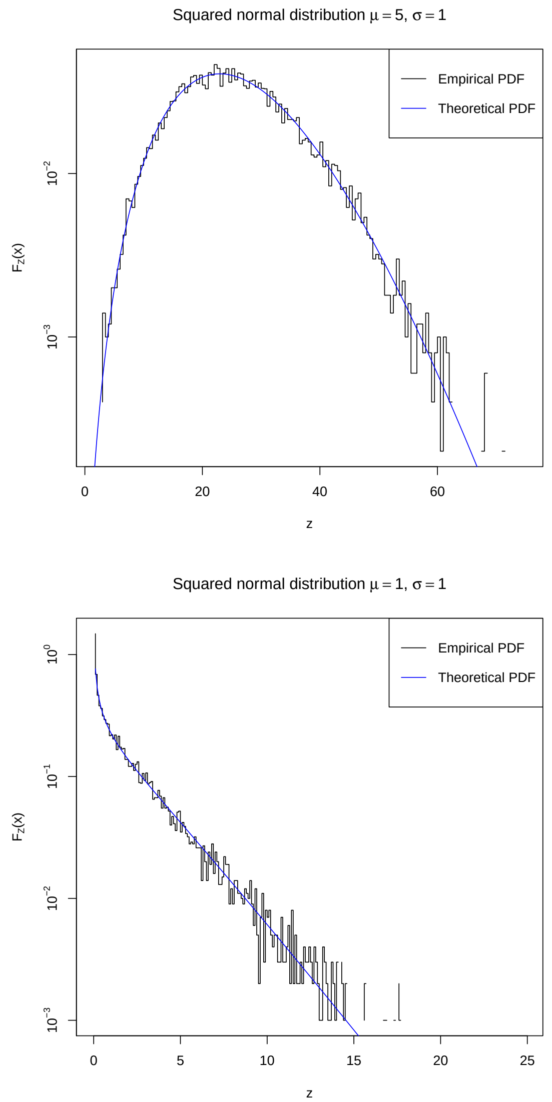
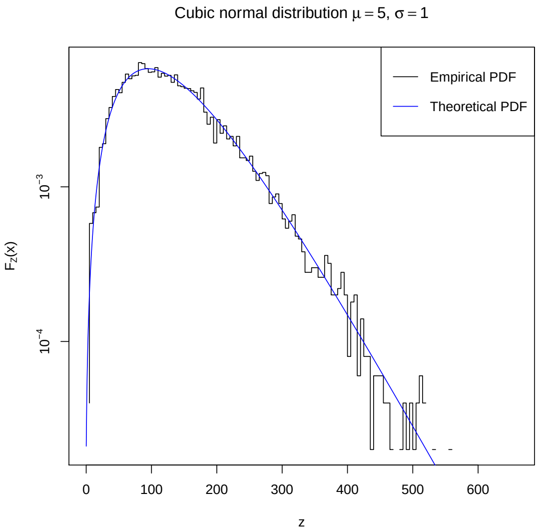


Figure 1: Comparison of Eq. (4) against empirical distribution obtained by sampling and squaring 10000 independent normal random variables.



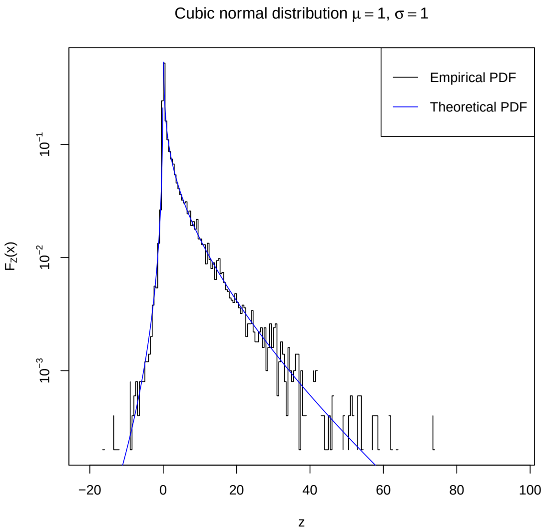


Figure 2: Comparison of Eq. (5) against empirical distribution obtained by sampling and cubing 10000 independent normal random variables.

4 PDF of an arbitrary power of the absolute value of a random variable

The CDF method also gives a general formula for any power, α, of the absolute value of a random variable. The absolute value is required since X^α is generally not defined for $X < 0$. In practice, this is likely to be most useful when we already know that X is a positive random variable. Let $Z = |X|^\alpha$. Then

$$\begin{aligned}
F_Z(z) &= \mathbb{P}(Z \leq z) \\
&= \mathbb{P}(|X|^\alpha \leq z) \\
&= \mathbb{P}(|X| \leq |z|^{\frac{1}{\alpha}}) \\
&= \int_{-\frac{1}{z}}^{0} dx\, f_X(x) + \int_0^{\frac{1}{z}} dx\, f_X(x) \\
&= F_X(a(z)) - F_X(-a(z)),
\end{aligned}$$

where $a(z) = \frac{1}{z}$. We can now get the PDF of z by differentiating:

$$\begin{aligned}
f_Z(z) &= \frac{dF_Z((a(z))}{dz} - \frac{dF_Z((-a(z))}{dz} \\
&= \left(F_X'(a(z)) - F_X'(-a(z))\right) \frac{da(z)}{dz} \\
&= \frac{1}{\alpha}\left[f_X\left(z^{\frac{1}{\alpha}}\right) - f_X\left(-z^{\frac{1}{\alpha}}\right)\right] z^{\frac{1-\alpha}{\alpha}}. \quad (6)
\end{aligned}$$

5 PDF of the product of two random variables

Let the random variables X and Y have joint PDF $F_{X,Y}(x,y)$ and let $Z = XY$. We need not assume that X and Y are independent. The CDF of Z is

$$\begin{aligned}
F_Z(z) &= \mathbb{P}(Z \leq z) \\
&= \mathbb{P}(XY \leq z) \\
&= \mathbb{P}(XY \leq z \,|\, X < 0) + \mathbb{P}(XY \leq z \,|\, X > 0) \\
&= \mathbb{P}\left(Y \geq \frac{z}{X} \,|\, X < 0\right) + \mathbb{P}\left(Y \leq \frac{z}{X} \,|\, X > 0\right) \\
&= \int_{-\infty}^{0} dx \int_{\frac{z}{x}}^{\infty} dy\, f_{X,Y}(x,y) + \int_0^\infty dx \int_0^{\frac{z}{x}} dy\, f_{X,Y}(x,y).
\end{aligned}$$

To get the PDF, we differentiate with respect to z. This is straightforward because the only z-dependence is via the combination $a(z) = \frac{z}{x}$ appearing in the respective lower and upper limits of the inner integrals. An almost trivial application of the chain rule then gives

$$f_Z(z) = \frac{dF_Z((a(z))}{dz} = \frac{dF_Z(a)}{da}\frac{da}{dz}.$$

The derivatives with respect to a come immediately from the Fundamental Theorem of Calculus. We end up with

$$\begin{aligned}
f_Z(z) &= -\int_{-\infty}^{0} \frac{dx}{x} f_{X,Y}(x, \frac{z}{x}) + \int_0^\infty \frac{dx}{x} f_{X,Y}(x, \frac{z}{x}) \\
&= \int_{-\infty}^{\infty} \frac{dx}{|x|} f_{X,Y}(x, \frac{z}{x}). \quad (7)
\end{aligned}$$

This is the formula attributed to Rohatgi [?] in [?]

6 PDF of error in the cube of a quantity with normally distributed error

This is motivated by understanding the distribution of the uncertainty in power, p, given noisy measurements of the streamwise velocity, v, assuming a cubic relationship between the two:

$$p = v^3. \quad (8)$$

Suppose that $v = u + \delta u$ where u is the true streamwise velocity and $\delta u \sim N(0, \sigma^2)$ is a normally distributed measurement error. Then the power is

$$p = u^3 + \delta p, \quad (9)$$

where

$$\delta p = 3u^2(\delta u) + 2u(\delta u)^2 + (\delta u)^3. \quad (10)$$

To find the distribution of δp we need to find the distribution of

$$Z = X^3 + 3aX^2 + 3a^2 X \tag{11}$$

where a is a constant. Proceeding as before:

$$\begin{aligned} F_Z(z) &= \mathbb{P}(Z \leq z) \\ &= \mathbb{P}(X^3 + 3aX^2 + 3a^2 X \leq z) \\ &= \mathbb{P}(X \leq r_z), \end{aligned}$$

where r_z is the root of the equation

$$x^3 + 3ax^2 + 3a^2 x = z. \tag{12}$$

There is a single real root:

$$r_z = -a + (a^3 + z)^{\frac{1}{3}}. \tag{13}$$

Thus we have

$$F_Z(z) = F_X(-a + (a^3 + z)^{\frac{1}{3}}) \tag{14}$$

Now differentiate to get the PDF, $f_Z(z)$:

$$\begin{aligned} f_Z(z) &= \frac{dF_Z(z)}{dz} \\ &= F_X'(r_z)\,\frac{dr_z}{dz}. \end{aligned}$$

Substituting everything in and simplifying, we get

$$f_Z(z) = \frac{1}{3\sqrt{2\,\pi\,\sigma^2}\;|a^3 + z|^{\frac{2}{3}}}\, e^{-\frac{1}{2\,\sigma^2}\left[-a + |a^3 + z|^{\frac{1}{3}}\right]^2}. \tag{15}$$

We can get the same formula from Eq. (5) by defining a shifted variable $W = Z - \mu^3$ representing the difference between the cube of the random variable and the cube of its mean and then substituting $Z = W + \mu$.

7 A better model the wind speed and power distributions

The power is the cube of a the speed and the speed, by definition, cannot be negative. The normal fluctuations assumed in the previous section therefore cannot be generally correct. To fix this, let us assume a 2-dimensional velocity vector with mean direction $(\mu, 0)$ and i.i.d. normal fluctuations in the parallel and transverse directions having mean 0 and standard deviation σ. The is, we model the x and$y$ components of the wind velocity as random variables, U and V, with distributions

$$\begin{aligned} f_U(u) &= \frac{1}{\sqrt{2\pi\sigma^2}}\,\exp\left[\frac{1}{2}\frac{(u-\mu)^2}{\sigma^2}\right] \\ f_V(v) &= \frac{1}{\sqrt{2\pi\sigma^2}}\,\exp\left[\frac{1}{2}\frac{v^2}{\sigma^2}\right]. \end{aligned}$$

The joint distribution is

$$f_{U,V}(u,v) = \frac{1}{2\pi\sigma^2}\exp\left[-\frac{1}{2\sigma^2}\left((u-\mu)^2 + v^2\right)\right].$$

The speed is

$$Z = \sqrt{U^2 + V^2}.$$

Let us now use the CDF method to obtain the PDF of speed.

$$\begin{aligned} F_Z(z) &= \mathbb{P}(Z \leq z) \\ &= \mathbb{P}\left(\sqrt{U^2+V^2} \leq z)\right) \\ &= \mathbb{P}\left(U^2 + V^2 \leq z^2\right) \\ &= \iint\limits_{u^2+v^2\leq z^2} \frac{du\,dv}{2\pi\sigma^2}\exp\left[-\frac{1}{2\sigma^2}\left((u-\mu)^2 + v^2\right)\right]. \end{aligned}$$

Writing this integral in polar coordinates,

$$\begin{aligned} u &= r\cos\theta \\ v &= r\sin\theta, \end{aligned}$$

we get

$$F_Z(z) = \int_0^{2\pi}\frac{d\theta}{2\pi\sigma^2}\int_0^z r\,dr\,\exp\left[-\frac{1}{2\sigma^2}\left((r\cos\theta-\mu)^2 + r^2\sin^2\theta\right)\right].$$

Now differentiating with respect to z, noting that z only enters as the upper limit of the integral over the radial coordinate, we get:

$$\begin{aligned} f_Z(z) &= \int_0^{2\pi}\frac{d\theta}{2\pi\sigma^2}\,z\exp\left[-\frac{1}{2\sigma^2}\left((z\cos\theta-\mu)^2 + z^2\sin^2\theta\right)\right] \\ &= \frac{z}{2\pi\sigma^2}\exp\left[-\frac{z^2+\mu^2}{2\sigma^2}\right]\int_0^{2\pi} d\theta\,\exp\left[\frac{z\mu}{\sigma^2}\cos\theta\right]. \end{aligned}$$

The integral representation of the modified Bessel function, $I_\nu(x)$, is

$$I_\nu(x) = \frac{1}{\pi}\int_0^\pi dt\,\exp\left[x\cos t\right]\cos t.$$

Using this, the integral over θ is seen to be $2\pi I_0\left(\frac{z\mu}{\sigma^2}\right)$. The speed distribution is therefore

$$f_Z(z) = \begin{cases} \frac{z}{\sigma^2}\exp\left[-\frac{z^2+\mu^2}{2\sigma^2}\right] I_0\left(\frac{z\mu}{\sigma^2}\right) & z \geq 0 \\ 0 & z < 0. \end{cases} \tag{16}$$

This PDF is nonzero only for positive values. It looks normal for μ much larger than σ and more like a Weibull for $\mu \sim \sigma$.

This allows it to capture some of the qualitative features of observed wind speed distributions.

We can now substitute Eq. (16) into Eq. (6) with $\alpha = 3$ to get a better model of the PDF of the power:

$$f_Z(z) = \begin{cases} \frac{1}{3\,\sigma^2 z^{\frac{1}{3}}} \exp\left[-\frac{z^{\frac{2}{3}}+\mu^2}{2\sigma^2}\right] I_0\left(\frac{z^{\frac{1}{3}}\mu}{\sigma^2}\right) & z \geq 0 \\ 0 & z < 0. \end{cases}$$

In this model, the power is roughly a stretched exponential distribution with an integrable singularity at 0.

References

[1] VK Rohatgi. *An Introduction to Probability Theory and Mathematical Statistics*. John Wiley & Sons, New York, 1976.

[2] Guolong Cui, Xianxiang Yu, Salvatore Iommelli, and Lingjiang Kong. Exact distribution for the product of two correlated Gaussian random variables. *IEEE Signal Processing Letters*, 23(11):1662–1666, 2016.