Why Is the Sum of Independent Normal Random Variables Normal?

BENNETT EISENBERG
Lehigh University
Bethlehem, PA 18015
BE01@lehigh.edu

ROSEMARY SULLIVAN
West Chester University
West Chester, PA 19383
rsullivan@wcupa.edu

The fact that the sum of independent normal random variables is normal is a widely used result in probability. Two standard proofs are taught, one using convolutions and the other moment generating functions, but neither gives much insight into why the result is true. In this paper we give two additional arguments for why the sum of independent normal random variables should be normal.

The convolution proof

The first standard proof consists of the computation of the convolution of two normal densities to find the density of the sum of the random variables. Throughout this article we assume that our normal random variables have mean 0 since a general normal random variable can be written in the form $\sigma Z + \mu$, where Z is standard normal and μ is a constant. One then finds the convolution of two normal densities to be

$$\int_{-\infty}^{\infty} \frac{\exp\left(-\frac{(y-x)^2}{2\sigma_1^2}\right)}{\sqrt{2\pi}\sigma_1} \frac{\exp\left(-\frac{x^2}{2\sigma_2^2}\right)}{\sqrt{2\pi}\sigma_2} dx = \frac{\exp\left(-\frac{y^2}{2(\sigma_1^2+\sigma_2^2)}\right)}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}}.$$

The computation is messy and not very illuminating even for the case of mean zero random variables.

The moment generating proof

The calculation of convolutions of probability distributions is not easy, so proofs using moment generating functions are often used. One uses the fact that the moment generating function of a sum of independent random variables is the product of the corresponding moment generating functions. Products are easier to compute than convolutions.

We have that the moment generating function of a mean zero normal random variable X with variance σ^2 is

$$M_X(t) = \int_{-\infty}^{\infty} \exp(tx) \frac{\exp\left(-\frac{x^2}{2\sigma^2}\right)}{\sqrt{2\pi}\,\sigma} dx = \exp\left(\frac{t^2\sigma^2}{2}\right).$$

Thus, if X_1 and X_2 are independent, mean zero, normal random variables with variances σ_1^2 and σ_2^2, respectively, then

$$M_{X_1+X_2}(t) = \exp\left(\frac{(\sigma_1^2+\sigma_2^2)t^2}{2}\right).$$

We see that the product of the moment generating functions of normal random variables is also the moment generating function of a normal random variable. The result then follows from the uniqueness theorem for moment generating functions, i.e., the fact that the moment generating function of a random variable determines its distribution uniquely.

This argument is a little more illuminating. At least we can see what is happening in terms of the moment generating functions. Of course, the fact that the moment generating function of a normal random variable takes this nice form is not obvious. We also must use the uniqueness theorem to make the proof complete.

The rotation proof

The geometric proof that we now present has the advantage that it is more visual than computational. It is elementary, but requires a bit more sophistication than the earlier proofs.

We begin with two independent standard normal random variables, Z_1 and Z_2. The joint density function is

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

which is rotation invariant (see FIGURE 1). That is, it has the same value for all points equidistant from the origin. Thus, $f(T(z_1, z_2)) = f(z_1, z_2)$, where T is any rotation of the plane about the origin.

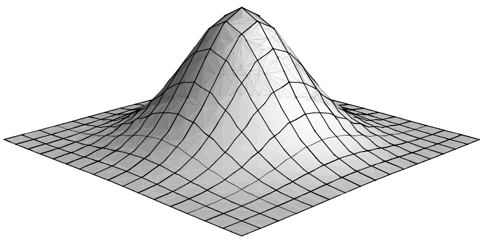


Figure 1 $f(z_1, z_2) = \frac{\exp\left(-\frac{z_1^2+z_2^2}{2}\right)}{2\pi}$ is rotation invariant.

It follows that for any set A in the plane $P((Z_1, Z_2) \in A) = P((Z_1, Z_2) \in TA)$, where T is a rotation of the plane. Now if X_1 is normal with mean 0 and variance σ_1^2 and X_2 is normal with mean 0 and variance σ_2^2, then $X_1 + X_2$ has the same distribution as $\sigma_1 Z_1 + \sigma_2 Z_2$. Hence

$$P(X_1 + X_2 \le t) = P(\sigma_1 Z_1 + \sigma_2 Z_2 \le t) = P((Z_1, Z_2) \in A),$$

where A is the half plane $\{(z_1, z_2) \mid \sigma_1 z_1 + \sigma_2 z_2 \le t\}$. The boundary line $\sigma_1 z_1 + \sigma_2 z_2 = t$ lies at a distance $d = \frac{|t|}{\sqrt{\sigma_1^2+\sigma_2^2}}$ from the origin.

It follows that the set A can be rotated into the set

$$TA = \left\{ (z_1, z_2) |\ z_1 < \frac{t}{\sqrt{\sigma_1^2 + \sigma_2^2}} \right\}.$$

(See FIGURE 2 for the case $t > 0$ and FIGURE 3 for the case $t < 0$.)

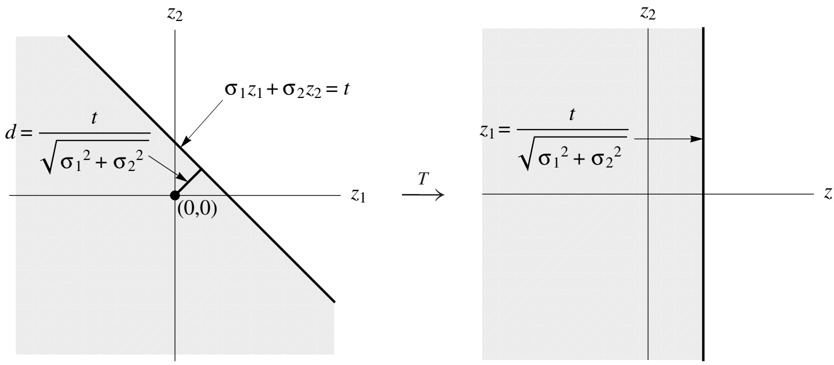


Figure 2 The half plane $\sigma_1 z_1 + \sigma_2 z_2 \leq t$, $t > 0$ is rotated into the half plane $z_1 \leq \frac{t}{\sqrt{\sigma_1^2+\sigma_2^2}}$.

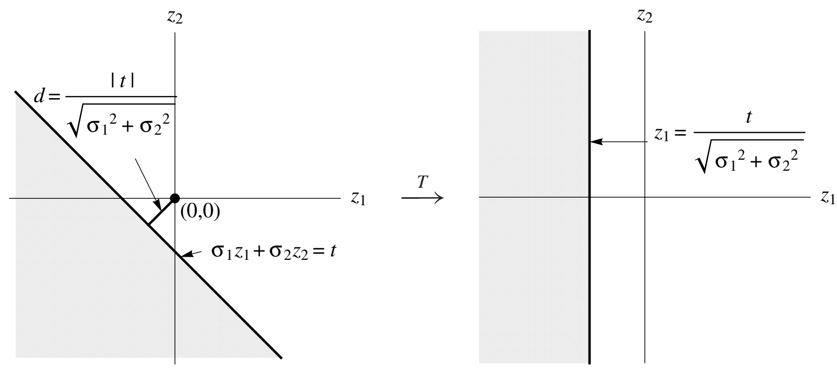


Figure 3 The half plane $\sigma_1 z_1 + \sigma_2 z_2 \leq t$, $t < 0$ is rotated into the half plane $z_1 \leq \frac{t}{\sqrt{\sigma_1^2+\sigma_2^2}}$.

Thus $P(X_1 + X_2 < t) = P(\sqrt{\sigma_1^2 + \sigma_2^2} Z_1 < t)$. It follows that $X_1 + X_2$ is normal with mean 0 and variance $\sigma_1^2 + \sigma_2^2$. This completes the proof.

In all the probability texts that we have surveyed, we have only found one [**2**, pp. 361–363] with an approach based on the rotation invariance of the joint normal density.

Generalizing the rotation argument

The next proposition follows from the same rotation argument used to show that $\sigma_1 Z_1 + \sigma_2 Z_2$ is equal to $\sqrt{\sigma_1^2 + \sigma_2^2} Z$ in the normal case.

PROPOSITION. *Assume that X and Y are random variables with rotation invariant joint distribution and X has density $f_X(x)$. Let $Z = aX + bY$. Then Z has density*

$$f_Z(z) = \frac{1}{\sqrt{a^2+b^2}} f_X\left(\frac{z}{\sqrt{a^2+b^2}}\right).$$

When X is normal, this implies that $aX + bY$ is normal. Another example occurs when (X, Y) is uniformly distributed over the unit disc. Then X has density $f(x) = \frac{2}{\pi}\sqrt{1-x^2}$ for $-1 \leq x \leq 1$. It follows that $aX + bY$ has density $f_c(x) = \frac{2}{c\pi}\sqrt{1-\frac{x^2}{c^2}}$, for $-c \leq x \leq c$, where $c = \sqrt{a^2+b^2}$. We note in this example that X and Y are not independent. It is a well known result [**1**, p. 78] that if X and Y are independent with rotation invariant joint density, then X and Y must be mean 0 normal random variables. Thus we cannot use this method to find the density of $aX + bY$ for independent X and Y except in the case where X and Y are normal.

The algebraic proof

It is possible to give a simple, plausible algebraic argument as to why the sum of independent normal random variables is normal if one is allowed to assume the central limit theorem. The central limit theorem implies that if $X_1, X_2, \ldots$ are independent, identically distributed random variables with mean 0 and variance 1, then

$$P\left(\frac{X_1+\cdots+X_n}{\sqrt{n}} \leq t\right) \to P(Z_1 \leq t),$$

and

$$P\left(\frac{X_{n+1}+\cdots+X_{2n}}{\sqrt{n}} \leq t\right) \to P(Z_2 \leq t),$$

where Z_1 and Z_2 are independent, standard normal random variables. Furthermore,

$$P\left(\frac{X_1+\cdots+X_{2n}}{\sqrt{2n}} \leq t\right) \to P(Z_3 \leq t),$$

where Z_3 is also standard normal. Since

$$\frac{X_1+\cdots+X_n}{\sqrt{n}} + \frac{X_{n+1}+\cdots+X_{2n}}{\sqrt{n}} = \frac{X_1+\cdots+X_{2n}}{\sqrt{n}} = \sqrt{2}\,\frac{X_1+\cdots+X_{2n}}{\sqrt{2n}},$$

it would seem reasonable that $Z_1 + Z_2$ has the same distribution as $\sqrt{2}\,Z_3$, i.e., $Z_1 + Z_2$ is normal with mean 0 and variance 2. This argument can be made rigorous using facts about convergence in distribution of random variables.

A similar argument using the fact that

$$P\left(\frac{X_1+\cdots+X_{[\sigma n]}}{\sqrt{n}} \leq t\right) \to P(\sigma Z \leq t)$$

would show why the sums of general independent normal random variables must be normal.

This algebraic argument is a nice conceptual argument for showing why the sum of independent normal random variables must be normal, but it assumes the central limit theorem, which is not obvious or easy to prove, as well as facts about convergence in distribution.

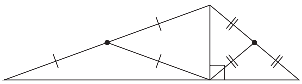
The rotation proof seems better to us than the others since it is elementary, self contained, conceptual, uses clever geometric ideas, and requires little computation. Whether it would give more insight to the average student is difficult to say. Nevertheless, with all of this in its favor, it ought to be more widely taught.

REFERENCES

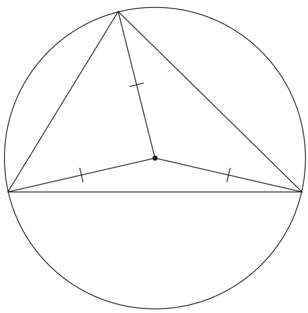
1. W. Feller, *An Introduction to Probability Theory and Its Applications, Vol. II*, John Wiley Sons, New York, 1971.
2. J. Pitman, *Probability*, Springer-Verlag, New York, 1993.

Proof Without Words: Isosceles Dissections

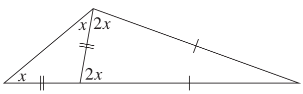
Every triangle can be dissected into four isosceles triangles:

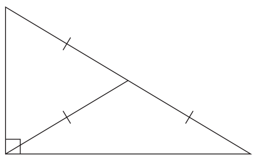


Every acute-angled triangle can be dissected into three isosceles triangles:



A triangle can be dissected into two isosceles triangles if and only if one of its angles is three times another or if the triangle is right angled:





REFERENCE

Angel Plaza, Proof without words: Every triangle can be subdivided into six isosceles triangles, this MAGAZINE **80** (2007).

DES MACHALE
Department of Mathematics
University College, Cork, Ireland