Physical interpretation of the first and second moments — the mean and variance — of a probability distribution

David Jon Furbish

Emeritus, Vanderbilt University

1 Initial remarks

The idea of a statistical moment comes from physics. Recall that there are two important moments used to describe rigid bodies: the *first moment of area* (sometimes incorrectly referred to as the first moment of inertia), and the *moment of inertia* (also referred to as the second moment of area or the rotational inertia). The first moment of area describes how the shape of an object is distributed relative to a coordinate axis, and the moment of inertia describes the tendency of an object to resist angular acceleration about an axis of rotation. Our objective is to show that the first and second moments of a probability distribution — its mean and variance — in fact are defined in a manner that is entirely consistent with the definitions of the first and second moments of area associated with rigid bodies, thus leading to the use of the term "moment" in probability and statistics.

2 The first moment (mean)

Consider a rigid but massless lever arm upon which we place small cubes, each with mass m, at arbitrary distances x from the origin $x = 0$ (Figure 1). A fulcrum is placed beneath the lever arm at the origin. Gravity is acting downward, normal to the lever arm. In order to keep the lever arm stable we apply an upward force F_+ to the lever arm at an arbitrary distance x from the fulcrum. (This force is merely an artifice to keep the physics right, but does not enter into our calculations.)

Suppose that we place N cubes on the lever arm. Let us denote the distance of the ith cube as x_i. In turn, let us calculate the torque on the lever arm about the origin produced by the ith cube. This is

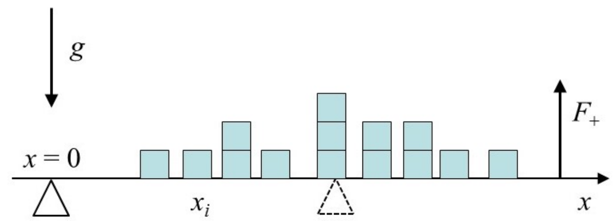

Figure 1: Definition diagram of lever arm with cubes, each of mass m, placed at distances x_i from the origin $x = 0$. The fulcrum is initially located at this position.

$$\tau_i = F_z x_i = -mgx_i \,. \tag{1}$$

This says that the torque τ_i is equal to the force F_z defined by the weight of the cube times the distance x_i of the lever arm that this force is acting on. In turn, the weight of the cube $F_z = -mg$, where g is the acceleration due to gravity. The negative sign indicates that this force is acting downward in the gravitational field. This also means that the torque τ_i is negative, implying that it is tending to produce clockwise rotation of the lever arm about the fulcrum.

Let us now calculate the total torque on the lever arm due to the N cubes. This is

$$\tau = \sum_{i=1}^{N} \tau_i = F_z \sum_{i=1}^{N} x_i = -mg \sum_{i=1}^{N} x_i \,. \tag{2}$$

With reference to Figure 1, the cubes resting on the lever arm produce a finite negative torque about the origin $x = 0$, tending to make the lever arm undergo clockwise rotation. Indeed, the force F_+ must provide a torque of equal magnitude and of opposite sign to keep the lever arm from rotating.

Let us now imagine placing the fulcrum at a finite distance from the origin such that some of the cubes are to the left of the fulcrum and some are to the right of it. Let us define the position of the fulcrum as x_0. Cubes at positions $x_i > x_0$ will provide a negative (clockwise) torque about the position x_0, and cubes at positions $x_i < x_0$ will provide a positive (counterclockwise) torque about the position x_0. Let us again calculate the torque produced by the ith cube. This is

$$\tau_i = -mg(x_i - x_0)\,. \tag{3}$$

Notice that the quantity $x_i - x_0$ is the length of the lever arm that the force $-mg$ is acting on, with respect to the position x_0 of the fulcrum. If $x_i > x_0$, then the torque τ_i is negative. If $x_i < x_0$, then the torque is positive.

Let us again calculate the total torque on the lever arm due to the N cubes. This is

$$\tau = \sum_{i=1}^{N} \tau_i = -mg \sum_{i=1}^{N} (x_i - x_0)\,. \tag{4}$$

We may then distribute the second sum in (4) giving

$$\tau = -mg \sum_{i=1}^{N} x_i + mg \sum_{i=1}^{N} x_0\,. \tag{5}$$

Because x_0 is a fixed value, this becomes

$$\tau = -mg \sum_{i=1}^{N} x_i + mgNx_0\,. \tag{6}$$

Let us now suppose that we placed the fulcrum at the specific position x_0 such that the positive torque produced by the cubes to the left of x_0 is exactly balanced by the negative torque produced by the cubes to the right of x_0. This is the same as saying that the total torque $\tau = 0$. Now (6) becomes

$$0 = -mg \sum_{i=1}^{N} x_i + mgNx_0\,. \tag{7}$$

With a little algebra we then discover that

$$x_0 = \frac{1}{N} \sum_{i=1}^{N} x_i\,. \tag{8}$$

Thus, the position x_0, which represents the center of mass of the cubes, is precisely the same as the mean position of the cubes, that is,

$$x_0 = \overline{x} = \frac{1}{N} \sum_{i=1}^{N} x_i\,. \tag{9}$$

In physics terms, this is the first moment of the system.

Let us now turn to the definition of the first moment (the mean) of a probability density function. Consider the probability density function $f_x(x)$ of the random variable x with mean μ_x. By definition

$$\int_{-\infty}^{\infty} f_x(x)\,\mathrm{d}x = 1\,. \tag{10}$$

Letting N denote a great number of the values of x, then (10) may be restated as

$$\int_{-\infty}^{\infty} n_x(x)\,\mathrm{d}x = N\,. \tag{11}$$

where $n_x(x) = Nf_x(x)$ is the number density. Then, for example, $n_x(x)\mathrm{d}x$ is the number of values of x falling within the small interval x to $x + \mathrm{d}x$, that is, the number of cubes within this interval. Because N is arbitrarily large, here we are imagining the cubes as being tiny — in effect point masses.

The small amount of torque $\mathrm{d}\tau$ about $x = 0$ due to the cubes within the interval x to $x + \mathrm{d}x$ is

$$\mathrm{d}\tau = -mgxn_x(x)\,\mathrm{d}x\,. \tag{12}$$

The total torque is

$$\tau = \int_{-\infty}^{\infty} \mathrm{d}\tau = -mg \int_{-\infty}^{\infty} xn_x(x)\,\mathrm{d}x\,. \tag{13}$$

With reference to Figure 2, the cubes resting on the lever arm produce a finite negative torque about the origin $x = 0$.

As before, let us now choose the position x_0 such that the total torque about this position is zero. We thus rewrite (13) as

$$\tau = \int_{-\infty}^{\infty} \mathrm{d}\tau = -mg \int_{-\infty}^{\infty} (x - x_0) n_x(x)\,\mathrm{d}x\,. \tag{14}$$

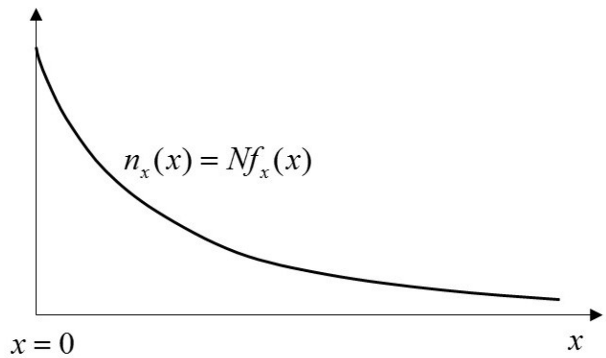

Figure 2: Example of number density function $n_x(x) = Nf_x(x)$.

In turn, setting $\tau = 0$ and distributing the integral,

$$0 = -mg \int_{-\infty}^{\infty} x n_x(x)\,\mathrm{d}x + mgx_0 \int_{-\infty}^{\infty} n_x(x)\,\mathrm{d}x\,. \tag{15}$$

Using (11) this becomes

$$\int_{-\infty}^{\infty} x n_x(x)\,\mathrm{d}x = N x_0\,. \tag{16}$$

Dividing both sides of (16) by N and using the fact that $f_x(x) = (1/N) n_x(x)$,

$$x_0 = \mu_x = \int_{-\infty}^{\infty} x f_x(x)\,\mathrm{d}x\,. \tag{17}$$

Thus, the position x_0, which represents the center of mass of the cubes, is precisely the same as their mean position, that is, the first moment of the distribution $f_x(x)$.

3 The second moment (variance)

Now that we have defined the first statistical moment (the mean) in physical terms, let us similarly define the second statistical moment (the variance) in terms of the moment of inertia. Once again we appeal to our lever arm upon which we place small cubes, each with mass m, at distances x_i from the origin.

We now know that the first moment, $x_0 = \overline{x}$, represents the center of mass such that the total toque τ about the fulcrum is zero and the system has no tendency to rotate. What we now need to imagine is a measure of the resistance of the system to being rotated about the fulcrum when a torque is applied to the system, independently of the gravitational field. This involves the rotational inertia (or moment of inertia).

With respect to rotation about the fulcrum positioned at x_0, the inertial moment of the ith cube is

$$I_i = m(x_i - x_0)^2\,. \tag{18}$$

The total inertial moment is

$$I = \sum_{i=1}^{N} I_i = m \sum_{i=1}^{N} (x_i - x_0)^2\,. \tag{19}$$

Let us use divide (19) by mN to give

$$\frac{I}{Nm} = \frac{1}{N} \sum_{i=1}^{N} (x_i - x_0)^2\,. \tag{20}$$

In turn, recalling that the first moment $x_0 = \overline{x}$, this becomes

$$\frac{I}{Nm} = s_x^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \overline{x})^2\,. \tag{21}$$

Note that the product Nm is the total mass of the system. The variance s_x^2 thus may be interpreted as the total inertial moment per system mass. And, we are free to set $Nm = 1$.

Consider two separate lever arms with cubes attached to each at various positions x_i, where in the first case the cubes are concentrated close to the mean position $\overline{x}$, and in the second case the cubes are relatively far away from the mean position. This is the same as saying that the variance is small in the first case and large in the second case. Now imagine trying to spin the lever arms about their center of mass (the mean position). Clearly it will be easier to spin the first lever arm with the smaller variance, that is, with the smaller moment of inertia.

Let us now turn to the definition of the second moment (the variance) of a probability density function. With respect to rotation about the fulcrum positioned at $x_0 = \mu_x$, the small inertial moment of cubes located within the interval x to $x + \mathrm{d}x$ is

$$\mathrm{d}I = m(x - \mu_x)^2 n_x(x)\mathrm{d}x\,. \tag{22}$$

The total inertial moment is

$$I = \int_{-\infty}^{\infty} \mathrm{d}I = m\int_{-\infty}^{\infty} (x - \mu_x)^2 n_x(x)\mathrm{d}x\,. \tag{23}$$

Dividing by Nm,

$$\frac{I}{Nm} = \sigma_x^2 = \int_{-\infty}^{\infty} (x - \mu_x)^2 f_x(x)\mathrm{d}x\,, \tag{24}$$

which is the continuous version of (21) above. And again, we are free to set $Nm = 1$.

4 Symmetry

In learning probability and statistics we are frequently asked for good reasons to think about the Gaussian distribution and its properties. It is easy to envision that the mean of this symmetrical distribution represents its center of "mass," and to grasp the idea that half of the probability represented by the Gaussian distribution falls to the left of the mean and half falls to its right. Based on our interpretation above, the torque of the system (distribution) measured about the mean is zero. But now let us briefly consider an asymmetrical distribution, the exponential distribution, whose number density is depicted in Figure 2.

Recall that the exponential distribution is given by

$$f_x(x) = \frac{1}{\mu_x} e^{-x/\mu_x}\,, \tag{25}$$

with mean μ_x. Much of the probability represented by this distribution is concentrated near the origin ($x = 0$). The cumulative distribution function is

$$F_x(x) = 1 - e^{-x/\mu_x}\,. \tag{26}$$

By setting $x = \mu_x$ we have

$$F_x(x) = 1 - e^{-1}\,, \tag{27}$$

which tells us that about 63% of the probability falls to the left of the mean and about 37% falls to its right — in sharp contrast to a symmetrical distribution. Nonetheless, our interpretation that the position given by the mean coincides with a condition of zero torque measured about this position holds. Recall that torque is the product of a force and a lever arm. Again imagining that the probability associated with $f_x(x)$ is represented by a great number of point masses, then to the left of the mean the lever arms $x - \mu_x$ associated with these masses are relatively short, but the number of point masses is large. To the right of the mean there are fewer point masses, but their associated lever arms $x - \mu_x$ are relatively large. Indeed, the magnitude of the largest lever arm to the left of the mean is μ_x, and the largest lever arm to the right of the mean approaches infinity! All is in good order.