

3.10: Statistics - the Mean and the Variance of a Distribution

There are two important statistics associated with any probability distribution, the **mean** of a distribution and the **variance** of a distribution. The mean is defined as the expected value of the random variable itself. The Greek letter μ is usually used to represent the mean. If $f(u)$ is the cumulative probability distribution, the mean is the expected value for $g(u) = u$. From our definition of expected value, the mean is

$$\mu = \int_{-\infty}^{\infty} u \left(\frac{df}{du} \right) du$$

The variance is defined as the expected value of $(u - \mu)^2$. The variance measures how dispersed the data are. If the variance is large, the data are—on average—farther from the mean than they are if the variance is small. The **standard deviation** is the square root of the variance. The Greek letter σ is usually used to denote the standard deviation. Then, σ^2 denotes the variance, and

$$\sigma^2 = \int_{-\infty}^{\infty} (u - \mu)^2 \left(\frac{df}{du} \right) du$$

If we have a small number of points from a distribution, we can estimate μ and σ by approximating these integrals as sums over the domain of the random variable. To do this, we need to estimate the probability associated with each interval for which we have a sample point. By the argument we make in [Section 3.7](#), the best estimate of this probability is simply $1/N$, where N is the number of sample points. We have therefore

$$\mu = \int_{-\infty}^{\infty} u \left(\frac{df}{du} \right) du \approx \sum_1^N u_i \left(\frac{1}{N} \right) = \bar{u}$$

That is, the best estimate we can make of the mean from N data points is \bar{u} , where \bar{u} is the ordinary arithmetic average. Similarly, the best estimate we can make of the variance is

$$\sigma^2 = \int_{-\infty}^{\infty} (u - \mu)^2 \left(\frac{df}{du} \right) du \approx \sum_{i=1}^N (u_i - \mu)^2 \left(\frac{1}{N} \right)$$

Now a complication arises in that we usually do not know the value of μ . The best we can do is to estimate its value as $\mu \approx \bar{u}$. It turns out that using this approximation in the equation we deduce for the variance gives an estimate of the variance that is too small. A more detailed argument (see [Section 3.14](#)) shows that, if we use \bar{u} to approximate the mean, the best estimate of σ^2 , usually denoted s^2 , is

$$\text{estimated } \sigma^2 = s^2 = \sum_{i=1}^N (u_i - \bar{u})^2 \left(\frac{1}{N-1} \right)$$

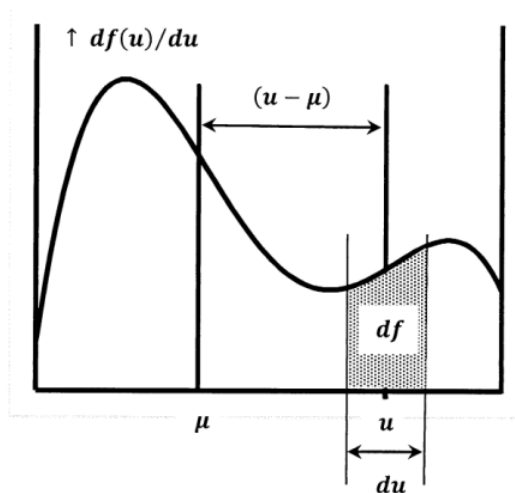


Figure 3.10.1: Variance is analogous to a moment of inertia.

Dividing by $N - 1$, rather than N , compensates exactly for the error introduced by using \bar{u} rather than μ .

The mean is analogous to a center of mass. The variance is analogous to a moment of inertia. For this reason, the variance is also called **the second moment about the mean**. To show these analogies, let us imagine that we draw the probability density function on a uniformly thick steel plate and then cut along the curve and the u -axis (Figure 3.10.1). Let M be the mass of the cutout piece of plate; M is the mass below the probability density curve. Let dA and dm be the increments of area and mass in the thin slice of the cutout that lies above a small increment, du , of u . Let ρ be the density of the plate, expressed as mass per unit area. Since the plate is uniform, ρ is constant. We have $dA = (df/du) du$ and $dm = \rho dA$ so that

$$dm = \rho \left(\frac{df}{du} \right) du$$

The mean of the distribution corresponds to a vertical line on this cutout at $u = \mu$. If the cutout is supported on a knife-edge along the line $u = \mu$, gravity induces no torque; the cutout is balanced. Since the torque is zero, we have

$$0 = \int_{m=0}^M (u - \mu) dm = \int_{-\infty}^{\infty} (u - \mu) \rho \left(\frac{df}{du} \right) du$$

Since μ is a constant property of the cut-out, it follows that

$$\mu = \int_{-\infty}^{\infty} u \left(\frac{df}{du} \right) du$$

The cutout's moment of inertia about the line $u = \mu$ is

$$\begin{aligned} I &= \int_{m=0}^M (u - \mu)^2 dm \\ &= \int_{-\infty}^{\infty} (u - \mu)^2 \rho \left(\frac{df}{du} \right) du \\ &= \rho \sigma^2 \end{aligned}$$

The moment of inertia about the line $u - \mu$ is simply the mass per unit area, ρ , times the variance of the distribution. If we let $\rho = 1$, we have $I = \sigma^2$.

We define the mean of $f(u)$ as the expected value of u . It is the value of u we should “expect” to get the next time we sample the distribution. Alternatively, we can say that the mean is the best prediction we can make about the value of a future sample from the distribution. If we know μ , the best prediction we can make is $u_{\text{predicted}} = \mu$. If we have only the estimated mean, \bar{u} , then \bar{u} is the best prediction we can make. Choosing $u_{\text{predicted}} = \bar{u}$ makes the difference, $|u - u_{\text{predicted}}|$, as small as possible.

These ideas relate to another interpretation of the mean. We saw that the variance is the second moment about the mean. The first moment about the mean is

$$\begin{aligned} 1^{\text{st}} \text{ moment} &= \int_{-\infty}^{\infty} (u - \mu) \left(\frac{df}{du} \right) du \\ &= \int_{-\infty}^{\infty} u \left(\frac{df}{du} \right) du - \mu \int_{-\infty}^{\infty} \left(\frac{df}{du} \right) du \\ &= \mu - \mu \\ &= 0 \end{aligned}$$

Since the last two integrals are μ and 1, respectively, **the first moment about the mean is zero**. We could have defined the mean as the value, μ , for which the first moment of u about μ is zero.

The first moment about the mean is zero. The second moment about the mean is the variance. We can define third, fourth, and higher moments about the mean. Some of these higher moments have useful applications.

This page titled [3.10: Statistics - the Mean and the Variance of a Distribution](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Paul Ellgen](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.