The Mean of a Random Variable

The *mean* of a random variable, also called its *expected value* or *expectation*, is its "average" value, accounting for the different probabilities of its possible values.

If X is a discrete random variable, its mean, written μ_X, is the following sum:

$$\mu_X \;=\; \sum_x x\, P(X=x)$$

where x ranges over all possible values of X.

So if X has n equally likely values, $x_1, \ldots, x_n$:

$$\mu_X \;=\; \frac{1}{n}\sum_{i=1}^{n} x_i$$

which is like the sample mean of data.

For a continuous random variable, with probability density $f(x)$, we use an integral:

$$\mu_X \;=\; \int_{-\infty}^{+\infty} x\, f(x)\, dx$$

(assuming the range of X is $(-\infty, +\infty)$).

The Variance of a Random Variable

We define the *variance* of a discrete random variable, X, to be

$$\sigma_X^2 \;=\; \sum_x (x-\mu_X)^2\, P(X=x)$$

So if X has n equally likely values, $x_1, \ldots, x_n$:

$$\sigma_X^2 \;=\; \frac{1}{n}\sum_{i=1}^{n} (x_i-\mu_X)^2$$

which is almost like the sample variance of data.

If X has a continuous distribution on $(-\infty, +\infty)$,

$$\sigma_X^2 \;=\; \int_{-\infty}^{+\infty} (x-\mu_X)^2 f(x)\, dx$$

Looked at another way, the variance of X is the mean of $(X-\mu_X)^2$ — the average squared distance of X from its mean value.

The *standard deviation* of a random variable X (written as σ_X) is the square root of its variance.

The Normal Probability Distribution

The *normal probability density function* with mean μ and standard deviation σ is given by the formula

$$f(x) \;=\; \frac{1}{\sqrt{2\pi}\sigma}\, e^{-(x-\mu)^2/2\sigma^2}$$

One could check that μ really is the mean of this distribution by verifying that

$$\int_{-\infty}^{+\infty} x\, \frac{1}{\sqrt{2\pi}\sigma}\, e^{-(x-\mu)^2/2\sigma^2}\, dx \;=\; \mu$$

and similarly for the variance. These integrals are doable, but we'll just take the result as given in this course.

Z-scores and the Standard Normal Distribution

There are many normal distributions, with different values for μ and σ, but they can all be related to the *standard normal distribution*, for which $\mu = 0$ and $\sigma = 1$.

To relate a random variable X to the standard normal, we define a corresponding random variable Z as follows:

$$Z \;=\; \frac{X-\mu_X}{\sigma_X}$$

If X has a normal distribution, this "Z-score" will have the standard normal distribution.

Why bother? Because questions involving the standard normal distribution can be answered using the table in Appendix B.2 of the book.

Example: Toronto Temperatures

Suppose that from historical data we know that the daily maximum temperature (in degrees Celsius) in Toronto on October 20 is normally distributed with mean 5 and standard deviation 8.

What is the probability that the maximum temperature on October 20 this year will be over 17 degrees Celsius?

Let X be the temperature. The Z-score is $Z = (X - 5)/8$. Z will have the standard normal distribution. We want to find

$$\begin{aligned} P(X > 17) &= P((X-5)/8 > (17-5)/8) \\ &= P(Z > 1.5) \\ &= 1 - P(Z \leq 1.5) \\ &= 1 - 0.9332 \;=\; 0.0668 \end{aligned}$$

The number 0.9332 comes from Table B.2 in the book.