THE VARIANCE OF A RANDOM VARIABLE

1. Motivation

Let $X \sim f(x)$ where

$$f(x) = \begin{cases} 2x & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Then, we have $\mathrm{E}(X) = \int_0^1 2x^2 \, \mathrm{d}x = \frac{2}{3}$. Now, let $Y \sim g(y) = \frac{2}{3} \ \forall \ y \in \mathbb{R}$. $\mathrm{E}(X) = \mathrm{E}(Y)$, but clearly X and Y are very different, since X can vary whereas Y can only take on its expected value, and therefore does so with 100% probability. This demonstrates that the expected value **alone** is not a good way to describe a random variable.

2. Defining variance

We want to have a measure of how much a random variable X can deviate from its expected value. We can define this deviation as $X - \mathrm{E}(X)$. However, it doesn't matter in which direction this difference is; that is, we want to treat positive and negative values of this the same. We can fix this by instead looking at $[X - \mathrm{E}(X)]^2$. We want to measure what value this is expected to take, that is, its *expected value*. With this in mind, let us define

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}([X - \mathrm{E}(X)]^2) \\
&= \mathrm{E}(X^2 + \mathrm{E}(X)^2 - 2X\mathrm{E}(X)) \\
&= \mathrm{E}(X^2) + \mathrm{E}(X)^2 - \mathrm{E}(2X\mathrm{E}(X)) \\
&= \mathrm{E}(X^2) + \mathrm{E}(X)^2 - 2\mathrm{E}(X)^2 \\
&= \mathrm{E}(X^2) - \mathrm{E}(X)^2
\end{aligned}$$

2.1. **Calculating variance.** To calculate this we can just apply the formula for the expected value of a variable. Let $\mu = \mathrm{E}(X)$. Then, for a discrete random variable,

$$\sum_x x^2 f(x) - \mu^2$$

and for a continuous random variable,

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} x^2 f(x) \, \mathrm{d}x - \mu^2$$

2.2. **The standard deviation.** Since the variance is defined as the expected value of the *square* of the deviation, its units are the units of the random variable squared. For practical purposes, we may want a measure that comes in the same units as our random variable. For this we can define the **standard deviation** σ as $\sqrt{\mathrm{Var}(X)}$. This is valid since the variance is always positive.

3. Properties of the variance

From its definition, we can derive properties of the variance function:

- $\text{Var}(X) \geq 0$ for all X
- Iff X is constant, that is $\mathbb{P}(X = c) = 1$ for some c, then $\text{Var}(X) = 0$
- $\text{Var}(aX + b) = a^2\text{Var}(X)$
- $\text{Var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \text{Var}(X_i)$ iff all the variables X_i are independent

3.1. **Jensen's inequality.** From the properties of the variance we can derive the following property: if g is a given *convex* function, then $\text{E}(g(X)) \geq g(\text{E}(X))$. This comes from the fact that $\text{Var}(X) = \text{E}(X^2) - \text{E}(X)^2 \geq 0$, defining $g(x) = x^2$. This holds for all convex functions, not just $g : x \mapsto x^2$.