Variance

Will Perkins

January 22, 2013

Variance

Definition

The *variance* of a random variable X is:

$$\text{var}(X) = \mathbb{E}[(X - \mathbb{E}X)^2]$$

Alternatively, (check using linearity of expectation),

$$\text{var}(X) = \mathbb{E}[X^2] - (\mathbb{E}X)^2$$

Variance

Variance is a measure of how far a random variable typically deviates from its mean.

Moments

Sometimes we refer to the mean and variance of a random variable as its first and second moments respectively.

The kth *moment* of a random variable X is $\mathbb{E}[X^k]$.

The kth *moment about the mean* or the kth *central moment* of X is $\mathbb{E}[(X - \mathbb{E}X)^k]$. The variance is techincally the 2nd moment about the mean.

Examples

Calculate the variance of the following random variables:

1. $X \sim Bin(n, p)$
2. $X \sim Uniform[0, 1]$
3. $X \sim N(0, 1)$

Variance of Sums

Unlike expectation, variance is not linear!

$$\text{var}[aX] = a^2\text{var}[X]$$

$$\text{var}(X + Y) = ?$$

Depends on the dependence.

Give examples where $\text{var}(X + Y)$ is as high and as low as possible, relative to $\text{var}(X)$ and $\text{var}(Y)$.

Covariance

We have one simple measurement of how two random variables depend on each other.

Definition

The *covariance* of X and Y is:

$$\text{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

Note:

1. Covariance can be positive or negative.
2. $\text{cov}(X, X) = \text{var}(X)$

Correlation

- If $\text{cov}(X, Y) = 0$ we say that X and Y are *uncorrelated*.
- If $\text{cov}(X, Y) > 0$ we say that X and Y are *positively correlated*.
- If $\text{cov}(X, Y) < 0$ we say that X and Y are *negatively correlated*.

Sometimes people mention the *correlation* of X and Y. This is defined as

$$corr(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}$$

Q: What are the units of correlation?

Variance of Sums

While variance is not linear, we have a useful formula for computing variance of sums:

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y)$$

[Check that this is correct when $X = Y$]

And in general,

$$\text{var}(X_1 + \ldots X_n) = \sum_{i=1}^{n} \text{var}(X_i) + \sum_{i \neq j} \text{cov}(X_i, X_j)$$

Note that the sum is over ordered pairs (which is why there is a factor 2 in the case of $X + Y$).

Variance of Counting Random Variables

When X is a counting random variable, we can use the decomposition of $X = X_1 + \cdots + X_n$ into indicator random variables to simplify the calculation of $\mathrm{var}(X)$.

Let $p_i = \Pr[X_i = 1] = \mathbb{E}X_i$. Then $\mathbb{E}X = \sum p_i$. And

$$\mathrm{var}(X) = \sum_i p_i - p_i^2 + \sum_{i \neq j} \mathrm{cov}(X_i, X_j)$$

Variance of Counting Random Variables

Using the definition of covariance,

$$\text{cov}(X_i, X_j) = \Pr[X_i = 1 \text{ AND } X_j = 1] - p_i p_j$$

So,

$$\text{var}(X) = \sum_i p_i - p_i^2 + \sum_{i \neq j} \Pr[X_i = 1 \text{ AND } X_j = 1] - p_i p_j$$

Examples

For the random graph $G(n, p)$, calculate

- The variance of the degree of a given vertex.
- The variance of the number of isolated vertices.