CS 70 Discrete Mathematics and Probability Theory
Fall 2026 Hug, Sabin

DIS 11

Variance, Covariance, and Correlation

Throughout, X and Y are discrete random variables.

Variance is the average squared distance from the mean:

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2,$$

where $\mathbb{E}[X^2] = \sum_x x^2 \, \mathbb{P}[X = x]$. It is nonnegative, and its square root is the standard deviation. For constants a, b,

$$\mathrm{Var}(aX + b) = a^2 \, \mathrm{Var}(X).$$

If $X_1, \ldots, X_n$ are independent, then $\mathrm{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathrm{Var}(X_i)$. Unlike linearity of expectation, this can fail for dependent variables.

Covariance measures whether X and Y tend to deviate from their means in the same direction (positive) or in opposite directions (negative):

$$\mathrm{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y],$$

where $\mathbb{E}[XY] = \sum_x \sum_y xy \, \mathbb{P}[X = x, Y = y]$. In general,

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{cov}(X, Y).$$

Independent variables have zero covariance, but zero covariance does not imply independence.

Correlation rescales covariance by the standard deviations. When both variances are positive (otherwise it is undefined),

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}} \in [-1, 1].$$

Correlation is unitless and measures linear association: it is ± 1 exactly when $Y = aX + b$ with probability 1, with $a > 0$ for $+1$ and $a < 0$ for -1. Variables with zero correlation are called *uncorrelated*; they need not be independent.

1 Binomial Variance

Note 21

Throw n balls into m bins uniformly at random. For a specific ball i, what is the variance of the number of roommates it has (i.e. the number of other balls that it shares its bin with)?

2 Diversify Your Hand

Note 20
Note 21

You are dealt 5 cards from a standard 52 card deck. Let X be the number of distinct values in your hand. For instance, the hand (A, A, A, 2, 3) has 3 distinct values.

(a) Calculate $\mathbb{E}[X]$. (Hint: Consider indicator variables X_i representing whether value i appears in the hand.)

(b) Calculate $\text{Var}(X)$. The answer expression will be quite involved; you do not need to simplify anything.

3 Covariance and Correlation Warm-Up

Note 22

Let X, Y be random variables where $\mathbb{P}[X=0, Y=0] = \frac{1}{4}$, $\mathbb{P}[X=1, Y=2] = \frac{1}{2}$, and $\mathbb{P}[X=-1, Y=2] = \frac{1}{4}$.

(a) What is $\text{cov}(X, Y)$?

(b) What is $\text{Corr}(X, Y)$?

(c) What is $\text{cov}(X+2Y, 3X-Y)$?

4 Correlation and Independence

Note 20
Note 22

(a) What does it mean for two random variables to be uncorrelated?

(b) What does it mean for two random variables to be independent?

(c) Are all uncorrelated variables independent? Are all independent variables uncorrelated? If your answer is yes, justify your answer; if your answer is no, give a counterexample.

5 Coupon Collector Variance

Note 21

It's that time of the year again (again)—Safeway is offering its Monopoly Card promotion. Each time you visit Safeway, you are given one of n different Monopoly Cards with equal probability. You need to collect them all to redeem the grand prize.

Let X be the number of visits you have to make before you can redeem the grand prize. Show that

$$\text{Var}(X) = n^2 \left(\sum_{i=1}^{n} \frac{1}{i^2} \right) - \mathbb{E}[X].$$