Math 141
Lecture 7: Variance, Covariance, and Sums

Albyn Jones

1 Library 304
jones@reed.edu
www.people.reed.edu/~jones/courses/141
Variance: expected squared deviation from the mean:

$$\text{Var}(X) = \mathbb{E}(X - \mu_X)^2 = \sigma^2$$

Standard Deviation:

$$\text{SD}(X) = \sqrt{\text{Var}(X)} = \sigma$$

Properties:

$$\text{Var}(a + bX) = b^2 \text{Var}(X)$$

$$\text{SD}(a + bX) = b \text{SD}(X)$$

Interpretation: For a symmetric distribution, the standard deviation may be considered a *typical* deviation from the mean. For a strongly asymmetrical distribution, it is better to work with quantiles (percentiles of the distribution).
**Definition: Covariance**

Let $X$ and $Y$ be two RV’s with means $\mu_X$ and $\mu_Y$, respectively. Covariance is the expected value of the products of deviations from the means:

$$\text{Cov}(X, Y) = \mathbb{E}((X - \mu_X)(Y - \mu_Y))$$

**Definition: Correlation**

Correlation is just scale free covariance:

$$\text{Cor}(X, Y) = \rho_{xy} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \text{Cov}(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y})$$
Covariance and Correlation

Extreme Cases: Covariance

If $X = Y$, i.e. the two RV’s are copies of each other, then

$$\text{Cov}(X, Y) = \text{Cov}(X, X) = \mathbb{E}((X - \mu_X)(X - \mu_X)) = \text{Var}(X)$$

If $X = -Y$, then

$$\text{Cov}(X, Y) = \text{Cov}(X, -X) = -\text{Var}(X)$$

Extreme Cases: Correlation

If $X = Y$, i.e. the two RV’s are copies of each other, then

$$\text{Cor}(X, Y) = \frac{\text{Cov}(X, X)}{\sigma_X \sigma_X} = \frac{\text{Var}(X)}{\sigma_X^2} = 1$$

If $X = -Y$, then

$$\text{Cor}(X, Y) = \frac{\text{Cov}(X, -X)}{\sigma_X^2} = \frac{-\text{Var}(X)}{\sigma_X^2} = -1$$
**Linear Association**

Covariance and Correlation are measures of the degree of linear association.

Independent RV’s have correlation 0, but correlation 0 does not guarantee independence!
Positive Covariance: linear association with positive slope.

\[ \text{Cov}(X, Y) > 0 \]
Negative Covariance: linear association with negative slope.

\[ \text{Cov}(X, Y) < 0 \]
Zero Covariance: no linear association!

$$\text{Cov}(X, Y) = 0$$
The variance of a sum

The expected value of a sum is the sum of the expected values:
\[ \mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y). \]
What about variances?

\[ \text{Var}(X + Y) = \mathbb{E}((X + Y) - (\mu_x + \mu_y))^2 \]

Collect the terms involving \( X \) and the terms involving \( Y \):

\[ \mathbb{E}((X - \mu_x) + (Y - \mu_y))^2 \]

Recalling that \((a + b)^2 = a^2 + 2ab + b^2\), we have

\[ \mathbb{E}((X - \mu_x)^2 + (Y - \mu_y)^2 + 2(X - \mu_x)(Y - \mu_y)) \]

Finally, the expected value of a sum is the sum of the expected values:

\[ \text{Var}(X + Y) = \mathbb{E}(X - \mu_x)^2 + \mathbb{E}(Y - \mu_y)^2 + 2\mathbb{E}((X - \mu_x)(Y - \mu_y)) \]
We have produced an expression for the variance of a sum:

\[ \text{Var}(X + Y) = \mathbb{E}(X - \mu_x)^2 + \mathbb{E}(Y - \mu_y)^2 + 2\mathbb{E}((X - \mu_x)(Y - \mu_y)) \]

The first two terms are \( \text{Var}(X) \) and \( \text{Var}(Y) \), the third is \( 2\text{Cov}(X, Y) \). Thus

\[ \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \]
The variance of a sum, part three

\[ \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \]

If \( \text{Cov}(X, Y) > 0 \), then

\[ \text{Var}(X + Y) > \text{Var}(X) + \text{Var}(Y) \]

If \( \text{Cov}(X, Y) < 0 \), then

\[ \text{Var}(X + Y) < \text{Var}(X) + \text{Var}(Y) \]
Fact: If two RV’s are independent, we can’t predict one using the other, so there is no linear association, and their covariance is 0.

Theorem: If $X$ and $Y$ are independent, then

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

In other words, if the random variables are independent, then the variance of their sum is the sum of their variances!

Remember Pythagorus? For independent RV’s, the SD of a sum works like Euclidean distance for right triangles! If $Z = X + Y$

$$\sigma_Z = \sqrt{\sigma_X^2 + \sigma_Y^2}$$
For the mathematically inclined, covariance is an inner product on the infinite dimensional vector space of random variables with finite variance.
Let $X$ be a Binomial($n, p$) RV. We know $\mathbb{E}(X) = np$. We also know that $X$ is the sum of $n$ independent Bernoulli($p$) trials $Y_i$, each with $\mathbb{E}(Y_i) = p$, and $\text{Var}(Y_i) = pq$ where $q = (1 - p)$ so

$$\text{Var}(X) = \text{Var}(Y_1 + Y_2 + \ldots + Y_n)$$

The $Y$’s are independent, so their variances add:

$$\text{Var}(X) = \text{Var}(Y_1) + \text{Var}(Y_2) + \ldots + \text{Var}(Y_n) = npq$$

Hence we have shown

$$\text{Var}(X) = \sum_{k=0}^{n} (k - np)^2 \binom{n}{k} p^k q^{n-k} = npq$$
Toss a fair coin independently 100 times, and let \( X \) be the number of Heads we get.

- What is the appropriate probability model?
  \( X \sim \text{Binomial}(100, \frac{1}{2}) \)

- What is the expected number of heads?

- What is \( \sigma^2_X \), the variance of \( X \)?

- What is \( \sigma_X \)?

- What are typical outcomes of this experiment?
Example: 1000 coin tosses

Toss a fair coin independently 1000 times, and let $X$ be the number of Heads we get.

- $X \sim \text{Binomial}(1000, 1/2)$
- What is the expected number of heads?
- What is $\sigma^2_{X}$, the variance of $X$?
- What is $\sigma_{X}$?
- What are typical outcomes of this experiment?
Suppose $X \sim \text{Poisson}(\mu)$. What are $\mathbb{E}(X)$ and $\text{Var}(X)$? One can compute them directly from the definition, by summing the infinite series

$$
\mathbb{E}(X) = \sum k \cdot \mu^k e^{-\mu} / k! = \mu
$$

and then

$$
\text{Var}(X) = \sum (k - \mu)^2 \cdot \mu^k e^{-\mu} / k!
$$

Summing those series makes use of ideas from Math 112. Can we guess the answer without using Math 112?
Let $X$ be a Binomial($n,p$) RV, where $n$ is large and $p$ is small.

- We know $\mathbb{E}(X) = np$, and $\text{Var}(X) = np(1 - p)$. If $p$ is small, $(1 - p)$ is close to 1, and so $\text{Var}(X) \approx np = \mathbb{E}(X)$.

- Since that Binomial $X$ is approximately a Poisson with parameter $\mu = np$, we should guess for $X \sim \text{Poisson}(\mu)$:
  - $\mathbb{E}(X) = \mu$
  - $\text{Var}(X) = \mu$

and
Variance of a sum of two RV’s:
\[ \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \]

Variance of a sum of independent RV’s:
\[ \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) \]

Pythagorus and the Standard Deviation of a sum of independent RV’s:
\[ \text{SD}(X + Y) = \sqrt{\sigma_X^2 + \sigma_Y^2} \]

Variance of \( X \sim \text{Binomial}(n, p) \):
\[ \text{Var}(X) = npq = np(1 - p) \]