More on Expectation and Variance

Probability Theory: Foundation for Data Science

with **Anne Dougherty**

Data Science
UNIVERSITY OF COLORADO **BOULDER**

Expectation, Variance, Covariance, and Correlation

At the end of this module, students should be able to

- **Compute the mean, variance, and standard deviation of a function of a random variable (i.e. $g(X)$).**
- Explain the concept of jointly distributed random variables, for two random variables X and Y.
- Define, compute, and interpret the covariance between two random variables X and Y.
- Define, compute, and interpret the correlation between two random variables X and Y.

Motivating Examples: In statistics and data science, we frequently collect data from several random variables and we want to understand and quantify the strength of their interactions.

- The length of time a student studies and their score on an exam.
- The relationship between male and female life expectancy in a certain country.
- The relationship between the quantity of two different products purchased by a consumer.

Recall:

- $E(X) = \sum_k kP(X = k)$ if X is discrete
- $E(X) = \int_{-\infty}^{\infty} xf(x)\,dx$ if X is continuous.

What can we say about $E(g(X))$?

$E(aX + b) =$

Example: Suppose a university has 15,000 students and let X equal the number of courses for which a randomly selected student is registered. The pmf is

x	1	2	3	4	5	6	7
p(x)	.01	.03	.13	.25	.39	.17	.02

If a student pays \$500 per course plus a \$100 per-semester registration fee, what is the average amount a student pays each semester?

Recall: $\sigma^2 = V(X) = E[(X-\mu)^2] = E(X^2) - (E(X))^2$.

- $V(X) = \sum_k (k-\mu)^2 P(X=k)$ if X is discrete
- $V(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx$ if X is continuous.

What about $V(g(X))$?

Find $V(aX + b)$

Example: Suppose a university has 15,000 students and let X equal the number of courses for which a randomly selected student is registered. The pmf is

x	1	2	3	4	5	6	7
p(x)	.01	.03	.13	.25	.39	.17	.02

If a student pays \$500 per course plus a \$100 per-semester registration fee, what is the average amount a student pays each semester?

We found $E(X) = 4.57$ and $E(500X + 100) = \$2,385$.

$V(X) =$

$V(500X + 100) =$