Jointly Distributed Random Variables

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.

Example: An insurance agency services customers who have both a homeowner's policy and an automobile policy. For each type of policy, a deductible amount must be specified. For an automobile policy, the choices are \$100 or \$250 and for the homeowner's policy, the choices are \$0, \$100, or \$200.

Suppose an individual, let's say Bob, is selected at random from the agency's files. Let X be the deductible amount on the auto policy and let Y be the deductible amount on the homeowner's policy.

We want to understand the relationship between X and Y.

Suppose the **joint probability table** is given by the insurance company as follows:

		y (home)		
		0	100	200
x (auto)	100	.20	.10	.20
	250	.05	.15	.30

Definition: Given two discrete random variables, X and Y, $p(x, y) = P(X = x, Y = y)$ is the **joint probability mass function** for X and Y.

		y (home)	
	0	100	200
x (auto) 100	.20	.10	.20
250	.05	.15	.30

Important property: X and Y are **independent random variables** if $P(X = x, Y = y) = P(X = x)P(Y = y)$ for all possible values of x and y.

Definition: If X and Y are continuous random variables, then $f(x, y)$ is the **joint probability density function** for X and Y if $P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f(x, y)\, dx\, dy$ for all possible a, b, c, and d.

Important property: X and Y are **independent random variables** if $f(x, y) = f(x)f(y)$ for all possible values of x and y.

Example: Suppose a room is lit with two light bulbs. Let X_1 be the lifetime of the first bulb and X_2 be the lifetime of the second bulb. Suppose $X_1 \sim Exp(\lambda_1 = 1/2000)$ and $X_2 \sim Exp(\lambda_2 = 1/3000)$. If we assume the lifetimes of the light bulbs are independent of each other, find the probability that the room is dark after 4000 hours.