Discrete Mathematics and Probability

Lecture 18
Joint Probability

Ian Stark

School of Informatics
The University of Edinburgh

Monday 20 November 2023

Week 10 Topics

Chapter 3: Continuous Random Variables and Probability Distributions

§3.7 Transformations of a Random Variable.

Chapter 4: Joint Probability Distributions and Their Applications

§4.1 Jointly Distributed Random Variables

§4.2 Expected Values, Covariance, and Correlation

§4.3 Properties of Linear Combinations (but not Moment Generating Functions)

The study guide and accompanying videos indicate the examinable course content. The corresponding sections in the book are to support this. Additional sections in the book are useful extension material but not required for this course.

Review: Jointly Distributed Random Variables

Suppose that we have an experiment with sample space $\mathcal{S}$ and two random variables X and Y both defined over that sample space. That is, for any outcome $s \in \mathcal{S}$ we have real numbers $X(s)$ and $Y(s)$. Then we can define the joint probability of the two random variables.

For discrete random variables we have a joint probability mass function:

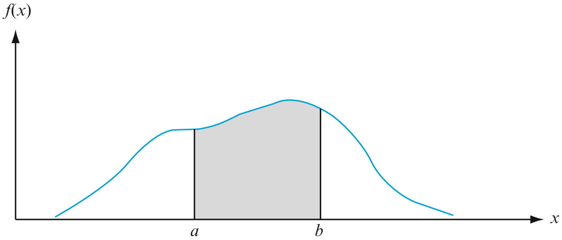
$$p(x, y) \;=\; P(X = x \text{ and } Y = y)\;.$$

For continuous random variables we have a joint probability density function $f(x, y)$.

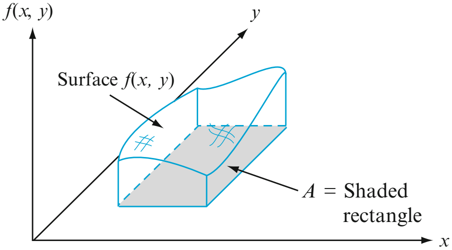
In many situations two random variables X and Y will be dependent: knowing information about one can tell us something about the other. Sometimes, though, variables are independent and the value of X says nothing about that of Y. Formally, random variables are defined as independent if their joint probability is the product of their individual marginal probabilities.

$$p(x, y) \;=\; p_X(x) \cdot p_Y(y) \quad \text{(discrete)} \qquad\qquad f(x, y) \;=\; f_X(x) \cdot f_Y(y) \quad \text{(discrete)}$$

If we treat the joint probability density function $f(x, y)$ as giving a height above the point (x, y) in a three-dimensional coordinate system then we get a surface in the same way that a single-variable PDF gives a curve.



Carlton & Devore Figure 3.2



Carlton & Devore Figure 4.1

For the curve on the left, probability is given by the area under the curve. For the surface on the right, probability is given by the volume under the surface.

For rectangle A on the right the volume under the surface is the probability $P((X, Y) \in A)$.

Expected Values

If X and Y are jointly distributed random variables and $h(X,Y)$ is some real-valued function, then $h(X,Y)$ is also a random variable. As with functions of a single random variable we can calculate its expected value without needing to directly know its probability distribution.

Proposition

For jointly distributed random variables X, Y the expected value of a function $h(X,Y)$ is given by summation or iterated integration.

$$\mu_{h(X,Y)} = E(h(X,Y)) = \begin{cases} \displaystyle\sum_x \sum_y h(x,y) \cdot p(x,y) & \text{if } X \text{ and } Y \text{ are discrete} \\ \displaystyle\int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} h(x,y) \cdot f(x,y)\, dx \right) dy & \text{if } X \text{ and } Y \text{ are continuous} \end{cases}$$

This can be extended to n multiple random variables $X_1, X_2, \cdots, X_n$ using repeated summation or integration.

Linearity of Expectation

Proposition

Given two random variables X, Y, functions h_1, h_2, and constants a_1, a_2 and b we have the following.

$$E(a_1 h_1(X,Y) + a_2 h_2(X,Y) + b) \;=\; a_1 E(h_1(X,Y)) + a_2 E(h_2(X,Y)) + b$$

Proposition

Given two independent random variables X, Y and a function $h(X,Y) = g_1(X) \cdot g_2(Y)$ for some functions g_1 and g_2, then

$$E(h(X,Y)) \;=\; E(g_1(X) \cdot g_2(Y)) \;=\; E(g_1(X)) \cdot E(g_2(Y)) \;.$$

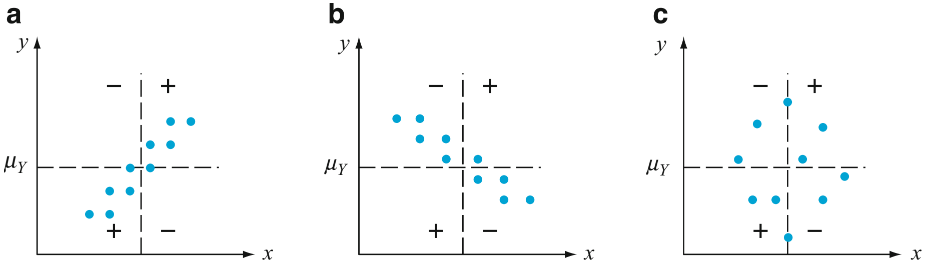
Covariance

Definition

The *covariance* between two random variables X and Y measures the extent to which they vary together (if positive) or in opposition (if negative).

$$
\begin{aligned}
\mathrm{Cov}(X,Y) &= E((X-\mu_X)(Y-\mu_Y)) \\
&= \begin{cases} \displaystyle\sum_x \sum_y (x-\mu_X)(y-\mu_Y)p(x,y) & \text{if } X \text{ and } Y \text{ are discrete} \\ \displaystyle\int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y)\cdot f(x,y)\,dx \right) dy & \text{if } X \text{ and } Y \text{ are continuous} \end{cases}
\end{aligned}
$$

Covariance



Carlton & Devore Figure 4.4

Properties of Covariance

Proposition

For any two random variables X and Y the following hold.

$$\begin{aligned}\mathrm{Cov}(X, Y) &= \mathrm{Cov}(Y, X) \\ \mathrm{Cov}(X, X) &= \mathrm{Var}(X) \\ \mathrm{Cov}(X, Y) &= E(XY) - \mu_X \cdot \mu_Y\end{aligned}$$

If Z is another random variable and a, b, c, d are constants then we also have the following.

$$\begin{aligned}\mathrm{Cov}(\, aX + bY + c,\ Z\,) &= a\,\mathrm{Cov}(X, Z) + b\,\mathrm{Cov}(Y, Z) \\ \mathrm{Cov}(\, aX + b,\ cY + d\,) &= ac\,\mathrm{Cov}(X, Y)\end{aligned}$$

Correlation Coefficient

The covariance of two random variables X and Y is like variance in that it scales as the square of the values involved, and also therefore depends on the units being used.

The following measure is based on covariance but is scale-independent.

Definition

The *linear correlation coefficient* of two random variables X and Y is defined as follows.

$$\rho_{X,Y} = \mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y}$$

It is common to simply describe this as the "correlation coefficient"; and where the random variables are known write it as just ρ.

If $\rho_{X,Y} > 0$ then X and Y are *positively correlated*. If $\rho_{X,Y} < 0$ then they are *negatively correlated*. If $\rho_{X,Y} = 0$ then they are *uncorrelated* (more specifically *not linearly correlated*).

Properties of the Correlation Coefficient

Proposition

For any two random variables X and Y the following are true.

$$\mathrm{Corr}(X, Y) = \mathrm{Corr}(Y, X)$$

$$\mathrm{Corr}(X, X) = 1$$

$$-1 \leqslant \mathrm{Corr}(X, Y) \leqslant 1$$

If a, b, c, d are constants with $ac > 0$ then we also have

$$\mathrm{Corr}(\, aX + b\, ,\, cY + d\,) = \mathrm{Corr}(X, Y)\ .$$

Correlation and Independence

Propositions

Correlation coefficient $\rho_{X,Y}$ is 0 if and only if $E(XY) = \mu_X \cdot \mu_Y$.

Correlation coefficient $\rho_{X,Y}$ is 1 if and only if $Y = aX + b$ for some constants $a > 0$ and b.

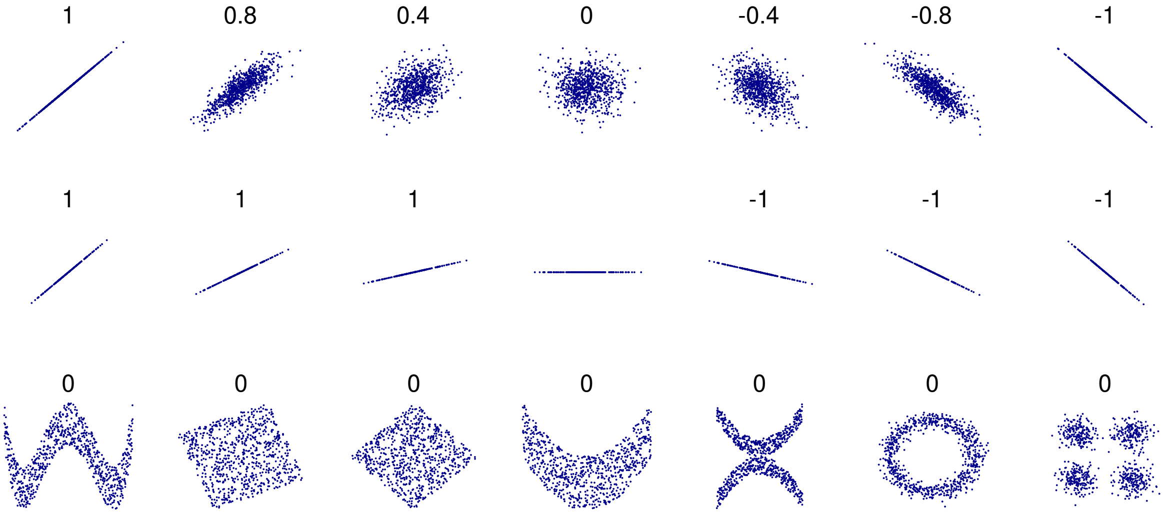
Correlation coefficient $\rho_{X,Y}$ is -1 if and only if $Y = aX + b$ for some constants $a < 0$ and b.

Two random variables X and Y are uncorrelated if $\rho_{X,Y} = 0$.

They are independent if their joint probability is the product of their marginal probabilities.

These things are not the same.

- If X and Y are independent then they are also uncorrelated (not linearly correlated).
- The reverse is not true: X and Y can be dependent even if they are not linearly correlated.


1
0.8
0.4
0
-0.4
-0.8
-1
1
1
1
-1
-1
-1
0
0
0
0
0
0
0

Correlation is Not Causation

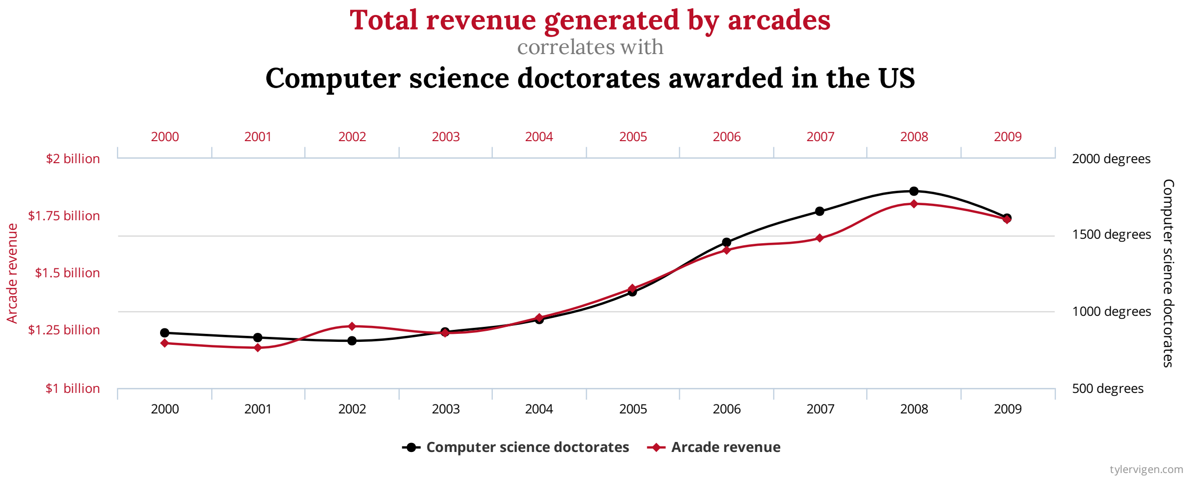
The correlation coefficient $\rho_{X,Y}$ measures one aspect of X and Y changing together. It says nothing about any possible reason *why* that may happen.

If a change in X *causes* a change in Y, then they will be correlated, $\rho_{X,Y} \neq 0$. If a change in Y *causes* a change in X, then they will be correlated. This is known as *causation*.

Correlation does not, however, necessarily imply causation. If sampling random variables X and Y shows evidence of correlation then there are several possible reasons.

- A change in X may cause a change in Y.
- A change in Y may cause a change in X.
- A change in something else may affect both X and Y.
- Chance: there is no causal connection, the correlation has no underlying cause.

In the last case, further sampling of the random variables may show no correlation after all.


Total revenue generated by arcades
correlates with
Computer science doctorates awarded in the US
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
$2 billion
$1.75 billion
$1.5 billion
$1.25 billion
$1 billion
Arcade revenue
2000 degrees
1500 degrees
1000 degrees
500 degrees
Computer science doctorates
Computer science doctorates
Arcade revenue
tylervigen.com

Week 10 Topics

Chapter 3: Continuous Random Variables and Probability Distributions

§3.7 Transformations of a Random Variable.

Chapter 4: Joint Probability Distributions and Their Applications

§4.1 Jointly Distributed Random Variables

§4.2 Expected Values, Covariance, and Correlation

§4.3 Properties of Linear Combinations (but not Moment Generating Functions)

The study guide and accompanying videos indicate the examinable course content. The corresponding sections in the book are to support this. Additional sections in the book are useful extension material but not required for this course.

Linear Combinations

If $X_1, X_2, \ldots, X_n$ are random variables then a *linear combination* is anything of the form $a_1X_1 + a_2X_2 + \cdots + a_nX_n + b$ for constants $a_1, a_2, \ldots, a_n, b$.

Proposition

$$E(a_1X_1 + \cdots + a_nX_n + b) = a_1E(X_1) + \cdots + a_nE(X_n) + b$$

$$\mathrm{Var}(a_1X_1 + \cdots + a_nX_n + b) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \,\mathrm{Cov}(X_i, X_j)$$

$$\mathrm{Var}(aX + bY) = a^2\,\mathrm{Var}(X) + b^2\,\mathrm{Var}(Y) + 2ab\,\mathrm{Cov}(X, Y)$$

Linear Combinations

If $X_1, X_2, \ldots, X_n$ are random variables then a *linear combination* is anything of the form $a_1X_1 + a_2X_2 + \cdots + a_nX_n + b$ for constants $a_1, a_2, \ldots, a_n, b$.

Proposition

If the random variables $X_1, \ldots, X_n$ are independent then we can say more.

$$\mathrm{Var}(a_1X_1 + \cdots + a_nX_n + b) = a_1^2\,\mathrm{Var}(X_1) + \cdots + a_n^2\,\mathrm{Var}(X_n)$$

$$\mathrm{SD}(a_1X_1 + \cdots + a_nX_n + b) = \sqrt{a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2}$$

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) = \mathrm{Var}(X-Y)$$

Sum of Random Variables

Proposition

Suppose X and Y are continuous random variables with JPDF $f(x, y)$. Then their sum, the random variable $W = X + Y$, has the following PDF:

$$f_W(w) \; = \int_{-\infty}^{\infty} f(x, w - x)\, dx \; .$$

If X and Y are independent then $f(x, y) = f_X(x) \cdot f_Y(y)$ for marginal PDFs $f_X(x)$ and $f_Y(y)$ giving the following:

$$f_W(w) \; = \int_{-\infty}^{\infty} f_X(x)\, f_Y(w - x)\, dx \; .$$

This method of combining two functions by integration is known as *convolution* $f_W = f_X \star f_Y$.

Example

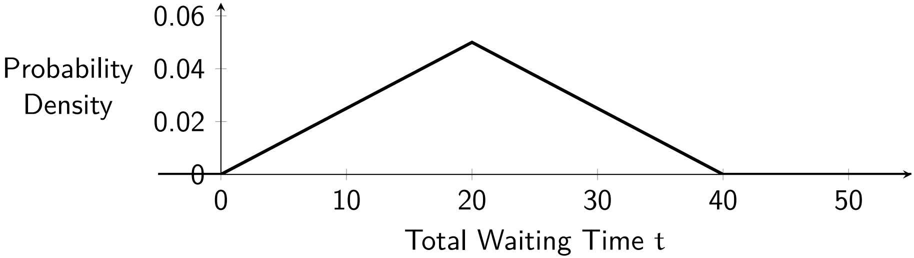
Week 9 Tutorial 7 Task 4

A commuter travels by train twice a day: heading into town in the morning and coming back home in the evening. Trains run every 20 minutes, but the commuter arrives at the station randomly and always takes the next train. This means waiting at the station between 0 and 20 minutes two times each day. These waits are distributed uniformly $\mathrm{Unif}(0, 20)$ and it turns out that the total waiting time each day in minutes is a random variable T with the following PDF.

$$f(t) = \begin{cases} \dfrac{t}{400} & 0 \leqslant t < 20 \\[2ex] \dfrac{40 - t}{400} & 20 \leqslant t < 40 \\[2ex] 0 & t < 0 \text{ or } t \geqslant 40 \end{cases}$$

Week 9 Tutorial 7 Task 4

A commuter travels by train twice a day: heading into town in the morning and coming back home in the evening. Trains run every 20 minutes, but the commuter arrives at the station randomly and always takes the next train. This means waiting at the station between 0 and 20 minutes two times each day. These waits are distributed uniformly $\mathrm{Unif}(0, 20)$ and it turns out that the total waiting time each day in minutes is a random variable T with the following PDF.



Sums of Standard Distributions

Sum of Independent Poisson

If $X_1 \dots X_n$ are independent Poisson random variables with means $\mu_1, \dots, \mu_n$ then their sum $Y = X_1 + \cdots + X_n$ also has a Poisson distribution, with mean $\mu_1 + \cdots + \mu_n$.

Sum of Independent Normal

If $N_1 \dots N_n$ are independent normal random variables with means $\mu_1, \dots, \mu_n$ and standard deviations $\sigma_1, \dots, \sigma_n$ then their sum $M = N_1 + \cdots + N_n$ is also normally distributed with mean $\mu_1 + \cdots + \mu_n$ and standard deviation $\sqrt{\sigma_1^2 + \cdots + \sigma_n^2}$.

Summary

Topics

- Expected values $E(h(X, Y))$
- Linearity of expectation
- Covariance $Cov(X, Y)$
- Correlation coefficient $Corr(X, Y)$
- Correlation is not causation!
- Linear combination of multiple random variables
- Special case of two random variables
- Special case of independent random variables
- PDF of a sum of random variables

Reading

Chapter 4, §4.2, 4.3, 4.3.1; pp. 255–270.

Exercises

Chapter 4, Exercises 23–61; pp. 262–264 and 272–276.