

Statistics for Data Science

Unit 5 Homework: Joint Distributions

September 27, 2016

1. Unladen Swallows

In the async lecture, we built a model consisting of two random variables: Let W represent the wingspan of a swallow, and V represents the velocity.

We assume W has a normal distribution with mean 10 and standard deviation 4.

We assume that $V = 0.5 \cdot W + U$, where U is a random variable (which we might call error). We assume that U has a standard normal distribution and is independent of W .

Using properties of variance and covariance, derive each element of the variance-covariance matrix for W and V .

2. Broken Rulers

You have a ruler of length 1 and you choose a place to break it using a uniform probability distribution. Let random variable X represent the length of the left piece of the ruler. X is distributed uniformly in $[0, 1]$. You take the left piece of the ruler and once again choose a place to break it using a uniform probability distribution. Let random variable Y be the length of the left piece from the second break.

- (a) Find the conditional expectation of Y given X , $E(Y|X)$.
- (b) Find the unconditional expectation of Y . One way to do this is to apply the law of iterated expectations, which states that $E(Y) = E(E(Y|X))$. The inner expectation is the conditional expectation computed above, which is a function of X . The outer expectation finds the expected value of this function.
- (c) Compute $E(XY)$. Hint: if you take an expectation conditional on a value of X , X is just a constant inside the expectation. This means that $E(XY|X) = XE(Y|X)$
- (d) Using the previous results, compute $cov(X, Y)$.

3. Great Time to Watch Async

Suppose your waiting time in minutes for the Caltrain in the morning is uniformly distributed on $[0, 5]$, whereas waiting time in the evening is uniformly distributed on $[0, 10]$. Each waiting time is independent of all other waiting times.

- (a) If you take the Caltrain each morning and each evening for 5 days in a row, what is your total expected waiting time?
- (b) What is the variance of your total waiting time?

- (c) What is the expected value of the difference between the total evening waiting time and the total morning waiting time over all 5 days?
- (d) What is the variance of the difference between the total evening waiting time and the total morning waiting time over all 5 days?

4. Maximizing Correlation

Show that if $Y = aX + b$ where X and Y are random variables and $a \neq 0$, $\text{corr}(X, Y) = -1$ or $+1$.

5. Optional Challenge Problem: Working with Poisson Variables

A Poisson random variable M is a discrete random variable with probability mass function given by

$$P_M(m) = \frac{\alpha^m}{m!} e^{-\alpha}, m = 0, 1, 2, \dots$$

Where α is a parameter.

Let N be another random variable that, conditional on $M = m$, is equally likely to take on any value in the set $0, 1, 2, \dots, m$

- Find the joint PMF of M and N
- Find the marginal PMF of N , $P_N(n)$
- Explain the significance of N in terms of a Poisson process.