Statistics for Data Science
Unit 5 Homework: Joint Distributions

September 27, 2017

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 $a \neq 0$, $corr(X, Y) = -1$ 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

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

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 $0, 1, 2, ..., m$

- 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.

W203 Unit 5 HW

Shan He

1. In order to compute the variance - covariance matrix for w and v, we need to calculate Var(w), Var(v), & Cov(w,v).

① $Var(w) = \sigma_w^2 = 4^2 = \boxed{16}$

② $Var(v) = Var(0.5w + u)$

$= E[(0.5w+u)^2] - [E(0.5w+u)]^2$

$= E(0.25w^2 + u^2 + wu) - E(0.5w)^2 - E(u)^2 - 2E(0.5w)\cdot E(u)$

since w & u are independent → $E(wu) = E(w)\cdot E(u)$

$= 0.25E(w^2) + \cancel{E(u^2)} + \cancel{E(w)}\cdot E(u) - 0.25E(w)^2 - \cancel{E(u)^2} - \cancel{E(w)\cdot E(u)}$

$= 0.25[E(w^2) - E(w)^2]$

$= 0.25\cdot Var(w)$

$= \boxed{4}$

③ $Cov(w,v) = E(w\cdot v) - E(w)\cdot E(v)$

$= E(0.5w^2 + uw) - E(w)\cdot E(0.5w+u)$

$= 0.5E(w^2) + \cancel{E(u)\cdot E(w)} - 0.5E(w)^2 - \cancel{E(w)\cdot E(u)}$

$= 0.5[E(w^2) - E(w)^2]$

$= 0.5\,Var(w)$

$= \boxed{8}$

Hence, we get a variance - covariance matrix as below:

$$\text{Matrix} = \begin{array}{c c} & \begin{array}{cc} w & v \end{array} \\ \begin{array}{c} w \\ v \end{array} & \begin{bmatrix} 16 & 8 \\ 8 & 4 \end{bmatrix} \end{array}$$

2. (a) we know that

$$f(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f(y|x) = \begin{cases} \frac{1}{1-x} & 0 \le y \le 1-x \\ 0 & \text{otherwise} \end{cases}$$

$$E(Y|X) = \int_0^{1-x} y\, f(y|x)\,dy = \int_0^{1-x} y\left(\frac{1}{1-x}\right) dy = \frac{1}{2}\left(\frac{1}{1-x}\right) y^2 \Big|_{y=0}^{y=1-x}$$

$$= \frac{1}{2}\left(\frac{1}{1-x}\right)(1-x)^2 = \boxed{\frac{1}{2}(1-x)}$$

(b) $$E(Y) = E(E(Y|X))$$

$$= E\left(\frac{1}{2}(1-x)\right)$$

$$= \int_{-\infty}^{\infty} \left(\frac{1}{2}(1-x)\right) f(x)\, dx$$

$$= \int_0^1 \left(\frac{1}{2} - \frac{x}{2}\right)\cdot 1\, dx = \frac{1}{2} - \frac{x^2}{4} \Big|_{x=0}^{x=1} = \boxed{\frac{1}{4}}$$

(c) $$E(XY) = E(E(XY|X))$$

$$= E(X\, E(Y|X))$$

$$= E\left(\frac{x}{2} - \frac{x^2}{2}\right)$$

$$= \int_0^1 \left(\frac{x}{2} - \frac{x^2}{2}\right)\cdot 1\, dx = \frac{x^2}{4} - \frac{x^3}{6} \Big|_{x=0}^{x=1} = \boxed{\frac{1}{12}}$$

(d) $$\operatorname{cov}(X,Y) = E(XY) - E(X)\cdot E(Y)$$

$$= \frac{1}{12} - \int_0^1 x \cdot 1\, dx \cdot \frac{1}{4}$$

$$= \frac{1}{12} - \frac{1}{2}\cdot\frac{1}{4} = \boxed{-\frac{1}{24}}$$

3. Let x be the waiting time in the morning
Y be the waiting time in the evening.

(a) $E[5(X+Y)] = 5E(X) + 5E(Y)$

$= 5\int_0^5 x \cdot \frac{1}{5} dx + 5 \int_0^{10} y \cdot \frac{1}{10} dy$

$= 5 \cdot (2.5) + 5(5)$

$= \boxed{37.5}$

(b) $Var[5(X+Y)] = 25 Var(X+Y) = 25 [Var(X) + Var(Y) + 2\cancel{Cov(X,Y)}]$ 0, X & Y are independent

$= 25 Var(X) + 25 Var(Y)$

$Var(X) = \int_0^5 x^2 \cdot \frac{1}{5} dx - (2.5)^2$

$= \frac{x^3}{15} \Big|_{x=0}^{x=5} - (2.5)^2 = \frac{25}{12}$

$Var(Y) = \int_0^{10} y^2 \frac{1}{10} dy - (5)^2$

$= \frac{y^3}{30} \Big|_{y=0}^{y=10} - 25$

$= \frac{25}{3}$

$Var[5(X+Y)] = 25 \cdot \frac{25}{12} + 25 \cdot \frac{25}{3} = \boxed{\frac{3125}{12}} = 260.4$

(c) $E(5Y-5X) = 5E(Y) - 5E(X) = 25 - 12.5 = \boxed{12.5}$

(d) $Var(5Y-5X) = 25 Var(Y-X) = 25 \cdot [Var(Y) + Var(X) - 2\cancel{Cov(-X,Y)}]$ 0, independent

$= 25 [Var(Y) + Var(X)]$

$= \boxed{\frac{3125}{12}}$

4. $\operatorname{corr}(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y}$; $Y = aX + b$

$\operatorname{cov}(X,Y) = E(XY) - E(X)E(Y)$.

$= E(aX^2 + bX) - E(X) \cdot E(aX + b)$.

$= aE(X^2) + bE(X) - E(X) \cdot (aE(X) + b)$

$= aE(X^2) + \cancel{bE(X)} - aE(X)^2 - \cancel{bE(X)}$

$= a(E(X^2) - E(X)^2)$

$= a \cdot \operatorname{Var}(X) = a \cdot \sigma_X^2$

$\sigma_Y = \sqrt{\operatorname{Var}(Y)} = \sqrt{\operatorname{Var}(aX + b)} = \sqrt{a^2 \operatorname{Var}(X)} = |a| \sigma_X$

hence $\operatorname{corr}(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{a \cdot \sigma_X^2}{\sigma_X \cdot (|a| \sigma_X)} = \frac{a \sigma_X^2}{|a| \sigma_X^2} = \frac{a}{|a|}$

when $a > 0$, $\operatorname{Corr}(X,Y) = 1$; when $a < 0$, $\operatorname{corr}(X,Y) = -1$

5. a. $P_M(m) = \frac{a^m e^{-a}}{m!}$ $\qquad P_{N|M}(n|m) = \frac{1}{m+1}$

$P(m,n) = P_M(m) \cdot P_{N|M}(n|m) = \frac{a^m e^{-a}}{m!} \cdot \frac{1}{m+1} = \frac{a^m e^{-a}}{(m+1)!}$

b. $P_N(n) = \sum_{m:\, P(m,n)>0} \frac{a^m e^{-a}}{(m+1)!} = \sum_0^M \frac{a^m e^{-a}}{(m+1)!}$

c. $P_N(n)$ in a poisson process is the possibility of the nth $(n<m)$ event of the m events that happen in a specified time interval, with expected of $m = a$, beng selected