Math 502AB - Lecture 14

Dr. Jamshidian

October 11, 2017

1 Lecture - Part 1

1.1 Section 4.5 - Covariances and Correlations

The idea now is that we are generally dealing with multivariate random variables. We want to come up with an index that measures the *relationship* between these two variables. As an example, you might want to examine the relationship between age and muscle mass. You might be interested in some sort of measure of association between these two.

Definition:

Let X and Y be two random variables with mean μ_X and μ_Y. Then:

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$

Definition:

The correlation between X and Y is given by:

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

Theorem:

For any X, Y

$$Cov(X,Y) = E(XY) - \mu_X\mu_Y$$

The proof of this is easy (and left to the reader).

Example:

Consider (X,Y) with the joint pdf:

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{20} & 0 < X < 10, \quad X-1 < Y < X+1 \\ 0 & o.w. \end{cases}$$

What is the correlation between X and Y?
We have:

$$
\begin{aligned}
E[XY] &= \int_0^{10}\int_{x-1}^{x+1} xy \cdot \frac{1}{20} dydx = \frac{100}{3} \\
E[X] &= \int_0^{10}\int_{x-1}^{x+1} x \cdot \frac{1}{20} dydx = 5 \\
E[Y] &= 5 \\
Cov(X,Y) &= \frac{100}{3} - 25 \\
E[X^2] &= \int_0^{10}\int_{x-1}^{x+1} x^2 \cdot \frac{1}{20} dydx = \frac{100}{3} \\
E[Y^2] &= \frac{101}{3} \\
Var(X) &= \frac{100}{3} - 25 \\
Var(Y) &= \frac{101}{3} - 25
\end{aligned}
$$

So, to calculate correlation, we have:

$$
\rho_{X,Y} = \frac{\frac{100}{3} - 25}{\sqrt{\left(\frac{100}{3} - 25\right)\left(\frac{101}{3} - 25\right)}} = 0.9806
$$

Theorem:

If X is independent of Y, then $Cov(X,Y) = 0$

Proof:

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

The converse of this is **untrue**. For example, Let (X,Y) be the coordinates of a randomly selected point on the unit circle.

We have the joint density as:

$$
f_{X,Y}(x,y) = \begin{cases} \frac{1}{\pi} & x^2 + y^2 \leq 1 \\ 0 & o.w. \end{cases}
$$

To get the *covariance* of X and Y, we have:

$$
\begin{aligned}
E(X,Y) &= \int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{xy}{\pi} dydx = 0 \\
E[X] &= E[Y] = 0 \\
Cov(X,Y) &= E[XY] - E[X]E[Y]
\end{aligned}
$$

But these are *not independent*! We know that because the joint density has an **indicator function** dependent on both X and Y.

Theorem:

If X and Y are two random variables, then for any a and b (constant), we have:

$$Var(aX + bY) = a^2 \cdot Var(x) + b^2 \cdot Var(Y) + 2ab \cdot Cov(X, Y)$$

The **proof** of this is pretty easy. We will start off with the definition that:

$$Var(aX + bY) = E\left[(aX + bY) - (a\mu_X + b\mu_Y)\right]^2$$

The rest is left as an exercise to the reader.

Theorem:

1. $-1 \leq \rho_{X,Y} \leq 1$
2. $\rho_{X,Y} = \pm 1 \iff P(Y = a + bX) = 1$ for some a and b

Proof:

1. Let $\sigma_X = Var(X)$ and $\sigma_Y = Var(Y)$

$$\begin{aligned} 0 \leq Var\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right) &= \frac{1}{\sigma_X^2} + \frac{1}{\sigma_Y^2} Var(Y) + \frac{2}{\sigma_X \sigma_Y} Cov(X, Y) \\ &= 2 + 2\rho_{X,Y} \end{aligned}$$

   But we know that
$$2\rho_{X,Y} \geq -2 \Rightarrow \rho_{X,Y} \geq -1$$

   To show that $\rho_{X,Y} \leq 1$, we use $Var\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right) \geq 0$ (This is left to the reader to verify).

2. ($\Rightarrow$) Suppose that $\rho_{X,Y} = 1$. We know that:

$$Var\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}\right) = 2 - 2\rho_{X,Y} = 0$$

   This implies that:
$$P\left(\frac{X}{\sigma_X} - \frac{Y}{\sigma_Y} = c\right) = 1$$

   A similar proof holds for $\rho_{X,Y} = -1$.

($\Leftarrow$) We start with:

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

Covariance is a **bilinear function**. This means that:

$$\begin{aligned} Cov(X,Y) &= Cov(X, a + bX) \\ &= Cov(X,a) + b \cdot Cov(X,X) \\ &= 0 + Var(X) \end{aligned}$$

In the denominator, you will repeat with $Var(Y)$, and the result follows.

1.1.1 Multivariate Normal Distribution

Suppose that:

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \right)$$

$$\begin{aligned} \sigma_{11} &= Var(X_1) \\ \sigma_{21} &= Cov(X_1, X_2) = \sigma_{12} \\ \sigma_{22} &= Var(X_2) \end{aligned}$$

And we have:

$$\rho = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}$$

But this equation is difficult to show in higher dimensions. So we have a generalization of the **normal density** as:

$$f(\vec{x}) = (2\pi)^{-1} \left|\Sigma\right|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2}(\vec{X} - \vec{\mu})^T \Sigma^{-1} (\vec{X} - \vec{\mu}) \right\}$$

Where Σ is the **covariate matrix** and $\vec{X}$ is a column vector.

1.1.2 Constant Value Contours

The solutions of the system of equations

$$(\vec{X} - \vec{\mu})^T \Sigma^{-1} (\vec{X} - \vec{\mu}) = c$$

are referred to a **constant value contours**. In the case of the *bivariate normal distribution*, these are easily seen as ellipses.

1.2 Section 4.6 - Multivariate R.V.

When we say *multivariate random variable* we are talking about a variable

$$\vec{X} = (X_1, X_2, ..., X_n)$$

Where:

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, ..., x_n) dx_1 dx_2 \cdots dx_n = 1$$
$$P\left(X \in A \subset \mathbb{R}^n\right) = \int_A \cdots \int_A f(x_1, x_2, ..., x_n) dx_1 dx_2 \cdots dx_n$$
$$E[g(\vec{X})] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\vec{X}) f(x_1, x_2, ..., x_n) dx_1 dx_2 \cdots dx_n = 1$$

We have the conditional density:

$$f(x_{k+1}, ..., x_n | x_1, ..., x_k) = \frac{f(x_1, ..., x_n)}{f(x_1, ..., x_k)}$$

Example:

Let X_1, X_2, X_3 be jointly distributed with pdf

$$\begin{aligned} f(x_1, x_2, x_3) = &\frac{2}{3}(x_1 + x_2 + x_3) \\ &0 < X_1 < 1 \\ &0 < X_2 < 1 \\ &0 < X_3 < 1 \end{aligned}$$

The **marginal** distribution of x_1 is as follows:

$$\begin{aligned} f_{X_1}(x_1) &= \int_0^1 \int_0^1 \frac{2}{3}(x_1 + x_2 + x_3) dx_2 dx_3 \\ &= \frac{2}{3}x_1 + \frac{2}{3} \end{aligned}$$

The **marginal** distribution of (x_1, x_2) is:

$$\begin{aligned} f_{X_1, X_2}(x_1, x_2) &= \int_0^1 \frac{2}{3}(x_1 + x_2 + x_3) dx_3 \\ &= \frac{2}{3}x_1 + \frac{2}{3}x_2 + \frac{1}{3} \end{aligned}$$

And we have:

$$\begin{aligned} P\left[0 < X_1 < \frac{1}{2}, 0 < X_2 + X_3 < \frac{1}{2}\right] &= \int_0^{\frac{1}{2}} \int_0^{\frac{1}{2}} \int_0^{\frac{1}{2} - x_2} \frac{2}{3}(x_1 + x_2 + x_3) dx_3 dx_2 dx_1 \\ &= \frac{7}{288} \end{aligned}$$

Finding expected values:

$$E[X_1^2X_2X_3 + 3X_1X_2^2X_3^2] = \int_0^1\int_0^1\int_0^1 \frac{2}{3}(x_1^2x_2x_3 + 3x_1x_2x_3^2)(x_1+x_2+x_3)dx_1dx_2dx_3$$
$$= \frac{15}{8}$$

Finding conditional values:

$$f_{x_2,x_3|x_1}(x_2,x_3|x_1) = \frac{\frac{2}{3}(x_1+x_2+x_3)}{\frac{2}{3}(x_1+1)}$$

Subject to $0 \leq X_2 \leq 1$ and $0 \leq x_3 \leq 1$. And finding

1.3 The Multinomial Distribution

Suppose that m indistinguishable items are to be distributed among n groups. Furthermore, suppose that the probability that an item falls into the i^{th} group is p_i. Let X_i be the number of items that fall into the i^{th} group. Then the random vector $(X_1, X_2, ..., X_n)$ is called a **multinomial random variable** with joint *pmf*

$$f(x_1, ..., x_m) = \binom{m}{x_1, ..., x_n} p_1^{x_1}p_2^{x_2}\cdots p_n^{x_n}$$

Where

$$\sum_{i=1}^{n} x_i = m$$
$$\sum_{i=1}^{n} p_i = 1$$
$$\binom{m}{x_1, ..., x_n} = \frac{m!}{x_1!\cdots x_n!}$$

Example:

Three fair dice are cast in 10 *independent* casts. Let:

$X_1 =$ # of times that none of the dice match.
$X_2 =$ # of times that exactly two dice match.
$X_3 =$ # of times that all three dice match.

Write the *pmf* of (X_1, X_2, X_3).
First, we note that:

$$p_3 = P(\text{all three match}) = \frac{6}{6^3} = \frac{1}{36}$$
$$p_2 = P(\text{exactly two match}) = \frac{\binom{3}{2} \cdot 6 \cdot 5}{6^3} = \frac{15}{36}$$
$$p_1 = P(\text{no matches}) = \frac{6 \cdot 5 \cdot 4}{6^3} = \frac{20}{36}$$

Thus, our *pmf* is:

$$f_{X_1,X_2,X_3}(x_1,x_2,x_3) = \binom{10}{x_1,x_2,x_3}\left(\frac{20}{36}\right)^{x_1}\left(\frac{15}{36}\right)^{x_2}\left(\frac{1}{36}\right)^{x_3}$$

with the constraints $x_1 + x_2 + x_3 = 10$ and $x_1, x_2, x_3 \geq 0$.

2 Lecture - Part 2

2.1 Marginal Distribution

Let $(X_1, ..., X_n) \sim MN(m, p_1, ..., p_n)$, a multinomial distribution. Obtain the marginal distribution of X_1. We know that X_1 is the number of balls in box 1. This follows a *binomial*(m, p_1) distribution.

Alternative Solution

We need to sum over

$$\mathcal{B} = \left\{(x_2, x_3, ..., x_n) : x_1 \subset \mathbb{Z}, x_i \geq 0, \sum_{i=2}^{n} x_i = m - x_1\right\}$$

Thus we have:

$$\begin{aligned}
f_{X_1}(x_1) &= \sum_{\mathcal{B}} \frac{m!}{x_1!x_2!\cdots x_n!} p_1^{x_1} p_2^{x_2} \cdots p_n^{x_n} \\
&= \frac{m!}{x_1!} p_1^{x_1} \sum_{\mathcal{B}} \frac{1}{x_2!\cdots x_n!} p_2^{x_2} \cdots p_n^{x_n} \\
&= \frac{m!}{x_1!(m-x_1)!} p_1^{x_1} \sum_{\mathcal{B}} \frac{(m-x_1)!}{x_2!\cdots x_n!} p_2^{x_2} \cdots p_n^{x_n} \\
&= \frac{m!}{x_1!(m-x_1)!} p_1^{x_1} \sum_{\mathcal{B}} \frac{(m-x_1)!}{x_2!\cdots x_n!} \frac{p_2^{x_2}}{(1-p_1)^{x_2}} \cdots \frac{p_n^{x_n}}{(1-p_1)^{x_n}}
\end{aligned}$$

Since our summation is $MN(m - x_1, \frac{p_2}{1-p_1}, ..., \frac{p_n}{1-p_1})$, it adds to 1 and our marginal distribution is:

$$f_{X_1}(x_1) = \frac{m!}{x_1!(m-x_1)!} p_1^{x_1} = \binom{m}{x_1} p_1^{x_1} (1-p_1)^{m-x_1}$$

2.2 Conditional Distribution

What is the *conditional distribution* of $X_2, ..., X_n | X_1 = x_1$?

$$
\begin{aligned}
f_{X_2,...,X_n|X_1}(x_2, ..., x_n|x_1) &= \frac{f(x_1, ..., x_n)}{f_{X_1}(x_1)} \\
&= \frac{\frac{m!}{x_1! \cdots x_n!} p_1^{x_1} \cdots p_n^{x_n}}{\frac{m!}{x_1!(m-x_1)!} p_1^{x_1} (1-p_1)^{m-x_1}} \\
&= \frac{(m-x_1)!}{x_2! \cdots x_n!} \left(\frac{p_2}{1-p_1}\right)^{x_2} \cdots \left(\frac{p_n}{1-p_1}\right)^{x_n}
\end{aligned}
$$

Definition:

Let $\vec{X}_1, ..., \vec{X}_n$ be random vectors with joint pdf's (or pmf's) $f(\vec{x_1}, ..., \vec{x_n})$. Let $f_{\vec{X}_i}(\vec{x}_i$ denote the marginal pdf (or pmf) of $\vec{X}_i$. Then $\vec{X}_1, ..., \vec{X}_n$ are called **mutually independent** if:

$$f(\vec{x}_1, ..., \vec{x}_n) = f_{\vec{x}_1}(\vec{x}_1) \cdots f_{\vec{x}_n}(\vec{x}_n)$$

But note that **pairwise independence** *does not imply* **mutual independence**.

Example:

Two dice are rolled. Let

1. $A =$ Getting a sum of 7
2. $B =$ Getting a 4 on the 1^{st} die
3. $C =$ Getting a 5 on the 2^{nd} die

$$X_1 = I_A \quad X_2 = I_B \quad X_3 = I_C$$

These random variables are pairwise independent.

$$P(X_1 = 1 \cap X_2 = 2) = P(X_1 = 1)P(X_2 = 1)$$
$$\frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6}$$

This is true! However the following equality **does not hold**

$$P(X_1 = 1, X_2 = 1, X_3 = 1) = P(X_1 = 1)P(X_2 = 1)P(X_3 = 1)$$

Theorem:

Let $X_1, ..., X_n$ be mutually independent random variables. Let $g_1, ..., g_n$ be n functions with each of the g_i's being only the function of X_i. Then:

$$E[g_1(X_1)g_2(X_2) \cdots g_n(X_n)] = E[g_1(X_1)]E[g_2(X_2)] \cdots E[g_n(X_n)]$$

Theorem:

Let $X_1, ..., X_n$ be mutually independent. Then, if $Z = X_1 + X_2 + ... + X_n$

$$M_Z(t) = M_{X_1}(t)M_{X_2}(t) \cdots M_{X_n}(t)$$

Example:

Let $X_1, ..., X_n \sim Poisson(\lambda)$ (**iid**). What is the distribution of $X_1+X_2+...+X_n$?

$$\begin{aligned} M_{X_1+...+X_n}(t) &= M_{X_1}(t)M_{X_2}(t) \cdots M_{X_n}(t) \\ &= [M_{X_1}(t)]^n = \left[e^{\lambda(e^t-1)}\right]^n \\ &= e^{n\lambda(e^t-1)} \sim poisson(n\lambda) \end{aligned}$$

Theorem:

Let $X_1, ..., X_n$ be mutually independent. Then for $Z = (a_1x_1 + b_1) + (a_2x_2 + b_2) + ... + (a_nx_n + b_n)$, we have:

$$M_Z(t) = e^{t\sum_{i=1}^n b_i} M_{X_1}(a_1t)M_{X_2}(a_2t) \cdots M_{X_n}(a_nt)$$

Example:

Let $X_1, ..., X_n$ be mutually independent with $X_i \sim N(\mu_i, \sigma_i^2)$. Then:

$$Z = \sum_{i=1}^n (a_ix_i + b_i) \sim N\left(\sum_{i=1}^n (a_i\mu_i + b_i), \sum_{i=1}^n a_i^2\sigma_i^2\right)$$

Proof:

$$\begin{aligned} M_Z(t) &= e^{t\sum_{i=1}^n b_i} e^{\mu_1 a_1 t + \sigma_1^2 a_1^2 \frac{t^2}{2}} \cdots e^{\mu_n a_n t + \sigma_n^2 a_n^2 \frac{t^2}{2}} \\ &= e^{t\left(\sum_{i=1}^n b_i + \sum_{i=1}^n \mu_i \sigma_i\right)} \exp\left\{\frac{t^2}{2}(\sigma_1^2 a_1^2 + ... + \sigma_n^2 a_n^2)\right\} \end{aligned}$$

This is the moment generating function where the term on the left is the *mean*, and the term on the right is the *variance*.

Theorem:

$$Var\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n Var(X_i) + \sum_{i=1}^n \sum_{j=1} Cov(X_i, X_j), \quad i \neq j$$

If $X_1, ..., X_n$ are independent, then:

$$Var\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n Var(X_i)$$

Example (Multivariate Transformation):

Let $Y_1, Y_2, Y_3, Y_4 \sim \exp(1)$ (**iid**). Let:

$$\begin{aligned} X_1 &= \text{ smallest value } = min(Y_1, Y_2, Y_3, Y_4) \\ X_2 &= \text{ second smallest value} \\ X_3 &= \text{ third smallest value} \\ X_4 &= \text{ fourth smallest value} \end{aligned}$$

These are referred to as *order statistics* (we will discuss this later). It can be shown that

$$f(x_1, x_2, x_3, x_4) = 24e^{-x_1-x_2-x_3-x_4} \quad 0 < x_1 < x_2 < x_3 < x_4$$

Consider the following random variables:

$$\begin{aligned} U_1 &= X_1 \leftarrow \text{ time of first death} \\ U_2 &= X_2 - X_1 \leftarrow \text{ time between } 1^{st} \text{ and } 2^{nd} \text{ death} \\ U_3 &= X_3 - X_2 \leftarrow \text{ time between } 2^{nd} \text{ and } 3^{rd} \text{ death} \\ U_4 &= X_4 - X_3 \leftarrow \text{ time between } 3^{rd} \text{ and } 4^{th} \text{ death} \end{aligned}$$

Obtain the joint distribution of $U_1, ..., U_4$.

$$f_{U_1,U_2,U_3,U_4}(u_1, u_2, u_3, u_4) = f_{X_1,X_2,X_3,X_4}(x_1, x_2, x_3, x_4)|J^{-1}(x_1, ..., x_4)|$$

Where $X_i = h(u_1, u_2, u_3, u_4)$ and the **jacobian** J is:

$$J = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix}$$

with $det(J) = 1$. If you take this and solve for X_i, you will get:

$$\begin{aligned} X_1 &= U_1 \\ X_2 &= U_2 + U_1 \\ X_3 &= U_3 + U_2 + U_1 \\ X_4 &= U_4 + U_3 + U_2 + U_1 \end{aligned}$$

So we have

$$f_{U_1,U_2,U_3,U_4}(u_1, u_2, u_3, u_4) = 24 \exp\{-4U_1 - 3U_2 - 2U_3 - U_4\}$$