

Week 6 – Random vectors and independence

Juan Pablo Lewinger

Last class

- Expectation

- $E[X] = \int_{-\infty}^{+\infty} x f_X(x) dx$ (Continuous)

- $E[X] = \sum_{x_i} x_i f_X(x_i)$ (Discrete)

- $E[X] = w_d \sum_{x_i} x_i f_X(x_i) + w_c \int_{-\infty}^{+\infty} x f_X(x) dx, \quad w_d + w_c = 1$ (mixed)

- Variance

- $Var[X] = E[(X - E[X])^2] = E[X^2] - E^2[X]$

- Change of variable formula/LOTUS

- $E[g(X)] = \int_{-\infty}^{+\infty} g(x) f_X(x) dx$ (Continuous)

- $E[g(X)] = \sum_{x_i} g(x_i) P_X(x_i)$ (Discrete)

Random vectors

We are typically interested in not one, but multiple related random variables defined on the same space.

- X, Y random variables or (X, Y) a random vector in \mathbb{R}^2
 - E.g. weight and height of a randomly selected person
- Geometrically, (X, Y) represents a random point in \mathbb{R}^2
- More generally, a random vector in \mathbb{R}^n , $\mathbf{X} = (X_1, \dots, X_n)$
 - E.g. expression levels of n genes for an individual
- We can have discrete, continuous, and mixed random vectors

Discrete random vectors

A random vector (X, Y) is discrete if it takes a finite or countable number of values

$$R_{X,Y} = \{(x_1, y_1), (x_2, y_2), \dots\}$$

Joint probability mass function:

- $P_{X,Y}(x, y) = P(X = x, Y = y) = P(X = x \cap Y = y)$
- $P(X \in A \cap R_X, Y \in B \cap R_Y) = \sum_{x_i \in A} \sum_{y_j \in B} P_{X,Y}(x_i, y_j)$

In general, if $C \subset \mathbb{R}^2$,

- $P((X, Y) \in C) = \sum_{(x_i, y_j) \in C \cap R_{X,Y}} P_{X,Y}(x_i, y_j)$

X and Y are discrete random variables $\iff (X, Y)$ is a discrete random vector

Continuous random vectors

A random vector is continuous if it has a joint probability density function:

- $f_{X,Y}(x, y) \geq 0$,
- $\iint_{\mathbb{R}^2} f_{X,Y}(x, y) dx dy = 1$
- $P((X, Y) \in C) = \iint_C f_{X,Y}(x, y) dx dy$

(X, Y) continuous as a random vector $\implies X$ and Y continuous as individual random variables

Converse is not true: X and Y are continuous $\not\Rightarrow$ (X, Y) continuous as a random vector (it may not have a density)

Random vectors

The cumulative distribution function (cdf) is defined for both discrete and continuous (and mixture) random vectors:

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y) = \begin{cases} \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(s, t) ds dt & \text{Continuous} \\ \sum_{x_i \leq x} \sum_{y_j \leq y} P_{X,Y}(x_i, y_j) & \text{Discrete} \end{cases}$$

Just like for random variables, the pdf (continuous), the pmf (discrete), or the cdf (both), completely characterize probabilistically a random vector

For a continuous random vector, $f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y)$

The distribution (pdf, pmf or cdf) of the component random variables X and Y are called the marginal distribution of X and Y respectively

Marginal distributions

(X, Y) a random vector

- $F_X(x) = \lim_{y \rightarrow +\infty} F_{X,Y}(x, y) \quad F_Y(y) = \lim_{x \rightarrow +\infty} F_{X,Y}(x, y)$

For a continuous random vector:

- $f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy \quad f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dx$

For a discrete random vector:

- $P_X(x_i) = \sum_{y_j} P_{X,Y}(x_i, y_j) \quad P_Y(y_j) = \sum_{x_i} P_{X,Y}(x_i, y_j)$

Example: discrete random vector

Let M and S be the minimum and the sum of two independent rolls of fair 3-faced die

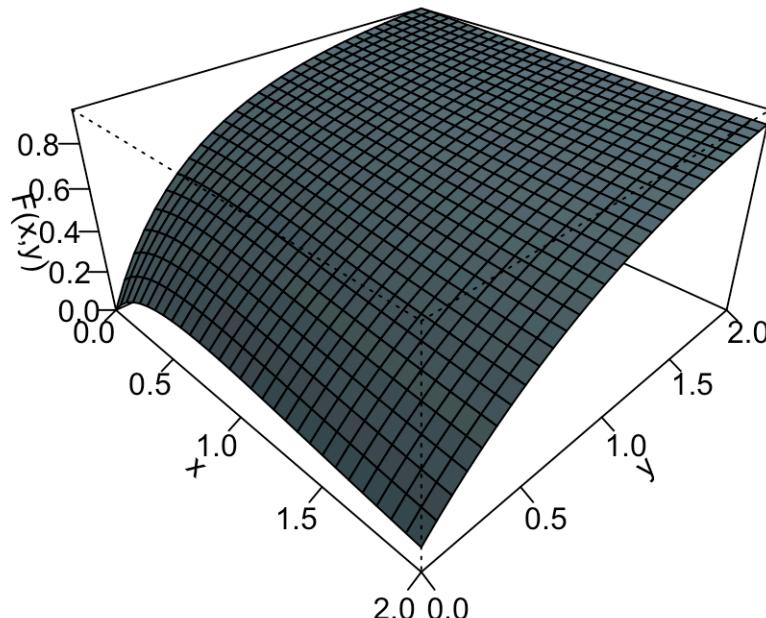
Determine:

- The joint pmf of M and S ,
- The marginal pmf of M and of S .

Example: continuous random vector

Suppose that the joint cumulative distribution function of (X, Y) is given by:

$$F_{X,Y}(x, y) = \begin{cases} 1 - e^{-2x} - e^{-y} + e^{-(2x+y)} & \text{if } x > 0, y > 0 \\ 0 & \text{Otherwise} \end{cases}$$



Example: continuous random vector

$$F_{X,Y}(x,y) = \begin{cases} 1 - e^{-2x} - e^{-y} + e^{-(2x+y)} & \text{if } x > 0, y > 0 \\ 0 & \text{Otherwise} \end{cases}$$

1. Determine the joint probability density function of X and Y .
2. Determine the marginal cumulative distribution functions of X and Y .
3. Determine the marginal probability density functions of X and Y .
4. Find out whether X and Y are independent.
5. Determine $Cov(X, Y)$ and $\rho(X, Y)$

Example: continuous random vector

Suppose that the joint probability density function of X and Y is given:

$$f_{X,Y}(x,y) = \begin{cases} x + cy^2 & \text{if } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{Otherwise} \end{cases}$$

1. Find the constant c .
2. Determine the joint cumulative distribution functions of (X, Y) .
3. Determine the marginal probability density functions of X and Y .
4. Find out whether X and Y are independent.
5. Determine $\text{Cov}(X, Y)$ and $\rho(X, Y)$

Independence of random variables

X and Y are independent if for any $A, B \subset \mathbb{R}$, $\{X \in A\}$ and $\{Y \in B\}$ are independent events:

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

- Equivalent to the factorization of the join cdf as a product of the marginal cdfs:
 $F_{X,Y}(x, y) = F_X(x)F_Y(y)$
- For continuous random vectors, also equivalent to the factorization of the join pdf as a product of the marginal pdfs: $f_{X,Y}(x, y) = f_X(x)f_Y(y)$
- For discrete random vectors, also equivalent to the factorization of the join pmf as a product of the marginal pmfs: $P_{X,Y}(x, y) = P_X(x)P_Y(y)$
- Definition of independence and factorization equivalences extend to multiple random variables X_1, \dots, X_n

Propagation of independence

NOTE this is important and not covered in the book

- If X , and Y are independent so are $g(X)$ and $h(Y)$ for $g, h : \mathbb{R} \rightarrow \mathbb{R}$
- If X_1, X_2, \dots, X_n are independent so are $h_1(X_1), \dots, h_n(X_n)$, for $h_i : \mathbb{R} \rightarrow \mathbb{R}$
- If $X_1, X_2, \dots, X_n, Y_1, Y_2, \dots, Y_m$ are independent then, $g(X_1, X_2, \dots, X_n), h(Y_1, Y_2, \dots, Y_m)$, where $g : \mathbb{R}^n \rightarrow \mathbb{R}, h : \mathbb{R}^m \rightarrow \mathbb{R}$

Examples:

1. $X_1, X_2, X_3, X_4, Y_1, Y_2$ independent $\Rightarrow Z_1 = \frac{\sin(X_1^2) + e^{X_2}}{X_3^5 + 1}, Z_2 = \cos(Y_1) - Y_2^3$ are independent

2. $X_1, X_2, \dots, X_n \stackrel{iid}{\sim} \text{Bernoulli}(p), Y_1, Y_2, \dots, Y_m \stackrel{iid}{\sim} \text{Bernoulli}(q)$

(iid stands for independent identically distributed)

Let $X = X_1 + X_2 + \dots + X_n, Y = Y_1 + Y_2 + \dots + Y_m$.

Then $X \sim \text{Binomial}(n, p)$ and $Y \sim \text{Binomial}(m, p)$ and X, Y are *independent*

Expectation of a random vector

$$E[(X, Y)] = (E[X], E[Y])$$

Interpretation is analog to that for random variables, ‘center’ of the two-dimensional distribution, center of mass if we think of probability as mass distributed on the surface of the plane \mathbb{R}^2

In general, $E[\mathbf{X}] = (E[X_1], \dots, E[X_n])$

Example: $X \sim \text{Exp}(\lambda_1)$ $Y \sim \text{Exp}(\lambda_2)$

$$E[(X, Y)] = (E[X], E[Y]) = \left(\frac{1}{\lambda_1}, \frac{1}{\lambda_2} \right)$$

Multi-dimensional LOTUS

$\mathbf{X} = (X_1, \dots, X_n)$ a random vector, $g : \mathbb{R}^n \rightarrow \mathbb{R}$ a function, then:

$$E[g(\mathbf{X})] = E[g(X_1, \dots, X_n)] = \begin{cases} \int \cdots \int_{\mathbb{R}^n} g(x_1, \dots, x_n) f_{\mathbf{X}}(x_1, \dots, x_n) dx_1 \dots dx_n & \text{if } \mathbf{X} \text{ continuous} \\ \sum_{\mathbf{x}_i} g(\mathbf{x}_i) f_{\mathbf{X}}(\mathbf{x}_i) & \text{if } \mathbf{X} \text{ discrete} \end{cases}$$

Consequences:

- Expectation is linear: $E[a_1 X_1 + a_2 X_2 + \dots + a_n X_n] = a_1 E[X_1] + a_2 E[X_2] + \dots + a_n E[X_n]$
- $E[XY] = E[X]E[Y]$ for independent random variables X, Y

Example of linearity: $X \sim \text{Binomial}(n, p)$

Then, $X = X_1 + \dots + X_n$, where $X_1, X_2, \dots, X_n \stackrel{iid}{\sim} \text{Bernoulli}(p)$

$$E[X] = E[X_1] + \dots + E[X_n] = \underbrace{p + \dots + p}_{n \text{ times}} = np$$

Covariance

- For arbitrary random variables X and Y :

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2E[(X - E[X])(Y - E[Y])]$$

- The term $\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$ is called the covariance of X and Y
 - It measures how much X and Y co-vary, i.e. vary together
 - $\text{Cov}(X, Y) > 0$ if whenever $X > E[X]$ is likely that also $Y > E[Y]$ and viceversa (and that when $X < E[X]$ also more likely that $Y < E[Y]$)
 - $\text{Cov}(X, Y) < 0$ if whenever $X > E[X]$ is likely that $Y < E[Y]$ and viceversa (and that when $X < E[X]$ also more likely that $Y > E[Y]$)
- X, Y independent $\implies \text{Cov}(X, Y) = 0$
- Converse is not true: $\text{Cov}(X, Y) = 0 \not\implies X, Y$ independent

Covariance

Example of additivity of variance for uncorrelated RVs

$$X \sim \text{Binomial}(n, p)$$

Then, $X = X_1 + \dots + X_n$, where $X_1, X_2, \dots, X_n \stackrel{iid}{\sim} \text{Bernoulli}(p)$

$$\begin{aligned} \text{Var}[X] &= \text{Var}[X_1] + \dots + \text{Var}[X_n] + \sum_{i < j} 2\text{Cov}(X_i, X_j) = \\ &= p(1-p) + \dots + p(1-p) = np(1-p) \\ &\quad \text{n times} \end{aligned}$$

Properties of covariance

- Covariance is linear in each of its terms:
 - $\text{Cov}(aX + bY, cU + dV) = ac\text{Cov}(X, U) + ad\text{Cov}(X, V) + bc\text{Cov}(Y, U) + bd\text{Cov}(Y, V)$
 - $\text{Cov}(X, X) = \text{Var}[X] (\Rightarrow \text{Var}[aX] = a^2 \text{Var}[X])$
- Cauchy-Schwartz inequality
 - $|E[XY]| \leq \sqrt{E[X^2]E[Y^2]}$
- From Cauchy-Schwartz inequality using $X - EX$ and $Y - E[Y]$ in place of X and Y , it follows that
 - $|\text{Cov}(X, Y)| \leq \sqrt{\text{Var}[X]\text{Var}[Y]} = \sigma_X \sigma_Y$ ($\sigma_X = \sqrt{\text{Var}[X]}, \sigma_Y = \sqrt{\text{Var}[Y]}$ are called the standard deviations of X and Y respectively)
 - Equivalently, $-1 \leq \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}[X]\text{Var}[Y]}} \leq 1$

Correlation

$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var[X]Var[Y]}}$ is the correlation between X and Y

- X, Y independent $\Rightarrow \rho(X, Y) = 0$
- Converse is not true: $\rho(X, Y) = 0 \not\Rightarrow X, Y$ independent
- $\rho(X, Y)$ is a standardized version of $Cov(X, Y)$
- $\rho(X, Y)$ is unaffected by changes of units: $\rho(aX + b, cY + d) = \rho(X, Y)$
- Covariance it's not invariant: $Cov(X, Y) = \rho(X, Y)\sigma_X\sigma_Y$, so the larger σ_X and σ_Y , the larger $Cov(X, Y)$ in absolute value (provided $\rho(X, Y) \neq 0$)

Next two weeks

- No class next week
- For class after next week read Ch 6.1.1-6.1.2, 6.2.0-6.2.3, 7.1.0-7.1.1 from 'Introduction to Probability, Statistics, and Random Processes'
- Do the corresponding exercises!