STAT 390 A
Statistical Methods in Engineering and Science

Week 5 Lectures – Part 1 – Spring 2023

Joint Distributions and Independence of Random Variables

Alexander Giessing

Department of Statistics

University of Washington

April 24, 2023

Outline

How to study many random variables?

- How to predict or control one random variable using another?
  - number of COVID-19 cases last week and this week.
  - education level and income.
  - SP500 and IBM stocks.
- What does it mean to say that two random variables are independent?
- How to compute the expectation and variance of several random variables?
- How to describe the dependence (i.e. covariance and correlation) between several random variables?

Example: Sum and Maximum of Two Dice

- Denote by S the sum and by M the maximum of two fair dice.
- To describe the **joint event**

$$\{(S,M)=(a,b)\} := \{S=a\}\cap\{M=b\},$$

we need to describe how the probability mass is distributed over the range of (S,M), i.e.

$$\big\{(a,b) : a\in\{1,2,\ldots,12\},\ b\in\{1,2,\ldots,6\}\big\}.$$

- The (marginal) pmfs of M and S are given by

M	1	2	3	4	5	6
p_M	1/36	3/36	5/36	7/36	9/36	11/36

S	2	3	4	5	6	7	8	9	10	11	12
p_S	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Example: Sum and Maximum of Two Dice

(Derivation of p_M and p_S.)

Example: Sum and Maximum of Two Dice (Cont.)

The joint probability mass function

$$p_{S,M}(a,b) = P(\{S = a\} \cap \{M = b\})$$

is given by

$p_{S,M}$		M					
		1	2	3	4	5	6
	2	1/36	0	0	0	0	0
	3	0	2/36	0	0	0	0
	4	0	1/36	2/36	0	0	0
	5	0	0	2/36	2/36	0	0
	6	0	0	1/36	2/36	2/36	0
S	7	0	0	0	2/36	2/36	2/36
	8	0	0	0	1/36	2/36	2/36
	9	0	0	0	0	2/36	2/36
	10	0	0	0	0	1/36	2/36
	11	0	0	0	0	0	2/36
	12	0	0	0	0	0	1/36

Joint Probability Mass Function

The joint distribution of two discrete random variables X and Y, defined on the same sample space Ω, is given by prescribing the probabilities of all possible values of the pair (X, Y).

Joint Probability Mass Function

Let X and Y be two discrete random variables defined on the same sample space Ω. Their joint probability mass function $p : \mathbb{R}^2 \to [0, 1]$ is defined by

$$p(a, b) := P(\{X = a\} \cap \{Y = b\}) = P(X = a, Y = b) \quad \text{for } -\infty < a, b < \infty.$$

- The joint pmf satisfies:
  - $0 \leq p(a, b) \leq 1$ for all $-\infty < a, b < \infty$.
  - $\sum_{a \in \mathrm{Range}(X)} \sum_{b \in \mathrm{Range}(Y)} p(a, b) = 1$.
  - $\sum_{(a,b) \in A} p(a, b) = P((X, Y) \in A)$ for all $A \subset \mathbb{R}^2$.

Marginal Probability Mass Function

Marginal Probability Mass Function

Let p be the joint probability mass function of random variables X and Y. The marginal probability mass functions of X and Y are defined by

$$p_X(a) := \sum_{b \in \mathrm{Range}(Y)} p(a,b) \quad \text{and} \quad p_Y(b) := \sum_{a \in \mathrm{Range}(X)} p(a,b).$$

- Recall the example of the sum and maximum of two dice. Since

$$\{S=6\} = \{S=6, M=1\} \cup \{S=6, M=2\} \cup \ldots \cup \{S=6, M=6\}$$

and because all six events on the RHS are mutually exclusive, we have

$$\begin{aligned} p_S(6) &= P(S=6) = P(S=6, M=1) + \ldots + P(S=6, M=6) \\ &= \ldots = 5/36. \end{aligned}$$

Example: Sum and Maximum of Two Dice (Cont.)

We retrieve the marginal probability mass functions p_S (and p_M) by summing over the columns (and rows) of the joint probability mass function.

$p_{S,M}$		M						p_S
		1	2	3	4	5	6	
	2	1/36	0	0	0	0	0	1/36
	3	0	2/36	0	0	0	0	2/36
	4	0	1/36	2/36	0	0	0	3/36
	5	0	0	2/36	2/36	0	0	4/36
	6	0	0	1/36	2/36	2/36	0	5/36
S	7	0	0	0	2/36	2/36	2/36	6/36
	8	0	0	0	1/36	2/36	2/36	5/36
	9	0	0	0	0	2/36	2/36	4/36
	10	0	0	0	0	1/36	2/36	3/36
	11	0	0	0	0	0	2/36	2/36
	12	0	0	0	0	0	1/36	1/36
p_M		1/36	3/36	5/36	7/36	9/36	11/36	1

Joint Distribution Function

Joint Distribution Function (Discrete Random Variables)

The joint (cumulative) distribution function F of two discrete random variables X and Y is the function $F : \mathbb{R}^2 \to [0, 1]$ defined by

$$F(a, b) := P(X \leq a, Y \leq b) = \sum_{x \leq a} \sum_{y \leq b} p(x, y) \quad \text{for } -\infty < a, b < \infty.$$

- If X and Y have range $a_1, a_2, \ldots$ and $b_1, b_2, \ldots$, respectively, then

$$F(a, b) = \sum_{a_i \leq a} \sum_{b_j \leq b} p(a_i, b_j).$$

- Recall the example of the sum and maximum of two dice. Verify at home that $F_{S,M}(5, 3) = 8/36$.

Marginal Distribution Function

Marginal Distribution Function

Let F be the joint (cumulative) distribution function of random variables X and Y. Then the marginal distribution function of X is given for each a by

$$F_X(a) := P(X \leq a) = F(a, +\infty) = \lim_{b \to \infty} F(a, b),$$

and the marginal distribution function of Y is given for each b by

$$F_Y(b) := P(Y \leq b) = F(+\infty, b) = \lim_{a \to \infty} F(a, b).$$

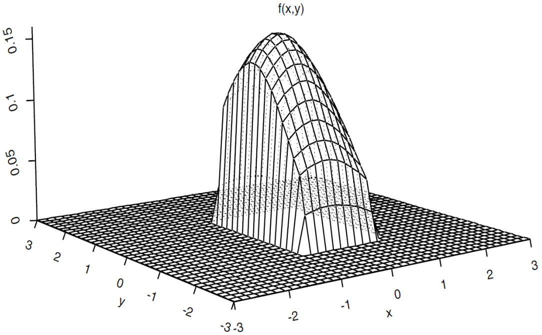
- Recall the example of the sum and maximum of two dice. Verify at home that $F_S(5) = F_{S,M}(5, +\infty) = F_{S,M}(5, 6) = 10/36$.
- This definition also applies to continuous random variables (see below)!

Outline

Joint Continuous Distribution

- Recall: *The probability that a single continuous RV X lies in the interval $[a, b]$ is equal to the area under the pdf of X over the interval.*

- Analgous for joint distribution of continuous RVs X and Y:
  The probability that the pair (X, Y) fall in the rectangle $[a_1, b_1] \times [a_2, b_2]$ is equal to the volume under the joint pdf of (X, Y) over the rectangle.



(Volume under a joint pdf f on the rectangle $[-0.5, 1] \times [-1.5, 1]$.)

Joint Continuous Distribution

Joint Continuous Distribution

Two random variables X and Y, defined on the same sample space Ω, have a joint continuous distribution if there exists a function $f : \mathbb{R}^2 \to [0, \infty]$ such that for all real numbers $a_1 \leq b_1$ and $a_2 \leq b_2$,

$$\begin{aligned} P(a_1 \leq X \leq b_1,\, a_2 \leq Y \leq b_2) &:= P(\{a_1 \leq X \leq b_1\} \cap \{a_2 \leq Y \leq b_2\}) \\ &= \int_{a_1}^{b_1} \int_{a_2}^{b_2} f(x, y) dx dy. \end{aligned}$$

The function f satisfies $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$ and $f(x, y) \geq 0$ for all $(x, y) \in \mathbb{R}^2$. We call f the joint probability density function of X and Y.

Marginal Probability Density Function

Marginal Probability Density Function

Let f be the joint probability density function of random variables X and Y. The marginal probability density functions of X and Y are defined by

$$f_X(x) := \int_{-\infty}^{\infty} f(x,y)dy \quad \text{and} \quad f_Y(y) := \int_{-\infty}^{\infty} f(x,y)dx.$$

- Thus, the marginal pdfs of each of the random variables X and Y can be obtained by "integrating out" the other variable.
- Integration is the analogous operation to the summation over the range in the case of discrete random variables.

Optional: Calculating Probabilities from a Joint PDF

Suppose that X and Y have joint probability density function

$$f(x,y) = \begin{cases} \frac{2}{75}(2x^2y + xy^2) & \text{for } 0 \leq x \leq 3, 1 \leq y \leq 2, \\ 0 & o/w. \end{cases}$$

$$\begin{aligned} P(1 \leq X \leq 2,\ 4/3 \leq Y \leq 5/3) &= \int_1^2 \int_{4/3}^{5/3} f(x,y)dxdy \\ &= \frac{2}{75}\int_1^2 \left(\int_{4/3}^{5/3} (2x^2y + xy^2)dy \right) dx \\ &= \frac{2}{75}\int_1^2 \left(x^2 + \frac{61}{81}x \right) dx \\ &= \frac{187}{2025}. \end{aligned}$$

Joint Distribution Function

Joint Distribution Function (Continuous Random Variables)

Let X and Y be two random variables with joint probability density function f. The joint (cumulative) distribution function $F : \mathbb{R}^2 \to [0,1]$ is defined by

$$F(a,b) := P(X \leq a, Y \leq b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f(x,y) dx dy \quad \text{for } -\infty < a, b < \infty.$$

- Marginal (cumulative) distribution functions F_X and F_Y can be found as in the case of discrete random variables.
- **Facts from calculus:**
  - $f(x,y) = \frac{\partial^2}{\partial x \partial y} F(x,y)$.
  - $P(a_1 \leq X \leq b_1, a_2 \leq Y \leq b_2) = F(b_1,b_2) - F(b_1,a_2) - F(a_1,b_2) + F(a_1,a_2)$.

(Derivations.)

Joint Distribution Function (Cont.)

(Graphical derivation of $P(a_1 \leq X \leq b_1, a_2 \leq Y \leq b_2) = \ldots$*)*

Optional: Joint and Marginal DF

Suppose that X and Y have joint probability density function

$$f(x,y) = \begin{cases} \frac{2}{75}(2x^2y + xy^2) & \text{for } 0 \le x \le 3, 1 \le y \le 2, \\ 0 & o/w. \end{cases}$$

Find the joint (cumulative) distribution function of (X, Y).
For $0 \le a \le 3$ and $1 \le b \le 2$ we have, since $f(x,y) = 0$ for $x < 0$ or $y < 1$,

$$\begin{aligned} F(a,b) &= \int_{-\infty}^{a} \left(\int_{-\infty}^{b} f(x,y)dy \right) dx \\ &= \frac{2}{75} \int_{0}^{a} \left(\int_{1}^{b} (2x^2y + xy^2)dy \right) dx \\ &= \frac{1}{225}(2a^3b^2 - 2a^3 + a^2b^3 - a^2). \end{aligned}$$

Optional: Joint and Marginal DF (Cont.)

Suppose that X and Y have joint probability density function

$$f(x,y) = \begin{cases} \frac{2}{75}(2x^2y + xy^2) & \text{for } 0 \leq x \leq 3, 1 \leq y \leq 2, \\ 0 & o/w. \end{cases}$$

For either $a \notin [0,3]$ or $b \notin [1,2]$ the expression for $F(a,b)$ is different:

- For $a \in [0,3]$ and $b > 2$, then, since $f(x,y) = 0$ for $y > 2$,

$$F(a,b) = F(a,2) = \frac{1}{225}(6a^3 + 7a^2).$$

- Verify at home that for $b \in [1,2]$ and $a > 3$,

$$F(a,b) = F(3,b) = \frac{1}{75}(3b^3 + 18b^2 - 21).$$

- Note that these are the marginal distribution functions of X and Y, respectively, i.e. $F_X(a) = F(a,2)$ and $F_Y(b) = F(3,b)$.

Optional: Joint and Marginal PDF

Suppose that X and Y have joint probability density function

$$f(x,y) = \begin{cases} \frac{2}{75}(2x^2y + xy^2) & \text{for } 0 \leq x \leq 3, 1 \leq y \leq 2, \\ 0 & o/w. \end{cases}$$

The pdf of X for $x \in [0,3]$ can be found in two ways:

- By differentiating F_X:

$$f_X(x) = \frac{d}{dx} F_X(x) = \frac{d}{dx}\left(\frac{1}{225}(6x^3 + 7x^2)\right) = \frac{2}{225}(9x^2 + 7x).$$

- By integrating $f(x,y)$ over $y \in [1,2]$:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \frac{2}{75}\int_1^2 (2x^2y + xy^2)dy = \frac{2}{225}(9x^2 + 7x).$$

The pdf of X for $x \notin [0,3]$ is $f_X(x) = 0$.

Optional: Joint and Marginal PDF (Cont.)

Suppose that X and Y have joint probability density function

$$f(x,y) = \begin{cases} \frac{2}{75}(2x^2y + xy^2) & \text{for } 0 \leq x \leq 3, 1 \leq y \leq 2, \\ 0 & o/w. \end{cases}$$

Verify at home (by differentiating F_Y or integrating out x in $f(x,y)$) that the pdf of Y for $y \in [1,2]$ is

$$f_Y(y) = \frac{1}{25}(3y^2 + 12y).$$

The pdf of Y for $y \notin [1,2]$ is $f_Y(y) = 0$.

Outline

How to model independent random variables?

- Recall that two events A and B are independent if and only if

$$P(A \cap B) = P(A)P(B).$$

- Therefore, we would like to define independence of RV as something like:

  "Two random variables X and Y are independent if and only if all events involving only X and all events involving only Y are independent."

- If X and Y are discrete RV, then an adequate definition for independence would be

$$P(X = a, Y = b) = P(X = a)P(Y = b) \quad \text{for all } -\infty < a, b < \infty.$$

- **Problem:** This definition is useless in the case of continiuous RV.

Independent Random Variables

Independent Random Variables

The random variables X and Y with joint distribution function F are independent if, for all $-\infty < a, b < \infty$,

$$F(a, b) = F_X(a)F_Y(b).$$

Random variables that are not independent are called dependent.

- If X and Y are discrete random variables with joint pmf p above definition is equivalent to

$$p(a, b) = p_X(a)p_Y(b) \quad \text{for all } -\infty < a, b < \infty.$$

- If X and Y are jointly continuous random variables with joint pdf f above definition is equivalent to

$$f(a, b) = f_X(a)f_Y(b) \quad \text{for all } -\infty < a, b < \infty.$$

Example: Verifying the Obvious.

We toss two fair coins. Show that the outcomes of the coin tosses are independent.

- Let $X \in \{H, T\}$ be the outcome of the first coin toss.
- Let $Y \in \{H, T\}$ be the outcomes of the second coin.
- The joint pmf of X and Y is given by

		y	
	$p(x, y)$	0	1
x	0	1/4	1/4
	1	1/4	1/4

- We need to check whether **for all** $a, b \in \{H, T\}$,

$$P(X = a, Y = b) = P(X = a)P(Y = b).$$

Obviously (?) this is the case.

Example: Dependent Random Variables

Show that the random variables X and Y are dependent!

- Joint pmf of X and Y:

	$p(x,y)$	y = 0	1	2
	0	0.01	0	0
x	1	0.09	0.09	0
	2	0	0	0.81

- Summing over the columns (the rows) of the joint pmf we obtain the marginal pmf of X (pmf of Y):

x	0	1	2
$p_X(x)$	0.01	0.18	0.81

and

y	0	1	2
$p_Y(y)$	0.10	0.09	0.81

- To show that X and Y are not independent it suffices to find **one pair** (x, y) which violates the independence property. One such pair is $(0, 0)$,

$$p(0,0) = 0.01 \neq 0.001 = p_X(0)p_Y(0).$$

Propagation of Independence

Are transformed independent random variables again independent?

- Let X and Y be two independent random variables with joint distribution function F.

- For $I = (a, b]$ an interval on the real line define the transformed random variables

$$U = \begin{cases} 1 & \text{if } X \in I, \\ 0 & \text{if } X \notin I, \end{cases} \qquad \text{and} \qquad V = \begin{cases} 1 & \text{if } Y \in I, \\ 0 & \text{if } Y \notin I. \end{cases}$$

- U and V are independent because for all values $u, v \in \{0, 1\}$,

$$P(U = u, V = v) = P(U = u)P(V = v).$$

- For example,

$$\begin{aligned} P(U = 0, V = 1) &= P(X \in I^c, Y \in I) \\ &= P(X \in I^c)P(Y \in I) = P(U = 0)P(V = 1). \end{aligned}$$

Propagation of Independence (Cont.)

Propagation of Independence

Let $X_1, X_2, \ldots, X_n$ be independent random variables. For functions $h_i : \mathbb{R} \to \mathbb{R}$, $i = 1, \ldots, n$, define the random variables

$$Y_i = h_i(X_i).$$

Then $Y_1, Y_2, \ldots, Y_n$ are also independent.

- **Important consequence:** We can simulate independent random variables of any distribution if we can generate independent uniform random variables!
  - Let $U_1, \ldots, U_n$ be n independent unform random variables. Recall from Week 4 Lecture, Part 2,

    $$X_i := -\frac{1}{\lambda}\ln(U_i) \sim Exp(\lambda).$$

  - By above result, $X_1, \ldots, X_n$ are independent exponentially distributed random variables with rate parameter $\lambda > 0$.

Modeling Extreme Events

- *Floodings.* Let $X_1, \ldots, X_{365}$ be the water levels of a river during the days of a particular year.
  - If the water levels exceed a certain hight h of the dykes, there will be a flooding. What is the probability of a flood?
  - Answer: $P(\max\{X_1, \ldots, X_{365}\} > h)$.
- *Drought.* Let $X_1, \ldots, X_{365}$ be the precipitation in a certain region during the days of a particular year.
  - If the precipitation falls below a certain amount h, there will be a drought. What is the probability of a drought?
  - Answer: $P(\max\{X_1, \ldots, X_{365}\} < h)$.
- *Financial markets.* Let $X_1, \ldots, X_{252}$ be the stock prices of the IBM stock during the trading days of a particular year.
  - Sell or buy stocks if they exceed (or fall below) a certain threshold t.
  - Answer: Sell with probability $P(\max\{X_1, \ldots, X_{252}\} > t)$ and buy with probability $P(\max\{X_1, \ldots, X_{252}\} < t)$.

Extremes of Independent Random Variables

Distribution of Maximum and Minimum.

Let $X_1, X_2, \ldots, X_n$ be independent random variables with the same distribution function F. Let

$$U = \max\{X_1, \ldots, X_n\} \quad \text{and} \quad V = \min\{X_1, \ldots, X_n\}.$$

Then,

$$F_U(a) = \big(F(a)\big)^n \quad \text{and} \quad F_V(a) = 1 - \big(1 - F(a)\big)^n.$$

(Derivations.)

Example: Uniform Athletes Competing for Gold.

Consider 10 amateur athletes, none of which is a clear favorite to win the 100m dash. We therefore assume that the final times are uniformly and independently distributed between 10 and 12 seconds. What is the probability that athlete 1 wins the race if he finishes in 11 seconds?

- X_k = time in seconds of athlete $k = 1, \dots, 10$ and $Z := \min\{X_1, \dots X_{10}\}$.
- If the smallest X_k is at least 11 seconds, then athlete 1 wins the race. Therefore the probability that we want to compute is
$$P(Z \geq 11).$$
- Since the X_k are uniformly distributed between 10 and 12 seconds, their distribution function is
$$F_{X_1}(x) = \dots F_{X_{10}}(x) \equiv F(x) = \begin{cases} 0 & \text{if } x < 10 \\ \frac{x-10}{2} & \text{if } 10 \leq x \leq 12 \\ 1 & \text{if } x > 12. \end{cases}$$
- By the independence and the previous theorem we have
$$P(Z \geq 11) = 1 - F_Z(11) = (1 - F(11))^{10} = \left(1 - \frac{1}{2}\right)^{10} = 0.5^{10} = 0.1\%.$$