Joint Probability Distributions Continuous Case

Section 6.2

STAT/MATH 395 Spring 2020

Vincent Roulet

Lecture 6, April 10th, 2020

Ask questions via chat on Zoom
Answer quiz via PollEverywhere (username: vincentroulet)

Anouncements

Office hours (After poll)

- Mondays 14:30 to 15:30 with T.A. Z. Yuan by Zoom
- Fridays 11:30 to 12:30 with instructor V. Roulet
- Register in advance to access the zoom session

Lecture material
Updated slides with solutions given at the end of the lecture

Answer Previous Exercise

Exercise

Roll a normal die 100 times. Find the probability that among the 100 rolls, we observe exactly 22 ones, 17 fives.

Answer Previous Exercise

Exercise

Roll a normal die 100 times. Find the probability that among the 100 rolls, we observe exactly 22 ones, 17 fives.

Solution Denote X_1, X_5 the number of times you get a 1 or a 5 resp. among 100 rolls
We have $\mathbb{P}(\text{"face is 1"}) = \mathbb{P}(\text{"face is 5"}) = 1/6$
We could model $X_1, X_2, X_3, X_4, X_5, X_6$ as a multinomial but that can be simplified
Denote $Y = X_2 + X_3 + X_4 + X_6$ the number of times you get any other face
We have $\mathbb{P}\,\text{"face is not 1 or 5"} = 4/6 = 2/3$
Then $(X_1, X_5, Y) \sim \text{Multinom}(100, 3, 1/6, 1/6, 2/3)$
So $\mathbb{P}(X_1 = 22, X_5 = 17, Y = 100 - (22 + 17)) = \frac{100!}{22!17!61!}\left(\frac{1}{6}\right)^{22}\left(\frac{1}{6}\right)^{17}\left(\frac{2}{3}\right)^{61} \approx 0.0037$

Joint Probability Density Functions

Definition (Joint probability density function)

Random variables $X_1, \ldots, X_n$ are **jointly continuous** if there exists a **joint probability density function** $f : \mathbb{R}^n \to \mathbb{R}$ such that for any[1] $B \subset \mathbb{R}^n$,

$$\mathbb{P}(X_1, \ldots, X_n \in B) = \int \underset{B}{\ldots} \int f(x_1, \ldots, x_n) dx_1 \ldots dx_n$$

Note:

- $f(x_1, \ldots, x_n) \geq 0$ and $\int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty} f(x_1, \ldots, x_n) dx_1 \ldots dx_n = 1$
- X and Y have a p.d.f. does not imply that (X, Y) is jointly continuous!

[1] Think of B as for example $[a, b]^n$. Again a rigorous definition requires B to belong to the Borel algebra of $\mathbb{R}^n$

Joint Probability Density Functions

Definition (Joint probability density function)

Random variables $X_1, \dots, X_n$ are **jointly continuous** if there exists a **joint probability density function** $f : \mathbb{R}^n \to \mathbb{R}$ such that for any[1] $B \subset \mathbb{R}^n$,

$$\mathbb{P}(X_1, \dots, X_n \in B) = \int \dots \int_B f(x_1, \dots, x_n) dx_1 \dots dx_n$$

Note:

- $f(x_1, \dots, x_n) \geq 0$ and $\int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} f(x_1, \dots, x_n) dx_1 \dots dx_n = 1$
- X and Y have a p.d.f. does not imply that (X, Y) is jointly continuous!

  Example: Take X any continuous r.v., define $Y = X$, s.t. $\mathbb{P}(X = Y) = 1$. If (X, Y) had a joint p.d.f. f, denoting $D = \{(x, y) : x = y\}$, we would have

  $$\mathbb{P}(X = Y) = \iint_D f(x, y) dx dy = \int_{-\infty}^{+\infty} \left(\int_x^x f(x, y) dy \right) dx = 0$$

[1] Think of B as for example $[a, b]^n$. Again a rigorous definition requires B to belong to the Borel algebra of $\mathbb{R}^n$

Joint Probability Density Functions

Lemma
Let $X_1, \ldots, X_n$ be n jointly continuous r.v.. Then for any subset $A \subset \mathbb{R}^n$ included in a linear subspace $E \subset \mathbb{R}^n$ of dimension $\dim(E) = m < n$,

$$\mathbb{P}((X_1, \ldots, X_n) \in A) = 0$$

Lemma

Let $X_1, \ldots, X_n$ be n jointly continuous r.v.. Then for any subset $A \subset \mathbb{R}^n$ included in a linear subspace $E \subset \mathbb{R}^n$ of dimension $\dim(E) = m < n$,

$$\mathbb{P}((X_1, \ldots, X_n) \in A) = 0$$

Proof General case requires change of variables, let's consider $A = [a, b]^m \subset \mathbb{R}^n$
Denote f the joint p.d.f. of $(X_1, \ldots, X_n)$,

$$\mathbb{P}((X_1, \ldots, X_n) \in A) = \underbrace{\int_a^b \ldots \int_a^b}_{m \ \ times} \underbrace{\int_0^0 \ldots \int_0^0}_{n-m \ \ times} f(x_1, \ldots, x_n) dx_1 \ldots dx_n = 0$$

Joint Probability Density Functions

Example (Synthetic)

Assume X, Y have a joint p.d.f.

$$f(x, y) = \begin{cases} \frac{3}{2}(xy^2 + y) & \text{if } 0 \leq x \leq 1,\ 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

1. Check that it is a valid joint p.d.f.

Example (Synthetic)

Assume X, Y have a joint p.d.f.

$$f(x,y) = \begin{cases} \frac{3}{2}(xy^2 + y) & \text{if } 0 \leq x \leq 1,\ 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

1. Check that it is a valid joint p.d.f.

Solution We have $f(x,y) \geq 0$ and

$$\begin{aligned} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y) dx dy &= \frac{3}{2} \int_0^1 \left(\int_0^1 xy^2 + y dx \right) dy \\ &= \frac{3}{2} \int_0^1 \left(\frac{1}{2} y^2 + y \right) dy = \frac{3}{2} \left(\frac{1}{6} + \frac{1}{2} \right) = 1 \end{aligned}$$

Example (Synthetic)

Assume X, Y have a joint p.d.f.

$$f(x,y) = \begin{cases} \frac{3}{2}(xy^2 + y) & \text{if } 0 \leq x \leq 1,\ 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

2. Compute $\mathbb{P}(X < Y)$

Example (Synthetic)

Assume X, Y have a joint p.d.f.

$$f(x,y) = \begin{cases} \frac{3}{2}(xy^2 + y) & \text{if } 0 \leq x \leq 1,\ 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

2. Compute $\mathbb{P}(X < Y)$

Solution

$$\begin{aligned} \mathbb{P}(X < Y) &= \frac{3}{2}\int_0^1 \left(\int_0^y (xy^2 + y)dx \right) dy \\ &= \frac{3}{2}\int_0^1 \left(\frac{1}{2}y^4 + y^2 \right) dy \\ &= \frac{3}{2}\left(\frac{1}{10} + \frac{1}{3} \right) = 0.65 \end{aligned}$$

Uniform Continuous Random Variable in higher dimensions

Definition (Uniform continuous random variable in dimension 2 or 3)

Let D be a bounded subset of $\mathbb{R}^2$ s.t. $\text{Area}(D) < +\infty$. The random point (X, Y) is **uniformly distributed on** D if its joint p.d.f. reads

$$f(x,y) = \frac{1}{\text{Area}(D)} \mathbf{1}_D(x,y) = \begin{cases} \frac{1}{\text{Area}(D)} & \text{if}(x,y) \in D \\ 0 & \text{otherwise} \end{cases}$$

Let D be a bounded subset of $\mathbb{R}^3$ s.t. $\text{Vol}(D) < +\infty$. The random point (X, Y, Z) is **uniformly distributed on** D if its joint p.d.f. reads

$$f(x,y,z) = \frac{1}{\text{Vol}(D)} \mathbf{1}_D(x,y) \begin{cases} \frac{1}{\text{Vol}(D)} & \text{if}(x,y.z) \in D \\ 0 & \text{otherwise} \end{cases}$$

We denote $(X, Y) \sim \text{Unif}(D)$ or $(X, Y, Z) \sim \text{Unif}(D)$.

Uniform Continuous Random Variable in higher dimensions

Lemma

Let $(X, Y) \sim \text{Unif}(D)$ *for* $D \subset \mathbb{R}^2$, *then for any* $G \subset D$, *(similar for* $\mathbb{R}^3$*)*

$$\mathbb{P}((X, Y) \in G) = \frac{\text{Area}(G)}{\text{Area}(D)}$$

Uniform Continuous Random Variable in higher dimensions

Lemma

Let $(X, Y) \sim \text{Unif}(D)$ *for* $D \subset \mathbb{R}^2$*, then for any* $G \subset D$*, (similar for* $\mathbb{R}^3$*)*

$$\mathbb{P}((X, Y) \in G) = \frac{\text{Area}(G)}{\text{Area}(D)}$$

Proof

$$\Pr((X, Y) \in G) = \frac{1}{\text{Area}(D)} \int \int \mathbf{1}_G(x, y)\, \mathbf{1}_D(x, y) dxdy = \int \int \mathbf{1}_G(x, y) dxdy = \frac{\text{Area}(G)}{\text{Area}(D)}$$

Uniform Continuous Random Variable in higher dimensions

Lemma

Let $(X, Y) \sim \text{Unif}(D)$ *for* $D \subset \mathbb{R}^2$*, then for any* $G \subset D$*, (similar for* $\mathbb{R}^3$*)*

$$\mathbb{P}((X, Y) \in G) = \frac{\text{Area}(G)}{\text{Area}(D)}$$

Proof

$$\Pr((X, Y) \in G) = \frac{1}{\text{Area}(D)} \int\int \mathbf{1}_G(x, y)\, \mathbf{1}_D(x, y) dxdy = \int\int \mathbf{1}_G(x, y) dxdy = \frac{\text{Area}(G)}{\text{Area}(D)}$$

Example

Denote $D_r = \{(x, y) : x^2 + y^2 < r^2\}$ a disk of radius r

Throw a dart uniformly at random on a disk of radius 2

What is the probability that the dart is in the central disk of radius one?

Uniform Continuous Random Variable in higher dimensions

Lemma

Let $(X, Y) \sim \text{Unif}(D)$ *for* $D \subset \mathbb{R}^2$*, then for any* $G \subset D$*, (similar for* $\mathbb{R}^3$*)*

$$\mathbb{P}((X, Y) \in G) = \frac{\text{Area}(G)}{\text{Area}(D)}$$

Proof

$$\Pr((X, Y) \in G) = \frac{1}{\text{Area}(D)} \int\int \mathbf{1}_G(x, y)\,\mathbf{1}_D(x, y) dxdy = \int\int \mathbf{1}_G(x, y) dxdy = \frac{\text{Area}(G)}{\text{Area}(D)}$$

Example

Denote $D_r = \{(x, y) : x^2 + y^2 < r^2\}$ a disk of radius r
Throw a dart uniformly at random on a disk of radius 2
What is the probability that the dart is in the central disk of radius one?

Solution $(X, Y) \sim \text{Unif}(D_2)$

$$\mathbb{P}((X, Y) \in D_1) = \frac{\pi 1^2}{\pi 2^2} = \frac{1}{4}$$

Joint Probability Density Functions

Lemma

Let $g : \mathbb{R}^n \to \mathbb{R}$ and let $X_1, \dots, X_n$ be jointly continuous r.v. with joint p.d.f. f,

$$\mathbb{E}[g(x_1, \dots x_n)] = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} g(x_1, \dots, x_n) f(x_1, \dots, x_n) dx_1 \dots dx_n$$

Example

Throw a dart uniformly at random on a square of edge size 2 centered on 0
Assume your score is equal to the square distance to the center
What is your average score?

Joint Probability Density Functions

Lemma

Let $g : \mathbb{R}^n \to \mathbb{R}$ and let $X_1, \ldots, X_n$ be jointly continuous r.v. with joint p.d.f. f,

$$\mathbb{E}[g(x_1, \ldots x_n)] = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} g(x_1, \ldots, x_n) f(x_1, \ldots, x_n) dx_1 \ldots dx_n$$

Example

Throw a dart uniformly at random on a square of edge size 2 centered on 0
Assume your score is equal to the square distance to the center
What is your average score?

Solution $(X, Y) \sim \mathrm{Unif}(S)$ with $S = \{(x, y) : -1 \leq x \leq 1, -1 \leq y \leq 1\}$
Score is $g(x, y) = x^2 + y^2$
Average score

$$\mathbb{E}[g(X, Y)] = \frac{1}{4} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) \mathbf{1}_S(x, y) dxdy = \frac{1}{4} \int_{-1}^{1} \int_{-1}^{1} (x^2 + y^2) dxdy = 1/3$$

Marginal Probability Density Function

Definition (Marginal probability density function)

Let X, Y be jointly continuous r.v. and denote $f_{X,Y}$ their joint p.d.f. then the p.d.f. of X exists and is given by

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

Marginal Probability Density Function

Definition (Marginal probability density function)

Let X, Y be jointly continuous r.v. and denote $f_{X,Y}$ their joint p.d.f. then the p.d.f. of X exists and is given by

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

Proof We have by definition of the joint p.d.f. an expression of the c.d.f. of X as

$$F_X(t) = \mathbb{P}(X \leq t) = \mathbb{P}(X \leq t, -\infty \leq Y \leq +\infty) = \int_{-\infty}^{t} \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dydx$$

Therefore $f_X(x) = F'_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)$

Example

Consider a disk of radius r, $D_r = \{(x, y) : x^2 + y^2 \leq r\}$ and $(X, Y) \sim \mathrm{Unif}(D_r)$. What is the marginal p.d.f. of X?

Marginal Probability Density Function

Example

Consider a disk of radius r, $D_r = \{(x,y) : x^2 + y^2 \leq r\}$ and $(X,Y) \sim \mathrm{Unif}(D_r)$.
What is the marginal p.d.f. of X?

Solution Joint p.d.f. is $f_{X,Y}(x,y) = \frac{1}{\pi r^2} \mathbf{1}_{D_r}(x,y)$ where $D_r = \{(x,y) : x^2 + y^2 \leq r^2\}$
Marginal density is then $f_X(x) = 0$ for $|x| > r$, and for $|x| \leq r$,

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy = \frac{1}{\pi r^2} \int_{-\sqrt{r^2-x^2}}^{\sqrt{r^2-x^2}} dy = \frac{2}{\pi r^2} \sqrt{r^2 - x^2}$$

Marginal Probability Density Function

Definition (Marginal probability density function)

Let $X_1, \ldots X_n$ be jointly continuous and denote f their joint p.d.f..
Then for any $j \in \{1, \ldots n\}$, X_j is a continuous random variable with p.d.f.

$$f_{X_j}(x) = \underbrace{\int_{-\infty}^{+\infty} \ldots \int_{-\infty}^{+\infty}}_{(n-1 \ \ integrals)} f(x_1, \ldots, x_{j-1}, x, x_{j+1}, \ldots, x_n) dx_1 \ldots dx_{j-1} dx_{j+1} \ldots dx_n$$

Joint Cumulative Distribution

Definition (Joint cumulative distribution)

The **joint cumulative distribution** of r.v. $X_1, \dots, X_n$ is defined as

$$F(t_1, \dots, t_n) = \mathbb{P}(\{X_1 \leq t_1\} \cap \dots \cap \{X_n \leq t_n\})$$
$$\triangleq \mathbb{P}(X_1 \leq t_1, \dots, X_n \leq t_n)$$

Lemma

1. *If (X, Y) are jointly continuous with joint p.d.f. f,*

$$F(t, s) = \int_{-\infty}^{t} \int_{-\infty}^{s} f(x, y) dxdy$$

2. *If (X, Y) are jointly continuous (i.e. there exists a joint p.d.f.) with joint c.d.f. F*

$$\frac{\partial^2}{\partial t \partial s} F(t, s)\Big|_{s=x, t=y} = f(x, y)$$

Borel algebra in $\mathbb{R}^n$*

Formal details

- Until now, we defined proba. distributions on any $B \subset \mathbb{R}^n$ for n=1 or n>1.
- Formal definitions require to restrict our focus to subsets $B \subset \mathbb{R}^n$ that form a σ-algebra $\mathcal{B}$

Definition (σ-algebra)

Let Ω be a set, a σ-algebra $\mathcal{F}$ on Ω is a subset of $2^\Omega = \{B \subset \Omega\}$ such that

1. $\Omega \in \mathcal{F}$
2. (Stable by complementarity) For any $A \in \mathcal{F}$, $A^c \triangleq \Omega \setminus A \in \mathcal{F}$
3. (Stable by countable union) For any $A_1, A_2, \ldots \in \mathcal{F}$, $\bigcup_{i\in\mathbb{N}} A_i \in \mathcal{F}$

Why introducing σ-algebra?
You want the probability measure to satisfy that

- the measure is non-negative
- the measure of the union of disjoint sets is the sum of the measure of union sets

Then you can build a union of sets V_k (see e.g. Vitali set on Wikipedia) s.t.

$$[0,1] \subset \bigcup_{k=1}^{+\infty} V_k \subset [-1,2] \qquad \mathbb{P}(V_k) = \lambda \geq 0 \quad \text{for all} \quad k$$

which leads to $1 \leq \sum_{k=1}^{+\infty} \mathbb{P}(V_k) \leq 3$ which is impossible

Borel algebra in $\mathbb{R}^n$*

Formally, we restrict our focus on the Borel algebra of $\mathbb{R}^n$

Definition (Borel algebra in $\mathbb{R}^n$)

The Borel algebra in $\mathbb{R}^n$, denoted $\mathcal{B}_n$, is the smallest σ-algebra (in terms of inclusion) that contains

- all product of intervals $[a_1, b_1] \times \ldots \times [a_n, b_n]$ for $a_i \leq b_i \in \mathbb{R}$

or equivalently defined as the smallest σ-algebra that contains

- all product of intervals of the form $(-\infty, a_1] \times \ldots \times (-\infty, a_n]$ for $a_i \in \mathbb{R}$.

Consequence

1. If we can measure all intervals of the form $(-\infty, a_1] \times \ldots \times (-\infty, a_n]$ for $a_i \in \mathbb{R}$, then we can measure all subsets of interests, i.e. all $B \in \mathcal{B}_n$,
   $\rightarrow$ we know all the information necessary to describe the proba distribution
2. All the information necessary to describe any r.v. is contained in its c.d.f.

Quiz for next lecture

Exercise

I am shooting an arrow on a target on a wall $W = \{(x, y) : -1 \leq x \leq 1, 0 \leq y \leq 1\}$*.*
A wind affects my shoot from the left and the gravity also affects my shoot such that the position of the arrow has a p.d.f. proportional to $\frac{e^x}{\sqrt{y+1}}$

What is the probability
that I touch the target $T = \{(x, y) : -0.1 \leq x \leq 0.1, 0.4 \leq y \leq 0.6\}$*?*