S&DS 241 Lecture 18

Conditional PDF & independence

B-H: 7.1

Recall: Independence of discrete random variables (Lec 6)

Let X, Y be a pair of discrete random variables.

Equivalent definitions of independence:

1. $\underbrace{P(X = x, Y = y)}_{p_{XY}(x,y)} = \underbrace{P(X = x)}_{p_X(x)} \underbrace{P(Y = y)}_{p_Y(y)}$, for all x, y

Recall: Independence of discrete random variables (Lec 6)

Let X, Y be a pair of discrete random variables.

Equivalent definitions of independence:

1. $\underbrace{P(X = x, Y = y)}_{p_{XY}(x,y)} = \underbrace{P(X = x)}_{p_X(x)} \underbrace{P(Y = y)}_{p_Y(y)}$, for all x, y
2. $F_{XY}(x, y) = F_X(x) F_Y(y)$

Recall: Independence of discrete random variables (Lec 6)

Let X, Y be a pair of discrete random variables.

Equivalent definitions of independence:

1. $\underbrace{P(X = x, Y = y)}_{p_{XY}(x,y)} = \underbrace{P(X = x)}_{p_X(x)} \underbrace{P(Y = y)}_{p_Y(y)}$, for all x, y
2. $F_{XY}(x, y) = F_X(x) F_Y(y)$

that is

$$\boxed{\text{Joint PMF/CDF} = \text{product of marginal PMFs/CDFs}}$$

Independence of continuous random variables

Let X, Y be a pair of continuous random variables.

Equivalent definitions of independence:

1. $f_{XY}(x, y) = f_X(x) f_Y(y)$, for all x, y
2. $F_{XY}(x, y) = F_X(x) F_Y(y)$

that is,

$$\boxed{\text{Joint PDF/CDF} = \text{product of marginal PDF/CDF}}$$

Example (continuation of last lecture)

Consider a randomly sampled individual.

- Let X be the indicator of being a smoker.
- Let Y be the indicator of developing lung cancer.

Suppose the joint PMF p_{XY} is:

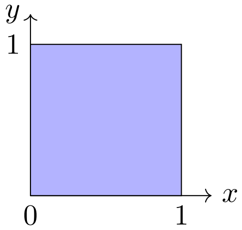
	$Y = 1$	$Y = 0$	Sum
$X = 1$	0.05	0.20	0.25
$X = 0$	0.01	0.74	0.75
Sum	0.06	0.94	1

$P(X = 1, Y = 1) \neq P(X = 1)P(Y = 1) \implies$ X and Y are dependent

Example (continuation of last lecture)

- Joint PDF of (X, Y), the arrival time of Alice and Bob:

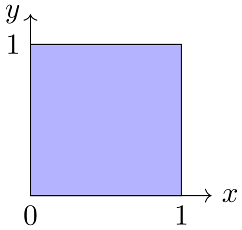
$$f_{XY}(x,y) = \begin{cases} 1 & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{else} \end{cases}$$



Example (continuation of last lecture)

- Joint PDF of (X, Y), the arrival time of Alice and Bob:

$$f_{XY}(x,y) = \begin{cases} 1 & 0 \le x \le 1, 0 \le y \le 1 \\ 0 & \text{else} \end{cases}$$



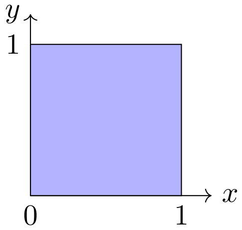
- Marginals:

$$f_X(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{else} \end{cases} \qquad f_Y(y) = \begin{cases} 1 & 0 \le y \le 1 \\ 0 & \text{else} \end{cases}$$

Example (continuation of last lecture)

- Joint PDF of (X, Y), the arrival time of Alice and Bob:

$$f_{XY}(x,y) = \begin{cases} 1 & 0 \le x \le 1, 0 \le y \le 1 \\ 0 & \text{else} \end{cases}$$



- Marginals:

$$f_X(x) = \begin{cases} 1 & 0 \le x \le 1 \\ 0 & \text{else} \end{cases} \quad f_Y(y) = \begin{cases} 1 & 0 \le y \le 1 \\ 0 & \text{else} \end{cases}$$

- Thus

$$f_{XY}(x,y) = f_X(x) \times f_Y(y), \quad \text{for any } x, y$$

X and Y are independent and identically distributed as Unif$(0,1)$

Consequence of independence

- For any set A and B:

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

- More generally, for any function g and h,

$$E(g(X)h(Y)) = E(g(X))E(h(Y))$$

Consequence of independence

Support of independent random variables must be a product set:

- Support of (X, Y):

$$S_{XY} = \{(x, y) : f_{XY}(x, y) > 0\}$$

- Support of X (and Y):

$$S_X = \{x : f_X(x) > 0\} = \text{ projection of } S_{XY} \text{ onto } x\text{-axis}$$
$$S_Y = \{y : f_Y(y) > 0\} = \text{ projection of } S_{XY} \text{ onto } y\text{-axis}$$

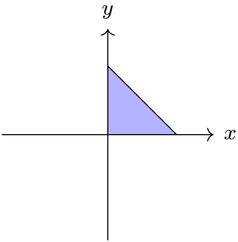
- If X and Y are independent, then we must have

$$S_{XY} = S_X \times S_Y$$

- $S_{XY} \neq S_X \times S_Y$ is a simple way to rule out independence

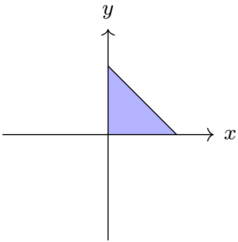
Examples

Consider the following support set of (X, Y):



Examples

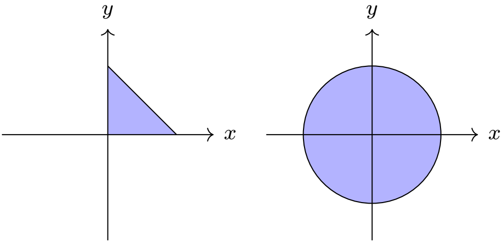
Consider the following support set of (X, Y):



dependent

Examples

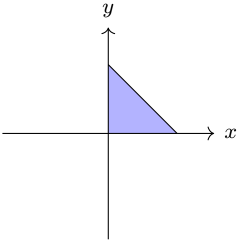
Consider the following support set of (X, Y):



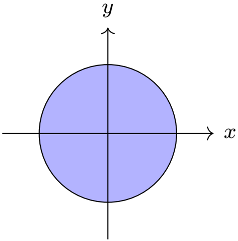
dependent

Examples

Consider the following support set of (X, Y):



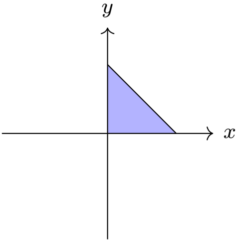
dependent



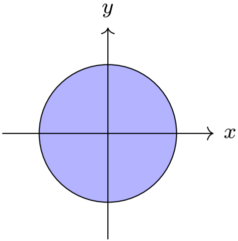
dependent

Examples

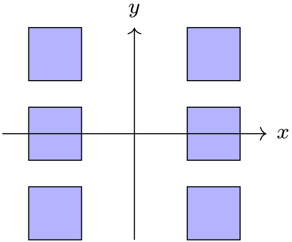
Consider the following support set of (X, Y):



dependent

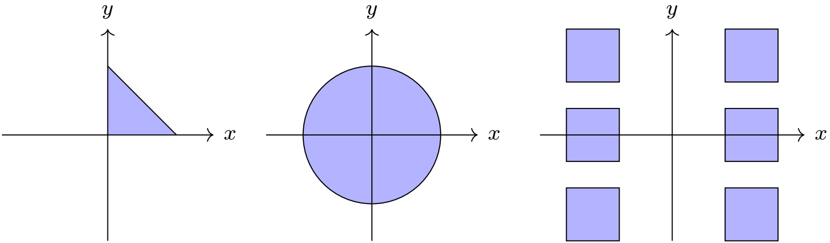


dependent



Examples

Consider the following support set of (X, Y):



dependent　　dependent　　possibly independent

Conditioning

Conditional PMF

Let's start from discrete:

- For discrete X, Y, if $P(X = x) > 0$, by definition,

$$P(Y = y \mid X = x) = \frac{P(X = x, Y = y)}{P(X = x)}$$

Conditional PMF

Let's start from discrete:

- For discrete X, Y, if $P(X = x) > 0$, by definition,

$$P(Y = y \mid X = x) = \frac{P(X = x, Y = y)}{P(X = x)}$$

- Denote by $p_{Y|X}(y|x) = P(Y = y \mid X = x)$ the conditional PMF of Y given $X = x$, which is related to joint PMF and marginal PMF via

$$p_{Y|X}(y|x) = \frac{p_{XY}(x, y)}{p_X(x)}, \quad \text{provided that } p_X(x) > 0$$

Conditional PMF

Let's start from discrete:

- For discrete X, Y, if $P(X = x) > 0$, by definition,

$$P(Y = y \mid X = x) = \frac{P(X = x, Y = y)}{P(X = x)}$$

- Denote by $p_{Y|X}(y|x) = P(Y = y \mid X = x)$ the conditional PMF of Y given $X = x$, which is related to joint PMF and marginal PMF via

$$p_{Y|X}(y|x) = \frac{p_{XY}(x, y)}{p_X(x)}, \quad \text{provided that } p_X(x) > 0$$

- Summary:

$$\boxed{\text{conditional PMF} = \frac{\text{joint PMF}}{\text{marginal PMF}}}$$

Conditional PDF

- For continuous (X, Y), the conditional PDF of Y given $X = x$ is defined as

$$f_{Y|X}(y|x) = \frac{f_{XY}(x, y)}{f_X(x)}, \quad \text{provided that } f_X(x) > 0,$$

i.e.,

$$\boxed{\text{conditional PDF} = \frac{\text{joint PDF}}{\text{marginal PDF}}}$$

Conditional PDF

- For continuous (X, Y), the conditional PDF of Y given $X = x$ is defined as

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)}, \quad \text{provided that } f_X(x) > 0,$$

i.e.,

$$\boxed{\text{conditional PDF} = \frac{\text{joint PDF}}{\text{marginal PDF}}}$$

- Interpretation of conditional PDF (in terms of conditional probability):

$$P(Y \in [y \pm \epsilon/2] \mid X \in [x \pm \epsilon/2]) \approx f_{Y|X}(y|x)\epsilon, \quad \epsilon \to 0$$

since

$$\text{LHS} = \frac{P(X \in [x \pm \epsilon/2], Y \in [y \pm \epsilon/2])}{P(X \in [x \pm \epsilon/2])} \approx \frac{f_{XY}(x,y)\epsilon^2}{f_X(x)\epsilon} = f_{Y|X}(y|x)\epsilon$$

Independence phrased in terms of conditioning

Independence means "conditioning does not change the distribution":

$$\begin{aligned} & X \text{ and } Y \text{ independent} \\ \Leftrightarrow\ & f_{XY}(x, y) = f_X(x) f_Y(y) \qquad \text{for all } x, y \end{aligned}$$

Independence phrased in terms of conditioning

Independence means "conditioning does not change the distribution":

$$
\begin{aligned}
& X \text{ and } Y \text{ independent} \\
\Leftrightarrow\ & f_{XY}(x,y) = f_X(x) f_Y(y) && \text{for all } x, y \\
\Leftrightarrow\ & f_{Y|X}(y|x) = f_Y(y) && \text{for all } x \text{ such that } f_X(x) > 0
\end{aligned}
$$

Independence phrased in terms of conditioning

Independence means "conditioning does not change the distribution":

$$
\begin{aligned}
& X \text{ and } Y \text{ independent} \\
\Leftrightarrow\ & f_{XY}(x,y) = f_X(x) f_Y(y) && \text{for all } x, y \\
\Leftrightarrow\ & f_{Y|X}(y|x) = f_Y(y) && \text{for all } x \text{ such that } f_X(x) > 0 \\
\Leftrightarrow\ & f_{X|Y}(x|y) = f_X(x) && \text{for all } y \text{ such that } f_Y(y) > 0
\end{aligned}
$$

Independence phrased in terms of conditioning

Independence means "conditioning does not change the distribution":

$$\begin{aligned}
& X \text{ and } Y \text{ independent} \\
\Leftrightarrow\ & f_{XY}(x,y) = f_X(x)f_Y(y) && \text{for all } x, y \\
\Leftrightarrow\ & f_{Y|X}(y|x) = f_Y(y) && \text{for all } x \text{ such that } f_X(x) > 0 \\
\Leftrightarrow\ & f_{X|Y}(x|y) = f_X(x) && \text{for all } y \text{ such that } f_Y(y) > 0
\end{aligned}$$

that is,

$$\boxed{\text{independence}} \Leftrightarrow \boxed{\text{conditional PDF} = \text{unconditional PDF}}$$

Independence phrased in terms of conditioning

Independence means "conditioning does not change the distribution":

$$\begin{aligned} & X \text{ and } Y \text{ independent} \\ \Leftrightarrow\ & f_{XY}(x,y) = f_X(x) f_Y(y) && \text{for all } x, y \\ \Leftrightarrow\ & f_{Y|X}(y|x) = f_Y(y) && \text{for all } x \text{ such that } f_X(x) > 0 \\ \Leftrightarrow\ & f_{X|Y}(x|y) = f_X(x) && \text{for all } y \text{ such that } f_Y(y) > 0 \end{aligned}$$

that is,

$$\boxed{\text{independence}} \Leftrightarrow \boxed{\text{conditional PDF} = \text{unconditional PDF}}$$

Similarly, for discrerte RVs

$$\boxed{\text{independence}} \Leftrightarrow \boxed{\text{conditional PMF} = \text{unconditional PMF}}$$

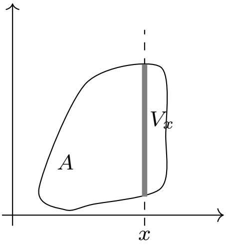
Conditional PDF $\rightarrow$ Joint PDF

Often we use marginal and conditional PDF (given by the problem or our statistical model) to find joint PDF

$$\boxed{f_{XY} = f_{Y|X} f_X}$$

Marginal PDF of 2D uniform

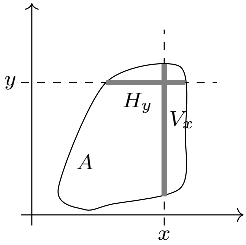
Let (X, Y) be uniformly distributed over some region A (Lec 17):



$$f_X(x) = \int_{V_x} f_{XY}(x,y)dy = \frac{\text{length}(V_x)}{\text{area}(A)}$$

Marginal PDF of 2D uniform

Let (X, Y) be uniformly distributed over some region A (Lec 17):

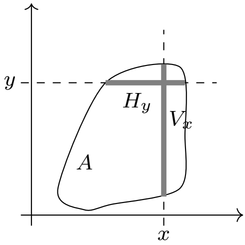


$$f_X(x) = \int_{V_x} f_{XY}(x, y)dy = \frac{\text{length}(V_x)}{\text{area}(A)}$$

$$f_Y(y) = \int_{H_y} f_{XY}(x, y)dx = \frac{\text{length}(H_y)}{\text{area}(A)}$$

where H_y and V_x denote the horizontal and vertical segment intersecting A, respectively.

Conditional PDF of 2D uniform



- Conditioned on $X = x$, Y is uniformly distributed over V_x, so

$$f_{Y|X}(y|x) = \begin{cases} \frac{1}{\text{length}(V_x)} & y \in V_x \\ 0 & \text{else} \end{cases}$$

- Conditioned on $Y = y$, X is uniformly distributed over H_y, so

$$f_{X|Y}(x|y) = \begin{cases} \frac{1}{\text{length}(H_y)} & x \in H_y \\ 0 & \text{else} \end{cases}$$

- Discrete counterpart: "Equiprobable experiment remains equiprobable after conditioning" (Lec03).

Example: Stick breaking

Given a stick of unit length, let's break it into three pieces as follows:

- First, choose X uniformly on the stick
- Next, choose Y uniformly on the remaining stick

Example: Stick breaking

Given a stick of unit length, let's break it into three pieces as follows:

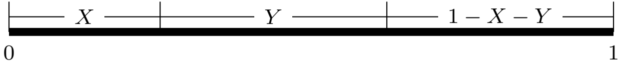
- First, choose X uniformly on the stick
- Next, choose Y uniformly on the remaining stick

In other words,

- X is the length of the 1st piece
- Y is the length of the 2nd piece

Find the joint PDF f_{XY}, the marginal PDF f_Y, and $E(Y)$.

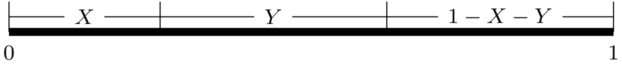
Example: Stick breaking



- We know $X \sim \text{Unif}(0, 1)$. So the PDF of X is:

$$f_X(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{else} \end{cases}$$

Example: Stick breaking



- We know $X \sim \text{Unif}(0,1)$. So the PDF of X is:

$$f_X(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{else} \end{cases}$$

- Conditioned on $X = x$, $Y \sim \text{Unif}(0, 1-x)$. So the conditional PDF of Y given X is

$$f_{Y|X}(y|x) = \begin{cases} \frac{1}{1-x} & 0 < y < 1-x \\ 0 & \text{else} \end{cases}$$

Example: Stick breaking

- Use $f_{XY} = f_{Y|X} f_X$ to find joint PDF:

$$f_{XY}(x, y) = \begin{cases} \frac{1}{1-x} & 0 < x + y < 1 \\ 0 & \text{else} \end{cases}$$

Example: Stick breaking

- Use $f_{XY} = f_{Y|X} f_X$ to find joint PDF:

$$f_{XY}(x,y) = \begin{cases} \frac{1}{1-x} & 0 < x+y < 1 \\ 0 & \text{else} \end{cases}$$

- Marginalize to get f_Y:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)dx = \int_0^{1-y} \frac{1}{1-x}dx = -\ln y$$

if $0 < y < 1$; otherwise, $f_Y(y) = 0$.

Example: Stick breaking

- Use $f_{XY} = f_{Y|X} f_X$ to find joint PDF:

$$f_{XY}(x,y) = \begin{cases} \frac{1}{1-x} & 0 < x+y < 1 \\ 0 & \text{else} \end{cases}$$

- Marginalize to get f_Y:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)dx = \int_0^{1-y} \frac{1}{1-x}dx = -\ln y$$

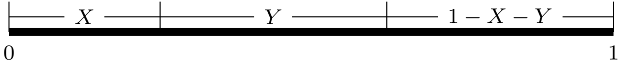
if $0 < y < 1$; otherwise, $f_Y(y) = 0$.

- Find the average length of the 2nd piece:

$$E(Y) = \int_{-\infty}^{\infty} y f_Y(y)dy = \int_0^1 -y \ln y dy = \frac{1}{4}$$

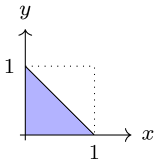
Compare: average length of the 1nd piece $E(X) = 1/2$.

Some observations



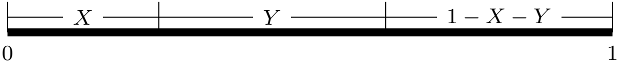
Joint PDF:

$$f_{XY}(x,y) = \begin{cases} \frac{1}{1-x} & 0 < x + y < 1 \\ 0 & \text{else} \end{cases}$$

- Support of $(X, Y) =$ 

- Support of X = Support of $Y = [0, 1]$
- So X and Y are dependent, clearly (if the 1st piece is longer, the 2nd piece tends to be shorter)

Example: Stick breaking

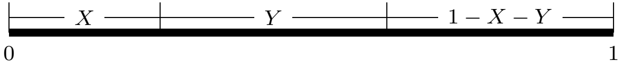


Joint PDF:

$$f_{XY}(x, y) = \begin{cases} \frac{1}{1-x} & 0 < x + y < 1 \\ 0 & \text{else} \end{cases}$$

- What is the chance that 2nd piece is longer than 1st piece?

Example: Stick breaking



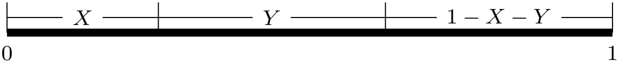
Joint PDF:

$$f_{XY}(x,y) = \begin{cases} \frac{1}{1-x} & 0 < x + y < 1 \\ 0 & \text{else} \end{cases}$$

- What is the chance that 2nd piece is longer than 1st piece?

$$P(Y > X) = \int_0^{1/2} \left(\int_x^{1-x} \frac{1}{1-x} dy \right) dx = \int_0^{1/2} \frac{1-2x}{1-x} dx$$
$$= 1 - \log 2 \approx 30.6\% < \frac{1}{2}$$

Example: Stick breaking



Joint PDF:

$$f_{XY}(x,y) = \begin{cases} \frac{1}{1-x} & 0 < x + y < 1 \\ 0 & \text{else} \end{cases}$$

- What is the chance that 2nd piece is longer than 1st piece?

$$P(Y > X) = \int_0^{1/2} \left(\int_x^{1-x} \frac{1}{1-x} dy \right) dx = \int_0^{1/2} \frac{1-2x}{1-x} dx$$
$$= 1 - \log 2 \approx 30.6\% < \frac{1}{2}$$

- On average, 2nd piece is shorter:

$$E(X) = \frac{1}{2} > E(Y) = \frac{1}{4}$$

Multiple random variables

Random vector

Let $X_1, \ldots, X_n$ be random variables.

- $(X_1, \ldots, X_n)$ is called a random vector, which represents a random point in the Euclidean space $\mathbb{R}^n$. For example, (X_1, X_2) is a random point on the plane.

Random vector

Let $X_1, \ldots, X_n$ be random variables.

- $(X_1, \ldots, X_n)$ is called a random vector, which represents a random point in the Euclidean space $\mathbb{R}^n$. For example, (X_1, X_2) is a random point on the plane.
- The concepts of joint PMF $p_{X_1,\ldots,X_n}$ (for discrete) or joint PDF $f_{X_1,\ldots,X_n}$ (for continuous) are similarly defined and applied, e.g., for the latter,

$$P((X_1, \ldots, X_n) \in A) = \int_A f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) dx_1 \ldots dx_n, \quad A \subset \mathbb{R}^n$$

Independence of multiple random variables

$X_1, X_2, \ldots, X_n$ are mutually independent if

- For discrete random variables, their joint PMF factorizes as product of marginal PMFs:

$$p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = p_{X_1}(x_1) \times \ldots \times p_{X_n}(x_n)$$

- For continuous random variables, their joint PDF factorizes as product of marginal PDFs:

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = f_{X_1}(x_1) \times \ldots \times f_{X_n}(x_n)$$

Example: (X_1, X_2, X_3) uniform over the cube $[0, 1]^3$. Then X_i are iid Unif$(0, 1)$.