Multiple discrete/continuous random variables

Andrew Thangaraj

IIT Madras

Subsection 1

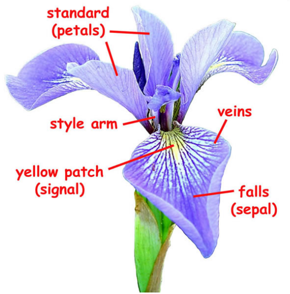
Motivation

Iris data set

- First used by R. A. Fisher
  - Wikipedia: https://en.wikipedia.org/wiki/Ronald_Fisher
    - "a genius who almost single-handedly created the foundations for modern statistical science"
    - "the single most important figure in 20th century statistics"

Iris data set

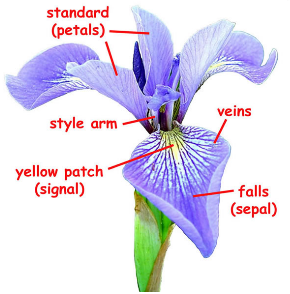
- First used by R. A. Fisher
  - Wikipedia: https://en.wikipedia.org/wiki/Ronald_Fisher
    - "a genius who almost single-handedly created the foundations for modern statistical science"
    - "the single most important figure in 20th century statistics"
- Iris flower
  - 3 classes of irises: 0, 1 and 2
    - 50 instances in each class
  - Data (cm)
    - sepal length (SL), sepal width (SW), petal length (PL), petal width (PW)
  - Classification
    - Given data, find class



(image source: fs.fed.us)

Iris data set

- First used by R. A. Fisher
  - Wikipedia: https://en.wikipedia.org/wiki/Ronald_Fisher
    - ★ "a genius who almost single-handedly created the foundations for modern statistical science"
    - ★ "the single most important figure in 20th century statistics"
- Iris flower
  - 3 classes of irises: 0, 1 and 2
    - ★ 50 instances in each class
  - Data (cm)
    - ★ sepal length (SL), sepal width (SW), petal length (PL), petal width (PW)
  - Classification
    - ★ Given data, find class



(image source: fs.fed.us)

How to statistically describe (class, SL, SW, PL, PW)?

Iris data

Class 0

SL	SW	PL	PW
5.1	3.5	1.4	0.2
4.9	3.0	1.4	0.2
4.7	3.2	1.3	0.2
4.6	3.1	1.5	0.2
5.0	3.6	1.4	0.2
⋮	⋮	⋮	⋮

Class 1

SL	SW	PL	PW
7.0	3.2	4.7	1.4
6.4	3.2	4.5	1.5
6.9	3.1	4.9	1.5
5.5	2.3	4.0	1.3
6.5	2.8	4.6	1.5
⋮	⋮	⋮	⋮

Class 2

SL	SW	PL	PW
6.3	3.3	6.0	2.5
5.8	2.7	5.1	1.9
7.1	3.0	5.9	2.1
6.3	2.9	5.6	1.8
6.5	3.0	5.8	2.2
⋮	⋮	⋮	⋮

Iris data

Class 0

SL	SW	PL	PW
5.1	3.5	1.4	0.2
4.9	3.0	1.4	0.2
4.7	3.2	1.3	0.2
4.6	3.1	1.5	0.2
5.0	3.6	1.4	0.2
⋮	⋮	⋮	⋮

Class 1

SL	SW	PL	PW
7.0	3.2	4.7	1.4
6.4	3.2	4.5	1.5
6.9	3.1	4.9	1.5
5.5	2.3	4.0	1.3
6.5	2.8	4.6	1.5
⋮	⋮	⋮	⋮

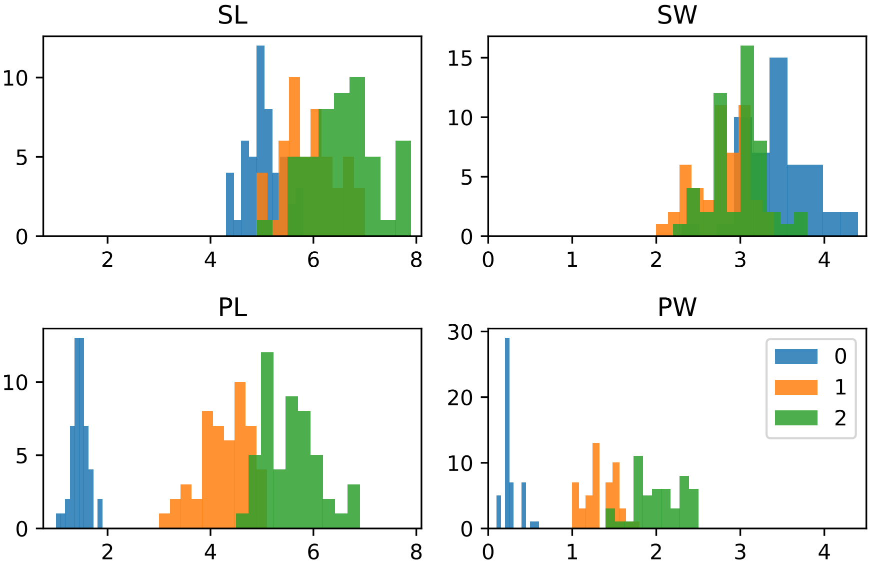
Class 2

SL	SW	PL	PW
6.3	3.3	6.0	2.5
5.8	2.7	5.1	1.9
7.1	3.0	5.9	2.1
6.3	2.9	5.6	1.8
6.5	3.0	5.8	2.2
⋮	⋮	⋮	⋮

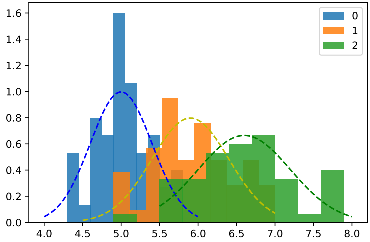
Summary: min-max, avg, stdev

	SL summary	SW summary	PL summary	PW summary
0	4.3-5.8,5.0,0.4	2.3-4.4,3.4,0.4	1.0-1.9,1.5,0.2	0.1-0.6,0.3,0.1
1	4.9-7.0,5.9,0.5	2.0-3.4,2.8,0.3	3.0-5.1,4.3,0.5	1.0-1.8,1.3,0.2
2	4.9-7.9,6.6,0.6	2.2-3.8,3.0,0.3	4.5-6.9,5.6,0.6	1.4-2.5,2.0,0.3

Histograms

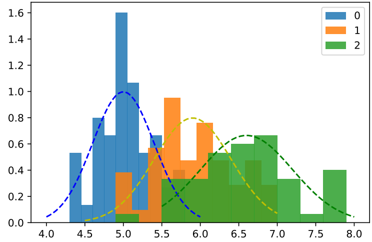


How to model sepal length and class of iris?



- density histograms of sepal length for three classes
- continuous approximations shown as dotted lines

How to model sepal length and class of iris?



- density histograms of sepal length for three classes
- continuous approximations shown as dotted lines

- Clearly, both are jointly distributed
- Class: discrete $\in \{0, 1, 2\}$
- Sepal length: continuous
  - distribution depends on class

Subsection 2

Joint distributions: Discrete and Continuous

Describing discrete-continuous joint distributions

- (X, Y): jointly distributed

Describing discrete-continuous joint distributions

- (X, Y): jointly distributed
- X: discrete with range T_X and PMF $p_X(x)$

Describing discrete-continuous joint distributions

- (X, Y): jointly distributed
- X: discrete with range T_X and PMF $p_X(x)$
- For each $x \in T_X$, we have a continuous random variable Y_x with density $f_{Y_x}(y)$

Describing discrete-continuous joint distributions

- (X, Y): jointly distributed
- X: discrete with range T_X and PMF $p_X(x)$
- For each $x \in T_X$, we have a continuous random variable Y_x with density $f_{Y_x}(y)$
- Y_x: Y given $X = x$, denoted $(Y|X = x)$

Describing discrete-continuous joint distributions

- (X, Y): jointly distributed
- X: discrete with range T_X and PMF $p_X(x)$
- For each $x \in T_X$, we have a continuous random variable Y_x with density $f_{Y_x}(y)$
- Y_x: Y given $X = x$, denoted $(Y|X = x)$
- $f_{Y_x}(y)$: conditional density of Y given $X = x$, denoted $f_{Y|X=x}(y)$

Describing discrete-continuous joint distributions

- (X, Y): jointly distributed
- X: discrete with range T_X and PMF $p_X(x)$
- For each $x \in T_X$, we have a continuous random variable Y_x with density $f_{Y_x}(y)$
- Y_x: Y given $X = x$, denoted $(Y|X = x)$
- $f_{Y_x}(y)$: conditional density of Y given $X = x$, denoted $f_{Y|X=x}(y)$
- Marginal density of Y

$$f_Y(y) = \sum_{x \in T_X} \overset{P(X=x)}{p_X(x)} \overbrace{f_{Y|X=x}(y)}^{\text{density}\,(Y|X=x)}$$

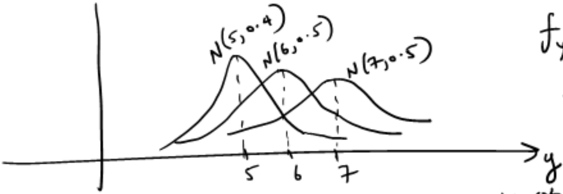
Problem

discrete uniform   conditional densities

Let $X \sim$ Uniform$\{0, 1, 2\}$. Let $Y|X{=}0 \sim$ Normal$(\underset{\mu=}{5}, \underset{\sigma=}{0.4})$,
$Y|X{=}1 \sim$ Normal$(\underset{\mu=}{6}, \underset{\sigma=}{0.5})$ and $Y|X{=}2 \sim$ Normal$(\underset{\mu=}{7}, \underset{\sigma=}{0.6})$.

N $\leftrightarrow$ Normal

- What is the marginal of Y?
- Suppose we observe Y to be around y_0. What can you say about X?



$$f_{Y|X=0}(y) = \frac{1}{\sqrt{2\pi}\times 0.4} e^{-\frac{(y-5)^2}{2(0.4)^2}}$$

$$f_{Y|X=1}(y) = \ ?$$

$$f_{Y|X=2}(y) = \ ?$$

$$f_Y(y) = \frac{1}{3} \cdot \frac{1}{\sqrt{2\pi}\cdot 0.4} e^{-\frac{(y-5)^2}{2(0.4)^2}} + \frac{1}{3} \cdot \frac{1}{\sqrt{2\pi}\cdot 0.5} e^{-\frac{(y-6)^2}{2(0.5)^2}} + \frac{1}{3} \cdot \frac{1}{\sqrt{2\pi}\cdot 0.6} e^{-\frac{(y-7)^2}{2(0.6)^2}}$$

"Not Gaussian"

called "Mixture of Gaussians"

Conditional probability of discrete given continuous

Definition

Suppose X and Y are jointly distributed with $X \in T_X$ being discrete with PMF $p_X(x)$ and conditional densities $f_{Y|X=x}(y)$ for $x \in T_X$. The conditional probability of X given $Y = y_0 \in \text{supp}(Y)$ is defined as

$$P(X{=}x|Y{=}y_0) = \frac{p_X(x) f_{Y|X=x}(y_0)}{\underbrace{f_Y(y_0)}_{= \sum_{x \in T_X} p_X(x) f_{Y|X=x}(y_0)}},$$

where f_Y is the marginal density of Y.

Conditional probability of discrete given continuous

Definition

Suppose X and Y are jointly distributed with $X \in T_X$ being discrete with PMF $p_X(x)$ and conditional densities $f_{Y|X=x}(y)$ for $x \in T_X$. The conditional probability of X given $Y = y_0 \in \text{supp}(Y)$ is defined as

$$P(X{=}x|Y{=}y_0) = \frac{p_X(x) f_{Y|X=x}(y_0)}{f_Y(y_0)},$$

where f_Y is the marginal density of Y.

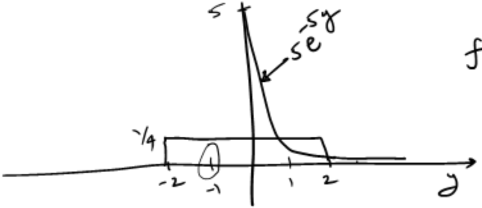
$P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$ $\rightarrow X=x$

- Similar to Bayes' rule: $P(X{=}x|Y{=}y_0) f_Y(y_0) = f_{Y|X=x}(y_0) p_X(x)$
- $X|Y{=}y_0$: "conditioned" discrete random variable
- When are X and Y independent? $f_{Y|X=x}$ is independent of x.
  - $f_Y = f_{Y|X=x}$ and $P(X{=}x|Y{=}y_0) = p_X(x)$

Problem

discrete   continuous

Let $X \sim \text{Uniform}\{-1, 1\}$. Let $Y|X{=}-1 \sim \text{Uniform}[-2, 2]$, $Y|X{=}1 \sim \text{Exp}(5)$. Find the distribution of X given $Y = -1$, $Y = 1$, $Y = 3$.



$$f_Y(y) = \frac{1}{2}\cdot \underbrace{f_{Y|X=-1}(y)}_{\frac{1}{4},\ y\in(-2,2)} + \frac{1}{2}\underbrace{f_{Y|X=1}(y)}_{5e^{-5y},\ y>0}$$

$$= \begin{cases} 0, & y<-2 \\ \frac{1}{2}\cdot\frac{1}{4}, & -2<y<0 \\ \frac{1}{2}\cdot\frac{1}{4}+\frac{1}{2}5e^{-5y}, & 0<y<2 \\ \frac{1}{2}5e^{-5y}, & y>2 \end{cases}$$

$X|Y=-1$: $P(X=-1|Y=-1) = \dfrac{p_X(-1)\cdot f_{Y|X=-1}(-1)}{f_Y(-1)} = \dfrac{\frac{1}{2}\cdot\frac{1}{4}}{\frac{1}{2}\cdot\frac{1}{4}} = 1$

$P(X=+1|Y=-1) = \dfrac{p_X(1)\, f_{Y|X=1}(-1)}{f_Y(-1)} = \dfrac{\frac{1}{2}\cdot 0}{\frac{1}{2}\cdot\frac{1}{4}} = 0$

$X|Y=1$: $P(X=-1|Y=1) = \dfrac{\frac{1}{2}\cdot\frac{1}{4}}{\frac{1}{2}\cdot\frac{1}{4}+\frac{1}{2}\cdot 5e^{-5}}$; $P(X=1|Y=1) = 1-P(X=-1|Y=1)$

$= \dfrac{\frac{1}{2}\cdot 5e^{-5}}{\frac{1}{2}\cdot\frac{1}{4}+\frac{1}{2}\cdot 5e^{-5}}$

$X|Y=3$: $P(X=-1|Y=3)=0$, $P(X=1|Y=3)=1$

Problem

Suppose 60% of adults in the age group of 45-50 in a country are male and 40% are female. Suppose the height (in cm) of adult males in that age group in the country is Normal(160, $\sigma=$ 10), and that of females is Normal(150, $\sigma=$ 5). A random person is found to have a height of 155 cm. Is that person more likely to be male or female?

$$X \sim \{\overset{0.6}{m}, \overset{0.4}{F}\}$$

$$Y|X=m \sim N(160, \sigma=10) \quad f_{Y|X=m}(y) = \frac{1}{\sqrt{2\pi}\cdot 10} e^{-\frac{(y-160)^2}{2(10)^2}}$$

$$Y|X=F \sim N(150, \sigma=5) \quad f_{Y|X=F}(y) = \text{"}$$

$$f_Y(y) = ?$$

$$X|Y=155: \quad P(X=m|Y=155) = \frac{0.6 \times \frac{1}{\sqrt{2\pi}\cdot 10} e^{-\frac{5^2}{2\cdot 10^2}}}{0.6 \times \frac{1}{\sqrt{2\pi}\cdot 10} \cdot e^{-\frac{5^2}{2\cdot 10^2}} + 0.4 \cdot \frac{1}{\sqrt{2\pi}\cdot 5} e^{-\frac{5^2}{2\cdot 5^2}}} = ?$$

$$P(X=F|Y=155) = 1 - P(X=m|Y=155) = ?$$

Problem

variance

Let $Y = X + Z$, where $X \sim \text{Uniform}\{-3,-1,1,3\}$ and $Z \sim \text{Normal}(0, \sigma^2)$ are independent. What is the distribution of Y? Find the distribution of $(X|Y = 0.5)$.

$f_{Y|X=-3}(y) = ?$

$Y|X=-3 \Leftrightarrow (-3+Z) \sim N(-3, \sigma^2)$
$Y|X=-1 \Leftrightarrow (-1+Z) \sim N(-1, \sigma^2)$
$Y|X=1 \Leftrightarrow N(1, \sigma^2)$; $Y|X=3 \Leftrightarrow N(3, \sigma^2)$

$\rightarrow$ rest is same as before.

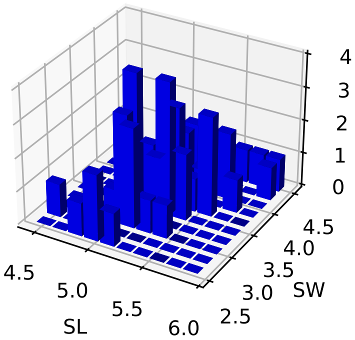
Exercise: $X \sim \text{Unif}\{-3,-1,1,3\}$, $Z \sim N(0, \sigma^2)$ indep

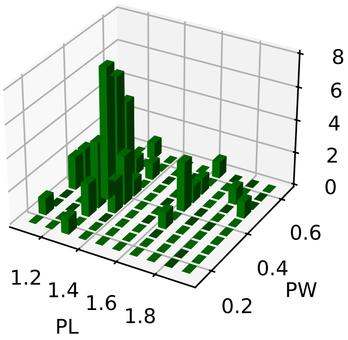
$Y = XZ$

Subsection 3

Jointly continuous random variables

2D histograms: (SL, SW) and (PL, PW) for Class 0





- Count the number of (x, y) falling into a rectangular bin
- (SL, SW): Both continuous and they have a joint distribution
  - Same for (PL, PW)

Joint density in two dimensions

Definition (Joint density)

A function $f(x, y)$ is said to be a joint density function if

- $f(x, y) \geq 0$, i.e. f is non-negative
- $\iint_{-\infty}^{\infty} f(x, y) dx dy = 1$
- Technical: $f(x, y)$ is piecewise continuous in each variable

Joint density in two dimensions

Definition (Joint density)

A function $f(x, y)$ is said to be a joint density function if

- $f(x, y) \geq 0$, i.e. f is non-negative
- $\iint_{-\infty}^{\infty} f(x, y)dxdy = 1$
- Technical: $f(x, y)$ is piecewise continuous in each variable

- For every joint density $f(x, y)$, there exist two jointly distributed continuous random variables X and Y such that, for any two-dimensional region A,

$$P((X, Y) \in A) = \iint_A f(x, y)dxdy$$

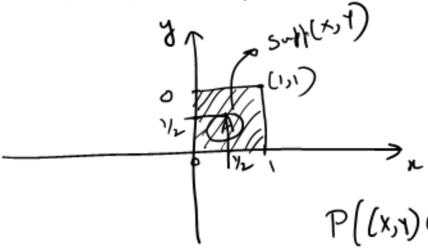
- $f(x, y)$, also denoted $f_{XY}(x, y)$, is called the joint density of X and Y
- $\text{supp}(X, Y) = \{(x, y) : f_{XY}(x, y) > 0\}$

Example: Uniform in the unit square

Let X and Y have joint density

$$f_{XY}(x,y) = \begin{cases} 1 & 0 < x < 1, 0 < y < 1, \\ 0 & \text{otherwise.} \end{cases}$$

- Picture the 3D plot of the joint density
- To compute probability, find the area of the region



$$\int_0^1\int_0^1 1\,dx\,dy = \int_0^1 \underbrace{x\Big|_0^1}_{=1} dy = \int_0^1 1\,dy$$

$$= y\Big|_0^1 = 1$$

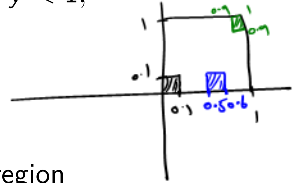
$$P\big((X,Y)\in A\big) = \iint_A 1\,dx\,dy = \text{Area of } A$$

$$P\left(0<X<\tfrac{1}{2}, 0<Y<\tfrac{1}{2}\right) = \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}$$

Example: Uniform in the unit square

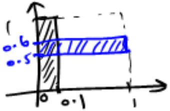
Let X and Y have joint density

$$f_{XY}(x,y) = \begin{cases} 1 & 0 < x < 1, 0 < y < 1, \\ 0 & \text{otherwise.} \end{cases}$$



- Picture the 3D plot of the joint density
- To compute probability, find the area of the region
- $P(0 < X < 0.1, 0 < Y < 0.1) \overset{0.01}{=}?$, $P(0.5 < X < 0.6, 0 < Y < 0.1) =?$, $P(0.9 < X < 1, 0.9 < Y < 1) =?$

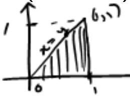
  $0.1 \times 0.1 = 0.01$ $0.1 \times 0.1 = 0.01$ $0.1 \times 0.1 = 0.01$



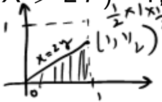
- $P(0 < X < 0.1) =?$, $P(0.5 < Y < 0.6) =?$

  $1 \times 0.1 = 0.1$ $1 \times 0.1 = 0.1$

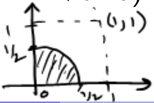
- $P(X > Y) =?$, $P(X > 2Y) =?$, $P(X^2 + Y^2 < 0.25) =?$



$\frac{1}{2} \times 1 \times 1 = \frac{1}{2}$



$\frac{1}{2} \times 1 \times \frac{1}{2} = \frac{1}{4}$



$\frac{1}{4} \times \pi \left(\frac{1}{2}\right)^2 = \frac{\pi}{16}$

2D uniform distribution

Fix some (reasonable) region D in $\mathbb{R}^2$ with total area $|D|$. We say that $(X, Y) \sim \text{Uniform}(D)$ if they have the joint density

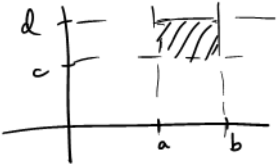
$$f_{XY}(x, y) = \begin{cases} \frac{1}{|D|} & (x, y) \in D, \\ 0 & \text{otherwise.} \end{cases}$$

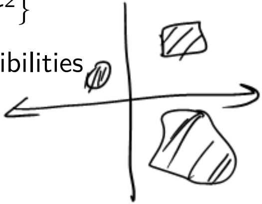
2D uniform distribution

Fix some (reasonable) region D in $\mathbb{R}^2$ with total area $|D|$. We say that $(X, Y) \sim \text{Uniform}(D)$ if they have the joint density

$$f_{XY}(x, y) = \begin{cases} \frac{1}{|D|} & (x, y) \in D, \\ 0 & \text{otherwise.} \end{cases}$$

- Rectangle: $D = [a, b] \times [c, d] = \{(x, y) : a < x < b, c < y < d\}$
- Circle: $D = \{(x, y) : (x - x_0)^2 + (y - y_0)^2 \leq r^2\}$
- Multiple disjoint areas and so many other possibilities





2D uniform distribution

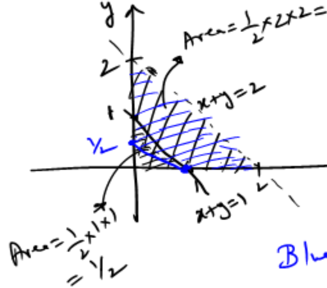
Fix some (reasonable) region D in $\mathbb{R}^2$ with total area $|D|$. We say that $(X, Y) \sim \text{Uniform}(D)$ if they have the joint density

$$f_{XY}(x, y) = \begin{cases} \frac{1}{|D|} & (x, y) \in D, \\ 0 & \text{otherwise.} \end{cases}$$

- Rectangle: $D = [a, b] \times [c, d] = \{(x, y) : a < x < b, c < y < d\}$
- Circle: $D = \{(x, y) : (x - x_0)^2 + (y - y_0)^2 \leq r^2\}$
- Multiple disjoint areas and so many other possibilities
- For any sub-region A of D, $P((X, Y) \in A) = |A|/|D| = \frac{\text{Area}(A)}{\text{Area}(D)}$
- Uniform distribution is a good approximation for flat histograms

Problem: 2D uniform

Let $(X, Y) \sim \text{Uniform}(D)$, where $D = \{(x, y) : x + y < 2, x > 0, y > 0\}$.
Sketch the support and compute $P(X + Y < 1)$, $P(X + 2Y > 1)$.



$$f_{XY}(x,y) = \begin{cases} \frac{1}{2}, & (x,y) \in D \\ 0, & \text{otherwise} \end{cases}$$

Blue area $= 2 - \frac{1}{2} \times 1 \times \frac{1}{2} = \frac{7}{4}$

$P(X+Y<1) = \frac{1/2}{2} = 1/4$

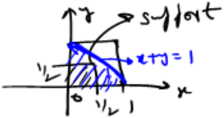
$P(X+2Y>1) = \frac{7/4}{2} = 7/8$

Problem: 2D non-uniform

Let (X, Y) have joint density

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

$f_{XY} \geq 0$



Show that the above is a valid density. Find $P(X < 1/2 \text{ and } Y < 1/2)$, $P(X + Y < 1)$.

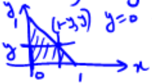
$$\int_{y=0}^{1} \underbrace{\int_{x=0}^{1} (x+y)\, dx}_{\text{Treat } y\text{: constant}}\, dy = \int_{y=0}^{1} \left(\frac{1}{2} + y\right) dy = \frac{y}{2} + \frac{y^2}{2}\Big|_0^1 = 1$$

$$\int_0^1 (x+y)\, dx = \frac{x^2}{2} + yx \Big|_0^1 = \frac{1}{2} + y$$

$$P(X<1/2, Y<1/2) = \int_0^{1/2}\int_0^{1/2} (x+y)\, dx\, dy = \int_0^{1/2} \left(\frac{x^2}{2} + yx\Big|_0^{1/2}\right) dy = \int_0^{1/2} \left(\frac{1}{8} + \frac{y}{2}\right) dy = \frac{y}{8} + \frac{y^2}{4}\Big|_0^{1/2} = 1/8$$

$$P(X+Y<1) = \int_{y=0}^{1}\int_{x=0}^{1-y} (x+y)\, dx\, dy = \int \frac{x^2}{2} + yx \Big|_0^{1-y} dy = \int_0^1 \frac{1}{2}(1-2y+y^2) + y - y^2\, dy$$

$$= \int_0^1 \left(\frac{1}{2} - \frac{y^2}{2}\right) dy = \frac{y}{2} - \frac{y^3}{6}\Big|_0^1 = \frac{1}{2} - \frac{1}{6} = \frac{1}{3}$$



Subsection 4

Marginal densities and independence

Marginal density

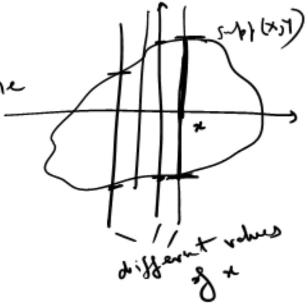
Theorem (Marginal density)

Suppose (X, Y) have joint density $f_{XY}(x, y)$. Then,

- *X has the marginal density $f_X(x) = \int_{y=-\infty}^{\infty} f_{XY}(x, y)dy$*
- *Y has the marginal density $f_Y(y) = \int_{x=-\infty}^{\infty} f_{XY}(x, y)dx$*

$$f_X(x) = \int_{y=-\infty}^{\infty} \underbrace{f_{XY}(x,y)}_{\text{constant}} \, \underbrace{dy}_{\text{variable}}$$

fix 'x'



Marginal density

Theorem (Marginal density)

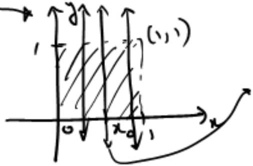
Suppose (X, Y) have joint density $f_{XY}(x, y)$. Then,

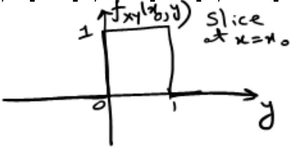
- *X has the marginal density $f_X(x) = \int_{y=-\infty}^{\infty} f_{XY}(x, y)dy$*
- *Y has the marginal density $f_Y(y) = \int_{x=-\infty}^{\infty} f_{XY}(x, y)dx$*

- The PDF of X and Y individually are called marginal densities
- The joint density exactly determines both the marginal densities

Examples: (Marginals do not determine joint)

- Uniform on unit square
- $(X, Y) \sim \text{Uniform}(D)$, where
  $D = [0, 1/2] \times [0, 1/2] \cup [1/2, 1] \times [1/2, 1]$

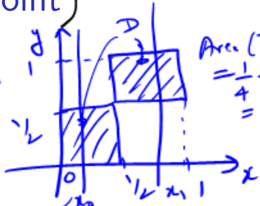




$$f_X(x_0) = \int_{y=0}^{1} \underbrace{f_{XY}(x_0,y)}_{1}\, dy = 1, \quad 0<x_0<1$$

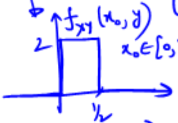
$X \sim \text{Uniform}[0,1]$

$Y \sim \text{Uniform}[0,1]$

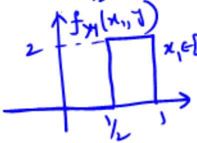


$\text{Area}(D) = \frac{1}{4}+\frac{1}{4} = \frac{1}{2}$

$$f_{XY}(x,y) = \begin{cases} 2, & (x,y)\in D \\ 0, & \text{otherwise} \end{cases}$$



$x_0 \in [0,1/2]$, $f_X(x_0)=1$



$x_1 \in [1/2,1]$, $f_X(x_1)=1$

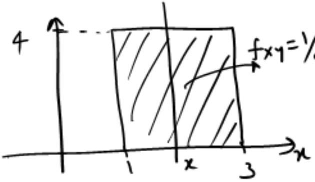
$f_X(x) = 1, \quad 0<x<1$

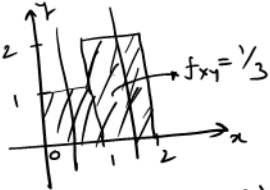
$X \sim \text{Uniform}[0,1]$

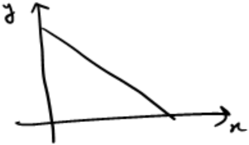
$Y \sim \text{Uniform}[0,1]$

Examples: More uniform

- $(X, Y) \sim \text{Uniform}(D)$, where $D = [1, 3] \times [0, 4]$
- $(X, Y) \sim \text{Uniform}(D)$, where $D = [0, 1] \times [0, 1] \cup [1, 2] \times [0, 2]$
- $(X, Y) \sim \text{Uniform}(D)$, where $D = \{(x, y) : x + y < 2, x > 0, y > 0\}$





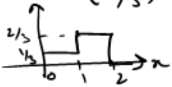


$f_X(x) = 1/2,\ 1 < x < 3$

$X \sim \text{Uniform}[1, 3]$

$Y \sim \text{Uniform}[0, 4]$

$$f_X(x) = \begin{cases} 1/3, & 0 < x < 1 \\ 2/3, & 1 < x < 2 \end{cases}$$



$$f_Y(y) = \begin{cases} 2/3, & 0 < y < 1 \\ 1/3, & 1 < y < 2 \end{cases}$$

Problem

Consider the joint density

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

Find the marginals.

Independence

Theorem (Independence)

(X, Y) *with joint density* $f_{XY}(x, y)$ *are independent if*

$$f_{XY}(x, y) = f_X(x) f_Y(y),$$

where $f_X(x)$ *and* $f_Y(y)$ *are the marginal densities.*

Independence

Theorem (Independence)

(X, Y) with joint density $f_{XY}(x, y)$ are independent if

$$f_{XY}(x, y) = f_X(x) f_Y(y),$$

where $f_X(x)$ and $f_Y(y)$ are the marginal densities.

- Given the joint density, the marginals can be computed
- If the joint density is the product of the marginal densities, then X and Y are independent
- So, if independent, the marginals determine the joint density

Examples

- Uniform on unit square
- $(X, Y) \sim \text{Uniform}(D)$, where
  $D = [0, 1/2] \times [0, 1/2] \cup [1/2, 1] \times [1/2, 1]$
- $(X, Y) \sim \text{Uniform}(D)$, where $D = [1, 3] \times [0, 4]$
- $(X, Y) \sim \text{Uniform}(D)$, where $D = [0, 1] \times [0, 1] \cup [1, 2] \times [0, 2]$
- $(X, Y) \sim \text{Uniform}(D)$, where $D = \{(x, y) : x + y < 2, x > 0, y > 0\}$
- $$f_{XY}(x, y) = \begin{cases} x + y, & 0 < x, y < 1, \\ 0, & \text{otherwise.} \end{cases}$$

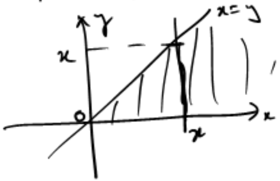
Problem

Suppose $X \sim \text{Exp}(\lambda_1)$, $Y \sim \text{Exp}(\lambda_2)$ are independent random variables. Find their joint density and compute $P(X > Y)$.

$f_X(x) = \lambda_1 e^{-\lambda_1 x}, \; x > 0$

$f_Y(y) = \lambda_2 e^{-\lambda_2 y}, \; y > 0$

$f_{XY}(x,y) = \lambda_1 e^{-\lambda_1 x} \cdot \lambda_2 e^{-\lambda_2 y}, \; x > 0, y > 0$



$$P(X>Y) = \int_{x=0}^{\infty}\left(\int_{y=0}^{x} \lambda_1 e^{-\lambda_1 x}\, \lambda_2 e^{-\lambda_2 y}\, dy\right) dx$$

$$= \int_{x=0}^{\infty} \lambda_1 e^{-\lambda_1 x}\left(-e^{-\lambda_2 y}\Big|_0^x\right) dx = \int_0^\infty \lambda_1 e^{-\lambda_1 x}\left(1 - e^{-\lambda_2 x}\right) dx$$

$$= \underbrace{\int_0^\infty \lambda_1 e^{-\lambda_1 x}\, dx}_{1} - \int_0^\infty \lambda_1 e^{-(\lambda_1+\lambda_2)x}\, dx$$

integral: $\dfrac{-e^{-(\lambda_1+\lambda_2)x}}{\lambda_1+\lambda_2}\Big|_0^\infty = \dfrac{1}{\lambda_1+\lambda_2}$

$$= 1 - \frac{\lambda_1}{\lambda_1+\lambda_2} = \frac{\lambda_2}{\lambda_1+\lambda_2}$$

Subsection 5

Conditional density

Definition of conditional density

Definition (Conditional density)

Let (X, Y) be random variables with joint density $f_{XY}(x, y)$. Let $f_X(x)$ and $f_Y(y)$ be the marginal densities.

- For a such that $f_X(a) > 0$ the conditional density of Y given $X = a$, denoted $f_{Y|X=a}(y)$, is defined as

$$\int_{-\infty}^{\infty} f_{Y|X=a}(y)\,dy = 1$$

$$f_{Y|X=a}(y) = \frac{f_{XY}(a, y)}{f_X(a)}.$$

slice of joint at $x=a$

$$f_X(a) = \int_{-\infty}^{\infty} f_{XY}(a,y)\,dy$$

- For b such that $f_Y(b) > 0$, the conditional density of X given $Y = a$, denoted $f_{X|Y=b}(x)$, is defined as

$$f_{X|Y=b}(x) = \frac{f_{XY}(x, b)}{f_Y(b)}.$$

Properties of conditional density

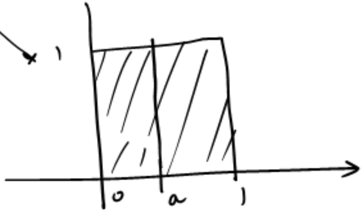
- Both the conditional densities are valid densities in one dimension. So, the "conditional" random variables $(Y|X = a)$ and $(X|Y = b)$ are well-defined.
- Joint = Marginal ~~times~~ $\times$ Conditional, for $x = a$ and $y = b$ such that $f_X(a) > 0$ and $f_Y(b) > 0$

$$f_{XY}(a, b) = f_X(a) f_{Y|X=a}(b) = f_Y(b) f_{X|Y=b}(a)$$

- The above is usually written as $f_{XY}(x, y) = f_X(x) f_{Y|X=x}(y) = f_Y(y) f_{X|Y=y}(x)$

Examples: Uniform

- Uniform on unit square
- $(X, Y) \sim \text{Uniform}(D)$, where
  $D = [0, 1/2] \times [0, 1/2] \cup [1/2, 1] \times [1/2, 1]$



$Y|X=a \sim \text{Uniform}[0,1]$
$0<a<1$

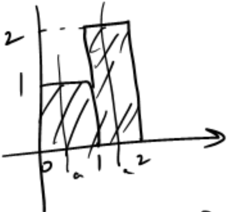
$X|Y=b \sim \text{Uniform}[0,1]$
$0<b<1$

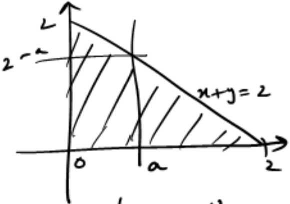
$Y|X=a \sim \text{Uniform}[0,1/2]$
$0<a<1/2$

$Y|X=a \sim \text{Uniform}[1/2,1]$
$1/2<a<1$

Examples: More uniform

- $(X, Y) \sim \text{Uniform}(D)$, where $D = [1, 3] \times [0, 4]$
- $(X, Y) \sim \text{Uniform}(D)$, where $D = [0, 1] \times [0, 1] \cup [1, 2] \times [0, 2]$
- $(X, Y) \sim \text{Uniform}(D)$, where $D = \{(x, y) : x + y < 2, x > 0, y > 0\}$





$Y \mid X=a \sim \text{Uniform}[0,4]$, $1<a<3$

$X \mid Y=b \sim \text{Uniform}[1,3]$, $0<b<4$

$Y \mid X=a \sim \text{Uniform}[0,1]$, $0<a<1$

$Y \mid X=a \sim \text{Uniform}[0,2]$, $1<a<2$

$Y \mid X=a \sim \text{Uniform}[0, 2-a]$, $0<a<2$

$$f_{Y \mid X=a}(y) = \begin{cases} \frac{1}{2-a}, & 0<y<2-a \\ 0, & \text{else.} \end{cases}$$

Problem

Consider the joint density

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

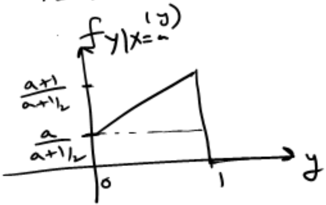
Find the conditionals.

$$f_X(x) = \int_0^1 (x+y)\,dy = \left(xy + \frac{y^2}{2}\right)\Big|_0^1 = x + 1/2, \quad 0 < x < 1$$

$$f_Y(y) = y + 1/2, \quad 0 < y < 1$$

$$f_{Y|X=a}(y) = \frac{a+y}{a+1/2}, \quad 0 < y < 1$$

$0 < a < 1$

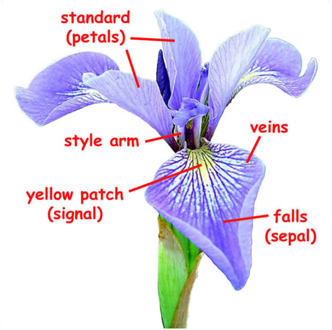


Subsection 6

From data to distribution

Iris data set

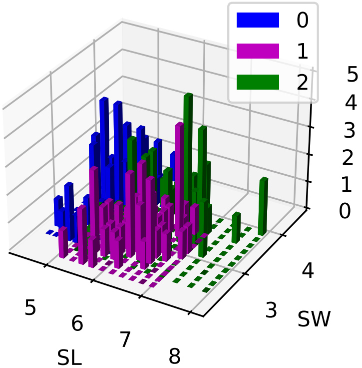
- Iris flower
  - 3 classes of irises: 0, 1 and 2
    - 50 instances in each class
  - Data (cm)
    - sepal length (SL), sepal width (SW), petal length (PL), petal width (PW)
  - Classification
    - Given data, find class

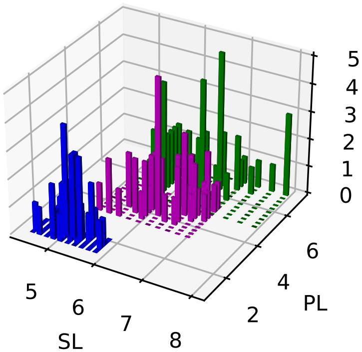


(image source: fs.fed.us)

How to statistically describe (class, SL, SW, PL, PW)?

2D histograms: Class with (SL, SW) and (SL, PL)





- Notice how (SL, PL) seems to almost entirely separate the classes
- Notice the number of points per bin - very low

One discrete and two continuous random variables

How to describe joint distributions involving one discrete and two continuous variables?

One discrete and two continuous random variables

How to describe joint distributions involving one discrete and two continuous variables?

- (X, Y, Z): X discrete and (Y, Z) continuous
- X has range T_X and PMF p_X
- For $x \in T_X$, "conditional" joint density $f_{YZ|X=x}(y, z)$
- Joint density $f_{YZ}(y, z) = \sum_{x \in T_X} p_X(x) f_{YZ|X=x}(y, z)$

Some issues in going from data to distribution

- When doing a histogram, there should be enough data points in each bin
- Typical situation: you will not have enough data to do stable histograms
  - Iris: only 50 instances per class
- In practice, the actual distribution is very difficult to know or estimate. However, it is good to have a sense of the distribution or the support at least.

Diabetes data set: 442 diabetes patients

- Ten baseline variables: age (yrs), sex (F/M), body mass index, average blood pressure, and six blood serum measurements
  - s1 tc, T-Cells (a type of white blood cells)
  - s2 ldl, low-density lipoproteins
  - s3 hdl, high-density lipoproteins
  - s4 tch, thyroid stimulating hormone
  - s5 ltg, lamotrigine
  - s6 glu, blood sugar level
- A quantitative measure of disease progression one year after baseline

Diabetes data set: 442 diabetes patients

- Ten baseline variables: age (yrs), sex (F/M), body mass index, average blood pressure, and six blood serum measurements
  - s1 tc, T-Cells (a type of white blood cells)
  - s2 ldl, low-density lipoproteins
  - s3 hdl, high-density lipoproteins
  - s4 tch, thyroid stimulating hormone
  - s5 ltg, lamotrigine
  - s6 glu, blood sugar level
- A quantitative measure of disease progression one year after baseline

How would you start to describe the "distribution" from above data?

- Data is not sufficient to be confident about joint distribution of several variables
- Usual goal (called regression): predict disease progression as a function of baseline variables
- Typically, some assumptions are made on the distribution to help in prediction
  - It is good to justify these assumptions from the data