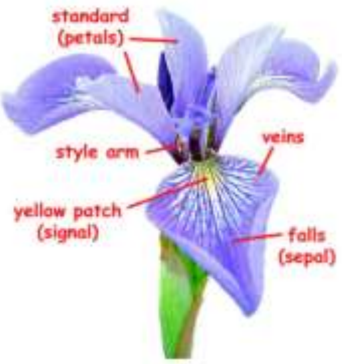
Multiple discrete/continuous random variables

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 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)?

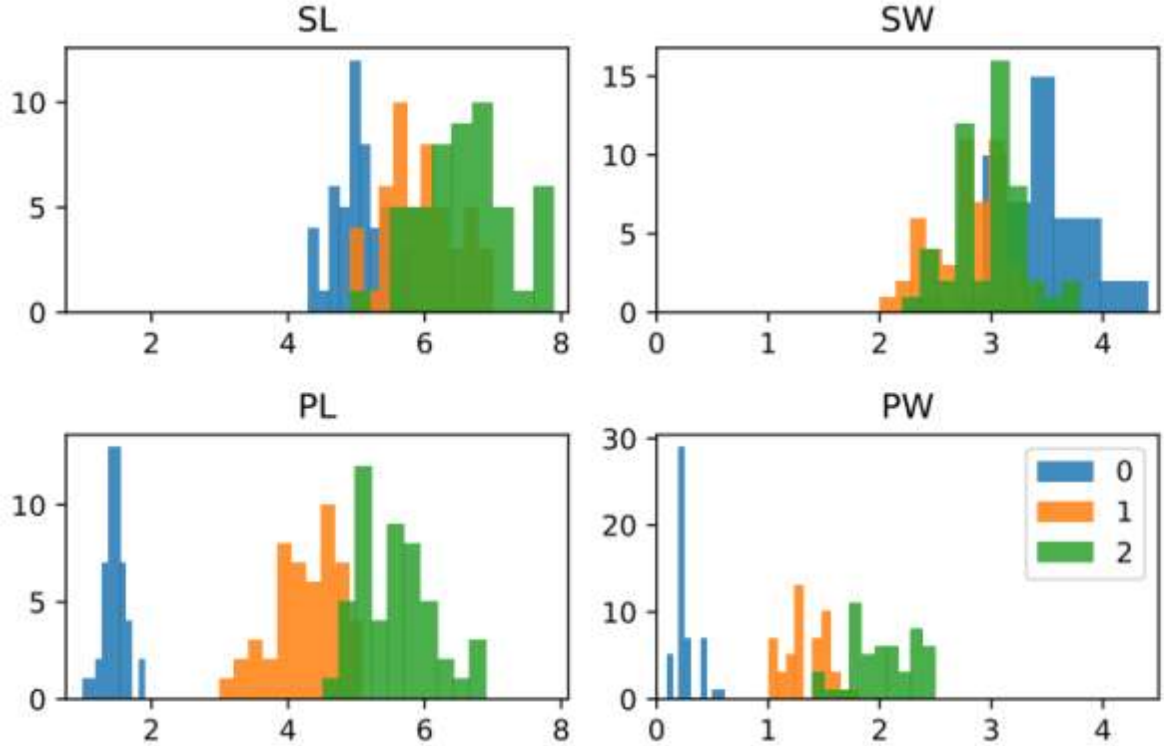
Iris data

Class 0				Class 1				Class 2			
SL	SW	PL	PW	SL	SW	PL	PW	SL	SW	PL	PW
5.1	3.5	1.4	0.2	7.0	3.2	4.7	1.4	6.3	3.3	6.0	2.5
4.9	3.0	1.4	0.2	6.4	3.2	4.5	1.5	5.8	2.7	5.1	1.9
4.7	3.2	1.3	0.2	6.9	3.1	4.9	1.5	7.1	3.0	5.9	2.1
4.6	3.1	1.5	0.2	5.5	2.3	4.0	1.3	6.3	2.9	5.6	1.8
5.0	3.6	1.4	0.2	6.5	2.8	4.6	1.5	6.5	3.0	5.8	2.2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

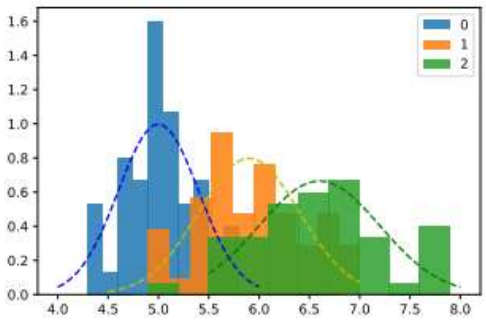
Summary: min-max, avg, stdev

	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

- 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
- 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

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

Problem

discrete uniform

conditional densities

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

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

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.

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

- 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

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, $\overset{\sigma=}{10}$), and that of females is Normal(150, $\overset{\sigma=}{5}$). A random person is found to have a height of 155 cm. Is that person more likely to be male or female?

$-\frac{(y-160)^2}{\dots}$

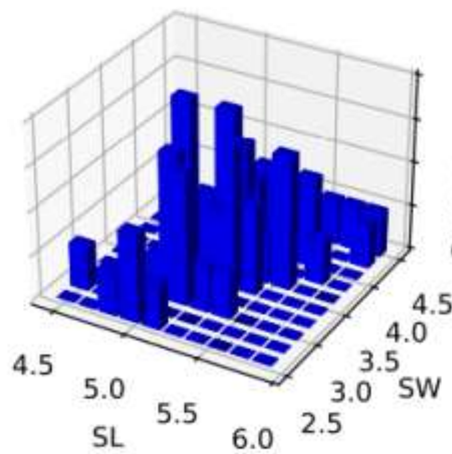
Problem

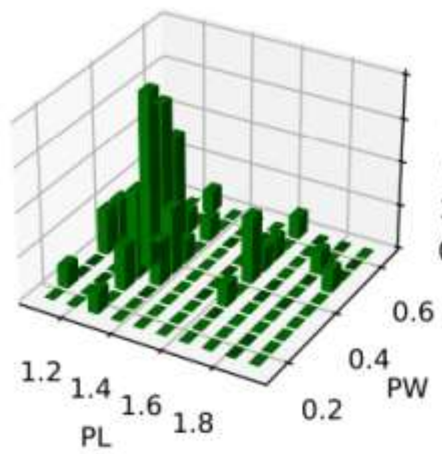
Let $Y - X + Z$, where $X \sim \text{Uniform}\{-3, -1, 1, 3\}$ and $Z \sim \text{Normal}(0, \overset{\text{variance}}{\sigma^2})$ are independent. What is the distribution of Y? Find the distribution of $(X|Y = 0.5)$.

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

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
- $\iint_{-\infty}^{\infty} f(x, y)dxdy = 1$
- 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,

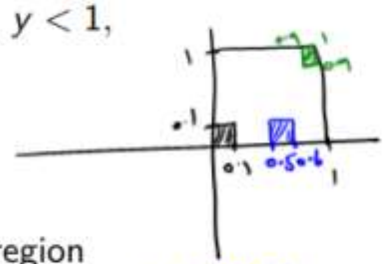
$$P((X, Y) \in A) = \iint_A f(x, y)dxdy$$

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

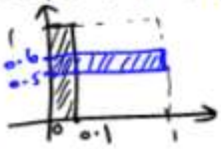
Example: Uniform in the unit square

Let X and Y have joint density

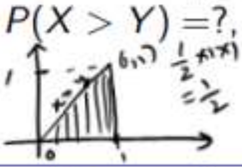
$$f_{XY}(x, y) = \begin{cases} 1 & 0 < x < 1, 0 < y < 1, \\ 0 & \text{otherwise.} \end{cases}$$

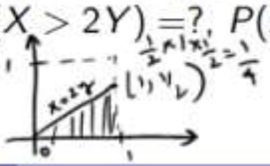


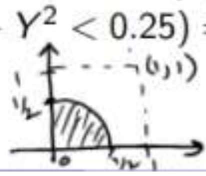
- Picture the 3D plot of the joint density
- To compute probability, find the area of the region
- $P(0 < X < 0.1, 0 < Y < 0.1) \overset{0.01}{=} ?$, $P(0.5 < X < 0.6, 0 < Y < 0.1) = ?$, $0.1 \times 0.1 = 0.01$
  $P(0.9 < X < 1, 0.9 < Y < 1) = ?$ $0.1 \times 0.1 = 0.01$



- $P(0 < X < 0.1) = ?$, $P(0.5 < Y < 0.6) = ?$
  $1 \times 0.1 = 0.1$ $\quad 1 \times 0.1 = 0.1$
- $P(X > Y) = ?$, $P(X > 2Y) = ?$, $P(X^2 + Y^2 < 0.25) = ?$ $\quad \frac{1}{4} \times \pi \left(\frac{1}{2}\right)^2 = \frac{\pi}{16}$







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

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

- 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

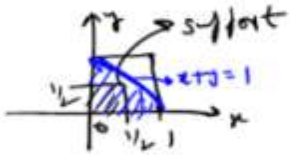
Let $(X, Y) \sim \text{Uniform}(D)$, where $D = \{(x, y) : x + y < 2, x > 0, y > 0\}$.
Sketch the support and compute $P(X + Y < 1)$, $P(X + 2Y > 1)$.

Problem: 2D non-uniform

Let (X, Y) have joint density

$f_{XY} \geqslant 0$

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



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

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 $f_Y(y) = \int_{x=-\infty}^{\infty} f_{XY}(x, y)dx$*

- 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
- $(X, Y) \sim$ Uniform(D), where
  $D = [0, 1/2] \times [0, 1/2] \cup [1/2, 1] \times [1/2, 1]$

1

1/2

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 $D = \{(x, y) : x + y < 2, x > 0, y > 0\}$

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.

- 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 $P(X > Y)$.

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

slice of joint at $x=a$

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

- 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 $D = [0, 1/2] \times [0, 1/2] \cup [1/2, 1] \times [1/2, 1]$

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 $D = \{(x, y) : x + y < 2, x > 0, y > 0\}$

Problem

Consider the joint density

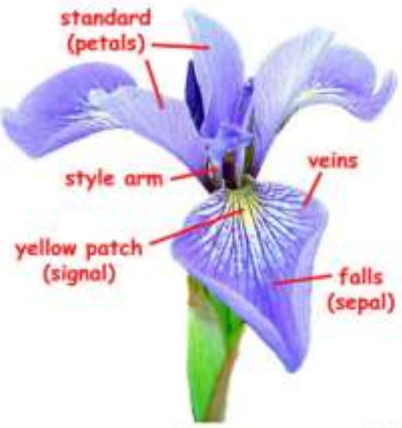
$$f_{XY}(x, y) = \begin{cases} x + y, & 0 < x, y < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Find the conditionals.

Subsection 6

From data to distribution
Iris data set

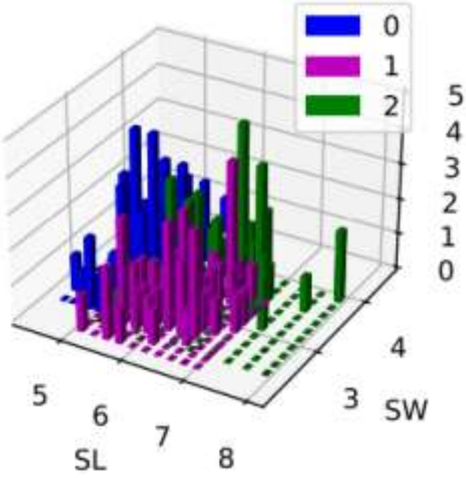
- Iris flower
  - 3 classes of irises: 0, 1 and 2
    - 50 instances in each class
  - Data (cm)
    - sepal length (SL), sepal width (SW), petal length (PL), petal width (PW)
  - Classification
    - Given data, find class

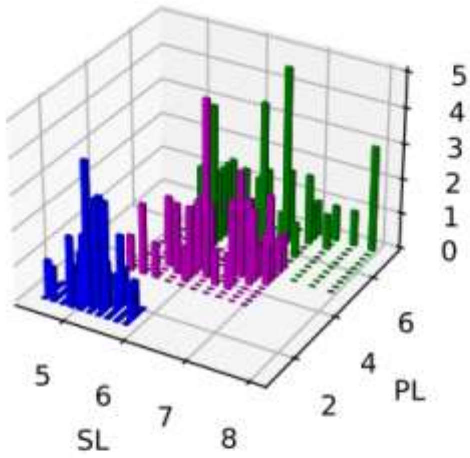


(image source: fs.fed.us)

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?

- (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