Computer Science

CSC380: Principles of Data Science

Probability 4

Xinchen Yu

Recap

- PDF of a transformation of a continuous RV
  - $X + b$ has a PDF that is a translation of X's PDF by b units
  - aX's PDF is X's PDF stretched by a factor of a horizontally
- Mean
  - $\mathrm{E[X]} = \int x\, f(x)\, dx$
  - $\mathrm{E}[r(X)] = \int r(x) f(x)\, dx$
- Variance
  - $\mathrm{Var}(X) = \sigma^2 = \mathrm{E}[(X-\mu)^2] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2$
- Properties
  - $\mathrm{E}[aX] = a\, \mathrm{E}[X]$
  - $\mathrm{Var}(aX) = a^2 \mathrm{Var}(X)$
  - $\mathrm{E}[X + b] = \mathrm{E}[X] + b$
  - $\mathrm{Var}(X + b) = \mathrm{Var}(X)$

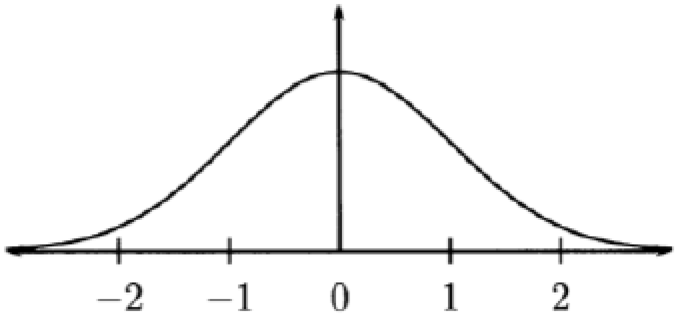
Outline

- Calculating probabilities about Gaussians

- Multivariate Random Variables
    - Joint distribution
    - Marginal distribution

The standard Gaussian distribution

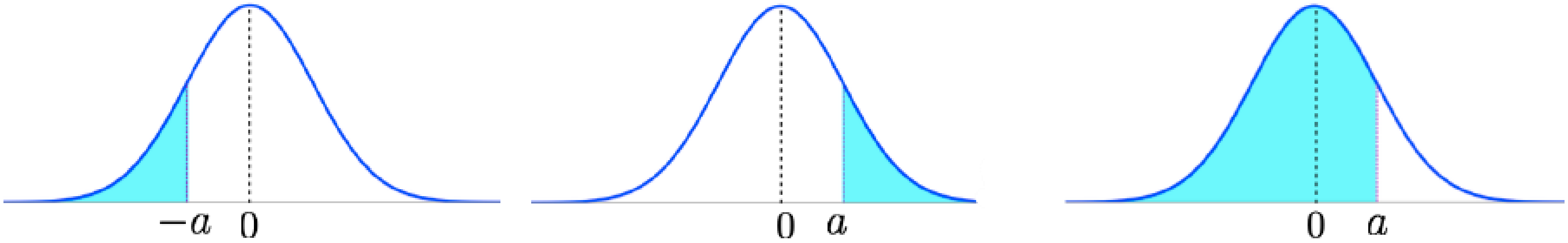
- Gaussian distribution with $\mu = 0$ and $\sigma^2 = 1$



- Denoted by $Z \sim N(0,1)$
- Its PDF denoted by $\phi(z)$, and CDF denoted by $\Phi(z)$

Calculating probabilities about Gaussians

- Symmetry of ϕ => $\Phi(-a) = 1 - \Phi(a)$



$\Phi(-a) = P(Z \leq -a)$ $= P(Z \geq a)$ $= 1 - P(Z \leq a) = 1 - \Phi(a)$

Calculating probabilities about Gaussians

- Suppose $X \sim N(5, 2^2)$, how can I calculate $P(1 < X < 8)$?

- Transform X into another variable:
  - $X \sim N(\mu, \sigma^2): E[X] = \mu, Var[X] = \sigma^2$

- What is mean and variance for the following transformations of X?

$\Rightarrow X - \mu$

$\Rightarrow \frac{X-\mu}{\sigma}$

$$\begin{gathered} \mathrm{E}[aX] = a\,\mathrm{E}[X] \\ \mathrm{Var}(aX) = a^2\mathrm{Var}(X) \\ \mathrm{E}[X + b] = \mathrm{E}[X] + b \\ \mathrm{Var}(X + b) = \mathrm{Var}(X) \end{gathered}$$

Calculating probabilities about Gaussians

- Suppose $X \sim N(5, 2^2)$, how can I calculate $P(1 < X < 8)$?

- Transform X into standard normal Z:
  - $X \sim N(\mu, \sigma^2)$

  $\Rightarrow X - \mu \sim N(0, \sigma^2)$

  $\Rightarrow Z = \frac{X-\mu}{\sigma} \sim N(0,1)$

$$\mathrm{E}[aX] = a\,\mathrm{E}[X]$$
$$\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$$
$$\mathrm{E}[X + b] = \mathrm{E}[X] + b$$
$$\mathrm{Var}(X + b) = \mathrm{Var}(X)$$

- We can write $\mathrm{P}(a < X < b)$ using $\mathrm{P}(c < Z < d)$, which in turn can be written in Φ.

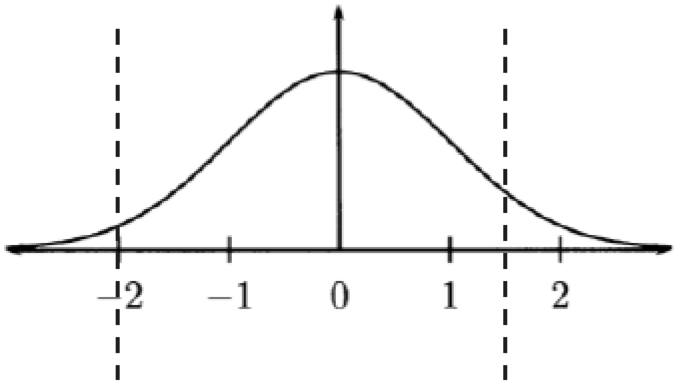
Calculating probabilities about Gaussians

- $\mathrm{P}(a < X < b)$

$$= P\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right)$$

$$= P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right)$$

$$= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$



Example Suppose $X \sim N(5, 2^2)$, calculate $P(1 < X < 8)$

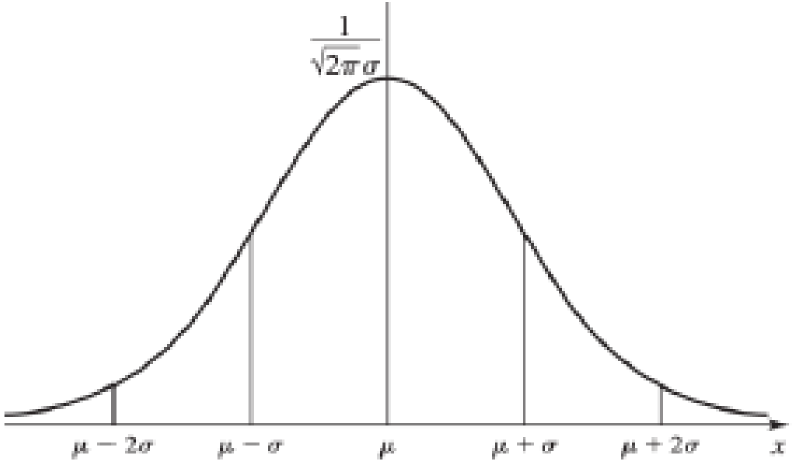
This is $\Phi\left(\frac{8-5}{2}\right) - \Phi\left(\frac{1-5}{2}\right) = \Phi(1.5) - \Phi(-2) = \Phi(1.5) - (1 - \Phi(2))$

```
from scipy.stats import norm
print(norm.cdf(1.5)-(1-norm.cdf(2)))
```

```
0.9104426667829627
```

Calculating probabilities about Gaussians

- What is the probability that a Gaussian RV X is within k ($k = 1, 2, \ldots$) std of its mean?



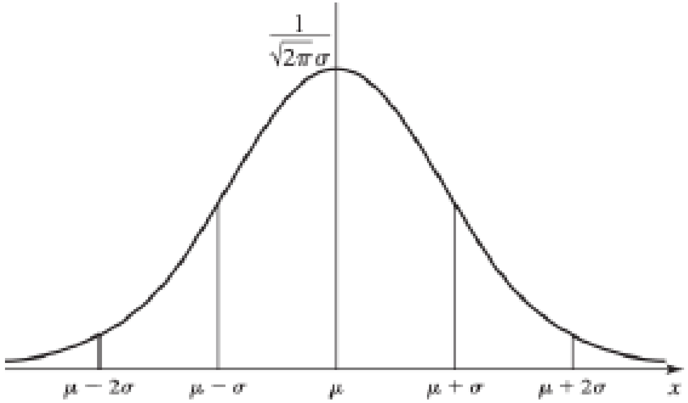
- $\mathrm{P}(\mu - k\sigma \leq X < \mu + k\sigma)$

Calculating probabilities about Gaussians

- $p_k = \mathrm{P}(\mu - k\sigma \le X < \mu + k\sigma)$

$$= P\left(-k < \frac{X-\mu}{\sigma} < k\right)$$
$$= P(-k < Z < k)$$
$$= \Phi(k) - (1 - (\Phi(k))$$
$$= 2\Phi(k) - 1$$

k	p_k
1	0.6826
2	0.9544
3	0.9974
4	0.99994

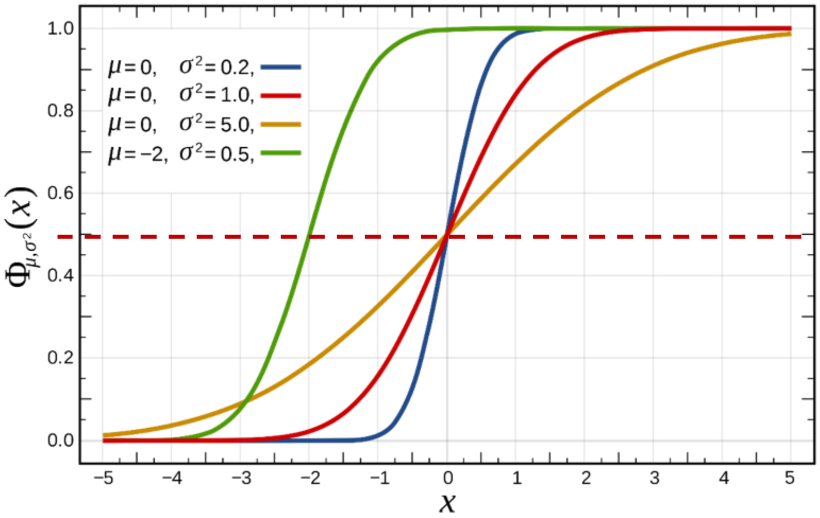


In words,

- With probability about 95%, X is within 2 std of its mean
- With overwhelming prob. (99.7%), X within 3 std of mean

CDF of Gaussian Distributions

- F: CDF of Gaussian $N(\mu, \sigma^2)$



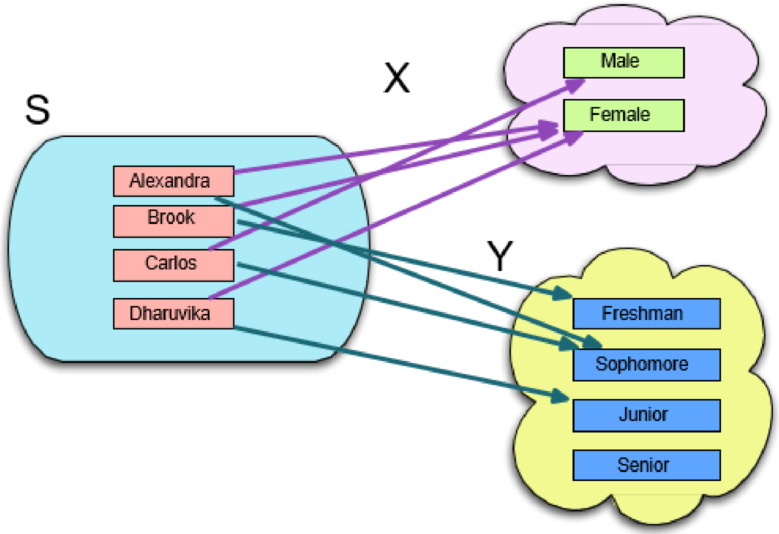
Q: what is $F(\mu)$?

- $F(\mu) = \frac{1}{2}$

- $F(x)$ changes fast near μ
- F's "sensitive range" is about $[\mu - 3\sigma, \mu + 3\sigma]$

Multivariate Random Variables

Multivariate RVs: example



- X: people -> their genders
- Y: people -> their class year
- We'd like to answer questions such as: does X and Y have a correlation?
  - I.e., is a student in higher class year more likely to be male?
- We call (X, Y) a multivariate RV, and will study its *joint* distribution

Joint distribution of discrete RVs

- The joint PMF (probability mass function) of discrete random variables X, Y:

$$f(x, y) = P(X = x, Y = y)$$

Examples

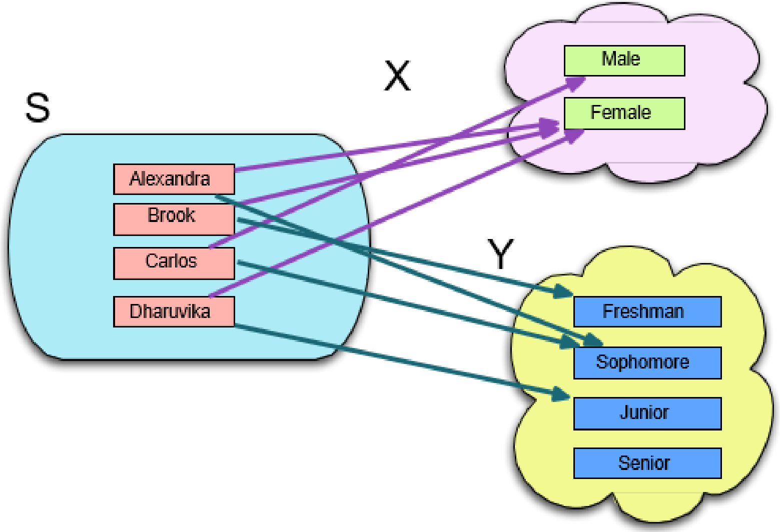
Alexandra

$$P(X = \text{Fem}, Y = \text{Soph}) = \frac{1}{4}$$

Dharuvika

$$P(X = \text{Fem}, Y = \text{Jun}) = \frac{1}{4}$$

…



Joint distribution of discrete RVs

- X: # of cars owned by a randomly selected household
- Y: # of computers owned by the same household
- Joint pmf shown with a table

x \ y	1	2	3	4
1	0.1	0	0.1	0
2	0.3	0	0.1	0.2
3	0	0.2	0	0

- Probability that a randomly selected household has ≥ 2 cars and ≥ 2 computers?
  - $P(X \geq 2, Y \geq 2) = 0.5$

Marginal distributions

Given joint distribution of (X, Y), need distribution of one of them, say X.
Named the ***marginal distribution*** of X.

- How to find $P(X = x)$?
- Using law of total probability:

$$f_1(x) = \sum_y f(x, y)$$

	y			
x	1	2	3	4
1	0.1	0	0.1	0
2	0.3	0	0.1	0.2
3	0	0.2	0	0

- This operation is called *marginalization* ('marginalizing out variable Y', or variable elimination)

Marginal distributions

x	y: 1	2	3	4	Total
1	0.1	0	0.1	0	0.2
2	0.3	0	0.1	0.2	0.6
3	0	0.2	0	0	0.2
Total	0.4	0.2	0.2	0.2	1.0

f_1: *marginal distribution* of X

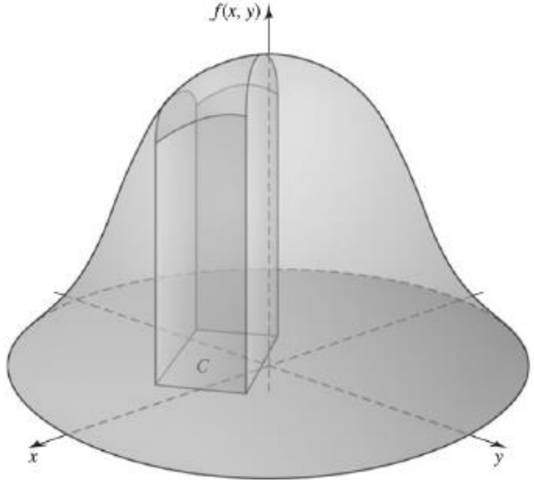
f_2: *marginal distribution* of Y

$$f_1(X=1) = \sum_y f(1,y) = 0.1 + 0 + 0.1 + 0 = 0.2$$

Joint distribution of continuous RVs

- Any continuous random vector (X,Y) has a *joint probability density function* (PDF) $f(x, y)$, such that for all C,

$$P\big((X,Y) \in C\big) = \iint_C f(x,y)\,dx\,dy$$



$f(x, y)$: represent a 2D surface

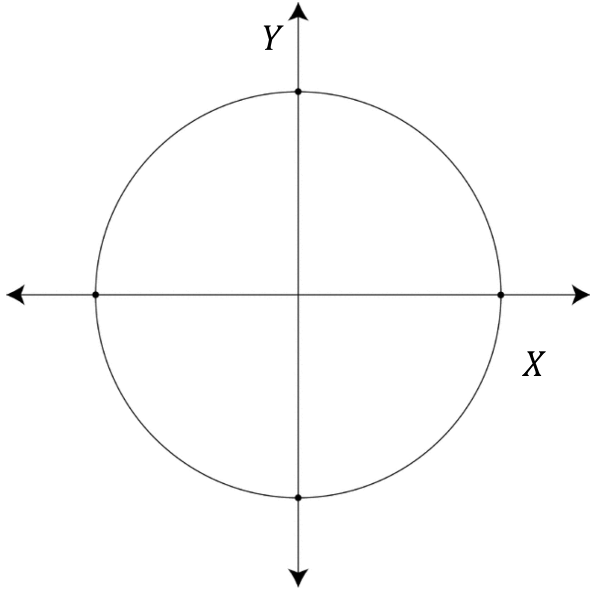
double integral: the *volume* under the surface

Properties:

- f is nonnegative
- $\iint_{R^2} f(x,y)\,dx\,dy = 1$ ($R^2 =$ the whole x-y plane)
  - $P\big((X,Y) \in R^2\big) = 1$

Example: dartboard

- Dartboard with center (0,0) and radius 1; dart lands uniformly at random on the board



- What is the joint PDF of (X, Y)?

- Fact: the PDF is

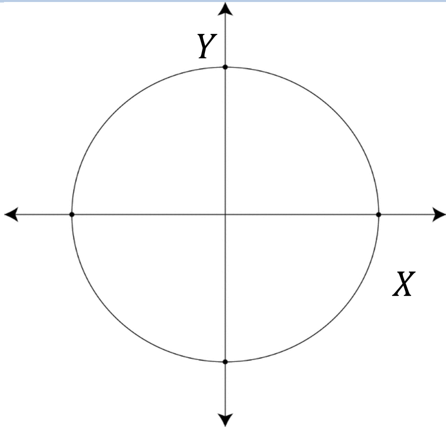
$$f(x,y) = \begin{cases} c, x^2 + y^2 \leq 1 \\ 0, \text{otherwise} \end{cases}$$

- This is called “the Uniform distribution over the unit disk”

Example: dartboard

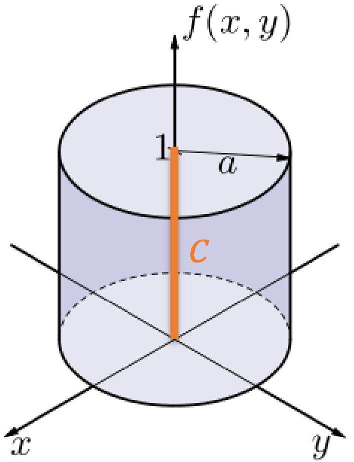
The PDF of X, Y is

$$f(x,y) = \begin{cases} c, x^2 + y^2 \leq 1 \\ 0, \text{otherwise} \end{cases}$$



Can we find c?

Observe: volume under $f(x,y)$ is πc (cylinder) which must also be 1



Therefore, $c = 1/\pi$

Marginal distribution of continuous RV

Given joint distribution of continuous RV (X, Y), how to find X's PDF f_1?

Fact (marginalization) $f_1(x) = \int_R f(x, y)\, dy$

Replacing summation with integration in the continuous case ('marginalizing / integrating out variable Y')

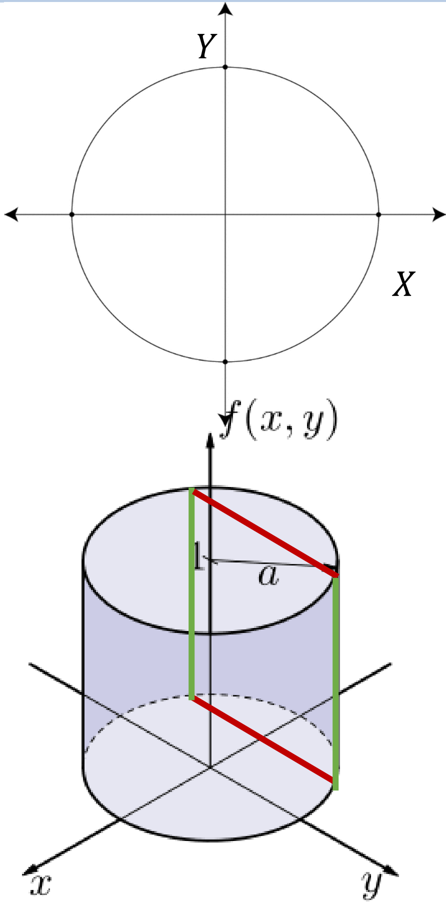
How about Y's PDF f_2?

- Marginalize out X

Example: dartboard

The PDF of X, Y is

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

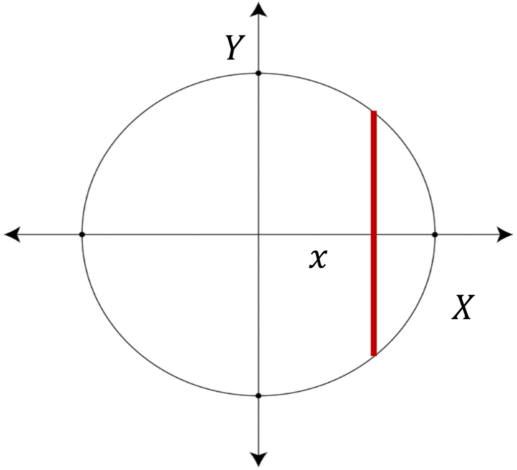


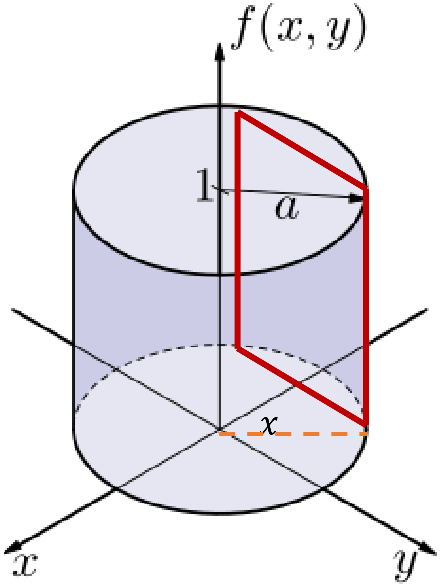
What is the marginal distribution over X?

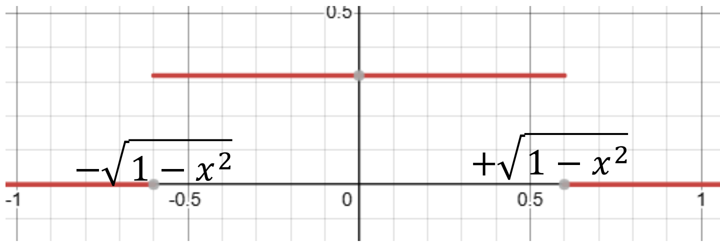
$$f_1(x) = \int_{-\infty}^{+\infty} f(x,y)\, dy$$

How to find this integral?

Example: dartboard



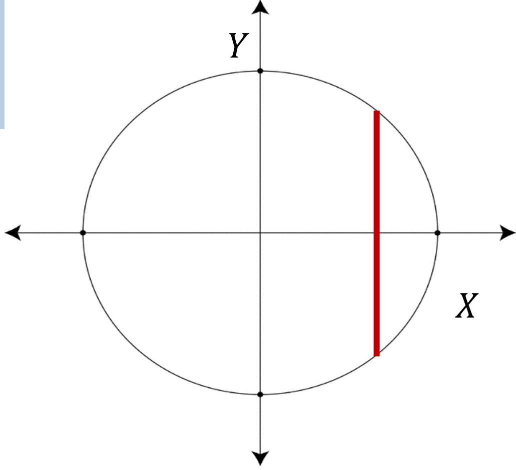




For a fixed $x \in [-1, 1]$, we can think of $f(x)$ is the area of the slice:

- height: $\frac{1}{\pi}$, width: $2 \cdot \sqrt{1-x^2}$
- $f_1(x) = \frac{2}{\pi} \cdot \sqrt{1-x^2}$

Example: dartboard



- In summary,

$$f(x) = \begin{cases} \frac{2}{\pi} \cdot \sqrt{1 - x^2}, x \in [-1,1] \\ 0, \qquad \text{otherwise} \end{cases}$$

X's distribution is NOT $\text{Uniform}([-1,1])$!

Actually makes sense: X closer to 1 is harder to be hit

Recap and Plan

- Multivariate RVs
  - Discrete
    - Marginal: $f_1(x) = \sum_y f(x, y)$
  - Continuous
    - Marginal: $f_1(x) = \int_R f(x, y)\, dy$
- Independence of RVs
- Conditional distribution of RVs

Joint distribution of more than 3 RVs

- We can consider the joint distribution of more than 3 random variables,
  - E.g. (A,B,C), A = gender, B = class year, C = blood type
- Discrete RVs: can still define joint PMFs

a	b	c	$P(A=a, B=b, C=c)$
0	0	0	0.06
0	0	1	0.09
0	1	0	0.08
0	1	1	0.12
1	0	0	0.06
1	0	1	0.24
1	1	0	0.10
1	1	1	0.25

Marginalization

a	b	c	$P(A = a, B = b, C = c)$
0	0	0	0.06
0	0	1	0.09
0	1	0	0.08
0	1	1	0.12
1	0	0	0.06
1	0	1	0.24
1	1	0	0.10
1	1	1	0.25

Given the joint distribution of (A, B, C)

- What is the distribution of A?
  - Need to find $P(A = 0)$ and $P(A = 1)$

$$P(A = 0) = \sum_{b,c} P(A = 0, B = b, C = c)$$

Marginalization: summing over irrelevant variables

- What is the joint distribution of (A, B)?
  - Need to find $P(A = 0, B = 0), \ldots,\ P(A = 1, B = 1)$

$$P(A = 0, B = 0) = \sum_{c} P(A = 0, B = 0, C = c)$$

Marginalization for continuous RVs

Suppose joint PDF of (A, B, C) is $f(a, b, c)$

- What is the PDF of A?

$$f_A(a) = \iint_{R^2} f(a, b, c)\ db\, dc$$

- What is the joint PDF of (A, B)?

$$f_{A,B}(a, b) = \int_R f(a, b, c) dc$$

Marginalization: summing over irrelevant variables

- These operations generalize to joint PDFs of more RVs..

Independence of RVs

Independence of two RVs

- RVs X, Y are independent (denoted by $X \perp\!\!\!\perp Y$) if

$$f(x,y) = f_1(x) \quad \cdot \quad f_2(y), \quad \textit{for all } x, y$$

PMF or PDF Marginal of X Marginal of Y

- E.g. for discrete X, Y,

$$P(X = 3, Y = 4) = P(X = 3) \cdot P(Y = 4)$$

Therefore, $\{X = 3\}$ and $\{Y = 4\}$ are independent events

In class activity: checking independence of RVs

- Which of these PMFs correspond to independent $X \perp\!\!\!\perp Y$?

	$Y=0$	$Y=1$	
X=0	1/4	1/4	1/2
X=1	1/4	1/4	1/2
	1/2	1/2	1

X, Y independent

Need to check:
$f_1(0)f_2(0) = f(0,0)$,
..
(4 equalities)

	$Y=0$	$Y=1$	
X=0	1/2	$\boxed{0}$	1/2
X=1	0	1/2	1/2
	1/2	1/2	1

X, Y not independent

E.g. $f_1(0)f_2(1) = \frac{1}{4}$, whereas $f(0,1) = 0$

only one counterexample suffices to disprove independence!

Independence is invariant under transformations

Fact If X, Y are independent, then $f(X), g(Y)$ are also independent

E.g. X = tomorrow's temperature (in Celsius); Y = tomorrow's NVIDIA stock price (in $)

f(X) = tomorrow's temperature (in Fahrenheit); g(Y) = tomorrow's NVIDIA stock price (in cents)

Independence of more than two RVs

- RVs $X_1, \ldots, X_n$ are independent if their joint PMF or PDF satisfy

$$f(x_1, x_2, \ldots, x_n) = f_1(x_1) f_2(x_2) \ldots f_n(x_n),$$

PMFs or PDFs Marginal for X_1 Marginal for X_n

for all $x_1, \ldots, x_n$

This captures many real-world applications:

- Independent trials: each X_i is $\mathrm{Bernoulli}(p)$
  - Flip 10 coins: $x_1, x_2, \ldots, x_{10}$

True or False?

- If I flip 10 coins independently, it is more likely that I see
  HTTTHTHHTHT
  than
  HHHHHHHHHH

- False

$$f(\text{HTTTHTHHTHT}) = f_1(H) \cdot \ldots \cdot f_{10}(T) = \frac{1}{2^{10}}$$

$$f(\text{HHHHHHHHHH}) = f_1(H) \cdot \ldots f_{10}(H) = \frac{1}{2^{10}}$$

Independence of more than two RVs

Fact If $X_1, \dots, X_n$ are independent, then

- any subset $X_{i_1}, \dots, X_{i_p}$ are independent
  - E.g. X_1, X_3, X_7 are independent

- any disjoint subset $(X_{i_1}, \dots, X_{i_m})$, $(X_{j_1}, \dots, X_{j_l})$ are independent, e.g.,
  - (X_1, X_2) is independent of X_3
  - (X_1, X_3) is independent of (X_2, X_4)

Conditional distributions of RVs

Conditional distributions (discrete)

- X, Y have joint PMF f. Y has marginal PMF f_2

- Conditional PMF of X given $Y = y$:

$$g_1(x|y) = \frac{f(x,y)}{f_2(y)}$$

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

- $g_1(x|y)$ is viewed as a function of x: “the conditional distribution of X given $Y = y$”

In-class activity (discrete case)

Example X=0: car not stolen, X=1: car stolen, which brand is the safest?

Joint PMF of X, Y, find $\mathrm{P}(\mathrm{X}=0|\mathrm{Y}=1)$

	Brand Y					
Stolen X	1	2	3	4	5	Total
0	0.129	0.298	0.161	0.280	0.108	0.976
1	0.010	0.010	0.001	0.002	0.001	0.024
Total	0.139	0.308	0.162	0.282	0.109	1.000

Solution

$$\mathrm{P}(\mathrm{X}=0|\mathrm{Y}=1)=\frac{\mathrm{P}(\mathrm{X}=0,\mathrm{Y}=1)}{\mathrm{P}(\mathrm{Y}=1)}=\frac{0.129}{0.139}=0.928$$

In-class activity (discrete case)

Example X=0: car not stolen, X=1: car stolen

Joint PMF of X, Y:

	Brand Y					
Stolen X	1	2	3	4	5	Total
0	0.129	0.298	0.161	0.280	0.108	0.976
1	0.010	0.010	0.001	0.002	0.001	0.024
Total	0.139	0.308	0.162	0.282	0.109	1.000

Find the table of the conditional PMF of X given Y

Solution

	Brand Y				
Stolen X	1	2	3	4	5
0	0.928	0.968	0.994	0.993	0.991
1	0.072	0.032	0.006	0.007	0.009

Brand 3 is the safest

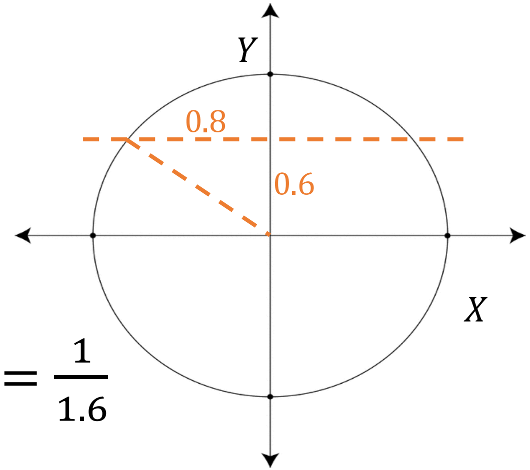
Conditional distributions (continuous)

- X, Y have joint PDF f. Y has marginal PDF f_2

- Conditional PDF of X given Y:

$$g_1(x|y) = \frac{f(x,y)}{f_2(y)}$$

Example Conditional distribution of X given $Y = 0.6$:

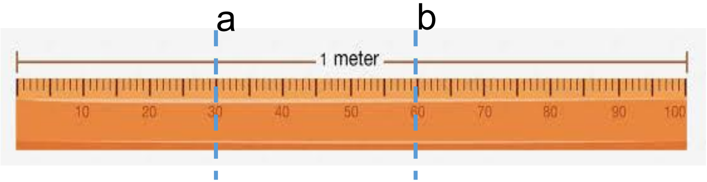


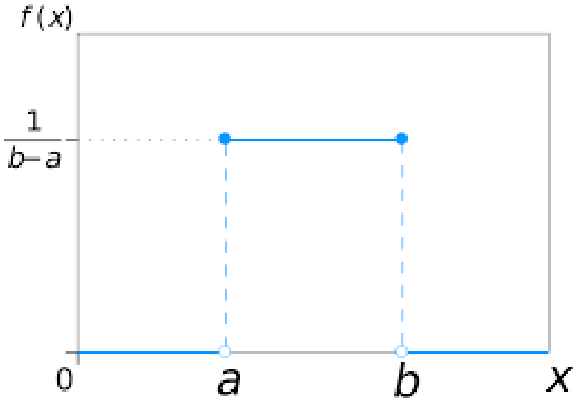
Answer: Uniform($[-0.8, +0.8]$), $g(x|Y = 0.6) = \frac{1}{0.8+0.8} = \frac{1}{1.6}$

Recap: Uniform Distribution

- $X \sim \text{Uniform}([a, b])$

$$f(x) = \begin{cases} 0, & y < a \\ \frac{1}{b-a}, & y \in [a, b] \\ 0, & y > b \end{cases}$$





Conditional distributions & independence

Fact X,Y are independent

$\Leftrightarrow$ for all y, $g(x|y)$ are all equal to $f(x)$

Here, g, f are PMF or PDF depending on the types of X,Y

Assume Y can only take the value 1, 2, and 3. We say X,Y are independent when

- $f(X = x) = g(X = x|Y = 1)$, and
- $f(X = x) = g(X = x|Y = 2)$, and
- $f(X = x) = g(X = x|Y = 3)$

X: ice cream sales
Y: weather (sunny, cloudy, rainy)

Not independent

In other words, knowing Y does not change our belief on X

In-class activity

Joint PMF

Stolen X	Brand Y 1	2	3	4	5	Total
0	0.129	0.298	0.161	0.280	0.108	0.976
1	0.010	0.010	0.001	0.002	0.001	0.024

$f(x)$

conditional PMF of X, Y

Stolen X	Brand Y 1	2	3	4	5
0	0.928	0.968	0.994	0.993	0.991
1	0.072	0.032	0.006	0.007	0.009

$g(x|1)$ $g(x|2)$

Question: are X,Y independent?

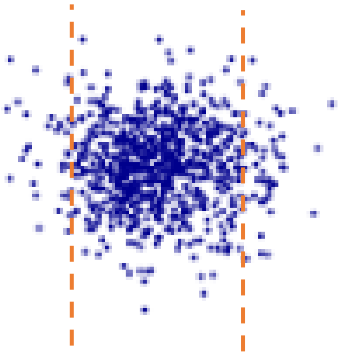
$$g(x = 0|1) = 0.928$$
$$f(x = 0) = 0.976$$

Not equal, so not independent

Independence: visualization

- Left: X, Y independent; Right: X, Y not independent

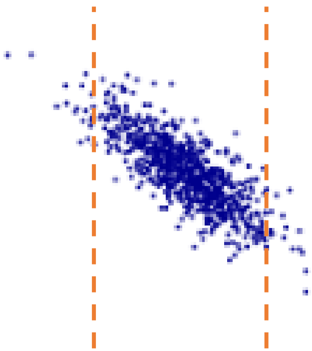
Y



X

$g_1(y|x=-1)$ $g_1(y|x=+1)$

Y



X

$g_1(y|x=-1)$ $g_1(y|x=+1)$

True or False?

- If I flip a fair coin repeatedly, and my first 2 trials are both tails. Then my third throw will have a higher chance of showing head.

- This is asking $g_3(\mathrm{H} \mid \mathrm{TT}) = P(X_3 = H | X_1 = T, X_2 = T)$

  Since X_3 is independent of X_1, X_2 $\quad = P(X_3 = H) = 1/2$

  so the claim is false

- This is known as the *gambler's fallacy*
  - Prior losses do not increase the chance of future win

Conditional expectation

Definition The mean of the conditional distribution of X given $Y = y$, is called the *conditional expectation* of X given $Y = y$, denoted as $\mathrm{E}[X \mid Y = y]$.

$\mathrm{E}[X \mid Y = y]$ can be found by:

- $\sum_x x \cdot g(x|y)$, if X is discrete

  Conditional PMF

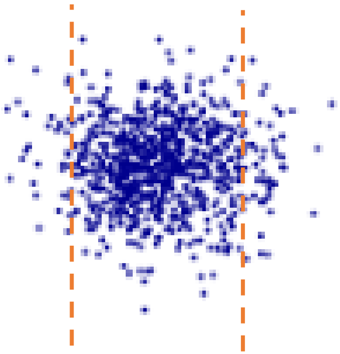
- $\int_{-\infty}^{+\infty} x \cdot g(x|y)\, dx$, if X is continuous

  Conditional PDF

Independence: visualization

- Left: X, Y independent; Right: X, Y not independent

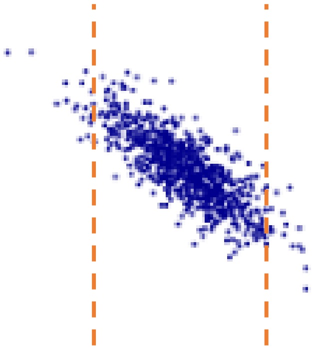
Y



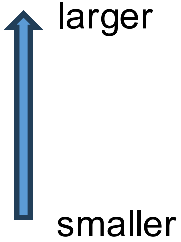
X

$g_1(y|x=-1)$ $g_1(y|x=+1)$

Y



X



$g_1(y|x=-1)$ $g_1(y|x=+1)$

Which one is larger, $\mathrm{E}[Y|X=-1]$ or $\mathrm{E}[Y|X=+1]$?
The former