STAT 1151
Introduction to Probability

Lecture 4 Joint Probability Distributions

Xiaomei Tan

Announcement

1. Midterm:

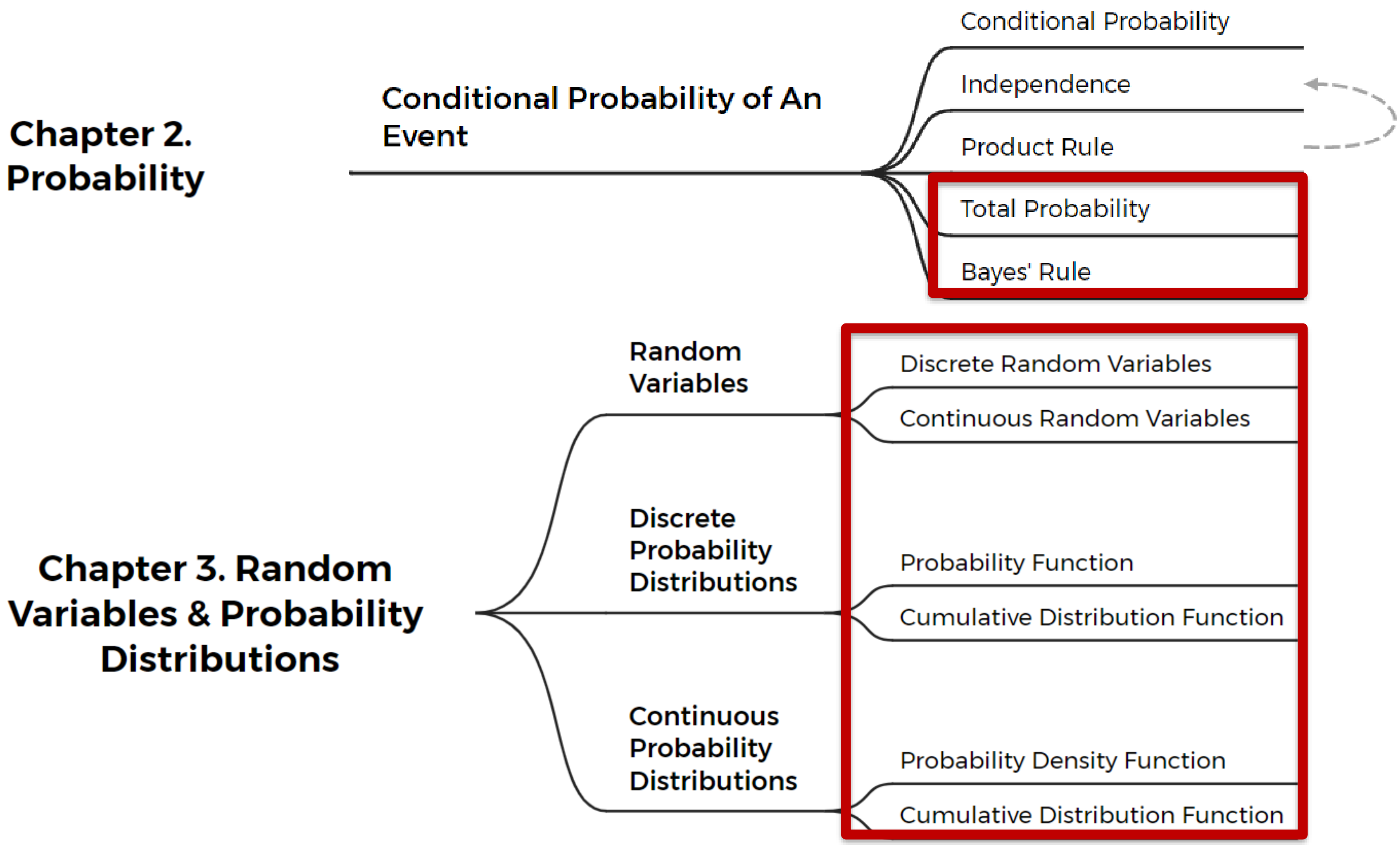
- Time: 18:30-20:30 November 3rd (Friday Week 9)
- Location: 综C 204

2. Office hour:

- No office hour this week
- Time: 13:30-16:30 October 7th (Saturday)

  9:00-12:00 October 8th (Sunday)
- Location: 3-317B

Last Lecture



Outline

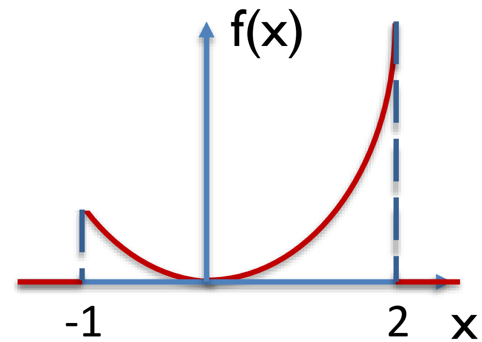
- Chapter 3 Random Variables and Probability Distributions
  - Continuous Probability Distributions
  - Joint Probability Distributions
    - Two discrete random variables
    - Two continuous random variables
    - More than two random variables

Probability Density Function

Example:

Suppose that the error in the reaction temperature (◦C) for a controlled laboratory experiment is a continuous random variable X having the probability density function

$$f(x) = \begin{cases} \dfrac{x^2}{3}, & -1 < x < 2 \\ 0, & elsewhere \end{cases}$$



(a) Verify that f(x) is a density function.

1. $f(x) \geq 0, for\ all\ x \in R$

2. $\int_{-\infty}^{\infty} f(x)dx = 1$

$$\int_{-\infty}^{\infty} \frac{x^2}{3} dx = \frac{x^3}{9}\bigg|_{-1}^{2} = \frac{8}{9} + \frac{1}{9} = 1$$

3. $P(a < X < b) = \int_a^b f(x)dx$

(b) Find *P*(0 < *X* ≤ 1).

$$P(0 < X \leq 1) = \int_0^1 \frac{x^2}{3} dx = \frac{x^3}{9}\bigg|_0^1 = \frac{1}{9}$$

(c) Find *P*(*X* ≤ 1).

? Cumulative distribution function

Cumulative Distribution Function

Definition:
The **cumulative distribution function F(x)** of a continuous random variable X with density function f(x) is

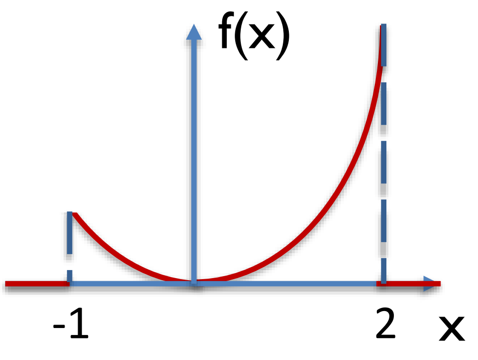
$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt, \qquad for\ -\infty < x < \infty$$

where $f(x) = \frac{dF(x)}{dx}$, if the derivative exists.

Example in last page:

- For -1 ≤ x < 2, $F(x) = \int_{-\infty}^{x} f(t)dt = \int_{-1}^{x} \frac{t^2}{3}dt = \frac{x^3+1}{9}$
- For x < -1, $F(x) = 0$
- For x ≥ 2, $F(x) = 1$

So $\mathrm{F(x)} = \Bigg\{$



Cumulative Distribution Function

Example:

The Department of Energy (DOE) puts projects out on bid and generally estimates what a reasonable bid should be. Call the estimate b. The DOE has determined that the density function of the winning (low) bid is

$$f(x) = \begin{cases} \frac{5}{8b}, & \frac{2b}{5} \le x \le 2b \\ 0, & elsewhere \end{cases}$$

Find F(x).

- For 2b/5 ≤ x ≤ 2b, $F(x) = \int_{-\infty}^{x} \frac{5}{8b} dt = \int_{2b/5}^{x} \frac{5}{8b} dt = \frac{5t}{8b}\Big|_{2b/5}^{x} = \frac{5x}{8b} - \frac{1}{4}$
- For x < 2b/5, $F(x) = 0$
- For x > 2b, $F(x) = 1$

$$F(x) = \begin{cases} 0, & x < \frac{2b}{5} \\ \frac{5x}{8b} - \frac{1}{4}, & \frac{2b}{5} \le x \le 2b \\ 1, & x > 2b \end{cases}$$

Using F(x) to Compute Probabilities

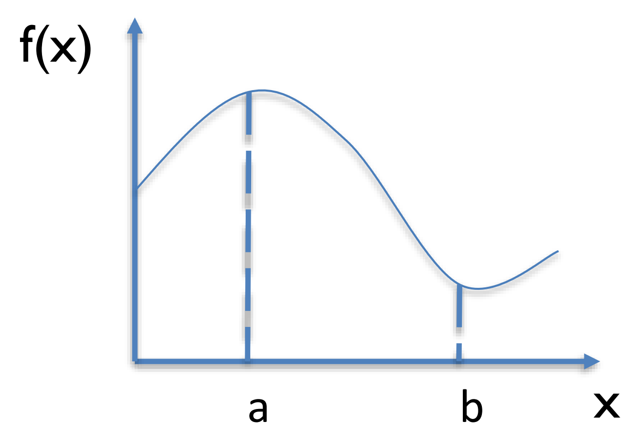
Proposition:
Let X be a continuous random variable with probability density function f(x) and cumulative distribution function F(x).

- For any number a,

$$P(X > a) = 1 - F(a)$$

- For any two numbers a and b with a < b,

$$P(a \leq X \leq b) = F(b) - F(a)$$



Using F(x) to Compute Probabilities

Example:
Suppose the probability density function of the magnitude X of a dynamic load on a bridge (in newtons) is given by

$$f(x) = \begin{cases} \frac{1}{8} + \frac{3}{8}x, & 0 \le x \le 2 \\ 0, & elsewhere \end{cases}$$

Find $P(X > 1)$ and $P(1 \le X \le 1.5)$.

- For 0 ≤ x ≤ 2, $F(x) = \int_{-\infty}^{x} (\frac{1}{8} + \frac{3}{8}t)dt = \frac{t}{8} + \frac{3}{16}t^2 \Big|_0^x = \frac{x}{8} + \frac{3}{16}x^2$
- For x < 0, $F(x) = 0$
- For x > 2, $F(x) = 1$

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{x}{8} + \frac{3}{16}x^2, & 0 \le x \le 2 \\ 1, & x > 2 \end{cases}$$

Joint Probability Distributions

Many problems in probability involve several random variables simultaneously.

Ex. - #successful free throws (F) and #successful three-point shots (T) of a basket player

- weight (W) and height (H)

- To determine the likelihood of success in college based on a student's high school data: college entrance exam score (E), high school class rank (R), and grade-point average at the end of freshman year (F)

Consider joint probability distributions for

- Two discrete random variables
- Two continuous random variables
- More than two random variables

Two Discrete Random Variables

Definition:
Consider two discrete random variables, X and Y.
The function $f(x, y)$ is a **joint probability distribution** if the following conditions hold:

1. $f(x, y) \geq 0 \; for \; all \; (x, y)$
2. $\sum_x \sum_y f(x, y) = 1$
3. $P(X = x, Y = y) = f(x, y)$

For any region A in the xy plane, $P[(X, Y) \in A] = \sum \sum_A f(x, y)$

Two Discrete Random Variables

Example:

Two ballpoint pens are selected at random from a box that contains 3 blue pens, 2 red pens, and 3 green pens. If X is the number of **blue** pens selected and Y is the number of **red** pens selected.

 $\times 3$  $\times 2$  $\times 3$

Possible (x, y):

(0, 0), (0, 1), (0, 2), (1, 0), (1, 1), and (2, 0)

Joint probability table

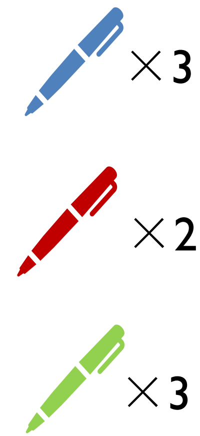
		y		
	f(x, y)	0	1	2
x	0	$\binom{3}{2}/\binom{8}{2}=\frac{3}{28}$	$\binom{2}{1}\binom{3}{1}/\binom{8}{2}$	$\binom{2}{2}/\binom{8}{2}$
	1	$\binom{3}{1}\binom{3}{1}/\binom{8}{2}$	$\binom{3}{1}\binom{2}{1}/\binom{8}{2}$	0
	2	$\binom{3}{2}/\binom{8}{2}$	0	0

Two Discrete Random Variables

y

x

f(x, y)	0	1	2
0	3/28	3/14	1/28
1	9/28	3/14	0
2	3/28	0	0

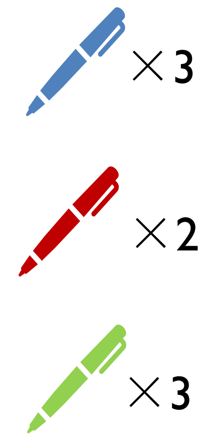


(a) Find the joint probability function $f(x, y)$.

(b) Find $P[(X, Y) \in A]$, where A is the region $\{(x, y) | x + y \leq 1\}$.

Two Discrete Random Variables

		y			
	f(x, y)	0	1	2	Row Totals
x	0	3/28	3/14	1/28	5/14
	1	9/28	3/14	0	15/28
	2	3/28	0	0	3/28
					1



(c) Find the probability that only one blue pen was selected.

The probability distribution **g(x)** of X alone results from holding x fixed and <u>summing</u> the joint probability distributions f(x, y) over the values of Y.

Marginal Distributions of X

Two Discrete Random Variables

Definition:

For discrete random variables X and Y,

- the **Marginal Distributions** of X alone is $g(x) = \sum_y f(x, y)$
- the **Marginal Distributions** of Y alone is $h(y) = \sum_x f(x, y)$

$g(0) = f(0,0) + f(0,1) + f(0,2) = 5/14$

$h(0) = f(0,0) + f(1,0) + f(2,0) = 15/28$

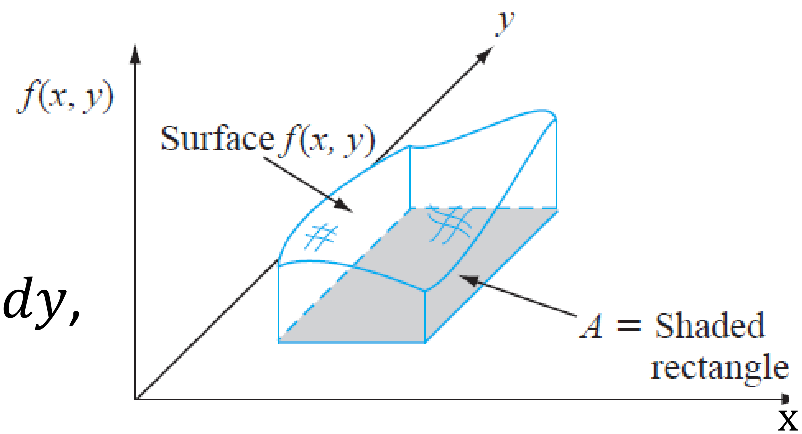
	f(x, y)	y = 0	y = 1	y = 2	Row Totals	
x	0	3/28	3/14	1/28	5/14	$g(0)$
x	1	9/28	3/14	0	15/28	$g(1)$
x	2	3/28	0	0	3/28	$g(2)$
	Column Totals	15/28	3/7	1/28	1	
		$h(0)$	$h(1)$	$h(2)$		

Two Continuous Random Variables

Definition:
Consider two continuous random variables, X and Y.
The function $f(x,y)$ is a **joint (probability) density function** if the following conditions hold:

1. $f(x,y) \geq 0 \; for \; all \; (x,y)$,
2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\, dxdy = 1$,
3. $P[(X,Y) \in A] = \int\int_A \; f(x,y)dxdy$,

for any region A in the *xy* plane.



In particular, if A is the two-dimensional rectangle {(x, y): a ≤ x ≤ b, c ≤ y ≤ d}, then

$$P[(X,Y) \in A] = P(a \leq X \leq b, c \leq Y \leq d) = \int_c^d \int_a^b f(x,y)dxdy$$

Two Continuous Random Variables

Example:
A privately owned business operates both a drive-in facility and a walk-in facility. On a randomly selected day, let X and Y, respectively, be the proportions of the time that the drive-in and the walk-in facilities are in use, and suppose that the joint density function of these random variables is

$$f(x,y) = \begin{cases} \frac{2}{5}(2x+3y), & 0 \le x \le 1, 0 \le y \le 1 \\ 0, & elsewhere \end{cases}$$

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

Two Continuous Random Variables

Example:
A privately owned business operates both a drive-in facility and a walk-in facility. On a randomly selected day, let X and Y, respectively, be the proportions of the time that the drive-in and the walk-in facilities are in use, and suppose that the joint density function of these random variables is

$$f(x,y) = \begin{cases} \frac{2}{5}(2x+3y), & 0 \le x \le 1, 0 \le y \le 1 \\ 0, & elsewhere \end{cases}$$

(b) Find $P[(X,Y) \in A]$, where $A = \{(x, y) | 0<x<0.5, 0.25<y<0.5\}$

Two Continuous Random Variables

Definition:

For continuous random variables X and Y,

- the **Marginal Distributions** of X alone is $g(x) = \int_{-\infty}^{\infty} f(x, y) dy$
- the **Marginal Distributions** of Y alone is $h(y) = \int_{-\infty}^{\infty} f(x, y) dx$

The marginal probability density function g(x) of X alone results from holding x fixed and <u>integrating</u> the joint density function f(x, y) over y.

Two Continuous Random Variables

Revisit example:
A privately owned business operates both a drive-in facility and a walk-in facility. On a randomly selected day, let X and Y, respectively, be the proportions of the time that the drive-in and the walk-in facilities are in use, and suppose that the joint density function of these random variables is

$$f(x,y) = \begin{cases} \frac{2}{5}(2x+3y), & 0 \le x \le 1, 0 \le y \le 1 \\ 0, & elsewhere \end{cases}$$

(c) Find g(x) and h(y).

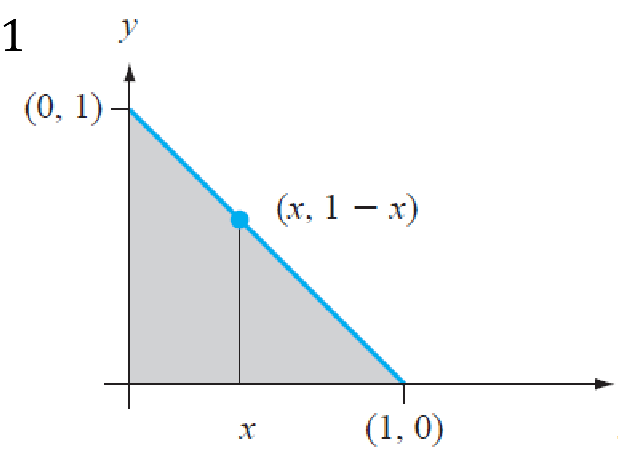
$$g(x) = \int_{-\infty}^{\infty} f(x,y)dy = \Big\{$$

$$h(y) = \int_{-\infty}^{\infty} f(x,y)dx = \Big\{$$

Two Continuous Random Variables

Consider a more complicated case that the region of positive density is not a rectangle:
A company sells fruit salad consisting of three types of fruits: apples, oranges, and grapes. Suppose the net weight of each salad is exactly 1 lb, but the weight contribution of each type of fruit is random. Because the three weights sum to 1, a joint probability model for any two gives all necessary information about the weight of the third type. Let X = the weight of apples in a selected salad and Y = the weight of oranges. Then the region of positive density is D = {(x, y): 0 ≤ x ≤ 1, 0 ≤ y ≤ 1, x + y ≤ 1}, the shaded region pictured in Figure. The joint density function is

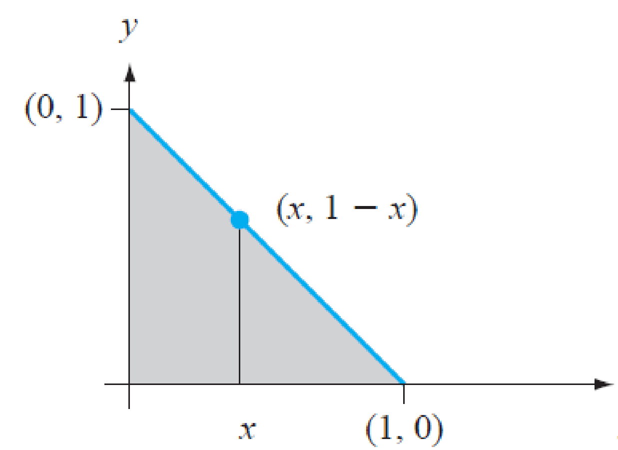
$$f(x,y) = \begin{cases} 24xy, & 0 \le x \le 1, 0 \le y \le 1, x+y \le 1 \\ 0, & elsewhere \end{cases}$$



Two Continuous Random Variables

$$f(x,y) = \begin{cases} 24xy, & 0 \le x \le 1, 0 \le y \le 1, x + y \le 1 \\ 0, & elsewhere \end{cases}$$

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



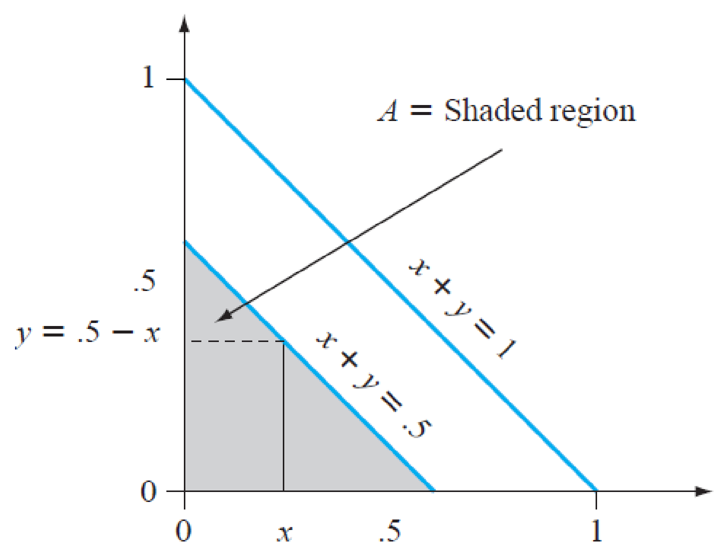
Two Continuous Random Variables

$$f(x,y) = \begin{cases} 24xy, & 0 \le x \le 1, 0 \le y \le 1, x + y \le 1 \\ 0, & elsewhere \end{cases}$$

(b) Find the probability that apples and oranges make up at most 50% of the salad.

Let $A = \{(x, y): 0 \le x \le 1, 0 \le y \le 1, \text{and } x + y \le 0.5\}$

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



Conditional Probability Distribution

Conditional probability of event B, given event A:

$$P(B|A) = \frac{P(B \cap A)}{P(A)}, provided\ P(A) > 0$$

If A and B are defined by discrete/continuous random variables X=x and Y=y, respectively:

$$P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)} = \boxed{\frac{f(x,y)}{g(x)}}, provided\ g(x) > 0$$

A function of y with x fixed (discrete rvs)
or joint density divided by marginal distribution (continuous rvs)

Conditional Probability Distribution

Definition:

Let X and Y be two random variables, discrete or continuous. The **conditional distribution of Y** given that X=x is

$$f(y|x) = \frac{f(x,y)}{g(x)}, provided\ g(x) > 0$$

Similarly, the **conditional distribution of X** given that Y=y is

$$f(x|y) = \frac{f(x,y)}{h(y)}, provided\ h(y) > 0$$

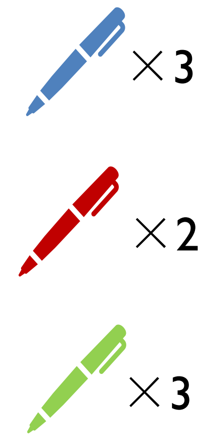
The probability that *X* falls between *a* and *b* when it is known that the *Y* = *y* is

$$P(a < X < b|Y = y) = \sum_{a<x<b} f(x|y), \qquad for\ discrete\ X\ and\ Y$$

$$P(a < X < b|Y = y) = \int_a^b f(x|y)dx, \qquad for\ continuous\ X\ and\ Y$$

Conditional Probability Distribution

Revisit example:
Two pens are randomly selected. X is the number of **blue** pens selected and Y is the number of **red** pens selected.



(a) Find the conditional distribution of X, given that Y = 1.

(b) Find P(X=0|Y=1).

		y				
	f(x, y)	0	1	2	Row Totals	
x	0	3/28	3/14	1/28	5/14	$g(0)$
	1	9/28	3/14	0	15/28	$g(1)$
	2	3/28	0	0	3/28	$g(2)$
	Column Totals	15/28	3/7	1/28	1	
		$h(0)$	$h(1)$	$h(2)$		

Conditional Probability Distribution

Example:
Given the joint density function

$$f(x,y) = \begin{cases} \dfrac{x(1+3y^2)}{4}, 0 < x < 2, 0 < y < 1 \\ 0, \qquad elsewhere \end{cases}$$

(a) Find g(x) and h(y).

(b) Find f(x|y).

(c) Find P(0.25 < X < 0.5 | Y = 1/3).

Independent Random Variables

If $f(x|y)$ does not depend on y, the outcome of Y has no impact on the outcome of X. We say X and Y are independent random variables.

Definition:
Let X and Y be two random variables, discrete or continuous, with joint probability distribution f(x, y) and marginal distributions g(x) and h(y), respectively. The random variables X and Y are said to be **statistically independent** if and only if

$$f(x, y) = \mathrm{g(x)h(y)} \iff f(x|y) = \mathrm{g(x)}$$

for all (x, y) within their range.

$$\iff f(y|x) = h(y)$$

Independent Random Variables

Checking for statistical independence of discrete rvs requires more attention. $f(x, y) = \mathrm{g}(\mathrm{x})\mathrm{h}(\mathrm{y})$ may apply to some but not all (x, y)!!

Example:

$f(1,1) = \mathrm{g}(1)\mathrm{h}(1)$

For other(x,y),

e.g., $f(0,0) \neq \mathrm{g}(0)\mathrm{h}(0)$

y

	f(x, y)	0	1	2	Row Totals	
x	0	1/4	0	0	1/4	$g(0)$
	1	0	1/4	1/4	1/2	$g(1)$
	2	0	1/4	0	1/4	$g(2)$
	Column Totals	1/4	1/2	1/4	1	
		$h(0)$	$h(1)$	$h(2)$		

More Than Two Random Variables

Let $X_1, X_2, \ldots, X_n$ be random variables.

- If $X_1, X_2, \ldots, X_n$ are all discrete random variables, the **joint probability function** of the rvs is

$$f(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

- If $X_1, X_2, \ldots, X_n$ are continuous random variables, the **joint density function** is

$$f(x_1, x_2, \ldots, x_n) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n) dx_1 \ldots dx_n$$

More Than Two Random Variables

- The **marginal distribution** of rv X_1, for example, is

$$g(x_1) = \begin{cases} \sum_{x_2} \dots \sum_{x_n} f(x_1, x_2, \dots, x_n), & for\ discrete\ case \\ \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_n) dx_2 \dots dx_n, & for\ continuous\ case \end{cases}$$

- The **joint marginal distribution** of rvs X_1 and X_2, for example, is

$$g(x_1, x_2) = \begin{cases} \sum_{x_3} \dots \sum_{x_n} f(x_1, x_2, \dots, x_n), & for\ discrete\ case \\ \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_n) dx_3 \dots dx_n, & for\ continuous\ case \end{cases}$$

More Than Two Random Variables

- The **joint conditional distribution** of rvs $X_1, X_2, and\ X_3$, given that $X_4 = x_4, \ldots, X_n = x_n$ for example, is

$$f(x_1, x_2, x_3 | x_4, \ldots, x_n) = \frac{f(x_1, x_2, \ldots, x_n)}{g(x_4, \ldots, x_n)}$$

- **Mutual statistical independence** of rvs $X_1, X_2, \ldots, X_n$, with marginal distributions $f_1(x_1), f_2(x_2), \ldots, f_n(x_n)$, respectively

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

for all $(x_1, x_2, \ldots, x_n)$ within their range.

More Than Two Random Variables

Example:
Suppose that the shelf life (years) of a certain perishable food product packaged in cardboard containers is a random variable whose probability density function is given by

$$f(x) = \begin{cases} e^{-x}, & x > 0 \\ 0, & elsewhere \end{cases}$$

Let $X_1, X_2, and\ X_3$ represent the shelf lives for three of these containers selected independently and find $P(X_1 < 2, 1 < X_2 < 3, X_3 > 2)$.

$$f(x_1, x_2, x_3) = \begin{cases} e^{-x_1}\, e^{-x_2} e^{-x_3} = e^{-x_1 - x_2 - x_3}, & x_1, x_2, x_3 > 0 \\ 0, & elsewhere \end{cases}$$

$$P(X_1 < 2, 1 < X_2 < 3, X_3 > 2) = \int_2^{\infty} \int_1^{3} \int_0^{2} e^{-x_1 - x_2 - x_3}\, dx_1 dx_2 dx_3$$

$$* \int e^{-x} dx = -e^{-x}$$