Chapter 3. Random Variables

He Wang

Contents

§3.2 Binomial Probabilities

Review: **Example.** Roll a biased (unfair) coin 8 times. Suppose the probability of getting Head is $P(H) = p$. Find: $P(\text{Exactly 3 Heads})$.

We know that the result is $\binom{8}{3} \cdot p^3(1-p)^5$. (Reason: There are $\binom{8}{3}$ possible outcomes that has exactly 3 heads. For each outcome, the probability is $p^3(1-p)^5$.)

This can be generalized to a series of n independent trials with 2 outcomes: "success" or "failure".

Theorem: (Binomial Distribution)

Given a series of n **independent** trials with **two** outcomes.
Suppose the probability of "success" for each trial is **constant** p. Then,

$$P(k \text{ successes}) = \binom{n}{k} \cdot p^k(1-p)^{n-k}$$

for $k = 0, 1, 2, \ldots, n$.

Remarks:

∘ Key for application: find correct n, k and p.

∘ When $n = 1$, the binomial distribution is called **Bernoulli** variable (distribution).

∘ The assumption means that there are n **independent** trails and each trail is identically the same distribution. (Here, each trail is Bernoulli distribution with probability p.) This assumption is called **IID** (Identical-Independent-Distributions).

∘ Recall the binomial formulas:

$$\begin{aligned}
(x+y)^2 &= x^2 + 2xy + y^2 \\
(x+y)^3 &= x^3 + 3x^2y + 3xy^2 + y^3 \\
(x+y)^4 &= x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4 \\
&\vdots
\end{aligned}$$

$$(x+y)^n = x^n + \binom{n}{n-1}x^{n-1}y + \cdots + \binom{n}{k}x^k y^{n-k} + \cdots + \binom{n}{1}xy^{n-1} + y^n$$

Example 1. A Roulette (a wheel gamble) has 18 red, 18 black, and 2 green.

(1) If you spin the wheel 20 times, what is the probability of getting 10 red.

Solution: For this question, we consider getting red as success and getting non-red as failure. So, it fits the binomial distribution with $n = 20$ and

$$p = P(red) = \frac{18}{38} = 9/19.$$

The probability of getting 10 red ($k = 10$)is

$$P(10\ red) = \binom{20}{10}(9/19)^{10}(10/19)^{10} \approx 0.171$$

(2) If you spin the wheel 10 times, what is the probability of getting 3 green.

Solution: For this question, we consider getting green as success and getting non-green as failure. So, it fits the binomial distribution with $n = 10$ and

$$p = P(green) = \frac{2}{38} = 1/19.$$

The probability of getting 3 green ($k = 3$)is

$$P(3\ green) = \binom{10}{3}(1/19)^3(18/19)^7 \approx 0.012.$$

Example 2. In NBA Final, the winner is the first team to get four victories. Suppose Boston Celtic will play with Team C for the final. For each game, Boston has 60% winning chance.

(1) What is the probability that Boston wins the championship within 5 games?

X: Number of games played until Boston win the championship.
We want to calculate the probability that $X = 4$ or $X = 5$.
For $X = 4$, the game result should be (B,B,B,B)
$P(X = 4) = (0.6)^4 \approx 0.1296$.
For $X = 5$, the game results should be $(\_,\_,\_,\_,B)$, where Boston need to win 3 in the first 4 games.
$P(X = 5) = \left[\binom{4}{3}(0.6)^3(0.4)\right](0.6) \approx 0.2074$.
So, the probability that Boston wins the championship within 5 games is $P(X \leq 5) = P(X = 4) + P(X = 5) \approx 0.337$

(2) What is the probability that Boston **wins** the championship using **7** games? (Win 3 of the first 6 games and win the 7th game)

X: Number of games played until Boston win the championship.
For $X = 7$, the game results should be $(\_, \_, \_, \_, \_, \_, B)$, where Boston need to win 3 in the first 6 games.
So, the probability that Boston wins the championship in exactly 7 games is

$$P(X = 7) = \left[\binom{6}{3}(0.6)^3(0.4^3)\right](0.6) \approx 0.166$$

Example 3. There are 100 marbles in a box: 60 **red**, 40 other colors.

(1) Choose 5 with replacement, find P(two **red**)

X: the number of red.
X fits the binomial distribution with $n = 5$ and $p = P(red) = \dfrac{60}{100} = 0.6$. (Here, "with replacement" is necessary to make sure X is binomial. Compare with the next question.)

$$P(two\ red) = P(X = 2) = \binom{5}{2}(0.6)^2(0.4)^3 \approx 0.2304$$

(2) Choose 5 without replacement, find P(two **red**).

The probability that the first ball is red is $60/100 = 0.6$. Because there is no replacement, once we get the first red marble, the probability of getting the next red marble is changed. For example,

$$P(RROOO) = \left(\frac{60}{100} \cdot \frac{59}{99}\right) \cdot \left(\frac{40}{98} \cdot \frac{39}{97} \cdot \frac{38}{96}\right)$$

$$P(ROROO) = \left(\frac{60}{100} \cdot \frac{40}{99}\right) \cdot \left(\frac{59}{98} \cdot \frac{39}{97} \cdot \frac{38}{96}\right)$$

There are $\binom{5}{2}$ possible outcomes with exactly 2 red marbles, and they all have the same probability.
So,

$$P(2\ red) = \binom{5}{2}\frac{60 \cdot 59 \cdot 40 \cdot 39 \cdot 38}{100 \cdot 99 \cdot 98 \cdot 97 \cdot 96}$$

(The general distribution for this question is called hypergeometric distribution.)

Example 4. Toss **10** fair 6-sided dice. What is the probability that at least two $6's$ appeared?

Solution: success=6 appeared; failure=6 did not appear.
X: the number of $6's$ appeared.
X fits the binomial distribution with $n = 10$ and $p = P(success) = 1/6$.
The probability that at least two $6's$ appeared is

$$\begin{aligned} P(k \geq 2) &= 1 - P(k < 2) \\ &= 1 - P(k = 0) - P(k = 1) \\ &= 1 - \binom{10}{0}(1/6)^0(5/6)^{10} - \binom{10}{1}(1/6)^1(5/6)^9 \\ &\approx 1 - 0.1615 - 0.3230 = 0.5155 \end{aligned}$$

You can use calculator TI-83/ TI-84 (plus) to verify the calculation.

$$\textbf{2ED} \to \textbf{Vars} \to A : \textbf{binompdf}$$

§3.2 Part 2. Hypergeometric Distribution.

Recall **Example 3**. There are 100 marbles in a box: 60 **red**, 40 other colors.

(2*) Choose 5 without replacement, find P(two **red**).

There are $\binom{5}{2}$ possible outcomes with exactly 2 red marbles, and they all have the same probability. So,

$$P(2\ red) = \binom{5}{2}\frac{(60 \cdot 59) \cdot (40 \cdot 39 \cdot 38)}{100 \cdot 99 \cdot 98 \cdot 97 \cdot 96} = \frac{5!}{3!2!}\frac{P_2^{60} \cdot P_3^{40}}{P_5^{100}} = \frac{\binom{60}{2}\binom{40}{3}}{\binom{100}{5}}$$

Theorem. Hypergeometric Distribution

Suppose an urn contains r red chips and w white chips, where $r + w = N$. If n chips are drawn out at random, **without replacement**, and if k denotes the number of red chips selected, then

$$P(k \text{ red chips are chosen}) = \frac{\binom{r}{k}\binom{w}{n-k}}{\binom{N}{n}}$$

where k varies over all the integers for which $\binom{r}{k}$ and $\binom{w}{n-k}$ are defined.

The probabilities appearing on the right-hand side of Equation are known as the **hypergeometric distribution**.

Example 5. Suppose we select 5 cards from a standard deck of playing cards. What is the probability of obtaining exactly 2 hearts?

This satisfies hypergeometric distribution with $N = 52$, $r = 13$, $w = 39$, $n = 5$ and $k = 2$. So the probability of obtaining exactly 2 hearts is

$$P(2 \text{ hearts}) = \frac{\binom{13}{2}\binom{39}{3}}{\binom{52}{5}} \approx 0.2743$$

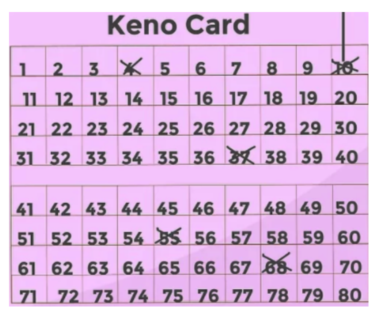
Example 6. In a TV show, the host randomly chooses 6 seat numbers for prizes from the 100 audiences. Suppose in the audiences, there are 40 men and 60 women. What is the probability that both men and women will be represented?

The number of men chosen are hypergeometric with $N = 100$, $r = 40$, $w = 60$, $n = 6$. We need to calculate the case $k = 0$ and $k = 6$.
The event that both men and women are represented is the **complement** of the event that 0 or 6 men will be chosen. The probability that both men and women will be represented is

$$1 - \frac{\binom{40}{0}\binom{60}{6}}{\binom{100}{6}} - \frac{\binom{40}{6}\binom{60}{0}}{\binom{100}{6}} = 0.9548$$

Example 7.

Keno is a casino game in which the player has a card with the numbers 1 through 80 on it. The player selects a set of k numbers from the card, where k can range from one to twenty. The "caller" announces twenty winning numbers, chosen at random from the eighty. The amount won depends on how many of the called numbers match those the player chose. (1) Suppose the player picks 5 numbers. What is the probability that all those 5 are winning numbers? (2) Suppose the player picks 20 numbers. What is the probability that among those twenty are five winning numbers?



(1) The probability is hypergeometric with $N = 80$, $r = 20$, $w = 60$, $n = 5$.

$$P(5 \text{ hits}) = \frac{\binom{20}{5}\binom{60}{0}}{\binom{80}{5}} \approx 0.0006645$$

(2) $P(5 \text{ hits in } 20) = \dfrac{\binom{20}{5}\binom{60}{15}}{\binom{80}{20}} \approx 0.233$

§3.3 Discrete Random Variables

Motivation Example. Toss **2** fair 6-sided dice. What is the probability that the **sum** of the numbers equal to 9?

The sample space S has 36 sample points given by

$$S = \left\{ \begin{matrix} (1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6) \\ (2,1), & (2,2), & (2,3), & (2,4), & (2,5), & (2,6) \\ (3,1), & (3,2), & (3,3), & (3,4), & (3,5), & (3,6) \\ (4,1), & (4,2), & (4,3), & (4,4), & (4,5), & (4,6) \\ (5,1), & (5,2), & (5,3), & (5,4), & (5,5), & (5,6) \\ (6,1), & (6,2), & (6,3), & (6,4), & (6,5), & (6,6) \end{matrix} \right\}$$

We know that the probability of each sample point is $\frac{1}{36}$.

The event A is given by $A = \{(6,3),(5,4),(4,5),(3,6)\}$

The probability of A is $P(A) = \frac{4}{36}$.

In this example, we only care about the sum of two numbers. So we want to **redefine** the sample space S to a "smaller sample space" in $\mathbb{R}$ using a **random variable** X assigning each outcome a real number,

$$X : S \to \mathbb{R}$$

In our example, we define $X : (a,b) \to a+b$. As a set $X(S) = \{2,3,4,5,6,7,8,9,10,11,12\} \subset \mathbb{R}$, which is called the **range** of X.

If the range of X is finite or countably subset of $\mathbb{R}$, then X is called a **discrete random variable**.

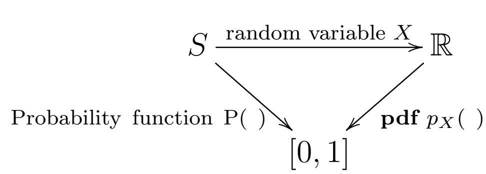
Definition.

For every discrete random variable X, we define a **probability density function (pdf)** by

$$p_X(k) = P(X=k) := P(\{s \in S | X(s) = k\})$$

(If $k \notin X(S)$, then $p_X(k) = 0$.)

Remark: 1. **Probability function** from sample space S to $[0,1] \subset \mathbb{R}$. 2. Random variable is a function from sample space S to $\mathbb{R}$. 3. **pdf** $p_X(k)$ is a function from $\mathbb{R}$ to $\mathbb{R}$.



Question: Find probability $p_X(k)$ for all k.

In our example, the probability $p_X(k)$ of each number k in the range is assigned by

k	2	3	4	5	6	7	8	9	10	11	12
$p_X(k)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

The probability that the sum of the numbers equal to 9 is given by $p_X(9) = \frac{4}{36}$.

Question: What is the probability that the sum of the numbers ≤ 4?

It is given by $P(X \leq 4) = p_X(2) + p_X(3) + p_X(4) = \frac{1}{36} + \frac{2}{36} + \frac{3}{36}$.

This is another useful function called **cumulative distribution function (cdf)**, defined as

$$F_X(t) = P(X \leq t) := P(\{s \in S | X(s) \leq t\})$$

In our example, find the cumulative distribution function (cdf).

t	2	3	4	5	6	7	8	9	10	11	12
$F_X(t)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{15}{36}$	$\frac{21}{36}$	$\frac{26}{36}$	$\frac{30}{36}$	$\frac{33}{36}$	$\frac{35}{36}$	$\frac{36}{36}$

Example 1. (**Binomial Distribution**) in §3.2

Let the random variable X denote the number of "successes" in n independent trials. Then, the Binomial Distribution can be stated as

$$P(X = k) = \binom{n}{k} \cdot p^k (1-p)^{n-k}$$

for $k = 0, 1, 2, \ldots, n$.

The cumulative distribution function (**cdf**) of X is

$$P(X \leq t) = \sum_{k=0}^{t} \binom{n}{k} \cdot p^k (1-p)^{n-k}$$

Theorem.

For any discrete random variable X, the **pdf** satisfies

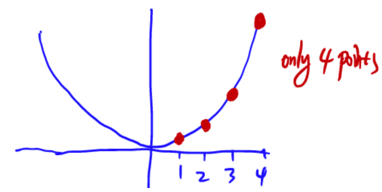
1. $p_X(k) \geq 0$.
2. $\sum_{x \in X(S)} p_X(k) = 1$.

Example 2. If a pdf is given by $p_X(k) = c \cdot k^2$ for $X(S) = \{1, 2, 3, 4\}$. Find c, and graph the pdf function.

By the above theorem,

$$p_X(1) + p_X(2) + p_X(3) + P_X(4) = 1,$$

So, $c + 4c + 9c + 16c = 1$, which implies $c = \frac{1}{30}$.
So $p_X(k) = (\frac{1}{30}) \cdot k^2$.



The definition of the cdf function $F_X(t) = \sum_{k \leq t} p_X(k)$ is calculated by the sum of the pdf functions $p_X(k)$.

We can also calculate the pdf $p_X(k)$ using the cdf function $F_X(t)$ by the following proposition.

Theorem.

For any discrete random variable X, the cdf function $F_X(t)$ and the pdf function $p_X(k)$ satisfy

$$p_X(k) = F_X(k) - F_X(k-1).$$

Example 3. Suppose the cumulative distribution function (cdf) is given by

t	1	3	5	6	8	9	10	12	15	16	19
$F_X(t)$	$\frac{1}{39}$	$\frac{1}{36}$	$\frac{2}{37}$	$\frac{3}{35}$	$\frac{1}{11}$	$\frac{2}{13}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{2}{3}$	$\frac{33}{35}$	1

What is the pdf $P(X = 9)$?

$$P(X = 9) = F_X(9) - F_X(8) = 2/13 - 1/11$$

Example 4. (HW 3.3.7 or 3.3.8)

Suppose you have \$10 and you go to gamble. Each time, you will either win or lose \$1. Each time, the probability that you will win is $\frac{1}{3}$.

What is the **pdf** of your money situation after 6 times gamble?

X: the number of times you win. $(X = 0, 1, 2, 3, 4, 5, 6)$
The number of losing times is $6 - X$.
In the end, the money in your pocket is $Y = 10 + X - (6 - X) = 4 + 2X$
The probability of win k times in 6 gambles is

$$P(X = k) = \binom{6}{k}(1/3)^k(2/3)^{6-k}$$

k	0	1	2	3	4	5	6
$P(X = k)$	0.0878	0.2634	0.3294	0.2195	0.0823	0.0164	0.0014
$P(Y = 4 + 2k)$	0.0878	0.2634	0.3294	0.2195	0.0823	0.0164	0.0014
$4 + 2k$	4	6	8	10	12	14	16

Or we write the pdf for Y as

$$P(Y = 4 + 2k) = \binom{6}{k}(1/3)^k(2/3)^{6-k}$$

for $k = 0, 1, 2, 3, 4, 5, 6$.

The mathematical model for this example called one dimensional **Random Walk**. Suppose that a dot sits on an integer number line. The dot starts in the center and start walk. For each step, it either forward or backward, with equal probability. We want to know where is the dot after it has taken k steps.

§3.4 Continuous Random Variables

The range of a **continuous random variable** X is a (piecewise) continuous interval of $\mathbb{R}$.

Motivation Example. Choose a real number randomly from the interval $[0, 2]$ (sample space). If we assume the numbers are equally likely, we have the following the probabilities:

- $P(X \leq 2) = 1$
- $P(X \leq 0.2) = 0.1$
- $P(X \leq 0.02) = 0.01$

  $\vdots$

- $P(X \leq x) = x/2$

We can continue this and $P(X = 0) = 0$.

In fact, $P(X = a) = 0$ for any real number. So, we care about the probability for a interval.

Definition.

The **probability density function (pdf)** of a **continuous** random variable X is a piecewise continuous function $f_X(x)$ satisfying

1. $f_X(x) \geq 0$
2. $\displaystyle\int_{-\infty}^{\infty} f_X(x)dx = 1.$

We also define $f_X(x) = 0$ if x in not in the range of X.

Definition.

The **probability** that X is in an interval $[a, b]$ is

$$P(a \leq X \leq b) = \int_a^b f_X(x)dx$$

Definition.

The **cumulative distribution function (cdf)** of a **continuous** random variable X is

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)dt$$

From the Fundamental Theorem of Calculus, we have the relation between cdf and pdf:

Theorem.

$$F'_X(x) = f_X(x)$$

We can use the cdf to find the probability

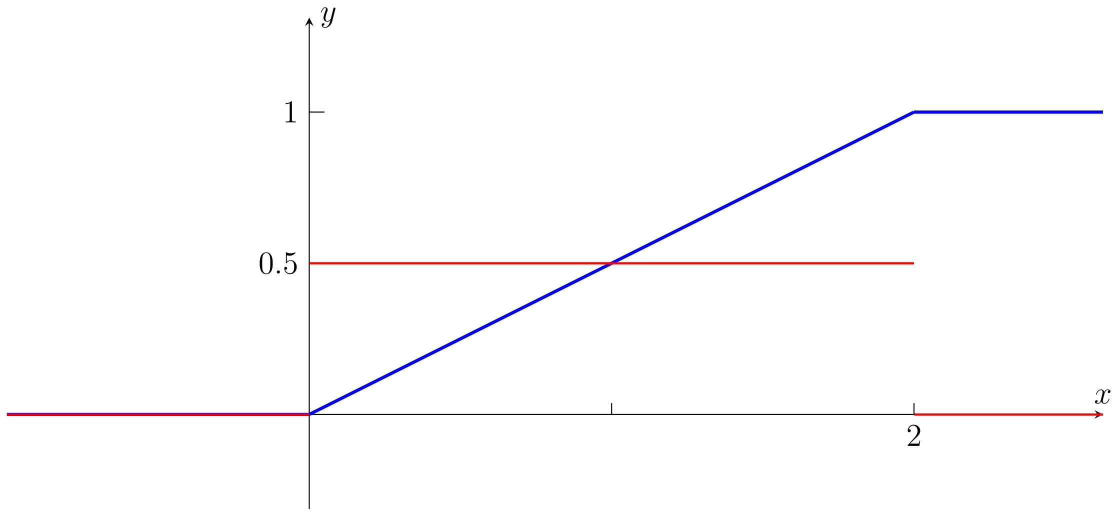
Theorem.

$$P(a \le X \le b) = F_X(b) - F_X(a)$$

In our motivation example, we have cdf $F_X(x) = P(X \le x) = \begin{cases} 0 & \text{for } x \in [-\infty, 0] \\ x/2 & \text{for } x \in [0, 2] \\ 1 & \text{for } x \in [2, \infty] \end{cases}$

So, the pdf $f_X(x) = F'_X(x) = \begin{cases} 1/2 & \text{for } x \in [0, 2] \\ 0 & \text{for others} \end{cases}$

The graphs for the pdf $f_X(x)$ and the cdf $F_X(x)$:



The **cdf** function is always a **continuous**, **increasing** function. The minimum is 0 and the maximum is 1.

Example 1. Choose a number randomly from the interval $[a, b]$. If we assume the numbers are equally likely. (This distribution is called **uniform distribution.**) Find pdf $f_X(x)$ and cdf $F_X(x)$.

The cdf function is

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & \text{for } x \in [-\infty, a] \\ \frac{x-a}{b-a} & \text{for } x \in [a, b] \\ 1 & \text{for } x \in [b, \infty] \end{cases}$$

The pdf function is

$$f_X(x) = F'_X(x) = \begin{cases} \frac{1}{b-a} & \text{for } x \in [a, b] \\ 0 & \text{for others} \end{cases}$$

Example 2. Suppose the pdf of a random variable Y is $f_Y(y) = c \cdot y^3$ for $0 \leq y \leq 2$.
(1) Find c and calculate $P(0 \leq Y \leq 1)$.

$$1 = \int_{-\infty}^{\infty} cy^3 dy = \int_0^2 cy^3 dy = \left[\frac{cy^4}{4}\right]_0^2 = 4c$$

So $c = 1/4$. The probability

$$P(0 \leq Y \leq 1) = \int_0^1 \frac{1}{4} y^3 dy = \left[\frac{y^4}{16}\right]_0^1 = 1/16$$

(2) Find the cdf $F_Y(y)$.

Solution: The cdf is

$$F_Y(y) = \int_{-\infty}^{y} f_Y(t) dy = \int_0^y \frac{1}{4} t^3 dt = \left[\frac{t^4}{16}\right]_0^y = y^4/16$$

where $0 \leq y \leq 2$.
$F_Y(y) = 0$ when $y < 0$, $F_Y(y) = 1$ when $y > 2$.

Example 3. An important continuous distribution is the **exponential distribution** defined as

$$f_X(x) = \lambda e^{-\lambda x} \text{ for } x \geq 0$$

where λ is a positive parameter.

(1) Check that $f_X(x) \geq 0$ and $\int_{-\infty}^{\infty} f_X(x) dx = 1$.

It is clear that $f_X(x) \geq 0$. For the second equality,

$$\int_{-\infty}^{\infty} f_X(x) dx = \int_0^{\infty} \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x}\right]_0^{\infty} = 1$$

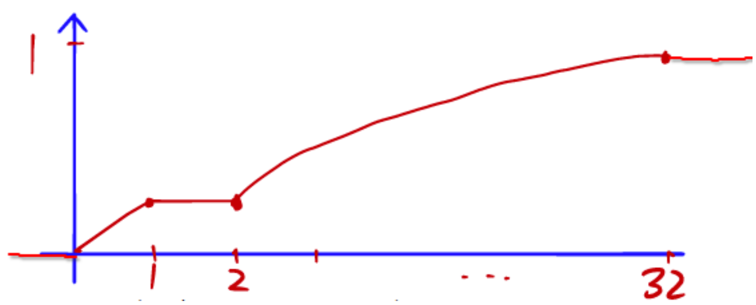
(2) Calculate the cdf $F_X(x)$ of X.

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \int_{0}^{x} \lambda e^{-\lambda t}dt = \left[-e^{-\lambda t}\right]_0^x = -e^{-\lambda x} - (-1) = 1 - e^{-\lambda x}$$

Example 4. The cdf of a random variable X is given by

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x/4 & 0 \leq x < 1 \\ 1/4 & 1 \leq x < 2 \\ \sqrt{2x}/8 & 2 \leq x < 32 \\ 1 & x \geq 32 \end{cases}$$

(1) Sketch the graph the cdf $F_X(x)$.



(2) Find $P(1/2 \leq X < 4)$.

$$P(\frac{1}{2} \leq x < 4) = F_X(4) - F_X(1/2) = \frac{\sqrt{8}}{8} - \frac{1}{8}$$

(3) Find $P(X > 4)$.

$$P(X > 4) = 1 - P(X \leq 4) = 1 - F_X(4) = 1 - \frac{\sqrt{8}}{8}$$

(4) Find the pdf of X.

The pdf of X is given by

$$f_X(x) = F'_X(x) = \begin{cases} 0 & x < 0 \\ 1/4 & 0 \leq x < 1 \\ 0 & 1 \leq x < 2 \\ \dfrac{\sqrt{2}}{8} \cdot \dfrac{1}{2} x^{-\frac{1}{2}} & 2 \leq x < 32 \\ 0 & x \geq 32 \end{cases}$$

§3.5 Expected Values

Expected value is a generalization of the concept "average".

Definition: Expected Value of **discrete** random variable

If X is a discrete random variable with probability function $p_X(k)$, then the **expected value** (or Mean) of X is

$$E(X) = \sum_{\text{all } k} k \cdot p_X(k).$$

Example 1. (In $ 3.3) Suppose you have $10 and you go to gamble. Each time, you will either win or lose $1. Each time, the probability that you will win is $\frac{1}{3}$.

We set k as the number of your winning times. The random variable Y as the amount of your money and we already calculated that $Y = 4 + 2k$. We already have the **pdf** of Y as

k	0	1	2	3	4	5	6
$P(X = k)$	0.0878	0.2634	0.3294	0.2195	0.0823	0.0164	0.0014
$P(Y = 4 + 2k)$	0.0878	0.2634	0.3294	0.2195	0.0823	0.0164	0.0014
$4 + 2k$	4	6	8	10	12	14	16

Q: What is the expect value of Y?

Solution:

$$E(Y) = 0.0878(4) + 0.2634(6) + 0.3292(8) + 0.2195(10) + 0.0823(12) + 0.0164(14) + 0.0014(16) \approx 8$$

Q: What is the expect value of X?

$E(X) = 2$

Proposition.

$$E(aX + b) = aE(X) + b$$

where a, b are real numbers.

Example 2. (Practice) Toss **2** fair 6-sided dice.

Let X be the **difference** of the two numbers (large−small).

The sample space S has 36 sample points given by

$$S = \left\{ \begin{array}{llllll} (1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6) \\ (2,1), & (2,2), & (2,3), & (2,4), & (2,5), & (2,6) \\ (3,1), & (3,2), & (3,3), & (3,4), & (3,5), & (3,6) \\ (4,1), & (4,2), & (4,3), & (4,4), & (4,5), & (4,6) \\ (5,1), & (5,2), & (5,3), & (5,4), & (5,5), & (5,6) \\ (6,1), & (6,2), & (6,3), & (6,4), & (6,5), & (6,6) \end{array} \right\}$$

So $X(S) = \{0, 1, 2, 3, 4, 5\}$.

What is the **pdf** of X?

k	0	1	2	3	4	5
$p_X(k)$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{6}{36}$	$\frac{4}{36}$	$\frac{2}{36}$

What is the expect value of X?

Solution:

$$E(X) = \frac{6}{36}(0) + \frac{10}{36}(1) + \frac{8}{36}(2) + \frac{6}{36}(3) + \frac{4}{36}(4) + \frac{2}{36}(5) = \frac{35}{18} \approx 1.94.$$

Theorem. Expected value of binomial distribution

Suppose X is a binomial random variable with parameters n and p. That is $p_X(k) = \binom{n}{k} \cdot p^k(1-p)^{n-k}$. Then

$$E(X) = np.$$

In a multiple-choice test, there are 100 questions and each with five possible answers. Let X be the number of correct answers just by guessing. Then X binomial random variable with parameters $n = 100$ and $p = \frac{1}{5}$. So, by Theorem, $E(X) = np = 20$.

We would "expect" to get 20 correct answers by "intuition".

Proof.

$$\begin{aligned}E(X) &= \sum_{k=0}^{n} k\binom{n}{k} \cdot p^k(1-p)^{n-k} \\ &= \sum_{k=0}^{n} k\frac{n!}{(n-k)!k!} \cdot p^k(1-p)^{n-k} \\ &= \sum_{k=0}^{n} \frac{n!}{(n-k)!(k-1)!} \cdot p^k(1-p)^{n-k} \\ &= np\sum_{k=1}^{n} \binom{n-1}{k-1} \cdot p^{k-1}(1-p)^{n-k} \\ &= np(p+1-p)^{n-1} \\ &= np\end{aligned}$$

Definition: (Expected Value of continuous random variable)

If X is a continuous random variable with probability function $p_X(k)$, then the **expected value** (or Mean) of X is

$$E(X) = \int_{-\infty}^{\infty} x \cdot p_X(x)\, dx.$$

Proposition.

$$E(aX+b) = aE(X) + b$$

where a, b are real numbers.

Proof:

$$E(aX+b) = \int_{-\infty}^{\infty} (ax+b)f_X(x)dx = a\int_{-\infty}^{\infty} xf_X(x)dx + b\int_{-\infty}^{\infty} f_X(x)dx = aE(X) + b$$

Example 3. The **pdf** for a continuous random variable Y is

$$p_Y(y) = \frac{3}{8}(y^2+1) \text{ for } -1 \le y \le 1.$$

Find the expected value (mean) $E(Y)$.

The expected value of Y is

$$\begin{aligned} E(Y) &= \int_{-\infty}^{\infty} y p_Y(y) dy \\ &= \int_{-1}^{1} y \frac{3}{8}(y^2+1) dy \\ &= \frac{3}{8}\int_{-1}^{1} y^3 + y dy \\ &= \left[\frac{3}{8}(\frac{y^4}{4}+\frac{y^2}{2})\right] \\ &= 0 \end{aligned}$$

Example 4. Let X be an exponential random variable,

$$f_X(x) = \lambda e^{-\lambda x} \text{ for } x \geq 0.$$

Find the expected value of X.

The expected value of X is

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f_X(x) dx \\ &= \int_{0}^{\infty} x \lambda e^{-\lambda x} dx \\ &= \left[-x e^{-\lambda x}\right]_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx \\ &= 0 + \left[\frac{e^{-\lambda x}}{-\lambda}\right]_0^{\infty} \\ &= \lim_{t\to\infty}\left(\frac{1}{-\lambda e^{\lambda t}} + \frac{1}{\lambda}\right) \\ &= \frac{1}{\lambda} \end{aligned}$$

For the third equality, we need integral by parts for $u = x$ and $v' = \lambda e^{-\lambda x}$. Then $u' = 1$ and $v = -e^{-\lambda x}$. So, $\int uv'dx = uv - \int u'vdx$.

Sometimes, the mean is not enough to describe the variable. Especially if there are extreme values on both sides.

Definition.

Let X be a discrete random variable. The **median** of X is the number m such that $P(X \leq m) \geq 0.5$ and $P(X \geq m) \geq 0.5$.

Example 5. Find the median of the random variable Example 3.

k	0	1	2	3	4	5
$p_X(k)$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{6}{36}$	$\frac{4}{36}$	$\frac{2}{36}$

$k = 2$ is the median.

Definition.

Let Y be a continuous random variable. The **median** of Y is the number m such that

$$\int_{-\infty}^{m} f_Y(y)\ dy = 0.5$$

Example 6. Find the median of

$$f_X(x) = e^{-x} \text{ for } x \geq 0.$$

Suppose m is the median.

$$0.5 = \int_0^m e^{-x}dx = \left[-e^{-x}\right]_0^m = -e^{-m} - (-1)$$

So, $e^{-m} = 0.5$ and then $m = -\ln(0.5) = \ln 2$.

More Practice Examples:

Example 7. (St. Petersburg paradox) A fair coin is flipped until the first Tail appears.
You win \$2 if it appears on the first toss.
You win \$4 if it appears on the second toss
In general, you win $\$2^k$ if it first occurs on the k-th toss.
Let the random variable X denote our winnings. How much should we have to pay in order for this to be a fair game?
(A fair game is one where the difference between the ante and $E(X)$ is 0.)

$p_X(2^k) = P(X = 2^k) = \frac{1}{2^k}$ for $k = 1, 2, ...$ So the expect value is

$$E(X) = \sum_{all\ \ k} 2^k p_X(2^k) = \sum_{k=1}^{\infty} 2^k(\frac{1}{2^k}) = 1 + 1 + 1 + \cdots = \infty$$

Example 8. Suppose 4 friends have exactly the same phone. They put their phone on the desk and pick one randomly.

(1) Let S_i be the even that the i-th person picked the correct phone. Denote X_i be the random variable that $X_i(S_i) = 1; X_i(S_i^c) = 0$. Find the pdf of X_i. Find the expected value of X_i

Answer: The probability that the i-th person picked the correct phone is $P(S_i) = \dfrac{P_3^3}{P_4^4} = 1/4$.

The **pdf** of X_i is **Bernoulli** distribution:

$X_i = k$	0	1
$P(X_i = k)$	3/4	1/4

The expect value of X_i is $E(X_i) = 1/4$.

(2) Is X is binomial distribution? What is the reason?

X is not binomial distribution.
The reason is that S_i are not independent.
The sample space S includes all permutations of $(1, 2, 3, 4)$. So the size (cardinality) of S is $|S| = P_4^4 = 4! = 24$. $P(S_1 \cap S_2) = \dfrac{2}{4!}$ since $S_1 \cap S_2 = \{(1, 2, 3, 4), (1, 2, 4, 3)\}$. So $P(S_1 \cap S_2) \neq P(S_1)P(S_2)$. So S_1 and S_2 are not independent.

(3) Let X be the number of correct pick-ups. Find the pdf of X. Find the expected value of X.

The range of X is $\{0, 1, 2, 3, 4\}$.
The probability that there are 4 correct pick-ups. $p_X(4) = P(X = 4) = \dfrac{1}{P_4^4} = 1/24$.
The probability that there are 3 correct pick-ups. $p_X(3) = P(X = 3) = 0$.
The probability that there are 2 correct pick-ups. $p_X(0) = P(X = 2) = \dfrac{\binom{4}{2}}{P_4^4} = 6/24$.
The probability that there are 1 correct pick-ups. $p_X(1) = P(X = 1) = \dfrac{\binom{4}{1}2}{P_4^4}$.
The probability that there are 0 correct pick-ups. $p_X(0) = P(X = 0) = \dfrac{9}{P_4^4}$.
The last two computations involves a technique called dearrangement, (denoted as $!n$) which means a permutation that no element appears in its original position. The calculation is by a recursive formula $!n = (n-1)(!(n-1)+!(n-2))$. So, $!1 = 0$; $!2 = 1$; $!3 = 2$; $!4 = 3(2+1) = 9$.
So the pdf of X is

$X = k$	0	1	2	3	4
$P(X = k)$	9/24	8/24	6/24	0	1/24

The expected value $E(X) = \frac{8}{24}(1) + \frac{6}{24}(2) + \frac{1}{24}(4) = 1$

(4) What is the relationship among X, X_1, X_2, X_3, X_4. What is the relation among their expected

values? Is there a relationship among their pdfs?

$X = X_1 + X_2 + X_3 + X_4$.
$E(X) = E(X_1) + E(X_2) + E(X_3) + E(X_4) = 1$.
There is no direct relation among the pdf functions of X, X_1, X_2, X_3, X_4.

§3.6 Variance

Theorem.

$E(g(X)) = \sum_{x \in X(S)} g(x)p_X(x)$ if X is a discrete random variable.

$E(g(Y)) = \int_{-\infty}^{\infty} g(y)f_Y(y)$ if Y is a continuous random variable.

In particular, we care about the case when $g(X) = X^2$.

Warning: $E(X^2) \neq (E(X))^2$.

Example 1. Calculate $E(X^2)$ for the random variable X in Example 2 in §3.5.

Solution: $E(X^2) = \frac{6}{36}(0) + \frac{10}{36}(1^2) + \frac{8}{36}(2^2) + \frac{6}{36}(3^2) + \frac{4}{36}(4^2) + \frac{2}{36}(5^2) = \frac{35}{6} \approx 5.83$.

Example 2. Calculate $E(Y^2)$ for the random variable in Example 3 in §3.5.

$$\begin{aligned} E(Y^2) &= \int_{-1}^{1} y^2 \frac{3}{8}(y^2+1)dy \\ &= \frac{3}{8}\int_{-1}^{1} y^4 + y^2 dy \\ &= \frac{3}{8}\left[\frac{y^5}{5} + \frac{y^3}{3}\right]_{-1}^{1} \\ &= 2/5 \end{aligned}$$

Definition. (Variance)

The **variance** of a random variable X is

$$\mathrm{Var}(X) := E\big((X-\mu)^2\big)$$

Here $\mu = E(X)$ is the mean of X.
The **standard deviation** is $\sigma := \sqrt{\mathrm{Var}(X)}$

Remark: This is the mean of the squared distance from the mean. It measures the spread of the data. The expect value $E\big(|X-\mu|\big)$ can measure the same property, but it is harder to calculate.

Theorem. Calculation of Variance

Let X be a random variable with mean $\mu = E(X)$.

$$\mathrm{Var}(X) = E(X^2) - \mu^2$$

Proof:

$$\begin{aligned}\mathrm{Var}(X) &= E((X-\mu)^2) \\ &= E(X^2 - 2\mu X + \mu^2) \\ &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - 2\mu^2 + \mu^2 \\ &= E(X^2) - \mu^2\end{aligned}$$

Example 3. Calculate the **variance** for the random variable X in Example 1.

$$\mathrm{Var}(X) = E(X^2) - \mu^2 = 5.83^2 - (1.94)^2 \approx 2.07$$

Example 4. Calculate the **variance** for the random variable in Example 2.

$$\mathrm{Var}(X) = E(X^2) - \mu^2 = 2/5 - 0 = 2/5.$$

Example 5. Calculate the standard deviation of X with pdf

$$f_X(x) = \begin{cases} 2-x, & 1 \le x \le 2 \\ 1/2 & 3 \le x \le 4 \end{cases}$$

$$\begin{aligned}E(X) &= \int_1^2 x(2-x)\,dx + \int_3^4 \frac{x}{2}\,dx \\ &= x^2 - \frac{x^3}{3}\Big|_1^2 + \frac{x^2}{4}\Big|_3^4 \\ &= \left(4 - \frac{8}{3} - \left(1 - \frac{1}{3}\right)\right) + \frac{16}{4} - \frac{9}{4} \\ &= \frac{29}{12} \approx 2.42\end{aligned}$$

$$\begin{aligned}E(X^2) &= \int_1^2 x^2(2-x)\,dx + \int_3^4 \frac{x^2}{2}\,dx \\ &= \frac{2x^3}{3} - \frac{x^4}{4}\Big|_1^2 + \frac{x^3}{6}\Big|_3^4 \\ &= \left(\frac{16}{3} - \frac{16}{4}\right) - \left(\frac{2}{3} - \frac{1}{4}\right) + \frac{64}{6} - \frac{27}{6} \approx 7.08\end{aligned}$$

$$\mathrm{Var}(X) = E(X^2) - (E(X))^2 = 7.08 - (2.42)^2 \approx 1.22$$

The standard deviation of X is $\sigma = \sqrt{\mathrm{Var}(X)} = \sqrt{1.22} \approx 1.1$

Example 6. (Homework 11) Let X be an exponential random variable with

$$f_X(x) = \lambda e^{-\lambda x} \text{ for } x \geq 0.$$

Find the variance and standard deviation of X.

From Example §3.5, we have $E(X) = \dfrac{1}{\lambda}$

$$E(X^2) = \int_0^\infty x^2 \lambda e^{-\lambda x}\,dx$$

$u = x^2 \quad v' = \lambda e^{-\lambda x}$

$u' = 2x \quad v = -e^{-\lambda x}$

$$= -x^2 e^{-\lambda x}\Big|_0^\infty - \int_0^\infty 2x(-e^{-\lambda x})\,dx$$

$$= \int_0^\infty 2x\, e^{-\lambda x}\,dx$$

$u = 2x \quad v' = e^{-\lambda x}$

$u' = 2 \quad v = \dfrac{e^{-\lambda x}}{-\lambda}$

$$= 2x\left(\frac{e^{-\lambda x}}{-\lambda}\right)\Big|_0^\infty - \int_0^\infty 2\left(-\frac{e^{-\lambda x}}{\lambda}\right)dx$$

$$= \int_0^\infty \frac{2}{\lambda} e^{-\lambda x}\,dx$$

$$= \frac{2}{\lambda}\,\frac{e^{-\lambda x}}{-\lambda}\Big|_0^\infty$$

$$= \frac{2}{\lambda}\,\frac{1}{\lambda} = \frac{2}{\lambda^2}$$

$$\mathrm{Var}(X) = E(X^2) - (E(X))^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

$$\sigma = \sqrt{\mathrm{Var}(X)} = \frac{1}{\lambda}$$

Theorem.

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

Proof.

$$\begin{aligned}
\mathrm{Var}(aX+b) &= E((aX+b)^2) - (E(aX+b))^2 \\
&= E((a^2X^2 + 2abX + b^2) - (a\mu + b)^2 \\
&= a^2E(X^2) + 2abE(X) + b^2 - a^2\mu^2 - 2ab\mu - b^2 \\
&= a^2E(X^2) - a^2\mu^2 \\
&= a^2\,\mathrm{Var}(X)
\end{aligned}$$

Theorem.

Let X be the binomial random variable

$$\text{Var}(X) = np(1-p)$$

Remark: The idea of mean and variance can be generalized to **the k-th moment**:

$$E(X^k)$$

and the **the k-th moment about the mean**:

$$E((X-\mu)^k)$$

Example 7. Let X be the **uniform distribution** on $[a, b]$. We already know the pdf function is

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{for } x \in [a,b] \\ 0 & \text{for others} \end{cases}$$

Find the expected value and variance of X.

The expected value is

$$E(X) = \int_a^b \frac{1}{b-a} x dx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b = \frac{a+b}{2}$$

$$E(X^2) = \int_a^b \frac{1}{b-a} x^2 dx = \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b = \frac{1}{3}(a^2+ab+b^2)$$

The variance is

$$\text{Var}(X) = E(X^2) - E(X)^2 = \frac{1}{3}(a^2+ab+b^2) - \frac{1}{4}(a^2+2ab+b^2) = \frac{1}{12}(b-a)^2.$$

We can also calculate the k-th moment of X

$$E(X^k) = \int_a^b \frac{1}{b-a} x^k dx = \frac{1}{b-a}\left[\frac{x^{k+1}}{k+1}\right]_a^b = \frac{1}{k+1}\frac{b^{k+1}-a^{k+1}}{b-a}$$

More about exponential random variable

Let X be an exponential random variable with **pdf** given by

$$f_X(x) = \lambda e^{-\lambda x} \text{ for } x \geq 0$$

where λ is a fixed positive number.

In §3.4, we verified that it is a **pdf** with **cdf** $F_X(x) = 1 - e^{-\lambda x}$.
In §3.5, we computed the mean of X, which is $E(X) = \dfrac{1}{\lambda}$
In §3.6 we computed the variance of X, which is $\text{Var}(X) = \dfrac{1}{\lambda^2}$

This is a very useful random variable to model the **life time** of some objects, i.e., computer parts, electric equipment, etc. Also including a customer spends in line at a bank tellers window, etc.

Geometric Distribution.

There is a discrete random variable works similarly. For example, if we flip a unfair (biased) coin with $P(Head) = p$. Let Y denote the times until we get our first Head. The **pdf** of Y is

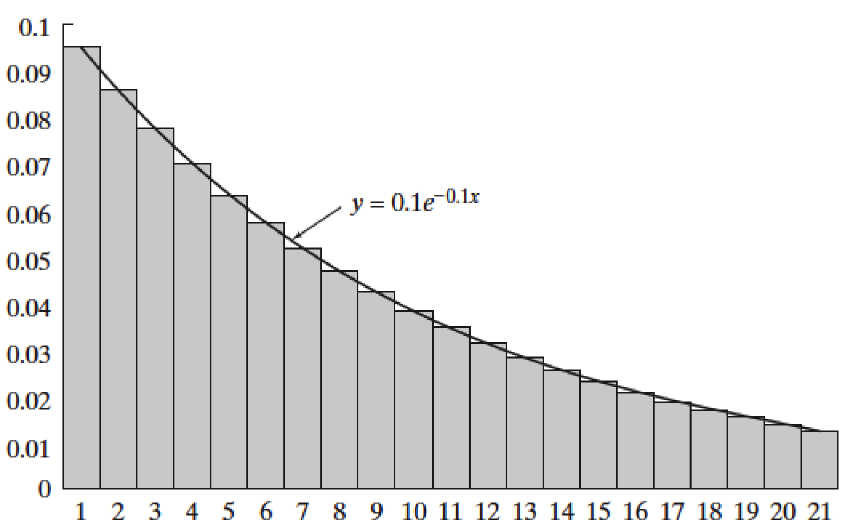
$$p_Y(k) = (1-p)^{k-1}p$$

for $k = 1, 2, 3, 4, ...$

The mean of Y is $\dfrac{1}{p}$.

(To compute this, we need to use the technique of Taylor series in Calculus II.)

Comparison. $p_Y(k) = (1 - 0.095)^{k-1}0.095$



§3.7 Joint Densities

Example 1. (Toss 2 fair 6-sided dice) Let X be the **difference** of the two numbers. Let Y be the **larger number** of the two numbers.

The sample space S has 36 sample points given by

$$S = \left\{ \begin{array}{cccccc} (1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6) \\ (2,1), & (2,2), & (2,3), & (2,4), & (2,5), & (2,6) \\ (3,1), & (3,2), & (3,3), & (3,4), & (3,5), & (3,6) \\ (4,1), & (4,2), & (4,3), & (4,4), & (4,5), & (4,6) \\ (5,1), & (5,2), & (5,3), & (5,4), & (5,5), & (5,6) \\ (6,1), & (6,2), & (6,3), & (6,4), & (6,5), & (6,6) \end{array} \right\}$$

From §3.5, the range of X is $X(S) = \{0, 1, 2, 3, 4, 5\}$ and the **pdf** of X is

$X = x$	0	1	2	3	4	5
$p_X(x)$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{6}{36}$	$\frac{4}{36}$	$\frac{2}{36}$

Similarly, range of Y is $Y(S) = \{1, 2, 3, 4, 5, 6\}$ and the **pdf** of Y is

$Y = y$	1	2	3	4	5	6
$p_Y(y)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

Definition. Discrete Joint Density

Let S is a **discrete** sample space. Let X and Y be two random variables on S. The **joint probability density function** (**joint pdf**) of X and Y is denoted by $p_{X,Y}(x, y)$ defined as

$$p_{X,Y}(x, y) := P(X = x, Y = y).$$

Here, $P(X = x, Y = y)$ is the probability when $X = x$ **and** $Y = y$.

Theorem.

The **joint pdf** $p_{X,Y}(x, y)$ satisfies

(1.) $p_{X,Y}(x, y) \geq 0$.

(2.) $\sum_{\text{All } x} \sum_{\text{All } y} p_{X,Y}(x, y) = 1$.

Question: Find the **joint pdf** of X and Y, $p_{X,Y}(x, y)$.

$Y=y$ \ $X=x$	0	1	2	3	4	5	$P_Y(y)$
1	$\frac{1}{36}$	0	0	0	0	0	$\frac{1}{36}$
2	$\frac{1}{36}$	$\frac{2}{36}$	0	0	0	0	$\frac{3}{36}$
3	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	0	0	0	$\frac{5}{36}$
4	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	0	0	$\frac{7}{36}$
5	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	0	$\frac{9}{36}$
6	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{11}{36}$
$P_X(x)=$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{6}{36}$	$\frac{4}{36}$	$\frac{2}{36}$	

$P(X=0, Y=1) = \frac{1}{36}$ (1,1) $\quad P(X=1, Y=1)=0 \quad P(X=2, Y=1)=0 \quad \cdots$

$P(X=0, Y=2) = \frac{1}{36}$ (2,2) $\quad P(X=1, Y=2)=\frac{2}{36}$ (2,1) (1,2) $\quad P(X=2, Y=2)=0 \quad \cdots$

$\vdots \qquad P(X=1, Y=3)=\frac{2}{36}$ (3,2) (2,3) $\quad P(X=2, Y=3)=\frac{2}{36}$ (3,1) (1,3) $\quad \cdots$

$\vdots \qquad \vdots$

Theorem.

Let $p_{X,Y}(x,y)$ be the **joint pdf** of X and Y. Then

$$p_X(x) = \sum_{\text{All } y} p_{X,Y}(x,y), \qquad \text{and} \qquad p_Y(y) = \sum_{\text{All } x} p_{X,Y}(x,y)$$

They are called the **marginal pdfs** of random variables X and Y respectively.

In particular,

$$p_X(a) = P(X=a) = \sum_{\text{All } y} P(X=a, Y=y)$$
$$p_Y(b) = P(Y=b) = \sum_{\text{All } x} P(X=x, Y=b)$$

Remark: In general, one can NOT recover joint pdf of X and Y, $p_{X,Y}(x,y)$ from the marginal pdfs $p_X(x)$ and $p_Y(y)$.

Question: Find the **marginal pdf**s for X and Y in Example 1.

Definition.

Two random variables X and Y are called **independent** if and only if $p_{X,Y}(x,y) = p_X(x)p_Y(y)$.

Definition.

If X and Y are **continuous** random variables. the **joint pdf** $f_{X,Y}(x,y)$ of X and Y is a piecewise continuous multi-variable function satisfying

(1.) $f_{X,Y}(x,y) \geq 0$.

(2.) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dxdy = 1$.

The **probability** that X and Y are in a region R in the xy-plane $\mathbb{R}^2$ is given by

$$P\big((X,Y) \in R\big) = \iint_R f_{X,Y}(x,y)\ dxdy$$

This involves the calculation the double integral $\iint_R f_{X,Y}(x,y)\ dxdy$ from Calculus 3. If you have not learned Calculus 3, we can learn and do some easy examples here.

Definition.

Let $f_{X,Y}(x,y)$ be the joint pdf of random variables X and Y. Then, the **marginal pdf** of X is

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dy$$

and the **marginal pdf** of Y is

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

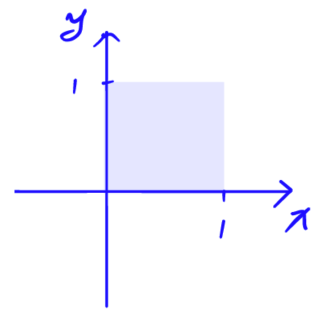
Example 2. Suppose the **pdf** function is $f_{X,Y}(x,y) = c(x+y)$ for $0 \leq x \leq 1$ and $0 \leq y \leq 1$.

(1) Find c.

$$1=\int_0^1\int_0^1 c(x+y)\,dy\,dx = \int_0^1 c\left(xy+\frac{y^2}{2}\right)\Big|_0^1 dx$$

$$=c\int_0^1 x+\frac{1}{2}\,dx = c\left(\frac{x^2}{2}+\frac{x}{2}\right)\Big|_0^1 = c\left(\frac{1}{2}+\frac{1}{2}\right)=c$$

$$\Rightarrow c=1$$

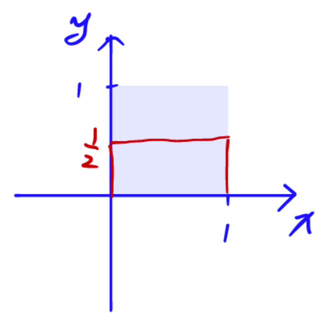


(2) Find $P(Y \leq \frac{1}{2})$

$$=\int_0^1\int_0^{\frac{1}{2}} x+y\,dy\,dx$$

$$=\int_0^1 xy+\frac{y^2}{2}\Big|_0^{\frac{1}{2}}\,dx$$

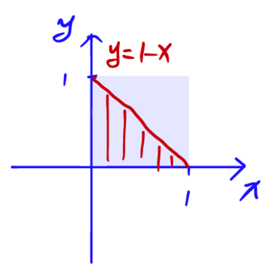
$$=\int_0^1 \frac{x}{2}+\frac{1}{8}\,dx = \frac{x^2}{4}+\frac{x}{8}\Big|_0^1 = \frac{1}{4}+\frac{1}{8} = \frac{3}{8}$$



(3) Find $P(X+Y \leq 1)$

$$=\int_0^1\int_0^{1-x} x+y\,dy\,dx$$

$$=\int_0^1 xy+\frac{y^2}{2}\Big|_0^{1-x}\,dx$$

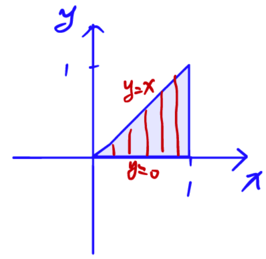


$$=\int_0^1 x(1-x)+\frac{(1-x)^2}{2}\,dx = \int_0^1 \frac{1}{2}-\frac{1}{2}x^2\,dx = \frac{1}{2}x-\frac{x^3}{6}\Big|_0^1 = \frac{1}{2}-\frac{1}{6} = \frac{1}{3}.$$

Example 3. Suppose the **pdf** function is $f_{X,Y}(x,y) = cxy$ for $0 \leq x \leq 1$, $0 \leq y \leq 1$ and $y < x$.

(1) Find c.

$$1 = \int_0^1 \int_0^x cxy\,dy\,dx = c\int_0^1 \frac{xy^2}{2}\Big|_0^x dx$$
$$= c\int_0^1 \frac{x^3}{2}\,dx$$
$$= c\cdot \frac{x^4}{8}\Big|_0^1 = \frac{c}{8} \quad \Rightarrow c = 8$$



(2) Find $P(X > 1/2)$

$$= \int_{\frac{1}{2}}^1 \int_0^x 8xy\,dy\,dx = \int_{\frac{1}{2}}^1 4xy^2\Big|_0^x dy = \int_{\frac{1}{2}}^1 4x^3\,dx$$
$$= x^4\Big|_{\frac{1}{2}}^1 = 1-\left(\frac{1}{2}\right)^6 = \frac{15}{16}$$

(3) Find marginal pdf of X and Y.

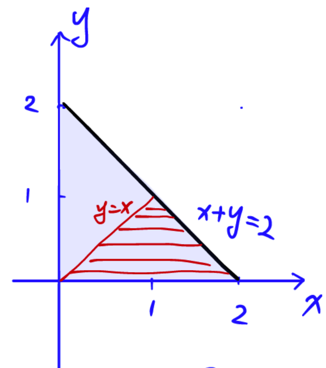
$$f_X(x) = \int_0^x 8xy\,dy = 4xy^2\Big|_0^x = 4x^3 \qquad 0 \le x \le 1$$

$$f_Y(y) = \int_{x=y}^{x=1} 8xy\,dx = 4x^2y\Big|_y^1 = 4y - 4y^3 \qquad 0 \le y \le 1$$

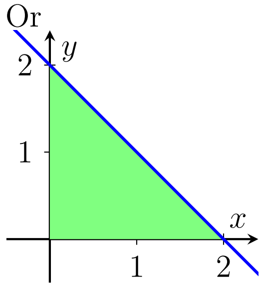
Example 4. (Practice at home.) Suppose the **pdf** function is $f_{X,Y}(x,y) = c(x+y)$ for $0 \le x \le 2$, $0 \le y \le 2$ and $x + y \le 2$.

(1) Find c.

$$1=\int_0^2\int_0^{2-x} c(x+y)\,dy\,dx$$
$$= c\int_0^2 xy+\frac{y^2}{2}\Big|_0^{2-x} dx$$
$$= c\int_0^2 x(2-x)+\frac{(2-x)^2}{2}\,dx = c\int_0^2 2-\frac{x^2}{2}\,dx = c\left(2x-\frac{x^3}{6}\right)\Big|_0^2 = \frac{8c}{3}$$
$$\Rightarrow\ c=\frac{3}{8}$$



Or



The marginal pdf of Y is

$$1=\int_0^2\int_0^{2-y}\frac{3}{8}(x+y)dxdy=\int_0^2\left[c(\frac{x^2}{2}+xy)\right]_0^{2-y}dy$$
$$=\int_0^2\frac{c}{2}(4-y^2)dy=\left[\frac{c}{2}(4y-y^3/3)\right]_0^2=\frac{8c}{3}$$

So, $c=\dfrac{3}{8}$

(2) Find $P(Y<X)$

$$=\int_0^1\int_{x=y}^{x=2-y}\frac{3}{8}(x+y)\,dx\,dy$$
$$=\frac{3}{8}\int_0^1 \frac{x^2}{2}+xy\Big|_y^{2-y} dy$$
$$=\frac{3}{8}\int_0^1 \frac{(2-y)^2}{2}+(2-y)y-\frac{y^2}{2}-y^2\,dy = \cdots$$

(3) Find marginal pdf of X and Y.

$$f_X(x) = \int_0^{2-x} \frac{3}{8}(x+y)\,dy = \frac{3}{8}\left(xy + \frac{y^2}{2}\right)\Big|_0^{2-x} = \frac{3}{8}\left(x(2-x) + \frac{(2-x)^2}{2}\right) = \frac{3}{8}\left(2 - \frac{x^2}{2}\right)$$

$$f_Y(y) = \int_0^{2-y} \frac{3}{8}(x+y)\,dx = \frac{3}{8}\left(\frac{x^2}{2} + xy\right)\Big|_0^{2-y} = \frac{3}{8}\left(2 - \frac{y^2}{2}\right)$$

Two random variables X and Y are called **independent** if and only if $f_{X,Y}(x,y) = f_X(x)f_Y(y)$.

If the pdf function $f_{X,Y}(x,y) = c$ in region R is constant, it is called bivariate **uniform** density. (HW3.7.10) This means all points are equally likely.

Example 5. Two **independent** random variables X and Y both have **uniform** distributions: X is uniform on $[0, 20]$, Y is uniform on $[5, 10]$.

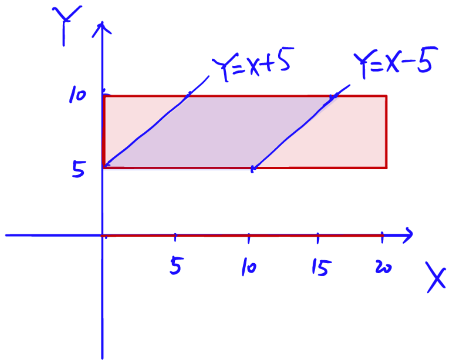
(1) Find the joint pdf $f_{X,Y}(x,y)$ for X and Y.

$f_X(x) = c$ s.t. $\int_0^{20} c\,dx = 1 \Rightarrow c = \frac{1}{20}$

$f_Y(y) = b$ s.t. $\int_5^{10} b\,dx = 1 \Rightarrow d = \frac{1}{5}$

$f_{XY}(x,y) = f_X(x)f_Y(y) = \frac{1}{20}\left(\frac{1}{5}\right) = \frac{1}{100}$

(2) Find the probability that $|X - Y| \leq 5$.



$X - Y \leq 5 \Rightarrow Y \geq X - 5$

$X - Y \geq -5 \Rightarrow Y \leq X + 5$

$$\text{probability} = \frac{\text{blue area}}{\text{red area}} = \frac{10 \times 5}{20 \times 5} = \frac{1}{2}$$

Method 2: $P(|X - Y| \leq 5) = \int_5^{10} \int_{y-5}^{y+5} \frac{1}{100} dx dy = \int_5^{10} \frac{1}{100}(10) dy = 0.5$

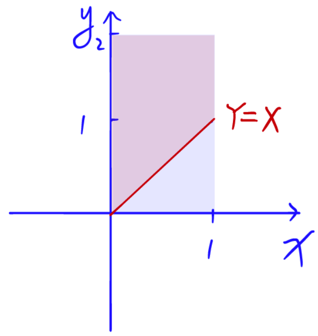
Theorem.

Two continuous random variables X and Y are **independent** if and only if there are functions $g(x)$ and $h(x)$ such that
$f_{X,Y}(x, y) = g(x)h(y), \quad f_X(x) = g(x) \;$ and $\; f_Y(y) = h(y)$.

It is easy to see that the pdfs in Example 2 and 4 are not independent.

Example 6. Suppose two random variables X and Y are independent and $f_X(x) = 3x^2$ for $0 \leq x \leq 1$ and $f_Y(y) = \frac{1}{2}y$ for $0 \leq y \leq 2$. Find $P(Y > X)$.

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y) = \frac{3}{2}x^2 y$$



$$P(Y>X) = \int_0^1 \int_x^2 \frac{3}{2}x^2 y \, dy \, dx$$

$$= \int_0^1 \left. \frac{3x^2y^2}{4} \right|_x^2 dx$$

$$= \int_0^1 3x^2 - \frac{3x^4}{4} \, dx$$

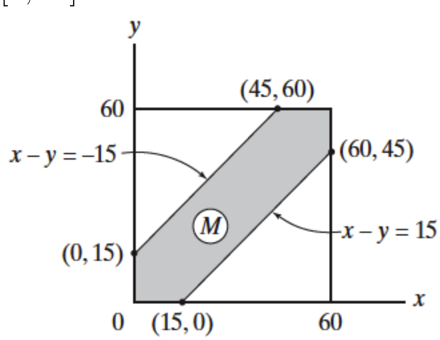
$$= \left. x^3 - \frac{3}{20}x^5 \right|_0^1$$

$$= 1 - \frac{3}{20}$$

$$= \frac{17}{20}$$

Example 7. Two friends agree to meet on Curry student center "sometime around 12:30pm". Each will arrive at random sometime from 12pm to 1pm. If one arrives and the other is not there, the first person will wait 15 minutes or until 1pm, whichever comes first, and then leave. What is the probability that the two will get together?

Let x and y denote the two arrival times. Two random variable are uniform distributions on $[0, 60]$. The two friends meet each other if and only if $|x-y| \leq 15$. That is $-15 \leq x-y \leq 15$.



So

$$P(\text{Meet}) = P(|X-Y| \leq 15) = \frac{\text{area M}}{60^2} = \frac{60^2 - 55^2}{60^2} = 0.44$$

§3.9 Further Properties of the Mean and the Variance

Theorem.

$$E(aX + bY) = aE(X) + bE(Y)$$

for **any** two random variables X and Y and numbers a and b.

Proof of continuous random variables:

$$E(aX + bY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (ax + by) f_{X,Y}(x, y) dxdy$$

$$= a \int_{-\infty}^{\infty} x f_X(x)dx + by \int_{-\infty}^{\infty} f_Y(y)dy = aE(X) + bE(Y).$$

The above formula can be generalized to more random variables.

The **expected value** of XY is

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy f_{XY}(x, y)dxdy$$

$$E(XY) = \sum_{all x,y} xy p_{XY}(x, y)dxdy$$

Recall that X and Y are **independent** if and only if $f_{XY}(x, y) = f_X(x)f_Y(y)$.

If X and Y are **independent**, then $E(XY) = E(X)E(Y)$.

The other direct is not true in general. (See homework 3.9.16)

Theorem.

If X and Y are **independent**, then

$$\operatorname{Var}(aX + bY) = a^2 \operatorname{Var}(X) + b^2 \operatorname{Var}(Y)$$

Proof.

$$\begin{aligned}\operatorname{Var}(aX+bY) &= E((aX+bY)^2) - (E(aX+bY))^2\\ &= E((a^2X^2+2abXY+b^2Y^2) - (aE(X)+bE(Y))^2\\ &= a^2E(X^2)+2abE(XY)+b^2E(Y^2) - a^2E(X)^2 - 2abE(X)E(Y) - b^2E(Y)^2\\ &= a^2\operatorname{Var}(X)+b^2\operatorname{Var}(Y)+2ab(E(XY)-E(X)E(Y))\end{aligned}$$

Since X and Y are independent, then $E(XY) - E(X)E(Y)$. Hence, $\operatorname{Var}(aX+bY) = a^2\operatorname{Var}(X)+b^2\operatorname{Var}(Y)$.

In general, we have the following theorem.

Theorem.

$$\operatorname{Var}(aX+bY) = a^2\operatorname{Var}(X)+b^2\operatorname{Var}(Y)+2ab\operatorname{Cov}(X,Y)$$

Here, $\operatorname{Cov}(X,Y)$ is the **covariance** of X and Y defined as

$$\operatorname{Cov}(X,Y) := E(XY) - E(X)E(Y)$$

Clearly, if X and Y are **independent**, then $\operatorname{Cov}(X,Y) = 0$ and hence $E(XY) = E(X)E(Y)$.

Example 1. Consider the experiment in homework 3.7.8 and 3.7.17.: Toss a fair coin 3 times. Let X denote the number of heads on the last flip. and let Y denote the total number of heads on the three flips. We already have the joint pdf $p_{X,Y}(x,y)$ given by

(x,y)	$p_{X,Y}$
(0,0)	1/8
(0,1)	2/8
(0,2)	1/8
(0,3)	0
(1,0)	0
(1,1)	1/8
(1,2)	2/8
(1,3)	1/8

or as

$y \backslash x$	0	1
0	1/8	0
1	2/8	1/8
2	1/8	2/8
3	0	1/8

(0)(3.7.17) Find the marginal pdfs of X and Y.

(1) Find the **mean** for X, Y, X^2, Y^2 and XY.

$E(X) = \frac{1}{2}$ $\quad E(Y) = \frac{3}{8} + 2\left(\frac{3}{8}\right) + 3\left(\frac{1}{8}\right) = \frac{12}{8} = \frac{3}{2}$

$E(X^2) = \frac{1}{2}$ $\quad E(Y^2) = \frac{3}{8} + 4\left(\frac{3}{8}\right) + 9\left(\frac{1}{8}\right) = \frac{24}{8} = 3$

$E(XY) = 1\left(\frac{1}{8}\right) + 2\left(\frac{2}{8}\right) + 3\left(\frac{1}{8}\right) = \frac{8}{8} = 1$

$Var(X) = E(X^2) - (E(X))^2 = \frac{1}{2} - \left(\frac{1}{2}\right)^2 = \frac{1}{4}$ $\quad Var(Y) = 3 - \left(\frac{3}{2}\right)^2 = \frac{3}{4}$

(2) Find the covariance of X and Y.

$Cov(X, Y) = E(XY) - E(X)E(Y) = 1 - \left(\frac{1}{2}\right)\left(\frac{3}{2}\right) = \frac{1}{4}$

(3) Find the variance of $X + Y$.

$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y) = \frac{1}{4} + \frac{3}{4} + 2\left(\frac{1}{4}\right) = \frac{3}{2}$

Suppose random variables $X_1, X_2, \cdots, X_n$ are independent and each X_i has the same distribution. Consider the **sample sum**

$$X = X_1 + X_2 + \cdots + X_n$$

Example 2. In binomial distribution, we consider a series of n **independent** trials with **two** outcomes. The probability of "success" for each trial is **constant** p. Let X be the number of successes in n trails.

Let X_i be the number of success in the i-th trial. So, $X = X_1 + X_2 + \cdots + X_n$. The **pdf** of X_i is called **Bernoulli** distribution:

$X_i = k$	0	1
$P(X_i = k)$	$1-p$	p

Clearly, $E(X_i) = p$. Then

$E(X) = E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n) = np.$

The variance of each X_i is $\text{Var}(X_i) = p - p^2$. So,

$\text{Var}(X) = \text{Var}(X_1 + X_2 + \cdots + X_n) = np(1 - p).$

Another important variable is the **sample mean**

$$\overline{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

As in our example, $E(\overline{X}) = p$ and $\text{Var}(\overline{X}) = \frac{1}{n}p(1 - p)$. This says the error gets smaller as the sample increases.

The covariance $\text{Cov}(X, Y)$ measures the association between X and Y. Another measure is the **correlation** of X and Y

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\,\text{Var}(Y)}}$$

The correlation is the normalized version of covariance, $-1 \le \text{Corr}(X, Y) \le 1$.

In Example 1, the correlation of X and Y is

$$\text{Corr}(X, Y) = \frac{1/4}{\sqrt{(\frac{1}{4})(\frac{3}{4})}} = \frac{1}{\sqrt{3}}$$

More question:

Example 3. Suppose you attend the high school graduation party with 200 people each wearing the same hat. All people throw the hats into the center of the room and then each person randomly select a hat. (1) Find the probability you selects your own hat. (2) Find the expected number of people who select their own hats.

Let $X_i = \begin{cases} 1 & \text{if } i\text{-th person select their own hat} \\ 0 & \text{other wise} \end{cases}$

pdf for X_i

$X_i=k$	0	1
prob.	$1-\frac{1}{200}$	$\frac{1}{200}$

$\Rightarrow E(X_i) = \frac{1}{200}$

Let X be the number of people who select their own hats.

$X = X_1 + X_2 + \cdots + X_{200}$

$E(X) = E(X_1) + E(X_2) + \cdots + E(X_{200}) = \frac{200}{200} = 1$

Warning: X is not binomial since X_i are not independent.

Example 4. Go back to the pickup phone example in §3.5. The pdf for X is

$X = k$	0	1	2	3	4
$P(X = k)$	9/24	8/24	6/24	0	1/24

What is the variance of X?

$$E(X^2) = \frac{8}{24}(1^2) + \frac{6}{24}(2^2) + \frac{1}{24}(4^2) = 2$$

$$\text{Var}(X) = E(X^2) - (E(X))^2 = 2 - 1 = 1$$

We know that variance of X_i is $\text{Var}(X_i) = \frac{1}{4} - (\frac{1}{4})^2 = \frac{3}{16}$.