L1&II : Introduction to Probability I & II

Distributive laws : $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ $\quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

DeMorgan's laws : $\overline{A_1 \cup A_2 \cup \ldots \cup A_n} = \bar{A}_1 \cap \bar{A}_2 \cap \ldots \cap \bar{A}_n$ $\quad \overline{A_1 \cap A_2 \cap \ldots \cap A_n} = \bar{A}_1 \cup \bar{A}_2 \cup \ldots \cup \bar{A}_n$

Axioms of probability : $P(\mathcal{S}) = 1.$ $\quad 0 \le P(A) \le 1.$ $\quad P(A \cup B) = P(A) + P(B).$

$P(\bar{A}) = 1 - P(A)$ $\quad P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$

The addition rule : $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

In general, if $A_1, \ldots, A_n$ are mutually exclusive events, such that $A_1 \cup A_2 \cup \ldots \cup A_n = \mathcal{S}$, then

Law of total probabilities $P(B) = P(A \cap B) + P(\bar{A} \cap B)$ $\quad P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \ldots + P(A_n \cap B)$

Conditional probability $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$

Independent events $P(A|B) = P(A)$, $P(B|A) = P(B)$ $\quad P(A \cap B) = P(A) \times P(B) \leftarrow$

The Bayes' rule : $P(A|B) = \dfrac{P(B|A)P(A)}{P(B)}$

$$P(B) = P(A \cap B) + P(A^c \cap B) = P(B|A)P(A) + P(B|A^c)P(A^c)$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

Example
Customers who purchase a certain make of car can order an engine in any of three sizes. Of all cars sold, 45% have the smallest engine, 35% have the medium-size and 20% have the largest one.

Of cars with the smallest engine 10% fail an emission test within two years of purchase, while only 12% of those with the medium-size and 15% of those with the largest engine fail the test.

What is the probability that a randomly selected car will fail an emissions test within two years?

~~test within two years?~~ P(B) ??

LET B = "CAR FAILS AN EMISSIONS TEST WITHIN 2 YEARS"
A_1 = "CAR HAS A SMALL ENGINE"
A_2 = "CAR HAS A MEDIUM-SIZE ENGINE"
A_3 = "CAR HAS A LARGE ENGINE"

THEN, $P(A_1) = 0.45$; $P(A_2) = 0.35$; $P(A_3) = 0.20$

ALSO, $P(B|A_1) = 0.10$; $P(B|A_2) = 0.12$; $P(B|A_3) = 0.15$.

WE HAVE

$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B)$, TOTAL PROBABILITIES
$= P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3)$, CONDITIONAL PROBABILITIES
$= (0.10)(0.45) + (0.12)(0.35) + (0.15)(0.20)$
$= 0.117$

L2 : Counting Techniques

Number of possible arrangements of size r out of n objects

	Without Replacement	With Replacement
P ← Ordered	$\frac{n!}{(n-r)!}$	n^r
C ← Unordered	$\binom{n}{r} = \frac{n!}{(n-r)!\,r!}$	$\binom{n+r-1}{r}$

L3 : Additional Examples

Example (Rolling a die)
Consider a gambler interested in the event that he could throw at least 1 six in 4 rolls of a die.

We have

$$P(\text{at least 1 six in 4 rolls}) = 1 - P(\text{no six in 4 rolls}) = 1 - \prod_{i=1}^{4} P(\text{no six in roll } i),$$

where for the last equality we are assuming independence. Now, on each roll, the probability of *not* getting a six is 5/6, so we get

$$P(\text{at least 1 six in 4 rolls}) = 1 - \left(\frac{5}{6}\right)^4 = 0.518$$

For this example we will consider *unordered outcomes*, so the sample space consists of all the 5-card hands that can be chosen from the 52-card deck. There are

$$\binom{52}{5} = 2,598,960$$

First, we need to figure out how many different hands are there with four aces. To do this, we notice that in order to have four aces in a five-card hand, then we have $52 - 4 = 48$ ways to chose the fifth card.

It follows that

$$P(\text{four aces}) = \frac{48}{2,598,960}.$$

$$P(\text{four of a kind}) = \frac{13 \times 48}{2,598,960} = \frac{624}{2,598,960}$$

Example (Independence of more than two events)
It might be tempting to say three events A, B, C are independent if $P(A \cap B \cap C) = P(A)P(B)P(C)$, however, we must be careful as this is not the case.

For instance, consider an experiment that consists of rolling two dice. The sample space for this experiment is

$$\mathcal{S} = \{(1,1),(1,2),\ldots,(1,6),(2,1),\ldots,(2,6),\ldots,(6,1),\ldots,(6,6)\},$$

where $\mathcal{S}$ is made of the 36 ordered pairs formed from the numbers 1 to 6.

Now, define the following events of interest
- $A = \{\text{doubles appear}\} = \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}$
- $B = \{\text{the sum is between 7 and 10}\}$
- $C = \{\text{the sum is 2 or 7 or 8}\}$

btain

$$P(A) = \frac{1}{6}, \quad P(B) = \frac{1}{2}, \quad P(C) = \frac{1}{3}.$$

$$\begin{aligned} P(A \cap B \cap C) &= P(\text{the sum is 8 composed of two 4s}) \\ &= \frac{1}{36} \\ &= \frac{1}{6} \times \frac{1}{2} \times \frac{1}{3} \\ &= P(A)P(B)P(C). \end{aligned}$$

L4 : Random Variables

Discrete random variables $\quad$ 1. $0 \le P(x) \le 1$ $\quad$ 2. $\sum_x P(x) = 1$

Expected value $E(X) = \sum_x x \cdot p(x)$ $\quad \mu = E(X)$

standard deviation

The population variance $Var(X) = \sum_x (x - \mu)^2 \cdot p(x) = \sigma^2$ $\quad \sigma = \sqrt{\sum (x - \mu)^2 \cdot p(x)}$

Example
All students in a class were asked how many times they had read the city newspaper in the past 5 days. The following table summarizes the results.

# Times read newspaper	Percentage
0	25%
1	5%
2	10%
3	10%
4	15%
5	35%

$$\mu = \sum x \cdot p(x) = 0 \times 0.25 + 1 \times 0.05 + \ldots + 5 \times 0.35 = 2.9$$

$$\begin{aligned} \sigma &= \sqrt{\sum (x-\mu)^2 \cdot p(x)} \\ &= \sqrt{(0-2.9)^2 \times 0.25 + \ldots + (5-2.9)^2 \times 0.35} \\ &= \sqrt{4.09} = 2.02 \end{aligned}$$

L5 : Expected Values

Properties of the expected value

1. For any constant c, $E(c) = c$.
2. For any constant c and function $g(X)$, such that $E[g(X)]$ exists,
$$E[cg(X)] = cE[g(X)].$$
3. For any functions $g_1(X), g_2(X), \ldots, g_k(X)$, such that their expectations exist,
$$E[g_1(X) + g_2(X) + \ldots + g_k(X)] = E[g_1(X)] + E[g_2(X)] + \ldots + E[g_k(X)].$$

Theorem (Tchebysheff's Theorem)

Let X be a random variable with mean μ and variance σ^2 (finite). Then, for any constant $k > 0$,

$$P(|X - \mu| < k\sigma) \ge 1 - \frac{1}{k^2},$$

or equivalently,

$$P(|X - \mu| \ge k\sigma) \le \frac{1}{k^2}$$

Example
The number of costumers per day at a sales counter Y has been observed for a long period of time and found to have a mean 20 and standard deviation 2.

If the probability distribution of Y is unknown, what can be said about the probability that on a given day Y will be between 16 and 24?

Observe that we want to find $P(16 < Y < 24)$, so from the Tchebysheff's inequality we have that for any $k > 0$,

$$P(|Y - \mu| < k\sigma) = P((\mu - k\sigma) < Y < (\mu + k\sigma)) \ge 1 - \frac{1}{k^2}$$

Then, for $\mu = 20$ and $\sigma = 2$, we have that $\mu - k\sigma = 16$ and $\mu + k\sigma = 24$ when $k = 2$.

It follows that,

$$P(|Y - \mu| < k\sigma) = P((\mu - 2\sigma) < Y < (\mu + 2\sigma)) \ge 1 - \frac{1}{2^2} = \frac{3}{4}$$

So the number of costumers at the sales counter on a given day is between 16 and 24 with a probability of (at least) 0.75.

Observe that if instead we had that $\sigma = 1$, then k would be equal to 4 and we would obtain

$$P(|Y - \mu| < k\sigma) \ge \frac{15}{16}$$

Definition
Let X be a random variable with mean $E(X) = \mu$. The, the *variance* $V(X)$ of the random variable X is defined as the expected value of $(X - \mu)^2$. That is,

$$V(X) = E[(X - \mu)^2].$$

The *standard deviation* of X is defined as the positive square root if $V(X)$.

When $p(x)$ represents the probability distribution of a population of interest, we will call $\mu = E(X)$ as the *population mean*, $\sigma^2 = V(X)$ the *population variance* and $\sigma = \sqrt{V(X)}$.

L6 Discrete Distributions

The Binomial distribution

A *Binomial experiment* is defined by the following conditions: $\quad X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases}$

1. The experiment consists of a fixed number n of identical trials.
2. Each trial results in one of two outcomes: success S, or failure F.
3. The probability of success on a single trial is equal to some value p which remains the same from trial to trial. Likewise, the probability of failure is equal to $q = (1 - p)$
4. The trials are independent.
5. The random variable of interest is defined as $X =$"the total number of successes observed in the n trials".

Definition
A random variable X is said to have a *binomial distribution* with n trials and success probability p if and only if

$$P(X = x) = p(x) = \binom{n}{x} p^x q^{(n-x)},$$

where $x = 0, 1, 2, \ldots, n$ and $0 \le p \le 1$. We write $X \sim$ binomial(n, p).

The mean $E(X) = np$. **variance** $V(X) = np(1-p) = npq.$

Example
A large employee pool (more than 1000 people) that can be used for selecting employees for management training program is half male and half female.

Suppose that since the program began, none of the 10 employees chosen for management training were female.

What would be the probability of 0 females in 10 selections, if there truly were no gender bias?

Since the proportion of male employees and the proportion of female employees are equal, the probability of randomly selecting a female is 0.5.

If X represents the number of female employees selected for management training in a random sample of 10 employees, we have:

$$P(X = 0) = \binom{10}{0}(0.50)^0(0.50)^{10} = 0.001$$

Read the problems carefully,
- "at least" means "$\ge$"
- "more than" means "$>$"
- "at most" means "$\le$"
- "less than" means "$<$"

L7 : Discrete Distributions - Cont'd

The Geometric distribution

Definition
A random variable X is said to have a Geometric distribution with success probability p, if and only if

$$p(x) = q^{x-1}p,$$

where $x = 1, 2, 3, \ldots$ and $0 \le p \le 1$. We write $X \sim$ geometric(p).

The mean $E(X) = \mu = \dfrac{1}{p}$. $\quad V(X) = \dfrac{1-p}{p^2} = \sigma$

The Negative Binomial distribution

Definition
A random variable X is said to have a Negative Binomial distribution if and only if

$$p(x) = \binom{x-1}{r-1} p^r q^{x-r}, x = r, r+1, , r+2, \ldots,$$

for a given integer r and $0 \le p \le 1$. We write $X \sim$ NegBin(r, p).

$$E(X) = \frac{r}{p} \text{ and } V(X) = \frac{r(1-p)}{p^2}.$$

Example

A geological study indicates that an exploratory oil well drilled in a particular region should strike oil with probability 0.2. What is the probability that the third oil strike comes on the fifth well drilled?

Assuming that the drillings are independent, we can define $X =$ "the number of the trial in which the third oil strike occurs" . Then, using the negative binomial distribution $r = 3$ and $p = 0.2$ we obtain

$$P(X=5) = p(5) = \binom{5-1}{3-1}(0.2)^3(0.8)^2 = 0.0307.$$

The Hypergeometric distribution

Definition

A random variable X is said to have a hypergeometric distribution if and only if

$$p(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}},$$

where $x = 0, 1, 2, \ldots, r$, $n \le N$ and $n - x \le N - r$.

$$E(X) = \frac{nr}{N} \text{ and } V(X) = n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right)$$

Example

Suppose that from a group 20 engineers applying to a position we need to hire only ten. The resumes of the applicants look similar in every way, so decide to make the selection at random.

What is the probability that our random selection includes all 5 engineers that are best qualified for the job in the group of applicants?

For example we are implicitly assuming that $N = 20$, $n = 10$ and $r = 5$, *i.e.* there are only 5 engineers in the group of 20 that are best qualified for the job.

If X denotes the number of best engineers among the 10 selected, we obtain

$$P(X=5) = p(5) = \frac{\binom{5}{5}\binom{15}{5}}{\binom{20}{10}} = \frac{21}{1292} = 0.0162.$$

L8: The Poisson Distribution

then X, the number of events in a fixed unit of time

$$X \sim \text{Poisson}(\lambda): \; p(x) = \frac{\lambda^x}{x!}e^{-\lambda};$$

large ~ np → small

$$E(X) = \lambda = V(X) = \sigma^2$$

Properties of a cdf

If $F(x)$ is a cumulative distribution function, then it satisfies the following properties:

1. $F(-\infty) = \lim_{x\to-\infty} F(x) = 0$
2. $F(\infty) = \lim_{x\to\infty} F(x) = 1$
3. $F(x)$ is non-decreasing in x. That is, if x_1, x_2 are any two values such that $x_1 < x_2$, then $F(x_1) \le F(x_2)$.

L9: Moment Generating Functions

mgf: $m(t) = E(e^{tX})$ $m(t)$ is finite for every $|t| < b$

Definition

The kth moment μ_k' of a random variable X is defined as

$$\mu_k' = E(X^k),$$

provided that the expectation exists.

Likewise, the kth central moment μ_k of a random variable X is defined as

$$\mu_k = E[(X-\mu)^k],$$

where $\mu = E(X)$, provided the expectations exist.

Theorem

If the moment generating function $m(t)$ exists, then for any positive integer k,

$$\left.\frac{d^k m(t)}{dt}\right|_{t=0} = m^{(k)}(0) = \mu_k'.$$

In words, the kth moment μ_k' is found as the kth derivative of $m(t)$ with respect to t, evaluated at $t = 0$.

L10. Continuous Random Variables

The cumulative distribution function (CDF)

$F(x) = P(X \le x)$, for $-\infty < x < \infty$.

The probability density function $\quad F(x) = \int_{-\infty}^{x} f(t)dt.$

Properties of a pdf

1. $f(x) \ge 0$ for all $x \in (-\infty, \infty)$.
2. $\int_{-\infty}^{\infty} f(x)dx = 1$

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

3

x	1	2	3	4	5 or above
$p(x)$	0.3	0.3	?	0.2	0

1. Determine $P(X=3)$. $P(X=3) = 0.2$
2. Find $P(X>1|X<4) = \frac{P(X>1, X<4)}{P(X<4)} = \frac{P(X=2)+P(X=3)}{P(X=1)+P(X=2)+P(X=3)}$
3. Find the variance of X. $\mathrm{Var}(X) = \mathrm{Exp}[X^2] - \mathrm{Exp}[X]^2$
4. Find the expected value of $1/X$. $\mathrm{Exp}\left[\frac{1}{X}\right] = \left(\frac{1}{1}\right)(0.3) + \left(\frac{1}{2}\right)(0.3) + \left(\frac{1}{3}\right)(0.2) + \left(\frac{1}{4}\right)(0.2) = 0.567$

Problem 4. A fair die has six faces, numbered 1 through 6. When it is rolled, all faces are equally likely to turn up. The die is rolled 5 times.

$P(\text{two or more the same}) = 1 - P(\text{none the same}) = 1 - 6!/6^5 = 0.907$

2. What is the probability that the number 3 has turned up at least twice in the first 5 rolls?

$$P(\text{3 at least twice}) = 1 - P(\text{3 zero times or one time})$$
$$= 1 - 2 \times (\tfrac{5}{6})^5 = 0.1962.$$

3. Now suppose you roll it a sixth time. What is the probability that it gives you a new number that had never turned up in the first 5 rolls? $\frac{6\times5^5}{6^6} = 0.4019$

4. The die is rolled until all 6 numbers have each turned up at least once. Let T be the number of rolls needed. Find $P(T=7)$. $\quad P(T=7) = \frac{6\times5\times\binom{6}{2}\times4!}{6^7} = 0.0386$

Problem 5. Y is a <u>continuous</u> random variable with c.d.f. F

1. $F(2) \ge F(1)$. 2. $f(2) \ge f(1)$. 3. $P(Y \ge 1) > P(Y > 1)$
4. $0 \le f(y) \le 1$. 5. $E(2Y) = 2E(Y)$. 6. $\mathrm{Var}(2Y) = 2\,\mathrm{Var}(Y)$.

Problem 6. Suppose the lifetime (in years) of a computer hard disk is modeled as a random variable Y with p.d.f. given by

$$f(y) = \begin{cases} cy, & 0 < y < 6, \\ 0. & \text{elsewhere,} \end{cases}$$

1. Find c. $1 = c\int_0^6 y dy = 1/18.$
2. Suppose the disk has operated for 3 years. What is the (conditional) probability that it will fail within the next (i.e. the fourth) year? $P(Y \le 4|Y \ge 3) = \frac{P(3 \le Y \le 4)}{P(Y \ge 3)} = \frac{\int_3^4 f(y)dy}{\int_3^6 f(y)dy}$
3. Suppose two such disks are used in an array: the array will operate if at least one disk operates (will fail if both disks fail). Assuming that the two disks operate independently, find the probability that the array will fail within 2 years.

$$P(Y_1 < 2, Y_2 < 2) = P(Y_1 < 2)P(Y_2 < 2) = \left[\frac{1}{18}\int_0^2 y_1 dy_1\right] \times \left[\frac{1}{18}\int_0^2 y_2 dy_2\right] \text{ (by indep.)}$$

Problem 2. You ask your neighbor to water a plant while you are on vacation. Without water, it will die with probability 0.8; with water, it will die with probability 0.1. You are 90% certain that your neighbor will remember to water it.

1. What is the probability that the plant will be alive when you return?

$P(\text{alive}) = P(\text{alive}|\text{water})\,P(\text{water}) + P(\text{alive}|\text{no water})\,P(\text{no water}) = (0.9)(0.9) + (0.2)(0.1) = 0.83$

2. If the plant is dead when you return, what is the conditional probability that your neighbor forgot to water it? $\quad P(\text{forgot}|\text{dead}) = \frac{P(\text{forgot} \cap \text{dead})}{P(\text{dead})}$

$$P(\text{forgot} \cap \text{dead}) = P(\text{dead}|\text{forgot})\,P(\text{forgot})$$
$$P(\text{dead}) = 1 - P(\text{alive})$$
$$\Longrightarrow P(\text{forgot}|\text{dead}) = \frac{(0.8)(0.1)}{1-0.83} = 0.471$$

Problem 1.

(a) $P(A \text{ occurs but } B \text{ does not occur}) = P(A) - P(B)$.

(b) $P(A \cup B) \ge P(A \cap B)$.

(c) If A and B are independent, then $P(A \cap \bar{B}) = P(A)\,P(\bar{B})$.

(d) If A and B are independent, then $P(A \cup B) = P(A) + P(B)$.