Statistical Methods: Lecture 3

Lecture Overview

Remark. These topics are treated more extensively than in the book (necessary for assignments and exam).

3.8 Law of Total Probability and Bayes theorem

Recall: conditional probability $P(B|A) = \frac{P(A \cap B)}{P(A)}$.

Selfies from a famous celebrity

40% posted only on Instagram (Inst), the rest both on Instagram and Facebook (FB).
70% from only Inst, and 90% posted on both Inst and FB: appear on gossip website.

What is the probability that a random selfie appears on a gossip website?

Let $B = \{\text{selfie on gossip website}\}$, $A = \{\text{selfie on Inst}\}$, $\bar{A} = \{\text{selfie on Inst and FB}\}$.

Compute $P(B)$: we know $P(B|A) = 0.7$, $P(B|\bar{A}) = 0.9$, $P(A) = 0.4$, $P(\bar{A}) = 0.6$.

- Selfie from Inst on gossip website

$$P(B \cap A) = P(B|A) \cdot P(A) = 0.7 \times 0.4 = 0.28$$

- Selfie from Inst and FB on gossip website

$$P(B \cap \bar{A}) = P(B|\bar{A}) \cdot P(\bar{A}) = 0.9 \times 0.6 = 0.54$$

Addition rule: $P(B) = 0.28 + 0.54 = 0.82$.

Law of Total Probability and 3.8 Bayes theorem

Addition rule for disjoint events ($A \cap B$ and $\bar{A} \cap B$):

$$P(B) = P(B \cap A) + P(B \cap \bar{A}).$$

Then, multiplication rule:

$$P(B) = P(B \cap A) + P(B \cap \bar{A}) = P(B|A) \cdot P(A) + P(B|\bar{A}) \cdot P(\bar{A}).$$

Simple law of total probability

Let A and B be events. Then

$$P(B) = P(B \cap A) + P(B \cap \bar{A}) = P(B|A) \cdot P(A) + P(B|\bar{A}) \cdot P(\bar{A}).$$

Law of Total Probability and 3.8 Bayes theorem

Combine with multiplication rule:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|\bar{A}) \cdot P(\bar{A})}.$$

Bayes' Theorem (simple)

Let A and B be events, then:

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|\bar{A}) \cdot P(\bar{A})}.$$

Caution:

$$P(B|A) + P(\bar{B}|A) = 1$$

But in general

$$P(B|A) + P(B|\bar{A}) \neq 1$$

Law of Total Probability and 3.8 Bayes theorem

Example: medical test for a disease

Suppose 0.1% of population has a disease. If someone has the disease, the medical test gives positive result with probability 0.98. If someone does not have the disease the medical test gives negative result with probability 0.99.

Suppose Alice conducts the test and the result is positive. What is the probability that Alice has the disease given the positive test outcome?

$B = \{\text{Positive}\}$ and $A = \{\text{Disease}\}$, interested in: $P(A|B)$.

Note $P(B|A) = 0.98$, but recall that $P(B|A) \neq P(A|B) \Rightarrow$, so use the Bayes theorem.

First compute $P(B|\bar{A})$, $P(A)$ and $P(\bar{A})$.

Denoting $\bar{A} = \{\text{no disease}\}$, compute $P(B|\bar{A}) = 0.01$, $P(A) = 0.001$ and $P(\bar{A}) = 1 - 0.001 = 0.999$.

$$\Rightarrow P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|\bar{A}) \cdot P(\bar{A})} = \frac{0.98 \cdot 0.001}{0.98 \cdot 0.001 + 0.01 \cdot 0.999} \approx 0.089.$$

Probability is only 8.9% that Alice has the disease!

Law of Total Probability and 3.8 Bayes theorem

Why did we get this counterintuitive result?

Consider following interpretation:

Suppose population: 100 000 people
and distribution of diseases and test results as follows:

	Positive	Negative	Total
Disease	98	2	100
No disease	999	98 901	99 900
Total	1 097	98 903	100 000

So, $98 + 999 = 1\ 097$ people with positive result, but thereof only 98 diseased!

$\Rightarrow$ fraction of people having disease when obtained positive test result: $\frac{98}{1097} \approx 0.089$.

Law of Total Probability and 3.8 Bayes theorem

Definition: Partition

Events $A_1, \ldots, A_m$ are called partition if

- pairwise disjoint: $A_i \cap A_j = \varnothing$, if $i \neq j$;
- union is entire sample space: $A_1 \cup A_2 \cup \ldots \cup A_m = \Omega$.

Let $A_1, \ldots, A_m$ be a partition, then also $B \cap A_1, \ldots, B \cap A_m$ disjoint. Then

$$\begin{aligned}
P(B) &= P(B \cap \Omega) = P(B \cap (A_1 \cup A_2 \cup \ldots \cup A_m)) \\
&= P((B \cap A_1) \cup (B \cap A_2) \cup \ldots \cup (B \cap A_m)) \\
&= \sum_{i=1}^{m} P(B \cap A_i) \quad \text{(General addition rule for disjoint events)} \\
&= \sum_{i=1}^{m} P(B|A_i) \cdot P(A_i). \quad \text{(multiplication rule)}
\end{aligned}$$

Law of total probability

Let $A_1, \ldots, A_m$ be a partition, then

$$P(B) = \sum_{i=1}^{m} P(B \cap A_i) = \sum_{i=1}^{m} P(B|A_i) \cdot P(A_i).$$

Law of Total Probability and 3.8 Bayes' theorem

Partition

A collection $A_1, \ldots, A_m$ of events is called a partition if

- They are pairwise disjoint: $A_i \cap A_j = \varnothing$, if $i \neq j$;
- The union is the entire sample space: $A_1 \cup A_2 \cup \ldots \cup A_m = \Omega$.

Law of total probability

Let $A_1, \ldots, A_m$ be a partition, then

$$P(B) = \sum_{i=1}^{m} P(B \cap A_i) = \sum_{i=1}^{m} P(B|A_i) \cdot P(A_i).$$

Now we can derive the Bayes theorem: let $A_1, \ldots, A_m$ be partition, then

$$P(A_r|B) = \frac{P(A_r \cap B)}{P(B)} = \frac{P(B|A_r) \cdot P(A_r)}{\sum_{i=1}^{m} P(B|A_i) \cdot P(A_i)}, \quad r = 1, \ldots, m.$$

Law of Total Probability and 3.8 Bayes theorem

Bayes theorem

Let $A_1, \ldots, A_m$ be partition, then for $r = 1, \ldots, m$

$$P(A_r|B) = \frac{P(B|A_r) \cdot P(A_r)}{\sum_{i=1}^{m} P(B|A_i) \cdot P(A_i)}.$$

Example: defect products in a factory

Machines 1, 2 and 3 produce resp. 30%, 45% and 25% of all products.
Resp. 2%, 3% and 2% thereof: defective. Suppose we selected a random product and it was defective. Probability that it came from Machine 2?

$A_i = \{\text{Machine } i \text{ made product}\}$, $B = \{\text{Product defect}\}$, so interested in $P(A_2|B)$.
We have $P(A_1) = 0.30, P(A_2) = 0.45$ and $P(A_3) = 0.25$. Furthermore,
$P(B|A_1) = 0.02, P(B|A_2) = 0.03$ and $P(B|A_3) = 0.02$. Hence,

$$P(B|A_1) \cdot P(A_1) + P(B|A_2) \cdot P(A_2) + P(B|A_3) \cdot P(A_3) = 0.0245.$$

$$P(A_2|B) = \frac{P(B|A_2) \cdot P(A_2)}{P(B|A_1) \cdot P(A_1) + P(B|A_2) \cdot P(A_2) + P(B|A_3) \cdot P(A_3)} = \frac{0.0135}{0.0245} \approx 0.55$$

4.2 Probability Distributions

Recall examples: sum of two dice / number of heads in three coin tosses.
Other real-life examples: height / income of random person.

We need a function of the outcomes.

Definition (Random variable)

A random variable is a variable that assigns a numerical value to each outcome of a probability experiment.
Notation: $X, Y, \ldots$

The book uses x both to denote a random variable AND values of random variable which is confusing. Here, X: random variable, x value of random variable.

Example: two coin tosses HH

Throw a fair coin twice. Let the random variable X be the number of heads.
Sample space: $\Omega = \{HH, HT, TH, TT\}$. Values of X for those outcomes:

$$X(HH) = 2, X(HT) = 1, X(TH) = 1, X(TT) = 0.$$

So X takes values 0, 1, 2. What is $P(X = 0) = ?, P(X = 1) = ?, P(X = 2) = ?$

4.2 Probability Distributions

Example: one die

Random variable $X =$ score on top

Sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$. Value of X for each outcome:

$$X(1) = 1, X(2) = 2, X(3) = 3, \dots, X(6) = 6.$$

So X takes values 1, 2, 3, 4, 5, 6.

What is $P(X = 1) = ?, P(X = 2) = ?, \dots, P(X = 6) = ?$

Example: Statistical Methods Favourite Colours

Select random person. Let random variable $X =$ favourite color of selected person.

Sample space is $\Omega = \{\text{all of us}\}$. Value of X for these outcomes is

$$X(\text{person}) = \text{favourite colour of person}, \;\; \text{for instance } X(\text{Sophia}) = \text{purple}.$$

So X takes values in all possible colours.

4.2 Probability Distributions

Definition (Probability distribution)

A probability distribution determines probabilities of values of a random variable.
Given by table, formula, or graph.

Definition (Discrete random variable)

A discrete random variable has finite (or countably) many different values. Its probability distribution: collection of all their individual probabilities.
Total sum of probabilities: 1.

Definition (A continuous random variable)

has uncountably many different values.
Its probability distribution: given by probability density function;
probabilities computed by area under this function.
Total area: 1.

For now: only discrete random variables (the first two examples).

4.2 Probability Distributions

Example: 2 coin tosses (fair) HH

Random variable X: number of heads.

$$\Rightarrow \quad X(HH) = 2,\; X(HT) = 1,\; X(TH) = 1,\; X(TT) = 0.$$

Aim: probability distribution of X: $\quad P(X = 0),\; P(X = 1),\; P(X = 2)$?

$$P(X = 0) = P(\{TT\}) = \frac{1}{4}, \qquad P(X = 1) = P(\{TH, HT\}) = \frac{2}{4} = \frac{1}{2},$$

$$P(X = 2) = P(\{HH\}) = \frac{1}{4}.$$

Presentation:

x	$P(X = x)$	num. $P(X = x)$
0	$1/4$	0.25
1	$1/2$	0.50
2	$1/4$	0.25

Check: $\quad P(X = 0) + P(X = 1) + P(X = 2) = \frac{1}{4} + \frac{1}{2} + \frac{1}{4} = 1.$

4.2 Probability Distributions

Outcomes ω in Ω and probability measure P determine probability distribution of X:

$$P(X = x) = P(\{\omega \in \Omega : X(\omega) = x\}).$$

Recipe: find probability distribution of discrete random variable

1. Determine sample space of underlying probability experiment and probabilities of outcomes ω (see Lecture 2).
2. List values $X(\omega)$ for all ω in Ω.
3. For each value x of X, find all simple events $\{\omega\}$ with value x. Unify: $\{X = x\} = \{\omega : X(\omega) = x\}$.
4. Probabilities $P(\{\omega\})$ determine probability of $\{X = x\}$:

$$P(X = x) = P(\{\omega : X(\omega) = x\}) = \sum_{\omega:\ X(\omega)=x} P(\{\omega\}).$$

5. Table. Left column: all values x of X and column with probabilities $P(X = x)$.

4.2 Probability Distributions

Example: X = maximum of two (fair) dice throughs

Probability distribution of X?

1. $\Omega = \{(1,1), \ldots, (1,6), (2,1), \ldots, (6,6)\}$ and $P(\{\omega\}) = \frac{1}{36}$ for all ω.
2. $$\begin{array}{llll} X((1,1)) = 1, & & & \\ X((1,2)) = 2, & X((2,2)) = 2, & X((2,1)) = 2, & \\ X((1,3)) = 3, & \ldots, & X((3,1)) = 3, & \\ \vdots & \vdots & \vdots, & \\ X((1,6)) = 6, & \ldots, & X((6,5)) = 6, & X((6,6)) = 6. \end{array}$$
3. Possible values of X: 1, 2, ..., 6 a d $\{X = 1\} = \{(1,1)\}$,
   $\{X = 2\} = \{(1,2),(2,2),(2,1)\}, \ldots, \{X = 6\} = \{(1,6), \ldots (6,6), \ldots, (6,1)\}$
4. $$\begin{aligned} P(X=1) &= P(\{(1,1)\}) = \tfrac{1}{36}, \\ P(X=2) &= P(\{(1,2),(2,2),(2,1)\}) = \tfrac{3}{36} = \tfrac{1}{12}, \\ P(X=3) &= P(\{(1,3),(2,3),(3,3),(3,2),(3,1)\}) = \tfrac{5}{36}, \\ P(X=4) &= P(\{(1,4),\ldots,(4,4),\ldots,(4,1)\}) = \tfrac{7}{36}, \\ P(X=5) &= P(\{(1,5),\ldots,(5,5),\ldots,(5,1)\}) = \tfrac{9}{36} = \tfrac{1}{4}, \\ P(X=6) &= P(\{(1,6),\ldots,(6,6),\ldots,(6,1)\}) = \tfrac{11}{36}, \end{aligned}$$

   Check: $\sum_{x=1}^{6} P(X = x) = \frac{1}{36} + \frac{3}{36} + \frac{5}{36} + \frac{7}{36} + \frac{9}{36} + \frac{11}{36} = 1.$

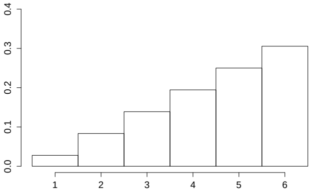
4.2 Probability Distributions

Example: maximum of two (fair) dice throughs

5. Tabulating results:

x	$P(X = x)$	num. $P(X = x)$
1	$1/36$	0.028
2	$1/12$	0.083
3	$5/36$	0.139
4	$7/36$	0.194
5	$1/4$	0.250
6	$11/36$	0.306

Also possible: present probability distribution graphically



4.2 Probability Distributions

Experiment vs. random variable

Experiment	Random variable
Possible outcomes of experiment	Possible values of random variable
Probability of outcome	Probability of value

Right column is determined by left column.

4.2 Probability Distributions

Example: 3 coin tosses (fair) HHH

How many H do you expect on average?

$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$.
Group outcomes with same amount of Heads:

$$
\begin{aligned}
P(X=0) &= P(\{TTT\}) = \frac{1}{8}, \\
P(X=1) &= P(\{HTT, THT, TTH\}) = \frac{3}{8}, \\
P(X=2) &= P(\{THH, HTH, HHT\}) = \frac{3}{8}, \\
P(X=3) &= P(\{HHH\}) = \frac{1}{8}.
\end{aligned}
$$

Expected value (expectation/mean): weighted average; weights are probabilities:

$$\mu = E(X) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \frac{12}{8} = 1.5$$

4.2 Probability Distributions

Definition (Expected value)

Expected value (expectation/mean) of a discrete random variable X with possible values $x_1, \ldots, x_k$, weighted average of all possible values of X:

$$\mu = E(X) = \sum_{i=1}^{k} x_i \cdot P(X = x_i).$$

NB: μ is not necessary a value of X.

Example: X = maximum of two (fair) dice throughs

What is $E(X)$? Probability distribution+weighted averages:

x	$P(X = x)$	num. $P(X = x)$	$x \cdot P(X = x)$
1	$1/36$	0.028	0.028
2	$1/12$	0.083	0.167
3	$5/36$	0.139	0.417
4	$7/36$	0.194	0.778
5	$1/4$	0.250	1.250
6	$11/36$	0.306	1.833

Thus $E(X) = \sum_{i=1}^{6} i \cdot P(X = i) \approx 4.472$.

4.2 Probability Distributions

Definition (Variance)

of a discrete random variable X with values $x_1, \ldots, x_k$:

$$\sigma^2 = Var(X) = E(X - EX)^2 = \sum_{i=1}^{k} \left[(x_i - \mu)^2 P(X = x_i)\right].$$

The standard deviation of X is

$$\sigma = SD(X) = \sqrt{Var(X)} = \sqrt{\sum_{i=1}^{k} [(x_i - \mu)^2 P(X = x_i)]}.$$

Remark. Alternative formula $Var(X) = EX^2 - (EX)^2 = \sum_{i=1}^{k} \left[x_i^2 P(X = x_i)\right] - \mu^2$.

4.2 Probability Distributions

Example: X = maximum of two (fair) dice throughs

What is $Var(X)$ and $SD(X)$?

Probability distribution+weighted averages:

x	$P(X = x)$	num. $P(X = x)$	$x \cdot P(X = x)$	$x^2 \cdot P(X = x)$
1	$1/36$	0.028	0.028	0.028
2	$1/12$	0.083	0.167	0.333
3	$5/36$	0.139	0.417	1.250
4	$7/36$	0.194	0.778	3.110
5	$1/4$	0.250	1.250	6.250
6	$11/36$	0.306	1.833	11.000

Thus $\sum_{i=1}^{6} i^2 \cdot P(X = i) \approx 21.972$. Hence,

$$Var(X) = \sigma^2 = \sum_{i=1}^{6} i^2 \cdot P(X = i) - \mu^2 \approx 21.972 - 20.000 = 1.972,$$

$$SD(X) = \sqrt{Var(X)} = \sigma =\approx \sqrt{1.972} \approx 1.404.$$

4.2 Probability Distributions

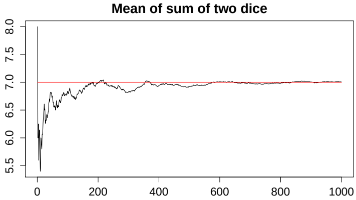
Theorem (Law of Large Numbers)

Let $X_1, \ldots, X_n$ be n independent versions of random variable X; let $\mu = E(X)$. Their mean $\frac{1}{n}(X_1 + \ldots + X_n)$ tends to approach μ.

Special version: LLN of Lect.2: random variable $X_i = 1$ if A occurs, $X_i = 0$ if A doesn't.

Law of large numbers: $X =$ sum of two (fair) dice throughs

We can find that $E(X) = 7$. Behaviour of mean of X_i's after $n(\to \infty)$ double rolls:



4.2 Probability Distributions: recap

General recipe to find probability distribution of discrete random variable

1. Determine sample space and probabilities of underlying probability experiment.
2. List the numerical values $X(\omega)$ for each outcome $\omega \in \Omega$.
3. Find the collection of outcomes which have the same numerical value x.
4. Determine $P(X = x) = P(\{\omega : X(\omega) = x\}) = \sum_{\omega : X(\omega)=x} P(\{\omega\})$.
5. Tabulate the results.

Expectation, standard deviation and variance

Let X be a discrete random variable attaining values $x_1, \ldots, x_k$. Then

$$\mu = E(X) = \sum_{i=1}^{k} x_i \cdot P(X = x_i),$$

$$\sigma^2 = Var(X) = \sum_{i=1}^{k} \left[x_i^2 P(X = x_i)\right] - \mu^2,$$

$$\sigma = SD(X) = \sqrt{Var(X)}.$$