Calculus-based Intro to Probability & Statistics

Math 11: Discussion 4

TA: Thu Nguyen
Tue. 04/21/2020

Recap of discussion 4:

1. Prepare for Midterm on Friday;
2. When given word problems, convert them into math using variables;
3. More importantly, in probability, half the work goes into defining the correct variables!

Contents

Reminder: Midterm this Friday, 04/24.

4.1 Upcoming assignments

Assignments	Chapters	Deadlines
Homework	Ch. 13	Wed. 04/22
Quiz	Ch. 13	Thu. 04/23
Homework	Ch. 14	Wed. 04/24
Quiz	Ch. 14	

Note: No Lab this week. Take the time to prepare for Midterm.

Chapters:

13. Experiments and Observational Studies
14. From Randomness to Probability

Key concepts (not exhaustive):

1. *Probability definitions*: events, complements, disjoint, independence, conditional probability
2. *Probability rules*: Bayes' rules, Law of Total Probability
3. *Expectation* and *variance*

4.2 Basic of Probability

4.2.1 Definitions

We first recall some basic definitions:

1. Probability space: which takes in 3 inputs: $(\Omega, \mathcal{F}, \mathbb{P})$:
   - Ω: the sample space, consisting of all the possible outcomes
   - $\mathcal{F}$: a collection of events, each of which is a set of some outcomes
   - $\mathbb{P}$: a probability measure, which gives each event a probability
2. Complement: given an event A, the *complement* of A, denoted A^C, is the set of all events not A.
3. Disjoint: two events A and B are *disjoint* if they cannot both happen.
4. Independent: an event A is *independent* of an event B if knowing B gives no information on determining A.
5. Conditional probability: an event A is said to be conditioned on B, denoted $A|B$, when we want to determine A given that B has already happened.

Notes: when in doubt, always think of the *naive* definition of probability:

$$\text{Probability} \;=\; \frac{\text{number of desired events}}{\text{total number of possible events}}$$

(You will be amazed by how often this approach can simplify the questions a lot.)

4.2.2 3 probability rules

There are 3 important rules when working with probability, which are true for all probability spaces:

1. $0 \;\leq\; \mathbb{P}(A) \;\leq\; 1$
2. $\mathbb{P}(A^C) \;=\; 1 - \mathbb{P}(A)$
3. $\mathbb{P}(A \cup B) \;=\; \mathbb{P}(A) \;+\; \mathbb{P}(B) \;-\; \mathbb{P}(A \cap B)$

where $A \cup B$ is the event that either A or B or possibly both happen, and $A \cap B$ is the event that both A and B happen. In particular, if A and B are *disjoint*, we have:

$$\mathbb{P}(A \cap B) \;=\; 0 \quad\Longrightarrow\quad \mathbb{P}(A \cup B) \;=\; \mathbb{P}(A) \;+\; \mathbb{P}(B)$$

Remarks:

1. Make sure you understand the notions of $\bigcap$ and $\bigcup$:

$$\begin{aligned} A \cap B &: \quad A \text{ and } B \text{ both happen} \\ A \cup B &: \quad A \text{ or } B \text{ happens, or possibly both} \end{aligned}$$

4.2.3 Bayes' rule

$$\mathbb{P}(A \cap B) \;=\; \mathbb{P}(A|B)\mathbb{P}(B) \;=\; \mathbb{P}(B|A)\mathbb{P}(A) \tag{1}$$

so that if $\mathbb{P}(A) > 0$, we have:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

In particular, if A and B are *independent*, we have:

$$\mathbb{P}(A|B) \;=\; \mathbb{P}(A) \quad \iff \quad \mathbb{P}(A \cap B) \;=\; \mathbb{P}(A) \cdot \mathbb{P}(B)$$

We can generalize the formula (1) to a case of many events:

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) \;=\; \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \mathbb{P}(A_3|A_1 \cap A_2) \ldots \mathbb{P}(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1})$$

4.2.4 Law of Total Probability

Suppose we want to compute the probability of an event A, say $\mathbb{P}(A)$. Instead of direct computation, we can break A into smaller events, similar to the Venn diagram below:

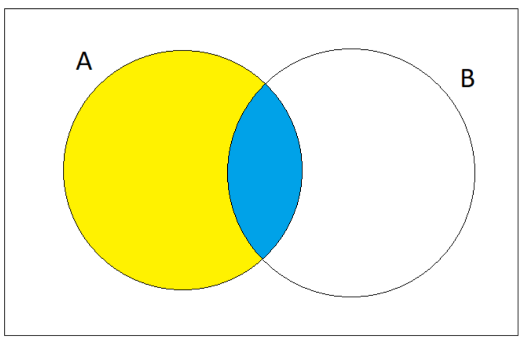

Figure 1: Here, the blue area is the event of both A and B happening, i.e. $A \cap B$, while the yellow area is that of A and not B, i.e. $A \cap B^C$, where B^C is the complement of B. Notice that $A = (A \cap B) \cup (A \cap B^C)$.

Law of Total Probability:

$$\mathbb{P}(A) \;=\; \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^C) \;=\; \mathbb{P}(A|B)\;\mathbb{P}(B) + \mathbb{P}(A|B^C)\;\mathbb{P}(B^C) \tag{1}$$

We can generalize the law: given a finite partition or countably infinitely partition $\{B_n : n = 1, 2, \ldots\}$ of the sample space Ω, we have:

$$\mathbb{P}(A) \;=\; \sum_i \mathbb{P}(A \cap B_i) \;=\; \sum_i \mathbb{P}(A|B_i)\;\mathbb{P}(B_i)$$

Recall the Bayes' rule, if $\mathbb{P}(A) \neq 0$, we can reexpress 1 as:

$$\mathbb{P}(B|A) \;=\; \frac{\mathbb{P}(A|B)\;\mathbb{P}(B)}{\mathbb{P}(A)} \;=\; \frac{\mathbb{P}(A|B)\;\mathbb{P}(B)}{\mathbb{P}(A|B)\;\mathbb{P}(B) + \mathbb{P}(A|B^C)}$$

Example 1. *Let's look at the reality we are facing and consider the likelihood that online learning will extend into the next school year. Suppose the scenario of:*

Event	*Probability*
Fall is online	3/5
Winter is online given that Fall is online	1/5
Winter is online given that Fall is in-person	1/10
Winter is in-person given that Fall is online	4/5

We note that each quarter is either online or in-person. Let's calculate the probability of these events:

(a) Fall and Winter are online;

(b) Fall is online and Winter is in-person;

(c) Winter is online;

(d) Fall is online given that Winter is online;

(e) Fall or Winter is online;

(f) Fall and Winter are in-person;

When working with word problems, always start with these steps:

1. Define variables;
2. Re-express hypothesis using those variables;
3. Re-express the questions using variables.

Given our example, let's define:

$$F: \text{ Fall is online;} \qquad W: \text{ Winter is online;}$$

The complements are:

$$F^C: \text{ Fall is in-person;} \qquad W^C: \text{ Winter is in-person;}$$

Re-expressing the given information gives:

$$\mathbb{P}(F) = \frac{3}{5}; \qquad \mathbb{P}(W|F) = \frac{1}{5}; \qquad \mathbb{P}(W|F^C) = \frac{1}{10}; \qquad \mathbb{P}(W^C|F) = \frac{4}{5};$$

We now re-express the questions:

$$(a)\ \mathbb{P}(F \cap W); \quad (b)\ \mathbb{P}(F \cap W^C); \quad (c)\ \mathbb{P}(W);$$
$$(d)\ \mathbb{P}(F|W); \quad (e)\ \mathbb{P}(F \cup W); \quad (f)\ \mathbb{P}(F^C \cap W^C);$$

In part (f) : $F^C \cap W^C = (F \cup W)^C$. Verify this by drawing a Venn diagram.

Solution:

(a) Fall and Winter are online;

$$\mathbb{P}(F \cap W) \;=\; \mathbb{P}(W|F)\mathbb{P}(F) \;=\; \frac{1}{5}\cdot\frac{3}{5} \;=\; \frac{3}{25}$$

(b) Fall is online and Winter is in-person;

$$\mathbb{P}(F \cap W^C) \;=\; \mathbb{P}(W^C|F)\mathbb{P}(F) \;=\; \frac{4}{5}\cdot\frac{3}{5} \;=\; \frac{12}{25}$$

(c) Winter is online;

$$\mathbb{P}(W) \;=\; \mathbb{P}(W \cap F) + \mathbb{P}(W \cap F^C) \;=\; \frac{3}{25} + \mathbb{P}(W|F^C)\mathbb{P}(F^C) \;=\; \frac{3}{25} + \frac{1}{10}\cdot\frac{2}{5} \;=\; \frac{4}{25}$$

(d) Fall is online given that Winter is online;

$$\mathbb{P}(F|W) \;=\; \frac{\mathbb{P}(F \cap W)}{\mathbb{P}(W)} \;=\; \frac{3/25}{4/25} \;=\; \frac{3}{4}$$

(e) Fall or Winter is online;

$$\mathbb{P}(F \cup W) \;=\; \mathbb{P}(F) + \mathbb{P}(W) - \mathbb{P}(F \cap W) \;=\; \frac{3}{5} + \frac{4}{25} - \frac{3}{25} \;=\; \frac{16}{25}$$

(f) Fall and Winter are in-person;

$$\mathbb{P}(F^C \cap W^C) \;=\; \mathbb{P}((F \cup W)^C) \;=\; 1 - \mathbb{P}(F \cup W) \;=\; 1 - \frac{16}{25} \;=\; \frac{9}{25}$$

Remarks:

1. It can get a lot more straightforward once we rewrite words into mathematical expressions, and then work with symbols rather than working directly with words, which can get confusing very quickly.
2. Define any variables that you use. Define precisely what you mean by x, y or {put your favorite letters here}. 1 definition per letter and use only that definition.
3. In particular, once you have defined a variable, keep that definition throughout. The moment you start adding additional meanings to already defined variables is the moment you will starting confusing yourself, and most likely will arrive at the wrong answers.

4.3 Random Variables

4.3.1 Definitions and Formulae

Random variables can be classified into 2 classes:

- discrete r.v.: which takes on single values $\{x_1, x_2, \dots\}$, each with *probability mass function* p_X such that:

$$\forall x_i \in \Omega : p_X(x_i) \geq 0; \qquad \text{and} \qquad \sum_{x_i \in \Omega} p_X(x_i) = 1$$

- continuous r.v.: which takes on values in some intervals, for example $[x_1, x_2]$ (note that in general, it does not matter if the intervals are inclusive of the end points), each of which has *probability density function* f_X:

$$\forall x \in \Omega : f_X(x) \geq 0; \qquad \text{and} \qquad \int_{x \in \Omega} f_X(x)\,\mathrm{d}x = 1$$

Analogous to the different measures introduced earlier in descriptive statistics, we define the probability over a range of values, *expected value*, and *variance* similarly:

	Discrete r.v.	*Continuous r.v.*
$\mathbb{P}(a \leq x \leq b)$	$\sum_{a \leq x_i \leq b} p_X(x_i)$	$\int_a^b f_X(x)\,\mathrm{d}x$
$\mathbb{E}[X] = \mu_X$	$\sum_{x_i \in \Omega} x_i p_X(x_i)$	$\int_\Omega x f_X(x)\,\mathrm{d}x$
$Var[X] = \sigma_X^2$	$\sum_{x_i \in \Omega} (x_i - \mu_X)^2 p_X(x_i)$	$\int_\Omega (x - \mu_X)^2 f_X(x)\,\mathrm{d}x$

Similarly, given $Var[X]$, the standard deviation of X is $sd[X] = \sqrt{Var[X]}$. Regarding the *variance*, it is sometimes more convenient to use this alternative formula when computing:

$$\boxed{Var[X] = \mathbb{E}[X^2] - \mathbb{E}^2[X] = \mathbb{E}[X^2] - \mu_X^2}$$

Note the use of notations. Traditionally, we use μ_X to denote the (theoretical) expected value of a random variable X, and $\bar{x}$ to denote the average value of the data points X.

4.3.2 Properties of Expected Value and Variance

Suppose X is a random variable (true for both *discrete* and *continuous* cases), and real numbesrs a, b, c:

$$\begin{aligned} \mathbb{E}[X+c] &= \mathbb{E}[X]+c \\ \mathbb{E}[cX] &= c\mathbb{E}[X] \\ \mathbb{E}[X+Y] &= \mathbb{E}[X]+\mathbb{E}[Y] \\ \mathbb{E}[aX+bY+c] &= a\mathbb{E}[X]+b\mathbb{E}[Y]+c \qquad \text{(Linearity of Expectation)} \end{aligned}$$

Suppose X is a random variable (true for both *discrete* and *continuous* cases), and real numbesrs a, b, c:

$$\begin{aligned} Var[X+c] &= VarE[X] \\ Var[cX] &= c^2 Var[X] \qquad (\implies SD[cX] = |c|SD[X]) \\ Var[X+Y] &= Var[X]+Var[Y] \qquad \text{(if } X \text{ and } Y \text{ are independent)} \\ Var[aX+bY+c] &= a^2 Var[X]+b^2 Var[Y] \qquad \text{(if } X \text{ and } Y \text{ are independent)} \end{aligned}$$

Example 2. *Let's consider a* 5*-sided die: five faces numbered* $1, 2, 3, 4, 5$*. Assume the die is fair, ie. the probability of getting any die is the same. Suppose we roll the die* n *times and sum up the results for each roll, denoted by* S_n*.*

(a) Let's roll 1 *time. Find the expected value and the variance of* S_1*.*

(b) Let's roll 5 *times. Assuming each roll is independent of the other. Find the expected value and the variance of* S_5*.*

Solution: Let's first define a random variable to denote the outcome of each roll:

$$X_i : \text{ outcome of roll } i$$

This gives us:

$$\mathbb{P}(X_i = 1) \;=\; \mathbb{P}(X_i = 2) \;=\; \mathbb{P}(X_i = 3) \;=\; \mathbb{P}(X_i = 4) \;=\; \mathbb{P}(X_i = 5) \;=\; \frac{1}{5}$$

We can now re-express the random variable S_n:

$$S_n \;=\; X_1 + X_2 + \cdots + X_n$$

(a) 1 roll:

$$\begin{aligned}
\mathbb{E}[S_1] \;&=\; \mathbb{E}[X_1] \;=\; \frac{1}{5}\Big(1 + \cdots + 5\Big) \;=\; 3 \\
Var[S_1] \;&=\; Var[X_1] \;=\; \mathbb{E}[X_1^2] - \mathbb{E}^2[X_1] \;=\; \frac{1}{5}\Big(1^2 + \cdots + 5^2\Big) - 3^2 \;=\; 2
\end{aligned}$$

(b) 5 rolls:

$$\begin{aligned}
\mathbb{E}[S_5] \;&=\; \mathbb{E}[X_1 + X_2 + \cdots + X_5] \\
&=\; \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_5] && \text{(Linearity of Expectation)} \\
&=\; 5\Big(\mathbb{E}[X_1]\Big) && \text{(since } \mathbb{E}[X_1] = \cdots = \mathbb{E}[X_5]\text{)} \\
&=\; 15 \\
Var[S_5] \;&=\; Var[X_1 + X_2 + \cdots + X_5] \\
&=\; Var[X_1] + \cdots + Var[X_5] && \text{(since each roll is independent)} \\
&=\; 5\Big(Var[X_1]\Big) && \text{(since } Var[X_1] = \cdots = Var[X_5]\text{)} \\
&=\; 10
\end{aligned}$$

Remarks:

1. As often seen in probability questions, half the work goes into defining the right variables. There might be more than one way to define relevant variables, in which case, choose the ones most straightforward and with the easiest interpretations.

2. Check for assumption of independence, if applicable. Indicate where you use the *independence* assumption.

Example 3. *Let's now roll an irregular 4-sided die. The possible outcomes for each roll is:*

Outcome	1	2	5	10
Probability	.5	.25	.2	.05

Suppose we roll the die n times and sum up the results for each roll, denoted by S_n.

(a) Let's roll 1 time. Find the expected value and the variance of S_1.

(b) Let's roll 4 times. Assuming each roll is independent of the other. Find the expected value and the variance of S_4.

Solution: Let's first define a random variable to denote the outcome of each roll:

$$X_i: \text{ outcome of roll } i$$

This gives us:

$$\mathbb{P}(X_i = 1) = .5; \quad \mathbb{P}(X_i = 2) = .25; \quad \mathbb{P}(X_i = 5) = .2; \quad \mathbb{P}(X_i = 10) = .05$$

We can now re-express the random variable S_n:

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

(a) 1 roll:

$$\begin{aligned}
\mathbb{E}[S_1] &= \mathbb{E}[X_1] \\
&= .5(1) + .25(2) + .2(5) + .05(10) \\
&= 2.5 \\
Var[S_1] &= Var[X_1] \\
&= \mathbb{E}[X_1^2] - \mathbb{E}^2[X_1] \\
&= .5(1) + .25(2^2) + .2(5^2) + .05(10^2) - 2.5^2 \\
&= 5.25
\end{aligned}$$

(b) 4 rolls:

$$\begin{aligned}
\mathbb{E}[S_5] &= \mathbb{E}[X_1 + X_2 + \cdots + X_5] \\
&= \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_5] && \text{(Linearity of Expectation)} \\
&= 4\Big(\mathbb{E}[X_1]\Big) && \text{(since } \mathbb{E}[X_1] = \cdots = \mathbb{E}[X_5]\text{)} \\
&= 10 \\
Var[S_1] &= Var[X_1 + X_2 + \cdots + X_5] \\
&= Var[X_1] + \cdots + Var[X_5] && \text{(since each roll is independent)} \\
&= 4\Big(Var[X_1]\Big) && \text{(since } Var[X_1] = \cdots = Var[X_5]\text{)} \\
&= 21
\end{aligned}$$