

Recap on Probability

IE231 - Lecture Notes 7

Nov 7, 2017

Definitions

Basic Definitions

Experiment

An experiment is simply a process that generates data. For example, “Tossing a coin” is an experiment.

Outcome

The result of an experiment is an outcome. As the experiment is repeated outcomes follow. For example, suppose a coin is tossed two times the first outcome can be “Heads” and the second outcome can also be “Heads”.

Sample Space

Sample space is the collection of all possible outcomes. For example, sample space of the coin toss experiments is $\mathbb{S} = \{H, T\}$. If the experiment were instead “tossing two coins consequently”, the sample space would be $\mathbb{S} = \{(T, T), (H, T), (T, H), (H, H)\}$.

Event

Event is a subset of the sample space. It should be defined as a possible outcome of an experiment. For example, with respect to the “two coin tosses” experiment, define event A as getting “1 Heads and 1 Tails in any order”. So $A = \{(H, T), (T, H)\}$.

Probability

Probability is the quantification of uncertainty. It is possible to know the exact probability of events if the sample space and events can be measured perfectly. The number of cases are very few (e.g. coin tosses, dice rolls, other forms of gambling) though. For example we know the size of event A and sample space in the two coin toss experiment. So probability of event A happening is $P(A) = 2/4 = 1/2$.

There are cases (not included in this course) where calculating the event size relative to sample size (i.e. probability) very hard or impossible even though some underlying probabilities or function are known. In this case, simulation can offer a good approximation. For the rest of the cases, approximations and assumptions are put in place and probability estimates are only as good as the validity of these assumptions.

Axioms of Probability

Any event that is that is a subset of the sample space $A_i \in \mathbb{S}$;

- a. $P(A_i) \geq 0$

- b. $P(\mathbb{S}) = 1$
- c. If these events are disjoint $P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + \dots + P(A_n)$.

Also remember, if you define A' as the complement of event A , $P(A') = 1 - P(A)$.

Counting Rules

Counting the possible outcomes is a way to measure probability. The basic counting rules are multiplication rule, permutation rule and combination rule. For example for the two coin tosses example we can count the sample space as 4 items (with multiplication rule).

Conditional Probability

Joint Events

If two events A and B are not independent (or disjoint), probability of both of the event happening is not zero $P(A \cap B) \neq 0$.

Example: In a group of 15 people, 8 people like red meat and 10 people like fish. If a person is randomly selected from the group what is the probability that the chosen person likes both red meat and fish.

Solution: Suppose A represents the red meat liking people and B represents fish liking people. $A + B > 15$, so some $A \cap B = 18 - 15 = 3$ people should love fish. Then $P(A \cap B) = \frac{n(A \cap B)}{n(S)} = 3/15 = 0.2$.

Conditional Events

Sometimes an event information is provided as it happened (i.e. given), therefore reducing the sample space. Then we are asked to calculate another event given this reduced sample space. It is called conditional probability and defined as $P(A|B)$, probability of event A happening given that B already happened.

Example: Suppose that the randomly chosen person is a likes fish. What is the probability that she doesn't like red meat?

Solution: $P(A'|B) = n(A' \cap B)/n(B)$. We know that $n(B) = 10$. We also know that $n(A \cap B) = 3$. So $n(A' \cap B) = 10 - 3 = 7$ and $P(A'|B) = 7/10 = 0.7$.

Independence

Two events are disjoint (or independent) if $P(A|B) = P(A)$ and $P(B|A) = P(B)$. Otherwise they are dependent.

Suppose a series of events A_1, A_2, \dots, A_n . If these events are not independent, then

$$P(A_1 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_2 \cap A_1)P(A_n|A_{n-1} \cap \dots \cap A_1)$$

Otherwise $P(A_1 \cap \dots \cap A_n) = P(A_1) \dots P(A_n)$.

Example: Suppose the difference between drawing cards and tossing coins.

Product Rule

If both $P(A) > 0$ and $P(B) > 0$, then $P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$.

Example: (Same example as above)

Solution: $P(A' \cap B) = P(A'|B)P(B)$. Rearrange $P(A'|B) = \frac{P(A' \cap B)}{P(B)} = \frac{7/18}{10/18} = 0.7$

Example: (Monty revisited). Suppose there are three doors A, B and C. There is a prize between one of those doors with equally likely odds (i.e. $1/3$). You choose one of them randomly (say A). Then one of the doors is opened for you (say C) and it is empty. Will you change the door? What are the odds?

Solution: Probabilities when you first choose are $P(A) = P(B) = P(C) = 1/3$. After door C is open, $P(C|C_{open}) = 0$ and $P(A \cup B|C_{open}) = P(A|C_{open}) + P(B|C_{open}) = 1$. $P(A) = P(A|C_{open}) = 1/3$, since you chose it. Therefore $P(B|C_{open}) = 2/3$.

Solution extended: Take it to the extreme of 100 doors. 98 of them are open, only A and B are left. You chose A previously. Would you change?

Theorem of Total Probability and Bayes' Rule

Suppose B_i are disjoint events, $P(B_i) > 0$ and together they comprise the sample space. Then for any event A, $P(A) > 0$

$P(A) = \sum_i P(B_i)P(A|B_i)$ (Theorem of Total Probability)

$P(B_r|A) = \frac{P(B_r|A)}{\sum_i^k P(B_i|A)} = \frac{P(B_r)P(A|B_r)}{\sum_i^k P(B_i)P(A|B_i)}$ (Bayes' Rule)

Example: At a pastry, customers like 50% of the cakes, 70% of the muffins and 40% of the pies. %60 of the pastry's products are cakes, %30 of them are muffins and 10% of the products are pies.

- What is the probability that she likes a product?
- If a customer likes the product she gets, what is the probability that she got a cake?

Solution: Define event A as she likes the product. And define B_1 , B_2 and B_3 as she purchases cake, muffin and pie, respectively. We know that $P(B_1) = 0.6$, $P(B_2) = 0.3$, $P(B_3) = 0.1$. Also

$$P(A|B_1) = 0.5$$

$$P(A|B_2) = 0.7$$

$$P(A|B_3) = 0.6$$

- $P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3) = 0.5 \cdot 0.5 + 0.3 \cdot 0.7 + 0.1 \cdot 0.6 = 0.52$.
- We would like to find $P(B_1|A)$. Using Bayes' Rule we calculate the region of A and get the joint

$$P(B_1|A) = \frac{P(B_1)P(A|B_1)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3)} = \frac{P(A \cap B_1)}{P(A)} = 0.25/0.52 = 0.48.$$

Random Variables and Probability Distributions

Random Variables

A random variable (usually defined with a capital letter or symbol i.e. X) is a quantity determined by the outcome of the experiment. Its realizations are usually symbolized with lowercase letter (x).

Example: Suppose there are 10 balls in an urn, 5 black and 5 red. Two balls are randomly drawn from the urn without replacement. Define the random variable X as the number of black balls. Then, $X = x$ can get the values of 0, 1 and 2. Let's enumerate $P(X = x)$.

$$\begin{aligned}
 P(X = 0) &= P(RR) = 5/10 * 4/9 = 2/9 \\
 P(X = 1) &= P(BR) + P(RB) = 5/10 + 5/9 + 5/10 * 5/9 = 5/9 \\
 P(X = 2) &= P(BB) = 5/10 * 4/9 = 2/9
 \end{aligned}$$

Discrete Random Variables and Distributions

If a sample space has finite number of possibilities or countably infinite number of elements, it is called a discrete sample space. Discrete random variable probabilities are shown as point probabilities $P(X = x)$. The probability distribution of discrete random variables is also called probability mass function (pmf).

$$\begin{aligned}
 P(X = x) &= f(x) \\
 \sum_x f(x) &= 1 \\
 f(x) &\geq 0
 \end{aligned}$$

Example: (Same as above) Enumerate the probability distribution.

Solution: Random variable X can take values (x) 0, 1 and 2. So $f(1) = 2/9$, $f(2) = 5/9$ and $f(3) = 2/9$.

Cumulative Distribution Function (CDF)

Cumulative distribution function is a special defined function yielding the cumulative probability of random variables up to a value. It is usually symbolised as $F(x)$

$$\begin{aligned}
 F(x) &= P(X \leq x) = \sum_{t \leq x} f(t) \\
 F(x) &= P(X \leq x) = \sum_{\infty} f(t) = 1
 \end{aligned}$$

Example: (Same as above) Enumerate the cdf.

$$\begin{aligned}
 F(0) &= P(X \leq 0) = P(X = 0) = 2/9 \\
 F(1) &= P(X \leq 1) = P(X = 0) + P(X = 1) = 7/9 \\
 F(2) &= P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = 1
 \end{aligned}$$

Continuous Random Variables and Distributions

If a sample space has uncountably infinite number of possibilities, it is called a continuous sample space. Continuous random variables' probabilities are defined in intervals $P(a < X < b)$. The probability distribution of continuous random variables is also called probability density function (pdf).

$$f(x) \geq 0$$
$$P(a < X < b) = \int_a^b f(x) dx$$
$$\int_{-\infty}^{\infty} f(x) = 1$$

Example (from the book): Suppose the probability function of a continuous distribution $f(x) = x^2/3$ defined between $-1 < x < 2$ and 0 everywhere else. Verify that it is a density function (i.e. the integral in the defined interval is 1) and calculate $P(0 < x < 1)$.

- a. $\int_{-1}^2 x^2/3 dx = x^3/9|_{-1}^2 = 8/9 - (-1/9) = 1$. Verified.
- b. $\int_0^1 x^2/3 dx = x^3/9|_0^1 = 1/9 - (0) = 1/9$. Verified.

Cumulative distribution function (CDF) for continuous random variables is defined with the integral.

$$F(x) = P(X < x) = \int_{-\infty}^x f(x) dx$$

Example: (same as above) Calculate the cdf $F(3/2)$

Solution: $F(1.5) = \int_{-\infty}^{3/2} f(x) = x^3/9|_{-1}^{3/2} = 3/8 - (-1/9) = 35/72$

Example: Calculate $P(X > 1)$.

Solution: $P(X > 1) = 1 - P(X < 1) = 1 - F(1) = 1 - \int_{-\infty}^1 = 1 - ((1/9) - (-1/9)) = 7/9$.

Joint, Marginal and Conditional Distributions

Joint Distribution

So far we had distributions with only one random variable. What if we had more than one random variable in a distribution? It is not that different from univariate distributions.

$$f(x, y) \geq 0$$
$$\sum_x \sum_y f(x, y) = 1$$
$$P(X = x, Y = y) = f(x, y)$$

Example (from the book): Two pens are selected at random from a box of 3 blue, 2 red and 3 green pens. Define X as the number of blue pens and Y as the red pens selected. Find

- a. Joint probability function $f(x, y)$
- b. $P[(X, Y) \in A]$ where A is the region $\{(x, y) | x + y \leq 1\}$.

Solution:

- a. The possible cases for (x, y) are $(0, 0), (0, 1), (0, 2), (1, 0), (2, 0), (1, 1)$. For instance $(0, 1)$ is one green and one red pen selected. There are a total of 8 pens. Then sample space size for two pens selected is $\binom{8}{2} = 28$. There are $\binom{2}{1}\binom{3}{1} = 6$ ways of selecting 2 pens from green and red pens. So the probability is $f(0, 1) = P(X = 0, Y = 1) = 6/28 = 3/14$. It is possible to calculate other possible outcomes in a similar way. A generalized formula would be as follows.

$$\frac{\binom{3}{x}\binom{2}{y}\binom{3}{2-x-y}}{\binom{8}{2}}$$

- b. Possible outcomes satisfying $A = (x, y) \leq 1$ are $(0, 0), (0, 1), (1, 0)$. So $P(X + Y \leq 1) = P(0, 0) + P(0, 1) + P(1, 0) = 9/14$.

In the continuous case it is similar. It is now called joint probability density function.

$$\begin{aligned} f(x, y) &\geq 0 \\ \int_x \int_y f(x, y) dx dy &= 1 \\ P(X = x, Y = y) \in A &= \int \int_A f(x, y) dx dy \end{aligned}$$

Example: (from the book)

A privately owned business operates both a drive in and a walk in facility. Define X and Y as the proportions of using the drive in and walk in facilities. Suppose the joint density function is $f(x, y) = 2/5 * (2x + 3y)$ where $0 \leq x \leq 1$ and $0 \leq y \leq 1$ (0, otherwise).

- Verify it is a distribution function.
- Find $P[(X, Y) \in A]$, where $A = \{(x, y) | 0 < x < 1/2, 1/4 < y < 1/2\}$.

Solution:

- $\int \int f(x, y) dx dy = \int 2/5 * (x^2 + 3xy) dy |_0^1 = \int (2/5 + 6/5 * y) dy = 2y/5 + 3/5 * y^2 |_0^1 = 2/5 + 3/5 = 1$.
- $\int_{1/4}^{1/2} \int_0^{1/2} f(x, y) dx dy = \int 2/5 * (x^2 + 3xy) dy |_0^{1/2} = \int_{1/4}^{1/2} (1/10 + 3/5 * y) dy, y/10 + 3y^2/10 |_{1/4}^{1/2} = 13/160$.

Marginal Distribution

In a joint distribution, marginal distribution is the probability distributions of individual random variables. Define $g(x)$ and $h(y)$ as the marginal distributions of X and Y .

$$\begin{aligned} g(x) &= \sum_y f(x, y), h(y) = \sum_x f(x, y) \\ g(x) &= \int_y f(x, y) dy, h(y) = \int_x f(x, y) dx \end{aligned}$$

Example: Go back to pen and walk in examples and calculate marginal probabilities.

Conditional Distribution

Remember the conditional probability rule $P(A|B) = P(A \cap B)/P(B)$ given $P(B) > 0$. We can define conditional distribution as $f(y|x) = f(x, y)/g(x)$, provided $g(x) > 0$ whether they are discrete or continuous.

Example: (pen example) Calculate $P(X = 0|Y = 1)$

Solution: We know that $h(1) = 3/7$. $f(x = 0, y = 1) = 3/14$. $f(x = 0|y = 1) = f(x = 0, y = 1)/h(1) = 1/2$

Statistical Independence

Two random variables distributions are statistically independent if and only if $f(x, y) = g(x)h(y)$.

Proof:

$$f(x, y) = f(x|y)h(y)$$
$$g(x) = \int f(x, y)dy = \int f(x|y)h(y)dy$$

If $f(x|y)$ does not depend on y we can write $f(x|y) \int h(y)dy$. $f(x|y) * 1$. Therefore, $g(x) = f(x|y)$ and $f(x, y) = g(x)h(y)$.

Any number of random variables $(X_1 \dots X_n)$ are statistically independent if and only if $f(x_1, \dots, x_n) = f(x_1) \dots f(x_n)$.

Mathematical Expectation

$$E[X] = \sum xf(x)$$

$$E[X] = \int xf(x)dx$$

Advanced Topics

Moment generating functions, Central Limit Theorem, convolutions, copulas, Sklar's Theorem.