RTPL Wiki

Probabilities

TABLE OF CONTENTS

This page is mainly to describe notations and rules regarding probabilities and expectations.

Probability [edit]

Random variables are denoted by big letters and constant values are denoted by small letters. For a discrete random variable X, the probability mass function (pmf) p of a particular value x gives the probability of the event that the random variable takes that value:

$$p(x) = P(X = x).$$

Sometimes we will talk about different mass functions, which we will distinguish with different symbols as $p_f(x), p_g(x)$ or $f(x), g(x)$.

Conditional probability [edit]

The conditional probability $P(X = x|Y = y)$ of an event $X = x$ given another event $Y = y$ is defined as

$$P(X = x|Y = y) = \frac{P(X = x \cap Y = y)}{P(Y = y)}.$$

In terms of mass functions:

$$p(x|y) = \frac{p(x, y)}{p(y)}.$$

Conditional probabilities can be related to marginal probabilities in the following way, which is known as **the law of total probability**:

$$p(x) = \sum_{y \in \mathcal{Y}} p(x, y) = \sum_{y \in \mathcal{Y}} p(x|y)p(y).$$

Expected value [edit]

The expected value of a discrete random variable is defined as:

$$E[X] = \sum_{x \in \mathcal{X}} x p(x),$$

where $\mathcal{X}$ is the sample space of X.

Note how this expectation can be written in terms of vectors:

$$E[X] = \mathbf{x}^\top \mathbf{p}_X,$$

where the vector $\mathbf{x} \in \mathbb{R}^{|\mathcal{X}|}$ contains different $x \in \mathcal{X}$ as its elements and the vector $\mathbf{p}_X \in \mathbb{R}^{|\mathcal{X}|}$ contains different $p(x)$ as its elements. For convenience, let's represent the values X takes as $x_1, x_2, \cdots, x_{|\mathcal{X}|}$. Then $[\mathbf{x}]_i = x_i$ and $[\mathbf{p}_X]_i = p(x_i)$.

The expected value of a continuous random variable is defined as:

$$E[X] = \int_{-\infty}^{\infty} x p(x) dx,$$

where p here is the probability density function.

Conditional expectation

[edit]

Conditional expectation of a random variable X w.r.t. an event $Y = y$ is written as $E[X|Y = y]$ and defined as:

$$E[X|Y = y] = \sum_{x \in \mathcal{X}} x p(x|y).$$

Conditional expectation of a random variable X w.r.t. another random variable Y is written as $E[X|Y]$ and is itself a random variable as it is a function of another random variable Y. In conditional expectations, the given random variable is not integrated over and thus does not give a constant. Different values this random variable takes are $E[X|Y = y], \forall y \in \mathcal{Y}$. We can take an expectation of this random variable as we obtain:

$$E[E[X|Y]] = \sum_{y \in \mathcal{Y}} E[X|Y = y] p(y) = \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} x p(x|y) p(y) = \sum_{x \in \mathcal{X}} x \sum_{y \in \mathcal{Y}} p(x, y) = \sum_{x \in \mathcal{X}} x p(x) = E[X].$$

This is known as the **law of total expectation** and also the tower rule.

这只求Y的积分，因为X的积分在E[X|Y]中已经被求出来了.

这个只求X的积分，不求Y的.

Again, the conditional expectation $E[X|Y]$ integrates only over the random variable X but not Y. On the other hand, in $E[E[X|Y]]$, the outer expectation integrates over the random variable Y but not X which is already integrated out.

Interesting results

$$E\left[E\left[f(X)|Y\right] g(Y)\right] = E\left[E\left[f(X) g(Y)|Y\right]\right].$$

$$E\left[f(X, Y) g(Y)\right] = E\left[E\left[f(X, Y)|Y\right] g(Y)\right].$$

Other rules

[edit]

$$E[aX + cY] = aE[X] + cE[Y]; \text{ linearity of expectation.}$$

$$\text{Generally, } E[XY] \neq E[X]E[Y]; \text{ non-multiplicativity of expectation.}$$

$$E[g(X)] = \sum_{x \in \mathcal{X}} g(x) p(x); \text{ the law of the unconscious statistician (LOTUS).}$$

Vector and matrix form

[edit]

$$\text{For scalar random variables: } E[X] = \mathbf{x}^\top \mathbf{p}_X, ; \text{ where } X \in \mathbb{R}, [\mathbf{x}]_i = x_i \text{ and } [\mathbf{p}_X]_i = p(x_i).$$

For vector random variables: $E[\mathbf{x}] = \sum_{i=1}^{|\mathcal{X}|} \mathbf{x}_i p(\mathbf{x}_i) = \mathbf{X}\mathbf{p}_{\mathbf{x}}$; where $\mathbf{x} \in \mathbb{R}^n, \mathbf{X} \in \mathbb{R}^{n \times |\mathcal{X}|} [\mathbf{X}]_{:,i} = \mathbf{x}_i$ and $[\mathbf{p}_{\mathbf{x}}]_i = p(\mathbf{x}_i)$.

Jensen's inequality

[edit]

If f is a convex function, then

就是这种:

concave 是

$$E[f(X)] \geq f(E[X]).$$

◀ Announcements

Jump to...

Schedule ▶

Probabilities and Expectations

Ashique Rupam Mahmood

September 9, 2015

a measure of uncertainty, 如 Rain (variable) → yes, → no, Rain = Y 即 having a rain today, R = N 即 not having a rain today. 若 P(Rain = Y) = 0.1, 則 P(Rain = N) = 1 − 0.1 = 0.9

Probabilities tell us about the likelihood of an event in numbers. If an event is certain to occur, such as sunrise, probability of that event is said to be 1. $\mathbf{Pr}(\texttt{sunrise}) = 1$. If an event will certainly not occur, then its probability is 0.

So, probability maps events to a number in $[0, 1]$. How do you specify an event? In the discussions of probabilities, events are technically described as a set. At this point it is important to go through some basic concepts of sets and maybe also functions.

Sets

A *set* is a collection of distinct objects. For example, if we toss a coin once, the set of all possible distinct outcomes will be $S = \{\texttt{head}, \texttt{tail}\}$, where `head` denotes a head and the `tail` denotes a tail. All sets we consider here are finite.

An *element* of a set is denoted as $\texttt{head} \in S$. A *subset* of a set is denoted as $\{\texttt{head}\} \subset S$. What are the possible subsets of S? These are: $\{\texttt{head}\}, \{\texttt{tail}\}, S = \{\texttt{head}, \texttt{tail}\}$, and $\phi = \{\}$. So, note that a set is a subset of itself: $S \subset S$. Also note that, an *empty set* (a collection of nothing) is a subset of any set: $\phi \subset S$. A *union* of two sets A and B is comprised of all the elements of both sets and denoted as $A \cup B$. An *intersection* of two sets A and B is comprised of only the common elements of both sets and denoted as $A \cap B$. A *complement* set of A in B is a set comprising the elements of B that are not in A and denoted as $B \setminus A$ or $B - A$. The *Cartesian product* of two sets A and B is a set denoted as $A \times B$ comprising all ordered pairs (a, b) where $a \in A$ and $b \in B$.

More about Sets (can be skipped)

A *power set* of a set S is a set of all the subsets of S. It is often denoted as 2^S. In our example, it is $2^S = \{\{\texttt{head}\}, \{\texttt{tail}\}, S, \phi\}$. Therefore, $S \in 2^S$. How many distinct elements we have in 2^S? It is 2^n, if n is the number of distinct elements in S.

Functions

A *function* is a map from a one set to another. A function takes an argument and associates it with exactly one quantity, which is often called the *output* of a function. The set of all the arguments a function takes is called the *domain*. The set of all the outputs that a function associates with is called a *codomain*, sometimes also known as *range*.

For example, if we want to associates the outcomes of a coin toss, `head` and `tail`, with numbers 1 and 0, we can do so by using functions. We can say, $f(\texttt{head}) = 1$ and $f(\texttt{tail}) = -1$. Then the domain of the function f is $S = \{\texttt{head}, \texttt{tail}\}$ and the codomain is $V = \{1, -1\}$. This function can also be denoted as $f : S \to V$.

Now, note that if a relation f associates more than one output with an argument, e.g., $f(\texttt{head}) = 1$ or $f(\texttt{head}) = 0$, then f is not a function. However, a function can map two different arguments to a single output. Hence, if $f(\texttt{head}) = 1$ and $f(\texttt{tail}) = 1$, then f is a function.

Sample Space, Events & Probabilities

Let us assume there is an *experiment* that is repeatable, such as rolling a dice. There are two important elements that make up a *probabilistic model*: a sample space and a probability distribution.

A *sample space* is the set of all possible outcomes of an experiment. Therefore, in the dice-rolling experiment, the sample space S is $\{1, 2, 3, 4, 5, 6\}$. An *event* is any subset of the sample space. For example, the event that a number more than two would appear in the dice-rolling experiment is $\{3, 4, 5, 6\}$. It makes sense to find the probability of such an event. An event can comprise only one outcome, for example, the event that 2 will appear: $\{2\}$ or simply 2. A *complementary event* of an event A is $A^c = S - A$.

Now, we are ready to talk about probabilities. A probability is a function that associates an event with a non negative number. Therefore, for any event A

$$\mathbf{Pr}(A) \geq 0. \tag{1}$$

A probability function has certain key properties. For example, the addition of probabilities of all the outcomes is always 1:

$$\sum_{e \in S} \mathbf{Pr}(e) = 1. \tag{2}$$

How the probability is distributed among the outcomes is defined by the *probability distribution*. For example, for the experiment of rolling an unbiased dice, the probability distribution can be uniform, that is, it is the same for all the outcomes: for each outcome $e \in S$, $\mathbf{Pr}(e) = 1/6$.

There are other properties of a probability function. For example, if two events A and B are disjoint (no common elements), then the probability of their union is the addition of their separate probabilities:

$$\mathbf{Pr}(A \cup B) = \mathbf{Pr}(A) + \mathbf{Pr}(B); \quad A \cap B = \phi. \tag{3}$$

The outcomes of an experiment are always mutually disjoint.

So, we talked about three properties of a probability function: nonnegativity (equation 1), normalization (equation 2) and additivity (equation 3).

If we are given to find the probability of an event, one way would be to find the sample space and the probability distribution among the outcomes. Then we need to specify the event in question. Then using the additive rule it becomes easy to find the probability of that event.

Now, we are ready to calculate the probability that a number more than 2 will appear in the unbiased dice rolling experiment. Using the properties in the above, we can write:

$$\begin{aligned}
\mathbf{Pr}(\text{observe a number more than 2}) &= \mathbf{Pr}(\{3,4,5,6\}) \\
&= \mathbf{Pr}(3) + \mathbf{Pr}(4) + \mathbf{Pr}(5) + \mathbf{Pr}(6);\ \text{additivity} \\
&= 4 \times 1/6 = 2/3.
\end{aligned}$$

More about Events & Probabilities (Can be skipped)

How many distinct events are possible in a dice rolling experiment? It is the number of all possible subsets of the sample space. The set of all possible distinct events is the power set of the sample space. The empty set ϕ is a possible event. An empty-set event is the event where nothing appears in a dice rolling experiment. Probability of an empty-set event is zero. This is an event that will certainly not occur. The sample space itself is also an event. It denotes the event that either of the possible outcomes would appear in the dice rolling experiment. This event will certainly occur. Probability of the sample space is always one: $\mathbf{Pr}(S) = 1$.

So, now we understand that probability is in fact a function that takes an event as the argument and gives a number between 0 and 1, inclusively, as an output. We can write $\mathbf{Pr} : 2^S \to [0, 1]$.

Random Variables

It is often more convenient to express events in relations with random variables. A random variable takes the possible values of the outcomes of an experiment. In the dice-rolling experiment, a random variable X would be such that $X \in \{1, 2, 3, 4, 5, 6\}$.

We can now write the event of having a number more than 2 as $X > 2$. Here, $X > 2$ stands for the event $\{3, 4, 5, 6\}$. For the events with only one outcome, we usually write as $X = 2$ or $X = 3$ or $X = x$ and so on. When we use letters, random variables will be denoted by capital letters and values of a random variable will be denoted by small letters.

Examples of Probabilities

In the dice rolling example, if the random variable X stands for the outcome, what is $\mathbf{Pr}(X > 1)$? Now,

$$\begin{aligned}
\mathbf{Pr}(1 \leq X \leq 6) &= 1 \\
\mathbf{Pr}([X = 1] \cup [2 \leq X \leq 6]) &= 1 \\
\mathbf{Pr}([X = 1] \cup [X > 1]) &= 1 \\
\mathbf{Pr}(X = 1) + \mathbf{Pr}(X > 1) &= 1 \\
\mathbf{Pr}(X > 1) &= 1 - \mathbf{Pr}(X = 1) \\
\mathbf{Pr}(X > 1) &= 5/6.
\end{aligned}$$

Conditional Probabilities => conditional events depend on other events.

Probability of an event A given another event B is defined as

$$\mathbf{Pr}(A|B) \doteq \frac{\mathbf{Pr}(A \cap B)}{\mathbf{Pr}(B)}.$$

Here $\doteq$ stands for equality by definition rather than by derivation. $\mathbf{Pr}(A \cap B)$ is also written as $\mathbf{Pr}(A, B)$ or $\mathbf{Pr}(A \& B)$.

When an event is already given or known to have happened, the uncertainty of another event might change. For example, in the dice rolling experiment, $\mathbf{Pr}(3)$ is $1/6$. However, if we already know that the outcome is an odd number, does the event of observing 3 remains uncertain in the same amount? Actually not. It is now more likely to see 3 than before. Conditional probability reflects such a case. It can be calculated in the following way:

$$\mathbf{Pr}(X = 3 | X \text{ is odd}) = \frac{\mathbf{Pr}(X = 3 \cap X \text{ is odd})}{\mathbf{Pr}(\text{X is odd})} = \frac{\mathbf{Pr}(3 \cap \{1, 3, 5\})}{\mathbf{Pr}(\{1, 3, 5\})} = \frac{\mathbf{Pr}(3)}{\mathbf{Pr}(\{1, 3, 5\})} = \frac{1/6}{3/6} = \frac{1}{3}.$$

The uncertainty of observing 3 has indeed changed. It has become more likely to be observed.

Another example. What is the probability to observe an odd number, when we know that it is a multiple of 3?

$$\mathbf{Pr}(X \text{ is odd}|X \text{ is a multiple of } 3) = \frac{\mathbf{Pr}(X \text{ is odd} \cap X \text{ is a multiple of } 3)}{\mathbf{Pr}(\text{ is a multiple of } 3)} = \\ \frac{\mathbf{Pr}(\{1,3,5\} \cap \{3,6\})}{\mathbf{Pr}(\{3,6\})} = \frac{\mathbf{Pr}(3)}{\mathbf{Pr}(\{3,6\})} = \frac{1/6}{2/6} = \frac{1}{2}.$$

When the likelihood of an event does not change after knowing another event, then those two events are said to be independent:

$$[A \text{ is independent of } B] \Leftrightarrow [\mathbf{Pr}(A|B) = \mathbf{Pr}(A)].$$

Following are also equivalent to the above:

$$\begin{aligned} \mathbf{Pr}(A,B) &= \mathbf{Pr}(A)\mathbf{Pr}(B) \\ \mathbf{Pr}(B|A) &= \mathbf{Pr}(B). \end{aligned}$$

When more than two events are involved in a conditional probability, then the following holds:

$$\mathbf{Pr}(A,B|C) = \mathbf{Pr}(A|B,C)\mathbf{Pr}(B|C).$$

Conditional probabilities are particularly important for compound experiments.

Examples

Let us say we have two urns. The first one contains 2 red balls and 1 black balls, and the second one contains 1 red ball and 1 black ball. If we pick one of the urns randomly and then choose one of the balls randomly, what is the probability that it will be red? What is the sample space here? It is $S = \{1r, 1b, 2r, 2b\}$. Here, $1r$, for example, stands for observing a red ball from urn 1.

What is the probability distribution among the outcomes? It is not straightforward to find as before anymore. In fact, this is a compound experiment comprising two experiments: 1) choosing the urn with sample space $S_1 = \{1,2\}$ and 2) choosing a ball within an urn with sample space $S_2 = \{r,b\}$. Here, S is a compound sample space constructed from the Cartesian product of S_1 and S_2 as $S = S_1 \times S_2 = \{1,2\} \times \{r,b\} = \{1r, 1b, 2r, 2b\}$.

Let us consider X_1 be the random variable for the outcome of choosing an urn (over S_1)

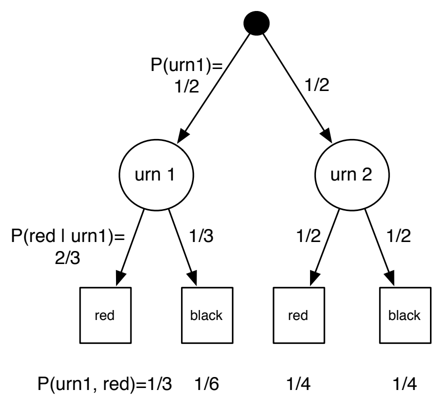
and X_2 be the random variable for the outcome of choosing a ball (over S_2). Then

$$\begin{aligned}\mathbf{Pr}(1r) = \mathbf{Pr}(X_1 = 1, X_2 = r) &= \mathbf{Pr}(X_1 = 1)\mathbf{Pr}(X_2 = r|X_1 = 1) \\ &= \mathbf{Pr}(1)\mathbf{Pr}(r|1) \\ &= 1/2 \times 2/3 \\ &= 1/3.\end{aligned}$$

In the same way, $\mathbf{Pr}(1b) = 1/2 \times 1/3 = 1/6, \mathbf{Pr}(2r) = 1/2 \times 1/2 = 1/4$ and $\mathbf{Pr}(2b) = 1/2 \times 1/2 = 1/4$.

Now, how do we specify the event in question (what is the probability that it will be red)? It is $A = \{1r, 2r\}$. So, $\mathbf{Pr}(A) = \mathbf{Pr}(1r) + \mathbf{Pr}(2r) = 7/12$.

It is often more convenient to use trees for conditional probabilities:



Bayes Theorem

Let us say we know the probability of an event B given another event A. Now we want to know the probability of A given B. Bayes theorem helps to find that. It is simply written as this:

$$\mathbf{Pr}(A|B) = \frac{\mathbf{Pr}(B|A)\mathbf{Pr}(A)}{\mathbf{Pr}(B)}. \tag{4}$$

It can be easily derived from the definition of conditional probability.

Now, let us say, there are n mutually disjoint events A_j, where $j = 1, 2, \ldots, n$ and their union is the sample space: $\cup_{j=1}^{n} A_j = S$. Then probability of any event $B \subset S$ can be written in the following way:

$$\begin{aligned}\mathbf{Pr}(B) &= \mathbf{Pr}(B \cap S);\ \text{as } B \text{ is completely contained in } S \\ &= \mathbf{Pr}(B \cap (\cup_j A_j));\ \text{as given above}\end{aligned}$$

$$
\begin{aligned}
&= \mathbf{Pr}(\cup_j (B \cap A_j)) \\
&= \sum_j \mathbf{Pr}(B \cap A_j); \text{ as all } B \cap A_j \text{ are disjoint} \\
&= \sum_j \mathbf{Pr}(B|A_j)\mathbf{Pr}(A_j). \text{ definition of conditional probability.}
\end{aligned}
$$

The above result is sometimes called *the law of total probability.* Replacing $P(B)$ in 4 with the above, we get a more general Bayes theorem:

$$
\mathbf{Pr}(A_i|B) = \frac{\mathbf{Pr}(B|A_i)\mathbf{Pr}(A_i)}{\sum_j \mathbf{Pr}(B|A_j)\mathbf{Pr}(A_j)}.
$$

Let us look at a classic example of Bayes theorem. Let us say that for a drug test, it returns positive result for a drug user 99% of the time and produces a negative result for a non-user 95% of the time. Suppose that 1% of the population uses drug. Then what is the probability that an individual is a drug user given that she tests positive?

Here, the sample space is constituted of $\{\texttt{user}+, \texttt{user}-, \texttt{nonuser}+, \texttt{nonuser}-\}$. Here, $\texttt{user}-$, for example, stands for the event that an individual is a drug user and he tests negative. We actually do not know the probability distribution here. Instead, we only know some of the conditional probabilities: $\mathbf{Pr}(+|\texttt{user}) = 0.99, \mathbf{Pr}(-|\texttt{nonuser}) = 0.95, \mathbf{Pr}(\texttt{user}) = 0.01$. We want to find $\mathbf{Pr}(\texttt{user}|+)$. Using Bayes theorem, we get

$$
\begin{aligned}
\mathbf{Pr}(\texttt{user}|+) &= \frac{\mathbf{Pr}(+|\texttt{user})\mathbf{Pr}(\texttt{user})}{\mathbf{Pr}(+|\texttt{user})\mathbf{Pr}(\texttt{user}) + \mathbf{Pr}(+|\texttt{nonuser})\mathbf{Pr}(\texttt{nonuser})} \\
&= \frac{0.99 \times 0.01}{0.99 \times 0.01 + (1 - \mathbf{Pr}(-|\texttt{nonuser})) \times (1 - \mathbf{Pr}(\texttt{user}))} \\
&= \frac{0.0099}{0.0099 + (1 - 0.95) \times (1 - 0.01)} \\
&= \frac{0.0099}{0.0099 + 0.05 \times 0.99} \\
&= \frac{0.0099}{0.0099 + 0.0495} \\
&\approx 0.167.
\end{aligned}
$$

this is somewhat surprising and interesting. Even though the stats on the test sounded good, it is still rather unlikely that the person is a drug user even when she tests positive. Bayes theorem helped us realize that.

Expectations

Random variables can have values that are numbers, like an urn number, or they may have non-numeric values, like the color of a ball. If a random variable has exclusively numeric outcomes, then we can talk about its expected value, or expectation. The expected value of a numeric random variable is a weighted average of its possible numeric outcomes, where the weights are the probabilities of the outcome occurring.

When all the outcomes are equally likely, the expectation is the same as the average of possible outcomes. For a discrete random variable (sample space is a finite set), its expectation is defined as

$$\mathbf{E}[X] \doteq \sum_{x \in S} x\mathbf{Pr}(X = x).$$

As any function of a random variable is also a random variable, it is possible to calculate the expectation of a function of a random variable, too. It can be calculated in the following way:

$$\mathbf{E}[g(X)] \doteq \sum_{x \in S} g(x)\mathbf{Pr}(X = x).$$

There are certain properties of expectations:

$$\begin{aligned} \mathbf{E}[X + c] &= \mathbf{E}[X] + c; \text{ where } c \text{ is not a r.v. of the model} \\ \mathbf{E}[X + Y] &= \mathbf{E}[X] + \mathbf{E}[Y] \\ \mathbf{E}[aX] &= a\mathbf{E}[X]; \text{ where } a \text{ is not a r.v. of the model.} \end{aligned}$$

In a certain lottery, it is 0.01% likely to win and the prize is 1000 dollars. The ticket price is 10 dollars. The expected monetary gain from the lottery is

$$\begin{aligned} \mathbf{E}[X] &= (1000 - 10) \times 0.0001 + (-10) \times 0.9999 \\ &= 0.099 - 9.999 \\ &= -9.9. \end{aligned}$$

It is almost the same as giving ten dollars away.

Conditional Expectations

A conditional expectation of a random variable is the expected value of the variable given that an event is already known to have happened. For discrete variables:

$$\mathbf{E}[X|Y=y] \doteq \sum_{x\in S} x\mathbf{Pr}(X=x|Y=y).$$

Examples

Let us say, in the two urns example, we get 10 dollars for observing a red ball and we get 5 dollars for observing a black ball. What is the expected gain of this experiment? Here, the sample space is $S = \{1r, 1b, 2r, 2b\}$.

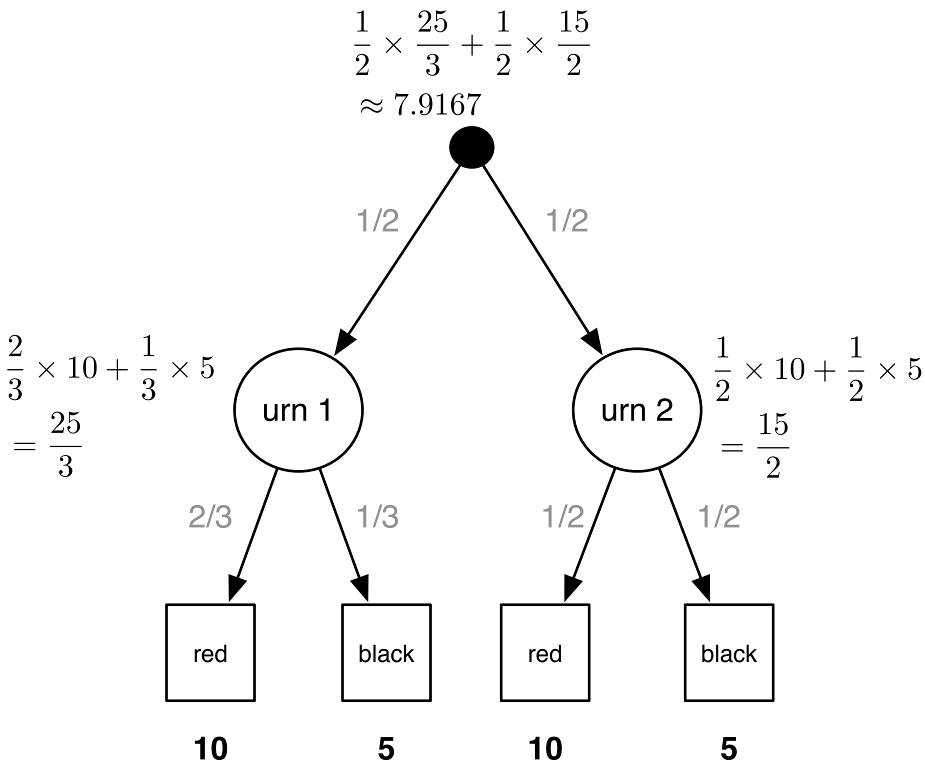
$$\begin{aligned}
\mathbf{E}[gain(X)] &= \sum_{x\in S} gain(x)\mathbf{Pr}(X=x) \\
&= gain(1r)\mathbf{Pr}(X=1r) + gain(1b)\mathbf{Pr}(X=1b) \\
&\quad +gain(2r)\mathbf{Pr}(X=2r) + gain(2b)\mathbf{Pr}(X=2b) \\
&= 10\times\frac{1}{3} + 5\times\frac{1}{6} + 10\times\frac{1}{4} + 5\times\frac{1}{4} \\
&\approx 7.9167.
\end{aligned}$$

Now, let us consider that we can decide which urn to choose. What is the expected gain if we choose urn 1? Let us define another random variable Y for the urn chosen.

$$\begin{aligned}
\mathbf{E}[gain(X)|Y=1] &= \sum_{x\in S} gain(x)\mathbf{Pr}(X=x|Y=1) \\
&= gain(1r)\mathbf{Pr}(X=1r|Y=1) + gain(1b)\mathbf{Pr}(X=1b|Y=1) \\
&\quad +gain(2r)\mathbf{Pr}(X=2r|Y=1) + gain(2b)\mathbf{Pr}(X=2b|Y=1) \\
&= 10\times 2/3 + 5\times 1/3 + 0 + 0 \\
&= 25/3.
\end{aligned}$$

Similarly, we can calculate that the expected gain of choosing urn 2 is $10\times\frac{1}{2} + 5\times\frac{1}{2} = 15/2$. So, it tells that choosing the first urn is more profitable.

Once again, it is often easier to use trees to find such values. The following tree shows the expected gain at each decision point:



Law of Total Expectation

Following is a very useful proposition often known as the law of total expectation:

$$\mathbf{E}[X] = \sum_y \mathbf{E}\left[X|Y=y\right]\mathbf{Pr}(Y=y).$$

It is analogous to the law of total probability that we derived for the Bayes theorem. The derivation of it as follows:

$$\begin{aligned}
\mathbf{E}[X] &= \sum_x x\mathbf{Pr}(X=x) \\
&= \sum_x x \sum_y \mathbf{Pr}\{[X=x] \cap [Y=y]\}; \text{see the derivation of the law of total probability} \\
&= \sum_x x \sum_y \mathbf{Pr}(X=x|Y=y)\mathbf{Pr}(Y=y); \text{definition of conditional probability} \\
&= \sum_y \left[\sum_x x\mathbf{Pr}(X=x|Y=y)\right]\mathbf{Pr}(Y=y) \\
&= \sum_y \mathbf{E}\left[X|Y=y\right]\mathbf{Pr}(Y=y).
\end{aligned}$$

CMPUT 397 Reinforcement Learning:

Probabilities & Expectations

Rupam Mahmood

January 10, 2020

Probabilities and intelligent systems

- ✓ Probability is a measure of uncertainty
- ✓ An intelligent system maximizes its “chances” of success
- ✓ Intelligent systems create a favorable future
- ✓ Probabilities and expectations are tools for reasoning about uncertain future events

- ✓ An intelligent system maximizes its “chances” of success

agents need to do decisions. they want some outcomes more likely than the other.
∴ They need to take actions to make some events occur more.

- ✓ Intelligent systems create a favorable future ⇒

i.e. make an event more likely to happen.
e.g. Rain = N more likely to happen.

After we quantify the probability, how can we make the favorable more likely?

- ✓ Probabilities and expectations are tools for reasoning about uncertain future events

Let's take the example of rolling a dice

- We say the probability of observing 3 is 1/6 If the dice is unbias.
- How to express it mathematically?
- Rolling a dice is, an **experiment**, a repeatable process with different possible results/outcomes
- One **outcome** is 3. Outcomes are mutually exclusive 互相排斥 (一次出现了一个，就不可能同时出现第二个. They don't have any common outcomes)
- The set of all outcomes is called a **sample space**: { 1, 2, 3, 4, 5, 6 } → for the experiment of rolling a dice
- An **event** is a set of outcomes. The event of observing 4 or more: {4, 5, 6}
- Define ***P*** as a function mapping from events to probabilities: $P(3) = 1/6$ → output is the probability in number.

input (an event)

prob-axioms.mp4

8:01 AM Wed Jan 13 66%

is a mapping takes an event as input.

Probability axioms

probability 永远大于0.

✓ **Non-negativity:** A probability is always non-negative

$$0 \leq P(A), \text{ for all } A$$

what kind of object is this?

must be an event (input)

✓ **Additivity:** If $A \cap B = \{\}$, then $P(A \cup B) = P(A) + P(B)$ $A \cup B$: the events of A and B happends

two events are exclusive (即 they don't have any common outcomes)

✓ **Unit measure:** $P(\Omega) = 1$, where Ω is the sample space

probability of whole sample space. (the sample space is certainty happend).

set of all possible outcomes (注: sample space is also an event. 即 an event of 所有可能事情都发生).

✓ What is the probability of observing 4 or more?

✓ $P(\{4, 5, 6\}) = P(4) + P(5) + P(6) = 3/6 = 1/2$

$4 \cap 5 = \{\}$
$5 \cap 6 = \{\}$
$4 \cap 6 = \{\}$

0:18 / 3:34

Random variables

- ✓ Random variables are a convenient way to express events
- ✓ A Random variable is a function mapping from outcomes to real values
- ✓ For coin-tossing experiment: it can be X(head) = 1 and X(tail) = -1

  Random variable
- ✓ For outcomes of dice-rolling experiment: $X(a) = a$ => 因为 outcome is already number, ∴ you can just map the outcome (like 掷到1) to the same number 1.

  ∴ the outcome of observing 1 is mapped to 1.
- ✓ It allows succinct expressions for events such as $[X \geq 4]$

  → event of observing 4 or more. (掷到4或4以上的数字)

  $\{4, 5, 6\}$

  which stands for $\{\ \omega \in \Omega : X(\omega) \geq 4\ \} = \{\ 4, 5, 6\ \}$

  choosing outcomes 满足 expression $X(\omega) \geq 4$

Random variables: example

- ✓ If we roll two dices, what is the probability of the sum being more than 2?
- ✓ Sample space: { (1,1), …, (1,6), (2,1), …, (2,6), …, (6,1), …, (6,6) } ⇒ containing all possible outcomes, 共36 different outcomes.
- ✓ We can define a random variable X standing for the sum
- ✓ Then the event of "the sum being more than 2" can be written as $[X > 2]$

  掷出来总数=2 加上 掷出来总数>2 (exclusive)
- ✓ Then $1 = P(\Omega) = P([X = 2] \cup [X > 2]) = P(X = 2) + P(X > 2)$ ∴ probability of that event is $P([X>2])$

the sample space. ∴ 总概率为1 (即一定会发生).

↳ 注: $P(X>2) = 1 - P(X=2)$, 因为 $1 = P(X=2) + P(X>2)$

$= 1 - \frac{1}{36} = \frac{35}{36}$.

↳ 只有 (1,1) 一个.

Conditional probabilities

✓ A conditional probability is a measure of an uncertain event when we know that another event has occurred

✓ In the single dice-rolling experiment, if the value is below 4, what is the probability that the value is more than 2

∴ value can't be 4, 5, 6

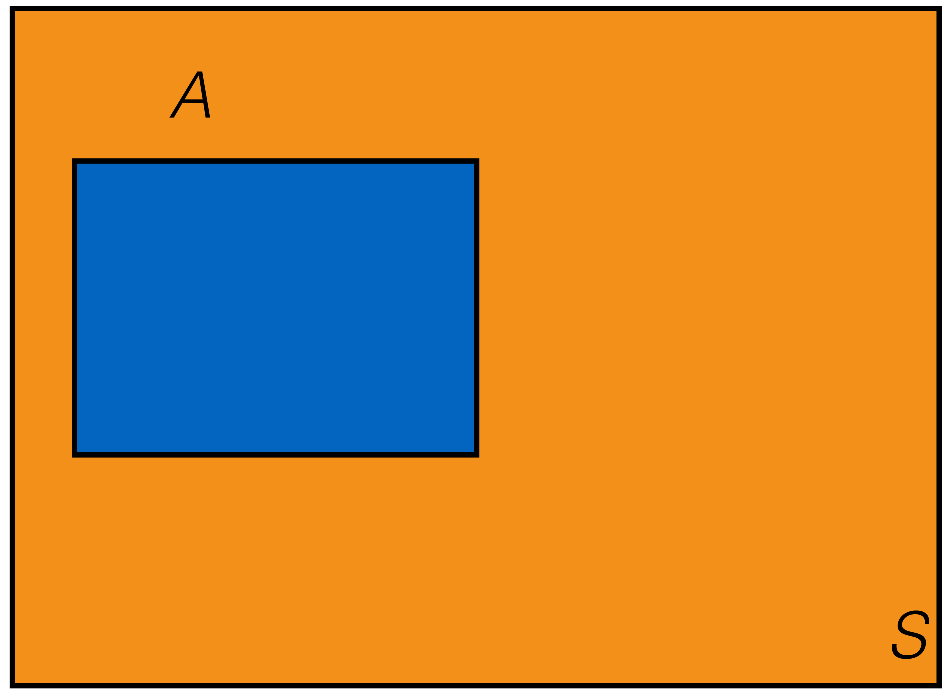
=> p (value > 2 | value < 4)

✓ Definition: $P(A \mid B) = P(A \cap B) \,/\, P(B) \neq P(A)$

probability of an unconditional event

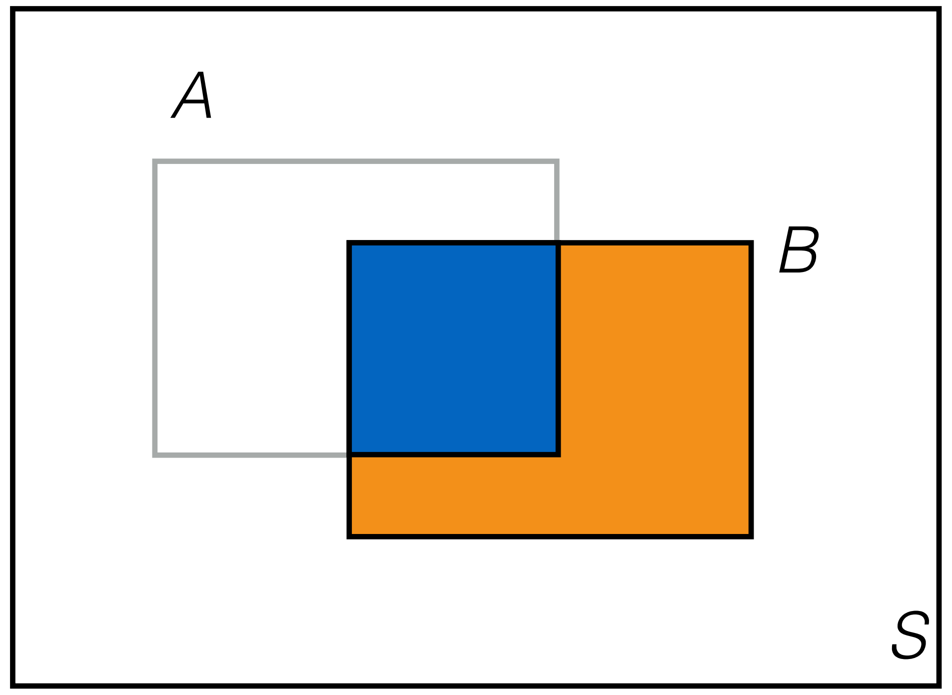
probability of A given B

probability of another unconditional event



这个是 P(A) : what is the probability of this particular event A which subsetted the sample space has happended?

unconditional probability



这个是 P (A|B) : what is the probability of event A happends given event B?

因为 B is given, ∴ not consider the whole sample space anymore, 只考虑 B. 此时 B 就是 the sample space. (因为 we already know B has happended. 所以不存在 outcome outside of B).

Conditional probabilities: example

✓ In the single dice-rolling experiment, if the value is below 4, what is the probability that the value is more than 2

given $Z<4$, ∴ Z只能在1, 2, 3里选. 此时 sample space 为{1, 2, 3}.

✓ $P(\,[Z > 2 \mid Z < 4]\,)$

需求 $Z>2$, 且 given $Z<4$. ∴ Z只能为3. 因只能在1, 2, 3里选.
∴ $P=\frac{1}{3}$.

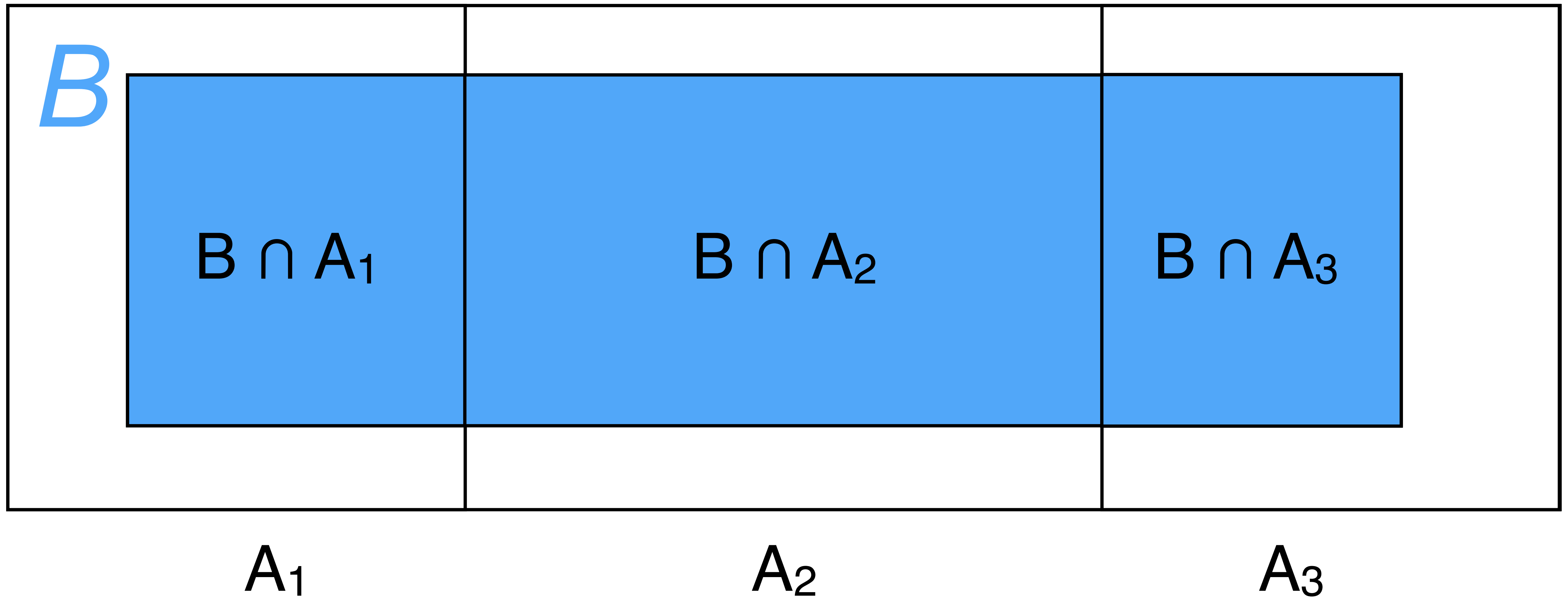
✓ $= P(\,[Z > 2] \cap [Z < 4]\,) \,/\, P(\,[Z < 4]\,)$

$=\frac{1}{2}$. (在1~6中选到1~3)

✓ $= P(\,[Z = 3]\,) \,/\, P(\,[Z < 4]\,)$

✓ $= (1/6) \,/\, (1/2) = 1/3$

Low of total probabilities



$$A_i \cap A_j = \phi,\ i \neq j, \qquad \cup_i A_i = \Omega$$

$$P(B) = \sum_k P(B \cap A_k)$$

$$= \sum_k P(B \mid A_k)\, P(A_k)$$

Expectations & conditional expectations

✓ An expected value of a random variable is a weighted sum of possible outcomes, where the weights are the probabilities of those outcomes

function

$$\mathbf{E}[X] = \sum_{x \in \mathcal{X}} x\, P(X=x)$$ ⇒ gives a real number as output

rv

RV 在这6个值中任取一个.

eg. dice rolling experiment 中, random variable $X \in \{1, 2, \dots, 6\}$

$P(X=1)$

$\therefore E[X] = \frac{1}{6}\times 1 + \frac{1}{6}\times 2 + \frac{1}{6}\times 3 + \frac{1}{6}\times 4 + \frac{1}{6}\times 5 + \frac{1}{6}\times 6$ ⇒ $\frac{1}{6}$ 是每个 event 的 weight, 就是它们出现的 probability.

$= \frac{1}{6}\times(1+2+\dots+6)$

$= 3.5$

✓ An expected value of a random variable conditional on another event is a weighted average of possible outcomes, where the weights are the conditional probabilities of those outcomes given the event

taken this event (a possible outcome).

main argument (a R.V.)

the weight of x from outcome space $\mathcal{X}$ is the conditional probability.

$$\mathbf{E}[X \mid Y=y] = \sum_{x \in \mathcal{X}} x\, P(X=x \mid Y=y)$$ ⇒ a real value.

second argument (an event)

2个 arguments 都是RV, 所以 E[X|Y] 的 output 也是RV.

✓ Expectation conditional on a random variable **E**[*X* | *Y*] itself is a random variable, which is a function of another random variable *Y*

$E[X|Y] = g(Y)$

a function of a second random variable as the second argument.

Properties of expectations

sum of two different R.V.

✓ Linearity: $\mathbf{E}[\underbrace{X + Y}] = \mathbf{E}[X] + \mathbf{E}[Y]$

可写为 $E[\underset{\text{R.V.}}{Y}]$, $Y = aX$

✓ Linearity: $\mathbf{E}[\underset{\text{constant}}{a}\underset{\text{R.V.}}{X}] = a\mathbf{E}[X]$

✓ Non-multiplicativity: $\mathbf{E}[XY] \neq \mathbf{E}[X]\,\mathbf{E}[Y]$

✓ Law of the unconscious statistician: $\mathbf{E}[\,g(X)\,] = \sum_{x \in \mathcal{X}} g(x)\,P(X=x)$

a R.V. which itself is a function of another R.V.

求 $E[Y]$, $Y = g(X)$.

$= \sum_{y \in Y} y\, P(Y=y)$

probability of the taken value of the R.V. Y is y.

a bigger sample space from which the R.V. Y can take values.

$= \sum g(x)\, P(g(X) = g(x))$

probability of g(X) taken value of g(x). 这也是 g(x) 的 weighting.

$= \sum_{x \in X} g(x)\, P(X=x)$

Expectations: example

- ✓ In the double dice-rolling experiment, What is the expected value of the sum of the two dice?

用正常方法解: $E[\underbrace{X}_{}]$

A random variable, which is the sum of two dices.

$\underbrace{X}(\omega) = \underbrace{c}$

the random variable takes an outcome as an input.

R.V. gets a real value (实数) as an output

outcome space = $\{(1,1), (1,2), \dots, (6,6)\}$ 共36个.

for outcome $(1,1)$, the R.V. $X(1,1) = 2$

$(2,5)$, $X(2,5) = 5$

all possible outcomes of the sum of two dices.

$X \in \{2, 3, 4, \dots, 12\}$

$(1,1)$ $(1,2)$ $(1,3)$ $(6,6)$

∴ probability $= \frac{1}{36}$

$$\therefore E[X] = P(1,1)\cdot 2 + [P(1,2) + P(2,1)]\cdot 3 + \dots + P(6,6)\times 12$$

$$= \frac{1}{36}\times 2 + \frac{2}{36}\times 3 + \dots + \frac{1}{36}\times 12$$

$$= \frac{1}{36}(2+12) + \frac{2}{36}(3+11) + \dots + \frac{5}{36}(6+8) + \frac{6}{36}\times 7$$

$$= 14 \times \frac{1}{36} \times (1+2+\dots+5) + \frac{7}{6}$$

$$= 7$$

用properties解: (简化的计算过程).

$$E[\underbrace{Y}_{\text{the sum}}] = E[\underbrace{X}_{\text{one dice}} + \underbrace{Z}_{\text{another dice}}] \Rightarrow Y = X + Z$$

$$= E[X] + E[Z]$$

$$= 3.5 + 3.5$$ ⇒ an expectation of one single dice 为 $3.5 = \frac{1}{6}\times(1+2+\dots+6)$.

$$= 7$$

Example of conditional Expectations:

Q. In the double dice-rolling experiment, what is the expected value of the sum of the two dice given that both dice have the same value?

A. X & Y be the random variables for the 2 dice.

We are asking $E[X+Y \mid X=Y]$

$= E[X \mid X=Y] + E[Y \mid X=Y]$; (linearity)

因为 given two dices give the same value, ∴ $X=Y$.

$= 2E[X \mid X=Y]$; (symmetry of the 2 dice)

$$= 2\sum_{x=1}^{6} x \, P(X=x \mid X=Y)$$

$$= 2 \boxed{\sum_{x=1}^{6} x \times \frac{1}{6}}$$

$\to (1+2+\cdots+6) \times \frac{1}{6}$

$= 3.5$

一共有 1~6 共6个 possible outcomes. (因为 two dices give the same value).
∴每个 outcome出现的可能性为 $\frac{1}{6}$.

$$= \frac{2}{6} \times \frac{6 \times 7}{2}$$

$$= 7.$$

Expectations: example

✓ Show that $\mathbf{E}[X] = \mathbf{E}[\,\mathbf{E}[X \mid Y]\,]$ ⇒ law of total expectation.

expectation of an conditional expectation.

R.V.

证：$E[E[X|Y]]$

2h arguments 都是 R.V.，∴ $E[X|Y]$ 也是 R.V.

有 $E[X|Y] = g(Y)$

$\therefore g(y) = E[X|Y=y]$

$= E[g(Y)]$

$= \sum_y g(y)\, P(Y=y)$ ⇒ due to LoTUS

Law of the unconscious statistician: $\mathbf{E}[\,g(X)\,] = \sum_{x \in \mathcal{X}} g(x)\, P(X=x)$

$= \sum_y E[X|Y=y]\, P(Y=y)$

$= \sum_y \sum_x x\, P(X=x|Y=y)\, P(Y=y)$ ⇒ $\mathbf{E}[X \mid Y=y] = \sum_{x \in \mathcal{X}} x\, P(X=x \mid Y=y)$ ⇒ a real value.

main argument (a R.V.)

second argument (an event)

taken this event (a possible outcome).

the weight of x from outcome space X is the conditional probability.

$= \sum_y \sum_x x\, P(X=x \cap Y=y)$ ⇒ $P(A|B) = P(A \cap B) \div P(B)$, $\therefore P(X=x|Y=y) \cdot P(Y=y)$

$= P(X=x|Y=y) \div P(Y=y) \times P(Y=y)$

$= P(X=x|Y=y)$.

$= \sum_x x \sum_y P(X=x, Y=y)$

$= \sum_x x\, P(X=x)$ ⇒ law of total probability

$= E[X]$.

大写 (Random Variable).

X ; $X = x$, $P(X=x)$

小写 (outcome)

$E[X|Y]$; $\underbrace{\boxed{E[X|Y]}}_{}$ $= \boxed{E[X|Y=y]}$;

$f(Y)$, $f(Y)$, $f(y)$

$P(E[X|Y] = E[X|Y=y])$

If we write a probability as P(A), what is A?

probability 的 input 是 event.

- A random variable
- ✓ An event
- A real number between 0 and 1

Send

Voting as Anonymous

If A is an event, what is P(A)?

Probability is a num ≥ 0, ≤ 1

A real number between 0 and 1

94%

A random variable

3%

An event

3%

Edit response

If we write a probability as P(X=x), what is X?

√A random variable ✓ 大写的是random variable

85%

An outcome

10%

An event

5%

Edit response

If we write a probability as P(X=x), what is x?

- An outcome ✓ 사용되는 outcome. — 61%
- A real-valued outcome — 26%
- A random variable — 6%
- An event — 6%

Edit response

app.sli.do

365w22...

Talking: Rupam Mahmood

If we write a probability as P(X=x), what is [X=x]?

✓ An event [illegible] an input of a probability function is an event.

76%

An outcome

24%

A random variable

0%

Edit response

Acceptable Use - Privacy Policy

app.sli.do

365w22...

Talking:

If X is a random variable, P(X) is a valid expression of a probability

can also be an event，但需写成 $X=x$，因为 random variable 中会有多个 events.

✓ False

70%

True

30%

Edit response

app.sli.do

365w22...

Talking: Rupam Mahmood

If we write a conditional probability as P(A|B), what is B?

Probability of event A happends given event B.

✓ An event

83%

A random variable

17%

A real number between 0 and 1

0%

Edit response

Ex: rolling a dice, $P(A) = P([X>3]) = \frac{1}{2}$

the outcome > 3

Acceptable Use - Privacy Policy

$P(A|B) = P([X>3] \mid X \text{ is even})$

$= \frac{2}{3}$

outcome 可为 4 或 6. 因为 given X is even, ∴ X 只能是在 2, 4, 6 里选.

且让 $X>3$. ∴ X 只能是 4 或 6.

∴ probability 比例为 在 2, 4, 6 中选 4, 6, 即 $\frac{2}{3}$.

$P(A|B) = P(X \text{ is even} \mid [X>3])$

$= \frac{2}{3}$

∴ X 只能在 4, 5, 6 里选.

在 4, 5, 6 中选 4, 6. ∴ 为 $\frac{2}{3}$.

app.sli.do

365w22...

Talking: Rupam Mahmood

If we write an expectation as E[X], what is X? Expectation takes an random variable as input and give a real number as an output.

✓ A random variable — the input of an expectation function is a random variable.

96%

An event

4%

A real number

0%

A real number between 0 and 1

0%

Edit response

app.sli.do

365w22...

Talking:

If X is a random variable, what is E[X]?

doesn't have to be between 0 and 1

A real number

97%

A random variable

3%

An event

0%

A real number between 0 and 1

0%

Edit response

app.sli.do

365w22...

Talking: Rupam Mahmood

If X is a random variable and x is an outcome, then E[X=x] is a valid expression of an expectation

an event can't be the input of the expectation function.

False

83%

True

17%

Edit response

Acceptable Use - Privacy Policy

app.sli.do

365w22...

Talking:

If we write a conditional expectation as E[Z|Y=y] where Z and Y are random variables and y is an outcome, what is E[Z|Y=y]?

an given event

A real number

100%

A random variable

0%

An event

0%

A real number between 0 and 1

0%

Edit response

Ex: $E[X] = 3.5 = (1+2+3+4+5+6)/6$

$E[X \mid X \text{ is even}] = 4. = (2+4+6)/3$

Acceptable Use - Privacy Policy

Deepak Ranganatha Sastry Mamillapalli 1:48 PM
to Everyone

D

think expectations are only defined for
numerical random vars @joel

app.sli.do

365w22...

Talking: Rupam Mahmood

If we write a conditional expectation as E[Z|Y] where Y and Z are random variables, what is E[Z|Y]?

A random variable

79%

A real number

18%

An event

4%

A real number between 0 and 1

0%

Edit response

If Y and Z are random variables, what is E[E[Z | Y]]?

왜냐 $E[Z|Y]$는 결국 random variable. $\therefore$ $E[E[Z|Y]]$ 결국 take an random variable as an input. $\therefore$ outcome 결국 real number.

A real number

77%

A random variable

23%

An event

0%

Edit response

If X, Y, and Z are random variables.
what is E[E[Z | Y] | X]? 因为 $E[Z|Y]$ 与 X 都为 RV，∴ $E[E[Z|Y]|X]$ 也为 RV.
因为 Z 与 Y 都为 RV，∴ $E[Z|Y]$ 也是 RV.

✓ A random variable ✓ 92%

A real number 8%

An event 0%

Edit response

app.sli.do

365w22...

Talking: Rupam Mahmood

real number. 因为 $E[c]=c$, where c is constant (real number). $\therefore E[E[Z|Y=y]]=E[Z|Y=y]$

If X, Y, and Z are random variables, and x and y are outcomes, what is E[E[Z | Y=y] | X=x]?

event

take an event as input, ∴ 该 expectation 的 output 必定为 real number 而不是 RV.

只有当 input 为 RV 时, 下的 output 才为 RV.

如 $E[10 | X=x]$ 的 outcome 就一定是 10.

如 $X=1$

E[Z | Y=y]

71%

A real number 它也对. 但 $E[Z|y=y]$ 更准确.

29%

A random variable

0%

An event

0%

A real number between 0 and 1

0%

Edit response

Acceptable Use - Privacy Policy

Worksheet questions

- PDF link
- Skip 2b, 2c,