Probability: Part III

Cunsheng Ding

HKUST, Hong Kong

October 23, 2015

Contents

Bayes Theorem

Sometimes, we want to assess the probability that a particular event occurred on the basis of partial evidence. This can be achieved with Bayes Theorem.

Theorem 1

Suppose that a sample space S is the union of mutually disjoint events B_1, B_2, B_3, …, B_n, suppose A is an event in S, and suppose A and all the B_i have nonzero probabilities. If k is an integer with $1 \leq k \leq n$, then

$$p(B_k|A) = \frac{p(A|B_k)p(B_k)}{\sum_{i=1}^{n} p(A|B_i)p(B_i)} \tag{1}$$

Proof of Bayes' Theorem

By the definition of conditional probability, $p(A \cap B_k) = p(A)p(B_k|A)$. But

$$p(A \cap B_k) = p(B_k \cap A) = p(B_k)p(A|B_k).$$

Therefore,

$$p(B_k|A) = \frac{p(A|B_k)p(B_k)}{p(A)}. \tag{2}$$

By assumption, $S = \cup_{i=1}^{n} B_i$, where $B_i \cap B_j = \emptyset$ for any (i,j) with $i \neq j$. Then

$$A = A \cap S = A \cap (\cup_{i=1}^{n} B_i) = \cup_{i=1}^{n}(A \cap B_i).$$

Note that $(A \cap B_i) \cap (A \cap B_j) = \emptyset$ for any (i,j) with $i \neq j$. We have

$$p(A) = p(\cup_{i=1}^{n}(A \cap B_i)) = \sum_{i=1}^{n} p(A \cap B_i) = \sum_{i=1}^{n} p(A|B_i)p(B_i). \tag{3}$$

Combining (2) and (3) yields the desired equality.

Bayes Theorem

When $k = 2$, Bayes' theorem becomes the following.

Corollary 2

Let A and B be two events from a sample space S such that $p(A) \neq 0$ and $p(B) \neq 0$. Then

$$p(B|A) = \frac{p(A|B)p(B)}{p(A|B)p(B) + p(A|B^c)p(B^c)}. \tag{4}$$

Bayes Theorem

Problem 3

We have two boxes. The first one contains 2 green balls and 7 red balls; the second one contains 4 green balls and 3 red balls. Bob selects a ball by first choosing one of the two boxes at random. He then selects one of the balls in this box at random. If Bob has selected a red ball, what is the probability that he selected a ball from the first box?

Solution

Let B be the event that Bob selected the ball from the first box, and A the event that he selected a red ball. Note that

$$p(B) = p(B^c) = \frac{1}{2}, \;\; p(A|B) = \frac{7}{9}, \;\; p(A|B^c) = \frac{3}{7}.$$

Then

$$p(B|A) = \frac{p(A|B)p(B)}{p(A|B)p(B) + p(A|B^c)p(B^c)} = \frac{49}{76}.$$

Random Variables

Definition 4

A probability space is a pair (S, p), where S is a sample space and p is probability function (or probability measure) on S.

Definition 5

A random variable is a **function** from the sample space S of a probability space (S, p) to the set $\mathbb{R}$ of real numbers.

Remark

A random variable is a function, **not a variable.**

Random Variables

Example 6

Flip a fair coin three times. Let $X(t)$ be the number of heads that occurs, where t is the outcome. Then

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

Distribution of a Random Variable

Definition 7

The distribution of a random variable X on a probability space (S, p) is the set of pairs:

$$(r, p(X = r)) \text{ for all } r \in X(S)$$

where $p(X = r)$ is the probability that X takes the value r.
A distribution is usually described by specifying $p(X = r)$ for each $r \in X(S)$.

Example 8

For the random variable X in Example 6, the distribution is

$$p(X = 0) = \frac{1}{8},\ p(X = 1) = p(X = 2) = \frac{3}{8},\ p(X = 3) = \frac{1}{8}.$$

Expected Value of a Random Variable

Definition 9

The expected value (also called, expectation) of a random $X(s)$ on a probability space (S, p) is defined by

$$E(X) = \sum_{s \in S} p(s)X(s) = \sum_{r} rp(X = r).$$

Remark

When there are infinitely many elements in S, the expectation is defined only when $\sum_{s \in S} p(s)X(s)$ is convergent.

Expected Value of a Random Variable

Example 10

Flip a fair coin three times. Let $X(s)$ be the number of heads that occurs, where s is the outcome. For the random variable X, the distribution is

$$p(X=0)=\frac{1}{8},\ \ p(X=1)=p(X=2)=\frac{3}{8},\ \ p(X=3)=\frac{1}{8}.$$

Hence, the expected value is

$$E(X)=0\times\frac{1}{8}+1\times\frac{3}{8}+2\frac{3}{8}+3\times\frac{1}{8}=\frac{3}{2}.$$

The Linearity of Expectation of a Random Variable

Theorem 11

Let X and X_i, $i = 1, 2, \ldots, n$, be random variables on a probability space (S, p), and let a and b be real numbers. Then

1. $E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i)$.
2. $E(aX + b) = aE(X) + b$.

Proof.

By definition, we have

$$\begin{aligned} E\left(\sum_{i=1}^{n} X_i\right) &= \sum_{s \in S} p(s) \sum_{i=1}^{n} X_i(s) = \sum_{i=1}^{n} \sum_{s \in S} p(s) X_i(s) = \sum_{i=1}^{n} E(X_i), \\ E(aX + b) &= \sum_{s \in S} p(s)(aX + b)(s) = \sum_{s \in S} p(s)(aX)(s) + \sum_{s \in S} p(s) b. \end{aligned}$$

$\square$

Independent Random Variables

Definition 12

Two random variables X and Y on a probability space (S, p) are independent if

$$p(X(s) = r_1 \text{ and } Y(s) = r_2) = p(X(s) = r_1) \times p(Y(s) = r_2)$$

for all real numbers r_1 and r_2.

Remark

The independence of two events in a sample space S is **usually** different from that of two random variables on S.

Independent Random Variables

Example 13

A pair of fair dice is rolled. Let X_1 and X_2 be the random variables denoting the numbers appearing on the first and second dice, respectively. Are X_1 and X_2 independent?

Solution

Since the pair of dice are fair and two rolling are independent of each other,

$$p(X_1 = r_1 \text{ and } X_2 = r_2) = \begin{cases} \frac{1}{36} & \text{if } r_1, r_2 \in \{1,2,3,4,5,6\}, \\ 0 & \text{otherwise.} \end{cases}$$

On the other hand, $p(X_i = r) = 1/6$ if $r \in \{1,2,3,4,5,6\}$ and $p(X_i = r) = 0$ otherwise. Hence

$$p(X_1 = r_1 \text{ and } X_2 = r_2) = p(X_1 = r_1) \times p(X_2 = r_2)$$

for all real numbers r_1 and r_2. Thus, X_1 and X_2 are independent.

Independent Random Variables

Proposition 14

Let X and Y be two independent random variables on a probability space (S, p). Then

$$E(XY) = E(X)E(Y).$$

Proof.

The proof is straightforward and left as an exercise. $\square$

Example 15

A pair of fair dice is rolled. Let X_1 and X_2 be the random variables denoting the numbers appearing on the first and second dice, respectively. Then $E(X_i) = \sum_{i=1}^{6} \frac{1}{6} i = \frac{7}{2}$. Since X_1 and X_2 are independent,

$$E(X_1 X_2) = E(X_1)E(X_2) = \frac{49}{4}.$$

Variance and Standard Deviation

Definition 16

Let X be a random variable on a probability space (S, p). The <u>variance</u> of X is

$$V(X) = \sum_{s \in S} [X(s) - E(X)]^2 p(s).$$

The <u>standard deviation</u> of X is defined to be $\sigma(X) = \sqrt{V(X)}$.

Remark

The expected value of a random variable informs us its average value, but does not give us information about how widely its values are distributed. The variance of a variable does this for us.

Variance and Standard Deviation

Proposition 17

If X is a random variable on a probability space (S, p), then $V(X) = E(X^2) - E(X)^2$.

Proof.

We have

$$
\begin{aligned}
V(X) &= \sum_{s \in S} [X(s) - E(X)]^2 p(s) \\
&= \sum_{s \in S} X(s)^2 p(s) - 2E(X) \sum_{s \in S} X(s) p(s) + E(X)^2 \sum_{s \in S} p(s) \\
&= E(X^2) - 2E(X)^2 + E(X)^2 \\
&= E(X^2) - E(X)^2.
\end{aligned}
$$

$\square$

Variance and Standard Deviation

Example 18

Flip a fair coin three times. Let $X(s)$ be the number of heads that occurs, where s is the outcome. For the random variable X, the distribution is

$$p(X=0)=\frac{1}{8},\ \ p(X=1)=p(X=2)=\frac{3}{8},\ \ p(X=3)=\frac{1}{8}.$$

Hence, the expected value is

$$E(X)=0\times\frac{1}{8}+1\times\frac{3}{8}+2\frac{3}{8}+3\times\frac{1}{8}=\frac{3}{2}.$$

and

$$E(X^2)=0\times\frac{1}{8}+1\times\frac{3}{8}+4\frac{3}{8}+9\times\frac{1}{8}=3.$$

Thus, $V(X)=E(x^2)-E(X)^2=\frac{3}{4}$ and $\sigma(X)=\frac{\sqrt{3}}{2}$.

Variance and Standard Deviation

Proposition 19

Let X and Y be two independent random variables on a probability space (S, p). Then

$$V(X+Y) = V(X) + V(Y).$$

Proof.

By Proposition 14, $E(XY) = E(x)E(Y)$. It then follows from Proposition 17 and Theorem 11 that

$$\begin{aligned} V(X+Y) &= E((X+Y)^2) - (E(X+Y))^2 \\ &= E(X^2 + Y^2 + 2XY) - (E(X) + E(Y))^2 \\ &= E(X^2) + E(Y^2) + 2E(XY) - E(X)^2 - E(Y)^2 - 2E(X)E(Y) \\ &= V(X) + V(Y). \end{aligned}$$

□

The Expected Value, Variance and Standard Deviation of Bernoulli Trials

Proposition 20

In n Bernoulli trials with probability p of success, the expected value is np, the variance is $np(1-p)$ and the standard deviation is $\sqrt{np(1-p)}$.

Proof.

We consider the case $n = 1$ and let X denote the number of successes in one Bernoulli trial. Then $p(X = 1) = p$ and $p(X = 0) = (1-p)$. Hence $E(X) = 1 \times p + 0 \times (1-p) = p$. Similarly, $E(X^2) = 1^2 \times p + 0^2 \times (1-p) = p$. It then follows that $V(X) = E(X^2) - E(X)^2 = p - p^2 = p(1-p)$.
Let X_i denote the number of success in the ith Bernoulli trial. Then X_1, X_2, $\cdots$, X_n are independent random variables. It then follows from Proposition 19 that

$$E(X_1 + X_2 + \cdots + X_n) = nE(X_1) = np,$$
$$V(X_1 + X_2 + \cdots + X_n) = nV(X_1) = np(1-p).$$

$\square$

The Central Limit Theorem

Theorem 21

*Let $X_1, X_2, \cdots, X_n$ be **independent and identically distributed** (in short, i.i.d.) random variables on a probability space (S, p) (i.e., they all have the same distribution and are mutually independent). Let μ and σ the expected value and the standard deviation of all X_i. Define*

$$Z = \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}.$$

Then

$$p(a \le Z \le b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx.$$

Remark

A proof of this theorem can be found in a textbook on probability.

The Normal Distribution

Definition 22

A random variable X on a sample space S is said to have the **normal distribution** if its probability distribution is given by

$$p(X=x)=\varphi_{(\mu,\sigma)}(x):=\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/2\sigma^2},\quad x\in\mathbb{R}.$$

Remarks

- The normal distribution has expected value μ and standard deviation σ.
- It is called the **standard normal distribution** when $\mu=0$ and $\sigma=1$.
- $\int_{-\infty}^{\infty}\varphi_{(\mu,\sigma)}(x)dx=1$.
- The Central Limit Theorem says that the probability distribution of

$$Z=\frac{X_1+X_2+\cdots+X_n-n\mu}{\sigma\sqrt{n}}$$

will tend to the standard normal distribution as n increases.

An Example of the Normal Distribution

Example 23

Consider n Bernoulli trials with probability p of success. Let X_i denote the number of success in the ith trial. Then $X_1, X_2, \cdots, X_n$ are i.i.d., and each has expected value p and standard deviation $\sqrt{p(1-p)}$.

One can verify that the probability distribution of the **standardized random variable**

$$Z = \frac{X_1 + X_2 + \cdots + X_n - np}{\sqrt{np(1-p)}} = \frac{X_1 + X_2 + \cdots + X_n - E(X_1 + X_2 + \cdots + X_n)}{\sigma(X_1 + X_2 + \cdots + X_n)}$$

converges to the standard normal distribution as n approaches to ∞.