Probability and Statistics (Pure)

Rui Yang

2024-01-05

Table of contents

1 From the book *Calculus with applications* by *Peter D. Lax*

1.1 Probability

Probability is the branch of mathematics that deals with **events** whose **individual outcomes** are **unpredictable**, but whose outcomes **on average** are **predictable**.

Experiments can be divided into two types:

- **Deterministic:** whose individual outcomes are predictable;

- **Nondeterministic (random):** whose individual outcomes are unpredictable.

1.1.1 Discrete probability

Next, all experiments we'll deal with are *repeatable* (it can be performed repeatedly any number of times) and *random* (any single performance of the experiment is unpredictable).

In this section, we'll deal with experiments having *a finite number of possible outcomes*. We denote the number of possible outcomes by n, and number them from 1 to n.

1.1.1.1 The probability of any single outcome

In a random and repeatable experiment, if we denote S_j by *the number of instances* among the first N experiments where the jth outcome was observed to occur, then the frequency $\frac{S_j}{N}$ with which the jth outcome has been observed to occur tends to a limit as N tends to infinity. We call this limit the probability of the jth outcome and denote it by p_j:

$$p_j = \lim_{N \to \infty} \frac{S_j}{N}$$

These probabilities have the following properties:

- $0 \le p_j \le 1$;

> **i Note**
>
> For $\frac{S_j}{N}$ lies between 0 and 1, and therefore so does its limit p_j.

- $\sum_{j=1}^{n} p_j = 1$.

> **i Note**
>
> We have
>
> $$S_1 + S_2 + \cdots + S_n = N$$
>
> Dividing by N, we get
>
> $$\frac{S_1}{N} + \frac{S_2}{N} + \cdots + \frac{S_n}{N} = 1$$
>
> As N tends to infinity, we have
>
> $$\lim_{N \to \infty} \sum_{j=1}^{n} \frac{S_j}{N} = \lim_{N \to \infty} \sum_{j=1}^{n} p_j = 1$$

1.1.1.2 The probability of an event

In fact very often, we are not interested in all the details of the outcome of an experiment, but merely in a particular aspect of it (e.g. throwing a die, we may be interested only in whether the outcome is even or odd).

An occurrence such as throwing an even number is called an *event*, which is defined as the following:

An event E is defined as **any collection of possible outcomes**.

Note: we say that an event E occurred whenever any outcome belonging to E occurred.

The probability $p(E)$ of an event E is

$$p(E) = \lim_{N \to \infty} \frac{S(E)}{N}$$

where $S(E)$ is the number of instances among the first N experiments when the event E took place.

> **i Note**
>
> We have
>
> $$S(E) = \sum_{j \text{ in } E} S_j$$
>
> Dividing by N
>
> $$\frac{S(E)}{N} = \sum_{j \text{ in } E} \frac{S_j}{N}$$
>
> As N tends to infinity, we have
>
> $$p(E) = \sum_{j \text{ in } E} p_j$$

1.1.1.2.1 The arithmetic rules of probability of some special events

1. Addition rule for disjoint events

Two events E_1 and E_2 are called *disjoint* if both cannot take place simultaneously (i.e. $E_1 \cap E_2 = \emptyset$).

Then we have

$$p(E_1 \cup E_2) = p(E_1) + p(E_2)$$

> **i** Note
>
> $$p(E_1 \cup E_2) = \sum_{j \text{ in } E_1 \text{ or } E_2} p_j$$
>
> Disjointness means that an outcome j may belong either to E_1 or to E_2 **but not to both**; therefore,
>
> $$p(E_1 \cup E_2) = \sum_{j \text{ in } E_1 \text{ or } E_2} p_j = \sum_{j \text{ in } E_1} p_j + \sum_{j \text{ in } E_2} p_j = p(E_1) + p(E_2)$$

2. Product rule for independent events

Two events E and F are called *independent* if the outcome of one cannot influence the other, nor is the outcome of both under the influence of a common cause.

Then we have

$$p(E \cap F) = p(E)p(F)$$

> **i** Note
>
> Among the first N experiments, count the number of times E has occurred ($S(E)$), F has occurred ($S(F)$), and $E \cap F$ has occurred ($S(E \cap F)$). Then we have
>
> $$p(E) = \lim_{N \to \infty} \frac{S(E)}{N}$$
>
> $$p(F) = \lim_{N \to \infty} \frac{S(F)}{N}$$
>
> $$p(E \cap F) = \lim_{N \to \infty} \frac{S(E \cap F)}{N}$$
>
> Suppose that we single out from the sequence of combined experiments the subsequence of those where E occurred. The frequency of occurrence of F in this subsequence is $\frac{S(E \cap F)}{S(E)}$. If the two events E and F are truly independent, the frequency with which F occurs in this subsequence should be the same as the frequency with which F occurs in the original sequence, i.e.
>
> $$\lim_{N \to \infty} \frac{S(E \cap F)}{S(E)} = \lim_{N \to \infty} \frac{S(F)}{N} = p(F)$$

We write the frequency of $\frac{S(E\cap F)}{N}$ as the product

$$\frac{S(E\cap F)}{N} = \frac{S(E\cap F)}{S(E)}\frac{S(E)}{N}$$

Then we have

$$p(E\cap F) = \lim_{N\to\infty}\frac{S(E\cap F)}{N} = \lim_{N\to\infty}\frac{S(E\cap F)}{S(E)}\cdot\lim_{N\to\infty}\frac{S(E)}{N} = p(F)p(E)$$

1.1.1.3 Characteristics of random variables

1. Numerical outcome

The numerical outcome of an experiment means the assignment of a real number x_j to each of the possible outcomes.

Note that different outcomes may be assigned the same number so we have to re-calculate the probability $p(x_j)$ for each x_j with which x_j occurred.

2. Expectation

In a random experiment with n possible outcomes of probability p_j and numerical outcome x_j, the average numerical outcome, called the mean of x or expectation of x, denoted by $\bar{x}$ or $E(x)$, is given by the formula

$$\bar{x} = E(x) = p_1 x_1 + p_2 x_2 + \cdots + p_n x_n$$

> **i Note**
>
> Among the first N experiments, denote by S_j the number of instances with which the jth outcome was observed. Then, the average numerical outcome is
>
> $$\frac{S_1 x_1 + S_2 x_2 + \cdots + S_n x_n}{N}$$
>
> As $N\to\infty$, we have
>
> $$\bar{x} = E(x) = \lim_{N\to\infty}\frac{S_1 x_1 + S_2 x_2 + \cdots + S_n x_n}{N} = p_1 x_1 + p_2 x_2 + \cdots + p_n x_n$$

3. Variance

Next we are tempted to know such a fact: by how much do the numerical outcomes differ on average from the mean?

This is characterized by the variance, the expectation of the square of the difference of the numerical outcome and its mean:

$$V = \overline{(x - \bar{x})^2} = E((x - \bar{x})^2)$$

> **i Note**
>
> $$
> \begin{aligned}
> V &= \overline{(x - \bar{x})^2} \\
> &= E((x - \bar{x})^2) \\
> &= \sum_{j=1}^{n} p_j (x_j - \bar{x})^2 \\
> &= p_1 x_1^2 + \cdots + p_n x_n^2 - 2(p_1 x_1 + \cdots + p_n x_n)\bar{x} + (\bar{x})^2 \\
> &= \overline{x^2} - (\bar{x})^2 \\
> &= E(x^2) - (E(x))^2
> \end{aligned}
> $$

Note: the square root of the variance $\sqrt{V}$ is called the standard deviation.

1.1.1.4 Some special distributions

1. The binomial distribution

Suppose a random experiment has two possible outcomes A and B, with probabilities p and q respectively, where $p + q = 1$.

Suppose we repeat the experiment N times, and the repeated experiments are independent of each other.

If we let k $(k = 0, 1, ..., N)$ denote the number of times with which A occurs, then the probability that A occurs exactly k times is given by the formula

$$b_k(N) = \binom{N}{k} p^k q^{N-k}$$

> **i Note**
>
> Since the outcomes of the experiments are independent of each other, the probability of a particular sequence of k A's and $N - k$ B's is $p^k q^{N-k}$.

> In addition, there are exactly $\binom{N}{k}$ arrangements of k A's and $N - k$ B's, which are disjoint of each other.

In addition, we have

$$E(x) = \sum_{k=0}^{N} kp(x = k) = Np$$

Note: the binomial theorem is $(a + b)^N = \sum_{k=0}^{N} \binom{N}{k} a^k b^{N-k}$.

1. The Poisson distribution

Suppose **each week** there are **a large number of vehicles** through a busy intersection and there are **on average** μ accidents.

Suppose the probability of a vehicle having an accident is independent of the occurrence of previous accidents.

Then we use a binomial distribution to determine the probability of k accidents in a week:

Setting $p = \frac{\mu}{N}$, then we have

$$
\begin{aligned}
b_k(N) &= \binom{N}{k} p^k q^{N-k} \\
&= \frac{N(N-1)\cdots(N-k+1)}{k!} p^k q^{N-k} \\
&= (1 - \frac{1}{N}) \cdots (1 - \frac{k-1}{N}) \frac{N^k p^k (1-p)^{N-k}}{k!} \\
&= \frac{(1 - \frac{1}{N}) \cdots (1 - \frac{k-1}{N})}{(1-p)^k} \frac{\mu^k}{k!} (1 - \frac{\mu}{N})^N
\end{aligned}
$$

Since $e^{-x} = \lim_{n \to \infty} (1 - \frac{x}{n})^n$ $(e^x = \lim_{n \to \infty} (1 + \frac{x}{n})^n)$, then we have

$$\lim_{N \to \infty,\ \mu=Np} b_k(N) = \frac{\mu^k}{k!} e^{-\mu}$$

This gives us an estimate for $b_k(N)$ when N is **large**, p is **small**, and $Np = \mu$.

The Poisson distribution is defined as

$$p_k(\mu) = \frac{\mu^k}{k!} e^{-\mu}$$

where μ is a parameter. p_k is the probability of k favorable outcomes, $k = 0, 1, ...$.

In addition, the combination of two Poisson processes is again a Poisson process.

Denote by $p_k(\mu)$ and $p_k(\nu)$ the probability of k favorable outcomes in these two processes. We claim that the probability of k favorable outcomes when both experiments are performed is $p_k(\mu + \nu)$.

> **i** Note
>
> There will be k favorable outcomes for the combined experiment if the first experiment has j ($j = 0, 1, ..., k$) favorable outcomes and the second experiment has $k - j$.
> If the experiments are independent, the probability of such a combined outcome is the product of the probabilities $p_j(\mu)p_{k-j}(\nu)$.
> So the probability of the combined experiment to have k favorable outcomes is the sum
>
> $$\sum_j p_j(\mu)p_{k-j}(\nu) = \sum_j \frac{\mu^j}{j!}e^{-\mu}\frac{\nu^{k-j}}{(k-j)!}e^{-\nu}$$
> $$= \frac{1}{k!}e^{-(\mu+\nu)}\sum_j \frac{k!}{j!(k-j)!}\mu^j\nu^{(k-j)}$$
> $$= \frac{(\mu+\nu)^k}{k!}e^{-(\mu+\nu)}$$
>
> which is the Poisson distribution $p_k(\mu + \nu)$.

1.1.2 Continuous probability

Suppose we have such an experiment making a physical measurement with an apparatus subject to random disturbances that can be reduced but not totally eliminated. Then every real number is a possible numerical outcome of such an experiment.

Repeat the experiment as many times as we wish and denote by $S(x)$ the number of instances among the first N performances for which the numerical outcome was **less than** x. Then the frequency $\frac{S(x)}{N}$ with which this event occurs tends to a limit as N tends to infinity. This limit is **the probability** that **the outcome is less than** x, and is denoted by $P(x)$:

$$P(x) = \lim_{N \to \infty} \frac{S(x)}{N}$$

The probability $P(x)$ has the following properties:

1. $0 \leq P(x) \leq 1$: $0 \leq S(x) \leq N \implies 0 \leq \frac{S(x)}{N} \leq 1 \implies 0 \leq P(x) \leq 1$.

2. $P(x)$ is a **nondecreasing** function of x: $S(x)$ is a **nondecreasing** function of x, so that $\frac{S(x)}{N}$ is a **nondecreasing** function of x; then so is the limit $P(x)$.

3. $P(x) \to 0 \ (x \to -\infty)$.

4. $P(x) \to 1 \ (x \to \infty)$.

5. $P(x)$ is a continuously differentiable function (denote the derivative of P by p): $\frac{\mathrm{d}P(x)}{\mathrm{d}x} = p(x)$.

The function $p(x)$ is called the **probability density function**.

> **ℹ Addition rule for disjoint events**
>
> Suppose E and F are two events with probabilities $P(E)$ and $P(F)$ respectively. Suppose they are **disjoint** (i.e. $E \cap F = \emptyset$).
> Then we have
>
> $$P(E \cup F) = P(E) + P(F)$$
>
> Let $E : x < a$, $F : a \leq x < b$, then we have $E \cup F : x < b$.
> Then we have
>
> $$P(E) = P(a), \ P(E \cup F) = P(b)$$
>
> We conclude that
>
> $$P(F) = P(b) - P(a)$$
>
> is the probability of $a \leq x < b$.

According to the **mean value theorem**, for every a and b, there is a number c lying between a and b such that

$$P(b) - P(a) = p(c)(b - a)$$

According to the **fundamental theorem of calculus**

$$P(b) - P(a) = \int_a^b p(x)\mathrm{d}x$$

According to $P(a) \to 0 \ (a \to -\infty)$

$$P(b) = \int_{-\infty}^b p(x)\mathrm{d}x$$

According to $P(b) \to 1$ $(b \to \infty)$

$$1 = \int_{-\infty}^{\infty} p(x)\mathrm{d}x$$

This is the continuous analogue of the basic fact that $p_1 + p_2 + \cdots + p_n = 1$ in discrete probability.

7. $p(x) \geq 0$: $P(x)$ is a nondecreasing function of x.

8. The expectation is:

$$\bar{x} = \int_{-\infty}^{\infty} xp(x)\mathrm{d}x$$

> **ℹ Note**
>
> Imagine the experiment performed as many times as we wish, and denote the sequence of outcomes by
>
> $$a_1, a_2, ..., a_N, ...$$
>
> Divide the interval I **in which all outcomes lie** into n subintervals $I_1, ..., I_n$. Denote the endpoints by
>
> $$e_0 < e_1 < \cdots < e_n$$
>
> The probability of $e_{j-1} \leq x < e_j$ (i.e. x lies in the interval I_j) is
>
> $$P_j = P(e_j) - P(e_{j-1}) = p(x_j)(e_j - e_{j-1})$$
>
> where x_j is a point in I_j guaranteed by the mean value theorem, and $e_j - e_{j-1}$ denotes the length of I_j.
> We now simplify the original experiment by recording merely the intervals I_j in which the outcome falls, and calling the numerical outcome in this case x_j, the point in I_j appears in the above formula; therefore, the actual outcome falling into I_j differs from x_j by at most $e_j - e_{j-1}$.
> Now consider the sequence of outcomes $a_1, a_2, ...$ of the **original experiment**. Denote the corresponding outcomes of the simplified experiment by $b_1, b_2, ...$. The simplified experiment has a finite number of outcomes. For such discrete experiments, we have the expectation
>
> $$\bar{x}_n = \lim_{N \to \infty} \frac{b_1 + \cdots + b_N}{N}$$

where n is the number of subintervals of I.

In fact, the expectation $\bar{x}_n$ of the simplified experiment can also be calculated by formula

$$\bar{x}_n = P_1 x_1 + \cdots + P_n x_n$$

Then we have

$$\bar{x}_n = p(x_1)x_1(e_1 - e_0) + \cdots + p(x_n)x_n(e_n - e_{n-1})$$

As $n \to \infty$, we have

$$\bar{x} = \lim_{n\to\infty} \sum_{i=1}^{n} x_i p(x_i)\Delta x_i = \int_{e_0}^{e_n} xp(x)\mathrm{d}x$$

So we conclude

$$\bar{x} = \int_{-\infty}^{\infty} xp(x)\mathrm{d}x$$

9. The probability density of a combined experiment of two experiments independent of each other:

- Case 1: the outcome of the first experiment may be any real number, but the second experiment can have only a finite number of outcomes.

Denote by $P(a)$ the probability of $x < a$. The second experiment has n possible outcomes $a_1, ..., a_n$ with probabilities $Q_1, ..., Q_n$.

We define *the numerical outcome of the combined experiment* to be *the sum of the separate numerical outcomes of the two experiments* that constitute it.

We denote by $E(x)$ the event that the numerical outcome of the combined experiment is less than x, and denote its probability by $U(x)$.

Then we have

$$U(x) = Q_1 P(x - a_1) + \cdots + Q_n P(x - a_n)$$

> **i Note**
>
> We denote by $E_j(x)$ the event that the numerical outcome of the second experiment is a_j. The numerical outcome of the combined experiment is then less than x if and only if the numerical outcome of the first experiment is less than $x - a_j$.
> Then we have

$$E(x) = E_1(x) \cup \cdots \cup E_n(x)$$

where the events $E_j(x)$ are disjoint.

It follows from **the addition rule for disjoint events** that

$$U(x) = P(E_1(x)) + \cdots + P(E_n(x))$$

Since the two experiments are independent, we have $P(E_j(x)) = Q_j P(x - a_j)$.

So we have

$$U(x) = Q_1 P(x - a_1) + \cdots + Q_n P(x - a_n)$$

- Case 2: both experiments can have any real number as outcome.

Denote by $P(a)$ and $Q(a)$ the probabilities that the outcome is less than a in each of the two experiments, respectively.

Assume the outcome of the second experiment always lies in some finite interval I. Then we subdivide I into a finite number n of subintervals $I_j = [e_{j-1}, e_j)$. Let Q_j denote the probability of the outcome of the experiment lying in I_j.

Suppose $Q(x)$ is continuously differentiable and denote its derivative by $q(x)$.

According to the mean value theorem, we have

$$Q_j = Q(e_j) - Q(e_{j-1}) = q(a_j)(e_j - e_{j-1})$$

where a_j is some point in I_j.

We discretize the second experiment by lumping together all outcomes that lie in I_j and redefine the numerical outcome in that case to be a_j.

Then we have

$$U_n(x) = q(a_1)P(x - a_1)(e_1 - e_0) + \cdots + q(a_n)P(x - a_n)(e_n - e_{n-1})$$
$$= \sum_{i=1}^{n} q(a_i)p(x - a_i)\Delta a_i$$

As $n \to \infty$, we have

$$U(x) = \lim_{n \to \infty} U_n(x)$$

$$= \lim_{n \to \infty} \sum_{i=1}^{n} q(a_i)p(x - a_i)\Delta a_i$$

$$= \int_I q(a)P(x - a)\mathrm{d}a$$

Then we have

$$U(x) = \int_{-\infty}^{\infty} q(a)P(x - a)\mathrm{d}a$$

Further, let us suppose $P(x)$ is continuously differentiable, and denote its derivative by $p(x)$.

Then we have

$$u(x) = \int_{-\infty}^{\infty} q(a)p(x - a)\mathrm{d}a$$

where $u(x)$ is the derivative of $U(x)$.

In a word, we have proved the following fact:

> **ℹ Note**
>
> Consider two independent experiments whose outcomes lie in some finite interval and have probability p and q respectively.
> In the combined experiment of the two experiments, define the outcome of the combined experiment to be the sum of the outcomes of the individual experiments.
> Then the combined experiment has the probability density:
>
> $$u(x) = \int_{-\infty}^{\infty} q(a)p(x - a)\mathrm{d}a$$

10. The convolution of the functions q and p:

The function u defined by $u(x) = \int_{-\infty}^{\infty} q(a)p(x - a)\mathrm{d}a$ is called the convolution of the functions q and p. This relation is denoted by $u = q * p$.

The convolution has the following properties:

> **ℹ Note**
>
> Let q_1, q_2, and p be continuous functions defined for all real numbers x, and assume the functions are 0 outside a finite interval. Then we have
>
> - Convolution is distributive: $(q_1 + q_2) * p = q_1 * p + q_2 * p$.
>
> - Let k be any constant. Then $(kq) * p = k(q * p)$.
>
> - Convolution is commutative: $q * p = p * q$.
>
> - The integral of the convolution is the product of the integrals of the factors:
>
> $$\int_{-\infty}^{\infty} u(x)\mathrm{d}x = \int_{-\infty}^{\infty} p(x)\mathrm{d}x \int_{-\infty}^{\infty} q(a)\mathrm{d}a$$