AE1MCS: Mathematics for Computer Scientists

Huan Jin University of Nottingham Ningbo China

Reading

Kenneth H. Rosen, *Discrete Mathematics and Its Applications*, 7th Edition, 2013.

- Chapter 7, Section 7.2 Probability Theory
- Chapter 7, Section 7.3 Bayes' Theorem
- Chapter 7, Section 7.4 Expected Value and Variance

Probability Distribution

Let s be the sample space of an experiment with a finite or countable number of outcomes. We assign a probability $p(s)$ to each outcome. We require that two conditions be met:

1. $0 \leq p(s) \leq 1$ for each $s \in S$
2. $\sum_{s \in S} p(s) = 1$.

The function p from the set of all outcomes of the sample space S is called a **probability distribution**.

Conditional Probability

Given an event F occurs, the probability that event E occurs is the **conditional probability** of E given F.

Let E and F be events with $p(F) > 0$. The **conditional probability** of E given F, denoted by $p(E|F)$, is defined as:

$$p(E|F) = \frac{p(E \cap F)}{p(F)}$$

.

Conditional Probability

A bit string of length four is generated at random so that each of the 16 bit strings of length four is equally likely. What is the probability that it contains at least two consecutive 0s, given that its first bit is a 0?

Solution: Let E be the event that a bit of length four contains at least two consecutive 0s,
Let F be the event that the first bit of a bit string of length four is a 0.
The probability that a bit string of length four has at least two consecutive 0s, given that its first bit is a 0, equals:

$$p(E|F) = \frac{p(E \cap F)}{p(F)}$$

Conditional Probability

A bit string of length four is generated at random so that each of the 16 bit strings of length four is equally likely. What is the probability that it contains at least two consecutive 0s, given that its first bit is a 0?

Solution: Let E be the event that a bit of length four contains at least two consecutive 0s,
Let F be the event that the first bit of a bit string of length four is a 0.
The probability that a bit string of length four has at least two consecutive 0s, given that its first bit is a 0, equals:

$$p(E|F) = \frac{p(E \cap F)}{p(F)}$$

Because $E \cap F = \{0000, 0001, 0010, 0011, 0100\}$, then $p(E \cap F) = \frac{5}{16}$. Because there are 8 bit strings of length four that start with a 0, we have $p(F) = \frac{8}{16} = \frac{1}{2}$.

$$p(E|F) = \frac{5/16}{1/2} = \frac{5}{8}$$

Independence

When two events are independent, the occurrence of one of the events gives no information about the probability of that the other event occurs.

The events E and F are independent **if and only if** $p(E \cap F) = p(E)p(F)$.

Independence

Suppose E is the event that a randomly generated bit string of length four begin with a 1 and F is the event that this bit string contains an even number of 1s. Are E and F independent, if the 16 bit strings of length four are equally likely?

Solution: There are eight bit strings of length four that begin with a one: 1000, 1001, 1010, 1011, 1100, 1101, 1110, and 1111. There are also eight bit strings of length four that contain an even number of ones: 0000, 0011, 0101, 0110, 1001, 1010, 1100, 1111. Because there are 16 bit strings of length four, it follows that

$$p(E) = p(F) = 8/16 = 1/2.$$

Because $E \cap F = 1111, 1100, 1010, 1001$, we see that $p(E \cap F) = 4/16 = 1/4$.

Because $p(E \cap F) = 1/4 = (1/2)(1/2) = p(E)p(F)$,we conclude that E and F are independent.

Bayes' Theorem

Suppose we know p(F), the probability that an event F occurs, but we have knowledge that an event E occurs.
The conditional probability that F occurs given that E occurs, $p(F|E)$

Bayes' Theorem

Bayes' Theorem

Suppose that E and F are events from a sample space S such that $P(E) \neq 0$ and $P(F) \neq 0$. Then

$$p(F|E) = \frac{p(E|F)p(F)}{p(E|F)p(F) + p(E|\overline{F})p(\overline{F})}$$

Bayesian Spam Filters

Suppose that we have found that the word "Rolex" occurs in 250 of 2000 messages known to be spam and in 5 of 1000 messages known not to be spam. Estimate the probability that an incoming message containing the word "Rolex" is spam, assuming that it is equally likely that an incoming message is spam or not spam. If our threshold for rejecting a message as spam is 0.9, will we reject such messages?

Solution: We use the counts that the word "Rolex" appears in spam messages and messages that are not spam to find that p(Rolex) = 250/2000 = 0.125 and q(Rolex) = 5/1000 = 0.005.
Because we are assuming that it is equally likely for an incoming message to be spam as it is not to be spam, we can estimate the probability that an incoming message containing the word "Rolex" is spam by

$$r(Rolex) = \frac{p(Rolex)}{p(Rolex) + q(Rolex)} = \frac{0.125}{0.125 + 0.005} = 0.962$$

Generalizing Bayes' Theorem

Suppose that E is an event from a sample space S and that $F_1, F_2, ..., F_n$ are mutually exclusive events such that $\cup_{i=1}^{n} F_i = S$. Assume that $p(E) \neq 0$ and $p(F_i) \neq 0$ for $i = 1, 2, ..., n$. Then

$$P(E) = \sum_{i=1}^{n} p(E|Fi)p(Fi)$$

$$p(F_j|E) = \frac{p(E|F_j)p(F_j)}{\sum_{i=1}^{n} p(E|Fi)p(Fi).}$$

Expected Value and Variance

The **expected value**, also called the expectation or mean, of the random variable X on the sample space S:

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

If X is a random variable and $p(X = r)$ is the probability that $X = r$, so that $p(X = r) = \sum_{s \in S, X(s) = r} p(s)$, then

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

Linearity of Expectations

If X_i, $i = 1, 2, ..., n$ with n a positive integer, are random variables on S, and if a and b are real numbers, then

1. $E(X_1 + X_2 + \cdots + Xn) = E(X_1) + E(X_2) + \cdots + E(X_n)$
2. $E(aX + b) = aE(X) + b.$

Independent Random Variables

The random variables X and Y on a sample space S are independent if

$$p(X = r1 \text{ and } Y = r2) = p(X = r1) \cdot p(Y = r2)$$

or, if the probability that $X = r1$ and $Y = r2$ equals the product of the probabilities that $X = r1 and Y = r2$, for all real numbers $r1$ and $r2$.

Corollary

If X is independent of Y, then

$$E(XY) = E(X) \cdot E(Y) \tag{1}$$

If $X_1, X_2, \ldots, X_n$ are mutually independent, then,

$$E(X_1 X_2 \ldots X_n) = E(X_1)E(X_2) \ldots E(X_n) \tag{2}$$

Bernoulli Trials and Binomial Distribution

Each performance of an experiment with two possible outcomes is called a **Bernoulli trial**.
In general, a possible outcome of a Bernoulli trial is called a **success** or a **failure**.

- Generate a bit, $\{0, 1\}$.
- Flip a coin, $\{$Heads, tails$\}$.

If p is the probability of a success and q is the probability of a failure, it follows that $p + q = 1$.

Binomial Distribution

- The probability of exactly k successes in n independent Bernoulli trials, with probability of success p and probability of failure $q = 1 - p$, is $C(n, k)p^k q^{n-k}$.
- We denote by $b(k; n, p)$ the probability of k successes in n independent Bernoulli trials with probability of success p and probability of failure $q = 1 - p$.
- Considered as a function of k, we call this function the **binomial distribution.**

$$b(k; n, p) = C(n, k)p^k q^{n-k}.$$

Binomial Distribution

Example

A coin is biased so that the probability of heads is 2/3. What is the probability that exactly four heads come up when the coin is flipped seven times, assuming that the flips are independent?

Expected Value: Extra Exercises (Not required!)

Bernoulli trials:
The expected number of successes when n mutually independent Bernoulli trials are performed, where p is the probability of success on each trial, is np.

Proof: Let X be the random variable equal to the number of successes in n trials. By Theorem 2 of Section 7.2 we see that $p(X = k) = C(n, k)p^k q^{n-k}$. Hence, we have

$$
\begin{aligned}
E(X) &= \sum_{k=1}^{n} k p(X = k) && \text{by Theorem 1} \\
&= \sum_{k=1}^{n} k C(n, k) p^k q^{n-k} && \text{by Theorem 2 in Section 7.2} \\
&= \sum_{k=1}^{n} n C(n-1, k-1) p^k q^{n-k} && \text{by Exercise 21 in Section 6.4} \\
&= np \sum_{k=1}^{n} C(n-1, k-1) p^{k-1} q^{n-k} && \text{factoring } np \text{ from each term} \\
&= np \sum_{j=0}^{n-1} C(n-1, j) p^j q^{n-1-j} && \text{shifting index of summation with } j = k-1 \\
&= np(p+q)^{n-1} && \text{by the binomial theorem} \\
&= np. && \text{because } p + q = 1
\end{aligned}
$$

This completes the proof because it shows that the expected number of successes in n mutually

Geometric Distribution

A random variable X has a geometric distribution with parameter p if $p(X = k) = (1 - p)^{k-1}p$ for $k = 1, 2, 3, \ldots,$ where p is a real number with $0 \leq p \leq 1$.

$$E(X) = 1/p$$

Geometric Distribution

Suppose that the probability that a coin comes up tails is p. This coin is flipped repeatedly until it comes up tails. What is the expected number of flips until this coin comes up tails?
The random variable X that equals the number of flips expected before a coin comes up tails is an example of a random variable with a geometric distribution.

System Fail Problem

A system fails with probability p at each step. we assume mutually independent between each step. What is the expected number of steps before the system Fail?

Variance

Betting game.
Betting 10RMB.
Betting 10,000RMB.

Variance

Variance provides a measure of how widely X is distributed about its expected value.

Definition

Let X be a random variable on a sample space S. The variance of X, denoted by $Var(X)$, is

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

That is, Var(X) is the weighted average of the square of the deviation of X. The standard deviation of X, denoted $\sigma(X)$, is defined to be $\sqrt{Var(X)}$.

Variance

Theorem

If X is a random variable on a sample space S, then
$Var(X) = E(X^2) - E(X)^2$

Corollary

If X is a random variable on a sample space S and $E(X) = \mu$, then
$Var(X) = E((X - \mu)^2)$.

How to prove it?

Example: Rolling a Die

Let X be the number that comes up when a fair die is rolled. What is the expected value and variance of X?

Variance for the sum of random variables

If X and Y are independent variable,

$$Var(X + Y) = Var(X) + Var(Y)$$

In addition,

$$Var(aX + b) = a^2 Var(X)$$

Poisson Distribution

Examples

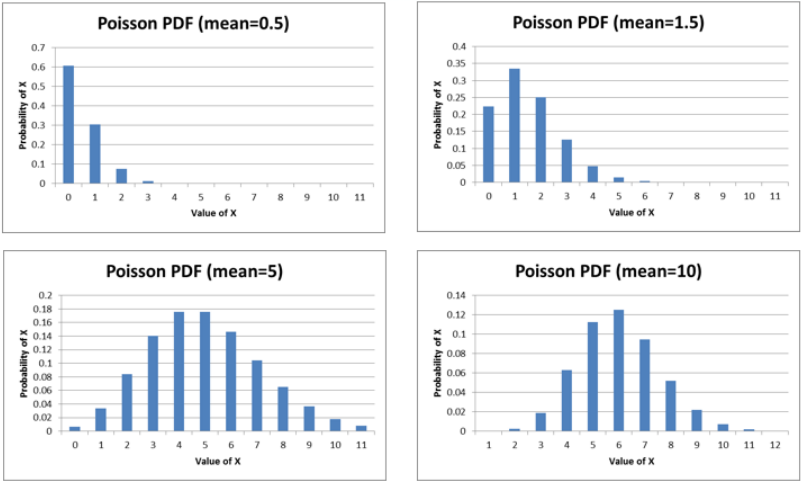
- Process with random arrivals - probability of seeing x events within a certain time period
- Number of slow moving items sold per day, week, month in a store

$$P(X = i; \lambda) = \begin{cases} \frac{e^{-\lambda}\lambda^i}{i!}, & \text{for } i = 0, 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

For Poisson Distribution, E(X)=Var(X)=λ

Source:
https://en.wikipedia.org/wiki/Poisson_distribution

Poisson Distribution (Not required!)



Example of Poisson Distribution (Not required!)

Suppose that trains arrive at Ningbo Station P(1.5) per 10 minute increments. That is, 1.5 trains on average arrive every 10 minutes and the variance is 1.5 and the distribution appears to be Poisson.

- What is the probability that I will see 3 or more trains in 10 minutes?
- What is the probability none will come by?

Discrete Probability

- Uniform (e.g. rolling a die)
- Bernoulli (Success or Failure)
- Binomial (Number of successes in fixed number of trials)
- Geometric (Number of trials until success)
- Poisson (Number of arrivals in fixed time interval)
- Many others...

Reading

Kenneth H. Rosen, *Discrete Mathematics and Its Applications*, 7th Edition, 2013.

- Chapter 7, Section 7.3 Bayes' Theorem
- Chapter 7, Section 7.4 Expected Value and Variance