AE1MCS: Mathematics for Computer Scientists

Huan Jin
University of Nottingham Ningbo China

Oct 2021

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

Bayes' Theorem

$$P(E) = P(E \cap F) + P(E \cap \bar{F})$$
$$= P(E|F)P(F) + P(E|\bar{F}) \cdot P(\bar{F})$$

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, $p(F|E)$

$$P(F|E) = \frac{P(E \cap F)}{P(E)} = \frac{P(E \cap F)}{P(E \cap F) + P(E \cap \bar{F})}$$

$$\boxed{= \frac{P(E|F)\,P(F)}{P(E|F) \cdot P(F) + P(E|\bar{F}) \cdot P(\bar{F})}}$$

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

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

Generalizing Bayes' Theorem

$F_i \cap F_j = \phi$

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

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

Bayesian Spam Filters

✓ $P(\text{Rolex} \mid \text{Spam}) = \frac{250}{2000} = 0.125$ $P(\text{spam}) = P(\text{Not spam}) = \frac{1}{2}$

✓ $P(\text{Rolex} \mid \text{not Spam}) = \frac{5}{1000} = 0.005$

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?

$$P(\text{Spam} \mid \text{Rolex}) = \frac{P(\text{Rolex} \mid \text{spam}) \cdot P(\text{spam})}{P(\text{Rolex} \mid \text{spam}) \cdot P(\text{spam}) + P(\text{Rolex} \mid \text{Not spam}) \cdot \underbrace{}_{P(\text{Not spam})}}$$

$$= \frac{0.125 \times \frac{1}{2}}{0.125 \times \frac{1}{2} + 0.005 \times \frac{1}{2}} = 0.962$$

Bayesian Spam Filters

Suppose that we train a Bayesian spam filter on a set of 2000 spam messages and 1000 messages that are not spam. The word “stock” appears in 400 spam messages and 60 messages that are not spam, and the word “undervalued” appears in 200 spam messages and 25 messages that are not spam. Estimate the probability that an incoming message containing both the words “stock” and “undervalued” is spam, assuming that we have no prior knowledge about whether it is spam. Will we reject such messages as spam when we set the threshold at 0.9?

Homework

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

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

Rolling a Die.

D	$P(D)$
1	$1/6$
2	$1/6$
3	$1/6$
4	$1/6$
5	$1/6$
6	$1/6$

$$E(D) = P(D=1)\cdot 1 + P(D=2)\cdot 2 + P(D=3)\cdot 3$$

$$+ \cdots\cdots + P(D=6)\cdot 6$$

$$= \frac{1}{6}(1+2+3+4+5+6)$$

$$= \underline{3.5}.$$

Expected Value

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.

X	P
0	$p(X=0) = C(n,0)\cdot p^0 \cdot (1-p)^n$
1	$C(n,1)\cdot p(1-p)^{n-1}$
2	
3	
$\vdots$	$\vdots$
n	

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

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

Expectation

Corollary

If X is a random variable and $P(X = i)$ is the probability that $X = i$, then

0

$$E(X) = \sum_{i=1}^{\infty} iP(X = i)$$

Theorem

If $X : S \to \mathbb{N}$, then

$$E(X) = \sum_{i=0}^{\infty} P(X > i) = \sum_{i=1}^{\infty} P(X \geq i)$$

System Fail Problem

$\sum_{i=0}^{\infty} a^i = \frac{1}{1-a}$.

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

$\underbrace{1-p \quad 1-p \quad \_ \quad \_}_{i} \quad \_ \quad \_ \quad \_ \quad \cdots \quad \_$

X = # of steps. when system fails.

$P(X > i) = P(\text{No failure in the first } i \text{ step})$

$= (1-p)^i$

$E(X) = \sum_{i=0}^{\infty} P(X > i) = \sum_{i=0}^{\infty} (1-p)^i$

$= \frac{1}{1-(1-p)} = \frac{1}{p}$

Baby Problem

Some couple wants to get a baby girl. If they get a baby boy, they keep trying until they get a girl. What is the expected number of boys before they get a baby girl?

$P(\text{boy}) = P(\text{girl}) = \frac{1}{2}$.

A = the # of boys before a girl.

$$E(A) = \frac{1}{P} - 1$$

$$= 2 - 1$$

$$= 1$$

Baby Problem

What is the expected number of children they have they want each of a kind?

1st kid + subprob (previous prob)

$= 1 + (E(A) + 1)$

$= 1 + (1 + 1)$

$= 3$

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

the # of the failures before the first success.

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.

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,

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

Rolling two dice

$X =$ the # of 1st die $\quad E(X) = \frac{1}{6}(1+2+3+\cdots+6) = \frac{7}{2}$

$Y =$ the # of 2nd die $\quad E(Y) = \frac{7}{2}$.

Roll two six-sided fair and independent dice. What's the expected product of the dice?

$$E(XY) = E(X)E(Y) = \frac{7}{2} \cdot \frac{7}{2} = \frac{49}{4} = 12.25.$$

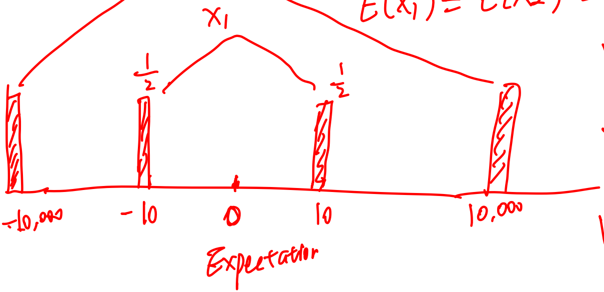
Variance

$$X_1 = \begin{cases} 10 \\ -10 \end{cases} \qquad P(X=10) = P(X=-10) = \frac{1}{2}$$

$$X_2 = \begin{cases} 10,000 \\ -10,000. \end{cases} \qquad P(X=10,000) = P(X=-10,000) = \frac{1}{2}$$

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

$$E(X_1) = E(X_2) = 0.$$



$$Var(X_1) = \frac{1}{2}\left[(10-0)^2 + (-10-0)^2\right]$$

$$Var(X_2) = \frac{1}{2}\left[(10,000-0)^2 + (-10,000-0)^2\right]$$

$$Var(X_1) < Var(X_2)$$

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?

$$\boxed{E(X) = \frac{7}{2} = 3.5}$$

$$\boxed{Var(X) = \frac{1}{6}}(1-3.5)^2 + \frac{1}{6}(2-3.5)^2 + \cdots + \frac{1}{6}(6-3.5)^2$$

$$= 2.917$$

X	$P(X)$	$E(X)$
1	1/6	3.5
2	1/6	
3	⋮	
4	⋮	
5		
6	1/6	

Example

What is the variance of the random variable $X((i,j)) = 2i$, where i is the number appearing on the first die and j is the number appearing on the second die, when two fair dice are rolled?

i	j	$X(i,j)$	X^2
1	1	2	2^2
2	2	4	4^2
3	3	6	6^2
		8	8^2
$\vdots$	$\vdots$	10	$\vdots$
6	6	12	12^2

$$Var(X) = E(X^2) - (E(X))^2$$

$$E(X^2) = \frac{1}{6}(2^2 + 4^2 + \cdots + 12^2) = \frac{182}{3}$$

$$E(X) = \frac{1}{6}(2 + 4 + 6 + \cdots + 12) = 7$$

$$Var(X) = \frac{182}{3} - (7)^2 = 35/3$$

Variance for the sum of random variables

If X and Y are independent variable,

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

In addition,

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

$$Var(Y + b) = Var(Y)$$

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

$\lambda = 1.5$

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

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

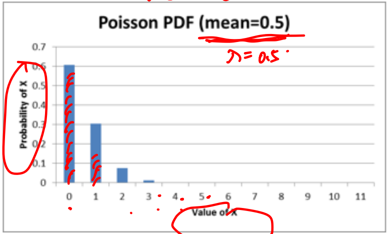
Source:

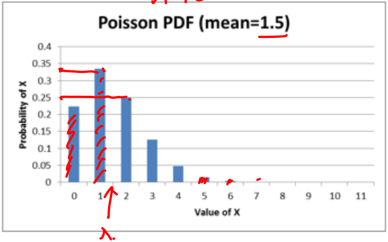
https://en.wikipedia.org/wiki/Poisson_distribution

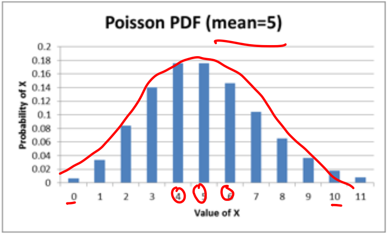
Poisson Distribution

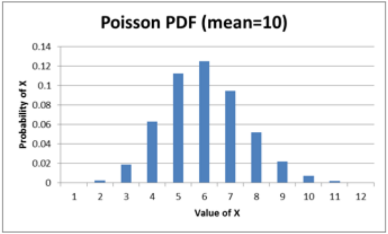
$\lambda = 0.5$

$\lambda = 1.5$









Example of Poisson Distribution

$N(\mu, \sigma^2)$

$P(\lambda)$

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? $P(X=0) = 22.3\%$

X = # of trains arrived in 10 mins.

$$P(X \geq 3) = \underline{P(X=3) + P(X=4) + P(X=5) + \cdots}$$

$$= 1 - P(X<3)$$

$$= 1 - (P(X=0) + P(X=1) + P(X=2))$$

$$= 19.1\%$$

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...

	Probability function	Expectation Mean.	Var.
Uniform	$f(k, a, b) = \frac{1}{b-a}$	$\frac{a+b}{2}$	$\frac{b^3-a^3}{3b-3a}$.
Bernouli	$f(k, p) = \begin{cases} p, & \text{if } k=1 \\ 1-p, & k=0 \end{cases}$	p	$p(1-p)$
Binomial	$f(k, n, p) = C(n, k)\, p^k (1-p)^{n-k}$.	np	$np(1-p)$
Geometric	$f(k, p) = (1-p)^{k-1} p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$.
Poisson.	$f(k, \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$, $k = 0, 1, 2, \ldots$	λ	λ

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