Lecture 8. Discrete Probability

Ref.: Rosen Section 6.1

Outline

- Discrete probability
- Probability of Combinations of events
- Conditional Probability
- Independence
- Binominal distribution
- Bayes' Theorem
- Random variables
- Expected Values and Variance

Definition: Finite Probability

If S is a **finite** sample space of equally likely outcomes and E is an event, that is a subset of S, then the probability of E is defined as $p(E) = \frac{|E|}{|S|}$

where $|E|$ and $|S|$ represent the number of events in E and S.

Example

What is the probability that when two dice are rolled, the sum of the numbers on the two dice is 7? (Each dice has 6 faces)

Solution: Each dice has 6 faces and there are total 36 equally like outcomes when two dice are rolled. There are only 6 successful outcomes, i.e., (1,6), (2,5), (3,4), (4,3), (5,2) and (6,1). So the probability is 6/36=1/6.

Theorems

A. Let E be an event in a sample space S. The probability of the event $\overline{E}$, the complimentary event of E is given by $p(\overline{E}) = 1 - p(E)$

B. Let $E1$ and $E2$ be events in a sample space S. Then

$$p(E1 \cup E2) = p(E1) + p(E2) - p(E1 \cap E2)$$

Example

A sequence of *10* bits is randomly generate, what is the probability that at least one of these bits is *0*?

Solution: Let E be the event that at least one of these bits is *0*. *Then the complementary event of E can be stated as the event that all the bits are 1s. Because the sample space is the set of all bit strings of length 10, so we have*

$$p(E) = 1 - p(\overline{E}) = 1 - \frac{1}{2^{10}} = 1 - \frac{1}{1024} = \frac{1023}{1024}$$

Example

What is the probability that a positive integer selected at random from the set of positive integers not exceeding 100 is divisible either by 2 or 5?

Solution: Let E1 be the event that the integer selected is divisible by 2 and E2 be the event that the integer selected is divisible by 5. Then $E1 \cap E2$ is the event that is divisible by both 2 and 5. That is to say $E1 \cap E2$ is the event divisible by 10. so $|E1 \cap E2| = 10$

Therefore $$p(E1 \cup E2) = p(E1) + p(E2) - p(E1 \cap E2)$$
$$= 50/100 + 20/100 - 10/100 = 3/5$$

Probability Theory

Let S be the sample space of an experiment with finite number of outcomes, we assign a probability p(si) for each outcome si, i=1,2,..n. We require the following two conditions are satisfied.

(1) $0 \leq p(s_i) \leq 1$

(2) $\sum_{i=1}^{n} p(s_i) = 1$

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

Example

What probability should we assign to the outcomes H (head) and T (tail) when a fair coin is flipped? What probabilities should be assigned to these outcomes when a coin is biased that head comes up twice as often as tails?

Solution: In fair flipping, p(H)=p(T)=1/2.

In second case, p(H)=2p(T),

Example

As $p(H)+p(T)=1$,

We can have $p(T)=1/3$, $p(H)=2/3$.

Uniform Distribution

Suppose S is a set with n elements. The uniform distribution assigns the probability 1/n to each element of S.

In other words:

The uniform distribution assigns the same probability to each element of sample space. The experiment of selecting an element from sample space with a uniform distribution is called selecting an element of S randomly.

Example

Suppose that a dice is biased so that 3 appears twice as often as each other numbers (1,2,4,5,6)and all other five numbers are equally appeared. What is the probability that an odd number appears when we roll this dice?

Solution: First we can derive

$$p(1)=p(2)=p(4)=p(5)=p(6)=1/7;\ p(3)=2/7$$

$$p(\text{odd number})=p(1)+p(3)+p(5)=4/7$$

Conditional Probability

Example: Suppose we flip a coin three times and all eight possibilities are equally likely. Moreover, suppose we know the even F that the first flip is T (tail). Given this information, what is the probability of the event E, that an odd number of tail appears? This is the conditional probability of E given F.

Conditional Probability

Solution: As the first flip is tail, there are only four possible outcomes: *TTT*,*TTH*,*THT*, *THH*. As there are only two cases with odd number of tails, *TTT* and *THH*, so the probability is 2/4=1/2.

Conditional Probability

Let E and F be two events with p(F)>0. The conditional probability of E given F is defined as

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

In other words:

We only count the outcomes after F happens, that the sample space utilized is shrunken.

Example

What is the conditional probability that a family with two children has two boys, given that they have at least one boy?

Let E be the event that a family with two children has two boys and let F be the event that a family has at least one boy. It follows that E={BB} and F={BB, BG, GB}. In this case,

$$E \cap F = \{BB\}$$

As four outcomes have same possibility, p(F)=3/4 and $p(E \cap F) = 1/4$

So p(E|F)=1/3

Independence

Definition: The events E and F are independent if and only if

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

In other words,

$$P(E|F) = p(E)$$

Example

Assume that each of the four ways a family can have two children is equally likely. Are the events E, that a family with two children has two boys, and F that a family with two children has at least one boy , independent?

Solution: As p(E)=1/4; p(F)=3/4.

$$p(E \cap F) = 1/4 \neq 3/16 = p(E)p(F)$$

So they are NOT independent.

Bernoulli Trials and the Binomial Distribution

Suppose that an experiment can only have two outcomes and in this case it is called a Bernoulli trial. In general, one possibility outcome with probability p and the other outcome will happen with probability q=1-p. If we have more experiments independently, we will have the following theorem.

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

$$C(n,k)\,p^k q^{n-k}$$

Random Variables

A random variable is a function from the sample space of an experiment to the set of real numbers. That is a random variable assigns a real number to each possible outcome.

Suppose that a coin is flipped three times. Let X(t) be the random variable that equals to the number of heads that appear when t is the outcome. Then X(t) is given as follows.

X(HHH)=3, X(HHT)=X(THH)=2, X(TTT)=0

Distribution of a RV.

The distribution of a RV X on a sample space is the set of pairs (r, p(X=r)) for all r in X(S). X(S) is all the possible values for the sample space S.

Example: Assume one coin is flipped three times and X(t) is defined as before. Then the distribution is as below.

P(X=3)=1/8;p(X=2)=3/8;p(X=1)=3/8;p(X=0)=1/8

Bayes' Theorem

In many cases, we need to estimate/assess the probability that a particular event happens when extra information is available.

Example: We have two boxes. The first contains two green balls and seven red balls; the second contains four green balls and three 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 selected a red ball, what is the probability that he selected a ball from the first box?

Example

Let E be the event that Bob has chosen a red ball; $\overline{E}$ is the event that has chosen a green ball. Let F be the event that Bob has chosen a ball from the first box; $\overline{F}$ is the event that Bob has chosen a ball from the second box. We want to compute p(F|E), the probability that the ball from the first box given that it is red.

As we know that $p(F \mid E) = p(E \cap F) / p(E)$

In this case, we need to know $p(E \cap F)$ and p(E).

Example

First we know that the first box contains 7 reds out of 9 balls, so p(E|F)=7/9, Similarly, $p(E \mid \overline{F}) = 3/7$. As we assume that Bob selects a box randomly, we have

$$p(F) = p(\overline{F}) = 1/2$$

As $p(E \mid F) = p(E \cap F) / p(F)$

We can compute $p(E \cap F) = p(E \mid F) p(F) = 7/9 * 1/2 = 7/18$

Next we need to compute p(E). As $E = (E \cap F) \cup (E \cap \overline{F})$

We need to compute $p(E \cap \overline{F}) = p(E \mid \overline{F}) p(\overline{F}) = 3/7 * 1/2 = 3/14$

Example

So we can compute $p(E) = p(E \cap F) + p(E \cap \overline{F}) = 7/18 + 3/14 = 38/63$

Finally we can compute

$$p(F \mid E) = p(E \cap F) / p(E) = 49/67$$

In this example, we used two important relations

A. $p(E \cap F) = p(F \cap E) = p(E \mid F) * p(F) = p(F \mid E) * p(E)$

B. $p(E) = p(E \cap F) + p(E \cap \overline{F}) = p(E \mid F)p(F) + p(E \mid \overline{F})p(\overline{F})$

Bayes' Theorem

Suppose that E and F are events from a sample space S such that $p(E) \neq 0 \text{ and } \mathrm{p(F)} \neq 0$

Then $$p(F \mid E) = \frac{p(E \mid F)\, p(F)}{p(E \mid F)\, p(F) + p(E \mid \overline{F})\, p(\overline{F})}$$

Suppose E is an event from a sample space S and that

$F_1, F_2, \ldots. F_n$ **are mutually exclusive events such that**

$\bigcup_{i=1}^{n} F_i = S$ **Assume** $p(E) \neq 0 \text{ and } \mathrm{p(F_i)} \neq 0$

Then $$p(F_j \mid E) = \frac{p(E \mid F_j)\, p(F_j)}{\sum_{i=1}^{n} p(E \mid F_i)\, p(F_i)}$$

Expected Value

The expected value (or expectation) of the RV X(s) on the sample space is defined as

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

When the sample space S has n elements $S = \{s_1, s_2, \ldots. s_n\}$

$$E(X) = \sum_{i=1}^{n} p(s_i) X(s_i)$$

The expected value of Bernoulli trials

The expected value of success when n independent Bernoulli trials are performed, where p is the success probability on each trial, is np. Let X be the RV equal to the number of success in n trials. As we know before, $p(X=k)=C(n,k)p^k q^{n-k}$

$$E(X)=\sum_{i=1}^{n} p(s_i)X(s_i)=\sum_{k=1}^{n} kp(X=k)$$

$$=\sum_{k=1}^{n} kC(n,k)p^k q^{n-k}=\sum_{k=1}^{n} nC(n-1,k-1)p^k q^{n-k}$$

$$=np$$

The expected value of a RV tells us its average value.

Many practical problems are based on the evaluation of this expectation as it has clear intuitive interpretations.

Average Case computational complexity.

Average Error

Linear Property of Expectations

If $X_i, i = 1,2,..n$ with n positive integers, are RVs on a sample space S and a and b are real numbers, then

$$E(X_1 + X_2 + ... + X_n) = E(X_1) + E(X_2) + ... + E(X_n)$$

$$E(aX + b) = aE(X) + b$$

Example

Let us roll two dices randomly, compute the expectation of the sum of the numbers appear on two dices.

Let X1 and X2 be two RVs with X1(i,j)=i and X2(i,j)=j, where X1 is the number on the first dice and X2 is the number on the second dice. So what we need to compute is E(X1+X2). In fact E(X1)=E(X2)=7/2 (do by yourself). So E(X1)+E(X2)=7

Independent RVs

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

$$p(X(s) = r \ and \ \mathrm{Y(s) = t) = p(X(s) = r)p(Y(s) = t)}$$

for ant real number r and t.

If X and Y are independent RVs on a space S, then

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

Variance

The expectation is only for the average value, it never tells us the distribution spread of the RV. The variance defined below is for this purpose.

Let X be a RV on a space S, the variance of X is defined as

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

The standard deviation of X, denoted as $\sigma(X)$, is defined to be $\sqrt{V(X)}$

In Fact

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

Example

What is the variance of the RV X with X(t)=1 if a Bernoulli trial is success and X(t)=0 if it is a failure, where p is the success probability?

Solution: As we need to compute $E(X^2)$, we need first to compute X^2 . From the definition of X, we can derive that

$$X^2 = X$$

Also we can know that E(X)=p. So V(X)=p(1-p)=pq.

One can see that it is an important step to compute X^2

Example

What is the variance of the RV X((i,j))=2i, where i is the number appearing on the first dice and j is the number appearing on the second dice, when two dice are rolled.

Solution: We compute E(X). As p(X=k)=1/6, for k=2,4,6,8,10,12.

Also we have $p(X^2 = k) = 1/6$ for k=4,16,36,64,100,144.

Therefor E(X)=7 and E(X^2)=182/3

We get V(X)=35/3.

Linearity of Variance

If X and Y are two independent RVs on a sample space S. Then

$$\mathbf{V(X+Y)=V(X)+V(Y)}$$

In general, the above is not true.

Chebyshev’s Inequality

The expectation value and variance are two important indexes for a RV. Expectation value represents its center in some sense and variance is how far away from the center.

We next will answer what is the chance for the random variable far away from its center. This is the very important Chebyshev’s inequality.

Chebyshev's Inequality

<u>Theorem</u> Let X be a RV on a sample space S with probability function p. If r is a positive real number, then

$$p(|X(s)-E(X)|\geq r)\leq V(X)/r^2$$

Proof

Let $A = \{s \in S \,||\, X(s) - E(X) \,|\geq r\}$

Now we decompose V(X) and then can finish the proof

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

$$= \sum_{s \in A} (X(s) - E(X))^2 p(s) + \sum_{s \notin A} (X(s) - E(X))^2 p(s)$$

$$\geq \sum_{s \in A} r^2 p(s) = r^2 p(A)$$

Summary

- Discrete Probability
- Conditional Probability
- Bayes' Theorem
- Random variables
- Expectation Values and Variance
- Chebyshev's Inequality