Probability

sourabhmadur

February 2019

1 Elements of Probability

1.1 Sample space and events

- **Sample space** Set of all outcomes of an experiment.
  - $S = \{b, g\}$ i.e boy or a girl
  - $S = \{1, 2, 3..6\}$ outcomes of roll of a die
  - $S = (-\infty, \infty)$ change in stock price tomorrow.
- **Event** Subset of sample space
  - $E = \{g\}$ i.e girl
  - $S = \{1\}$ outcomes of roll of a die is 1
- Define **Union of Events**, $E \cup F$ is the event which occurs if E or F occur.
  - $E = \{g\}, F = \{b\}, E \cup F$ is the event where a boy or a girl is born
- Define **Intersection of Events**, $E \cap F$ is the event which occurs only if both E and F occur.
  - $E = \{\text{odd rolls of dice}\}, F = \{\text{ prime number}\}, E \cap F$ is the event where the outcome is prime and odd.
- E and F are **Mutually Exclusive events**, if $E \cap F = \phi$
  - $E = \{boy\}, F = \{girl\}$, then,$E \cap F = \phi$
- **Complement** E^c is the set of all outcomes in sample space not in E
- **Algebra of Intersections and complements**follow the same algebra of addition and multilpication:(can be graphically seen by venn diagrams)
  - $a + b = b + a$: Commutative
  - $(a + b) + c = a + (b + c)$
  - $(a + b).c = a.c + b.c$

where $(+, .)$ are $(\cup, \cap)$ or vice versa

- **DeMorgans Laws**
  - $(E \cup F)^c = E^c F^c$
  - $(EF)^c = E^c \cup F^c$

1.2 Axioms Of Probability

- **Axioms** For any event, the probability of it can be viewed as the relative frequency of it occcuring when experiment performed large number of times. Say there is a number $P(E)$ for any event in sample space. Then, we have the following axioms
  1. $0 \leq P(E) \leq 1$
  2. $P(S) = 1$
  3. For mututally exclusive events $E_1, E_2, ..E_n$,
  $$P(\cup_{i=1}^{n} E_i) = \sum_{i=1}^{n} P(E_i)$$
- Corollaries of these axioms
  1. $P(E^c) = 1 - P(E)$
  2. $P(E \cup F) = P(E) + P(F) - P(EF)$
- **Odds** of event A is defined as $\frac{P(A)}{P(A_c)} = \frac{P(A)}{1-P(A)}$

1.3 Principle of counting

- Number of ways of permuting n distinct objects $= n(n-1)(n-2)..2.1 = n!$
- Selecting a group of k distinct items from a group of n is
  $$\frac{n(n-1)...(n-k+1)}{k!} = \binom{n}{k}$$
  . We divide by k! as the each selection of k distinct items ABC is counted k! times (3! times for permuting ABCgggff)

1.4 Conditional Probability

- $$P(E|F) = \frac{P(EF)}{P(F)} \tag{1}$$
  $$P(EF) = P(E|F)P(F) \tag{2}$$
- **Note**
  $$P(A^c|B) = 1 - P(A|B)$$

1.5 Baye's Formula

- **Law of Total Probability**

$$E = EF \cup EF^c$$
$$P(E) = P(EF) + P(EF_c)$$

As they are mutually exclusive(see from venn diagram). So,

$$P(E) = P(E|F)P(F) + P(E|F^c)P(F^c) \tag{3}$$

Therefore, when finding the probability of an event is hard to find directly, it can be found from the probabilities of that event conditioned on some other event.

For a general case,

$$E = \cup_{i=1}^{n} EF_i$$

where F_i are mutually exclusive events and $\cup_{i=1}^{n} F_i = S$. Then,

$$\begin{aligned} P(E) &= \sum_{i=1}^{n} P(EF_i) \\ &= \sum_{i=1}^{n} P(E|F_i)P(F_i) \end{aligned}$$

This is known as the law of total probability

- **Bayes Theorem** Suppose E has occured and we want to determine which of the $F_j's$ has occured(i.e we want to update the probabilities $P(F_i)'s$), then, from the above equation, we have,

$$P(F_j|E) = \frac{P(EF_j)}{P(E)} \tag{4}$$
$$= \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)} \tag{5}$$

This is the Bayes formula. It is used to update the probabilities of the hypothesis $(P(F_j)'s)$ give the new information that event E has occured.

1.6 Independent Events

- Two events are independent if

$$P(A|B) = P(A)$$

Or,

$$P(AB) = P(A)P(B) \tag{6}$$

- **Corollary** If E anf F are independent, so are E and F^c.
- For a general case, events $E_1, E_2, E_3, ..E_n$ are said to be independent if for every subset $E_{1'}, E_{2'}, E_{3'}, ..E_{r'}$, $r \leq n$

$$P(E_{1'}E_{2'}E_{3'}..E_{r'}) = P(E_{1'})P(E_{2'})P(E_{3'})..P(E_{r'})$$

2 Random Variables and expectation

2.1 Random Variables

Whenever an experiment is performed, every outcome in the sample space can be mapped to a real number. A random variable takes these different values.

- **Cumulative Distribution Function:** CDF/ distribution funciton of a random variable is defined for any real number x by

$$F(x) = P\{X \leq x\}$$

Notation:

$$X \sim F$$

Meaning, X has the distribution function F

2.2 Types of Random Variables

2.2.1 Discrete Random Variable:

Set of possible values is a sequence. $x_1, x_2, x_3...$ (ex: sum of dice)

- Its probability mass function for $p(a)$ is defined as :

$$p(a) = P\{X = a\}$$

- Also, we have

$$\sum_{i=1}^{\infty} p(x_i) = 1$$

- The CDF for a Discrete Random Variable can be defined as

$$F(a) = \sum_{all x \leq a} p(x)$$

2.2.2 Continuous Random Variable:

Set of possible values is all real numbers(ex: stock price next day). We say X is a continuous random variable if there exists a non-negative function $f(x)$, defined for all real x, having the property that for any set of real numbers

$$P\{X \in B\} = \int_B f(x)dx$$

$f(x)$ is called the probability density function.

- $$P\{a \leq X \leq b\} = \int_a^b f(x)dx$$
- Not that
$$\frac{dF(a)}{da} = f(a)$$

2.3 Jointly Distributed Random Variables

- We define joint cumulative distribution function F of X and Y as
$$F(x,y) = P\{X \leq x, Y \leq y\}$$
$$F_X(x) = P\{X \leq x, Y \leq \infty\}$$
$$F_Y(y) = P\{X \leq \infty, Y \leq y\}$$
- For discrete random variables X and Y, Joint probability mass function is defined as
$$p(x_i, y_j) = P\{X = x_i, Y = y_i\}$$
- We say that X and Y are jointly continuous if there exists a function $f(x,y)$ defined for all real x, y having the property that for every set C of pairs of real numbers,
$$P\{(X,Y) \in C\} = \iint_{(x,y)\in C} f(x,y)dxdy$$
$f(x,y)$ is the joint probability density function.
$$P\{X \in A, Y \in B\} = \int_B \int_A f(x,y)dxdy$$
Also,
$$f(a,b) = \frac{\partial^2}{\partial a \partial b} F(a,b)$$

- Individual probability density functions are

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

- **Independence of Random Variables**

  Random variables X and Y are said to be independent if for any two sets of real numbers A and B

$$P\{X \in A, Y \in B\} = P\{X \in A\}P\{Y \in B\}$$

  or,

$$P\{X \leq a.Y \leq b\} = P\{X \leq a\}P\{Y \leq b\}$$

  i.e

$$F(a,b) = F_X(a)F_Y(b)$$

  This condition is equivalent to mass and density functions being equal to:

  - **Discrete**

$$p(x,y) = p_X(x)p_Y(y)$$

  - **Continuous**

$$f(x,y) = f_X(x)f_Y(y)$$

  - Thm: If the joint probability $f(x,y)$ can be factored into $f(x,y) = k(x)l(y)$, then, X and Y are independent.

- **Conditional Dependence of discrete random variables** If X and Y are discrete random variables, conditional probability mass function of X given $Y = y$

$$p_{X|Y}(x|y) = P\{X = x|Y = y\}$$

  or

$$p_{X|Y}(x|y) = \frac{p(x,y)}{p_Y(y)}$$

- **Conditional Dependence of discrete random variables** If X and Y have a joint probability density function $f(x,y)$, then the conditional probability density function of X , given that $Y = y$, is defined for all values of y such that $f_Y(y) > 0$, by

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$$

2.4 Generalization

- CDF of set of n random variables is

$$F(a_1, a_2...a_n) = P\{X_1 \leq a_1, X_2 \leq a_2, .., X_n \leq a_n\}$$

- Probability mass function of n discrete random variables is

$$p(a_1, a_2...a_n) = P\{X_1 = a_1, X_2 = a_2, .., X_n = a_n\}$$

- We say that $X_1, X_2, ..X_n$ are jointly continuous if there exists a function $f(x_1, x_2, ...x_n)$ defined for all real $\vec{x}$ having the property that for every set C of n-space,

$$P\{\{X_1, X_2, ..X_n\} \in C\} = \iint_{(x_1, x_2..x_n)\in C} ... \int f(x_1, x_2..x_n)dx_1 dx_2...dx_n$$

- The n random variables $X_1, X_2, ..., X_n$ are said to be independent if, for all sets of real numbers $A_1, A_2, ..., A_n$,

$$P\{X_1 \in A, X_2 \in A_2, ..X_n \in A_n\} = \prod_{i=1}^{n} P\{X_i \in A_i\}$$

2.5 Expectation

- For discrete random variables, expectation is defined as

$$E[X] = \sum_i x_i P\{X = x_i\}$$

- Indicator random variable for an event A is defined as

$$I = \begin{cases} 1 \; if \;\; A \; occurs \\ 0 \; if \;\; A \; does \; not \; occur \end{cases}$$

$$E[I] = P(A)$$

- Expectation of a function of a random variable is

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$

- Expectation of sum of random variables is

$$E[X_1 + X_2 + X_3...X_n] = \sum_{i=1}^{n} E[X_i]$$

- **Variance of a random variable X** is defined as

$$Var(X) := E[(X-\mu)^2]$$

or

$$Var(X) = E[X^2] - (E[X])^2 \tag{7}$$

- For constants a and b

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

- **Standard deviation** of a random variable X is defined as

$$\sigma_X = \sqrt{Var(X)}$$

- **Covariance** of two random variables X and Y is defined as

$$\begin{aligned} Cov(X,Y) &= E[(X-\mu_x)(Y-\mu_y)] \\ &= E[XY] - E[X]E[Y] \end{aligned}$$

- Properties of covariance
  - $Cov(X,Y) = Cov(Y,X)$
  - $Cov(X,X) = Var(X)$
  - $Cov(aX,Y) = aCov(Y,X)$
  - $Cov(X+Z,Y) = Cov(Y,X) + Cov(Z,Y)$
  - $Cov(\sum_{i=1}^{n} X, \sum_{i=1}^{n} Y) = \sum_{i=1}^{n}\sum_{j=1}^{n} Cov(X_i, Y_j)$
  - Variance of sum of random variables is

$$Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i) + \sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} Cov(X_i, X_j)$$

  - If X and Y are independent random variables,

$$Cov(X,Y) = 0$$

  - For independent random variables $X_1, X_2, ...X_n$

$$Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i)$$

- **Correlation** between two random variables X and Y is defined as

$$Corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

  It has value between -1 and +1 and it indicates the strength of relationship between X and Y.

- **Moment generating function** $\phi(t)$ of the random variable X is defined as
$$\phi(t) = E[e^{tX}]$$
It has the property that the n^{th} derivative of ϕ is equal to the n^{th} moment of X i.e.
$$\phi^n(0) = E[X^n]$$

2.6 Chebyshev's Inequatily and weak law of large numbers

These inequalities help us put bounds on probabilities when the mean and the variance of the distribution are known and actual distribution is unknown.

- **Markov's Inequality**
$$P\{X \geq a\} \leq \frac{E[X]}{a}$$

- **Chebyshev's Inequality**
$$P\{|X - \mu| \geq k\} \leq \frac{\sigma^2}{k^2}$$

- **Weak Law of Large Numbers** The probability that the average of the first n terms in a sequence of independent and identically distributed random variables differs by its mean by more than ϵ, goes to 0 as n goes to infinity.
$$P\{|\frac{X_1 + X_2 + ..X_n}{n} - \mu| > \epsilon\} \rightarrow 0 \ as \ n \rightarrow \infty$$