W1211 Introduction to Statistics Lecture 6

Wei Wang

Sep 24th, 2012

The Law of Total Probability

- The Law of Total Probability states, Let $A_1, \ldots, A_k$ be mutually exclusive and *exhaustive* events. Then for any other event B.

$$P(B) = P(B|A_1)P(A_1) + \ldots + P(B|A_k)P(A_k)$$
$$= \sum P(B|A_i)P(A_i)$$

- $A_1, \ldots, A_k$ are *exhaustive*, if one A_i must occur, so that $A_1 \cup \ldots \cup A_k = S$.

- Proof: when k=2,

$$P(B) = P((B \cap A) \cup (B \cap A^c))$$
$$= P(B \cap A) + P(B \cap A^c)$$
$$= P(B|A)P(A) + P(B|A^c)P(A^c)$$

Bayes Theorem

- With the help of the Law of Total Probability, we can state the Bayes Rule, which says, let $A_1, \ldots, A_k$ be a collection of *k* mutually exclusive and exhaustive events with *prior* probabilities $P(A_i)$ ($i=1,\ldots,k$). Then for any other event B for which P(B) >0, the *posterior* probability of A_j given that B has occurred is,

$$\mathrm{P}(A_j|B) = \frac{\mathrm{P}(A_j \cap B)}{\mathrm{P}(B)} = \frac{\mathrm{P}(B|A_j)\mathrm{P}(A_j)}{\sum_{i=1}^{k} \mathrm{P}(B|A_i)\mathrm{P}(A_i)}$$

- When *k*=2, we have,

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(B|A)\mathrm{P}(A)}{\mathrm{P}(B|A)\mathrm{P}(A) + \mathrm{P}(B|A^c)\mathrm{P}(A^c)}$$

- Bayes Rule can be used to "*reverse*" the probability from the conditional probability that was originally given, or *to find the cause given the result*.

Bayes Theorem Example

- One percent of all individuals in a certain population are carriers of a particular disease. A diagnostic test for this disease has a 90% detection rate for carriers and a 5% detection rate for noncarriers. If a person is tested positive, what's the probability that this person is a carrier?

Bayes Theorem Example

- One percent of all individuals in a certain population are carriers of a particular disease. A diagnostic test for this disease has a 90% detection rate for carriers and a 5% detection rate for non-carriers. If a person is tested positive, what's the probability that this person is a carrier?

- $$\begin{aligned} & P(\text{is a carrier}|\text{tested positive}) \\ & = \frac{P(\text{carrier} \cap \text{tested positive})}{P(\text{tested positive})} \\ & = \frac{P(\text{positive}|\text{carrier})P(\text{carrier})}{P(\text{positive}|\text{carrier})P(\text{carrier}) + P(\text{positive}|\text{non-carrier})P(\text{non-carrier})} \end{aligned}$$

Independence

- Definition: Two events A and B are independent if $P(A|B) = P(A)$ (or alternatively $P(B|A) = P(B)$).
- A and B are independent if and only if

$$P(A \cap B) = P(A) \cdot P(B)$$

- Independent Events $\neq$ Disjoint Events.

When will we have independence

- Well, in the context of exam or homework problems, it is often given as the conditions.

When will we have independence

- Well, in the context of exam or homework problems, it is often given as the conditions.
- Finite Population v.s. Infinite Population

Multiple Events

- Events $A_1, \ldots, A_n$ are mutually independent if for every k ($k = 2, 3, \ldots, n$) and every subset of indices $i_1, i_2, \ldots, i_k$,

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = P(A_{i_1})\, P(A_{i_2}) \ldots P(A_{i_k}).$$

- Independence is very very important!

Example

Ex. You recently bought a new set of tires from a manufacturer who just announced a recall because 2% of that particular brand were defective. What is the probability that at least one of your tires is defective? You may assume that the tires are defective independently of one another.

P(at least one defective tire) = 1 – P(no defective tire)

Let A_i = tire *i* is not defective

$P(A_i) = 1\text{-}0.02 = 0.98$

$$\begin{aligned} P(\text{no defective tire}) &= P(A_1 \cap A_2 \cap A_3 \cap A_4) \\ &= P(A_1)\, P(A_2)\, P(A_3)\, P(A_4) = (0.98)^4 \end{aligned}$$

P(at least one defective tire) = $1\text{-}(0.98)^4 = 0.0776$

Random Variables

- A random variable is a variable whose value is a numerical outcome of a random phenomenon.
- For a given sample space S of some experiment, a random variable is any rule that associates a number with each outcome in S.
- To put it more mathematically, a random variable is a function whose domain is the sample space and whose range is the set of real numbers.

Random Variables v.s. Experiments

- An experiment is a physical setup in real world that provides us intuition about randomness.
- A random variable is a mathematical abstraction that describes randomness.
- When the outcome of the experiment can be seen as numerical, e.g., roll a die, we can effectively treat the experiment as a random variable.
- But for most RVs, especially continuous one, it is difficult to find some experiment that provides physical setup and intuition.

Discrete vs. Continuous

- X is a discrete random variable if its possible values either constitute a finite set or else can be listed in an infinite sequence in which there is a first element, a second element, and so on (“countably” infinite).

- X is a continuous random variable if it takes all possible values in an interval of numbers or all numbers in a disjoint union of such intervals. No possible value of the variable has positive probability, that is, $P(X=c) = 0$ for any possible value c.

- X can also be a random variable with a mixture distribution of both discrete and continuous components.

PMF

- The probability model for a discrete random variable X, lists its possible values and their probabilities.

Value of X	x_1	x_2	……..	x_k
Probability	p_1	p_2	……..	p_k

- Every probability, p_i, is a number between 0 and 1.
- $p_1+p_2+\ldots+p_k=1$

- The probability distribution or probability mass function (pmf) of a discrete rv is defined for every number x by $p(x) = P(X=x) = P(\text{all } s \in S\text{: } X(s)=x)$.

- How to check if some function $p(x)$ is a proper PMF?

Bernoulli RV

- The arguably simplest probability model is Bernoulli. Any random variable whose possible values are only 0 and 1 is called a Bernoulli random variable.
- Ex. Flip a coin. $S = \{H, T\}$. We can define a Bernoulli random variable, $X(H) = 1, X(T) = 0$. Then the distribution of X is
$$P(X = 1) = .5, P(X = 0) = .5$$
- Ex. Roll a die. $S = \{1, 2, 3, 4, 5, 6\}$. We can define a bernoulli random variable, $X(1) = X(2) = \mathbf{1}, X(3) = X(4) = X(5) = X(6) = \mathbf{0}$. Then the distribution is
$$P(X = \mathbf{1}) = 1/3, P(X = \mathbf{0}) = 2/3$$

Example

<u>Ex.</u> Flip three fair coins. (*Binomial*)

S = {HHH, HHT, HTH, HTT, THT, THH, TTH, TTT}. Let's define random variable X to be the number of heads in the experiment, i.e., X(HHH)=3, X(THT)=1, etc.

X	
0	TTT
1	TTH THT HTT
2	THH HTH HHT
3	HHH

Value of X	0	1	2	3
Probability	0.125	0.375	0.375	0.125

One can calculate the probability of an event by adding the probabilities p_i of the particular values of x_i that make up the event. For example, if we want to know the probability of getting less than 2 heads, we can use

$$P(X<2) = P(X=0) + P(X=1) = 0.125 + 0.375 = 0.5$$

$$\text{Note: } P(X\leq 2) = P(X=0) + P(X=1) + P(X=2) = 0.875$$

CDF

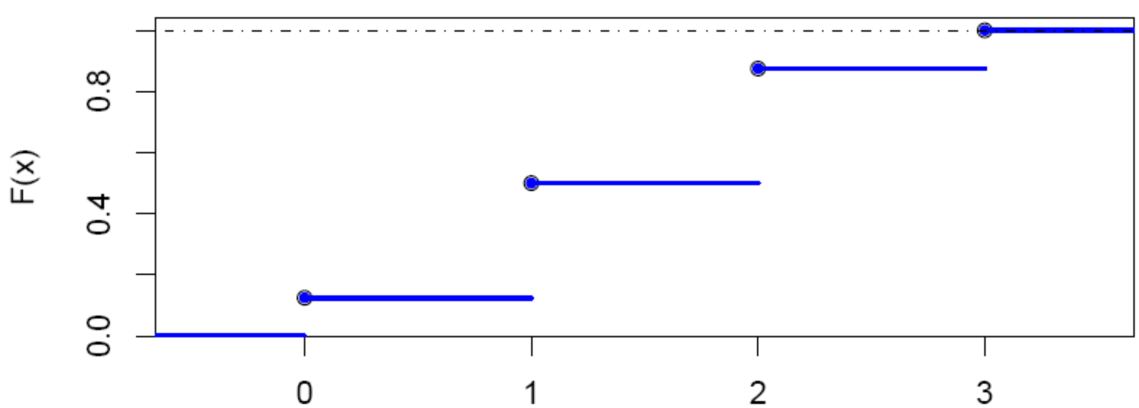
- The cumulative distribution function (cdf) F(*x*) of a discrete rv variable X with pmf *p*(*x*) is defined for every number *x* by

$$\mathrm{F}(x) = \mathrm{P}(\mathrm{X} \leq x) = \sum_{y:y\leq x} p(y).$$

  For any number *x*, F(*x*) is the probability that the observed value of X will be at most *x*.

- For X a discrete rv, the graph of F(*x*) will have a jump at every possible value of X and will be flat between possible values. Such a graph is called a step function.

The three coin flips example



Parameter and Family

- Suppose $p(x)$ depends on a quantity that can be assigned any one of a number of possible values, with each different value determining a different probability distribution. Such a quantity is called a parameter of the distribution. The collection of all probability distributions for different values of the parameter is called a family of probability distributions.

Ex. For Bernoulli rv's, the parameter is the probability of being 1 (or 0), that is,

$$p = P(X=1)$$

Expectation and Variance

- Random variables have distributions, so they have centers and spreads.

- The expected value (mean value or expectation) of a random variable describes its theoretical long-run average value.

- We typically use μ or $E(X)$ to denote the mean, $Var(X)$ to denote the variance and σ or $SD(X)$ to denote the standard deviation of a rv X.

Motivating examples

Ex. How many heads would you expect if you flipped a fair coin twice?

S = {HH, HT, TH, TT}.

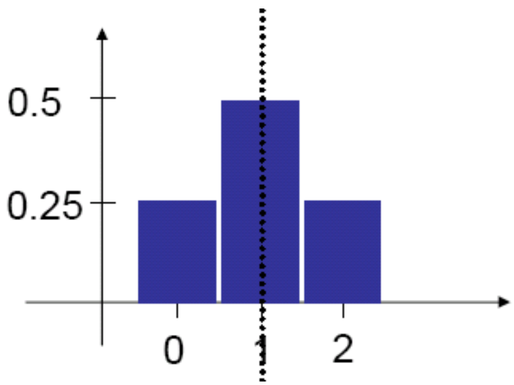
X = number of heads.

0 TT

1 HT TH

2 HH

$p(X=0) = 0.25$; $p(X=1) = 0.5$; $p(X=2) = 0.25$.



Each outcome is weighted by its probability.

$\mu = 0 \times 0.25 + 1 \times 0.5 + 2 \times 0.25 = 1$

Example

Ex. How many heads would you expect if you flipped a coin three times?

$$\mu = 0 \times 0.125 + 1 \times 0.375 + 2 \times 0.375 + 3 \times 0.125 = 1.5$$

This can never occur in a single trial of 3 flips. However, on average we would expect to get 1.5 heads if we repeated the experiment many times.

Definition

- Suppose X is a discrete random variable whose probability model is given by

Value of X	x_1	x_2	……..	x_k
Probability	p_1	p_2	……..	p_k

The expected value of X is given by

$$\mathrm{E}(\mathrm{X}) = \mu_X = \sum_{x \in D} x \cdot p(x) = x_1 p_1 + x_2 p_2 + \cdots x_k p_k$$

Example

Ex. Expectation of a Bernoulli rv.

$$p(x) = \begin{cases} 1-p & x=0 \\ p & x=1 \\ 0 & x \neq 0,1 \end{cases}$$

μ = 0 x (1-*p*) + 1 x *p* = *p*.