Discrete probability distributions - Class 4

January 16, 2014   Debdeep Pati

Random Variables

1. Events are not very convenient to use.

2. In a clinical trial of certain new treatment, we may be interested in the proportion of patients cured.

3. We are interested in some functions of the outcome of an experiment.

4. Random variables: A real valued function defined on sample space. A numeric function that assigns probabilities to different events in a sample space.

5. Example: New drugs have been introduced to bring hypertension under control. Suppose a physician agrees to use a new antihypertensive drug on a trial basis on the first 4 untreated hypertensives she encounters, before deciding whether to adopt the drug for routine use. Let X = the number of patients out of 4 who are brought under control. Then X is a random variable, which takes on the values 0, 1, 2, 3, 4. We can use expressions $X = 0, X = 3, X > 0, \ldots$ to represent events.

6. Example: Flip a coin three times and record the flips. Then S = HHH, HHT, HTH, THH, HTT, THT, TTH, TTT. Define a function X on S by X(s) = the # of heads in the three flips. So X(HHH) = 3, X(HTT) = 1, X(THT) = 1, and X(TTT) = 0. The range of the random variable X is $\{0, 1, 2, 3\}$. P(X=0) + P(X=1) + P(X=2) + P(X=3) = 1/8 + 3/8 + 3/8 + 1/8 = 1

7. Discrete random variables: A random variable for which there exists a discrete set of values with specified probabilities. A random variable is discrete if its range is finite or countably infinite. Countably infinite it is possible to make a list of the elements even though they are infinite. For instance the set of positive even numbers is countably infinite: 2, 4, 6, 8,. The set of positive numbers of no more than two decimal digits is countably infinite: $0.01, 0.02, \ldots, 0.99, 1.00, 1.01, \ldots$ An interval like $[0, 1]$ is uncountably infinite; it is impossible to make a list even an infinite list of the real numbers between zero and one.

Probability Mass function

1. Given a discrete random variable X, the associated probability mass function is defined as $P(X = r)$ for all values of r that have positive probability. It is also called the probability distribution.

2. Any value of the function is between 0 and 1 $0 < P(X = r) \leq 1$.

3. The sum of the probability of all possible values of X is 1. $\sum_{allr} P(X = r) = 1$.

4. Ex. (Hypertension) Suppose from previous experience with a drug, the drug company expects that for any clinical practice the probability that 0 patients out of 4 will be brought under control is .008, 1 patient out of 4 is .076, 2 patient out of 4 is .265, 3 patients out of 4 is .411, and all patients is .240. This probability mass function can be displayed in the table below.

Table 1: default

r	0	1	2	3	4
$P(X = r)$	0.008	0.076	0.265	0.411	0.240
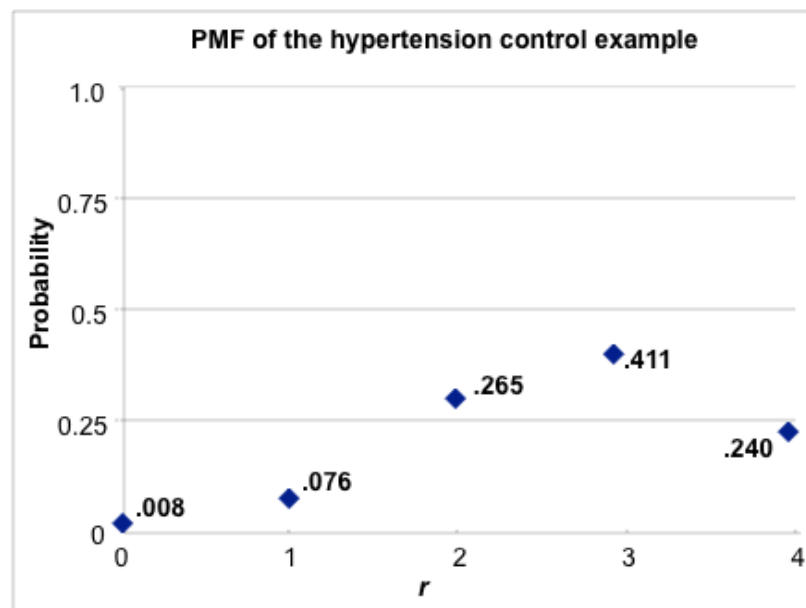
Expected Value

1. Expected value of X is a weighted average of the possible values that X can take on, each value being weighted by the probability that X assumes it. $E(X) = \mu =$

$\sum_{all\,x} xp(X = x)$ Expected value is what one should expect if the experiment is repeated many times. $E(f(X)) = \sum_{all\,x} f(x)P(X = x)$

2. Find the expected value for the random variable in hypertension example. $P(0) = .008$, $P(1) = .076$, $P(2) = .265$, $P(3) = .411$, $P(4) = .240$. $E(X) = 0(.008) + 1(.076) + 2(.265) + 3(.411) + 4(.240) = 2.8$

Variance

1. Sometimes, the expected value of a random variable X is not informative enough for decision making.

2. Definition: If X is random variable with mean μ, then the variance of X, denoted by $Var(X)$, is defined by $\text{Var(X)} = \sigma^2 = E[(X - \mu)^2] = \sum_{i=1}^{\infty}(x_i - \mu)^2 P(X = x_i)$

3. Relation: $Var(X) = E(X^2) - E(X)^2$

4. Example: Probability mass function for the number of episodes of otitis media in the first 2 years of life:

Table 2: default

r	0	1	2	3	4	5	6
$P(X = r)$	0.129	0.264	271	0.185	0.095	0.039	0.017

5. $E(X) = 2.038$, $V(X) = 1.967$.

Permutations

1. The number of permutations of n things taken k at a time is $P_k^n = n(n-1) \times \ldots \times (n-k+1) = n!/(n-k)!$ It represents the number of ways of selecting k items out of n, where the order of selection is important.

2. Ex. Suppose 3 diseased people and 6 eligible controls live in the same community. How many ways are there of selecting three controls? $P_3^6 = 6 \times 5 \times 4 = 120$

Combinations

1. The number of combinations of n things taken k at a time is $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

2. It represents the number of ways of selecting k items out of n, where the order of selection is not important.

Bernoulli Random variable

1. Outcomes of an experiment can be classified as either positive or negative, success or failure...

2. If we let X = 1 when the outcome is a positive and X = 0 when it is a negative, we can have the following pmf:

3. p(0) =1-p, p(1) = p, where p is the probability that the outcome is a positive.

4. If a random variable has the above pmf, it is a Bernoulli random variable.

5. Example: Ex. Infectious Disease One of the most common laboratory tests performed on any routine medical examination is a blood count. The two main aspects to a blood count are (1) counting the number of white blood cells (the white count) and (2) differentiating white blood cells that do exist into five categories namely, neutrophils, lymphocytes, monocytes, eosinophils, and basophils (the differential). Here we look at the number of neutrophils. What is the probability that any 2 cells out of 5 will be neutrophils, given the probability that any one cell is a neutrophil is .6? We can list all the possible outcomes with 2 cells being neutrophils, where x represents a neutrophil and o as others: xxooo, xoxoo, xooxo, xooox, oxxoo, oxoxo, oxoox, ooxxo, ooxox, oooxx There are totally 5 choose 2 ways of selecting 2 cells to be neutrophils out of 5 cells. The probability of each outcome is $(.6)^2(.4)^3 = .023$ The probability of seeing 2 neutrophils out of 5 is $10(.023) = 0.23$.

Binomial Random variable

1. If we perform n Bernoulli trials and let X represent the number of positives that occur in the n trials, then X is a binomial random variable with parameter (n, p).

2. The pmf of a binomial random variable with parameter (n, p) is
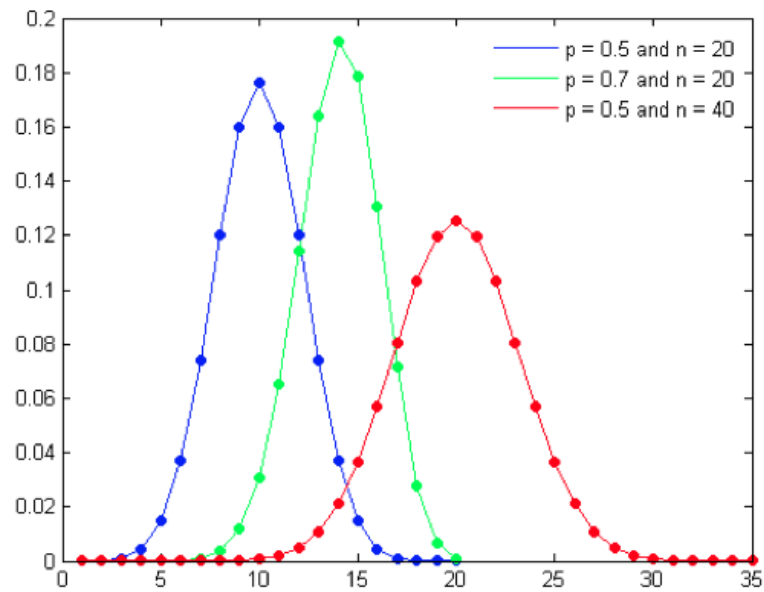
$$p(i) = \binom{n}{i} p^i (1-p)^{n-i}$$

3. Ex. What is the probability of obtaining 2 boys out of 5 children if the probability of a boy is .51 at each birth and the sexes of successive children are considered independent random variables?

$$P(X = 2) = \binom{5}{2}(0.51)^2(0.49)^3 = 0.306$$

4. $E(X) = np, Var(X) = np(1-p)$.

Binomial Distribution

Poisson distribution

1. Poisson distribution is widely used in statistics for modeling rare events.

2. Ex. Infectious Disease The number of deaths attributed to typhoid fever over a long period of time, for example, 1 year, follow a Poisson distribution if:

   (a) The probability of a new death from typhoid fever in any one day is very small.

   (b) The number of cases reported in any two distinct periods of time are independent random variables.

3. Ex. Rare events occurring on a surface area. The distribution of number of bacterial colonies growing on an agar plate. The number of bacterial colonies over the entire agar plate follow a Poisson distribution if we assume:

   (a) The probability of finding any bacterial colonies in a small area is very small.

   (b) The events of finding bacterial colonies at any two areas are independent.

4. The probability of k events in a time period t for a Poisson random variable with parameter μ is

$$P(X = k) = e^{-\mu}\frac{\mu^k}{k!},$$

   where $\mu = \lambda t$.

5. Parameter λ represents expected number of events per unit time.

6. Parameter μ represents expected number of events over time period t.

7. Difference between Binomial and Poisson distribution

   (a) There are a finite number of trials n in Binomial distribution

   (b) The number of events can be infinite for Poisson distribution

8. $E(X) = \mu, Var(X) = \mu.$

Poisson Distribution