Chapter 4 *Probability*

Discrete Distributions

Donald E. Mercante, PhD

Biostatistics

May 2010

Discrete Probability Distributions

Definition

Random Variable - variable whose outcome cannot be predicted with certainty in advance.

Two Types of Random Variables:
Discrete

- Binomial
- Poisson

Continuous

- Normal
- t
- F
- χ^2

Probability Mass Function

Probability Mass Function (pmf) for a discrete random variable

PMF - Rule that assigns probability, Pr(X=x), to each value of a discrete random variable

Discrete Probability Distribution

Example

Let X = Blood Type

A random sample from a certain population was used to construct the empirical distribution of blood types displayed in the following table.

X = Blood Type	O (x=1)	A (x=2)	B (x=3)	AB (x=4)
P(X=x)	0.40	0.35	0.15	0.10

Note: for any discrete distribution, $\sum_{x=1}^{N} \Pr(X = x) = 1$

Probability Mass Function

Two Properties of Discrete Distributions

1. $0 \leq \Pr(X = x) \leq 1$

2. $\sum_{all\ x} \Pr(X = x) = 1$

Expected Value (Mean) of a Discrete Random Variable

Let X be a discrete random variable. The Expected Value or Population Mean of X is,

$$E(X) = \mu = \sum_{i=1}^{N} x_i \Pr(X = x_i)$$

Example

Number of Episodes of Ottitis Media in infants first two years.

x_i	0	1	2	3	4	5	6
$\Pr(X = x_i)$	0.129	0.264	0.271	0.185	0.095	0.039	0.017

Expected Value (Mean) of a Discrete Random Variable

$$E(X) = \mu = \sum_{i=1}^{7} x_i \Pr(X = x_i) =$$

$$(0)(.129) + (1)(.264) + \cdots + (6)(.017) = 2.038$$

Thus, we would expect, on average, a child would have about 2 episodes of O.M. in the first 2 years of life.

Variance of a Discrete Random Variable

$$Var\left(X\right) = \sigma^2 = \sum_{i=1}^{7} x_i^2 \Pr\left(X = x_i\right) - \mu^2 =$$

$$\left(0^2\right)\left(.129\right) + \left(1^2\right)\left(.264\right) + \cdots + \left(6^2\right)\left(.017\right) - 2.038^2$$
$$= 6.12 - 2.038^2 = 1.967$$

$$\sigma = \text{Standard Deviation} = \sqrt{1.967} = 1.402$$

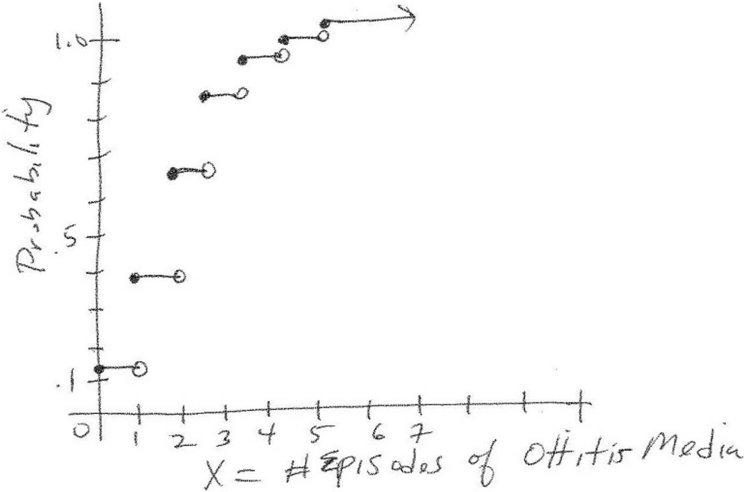
Cumulative Distribution Function of a Discrete Random Variable (CDF)

CDF: $F(x) = \Pr(X \leq x)$

x	$\Pr(X = x)$	$\Pr(X \leq x)$
0	0.129	0.129
1	0.264	0.393
2	0.271	0.664
3	0.185	0.849
4	0.095	0.944
5	0.039	0.983
6	0.017	1.000

Cumulative Distribution Function of a Discrete Random Variable (CDF)

CDF: $F(x) = \Pr(X \leq x)$ Plot of CDF is a Step Function



Binomial Distribution

Definition

Bernoulli Experiment: Outcome is either one of two mutually exclusive outcomes (usually referred to as Success or Failure)

Let $p = \Pr(Success)$; and $1 - p = \Pr(Failure)$, then for the i^{th} Bernoulli trial, define

$$X_i = \begin{Bmatrix} 0 & Failure \\ 1 & Success \end{Bmatrix}$$

- Let X= Number of successes in N-trials.
- $\Pr(S)$ remains constant from trial to trial.
- N-trials are independent

Binomial Distribution

Definition

Binomial Random Variable: Let $X = \sum_{i=1}^{N} X_i$, then X is a binomial random variable if

- X = number of successes in N-trials.
- $\Pr(S)$ remains constant from trial to trial.
- The N trials are independent

Binomial Distribution

Example

Observing n randomly chosen live births.

Let $X_1 = \{M\} \ \ or \ \ \{F\}$ for first observed birth.

Let $X_2 = \{M\} \ \ or \ \ \{F\}$ for second observed birth.

$\vdots$

Let $X_n = \{M\} \ \ or \ \ \{F\}$ for nth observed birth.

Also, assume that $\Pr\{M\} = \Pr\{F\} = \frac{1}{2}$

Binomial Distribution

Let X = number of males born in our sample of n live births.

Then, X is a Binomial distributed random variable, and

$X \sim B(n, p)$, where $p = \Pr(M)$

Since the trials are independent, the probability of joint events are simply the product of the individual probabilities (as shown in the following example).

Example

Since $\Pr(M) = \Pr(F) = \frac{1}{2}$,
$\Pr(M_1 \cap M_2 \cap M_3) = \Pr(M_1)\Pr(M_2)\Pr(M_3) = \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \left(\frac{1}{8}\right)$

Binomial Distribution

From the previous slide, the probability that the first 3 babies observed are males

i.e., male on first, second, and third observed births $=$ $\Pr\left(M_1 \cap M_2 \cap M_3\right)$,

is simply the product of the individual probabilites, namely $\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right) = \left(\frac{1}{8}\right)$.

Distribution of X = # male live births in a random sample of n=3 births.

X	**Select 3 live births at random**			
X=0	F	F	F	$(1-p)\cdot(1-p)\cdot(1-p)$
	M	F	F	$p\cdot(1-p)\cdot(1-p)$
X=1	F	M	F	$(1-p)\cdot p\cdot(1-p)$
	F	F	M	$(1-p)\cdot(1-p)\,p$
	M	M	F	$p\cdot p\cdot(1-p)$
X=2	M	F	M	$p\cdot(1-p)\cdot p$
	F	M	M	$(1-p)\cdot p\cdot p$
X=3	M	M	M	$p\cdot p\cdot p$

Distribution of X = # male live births in a random sample of n=3 births.

x	$\Pr(X = x)$
0	$(1-p)^3 = \frac{1}{8}$
1	$\mathbf{3}(p)(1-p)^2 = 3\left(\frac{1}{8}\right) = \frac{3}{8}$
2	$\mathbf{3}(p^2)(1-p) = 3\left(\frac{1}{8}\right) = \frac{3}{8}$
3	$p^3 = \frac{1}{8}$

$\Pr(X = x)$ = (# ways event can occur)*(Probability of Event)

Combinations

The multiplier on the probability of occurrence of an event can be determined by the combinations formula:

$\binom{n}{x} = \frac{n!}{x!(n-x)!}$, where $n! = n \cdot (n-1) \cdot (n-2) \cdots (2)(1)$ is called "n factorial".

For example, the multiplier on the probability of obtaining exactly 2 males in 3 live births is $\binom{3}{2} = \frac{3!}{2!1!} = 3$

Binomial Probability Mass Function (pmf)

Fact

Probabilities of events for binomially distributed random variables are easily found using the pmf:

$$\Pr\left(X = x\right) = \binom{n}{x} p^x \left(1 - p\right)^{n-x}$$

Example

To find the probability that exactly 2 males (successes) are observed in 3 live births:

$$\Pr\left(X = 2\right) = \binom{3}{2} 0.5^2 \left(1 - 0.5\right)^{3-2} = \left(3\right) 0.5^3 = \frac{3}{8}.$$

Binomial Probability Mass Function (pmf)

Example

Let probability that a randomly chosen cell in a field of cells is a neutrophil be $p = 0.6$.

Define X = # of neutrophils in a random sample of 4 cells. Then $X \sim B\left(n = 4, p = 0.6\right)$, where $p = \Pr\left(Neutrophil\right)$.

We may ask: What is the probability that exactly 2 of the 4 cells will be neutrophils?

$$\Pr\left(X = 2\right) = \binom{4}{2} 0.6^2 \left(1 - 0.6\right)^{4-2} = \left(\frac{4!}{2!2!}\right) 0.6^2 0.4^2 =$$

$$\left(6\right)\left(0.36\right)\left(0.16\right) = 0.3456$$

Binomial Probability Mass Function (pmf)

Example

Or, we may ask: What is the probability that no more than 1 of the 4 cells will be neutrophils?

$$\Pr(X \leq 1) = \Pr(X = 0) + \Pr(X = 1) =$$

$$\binom{4}{0} 0.6^0 (0.4)^4 + \binom{4}{1} 0.6^1 (0.4)^3 = \left(\frac{4!}{0!4!}\right) 0.6^0 0.4^4 + \left(\frac{4!}{1!3!}\right) 0.6^1 0.4^3 =$$
$$(1)(1)(0.0256) + (4)(0.6)(0.064) = 0.0256 + 0.1536 = 0.1792$$

Binomial Tables

Fact

Probabilities of events for binomially distributed random variables are even more easily found using the Binomial Tables (Table 1):

The tables allow us to compute probabilities for binomial variables without

using the formula when the sample size (n) is 20 or less.

Example

Let X=# of male births out of a random sample of 3 live births, assuming the probability that a live birth will be male = 0.5. What is $\Pr(X = 2)$?

$n = 3$, $p = 0.5$, From Table 1 on page 817, with k= 2. The answer is 0.375 or $3/8$.

Binomial Tables

Example

Let X = # of petri dishes where bacterial colonies appear two days after innoculation. Assume the probability that a colony will form on a petri dish within two days is 0.7. Now suppose in a random sample of 10 innoculated petri dishes we ask: What is the probability that exactly 8 petri dishes will have colonies forming within two days after innoculation? i.e., $\Pr(X = 8)$?

Our parameters are $n = 10$, $p = 0.7$, $k = 8$. We see from Table 1 starting on page 817, that n=10 is in the Table, but there is no $p = 0.7$.

Binomial Tables

- We can still, however, use the tables when p>0.5 by re-writing the problem in terms of the number of failures, instead of the number of successes.
- We would then use 1-p rather than p.
- For this example, If there are 8 successes, then there are 10-8 = 2 failures.
- If the success rate is 0.7, then the failure rate is 1 - 0.7 = 0.3.
- Thus, if $X \sim B\left(n = 10, p = 0.7\right)$ where X = #successes,

then Y = n - X, would represent the number of failures, and $Y \sim B\left(n = 10, p = 0.3\right)$.

- The original problem $\Pr(X = 8)$ would be rewritten in terms of failure as $\Pr(Y = n - x) = \Pr(Y = 10 - 8) = \Pr(Y = 2)$.
- We go to the tables on page 818 and for n=10, p=0.3, and k=2, the probability is **0.2335**.

Binomial Mean and Variance

Theorem

Let $X \sim B(n, p,)$. The mean of X is given as $\mu = np$; and the variance of X is given as $\sigma^2 = np(1-p) = npq$.

Example

$X \sim B(n = 10, p = 0.7)$. The mean of X is $\mu = np = (10)(0.7) = 7$; and the variance of X is $\sigma^2 = npq = (10)(0.7)(0.3) = 2.1$.

Poisson Distribution

Definition

Poisson Random Variable: Let $X = \#$ events in some interval of time.

- X = number of events in some interval of time or space.
- Usually associated with rare events, such as incidence of a rare disease.

Poisson Distribution

Assumptions:

- The probability that an event occurs is directly proportional to the length of the time interval.
- The probability of observing more than one event in a small interval of time (or area) is essentially zero.
- If an event occurs in some interval of time (or area), it does not affect the probability of an event occurring in the next interval (independence).

Poisson Distribution

Example

Infectious Diseases Consider the distribution of number of deaths attributed to typhoid fever over a long period of time, for example 1 year. Assuming the probability of a new death by typhoid fever in any one day is very small and the number of cases reported in any two distinct periods of time are independent random variables, then the number of deaths over a 1-year period will follow a Poisson distribution.

Poisson Distribution

Example

Bacteriology Consider a 100 cm^2 agar plate. the probability of finding any bacterial colonies at any one point in time is very small, and the events of finding bacterial colonies at any two points in time are independent. Then, the number of bacterial colonies over the entire agar plate will follow a Poisson distribution.

Poisson Distribution

Definition

Poisson Distribution

The probability of k events occurring in a time period t for a Poisson random variable with parameter λ is

$$\Pr(X = k) = \frac{e^{-\mu}\mu^k}{k!}, \quad k = 0, 1, 2, \ldots$$

where $\mu = \lambda t$ and e is aprroximately 2.71828.

The Poisson distribution has one parameter, $\mu = \lambda t$.
Note: λ represents expected number of events per unit time, while μ represents expected number of events over time period t.

Poisson Distribution

Example

Infectious Diseases Consider the typhoid fever example. Suppose the number of deaths from typhoid fever over a one year period is Poisson distributed with parameter $\mu = 4.6$. What is the probability distribution of the number of deaths over a 6-month period?

$X = \#$ deaths from typhoid fever in 1-year, then $\mu = \lambda t = 4.6$ deaths per *year*, implies $\lambda = 4.6$ deaths per year. Then, for a 6-month period, $t = 0.5$, and $\mu = \lambda(0.5) = 2.3$ deaths per 0.5 year (6-months).

Thus, $X \sim P(\mu = 2.3)$, and $\Pr(X = 0) = \frac{e^{-2.3}(2.3)^0}{0!} = e^{-2.3} = 0.1003$
$\Pr(X = 1) = \frac{e^{-2.3}(2.3)^1}{1!} = e^{-2.3}(2.3) = 0.2306$, etc.

Poisson Distribution

Example

Bacteriology If $\mathbf{A} = 100$ cm^2 and $\lambda = 0.02$ colonies per cm^2, calculate the probability distribution of the number of bacterial colonies.

We have $\mu = \lambda A = 100\,(0.02) = 2$. Let X = the number of colonies, then $X \sim P\,(\mu = 2)$.

$\Pr(X = 0) = \frac{e^{-2}(2)^0}{0!} = e^{-2} = 0.1353.$

$\Pr(X = 2) = \frac{e^{-2}(2)^2}{2!} = \frac{4e^{-2}}{2} = 0.2707.$

Poisson Tables, expected Value, and Variance

Fact

Probabilities of events for Poisson distributed random variables are even more easily found using the Poisson Tables (Table 2).

Fact

Expected value, μ, *and the Variance,* σ^2, *are both equal to* $\mu = \lambda t$.