Chapter 3. Discrete Random Variables and Probability Distributions

Peijia Zheng (郑培嘉)

School of Data & Computer Science

Sun Yat-Sen University

Email：zhpj@mail.sysu.edu.cn

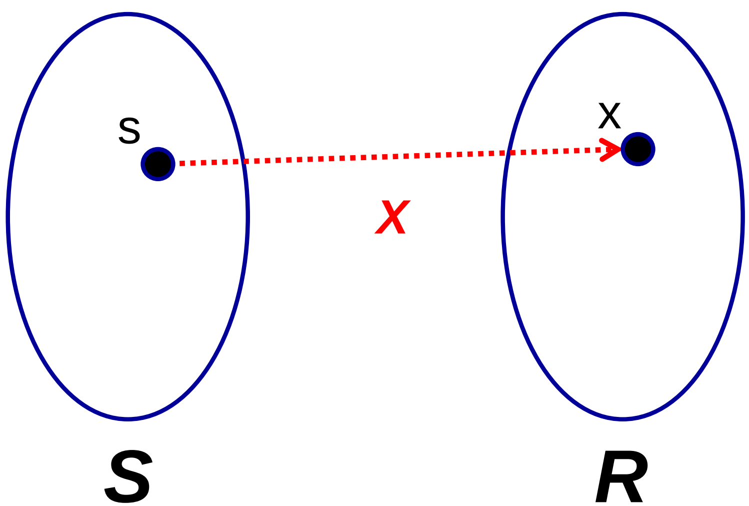
Chapter three:
Discrete Random Variables and Probability Distributions

- 3.1 Random Variables
- 3.2 Probability Distributions for Discrete Random Variables
- 3. 3 Expected Values
- 3.4 The Binomial Probability Distribution
- 3.5 Hypergeometric and Negative Binomial Distributions
- 3.6 The Poisson Probability Distribution

3.1 Random Variables

- Random Variable (rv)

  For a given sample space S of some experiment, a random variable is any rule that associates a number with each outcome in S. In mathematical language, a random variable is a function whose domain is the sample space and whose range is the set of real number.



3.1 Random Variables

- Two Types of Random Variables

➢ Discrete Random Variable (Chap. 3)

A discrete random variable is an rv whose 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.

➢ Continuous Random Variable (Chap. 4)

A random variable is continuous if its set of possible values consists of an entire interval on the number line.

Note: there is no way to create an infinite listing them! (why?)

3.1 Random Variables

- Example 3.3 (Ex. 2.3 Cont’)

  X = the total number of pumps in use at the two stations

  Y = the difference between the number of pumps in use at station 1 and the number in use at station 2

  U = the maximum of the numbers of pumps in use at the two stations.

  Note: X, Y, and U have finite values.

- Example 3.4 (Ex. 2.4 Cont’)

  X = the number of batteries examined before the experiment terminates

  Note: X can be listed in an infinite sequence

3.1 Random Variables

- Example 3.5

  suppose that in some random fashion, a location (latitude and longitude) in the continental United States is selected. Define an rv Y by

  Y= the height above sea level at the selected location

Note: Y contains all values in the range [A, B]

A: the smallest possible value

B: the largest possible value

3.1 Random Variables

- Bernoulli Random Variable

Any random variable whose only possible are 0 and 1 is called Bernoulli random variable.

- Example 3.1

When a student calls a university help desk for technical support, he/she will either immediately be able to speak to someone (S, for success) or will be placed on hold (F, for failure). With ***S***={S,F}, define an rv X by

$$X(S) = 1, X(F) = 0$$

The rv X indicates whether (1) or not (0) the student can immediately speak to someone.

3.1 Random Variables

- Example 3.2

  Consider the experiment in which a telephone number in a certain area code is dialed using a random number dialer, and define an rv Y by

  Y(n)=1 if the selected number n is unlisted

  Y(n)=0 otherwise

e.g.

if 5282966 appears in the telephone directory, then

Y(5282966) = 0, whereas Y(7727350)=1 tells us that the number 7727350 is unlisted.

3.1 Random Variables

- Homework

  Ex. 5, Ex. 6, Ex 8, Ex. 10

- Probability Distribution

The probability distribution or probability mass function (pmf) of a discrete rv is defined for every number x by

$$p(x) = P(X=x) = P(\text{all } s \text{ in } S: X(s)=x)$$

Note:

$$p_i = p(x_i) \geq 0, \sum_{i=1}^{n} p_i = 1$$

x	x_1	x_2	…	X_n
$p(x)$	p_1	p_2	…	p_n

- Example 3.7

The Cal Poly Department of Statistics has a lab with six computers reserved for statistics majors. Let X denote the number of these computers that are in use at a particular time of day. Suppose that the probability distribution of X is as given in the following table; the first row of the table lists the possible X values and the second row gives the probability of each such value.

x	0	1	2	3	4	5	6
$p(x)$	.05	.10	.15	.25	.20	.15	.10

$$P(X \le 2) = P(X = 0 \text{ or } 1 \text{ or } 2) = p(0) + p(1) + p(2) = .05 + .10 + .15 = .30$$

$$P(X \ge 3) = 1 - P(X \le 2) = 1 - .30 = .70$$

$$P(2 < X < 5) = P(X = 3 \text{ or } 4) = .25 + .20 = .45$$

- Example 3.8

Six lots of components are ready to be shipped by a certain supplier. The number of defective components in each lot is as follows:

Lot	1	2	3	4	5	6
Number of defectives	0	2	0	1	2	0

One of these lots is to be randomly selected for shipment to a particular customer. Let X=the number of defectives in the selected lot.

X	0 (lot 1,3 or 6)	1 (lot 4)	2 (lot 2 or 5)
Probability	0.5	0.167	0.333

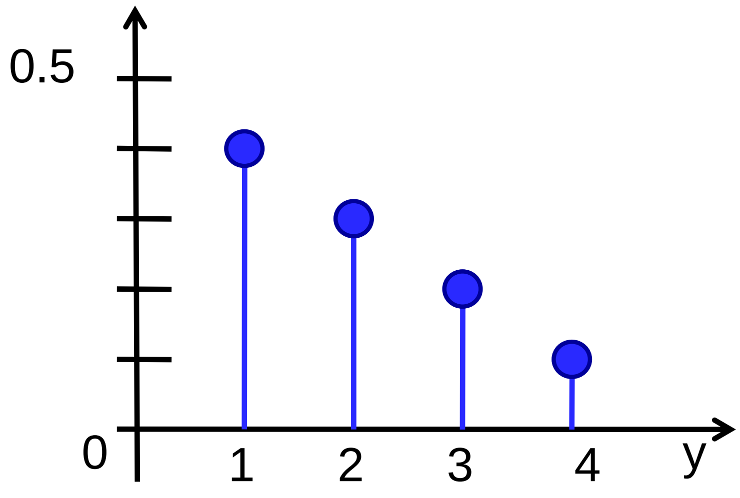
- Example 3.10

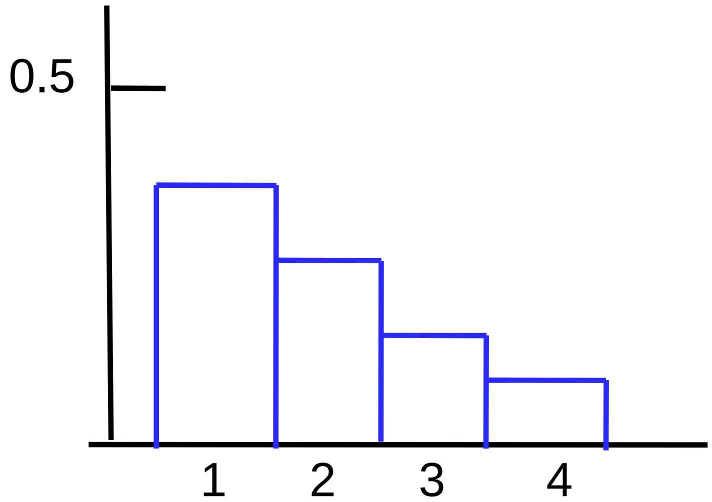
Consider a group of five potential blood donors—A,B,C,D, and E—of whom only A and B have type O+ blood. Five blood samples, one from each individual, will be typed in random order until an O+ individual is identified. Let the rv Y= the number of typings necessary to identify an O+ individual. Then the pmf of Y is

y	1	2	3	4
$p(y)$	0.4	0.3	0.2	0.1

- Line Graph and Probability Histogram

y	1	2	3	4
$p(y)$	0.4	0.3	0.2	0.1





- Example 3.9

Consider whether the next person buying a computer at a certain electronics store buys a laptop or a desktop model. Let

$$X = \begin{cases} 1 & \text{If the customer purchase a desktop computer} \\ 0 & \text{If the customer purchase a laptop computer} \end{cases}$$

If 20% of all purchasers during that week select a desktop, the pmf for X is

$$p(x) = \begin{cases} 0.8, & if \quad x = 0 \\ 0.2, & if \quad x = 1 \\ 0, & if \quad x \neq 0 \; or \; 1 \end{cases}$$

x	0	1
$p(x)$	0.8	0.2

A Parameter of a Probability Distribution

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.**

e.g. **Ex. 3.9**

x	0	1
$p(x)$	0.8	0.2

x	0	1
$p(x)$	$1-\alpha$	$\boldsymbol{\alpha}$

- Example 3.12

Starting at a fixed time, we observe the gender of each newborn child at a certain hospital until a boy (B) is born. Let p=P(B), assume that successive births are independent, and define the rv X by X=number of births observed. Then

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

$$p(2) = P(X=2) = P(GB) = P(G)\,P(B) = (1-p)p$$

$$\ldots$$

$$p(k) = P(X=k) = P(G\ldots GB) = (1-p)^{k-1}p$$

- Cumulative Distribution Function

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

$$F(x) = P(X \le x) = \sum_{y: y \le x} p(y)$$

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

- Example 3.13

A store carries flash drives with either 1 GB, 2 GB, 4 GB, 8 GB, or 16 GB of memory. The accompanying table gives the distribution of Y= the amount of memory in a purchased drive:

y	1	2	4	8	16
$p(y)$	.05	.10	.35	.40	.10

$$F(1) = P(Y \leq 1) = P(Y = 1) = p(1) = .05$$
$$F(2) = P(Y \leq 2) = P(Y = 1 \text{ or } 2) = p(1) + p(2) = .15$$
$$F(4) = P(Y \leq 4) = P(Y = 1 \text{ or } 2 \text{ or } 4) = p(1) + p(2) + p(4) = .50$$
$$F(8) = P(Y \leq 8) = p(1) + p(2) + p(4) + p(8) = .90$$
$$F(16) = P(Y \leq 16) = 1$$

- Example 3.14 (Ex. 3.12 cont')

$$p(x) = \begin{cases} (1-p)^{x-1}p, & x = 1,2,3,\ldots \\ 0 & otherwise \end{cases}$$

For a positive integer x,

$$F(x) = \sum_{y \le x} p(y) = \sum_{y=1}^{x} (1-p)^{y-1} p = p \sum_{y=1}^{x} (1-p)^{y-1}$$

$$F(x) = p \cdot \frac{1-(1-p)^x}{1-(1-p)} = 1-(1-p)^x$$

For any real value x,

$$F(x) = 1-(1-p)^{\lfloor x \rfloor}$$

$\lfloor x \rfloor$ is the largest integer ≤ x

- Example 3.14 (cont')

$$F(x) = \begin{cases} 0 & x < 1 \\ 1 - (1-p)^{[x]} & x \geq 1 \end{cases}$$

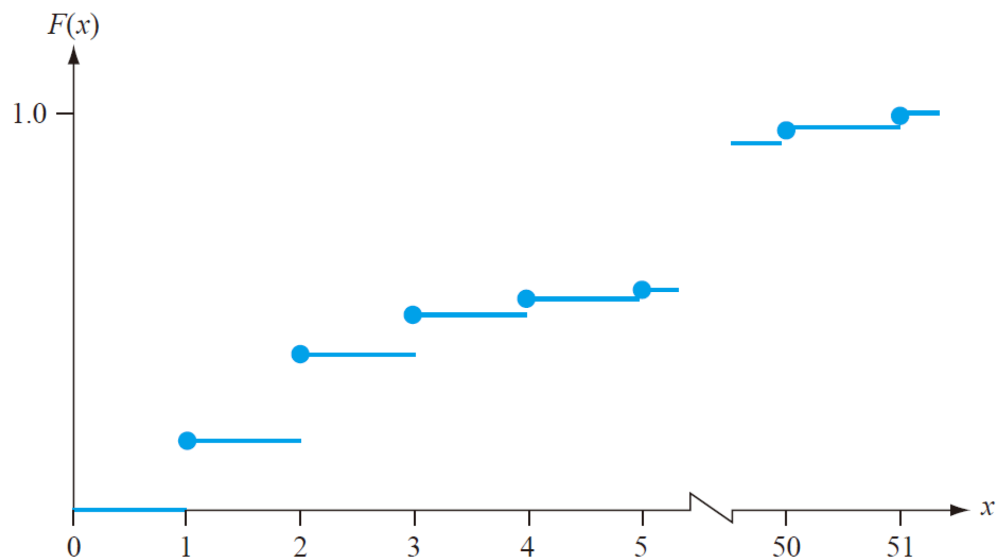


Figure 3.6 A graph of $F(x)$ for Example 3.14

- Proposition

For any two numbers a and b with $a \le b$,

$$P(a \le X \le b)=F(b)-F(a-)$$

where "a-" represents the largest possible X value that is strictly less than a.

In particular, if the only possible values are integers and if a and b are integers, then

$$P(a \le X \le b)= P(X=a \text{ or } a+1 \text{ or } \dots \text{ or } b)$$
$$= F(b) - F(a-1)$$

Taking a=b yields $P(X = a) = F(a)-F(a-1)$ in this case.

- Example 3.15

  Let X= the number of days of sick leave taken by a randomly selected employee of a large company during a particular year. If the maximum number of allowable sick days per year is 14, possible values of X are 0, 1, …, 14. With $F(0)=0.58$, $F(1)=0.72$, $F(2)=0.76$, $F(3)=0.81$, $F(4)=0.88$, and $F(5)=0.94$

$$P(2 \le X \le 5) = P(X=2,3,4 \text{ or } 5) = F(5) - F(1) = 0.22$$

and $P(X=3) = F(3) - F(2) = 0.05$

- Three Properties of cdf (discrete/continuous cases)

1. Non-decreasing, ***i.e.*** if $x_1<x_2$ then $F(x_1) \leq F(x_2)$

2. $$F(-\infty)=\lim_{x\to-\infty}F(x)=0$$

$$F(+\infty)=\lim_{x\to+\infty}F(x)=1$$

3. $F(x+0)=F(x)$

Note: Any function that satisfies the above properties would be a cdf.

- Homework

  Ex. 13, Ex. 23, Ex. 24, Ex. 27

3.3 Expected Values of Discrete Random Variables

- The Expected Value of X

Let X be a discrete rv with set of possible values *D* and pmf p(x). The expected value or mean value of X, denoted by E(X) or μ_X (or μ for short), is

$$E(X) = \mu_X = \sum_{x \in D} x \cdot p(x)$$

Note: When the sum does not exist, we say the expectation of *X* does not exist. (finite or infinite case?)

- Example 3.16

Consider selecting at random a student who is among the 15,000 registered for the current term at Mega University. Let X= the number of course for which the selected student is registered, and suppose that X has the pmf as following table

X	1	2	3	4	5	6	7
$P(x)$	0.01	0.03	0.13	0.25	0.39	0.17	0.02
Number registered	150	450	1950	3750	5850	2550	300

$$\mu_X = 1 \cdot p(1) + 2 \cdot p(2) + ... + 7 \cdot p(7)$$
$$= (1)(.01) + 2(.03) + ... + (7)(.02)$$
$$= .01 + .06 + ... + .14 = 4.57$$

- Example 3.17

Just after birth, each newborn child is rated on a scale called the Apgar scale. The possible ratings are 0, 1, . . . , 10, with the child's rating determined by color, muscle tone, respiratory effort, heartbeat, and reflex irritability

x	0	1	2	3	4	5	6	7	8	9	10
$p(x)$	.002	.001	.002	.005	.02	.04	.18	.37	.25	.12	.01

Then the mean value of X is

$$\begin{aligned} E(X) = \mu &= 0(.002) + 1(.001) + 2(.002) \\ &\quad + \cdots + 8(.25) + 9(.12) + 10(.01) \\ &= 7.15 \end{aligned}$$

- Example 3.18

  Let X=1 if a randomly selected vehicle passes an emissions test and otherwise. Then X is a Bernoulli rv with pmf

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

then E(x) = 0 p(0) + 1 p(1) = p(1) =p.

- Example 3.19

The general form for the pmf of X=number of children born up to and including the first boy is

$$p(x)=\begin{cases} p(1\text{-}p)^{x\text{-}1} & x=1,2,3,\ldots \\ 0 & \text{otherwise} \end{cases}$$

$$E(x)=\sum_{D} x\cdot p(x)=\sum_{x=1}^{\infty} xp(1-p)^{x-1}=p\sum_{x=1}^{\infty}\left[-\frac{d}{dp}(1-p)^{x}\right]$$

$$=-p\frac{d}{dp}\sum_{x=1}^{\infty}(1-p)^{x}=-p\frac{d\left[\dfrac{1-p}{1-(1-p)}\right]}{dp}=-p\left[\frac{1-p}{p}\right]'=\frac{1}{p}$$

- Example 3.20

Let X, the number of interviewers a student has prior to getting a job, have pmf

$$p(x) = \begin{cases} k/x^2 & x = 1,2,3,\ldots \\ 0 & \text{otherwise} \end{cases}$$

Where k is chosen so that $\sum_{X=1}^{\infty}(k/x^2) = 1$. (In a mathematics course on infinite series, it is shown that $\sum_{x=1}^{\infty}(1/x^2) < \infty$, which implies that such a k exists, but its exact value need not concern us).The expected value of X is

$$\mu = E(X) = \sum_{X=1}^{\infty} x \cdot \frac{k}{x^2} = k\sum_{x=1}^{\infty}\frac{1}{x}$$

Harmonic Series!

- Example 3.21

Suppose a bookstore purchases ten copies of a book at \$ 6.00 each, to sell at \$12.00 with the understanding that at the end of a 3-month period any unsold copy can be redeemed for \$2.00. If X=the number of copies sold, then

$$\text{Net revenue}=h(X)=12X+2(10-X)-60=10X-40.$$

Here, we are interested in the expected value of the net revenue (h(X)) rather than X itself.

- The Expected Value of a function

Let X be a discrete rv with set of possible values D and pmf p(x). Then the expected values or mean value of any function $h(X)$, denoted by $E[h(X)]$ or $\mu_{h(X)}$, is computed by

$$E[h(X)] = \sum_{x \in D} h(x) \cdot p(x)$$

- Example 3.22

The cost of a certain vehicle diagnostic test depends on the number of cylinders *X* in the vehicle's engine. Suppose the cost function is given by h(x) = $20+3X+0.5X^2$. Since X is a random variable, so is Y = h(X). The pmf of X and derived pmf of Y are as follows:

x	4	6	8
$p(x)$	.5	.3	.2

$\Rightarrow$

y	40	56	76
$p(y)$	.5	.3	.2

$$
\begin{aligned}
E(Y) = E[h(X)] &= \sum_{D^*} y \cdot p(y) \\
&= (40)(.5) + (56)(.3) + (76)(.2) \\
&= h(4) \cdot (.5) + h(6) \cdot (.3) + h(8) \cdot (.2) \\
&= \sum_{D} h(x) \cdot p(x)
\end{aligned}
$$

- Example 3.23

A computer store has purchased three computers of a certain type at \$500 apiece. It will sell them for \$1000 apiece. The manufacturer has agree to repurchase any computers still unsold after specified period at \$200 apiece. Let X denote the number of computers sold, and suppose that p(0)=0.1, p(1)=0.2, p(2)=0.3 and p(3)=0.4. With h(x) denoting the profit associated with selling X units, the given information implies that h(X) =revenue- cost =1000X+200(3-X)-1500 =800X-900. The expected profit is then

$$E(h(X)) = h(0)p(0)+h(1)p(1)+h(2)p(2)+h(3)p(3)$$
$$= (-900)(0.1)+(-100)(0.2)+(700)(0.3)+(1500)(0.4)$$
$$= 700$$

- Rule of Expected Value

$$E(aX+b) = a\,E(X) + b$$

Proof:

$$\begin{aligned} E(aX+b) &= \sum_D (ax+b)\,p(x) \\ &= a\sum_D xp(x) + b\sum_D p(x) \\ &= aE(x) + b \end{aligned}$$

1. For any constant a, E(aX)=aE(X) (b=0)
2. For any constant b, E(X+b)=E(X)+b (a=1)

- The Variance of X

Let X have pmf p(x) and the expected value μ. Then the variance of X, denoted by V(X) or δ^2_x , or just δ^2, is

$$V(X) = \sum_{x \in D} (x-\mu)^2 \cdot p(x) = E[(X-\mu)^2]$$

The standard deviation (SD) of X is

$$\sigma_X = \sqrt{\sigma_X^2}$$

- Example 3.24

x	1	2	3	4	5	6
$p(x)$	.30	.25	.15	.05	.10	.15

The expected value of X is easily seen to be $\mu = 2.85$. The variance of X is then

$$V(X) = \sigma^2 = \sum_{x=1}^{6} (x - 2.85)^2 \cdot p(x)$$

$$= (1 - 2.85)^2(.30) + (2 - 2.85)^2(.25) + \cdots + (6 - 2.85)^2(.15) = 3.2275$$

The standard deviation of X is $\sigma = \sqrt{3.2275} = 1.800$.

- A short formula for δ^2

$$V(X)=\sigma^2=[\sum_D x^2 p(x)]-\mu^2=E(X^2)-[E(X)]^2$$

Proof:

$$V(X)=\sum_{x\in D}(x-\mu)^2\cdot p(x)=E[(X-\mu)^2]$$

$$=\sum_{x\in D}x^2\cdot p(x)-2\mu\sum_D xp(x)+\mu^2\sum_D p(x)$$

$$=E(X^2)-2\mu\mu+\mu^2=E(X^2)-\mu^2$$

$$=E(X^2)-[E(X)]^2$$

- Example 3.25 (Ex. 3.24 cont')

$$E(X^2) = \sum_{x=1}^{6} x^2 \cdot p(x) = (1^2)(.30) + (2^2)(.25) + \cdots + (6^2)(.15) = 11.35$$

Thus $\sigma^2 = 11.35 - (2.85)^2 = 3.2275$ as obtained previously from the definition. ■

- Rules of Variance

$$V(aX+b)=\sigma_{aX+b}^2=a^2\sigma_X^2 \quad \& \quad \sigma_{aX+b}=|a|\sigma_X$$

1. $\sigma_{aX}^2=a^2\sigma_X^2, \sigma_{aX}=|a|\sigma_X$
2. $\sigma_{X+b}^2=\sigma_X^2$

- Example 3.26

In the computer sales problem of Example 3.23, $E(X)=2$ and

$$E(X^2)=(0)^2(0.1)+(1)^2(0.2)+(2)^2(0.3)+(3)^2(0.4)=5$$

so $V(X)=5-(2)^2=1$. The profit function $h(X)=800X-900$ then has variance $(800)^2V(X)=(640{,}000)(1)=640{,}000$ and standard deviation 800.

3.3 Expected Values of Discrete Random Variables

- Homework

  Ex. 30, Ex. 37, Ex. 41, Ex. 45

3.4 The Binomial Probability Distribution

- The requirements for a **binomial experiment**

1. The experiment consists of a sequence of n smaller experiments called trials, where n is fixed in advance of the experiment.
2. Each trail can result in one of the same two possible outcomes (dichotomous trials), which we denote by success (S) or failure (F).
3. The trails are independent, so that the outcome on any particular trail does not influence the outcome on any other trail.
4. The probability of success is constant from trail to trail; we denote this probability by p.

- Example 3.27

  The same coin is tossed successively and independently n times. We arbitrarily use S to denote the outcome H(heads) and F to denote the outcome T(tails). Then this experiment satisfies Condition 1-4. Tossing a thumbtack n times, with S=point up and F=point down, also results in a binomial experiment.

3.4 The Binomial Probability Distribution

- Example 3.28

The color of pea seeds is determined by a single genetic locus. If the two alleles at this locus are AA or Aa (the genotype), then the pea will be yellow (the phenotype), and if the allele is aa, the pea will be green. Suppose we pair off 20 Aa seeds and cross the two seeds in each of the ten pairs to obtain ten new genotypes. Call each new genotype a success S if it is aa and a Failure otherwise. Then with this identification of S and F, the experiment is binomial with n=10 and p=P(aa genotype). If each member of the pair is equally likely to contribute a or A, then p=P(a)P(a)=(1/2)(1/2)=1/4

Many experiments involve a sequence of independent trials for which there are more than two possible outcomes on any one trial.

AA Aa aA aa

F S

- Example 3.29

Suppose a certain city has 50 licensed restaurants, of which 15 currently have at least one serious health code violation and the other 35 have no serious violations. There are five inspectors, each of whom will inspect one restaurant during the coming week. The name of each restaurant is written on a different slip of paper, and after the slips are thoroughly mixed, each inspector in turn draws one of the slips **without replacement**. Label the ith trail as success if the ith restaurant selected (i=1,…5) has no serious violations. Then

P(S on first trail) = 35/50 = 0.7 &

P(S on second trial) = P(SS) + P(FS)

= P(second S | first S) P(first S) + P(second S| first F) P(first F)

=(34/49)(35/50) +(35/49)(15/50)= (35/50)(34/49+15/49) = 0.7

Similarly, P(S on ith trail) = 0.7 for i=3,4,5.

- Example 3.28 (Cont')

P(S on fifth trail | SSSS)

= (35-4) / (50-4) = 31/46

$\neq$

P(S on fifth trail | FFFF)

= 35 / (50-4) = 35/46

Thus the experiment is not binomial because the trials are not independent. In general, if sampling is **without replacement**, the experiment will not yield independent trials.

- Example 3.30

Suppose a certain state has 500,000 licensed drivers, of whom 400,000 are insured. A sample of 10 drivers is chosen without replacement. The ith trial is labeled S if the ith driver chosen is insured.

Although this situation would seem identical to that of Example 3.28, the important difference is that the size of the population being sampled is very large relative to the sample size. In this case

$P(S \text{ on } 2 \mid S \text{ on } 1) = 3999{,}999/4999{,}999 = 0.8$ &

$P(S \text{ on } 10 \mid S \text{ on first } 9) = 399{,}991/499{,}991 = 0.799996 \approx 0.8$

These calculations suggest that although the trials are not exactly independent, the conditional probabilities differ so slightly from one another that for practical purposes the trials can be regarded as independent with constant P(S)=0.8. Thus, to a very good approximation, the experiment is binomial with n =10 and p=0.8.

3.4 The Binomial Probability Distribution

- Rule

Consider sampling without replacement from a dichotomous population of size N. If the sample size (number of trials) n is at most 5% of the population size, the experiment can be analyzed as though it were exactly a binomial experiment.

In Ex. 3.30, the sample size n is 10, and the population size N is 500,000, $10/500000<0.05$.

However, in Ex. 3.29, the sample size n =5, and the population size N is 50, $5/50 > 0.05$.

3.4 The Binomial Probability Distribution

- Binomial random variable

Given a binomial experiment consisting of n trails, the binomial random variable X associated with this experiment is defined as

X = the number of S's among the n trials

Suppose, for instance, that n=3. Then there are eight possible outcomes for the experiment:

SSS SSF SFS SFF FSS FSF FFS FFF

X(SSS) = 3, X(SSF) =2, … X(FFF)=0

3.4 The Binomial Probability Distribution

- X~ Bin(n,p)

Possible values for *X* in an *n*-trial experiment are *x* = 0,1,2,…,*n*. we will often write X~Bin(n,p) to indicate that X is a binomial rv based on n trials with success probability *p*.

Because the pmf of a binomial rv depends on the two parameters *n* and p, we denote the pmf by b(x;n,p)

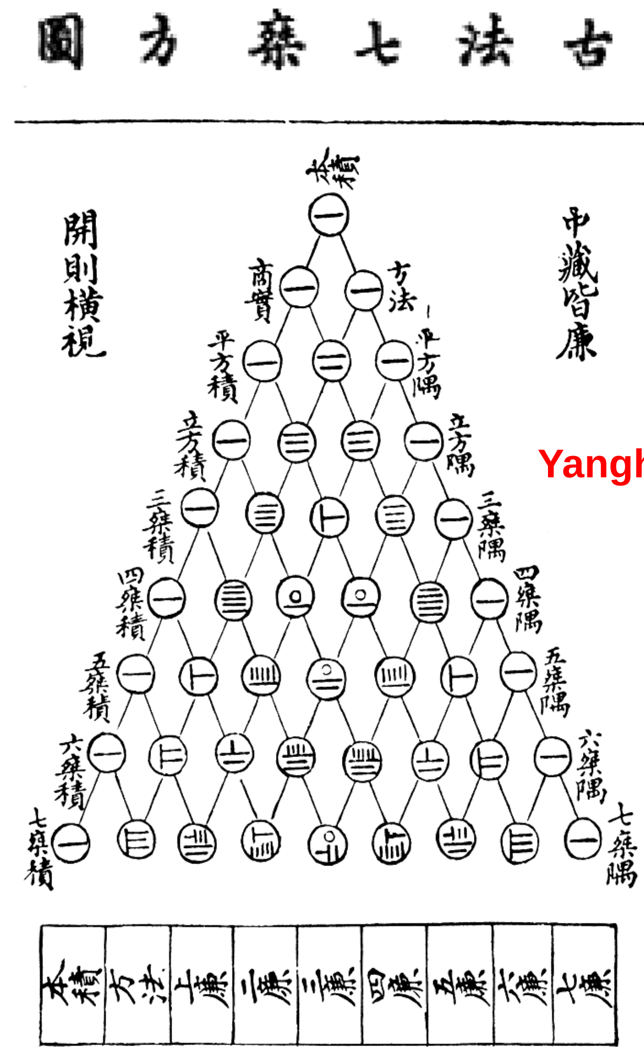
$$b(x;n,p)=\begin{cases}\binom{n}{x}p^x(1-p)^{n-x}, & x=0,1,2,\ldots,n\\ 0, & otherwise\end{cases}$$

3.4 The Binomial Probability Distribution

$$(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x}$$

$$\sum_{x=0}^{n} b(x;n,p) = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x}$$

$$= [p + (1-p)]^n = 1$$



Yanghui_triangle

3.4 The Binomial Probability Distribution

- The outcomes and probabilities for a binomial experiment with 3 trails

Outcomes	x	Probability	Outcomes	X	Probability
SSS	3	p^3	FSS	2	$p^2(1-p)$
SSF	2	$p^2(1-p)$	FSF	1	$p(1-p)^2$
SFS	2	$p^2(1-p)$	FFS	1	$p(1-p)^2$
SFF	1	$p(1-p)^2$	FFF	0	$(1-p)^3$

$$b(2;3,p) = P(SSF) + P(SFS) + P(FSS)$$

$$= \binom{3}{2} p^2 (1-p)^{3-2} = 3p^2(1-p),$$

3.4 The Binomial Probability Distribution

- Example 3.31

Each of six randomly selected cola drinkers is given a glass containing cola S and one containing cola F. The glasses are identical in appearance except for a code on the bottom to identify the cola. Suppose there is actually no tendency among cola drinkers to prefer one cola to the other. Then p=P(a selected individual prefers S) =0.5, so with X=the number among the six who prefer S, X~Bin(6,0.5).

$$P(X=3)=b(3;6,0.5)=\binom{6}{3}(0.5)^3(0.5)^3=20(0.5)^6=0.313$$

$$P(3\le X)=\sum_{x=3}^{6}b(x;6,0.5)=\sum_{x=3}^{6}\binom{6}{x}(0.5)^x(0.5)^{6-x}=0.656$$

3.4 The Binomial Probability Distribution

- Notation

  For X~Bin(n,p), the cdf will be denoted by

$$P(X \le x) = B(x; n, p) = \sum_{y=0}^{x} b(y; n, p), x = 0, 1, \ldots, n$$

- Binomial Table

  Refer to Appendix Table A.1

3.4 The Binomial Probability Distribution

- B(n,p) with n=5, p=0.1,0.3,0.5,0.7 and 0.9

	p					
x		0.1	0.3	0.5	0.7	0.9
	0	0.590	0.168	0.031	0.002	0.000
	1	0.919	0.528	0.188	0.031	0.000
	2	0.991	0.837	0.500	0.163	0.009
	3	1.000	0.969	0.812	0.472	0.081
	4	1.000	0.998	0.969	0.832	0.410
	5	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**

B(3; 5, 0.5)

B(2; 5, 0.7)

- Example 3.32

Suppose that 20% of all copies of a particular textbook fail a certain binding strength test. Let X denote the number among 15 randomly selected copies that fail the test. Then X has a binomial distribution with n=15 and p=0.2.

1. The probability that at most 8 fail the test is

$$P(X \le 8) = \sum_{y=0}^{8} b(y;15,0.2) = B(8;15,0.2) = 0.999$$

2. The probability that exactly 8 fail is

$$P(X = 8) = P(X \le 8) - P(X \le 7) = B(8;15,0.2) - B(7;15,0.2) = 0.999 - 0.996 = 0.003$$

3. The probability that at least 8 fail is

$$P(X \ge 8) = 1 - P(X \le 7) = 1 - B(7;15,0.2) = 1 - 0.996 = 0.004$$

4. The probability that between 4 and 7

$$P(4 \le X \ge 7) = P(X \le 7) - P(X \le 3) = B(7;15,0.2) - B(3;15,0.2) = 0.996 - 0.648 = 0.348$$

- Example 3.33

An electronics manufacturer claims that at most 10% of its power supply units need service during the warranty period. To investigate this claim, technicians at a testing laboratory purchase 20 units and subject each one to **accelerated testing** to simulate use during the warranty period. Let p denote the probability that a power supply unit needs repair during the period. The laboratory technicians must decide whether the data resulting from the experiment supports the claim that $p \le 0.1$. Let X denote the number among the 20 sampled that need repair, so $X \sim \text{Bin}(20, p)$. Consider the decision rule

Reject the claim that $p \le 0.1$ in favor of the conclusion that $p > 0.1$ if $x \ge 5$

and consider the claim plausible if $x \le 4$

3.4 The Binomial Probability Distribution

- Example 3.33 (Cont')

The probability that the claim is rejected when p=0.10 (an incorrect conclusion) is

$$P(X \geq 5 \; when \; p = 0.1) = 1 - B(4; 20, 0.1) = 1 - 0.957 = 0.043$$

The probability that the claim is not rejected when p=0.20 (a different type of incorrect conclusion) is

$$P(X \leq 4 \; when \; p = 0.2) = B(4; 20, 0.2) = 0.63$$

The first probability is rather small, but the second is intolerably large. When p=0.20, so that the manufacturer has grossly understate the percentage of units that need service, and the stated decision rule is used, 63% of all samples will result in the manufacturer's claim being judges plausible!

3.4 The Binomial Probability Distribution

- Proposition

  If X~Bin(n,p), then E(X)=np, V(X)=np(1-p)=npq, and

$$\delta_X = \sqrt{npq}$$

where q=1-p

- Example 3.34

  If 75% of all purchases at a certain store are made with a credit card and X is the number among the randomly selected purchases made with a credit card, then $X \sim Bin(10, 0.75)$. Thus $E(X)=np=(10)(0.75)=7.5$, $V(X)=npq=10(0.75)(0.25)=1.875$.

  If we perform a large number of independent binomial experiments, each with n=10 trails and p=0.75, then the average number of S's per experiment will be close to 7.5.

3.4 The Binomial Probability Distribution

- Homework

  Ex. 48, Ex. 50, Ex.52, Ex. 60, Ex. 64

- The assumptions leading to the hypergeometric distribution are as follows:

1. The population or set to be sampled consists of N individuals, objects, or elements (a finite population).
2. Each individual can be characterized as a success (S) or a failure (F), and there are M successes in the population.
3. A sample of n individuals is selected without replacement in such a way that each subset of size n is equally likely to be chosen.

Consider X = the number of S's in the sample,

the probability distribution of X depends on the parameters n, M and N, $P(X=x) = h(x; n,M,N)$

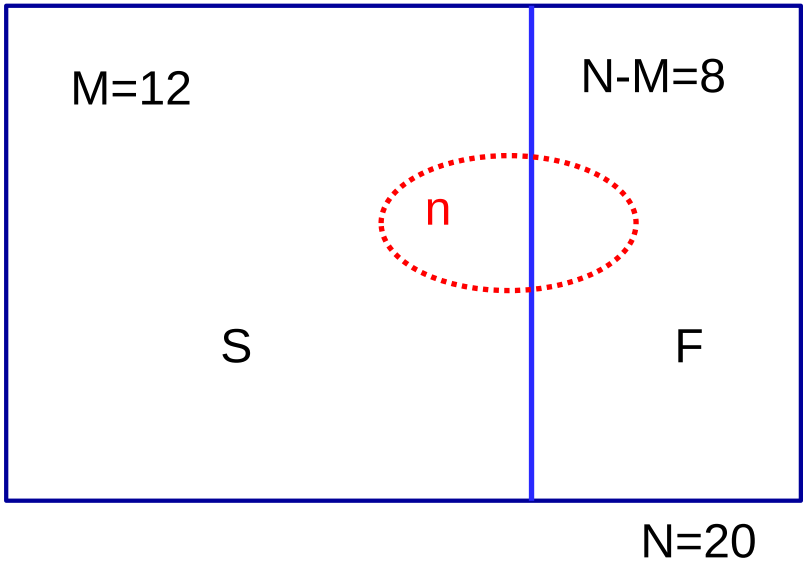
- Example 3.34

A office received 20 service orders for problems with printers, of which 8 were laser printers and 12 were inkjet models. A sample of 5 of there service orders is to be selected for inclusion in a customer satisfaction survey. Suppose that the 5 are selected randomly, what is the probability that exactly x of the selected service orders were for inkjet printers?

In this example, N = 20, M = 12, n = 5

$$P(X = x) = h(x; 5, 12, 20) = \frac{\#\,of\,outcomes\,X = x}{\#\,of\,possible\,outcomes}$$

3.5 Hypergeometric and Negative Binomial Distributions



# of Possible outcomes: $\binom{N}{n} = \binom{20}{5}$

$$P(X = x) = \frac{\binom{12}{x}\binom{8}{5-x}}{\binom{20}{5}}$$

# of outcomes having X=x

Step 1: Choosing x elements from subset S $\binom{M}{x} = \binom{12}{x}$

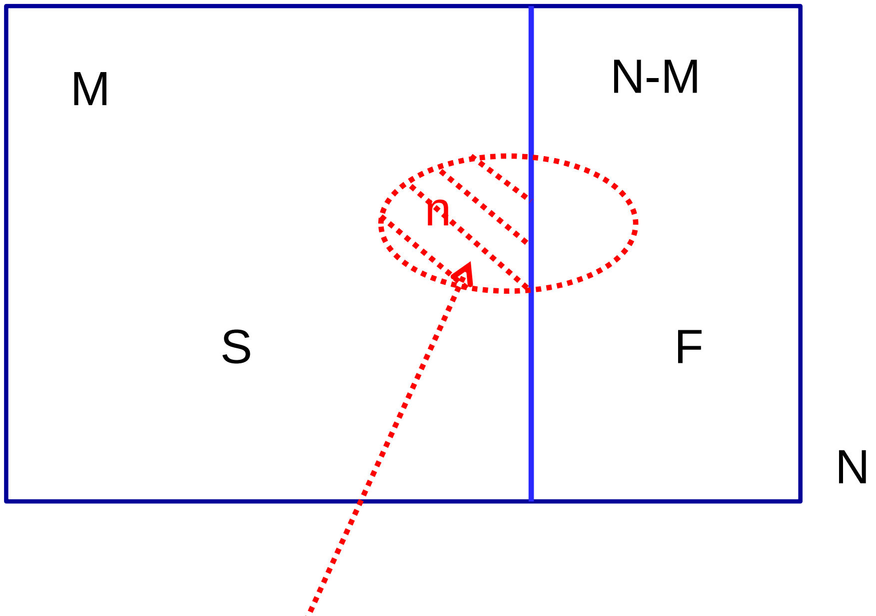
Step 2: Choosing 5-x elements from subset F $\binom{N-M}{n-x} = \binom{8}{5-x}$

- Hypergeometric Distribution

If X is the number of S's in a completely random sample of size n drawn from a population consisting of M S's and (N-M) F's, then the probability distribution of X, called the hypergeometric distribution, is given by

$$P(X = x) = h(x; n, M, N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

- The range of rv X



X = the number of S's in a randomly selected sample of size n

$$\text{Max}(0, n-(N-M)) \leq X \leq \text{Min}(n, M)$$

- # Example 3.36

Five individuals from an animal population thought to be near extinction in a certain region have been caught, tagged, and released to mix into the population. After they have had an opportunity to mix, a random sample of 10 of these animals is selected . Let X=the number of tagged animals in the second sample. If there are actually 25 animals of this type in the region, what is the probability that (a) X=2? (b) $X \le 2$

In this example, N=25, M=5, n =10

$$P(X=x)=\frac{\binom{5}{x}\binom{20}{10-x}}{\binom{25}{10}}, x=0,1,2,3,4,5$$

a) P(X=2)=0.385
b) P(X=0,1,2)=0.699

- Proposition

The mean and variance of the hypergeometric rv X having pmf h(x;n,M,N) are

$$E(X) = np; V(X) = \left(\frac{N-n}{N-1}\right) \cdot n \cdot p \cdot (1-p)$$

≤ 1

where p=M/N

Note: the means of the binomial and hypergeometric rv's are equal, while the variances of the two rv's differ by the factor (N-n)/(N-1) (called ***finite population correction factor***)

- Example 3.37 (Ex. 3.36 Cont')

In the animal-tagging example, $n=10$, $M=5$, and $N=25$, so $p=5/25=0.2$ and

$$E(X) = 10(0.2)=2$$

$$V(X) = (15/24)\,(10)(0.2)(0.8) = 1$$

If the sampling was carried out with replacement, $V(X)=1.6$ (Binomial Distribution)

- The Negative Binomial Distribution

The negative binomial rv and distribution are based on an experiment satisfying the following conditions:

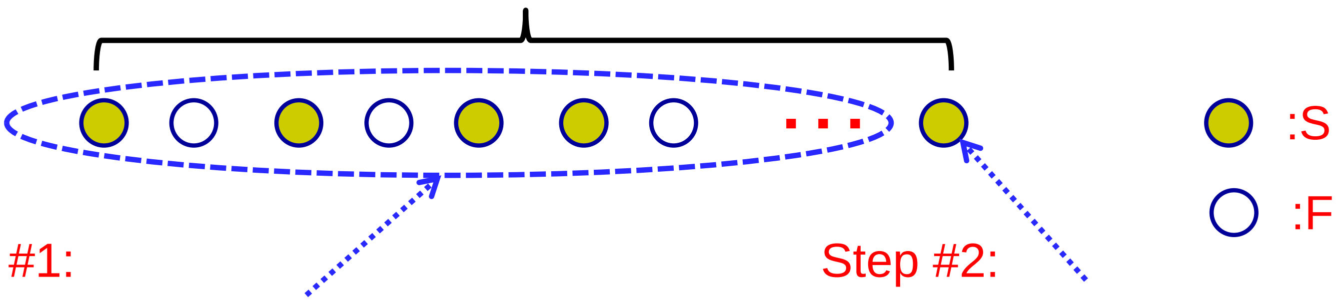
1. The experiment consists of a sequence of independent trials.
2. Each trial can result in either a success (S) or a failure (F).
3. The probability of success is constant from trial to trial, so P(S on trial i)=p for i=1,2,3…
4. The experiment continues until a total of r successes have been observed, where r is a specified positive integer.

The random variable of interest is X= the number of failures that precede the rth success. X is called a negative binomial variable (Here: the number of success is fixed, while the number of trials is random).

3.5 Hypergeometric and Negative Binomial Distributions

Fixed Random

Total number: r (S) + x (F)



Step #1:

Arrage (r-1) S in the first r+x-1 trails

$$\binom{x+r-1}{r-1} p^{r-1}(1-p)^{x}$$

Step #2:

Fixed the final S

$$p$$

$$nb(x;r,p) = \binom{x+r-1}{r-1} p^{r}(1-p)^{x}, x = 0,1,2,\ldots$$

- Example 3.38

A pediatrician wishes to recruit 5 couples, each of whom is expecting their first child, to participate in a new natural childbirth regimen. Let p = P(a randomly selected couple agrees to participate). If p = 0.2, what is the probability that 15 couples must be asked before 5 are found who agree to participate? That is, with S={agrees to participate}, what is the probability that 10 F's occur before the fifth S?

Substituting r=5, p=0.2, and x=10 into $nb(x;r,p)$ gives

$$nb(10;5,.2) = \binom{14}{4}(0.2)^5(0.8)^{10} = 0.034$$

The probability that at most 10 F's are observed (at most 15 couples are asked) is

$$p(X \le 10) = \sum_{x=0}^{10} nb(x;5,0.2) = (0.2)^5 \sum_{x=0}^{10} \binom{x+4}{4}(0.8)^x = 0.164$$

- Proposition

If X is a negative binomial rv with pmf bn(x;r,p), then

$$E(X) = \frac{r(1-p)}{p}\,;\ V(X) = \frac{r(1-p)}{p^2}$$

- Homework

Ex. 68, Ex. 72, Ex. 74, Ex. 76

3.6 The Poisson Probability Distribution

- Poisson Distribution

A random variable X is said to have a Poisson distribution with parameter λ (λ>0) if the pmf of X is

$$p(x;\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!} \qquad x = 0,1,2,3,....$$

The value of λ is frequency a rate per unit time or per unit area. The constant e is the base of the natural logarithm system.

3.6 The Poisson Probability Distribution

- The Maclaurin infinite series expansion of e^{λ}

$$e^{\lambda} = 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}$$

Thus, we have

$$1 = \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^x}{x!}$$

- Proposition

If X has a Poisson distribution with parameter λ, then E(X)=V(X)= λ.

Proof:

$$E(X)=\sum_{x=0}^{\infty} x\frac{e^{-\lambda}\lambda^{x}}{x!}=\sum_{x=1}^{\infty}\frac{e^{-\lambda}\lambda^{x}}{(x-1)!}=\sum_{y=0}^{\infty}\frac{e^{-\lambda}\lambda^{y+1}}{y!}=\lambda e^{-\lambda}\sum_{y=0}^{\infty}\frac{\lambda^{y}}{y!}=\lambda$$

$$E(X^2)=\sum_{x=0}^{\infty} x^2\frac{e^{-\lambda}\lambda^{x}}{x!}=\lambda e^{-\lambda}\sum_{x=1}^{\infty} x\frac{\lambda^{x-1}}{(x-1)!}$$

$$=\lambda e^{-\lambda}\{\sum_{x=1}^{\infty}[(x-1)\frac{\lambda^{x-1}}{(x-1)!}]+[\frac{\lambda^{x-1}}{(x-1)!}]\}=\lambda e^{-\lambda}[\lambda\sum_{x=2}^{\infty}\frac{\lambda^{x-2}}{(x-2)!}+\sum_{x=1}^{\infty}\frac{\lambda^{x-1}}{(x-1)!}]$$

$$=\lambda e^{-\lambda}[\lambda e^{\lambda}+e^{\lambda}]=\lambda^2+\lambda$$

$$V(X)=E(X^2)-E(X)^2=\lambda^2+\lambda-\lambda^2=\lambda$$

3.6 The Poisson Probability Distribution

- Example 3.39

  Let X denote the number of creatures of a particular type captured in a trap during a given time period. Suppose that X has a Poisson distribution with λ=4.5, so on average traps will contain 4.5 creatures. The probability that a trap contains exactly five creatures is

$$P(X=5)=\frac{e^{-4.5}(4.5)^5}{5!}=0.1708$$

The probability that a trap has at most five creatures is

$$P(X\le 5)=\sum_{x=0}^{5}\frac{e^{-4.5}(4.5)^x}{x!}=e^{-4.5}\left[1+4.5+\frac{(4.5)^2}{2!}+\ldots+\frac{(4.5)^5}{5!}\right]=0.7029$$

3.6 The Poisson Probability Distribution

- Example 3.40 (Ex. 3.38 Cont')

  Both the expected number of creatures trapped and the variance of the number trapped equal 4.5, and $\delta_x = (4.5)^{1/2} = 2.12$

3.6 The Poisson Probability Distribution

- The Poisson Distribution as a Limit

Suppose that in the binomial pmf $b(x;n,p)$, we let $n \rightarrow \infty$ and $p \rightarrow 0$ in such a way that *np* approaches a value $\lambda > 0$. Then $b(x;n,p) \rightarrow p(x;\lambda)$ **Proof ?**

According to this proposition, in any binomial experiment in which n is large and p is small,

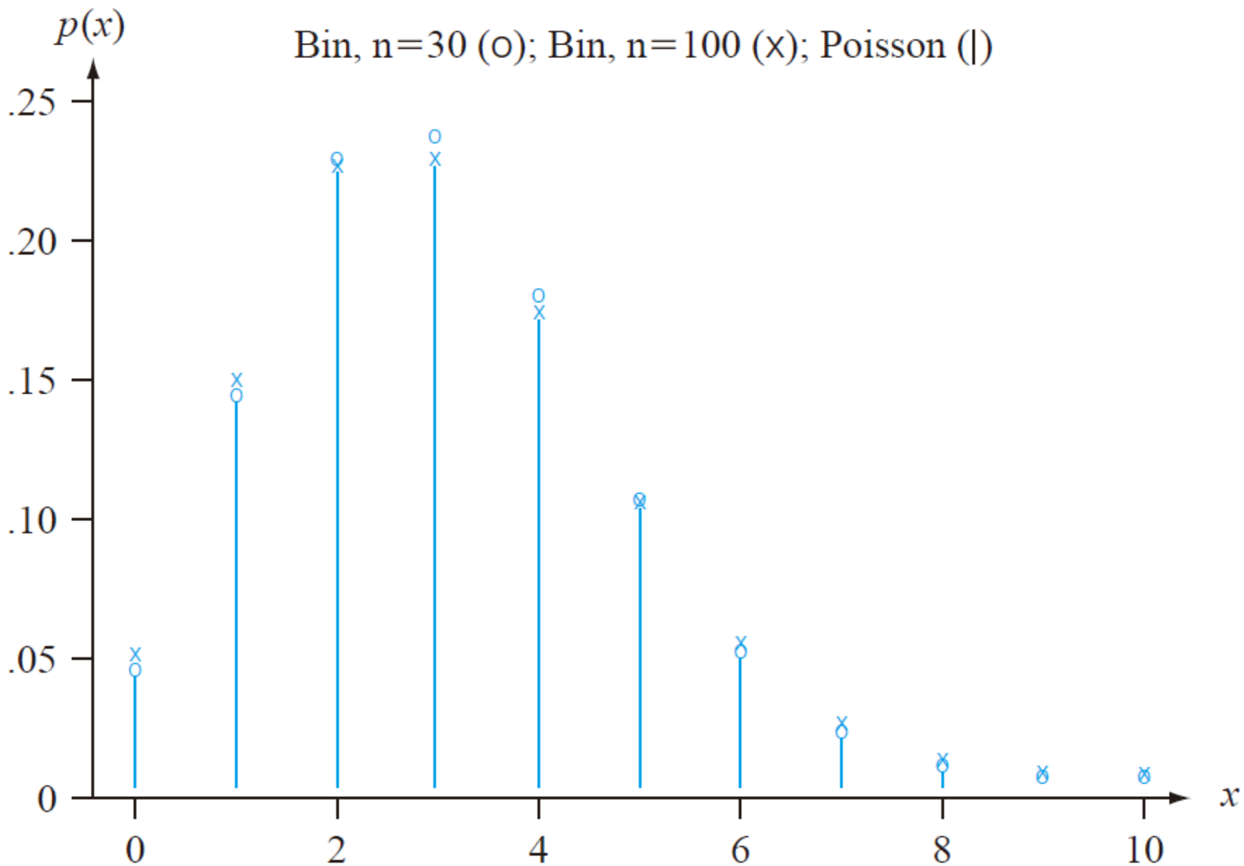
$$b(x;n,p) \approx p(x;\lambda)$$

As a rule, this approximation can safely be applied if $n \geq 100$, and $p \leq 0.01$ and $np \leq 20$

3.6 The Poisson Probability Distribution

Table 3.2 Comparing the Poisson and Three Binomial Distributions

x	$n = 30, p = .1$	$n = 100, p = .03$	$n = 300, p = .01$	Poisson, $\mu = 3$
0	0.042391			
1	0.141304			
2	0.227656			
3	0.236088			
4	0.177066			
5	0.102305			
6	0.047363			
7	0.018043			
8	0.005764			
9	0.001565			
10	0.000365			



3.6 The Poisson Probability Distribution

- Example 3.40

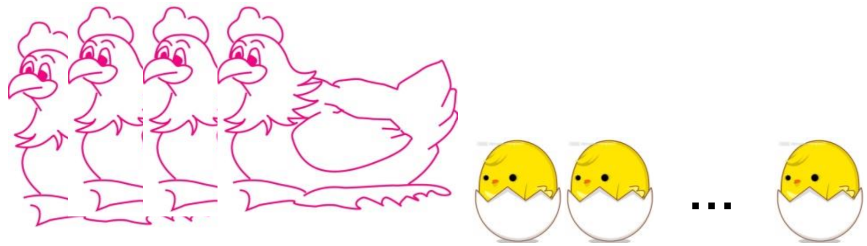
If a publisher of nontechnical books takes great pains to ensure that its books are free of typographical errors, so that the probability of any given page containing at least one such error is 0.005 and errors are independent from page to page, what is the probability that one of its 400-page novels will contain exactly one page with errors? At most three pages with errors?

With S denoting a page containing at least one error and F an error-free page, the number X of pages containing at least one error is a binomial rv with n = 400 and p = 0.005, so np=2. We wish

$$P(X=1)=b(1;400,0.005)\approx p(1;2)=\frac{e^{-2}(2)^1}{1!}=0.271$$

$$\&P(X\le 3)\approx\sum_{x=0}^{3}p(x;2)=\sum_{x=0}^{3}e^{-2}\frac{2^x}{x!}=0.135+0.271+0.271+0.180=0.857$$

3.6 The Poisson Probability Distribution



A time unit: *e.g.* 1 year

Poisson Distribution (# of egg)

| Δt | Δt | Δt | Δt | ... | Δt |

1. Divided it into many (or infinite) independent n small trials: *e.g.* 1 day or less
2. For each trial, the outcome is either S (one egg) or F (none); and P(S) =p_n
3. $np_n \rightarrow \lambda$

Binomial Distribution → Poisson Distribution

3.6 The Poisson Probability Distribution

- Homework

  Ex. 79, Ex. 81, Ex. 82, Ex. 83