Useful Discrete Probability Distributions

Bernoulli, Binomial and Poisson Probability Distributions

Dr.Mamatha.H.R
Professor
Department of Computer Science and Engineering
PES University
Bangalore

Course material created using various Internet resources and text book

Note 6 of 5E

Review

I. What we learnt?

Experiment, Event, Sample space, Probability, Counting rules, Conditional probability, Bayes's rule, random variables, mean, variance.

II. What's in this lecture?

Bernoulli, Binomial, and Poisson Probability Distributions.

Introduction

- **Discrete random variables take on only a finite or countable number of values.**
- **There are several useful discrete probability distributions. We will learn Bernoulli ,Binomial and Poisson distributions.**

Bernoulli Trials

http://www.math.wichita.edu/history/topics/probability.html#bern-trials

- **Boy? Girl? Heads? Tails? Win? Lose? Do any of these sound familiar? When there is the possibility of only two outcomes occuring during any single event, it is called a Bernoulli Trial. Jakob Bernoulli, a profound mathematician of the late 1600s, from a family of mathematicians, spent 20 years of his life studying probability. During this study, he arrived at an equation that calculates probability in a Bernoulli Trial. His proofs are published in his 1713 book *Ars Conjectandi* (Art of Conjecturing).**

What constitutes a Bernoulli Trial?
http://www.math.wichita.edu/history/topics/probability.html#bern-trials

- **To be considered a** Bernoulli trial, **an experiment must meet each of three criteria:**

- **There must be only 2 possible outcomes, such as: black or red, sweet or sour. One of these outcomes is called a** success, **and the other a** failure. **Successes and Failures are denoted as S and F, though the terms given do not mean one outcome is more desirable than the other.**

- **Each outcome has a fixed probability of occurring; a success has the probability of** p**, and a failure has the probability of** $1 - p$**.**
- **Each experiment and result are completely independent of all others.**

Some examples of Bernoulli Trials

http://en.wikipedia.org/wiki/Bernoulli_trial

- Flipping a coin. In this context, obverse ("heads") denotes success and reverse ("tails") denotes failure. A fair coin has the probability of success 0.5 by definition.
- Rolling a die, where for example we designate a six as "success" and everything else as a "failure".
- In conducting a political opinion poll, choosing a voter at random to ascertain whether that voter will vote "yes" in an upcoming referendum.
- Call the birth of a baby of one gender "success" and of the other gender "failure." (Take your pick.)

- **For any Bernoulli trial,**

we define a random variable X as follows:

- **If the experiment results in success,**
- **then X = 1. Otherwise X = 0.**

- **It follows that X is a discrete random**
- **variable, with probability mass function p(x) defined by**
- **$p(0) = P(X = 0) = 1 - p$**
- **$p(1) = P(X = 1) = p$**

- $p(x) = 0$ for any value of x other than 0 or 1
- The random variable X is said to have the Bernoulli distribution with parameter p.
- The notation is $X \sim \text{Bernoulli}(p)$.

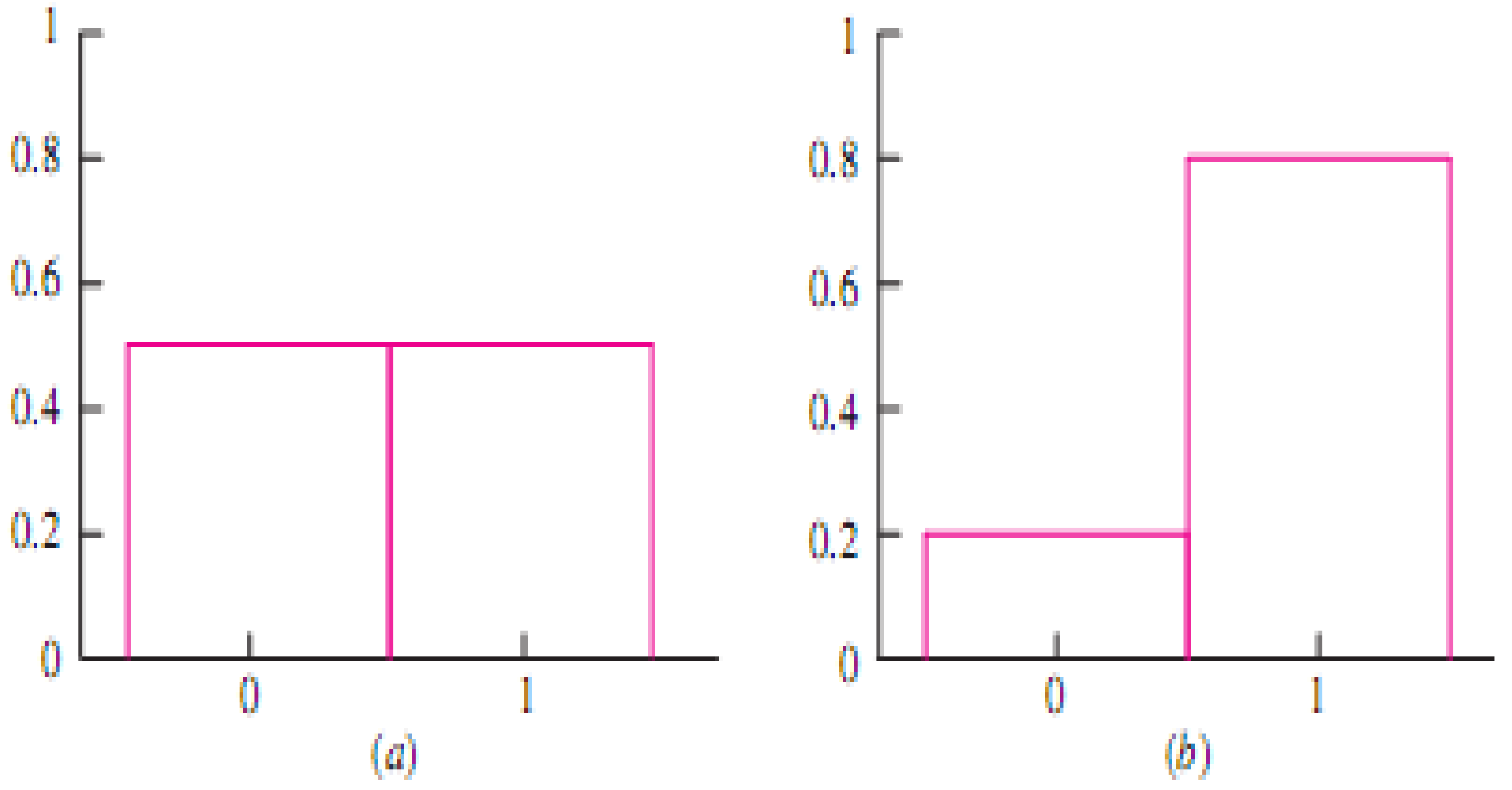


FIGURE 4.1 *(a)* The Bernoulli(0.5) probability histogram. *(b)* The Bernoulli(0.8) probability histogram.

- **Example:**
- **A coin has probability 0.5 of landing heads when tossed. Let X = 1 if the coin comes up heads, and X = 0 if the coin comes up tails. What is the distribution of X?**

- Solution
- Since $X = 1$ when heads comes up, heads is the success outcome.
- The success
- probability, $P(X = 1)$, is equal to 0.5. Therefore $X \sim \text{Bernoulli}(0.5)$.

- **A die has probability 1/6 of coming up 6 when rolled. Let X = 1 if the die comes up 6, and X = 0 otherwise. What is the distribution of X?**

- **Solution**
- **The success probability is $p = P(X = 1) = 1/6$.**
- **Therefore $X \sim \text{Bernoulli}(1/6)$.**

- Ten percent of the components manufactured by a certain process are defective. A component is chosen at random. Let $X = 1$ if the component is defective, and $X = 0$ otherwise. What is the distribution of X?

- **Solution**
- **The success probability is $p = P(X = 1) = 0.1$.**
- **Therefore $X \sim \text{Bernoulli}(0.1)$.**

Mean and Variance of a Bernoulli Random Variable

Mean and Variance of a Bernoulli Random Variable

It is easy to compute the mean and variance of a Bernoulli random variable. If $X \sim \text{Bernoulli}(p)$, then, using Equations (2.29) and (2.30) (in Section 2.4), we compute

Mean : $\mu = \sum xp(x)$

Variance : $\sigma^2 = \sum (x-\mu)^2 p(x)$

Standard deviation : $\sigma = \sqrt{\sigma^2}$

$$\mu_X = (0)(1-p) + (1)(p)$$
$$= p$$

$$\sigma_X^2 = (0-p)^2(1-p) + (1-p)^2(p)$$
$$= p(1-p)$$

Summary

If $X \sim \text{Bernoulli}(p)$, then

$$\mu_X = p \qquad (4.1)$$

$$\sigma_X^2 = p(1-p) \qquad (4.2)$$

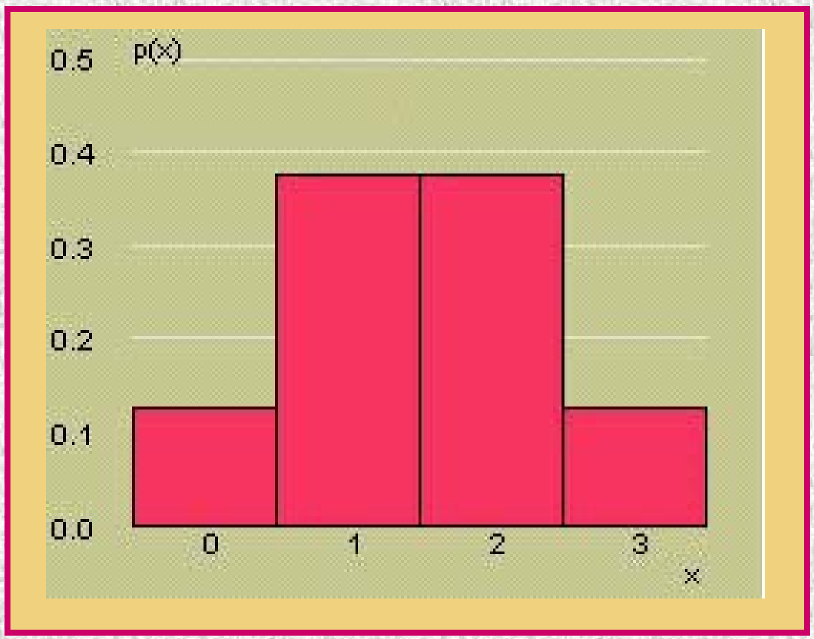
- **Ten percent of the components manufactured by a certain process are defective. A component is chosen at random. Let X = 1 if the component is defective, and X = 0 otherwise. What is the distribution of X? find the mean and variance.**

Solution

Since $X \sim \text{Bernoulli}(0.1)$, the success probability p is equal to 0.1. Using Equations (4.1) and (4.2), $\mu_X = 0.1$ and $\sigma_X^2 = 0.1(1 - 0.1) = 0.09$.

The Binomial Random Variable

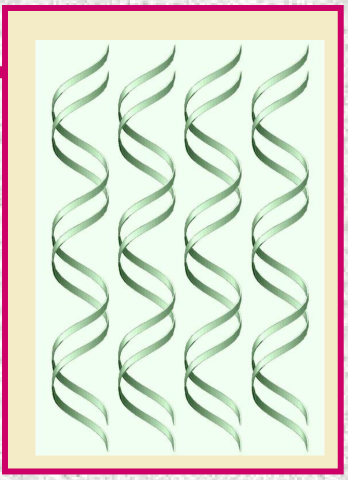
The coin-tossing experiment **is a simple example of a** binomial random variable. **Toss a fair coin *n* = 3 times and record *x* = number of heads.**



x	$p(x)$
0	1/8
1	3/8
2	3/8
3	1/8

The Binomial Random Variable

- **Many situations in real life resemble the coin toss, but the coin is not necessarily fair, so that P(H) $\neq$ 1/2.**



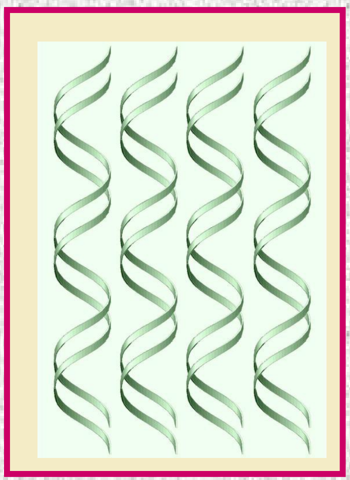
- Example: **A geneticist samples 10 people and counts the number who have a gene linked to Alzheimer's disease.**

- Coin: **Person**
- Head: **Has gene**
- Tail: **Doesn't have gene**
- Number of tosses: **n = 10**
- P(H): **P(has gene) = proportion in the population who have the gene.**

The Binomial Experiment

1. **The experiment consists of** n identical trials.
2. **Each trial results in** one of two outcomes, **success (S) or failure (F).**
3. **The probability of success on a single trial is** p **and** remains constant **from trial to trial. The probability of failure is** $q = 1 - p$.
4. **The trials are** <u>independent</u>**.**
5. **We are interested in** x, the number of successes in n trials.

Binomial or Not?



The independence is a key assumption that often violated in real life applications

- **Select two people from the U.S. population, and suppose that 15% of the population has the Alzheimer's gene.**
  - **For the first person, p = P(gene) = .15**
  - **For the second person, $p \approx$ P(gene) = .15, even though one person has been removed from the population.**

Binomial or Not?

2 out of 20 PCs are defective. We randomly select 3 for testing. Is this a binomial experiment?

1. The experiment consists of n=3 identical trials
2. Each trial result in one of two outcomes
3. The probability of success (finding the defective) is 2/20 and remains the same
4. The trials are not independent. For example,

   P(success on the 2nd trial | success on the 1st trial) = 1/19, not 2/20

Rule of thumb: if the sample size n is relatively large to the population size N, say n/N >= .05, the resulting experiment would not be binomial.

- **A biased coin has probability 0.6 of coming up heads. The coin is tossed three times.**

- **Let *X be the number of heads. Then X ~ Bin(3, 0.6).***

- ***We will compute P(X = 2).***

- There are three arrangements of two heads in three tosses of a coin, HHT, HTH, and THH.
- We first compute the probability of HHT. The event HHT is a sequence of independent events: H on the first toss, H on the second toss, T on the third toss.
- We know the probabilities of each of these events separately:
- *P(H on the first toss)=0.6, P(H on the second toss)=0.6, P(T on the third toss)=0.4*
- Since the events are independent, the probability that they all occur is equal to the product of their probabilities

$$P(\text{HHT}) = (0.6)(0.6)(0.4) = (0.6)^2(0.4)^1$$

Similarly, $P(\text{HTH}) = (0.6)(0.4)(0.6) = (0.6)^2(0.4)^1$, and $P(\text{THH}) = (0.4)(0.6)(0.6) = (0.6)^2(0.4)^1$. It is easy to see that all the different arrangements of two heads and one tail have the same probability. Now

$$\begin{aligned} P(X = 2) &= P(\text{HHT or HTH or THH}) \\ &= P(\text{HHT}) + P(\text{HTH}) + P(\text{THH}) \\ &= (0.6)^2(0.4)^1 + (0.6)^2(0.4)^1 + (0.6)^2(0.4)^1 \\ &= 3(0.6)^2(0.4)^1 \end{aligned}$$

- **we see that the number 3 represents the number of arrangements**
- **of two successes (heads) and one failure (tails), 0.6 is the success probability *p, the* exponent 2 is the number of successes, 0.4 is the failure probability 1 − *p, and the* exponent 1 is the number of failures.**
- **We can now generalize this result to produce a formula for the probability of *x* successes in *n independent Bernoulli trials with success probability p, in terms of x,n,* and *p. In other words, we can compute P(X = x) where X ~ Bin(n, p)***

$$P(X = x) = (\text{number of arrangements of } x \text{ successes in } n \text{ trials}) \cdot p^x (1-p)^{n-x}$$

If $X \sim \text{Bin}(n, p)$, the probability mass function of X is

$$p(x) = P(X = x) = \begin{cases} \dfrac{n!}{x!(n-x)!} p^x (1-p)^{n-x} & x = 0, 1, \ldots, n \\ 0 & \text{otherwise} \end{cases} \qquad (4.4)$$

The Binomial Probability Distribution

SticiGui

For a binomial experiment with *n* trials and probability *p* of success on a given trial, the probability of *k* successes in *n* trials is

$$P(x = k) = C_k^n p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k} \text{ for } k = 0,1,2,\ldots n.$$

Recall $C_k^n = \dfrac{n!}{k!(n-k)!}$

with $n! = n(n-1)(n-2)\ldots(2)1$ and $0! \equiv 1$.

The Mean and Standard Deviation

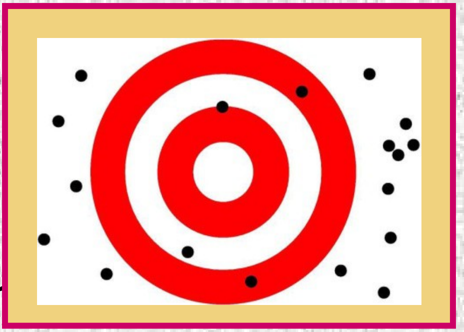
For a binomial experiment with *n* trials and probability *p* of success on a given trial, the measures of center and spread are:

$$\text{Mean}: \mu = np$$

$$\text{Variance}: \sigma^2 = npq$$

$$\text{Standard deviation}: \sigma = \sqrt{npq}$$

Example



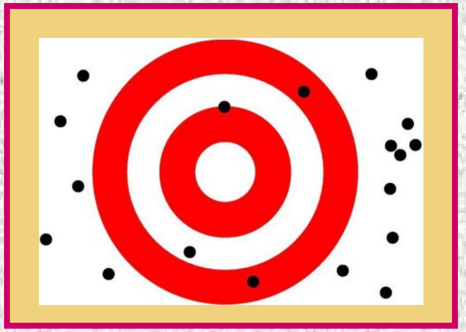
A marksman hits a target 80% of the time. He fires five shots at the target. What is the probability that exactly 3 shots hit the target?

n = 5 *success* = hit *p* = .8 *x* = # of hits

$$P(x=3)=C_3^n p^3 q^{n-3} = \frac{5!}{3!2!}(.8)^3(.2)^{5-3}$$

$$=10(.8)^3(.2)^2=.2048$$

Example



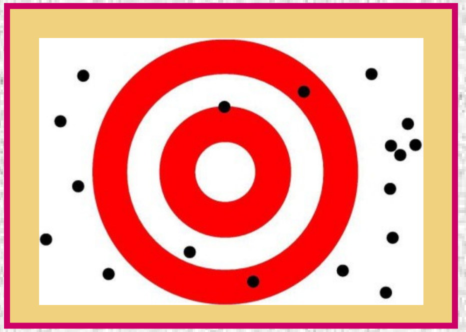
What is the probability that more than 3 shots hit the target?

$$P(x>3)=C_4^5 p^4 q^{5-4}+C_5^5 p^5 q^{5-5}$$

$$=\frac{5!}{4!1!}(.8)^4(.2)^1+\frac{5!}{5!0!}(.8)^5(.2)^0$$

$$=5(.8)^4(.2)+(.8)^5=.7373$$

Cumulative Probability Tables

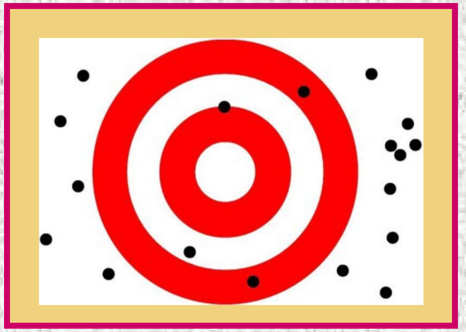


You can use the cumulative probability tables **to find probabilities for selected binomial distributions.**

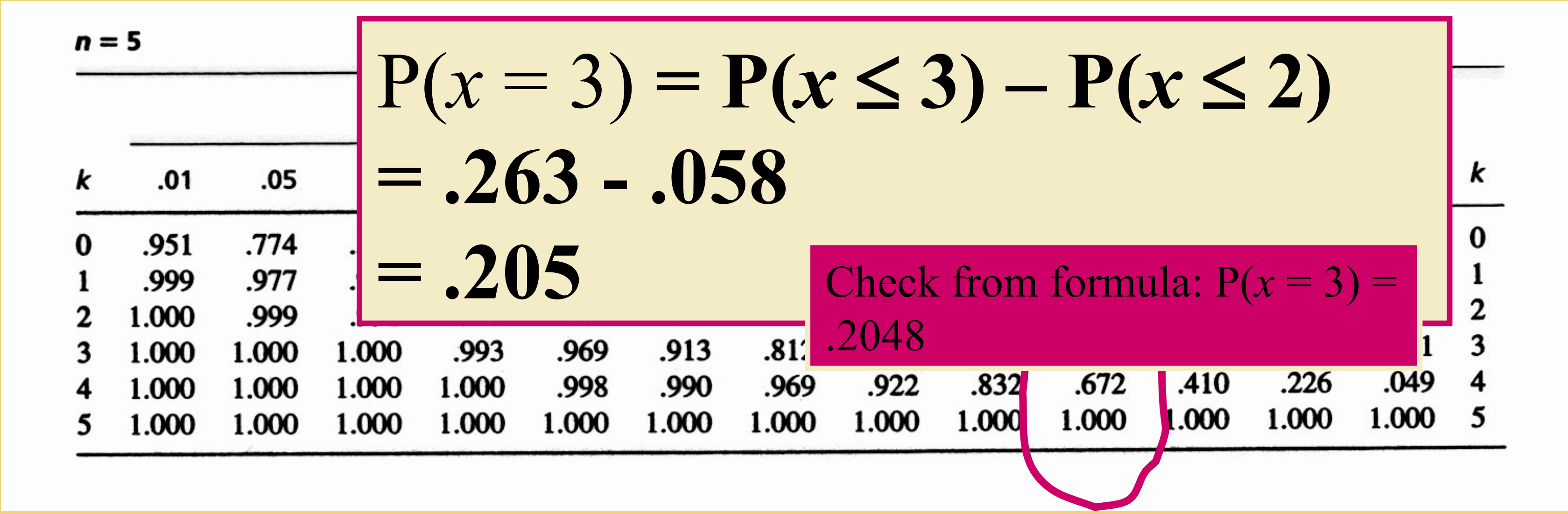
- ✓Find the table for the correct value of *n*.
- ✓Find the column for the correct value of *p*.
- ✓**The row marked "*k*" gives the cumulative probability, $P(x \leq k) = P(x = 0) + \ldots + P(x = k)$**

Example

k	*p* = .80
0	.000
1	.007
2	.058
3	.263
4	.672
5	1.000



What is the probability that exactly 3 shots hit the target?



k	.01	.05												*k*
0	.951	.774												0
1	.999	.977												1
2	1.000	.999												2
3	1.000	1.000	1.000	.993	.969	.913								3
4	1.000	1.000	1.000	1.000	.998	.990	.969	.922	.832	.672	.410	.226	.049	4
5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	5

$n = 5$

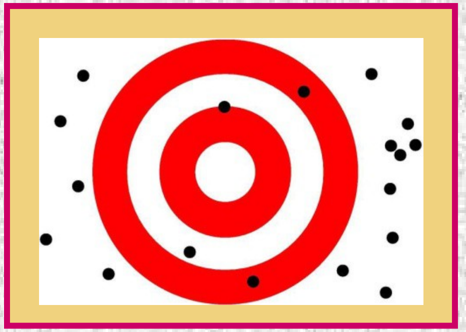
$P(x = 3) = \mathbf{P(x \le 3) - P(x \le 2)}$

$\mathbf{= .263 - .058}$

$\mathbf{= .205}$

Check from formula: $P(x = 3) = .2048$

Example



k	*p* = .80
0	.000
1	.007
2	.058
3	.263
4	.672
5	1.000

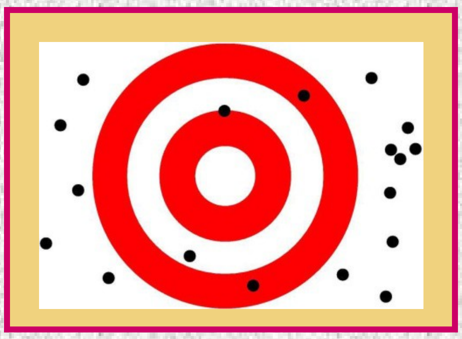
What is the probability that more than 3 shots hit the target?

$P(x > 3) = \mathbf{1 - P(x \leq 3)}$

$\mathbf{= 1 - .263 = .737}$

Check from formula: $P(x > 3) = .7373$

Example



Would it be unusual to find that none of the shots hit the target?

$\mathrm{P}(x = 0) = \mathbf{P}(\boldsymbol{x} \leq \mathbf{0}) = \mathbf{0}$

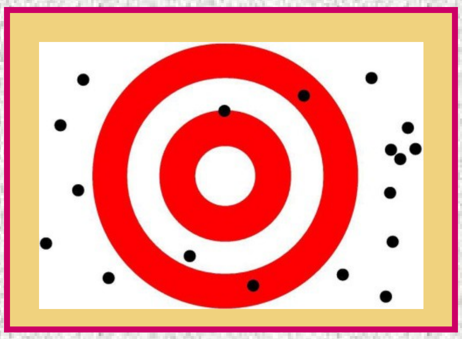
What is the probability that less than 3 shots hit the target?

$\mathrm{P}(x < 3) = \mathbf{P}(\boldsymbol{x} \leq \mathbf{2}) = \mathbf{0.058}$

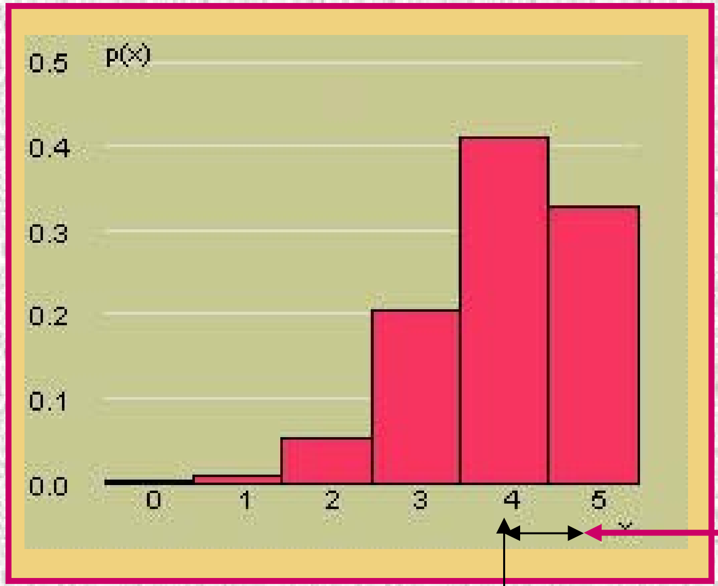
What is the probability that less than 4 but more than 1 shots hit the target?

$\mathrm{P}(1 < x < 4) = \mathbf{P}(\boldsymbol{x} \leq \mathbf{3}) - \mathbf{P}(\boldsymbol{x} \leq \mathbf{1})$
$= \mathbf{.263 - .007 = .256}$

Example



Here is the probability distribution for x = number of hits. **What are the mean and standard deviation for *x*?**



μ

Mean: $\mu = np = 5(.8) = 4$

Standard deviation: $\sigma = \sqrt{npq}$

$= \sqrt{5(.8)(.2)} = .89$

- **A fair coin is tossed 10 times. Let X *be the number of heads that appear. What is the* distribution of X?**
- Solution
- **There are 10 independent Bernoulli trials, each with success probability**
- ***$p = 0.5$.***
- **The random variable X *is equal to the number of successes in the 10 trials. Therefore* $X \sim Bin(10, 0.5)$.**

- **A lot contains several thousand components, 10% of which are defective. Seven components are sampled from the lot. Let** ***X represent the number of defective components*** **in the sample. What is the distribution of** ***X*****?**
- Solution
- **Since the sample size is small compared to the population (i.e., less than 5%), the**
- **number of successes in the sample approximately follows a binomial distribution.**
- **Therefore we model** ***X with the Bin(7, 0.1)***

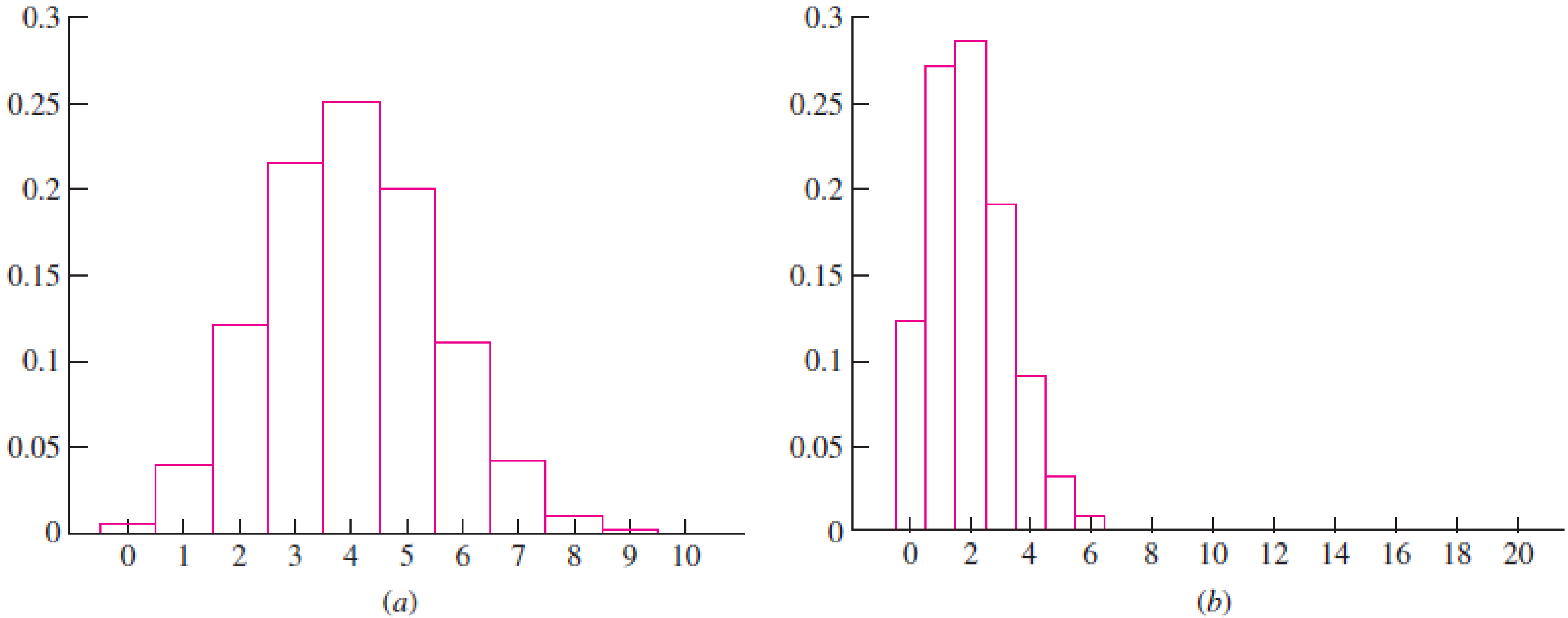


FIGURE 4.2 *(a)* The Bin(10, 0.4) probability histogram. *(b)* The Bin(20, 0.1) probability histogram.

- **Find the probability mass function of the random variable *X if X ~ Bin(10, 0.4).***
- **Find *P(X = 5).***
- Solution
- **We use Equation with *n = 10 and p = 0.4. The probability mass function is***

$$p(x) = \begin{cases} \dfrac{10!}{x!(10-x)!}(0.4)^x(0.6)^{10-x} & x = 0, 1, \ldots, 10 \\ 0 & \text{otherwise} \end{cases}$$

$$P(X = 5) = p(5) = \frac{10!}{5!(10-5)!}(0.4)^5(0.6)^{10-5}$$

$$= 0.2007$$

Using a Sample Proportion to Estimate a Success Probability

- **In many cases we do not know the success probability *p associated with a certain***
- **Bernoulli trial, and we wish to estimate its value. A natural way to do this is to conduct *n independent trials and count the number X of successes. To estimate the success* probability *p***
- ***we compute the sample proportion p***

$$\hat{p} = \frac{\text{number of successes}}{\text{number of trials}} = \frac{X}{n}$$

Note 6 of 5E

- **A quality engineer is testing the calibration of a machine that packs ice cream into containers. In a sample of 20 containers, 3 are underfilled. Estimate the probability *p***

  that the machine underfills a container.
- Solution
- **The sample proportion of underfilled containers is $p = 3/20 = 0.15$.**
- ***We estimate* that the probability *p that the machine underfills a container is 0.15 as well.***

Uncertainty in the Sample Proportion

- **It is important to realize that the sample proportion hat(*p*)** ***is just an estimate of the success*** **probability** ***p, and in general, is not equal to p.***
- ***If another sample were taken, the value of***
- **hat(*p*)** ***would probably come out differently.***
- ***In other words, there is uncertainty in hat(p).***
- ***For ^p*** **to be useful, we must compute its bias and its uncertainty**

- **Let *n denote* the sample size, and let *X* *denote the number of successes, where X ~Bin(n, p).***

The bias is the difference $\mu_{\hat{p}} - p$. Since $\hat{p} = X/n$,

$$\mu_{\hat{p}} = \mu_{X/n} = \frac{\mu_X}{n} = \frac{np}{n} = p$$

Since $\mu_{\hat{p}} = p$, $\hat{p}$ is unbiased; in other words, its bias is 0.

The uncertainty is the standard deviation $\sigma_{\widehat{p}}$.

deviation of X is $\sigma_X = \sqrt{np(1-p)}$. Since $\widehat{p} = X/n$

$$\sigma_{\widehat{p}} = \sigma_{X/n} = \frac{\sigma_X}{n}$$

$$= \frac{\sqrt{np(1-p)}}{n} = \sqrt{\frac{p(1-p)}{n}}$$

In practice, when computing the uncertainty in $\widehat{p}$, we don't know the success probability p, so we approximate it with $\widehat{p}$.

- **The safety commissioner in a large city wants to estimate the proportion of buildings in the city that are in violation of fire codes. A random sample of 40 buildings is chosen for inspection, and 4 of them are found to have fire code violations. Estimate the proportion of buildings in the city that have fire code violations, and find the**

uncertainty in the estimate.

- Solution
- **Let *p denote the proportion of buildings in the city that have fire code violations. The* sample size (number of trials) is *n = 40. The number of buildings with violations* (successes) is *X = 4.***
- ***We estimate p with the sample proportion hat(p):***

$$\hat{p} = \frac{X}{n} = \frac{4}{40} = 0.10$$

the uncertainty in $\widehat{p}$ is

$$\sigma_{\widehat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Substituting $\widehat{p} = 0.1$ for p and 40 for n, we obtain

$$\sigma_{\widehat{p}} = \sqrt{\frac{(0.10)(0.90)}{40}}$$

$$= 0.047$$

- **it tu s rather large. We can reduce the uncertainty by increasing the sample size.**

- In the previous Example, approximately how many additional buildings must be inspected
- so that the uncertainty in the sample proportion of buildings in violation will be only 0.02?

Solution

We need to find the value of n so that $\sigma_{\widehat{p}} = \sqrt{p(1-p)/n} = 0.02$. Approximating p with $\widehat{p} = 0.1$, we obtain

$$\sigma_{\widehat{p}} = \sqrt{\frac{(0.1)(0.9)}{n}} = 0.02$$

Solving for n yields $n = 225$. We have already sampled 40 buildings, so we need to sample 185 more.

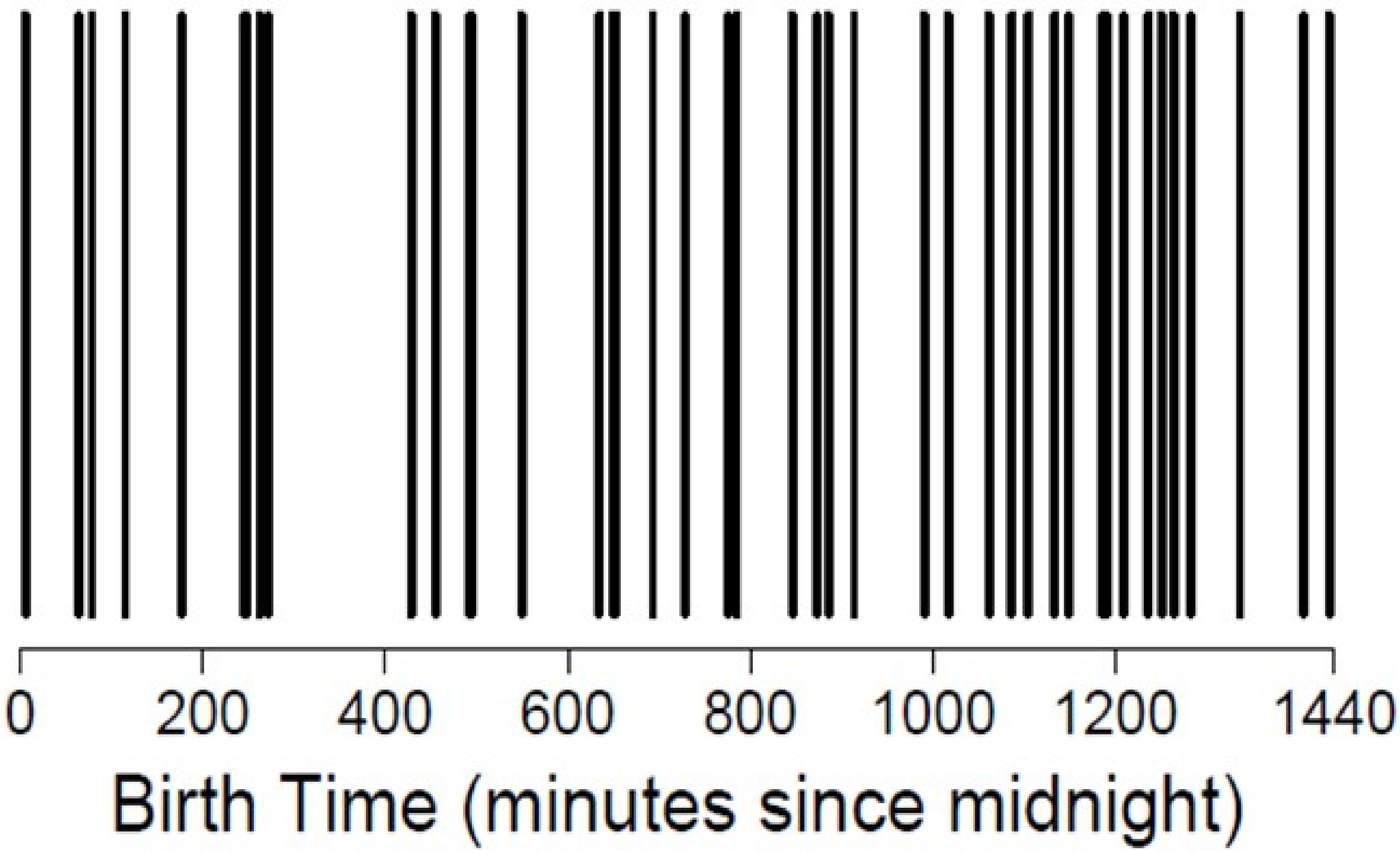
Random events in time and space

Many experimental situations occur in which we observe the counts of events within a set unit of time, area, volume, length etc. For example,

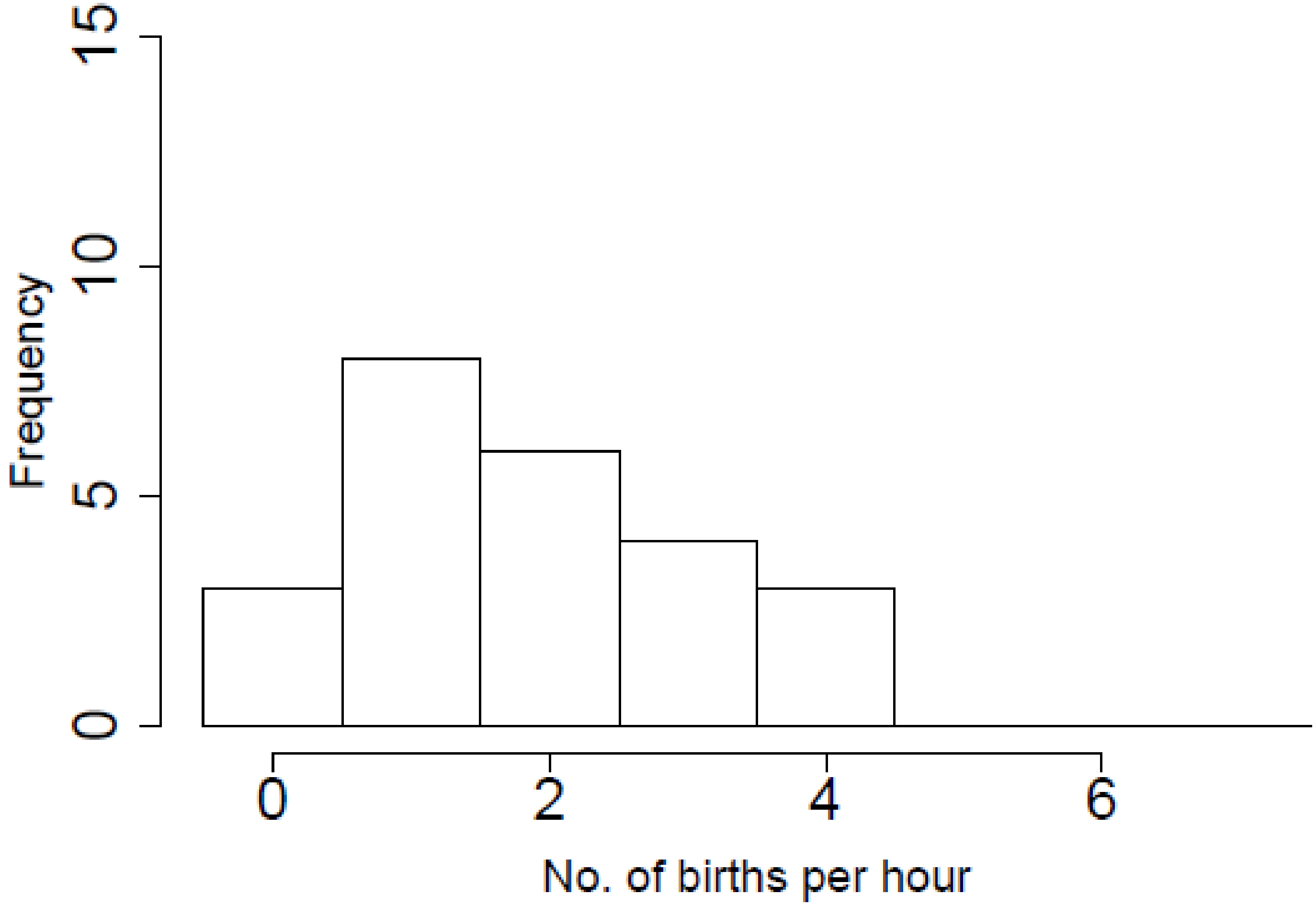
- The number of cases of a disease in different towns
- The number of mutations in set sized regions of a chromosome

In such situations we are often interested in whether the events occur randomly in time or space or not.

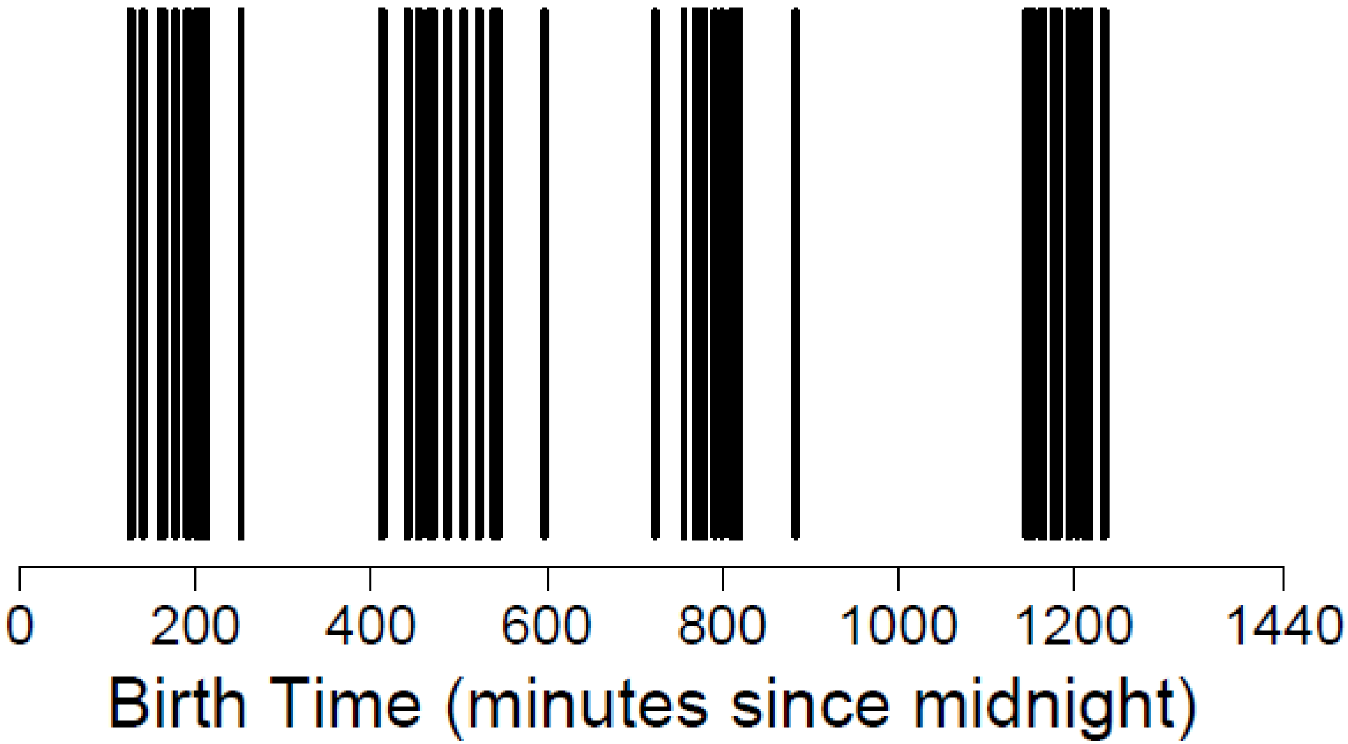
Consider the birth times from the Babyboom dataset



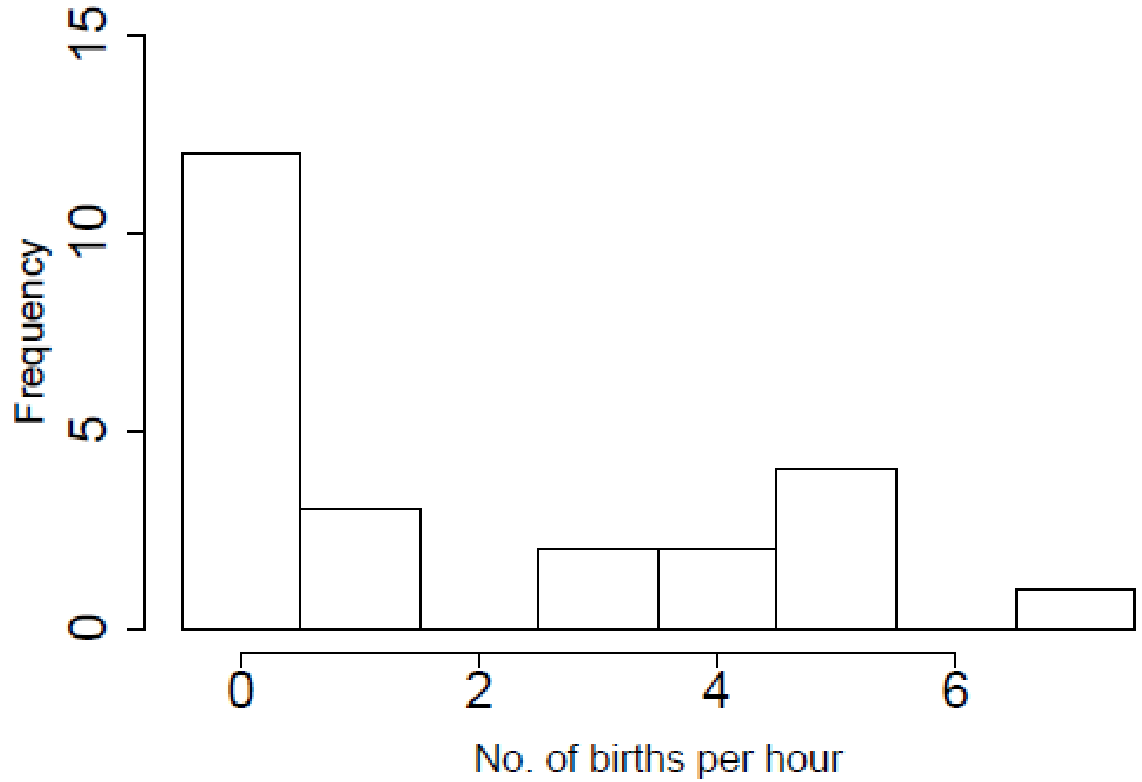
We can plot a histogram of births per hour.



Consider the following sequence of birth times that are obviously not random.



We observe a very different pattern in the histogram of these birth times per hour.



- **This example illustrates that the distribution of counts**

is useful in uncovering whether the events might occur randomly or non-randomly in time (or space).

- **Simply looking at the histogram isn't sufficient if we want to ask the question whether the events occur randomly or not.**
- **To answer this question we need a probability model for the distribution of counts of random events that dictates the type of distributions we should expect to see.**

Some events are rather <u>rare</u> - they don't happen that often. For instance, car accidents are the exception rather than the rule. Still, over a period of time, we can say something about the nature of rare events.

An example is the improvement of traffic safety, where the government wants to know whether seat belts reduce the number of death in car accidents. Here, the Poisson distribution can be a useful tool to answer questions about benefits of seat belt use.

Other phenomena that often follow a Poisson distribution are death of infants, the number of misprints in a book, the number of customers arriving.

The distribution was derived by the French mathematician Siméon Poisson in 1837, and the first application was the description of the number of deaths by horse kicking in the Prussian army.

The Poisson Random Variable

- **The Poisson random variable *x* is often a model for data that represent the number of occurrences of a specified event in a given unit of time or space.**

- Examples:
  - **The number of calls received by a switchboard during a given period of time.**
  - **The number of machine breakdowns in a day**
  - **The number of traffic accidents at a given**

- **One way to think of the Poisson distribution is as an approximation to the binomial distribution when *n is large* and *p is small.***
- *We illustrate with an example.*
- **A mass contains 10,000 atoms of a radioactive substance. The probability that a given atom will decay in a one-minute time period is 0.0002. Let *X represent the number of* atoms that decay in one minute. Now each atom can be thought of as a Bernoulli trial, where success occurs if the atom decays. Thus *X is the number of successes in 10,000* independent Bernoulli trials, each with success probability 0*.0002, so the distribution of***
- ***X is Bin(10,000, 0.0002). The mean of X is μX = (10,000)(0.0002) = 2.***

- **Another mass contains 5000 atoms, and each of these atoms has probability 0.0004 of decaying in a one-minute time interval. Let *Y represent the number of atoms* that decay in one minute from this mass. By the reasoning in the previous paragraph,**
- ***$Y \sim Bin(5000, 0.0004)$ and $\mu Y = (5000)(0.0004) = 2$.***

- **In each of these cases, the number of trials *n and the success probability p are* different, but the mean number of successes, which is equal to the product *np, is the* same.**
- **compute the probability that exactly three atoms decay in one minute for each of these masses using the binomial probability mass function**

$$P(X = 3) = \frac{10{,}000!}{3!\,9997!}(0.0002)^3(0.9998)^{9997} = 0.180465091$$

$$P(Y = 3) = \frac{5000!}{3!\,4997!}(0.0004)^3(0.9996)^{4997} = 0.180483143.$$

- **when *n is large and p is small the mass function depends almost entirely on the mean np, and very little on the specific values of n and p.***
- ***We can therefore approximate the* binomial mass function with a quantity that depends on the product *np only.***
- ***Specifically, if n is large and p is small, and we let λ = np, it can be shown by advanced methods* that for all *x,***

$$\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x} \approx e^{-\lambda}\frac{\lambda^x}{x!} \tag{4.8}$$

We are led to define a new probability mass function, called the Poisson probability mass function. The Poisson probability mass function is defined by

$$p(x) = P(X = x) = \begin{cases} e^{-\lambda}\dfrac{\lambda^x}{x!} & \text{if } x \text{ is a non-negative integer} \\ 0 & \text{otherwise} \end{cases} \tag{4.9}$$

If X is a random variable whose probability mass function is given by Equation (4.9), then X is said to have the **Poisson distribution** with parameter λ. The notation is $X \sim \text{Poisson}(\lambda)$.

The Poisson Probability Distribution

Let x a Poisson random variable. The probability of k occurrences of this event is

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

For values of k = 0, 1, 2, … The mean and standard deviation of the Poisson random variable are

Mean: μ

Στανδαρδ δεϖιατιον:

$$\sigma = \sqrt{\mu}$$

Example

The average number of traffic accidents on a certain section of highway is two per week. Find the probability of exactly one accident during a one-week period.

$$P(x=1)=\frac{\mu^{k}e^{-\mu}}{k!}=\frac{2^{1}e^{-2}}{1!}=2e^{-2}=.2707$$

Cumulative Probability Tables

You can use the cumulative probability tables **to find probabilities for selected Poisson distributions.**

✓Find the column for the correct value of μ.

✓**The row marked "k" gives the cumulative probability, $P(x \le k) = P(x = 0) + \ldots + P(x = k)$**

Example

k	$\mu = 2$
0	.135
1	.406
2	.677
3	.857
4	.947
5	.983
6	.995
7	.999
8	1.000

What is the probability that there is exactly 1 accident?

k	2.0	2.5	3.0	3.5	4.0	4.5
0	.135	.082	.055	.033	.018	.011
8	1.000	.999				
9		1.000	.999	.997	.992	.983
10			1.000	.999	.997	.993
11				1.000	.999	.998
12					1.000	.999
13						1.000

$P(x = 1) = \mathbf{P(x \le 1) - P(x \le 0)}$

$= \mathbf{.406 - .135}$

$= \mathbf{.271}$

Check from formula: $P(x = 1) = .2707$

Example

k	$\mu = 2$
0	.135
1	.406
2	.677
3	.857
4	.947
5	.983
6	.995
7	.999
8	1.000

What is the probability that 8 or more accidents happen?

$$P(x \geq 8) = 1 - P(x < 8)$$
$$= 1 - P(x \leq 7)$$
$$= 1 - .999 = .001$$

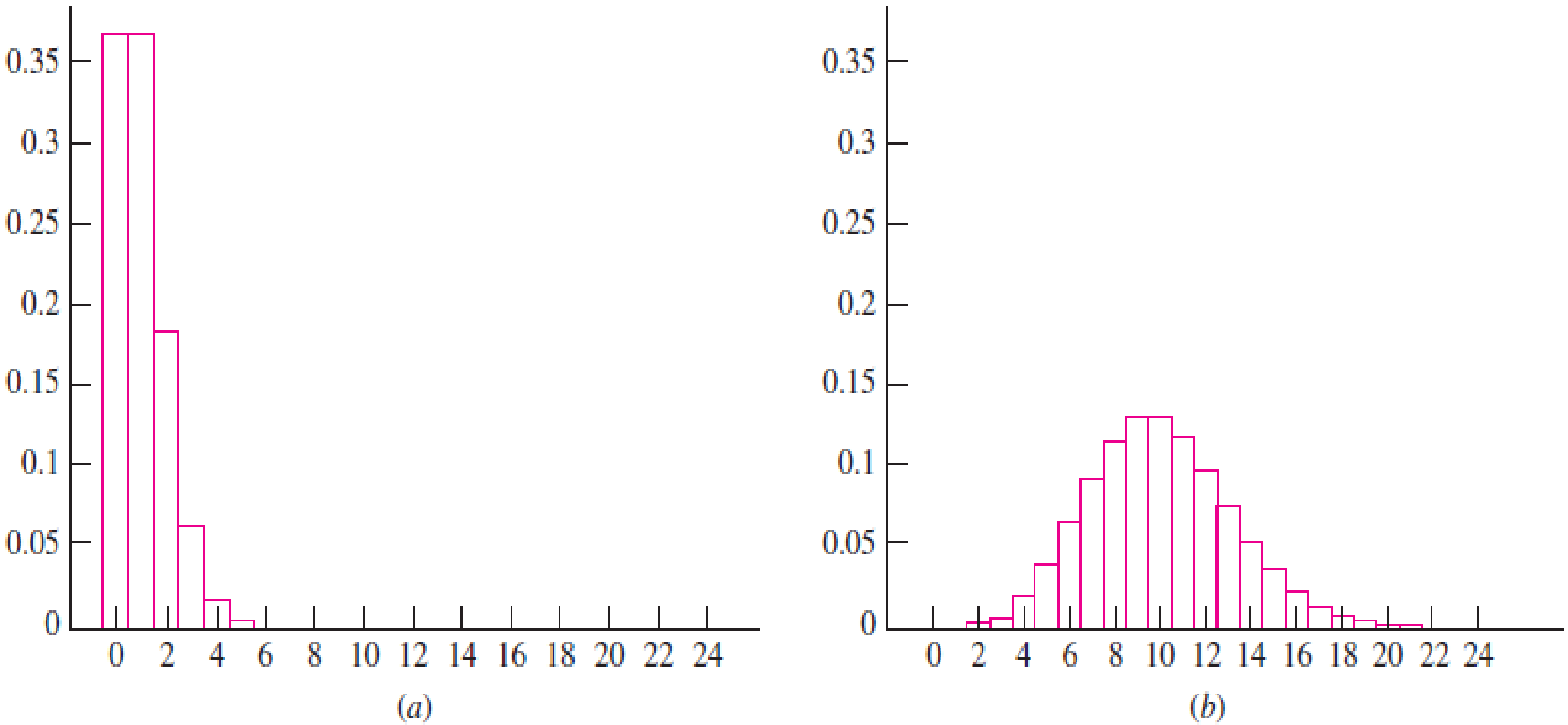


FIGURE 4.3 *(a)* The Poisson(1) probability histogram. *(b)* The Poisson(10) probability histogram.

- Using the Poisson Distribution to Estimate a Rate

- **Often experiments are done to estimate a rate λ *that represents the mean number of* events that occur in one unit of time or space.**
- **In these experiments, the number of events *X that occur in t units is counted, and the rate λ is estimated with the quantity hat(λ) = X/t.***
- ***If the numbers* of events in disjoint intervals of time or space are independent, and if events cannot occur simultaneously, then *X follows a Poisson distribution.***
- ***A process that produces* such events is called a** Poisson process. Since the mean number of events that occur in ***t units of time or space is equal to λt, X ~ Poisson(λt).***

- **A suspension contains particles at an unknown concentration of *λ per mL. The suspension* is thoroughly agitated, and then 4 mL are withdrawn and 17 particles are**

  counted. Estimate *λ*.
- Solution
- **Let *X = 17 represent the number of particles counted, and let t = 4 mL be the* volume of suspension withdrawn.**
- **Then hat(*λ*)= *X/t = 17/4 = 4.25 particles per mL***

Summary

If $X \sim \text{Poisson}(\lambda)$, then the mean and variance of X are given by

$$\mu_X = \lambda \tag{4.10}$$

$$\sigma_X^2 = \lambda \tag{4.11}$$

Uncertainty in the Estimated Rate

- **Let** ***X be the number of events counted in t units of time or space, and*** **assume that** ***X*** **~** ***Poisson(λt).***

The bias is the difference $\mu_{\hat{\lambda}} - \lambda$. Since $\hat{\lambda} = X/t$,

$$\mu_{\hat{\lambda}} = \mu_{X/t} = \frac{\mu_X}{t}$$
$$= \frac{\lambda t}{t} = \lambda$$

Since $\mu_{\hat{\lambda}} = \lambda$, $\hat{\lambda}$ is unbiased.

The uncertainty is the standard deviation $\sigma_{\hat{\lambda}}$. Since $\hat{\lambda} = X/t$, $\sigma_{\hat{\lambda}} = \sigma_X/t$. Since $X \sim$ Poisson(λt), it follows from Equation (4.11) uiat $\sigma_X = \sqrt{\lambda t}$. Therefore

$$\sigma_{\hat{\lambda}} = \frac{\sigma_X}{t} = \frac{\sqrt{\lambda t}}{t} = \sqrt{\frac{\lambda}{t}}$$

E

In practice, the value of λ is unknown, so we approximate it with $\hat{\lambda}$.

A 5 mL sample of a suspension is withdrawn, and 47 particles are counted. Estimate the mean number of particles per mL, and find the uncertainty in the estimate.

Solution

The number of particles counted is $X = 47$. The volume withdrawn is $t = 5$ mL. The estimated mean number of particles per mL is

$$\hat{\lambda} = \frac{47}{5} = 9.4$$

The uncertainty in the estimate is

$$\begin{aligned} \sigma_{\hat{\lambda}} &= \sqrt{\frac{\lambda}{t}} \\ &= \sqrt{\frac{9.4}{5}} \qquad \text{approximating } \lambda \text{ with } \hat{\lambda} = 9.4 \\ &= 1.4 \end{aligned}$$

BASIS FOR COMPARISON	BINOMIAL DISTRIBUTION	POISSON DISTRIBUTION
Meaning	**Binomial distribution is one in which the probability of repeated number of trials are studied.**	**Poisson Distribution gives the count of independent events occur randomly with a given period of time.**
Nature	**Biparametric**	**Uniparametric**
Number of trials	**Fixed**	**Infinite**
Success	**Constant probability**	**Infinitesimal chance of success**
Outcomes	**Only two possible outcomes, i.e. success or failure.**	**Unlimited number of possible outcomes.**
Mean and Variance	**Mean > Variance**	**Mean = Variance**
Example	**Coin tossing experiment.**	**Printing mistakes/page of a large book.**

Key Concepts

The Binomial Random Variable

1. Five characteristics: *n* identical trials, each resulting in either success *S* or failure *F*; probability of success is *p* and remains constant from trial to trial; trials are independent; and *x* is the number of successes in *n* trials.

2. Calculating binomial probabilities

$$P(x = k) = C_k^n p^k q^{n-k}$$

a. Formula:

b. Cumulative binomial tables

3. Mean of the binomial random variable: $\mu = np$

4. Variance and standard deviation: $\sigma^2 = npq$ and

$$\sigma = \sqrt{npq}$$

Key Concepts

II. The Poisson Random Variable

1. The number of events that occur in a period of time or space, during which an average of μ such events are expected to occur

2. Calculating Poisson probabilities

a. Formula: $P(x=k)=\frac{\mu^k e^{-\mu}}{k!}$

b. Cumulative Poisson tables

3. Mean of the Poisson random variable: $E(x) = \mu$

4. Variance and standard deviation: $\sigma^2 = \mu$ and $\sigma=\sqrt{\mu}$