

PSTAT 5E
Statistics with Economics and Business
Applications

Chapter 5
Discrete Probability Distribution

Yan Xu

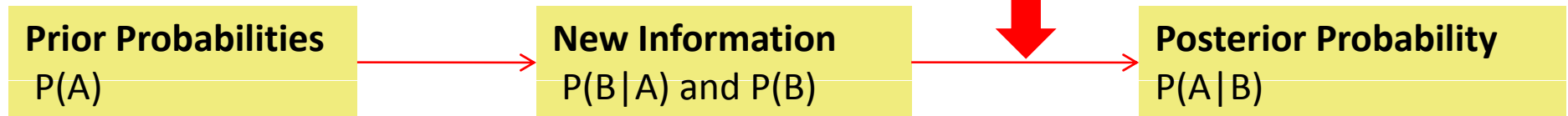
4.5 Bayes' Theorem

Recall:

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

Simple version of
Bayes Formula:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$



Comments: Bayes' theorem is used extensively in making decisions. Prior information are often subjective estimates provided by a decision maker. "New" information is obtained from certain study/report/survey, etc., and the posterior probabilities can be computed for use in choosing the best decision.

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

What if $P(B)$ is not given and instead we observe $P(B|A)$ and $P(B|A^c)$ instead?

The probability of B can be written as

$$\begin{aligned} P(B) &= P(A \cap B) + P(A^c \cap B) \\ &= P(A)P(B|A) + P(A^c)P(B|A^c) \end{aligned}$$

(The Law of Total Probability)

Example:

According to the promotion record of 80 male and 20 Female officers in a large firm in US, we have the following information:

$$P(\text{Promoted} | \text{Male}) = 0.3 \quad P(\text{Promoted} | \text{Female}) = 0.1$$

$$P(\text{Promoted}) = ?$$

Bayes' Th (two-event case)

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

$P(A)$ is the **prior probability** -- initial information about A

$P(B|A)$ and $P(B|A')$ provides some “**additional information**”

$P(A|B)$ is the **posterior probability**— revised probability of event A based on the additional information

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

Example:

According to the promotion record of 80 male and 20 female officers in a large firm in US, we have the following information:

$$P(\text{Promoted} | \text{Male}) = 0.3$$

$$P(\text{Male} | \text{Promoted}) = ?$$

$$P(\text{Promoted} | \text{Female}) = 0.1$$

$$P(\text{Female} | \text{Promoted}) = ?$$

Example

Suppose a rare disease infects one out of every 1000 people in a population. And suppose that there is a good, but not perfect, test for this disease: if a person has the disease, the test comes back positive **99%** of the time. On the other hand, the test also produces some false positives: **2%** of uninfected people are also test positive. For someone just tested positive, what are his chances of having this disease?

Example

Define A: has the disease B: test positive

Information we know:

$$P(A) = .001 \quad P(A^c) = .999$$

$$P(B|A) = .99 \quad P(B|A^c) = .02$$

We want to know $P(A|B)=?$

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A^c)P(B|A^c)}$$
$$= \frac{.001 \times .99}{.001 \times .99 + .999 \times .02} = .0472$$



What if A can be separated into more than two classes, such as, A1-“poor”, A2-“Average”, A3-“good”?

Bayes' Rule (multiple events)

$$\begin{aligned} P(B) &= P(A_1 \cap B) + \dots + P(A_k \cap B) \\ &= P(A_1)P(B | A_1) + \dots + P(A_k)P(B | A_k) \\ &= \sum P(A_i)P(B | A_i) \end{aligned}$$

$$P(A_i | B) = \frac{P(A_i)P(B | A_i)}{\sum P(A_i)P(B | A_i)} \text{ for } i = 1, 2, \dots, k$$

5.1 Random Variables

A random variable(RV) is a **numerical** description of the outcomes of an experiment, usually referred as X .

Discrete random variables take on only a finite or countable number of values, such as, 0,1,2,3....

Continuous random variables may assume any numerical value in an interval, for example, $0 < X < 20$.

| <i>Experiment</i> | <i>RV X</i> | <i>Possible values for X</i> |
|------------------------|----------------------------------|------------------------------|
| Toss a coin | # of head | 0,1 |
| Contact five customers | # of customers who place a order | 0,1,2,3,4,5 |
| Operate a restaurant | # of customers for one day | 0,1,2,3,..... |

Exercise:

| Experiment | Random Variable X | Type |
|-----------------------------|--|----------|
| Family size | x = Number of dependents reported on tax return | Discrete |
| Distance from home to store | x = Distance in miles from home to the store site | Cont |
| Own dog or cat | x = 1 if own no pet; = 2 if own dog(s) only; = 3 if own cat(s) only; = 4 if own dog(s) and cat(s) | Discrete |

5.2 Discrete Probability Distribution

The **probability distribution** for a random variable X describes how probabilities are distributed over the values of the random variable X .

Define $f(x)$ as the **probability function**, which provides the probability for each value of the random variable X .

Conditions for $f(x)$

$$0 \leq f(x) \leq 1 \text{ and } \sum f(x) = 1$$

The possible values of a RV X and the associated probabilities $f(x)$ can be shown in a table as belows:

Example1 : Roll a fair die

$X = \#$ of dots on the upward face

$X = \{1, 2, 3, 4, 5, 6\}$

| X | $f(x)$ |
|-----|--------|
| 1 | $1/6$ |
| 2 | $1/6$ |
| 3 | $1/6$ |
| 4 | $1/6$ |
| 5 | $1/6$ |
| 6 | $1/6$ |

total=1

Example2:

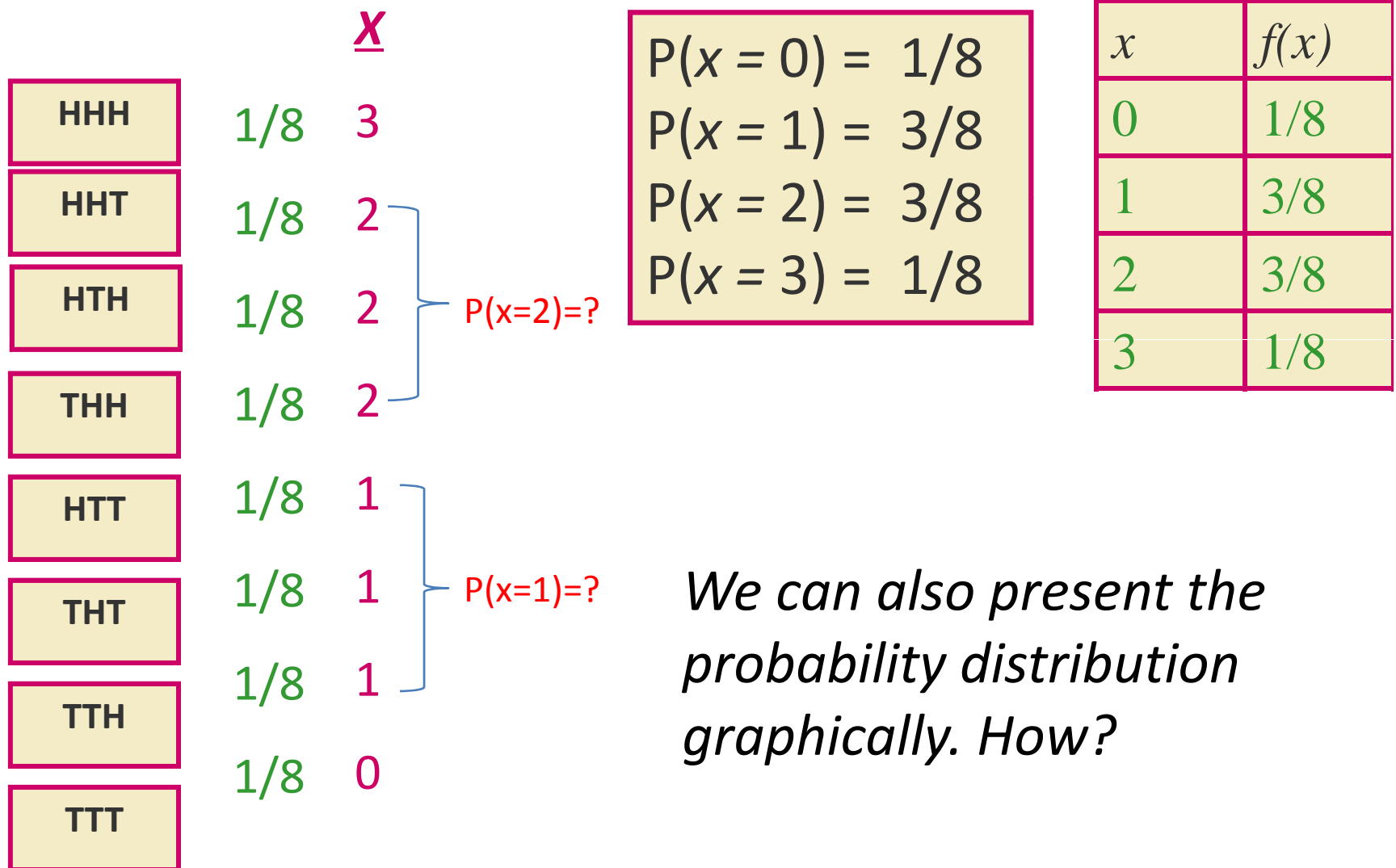
$X = \#$ of iphone's sold per day in a particular store

Suppose $X = \{5, 10, 15, 20\}$

| X | $f(x)$ |
|-----|--------|
| 5 | $4/10$ |
| 10 | $3/10$ |
| 15 | ? |
| 20 | $1/10$ |

Example

Toss a fair coin 3 times and $x = \text{number of heads}$.



Discrete Uniform Distribution

ALL values of the random variable are **equally likely**.

$f(x)=1/n$, $n=$ # of values X may assume

Example:

| X | $f(x)$ |
|-----|--------|
| 10 | 1/8 |
| 20 | 1/8 |
| 30 | 1/8 |
| ... | ... |
| 70 | 1/8 |
| 80 | 1/8 |

Questions?

1. Check this is a correct probability function.

2. How to draw the distribution plot ?

(i.e., $f(x)$ against X)

Expected Value and Variance

Let X be a discrete random variable with probability distribution $f(X)$. Then the expected value and variance for the RV X are defined as:

$$E(X) = \mu = \sum xf(x)$$

$$Var(X) = \sigma^2 = \sum (x - \mu)^2 f(x)$$

Standard Deviation is σ , which is the positive root of variance.

Example:

X = # of TVs sold in a day and the associated probability $f(X)$ are shown below:

| x | $f(x)$ | $xf(x)$ |
|-----|--------|---------|
| 0 | .40 | .00 |
| 1 | .25 | .25 |
| 2 | .20 | .40 |
| 3 | .05 | .15 |
| 4 | .10 | .40 |

$$E(x) = 1.20$$

$$E(X) = \mu = \sum xf(x)$$

expected number of
TVs sold in a day

Variance

| x | $x - \mu$ | $(x - \mu)^2$ | $f(x)$ | $(x - \mu)^2 f(x)$ |
|-----|-----------|---------------|--------|--------------------|
| 0 | -1.2 | 1.44 | .40 | .576 |
| 1 | -0.2 | 0.04 | .25 | .010 |
| 2 | 0.8 | 0.64 | .20 | .128 |
| 3 | 1.8 | 3.24 | .05 | .162 |
| 4 | 2.8 | 7.84 | .10 | .784 |

Variance of daily sales = $\sigma^2 = 1.660$

$$\text{Var}(X) = \sigma^2 = \sum (x - \mu)^2 f(x)$$

Standard deviation of daily sales = 1.2884

5.3 Binomial Distribution

Binomial Distribution is one of the most important discrete distributions. It is associated with a multiple-step experiment called Binomial Experiment.

- **Four Properties of a Binomial Experiment**
 1. The experiment consists of a sequence of n identical trials.
 2. Two outcomes, success and failure, are possible on each trial.
 3. The probability of a success, denoted by p , does not change from trial to trial. (p is fixed)
 4. The trials are independent.

Binomial Probability Function

For a binomial experiment with n trials and probability p of success on any one trial, we're interested in **the probability of k successes in n trials.**

Define $X = \#$ of successes occurring in the n trials:

$$f(k) = P(X = k) = C_k^n p^k q^{(n-k)} = \frac{n!}{k!(n-k)!} p^k q^{(n-k)}$$

for $k = 0, 1, 2, 3, \dots, n$

where $q=1-p$, the probability of failure on any trial.

Previous Example

Toss a fair coin **3** times and **x = number of heads**.

X can take **0,1,2** and **3**. **P(x=1)=?** And **P(x=3)=?**

$n = 3$ *success* = **head** $p = 1/2$ $X = \#$ of heads

Apply the formula:

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

$$P(x = 0) = 1/8$$

$$P(x = 1) = 3/8$$

$$P(x = 2) = 3/8$$

$$P(x = 3) = 1/8$$

Exercise

A marksman hits a target **80%** of the time. He fires **5** shots at the target. What is the probability that **exactly 3** shots hit the target?

$n =$ **5**

success = **hit**

$p =$ **.8**

$x =$ **# of hits**

$$P(x = 3) = C_3^n p^3 (1 - p)^{n-3} = \frac{5!}{3!2!} (.8)^3 (.2)^{5-3}$$

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

Exercise

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

$$\begin{aligned}P(x > 3) &= P(x = 4) + P(x = 5) \\&= C_4^5 p^4 q^{5-4} + C_5^5 p^5 q^{5-5} \\&= \frac{5!}{4!1!} (.8)^4 (.2)^1 + \frac{5!}{5!0!} (.8)^5 (.2)^0 \\&= 5(.8)^4 (.2) + (.8)^5 = .7373\end{aligned}$$

For a binomial experiment with n trials and probability p of success on a given trial, the **expected value** and **variance** are of the following close forms:

$$E(X) = \mu = np$$

$$\text{Var}(X) = \sigma^2 = npq$$

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

5.4 Poisson Random Variable

- The Poisson random variable X is used for the number of occurrences of a specified event in a particular time interval.

- Examples:

- # of arrivals at a car wash in one hour
- # of machine breakdowns in a day
- # of traffic accidents at a given intersection during a given time period.

Poisson Probability Distribution

Let X a Poisson random variable. The probability of k occurrences in an interval is

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

For values of $k = 0, 1, 2, \dots$

The expected value and variance:

$$E(X) = \mu$$

$$\text{Var}(X) = \mu$$

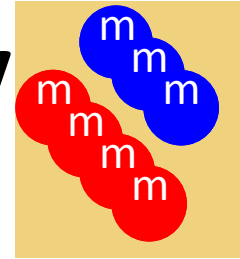
Example

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**.

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

Other time interval can be used. For example, what is the probability of exactly one accident during **two weeks**?

5.5 Hypergeometric Probability Distribution



A bowl contains M red M&M[®] candies and $N - M$ blue M&M[®] candies. Select n candies from the bowl and record x the number of red candies selected. Define a “red M&M[®]” to be a “success”.

The probability of exactly k successes in n trials is

$$P(x = k) = \frac{C_k^M C_{n-k}^{M-N}}{C_n^N}$$

The Mean and Variance

$$\text{Mean : } \mu = n \left(\frac{M}{N} \right)$$

$$\text{Variance : } \sigma^2 = n \left(\frac{M}{N} \right) \left(\frac{N - M}{N} \right) \left(\frac{N - n}{N - 1} \right)$$

Example

A package of 8 AA batteries contains 2 batteries that are defective. A student randomly selects four batteries and replaces the batteries in his calculator. What is the probability that all four batteries work?

$$\begin{array}{l} \text{Success} = \text{working battery} \\ N = 8 \\ M = 6 \\ n = 4 \end{array} \quad P(x = 4) = \frac{C_4^6 C_0^2}{C_4^8} = \frac{6(5) / 2(1)}{8(7)(6)(5) / 4(3)(2)(1)} = \frac{15}{70}$$

Example



What are the mean and variance for the number of batteries that work?

$$\mu = n \left(\frac{M}{N} \right) = 4 \left(\frac{6}{8} \right) = 3$$

$$\begin{aligned} \sigma^2 &= n \left(\frac{M}{N} \right) \left(\frac{N-M}{N} \right) \left(\frac{N-n}{N-1} \right) \\ &= 4 \left(\frac{6}{8} \right) \left(\frac{2}{8} \right) \left(\frac{4}{7} \right) = .4286 \end{aligned}$$

Summary

Discrete Random Variables and Probability

1. Random variables, discrete and continuous
2. Properties of probability distributions

$$0 \leq p(x) \leq 1 \text{ and } \sum p(x) = 1$$

$$E(X) = \sum xf(x)$$

3. Expected value of a discrete random variable
4. Variance and standard deviation of a discrete random variable:

$$\text{Variance: } \sigma^2 = \sum (x - \mu)^2 f(x)$$

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

5. Binomial Distribution
Poisson Distribution