BME 1132
Probability and Biostatistics

Instructor: Ali AJDER, *Ph.D.*

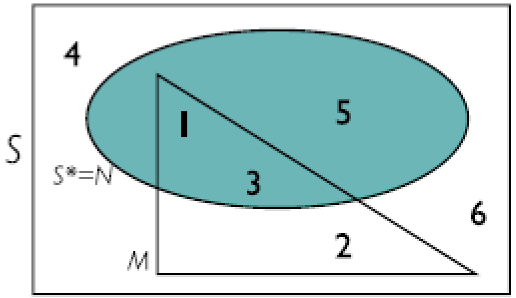
Week-9

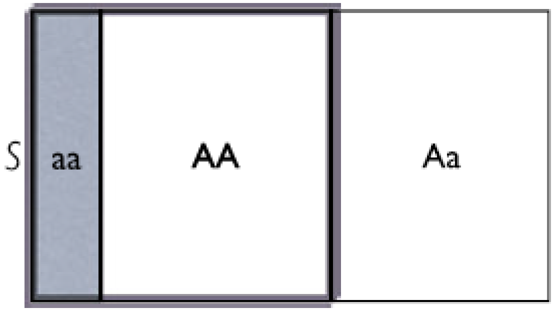
- Introduction
- The Law of Total Probability
- Independent & Dependent events
- Bayes' Theorem
- Concept of Random Variable (RV)
- Probability Distribution
- Discrete & Continuous RVs
- Discrete Probability Distribution
- PMF & CDF
- Binomial Distribution
- Poisson Distribution

Conditional Probability

The conditional probability of event E_1 given event E_2 can be calculated as follows:
(assuming $P(E_2) \neq 0$)

$$P(E_1|E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)}$$





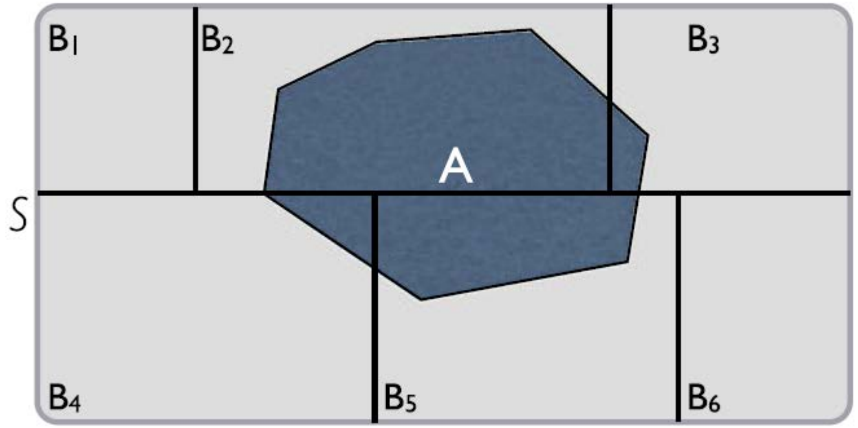
The Law of Total Probability-1

By rearranging $P(E_1|E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)}$ (i.e., moving $P(E_2)$ to the other side),

We obtain the following useful equation:

$$P(E_1 \cap E_2) = P(E_1|E_2)\, P(E_2)$$

Now suppose that a set of **K** events $\boldsymbol{B_1, B_2, \ldots, B_K}$ forms a partition of the sample space. (See Figure, where **K** = 6.)



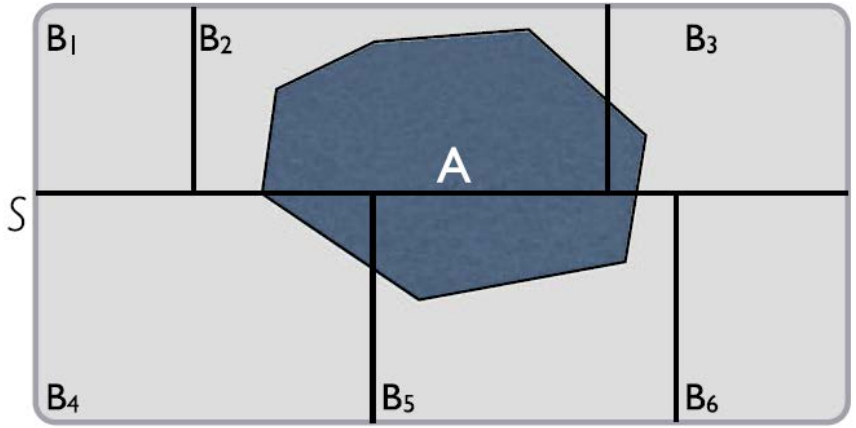
$$P(A \cap B_k) = P(A|B_k)\, P(B_k)$$

The Law of Total Probability-2

The above rule is known as the **law of total probability**, which can be written as

$$P(A) = \sum_{k=1}^{K} P(A|B_k)\, P(B_k)$$

where $\boldsymbol{B_1, B_2, \ldots, B_K}$ form a partition of the sample space, and $\boldsymbol{A}$ is an event in the sample space.



$$P(A \cap B_k) = P(A|B_k)\, P(B_k)$$

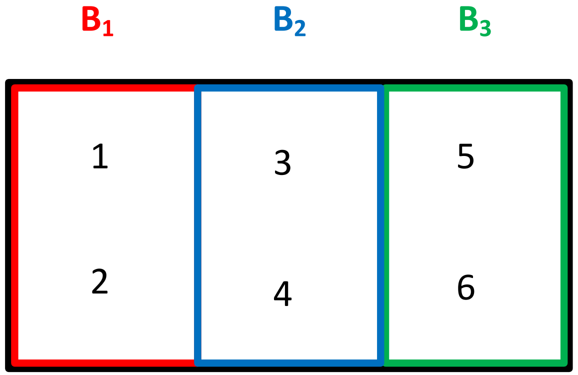
The Law of Total Probability: Example

For die rolling example, consider the three events

$\mathbf{B_1}$ = {1, 2},
$\mathbf{B_2}$ = {3, 4},
$\mathbf{B_3}$ = {5, 6},

whose probabilities are P($\mathbf{B_1}$) = P($\mathbf{B_2}$) = P($\mathbf{B_3}$) = 1/3.

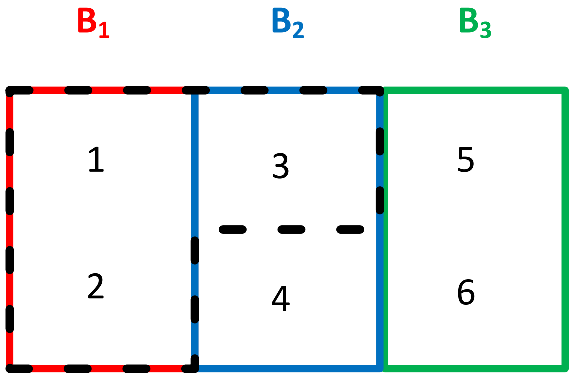
These events form a partition of the sample space.



Sample Space

The conditional probabilities of **M** (**<u>outcome less than four</u>**) given either of these three events are

$$P(\boldsymbol{M}|\boldsymbol{B_1}) = 1, P(\boldsymbol{M}|\boldsymbol{B_2}) = 1/2\,, P(\boldsymbol{M}|\boldsymbol{B_3}) = 0.$$



M

Using the law of total probability rule, we have

$$P(\boldsymbol{M}) = P(\boldsymbol{M}|\boldsymbol{B_1})P(\boldsymbol{B_1}) + P(\boldsymbol{M}|\boldsymbol{B_2})P(\boldsymbol{B_2}) + P(\boldsymbol{M}|\boldsymbol{B_3})P(\boldsymbol{B_3})$$

$$P(\boldsymbol{M}) = 1 * \frac{1}{3} + \frac{1}{2} * \frac{1}{3} + 0 * \frac{1}{3} = \frac{\mathbf{1}}{\mathbf{2}}$$

<u>which is the same as the probability we found directly based on the outcomes included in M.</u>

Independent Events-1

Two events E_1 and E_2 are **independent** if our knowledge of the occurrence of one event does **NOT** change the probability of occurrence of the other event.

That is, if E_1 and E_2 are **independent**,
then the conditional probability of E_1 given E_2 (i.e., probability of E_1 knowing E_2 has occurred)
is the **SAME** as the unconditional (or marginal) probability of E_1 (i.e., probability of E_1 regardless of E_2).

Therefore, for two independent events,

$$P(\boldsymbol{E_1}|\boldsymbol{E_2}) = P(\boldsymbol{E_1})$$

Example: Independent Events Suppose that we toss two dice simultaneously. Knowing that the outcome of one of them is less than 4 does not change the probability that the outcome of the other one is an odd number. In this case, we say that the two events, “less than 4” for one die and “odd number” for the other one, are independent.

Example: Dependent Events For our running example, where we are rolling one die only, the two events M and N are not independent. In this case, as we showed, knowing that the outcome is an odd number, i.e., event N has occurred, increases the probability of M from 1/2 to 2/3.

For the gene-disease example, we also showed that our knowledge that the genotype is homozygous increased the probability of the disease from P(D) = 0.09 (the unconditional probability) to P(D|HM) = 0.16 (the conditional probability). Therefore, the two events are dependent. This is of course consistent with our assumption that the disease is caused by gene A.

Independent Events-2

When two events $\boldsymbol{E_1}$ and $\boldsymbol{E_2}$ are **independent**, the probability that $\boldsymbol{E_1}$ and $\boldsymbol{E_2}$ occur simultaneously, i.e., their joint probability, is the product of their marginal probabilities:

$$P(\boldsymbol{E_1} \cap \boldsymbol{E_2}) = P(\boldsymbol{E_1}) \times P(\boldsymbol{E_2})$$

In general, if events E_1, E_2, …, E_n are **independent**, then

$$P(E_1 \cap E_2 \cap \cdots \cap E_n) = P(E_1) \times P(E_2) \times \cdots \times P(E_n)$$

Using the above rule along with $P(\boldsymbol{E_1} \cup \boldsymbol{E_2})$,
we obtain the probability of the union of two **independent** events as follows:

$$P(\boldsymbol{E_1} \cup \boldsymbol{E_2}) = P(\boldsymbol{E_1}) + P(\boldsymbol{E_2}) - P(\boldsymbol{E_1}) \times P(\boldsymbol{E_2})$$

The probability of the union $P(\boldsymbol{E_1} \cup \boldsymbol{E_2})$ is the sum of their marginal probabilities minus their joint probability.

$$P(\boldsymbol{E_1} \cup \boldsymbol{E_2}) = P(\boldsymbol{E_1}) + P(\boldsymbol{E_2}) - P(\boldsymbol{E_1} \cap \boldsymbol{E_2})$$

Bayes' Theorem

In some situations, we know the conditional probability of $\mathbf{E_1}$ given $\mathbf{E_2}$, but we are interested in the conditional probability of $\mathbf{E_2}$ given $\mathbf{E_1}$.

For example, suppose that the probability of having lung cancer is $P(\boldsymbol{C}) = 0.001$ and that the probability of being a smoker is $P(\boldsymbol{SM}) = 0.25$.
Further, suppose we know that if a person has lung cancer, the probability of being a smoker increases to $P(\boldsymbol{SM}|\boldsymbol{C}) = 0.40$.
We are, however, interested in the probability of developing lung cancer if a person is a smoker, $P(\boldsymbol{C}|\boldsymbol{SM})$.

$$P(E_1|E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)}$$

Using $P(E_1|E_2)$, this conditional probability is

$$P(\boldsymbol{C}|\boldsymbol{SM}) = \frac{P(\boldsymbol{SM} \cap \boldsymbol{C})}{P(\boldsymbol{SM})}$$

$$P(\boldsymbol{C}|\boldsymbol{SM}) = \frac{P(\boldsymbol{SM}|\boldsymbol{C})P(\boldsymbol{C})}{P(\boldsymbol{SM})}$$

$$P(\boldsymbol{C}|\boldsymbol{SM}) = \frac{0.40 \times 0.001}{0.25} = 0.0016$$

Therefore, the probability of lung cancer $P(\boldsymbol{C}) = 0.001$ increases to 0.0016 for smokers. <u>That is, the probability becomes 60% higher than the overall probability of lung cancer.</u>

Bayes’ Theorem

In general, for two events $\mathbf{E_1}$ and $\mathbf{E_2}$, the following equation shows the relationship between $P(\boldsymbol{E_2}|\boldsymbol{E_1})$ *and* $P(\boldsymbol{E_1}|\boldsymbol{E_2})$:

$$P(\boldsymbol{E_2}|\boldsymbol{E_1}) = \frac{P(\boldsymbol{E_1}|\boldsymbol{E_2})P(\boldsymbol{E_2})}{P(\mathbf{E_1})}$$

This formula is known as **<u>Bayes’ theorem</u>** or **<u>Bayes’ rule</u>**.

Bayes' Theorem & Inference

Bayes' Rule is a systematic approach for incorporating new evidence.

- ✓ Initial beliefs $P(\boldsymbol{B_i})$ on possible causes of an observed event **M**
- ✓ Model of the world under each $\boldsymbol{B_i}$: $\boldsymbol{P(M|B_i)}$

$$\boldsymbol{B_i} \xrightarrow[\boldsymbol{P(M|B_i)}]{\textbf{model}} \mathbf{M}$$

- ✓ Draw conclusions about causes

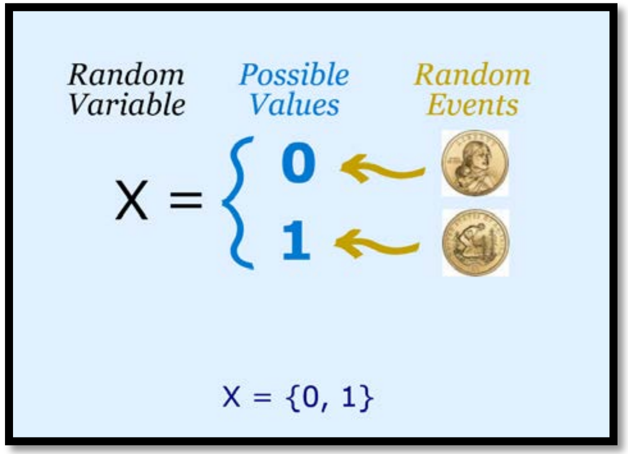
$$\mathbf{M} \xrightarrow[\boldsymbol{P(B_i|M)}]{\textbf{inference}} \boldsymbol{B_i}$$

Concept of Random Variables

The concept of a random variable provides to evaluate different types of events having the *same probabilistic structure* with the same perspective.

A **random variable** is a function that assigns numeric values to different events in a sample space.

Flip a coin example:



Genotypes example: we can define ***X*** based on possible genotypes of a bi-allelic gene **A** as follows:

$$X = \begin{cases} 0 & \text{for genotype } AA, \\ 1 & \text{for genotype } Aa, \\ 2 & \text{for genotype } aa. \end{cases}$$

Probability Distribution

The **probability distribution** of a random variable specifies its possible values (i.e., its range) and their corresponding probabilities.

Genotypes example: we can define ***X*** based on possible genotypes of a bi-allelic gene **A** as follows:

$$X = \begin{cases} 0 & \text{for genotype } AA, \\ 1 & \text{for genotype } Aa, \\ 2 & \text{for genotype } aa. \end{cases}$$

For the random variable ***X*** defined based on genotypes, the probability distribution can be simply specified as follows:

$$P(X = x) = \begin{cases} 0.49 & \text{for } x = 0, \\ 0.42 & \text{for } x = 1, \\ 0.09 & \text{for } x = 2. \end{cases}$$

Discrete VS. Continuous

The grouping of random variables into **discrete** and **continuous** is based on their range.

Discrete random variables can take a countable set of values.

These variables can be categorical (nominal or ordinal), such as genotype, gender, disease status, or pain level.

They can also be counts, such as the number of patients visiting an emergency room per day, or the number of lymph nodes containing evidence of cancer.
For all these examples, we can count the number of possible values the random variable can take. In the above genotype examples, X is a discrete random variable since it can take 3 possible values only.

Continuous random variables can take an uncountable number of possible values.

Examples include weight, body temperature, BMI, and blood pressure.

Consider the random variable Y for birthweight, which is a random phenomenon. In this case, the values of the random variables (i.e., numbers it assigns to each possible outcome) are the same as the corresponding outcomes; Y = 3.54 kg. if the birthweight is 3.54 kg. In this example, we cannot count the possible values of Y. For any two possible values of this random variable, we can always find another value between them.

Discrete Probability Distribution
Probability Mass Functions (PMF)

The values taken by a discrete random variable and its associated probabilities can be expressed by a rule or relationship called a ***probability-mass function*** **(pmf)**.

A **probability-mass function** is a mathematical relationship, or rule, that assigns to any possible value r of a discrete random variable X the probability $Pr(X = r)$. This assignment is made for all values r that have positive probability. The probability-mass function is sometimes also called a **probability distribution.**

Example:

Suppose ***Y*** is a random variable that is equal to **1** when a newborn baby has low birthweight, and is equal to **0** otherwise.

We say ***Y is a binary random variable.***

Further, assume that the probability of having a low birthweight for babies is 0.3.

Then the **PMF** for the random variable ***Y*** is

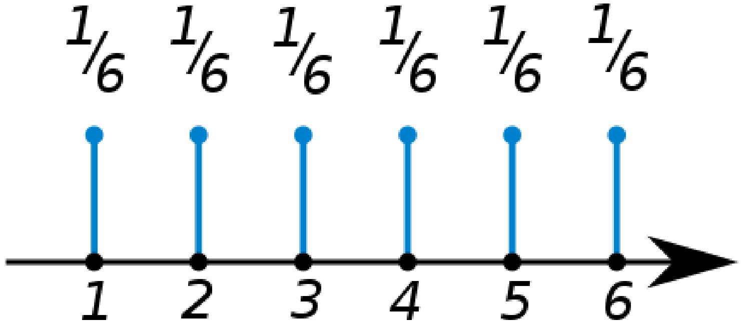
$$P(Y = y) = \begin{cases} 0.7 & \text{for } y = 0, \\ 0.3 & \text{for } y = 1. \end{cases}$$

Probability Mass Functions (PMF)

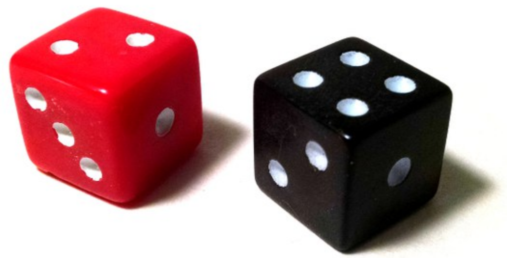
For a discrete random variable $\boldsymbol{X}$ with possible values $\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n}$, a **probability mass function** is a function such that

1. $f(x_i) \geq 0$
2. $\sum_{i=1}^{n} f(x_i) = 1$
3. $f(x_i) = P(X = x_i)$

Roll a dice



Sum of rolling two dice



The Expected Value of a Discrete RV

The **expected value of a discrete random variable** is defined as

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

where the x_i's are the values the random variable assumes with positive probability.

<u>Example</u> Hypertension

Find the expected value for the random variable shown in Table.

$Pr(X = r)$	.008	.076	.265	.411	.240
r	0	1	2	3	4

Solution: $E(X) = 0(.008) + 1(.076) + 2(.265) + 3(.411) + 4(.240) = 2.80 = \mu$

Thus, on average about 2.8 hypertensives would be expected to be brought under control for every 4 who are treated.

The **<u>mean</u>** of a random variable is also called its ***<u>expected value</u>*** even.

The Variance of a Discrete RV

The **variance of a discrete random variable,** denoted by $Var(X)$, is defined by

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

where the x_i's are the values for which the random variable takes on positive probability. The **standard deviation of a random variable** X, denoted by $sd(X)$ or σ, is defined by the square root of its variance.

Note: The standard deviation of a random variable X, denoted by $sd(X)$ or σ, is defined by the square root of its variance.

The population variance can also be expressed in a different (**"short"**) form as follows:

A **short form for the population variance** is given by

$$\sigma^2 = E(X - \mu)^2 = \sum_{i=1}^{R} x_i^2 Pr(X = x_i) - \mu^2$$

Cumulative Distribution Functions (CDF)

The basic idea is to assign to each individual value the sum of probabilities of all values that are no larger than the value being considered.
This function is defined as follows:

The **cumulative-distribution function (cdf)** of a random variable X is denoted by $F(X)$ and, for a specific value x of X, is defined by $Pr(X \leq x)$ and denoted by $F(x)$.

Example **Otolaryngology**
Compute the **CDF** for the otitis-media random variable in Table and display it graphically.

Probability-mass function for the number of episodes of otitis media in the first 2 years of life

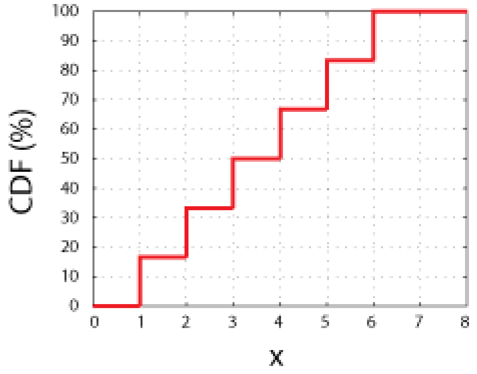
r	0	1	2	3	4	5	6
$Pr(X = r)$	.129	.264	.271	.185	.095	.039	.017

Cumulative Distribution Functions (CDF)

The **cumulative distribution function** $F(x)$ of a discrete random variable X with probability distribution $f(x)$ is

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$$

Roll a dice



CDF for a fair 6-sidded dice.
Note that each step is a height of $\frac{1}{6}$ or 16.67 %.

There are many problems where we may wish to compute the probability that the observed value of a random variable *X* will be less than or equal to some real number *x*.

The Binomial Distribution

All examples involving the binomial distribution have a common structure:

- ✓ A sample of n independent trials, each of which can have only two possible outcomes, which are denoted as "***<u>success</u>***" and "***<u>failure</u>***."
- ✓ The probability of a success at each trial is assumed to be some constant $\boldsymbol{p}$, and hence, the probability of a failure at each trial is $\mathbf{1 - \boldsymbol{p} = \boldsymbol{q}}$.
- ✓ The term "success" is used in a general way, without any specific contextual meaning.

The distribution of the number of successes in n statistically independent trials, where the probability of success on each trial is p, is known as the **binomial distribution** and has a probability-mass function given by

$$Pr(X = k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, \ldots, n$$

The Binomial Distribution

<u>Example</u> Infectious Disease

One of the most common laboratory tests performed on any routine medical examination is a blood count.

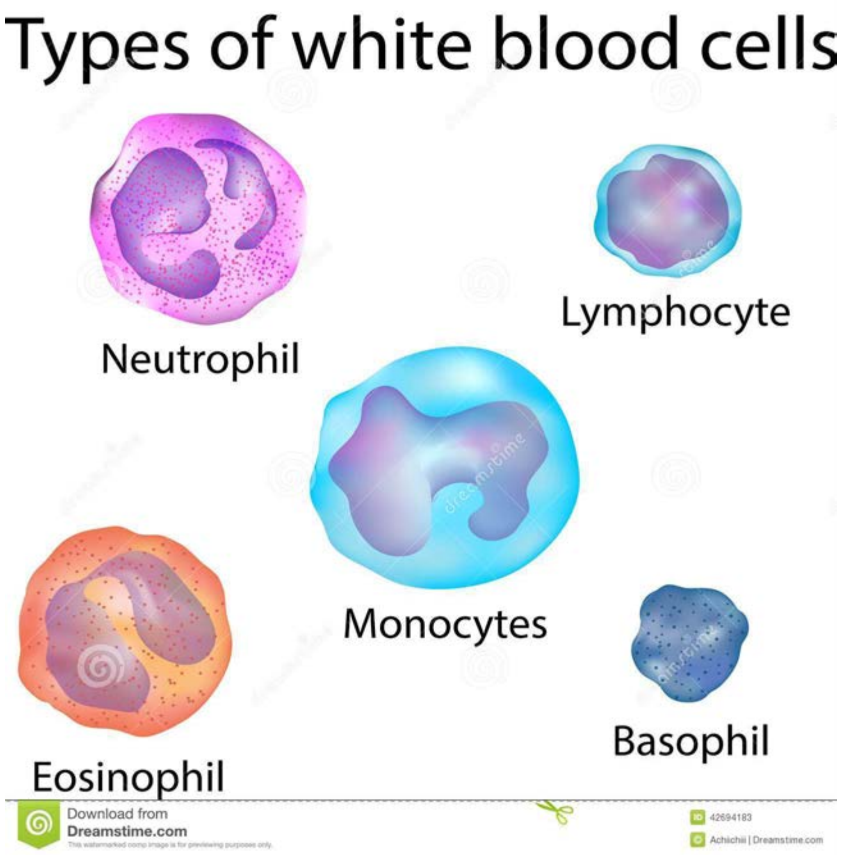
The two main aspects of a blood count are

(1) counting the number of white blood cells (the "white count") and

(2) differentiating the white blood cells that do exist into five categories—namely, neutrophils, lymphocytes, monocytes, eosinophils, and basophils (called the "differential").

Both the white count and the differential are used extensively in making clinical diagnoses. We concentrate here on the differential, particularly on the distribution of the number of neutrophils k out of 100 white blood cells (which is the typical number counted).

We will see that the number of neutrophils follows a *binomial distribution*.

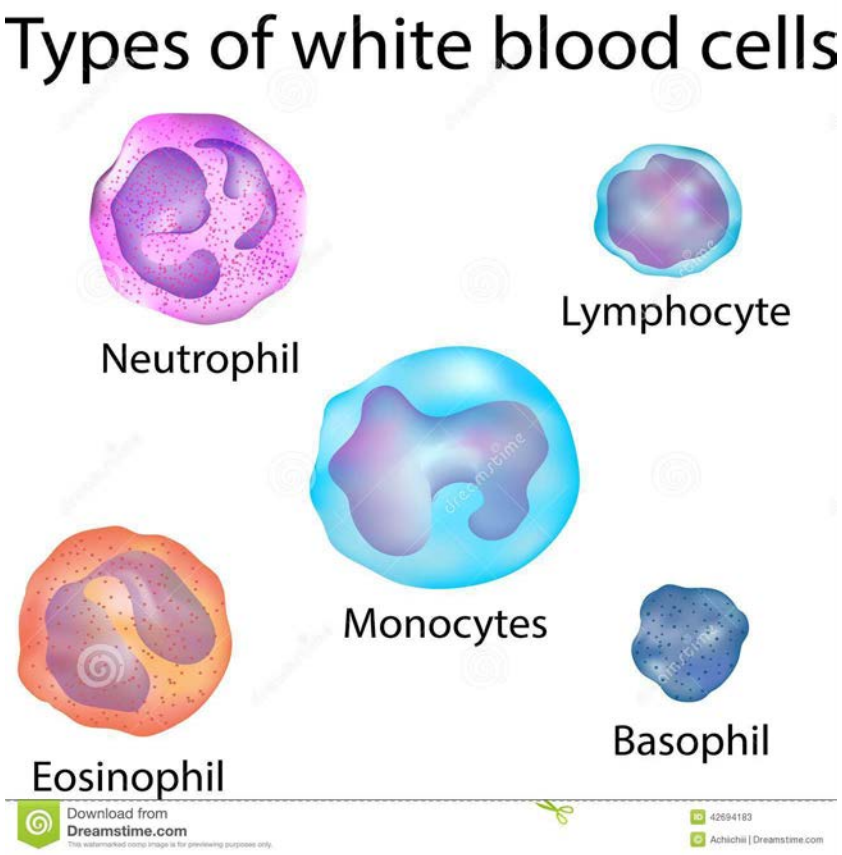


The Binomial Distribution

<u>Example</u> Infectious Disease

Consider with 5 cells rather than 100, and ask the more limited question:

1. What is the probability that the second and fifth cells considered will be neutrophils and the remaining cells non-neutrophils, given a probability of .6 that any one cell is a neutrophil?
2. Now consider the more general question: What is the probability that any 2 cells out of 5 will be neutrophils?



Expected Value and Variance of the Binomial Distribution

$$E(X) = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k} \longrightarrow \boldsymbol{np}$$

$$Var(X) = \sum_{k=0}^{n} (k - np)^2 \binom{n}{k} p^k q^{n-k} \longrightarrow \boldsymbol{npq}$$

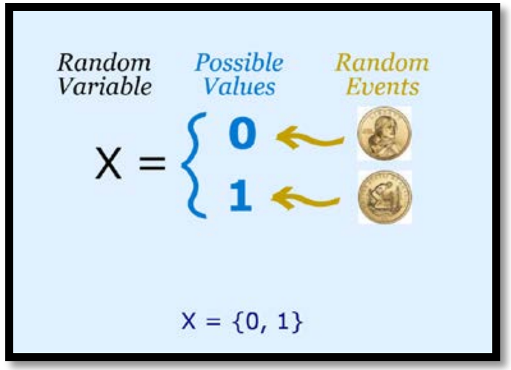
The **expected value** and the **variance of a binomial distribution** are np and npq, respectively.

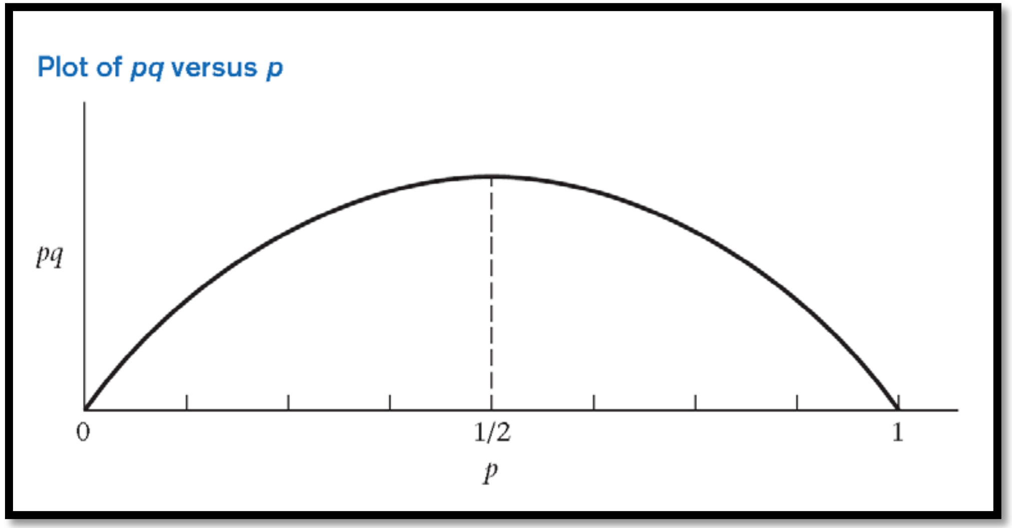
Expected Value and Variance of the Binomial Distribution

$$E(X) = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k} \longrightarrow \boldsymbol{np}$$

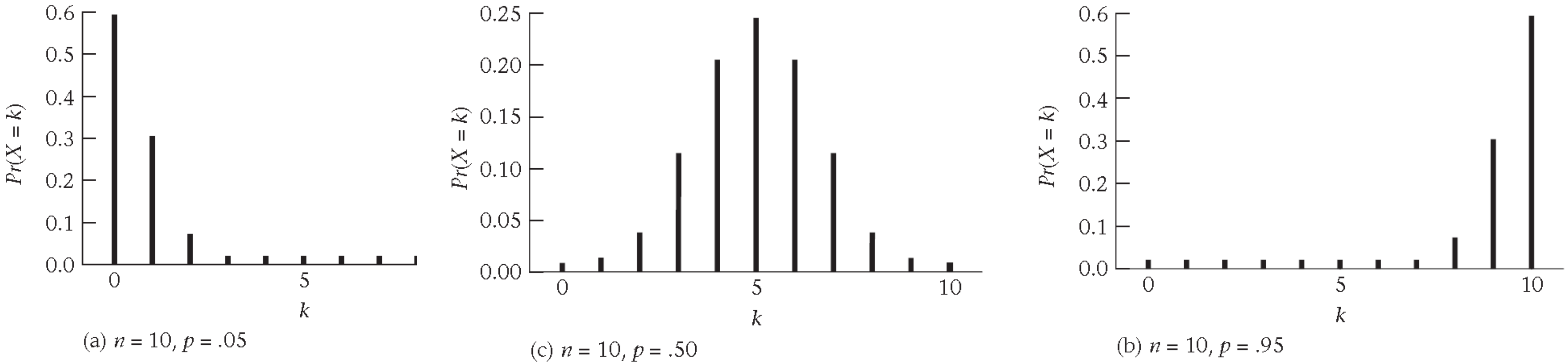
$$Var(X) = \sum_{k=0}^{n} (k - np)^2 \binom{n}{k} p^k q^{n-k} \longrightarrow \boldsymbol{npq}$$

For a given number of trials n, the binomial distribution has the highest variance when p = 1/2, as shown in Figure.





The binomial distribution for various values of p when n = 10



When p is near 0 or near 1, the distribution of the number of successes is clustered near 0 and n, respectively, and there is comparatively little variability as compared with the situation when $p = 1/2$.
This point is illustrated in Figure.

The Poisson Distribution

The Poisson distribution is perhaps the second most frequently used discrete distribution after the binomial distribution.
This distribution is usually associated with ***rare events***.

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

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

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

The Poisson distribution depends on a single parameter $\mu = t$.
The parameter $\boldsymbol{\lambda}$ represents *the expected number of events per unit time*,
whereas the parameter $\boldsymbol{\mu}$ represents *the expected number of events over time period t*.

The Poisson Distribution

The Poisson distribution is perhaps the second most frequently used discrete distribution after the binomial distribution.
This distribution is usually associated with ***rare events***.

Example Infectious Disease
Suppose the number of deaths from typhoid fever over a 1-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? A 3-month period?

The Poisson Distribution

Example Infectious Disease

Suppose the number of deaths from typhoid fever over a 1-year period is Poisson distributed with parameter μ = 4.6.

What is the probability distribution of the number of deaths over a 6-month period? A 3-month period?

Solution: Let X = the number of deaths in 6 months. Because $\mu = 4.6$, $t = 1$ year, it follows that $\lambda = 4.6$ deaths per year. For a 6-month period we have $\lambda = 4.6$ deaths per year, $t = .5$ year. Thus, $\mu = \lambda t = 2.3$. Therefore,

$$Pr(X=0) = e^{-2.3} = .100$$

$$Pr(X=1) = \frac{2.3}{1!}e^{-2.3} = .231$$

$$Pr(X=2) = \frac{2.3^2}{2!}e^{-2.3} = .265$$

$$Pr(X=3) = \frac{2.3^3}{3!}e^{-2.3} = .203$$

$$Pr(X=4) = \frac{2.3^4}{4!}e^{-2.3} = .117$$

$$Pr(X=5) = \frac{2.3^5}{5!}e^{-2.3} = .054$$

$$Pr(X \ge 6) = 1-(.100+.231+.265+.203+.117+.054) = .030$$

Let Y = the number of deaths in 3 months. For a 3-month period, we have $\lambda = 4.6$ deaths per year, $t = .25$ year, $\mu = \lambda t = 1.15$. Therefore,

$$Pr(Y=0) = e^{-1.15} = .317$$

$$Pr(Y=1) = \frac{1.15}{1!}e^{-1.15} = .364$$

$$Pr(Y=2) = \frac{1.15^2}{2!}e^{-1.15} = .209$$

$$Pr(Y=3) = \frac{1.15^3}{3!}e^{-1.15} = .080$$

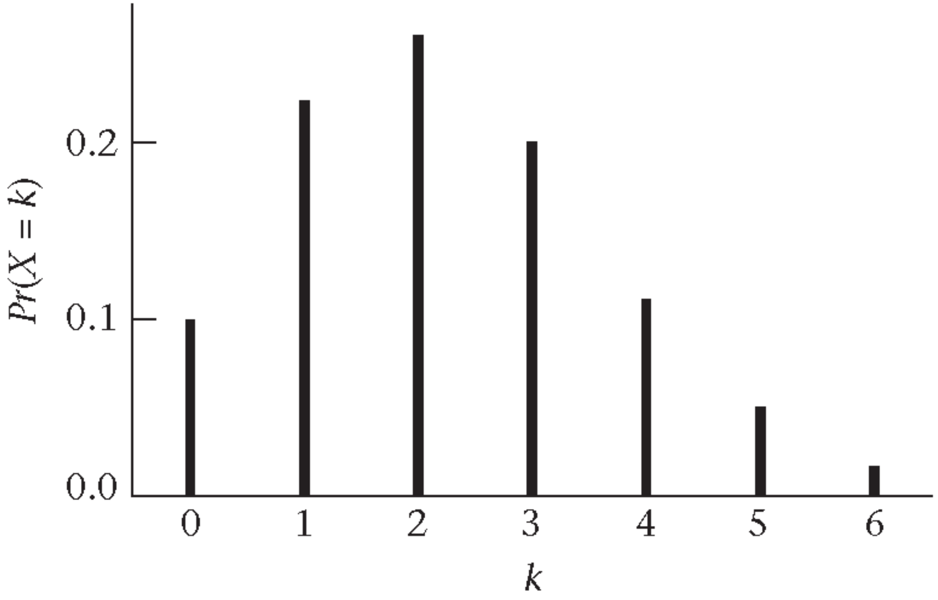
$$Pr(Y \ge 4) = 1-(.317+.364+.209+.080) = .030$$

The Poisson Distribution

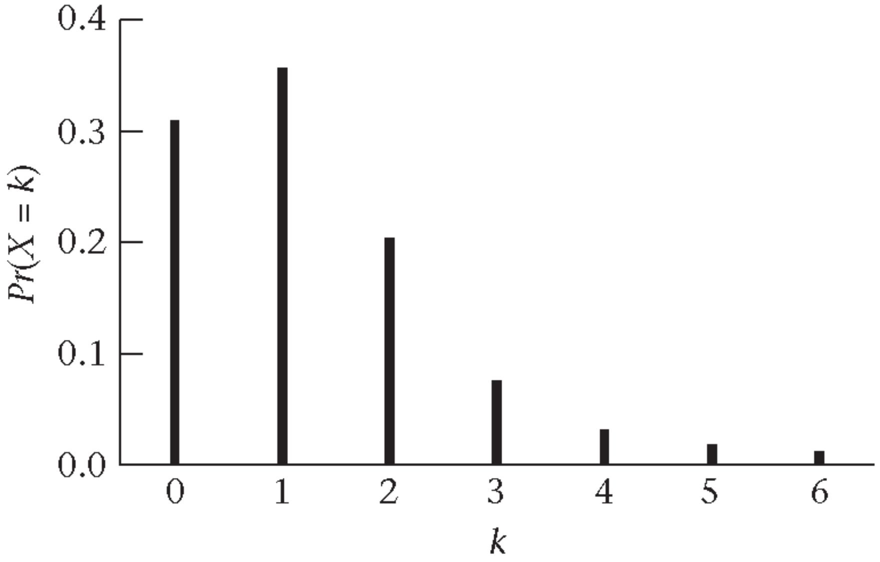
<u>Example</u> Infectious Disease

Suppose the number of deaths from typhoid fever over a 1-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? A 3-month period?



(a) 6 months



(b) 3 months

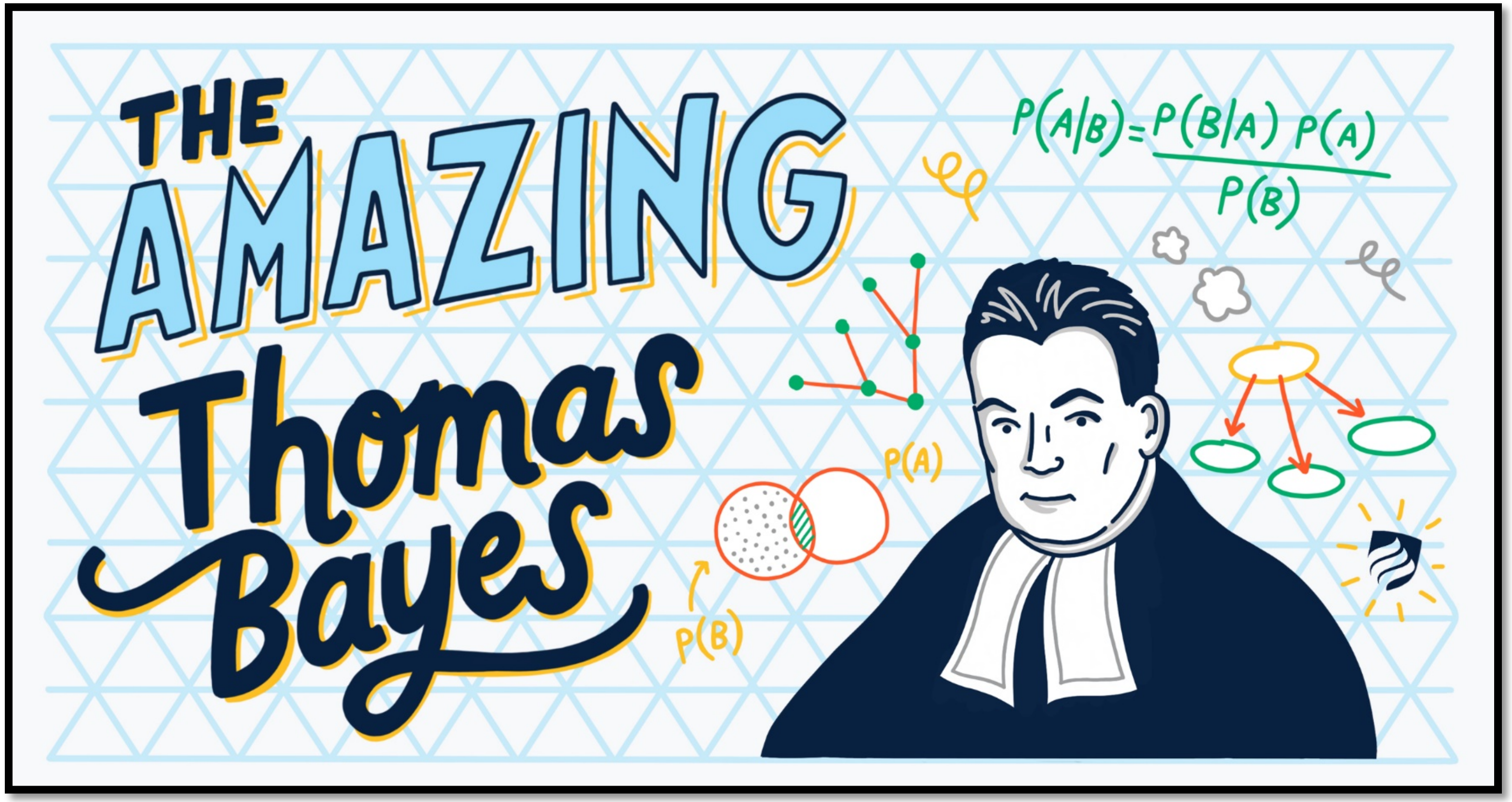
Note that the distribution tends to become more symmetric as the time interval increases or, more specifically, as μ increases.

Expected Value and Variance of the Poisson Distribution

For a Poisson distribution with parameter μ, the mean and variance are both equal to μ.

This fact is useful, because if we have a data set from a discrete distribution where the *mean and variance are about the same*, then we can preliminarily identify it as a Poisson distribution and use various tests to confirm this hypothesis.

Questions?



Thomas Bayes *(1701 – 7 April 1761) was an English statistician, philosopher and Presbyterian minister who is known for formulating a specific case of the theorem that bears his name: [Bayes' theorem](#).*