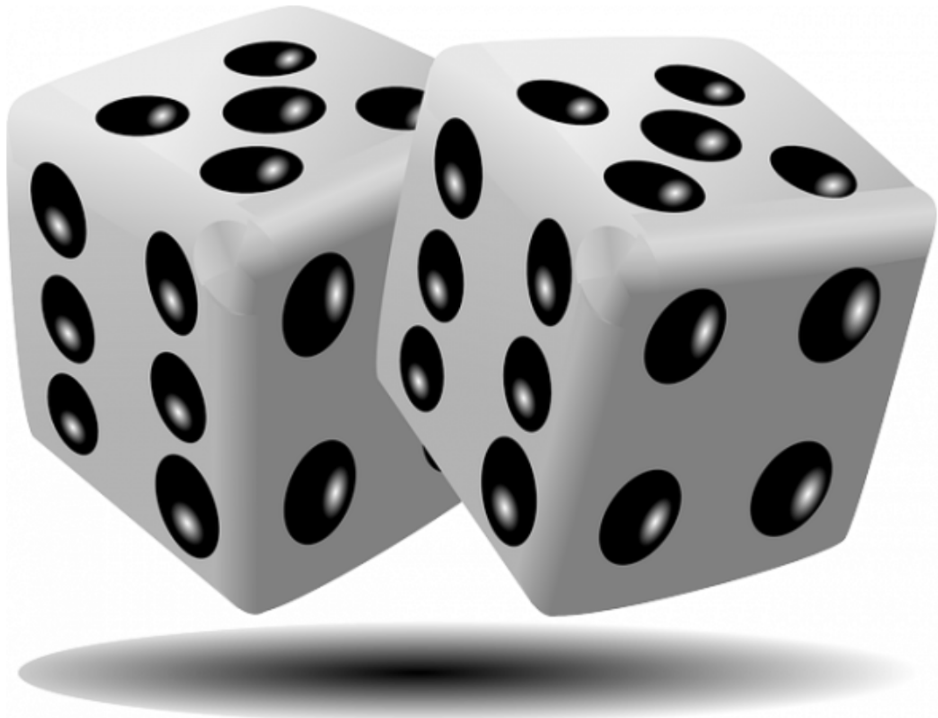
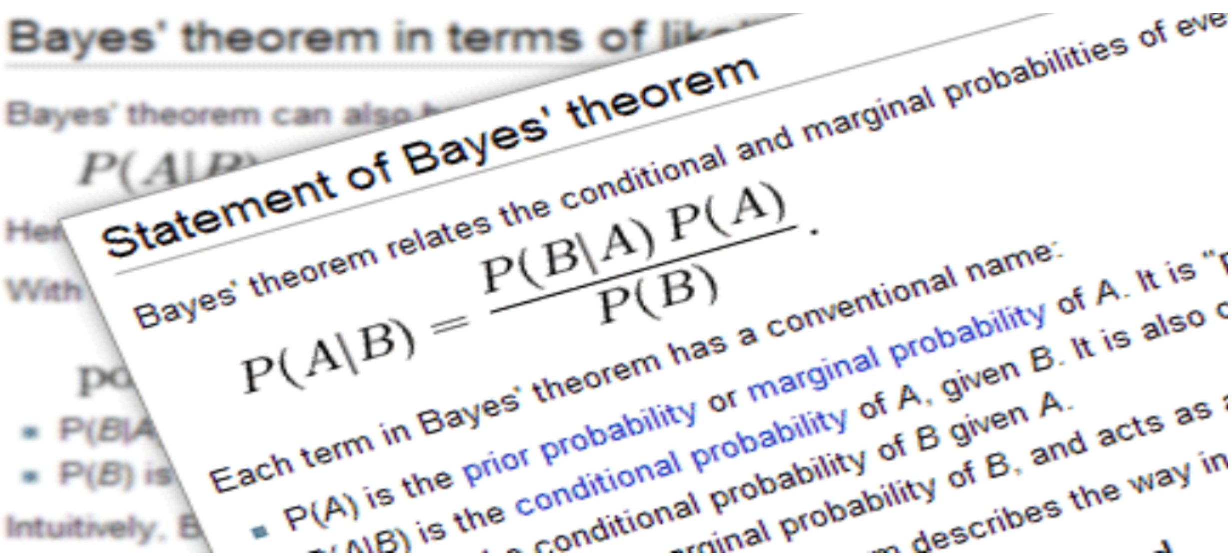




STAT 4444/5444G

Applied Bayesian Statistics

Unit 2

Review of Probability

Hamdy F. F. Mahmoud, PhD
Collegiate Assistant Professor
Statistics Department @ VT
Ehamdy@vt.edu

Spring 2021

This unit covers

- Events and Sample Space
- Unions, Intersections, Complements
- The Addition Rule
- Marginal and Conditional Probabilities
- The Multiplication Rule
- The Law of Total Probability
- Bayes' Rule in the Discrete Case
- Random Variables and Probability Distributions
- What is to come?

❑ Events and Sample Space

Statisticians and probabilists use the term **event** to refer to any outcome or set of outcomes of a random phenomenon.

For example, you randomly select a patient at random from a huge database of patients.

- The phenomenon here is drawing a patient.
- Random means each person in the population has the same probability to be selected.

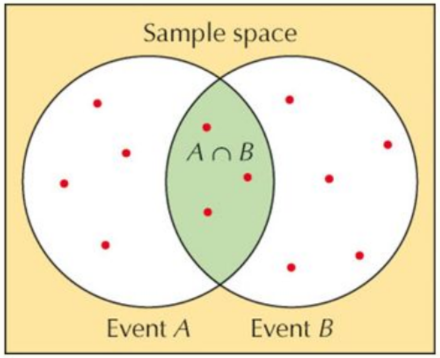
We can define infinity number of evens to this experiment. For example, let event A that the patient we draw is under 6 years of age. As in Unit 1, we will denote the probability of event A as P(A).

The sample space is the set of all possible outcomes of a random phenomenon. A commonly used symbol for it is S. If a random phenomenon occurs, one of the outcomes in S has to happen. Thus P(S)=1.

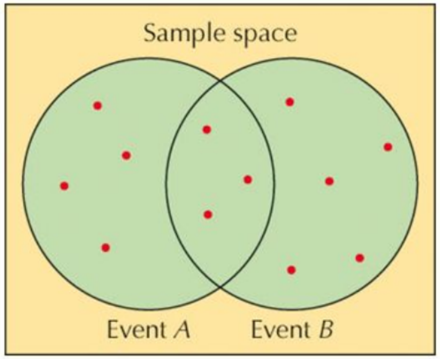
Note: this is an example of a continuous random variable.

❑ Unions, Intersections, Complements

The intersection of two events A and B is the event "both A and B," which is represented by $A \cap B$. For example, if event B is the event that the patient we draw weighs at least 150 pounds, then $A \cap B$ is the event that the patient we draw is under 6 years of age and weighs at least 150 pounds

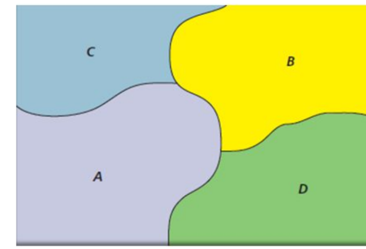


The union of two events A and B is the event "either A or B or both," or, in symbols, $A \cup B$. In our example with A and B defined as above, $A \cup B$ is the event that the patient we draw either is under 6 years of age, or weighs at least 150 pounds, or both.

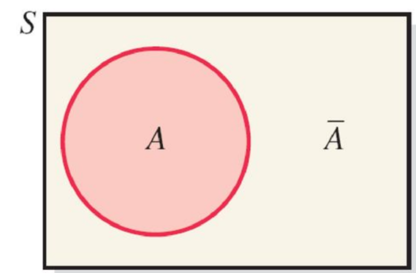


❑ Unions, Intersections, Complements

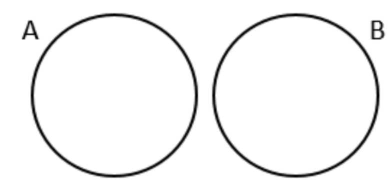
A set of events is said to be **exhaustive** if, taken together, the events encompass the entire sample space. A set of events A1,A2,A3,... is exhaustive if
$A1 \cup A2 \cup A3 \cup ... = S$.



The complement of an event A is the event "everything else that could possibly happen except A," notated A^C or $\bar{A}$. Clearly, $A \cup \bar{A} = S$.



The null event, represented by the symbol ∅, is an event that can never happen. Two events A and B are **disjoint** or equivalently mutually exclusive if they cannot occur together. If A and B are **mutually exclusive**, then $A \cap B = \emptyset$.



P(A or B) = P(A) + P(B)

❑ The Addition Rule

The addition rule of probability states that if two events A and B are mutually exclusive, then the probability that one or the other happens is just the sum of the probabilities of each event individually:

$$P(A \cup B) = P(A) + P(B)$$

If two events A and B are not mutually exclusive, then the union probability is

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Since any event A and its complement $\bar{A}$ are mutually exclusive by definition, and $A \cup \bar{A}$ is the sample space S, one implication of the addition rule is $P(\bar{A}) = 1 - P(A)$.

Practice

In the experiment of rolling a single die, find the probabilities of the following events:

- A: the number rolled is even
- B: the number rolled is odd
- C: the number rolled is greater than two
- $A \cap B$

- $A \cup B$

- $A \cap C$

- $A \cup C$

Marginal and Conditional Probabilities

For any two events A and B, P(A) and P(B) are called marginal probabilities and P(B|A) is called the conditional probability of B given A. P(B|A) is the probability that event B will occur given that we already know that event A has occurred.

If A and B are **dependent**, the intersection probability is

$$P(A\cap B)=P(A)P(B|A) \quad \text{and} \quad P(A\cap B)=P(B)P(A|B)$$

So,

$$P(B|A)=\frac{P(A\cap B)}{P(A)}, \quad P(A)>0$$

Practice

Data are collected and summarized in the table below

	< 150 pounds	≥ 150 pounds	Total
under 6	798	2	800
≥ 6	4,702	4,498	9,200
Total	5,500	4,500	10,000

If one patient is randomly selected, what is

a. the probability she weighs less than 150 pounds?

b. If a patient is randomly selected and found that she is less than 6 years, what is the probability that she is less than 150 pounds?

c. If a patient is randomly selected and found that she is older than 6 years, what is the probability that she is less than 150 pounds?

- What is your conclusion?

❑ The Multiplication Rule

Two events are independent if the occurrence (or nonoccurrence) of one of them does not affect the probability that the other one occurs. That is, events A and B are independent if

$$P(A|B)=P(A) \text{ and } P(B|A)=P(B)$$

There is a special form of the multiplicative rule of probability for independent events. If A and B are independent, then

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

Practice

Data are collected about weight and eye color from 10,000 of people and summarized below.

	< 150 pounds	≥ 150 pounds	Total
Green eyes	440	360	800
Not green	5,060	4,140	9,200
Total	5,500	4,500	10,000

Assume a person is randomly selected. Find the probabilities of the following events:

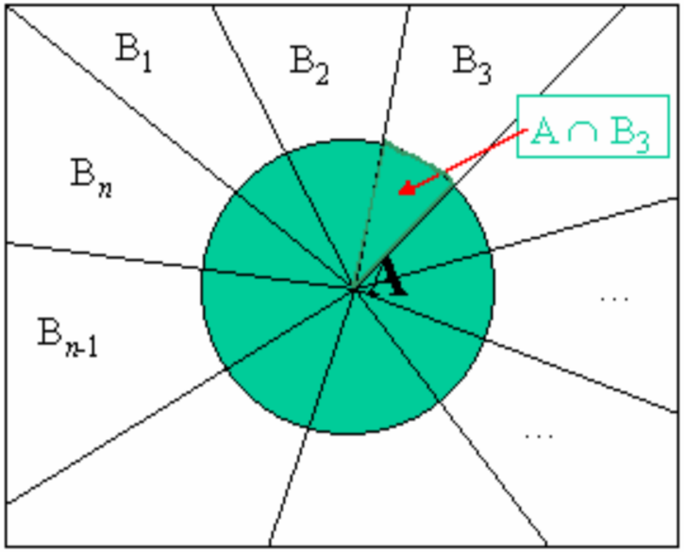
- A: has weight greater than 150 pounds

- B: has green eyes.

➢ Show that these two events are independent.

❑ The Law of Total Probability

The law of total probability comes into play when you wish to know the marginal (unconditional) probability of some event, **A**, but you only know its probability under some conditions, $B_1, B_2, \ldots\ldots, B_n$. If $B_1, B_2, \ldots\ldots, B_n$ are mutually exclusive and exhaustive events, then



$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \ldots\ldots + P(A \cap B_n)$$
$$= P(A|B_1)\,P(B_1) + P(A|B_2)\,P(B_2) + \ldots\ldots\ldots + P(A|B_n)P(B_n)$$

❑ Bayes' Rule in the Discrete Case

Bayes' rule holds if event A could happen conditional on a number of different other events, $B_1, B_2, \ldots\ldots, B_n$. To apply Bayes' rule, we must know the conditional probabilities $P(A|B_1), P(A|B_2)$, ….., $P(A|B_n)$, as well as the marginal probabilities $P(B_1), P(B_2)$, ….., $P(B_n)$.

- After the event A has occurred, we want to assess the conditional probability of one of the events B_j, $P(B_j|A)$. If $B_1, B_2, \ldots\ldots, B_n$ are mutually exclusive and exhaustive events, then

$$P(B_j|A) = \frac{P(B_j \cap A)}{P(A)} = \frac{P(A|B_j)P(B_j)}{P(A|B_1)\ P(B_1) + P(A|B_2)\ P(B_2) + \ldots\ldots + P(A|B_n)P(B_n)}$$

Industry application

The entire output of a factory is produced on three machines. The three machines account for 20%, 30%, and 50% of the factory output. The fraction of defective items produced is 5% for the first machine; 3% for the second machine; and 1% for the third machine.

- If an item is chosen at random from the total output, what is the probability that it was defective?

If an item is chosen at random from the total output and is found to be defective, what is the probability that it was produced by the 3rd machine?

Biology application

An entomologist spots what might be a rare subspecies of beetle, due to the pattern on its back. In the rare subspecies, 98% have the pattern, or $P(\text{Pattern} \mid \text{Rare}) = 98\%$. In the common subspecies, 5% have the pattern. The rare subspecies accounts for only 0.1% of the population. How likely is the beetle having the pattern to be rare, or what is $P(\text{Rare} \mid \text{Pattern})$?

❑ Random Variables and Probability Distributions

A random variable may be defined as a function that assigns one real number to each outcome in the sample space of a random phenomenon.

For example, drawing a household at random from among all the households in Blacksburg, Virginia, and recording the number of people living in the household. The sample space of this random experiment is numeric, consisting of the integers X=1, 2, ...

In this example, X is ***discrete random variable*** because it has a discrete set of possible values that it can take on.

❑ Random Variables and Probability Distributions

By contrast, *a continuous random variable* is not restricted to a discrete set of possible values, but instead may take on any value in a continuum.

- For example, we might define the random variable Y as the height of a woman drawn at random from among female first year students at VT.

- Why Y is a *continuous random variable?*

❑ Random Variables and Probability Distributions

- **Probability distributions describe the behavior of random variables.**

- The probability distribution of a discrete random variable identifies the possible values the variable can take on and associates a numeric probability with each.

- Each of these probabilities must be between 0 and 1, and the probabilities of all the possible values in the space of the random variable must sum to 1.

- Presenting a probability distribution can be by a ***table*** or a ***function***.

Random Variables and Probability Distributions

- Flipping a ***fair*** coin one time can be presented by a table

x	Pr(X=x)
0	0.5
1	0.5

Or by Bernoulli probability mass function (*pmf*)

$$P(X = x) = p^x(1 - p)^{1-x}, \quad x = 0,1$$

- The height of a woman drawn at random from among female first year students at VT cab be represented by the probability density function of normal distribution.

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{1}{-2\sigma^2}(x-\mu)^2},$$

where $\quad -\infty < x < \infty, \quad -\infty < \mu < \infty, \quad \sigma > 0$

Probability distributions

Table A.1 Discrete distributions

Distribution	Probability mass function	Mean	Mode	Variance
Binomial	$p(y\|\pi) = \binom{n}{y} \pi^y (1-\pi)^{n-y}$	$n\pi$	$\lfloor (n+1)\pi \rfloor$	$n\pi(1-\pi)$
$Y \sim Bin(n, \pi)$	$y = 0, 1, \ldots, n$			
$0 < \pi < 1$				
Poisson	$p(y\|\lambda) = \frac{\exp(-\lambda)\lambda^y}{y!}$	λ	$\lfloor \lambda \rfloor$	λ
$Y \sim Pois(\lambda)$	$y = 0, 1, \ldots$			
$\lambda > 0$				

Table A.2 Univariate continuous distributions

Distribution	Density	Mean Mode	Variance
Beta	$p(y\|\alpha,\beta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}y^{\alpha-1}(1-y)^{\beta-1}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
$Y\sim Beta(\alpha,\beta)$	$0<y<1$	$\frac{\alpha-1}{\alpha+\beta-2}$	
$\alpha,\beta>0$			
Gamma	$p(y\|\alpha,\beta)=\frac{\beta^{\alpha}}{\Gamma\alpha}y^{\alpha-1}\exp(-\beta y)$	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$
$Y\sim Gamma(\alpha,\beta)$	$0<y<\infty$	$\frac{\alpha-1}{\beta},\ \alpha\geq 1$	
$\alpha,\beta>0$			
Inverse gamma	$p(y\|\alpha,\beta)=\frac{\beta^{\alpha}}{\Gamma(\alpha)}\frac{1}{y^{\alpha+1}}\exp(-\frac{\beta}{y})$	$\frac{\beta}{\alpha-1},\ \alpha>1$	$\frac{\beta^2}{(\alpha-1)^2(\alpha-2)},\ \alpha>2$
$Y\sim IG(\alpha,\beta)$	$0<y<\infty$	$\frac{\beta}{\alpha+1}$	
$\alpha,\beta>0$			
Normal	$p(y\|\mu,\sigma^2)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\frac{(y-\mu)^2}{2\sigma^2})$	μ	σ^2
$Y\sim N(\mu,\sigma^2)$	$-\infty<y<\infty$	μ	
$\sigma^2>0$			
Student's t	$p(y\|\mu,\sigma^2,\nu)=\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\nu\pi\sigma^2}}\frac{1}{(1+\frac{1}{\nu}(\frac{y-\mu}{\sigma})^2)^{\frac{\nu+1}{2}}}$	$\mu,\ \nu>1$	$\frac{\nu}{\nu-2}\sigma^2,\ \nu>2$
$Y\sim t(\mu,\sigma^2,\nu)$	$-\infty<y<\infty$	μ	
$\sigma^2>0,\nu\geq 1$			
Uniform	$p(y\|a,b)=\frac{1}{b-a}$	$\frac{b-a}{2}$	$\frac{(b-a)^2}{12}$
$Y\sim U(a,b)$	$a\leq y\leq b$	none	

Table A.3 Multivariate continuous distributions

Distribution	Density	Mean Mode	Variance
Normal	$p(\mathbf{y}\vert\mu,\Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}}\vert\Sigma\vert^{\frac{1}{2}}}\exp(-\frac{(\mathbf{y}-\mu)^T\Sigma^{-1}(\mathbf{y}-\mu)}{2})$	μ	Σ
$\mathbf{Y} \sim N_d(\mu,\Sigma)$	$\mathbf{y}$ a vector of length d	μ	
Σ pos def symm matrix			
Wishart	$p(T\vert S,\nu) \propto \vert T\vert^{\vert}S\vert^{\frac{\nu}{2}}\,\frac{\nu-d-1}{2}\exp(-\frac{tr(ST)}{2})$	$E(T_{ij}) = \nu S_{ij}^{-1}$	
$T \sim W_d(S,\nu)$	T pos def symm matrix		
S pos def symm matrix			

❑ What is to come?

In this unit, basic concepts of probability are reviewed. In the remainder chapters, we will build on these foundations in order to understand Bayesian modeling and inference for increasingly realistic problems.