Statistics Workshop Part I

Probability, Counting Techniques, Bayes' Theorem

Silvia Heubach

Department of Mathematics

California State University Los Angeles

Basic Probability

- **experiment** with outcome not known beforehand
- **S** = **sample space** = set of possible outcomes
- **A** $\subseteq S$ = **event**

Example 1: Two dice are rolled.

- Sample space $S =$
$\{(1,1),(1,2),\ldots,(1,6),(2,1),\ldots,(2,6),\ldots(6,6)\} =$
$\{(i,j)|i,j \in \{1,2,3,4,5,6\}\}$
- Event $A =$ sum of dice is seven
$= \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$

Basic Set Review

- $A \cap B$ = set of elements that are in both A and B
- $A \cup B$ = set of elements that are in either A or B
- A', $\bar{A}$ or A^c = set of elements in the sample space S that are not in A

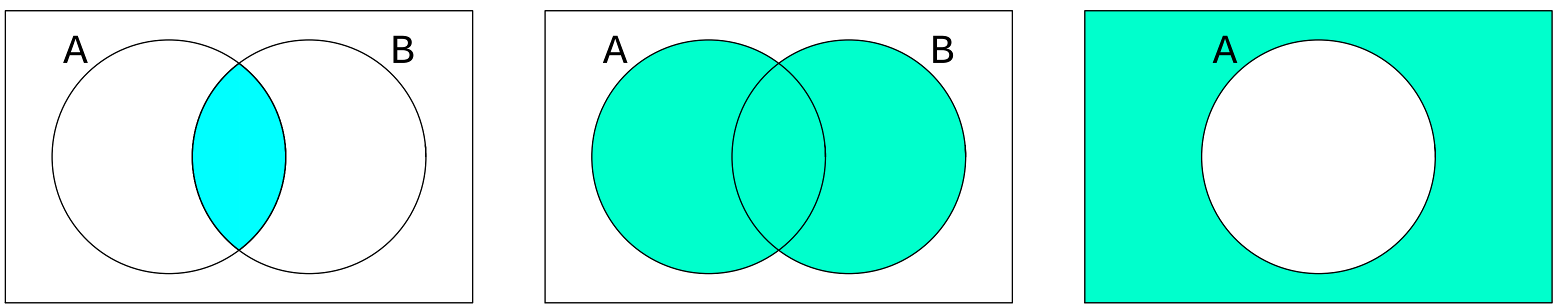


Figure 1: $A \cap B$,$A \cup B$, and A^c

Axioms of Probability

For $A \subseteq S$,

- $0 \leq P(A)$
- $P(S) = 1$
- If $\{A_1, A_2, \ldots\}$ is a sequence of mutually exclusive events, i.e., $A_i \cap A_j = \emptyset$ if $i \neq j$, then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

Consequences

- $P(\emptyset) = 0$; $0 \leq P(A) \leq P(S) = 1$
- If $S = \{s_1, s_2, ..., s_N\}$ is finite and all outcomes are equally likely, then

$$P(\{s_i\}) = \frac{1}{N} \quad \text{for all } i$$

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

where $|A|$ = number of elements in A

In Example 1, $P(A) = P(\text{sum of dice is seven}) = \frac{6}{36} = \frac{1}{6}$

Basic Theorems

- For any event A, $P(A^c) = 1 - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

These rules can be used to compute probabilities for events from given probabilities. In the case of equally likely outcomes, one important ingredient to establish probabilities for events is counting.

Multiplication Principle

Let $E_1, E_2, .., E_k$ be sets with $n_1, n_2, \ldots, n_k$ elements, respectively. Then there are $n_1 \cdot n_2 \cdots n_k$ ways in which we can first choose an element of E_1, then an element of $E_2, \ldots$ and finally an element of E_k.

Addition Principle

When counting different cases that have no overlap, the probabilities add up (follows from Axiom 3 for the finite case).

Example 2: You have 4 pants and 3 T-shirts. How many different outfits can you create?

$$\underset{\text{Pants}}{\underline{\quad \mathbf{4} \quad}} \cdot \underset{\text{T-Shirts}}{\underline{\quad \mathbf{3} \quad}} = \mathbf{12}$$

Example 3: Given the nucleotides A, C, G, and T.

a) How many different DNA sequences of length 4 can be created?

$$\underset{\text{A,C,G,T}}{\underline{\quad \mathbf{4} \quad}} \cdot \underset{\text{A,C,G,T}}{\underline{\quad \mathbf{4} \quad}} \cdot \underset{\text{A,C,G,T}}{\underline{\quad \mathbf{4} \quad}} \cdot \underset{\text{A,C,G,T}}{\underline{\quad \mathbf{4} \quad}} = 4^4$$

Example 3 continued:

b) How many that start with **T** and end with **C**?

$$\underset{\text{T}}{\underline{\mathbf{1}}} \cdot \underset{\text{A,C,G,T}}{\underline{\mathbf{4}}} \cdot \underset{\text{A,C,G,T}}{\underline{\mathbf{4}}} \cdot \underset{\text{C}}{\underline{\mathbf{1}}} = 4^2$$

c) How many 4 letter sequences with no repeated nucleotide?

$$\underset{\text{A,C,G,T}}{\underline{\mathbf{4}}} \cdot \underset{\text{all but first}}{\underline{\mathbf{3}}} \cdot \underset{\text{last two}}{\underline{\mathbf{2}}} \cdot \underset{\text{last}}{\underline{\mathbf{1}}} = 4! = 24$$

Special Formulas

- Selecting k items from a set of n elements, repetition allowed (see Example 3a), $n = k = 4$) can be done in n^k ways

- Selecting k items from a set of n elements, repetition **not** allowed (see Example 3c), $n = k = 4$) can be done in

$$P_{n,k} = n \cdot (n-1) \cdot (n-2) \cdots (n-k+1) = \frac{n!}{(n-k)!}$$

ways

Note: $P_{n,k}$ is the number of **k-element permutations of a set of n elements**. Order matters!

Example 3 continued:

d) How many that have **at least** one **G**?

Hard to count - look at complement

A = at least one G; A^c = no G

$$\underset{\text{A,C,T}}{\underline{\mathbf{3}}} \cdot \underset{\text{A,C,T}}{\underline{\mathbf{3}}} \cdot \underset{\text{A,C,T}}{\underline{\mathbf{3}}} \cdot \underset{\text{A,C,T}}{\underline{\mathbf{3}}} = 3^4$$

$$|A| = |S| - |A^c| = 4^4 - 3^4 = 256 - 81 = \mathbf{175}$$

Note: When counting **at least, at most, more than, less than** $\Rightarrow$ try **complement**

Amino acids are codons of length 3 made up from C, A, G, U $\Rightarrow 4^3 = \mathbf{64}$ possibilities.

Why are there only 20 amino acids?

(Early) Combinatorial Explanation:

- Reading frame of width 3 should recognize amino acid without ambiguity.
- Amino acid of the form **AAA** cannot happen - AAAAAA is ambiguous $\Rightarrow 64 - 4 = \mathbf{60}$
- Amino acid of the form **XYZ**: to read XYZXYZ unambiguously, YZX and ZXY have to be excluded $\Rightarrow \frac{60}{3} = \mathbf{20}$.

Generalization of Permutations

The number of distinguishable permutations of n objects of k different types, where n_1 are alike, n_2 are alike, ..., n_k are alike and $n = n_1 + n_2 + \cdots + n_k$ is given by

$$\tilde{P}_{n_1,n_2,\ldots,n_k} = \frac{n!}{n_1! \cdot n_2! \cdots n_k!}$$

Special case $k = 2$:

$$\tilde{P}_{n_1,n_2} = \tilde{P}_{n_1,n-n_1} = \frac{n!}{n_1! \cdot (n-n_1)!}$$

Combinations

The number of distinguishable permutations of n objects of 2 different types with k and $n-k$ objects of each type is given by $\tilde{P}_{n-k,k} = \frac{n!}{(n-k)! \cdot k!}$. This also counts

$C_{n,k}$ = the number of **k-element combinations of a set of n elements**. Order does NOT matter!

How does that work? Why is the number of permutations of two sets of alike objects the same as the number of k-element combinations? Permutations involve order, whereas combinations do not. What is going on ?????

Explanation

Let's consider the problem of counting binary strings of length 8 which have exactly 3 ones (and 5 zeros).

- We can think of this as permutations (since order matters) of two sets of alike objects (zeros and ones), with $n_1 = 3$ and $n_2 = 5$. The number of such strings is given by $\tilde{P}_{3,5} = \frac{8!}{3! \cdot 5!}$
- Alternatively, imagine a hat containing slips of paper with numbers 1 through 8. Select 3 slips from the hat (order does not matter); the selected numbers tell us at which positions in the string the ones occur. This can be done in $C_{8,3} = \frac{8!}{3! \cdot 5!}$ ways.

Independence

When computing intersection probabilities, there are two cases:

- the events are independent
- the events are dependent

Independence captures the idea that knowing about one event does not influence the likelihood for the second event.

Example 3 (continued): For DNA sequences of length 4 with certain properties we computed

- Total number of sequences $= 4^4$
- # of sequences starting with T and ending with C $= 4^2$

Equally likely events implies that

$$P(\text{sequence starts with T and ends with C}) = \frac{4^2}{4^4} = \frac{1}{16}$$

Note that this is an intersection probability: Let $A_i = L$ denote the event that the i^{th} letter of the seqence is L, where $L =$A, C, G, or T. Then

$P(\text{starts with T and ends with C}) =$

$$P(A_1 = \text{T} \cap A_2 = \text{any} \cap A_3 = \text{any} \cap A_4 = \text{C})$$

Also,

$P(\text{starts with T and ends with C}) = \frac{1 \cdot 4 \cdot 4 \cdot 1}{4 \cdot 4 \cdot 4 \cdot 4} = \frac{1}{4} \cdot \frac{4}{4} \cdot \frac{4}{4} \cdot \frac{1}{4} =$

$$P(A_1 = \text{T}) \cdot P(A_2 = \text{any}) \cdot P(A_3 = \text{any}) \cdot P(A_4 = \text{C})$$

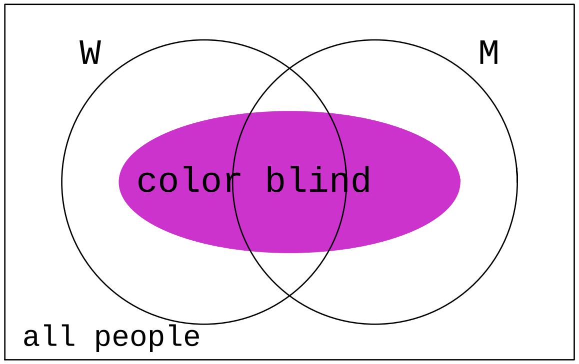
Definition: Two events A and B are independent if and only if $P(A \cap B) = P(A)P(B)$.

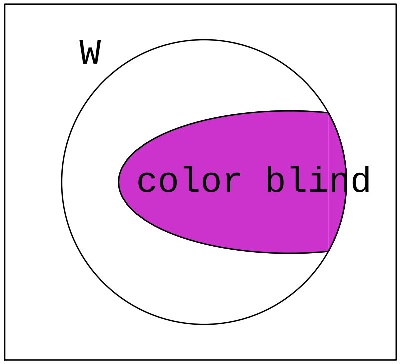
Independence of events can be checked using this definition. Most often, however, independence is derived from the context, and then used to easily compute intersection probabilities. Typical examples of independent events include

- tosses of a **fair** coin
- rolls of **fair** dice
- gene formation - genes received from father and mother are considered independent

Conditional Probability

Probabilities given for women, men and color blind people. Want to find percentage of women that are color blind.





$$P(\text{color blind}|\text{W}) = \frac{P(\text{color blind} \cap \text{W})}{P(\text{W})}$$

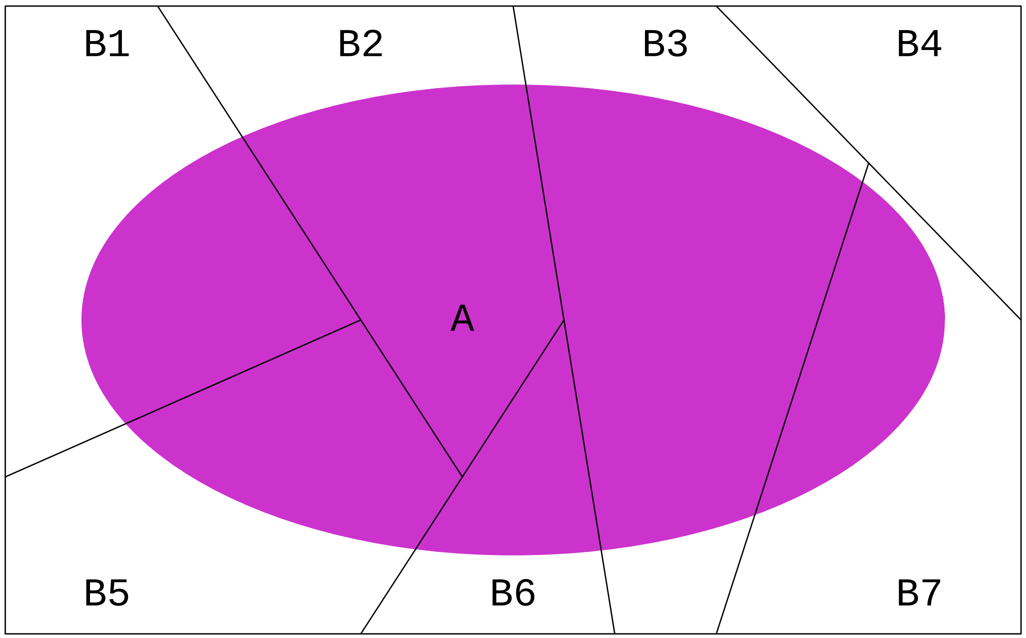
The **conditional probability of** an event B **given** that an event A has occurred (for $P(A) \neq 0$) is given by

$$\mathbf{P(B|A)} = \frac{\mathbf{P(A \cap B)}}{\mathbf{P(A)}}$$
$$= \frac{|\mathbf{A \cap B}|}{|\mathbf{A}|} \quad \text{(if equally likely outcomes)}$$

$\rightsquigarrow$ Multiplication Rule: $\mathbf{P(A \cap B) = P(A) \cdot P(B|A)}$

Idea: Look at intersection as 2-stage process - first A occurs, then B occurs

Law of Total Probability



A **partition** $\{B_1, B_2, \ldots, B_k\}$ of the sample space S is a collection of pairwise disjoint sets with $B_i \neq \emptyset$ such that $\bigcup_{i=1}^{k} B_i = S$.

For any event A and partition $\{B_1, B_2, \ldots, B_k\}$

$$\begin{aligned} A &= (A \cap B_1) \cup (A \cap B_2) \cup \ldots \cup (A \cap B_k) \\ \Rightarrow \quad P(A) &= P(A \cap B_1) + P(A \cap B_2) + \ldots + P(A \cap B_k) \end{aligned}$$

Law of Total Probability

$$\mathbf{P(A) = \sum_{i=1}^{k} P(A \cap B_i) = \sum_{i=1}^{k} P(B_i) P(A|B_i)}$$

Note: In applications, the B_i often correspond to different cases.

Example 4: An insurance company rents 35% of cars for its customers from agency I and 65% from agency II. If 8% of cars from agency I and 5% of cars from agency II break down during the rental period, what is the probability that a car rented by the insurance company breaks down during the rental period?

$$P(\text{I}) = .35 \quad P(\text{bd}|\text{I}) = .08 \quad P(\text{II}) = .65 \quad P(\text{bd}|\text{II}) = .05$$

$$\begin{aligned} P(\text{bd}) &= P(\text{I})P(\text{bd}|\text{I}) + P(\text{II})P(\text{bd}|\text{II}) \\ &= (.08)(.35) + (.05)(.65) \approx \mathbf{.06} \end{aligned}$$

Example 5: In answering a question on a multiple choice test, a student either knows the answer or guesses randomly among the five answers. Assume that the student knows the answer with probability 1/2. What is the probability that the student actually knew the answer, given that he/she answered correctly?

Given: $P(c|k) = 1$, $P(k) = 1/2$; wanted: $P(k|c) = ?$

$$\begin{aligned} P(k|c) &= \frac{P(k \cap c)}{P(c)} = \frac{P(c|k)P(k)}{P(c|k)P(k) + P(c|nk)P(nk)} \\ &= \frac{1 \cdot (0.5)}{1 \cdot (0.5) + (0.2)(0.5)} = \mathbf{5/6} \end{aligned}$$

Bayes' Theorem

For any event A and partition $\{B_1, B_2, \ldots, B_k\}$ of S,

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{k} P(A|B_j)P(B_j)}$$

Special case for $k = 2$:

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{P(A|B_1)P(B_1) + P(A|B_1^c)P(B_1^c)}$$

Bayes' Theorem in Statistics

- $B_i \rightsquigarrow H_i$, with H_i hypotheses to be tested by observing data D
- know **prior** probabilities $P(H_i)$ and also know $P(D|H_i)$
- compute **posterior** probabilities $P(H_i|D)$ using Bayes' Theorem

$$P(H_i|D) = \frac{P(D|H_i)P(H_i)}{\sum_{j=1}^{k} P(D|H_j)P(H_j)}$$

Note: Often, the prior probabilities are not known and therefore all hypotheses are assumed to be equally likely $\Rightarrow P(H_i) = 1/k$.

- $(P(H_1), P(H_2), \ldots, P(H_k))$ is called the prior distribution
- $(P(D|H_1), P(D|H_2), \ldots, P(D|H_k))$ is called the likelihood function
- $(P(H_1|D), P(H_2|D), \ldots, P(H_k|D))$ is called the posterior distribution

If another observation is made, then the posterior probabilities from the previous observation can serve as the prior probabilities for the new observation.

Example 6 - Spam Filters

- The idea of spam filters is to classify an incoming message as either spam or "ham" , i.e., we want to compute $P(spam|message)$
- Use training set of email messages to compute the prior probabilities $P(spam)$ and $P(ham)$
- Messages are represented as vectors of words; thus, liklihoods for spam and ham are computed as: $P(message|spam) = \prod P(word_i|spam)$, $P(message|ham) = \prod P(word_i|ham)$
- $P(message)$ can be thought of as normalizing constant to make left side a probability

Example 7

One coin is selected at random from a fair coin and a biased coin, with $P(H|biased) = 2/3$. Then the coin is tossed and a **H** shows. What are the posterior probabilities for $H_1 =$ "fair coin was chosen" and $H_2 =$ "biased coin was chosen"?

Given:

- Prior distribution $(\frac{1}{2}, \frac{1}{2})$
- Likelihood function $(P(H|f), P(H|b)) = (\frac{1}{2}, \frac{2}{3})$

Wanted: Posterior distribution $(P(f|H), P(b|H)) = ?$

$$P(f|H) = \frac{P(H|f)P(f)}{P(H|f)P(f) + P(H|b)P(b)}$$

$$= \frac{(1/2)(1/2)}{(1/2)(1/2) + (2/3)(1/2)} = \frac{1/4}{7/12} = \mathbf{3/7}$$

$$P(b|H) = \frac{P(H|b)P(b)}{P(H|f)P(f) + P(H|b)P(b)}$$

$$= \frac{(2/3)(1/2)}{7/12} = \mathbf{4/7}$$

Note: If there are only two hypotheses, then $P(H_2|D) = 1 - P(H_1|D)$.

```
The fair coin was chosen
```

Out[28]//TableForm=

Obs	Prob(fair coin)
	0.5
H	0.428571
H	0.36
H	0.296703
H	0.240356
H	0.191792
H	0.151088
H	0.117764
H	0.0910024
H	0.0698407
T	0.101226
H	0.0778908
T	0.112457
H	0.0867822
H	0.0665301
T	0.0965823
T	0.1382
T	0.193901
H	0.152834
H	0.119179
H	0.0921296
T	0.132109
T	0.185885
H	0.146208
T	0.204372
T	0.278136