Lecture 1: Probability

MATH 697

Omidali Jazi

September 4, 2018

McGill University

Goals for this Chapter

- Introduction to probability
- Methods for calculating probabilities (Sample-point method, Event-composition method)
- Counting tools (Combinations, permutations, etc)
- Conditional probability and independence
- Probability laws and rules (Multiplicative and additive laws, Law of Total Probability and Bayes' Rule)
- Random sampling

Defining Probability

- In every day conversation, probability is a measure of one's belief in the occurrence of a future event.
  What is the chance of wining the election?
  In this sense, probability is completely subjective.
- More quantitatively, we define probability as the fraction of times an event occurs if it's repeated over and over (infinitely).
  This is the relative frequency definition of probability.

Probability Quantifies Outcome Uncertainty

Example 1

Think about someone flipping a coin:
With enough practice, can learn to flip a coin and have it come up heads every time.

In this case, there is no uncertainty in the outcome.
But without practice, the coin is going to come up heads and tails "randomly".

There is randomness because the particular outcomes are uncontrollable and/or unpredictable.

Probability helps you quantify the likelihood of various outcomes.

Probability and Inference

- In statistics, inference is using a sample of data to say something about an entire population. That is,
  You will learn formal methods for doing inference
  Hypothesis testing, Model fitting.
- But first, you need to have to understand the theory of probability.
  Both what it is and, perhaps more importantly, how to calculate?

Probability vs. Statistics

- What is the difference between probability and statistics inference?
- In probability, the randomization mechanism is assumed/known and calculations follow.
  E.g., Given a fair coin, what's the chance of observing 8 heads out of 10 coin flips?
- In statistics, the outcome is known and used to infer something about the mechanism.
  E.g., Given that I've just observed 8 heads out of 10 coin flips, is the coin fair?

Set Terminology and Notation

- To begin, we need some concepts and notation from set theory.
- Capital letters denote the sets of objects: A, B, etc.
  If the set A consists of objects a_1, a_2, a_3, we write $A = \{a_1, a_2, a_3\}$
- S is the unervsal set (i.e, The set of all possible objects)
- ϕ the null or empty set (i.e, $\phi = \{\}$)

Set Terminology and Notation

- For any two sets A and B, A is a subset of B if every object in A is in B. Notation: $A \subseteq B$.
- Example: If $B = \{a_1, a_2, a_3, a_4\}$ and $A = \{a_1, a_2, a_3\}$ then A is a subset of B.
- The null set is a subset of every set.
- The union of A and B is the set of all objects in A or B or both. Notation: $A \cup B$.
- Example: If $A = \{a_1, a_2\}$ and $B = \{b_1, b_2\}$ then $A \cup B = \{a_1, a_2, b_1, b_2\}$.

Venn Diagrams

FIGURE 2.2
Venn diagram for
$A \subset B$

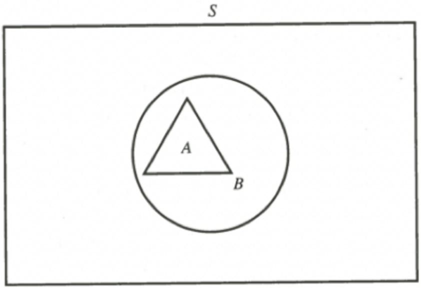
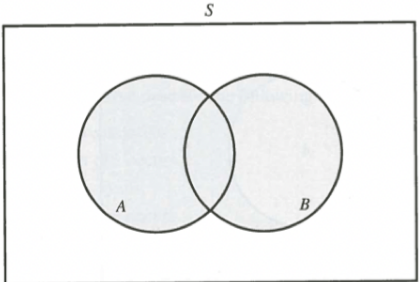


FIGURE 2.3
Venn diagram for
$A \cup B$

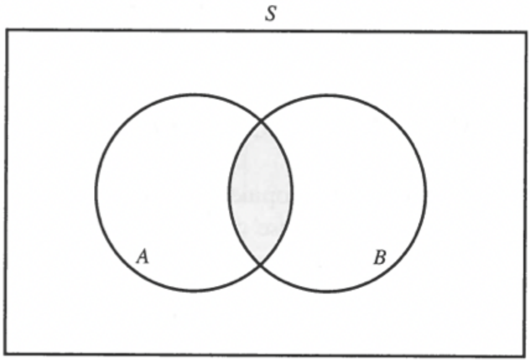


Set Terminology and Notation

- The intersection of sets A and B is the set of all objects in both A and B. Notation: $A \cap B$ or just AB
- Example: If $B = \{a_1, a_2, a_3, a_4\}$ and $A = \{a_1, a_2, a_3\}$ then $A \cap B = \{a_1, a_2, a_3\}$
- The key word for union is <u>or</u> while the key word for intersection is <u>and</u>

Venn Diagrams

FIGURE **2.4**
Venn diagram for AB

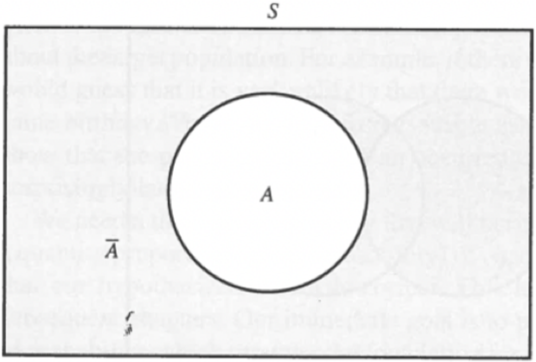


Set Terminology and Notation

- If A is a subset of S, then the complement of A is denoted by $\overline{A}$.
- The complement of A is the set of all objects in S that are not in A.
  Note that $A \cup \overline{A} = S$

Venn Diagrams

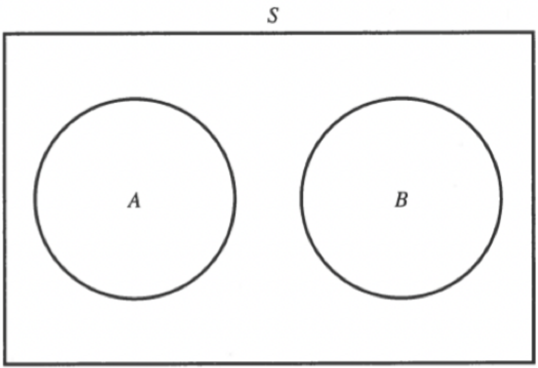
FIGURE 2.5
Venn diagram for $\overline{A}$



Set Terminology and Notation

- Two sets A and B are said to be disjoint or mutually exclusive if they have no object in common. That is, A and B are disjoint if $A \cap B = \phi$
  Note that, by definition, $A \cap \overline{A} = \phi$
- Distributive laws:
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$
- DeMorgan's Laws:
$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$
$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$

FIGURE 2.6
Venn diagram for mutually exclusive sets A and B



Definition 2

Simple event is the most basic outcome of an expriment and cannot be further decomposed, while compound events consist of two or more simple events.

Probability Terminology

- Example: The event “observe an odd outcome” on the six-sided die experiment is a compound event.
  and “observe a 5”.
  However, “observe a 1,” “observe a 3,” and “observe a 5” are all simple events since they cannot be decomposed any further.

- Each simple event corresponds to only one sample point. The letter E with a subscript will be used to denote a simple event or the corresponding sample point.

Probability Terminology

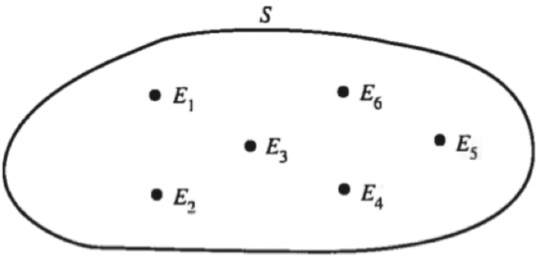
Definition 3

The sample space of an experiment is the set of all possible outcome (sample points).

- The sample space is denoted by S.
- In the six-sided die tossing experiment, $S = \{E_1, E_2, E_3, E_4, E_5, E_6\}$ where for example, $E_1 =$ "roll a 1"

Example

FIGURE 2.7
Venn diagram for the sample space associated with the die-tossing experiment



Probability Terminology

Definition 4

A discrete sample space is one that contains either a finite or countable number of distinct sample points.

- When an experiment is conducted, you will observe one and only one simple event.
  All simple events are, by definition, mutually exclusive.
  Die example: You cannot roll a 1 and a 2 at the same time.
- Thus, $E_1 \cap E_2 = \phi$.

Probability Terminology

Definition 5

An event in a discrete sample space S is a collection of sample points. That is, any subset of S.

- Thus, the general term "event" can mean a simple, event, a compound event, or even a non-event.
- The rule for determining which simple events are part of a compound event is:
  A simple event E_i is included in event A if and only if A occurs whenever E_i occurs.

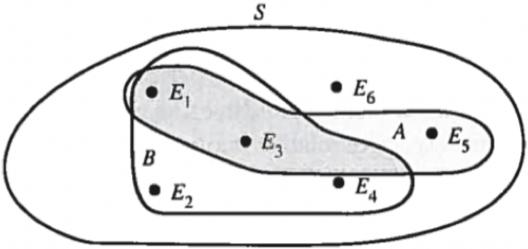
Example

Example 6

Back to tossing the six-sided die, Let E_i denote the event "roll the die and observe i" Then, for example, the compound event A, "observe an odd number," is $A = \{E_1, E_3, E_5\}$. We could also write $A = E_1 \cup E_3 \cup E_5$. Define event B to be "observe a number less than 5" or $B = \{E_1, E_2, E_3, E_4\}$

Example

FIGURE 2.8
Venn diagram for the die-tossing experiment



Probability Axioms

1. The relative frequency of any event must be greater than or equal to zero. Hence, negative frequency does not make sense.
2. The relative frequency of the whole sample space S must be 1 Since all possible experimental outcomes are in S, something in S must occur every time the experiment is conducted.
3. If two events are mutually exclusive, then the probability of their union is the sum of their respective frequencies.

Probability Axioms, Mathematically

Definition 7

Suppose S is a sample space. To every event A in S, we assign a number $P(A)$, called the probability of A such that

- Axiom 1: $P(A) \geq 0$
- Axiom 2: $P(S) = 1$
- Axiom 3: If $A_1, A_2, A_3, ...$ are a sequence of pairwise mutually exclusive events in S (i.e, $A_i \cap A_j = \phi$ if $i \neq j$)

$$P(A_1 \cup A_2 \cup ...) = P(A_1) + P(A_2) + ...$$

The Sample-Point Method

- With discrete sample spaces, two ways to find probabilities: sample-point and event- composition methods Methods very similar, both using the sample space ideas we've been developing
- We first start with sample-point. The idea is much like what we have just done.
- We will then come back to event-composition approach.

Example

Consider a "fair" six-sided die. Let E_i $(i = 1, .., 6)$ denote the event "roll the die and observe i". Then, by Axioms 2 and 3:

$$\begin{aligned} P(S) &= P(E_1 \cup E_2 \cup E_3 \cup E_4 \cup E_5 \cup E_6) \\ &= P(E_1) + P(E_2) + P(E_3) + P(E_4) + P(E_5) + P(E_6) \\ &= 1 \end{aligned}$$

The Sample-Point Method Steps

1. First, define the experiment and clearly determine how to describe one simple event.
2. List the simple events and test to make sure they cannot be decomposed. This defines S.
3. Assign reasonable probabilities to the sample points in S making sure $P(E_i) \geq 0$ and $P(\sum_i E_i) = 1$.
4. Define the event of interest, A, as a specific collection of sample points.
5. Find $P(A)$ by summing the probabilities of the sample points in A.

Pros & Cons of Sample-Point Approach

- Sample-point approach is direct and powerful, But it is prone to human error.
- Mistakes include mis-defining simple events and missing simple events. For sample spaces with large number of events, can be challenge (perhaps impossible) to count all the events.
- Leads us to learn about combinatorial analysis. (combinations and permutations)

Counting Sample Points

- Let's look at some useful and powerful ways to count numbers of events.
- A sample space contains N equally likely events.
  An event A contains exactly n_a sample points, then we have
  $$P(A) = \frac{n_a}{N}$$
- So, if we know N and can count n_a, then we can calculate the probability of event A.

The mn Rule

Theorem 8

With m elements $\{a_1, a_2, ..., a_m\}$ and n elements $\{b_1, b_2, ..., b_n\}$, it is possible to form $m \times n$ pairs containing one element from each group.

Permutations (Counting Ordered Arrangements)

Definition 9

An ordered arrangement of r distinct objects is called a permutation. The number of ways of ordering n distinct objects taken r at a time is denoted by ${}_nP_r$.

- Simple example: I have $n = 3$ people (A, B, and C) and $r = 2$ job positions. How many ways (permutations) are there to assign the positions to the people?
  There are six orderings: $\{A, B\}$, $\{A, C\}$, $\{B, A\}$, $\{B, C\}$, $\{C, A\}$, $\{C, B\}$

Permutations (Counting Ordered Arrangements)

Theorem 10

To calculate the number of ways r items can be ordered out of n

$$_nP_r = n(n-1)(n-2)\cdots(n-r+1) = \frac{n!}{(n-r)!}$$

Combination

Theorem 11

The number of ways of partitioning n distinct objects into k distinct groups containing n_1, n_2, ..., n_k objects, respectively, where each object appears in $\sum_{i=1}^{n} n_i = n$ is

$$\begin{pmatrix} n \\ n_1, n_2, ..., n_k \end{pmatrix} = \frac{n!}{n_1! n_2! ... n_k!}$$

The term $\begin{pmatrix} n \\ n_1, n_2, ..., n_k \end{pmatrix}$ is called the multinomial coefficients.

Example:

Example 12

You need to assign 20 aircraft to four sorties of size 6, 5, 5, and 4 aircrafts. How many ways can the 20 aircraft be assigned to the four groups? Solution: $\frac{20!}{6!5!5!4!}$

Combinations (Counting Unordered Arrangements)

Definition 13

The number of unordered subsets of r objects out of n is called the number of combinations. That is, it is the number of subsets, of size r, that can be formed from n objects.

- The notation is ${}_nC_r$
- Simple example: I have $n = 3$ people (A, B, and C) and pick $r = 2$ people to create a committee. How many combination of two types of three?
  There are only three: $\{A, B\}$, $\{A, C\}$,$\{B, C\}$.

Theorem 14

The number of unordered subsets of size r chosen (without replacement) from n objects is

$$ {}_nC_r = \binom{n}{r} = \frac{{}_nP_r}{r!} = \frac{n!}{r!(n-r)!} $$

The term $\binom{n}{r}$ is generally referred as the binomial coefficient.

Conditional Probability and Independence

- The idea of conditional probability is that the probability of an event depends on our knowledge of other events that have occurred.

- Example:
  Imagine I am holding a card I picked at random out of a standard deck of playing. What is the probability it is the ace of hearts?

- It is $1/52$ or just under 0.02.
  Now, if I tell you the card I am holding is an ace, what is the chance I'm holding the ace of hearts?

- It's $1/4$ (or 25%) since we know it has to be one of only 4 aces in the deck.

Conditional Probability

Definition 15

The conditional probability of an event A, given that an event B has occurred, is equal to

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

provided that $P(B) > 0$.

Back to the Ace of Hearts Example

- Define the events: $A =$ choose the ace of hearts and B $=$ choose an ace (randomly out of 52 cards)

- Calculate the probabilities: Let's start with $P(B)$

- $P(B) = 4/52$ (there are 4 aces in 52 cards)
  Now, $P(A \cap B)$ means "the probability of choosing the ace of hearts <u>and</u> choosing an ace"

- In this case, that is the same as "the probability of choosing the ace of hearts"

- So, $P(A \cap B) = P(A) = 1/52$. Then

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/52}{4/52} = 1/4$$

Independence

- The idea of independent events is that the probability of occuring event A is unaffected by whether B occurs or not.
  Another way to think about it is that knowledge about event B does not affect your probabilistic assessment of A.
  This is the intuitive notion that A and B are independent.
  But, if knowledge of B tells you something about the probability of A, then the two events are dependent.

Independent Events

Definition 16

Two events A and B are said to be independent if any one of the following holds:

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

Otherwise, the events are dependent.
Note that each condition follows from either of the other two conditions and the definition of conditional probability.

Exmaple

Consider the following events in the toss of a single balanced six-sided die:

- A=Observe an odd number
- B= Observe an even number
- C= Observe a 1 or a 2.
- Question 1: Are A and B independent events?
  Solution: No, because

  $$P(A \cap B) = 0 \neq P(A)P(B) = 9/36$$

  Conclusion: Two disjoint events with positive probabilities are always dependent.

Example

- Question 2: Are A and C independent events?
  Solution: Yes, because

  $$P(A \cap C) = 1/6 = P(A)P(C) = (3/6)(2/6) = 1/6$$

  Conclusion: Disjoint events and independent events are different notions and should not be mixed up. In this case, two events share a sample point, but they are independent.

Two Laws of Probability

Theorem 17

(Multiplicative Law of Probability): The probability of the intersection of two events A and B is

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

If A and B are independent, then

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

Extending the Multiplicative Law

By twice applying the Theorem

$$\begin{aligned} P(A \cap B \cap C) &= P[(A \cap B) \cap C] \\ &= P(A \cap B)P(C|A \cap B) \\ &= P(A)P(B|A)P(C|A \cap B) \end{aligned}$$

And, this can be extended to any number, say k, events:

$$\begin{aligned} &P(A_1 \cap A_2 \cap .... \cap A_k) = \\ &\qquad P(A_1)P(A_2|A_1)...P(A_k|A_1 \cap A_2 \cap ... \cap A_{k-1}) \end{aligned}$$

Application Example

- The Multiplicative Law is very useful for calculating complicated probabilities when conditional probabilities are easier to determine.
- Example: What is the probability of choosing a female local undrgraduate stuent when the university consists of 60% female and 40% male students. If a student is female, 70% and 30% are the chances of being local and international respectively. Further, half of of the female local students are undergraduate, while 30% and 20% are master and PhD respectively.
  Solution: $0.6 \times 0.7 \times 0.5 = 0.21$

The Addition and Complement Rules

Theorem 18

(The Additive Law of Probability): The probability of the union of two events A and B is

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

If A and B are mutually exclusive, then

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

Theorem 19

For event A,

$$P(A) = 1 - P(\overline{A})$$

Example Application: Gene Expression Profiling

Example 20

Gene expression profiling is a state-of-the-art method for determining the biology of cells. In Briefings in Functional Genomics and Proteomics (Dec. 2006), biologists at Pacific Northwest National Laboratory reviewed several gene expression profiling methods. The biologists applied two of the methods (A and B) to data collected on proteins in human mammary cells. The probability that the protein is cross-referenced (i.e., identified) by method A is .41, the probability that the protein is cross-referenced by method B is .42, and the probability that the protein is cross-referenced by both methods is .40.

Example Application

- Find the probability that the protein is cross-referenced by either method A or method B.
  Solution: $0.41 + 0.42 - 0.40 = 0.43$
- On the basis of your answer to part a, find the probability that the protein is not cross-referenced by either method.
  Solution: $1 - 0.43 = 0.57$

Example Application

Example 21

It is known that a patient with a disease will respond to treatment with probability 0.9. If three patients with the disease are treated and respond independently, find the probability that at least one will respond.

Solution: $1 - P(No\ Response) = 1 - 0.1^3 = 0.999$

Extension: Iclusion-Exlusion Formula

The Additive Law of Probability can be extended to k events, $A_1, A_2, ..., A_k$ as

$$P\left(\bigcup_{i=1}^{k} A_i\right)$$
$$= \sum_{i=1}^{k} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + ... + (-1)^{k-1} P(A_1 \cap ... \cap A_k)$$

Calculating Probabilities Via the Event-Composition Method

- The event-composition method uses the various laws of probability that we have just learned to calculate the probability of some event A.
- The idea is to express A as the union, intersection, or compliment of other sets.
- Then, calculate the probabilities of those events and appropriately combine the probabilities to obtain $P(A)$.

Law of Total Probability and Bayes' Rule

It is often useful to view a sample space as the union of mutually exclusive subsets.

Definition 22

For some positive integer k, let the sets $\{B_1, B_2, ..., B_k\}$ be such that

1. $S = B_1 \cup B_2 \cup ... \cup B_k$
2. $B_i \cap B_j = \phi$ for $i \neq j$.

Then the collection of sets $\{B_1, B_2, ..., B_k\}$ is said to be a partition of S.

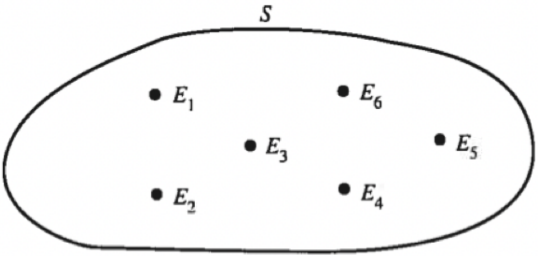
Using the Partition to Decompose a Set

Now, if A is any subset of S and $\{B_1, B_2, ..., B_k\}$ is a partition of S, then A can be decomposed as follows:

$$A = (A \cap B_1) \cup (A \cap B_2) \cup ... \cup (A \cap B_k)$$

An illustration for $k = 3$

FIGURE 2.7
Venn diagram for the sample space associated with the die-tossing experiment



Law of Total Probability

Theorem 23

Assume that that $\{B_1, B_2, ..., B_k\}$ *is a partition of* S *such that* $P(B_i) > 0$ *for* $i = 1, ..., k$. *Then for any event* A,

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

Example Application

Example 24

Of the voters in a city, 40% are Republican and 60% are Democrats. Among the Republicans, 70% are in favor of a bond issue, while 80% of the Democrats favor this issue. If a voter is selected at random, what is the probability that he/she will favor the bond issue?

Solution: $P(F) = 0.7 \times 0.4 + 0.8 \times 0.6 = 0.76$

Bayes' Rule

Theorem 25

Assume that $\{B_1, B_2, ..., B_k\}$ *is a partition of* S *such that* $P(B_i) > 0$ *for* $i = 1, ..., k$. *Then,*

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

Applications of the Bayes' Rule

- Invert conditional probabilities
  You know the various conditional probabilities $P(A|B_1)$, $P(A|B_2)$,..., $P(A|B_k)$ but you want to know $P(B_j|A)$ for some B_j

- Update probabilities $P(B_j|A)$
  You know or obtain $P(B_1)$, $P(B_2)$,...,$P(B_k)$ then subsequently observe/collect some information $P(A|B_1)$, $P(A|B_2)$,..., $P(A|B_k)$.
  Then update your probabilities conditioned on what you've observed by calculating $P(B_1|A)$, $P(B_2|A)$, ...,$P(B_k|A)$

Application: A Drug Testing Example

Example 26

Based on an investigation, it is believed that 1% of the sailors use illegal drugs. A test detects drug use as follows:

- If someone has used illegal drugs, there is a 99% chance it will detect it (true positive).
- If someone does not use illegal drugs, there is a 0.5% chance the test will indicate it (false positive).

What is the chance that a positive drug test signals a drug user?

Solution

- U_+ is the event an individual used illegal drugs, U_- is the event he/she did not
- T_+ is the event the drug test is positive for the individual
- We know $P(T_+|U_+) = 0.99$ and we want to calculate $P(U_+|T_+)$.

$$\begin{aligned}
P(U_+|T_+) &= \frac{P(T_+|U_+)P(U_+)}{P(T_+|U_+)P(U_+) + P(T_+|U_-)P(U_-)} \\
&= \frac{0.99 \times 0.01}{0.99 \times 0.01 + 0.005 \times (1 - 0.01)} \\
&= \frac{2}{3}
\end{aligned}$$

Interpretation

- Say, there are 10,000 sailors: 100 are drug users (1%) and 9,900 are not.
  99 will be detected (99% of 100).
  49.5 will be false positives (0.5% of 9,900)
- Total of 148.5 positives. Of those, 99/148.5 (2/3) are drug users while 49.5/148.5 (1/3) are clean.
- Conclusions:
  Don't use drug tests with high false positives.
  No point in testing clean population (all you get are false positives).

Example Application (Maize seeds)

Example 27

The genetic origin and properties of maize (modern-day corn) were investigated in Economic Botany. Seeds from maize ears carry either single spikelets or paired spikelets, but not both. Progeny tests on approximately 600 maize ears revealed the following information: Forty percent of all seeds carry single spikelets, while 60% carry paired spikelets. A seed with single spikelets will produce maize ears with single spikelets 29% of the time and paired spikelets 71% of the time. A seed with paired spikelets will produce maize ears with single spikelets 26% of the time and paired spikelets 74% of the time.

Example Application (Maize seeds)

- Find the probability that a randomly selected maize ear seed carries a single spikelet and produces ears with single spikelets.
  Solution: Exercise
- Find the probability that a randomly selected maize ear seed produces ears with paired spikelets.
  Solution: Exercise

Random Sampling

One application of probability is in selecting random samples from a population

- Important when executing various types of experiments as well as in collecting survey data.
- Random selection of a sample from a population is critical for being able to do valid inference. Lots of ways to draw "random samples". Two broad categories include:
    - Sampling without replacement: Population elements cannot be sampled more than once.
    - Sampling with replacement: Population elements can be sampled multiple times.

(Simple) Random Sampling

Definition 28

Let N be the number of elements in the population and n be the number in the sample. A (simple) random sample is a sample such that each of the $\binom{N}{n}$ possible samples has an equal probability of being selected.

This is just one possible type of sample. Others include stratified sampling, cluster sampling, and complex sampling that combines two or more of these methods.

What We Have Just Learned

- Introduction to Probability
- Methods for Calculating Probabilities (Sample-point Method, Event-composition Method)
- Counting Tools (Combinations, Permutations, etc)
- Conditional Probability and Independence
- Probability Laws and Rules (Multiplicative and Additive Laws, Law of Total Probability and Bayes' Rule)
- Random sampling