3. Application of Bayes’ Rule to Statistical Inference
   a. $B$ represents event of seeing the observed data set.
   b. $A_i$ represents the event that a certain model generated the data
      i. Ex., coin example: "Which coin" is the model.
      ii. Requires probabilities associated with various models.
         • Probabilities are called the prior.
         WMS: 2.11

H. Numerical summaries of Experiments
   1. Random variable notation:
      a. Definition: one or more numerical summaries of experimental results.
         i. Data analysts might talk of categorical random variables, but for this course restrict these to numerical
      b. They are usually written as capital letters (often $X$).
      c. They are functions of $s$.
      d. Note that the same kind of symbol (upper case Latin letter) is used for both sets and random variables.
         i. Which one is under consideration is hopefully clear from context
         ii. Additionally, early letters ($A$, $B$, $C$, etc.) will generally indicate events, and later letters ($X$, $Y$, $Z$, etc.) will generally indicate random variables.

2. Mechanistic Examples
   a. Coin flips: Observed the entire results of the experiment.
      i. Random variables representing heads or tails on successive flips represent all there is to know about
         the outcome.
      ii. With seven successive flips of a coin, there are exactly $2 \times 2 \times 2 \times 2 \times 2 \times 2 = 2^7 = 128$ points $s \in S$.
      iii. We might only be interested in a summary reflecting whether heads and tails tend to come up with equal frequency;
         • Consider the random variable representing the total # of heads.
         $X(HTHTHTH) = 4$.
   b. Dice: Report $X =$ total number of spots
      i. Sample space is
         \[
         \begin{align*}
         &R1G1 \quad R1G2 \quad R1G3 \quad R1G4 \quad R1G5 \quad R1G6 \quad R2G1 \quad R2G2 \quad R2G3 \quad R2G4 \quad R2G5 \quad R2G6 \quad R3G1 \quad R3G2 \quad R3G3 \quad R3G4 \quad R3G5 \quad R3G6 \quad R4G1 \quad R4G2 \quad R4G3 \quad R4G4 \quad R4G5 \quad R4G6 \quad R5G1 \quad R5G2 \quad R5G3 \quad R5G4 \quad R5G5 \quad R5G6 \quad R6G1 \quad R6G2 \quad R6G3 \quad R6G4 \quad R6G5 \quad R6G6
         \end{align*}
         \]
         ii. $X(R3G1) = 4$.

3. Real life examples.
   a. Medical Experiment:
      i. It is impossible to measure all aspects of a person’s health, or to make sense of an enormous # of even those measurements possible.
      ii. Usually researchers confine their attentions to a few measurements indicating the severity of a specific disease that’s being addressed, and a few specific risk factors.
         • when studying lung cancer one might measure

   b. Economics:
      i. Here it is also impossible to measure or analyze all of all individuals’ decision.
      ii. In the study of companies’ investment and dividend policies, one might focus on measuring total investment and dividends in a certain set of companies, as well as covariates like size and type of business.

WMS: 2.12

4. Random variables often come from Random Samples.
   a. A simple random sample is a process of generating experimental outcomes
      i. by selecting items sequentially from a population
      ii. such that each item has an equal choice of being chosen,
      iii. such that the probability that the next item chosen does not depend on previous choices,
      iv. and usually such that chosen items are removed from further selection: sampling without replacement
   b. Sampling schemes other than simple random samples are sometimes used
      i. Ex., a scheme to survey opinions that may sample households and then sample multiple people in households.
      ii. Ex., a scheme to make sure some rare traits are reflected in a sample might make subjects with these traits more likely for inclusions
   c. They are functions of $s$.
   d. Note that the same kind of symbol (upper case Latin letter) is used for both sets and random variables.
      i. Which one is under consideration is hopefully clear from context
      ii. Additionally, early letters ($A$, $B$, $C$, etc.) will generally indicate events, and later letters ($X$, $Y$, $Z$, etc.) will generally indicate random variables.

WMS: 2.11

III. Discrete Probability Distributions
A. Change focus
   1. Focus on events defined in terms of random variables
      a. Allows us to add structure to questions
         i. Order to possible values
         ii. Describe distance between points
         iii. Allow for totaling wins and losses
         iv. Next semester will lead to identification of data summaries to aid with inference.

B. Discrete Probability distns.
   1. Discrete Case: Possible data values can be put into a list
      a. Specific mathematical term is that possible values are countable, and that the set of values can be put into a list $X = \{x_1, x_2, x_3, \ldots \}$.
      b. Possible data values are called probability atoms.
      c. Result is a probability structure like before, with the measurement indicating the severity of a specific disease that’s being addressed, and a few specific risk factors.

   2. Examples of Random Variables with a countable number of outcomes
      a. The # of heads in first 10 flips,
         i. Sample points $s$ might be taken to be infinite strings of $H$ and $T$
         ii. Sample space $S$ is the set of all such infinite strings
         iii. Set of potential values is $X = \{0, 1, 2, \ldots, 10\}$.
      b. The # of flips needed to get 10 heads.
i. Sample points \( s \), sample space \( S \) the same.
ii. Set of potential values is \( \mathcal{X} = \{10, 11, 12, \ldots\} \).
iii. The \# of patients recovering when on a certain medication
iv. Counterexample: Set of all real numbers.

3. Get probabilities via summation.
   a. Probability of seeing a one of a set of possible results here is the sum of the probabilities associated with each value in the set.
   b. Call the probabilities associated with each element a probability function at that possible value.
      i. \( p_X(x) = P(X = x) = P(\{s : X(s) = x\}) \)
      ii. Probabilities for events defined in terms of \( X \) can be calculated by summing the \( p_X(x) \) over various sets of \( x \).
         • Example: probability of six or more favorable outcomes in nine trials add the probabilities associated with six, seven, eight, and nine successes.
   c. Hence probability calculations involving \( X \) can be performed without considering the underlying sample points \( s \).
      i. And so reference to sample points will not be prominent.

C. The distribution function contains information about probabilities.
   1. What happened to \( S \)?
      a. For now, consider \( S \) as the set of all possible values of \( X \).
      b. Later on, we will want to consider multiple random variables moving together.
      c. This is the context when we will rely on some underlying ties (coming from the same sample point) to allow for dependence in these values.

2. Two Consequences of Summability:
   a. If \( A_j \) are such that \( A_j \subset A_{j+1} \), then \( \lim_{j \to \infty} P(A_j) = P(\cup_{j=1}^{\infty} A_j) \).
      i. Let \( B_1 = A_1 \), \( B_j = A_j \cap \overline{A_{j-1}} \) represent the original sets split into disjoint sets.
      ii. Then \( B_j \cap B_k = \emptyset \) for \( j \neq k \).
      iii. Also \( \cup_{j=1}^{\infty} A_j = \cup_{j=1}^{\infty} B_j \).
      iv. So \( P(\cup_{j=1}^{\infty} A_j) = P(\cup_{j=1}^{\infty} B_j) = \lim_{n \to \infty} \sum_{j=1}^{n} P(B_j) = \lim_{n \to \infty} P(A_n) \).
   b. If \( C_j \) are such that \( C_{j+1} \subset C_j \), then \( \lim_{j \to \infty} P(C_j) = P(\cap_{j=1}^{\infty} C_j) \).
      i. Apply the above to \( A_j = C_j \).
      ii. So \( P(\cap_{j=1}^{\infty} C_j) = P(\cap_{j=1}^{\infty} A_j)) = 1 - P(\cup_{j=1}^{\infty} A_j) = 1 - \lim_{j \to \infty} P(A_j) = \lim_{j \to \infty} P(C_j) \).

3. Definition and Properties of the Distribution Function
   a. Define the distribution function:
      \( F_X(x) = P(X \leq x) \). See Fig. 11.
   b. \( F_X(x) \) is non-decreasing, since
      i. Take \( x_1 < x_2 \)
      ii. Nondecreasing means \( F_X(x_1) \leq F_X(x_2) \)
      iii. To see this, let \( A = \{s : X(s) \leq x_1\} \) and \( B = \{s : X(s) \leq x_2\} \)
   c. \( F_X(x) \) is finite, values as \( x \to \pm \infty \) are obvious, since
      i. \( F_X(x) = 0 \) below the minimal value.
      ii. \( F_X(x) = 1 \) above the maximal value.
      iii. Infinite \( \mathcal{X} \) requires the above investigation.

   f. Let \( F_X(x^+) = \lim_{y \to x, y > x} F_X(y) \), \( F_X(x^-) = \lim_{y \to x, y < x} F_X(x) \).
   g. Then \( F_X(x^+) = F_X(x) \), and \( F_X(x) \) is continuous from the right.
      i. Holds because if \( x_j \) is a monotonic sequence such that \( x_j > x \) and \( x_j \to x \), then
        \( F_X(x) = P(X \leq x) = P(\cap_{j=1}^{\infty} (X \leq x_j)) = \lim_{j \to \infty} P(X \leq x_j) = \lim_{j \to \infty} F_X(x_j) \).
      ii. Other direction doesn’t hold, because if \( x_j < x \), \( x_j \to x \), \( \cup_{j=1}^{\infty} (X \leq x_j) \neq \{X \leq x\} \).
      iii. Note, however, that this argument shows \( F_X(x^-) = P(X < x) \).
   h. For a discrete distribution, distribution function consists of flat bits and jumps. See Fig. 12.

4. Probabilities of Intervals from the distribution function.
   a. Take \( x_2 \geq x_1 \).
   b. \( \{s : X(s) \leq x_2\} = \{s : x_1 < X(s) \leq x_2\} \cup \{s : X(s) \leq x_1\} \).
      i. \( P(X \leq x_2) = P(x_1 < X \leq x_2) + P(X \leq x_1) \).
      ii. \( P(x_1 < X \leq x_2) = P(X \leq x_2) - P(X \leq x_1) \).
   c. \( P(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1) \), and so \( P(x_1 < X) = 1 - F_X(x_1) \).
\( F_X(2.5) = F_X^+(2.5) = F_X^-(2.5) \)

\[ F_X(2) = F_X^+(2) \]
\[ F_X^-(2) \]

\[ F_X(2) = F_X^+(2) \]
\[ F_X^-(2) \]

\[ F_X(2) = F_X^+(2) \]
\[ F_X^-(2) \]

- For other patterns of \(<\), \(\leq\),
  \[ P(x_1 < X < x_2) = F_X(x_2) - F_X(x_1) \]
  \[ \rightarrow P(x_1 \leq X) = 1 - F_X(x_1) \]
  \[ P(x_1 \leq X \leq x_2) = F_X(x_2) - F_X(x_1) \]
  \[ \rightarrow P(x_1 = X) = F_X(x_1) - F_X(x_1^-) \]
  \[ P(x_1 \leq X < x_2) = F_X(x_2) - F_X(x_1^-) \]