Recap

- Reliability block diagrams
- Conditional probability.
- Law of total probability
- Bayes theorem
- Conditioning examples
- Gambler’s ruin problem
Gambler’s Ruin

Example

• The gambler’s ruin problem is another famous problem in probability.
• It is related to sequential stopping rules in clinical trials.
• Players $A$ and $B$ have a total fortune of $M$ dollars between them.
• Player $A$ starts with $n$ dollars and $B$ with $M - n$.
• Each turn a coin is toss independently:
  • The probability of heads is $p$.
  • If a head occurs then $B$ pays one dollar to $A$.
  • If a tail occurs then $A$ pays one dollar to $B$.
• The game ends when one player has all $M$ dollars.
• What is the probability that $A$ ends up with all the fortune, or $B$ is ruined?
• Some simulations:
  
  http://www.stat.uiowa.edu/~luke/classes/193/ruin.R.
**Example (gambler’s ruin. continued)**

- Let $p_n$ be the probability that $B$ is ruined when $A$ starts with $n$ dollars.
- Then
  
  \[
  p_0 = 0 \\
  p_M = 1.
  \]

- For $0 < n < M$

  \[
  p_n = P(B \text{ is ruined} | \text{first toss is } H)p + P(B \text{ is ruined} | \text{first toss is } T)(1 - p)
  \]

- Now

  \[
  P(B \text{ is ruined} | \text{first toss is } H) = p_{n+1} \\
  P(B \text{ is ruined} | \text{first toss is } T) = p_{n-1}
  \]

- So $p_n$ satisfies the second order difference equation

  \[
  p_n = p_{n+1}p + p_{n-1}(1 - p).
  \]
Example (gambler’s ruin. continued)

• Subtracting \( p_n p \) from both sides and rearranging yields

\[
p_{n+1} - p_n = \frac{1-p}{p} (p_n - p_{n-1}).
\]

• Since \( p_0 = 0 \) this implies that

\[
p_{n+1} - p_n = \left(\frac{1-p}{p}\right)^n p_1
\]

• Therefore for \( 0 < n \leq M \)

\[
p_n = p_1 \sum_{k=1}^{n} \left(\frac{1-p}{p}\right)^{k-1}
\]
Example (gambler’s ruin. continued)

• The sum is

\[
\sum_{k=1}^{n} \left( \frac{1-p}{p} \right)^{k-1} = \begin{cases} 
1 - \left( \frac{1-p}{p} \right)^{n} & \text{if } p \neq \frac{1}{2} \\
\frac{1}{1 - \left( \frac{1-p}{p} \right)^{M}} & \text{if } p = \frac{1}{2}.
\end{cases}
\]

• Using the fact that \( p_{M} = 1 \) this gives

\[
p_{n} = \begin{cases} 
\frac{1 - \left( \frac{1-p}{p} \right)^{n}}{1 - \left( \frac{1-p}{p} \right)^{M}} & \text{if } p \neq \frac{1}{2} \\
n/M & \text{if } p = \frac{1}{2}.
\end{cases}
\]

• This holds for \( 0 \leq n \leq M \).
Continuous Sample Spaces

• Some experiments require sample spaces that are continuous ranges:
  • Measuring the melting point of a substance.
  • Measuring the angle at which a particle leaves a point source.
  • Recording the time it takes to service a customer at a bank.

• Some experiments may require several continuous ranges:
  • Recording height and weight of a randomly selected person.
  • Recording the service times for each of the first 10 customers.

• There can be a continuous range of continuous ranges:
  • Recording the trajectory of a migrating bird.

• Continuous ranges are uncountable and require some different approaches.
Example

- We would like to model the direction in which a particle leaves a point source.
- This has a number of applications:
  - Photon emission in optics.
  - Gamma particles in nuclear medicine and medical imaging.
- We would like to capture the idea that the particle is equally likely to leave in any direction.
- The angle will be in the range \([0, 2\pi)\).
- Equally likely to leave in any direction would mean
  - The probability of the angle falling in the ranges \([0, \pi) \) and \([\pi, 2\pi)\) should be \(\frac{1}{2}\) each.
  - The probability of the angle falling in the range \([\pi, 3\pi/2)\) should be \(\frac{1}{4}\).
  - The probability of the angle falling in an interval \([a, b)\) should be proportional to the length \(b - a\).
  - This means the probability would be \(\frac{b{-a}}{2\pi}\).
Example (continued)

- What is the probability the particle leaves in a particular direction $a$; what is $P(\{a\})$?
  - Let $A_n = [a, a + 1/n]$.
  - Then $A_1 \supset A_1 \supset \ldots$ and $\bigcap_{n=1}^{\infty} A_n = \{a\}$.
  - So by the continuity property of probability
    
    $$P(\{a\}) = \lim_{n \to \infty} P(A_n) = \lim_{n \to \infty} \frac{1}{2\pi n} = 0.$$ 

- A consequence is that
  
  $$P([a, b)) = P((a, b)) = P((a, b]) = P([a, b])$$

- We can compute the probability density near a direction $a$ as
  
  $$f(a) = \lim_{h \downarrow 0} \frac{P([a, a+h))}{h} = \lim_{h \downarrow 0} \frac{P([a-h, a])}{h} = \lim_{h \downarrow 0} \frac{1}{h} \frac{1}{2\pi} = \frac{1}{2\pi}.$$
• Does a probability with the properties of the previous example exist?

• Equivalent question: is it possible to compute a meaningful size, or length, for any subset of the real line?

• Unfortunately: no.

• There exist some very strange subsets of the real line.

• Reasonable subsets, even many very complicated ones, do allow their size to be computed.

• To move forward we need to be able to restrict our probabilities to a collection $\mathcal{B}$ of reasonable subsets.

• For our axioms to make sense we need our collection of reasonable subsets $\mathcal{B}$ to satisfy some closure conditions:

  1. $\emptyset \in \mathcal{B}$.
  2. If $A \in \mathcal{B}$ then $A^c \in \mathcal{B}$.
  3. If $A_1, A_2, \cdots \in \mathcal{B}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{B}$.

• A collection $\mathcal{B}$ of subsets of a sample space $S$ that satisfies these conditions is called a sigma-algebra of subsets.
Examples

• The smallest sigma-algebra on a sample space $S$ is
  \[ \mathcal{B} = \{ \emptyset, S \} \]

• The largest sigma-algebra on a sample space $S$ is
  \[ \mathcal{B} = \{ \text{all subsets of } S \} = 2^S \]

• If $S$ is countable and $\{s\} \in \mathcal{B}$ for all $s \in S$, then
  \[ \mathcal{B} = \{ \text{all subsets of } S \} \]

• If $S$ is not countable, then we often need to use a $\mathcal{B}$ that does not contain all subsets of $S$.

• Even for countable state spaces it is sometimes useful to use a $\mathcal{B}$ that does not contain all subsets of $S$. 
Example

Let $S = \mathbb{R} = (-\infty, \infty)$ and let $\mathcal{B}$ be the smallest sigma-algebra containing all open intervals $(a, b)$, $a, b \in \mathbb{R}$.

- This is well-defined.
- $\mathcal{B}$ is called the Borel sigma-algebra on $S$.
- The sets in $\mathcal{B}$ are called Borel sets.
- Non-Borel sets do exist but are quite strange.
- $\mathcal{B}$ is also the smallest sigma-algebra containing all intervals of the form $(-\infty, a]$ for all $a \in \mathbb{R}$.
- Other characterizations are possible.
- Analogous definitions apply when $S$ is an interval.
- Analogous definitions apply when $S$ is $\mathbb{R}^2$ or $\mathbb{R}^n$. 
A Final Definition of Probability

Definition

A probability, or probability function, is a real-valued function defined on a sigma-algebra $\mathcal{B}$ of subsets of a sample space $S$ such that

(i) $P(A) \geq 0$ for all $A \in \mathcal{B}$

(ii) $P(S) = 1$

(iii) If $A_1, A_2, \ldots \in \mathcal{B}$ are pairwise disjoint, then

$$P \left( \bigcup_{i=1}^{\infty} A_i \right) = \sum_{i=1}^{\infty} P(A_i).$$

• This is the definition introduced by Kolmogorov in 1933 that is now the standard.
• A triple $(S, \mathcal{B}, P)$ is called a probability space.
• If $S$ is a subset of the real line then a probability is sometimes called a probability distribution.
With this definition we can use the probability we developed to model the angle of a particle leaving a point source:

**Theorem**

For any bounded interval \([L, U]\) with \(L < U\) there exists a unique probability defined on the Borel subsets of the interval such that

\[
P((a, b)) = \frac{b - a}{U - L}
\]

for \(L \leq a < b \leq U\). This is called the **uniform distribution** on \([L, U]\).
Random Variables

• Random variables are numerical quantities with values we are uncertain about.

• In our framework we can think of each outcome in our sample space as producing a particular value for the random variable.

• We can think of a random variable as a real-valued function defined on our sample space.

• Sometimes we are interested in only one random variable, sometimes in many.
Examples

- Outcomes are all possible subsets of 3 out of 100 items.
  - We are mainly interested in the number of defectives in the sample.

- Outcomes are all possible past and future prices of a stock.
  - We are mainly interested in the stock today and one month from now.

- A student is selected at random. Outcomes are all possible students.
  - We are interested in the height and weight of the student selected.
Recap

- Conditioning examples
  - reliability block diagram
  - matching problem
  - gambler’s ruin problem
- Continuous sample spaces
  - sigma algebras
  - Borel sets
- A final definition of probability
- Random variables
A random variable represents a numerical measurement that can be computed for each outcome.

Informally: A random variable is a real-valued function defined on a sample space.

Random variables are usually denoted by upper case letters from the end of the alphabet:

\[ X(s) : S \rightarrow \mathbb{R} \]

Lower case letters usually denote generic realized values: \( x = X(s) \).

- Two dice are rolled; the sample space is \( S = \{(1, 1), (1, 2), \ldots, (6, 6)\} \).
- \( X \) is the sum of the values on the two dice.
- The experiment is performed and results in the outcome \( s = (3, 2) \).
- The realized value of \( X \) is \( x = X(s) = X((3, 2)) = 5 \).

Often the argument is dropped:

\[ P(\{s \in S : X(s) \leq 5\}) = P(\{X \leq 5\}) = P(X \leq 5) \]
Random Variables

- When our probability is defined on all subsets of the sample space then any real-valued function on the sample space is a random variable.
- If our probability is defined only for some subsets of the sample space, the subsets in a sigma-algebra $\mathcal{B}$, then to be able to compute $P(X \leq 5) = P(\{s \in S : X(s) \leq 5\})$ we need to have $\{s \in S : X(s) \leq 5\} \in \mathcal{B}$.
- This property is called measurability with respect to $\mathcal{B}$.
- A random variable is a $\mathcal{B}$-measurable real-valued function on the sample space $S$. 
**Definition**

Let $S$ be a sample space and $\mathcal{B}$ a sigma-algebra of events on $S$. A real-valued function $X = X(s) : S \to \mathbb{R}$ is a random variable if

$$\{ s : X(s) \leq t \} \in \mathcal{B}$$

for all $t \in \mathbb{R}$. An equivalent condition is that

$$X^{-1}(A) = \{ s : X(s) \in A \} \in \mathcal{B}$$

for any Borel set $A$. 
Examples

• Toss a coin $n$ times, $X =$ number of heads.
• Roll two dice,
  $X =$ sum,
  $Y =$ difference.
• Sample 1000,
  $X =$ number who favor some issue,
  $Y =$ number who oppose the issue,
  $Z = 1000 − X − Y$. 
Example

- A coin is tossed twice; the sample space is $S = \{HH, HT, TH, TT\}$.
- Let $X_i$ be the number of heads on toss $i$ and let $Y = X_1 + X_2$ be the total number of heads.
- Let $\mathcal{B}$ be all subsets of $S$.
- and let
  
  $\mathcal{B}_1 = \{\emptyset, S, \{HH, HT\}, \{TH, TT\}\}$.

- Both $\mathcal{B}$ and $\mathcal{B}_1$ are sigma-algebras.
- $X_1$, $X_2$, and $Y$ are all measurable with respect to $\mathcal{B}$.
- $X_1$ is measurable with respect to $\mathcal{B}_1$.
- $X_2$ and $Y$ are not measurable with respect to $\mathcal{B}_1$:
  
  \[
  \{X_2 \leq 0\} = \{HT, TT\} \notin \mathcal{B}_1 \\
  \{Y \leq 0\} = \{TT\} \notin \mathcal{B}_1
  \]
Example (continued)

- $\mathcal{B}_1$ is called the sigma-algebra generated by $X_1$.
- $\mathcal{B}_1$ is the smallest sigma-algebra with respect to which $X_1$ is measurable.
- $\mathcal{B}_1$ consists of those subsets of $S$ for which one can determine whether the outcome that has occurred is in the subset or not by knowing only the value of $X_1$.
- Conversely, if we know for each event $A \in \mathcal{B}_1$ whether $A$ has occurred or not, then we know the value of $X_1$. 
Distributions

• Let $X$ be a random variable on a probability space $(S, \mathcal{B}, P)$.
• For any Borel set $A$, let

$$P_X(A) = P(\{X \in A\}) = P(\{s \in S : X(s) \in A\}).$$

• $P_X$ is a probability on $\mathbb{R}$, the Borel sigma-algebra on $\mathbb{R}$.
• A random variable induces the probability $P_X$ on the real line $\mathbb{R}$.
• This induced probability is called the distribution of $X$.
• The triple $(\mathbb{R}, \mathcal{B}, P_X)$ is the probability space induced by $X$ on the real line.
• Similarly, a pair of random variables $X$ and $Y$ induces a probability on $\mathbb{R}^2$.
• This is called the joint probability distribution of $X$ and $Y$. 
Recap

- Random variables
- Measurability
- Distributions, induced probabilities
- Cumulative Distribution Function (CDF)
Cumulative Distribution Functions

- Writing down probabilities for all Borel sets would be a daunting task.
- We need some simpler tools for specifying distributions.
- One tool that is available for any random variable is the cumulative distribution function:

**Definition**

The cumulative distribution function (CDF) of a random variable $X$ is

$$F_X(x) = P(X \leq x)$$

for all $x \in \mathbb{R}$.

- A few authors define the CDF as $F_X(x) = P(X < x)$.
- The *survival function* $S_X(x) = P(X > x)$ is sometimes used for lifetime distributions.
- The survival function satisfies $S_X(x) = 1 - F_X(x)$. 
Example

For $X =$ number of heads in two coin flips:

$$F_X(x) = \begin{cases} 
0 & x < 0 \\
\frac{1}{4} & 0 \leq x < 1 \\
\frac{3}{4} & 1 \leq x < 2 \\
1 & x \geq 2
\end{cases}$$

The plot was created with

```r
s <- stepfun(c(0, 1, 2), c(0, .25, .75, 1))
plot(s, verticals = FALSE, pch = 19, main="", ylab=expression(F[X](x)))
```
Example

• Let $X$ be the angle in which a particle leaves a point source.
• Suppose $X$ has a uniform distribution on $[0, 2\pi)$.
• This means for $0 \leq a \leq b \leq 2\pi$

$$P(a < X < b) = \frac{b - a}{2\pi}.$$ 

• Therefore

$$F_X(x) = \begin{cases} 
0 & x < 0 \\
\frac{x}{2\pi} & 0 \leq x < 2\pi \\
1 & x \geq 2\pi.
\end{cases}$$
Some properties of the CDF:

(a) \( \lim_{x \to -\infty} F(x) = 0 \)

(b) \( \lim_{x \to \infty} F(x) = 1 \)

(c) \( F \) is nondecreasing

(d) \( F \) is right-continuous

Properties (a), (b), and (d) follow from the continuity of probability.

If a function \( F \) has these four properties, then it is the CDF of some random variable.

If \( F \) has a jump at \( x \), then \( P(X = x) \) is the size of the jump.
Computing Some Probabilities

\[ P(a < X \leq b) = F(b) - F(a) \] for any \( a < b \).

**Proof.**

The event \( \{X \leq b\} \) can be written as the disjoint union

\[ \{X \leq b\} = \{X \leq a\} \cup \{a < X \leq b\} \]

By finite additivity,

\[
F(b) = P(X \leq b) = P(X \leq a) + P(a < X \leq b) = F(a) + P(a < X \leq b)
\]
Computing Some Probabilities

\[ P(X < a) = F(a-) = \lim_{x \uparrow a} F(x) \] for any \( a \in \mathbb{R} \).

**Proof.**

Let \( A_n = \{a - \frac{1}{n} < X < a\} \). Then for each \( n \geq 1 \)

\[ P(X < a) = F \left( a - \frac{1}{n} \right) + P(A_n) \]

Furthermore,

\[ \bigcap_{n=1}^{\infty} A_n = \emptyset. \]

So by the continuity property \( \lim_{n \to \infty} P(A_n) = 0 \), and

\[ P(X < a) = F(a-) + \lim_{n \to \infty} P(A_n) = F(a-). \]
Computing Some Probabilities

\[ P(X = a) = F(a) - F(a-) = \text{size of the jump of } F \text{ at } a. \]

\textbf{Proof.}

The event \{X = a\} can be written as

\[ \{X = a\} = \{X \leq a\} \setminus \{X < a\}. \]

So

\[ P(X = a) = P(X \leq a) - P(X < a) = F(a) - F(a-) \]

\[ P(a \leq X \leq b) = F(b) - F(a-). \]

\textbf{Proof.}

Follows similarly from observing that

\[ \{a \leq X \leq b\} = \{X \leq b\} \setminus \{X < a\}. \]
Joint CDF of Two Random Variables

- The joint CDF of two random variables $X$ and $Y$ is defined as

$$F_{XY}(x, y) = P(\{X \leq x\} \cap \{Y \leq y\}).$$

- The probability of any rectangle with sides parallel to the axes can be calculated from the joint CDF.
- Other probabilities are harder to obtain from the joint CDF.
Comparing Distributions

**Definition**
Two random variables $X$ and $Y$ are *identically distributed* if their distributions $P_X$ and $P_Y$ are identical, i.e. if

$$P_X(A) = P(X \in A) = P(Y \in A) = P_Y(A)$$

for all Borel sets $A$.

**Theorem**
Two random variables $X$ and $Y$ are identically distributed if and only if their CDF’s are identical.
Classifying Distributions

- A random variable is \textit{discrete} if the set of its possible values is finite or countable.
- Equivalently, a random variable is discrete if its CDF is a step function.
- A random variable is \textit{continuous} if its CDF is continuous and there is a nonnegative function $f$ such that
  \[ F(x) = \int_{-\infty}^{x} f(u) \, du \]
  for all $x \in \mathbb{R}$.
  - $F$ will be differentiable “almost everywhere” with $F' = f$.
  - $f$ is called a \textit{probability density function}.
- Continuous distributions are sometimes called \textit{absolutely continuous}.
• Combinations of discrete and continuous features are possible.
  • The CDF of the waiting time at a bank might look like this:
  • The CDF of the amount of rainfall on a day would also look like this.
  • There also exist weird random variables that have a CDF that is continuous but nowhere differentiable.
Characterizing Discrete and Continuous Distributions

• The CDF can be used to characterize any probability distribution on the real line.
• Simpler characterizations are available if the RV’s are discrete or continuous.
• For discrete random variables we can use the probability mass function (PMF):

Definition
The probability mass function (PMF) of a discrete random variable is given by

\[ f_X(x) = P(X = x) \]

for all \( x \in \mathbb{R} \).
• If \( \mathcal{X} \) is the discrete set of possible values of \( X \), then \( f_X(x) = 0 \) for all \( x \notin \mathcal{X} \).
• Using the PMF we can compute any probabilities we want:

\[
P(a \leq X \leq b) = \sum_{a \leq x \leq b} f_X(x)
\]

\[
P(X \in A) = \sum_{x \in A} f_X(x)
\]

• In particular,

\[
F_X(x) = \sum_{y \leq x} f_X(y)
\]

• So the PMF completely determines the distribution of a RV.
• The PMF is defined for continuous random variables but isn’t useful.
• Instead, we use the *probability density function* (PDF):

\textbf{Definition}

If $X$ is a random variable and $f_X(x)$ is a nonnegative real-valued function on $\mathbb{R}$ such that

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(u) du$$

for all $x \in \mathbb{R}$ then $X$ is continuous and $f_X(x)$ is a probability density function (PDF) for $X$ (or for $F_X$).
Notes

• If \( f_X(x) \) is continuous at \( x \), then \( f_X(x) = F'_X(x) \), i.e.

\[
 f_X(x) = \lim_{\varepsilon \downarrow 0} \frac{F_X(x + \varepsilon) - F_X(x)}{\varepsilon} \\
= \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon} P(x \leq X \leq x + \varepsilon)
\]

• So \( f_X(x) \) is the “density of probability” at \( x \).
• \( f_X(x) > 1 \) is possible.
• \( P(X = x) = 0 \) for all \( x \) if \( X \) is continuous.
• For any Borel subset \( A \) of \( \mathbb{R} \),

\[
 P(X \in A) = \int_A f_X(x) \, dx
\]
**Theorem**

A function $f_X(x)$ is a PMF (or PDF) of some random variable if and only if

(i) $f_X(x) \geq 0$ for all $x \in \mathbb{R}$.

(ii) $\sum_x f_X(x) = 1$ \hspace{1cm} (PMF)

$\int_{-\infty}^{\infty} f_X(x) dx = 1$ \hspace{1cm} (PDF)
Example

• A random variable with density

\[ f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases} \]

for some \( \lambda > 0 \) is called an exponential random variable.

• This is a useful simple model for life times.

• We have \( f_X(x) \geq 0 \) for all \( x \) and

\[
\int_{-\infty}^{\infty} f_X(x) \, dx = 0 + \int_{0}^{\infty} \lambda e^{-\lambda x} \, dx = 0 + 1 = 1
\]

So \( f_X \) is a density for any \( \lambda > 0 \).

• The CDF is

\[
F_X(x) = \int_{-\infty}^{x} f_X(y) \, dy = \begin{cases} 0 & \text{for } x \leq 0 \\ 1 - e^{-\lambda x} & \text{for } x > 0 \end{cases}
\]