Lecture #12 – Wednesday, October 15, 2003

PARTITIONS

1) The sets \{A_1, A_2, ..., A_n\} are mutually exclusive

2) \( \bigcup_{i=1}^{n} A_i = S \)

   - If \( A_1, ..., A_n \) is a partition then for any event \( B \subset S \), \( B = \bigcup_{i=1}^{n} (B \cap A_i) \)

Example

   - \( B = (B \cap A) \cup (B \cap A^c) \)

LAW OF TOTAL PROBABILITY

- If \( \{B, B^c\} \) is a partition, then \( P(A) = P(A \mid B) P(B) + P(A \mid B^c) P(B^c) \)

- If \( \{B_1, B_2, B_3\} \) is a partition of \( S \) and \( P(B_i) > 0 \) for \( i = 1, 2, 3, ..., n \), then
  \[
  P(A) = P(A \mid B_1) P(B_1) + P(A \mid B_2) P(B_2) + P(A \mid B_3) P(B_3)
  \]

Generalization

\[
P(A) = \sum_{i=1}^{n} P(A \mid B_i) P(B_i)
\]

Example

A transport company needs to do safety checking for its cars. Suppose there are only two repair stores in town. One of them is owned by a rival firm and the other is owned by an allied firm so the diagnostics will not be objective. The friendly garage will lie with probability 20% if a car is unsafe and will tell the truth if the car is safe. The “rival” garage with probability 35% will declare a car unsafe even if the car is safe and will always declare the truth if the car is unsafe. Assume that 20% of the company’s cars are actually unsafe and assume that the company will obtain a certificate from each garage for each of its cars. What is the percentage of cases in which the two conclusions are contradictory?

- \( G = \{ \text{Good cars} \} \), \( B = \{ \text{Bad cars} \} \)
- \( C = \{ \text{Conclusions are contradictory} \} \)

\[
P(C) = P(C \cap G) + P(C \cap B) \text{ because } G \text{ and } B \text{ forms a partition of all cars}
\]

\[
P(C) = P(C \mid G) P(G) + P(C \mid B) P(B)
\]

- \( P(C \mid G) = 35\% \) because the friendly garage tells the truth and the rival garage tells the lie
- \( P(C \mid B) = 20\% \) because the rival garage tells the truth and the friendly garage tells the lie

\[
P(C) = \frac{35}{100} \times \frac{80}{100} + \frac{20}{100} \times \frac{20}{100} = \frac{32}{100}
\]

BAYES RULE

- Start with a partition \( \{B_1, B_2, B_3, \} \)
- Take an event \( A \)
Instead of being interested in $P(A \mid B_1)$, we are now interested in $P(B_1 \mid A)$.

$$P(B_1 \mid A) = \frac{P(B_1 \cap A)}{P(A)}$$

$$= \frac{P(A \mid B_1)P(B_1)}{P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3)}$$

$$= \frac{P(A \mid B_1)P(B_1)}{P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + P(A \mid B_3)P(B_3)}$$

**Experiment In 2 Stages**

- First stage – one of the $B$’s happens
- Second stage – we check whether $A$ happens

### Diagram

- Stage 1
  - $P(B_1)$
  - $P(B_2)$
  - $P(B_3)$

- Stage 2
  - $P(A \mid B_1)$
  - $P(A \mid B_2)$
  - $P(A \mid B_3)$

### I know $A$ happened. What is the probability that $B_1$ happened?

- $P(B_1 \mid A) = \frac{P(B_1 \cap A)}{P(A)} = \frac{P(A \mid B_1)P(B_1)}{P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + P(A \mid B_3)P(B_3)}$
- $P(A) = P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + P(A \mid B_3)P(B_3)$
- $P(B_1 \cap A) = P(A \cap B_1) = P(A \mid B_1)P(B_1)$

**Example**

Say I have a wallet that contains either a $5 bill or a $20 bill (with equal probability), but I don’t know which one. I add a $5 bill. Later, I reach into my wallet (without looking) and remove a bill. It’s a $5 bill. What is the probability that the bill left in the wallet is a $5 bill?

### Diagram

- Start
  - $1/2$ $1/2$ $5$ $20$

- Add
  - $2 \times 5$ $5, 20$

- Extract
  - $1/2$ $1/2$ $1/2$ $0$
  - $5$ $20$ $5$ $20$
• \( P(\text{start with } $5 \mid \text{extract a } $5) = \frac{1 \cdot 1}{1 \cdot 1 + 2 \cdot 1} = \frac{2}{3} \)

Example

A lab blood test is 95% effective in detecting a certain disease when it is in fact, present. However the test also yields a “false positive” result for 1% of the healthy people tested. If 0.5% of the population has the disease, what is the probability that a person who was tested positive actually has the disease?

\[
\begin{array}{c|c|c}
\text{Initial Condition} & 0.005 & 0.995 \\
\hline
\text{sick} & 0.95 & 0.05 \\
\text{healthy} & 0.01 & 0.95 \\
\end{array}
\]

\[ P(\text{sick} \mid \text{positive}) = \frac{P(\text{sick} \cap \text{positive})}{P(\text{positive})} = \frac{P(\text{positive} \mid \text{sick})P(\text{sick})}{P(\text{positive})} \]

\[ P(\text{positive}) = 0.005 \times 0.95 + 0.995 \times 0.01 \]

\[ P(\text{sick} \cap \text{positive}) = 0.005 \times 0.95 \]

\[ P(\text{sick} \mid \text{positive}) = \frac{0.005 \times 0.95}{0.005 \times 0.95 + 0.995 \times 0.01} = 32.31\% \]

Lecture #13 – Wednesday, October 22, 2003

INDEPENDENCE

• Two events, \( A \) and \( B \) are independent if \( P(A \cap B) = P(A) \cdot P(B) \).

• In general, \( P(A \cap B) = P(A \mid B) \cdot P(B) \).
  • If \( A \) and \( B \) are independent, \( P(A \mid B) = P(A) \) – The fact that \( B \) happened doesn’t change the chance of the occurrence of \( A \).

• In general, \( \{A_1, A_2, ..., A_n\} \) are independent if any two \( A_i, A_j \) are independent.

Example

An urn contains five red and seven blue balls. Suppose that two balls are selected at random and with replacement. Let \( A \) and \( B \) be the events that the first and the second balls are red, respectively. Check whether \( A \) and \( B \) are independent or not. Redo the calculation for the case of random selection without replacement.

• With replacement:
  • \( P(A \cap B) = \frac{5 \times 5}{12 \times 12} \).
• \( P(A) = \frac{5}{12}, \ P(B) = \frac{5}{12} \).
• So \( A, B \) are independent \( (A \perp B) \).

• Without replacement:
  \[
  P(A \cap B) = \frac{5 \times 4}{12 \times 11} = \frac{20}{12 \times 11}
  \]
  \[
  P(A) = \frac{5}{12}, \ P(B) = P(B \mid A)P(A) + P(B \mid A^c)P(A^c) = \frac{4}{11} \cdot \frac{5}{12} + \frac{5}{11} \cdot \frac{7}{12} = \frac{55}{11 \times 12} = \frac{5}{12}.
  \]
• So \( A \) and \( B \) are not independent.

**Exercises**

1) Show that if \( E \) and \( F \) are independent then \( E \) and \( F^c \) are independent. What about then \( E \) and \( F^c \)?
   • I want to show \( P(E \cap F^c) = P(E) \cdot P(F^c) \).
   \[
   P(E \cap F^c) = P(E \cap F^c) + P(E \cap F) - P(E \cap F)
   \]
   \[
   = P(E) - P(E \cap F)
   \]
   \[
   = P(E) - P(E) \cdot P(F)
   \]
   \[
   = P(E) \cdot (1 - P(F))
   \]
   \[
   = P(E) \cdot P(F^c)
   \]
   • \( E \perp F \Rightarrow E \perp F^c \Rightarrow E^c \perp F^c \)

2) Show that if \( E \) and \( F \) are mutually exclusive with \( P(E) > 0 \) and \( P(F) > 0 \), then \( E, F \) cannot be independent.
   • \( P(E \cap F) = 0 \).
   • If \( E \) and \( F \) are independent, \( P(E \cap F) = P(E) \cdot P(F) > 0 \).
   • So \( A \) and \( B \) are not independent if they are mutually exclusive.

3) True or False: If \( E \) and \( F \) are independent and \( E \) and \( G \) are independent, then \( E \) and \( F \cap G \) are independent.
   • False!

4) True or False: In the case of 3 events \( E, F, G \), the equality \( P(E \cap F \cap G) = P(E) \cdot P(F) \cdot P(G) \) doesn’t imply that \( E, F, G \) are independent.
   • True!

**Probability as a Continuous Function**

• \( f : \mathbb{R} \rightarrow \mathbb{R} \).
• \( x_n \xrightarrow{n \to \infty} x \).
• \( f \) is continuous in \( x \) iff \( \lim_{x \to x_0} f(x_n) = f(x) \).
• \( P : \) sets from \( S \rightarrow [0,1] \).
Increasing Sequence of Events
- \( \{E_1, E_2, \ldots, E_n\} \) infinite sequence is increasing. \( E_k \subset E_{k+1}, \forall k \geq 1. \)

![Diagram of increasing sequence of events](image)

- The “limit” is the \( \bigcup_{k=1}^{\infty} E_k. \)

Decreasing Sequence of Events
- \( \{E_1, E_2, \ldots, E_n\} \) infinite sequence is decreasing. \( E_k \supseteq E_{k+1}, \forall k \geq 1. \)

![Diagram of decreasing sequence of events](image)

- The “limit” is the \( \bigcap_{k=1}^{\infty} E_k. \)

- \( \{E_n\} \) is decreasing or increasing. I can define \( \lim_{n \to \infty} E_n = \bigcap_{n=1}^{\infty} E_n \).

- In both cases, \( \lim_{n \to \infty} P(E_n) = P\left( \lim_{n \to \infty} E_n \right) \).

Lecture #14 – Friday, October 24, 2003 10 26
- If sets are increasing, then \( \lim_{n \to \infty} E_n = \bigcup_{n=1}^{\infty} E_n \) – the smallest set containing all the sets.

- If sets are increasing, then \( \lim_{n \to \infty} E_n = \bigcap_{n=1}^{\infty} E_n \) – the smallest set containing all the sets.

Example

If the probability that the entire population will die before having offspring in the \( n \)th is \( e^{\frac{2n^3}{6n^4+3}} \), what is the probability that it will survive forever?
• Let $E_n = \{\text{population wiped out by the } n^{th} \text{ generation}\}$.
• So $E_1 \subset E_2 \subset \ldots \subset E_n \subset E_{n+1}$.

\[
P(\text{survives forever}) = 1 - P(\text{dies out sometime in the future}) = 1 - P\left(\bigcup_{n=1}^{\infty} E_n\right) = 1 - \lim_{n \to \infty} P(E_n)
\]

\[
P(E_n) = \exp\left[-\frac{2n^2}{6n^2 + 3}\right]
\]

\[
\lim_{n \to \infty} P(E_n) = \lim_{n \to \infty} \exp\left[-\frac{2n^2}{6n^2 + 3}\right] = \exp\left[-\frac{2}{6}\right] = e^{-\frac{1}{3}}
\]

**Lecture #15 – Monday, October 27, 2003**

**RANDOM VARIABLES**

• In many situations when an experiment is performed the interest is in some numerical function of the outcome rather than the actual outcome itself.
• If $S$ is the sample space of an experiment then a map $X: S \to \mathbb{R}$ is called a random variable.
• Additional Requirement: For any interval $I \subset \mathbb{R}$, $X^{-1}(I)$ is an event in $S$.

**Example**

• Tomorrow: rain, snow, shine.
• I bet $5 that it’s rain.
• $X = \text{my profit} = \{-$5, $5\}$
• $P(X = $5)$
• $Y = \text{profit from betting $100 on snow}$

• If $S$ is a discrete sample space, then $X: S \to \mathbb{R}$ is a discrete random variable. The set of possible values of $X$ is also discrete – $X: S \to \{a_1, a_2, \ldots, a_n\}$
• Some $S$’s may not be discrete.

**Example**

Suppose that three cards are drawn from an ordinary deck of 52 cards one-by-one at random and with replacement. Let $X$ be the number of spades drawn. Find $P(X = i), i = 0, 1, 2, 3$.

• Possible values for random variable $X$ are $\{0, 1, 2, 3\}$. $X: S \to \{0, 1, 2, 3\}$. 
Random Selection of Points In Intervals

- Fix \( a < b \) and \( \alpha, \beta \) such that \( a \leq \alpha < \beta \leq b \). The probability that a point is randomly selected in the interval \((\alpha, \beta)\) is \( \frac{\beta - \alpha}{b - a} \).

- Sample space is not discrete but continuous (interval).

- Let \( C \) be a fixed point in the interval \((a, b)\). If \( X \) is a point randomly selected in the interval \((a, b)\) then the probability that \( X \) is selected to be exactly \( C \) is \( P(X = C) = 0 \).

- If \( (a, b) \) was discrete \( \{a_1, a_2, \ldots, a_n\} \), then \( P(X = a_i) = \frac{1}{n} \).

- Define \( E_n = \left( C - \frac{1}{n}, C + \frac{1}{n} \right) \).

- \( E_n \Rightarrow E_{n+1} \), so \( E_{n+1} = \left( C - \frac{1}{n+1}, C + \frac{1}{n+1} \right) \). Sequence of events \( E_n = \left\{ C - \frac{1}{n}, C + \frac{1}{n} \right\} \) which is decreasing.

- \( P(X = C) = P(X \in E_1 \text{ and...and } X \in E_n) = P\left( \bigcap_{n=1}^\infty X \in E_n \right) = \lim_{n \to \infty} P(X \in E_n) = \lim_{n \to \infty} \frac{2}{b - a} = 0 \)

Example

A train passes through a town at random time between 10:00 am and 10:12 pm. If I drive through town between 9:55 am and 10:05 am, what is the probability that I see the train?

- \((a, b) = (10:00 \text{ am}, 10:12 \text{ am})\)

- \( P(\text{I see the train}) = P(\text{train goes through between 10:00 and 10:05}) = \frac{5}{12} \) — does not depend on the scale of measurement.
Example
A patient with flu may have a fever between 39°C and 42°C. Let X be the temperature of a randomly selected flu patient. What is the probability that the temperature is less than 40°C?

- \[ S = \{ \text{all possible temperatures} \} = \{39°, 40°\} \]
- \[ (a, b) = (39, 42), \ (\alpha, \beta) = (39, 40) \]
- \[ P(X \in (39, 40)) = \frac{40 - 39}{42 - 39} = \frac{1}{3} \]
- \[ P(X = 39.5) = 0 \]
- \[ P(X \in (39.49, 39.51)) = \frac{0.02}{3} \]

- For continuous random variable, we look at \( P(\text{interval}) \) instead of \( P(\text{point}) \).

Definition
- For a discrete random variable \( X : S \rightarrow \{a_1, a_2, ..., a_n\} \), we can define a function \( P : \{a_1, a_2, ..., a_n\} \rightarrow [0, 1] \) such that \( P(a_i) = P(X = a_i) \).
- \( P \) is called probability function of the random variable \( X \).

Lecture #16 – Wednesday, October 29, 2003

Distribution Functions
- The distribution function \( F : \mathbb{R} \rightarrow [0, 1] \) of a random variable \( X \) is defined as \( F(t) = P(X \leq t) \).
- For example, \( F(1) = P(X \leq 1) \).

Properties of \( F \)
1) \( F \) is non-decreasing – i.e. \( t_1 \leq t_2 \Rightarrow F(t_1) \leq F(t_2) \). So \( F \) is increasing or constant.
2) \( \lim_{t \to \infty} F(t) = 1 \)
3) \( \lim_{t \to -\infty} F(t) = 0 \)
4) \( F \) is continuous to the right.

Proof of (1)
\[ t_1 \leq t_2 \] I want to show \( F(t_1) \leq F(t_2) \)

- \[ F(t_1) = P(X \leq t_1), \ F(t_2) = P(X \leq t_2) \]
- \( E = \{X \leq t_1\}, \ F = \{X \leq t_1\} \)
- \( E \subset F \Rightarrow P(E) \leq P(F) \)

- \( P(E) = P(X \leq t_1) = P(t_1) \)
- \( P(F) = P(X \leq t_2) = P(t_2) \) \( \Rightarrow F(t_1) \leq F(t_2) \)

- Attention: \( t_1 < t_2 \) does not imply \( F(t_1) < F(t_2) \); \( t_1 < t_2 \) implies \( F(t_1) \leq F(t_2) \)
**Proof of (2)**

- $X : S \rightarrow \mathbb{R}$ – real random variable.
- Always, there will be $t$ large enough to have $P(X \leq t)$ is almost 1.
- Let $E_n = \{X \leq n\}, n \in \mathbb{Z}$.
- $\lim_{t \to \infty} F(t) = \lim_{n \to \infty} F(n) = \lim_{n \to \infty} P(n)$.
- $E_n \subseteq E_{n+1}$ implies the sequence $E_n$ is increasing. I can use the continuity property of the probability function.
- $\lim_{n \to \infty} P(E_n) = P\left(\lim_{n \to \infty} E_n\right) = P\left(\bigcup_{n=1}^{\infty} E_n\right) = P\left(\bigcup_{n=1}^{\infty} X \leq n\right) = 1$

**Proof of (3)**

- The proof of $\lim_{t \to -\infty} F(t) = 0$ is the mirror of the proof of $\lim_{t \to \infty} F(t) = 1$.

**Proof of (4)**

- Convergence from the left means for all $x_n \leq x_0$ and $x_n \to x_0$. In fact, for all practical purposes, one can consider only increasing sequences. $x_n < x_{n+1} \forall n \in \mathbb{Z}$ and $x_n \leq x_0$ and $\lim_{n \to \infty} x_n = 0$.
- Notation: $x_n \uparrow x_0$ – converges to $x_0$ from the left. $\lim_{x_n \uparrow x_0} f(x_n) = f(x_0 -)$.
- If $f(x_0 -) = f(x_0)$, then $f$ is continuous to the left.
- Continuity to the right.
- $x_n \geq x_{n+1}$
- $x_n \geq x_0$ \(\forall n \in \mathbb{Z}\)

- The limit to the right is denoted $x_n \downarrow x_0$ – $\lim_{x_n \downarrow x_0} f(x_n) = f(x_0 +)$.
- If $f(x_0 +) = f(x_0)$, then $f$ is called continuous to the right.

- $f(x_0) = b$
- $f(x_0 -) = \lim_{x_n \downarrow x_0} f(x_n) = a$
- $f(x_0 +) = \lim_{x_n \downarrow x_0} f(y_n) = b$
- $f(x_0) = f(x_0 +)$ \(\Rightarrow f\) is continuous to the right, but $f$ is not continuous to the left.
Generic $F(t)$

- Multiple jumps (or none).

**Example**

Suppose we flip twice a coin that has probability to land heads equal to 0.4. Let $X$ be the number of tails. Calculate $F(t)$, the distribution function of $X$.

- $X : S \to \{0, 1, 2\}$
- $P(X = 0) + P(X = 1) + P(X = 2) = 1$
  - $P(X = 0) = 0.4 \cdot 0.4 = 0.16$
  - $P(X = 0) = 0.4 \cdot 0.6 + 0.6 \cdot 0.4 = 0.48$
- Probability function of $X$.
  - $P(X = 2) = 0.36$

- $F(t) = P(X \leq t)$ for any $t \in \mathbb{R}$:
  - If $t < 0$, $P(X \leq t) = F(0) = 0$.
  - If $t = 0$, $P(X \leq 0) = P(X = 0) = 0.16$.
  - If $t \in (0, 1)$, $P(X \leq t) = 0.16$.
  - If $t = 1$, $P(X \leq 1) = P(X = 0) + P(X = 1) = 0.64$.
  - If $t \in (1, 2)$, $P(X \leq t) = 0.64$.
  - If $t \geq 2$, $P(X \leq t) = 1$.

- Sometimes $F$ called cumulative distributive function.
- Given a probability function for a discrete random variable $X$, one should be able to construct the distribution function for $X$.

**Lecture #17 – Friday, October 31, 2003**

**Example**

From 18 potential women jurors and 28 potential men jurors, a jury of 12 is chose at random. Let $X$ be the number of women selected. Find the probability function of $X$.

- $S = \text{set of all possible jurors}$
- $X : S \to \{0, 1, \ldots, 12\}$
• Probability function of $X$ is $P : \{0,1,2,\ldots,12\} \rightarrow [0,1]$

  \[ \begin{pmatrix} 18 & 28 \\ 0 & 12 \end{pmatrix} \]

  $P(0) = P(X = 0) = \begin{pmatrix} 46 \\ 12 \end{pmatrix} = 0.000782$

  \[ \begin{pmatrix} 18 & 28 \\ 1 & 11 \end{pmatrix} \]

  $P(1) = P(X = 1) = \begin{pmatrix} 46 \\ 12 \end{pmatrix} = 0.001$

  \[ \begin{pmatrix} 18 & 28 \\ 4 & 8 \end{pmatrix} \]

  $P(4) = P(X = 4) = \begin{pmatrix} 46 \\ 12 \end{pmatrix} = 0.244$

### Connection Between Probability and Distribution

<table>
<thead>
<tr>
<th>Event concerning $X$</th>
<th>Probability of the event in terms of $F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X \leq a$</td>
<td>$F(a)$</td>
</tr>
<tr>
<td>$X &gt; a$</td>
<td>$1 - F(a)$</td>
</tr>
<tr>
<td>$X \geq a$</td>
<td>$1 - F(a -)$</td>
</tr>
<tr>
<td>$X = a$</td>
<td>$F(a) - F(a -)$</td>
</tr>
<tr>
<td>$a &lt; X \leq b$</td>
<td>$F(b) - F(a)$</td>
</tr>
<tr>
<td>$a &lt; X &lt; b$</td>
<td>$F(b -) - F(a)$</td>
</tr>
<tr>
<td>$a \leq X \leq b$</td>
<td>$F(b -) - F(a -)$</td>
</tr>
<tr>
<td>$a \leq X &lt; b$</td>
<td>$F(b -) - F(a -)$</td>
</tr>
</tbody>
</table>

- $P(a < X \leq b) = P([X \leq b \text{ but not } \{X \leq a\}]) = P([X \leq b] - \{X \leq a\})$
- $P(X \in [a,b]) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$
- $P(a \leq X < b) = P(X \in [a,b]) = P([X \leq b \text{ but not } \{X \leq a\}]) = P([X \leq b] - \{X \leq a\}$
- $P(a \leq X < b) = P(X \leq b) - P(X \leq a) = F(b -) - F(a -)$

### Example

The distribution function of a random variable $X$ is given by:

\[
F(X) = \begin{cases} 
0 & \text{if } x < -2 \\
\frac{1}{2} & \text{if } -2 \leq x < -2 \\
\frac{3}{4} & \text{if } 2 \leq x < 4 \\
\frac{8}{9} & \text{if } 4 \leq x < 6 \\
1 & \text{if } 6 \leq x 
\end{cases}
\]

Calculate the probability function of $X$. 
· $P(X = -3) = 0$. Why?
  · Take a sequence that converges to -3 from the left.
  · $P(X = -3) = P(X \leq -3) - P(X < -3) = F(-3) - F((-3)-) = F(-3) - \lim_{x_n \to -3} F(X_n) = 0 - 0 = 0$.

· The set of all possible values that $X$ can take is given, in general, by those points $t$ such that $F(t-) \neq F(t+) = F(t)$.
· In this example the set is $\{-2, 2, 4, 6\}$.
  · $P(X = -2) = F(-2) - F((-2)-) = \frac{1}{2}$.
  · $P(X = 2) = F(2) - F(2-) = \frac{3}{5} - \frac{1}{2} = \frac{1}{10}$.
  · $P(X = 4) = \frac{8}{9} - \frac{3}{5} = \frac{13}{45}$.
  · $P(X = 6) = \frac{1}{9}$.

**Summary**
· Both probability function and distribution function characterize completely the random variable $X$.
· Given the probability function $P$, we can construct the distribution function $F$ and vice-versa.

---

**Lecture #18 – Monday, November 3, 2003**

**EXPECTATION OF A DISCRETE RANDOM VARIABLE**

**Example**

If I flip a coin (fair: $P(T) = P(H) = \frac{1}{2}$). If T, I gain $1. If H, I lose $1.

· $X = \{\text{my gain}\}$
· $X : \{H, T\} \to \{-1, 1\}$
· If I repeat many times, the expected gain is $0 - P(T) \cdot 1 + P(H) \cdot (-1) = \frac{1}{2} - \frac{1}{2} = 0$. 
Definition

- \( X : S \rightarrow \{ a_1, a_2, \ldots, a_n \} \).
- The expected value of the discrete random variable \( X \) is defined \( E[X] = \sum_{i=1}^{\infty} P(X = a_i) \cdot a_i = \sum_{i=1}^{\infty} P(a_i) \cdot a_i \).

Properties of \( E[X] \)

1) If \( X \) is constant, \( \mathbb{E}[X] = c \cdot P(X = c) = c \cdot 1 = c \).

2) \( \mathbb{E}[X + c] = \sum_{i=1}^{\infty} P(a_i) \cdot (a_i + c) \).
   - Define \( Y = X + c \).
   - \( Y \) is also discrete and every time \( X \) takes value \( a_i \), \( Y \) takes value \( a_i + c \).
   \[ E[X + c] = E[Y] = \sum_{i=1}^{\infty} (a_i + c) \cdot P(Y = a_i + c) = \sum_{i=1}^{\infty} (a_i + c) \cdot P(X = a_i) = \sum_{i=1}^{\infty} (a_i + c) \cdot P(a_i) \]
   \[ = \sum_{i=1}^{\infty} (a_i) \cdot P(a_i) + c \sum_{i=1}^{\infty} P(a_i) = E[X] + c \]

3) \( \mathbb{E}[X \cdot C] = c \cdot \mathbb{E}[X] \).
   - Define \( Y = X \cdot c \).
   - If \( X = a_i \Leftrightarrow Y = c \cdot a_i \).
   \[ E[X \cdot C] = E[Y] = \sum_{i=0}^{\infty} P(Y = c \cdot a_i) \cdot (c \cdot a_i) = c \sum_{i=0}^{\infty} P(X = a_i) \cdot a_i = c \cdot E[X] \]

Example

In a certain part of downtown Toronto parking lots charge $10 per day. A car that is illegally parked on the street will be fined $20 if caught and the chance of being caught is 70%. If money is the only concern, should one park illegally or not?

- \( X = \) What I pay per day if I park legally \( \quad Y = \) What I pay per day if I park illegally.
- \( X = $10 \) is a constant variable.
- \( Y = \{0, 20\} \quad P(Y = 0) = 30\% = \frac{3}{10}, \quad P(Y = 20) = 70\% = \frac{7}{10} \).
- \( E[X] = 10 \)
- \[ E[Y] = P(Y = 0) \cdot 0 + P(Y = 20) \cdot 20 = \frac{3}{10} \cdot 0 + \frac{7}{10} \cdot 20 = 14 \].
- On average, we are better off if we park legally.

Example

In a lottery, a player pays $1 and selects four distinct numbers from 0 to 9. Then from an urn containing 10 identical balls numbered 0 to 9, four balls are drawn at random and without replacement. If the numbers of three or all four of these balls matches the player’s numbers, he wins $5 and $10, respectively. Otherwise he loses. On average, how much money does the player gain per game (gain = win – loss).

- \( X = \) gain after one round of the game.
- \( X : S \rightarrow \{-1, 4, 9\} \).
\[ P(X = 9) = \binom{4}{4} = 0.005, \quad P(X = 4) = \binom{3}{4} = 0.114, \quad P(X = -1) = 1 - P(X = 9) - P(X = 4) = 0.881. \]

- \[ E[X] = -1 \cdot 0.881 + 4 \cdot 0.114 + 9 \cdot 0.005 = -0.38 \]

- On average, he loses 38¢ per game.

**Example**

In the US the number of twin births is approximately 1 in 90. Let \( X \) be the number of births until the first twins are born. Calculate distribution function of \( X \).

- \( X : S \to \{0,1,2,\ldots,n\} \)
- \[ P(X = 0) = \frac{1}{90}, \quad P(X = 1) = \frac{89}{90} \cdot \frac{1}{90}, \quad P(X = 2) = \left( \frac{89}{90} \right)^2 \cdot \frac{1}{90}. \] So \( P(X = n) = \left( \frac{89}{90} \right)^n \cdot \frac{1}{90}. \)

- \[ \sum_{n=1}^{\infty} \frac{89}{90} = \lim_{n \to \infty} \sum_{n=1}^{n} \left( \frac{89}{90} \right)^n \]
- \[ 
\begin{align*}
\frac{89}{90} + \left( \frac{89}{90} \right)^2 + \ldots + \left( \frac{89}{90} \right)^n = R_n & \Rightarrow \left( \frac{89}{90} \right)^2 + \left( \frac{89}{90} \right)^3 + \ldots + \left( \frac{89}{90} \right)^{n+1} = \frac{89}{90} \cdot R_n \\
\Rightarrow R_n \left( 1 - \frac{89}{90} \right) = \frac{89}{90} - \left( \frac{89}{90} \right)^{n+1} & \Rightarrow R_n = \frac{\frac{89}{90} - \left( \frac{89}{90} \right)^{n+1}}{1 - \frac{89}{90}}
\end{align*}
\]
- \[ \lim_{n \to \infty} R_n = \frac{89}{90} \]
- \[ \lim_{n \to \infty} R_n = \frac{89}{90} = 89 \]

- \[ E[X] = 0 \cdot \frac{1}{90} + \left( \frac{89}{90} \right) \cdot \frac{1}{90} + \ldots + n \left( \frac{89}{90} \right)^n \cdot \frac{1}{90} = \sum_{n=1}^{\infty} \left( \frac{89}{90} \right)^n \cdot \frac{n}{90} = \frac{1}{90} \cdot 89 \sum_{n=1}^{\infty} \left( \frac{89}{90} \right)^{n-1} \cdot n = \lim_{n \to \infty} \sum_{n=1}^{\infty} \left( \frac{89}{90} \right)^{n-1} \cdot n.
\]

**Lecture #19 – Wednesday, November 5 2003**

- Sometimes we want \( E[g(x)] \) where \( g = \{a_1, a_2, \ldots, a_n, \ldots\} \to \mathbb{R} \) (a real function).
- \( g(x) = x^2 - E[X^2] \)
- \( g(x) = \log x - E[\log X] \)

**Law of the Unconscious Probabilist**

- \[ E[g(x)] = \sum_{i=1}^{\infty} g(a_i) \cdot p(a_i) \]
- \[ E[X] = \sum_{i=1}^{\infty} a_i \cdot p(a_i) \]
• If $X$ is a random variable $X : S \rightarrow \{a_1, a_2, \ldots, a_n, \ldots\}$ with probability mass function $p(\cdot)$ and $g$ is a map $g : \{a_1, a_2, \ldots, a_n, \ldots\} \rightarrow \mathbb{R}$ then $E[g(X)] = \sum_{i=1}^{\infty} g(a_i) \cdot p(a_i)$.

**Example: 3 Gamblers**

1) Risk-lover: Flips a coin. If T, wins $100; if H, loses $100. ($X_1$).
2) Conservative-gambler: If T, wins $1; if H, loses $1. ($X_2$).
3) Boring: Does not gamble. ($X_3$).

• $X_i =$ gain by player $i$.

$E[X_1] = 100 \cdot \frac{1}{2} - 100 \cdot \frac{1}{2} = 0$

$E[X_2] = 1 \cdot \frac{1}{2} - 1 \cdot \frac{1}{2} = 0$

$E[X_3] = 0 \cdot \frac{1}{2} - 0 \cdot \frac{1}{2} = 0$

**VARIANCE**

• Another interesting quantity we can look at is, beside the center, the spread of a distribution. How far apart the probability function “spikes” are.

• Formally, this is measured using the variance of a random variable. It is denoted $\text{Var}(X)$.

• $\text{Var}(X) = E[(X - E(X))^2]$

• Let $E[X] = \mu$.

• $\text{Var}(X) = E[(X - E(X))^2] = E[X^2 - 2\mu X + \mu^2] = E[X^2] - 2E[\mu X] + E[\mu^2] = E[X^2] - 2\mu^2 + \mu^2$

1) $X = c$ a constant:

• $\text{Var}(X) = E[(c - E(c))^2] = 0$.

• The variance is never negative.

2) $\text{Var}(X + c) = \text{Var}(X)$

• $\text{Var}(X + c) = E[((X + c) - E[X + c])^2] = E[(X - E[X])^2] = \text{Var}(X)$.

• Spread remains the same.

3) $\text{Var}(c \cdot X) = c^2 \text{Var}(X)$
\[
\text{Var}(c \cdot X) = E[(c \cdot X)^2] - (E[c \cdot X])^2 = E[c^2 \cdot X^2] - (c \cdot E[X])^2 = c^2 \cdot E[X^2] - c^2 \cdot (E[X])^2 = c^2 \cdot \text{Var}(X)
\]

**Example**

The distribution function of a random variable \( X \) is given by:

\[
F(x) = \begin{cases} 
0 & \text{if } x < -3 \\
\frac{3}{8} & \text{if } -3 \leq x < 0 \\
\frac{1}{2} & \text{if } 0 \leq x < 3 \\
\frac{3}{4} & \text{if } 3 \leq x < 4 \\
1 & \text{if } x \geq 4
\end{cases}
\]

Calculate \( E[X] \), \( E[X^2 - 2|X|] \), and \( \text{Var}[X \mid |X|] \).

- \( \lim_{t \to -\infty} F(t) = 0 \)
- \( \lim_{t \to \infty} F(t) = 1 \)
- \( F \) is non-decreasing
- \( F \) is continuous from the right

The values \( X \) can take are the values where \( F \) has different limits at the left and the right.

\( X : S \to \{-3, 0, 3, 4\} \).

- \( P(X = -3) = \) difference between limit to the right and limit to the left = \( \frac{3}{8} \).
- \( P(X = 0) = \frac{1}{2} - \frac{3}{8} = \frac{1}{8} \).
- \( P(X = 3) = \frac{3}{4} - \frac{1}{2} = \frac{1}{4} \).
- \( P(X = 4) = 1 - \frac{3}{4} = \frac{1}{4} \).

\[
E[X] = \frac{-3 \cdot 3}{8} + \frac{0 \cdot 1}{8} + \frac{3 \cdot 1}{4} + \frac{4 \cdot 1}{4} = -\frac{5}{8}.
\]

\[
E[X^2 - 2|X|] = 3 \cdot \frac{3}{8} + 0 \cdot \frac{1}{8} + 3 \cdot \frac{1}{4} + 4 \cdot \frac{1}{4} = \frac{31}{8}
\]
\[ \text{Var}[X \mid X] = \text{always boils down to computing some expectations} = E[(X - E[X])^2] - (E[X \mid X])^2 \]

\[ h(r) = (r - \mu)^2, \quad h : \mathbb{R} \rightarrow \mathbb{R} \]

\[ E[h(X)] = \sum h(a_i) \cdot p(a_i) \]

\[ h(3) = 81, \quad h(0) = 0, \quad h(3) = 81, \quad h(4) = 256 \]

\[ E[(X - E[X])^2] = 81 \cdot \frac{3}{8} + 0 \cdot \frac{1}{8} + 81 \cdot \frac{1}{4} + 256 \cdot \frac{1}{4} = \frac{755}{8} \]

\[ c(v) = v \cdot \mid v \mid \]

\[ v(-3) = -9, \quad v(0) = 0, \quad v(3) = 9, \quad g(4) = 16 \]

\[ E[X \mid X] = -9 \cdot \frac{3}{8} + 0 \cdot \frac{1}{8} + 9 \cdot \frac{1}{4} + 16 \cdot \frac{1}{4} = \frac{23}{8} \]

\[ \text{Var}[X \mid X] = \frac{755}{8} - \left( \frac{23}{8} \right)^2 = \frac{5511}{64} \]