Scientific Determinism

Scientific determinism holds that we can exactly predict how a system will develop over time if we know the laws governing the system and the state of the system at some time. For example, the early successes of Newtonian mechanics led P.-S. Laplace to theorize that the entire future and history of the universe could be determined exactly by an entity (Laplace’s demon) that knew the current positions and momenta of all particles.

\[
m \frac{d^2x}{dt^2} = F(x(t))
\]

\[
x(0) = x_0
\]

\[
x'(0) = v_0.
\]

In a similar vein, Einstein famously wrote that “God doesn’t play dice with the universe.”
Multiple Source of Uncertainty

- Imprecise knowledge of the state of a system (measurement error).
- Imprecise knowledge of model parameters.
- Model uncertainty: “All models are wrong, but some are useful.” (George Box)
- Numerical solutions are usually approximate due to round-off and truncation errors.
- Some models are ill-posed in the sense that they either don’t have classical solutions or they have multiple solutions.
- Some models (chaotic dynamical systems) are known to be exquisitely sensitive to their initial conditions.
- According to some interpretations, quantum mechanics provides an intrinsically stochastic description of nature.

Advances in computing have helped to shift attention to problems involving large numbers of variables (“big data”) where uncertainty and error cannot be neglected.
Motivation

Probabilistic Forecasting of Elections

Figures show Nate Silver’s state-by-state predictions of the outcome of the 2012 US presidential elections, along with the \( p \)-values for each state. (Source: A. C. Thomas)
Moneyball: Probabilistic Modeling Applied to Baseball
Probabilistic Forecasting of Storm Trajectories

Tropical Storm Force Wind Speed Probabilities
For the 120 hours (5 days) from 8 AM EDT Tue Aug 23 to 8 AM EDT Sun Aug 28

Probability of tropical storm force surface winds (1-minute average >= 39 mph) from all tropical cyclones

Hurricane Irene forecast, August 23, 2011 (NOAA)
Probabilistic Forecasting of Climate Change

Source: IPCC
Motivation

Chance and Contingency Influence Biological Processes at all Scales

Stochastic X-inactivation produces the mottled coat pattern on tortoiseshell cats.

Mass extinctions have had profound and often unpredictable consequences for life on earth. (Figure: Sepkoski 1984)
Uncertainty and Mathematics

Different mathematical frameworks have been proposed for uncertainty quantification:

- interval analysis: replace numbers by intervals
- fuzzy set theory and fuzzy logic
- Dempster-Shafer theory: plausibility is represented by belief functions
- probability theory

Of these approaches, probability theory is by far the most widely-accepted. What it means, however, is less clear.

Bertrand Russell (apocryphal): “Probability is among the most important sciences, not least because no one understands it.”
Multiple Interpretations of Probability

- **Frequentist interpretation** (J. Venn): The probability of an event is equal to its limiting frequency in an infinite series of independent, identical trials.

- **Propensity interpretation** (C. Pierce, K. Popper): The probability of an event is equal to the physical propensity of the event.

- **Logical interpretation** (J. M. Keynes, R. Cox, E. Jaynes): Probabilities quantify the degree of support that some evidence $E$ gives to a hypothesis $H$. Given the same evidence, rational individuals should assign the same probability to $H$.

- **Subjective interpretation** (F. Ramsey, J. von Neumann, B. de Finetti): The probability of a proposition is equal to the strength of one’s personal belief that the proposition is true. Rational individuals may assign different probabilities to the same proposition, even if they share the same evidence.

The frequentist and propensity interpretations of probability are *physical theories*: they regard the probability of an event as a unique quantity that is determined by physical reality.
Interlude: A Brief Review of Set Theory

We recall the following notation. Let $A$ and $B$ be subsets of $S$.

- $A \cup B$ will be used to denote the union of $A$ and $B$, i.e., the set that contains all of the elements that are found in either $A$ or $B$ or both $A$ and $B$. Similarly, $\bigcup_{i=1}^{n} A_i$ will denote the union of the $n$ sets $A_1, \ldots, A_n$. If $A$ and $B$ are events, then $A \cup B$ is the event that either $A$ or $B$ or both $A$ and $B$ occurs.

- $A \cap B$ or $AB$ will be used to denote the intersection of $A$ and $B$, i.e., the set that contains all of the elements that are found in both $A$ and $B$. Similarly, $\bigcap_{i=1}^{n} A_i$ or $\prod_{i=1}^{n} A_i$ will denote the intersection of all $n$ sets. If $A$ and $B$ are events, then $A \cap B$ is the event that both $A$ and $B$ occur.

- $\bar{A}$ or $S \setminus A$ will be used to denote the complement of $A$ in $S$, i.e., the set that contains all of the elements that are found in $S$ but not $A$. If $A$ is an event, then $\bar{A}$ is the event that $A$ does not occur.
Unions and intersections satisfy the following algebraic laws:

- **Commutativity:**
  \[ A \cup B = B \cup A \quad A \cap B = B \cap A \]

- **Associativity:**
  \[ A \cup (B \cup C) = (A \cup B) \cup C \quad A \cap (B \cap C) = (A \cap B) \cap C \]

- **Distributive laws:**
  \[ A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \]
DeMorgan’s Laws

Suppose that $A_1, A_2, \cdots, A_n$ are subsets of some common set $S$. Then DeMorgan’s laws assert that:

\[
\bigcup_{i=1}^{n} A_i = \bigcap_{i=1}^{n} \overline{A_i}
\]

\[
\bigcap_{i=1}^{n} A_i = \bigcup_{i=1}^{n} \overline{A_i}.
\]

In words, the complement of a union is the intersection of the complements, while the complement of an intersection is the union of the complements. These identities will be used repeatedly in this class.
A Behavioristic Derivation of the Laws of Probability

We consider a two-player game, with players $P_1$ and $P_2$, that place bets on an event with a then-unknown outcome. Here we will assume that the event can result in one of $n$ mutually-exclusive and exhaustive outcomes in the set $S = \{e_1, \cdots, e_n\}$. The game takes place in three steps.

1. First, player $P_1$ assigns a unit price $Pr(A)$ to each possible event $A \subset S$. This is the amount that $P_1$ is willing to pay for a $1$ bet on the event that $A$ happens, as well as the amount that $P_1$ is willing to accept from $P_2$ in exchange for such a bet. The player selling such a wager agrees to pay the purchaser $1$ if $A$ occurs.

2. Player $P_2$ then decides the amount $W(A)$ that they wish to bet on each event $A$. If $W(A)$ is positive, then $P_2$ purchases a bet worth $W$ for $W(A) \cdot Pr(A)$ from $P_1$, while if $W(A)$ is negative, then $P_2$ sells such a bet to $P_1$ for the same amount.

3. The outcome of the event is then determined and the two players settle their wagers.
Example

For concreteness, assign prices to $1 bets on the following events:

- $A = \text{‘it will rain in Tempe tomorrow’}$
- $B = \text{‘it will not rain in Tempe tomorrow’}$
- $C = \text{‘it will rain in Tempe tomorrow but not the next day’}$
- $D = \text{‘it will rain in Tempe both tomorrow and the next day’}$

Keep in mind that these are the prices at which you would be willing to either purchase or sell such a bet.
Although player $P_1$ may assign any set of prices to the various possible wagers, many assignments can be considered irrational insofar as they will lead to certain loss for $P_1$. Indeed, a set of prices $\{Pr(A) : A \subseteq S\}$ and a set of wagers $\{W(A) : A \subseteq S\}$ is said to be a Dutch book if these guarantee that $P_2$ will earn a profit from $P_1$ no matter which outcome occurs. The following are examples of Dutch books.

**Case 1:** Suppose that $Pr(A) < 0$ for some event $A$. In this case, $P_2$ can agree to purchase a $1$ wager from $P_1$ at a cost of $Pr(A)$, i.e., $P_1$ will pay $P_2$ an amount $|Pr(A)|$ to place such a bet. Then $P_2$ will have earned $1 + |Pr(A)|$ if the event $A$ occurs and will have earned $|Pr(A)|$ if the event $A$ does not occur. Thus $P_2$ makes a profit no matter what occurs and thus this is a Dutch book.

**Case 2:** Suppose that $Pr(S) < 1$. In this case, $P_2$ can earn a profit of $1 − Pr(S) > 0$ by placing a $1$ bet on $S$. Since one of the events $e_1, \ldots, e_n$ is certain to happen, it follows that $S$ itself is certain occur so that $P_2$ is certain to make a profit of $1 − Pr(S) > 0$. Similarly, if $Pr(S) > 1$, then $P_2$ should sell a $1$ bet on $S$ to $P_1$ and this will guarantee that $P_2$ will earn a profit of $Pr(S) − 1 > 0$. 
Case 3: Now suppose that \( A \) and \( B \) are disjoint subsets of \( S \), i.e., \( AB \equiv A \cap B = \emptyset \), and that \( Pr(A \cup B) < Pr(A) + Pr(B) \). Assume that \( P_2 \) purchases from \( P_1 \) a $1 wager on the event \( A \cup B \) and sells to \( P_1 \) both a $1 wager on \( A \) and a $1 wager on \( B \). The profits and losses incurred by \( P_2 \) are summarized in the following table:

<table>
<thead>
<tr>
<th>( A )</th>
<th>( B )</th>
<th>( A \cup B )</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>( Pr(A) + Pr(B) - Pr(A \cup B) )</td>
</tr>
<tr>
<td>( \bar{A} \bar{B} )</td>
<td>( Pr(A) - 1 )</td>
<td>( Pr(B) )</td>
<td>( 1 - Pr(A \cup B) )</td>
</tr>
<tr>
<td>( \bar{A}B )</td>
<td>( Pr(A) )</td>
<td>( Pr(B) - 1 )</td>
<td>( 1 - Pr(A \cup B) )</td>
</tr>
<tr>
<td>( A \bar{B} )</td>
<td>( Pr(A) )</td>
<td>( Pr(B) )</td>
<td>( -Pr(A \cup B) )</td>
</tr>
</tbody>
</table>

Notice that we do not consider the outcome \( AB \) since we are assuming that \( A \) and \( B \) represent mutually exclusive events. It follows that \( P_2 \) earns a profit no matter which outcome occurs and so \( P_2 \) has made Dutch book against \( P_1 \). Similarly, if \( Pr(A \cup B) > Pr(A) + Pr(B) \), then \( P_2 \) can make a Dutch book by selling a $1 wager on \( A \cup B \) and purchasing a $1 wager on each event \( A \) and \( B \).
The above examples show that a Dutch book can be made against $P_1$ whenever the prices assigned by $P_1$ to events in $S$ satisfy certain inequalities. This motivates the following definition. The set of prices $\{Pr(A) : A \subset S\}$ is said to be **coherent** if it satisfies the following conditions:

1. $Pr(A) \geq 0$ for each $A \subset S$;
2. $Pr(S) = 1$;
3. $Pr(A \cup B) = Pr(A) + Pr(B)$ whenever $A$ and $B$ are mutually exclusive events.

Thus our previous examples show that any set of prices that is not coherent is vulnerable to a Dutch book. In fact, the converse is also true: if the prices assigned by $P_1$ are coherent, then $P_2$ cannot make a Dutch book under the rules of this game. We will prove this result next week.
Condition (3) implies that a coherent price assignment is **finitely-additive**. If \( A_1, \cdots, A_m \) are mutually exclusive, i.e., disjoint subsets of \( S \), then

\[
Pr \left( \bigcup_{i=1}^{m} A_i \right) = \sum_{i=1}^{m} Pr(A_i).
\]

Indeed, this follows by recursion on \( m \) since we can use (3) to peel off sets \( A_m, A_{m-1}, \cdots \) one by one from the union inside the price function:

\[
Pr \left( \bigcup_{i=1}^{m} A_i \right) = Pr \left( \left( \bigcup_{i=1}^{m-1} A_i \right) \cup A_m \right) = Pr \left( \bigcup_{i=1}^{m-1} A_i \right) + Pr(A_m)
\]

\[
= Pr \left( \bigcup_{i=1}^{m-2} A_i \right) + Pr(A_{m-1}) + Pr(A_m)
\]

\[
\cdots
\]

\[
= \sum_{i=1}^{m} Pr(A_i).
\]
If we identify subjective probabilities with wagers and we accept that rational actors should assign probabilities in such a way that prevents a Dutch book from being made against them, then it follows that subjective probabilities should obey the usual axioms of frequentist probability. In this case, we will use the notation $P(A)$ and we will refer to $P(A)$ as the **probability of event** $A$. This motivates the following definition:

**Definition**

Suppose that $S = \{e_1, \cdots, e_n\}$ is a finite set and let $\mathcal{P}(S)$ denote the collection of all subsets of $S$, i.e., the power set of $S$. A **probability distribution on** $S$ **is a function** $P: \mathcal{P}(S) \rightarrow [0,1]$ **that satisfies the following three conditions:**

1. $P(A) \geq 0$ for every $A \subset S$;
2. $P(S) = 1$;
3. $P(A \cup B) = P(A) + P(B)$ whenever $A$ and $B$ are disjoint subsets of $S$.

**Remark:** The three conditions in this definition are sometimes known as the first, second and third laws of probability. The set $S$ is sometimes known as the **sample space**.
Limitations of the Behaviorist Derivation of Subjective Probability

- Our arguments in favor of coherence presume that rational individuals should act in such a way to avoid being made a sure looser. However, there may be situations in which this is not true, e.g., when individuals act altruistically and make sacrifices to benefit others.

- Our arguments also assume that it is possible for the players to eventually know what the true outcome is. In many instances where we wish to apply probabilistic reasoning, such knowledge might never be available.

- We have restricted attention to settings with only finitely-many possible outcomes. However, it is often useful to consider probabilities on infinite sets. We will return to this matter later in the course.
The laws of probability imply several useful properties:

1. $P(\bar{A}) = 1 - P(A)$;
2. $P(\emptyset) = 0$;
3. If $A \subset B$, then $P(A) \leq P(B)$;
4. $P(A \cup B) = P(A) + P(B) - P(AB)$.

**Proof of (1):** If $A$ is an event in $S$, then $S = A \cup \bar{A}$, and $A$ and its complement $\bar{A}$ are disjoint. Consequently, the finite additivity of $P$ implies that

$$1 = P(S) = P(A \cup \bar{A}) = P(A) + P(\bar{A}),$$

and rearranging the equation gives (1).

**Proof of (2):** Since $\bar{S} = \emptyset$, taking $A = S$ in (1) gives $P(\emptyset) = P(\bar{S}) = 1 - P(S) = 0$.

**Proofs of (3-4):** In-class Exercises (or see following slides).
Proof of (3): If $A \subset B$, then we can write $B$ as the disjoint union of $A$ and $B \setminus A$:

$$B = A \cup B \setminus A.$$ 

Consequently,

$$P(B) = P(A) + P(B \setminus A) \geq P(A),$$

where the inequality follows from the fact that $P(B \setminus A) \geq 0$. 

---

Jay Taylor (ASU) Fall 2013 23 / 62
Proof of (4): We first observe that \( A, B \) and \( A \cup B \) can be written as the following disjoint unions

\[
A = (A \setminus B) \cup (A \cap B)
\]
\[
B = (B \setminus A) \cup (A \cap B)
\]
\[
A \cup B = (A \setminus B) \cup (A \cap B) \cup (B \setminus A).
\]

Then, because \( \mathbb{P} \) is finitely additive, we have

\[
\mathbb{P}(A) = \mathbb{P}(A \setminus B) + \mathbb{P}(A \cap B)
\]
\[
\mathbb{P}(B) = \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B)
\]
\[
\mathbb{P}(A \cup B) = \mathbb{P}(A \setminus B) + \mathbb{P}(A \cap B) + \mathbb{P}(B \setminus A).
\]

It follows that

\[
\mathbb{P}(A) + \mathbb{P}(B) = \mathbb{P}(A \setminus B) + \mathbb{P}(A \cap B) + \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B)
\]
\[
= \mathbb{P}(A \cup B) + \mathbb{P}(A \cap B),
\]

and rearranging the terms in this identity gives (4). In effect, (4) corrects for the fact that \( \mathbb{P}(A \cap B) \) is double counted when we add \( \mathbb{P}(A) \) and \( \mathbb{P}(B) \).
**Theorem**

A probability distribution \( \mathbb{P} \) on a finite set \( S = \{e_1, \ldots, e_n\} \) is completely determined by the probabilities of the singleton sets \( \{e_i\} \).

**Proof:** We need to show that the probability of every event \( A \subset S \) can be determined from the values \( \mathbb{P}(\{e_1\}), \ldots, \mathbb{P}(\{e_n\}) \). However, because \( A \) is a disjoint union of the singleton sets contained in \( A \),

\[
A = \bigcup_{x \in A} \{x\}
\]

it follows from the finite additivity of \( \mathbb{P} \) that

\[
\mathbb{P}(A) = \sum_{x \in A} \mathbb{P}(\{x\}).
\]

Thus, as soon as we know the probabilities of the singleton sets, we can calculate the probability of any event. \( \Box \)
Example: Suppose that \( S = \{1, 2, 3, 4, 5, 6\} \) is the set of possible outcomes when a die is rolled. If we assume that the die is fair, then each outcome is equally likely and so

\[
1 = P(S) = P(\{1\}) + P(\{2\}) + P(\{3\}) + P(\{4\}) + P(\{5\}) + P(\{6\})
\]

\[
= 6 \times P(\{1\}) = 6 \times P(\{2\}) = \cdots.
\]

This shows that the probability of each possible outcome is \( \frac{1}{6} \). Furthermore, the probability of an event \( A \subset S \) is equal to

\[
P(A) = \frac{|A|}{6}
\]

where \( |A| \) denotes the cardinality of the set \( A \), i.e., the number of elements in \( A \).
(Boole’s Inequality) Let $A_1, A_2, \cdots, A_n$ be events. Then

$$P\left(\bigcap_{i=1}^{n} A_i\right) \geq 1 - \sum_{i=1}^{n} P(\overline{A_i}).$$

Proof: We proceed by induction on $n$. First, if $n = 1$, then the theorem asserts that

$$P(A_1) \geq 1 - P(\overline{A_1}),$$

and we know that this is true since equality holds between these two expressions by result (1) on the previous slide. Similarly, if $n = 2$, then

$$P(A_1A_2) = P(A_1) + P(A_2) - P(A_1 \cup A_2)$$
$$= 1 - P(\overline{A_1}) + 1 - P(\overline{A_2}) - (1 - P(\overline{A_1} \cup \overline{A_2}))$$
$$= 1 - P(\overline{A_1}) - P(\overline{A_2}) + P(\overline{A_1} \cup \overline{A_2})$$
$$\geq 1 - P(\overline{A_1}) - P(\overline{A_2}),$$

where the inequality on the last line follows from the fact that probabilities are non-negative.
Now suppose that Boole’s inequality holds for some value \( n \geq 2 \) (the induction hypothesis). Then,

\[
\Pr \left( \bigcap_{i=1}^{n+1} A_i \right) = \Pr \left( \left( \bigcap_{i=1}^{n} A_i \right) \cap A_{n+1} \right) \\
\geq 1 - \Pr \left( \bigcap_{i=1}^{n} A_i \right) - \Pr(\bar{A}_{n+1}) \\
= \Pr \left( \bigcap_{i=1}^{n} A_i \right) - \Pr(\bar{A}_{n+1}) \\
\geq 1 - \sum_{i=1}^{n} \Pr(\bar{A}_i) - \Pr(\bar{A}_{n+1}) \\
= 1 - \sum_{i=1}^{n+1} \Pr(\bar{A}_i),
\]

which shows that Boole’s inequality also holds for \( n + 1 \). This completes the induction. \( \square \)
Random Variables

A quantity $X$ is said to be a **random variable** if the possible values of $X$ are real numbers and if the actual value of $X$ is uncertain. Some examples of random variables include:

- The number of heads obtained when a coin is tossed 10 times.
- The height of a randomly sampled individual.
- The number of colds that you will contract this year.
- The speed of light (uncertain due to measurement error).

Transformations of random variables are also random variables: If $X$ is a random variable and $f : \mathbb{R} \to \mathbb{R}$, then $Y = f(X)$ is also a random variable.
Definition

Suppose that $Z$ is a random variable that takes values in a finite set $E = \{z_1, \cdots, z_n\} \subset \mathbb{R}$. The distribution of $Z$ is the probability distribution on $E$ defined by

$$P(A) = P(Z \in A),$$

where $A \subset E$ is an event.

In other words, the distribution of a random variable specifies the probabilities of events that depend on the value of $Z$. Furthermore, because $E$ is finite, the distribution of $Z$ is completely determined by the probabilities of the individual values:

$$p_i = P(Z = z_i).$$

For this reason, we sometimes identify the distribution of a random variable $Z$ with the vector $(p_1, \cdots, p_n)$. Recall that these probabilities satisfy $0 \leq p_i \leq 1$ and $p_1 + \cdots + p_n = 1$. 
Example: Suppose that $A$ is an event that occurs with probability $p = \mathbb{P}(A)$. The 
**indicator variable** $I_A$ of $A$ is the random variable defined to be equal to 1 if $A$ occurs and 0 if $A$ does not occur. Since $I_A$ can only assume one of two possible values, the 
distribution of $I_A$ is particularly simple:

\[
\begin{align*}
p_1 &= \mathbb{P}(I_A = 1) = \mathbb{P}(A) = p \\
p_0 &= \mathbb{P}(I_A = 0) = \mathbb{P}(\overline{A}) = 1 - p.
\end{align*}
\]

For example, if $A$ is the event that we roll an even number using a fair six-sided die, 
then $I_A$ will be equal to 1 if we roll a 2, 4 or 6, and zero otherwise. The distribution of 
$I_A$ in this case is simply $p_1 = p_0 = 1/2$. 
Although the distribution of a random variable with $n$ possible values can be completely specified by listing the probabilities $p_1, \cdots, p_n$, it is often possible to summarize the key features of a distribution in terms of a small number of quantities. One quantity that is frequently used in this way is the expectation.

**Definition**

Suppose that $Z$ is a random variable that takes values in a finite set $E = \{z_1, \cdots, z_n\} \subset \mathbb{R}$ with probabilities $p_i = \mathbb{P}(Z = z_i)$. The expectation (also called the expected value or the mean) of $Z$ is the quantity

$$
\mathbb{E}[Z] = \sum_{i=1}^{n} p_i \cdot z_i.
$$

Notice that the expected value of a random variable is a weighted average of the possible values of that variable where the weights are just the probabilities of each of the values.
**Example:** If $I_A$ is the indicator variable for event $A$ and $A$ has probability $p = \mathbb{P}(A)$, then the expectation of $I_A$ is

$$\mathbb{E}[I_A] = p \cdot 1 + (1 - p) \cdot 0 = p,$$

i.e., the expected value of an indicator variable is equal to the probability of the event that is indicated. This is a simple but very useful result.

**Example:** Suppose that $X$ is the number obtained when a fair six-sided die is rolled. Then $p_1 = p_2 = p_3 = p_4 = p_5 = p_6 = 1/6$ and so the expected value of $X$ is

$$\mathbb{E}[X] = \sum_{i=1}^{6} \frac{1}{6} \cdot i = \frac{1}{6} (1 + 2 + 3 + 4 + 5 + 6) = \frac{7}{2}.$$

**Remark:** Notice that in both of these examples, the expected value of $X$ is not actually a value that can be taken by $X$. 
Lemma

Suppose that $Z$ is a random variable that takes values in $E = \{z_1, \cdots, z_n\}$ with probabilities $p_1, \cdots, p_n$. If $k$ and $b$ are constants and we let $W = kZ + b$, then

$$
\mathbb{E}[W] = k \cdot \mathbb{E}[Z] + b.
$$

Proof: Notice that $W$ is a random variable that takes values in the set $\{kz_1 + b, \cdots, kz_n + b\}$ with probabilities $p_1, \cdots, p_n$. Consequently,

$$
\mathbb{E}[W] = \sum_{i=1}^{n} p_i \cdot (kz_i + b)
$$

$$
= \sum_{i=1}^{n} p_i \cdot kz_i + \sum_{i=1}^{n} p_i \cdot b
$$

$$
= k \sum_{i=1}^{n} p_i \cdot z_i + b \sum_{i=1}^{n} p_i
$$

$$
= k \mathbb{E}[Z] + b,
$$

where we have used the fact that $\sum_{i=1}^{n} p_i = 1$ to pass to the last line. □
Theorem

Suppose that $X$ is a random variable that takes values in the set $\{x_1, \cdots, x_n\}$ and that $Y$ is a random variable that takes values in the set $\{y_1, \cdots, y_m\}$. Then

$$E[X + Y] = E[X] + E[Y]$$

In other words, the expectation of a sum is equal to the sum of the expectations.

Proof: First notice that the sum $X + Y$ is a random variable that takes the values $x_1 + y_1, x_1 + y_2, \cdots, x_n + y_m$ with probabilities $P(X = x_i, Y = y_j)$. Consequently, the expected value of $X + Y$ is equal to

$$E[X + Y] = \sum_{i,j} P(X = x_i, Y = y_j) \cdot (x_i + y_j).$$

Similarly, the expectations on the right-hand side of the identity can be written as

$$E[X] + E[Y] = \sum_{i=1}^{n} P(X = x_i) \cdot x_i + \sum_{j=1}^{m} P(Y = y_j) \cdot y_j.$$
To show that these two expressions are equal we need to know how the probabilities

\[ p_{i,j} \equiv \mathbb{P}(X = x_i, Y = y_j) \]

that appear on the left-hand side are related to the probabilities

\[ p_i^X \equiv \mathbb{P}(X = x_i) \quad \text{and} \quad p_j^Y \equiv \mathbb{P}(Y = y_j) \]

that appear on the right-hand side.

Since

\[ \{X = x_i\} = \bigcup_{j=1}^{m} \{X = x_i, Y = y_j\} \]

is a finite union of disjoint events, the finite additivity of probabilities implies that

\[ p_i^X = \mathbb{P}(X = x_i) = \sum_{j=1}^{m} \mathbb{P}(X = x_i, Y = y_j) = \sum_{j=1}^{m} p_{ij}. \]
Similarly,

\[ \{ Y = y_j \} = \bigcup_{i=1}^{n} \{ X = x_i, Y = y_j \} \]

is a finite union of disjoint events and so

\[ p_j^Y = \mathbb{P}(Y = y_j) = \sum_{i=1}^{n} \mathbb{P}(X = x_i, Y = y_j) = \sum_{i=1}^{n} p_{ij}. \]

**Aside:** The probabilities \( p_{i,j} \) specifying the values of both random variables \( X \) and \( Y \) are said to be **joint probabilities**, while the probabilities \( p_{i}^{X} \) and \( p_{j}^{Y} \) specifying the value of only one of the two variables are called **marginal probabilities**.
This shows that

\[ \mathbb{E}[X + Y] = \sum_{i=1}^{n} \sum_{j=1}^{m} p_{i,j} \cdot (x_i + y_j) \]

\[ = \sum_{i=1}^{n} \sum_{j=1}^{m} p_{i,j} \cdot x_i + \sum_{i=1}^{n} \sum_{j=1}^{m} p_{i,j} \cdot y_j \]

\[ = \sum_{i=1}^{n} x_i \cdot \left( \sum_{j=1}^{m} p_{i,j} \right) + \sum_{j=1}^{m} y_j \cdot \left( \sum_{i=1}^{n} p_{i,j} \right) \]

\[ = \sum_{i=1}^{n} x_i \cdot p_i^X + \sum_{j=1}^{m} y_j \cdot p_j^Y \]

\[ = \mathbb{E}[X] + \mathbb{E}[Y], \]

which completes the proof. \( \square \)
Several useful results follow from the preceding theorem. First, a simple induction argument shows that

$$E[X_1 + X_2 + \cdots + X_q] = E[X_1] + E[X_2] + \cdots + E[X_q].$$

Secondly, if $c_1, \cdots, c_n$ are constants, then by combining the previous lemma with this theorem, we find that

$$E\left[ \sum_{i=1}^{n} c_i X_i \right] = \sum_{i=1}^{n} c_i E[X_i].$$

In other words, the expectation of a linear combination of random variables is equal to the linear combination of the expected values of those variables.
We can use the linearity of expectations to re-derive an earlier result concerning the probability of a union of events. To this end, let $A$ and $B$ be events, not necessarily disjoint, and let $\text{I}_A$ and $\text{I}_B$ be the corresponding indicator variables. In one of the exercises, you are asked to show that the following identities are true

$$\text{I}_{AB} = \text{I}_A \text{I}_B \quad \text{I}_{\bar{A}} = 1 - \text{I}_A.$$ 

However, since $\bar{A} \cup B = AB$,

$$\text{I}_{A \cup B} = 1 - \text{I}_{\bar{A} \cup B} = 1 - \text{I}_{\bar{AB}} = 1 - \text{I}_{\bar{A}} \text{I}_B = 1 - (1 - \text{I}_A)(1 - \text{I}_B) = \text{I}_A + \text{I}_B - \text{I}_{AB}.$$ 

Upon taking expectations of both sides of this identity, we obtain

$$\mathbb{P}(A \cup B) = \mathbb{E}[\text{I}_{A \cup B}] = \mathbb{E}[\text{I}_A + \text{I}_B - \text{I}_{AB}] = \mathbb{E}[\text{I}_A] + \mathbb{E}[\text{I}_B] - \mathbb{E}[\text{I}_{AB}] = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB).$$
Indicator variables can also be used to derive an expression for the probability of the union of \( n \) events \( A_1, \ldots, A_n \). Recall that DeMorgan’s laws assert that the complement of a union of sets is equal to the intersection of the complements:

\[
\bigcup_{i=1}^{n} A_i = \bigcap_{i=1}^{n} \overline{A_i}.
\]

Consequently,

\[
l_{\bigcup_{i=1}^{n} A_i} = 1 - l_{\bigcap_{i=1}^{n} \overline{A_i}} = 1 - \prod_{i=1}^{n} l_{\overline{A_i}}
\]

\[
= 1 - \prod_{i=1}^{n} (1 - l_{A_i})
\]

\[
= 1 - (1 - l_{A_1})(1 - l_{A_2}) \cdots (1 - l_{A_n})
\]

\[
= \sum_{i=1}^{n} l_{A_i} - \sum_{i<j} l_{A_i}l_{A_j} + \sum_{i<j<k} l_{A_i}l_{A_j}l_{A_k} - \cdots + (-1)^{n-1} \prod_{i=1}^{n} l_{A_i}.
\]
Taking expectations of both sides of this identity gives:

\[
\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \mathbb{E}\left[\mathbb{I}_{\bigcup_{i=1}^{n} A_i}\right] = \mathbb{E}\left[\sum_{i=1}^{n} \mathbb{I}_{A_i} - \sum_{i<j} \mathbb{I}_{A_i A_j} + \cdots + (-1)^{n-1} \prod_{i=1}^{n} \mathbb{I}_{A_i}\right] 
\]

\[
= \sum_{i=1}^{n} \mathbb{E}[\mathbb{I}_{A_i}] - \sum_{i<j} \mathbb{E}[\mathbb{I}_{A_i A_j}] + \cdots + (-1)^{n-1} \mathbb{E}[\mathbb{I}_{A_1 A_2 \cdots A_n}] 
\]

\[
= \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i A_j) + \cdots + (-1)^{n-1} \mathbb{P}(A_1 A_2 \cdots A_n). 
\]

The resulting identity is sometimes known as the inclusion-exclusion formula.
When $n = 3$ and $n = 4$, the inclusion-exclusion formula gives:

$$\Pr(A_1 \cup A_2 \cup A_3) = \left( \Pr(A_1) + \Pr(A_2) + \Pr(A_3) \right)$$
$$- \left( \Pr(A_1 A_2) + \Pr(A_1 A_3) + \Pr(A_2 A_3) \right) + \Pr(A_1 A_2 A_3)$$

and

$$\Pr(A_1 \cup A_2 \cup A_3 \cup A_4) = \left( \Pr(A_1) + \Pr(A_2) + \Pr(A_3) + \Pr(A_4) \right)$$
$$- \left( \Pr(A_1 A_2) + \Pr(A_1 A_3) + \Pr(A_1 A_4) + \Pr(A_2 A_3) \right)$$
$$+ \Pr(A_2 A_4) + \Pr(A_3 A_4))$$
$$+ \left( \Pr(A_1 A_2 A_3) + \Pr(A_1 A_2 A_4) + \Pr(A_1 A_3 A_4) + \Pr(A_2 A_3 A_4) \right)$$
$$- \Pr(A_1 A_2 A_3 A_4).$$
The Matching Problem

The power of the inclusion-exclusion formula can be illustrated by the matching problem, which was first posed by the French mathematician Pierre Rémond de Montmort in a 1708 treatise on gambling. There are many equivalent formulations of this problem, but here we give one that is close to the original version.

**Problem:** Suppose that a deck of \( n \) distinct cards is shuffled at random. What is the probability \( P_{0,n} \) that no card remains in its original position in the deck?

**Interpretation:** Before we can solve this problem, we need to decide what we mean when we say that the deck is shuffled at random. Here we will assume that that each of the \( n! \) possible ways of re-ordering the \( n \) cards is equally likely, i.e., each possible re-ordering of the deck has probability \( 1/n! \).
**Solution:** Let $A_i$ be the event that the $i$'th card remains in its original position. Notice that

$$\bigcup_{i=1}^{n} A_i = \text{‘at least one of the } n \text{ cards remains in its original position’}$$

and so the event that none of the cards remain in their original position is the complement of this union. This shows that

$$P_{0,n} = 1 - \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right),$$

and so we can solve for $P_{0,n}$ by calculating the probability of the union and subtracting this quantity from 1. This is what allows us to make use of the inclusion-exclusion identity.
By the inclusion-exclusion identity, we know that

$$
\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_iA_j) + \sum_{i<j<k} \mathbb{P}(A_iA_jA_k) - \cdots + (-1)^{n-1} \mathbb{P}(A_1A_2\cdots A_n),
$$

which we can evaluate provided that we can calculate the probabilities of the various intersections of the events $A_1, \cdots, A_n$.

Consider $\mathbb{P}(A_i)$. This is the probability that the $i$'th card remains in its original position. Notice that we make no assumptions about the positions of the other cards. Since there are $(n-1)!$ ways of shuffling the deck so that card $i$ remains in its original position and since each of these has probability $1/n!$, it follows that

$$
\mathbb{P}(A_i) = (n-1)! \times \frac{1}{n!} = \frac{(n-1)(n-2)\cdots 2 \cdot 1}{n(n-1)(n-2)\cdots 2 \cdot 1} = \frac{1}{n},
$$

and this probability does not depend on $i$. Accordingly,

$$
\sum_{i=1}^{n} \mathbb{P}(A_i) = \sum_{i=1}^{n} \frac{1}{n} = n \times \frac{1}{n} = 1.
$$
Next, if $i \neq j$, $A_i A_j$ is the event that both the $i$’th card and the $j$’th card remain in their original position. Again, we make no assumptions about the positions of the other $n-2$ cards. It follows that there are $(n-2)!$ ways of shuffling the deck so that cards $i$ and $j$ do not move to new locations and so

$$P(A_i A_j) = (n-2)! \times \frac{1}{n!} = \frac{(n-2) \cdots 2 \cdot 1}{n(n-1)(n-2) \cdots 2 \cdot 1} = \frac{1}{n(n-1)}.$$ 

Since these probabilities do not depend on $i$ and $j$ (apart from the requirement that $i \neq j$), the second sum is equal to

$$\sum_{i<j} P(A_i A_j) = \sum_{i<j} \frac{1}{n(n-1)} = \binom{n}{2} \frac{1}{n(n-1)} = \frac{n(n-1)}{2} \frac{1}{n(n-1)} = \frac{1}{2}.$$

Here $\binom{n}{2} = \frac{n!}{(n-2)!2!}$ is the number of ways of choosing 2 distinct cards from a deck containing $n$ cards.
Similarly, if $i < j < k$, $A_i A_j A_k$ is the event that the $i$’th, the $j$’th and the $k$’th cards remain in their original positions. Since there are $(n - 3)!$ ways of shuffling the deck so that this is true, we see that $\mathbb{P}(A_i A_j A_k) = (n - 3)!/n! = 1/(n(n - 1)(n - 2))$

$$
\sum_{i<j<k} \mathbb{P}(A_i A_j A_k) = \sum_{i<j<k} \frac{1}{n(n - 1)(n - 2)}
$$

$$
= \binom{n}{3} \frac{1}{n(n - 1)(n - 2)}
$$

$$
= \frac{n!}{(n - 3)!3!} \times \frac{1}{n(n - 1)(n - 2)}
$$

$$
= \frac{n(n - 1)(n - 2)}{6} \times \frac{1}{n(n - 1)(n - 2)}
$$

$$
= \frac{1}{6}.
$$
For the general case, we need to calculate the probability that the $r$ cards $i_1 < i_2 < \cdots < i_r$ remain in their original positions. There are $(n - r)!$ ways of shuffling the deck so that this is true, which shows that

$$\mathbb{P}(A_{i_1} A_{i_2} \cdots A_{i_r}) = \frac{(n - r)!}{n!}.$$

Consequently,

$$\sum_{i_1 < i_2 < \cdots < i_r} \mathbb{P}(A_{i_1} A_{i_2} \cdots A_{i_r}) = \binom{n}{r} \frac{(n - r)!}{n!} = \frac{n!}{(n - r)!r!} \frac{(n - r)!}{n!} = \frac{1}{r!}.$$
Substituting these results back into the original inclusion-exclusion formula, we have

\[ P\left( \bigcup_{i=1}^{n} A_i \right) = 1 - \frac{1}{2} + \frac{1}{3!} - \frac{1}{4!} + \cdots + (-1)^{n-1} \frac{1}{n!} \]

\[ = \sum_{r=1}^{n} (-1)^{r-1} \frac{1}{r!}. \]

However, since we are interested in the probability that no card remains in its original position, we need to subtract this quantity from 1:

\[ P_{0,n} = 1 - \sum_{r=1}^{n} (-1)^{r-1} \frac{1}{r!} \]

\[ = 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^n \frac{1}{n!} \]

\[ = \sum_{r=0}^{n} (-1)^r \frac{1}{r!}, \]

which is the solution to the matching problem.
When $n$ is large, it turns out that $P_{0,n} \approx e^{-1} \approx 0.36787944$. This follows from the fact that the Taylor series expansion for the exponential function $e^x$ is

$$
\sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots.
$$

If we take $x = -1$ in this formula, then we arrive at the identity

$$
e^{-1} = \sum_{n=0}^{\infty} (-1)^n \frac{1}{n!} = \lim_{n \to \infty} P_{0,n}.
$$

Inspection of $P_{0,n}$ for $n \leq 10$ demonstrates that convergence is quite rapid:

<table>
<thead>
<tr>
<th>$n$</th>
<th>$P_{0,n}$</th>
<th>$n$</th>
<th>$P_{0,n}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>6</td>
<td>0.36805556</td>
</tr>
<tr>
<td>2</td>
<td>0.5</td>
<td>7</td>
<td>0.36785714</td>
</tr>
<tr>
<td>3</td>
<td>0.33333333</td>
<td>8</td>
<td>0.36788194</td>
</tr>
<tr>
<td>4</td>
<td>0.375</td>
<td>9</td>
<td>0.36787919</td>
</tr>
<tr>
<td>5</td>
<td>0.36666667</td>
<td>10</td>
<td>0.36787946</td>
</tr>
</tbody>
</table>
It is also of interest to calculate the expected number of cards that remain in their original positions when the deck is shuffled. Although the distribution of this number is fairly complicated, we can use the linearity of expectations and the properties of indicator variables to arrive at a fairly easy solution.

To this end, let $I$ denote the (random) number of cards that remain in their original position and let $I_{A_i}$ denote the indicator variable for the event $A_i$ that the $i$'th card remains in its original position. Then

$$I = \sum_{i=1}^{n} I_{A_i}.$$ 

However, since $\mathbb{E}[I_{A_i}] = \mathbb{P}(A_i) = 1/n$ for all $i = 1, \cdots, n$, it follows that

$$\mathbb{E}[I] = \mathbb{E} \left[ \sum_{i=1}^{n} I_{A_i} \right] = \sum_{i=1}^{n} \mathbb{E}[I_{A_i}] = \sum_{i=1}^{n} \frac{1}{n} = 1.$$

Thus, on average, exactly 1 card will remain in its original position and this holds true for all $n \geq 1$. 
Our next result tells us how to calculate the expectation of a function of random variable and is used so frequently that it has come to be known as the *Law of the Unconscious Statistician*.

**Theorem**

*Suppose that $X$ is a random variable with values in the set $S = \{x_1, \ldots, x_n\}$ and let $Y = g(X)$, where $g : S \to \mathbb{R}$ is a function from $S$ into the real numbers. Then*

$$
E[Y] = E[g(X)] = \sum_{i=1}^{n} P(X = x_i)g(x_i).
$$

**Proof:** Let the possible values of $Y$ be denoted $y_1, \ldots, y_m$, where $m \leq n$, and define the disjoint sets $S_1, \ldots, S_m$ by

$$
S_j = g^{-1}(y_j) = \{x_k \in S : g(x_k) = y_j\}.
$$

Then $y_j = g(x_k)$ if and only if $x_k \in S_j$. 

---
Notice that the sets $S_1, \cdots, S_m$ are disjoint (since $Y$ can only be equal to one of the $y_j$'s) and also that

$$S = \bigcup_{j=1}^{m} S_j.$$ 

In other words, $S_1, \cdots, S_m$ is said to be a **partition** of $S$.

To see how the distributions of $X$ and $Y$ are related, observe that

$$\{Y = y_j\} = \{X \in S_j\} = \bigcup_{k: x_k \in S_j} \{X = x_k\}$$

expresses the event on the left-hand side as the disjoint union of events on the right-hand side. Then, because probabilities are finitely additive, we have

$$\mathbb{P}(Y = y_j) = \sum_{k: x_k \in S_j} \mathbb{P}(X = x_k).$$
By the definition of the expectation of $Y$, we have

$$
\mathbb{E}[Y] = \sum_{j=1}^{m} \mathbb{P}(Y = y_j) \cdot y_j
$$

$$
= \sum_{j=1}^{m} \left( \sum_{k: x_k \in S_j} \mathbb{P}(X = x_k) \right) \cdot y_j
$$

$$
= \sum_{j=1}^{m} \left( \sum_{k: x_k \in S_j} \mathbb{P}(X = x_k) \cdot y_j \right)
$$

$$
= \sum_{j=1}^{m} \left( \sum_{k: x_k \in S_j} \mathbb{P}(X = x_k) \cdot g(x_k) \right)
$$

$$
= \sum_{i=1}^{n} \mathbb{P}(X = x_i) \cdot g(x_i),
$$

where the final line follows from the fact that the sets $S_1, \cdots, S_m$ form a partition of $S$, so that every value $x_i \in S$ belongs to exactly one of the $S_j$'s. □
Example: Suppose that $X$ is a random variable with the following distribution:

$$
P(X = -1) = 0.3, \ P(X = 0) = 0.2, \ P(X = 1) = 0.1, \ P(X = 2) = 0.4.
$$

Calculate the expectations of $X$ and $X^2$. 

Random Variables

Example: Suppose that $X$ is a random variable with the following distribution:

$$P(X = -1) = 0.3, P(X = 0) = 0.2, P(X = 1) = 0.1, P(X = 2) = 0.4.$$ 

Calculate the expectations of $X$ and $X^2$.

Solution: The expected value of $X$ is

$$E[X] = \sum_{i=-1}^{2} P(X = i) \cdot i$$

$$= 0.3 \cdot (-1) + 0.2 \cdot 0 + 0.1 \cdot 1 + 0.4 \cdot 2 = 0.6.$$ 

Also, to calculate $E[X^2]$, take $g(x) = x^2$ and use the preceding theorem to obtain:

$$E[X^2] = \sum_{i=-1}^{2} P(X = i)g(i)$$

$$= 0.3 \cdot (-1)^2 + 0.2 \cdot 0^2 + 0.1 \cdot 1^2 + 0.4 \cdot 2^2 = 2.$$
Expectations and Coherence

In this section we will complete our discussion of coherence and probabilities by showing that a Dutch book cannot be made against a coherent price assignment under the rules of the betting game introduced previously.

To do so, we will need the following result which relates the expectation of a random variable to the minimum and maximum values that the variable can assume. Let us say that a random variable is trivial if there is a number \( c \) such that \( \mathbb{P}(X = c) = 1 \), i.e., \( X \) is trivial if there is no uncertainty about its value. Otherwise, \( X \) is said to be non-trivial.

**Theorem**

Suppose that \( X \) is a non-trivial random variable with values in the set \( \{x_1, \ldots, x_n\} \), where \( x_1 < \cdots < x_n \). Then

\[
\min X \equiv x_1 < \mathbb{E}[X] < \max X = x_n
\]

where the lower and upper inequalities are strict.
Proof: Let \( p_i = \mathbb{P}(X = x_i) \). Without loss of generality, we may assume that each of the probabilities \( p_i \) is positive. Indeed, since \( X \) is non-trivial, we know that there are at least two values with positive probabilities. Furthermore, because the expected value of a random variable is the weighted sum of these values with weights supplied by the probabilities \( p_i \), adding or removing values with zero probabilities does not change the expected value.

Then, since \( \sum_{i=1}^{n} p_i = 1 \), we have

\[
\min X = x_1 = \left( \sum_{i=1}^{n} p_i \right) x_1 < \sum_{i=1}^{n} p_i x_i < \left( \sum_{i=1}^{n} p_i \right) x_n = x_n = \max X.
\]

\[\square\]
Our proof that coherence precludes a Dutch book will depend on the following corollary to the preceding theorem.

**Corollary**

*If* $X$ *is a non-trivial random variable, then there is both a positive probability* $\epsilon_1 > 0$ *that* $X$ *will exceed its expectation by some positive quantity* $\eta_1 > 0$ *and a positive probability* $\epsilon_2 > 0$ *that* $X$ *will be less than its expectation by some positive quantity* $\eta_2 > 0$.

**Proof:** The corollary follows from the preceding theorem if we let $\eta_1 = x_n - \mathbb{E}[X] > 0$ and $\epsilon_1 = p_n > 0$, as well as $\eta_2 = \mathbb{E}[X] - x_0 > 0$ and $\epsilon_2 = p_1$. □
Theorem

(Fundamental Theorem of Coherence:) A Dutch book is possible in the betting game introduced earlier if and only if the price assignment \( \{ Pr(A) : A \subset S \} \) is not coherent.

Proof: We have already shown that a Dutch book can be made whenever the prices are not coherent. Here we will prove the converse, i.e., if the prices are coherent, then a Dutch book cannot be made.

Let \( A \) be an event in \( S \) and let \( I_A \) be the indicator variable of \( A \). Then the amount earned by player 2 if they wager \( W(A) \) on \( A \) at a price of \( Pr(A) \) per $1 wager is:

\[
W(A) \cdot (I_A - Pr(A)).
\]

Indeed, if \( A \) occurs, then \( I_A = 1 \) and player 2 earns \( 1 - Pr(A) \) per $1 wager, while if \( A \) does not occur then \( I_A = 0 \) and player 2 loses \( -Pr(A) \) per $1 wager.
Now denote the events in $S$ by $A_1, \cdots , A_n$ (notice that these need not be disjoint) and suppose that player 2 wagers $w_i = W(A_i)$ on each event $A_i$ at a price of $Pr(A_i)$ per $1$ wager. Then player 2’s total wins or losses, once all bets have been settled, is equal to

$$W \equiv \sum_{i=1}^{n} w_i (I_{A_i} - Pr(A_i)).$$

Also, $W$ is a random variable that can take on finitely many possible values with expectation

$$\mathbb{E}[W] = \mathbb{E} \left[ \sum_{i=1}^{n} w_i (I_{A_i} - Pr(A_i)) \right]$$

$$= \sum_{i=1}^{n} \mathbb{E} [w_i (I_{A_i} - Pr(A_i))]$$

$$= \sum_{i=1}^{n} w_i \cdot \mathbb{E} [I_{A_i} - Pr(A_i)].$$
Furthermore, for every $i = 1, \cdots, n$, we have

$$\mathbb{E} [I_{A_i} - Pr(A_i)] = \mathbb{E} [I_{A_i}] - Pr(A_i) = Pr(A_i) - Pr(A_i) = 0,$$

since the expectation of the indicator variable $I_{A_i}$ is equal to the probability of the event $A_i$, which is given by the price $Pr(A_i)$ that player 1 assigns to this event. This shows that $\mathbb{E}[W] = 0$ and so, according to the preceding corollary there are just two possibilities:

1. $W$ is trivial, in which case $W = 0$ with probability 1 and so neither player can be a sure loser (i.e., there is no Dutch book).

2. If $W$ is not trivial, then there is a positive probability that $W$ will be negative as well as a positive probability that $W$ will be positive. Again, neither player can be a sure loser and so there is no Dutch book.