Summer School on Mathematical Philosophy
for Female Students

Introduction to Probability Theory, Algebra, and Set Theory

Catrin Campbell-Moore and Sebastian Lutz

Munich Center for Mathematical Philosophy

July 28, 2014

Supported by

Alexander von Humboldt
Stiftung/Foundation

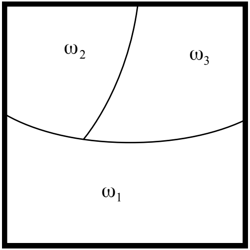
Outline

Outline

Venn Diagrams

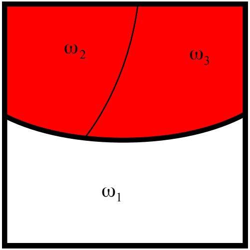
The domain, Ω, is a set

Example: $\{\omega_1, \omega_2, \omega_3\}$



A subset of the domain

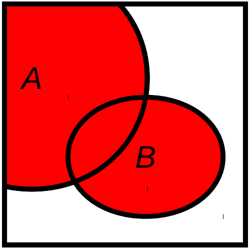
$\{\omega_2, \omega_3\}$



Operations on Sets: Union and Intersection

Union $A \cup B$

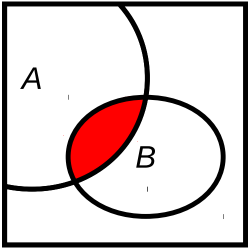
$$\{\omega_1, \omega_2\} \cup \{\omega_2, \omega_3\} = \{\omega_1, \omega_2, \omega_3\}$$



Disjunction $\mathbf{A} \vee \mathbf{B}$

Intersection $A \cap B$

$$\{\omega_1, \omega_2\} \cap \{\omega_2, \omega_3\} = \{\omega_2\}$$

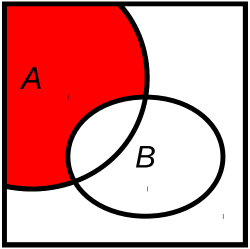


Conjunction $\mathbf{A} \wedge \mathbf{B}$

Operations on Sets: Subtraction and Complement

Subtraction $A \backslash B$

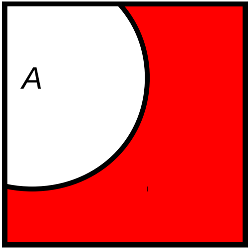
$\{\omega_1, \omega_2\} \backslash \{\omega_2, \omega_3\} = \{\omega_1\}$



$\mathbf{A} \wedge \neg \mathbf{B}$

Complement $A^c = \Omega \backslash A$

$\{\omega_1, \omega_2\}^c = \{\omega_3\}$

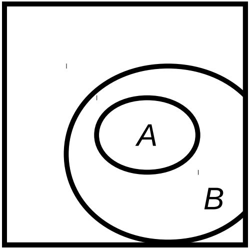


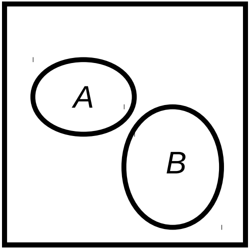
Negation $\neg \mathbf{A}$

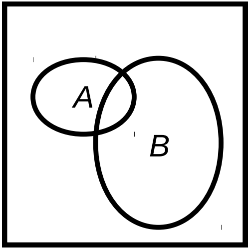
Relations Between Sets

$A \subseteq B$ | $A \cap B = \varnothing$ | $A \cap B \neq \varnothing$

$\{\omega_1\} \subseteq \{\omega_1, \omega_2\}$ | $\{\omega_1, \omega_2\} \cap \{\omega_3\} = \varnothing$ | $\{\omega_1, \omega_2\} \cap \{\omega_2\} = \{\omega_2\}$



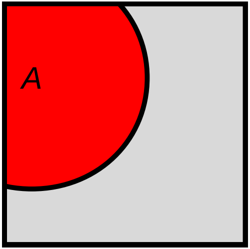




$\mathbf{A} \rightarrow \mathbf{B}$ | $\mathbf{A} \wedge \mathbf{B} \rightarrow \bot$ | $\neg(\mathbf{A} \wedge \mathbf{B} \rightarrow \bot)$

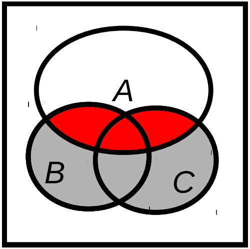
Inferences With Venn Diagrams

$(A^c)^c = A$



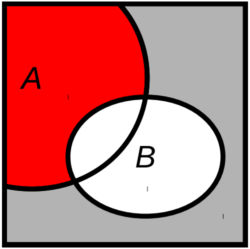
$\neg\neg\mathbf{A} \leftrightarrow \mathbf{A}$

$A \cap (B \cup C) =$
$(A \cap B) \cup (A \cap C)$



$\mathbf{A} \wedge (\mathbf{B} \vee \mathbf{C})$
$\leftrightarrow (\mathbf{A} \wedge \mathbf{B}) \vee (\mathbf{A} \wedge \mathbf{C})$

$A \backslash B = A \cap B^c$

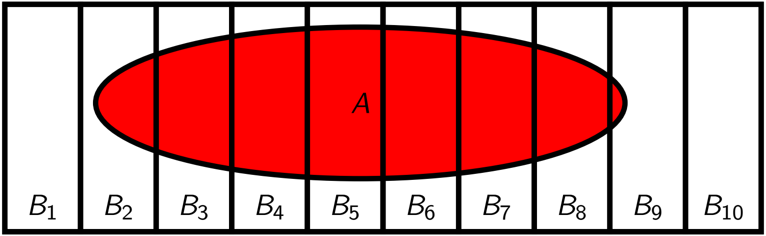


$\mathbf{A} \wedge \neg\mathbf{B}$

Partitions

$B_1, B_2, \ldots, B_k$ is a *partition* of Ω if and only if

$$B_1 \cup B_2 \cup \cdots \cup B_k = \Omega \text{ and } B_i \cap B_j = \varnothing \text{ for } i \neq j.$$



If $B_1, B_2, \ldots, B_k$ is a partition, then for every A,

- $(A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cup B_k) = A$.
- $(A \cap B_i) \cap (A \cap B_j) = \varnothing$ for $i \neq j$.

$\Rightarrow$ A partition of Ω partitions every subset of Ω.

The Set of All States

- A *state*: A way in which the world could be.
- We call the set of all possible states Ω.
- Examples for Ω:
  - The set of possible entire past, present and futures of the universe.
  - $\{\text{heads}, \text{tails}\}$
  - $\{\text{egg rotten}, \text{egg good}\}$
  - $\{\text{egg good and Jo hungry}, \text{egg good and Jo not hungry}, \text{egg rotten and Jo hungry}, \text{egg rotten and Jo not hungry}\}$
  - $\{\text{Jo has height } r\mathrm{m} : r \in \mathbb{R}^+\}$
  - $\{\text{The center of the vase is at } x : x \text{ is a point on the tabletop}\}$
  - The set of infinite sequences of tosses of a coin.
  - The set of models of a language L

Events as Sets of States: Basic Idea

Roughly, subsets of Ω are called *events*.

- $\Omega = \{\langle H, H\rangle, \langle H, T\rangle, \langle T, H\rangle, \langle T, T\rangle\}$
- The proposition $\mathbf{A}$ = "The first coin lands heads" describes the *event* $A = \{\langle H, H\rangle, \langle H, T\rangle\}$
- The proposition $\mathbf{B}$ = "At least one coin lands heads", describes the *event* $B = \{\langle H, H\rangle, \langle H, T\rangle, \langle T, H\rangle\}$

- $\Omega = \{\text{Jo has height } r\mathrm{m} : r \in \mathbb{R}\}$
- Is Jo taller than $2\mathrm{m}$?
- Events of interest:
  $\{\text{Jo has height } r\mathrm{m} : r \leqslant 2\}$ and $\{\text{Jo has height } r\mathrm{m} : r > 2\}$

Events as Sets of States: Formalism

$\Omega = \{\text{Jo has height } r\mathrm{m} \text{ and Ed has height } t\mathrm{m} : r, t \in \mathbb{R}\}$

Suppose I'm interested in

- A: Jo is taller than $2\mathrm{m}$
- B: Ed is taller than Jo

We will also then be interested in events which can be formed from combining A and B, e.g.

- $A \cap B$: Jo is taller than $2\mathrm{m}$ and Ed is taller than Jo
- A^c: Jo is not taller than $2\mathrm{m}$

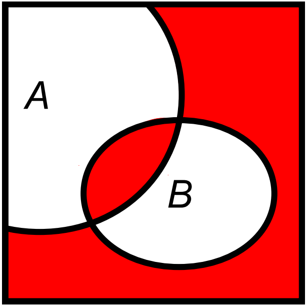
We call the set of the events that we're interested in $\mathcal{F}$.

We assume that $\mathcal{F}$ is a *Boolean algebra*, i.e.

- $\varnothing \in \mathcal{F}$ and $\Omega \in \mathcal{F}$.
- If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$.
- If $A, B \in \mathcal{F}$ then $A \cup B \in \mathcal{F}$.

Boolean Algebra

A Boolean algebra which contains A and B will also contain all the subsets which you can draw lines around.



The *Boolean algebra generated by* $A_1, \ldots, A_n$ is just the smallest Boolean algebra containing all of A_1 to A_n.

Events as Sets of States: Some More

For example $\Omega = \{\langle H, H\rangle, \langle H, T\rangle, \langle T, H\rangle, \langle T, T\rangle\}$, the following are Boolean algebras over Ω

- $\{\varnothing, \Omega\}$
- $\{\varnothing, \{\langle H, H\rangle, \langle H, T\rangle\}, \{\langle T, H\rangle, \langle T, T\rangle\}, \Omega\}$
- $\mathcal{P}(\Omega)$

Consequences of the formalism:

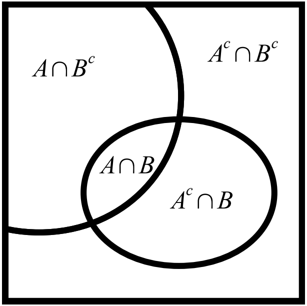
- If $A, B \in \mathcal{F}$ then $A \cap B \in \mathcal{F}$.
- If $A, B \in \mathcal{F}$ then $A \backslash B \in \mathcal{F}$.

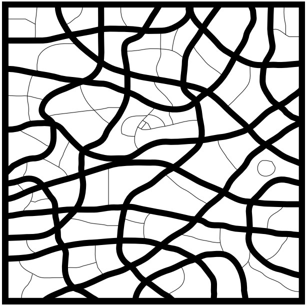
Sometimes it is asked that the event space is a σ-*algebras*:

- $\varnothing \in \mathcal{F}$ and $\Omega \in \mathcal{F}$.
- If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$.
- If $A_1, A_2, A_3, \ldots \in \mathcal{F}$ then $A_1 \cup A_2 \cup A_3 \cup \ldots \in \mathcal{F}$.

Atoms in a Boolean algebra

A is an *atom* of a Boolean algebra $\mathcal{F}$ if there is no $B \in \mathcal{F}$ with $\varnothing \subset B \subset A$.



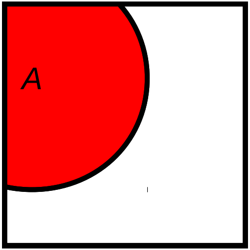


- If $\mathcal{F}$ is finite we can always partition Ω into atoms like this.
- All other events in $\mathcal{F}$ are unions of the atoms.
- Note: Atoms can be *sets* of states.
- Note: The atoms form a partition of $\mathcal{F}$.

Outline

What is Probability?

- $P : \mathcal{F} \to \mathbb{R}$
- How likely the event is to happen.
- We can think of this by taking the size of the areas in the diagrams into account.
- We stipulate that the size of the diagram is 1.
- $P(A)$ measures the area A.

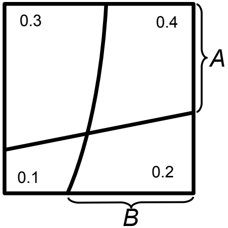


Just Look at the Atoms

We want to calculate the size of each of $A \in \mathcal{F}$.

- To do this we can just look at the size of the atoms.
- Since the atoms partion Ω, $\sum_{A \text{ is an atom}} P(A) = 1$.

Atom C	$P(C)$
$A \cap B$	0.4
$A \cap B^c$	0.3
$A^c \cap B$	0.2
$A^c \cap B^c$	0.1



This allows us to work out the other probabilities of $B \in \mathcal{F}$:

$P(D) = \sum_{C \text{ is an atom and } C \subseteq D} P(C)$ (Note: $\sum_{\varnothing} P(C) = 0$)

$P(A) = P(A \cap B) + P(A \cap B^c) = 0.4 + 0.3 = 0.7$

The Axiomatic Approach

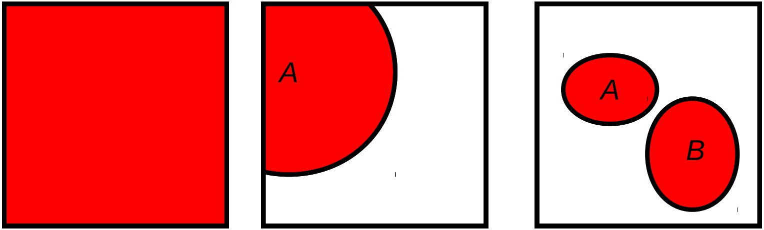
In general we might not have atoms so we give axioms that don't presuppose atoms.

$P : \mathcal{F} \to \mathbb{R}$ satisfying:

Normalisation: $P(\Omega) = 1$

Positivity: $P(A) \geqslant 0$

Finite Additivity: If $A \cap B = \varnothing$ then $P(A \cup B) = P(A) + P(B)$

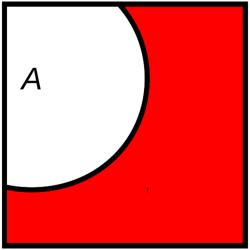


When we have infinite spaces and a σ-algebra we sometimes add:

- *σ-Additivity:* If each $A_i \in \mathcal{F}$ and $A_i \cap A_j = \varnothing$ for all $i \neq j$ then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

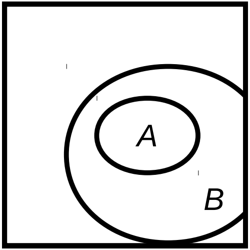
Consequences of the Axioms

$P(A^c) = 1 - P(A)$



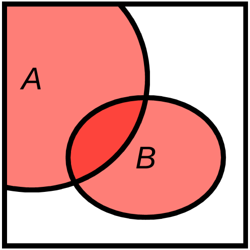
If $A \subseteq B$

then $P(A) \leqslant P(B)$



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

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

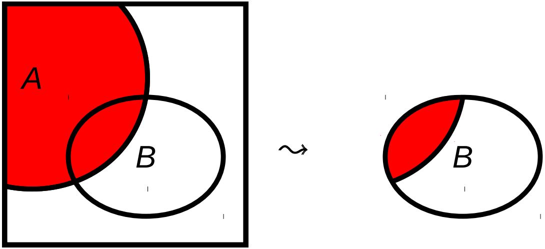


These can also be derived from the axioms.

- $A \cap A^c = \varnothing$ so $1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$

Conditional Probabilities

- $P(A|B)$: "The probability of A given B"
- Remove the area outside B, pretend that B has size 1.



- This should satisfy the *ratio formula*:

$$\text{If } P(B) > 0 \text{ then } P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- The ratio formula can be read as a definition or as a restriction.

Probabilistic Independence

A is *probabilistically independent* from B if and only if $P(A|B) = P(A)$.

Equivalently: $P(A \cap B) = P(A) \cdot P(B)$

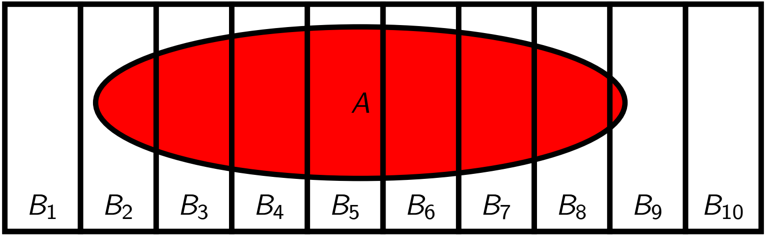
- because
  $P(A \cap B) = \frac{P(A \cap B)}{P(B)} P(B) = P(A|B) \cdot P(B) = P(A) \cdot P(B)$

If A and B are independent then from knowing $P(A)$ and $P(B)$ one can find the probabilities of all the events in the Boolean algebra generated by A and B.

Law of total probability

The *law of total probability* says that if $B_1, \ldots, B_k$ is a partition of Ω then

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



Bayes' Theorem

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

Use the law of total probability: If $B_1, \ldots, B_k$ is a partition of Ω, then

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

Example:

- Jo knows that she has one of three biased coins: $P(\text{Head}|B_1) = 0.6$, $P(\text{Head}|B_2) = 0.7$, $P(\text{Head}|B_3) = 0.6$.
- $P(B_1) = 0.5$, $P(B_2) = 0.3$, $P(B_3) = 0.2$

Then

$$P(B_2|\text{Head}) = \frac{0.7 \times 0.3}{0.6 \times 0.5 + 0.7 \times 0.3 + 0.6 \times 0.2} = \frac{0.21}{0.63} = \frac{1}{3}$$

Outline

What is a Random Variable?

A *random variable* is a function from Ω to $\mathbb{R}$,

$$\begin{aligned} X : \Omega &\to \mathbb{R} \\ \omega &\mapsto X(\omega) \end{aligned}$$

such that

$$\{\omega : X(\omega) \leq r\} \in \mathcal{F} \text{ for all } r \in \mathbb{R} \quad .$$

- $\{\omega : X(\omega) \leq r\} =: \{X \leq r\}$
- $\{\omega : X(\omega) = r\} =: \{X = r\}$ etc.

- Note: Random variables are neither variables nor random.

Examples

- The outcome of a roll of a die.
- $\Omega = \{1 \text{ on top}, 2 \text{ on top}, \ldots, 6 \text{ on top}\}$
- $X(\{1 \text{ on top}\}) = 1, \ldots, X(\{6 \text{ on top}\}) = 6$

- $\Omega = \{\text{Jo has height } r\mathrm{m} \text{ and Ed has height } t\mathrm{m} : r, t \in \mathbb{R}\}$
- $X(\omega) =$ Jo's height
- $Y(\omega) =$ Ed's height

- $\Omega =$
  The set of entire past, present and futures of the universe
- $X(\omega) =$ how rich I am at time t_0 in ω, measured in Euro

- $\Omega = \{\text{it rains today, it does not rain today}\}$
- $X(\omega) =$ how happy I am if I take my umbrella today

Algebraic Operations on Random Variables

$\Omega = \{\text{Jo has height } r\mathrm{m} \text{ and Ed has height } t\mathrm{m} : r, t \in \mathbb{R}\}$
$X(\omega) = $ Jo's height
$Y(\omega) = $ Ed's height
$(X - Y)(\omega) = X(\omega) - Y(\omega)$: how much taller Jo is than Ed

- Let X and Y be random variables.
- Then we also can consider random variables:
  - $(X + Y)(\omega) = X(\omega) + Y(\omega)$
  - $(X \cdot Y)(\omega) = X(\omega) \cdot Y(\omega)$
  - $(-X)(\omega) = -(X(\omega))$
  - $(\lambda X)(\omega) = \lambda(X(\omega)), \lambda \in \mathbb{R}$

Example: Roll of an Eight-sided and a Six-sided Die

- $\Omega = \{\langle i \text{ on top}, j \text{ on top}\rangle : 1 \leqslant i \leqslant 8, 1 \leqslant j \leqslant 6\}$
- $X(\langle i \text{ on top}, j \text{ on top}\rangle) = i$: Result of the eight-sided die.
- $Y(\langle i \text{ on top}, j \text{ on top}\rangle) = j$: Result of the six-sided die.
- $\max\{X, Y\}(\omega) = \max\{X(\omega), Y(\omega)\}$: The maximum score.
- $(X + Y)(\omega) = X(\omega) + Y(\omega)$: The total score.
- $\{X + Y = 3\} = \{\omega : X(\omega) + Y(\omega) = 3\} =$
  $\{\langle 1 \text{ on top}, 2 \text{ on top}\rangle, \langle 2 \text{ on top}, 1 \text{ on top}\rangle\}$

Expectation Value of a Discrete Random Variable

Probability of X having value r:

$$P(\{X = r\}) = P(\{\omega : X(\omega) = r\})$$

Expected value of X:

$$E[X] = \sum_r r.P(\{X = r\}) = \sum_r r.P(\{\omega : X(\omega) = r\})$$

Example: Roll of a fair die.

- $\Omega = \{1 \text{ on top}, 2 \text{ on top}, \ldots, 6 \text{ on top}\}$
- $X(\{1 \text{ on top}\}) = 1, \ldots, X(\{6 \text{ on top}\}) = 6$
- $P(1) := P(\{X = 1\}) = P(\{1 \text{ on top}\}) = \frac{1}{6}, \ldots$
  $P(6) := P(\{X = 6\}) = P(\{6 \text{ on top}\}) = \frac{1}{6}$

$$E[X] = \sum_{i=1}^{6} i.P(\{X = i\}) = 1.\frac{1}{6} + \cdots + 6.\frac{1}{6} = 3.5$$

Expectation Values of Functions of Random Variables

A: random variable
U: function on the real numbers

$U \circ A$: $U \circ A(\omega) = U(A(\omega))$
$U \circ A$ is a random variable.

$$\begin{tikzcd} \Omega \arrow[r, "A"] \arrow[rr, bend right, "U \circ A"'] & \mathbb{R} \arrow[r, "U"] & \mathbb{R} \end{tikzcd}$$

$(*)$ The law of total probability: $P(C) = \sum_o P(C|D_o) \cdot P(D_o)$.
$(**)$ $P(\{U \circ A = x\}|\{A = o\}) = 1$ iff $x = U(o)$ and 0 otherwise.

$$\begin{aligned} E(U \circ A) &= \sum_x x.P(\{U \circ A = x\}) \\ &\overset{(*)}{=} \sum_o \sum_x x.P(\{U \circ A = x\}|\{A = o\})P(\{A = o\}) \\ &\overset{(**)}{=} \sum_o U(o) \cdot P(\{A = o\}) \end{aligned}$$

Independent, Identically Distributed Random Variables

Can we determine probabilities from frequencies?

A sequence $X_1, X_2, \dots$ of random variables is independent and identically distributed (i. i. d.) if and only if

- X_i is probabilistically independent from X_j for $i \neq j$,
  - i. e. for all $(r_1, r_2), (r_3, r_4) \subseteq \mathbb{R}$, $P(\{X_i \in (r_1, r_2)\} | \{X_j \in (r_3, r_4)\}) = P(\{X_i \in (r_1, r_2)\})$, and
- the probability distribution for X_i is identical to that of X_j,
  - i. e. for all $(r_1, r_2) \subseteq \mathbb{R}, P(\{X_i \in (r_1, r_2)\}) = P(\{X_j \in (r_1, r_2)\})$.

For example a repeated sequence of coin tosses

Sample mean of the initial sequence of a sequence of i. i. d. variables:

$$\overline{X}_n := \frac{X_1 + X_2 + \dots X_n}{n}$$

Laws of Large Numbers

Expectation value (real mean, population mean) of the i. i. d.:
$\mu = E[X_1] = E[X_2] = E[X_3] = \dots$

- Strong law of large numbers (for finite variance):

$$P(\{\lim_{n\to\infty} \overline{X}_n = \mu\}) = 1$$

"The probability of getting to the real mean through infinitely many observations is 1."

- Weak law of large numbers:

$$\text{For all } \varepsilon > 0,\ \lim_{n\to\infty} P(\{|\overline{X}_n - \mu| \leqslant \varepsilon\}) = 1$$

"For any ε, you can improve your chance of getting that close to the real mean through measurement arbitrarily by further observations."