N.I. Lobachevsky State University
of Nizhni Novgorod

Probability theory and mathematical statistics:

Geometric Probability, Statistical Probability, Axiomatic Probability

Associate Professor
A.V. Zorine

Outline

1. Geometric probabilities. Examples.
2. Properies of geometric probability.
3. Statistical probability and its properties. Comparision to geometric probability.
4. Axiomatic definition of probability.
5. General properties of probabilities.
6. Summation theorem for probabilities.
7. Discrete probability space.

Definition

Classical probability is applicable only for experiments with finite number of elementary outcomes. When the number of elementary outcomes is infinite and non-denumerable, it is impossible to assign equal positive probability to each outcome.

Assume Ω can be represented as a segment, or flat figure on a plane, or a body in space. Denote by G its geometric representation.

Definition

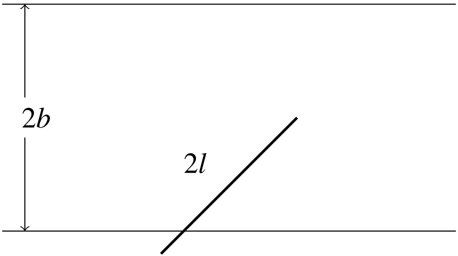
Let $g \subset G$ be a geometric representation of A. Geometric probability $\mathrm{P}(A)$ of event A is defined by

$$\mathrm{P}(A) = \frac{\text{measure of } g}{\text{measure of } G}$$

A measure is either length, or area, or volume depending on G.

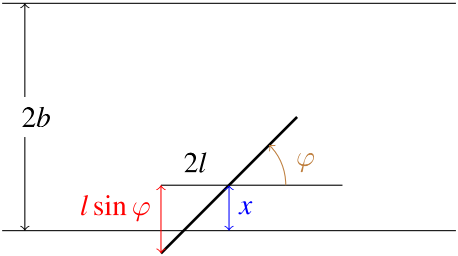
Buffon's needle (1733, 1777)

A needle of length $2l$ is dropped on a floor with equally spaced parallel lines at a distance $2b$ apart. What's the probability that the needle crosses a line?



Buffon's needle (1733, 1777)

A needle of length $2l$ is dropped on a floor with equally spaced parallel lines at a distance $2b$ apart. What's the probability that the needle crosses a line?

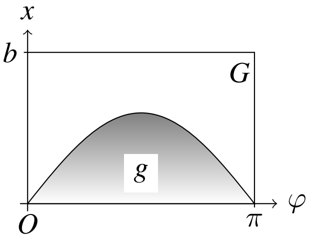


Let x denote the distance between the needle's center and the nearest line, let φ equal the angle between the needle and the nearest line. The needle crosses the line if and only if $x < l\sin\varphi$.

Buffon's needle

Here

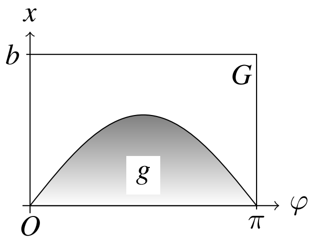
$$\Omega = G = \{(\varphi, x) \colon 0 \leqslant \varphi \leqslant \pi,\ 0 \leqslant x < b\},$$
$$A = g = \{(\varphi, x) \colon 0 \leqslant \varphi \leqslant \pi,\ 0 \leqslant x < l \sin \varphi\}$$



Buffon's needle

Here

$$\Omega = G = \{(\varphi, x) \colon 0 \leqslant \varphi \leqslant \pi,\ 0 \leqslant x < b\},$$
$$A = g = \{(\varphi, x) \colon 0 \leqslant \varphi \leqslant \pi,\ 0 \leqslant x < l \sin\varphi\}$$



The measure of an event is the area of its geometric representation, area of G is $b\pi$, area of g is

$$\int_0^\pi l \sin\varphi \, d\varphi = -l \cos\varphi \Big|_0^\pi = 2l,$$

$$\mathrm{P}(A) = \frac{2l}{\pi b}.$$

Dating problem

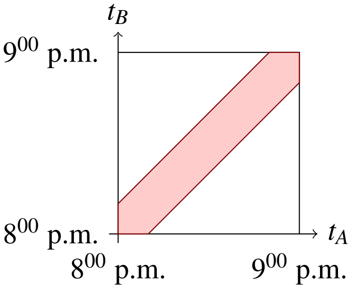
Ann and Bart have a date tonight. They are to meet each other between 8 p. m. and 9 p. m. Each of them waits for 10 minutes for another and then leaves. What's the probability they'll meet?

Dating problem

Ann and Bart have a date tonight. They are to meet each other between 8 p. m. and 9 p. m. Each of them waits for 10 minutes for another and then leaves. What's the probability they'll meet?

Let t_A denote Ann's arrival time (measured in minutes), t_B denote Bart's arrival time (in minutes, too). An elementary outcome is $\omega = (t_A, t_B)$, and the meet each if $|t_A - t_B| \leqslant 10$.

Dating problem



$$\Omega = G = \{(t_A, t_B) \colon 8^{00} \leqslant t_A \leqslant 9^{00},$$
$$8^{00} \leqslant t_B \leqslant 9^{00}\},$$
$$A = g = \Big\{(t_A, t_B) \colon 8^{00} \leqslant t_A \leqslant 9^{00},$$
$$8^{00} \leqslant t_B \leqslant 9^{00}, |t_A - t_B| \leqslant 10\Big\},$$
$$\text{area of } g = 60^2 - 50^2 = 1100,$$
$$\text{area of } G = 60^2,$$
$$\mathrm{P}(A) = \frac{60^2 - 50^2}{60^2} = \frac{11}{36}.$$

Properties of geometric probability

Corollary

For any elementary outcome $E = \{\omega\}$, $\mathrm{P}(A) = 0$.

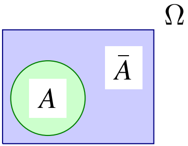
Proof. A length of a point is 0, an area of a point is 0, etc.

Corollary

For any event A, $\mathrm{P}(\bar{A}) = 1 - \mathrm{P}(A)$.

Proof.

$$\text{measure of } \bar{A} = \text{measure of } \Omega - \text{measure of } A$$



Corollary

For any events A and B,

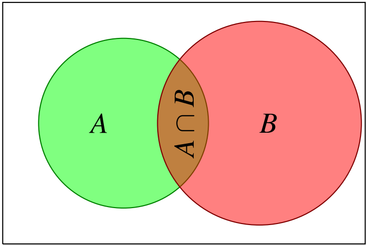
$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

If A and B are mutually exclusive, then

$$P(A \cup B) = P(A) + P(B).$$

Proof.

$$\text{measure of } A \cup B = \text{measure of } A + \text{measure of } B - \text{measure of } A \cap B$$



Event A is not necessarily impossible when $\mathrm{P}(A) = 0$.
Events A and B are not necessarily mutually exclusive when $\mathrm{P}(A \cap B) = 0$.

Definition

Denote by $\mu(A,N)$ the number of occurrences of event A in N repetitions of an experiment. A statistical probability of event A is defined by

$$\mathrm{P}(A) = \frac{\mu(A,N)}{N}$$

Properties

$$0 \leqslant \mathrm{P}(A) \leqslant 1, \quad \mathrm{P}(\Omega) = 1,$$
$$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B),$$
$$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) \quad \text{when } A \cap B = \varnothing.$$

Statistical probability for Buffon's needle

For a statistically stable random experiment,

$$\mathrm{P}_{\text{classical}}(A) \approx \mathrm{P}_{\text{statistical}}(A)$$

or

$$\mathrm{P}_{\text{geometric}}(A) \approx \mathrm{P}_{\text{statistical}}(A)$$

Wolf, 1849: needle's length $2l = 36$ mm, distance between lines $2b = 45$ mm, $N = 500$ trials, $\mu(A, N) = 2532$, $\mathrm{P}(A) \approx 0{,}5064$. From

$$\frac{2l}{\pi b} \approx 0{,}5064$$

we obtain

$$\pi \approx 3{,}159$$

Exact value of π is $3{,}14159\ldots$

Buffon's needle

Who?	When?	How many times?	Observed π
Wolf	1850	5000	3,1596
Smith	1855	3204	3,1553
Fox	1894	1120	3,1419
Lazzarini	1901	3408	3,1415929
π with 8 decimal places			3,1415927

Property	Classical probability	Geometric probability	Statistical probability
$\mathrm{P}(A) \geqslant 0$			

Property	Classical probability	Geometric probability	Statistical probability
$\mathrm{P}(A) \geqslant 0$	yes	yes	yes
$\mathrm{P}(\Omega) = 1$			

Property	Classical probability	Geometric probability	Statistical probability
$\mathrm{P}(A) \geqslant 0$	yes	yes	yes
$\mathrm{P}(\Omega) = 1$	yes	yes	yes
$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$ when $A \cap B = \varnothing$			

Property	Classical probability	Geometric probability	Statistical probability
$\mathrm{P}(A) \geqslant 0$	yes	yes	yes
$\mathrm{P}(\Omega) = 1$	yes	yes	yes
$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$ when $A \cap B = \varnothing$	yes	yes	yes

Property	Classical probability	Geometric probability	Statistical probability
$\mathrm{P}(A) \geqslant 0$	yes	yes	yes
$\mathrm{P}(\Omega) = 1$	yes	yes	yes
$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$ when $A \cap B = \varnothing$	yes	yes	yes

A.N. Kolmogorov proposed in 1933 a set of axioms for probability in general case. His main idea: any probability should behave as probabilities above.

A set of observable events should be closed under main operations: union, intersection, etc.

Definition

A set $\mathfrak{F}$ of events is called sigma-algebra (σ-algebra) if

- $\Omega \in \mathfrak{F}$
- $\bar{A} \in \mathfrak{F}$ if $A \in \mathfrak{F}$
- $A_1 \cup A_2 \cup \ldots \in \mathfrak{F}$ if $A_1 \in \mathfrak{F}, A_2 \in \mathfrak{F}, \ldots$

A set of observable events should be closed under main operations: union, intersection, etc.

Definition

A set $\mathfrak{F}$ of events is called sigma-algebra (σ-algebra) if

- $\Omega \in \mathfrak{F}$
- $\bar{A} \in \mathfrak{F}$ if $A \in \mathfrak{F}$
- $A_1 \cup A_2 \cup \ldots \in \mathfrak{F}$ if $A_1 \in \mathfrak{F}, A_2 \in \mathfrak{F}, \ldots$

Theorem

$A_1 \cap A_2 \cap \ldots \in \mathfrak{F}$ if $A_1 \in \mathfrak{F}, A_2 \in \mathfrak{F}, \ldots$

Proof. $A_1 \cap A_2 \cap \ldots = \overline{\bar{A}_1 \cup \bar{A}_2 \cup \ldots} \in \mathfrak{F}$, because $\bar{A}_1 \in \mathfrak{F}, \bar{A}_2 \in \mathfrak{F}$, $\ldots$, and $\bar{A}_1 \cup \bar{A}_2 \cup \ldots \in \mathfrak{F}$.

Axiomatic definition of probability

Suppose every event $A \in \mathfrak{F}$ is assigned with a number $\mathrm{P}(A)$ so that

1. $\mathrm{P}(A) \geqslant 0$ for any $A \in \mathfrak{F}$
2. $\mathrm{P}(\Omega) = 1$
3. $\mathrm{P}(A_1 \cup A_2 \cup \ldots) = \mathrm{P}(A_1) + \mathrm{P}(A_2) + \ldots$ if $A_i \cap A_j = \varnothing$ for $i \neq j$

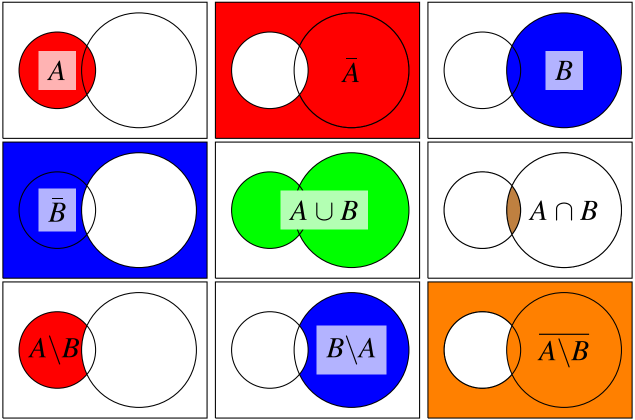
Then $\mathrm{P}(A)$ is called a *probability of event A*.

Here $\mathfrak{F}$ is a domain of P. Axioms 1)–3) are not contradictious, they are incomplete as well. They don't define a unique probability, and we know that some probabilities satisfying the axioms exist.

A triple $(\Omega, \mathfrak{F}, \mathrm{P})$ is called a *probability space*.

The third axiom is called *axiom of denumerable additivity*. A weaker condition of *finite additivity* $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B)$ for mutually exclusive events A and B is not sufficient to construct deep theory.

In what follows it is useful to recall definitions of events operations.



Theorem

$$P(\varnothing) = 0$$

Proof. Since $\Omega \cup \varnothing = \Omega$ and events Ω and $\varnothing$ are mutually exclusive, by axiom 3 obtain

$$P(\Omega \cup \varnothing) = P(\Omega) + P(\varnothing)$$

Using axiom 1, we have

$$1 = 1 + P(\varnothing)$$

Thus, $P(\varnothing) = 0$.

Theorem

$$\mathrm{P}(\bar{A}) = 1 - \mathrm{P}(A)$$

Proof. Since A and $\bar{A}$ are mutually exclusive and $A \cup \bar{A} = \Omega$, we have

$$\mathrm{P}(A) + \mathrm{P}(\bar{A}) = \mathrm{P}(\Omega)$$

and

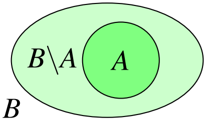
$$\mathrm{P}(\bar{A}) = \mathrm{P}(\Omega) - \mathrm{P}(A) = 1 - \mathrm{P}(A).$$

Theorem (monotonicity)

Assume $A \subset B$ (A implies B), then

$$\mathrm{P}(A) \leqslant \mathrm{P}(B) \qquad \text{and} \qquad \mathrm{P}(B \backslash A) = \mathrm{P}(B) - \mathrm{P}(A)$$

Proof. Observe that $B = A \cup (B \backslash A)$ and $A \cap (B \backslash A) = \varnothing$.



$$\mathrm{P}(B) = \mathrm{P}(A) + \mathrm{P}(B \backslash A) \geqslant \mathrm{P}(A)$$

by axiom 1. Now,

$$\mathrm{P}(B \backslash A) = \mathrm{P}(B) - \mathrm{P}(A)$$

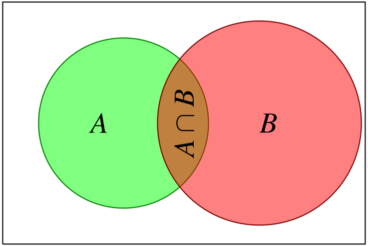
Theorem

For any not mutually exclusive events A, B one has

$$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$$

Proof.

$$A \cup B = A \cup (B \backslash A) = A \cup (B \backslash (A \cap B)),$$
$$A \cap (B \backslash (A \cap B)) = \varnothing,$$
$$A \cap B \subset B,$$
$$\begin{aligned}\mathrm{P}(A \cup B) &= \mathrm{P}(A) + \mathrm{P}(B \backslash (A \cap B)) \\ &= \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)\end{aligned}$$



Extensions of summation theorem

Theorem

For any events A, B, C

$$\begin{aligned} \mathrm{P}(A \cup B \cup C) = \mathrm{P}(A) + \mathrm{P}(B) + \mathrm{P}(C) - \mathrm{P}(A \cap B) \\ - \mathrm{P}(A \cap C) - \mathrm{P}(B \cap C) + \mathrm{P}(A \cap B \cap C) \end{aligned}$$

Extensions of summation theorem

Theorem

For any events A, B, C

$$\begin{aligned}\mathrm{P}(A\cup B\cup C) = \mathrm{P}(A)+\mathrm{P}(B)+\mathrm{P}(C)-\mathrm{P}(A\cap B)\\ -\mathrm{P}(A\cap C)-\mathrm{P}(B\cap C)+\mathrm{P}(A\cap B\cap C)\end{aligned}$$

Proof.

$$\begin{aligned}\mathrm{P}(A\cup B\cup C) &= \mathrm{P}(A)+\mathrm{P}(B\cup C)-\mathrm{P}(A\cap(B\cup C))\\ &= \mathrm{P}(A)+\mathrm{P}(B)+\mathrm{P}(C)-\mathrm{P}(B\cap C)-\mathrm{P}((A\cap B)\cup(A\cap C))\\ &= \mathrm{P}(A)+\mathrm{P}(B)+\mathrm{P}(C)-\mathrm{P}(B\cap C)-\mathrm{P}(A\cap B)\\ &\quad -\mathrm{P}(A\cap C)+\mathrm{P}(A\cap B\cap A\cap C).\end{aligned}$$

Extensions of summation theorem (Cont.)

Theorem

For any events $A_1, A_2, \ldots, A_n$,

$$\begin{aligned}
P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n) \\
- P(A_1 \cap A_2) - P(A_1 \cap A_3) \ldots - P(A_{n-1} \cap A_n) \\
+ P(A_1 \cap A_2 \cap A_3) + \ldots + P(A_{n-2} \cap A_{n-1} \cap A_n) \\
+ (-1)^{n-1} P(A_1 \cap A_2 \cap \ldots \cap A_n)
\end{aligned}$$

Proof by induction.

Problem. If n dice are thrown at a time, what is the probability of having each of the points 1, 2, 3, 4, 5, 6 appear at least once?

Problem. If n dice are thrown at a time, what is the probability of having each of the points 1, 2, 3, 4, 5, 6 appear at least once?

Denote $A_i =$ "i points appear at least once", $i = 1, 2, \ldots, 6$; We want $\mathrm{P}(A_1 \cap A_2 \cap \ldots \cap A_6)$.

$$\mathrm{P}(A_1 \cap A_2 \cap \ldots \cap A_6) = 1 - \mathrm{P}(\bar{A}_1 \cup \bar{A}_2 \cup \ldots \cup \bar{A}_6)$$

$$= 1 - \Big(\mathrm{P}(\bar{A}_1) + \ldots + \mathrm{P}(\bar{A}_6) - \mathrm{P}(\bar{A}_1 \cap \bar{A}_2) - \ldots - \mathrm{P}(\bar{A}_5 \cap \bar{A}_6)$$

$$+\mathrm{P}(\bar{A}_1 \cap \bar{A}_2 \cap \bar{A}_3) + \ldots + (-1)^6 \mathrm{P}(\bar{A}_1 \cap \bar{A}_2 \cap \ldots \cap \bar{A}_6)\Big)$$

$$\mathrm{P}(\bar{A}_i) = \frac{5^n}{6^n}, \qquad \mathrm{P}(\bar{A}_i \cap \bar{A}_j) = \frac{4^n}{6^n}, \qquad \ldots$$

$$\mathrm{P}(A_1 \cap A_2 \cap \ldots \cap A_6) = 1 - 6\left(\frac{5}{6}\right)^n + 15\left(\frac{4}{6}\right)^n$$

$$-20\left(\frac{3}{6}\right)^n + 15\left(\frac{2}{6}\right)^n - 6\left(\frac{1}{6}\right)^n$$

Discrete probability space

Let $\Omega = \{\omega_1, \omega_2, \ldots\}$ be a finite or denumerable infinite set, each outcome $\{\omega_i\}$, $i = 1, 2, \ldots$ is assigned a non negative number $p_i = p(\omega_i)$ in such a way that

$$p_1 + p_2 + \ldots = 1$$

Put by definition

$$\mathrm{P}(A) = \sum_{\omega \in A} p(\omega)$$

Then P is (discrete) probability and $(\Omega, \mathfrak{F}, \mathrm{P})$ is called a *discrete probability space*. For example, when Ω is finite, putting

$$p(\omega) = \frac{1}{N(\Omega)}$$

we obtain classical probability. Another example: this kind of probability spaces is often used to model infinite sequences of repeated trials.