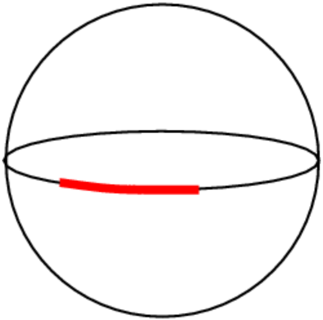
The Borel-Kolmogorov Paradox and conditional expectations

Zalán Gyenis
Rényi Institute of Mathematics, Budapest
Gábor Hofer-Szabó
Research Centre for the Humanities, Budapest
Miklós Rédei
London School of Economics

What is the conditional probability that a randomly chosen point is on an arc of a great circle of the sphere on the condition that it lies on that great circle?

Tension:

1. **Intuition:** the conditional probability is proportional to the length of the arc.
2. **Fact**: since a great circle has surface measure zero, Bayes' formula cannot be used to calculate the conditional probability.



Main messages

1. Use **conditional expectation** to conditionalize.
2. This will allow for conditionalizing on **probability zero** events, as in case of the Borel-Kolmogorov Paradox.
3. Using conditional expectation the Borel-Kolmogorov Paradox is **not paradoxical**.

Historical remarks

- The Borel-Kolmogorov Paradox was formulated by Borel in 1909 before Kolmogorov's (1933) measure-theoretic probability theory.
- Kolmogorov's own resolution of the Paradox is based on the theory of conditional expectation.
- Since Kolmogorov's work, conditional expectation is the standard device for conditionalization in probability theory.

Outline

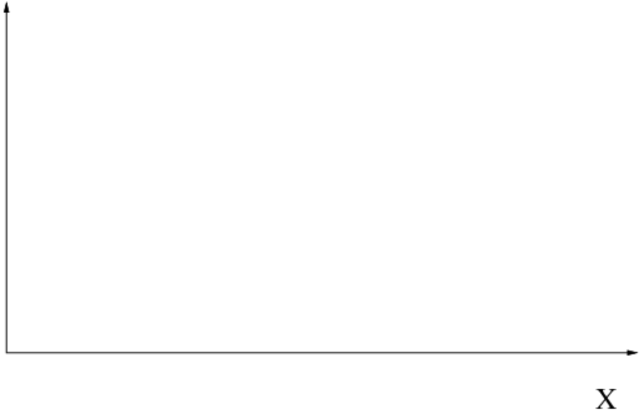
Conditioning using conditional expectation

Conditional expectation

- Conditional expectation is a coarse-graining.

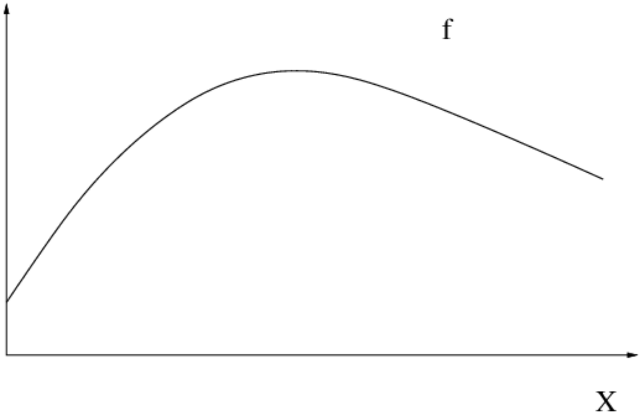
Conditional expectation

- Conditional expectation is a coarse-graining.
- $(X, \mathcal{S}, p)$: probability measure space



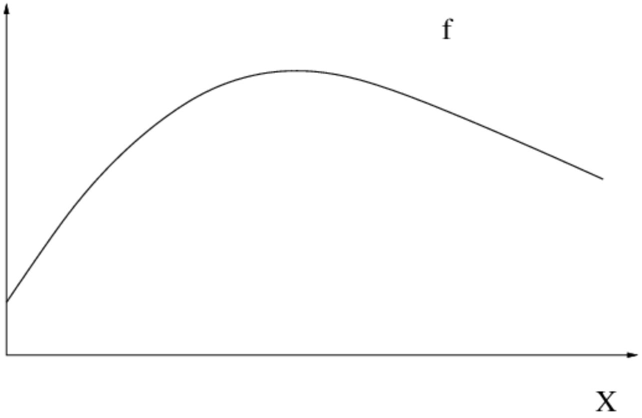
Conditional expectation

- $\mathcal{L}^1(X, \mathcal{S}, p)$: set of p-integrable functions



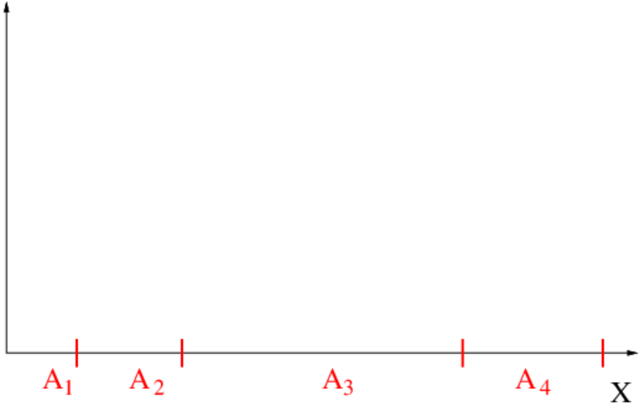
Conditional expectation

- $\mathcal{L}^1(X, \mathcal{S}, p)$: set of p-integrable functions
- p defines a functional $\phi_p(f) \doteq \int_X f\, dp, \qquad f \in \mathcal{L}^1(X, \mathcal{S}, p)$



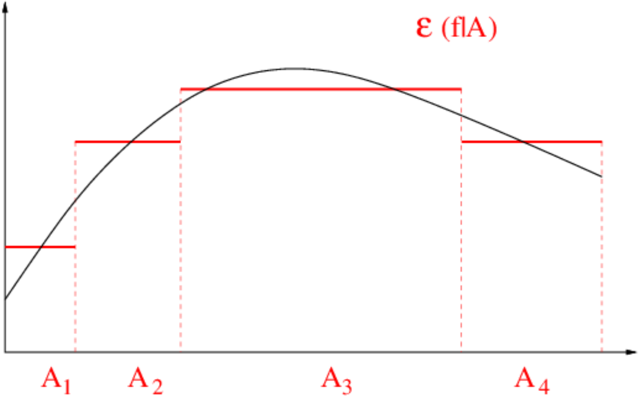
Conditional expectation

- $(X, \mathcal{A}, p_{\mathcal{A}})$: coarse-grained probability measure space



Conditional expectation

- $\mathscr{E}(\cdot|\mathcal{A}) : \mathcal{L}^1(X, \mathcal{S}, p) \to \mathcal{L}^1(X, \mathcal{A}, p_{\mathcal{A}})$: conditional expectation



Conditional expectation

Definition

A map

$$\mathscr{E}(\cdot|\mathcal{A})\colon \mathcal{L}^1(X,\mathcal{S},p) \to \mathcal{L}^1(X,\mathcal{A},p_{\mathcal{A}})$$

is called an **$\mathcal{A}$-conditional expectation** if:

(i) $\mathscr{E}(f|\mathcal{A})$ is $\mathcal{A}$-measurable for all $f \in \mathcal{L}^1(X,\mathcal{S},p)$;

(ii) $\mathscr{E}(\cdot|\mathcal{A})$ preserves the integration on elements of $\mathcal{A}$:

$$\int_Z \mathscr{E}(f|\mathcal{A})dp_{\mathcal{A}} = \int_Z f dp \qquad \forall Z \in \mathcal{A}.$$

Bayesian statistical inference

Bayesian statistical inference

Problem of statistical inference (informal)

Given a probability measure $p'_{\mathcal{A}}$ on $\mathcal{A}$, what is its **extension** to $\mathcal{S}$?

Bayesian statistical inference

Problem of statistical inference (informal)

Given a probability measure $p'_{\mathcal{A}}$ on $\mathcal{A}$, what is its **extension** to $\mathcal{S}$?

Reformulated in terms of functionals:

Problem of statistical inference (informal)

Given a probability measure $p'_{\mathcal{A}}$ on $\mathcal{A}$, what is its **extension** to $\mathcal{S}$?

Reformulated in terms of functionals:

Problem of statistical inference

Given a continuous linear functional $\phi'_{\mathcal{A}}$ on $\mathcal{L}^1(X, \mathcal{A}, p_{\mathcal{A}})$, what is its **extension** to $\mathcal{L}^1(X, \mathcal{S}, p)$?

Problem of statistical inference (informal)

Given a probability measure $p'_{\mathcal{A}}$ on $\mathcal{A}$, what is its **extension** to $\mathcal{S}$?

Reformulated in terms of functionals:

Problem of statistical inference

Given a continuous linear functional $\phi'_{\mathcal{A}}$ on $\mathcal{L}^1(X, \mathcal{A}, p_{\mathcal{A}})$, what is its **extension** to $\mathcal{L}^1(X, \mathcal{S}, p)$?

There is no unique answer, but:

Bayesian statistical inference

Problem of statistical inference (informal)

Given a probability measure $p'_{\mathcal{A}}$ on $\mathcal{A}$, what is its **extension** to $\mathcal{S}$?

Reformulated in terms of functionals:

Problem of statistical inference

Given a continuous linear functional $\phi'_{\mathcal{A}}$ on $\mathcal{L}^1(X, \mathcal{A}, p_{\mathcal{A}})$, what is its **extension** to $\mathcal{L}^1(X, \mathcal{S}, p)$?

There is no unique answer, but:

Bayesian statistical inference

Let the **extension** ϕ' be:

$$\phi'(f) \doteq \phi'_{\mathcal{A}}(\mathscr{E}(f|\mathcal{A})) \qquad \forall f \in \mathcal{L}^1(X, \mathcal{S}, p)$$

Definition

The **conditional probability** $p'(B)$ is simply the application of the Bayesian statistical inference to the characteristic function χ_B:

$$p'(B) \doteq \phi'_{\mathcal{A}}(\mathscr{E}(\chi_B|\mathcal{A}))$$

Conditional probability

Remarks

1. $p'(B)$ **depends** on three factors: $\mathcal{A}$, $\mathcal{E}(\cdot|\mathcal{A})$ and $\phi'_{\mathcal{A}}$

Conditional probability

Remarks

1. $p'(B)$ **depends** on three factors: $\mathcal{A}$, $\mathcal{E}(\cdot|\mathcal{A})$ and $\phi'_{\mathcal{A}}$
2. In the special case when
   - $\mathcal{A}$ is generated by $\{A, A^{\perp}\}$,
   - $p(A) \neq 0$;
   - $p'_{\mathcal{A}}(A) = 1$,

   the conditional probability can be given by the **Bayes formula**:

$$p'(B) = \frac{p(B \cap A)}{p(A)}$$

Conditional probability

Remarks

1. $p'(B)$ **depends** on three factors: $\mathcal{A}$, $\mathcal{E}(\cdot|\mathcal{A})$ and $\phi'_{\mathcal{A}}$
2. In the special case when
   - $\mathcal{A}$ is generated by $\{A, A^{\perp}\}$,
   - $p(A) \neq 0$;
   - $p'_{\mathcal{A}}(A) = 1$,

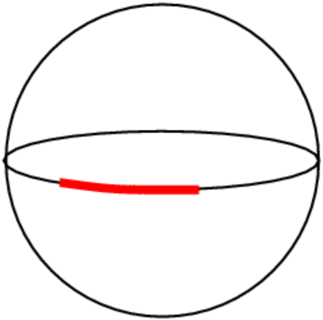
   the conditional probability can be given by the **Bayes formula**:

$$p'(B) = \frac{p(B \cap A)}{p(A)}$$

3. Elements of $\mathcal{A}$ can also have **zero prior probability** p. Hence, it is possible to obtain conditional probabilities with respect to probability zero conditioning events if one uses conditional expectations.

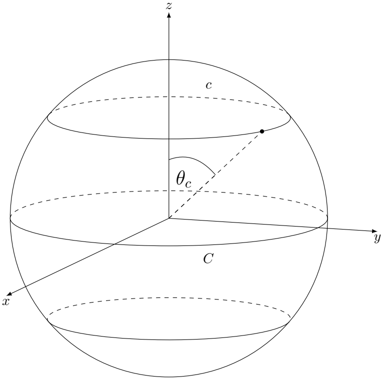
The Borel-Kolmogorov Paradox

Let $(S, \mathcal{B}(S), p)$ be the probability measure space on the sphere S with the uniform probability p on S. What is the conditional probability of being on an arc on a great circle C on condition of being on a great circle C of S?



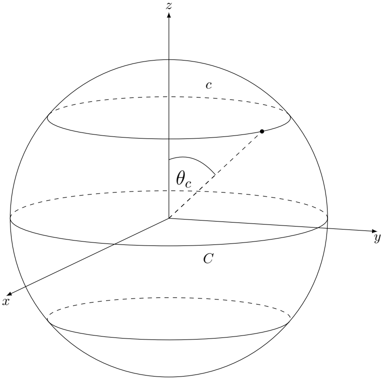
The Borel-Kolmogorov Paradox

$\mathcal{O}$: generated by Borel measurable sets of circles **parallel** to C



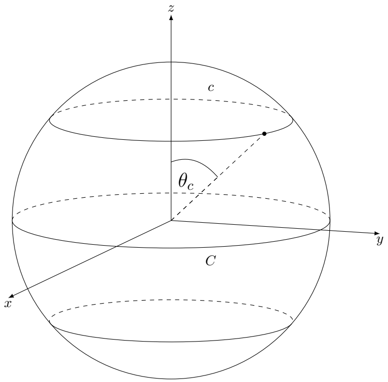
The Borel-Kolmogorov Paradox

$\mathcal{O}$: generated by Borel measurable sets of circles **parallel** to C

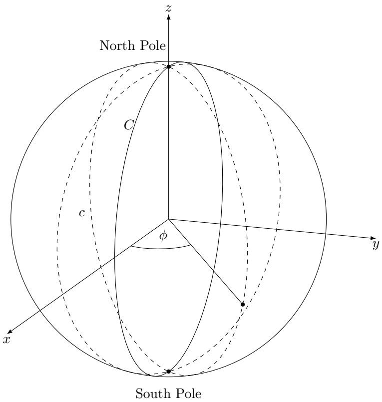


Uniform conditional probability

$\mathcal{O}$: generated by Borel measurable sets of circles **parallel** to C

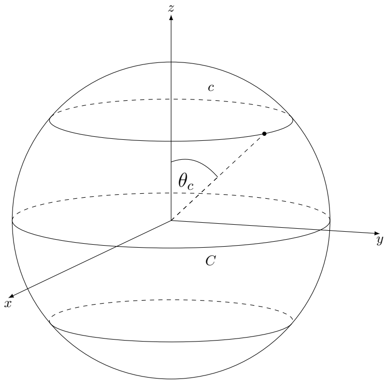


$\mathcal{M}$: generated by Borel measurable sets of **meridian** circles containing C



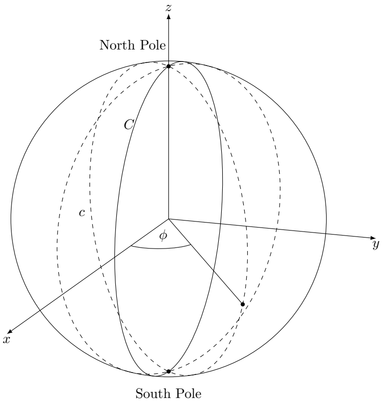
Uniform conditional probability

$\mathcal{O}$: generated by Borel measurable sets of circles **parallel** to C



Uniform conditional probability

$\mathcal{M}$: generated by Borel measurable sets of **meridian** circles containing C



Non-uniform conditional probability

Remarks

Remark 1

Standard view

> *"... we have different conditional distributions depending on how we describe the circle." (Myrvold, 2014)*

The Borel-Kolmogorov Paradox is paradoxical because the conditional probabilities of the **same event** on the **same conditioning events** should not depend on the different parametrizations (violation of the Labelling Irrelevance).

Remark 1

Our view

- Two Boolean algebras of random events are the **same** if there is an isomorphism between them.

Remark 1

Our view

- Two Boolean algebras of random events are the **same** if there is an isomorphism between them.
- **Proposition**. There exists no isomorphism between $\mathcal{O}$ and $\mathcal{M}$.

Remark 1

Our view

- Two Boolean algebras of random events are the **same** if there is an isomorphism between them.
- **Proposition**. There exists no isomorphism between $\mathcal{O}$ and $\mathcal{M}$.

The Borel-Kolmogorov Paradox merely displays a sensitive dependence of conditional probabilities of the **same event** on **different conditioning Boolean subalgebras** with respect to which conditional probabilities are defined in terms of conditional expectations. These conditional probabilities are answers to different questions – not different answers to the same question.

Remark 2

Another standard view

Only the uniform conditional distribution is intuitively correct.

Remark 2

Our view

Both conditional distributions are intuitively correct.

Remark 2

Our view

Both conditional distributions are intuitively correct.

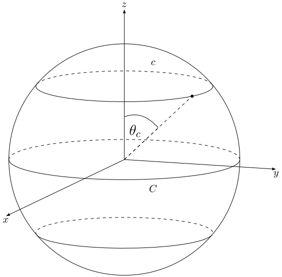
One obtains a **uniform** distribution on the sphere by generating points:

Remark 2

Our view

Both conditional distributions are intuitively correct.

One obtains a **uniform** distribution on the sphere by generating points: **uniformly** on all circles in $\mathcal{O}$.



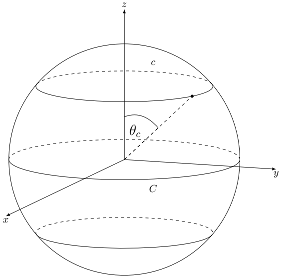
Remark 2

Our view

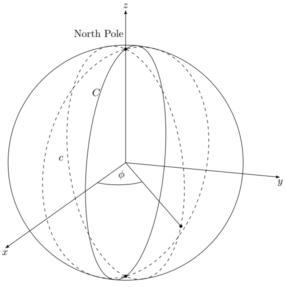
Both conditional distributions are intuitively correct.

One obtains a **uniform** distribution on the sphere by generating points:

uniformly on all circles in $\mathcal{O}$.



non-uniformly on all meridian circles in $\mathcal{M}$.



Remark 2

Our view

- The intuition – that only the uniform conditional distribution is correct – might come from the intutition that the uniform length measure is singled out by pure **probabilistic means**.

Remark 2

Our view

- The intuition – that only the uniform conditional distribution is correct – might come from the intutition that the uniform length measure is singled out by pure **probabilistic means**.
- To single out the uniform length measure one needs to use **further non-probabilistic mathematical conditions**, e.g.

Remark 2

Our view

- The intuition – that only the uniform conditional distribution is correct – might come from the intutition that the uniform length measure is singled out by pure **probabilistic means**.
- To single out the uniform length measure one needs to use **further non-probabilistic mathematical conditions**, e.g.
  - **group-theoretic**: it is the unique measure invariant with respect to the subgroup of rotations on the sphere;
  - **geometric**: it is the restriction of the Lebesgue measure to the circle as differentiable manifold.

Summary

1. The proper mathematical device to handle conditional probabilities is the theory of **conditional expectation**.
2. This theory makes it possible to conditionalize on **probability zero events**, as in the case of the Borel-Kolmogorov Paradox.
3. Obtaining different conditional probabilities is **not paradoxical** because
   - they cannot be regarded as conditional probabilities of the **same event** with respect to the same conditioning events;
   - both are intuitively correct if seen as describing **generation of points** on sphere with uniform distribution.

References

- Z. Gyenis, G. Hofer-Szabó, and M. Rédei, "Conditioning using conditional expectation: the Borel-Kolmogorov Paradox," (submitted).
- A.N. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, (Springer, Berlin, 1933); English translation: *Foundations of the Theory of Probability*, (Chelsea, New York, 1956).
- W. Myrvold, "You can't always get what you want: Some considerations regarding conditional probabilities," *Erkenntnis*, **80**, 573-603 (2015).
- M. Rescorla, "Some epistemological ramifications of the Borel-Kolmogorov Paradox," *Synthese*, **192**, 735-767 (2015).