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.
Reactions

- Radical: the ‘ratio analysis’ of the conditional probability is wrong; one should take conditional probability as a primitive notion.
- Conservative: the Borel-Kolmogorov Paradox can be resolved in the measure-theoretic probability theory.
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**.
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.
1. Conditioning using conditional expectation

2. The Borel-Kolmogorov Paradox

3. Remarks
Conditioning using conditional expectation
Conditional expectation

- Conditional expectation is a coarse-graining.
Conditional expectation is a coarse-graining.

\((X, S, \rho)\): probability measure space
Conditional expectation

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

- $L^1(X, S, p)$: set of $p$-integrable functions
- $p$ defines a functional $\phi_p(f) = \int_X f \, dp$, $f \in L^1(X, S, p)$
Conditional expectation

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

\[ \mathcal{E}(\cdot | A) : \mathcal{L}^1(X, \mathcal{S}, \mu) \to \mathcal{L}^1(X, \mathcal{A}, \mu_A) : \text{conditional expectation} \]
Conditional expectation

Definition

A map

\[ \mathcal{E}(\cdot | \mathcal{A}) : \mathcal{L}^1(X, S, \rho) \rightarrow \mathcal{L}^1(X, \mathcal{A}, \rho_A) \]

is called an \textbf{A-conditional expectation} if:

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

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

\[ \int_Z \mathcal{E}(f | \mathcal{A}) d\rho_A = \int_Z f d\rho \quad \forall Z \in \mathcal{A}. \]
Remarks

1. The conditional expectation always exists.
2. It is **unique** only **up to measure zero**. Different conditional expectations equal up to measure zero are called **versions**.
3. If generated by a **countable partition** \( \{A_i\} \) such that \( p(A_i) \neq 0 \) for any \( i \), then:

\[
\mathcal{E}(\chi_B|A) = \sum_i \frac{p(B \cap A_i)}{p(A_i)} \chi_{A_i} \quad \forall B \in S
\]
Bayesian statistical inference

Problem of statistical inference (informal)

Given a probability measure $p'$ on $A$, what is its extension to $S$?

Reformulated in terms of functionals:

Problem of statistical inference

Given a continuous linear functional $\phi'$ on $L_1(X, A, p_A)$, what is its extension to $L_1(X, S, p)$?

There is no unique answer, but:

Bayesian statistical inference

Let the extension $\phi'$ be:

$$
\phi'(f) = \phi'(A)(E(f|A))
$$

for all $f \in L_1(X, S, p)$.
Bayesian statistical inference

Problem of statistical inference (informal)

Given a probability measure $p'_A$ on $A$, what is its extension to $S$?
Bayesian statistical inference

Problem of statistical inference (informal)

Given a probability measure $p'_A$ on $A$, what is its extension to $S$?

Reformulated in terms of functionals:
Problem of statistical inference (informal)

Given a probability measure $p'_A$ on $A$, what is its extension to $S$?

Reformulated in terms of functionals:

Problem of statistical inference

Given a continuous linear functional $\phi'_A$ on $L^1(X, A, p_A)$, what is its extension to $L^1(X, S, p)$?
Problem of statistical inference (informal)

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

Reformulated in terms of functionals:

Problem of statistical inference

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

There is no unique answer, but:

\[ \phi'_A(f) = \phi'_A(E(f | A)) \forall f \in L^1(X, S, p) \]
Problem of statistical inference (informal)

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

Reformulated in terms of functionals:

Problem of statistical inference

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

There is no unique answer, but:

Bayesian statistical inference

Let the extension $\phi'$ be:

$$\phi'(f) \doteq \phi'_A(\mathbb{E}(f|\mathcal{A})) \quad \forall f \in L^1(X, S, p)$$
Definition

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

$$p'(B) = \phi'_A(\mathcal{E}(\chi_B|A))$$
1. $p'(B) = \phi'_A(\mathcal{E}(\chi_B|A))$ should properly be called the $(A, p'_A)$-conditional probability of $B$, since it depends on three factors: $A$, $\mathcal{E}(\cdot|A)$ and $\phi'_A$. 

Remarks
Remarks

1. $p'(B) \doteq \phi'_A(\mathcal{E}(\chi_B|A))$ should properly be called the 
   \((A, p'_A)\)-conditional probability of \(B\), since it depends on three 
   factors: \(A\), \(\mathcal{E}(\cdot|A)\) and \(\phi'_A\).

2. In the special case when
   - \(A\) is generated by \(\{A, A^\perp\}\),
   - \(p(A) \neq 0\); 
   - \(p'_A(A) = 1\),
   the conditional probability can be given by the Bayes formula:

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

1. $p'(B) := \phi'_A(\mathcal{E}(\chi_B|A))$ should properly be called the $(A, p'_A)$-conditional probability of $B$, since it depends on three factors: $A$, $\mathcal{E}(\cdot|A)$ and $\phi'_A$.

2. In the special case when
   - $A$ is generated by $\{A, A^\perp\}$,
   - $p(A) \neq 0$;
   - $p'_A(A) = 1$,

the conditional probability can be given by the Bayes formula:

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

3. Elements of $A$ 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 \(S\)?
\( \mathcal{O} \): generated by Borel measurable sets of circles parallel to \( C \)
\( \mathcal{O} \): generated by Borel measurable sets of circles parallel to \( C \) 

\((\mathcal{O}, \rho'_{\mathcal{O}})\)-conditional probability:

The \((\mathcal{O}, \rho'_{\mathcal{O}})\)-conditional probability is uniform on the great circle.
\[ \mathcal{O} \]: generated by Borel measurable sets of circles parallel to \( C \)

\((\mathcal{O}, p'_\mathcal{O})\)-conditional probability:

- Consider the canonical version of the \( \mathcal{O} \)-conditional expectation \( E_\mathcal{O}(\cdot | \mathcal{O}) \)
**The Borel-Kolmogorov Paradox**

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

\((\mathcal{O}, p'_\mathcal{O})\)-**conditional probability**:
- Consider the canonical version of the \( \mathcal{O} \)-conditional expectation \( \mathcal{E}_\mathcal{O}(\cdot | \mathcal{O}) \)
- Fix \( p'_\mathcal{O} \) on \( \mathcal{O} \) by
  
  \[
  p'_\mathcal{O}(C) = 1 \quad \text{and} \quad p'_\mathcal{O}(C^\perp) = 0
  \]
The Borel-Kolmogorov Paradox

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

\((\mathcal{O}, p'_\mathcal{O})\)-conditional probability:
- Consider the canonical version of the \( \mathcal{O} \)-conditional expectation \( \mathcal{E}_\mathcal{O}(\cdot|\mathcal{O}) \)
- Fix \( p'_\mathcal{O} \) on \( \mathcal{O} \) by
\[
\begin{align*}
p'_\mathcal{O}(C) &= 1 \quad \text{and} \quad p'_\mathcal{O}(C^\perp) = 0
\end{align*}
\]
- Compute the \((\mathcal{O}, p'_\mathcal{O})\)-conditional probability
The Borel-Kolmogorov Paradox

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

\((\mathcal{O}, p'_\mathcal{O})\)-conditional probability:

- Consider the canonical version of the \( \mathcal{O} \)-conditional expectation \( \mathcal{E}_\mathcal{O}(\cdot|\mathcal{O}) \)
- Fix \( p'_\mathcal{O} \) on \( \mathcal{O} \) by
  
  \[
  p'_\mathcal{O}(C) = 1 \quad \text{and} \quad p'_\mathcal{O}(C^\perp) = 0
  \]
- Compute the \((\mathcal{O}, p'_\mathcal{O})\)-conditional probability

**Result:**

- The \((\mathcal{O}, p'_\mathcal{O})\)-conditional probability is **uniform** on the great circle.
\( \mathcal{M} \): generated by Borel measurable sets of meridian circles containing \( C \)
\[ \mathcal{M}: \text{generated by Borel measurable sets of meridian circles containing } C \]

\[(\mathcal{M}, p'_\mathcal{M})\text{-conditional probability:}\]
- Consider the canonical version of the \(\mathcal{M}\)-conditional expectation \(E_\mathcal{M}(\cdot|\mathcal{M})\)
- Fix \(p'_\mathcal{M}\) on \(\mathcal{M}\) by
  \[ p'_\mathcal{M}(C) = 1 \quad \text{and} \quad p'_\mathcal{M}(C^\perp) = 0 \]
- Compute the \((\mathcal{M}, p'_\mathcal{M})\)-conditional probability

\textbf{Result:}
- The \((\mathcal{M}, p'_\mathcal{M})\)-conditional probability is not uniform on the great circle.
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.
Our view

- **Remark 1**: Our view

- **Remark 2**: Two Boolean algebras of random events are the same if there is an isomorphism between them.

- **Proposition**: There exists no isomorphism between $O$ and $M$. 

- **The Borel-Kolmogorov Paradox** merely displays a sensitive dependence of conditional probabilities of the same event on different 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 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 $O$ and $M$.

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.
Another standard view

Only the uniform conditional distribution is intuitively correct.
Remark 2

Our view

Both the \((O, p'_O)\)- and the \((M, p'_M)\)-conditional distributions are intuitively correct.
Remark 2

**Our view**

Both the \((O, p'_O)\)- and the \((M, p'_M)\)-conditional distributions are intuitively correct.

By generating points:
Remark 2

Our view

Both the \((\mathcal{O}, p'_O)\)- and the \((\mathcal{M}, p'_M)\)-conditional distributions are intuitively correct.

By generating points:

**uniformly** on all circles in \(\mathcal{O}\).
Remark 2

Our view

Both the \((\mathcal{O}, p'_\mathcal{O})\)- and the \((\mathcal{M}, p'_\mathcal{M})\)-conditional distributions are intuitively correct.

By generating points: uniformly on all circles in \(\mathcal{O}\). by the \(\cos \theta\) density on all meridian circles in \(\mathcal{M}\).
Remark 2

Our view

Both the \((\mathcal{O}, p'_O)\)- and the \((\mathcal{M}, p'_M)\)-conditional distributions are intuitively correct.

By generating points: **uniformly** on all circles in \(\mathcal{O}\). **by the \(\cos \theta\) density** on all meridian circles in \(\mathcal{M}\).

One obtains a **uniform** distribution on the sphere.
Remark 2

Our view

- The (wrong) intuition that only the uniform conditional distribution is correct might come from the (wrong) intuition that the uniform length measure is singled out by pure probabilistic means.
Remark 2

Our view

- The (wrong) intuition – that only the uniform conditional distribution is correct – might come from the (wrong) intuition 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.
The (wrong) intuition – that only the uniform conditional distribution is correct – might come from the (wrong) intuition 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.
The proper mathematical device to handle conditional probabilities is the theory of conditional expectation.

This theory makes it possible to conditionalize on probability zero events, as in the case of the Borel-Kolmogorov Paradox.

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.

