Negative probabilities and counterfactual reasoning

J. Acacio de Barros
(with Gary Oas and Patrick Suppes)

Liberal Studies Program
San Francisco State University

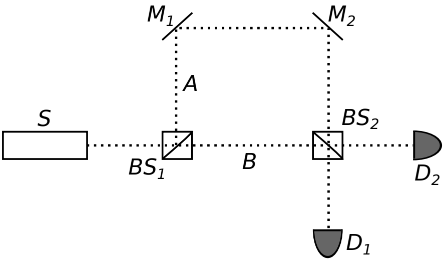
June 13, 2013

Outline

Outline

The mystery of QM (Feynman)

- The two-slit experiment has the essence of QM: it makes no sense.
- A simple version of the two-slit experiment is the Mach-Zehnder interferometer.



What does it mean to make no sense?

- It is non-monotonic.
- There is no joint probability distribution.
- We can't tell a story: consistent histories require proper measure.

What we cannot talk about we must pass over in silence

- We don't have consistent histories: so what can we talk about the two slit?
- Are there non-monotonic quasi-probabilities that we could use?
- If so, how do we interpret them?

Outline

Why Kolmogorov probabilities?

- Kolmogorov axiomatized probability in a set-theoretic way, with the following simple axioms.

$$\text{A1. } 1 \geq P(A) \geq 0$$

$$\text{A2. } P(\Omega) = 1$$

$$\text{A3. } P(A \cup B) = P(A) + P(B)$$

- Most ways to think rationally lead to probability measures a la Kolmogorov. For example,
  - Cox, Jaynes, Ramsey, de Finneti.
  - Venn, von Mises.

We can't always assign probabilities

There are systems that do not allow for a Kolmogorovian measure.

Example 1:

- $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$ are ± 1-valued random variables.

$$E(\mathbf{XY}) = E(\mathbf{XZ}) = E(\mathbf{YZ}) = -1.$$

- We can easily see why:

$$(\mathbf{X} = 1) \to (\mathbf{Y} = -1) \to (\mathbf{Z} = 1) \to (\mathbf{X} = -1).$$

Probabilities are monotonic

- Monotonicity is one a consequence of Kolmogorov's axioms.
- Say $C \subseteq D$. Then define $C' = D \setminus C$. Then C and C' are disjoint, and $C \cup C' = D$.
- But for disjoint sets

$$P\left(C \cup C'\right) = P(C) + P(C') = P(D).$$

From axiom A1 (positivity), it follows that if $C \subseteq D$ then $P(C) \leq P(D)$.

Double slit is not monotonic

- Feynman's remark on the two-slit are about the non-monotonicity of probabilities: how can we have a brighter spot when we actually close one of the slits?
- In other words, how can we have

$$P(A) > P(A \cup B)?$$

People reason nonmonotonically

- Busemeyer showed that we use nonmonotonic reasoning, as we violate monotonicity.
- Examples are
    - Conjunction fallacy
    - Violation of Savage's sure-thing principle (disjunction effect).

Upper and lower probabilities

- How do we get nonmonotonicity?
- de Finetti: relax Kolmogorov's axiom A2:

$$P^*(A \cup B) \geq P^*(A) + P^*(B)$$

or

$$P_*(A \cup B) \leq P_*(A) + P_*(B).$$

- Subjective meaning: bounds of best measures for inconsistent beliefs (imprecise probabilities).

Upper and lower probabilities

- Consequence:

$$M^* = \sum_i P_i^* > 1,$$

$$M_* = \sum_i P_{*i} < 1.$$

- M^* and M_* should be as close to one as possible.
- Inequalities and nonmonotonicity make it hard to compute upper and lowers for practical problems.

Workaround?

- Instead of violating A2, let us relax A1.

$$\text{A1.}\;\; M^T = \sum_i |p_i| \;\text{ is minimum}$$

$$\text{A2.}\;\; p(A \cup B) = p(A) + p(B)$$

$$\text{A3.}\;\; p(\Omega) = 1.$$

- p_i can now be negative.
- p defines an optimal upper probability distribution by simply setting all negative probability atoms to zero.
- Atoms with negative probability are thought subjectively as impossible events.

Why negative probabilities?

- We can compute them easily (compared to uppers/lowers).
- May be helpful to think about certain contextual problems (Oas and Al-Safi talks on Monday).
- They have a meaning in terms of subjective probability.

Outline

What doesn't work?

- Let us start with a simple way to think about the Mach-Zehnder interferometer.
- For the Mach-Zehnder, we have four possible experimental conditions:
  - D_1, D_2 only.
  - D_1, D_2, D_A.
  - D_1, D_2, D_B.
  - D_1, D_2, D_A, and D_B.

D_1, D_2 only. We have that $p_{d_1\overline{d}_2} = 1$ and $p_{d_1 d_2} = p_{\overline{d}_1 d_2} = p_{\overline{d}_1\overline{d}_2} = 0$.

D_1, D_2, D_A. Detection on D_A implies no detection on D_1 and D_2; no detection on D_A implies D_1 and D_2 are equiprobable.

$p_{d_a d_1 d_2} = p_{d_a d_1 \overline{d}_2} = p_{d_a \overline{d}_1 \overline{d}_2} = p_{d_a \overline{d}_1 d_2} = 0$, $p_{\overline{d}_a d_1 d_2} = p_{\overline{d}_a \overline{d}_1 \overline{d}_2} = 0$, and $p_{\overline{d}_a d_1 \overline{d}_2} = p_{\overline{d}_a \overline{d}_1 d_2} = \frac{1}{2}$.

D_1, D_2, D_B. $p_{d_b d_1 d_2} = p_{d_b d_1 \overline{d}_2} = p_{d_b \overline{d}_1 \overline{d}_2} = p_{d_b \overline{d}_1 d_2} = 0$, $p_{\overline{d}_b d_1 d_2} = p_{\overline{d}_b \overline{d}_1 \overline{d}_2} = 0$, and $p_{\overline{d}_b d_1 \overline{d}_2} = p_{\overline{d}_b \overline{d}_1 d_2} = \frac{1}{2}$.

D_1, D_2, D_A, and D_B.

$p_{d_a d_b d_1 d_2} = p_{d_a d_b \overline{d}_1 d_2} = p_{d_a d_b d_1 \overline{d}_2} = p_{d_a d_b \overline{d}_1 \overline{d}_2} = 0$, $p_{\overline{d}_a d_b d_1 d_2} = p_{\overline{d}_a d_b \overline{d}_1 d_2} = p_{\overline{d}_a d_b d_1 \overline{d}_2} = 0$, $p_{d_a \overline{d}_b d_1 d_2} = p_{d_a \overline{d}_b \overline{d}_1 d_2} = p_{d_a \overline{d}_b d_1 \overline{d}_2} = 0$, $p_{\overline{d}_a \overline{d}_b d_1 d_2} = p_{\overline{d}_a \overline{d}_b \overline{d}_1 d_2} = p_{\overline{d}_a \overline{d}_b d_1 \overline{d}_2} = p_{\overline{d}_a \overline{d}_b \overline{d}_1 \overline{d}_2} = 0$, and $p_{\overline{d}_a d_b \overline{d}_1 \overline{d}_2} = p_{d_a \overline{d}_b \overline{d}_1 \overline{d}_2} = \frac{1}{2}$.

What works?

- No probability can be defined (not even negative probabilities) for the case above.
- For the which path version, we observe

$$P(ab) = 0, \quad (1)$$

$$P(\bar{a}b) = \frac{1}{2}, \quad (2)$$

$$P(a\bar{b}) = \frac{1}{2}, \quad (3)$$

$$P(\bar{a}\bar{b}) = 0. \quad (4)$$

- Interference (maximum visibility) requires that we only observe in D_1:

$$P(d_1\overline{d}_2) = 1 \tag{5}$$

$$P(\overline{d}_1 d_2) = P(d_1 d_2) = P(\overline{d}_1\overline{d}_2) = 0. \tag{6}$$

- If we put a detector in $P_i \in \{A, B\}$, we "infer" that whenever we observe the particle *not* being in A, then the particle must be (probability 1) in B.
- But when we block the path, the probabilities are

$$P(\overline{a}d_1\overline{d}_2) = P(\overline{a}\overline{d}_1 d_2) = \frac{1}{2}, \tag{7}$$

$$P(\overline{b}d_1\overline{d}_2) = P(\overline{p}\overline{d}_1 d_2) = \frac{1}{2}. \tag{8}$$

- This is different from the previous example, where we talked about detections in A, and not lack of detections.

Negative joint: general solution

$$
\begin{aligned}
p_{abdd} &= -\frac{3}{4} + b, & p_{\bar{a}bdd} &= \frac{1}{2} - a + c - d,\\
p_{abd\bar{d}} &= \frac{3}{4} + c, & p_{\overline{ab}dd} &= b,\\
p_{a\overline{bd}d} &= a, & p_{a\bar{b}d\bar{d}} &= -c,\\
p_{\bar{a}b\overline{dd}} &= d, & p_{\overline{ab}d\bar{d}} &= c,\\
p_{a\bar{b}dd} &= \frac{1}{4} + a - 2b - c + d, & p_{ab\bar{d}d} &= \frac{1}{4} - a,\\
p_{\bar{a}b\bar{d}d} &= a, & p_{\bar{a}bd\bar{d}} &= -c,\\
p_{ab\overline{dd}} &= -\frac{1}{4} + a - b - c, & p_{a\overline{bdd}} &= \frac{1}{4} - 2a + 2b + 2c - d,\\
p_{\overline{abd}d} &= -a, & p_{\overline{ppdd}} &= a - b - c.
\end{aligned}
$$

Negative joint: a M^T minimizing solution

$$
\begin{array}{ll}
p_{abdd} = -\frac{3}{4}, & p_{\bar{a}bdd} = \frac{1}{2}, \\
p_{abd\bar{d}} = \frac{3}{4}, & p_{\overline{ab}dd} = 0, \\
p_{a\overline{bd}d} = 0, & p_{a\bar{b}d\bar{d}} = 0, \\
p_{\bar{a}b\overline{dd}} = 0, & p_{\overline{ab}d\bar{d}} = 0, \\
p_{a\bar{b}dd} = \frac{1}{4}, & p_{ab\bar{d}d} = \frac{1}{4}, \\
p_{\bar{a}b\bar{d}d} = 0, & p_{\bar{a}bd\bar{d}} = 0, \\
p_{ab\overline{dd}} = -\frac{1}{4}, & p_{a\overline{bdd}} = \frac{1}{4}, \\
p_{\overline{abd}d} = 0, & p_{\overline{ppdd}} = 0.
\end{array}
$$

Some conditionals

Since we have a "joint", we can compute conditional probabilities. E.g.

$$p(a|d_1\overline{d_2}) = \frac{1}{3},$$

$$p(b|d_1\overline{d_2}) = \frac{2}{3}.$$

To summarize:

- Random variables with inconsistent correlations can be thought in terms of negative probabilities. Such probabilities can be given a subjective interpretation in terms of upper bounds for imprecise subjective probabilities.
- In the two slit experiment, by reducing the amount of counterfactuals, it is possible to obtain a negative joint probability for the atoms.
- Such negative joint probability can then be used to compute unmeasured correlations of conditional probabilities, consistent with the marginals.

Thank you!