Phil. 4400
Notes #13: The Doomsday Argument

I. Doomsday Possibilities

- Perhaps mankind will soon go extinct. How could this happen?
  1. Nuclear war.
  2. Environmental destruction.
  3. Asteroid impact.
  4. Disease, genetically engineered or otherwise.
  5. Misuse of nanotechnology.
  6. We're living in a simulation and it gets shut down.
  7. Badly programmed superintelligence.
  8. Physics experiments.
  9. Something unforeseen.
- Consider two hypotheses:
  (S) *Doom Soon:* Mankind will soon go extinct. This could happen from any of the above causes.
  (L) *Doom Late:* Mankind will survive far into the future. Maybe humans will figure out how to avoid all those threats.
- Other possibilities: There are infinitely many hypotheses as to when the species will go extinct. We consider just these two for the sake of simplicity.
- Mankind will go extinct. The question is when.

II. The Doomsday Argument

A. The Self-Sampling Assumption

- You should treat yourself as a random selection from all observers.
- Example: 100 people, including you, are drugged and kidnaped. Each is taken to a hotel room. 90 of the rooms are painted blue on the outside, 10 are red. Q: Given this information, what is the probability that your room is blue?

B. Another Example

- You are drugged, kidnaped, and taken to a hotel room. You were either taken to a small hotel (with 10 rooms), or a large hotel (with 100 rooms), depending on the toss of a fair coin. Each hotel has its rooms labeled with integers, starting from 1 (room #1, room #2, etc.).
  Q1: Given just this information, what is the probability that you are in the small hotel?
  Q2: Suppose you learn that you are in room #7. *Now* what is the probability that you are in the small hotel?
- Bayes' Theorem: $P(h \mid e) = \dfrac{P(h) \cdot P(e \mid h)}{P(h) \cdot P(e \mid h) + P(\sim h) \cdot P(e \mid \sim h)}$
- In this example, let:
  S = You are in the Small hotel.
  L = You are in the Large hotel.
  E = Your room number is 7.

- Applying Bayes' Theorem:

$$P(S \mid E) = \frac{P(S) \cdot P(E \mid S)}{P(S) \cdot P(E \mid S) + P(\sim S) \cdot P(E \mid \sim S)}$$

P(S) = .5 P(~S) = .5

P(E|S) = 1/10 P(E|~S) = P(E|L) = 1/100

Hence:

$$P(S \mid E) = \frac{(.5)(.1)}{(.5)(.1) + (.5)(.01)} = \frac{.05}{.05 + .005} \approx .91$$

C. The Doomsday Analogy

- You find yourself living as a member of the human species. It is either a short-lived species (per "Doom Soon") or a long-lived species (per "Doom Late").
  - Q1: What is the initial probability of Doom Soon?
  - Q2: You learn that you are approximately the 60 billionth human born. Now what is the probability of Doom Soon?
- Let
  - S = Doom Soon. Suppose this means the total # of humans ever to have lived will be ~100 billion.
  - L = Doom Late. Suppose this means the total # of humans ever to have lived will be ~1 trillion.
  - E = Your birth rank is ~60 billion.
- What is P(S)? Hard to calculate. Seems like a pretty good chance, somewhere between 10% and 90%. Let's leave it as an unknown, *s*.
- What is P(S|E)? By Bayes' Theorem,

$$P(S \mid E) = \frac{P(S) \cdot P(E \mid S)}{P(S) \cdot P(E \mid S) + P(\sim S) \cdot P(E \mid \sim S)}$$

P(S) = *s* P(~S) = 1 - *s*

$$P(E \mid S) = \frac{1}{100 \text{ billion}} \qquad P(E \mid \sim S) = P(E \mid L) = \frac{1}{1 \text{ trillion}}$$

Hence:

$$P(S \mid E) = \frac{(s)\left(\frac{1}{100 \text{ billion}}\right)}{(s)\left(\frac{1}{100 \text{ billion}}\right) + (1-s)\left(\frac{1}{1 \text{ trillion}}\right)}$$

Multiplying numerator and denominator by 1 trillion gives us:

$$P(S \mid E) = \frac{10s}{10s + (1-s)} = \frac{10s}{9s+1}$$

- Try plugging in some hypothetical values for *s*:

When $s = 0$: $P(S \mid E) = \frac{10(0)}{9(0)+1} = 0$

When $s = 1$: $P(S \mid E) = \frac{10(1)}{9(1)+1} = 1$

When $s = .1$: $P(S \mid E) = \frac{10(.1)}{9(.1)+1} \approx .53$

When $s = .5$: $P(S \mid E) = \frac{10(.5)}{9(.5)+1} \approx .91$

- Conclusion: E supports S for all values of s between 0 and 1.
  - The support is stronger, the larger we make the population in Doom Late. If Doom Late has 100 trillion people, then we get:

    $$P(S \mid E) = \frac{1000s}{999s+1}$$

    When $s = .5$: $P(S \mid E) = \frac{1000(.5)}{999(.5)+1} \approx .999$

III. Criticisms of the Doomsday Argument

A. A disanalogy between IIB and IIC:

- In the analogy in IIB:
  - I already exist before the hotel is chosen. Which of {S,L} is the case only affects *where* I am. If the coin flip came up the other way, I would just be somewhere else.
  - If the Large hotel had been chosen, I would probably now have a higher room #.
- In IIC,
  - Whichever of {S,L} is the case, I would *not* be somewhere else if the other one were the case.
  - It is *false* that if L were the case, I would probably now have had a higher birth order: Everything up to now would be exactly as it is.

B. Whether S or L were the case,* someone *would have birth order 60 billion (or whatever my birth order is).

- Therefore, the fact that someone has birth order 60 billion does not support S over L or vice versa.
  - If $P(E|H_1) = P(E|H_2)$, then E supports H_1 if and only if E supports H_2.
  - In this case, P(Someone has birth order 60 billion | S) = 1 = P(Someone has birth order 60 billion | L).
- Surely the fact that that someone is *me*, rather than someone else, is not relevant either.
  - S and L do not make any different predictions about how likely it is that the person in question would "be me".

C. My existence is evidence for L

- Intuitive idea:
  - *Given that I exist,* S makes it ten times more likely that my birth order would be < 100 billion than L does.
  - So "My birth order is < 100 billion" is evidence for S over L.
  - But L makes it ten times more likely that I would exist at all than S does, because in L there would be ten times as many people.

- So my existence is evidence for L.
- These two pieces of evidence exactly counterbalance one another.
- Thus, "I exist and have birth order < 100 billion" is overall neutral between S and L.

• The probability calculations:
Let B = That I exist at some time in the history of the species.

$$P(S \mid E \,\&\, B) = \frac{P(S) \cdot P(E \,\&\, B \mid S)}{P(S) \cdot P(E \,\&\, B \mid S) + P(\sim S) \cdot P(E \,\&\, B \mid \sim S)}$$

P(S) = *s*
P(~S) = 1 - *s*
P(B|~S) = 10·P(B|S)

$$P(E \,\&\, B \mid S) = P(B \mid S) \cdot P(E \mid B \,\&\, S) = P(B \mid S) \cdot \frac{1}{100 \text{ billion}}$$

$$P(E \,\&\, B \mid \sim S) = P(B \mid \sim S) \cdot P(E \mid B \,\&\, \sim S) = 10 \cdot P(B \mid S) \cdot \frac{1}{1 \text{ trillion}} = P(B \mid S) \cdot \frac{1}{100 \text{ billion}}$$

Thus, plugging those things into the first equation:

$$P(S \mid E \,\&\, B) = \frac{s \cdot P(B \mid S) \cdot \frac{1}{100 \text{ billion}}}{s \cdot P(B \mid S) \cdot \frac{1}{100 \text{ billion}} + (1-s) \cdot P(B \mid S) \cdot \frac{1}{100 \text{ billion}}} = \frac{s}{s + (1-s)} = s$$