The Two Envelope Problem

Gary Glonek

School of Mathematical Sciences, University of Adelaide

11 August 2010

The game

- As contestant on a game show, you are presented with two identical envelopes each containing a cheque:
    - You have no information about what the amounts might be;
    - But you are told that the amount on one of the cheques is twice as much as that on the other.
- You then choose one of the envelopes that you may open to reveal the amount on the cheque.
- Finally you are offered the choice of keeping the cheque or swapping it for the unseen amount in the second envelope.

What is the best strategy?

The paradox

Suppose the envelope is allocated to the contestant at random. For example, by a coin toss.

One argument is as follows:

- Prior to opening the envelope, I can say the chance that it contains the larger amount is 50%.
- Suppose now that I open the envelope and see an amount, say $100.
    - I can now be certain that the amount in the other envelope is either $50 or $200.
- *If I really had no information about how the amounts in the envelopes were generated, then seeing $100 gives me no new information about whether I have received the larger or smaller amount.*

- If that argument is accepted then the conclusion is that there is a 50% chance that the unopened envelope contains $50 and a 50% chance that it contains $200.
- Therefore, if I swap, there is 50% chance that I will win an extra $100 and a 50% chance that I will lose $50.

Therefore, it is in my interest to swap.

BUT...

- The fact that I saw $100 in the envelope was irrelevant.
- I would reach the same conclusion irrespective of the amount I saw.
- Therefore I am claiming that it is a good bet to swap, even without opening the envelope.
- But if the allocation of the envelopes was by coin toss, there can be no systematic advantage to swapping.

Conditional probability

The standard framework for updating probabilistic beliefs in the light of partial information is through the calculation of conditional probabilities.

For events A and B with $P(B) > 0$ the conditional probability of A given B is defined by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

For example, suppose a fair die is rolled. Let $A = \{2, 4, 6\}$ be the event that an even face turns up and let $B = \{1, 2, 3\}$ be the event that the outcome is no greater than 3.

It is easy to check that $P(A) = 1/2$ and $P(A|B) = 1/3$.

This can be interpreted as a rule for updating probabilities.

Bayes' theorem

Conditional probabilities are sometimes calculated via Bayes' theorem.

In its simplest form:

$$P(B|A) = \frac{P(B)P(A|B)}{P(B)(A|B) + P(B^c)P(A|B^c)}$$

For example, suppose we have two coins: a fair coin and a two-headed coin. A coin is chosen at random and tossed once. Given the outcome is a head, what is the probability that the two-headed coin was chosen?

Taking B to be the event that the two-headed coin was chosen and A to be the event that a head was tossed, we have $P(B) = P(B^c) = 1/2$, $P(A|B) = 1$ and $P(A|B^c) = 1/2$ whereby Bayes' theorem gives

$$P(B|A) = 2/3.$$

Bayes' theorem and the two envelopes

Back to the two-envelope problem...

- Let θ denote the larger of the two amounts so that one envelope contains θ and the other contains $\theta/2$.
- Let X denote the amount in the envelope opened.
- Let A be the event that the envelope opened contains the larger amount.

Using Bayes' theorem, it follows that

$$
\begin{aligned}
P(A|X=x) &= \frac{P(A)P(X=x|A)}{P(A)P(X=x|A)+P(A^c)P(X=x|A^c)} \\
&= \frac{0.5P(\theta=x)}{0.5P(\theta=x)+0.5P(\theta=2x)} \\
&= \frac{P(\theta=x)}{P(\theta=x)+P(\theta=2x)}.
\end{aligned}
$$

The prior for θ

To evaluate $P(A|X = x)$ using Bayes' theorem, we need to be able to calculate $P(\theta = x)$ and $P(\theta = 2x)$.

This introduces *prior information* about the content of the envelopes to the problem.

In this problem we need to deal with a continuous prior distribution for θ specified by its probability density function, $f_\theta()$.

It can be shown in this case that Bayes' theorem yields

$$P(A|X = x) = \frac{f_\theta(x)}{f_\theta(x) + 2f_\theta(2x)}.$$

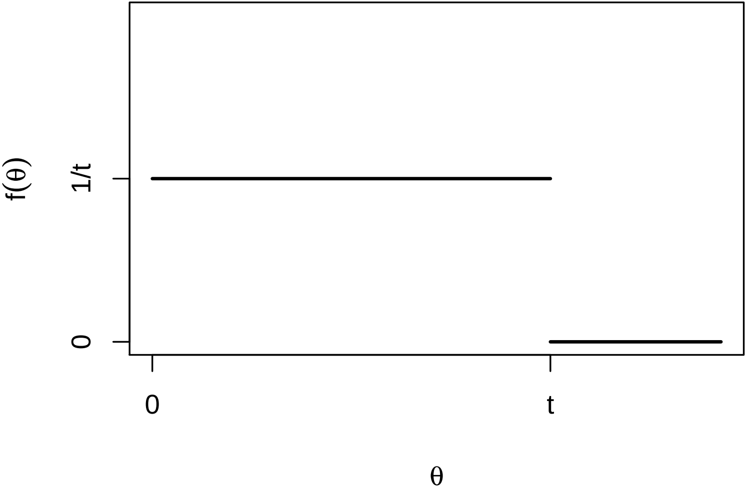
Example: the uniform prior

Suppose we have prior information that θ is uniformly distributed on the interval $(0, t)$.

That is

$$f(\theta) = \begin{cases} \frac{1}{t} & 0 < \theta < t \\ 0 & \text{otherwise.} \end{cases}$$

The density for the uniform prior



Suppose now that we observe an amount $\$x$ inside our envelope.

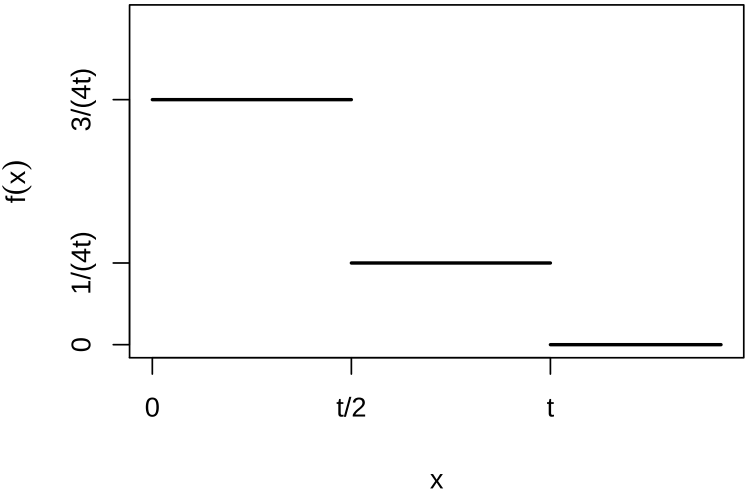
- As before we can be 100% certain that the other envelope contains either $\$x/2$ or $\$2x$.
- However, based on the uniform prior assumption for θ the probability that we have drawn the larger amount is

$$P(A|X=x) = \begin{cases} \frac{1}{3} & 0 < x < t/2 \\ 1 & t/2 \leq x < t. \end{cases}$$

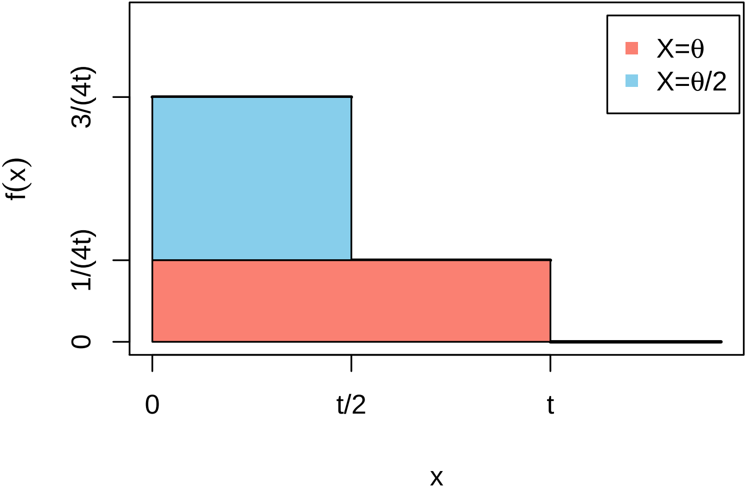
- That is, we should swap if our observed value is less than $t/2$ but definitely not otherwise.

This is very different from the original “solution”!

The density for X implied by the uniform prior



The density for X insightfully coloured



Conclusion

A simple Bayesian calculation leads to a coherent result with no paradoxes, *if we can assume a prior distribution for* θ.

But, that was not the original problem.

We might now ask:

1. Is there a prior distribution that captures the concept of "no information about θ"?
2. Is there a prior distribution that would lead to conclusion that there is a 50% that $X = \theta$ irrespective of the value of X that I observed?
3. Can we argue that the paradox does not really exist because, in real life, there must always be *some* prior information about θ?

1. No

Surprisingly this is not as easy as it might first appear.

- All prior distributions carry *some* information.
- There are technical obstacles that prevent us from even defining a proper prior distribution that would make all positive real numbers "equally likely".
- Even if we consider the less ambitious problem of specifying a prior that carries no information except that $0 < \theta < 1$, it turns out there are ambiguities that cannot be resolved.

2. No

A probability density for θ has to satisfy

$$f(\theta) \geq 0 \text{ and } \int_0^\infty f(\theta)d\theta = 1.$$

Now consider the requirement that

$$P(A|X = x) = \frac{f_\theta(x)}{f_\theta(x) + 2f_\theta(2x)} = 0.5$$

for all possible x.

Clearly this requires $2f(2\theta) = f(\theta)$ for all θ.

It is easy to check that there is no such function can be a probability density.

Hint: Obtain an expression for

$$\int_1^{2^n} f(\theta)d\theta \text{ in terms of } \int_1^2 f(\theta)d\theta.$$

3. Not really

It is sometimes argued that there are real world constraints that we need to recognise.

For example, there is an upper limit on the total amount of money in the world and this represents prior information that needs to be accounted for.

Similarly there is also minimum denomination of currency.

These points can be avoided if we assume the units of currency to be unspecified and arbitrary.

For example, X could be in units of \$$10^{-250}$.

Conclusion

The two-envelope paradox cannot occur within a Bayesian framework.

That is, starting with a well defined prior distribution for θ.

One position might then be that the 50/50 calculation in the original formulation is simply nonsense.

However, this is unsatisfying in that the fallacy in the argument is not revealed.

So why should anyone care?

The problematic 50/50 conclusion can be reached in a Bayesian framework if we allow for an *improper* prior distribution for θ.

That is, $f(\theta) \geq 0$ with

$$\int_0^\infty f(\theta) d\theta = +\infty.$$

Improper priors are advocated in many treatments of Baysesian inference so there is a question to be asked about whether there might be similar consequences in other contexts.

The frequency interpretation of probability is a relatively well understood and accepted framework.

However, subjective interpretations are also used in various contexts to represent “strengths of belief”.

If we interpret the 50/50 conclusion in the two envelope problem as a subjective probability then:

- This would appear to be as well founded as many other claims of subjective probability;
- It shows that probabilities obtained outside of a strict frequency framework sometimes need to be treated very carefully.

The two-envelope problem, as posed here, can also be seen as an attempt to induce a posterior distribution for θ based on an initial position of "complete ignorance" rather than a prior distribution.

This type of inference has been considered many times including by statisticians as distinguished as RA Fisher (fiducial inference). It turns out that even Fisher wasn't able to come to a simple conclusion on this matter.

It is fun to mess about with this problem, provided you define fun appropriately.

Student challenge

You are presented with an envelope containing a cheque for an amount \$$x$ and you know the amount in the other envelope is either \$$2x$ or \$$x/2$.

Are the following equivalent?

1. A number θ is obtained from somewhere and the amounts θ and $\theta/2$ are placed in two envelopes. You are presented with one of the envelopes chosen at random.
2. A number x is obtained from somewhere and placed in an envelope. A coin is tossed and the value $2x$ is placed in the second envelope if for a Head and $x/2$ for a Tail. You are given the envelope containing x.

Solutions to `gary.glonek@adelaide.edu.au` by Friday 20/8.

Best answer by an undergraduate wins a chance to play the two envelope game (once) at my expense.