Is there a two-envelope paradox?

Federico O'Reilly
IIMAS, UNAM, México, updated January, 2010

Abstract

We address the two-envelope paradox, studied over a number of years. A statistical analysis is provided based on a classical inference point of view as well as from a Bayesian perspective. In the classical analysis it is stressed that there is no paradox but rather an erroneous use of an expected value, something that has been said before by some authors. A different, related problem is discussed briefly where the issues of "utility" and "personal preferences" arise naturaly. From a Bayesian point of view, the analysis centers on the degeneracy caused by the implied (though correct), improper prior which arises from the explicit argument used when stating the paradoxical result. The implied prior coincides with that in many articles that claim that underneath the paradox is the mistaken interpretation of infinite quantities. The classical analysis of this note is based on some ideas presented in O'Reilly (2006) and the Bayesian analysis on Quintana and O'Reilly (2008). The aim of this working document is to have both analyses in the same manuscript.

Key Words: Expected value, likelihood, utility, improper priors. repeatable essays.

1. INTRODUCTION

Christensen and Utts (1992) addressed the "Exchange Paradox" and many letters and articles have been written on this problem since. A description and a brief account of literature, besides references in the article of Christensen and Utts appear in the book by Casella and Berger (2002). Other articles have appeared after, mostly in Philosophy journals, where according to us, the basic inference problem, is still mostly overlooked. Also according to us, in the Bayesian analysis one has to identify a prior implied by the "paradoxical" reasoning in the original setting. Other approaches suggesting different priors, deal with a variant of the original two envelope setting which we do not address.

The problem of the two-envelope paradox, refered also as the exchange paradox or the problem of the two wallets, is as follows:

A Benefactor places a certain ammount of money in one envelope and places twice that quantity in another envelope. The initial ammount and the identification of the envelope where the larger quantity was placed, are unknown to you. Then, the Benefactor selects randomly (equal probability) one envelope and hands it to you and the contents is yours. The mental reasoning to get the "paradoxical result" is as follows:

If X is inside your envelope, then even before opening it, you realize that the contents of the other envelope, call it Y, will be either $X/2$ or $2X$, which due to the random selection yields an "expected value" (denoted E_X at this moment) for the contents of the other envelope:

$$E_X(Y) = 1/2(X/2) + 1/2(2X) = (5/4)X.$$

So you reason that you must switch envelopes if given the opportunity to do so, because the expected value if trading your envelope for the other one, which turned to be $(5/4)X$ *is larger than X. But the same reasoning applies have you been handed the other envelope, so something is definetely wrong because of the symmetry in the problem.* This is the famous two-envelope paradox.

Denote by θ, the smaller of the two quantities assigned to the envelopes, to stress the role of that ammount of money as a parameter. We will see that from a classical point of view, the computation of the expected value is incorrect. An explicit referal to this fact has been done before in Bruss (1996), terming it "a fallacy of notation". And we will see from a Bayesian perspective that the implied prior makes the possible values of θ essentially either 0 or ∞.

Due to the huge ammount of articles and sites (internet), there is no claim at all of having made a thorough review of literature on the subject. Looking at the web there are thousands (if not tens of thousands) of sites related to this problem. We looked at some and found as an example, that the manuscript by E. Schwitzgebel and J. Dever is closely connected to the key issue in Bruss (1996) which is also very important to our classical analysis. In our opinion, the problem specifies that one faces the dilemma after an unknown quantity θ has been fixed so we take that, as said colloquially, at "face-value".

2. BASIC CLASSICAL ANALYSIS

If you have X in your closed envelope, the possibility that the other envelope has X/2 inside, since θ is fixed, means that your $X = 2\theta$, because the Benefactor has already decided the amount 3θ to be split in proportions 1:2 in the envelopes. On the other hand if you analyze the case where the other envelope has $2X$, this means without any ambiguity that you have $X = \theta$ in your envelope. Observe that the value of X is not the same in these two distinct cases.

The expected value for the contents of the envelope you were given, depends on the fixed amount decided by the Benefactor, θ and is:

$$E(X;\theta) = (1/2)\theta + (1/2)2\theta = (3/2)\theta.$$

(The notation $E(X|\theta)$ will be also used, but when dealing with the Bayesian analysis; that is, the ";" will later be replaced by the "|").
If one analyzes the "pseudo-expected value" $E_X(Y) = (5/4)X$, this is meaningless. It has been incorrectly computed since it added the product of 1/2 times $X/2$, the first summand (which was θ in the first possibility, when Y was allowed to be $X/2$), with the product of 1/2 times $2X$, the second summand (which must be 2θ in this second possibility when Y would be $2X$), so the correct expected value, substituting the quantity $X/2$ for its true value θ and the value $2X$ for its true value 2θ is, $E(Y;\theta) = (3/2)\theta$.

This last expected value coincides with the expected value of the contents of the envelope handed to you originally, described in terms of θ. The incorrect use of the concept of expected value (conditioning on something which is not constant) for the contents in the second envelope, in terms of X, is responsible for the erroneous finding. The quantity θ is fixed and X is a quantity that varies according to the choice of the envelope; which in turn defines the two possibilities for Y (recall that the sum of the contents in both envelopes is constant, that is $X + Y = 3\theta$). The mistake is "adding apples and oranges" as mentioned by Scwitzgebel and Dever in their manuscript. So that part of the reasoning that says "even before opening it, you realize that...", in our opinion, is satisfactorily shown to be wrong. You ask yourself if you would exchange your envelope, after θ was fixed.

3. CLASSICAL INFERENTIAL ANALYSIS

Now, from the inferential point of view, if one opens the envelope and finds out that $X = x$, it is true that logically, there are two possibilities for the other envelope in terms of x, now observed and considered fixed . These possibilities are that either its contents is $Y = x/2$, which occurred if $\theta = x/2$, or its contents is $Y = 2x$ which happened if $\theta = x$. Of course implying also, that in the first case, the random selection of the envelope lead to the larger quantity in your envelope and in the second case, to the envelope with the smaller quantity. It must be observed however, that in the experiment, only one value of θ is present. In the analysis one must consider not only one expected value for Y in terms of x; there are two cases to be considered after $X = x$ was observed according to the two distinct logical possibilities for the value that θ could have had. In Christensen and Utts (1992) this is said clearly in their Section 3.

With $X = x$ observed, if θ equals $x/2$ then there is certainty that $Y = x/2$, and if θ is x one is sure that $Y = 2x$. These are the two expected values. It is here that

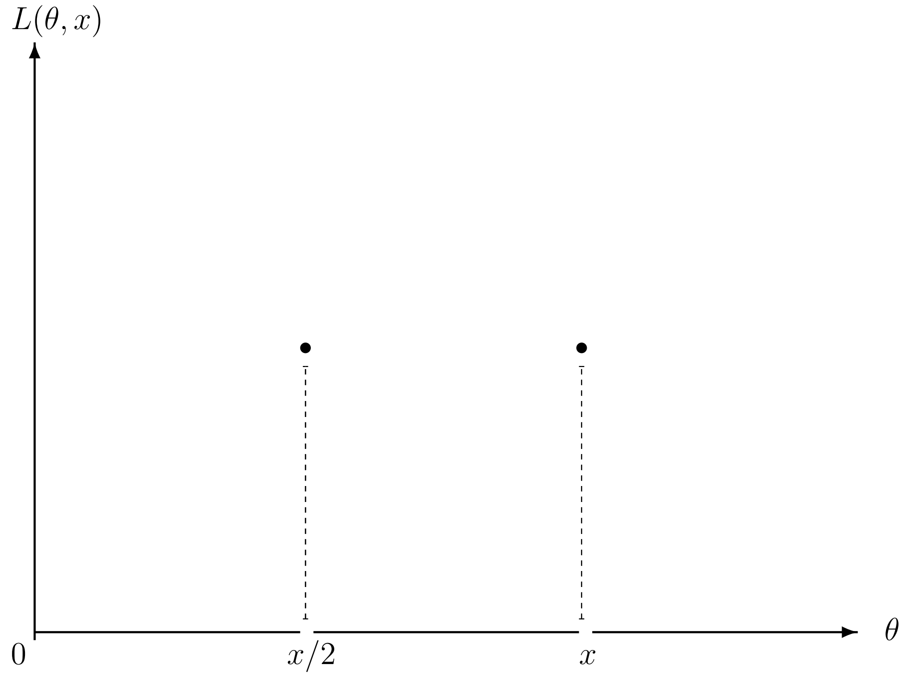

Figure 1: Likelihood Function

reasoning beyond this point, from a classical perspective, seems to produce confusion because from the setting of the problem it is just stated that "A Benefactor places a certain amount of money ...", and no one says that the quantity, θ is an outcome of a previous experiment (which well may be, but has no relevance) nor that there will be a future experiment to decide its value, which definetely would be a different problem altogether.

The value x observed, implies that θ, originally an unknown positive parameter, is now, known to belong to the set $\{x/2, x\}$, but that is all we know about it. Incidentally, the likelihood function for θ is equal at both points of the set, and zero outside. So one reasons, from the inferential point of view, that x is a realization of X which has exactly the same distribution as Y. Also, knowledge of $X = x$ even though "intriguing as to sharpen our appraisal" for Y, is absolutely useless as to which value of Y amongst the two, might be more likely. We believe hard to grasp from the classical point of view, any attempt to exhibit an expected value for Y given $X = x$, ignoring θ, in fact "averaging" the possibilities for Y as if one could sample θ out of $\{x/2, x\}$. That would be another setting, as if one asks the Benefactor to remove the contents of the unselected envelope, and then place inside it, either twice or half the amount x found in the envelope selected originally, which is a a different problem. In Ridgeway (1993) this related problem is mentioned.

So the classical reasoning is, that since no information on Y is obtained after observing that $X = x$, other than the fact that now θ is pinned down to two equally likely values, and that in the experiment performed, both X and Y materialize with θ fixed, and further more, that they do have exactly the same distribution; the quantity Y needs not be preferable over X, observed or not. Knowing that $X = x$ leaves us as ignorant as before opening the envelope in terms of the odds (1:1) of having being handed, the envelope with the larger quantity. So from a classical standpoint, there is no reason to assert that having observed $X = x$, you must switch envelopes.

4. DIFFERENT PROBLEM

At this point we believe it is worth to analize some related concepts that appear in some of the discussions of the two-envelope problem but stating very clearly that these concepts, according to us, apply to the different related problem for which there is no paradoxical result.

In this related problem (see Ridgeway, 1993), an explicit referal to the "utility of money" is made; as mentioned previously in articles like that of Linzer (1994).

If one looks at the different surrogate problem mentioned in Section 3, namely that the Benefactor places an amount θ (now define $\theta = x$) in an envelope and hands it to you, and then with probability ï¿$\frac{1}{2}$ each, he either places twice that amount or half that amount in a second envelope, then the expected value for the contents in the second envelope is indeed $(5/4)x$, a legitimate expected value in this different problem. So the question arises of whether you should switch envelopes or if you are indifferent.

In this setting there is an original parameter space, which is the positive real axis for θ. After observing $X = x$, the parameter space shrinks to a set of only one possible value so there is no uncertainty regarding the parameter value after observing x. In this case,

$$E(Y|x) = E(Y|x, \theta = x) = (5/4)x > x.$$

And there is no inference problem.

We may decide, for example, to switch in this different problem, if allowed to play it a large number of times and keep what we get overall, divided by the number of times played; that is, if we may keep what we get on "the average". The reasoning is because probabilistic results imply that one will get on the average, a quantity very close to the expected value. This type of reasoning keeps the Casinos making money, but relays on the repeatabilty of the experiment.

We realize however, that if faced once in our lifetime with the dilemma of staying with the known x in our pocket or gambling to either loose half of x or win another x, our decision, will certainly depend on x, our income, and perhaps our mood. This is the favorite subjective part of this type of problems. But recall this is not the original two-envelope problem.

There is a notion used in Economics and in Decision Theory related to the utility of money, which allows first, for the personal appraisal related to an economic gain; that is, the utility of money is perceived differently by different individuals so it is subjective. Also one should allow for the possible non-linearity of the "utility", as a function: more precisely concavity is suggested, meaning that as the amount of our wealth increases, the "utility" (the increase of our feeling of extra gained wealth or comfort) increases also, but much more slowly if our wealth is big; and the bigger it gets, the slower the utility grows.

In the very interesting book by Bernstein (1996), reference is made on a direct mention of utility done long ago by D. Bernoulli, (1738), where we quote from the cited reference, *utility resulting from any small increase in wealth will be inversely proportionate to the quantity of goods previously possessed*, meaning that the increase (relative; which is the derivative of the utility function), should be proportional to the inverse of the amount of wealth already possessed; that is, the utility suggested is a logarithmic function.

Observe, just as a curiosity, that in the different problem analysed in this section, the expected value of the logarithm of the contents in the second envelope is the same as the logarithm of x. So if we accept the use of a logarithmic function as our measure of utility, in this different problem we would be indifferent to the exchange of envelopes.

$$E(\log(Y)|x) = E(\log(Y)|x, \theta = x) = \log(x).$$

If we do not accept the logarithmic utility and use, say, a linear utility function, then we must decide to exchange the envelope. But in this problem there is no paradoxical reasoning, since the roles of the two envelopes are **not** symmetrical. If you were a second individual and were to receive the second envelope instead, you certainly would have an advantage and would not want to exchange, under a linear utility.

5. BAYESIAN INFERENTIAL ANALYSIS

From a Bayesian perspective, unknown quantities, as parameters, are given a prior or initial distribution which models the initial knowledge that the decision-maker has, regarding these unknowns.

In the original setting of the two-envelope paradox; and more specifically in the description of the "expected value" for the contents of the other envelope in terms of the contents of the envelope handed to you, there appears to be agreement in most Bayesian papers dealing with this problem, in that what is being implied is the conditional expectation of the contents of the other envelope given the contents of your envelope.

In order to use part of the notation already presented, with the obvious changes for the Bayesian description, having observed x in your envelope, the conditional expected value for Y given x may be expressed as:

$$E(Y|x) = E(Y|x, \theta = x/2)p(\theta = x/2|x) + E(Y|x, \theta = x)p(\theta = x|x),$$

with, $p(\cdot|x)$ standing for the posterior (final) distribution for θ.

There is also agreement in that, $E(Y|x, \theta = x/2) = x/2$ since under such conditions the distribution of Y is degenerate. Also, $E(Y|x, \theta = x) = 2x$, as was mentioned already in Section 3. Therefore, the "expected value" for the contents of the other envelope, given x, mentioned in the setting of the problem, implies, without ever making it explicit, that the posterior distribution for θ is uniform on the set $\{x/2, x\}$.

One way in which such posterior arises in this setting is when the prior for θ, say, $p(\theta) \propto 1/\theta$. In the derivation of the posterior, many authors have not implied a correct prior, applying naively Bayes formula, and ending for example, with a uniform for θ instead (see observations made in Blachman *et.al*, 1996 and also in Broome, 1995).

An incorrect imputation of prior appears for example in Zabell, (1988) page 234, where it is asserted, in our notation, that the posterior probabilities in $\{x/2, x\}$ should be proportional to $p(x/2)$ and $p(x)$, instead of proportional to $(1/2)p(x/2)$ and $p(x)$, which is the correct assertion.

Up to here, the paradox seems to remain, since indeed, $E(Y|x) = (5/4)x$ despite the symmetry in the setting of the problem. It is at this point that many authors mention that the problem lies in dealing with infinite expected values; for example when dealing with the improper prior, $p(\theta) \propto 1/\theta$.

It is fairly direct to recognize that because Y and X have exactly the same conditional expectation given θ (in fact, $(3/4)\theta)$), then it follows taking expectations in both sides, that $E(Y) = E(X)$. On the other hand, as already shown, $E(Y|X) = (3/4)X$, so again by taking expectations in both sides, it follows that $E(Y) = (3/4)E(X)$. This two equalities are thus true if and only if, both expected values are either 0 or ∞ ($-\infty$ is ruled out).

It does not take long, to show that when the positive real line $\Theta = (0, \infty)$, the original parameter space, is assigned the "improper measure" $p(\theta) \propto 1/\theta$, then for practical purposes the measure is degenerate. Suppose we aproach Θ with a sequence of compact parameter spaces $\Theta_n = [a_n, b_n]$ (with $0 < a_n < b_n < \infty$), defining for each n, the measure p_n, as the restriction of p to Θ_n. Then no matter how fast (slow) a_n approaches 0 relative to how fast b_n approaches ∞, in the limit, the sequence $\{p_n\}$ of proper measures "converges" to a degenerate distribution on the set of two points $\{0, \infty\}$, a proper subset of the non-negative extended reals. That degenerate distribution may assign part or all of the unitary mass to zero, and the remainder to ∞ (this depends on the selected rate at which both, a_n and b_n approach their limits). If the rate is very fast for a_n relative to b_n, all mass goes to 0.

In recent work, Berger, *et.al* (2009), have used a criterion to verify if an improper distribution, candidate to be a "reference" prior should be termed so. It is no surprise that $p(\theta) \propto 1/\theta$, in the two-envelope inference setting, satisfies their first criterion of being permissible, however their second crterion of maximizing the missing information, does not apply in this case.

From other perspective, if one were to look for a possible "non-informative" prior, for the two envelope setting, using the concept of a data translated likelihood (Box and Tiao, 1973 page 32), one gets, with the reparametrization $\eta = log(\theta)$, that the likelihood of η is "exactly" data-translated relative to the new variable $z = log(x)$. And the uniform prior for η corresponds precisely to the already mentioned $p(\theta) \propto 1/\theta$.

So in the setting of the two-envelope paradox, steming from the computation of $E(Y|x)$, the unknown quantity which has been referred as θ, must have expected value either zero or ∞. Due to degeneracy of the prior on $\{0, \infty\}$, there is no reason to exchange, because x is either zero or ∞, so there is no paradox. It is interesting to observe that published arguments associated to the use of the improper prior, always refer to an infinite value and no mention is made of a (possible) zero value for $E(\theta)$, which we find here as the only other possibility.

6. COMMENTS

According to the classical point of view, the ilusion created by computing an incorrect expected value ignoring that in the problem the amount split in both envelopes is fixed, is responsible for the apparent paradox. Moreover, the fact that θ is fixed when you might consider switching or not, implies that both, X and Y are identically distributed.

If opening the envelope, knowledge of its contents x, leaves the odds of having in one's hand the envelope with the larger quantity, exactly as before knowing its contents. The arguments in the classical analysis, ignore the statement regarding the expected value, termed E_X, in Section 1, and hence apply for any quantity θ decided by the Benefactor. These arguments are independent of the finiteness, or not, of θ or its expected value!.

From an orthodox Bayesian perspective, the problem seems more complex, but further analysis shows that the implied prior leaves the expected contents of the envelopes to be equal, and either zero or ∞. Moreover the value of x is to be either zero or ∞, so there is clearly no preference for switching or keeping the envelope.

According to a classical or a Bayesian point of view, it is reassuring to find agreement in that there is no paradox.

The original problem is quite different from the surrogate problem mentioned in Section 4, just for completeness of the discussion. In this problem, considerations beyond statistics, need be made, once the inference problem, as shown, disappears after observing x. Also, in the surrogate problem there is clearly no paradox since the roles played by the first and second envelopes are not symmetrical; except when using the logarithmic utility, but then, there is indifference to exchanging.

ACKNOWLEDGMENTS. The author wants to thank, Columbia University for the time spent there while on sabatical leave, and also J. M. Quintana regarding the very interesting Thursday's discussions on Bayesian thinking.

REFERENCES

Berger, J.O, Bernardo, J.M. and Sun, D. (2009). "The formal definition of reference priors". *Ann. Statist.* Vol 37 No 2, pp 905-938.

Bernoulli, D. (1733), "Specimen Theoriae Novae de Mensura Sortis (Exposition of a New Theory of the Measurement of Risk)."Transalated from the Latin by Louise Sommer in *Econometrica*, Vol 22, 1954, pp 23-36.

Bernstein, P. L. (1996), "Against the Gods: The Remarkable Story of Risk". Wiley

Blachman, N.M, Christensen, R., Uts, J.M., and Finney, D.J. (1996). Letter *The American Statistician*, 50(1), pp 274-276.

Box, G.E.P. and Tiao, G.C. (1973), *Bayesian Inference in Statistical Analyis.* Addison Wesley

Broome, J. (1995). "The Two-Envelope Paradox". *Analysis*, Vol 55, No 1, pp 6-11.

Bruss, F.T. (1996), "The fallacy of the two-envelopes problem", *Math. Scientist* 21 112-119.

Casella, G., and Berger, R.L. (2002), *Statistical Inference* (2nd ed.), Duxbury Press.

Christensen, R., and Utts, J. (1992), "Bayesian Resolution of the Exchange Paradox," *The American Statistician*, 46, 274-276.

Linzer, E. (1994), "The Two-Envelope Paradox," *The American Mathematical Monthly*, 101, 417-419.

O'Reilly, F. (2006), "A Look at the Two-envelope Paradox"*STATS* Issue 45, Spring, 15-17.

Quintana, J.M., O'Reilly, F. (2008), "Demystifying the Two-Envelope Paradox?". Preimpreso No. 148 Septiembre de 2008, IIMAS, UNAM, Mexico.

Ridgeway, T. (1993), "Letter to the Editor about The Exchange Paradox," *The American Satistician*, 47, 311.

Schwitzgebel. www.faculty.ucr.edu/ eschwitz/SchwitzAbs/TwoEnvelopeSimple.htm

Zabell, S.L. (1988), Discussion of "De Finneti's theorem, induction, and $A_{(n)}$ a Bayesian non-parametric predictive inference" *Bayesian Statistics 3* pp 211-241. J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith (Eds). Oxford Univ. Press.