The Switching Problem and Conditionally Specified Distributions

F. Thomas Bruss
Université Libre de Bruxelles*

Ludger Rüschendorf
University of Freiburg†

Abstract

This article re-examines switching in the two envelopes problem with a different focus. The question is presented as a problem of the existence of a joint probability model with conditionally specified distributions. We prove the nonexistence of a solution for the classical two envelopes specification in terms of conditional distributions. Then we introduce a generalized version of the problem and, within this framework, characterize those distributions which support the switching paradox. Finally we discuss conditionally specified distributions, in connection with the two envelopes problem of Cover, and consider possible misinterpretations.

Keywords: Two envelopes problem, Fallacy, Paradox, Markov transition kernel, Characteristic function, Cover's problem.

AMS Subject Classification: 60-01 (60A99; 90D05)

1 Introduction

Various aspects of the 'two envelopes' or 'switching' problem have been discussed in several papers (see [2] – [6], [10], [14]). In the two envelopes problem a player (player I) is confronted with the question of deciding between two envelopes, one containing an amount S, and the other the amount $L = 2S$, where S is chosen by a player II.

Player I knows that player II has written down a number S on one slip and $2S$ on the other slip. In the classical formulation, S is any positive number unknown to player I. In an extended version, $S > 0$ is chosen randomly by player II randomly according to a distribution Q which is known to player I. Player I chooses at random (with probability $\frac{1}{2}$ each) either of the two envelopes with amount X, say. Player II then suggests to player I that he switch to the other envelope containing the amount Y, say.

*Dépt. Math. et ISRO, Campus Plaine CP 210, B-1050 Brussels; e-mail: tbruss@ulb.ac.be
†Math. Stochastik, Eckerstr. 1, D-79104 Freiburg; e-mail: ruschen@stochastik.uni-freiburg.de

Before we recall the 'argument' for switching, we note that the envelopes are not marked. Player I is not informed which envelope contains the first number S chosen by player II. We refer to this as the *classical* version of the two envelopes problem (also in the extended case).

The 'argument' for the switch in this classical version is that, given $X = x$, there are equal chances $\frac{1}{2}$ that Y takes the larger value $2x$ or the smaller value $\frac{x}{2}$. So the conditional distribution of Y given $X = x$ should be

$$K^*(x, \cdot) = \frac{1}{2}\varepsilon_{2x} + \frac{1}{2}\varepsilon_{\frac{x}{2}}, \tag{1.1}$$

where ε_u denotes the unit-mass distribution concentrated at the point u. This implies the conditional expectation argument for the switch

$$E(Y|X) = \frac{1}{2}2X + \frac{1}{2}\frac{X}{2} = \frac{5}{4}X. \tag{1.2}$$

Similarly, by symmetry one would also argue that the conditional distribution of X given $Y = y$ is given by

$$K^*(y, \cdot) = \frac{1}{2}\varepsilon_{2y} + \frac{1}{2}\varepsilon_{\frac{y}{2}} \tag{1.3}$$

and so $E(X|Y) = \frac{5}{4}Y$. Hence Player I would like to switch again, even without seeing either the amounts X or Y.

The switching argument based on the equality (1.2) was analyzed in detail for the non-random version by Bruss (1996) and dismissed as a fallacy of notation: X cannot serve *simultaneously* as a point of reference (i.e. a fixed value of comparison) and a random variable, because such an X cannot be measurable. This fallacy of notation can be looked upon as the most interesting part of the switching problem. Another way to express this is that the fallacy is based on an incorrect specification of the conditional distributions. Consequently, this problem does not have the authority of an established mathematical paradox.

In contrast to this discussion however, Christensen and Utts (1992) and Brams and Kilgour (1995) argue that the paradox can be established and resolved in a mathematically correct manner within an extended Bayesian framework with a random amount S.

In this paper we discuss in more detail the relation of these two points of view (see also the discussion between Brams and Kilgour (1998) and Bruss (1998)) and extend the class of models to those *allowing* for a paradoxical switching behavior.

Let S be a random variable denoting the smaller amount in the first envelope and let $L = 2S$ be the larger amount in the second envelope. The random choice X of player I can be described as

$$X = US + (1 - U)L \tag{1.4}$$

where U is independent of S, and such that $P(U = 1) = P(U = 0) = \frac{1}{2}$. The alternative choice (the switch) Y is given by

$$Y = (1-U)S + UL = 3S - X. \tag{1.5}$$

Brams and Kilgour (1995) constructed and characterized discrete and absolutely continuous distributions for S such that the switching conditions

$$E(Y|X) > X \qquad \text{and} \qquad E(X|Y) > Y \tag{1.6}$$

both hold almost surely. So 'paradoxically' Y is predicted to be larger than X and simultaneously X is predicted to be larger than Y whether the content of the chosen envelope is uncovered or not. Obviously the switching conditions in (1.6) imply that $E(X) = E(Y) = \infty$. Thus, in a framework of formal games involving real payoffs there is no place for a paradox. But even if $E(X) = E(Y) = \infty$ one could think of a finite payoff $E(Y-X)^+$, as for instance if X has a Pareto distribution with parameter $\alpha = 1$, and $Y = X + 10$. Then the condition $E(Y|X) > X$ implies that $E(Y - X) > 0$, suggesting that a switch from X to Y is desirable.

On the other hand, the problem of predicting Y based on X and conversely, predicting X based on Y, makes sense if we use mean squared errors with the conditional expectation as optimal predictors. Under the switching conditions (1.6), this leads to paradoxical conclusions. In this respect the switching paradox is similar to classical paradoxes like the St. Petersburg paradox, with infinite expectation in the background.

We may remark that the switching paradox is not specifically limited to a 'Bayesian' construction; one could imagine obtaining an offer of two envelopes with random amounts X and Y satisfying the switching property (1.6). In this formulation one obtains a pure probabilistic problem with several possible interpretations which, at the same time, give rise to some new questions.

For discretely distributed $S \geq 0$ the general switching condition (1.6) holds if and only if

$$f_S(\frac{x}{2}) < 2f_S(x) \quad \text{ for all } x, \tag{1.7}$$

where $f_S(x) = P(S = x)$, i.e. the probabilities decrease slowly enough as x grows (see Brams and Kilgour (1995)). In (1.7) we assume that at least one of the two sides of the inequality is nonzero. A similar formula holds for the case with Lebesgue densities; the case of general distributions has not been considered so far. Based on this simple characterization, it is easy to construct examples which illustrate switching behaviour.

2 The switching problem and specified conditionals

The arguments in Christensen and Utts (1992), Brams and Kilgour (1995) and Bruss (1996) show that the conditional distributions $P^{Y|X=x}, P^{X|Y=y}$ are not correctly specified in the original statement of the 'switching paradox', where they were 'suggested' as being given symmetrically by the Markov kernel defined by $K^*(x,\cdot) = \frac{1}{2}\varepsilon_{2x} + \frac{1}{2}\varepsilon_{x/2}$ and the corresponding $K^*(y,\cdot)$. Instead one obtains by simple calculations, as in Brams and Kilgour's (1995) discrete case,

$$P^{Y|X=x} = K_S(x,\cdot) = \frac{f_S(x)}{f_S(x)+f_S(2x)}\varepsilon_{\frac{x}{2}} + \frac{f_S(2x)}{f_S(x)+f_S(2x)}\varepsilon_{2x}, \tag{2.1}$$

when calculating the conditional distributions in the random choice two envelopes model correctly. A similar formula has been established for absolutely continuous distributions.

Brams and Kilgour (1995) constructed and characterized distributions in this random choice experiment, i.e. with conditional kernels given by (2.1) that display the paradoxical switching property (1.6). The question remains as to whether one can think of an experiment described by random variables X, Y such that the conditional distributions are specified by

$$K^*(x,\cdot) = P^{Y|X=x} \quad \text{and} \quad K^*(y,\cdot) = P^{X|Y=y},$$

as in the classical switching paradox. Bruss (1996) pointed out that an experiment with conditional distribution $P^{Y|X=x}$ described by the kernel K^* is different from the random choice two envelopes model, and hence also from the classical version. In this experiment an amount $X = S > 0$ is put in an envelope marked A. Then player II flips a fair coin, and according to whether it results in a head or a tail, puts $Y = 2S$ or $Y = \frac{1}{2}S$ in an envelope marked B. Player I is then offered the envelope marked A. In this case it is advantageous for playerI to switch to the envelope marked B since $P^{Y|X=x} = K^*(x,\cdot)$ and so $E(Y|X) = \frac{5}{4}X$. Thus the $\frac{5}{4}X$ argument is correct in this modified problem.

What is still unresolved in the classical version of the two envelopes problem is the following question. Can we, in the extended framework, find two random variables X and Y defined on the same probability space, such that the conditional distributions of X given Y, and Y given X, are given by K^* as in the classical two envelope specification?

The answer is negative.

Proposition 1 *There do not exist two real random variables X and Y on a probability space $(\Omega, \mathcal{A}, P)$ such that the conditional distributions of X given Y and of Y given X are given by K^*, i.e. satisfying*

$$P^{X|Y=y} = K^*(y,\cdot), \qquad P^{Y|X=x} = K^*(x,\cdot) \tag{2.2}$$

for all x, y.

Proof: Suppose first that $X > 0$ and $Y > 0$, and that X and Y fulfill condition (2.2). Let A_1, A_2 be independent random variables independent of X and Y with $P(A_i = 2) = P(A_i = \frac{1}{2}) = \frac{1}{2}$. Then $Y \stackrel{d}{=} A_1 X$ and $X \stackrel{d}{=} A_2 Y$ where $\stackrel{d}{=}$ denotes equality in distribution. Therefore, $\ln X \stackrel{d}{=} \ln A_2 + \ln Y \stackrel{d}{=} \ln A_1 + \ln A_2 + \ln X$. We obtain by independence of A_1 and A_2 for the characteristic function $\varphi_{\ln X}$ of $\ln X$

$$\varphi_{\ln X}(t) = \varphi_{\ln X}(t)(\varphi_{\ln A_1}(t))^2. \tag{2.3}$$

Observe that $\varphi_{\ln A_1}(t) = \frac{1}{2}(e^{it\ln 2}+e^{-it\ln 2}) = \cos(t\ln 2)$. This leads to a contradiction in (2.3).

The general case can be reduced to the positive case since the kernel K^* leaves $\mathbb{R}^+$ and $\mathbb{R}^-$ invariant, and so we can split the problem into two problems on $\mathbb{R}^+$ and $\mathbb{R}^-$ which can be treated as above separately. □

In Proposition 1 the question of the existence of a paradoxical two envelopes model is posed mathematically as the problem of existence of a joint distribution of (X, Y) with specified conditional distributions. Since

$$P^{(X,Y)} = P^{X|Y=y} \times P^Y(dy) = P^{Y|X=x} \times P^X(dx) \tag{2.4}$$

this question leads directly to the question of the existence of invariant probability measures with respect to a Markov kernel. If Markov kernels $K_1(y,\cdot) = P^{X|Y=y}$, $K_2(x,\cdot) = P^{Y|X=x}$ are specified, then for a distribution of (X, Y) with conditionals K_1, K_2, the conditions

$$P^X = K_1 P^Y = (K_1 \cdot K_2)P^X \qquad P^Y = (K_2 \cdot K_1)P^Y \tag{2.5}$$

must hold, where $K_1 \cdot K_2$ is the usual product of kernels and KP is the product of a kernel and a measure. From the uniqueness theorem for invariant measures one obtains that P^X and P^Y are uniquely determined by equation (2.5) if solutions of (2.5) exists, and if $K_1 \cdot K_2$ and $K_2 \cdot K_1$ are indecomposable.

In the case that conditional densities k_1, k_2 of K_1, K_2 exist w.r.t. a product measure $\mu_1 \otimes \mu_2$ a simple existence condition is known (see Arnold et al. (1992)): A joint density $f(x, y)$ with conditional densities k_1, k_2 exists if and only if

1) $\{(x, y) : k_1(x, y) > 0\} = \{(x, y) : k_2(x, y) > 0\} \quad [\mu_1 \otimes \mu_2]$

and

2) There exists functions $u(x)$ and $v(y)$ such that $k_1(x, y)/k_2(x, y) = u(x)v(y)$.

General abstract conditions (Doeblin's conditions) are well understood. They yield existence results on invariant measures in the theory of Markov chains (see Revuz (1975)). Hence, if we can solve the problem of determining the invariant measures in (2.5), then we can explicitly construct a two envelopes model with transitions K_1 and K_2 from X to Y and Y to X.

These ideas have been used extensively in characterization problems of models with specified conditional distributions (see Arnold et al. (1992)). So we could try to apply this theory to prove Proposition 1 or solve related envelope problems. Note that for the proof of Proposition 1, we used instead a simple direct argument for the nonexistence of a model conditionally specified by the classical two envelopes description.

3 A generalization of the switching problem

In this section we introduce a generalization of the two envelopes problem where the larger amount L in one of the envelopes need not be given by $L = 2S$. As before, S may be thought of as being chosen according to some known distribution Q.

Let $T := \mathbb{R}_+ \to \mathbb{R}_+$ be a continuous strictly increasing function $T(x) > x$ for all x, so that $T^k(x) \to \infty$ and $T^{-k}(x) \to 0$ for all $x > 0$ as $k \to \infty$. Consider two envelopes, one containing a random amount S and the other the larger amount $L = T(S)$. Let X and Y denote random choices between the two envelopes, so that

$$\begin{aligned} X &= US + (1-U)T(S) \\ Y &= (1-U)S + UT(S) \\ &= S + T(S) - X, \end{aligned} \tag{3.1}$$

where U is, by analogy with equation (1.4), independent of S, and such that $P(U = 1) = P(U = 0) = 1/2$.

We first discuss the case where S follows a discrete distribution denoted by $f_S(x) = P(S = x)$. Then as for the standard case we obtain, as in the paper of Brams and Kilgour (1995), the following formulae for the conditional distribution $f_{Y|x} = P^{Y|X=x}$ of Y given $X = x$:

$$\begin{cases} f_{Y|x}(T(x)) = \dfrac{f_S(x)}{f_S(T^{-1}(x)) + f_S(x)} \\[2ex] f_{Y|x}(T^{-1}(x)) = \dfrac{f_S(T^{-1}(x))}{f_S(T^{-1}(x)) + f_S(x)}. \end{cases} \tag{3.2}$$

These equations imply that $E(Y|X = x) > x$ if and only if

$$f_{Y|x}(T^{-1}(x)) = P(X = L|X = x) < \frac{T(x) - x}{T(x) - T^{-1}(x)}. \tag{3.3}$$

As for the standard case with $T(x) = 2x$, it is not difficult to construct examples where the switching condition (1.6) is fulfilled. Condition (3.3) specifies that the distribution does not decrease too fast in relation to the increase of T. This condition has a simple intuitive interpretation.

Note that condition (3.3) coincides in the standard case $T(x) = 2x$ with the switching condition (1.7). Hence, as in the classical case, it is not difficult to construct for the generalized two envelopes problem with amounts $(S, T(S))$ distributions of S which sustain the paradoxical switching phenomenon $E(Y|X) > X$ and $E(X|Y) > Y$.

In the case of general (nondiscrete) distributions Q of S it is useful to identify the distribution in the following way. Define a basic interval $A := [1, T(1))$. Then any real $y > 0$ can be represented uniquely as $T^k(m)$ for some $k \in \mathbb{Z}$ and $m \in A$.

Therefore, the random amount S can be identified with a pair of random variables $(K \in \mathbb{Z},\ M \in A)$ by

$$S = T^K(M). \tag{3.4}$$

Consider the conditional distribution of $S|M = m$ which is supported on the discrete set $\{T^k(m), k \in \mathbb{Z}\}$. We can now use the formulae for the discrete case ((3.2)) to obtain the conditional distribution of $Y|X = x$ by replacing the distribution of S by the conditional distribution of $f_{S|M=m}$ in (3.2). Here $x = T^k(m)$. Therefore, also based on the representation (3.4), the conditional probability (3.3) extends directly to the general framework. In consequence the paradoxical switching condition can be formulated for any distribution on $\mathbb{R}$, since the switching result is just based on the discrete structure of the set $\{S^k(m), k \in \mathbb{Z}\}$.

4 Conditionally specified distributions for Cover's problem

To complete our discussion, we should indicate that the switching problem in terms of conditionally specified distributions studied in this paper is also related to the two-numbers or two envelopes problem of Cover (1987). Here the problem is to choose, with maximum probability, after inspection of the first chosen number, the larger of the two numbers. No knowledge is available about the relationship between the larger and the smaller amounts, L and S (other than their being different a.s.).

For ease of reference we recall Cover's formulation of the problem:

> *Pick the largest number.* (Problem 5.1, Cover (1987)). Player I writes down any two distinct numbers on separate slips of paper. Player II randomly chooses one of these slips of paper and looks at the number. Player II must decide whether the number in his hand is the larger of the two numbers. He can be right with probability one-half. It seems absurd that he can do better.

Cover argues that player I can assure himself of a success probability strictly greater than 1/2 by looking at the first chosen amount before deciding whether or not to switch. Indeed, it suffices to obtain a random number Z (for instance from an exponential distribution) and to switch after observing X if and only if $X < Z$. This interesting observation is sometimes misinterpreted, with the wrong conclusion that Cover's statement is not true in general.

However, Cover's statement is correct. What is wrong with the 'counterexamples' we have seen is that they are based on wrongly specified conditionals, or, equivalently, the specified conditionals do not meet the formulation of Cover's problem, as exemplified in the following elementary case.

Let X and Y be two random variables defined on the same probability space by $P(X = a) = 1$ and $P(Y = 0|X) = P(Y = 2a|X) = 1/2$, where $a > 0$. The decision maker may randomize his threshold Z, and look at X. Clearly, whatever $Z = z$, the success probability equals 1/2, which seems to contradict Cover's statement. However, specifying the conditionals in this way implies that (X, Y) is not symmetrically distributed. The second sentence of Cover's hypothesis in his problem 5.1 states that the first number is chosen at random, which is common language for an equiprobable choice of each. Thus, in terms of conditional distributions, symmetry is an intrinsic assumption for Cover's problem. It is violated in the example above, because the decision maker first looks at X. Randomizing the choice of the first envelope means that the specification of X and Y switches with probability 1/2, and it is easy to check that the Z-threshold strategy (Z exponential, say) now yields a success probability p strictly greater than 1/2. The best choice is clearly any Z with $0 \le Z < 2a$, with success probability $p = 3/4$.

If the conditionals are specified, then the decision maker can do better than just randomize Z from an arbitrary continuous distribution. If, for example, the median m_x of the conditional distribution of Y given $X = x$ is known then an optimal decision rule is based on the splitting variable $Z \equiv m_x$ and the optimal success probability is $p = 3/4$ (for continuous distributions). This follows from a simple direct argument. A discussion of this problem can be found in Müller-Gronbach (1998).

Other results related to Cover's problem are given in Silverman and Nadas (1992) who study 3 observations and unconditional distributions. The paper by Gnedin and Krengel (1996) may be regarded as the most general study of Cover's Problem for unconditional distributions, since it analyses and compares the distribution-specific content of information contained in relative ranks (only) and of the (complete) observation of the values of variables. And this is indeed the essence of Cover's problem.

References

[1] Arnold, B.C., Castillo, E., and Sarabia, J.M. (1992): Conditionally specified distributions *Lecture Notes in Statistics 73*

[2] Brams, S.J.; Kilgour, D.M. (1995): The box problem: to switch or not to switch. *Math. Mag.* 68, 27-34.

[3] Brams, S.J.; Kilgour, D.M. (1998): Letter to the editor. *Math. Scientist* 23, 58-59.

[4] Bruss, F.T. (1996): The fallacy of the two-envelopes problems. *Math. Scientist* 21, 112-119.

[5] Bruss, F.T. (1998): Letter to the editor. *Math. Scientist* 23, 60-62.

[6] Christensen, R. and Utts, J. (1992): Bayesian solution of the exchange paradox. *Amer. Statist.* 46, 274-276.

[7] Cover, T. (1987): Open Problems in Communication and Computation. Problem 5.1: Pick the largest number. *Springer Verlag*, New York.

[8] Gnedin, A.V. and Krengel, U. (1996): Optimal selection based on exchangeable trials. *Ann. Appl. Probab.*, Vol. 6, N0 3, 862-882

[9] Müller-Gronbach, T. (1998): Strategien zur Bestimmung des Maximums zweier Zahlen. *Preprint*, F.U. Berlin.

[10] Rawling, P. (1994): A note on the two envelopes problem. *Theory and Decision* 36, 97-102.

[11] Revuz, D. (1975): Markov Chains. North-Holland.

[12] Samuels, S.M. (1991): Secretary problems. In *Handbook of Sequential Analysis*, ed. B.K. Gosh and P.K. Sen, Marcel Dekker, New York, 381-405.

[13] Silverman, S. and Nadas, N. (1992): On the game of Googol as *the* secretary problem *Contemporary Mathematics*, Vol. 125, 77-88.

[14] Sobel, J.H. (1994): Two envelopes. *Theory and Decision* 36, 68-96.