Thesis

Mathijs Kolkhuis Tanke

July 4, 2019

Abstract

This is my thesis.

Contents

Chapter 1

Introduction

Probability theory is one of the most important and most researched fields in mathematics. It is in essence based on just three axioms first stated by Andrey Kolmogorov [Kol33], namely that the probability of an event is non-negative, the probability measure is a unit measure and the probability measure is countably additive on disjoint sets. These three axioms however do not prevent the existence of certain paradoxes, like the Borel-Kolmogorov paradox, Monty Hall's problem and the Sleeping Beauty problem. Some paradoxes arise from wrongly using probability theory, others are a misinterpretation of results.

This thesis focuses mainly paradoxes arising from conditional probability, such as the Borel-Kolmogorov paradox, Monty Hall's problem and the two envelope problem. We study will these problems and address their paradoxical nature. Ultimately it will be shown that all paradoxes arise by incorrectly applying conditional probability.

Take a look at the Borel-Kolmogorov paradox first. A sphere is equipped with the uniform probability measure. Suppose a point exists on a great circle of the sphere, but you do not know where that point is. Is there a probability distribution on this great circle for that point? If one models this problem using latitudes, the conditional distribution on the great circle is uniform. However, if one models this problem using meridians, the conditional distribution is a cosine.

Borel [Bor09] and Kolmogorov [Kol33] both addressed this problem and Kolmogorov gave the following answer:

> Dieser Umstand zeigt, daß der Begriff der bedingten Wahrscheinlichkeit in bezug auf eine isoliert gegebene Hypothese, deren Wahrscheinlichkeit gleich Null ist, unzulässig ist: nur dann erhält man auf einem Meridiankreis eine Wahrscheinlichkeitsverteilung für [Breite] Theta, wenn dieser Meridiankreis als Element der Zerlegung der ganzen Kugelfläche in Meridiankreise mit den gegebenen Polen betrachtet wird. (Andrey Kolmogoroff, [Kol33])

In summary, Kolmogorov states that the concept of conditional probability on a great circle is inadmissible, since the event that the point lies on a great circle of the sphere has zero measure. Furthermore, the sole reason of a cosine distribution on a great circle arising when considering meridians is that the

meridian circle serves as an element of the decomposition of the whole spherical surface in meridian circles with the poles considered.

Despite Kolmogorov's explanation this problem is still upon debate. Recently Gyenis, Hofer-Szabó and Rédei [GHSR17] studied this problem and provided more insight and an in my opinion satisfying discussion to the problem, eventually drawing the same conclusion as Kolmogorov already did. This problem is more elaborately discussed in Chapter 2 where we will expand the analysis of [GHSR17] and not only consider latitudes and meridians, but their combination as well. Furthermore, we will look whether a conditional probability on a zero set can be uniquely approached by traditional conditional probability on sets of decreasing measure as stated in Conjecture 2.3.5. This conjecture turns out to be false, again affirming that conditional probabilities on sets of measure zero are not uniquely defined.

Another paradox arising from conditional probability is Monty Hall's problem. In Monty Hall's problem, a player is facing three doors called a, b and c. One door has a car behind and the other two have goats. Suppose the player initially chooses door a. The game master then opens either door b or c, but always a door with a goat. The player now faces two doors and is asked whether he wants to switch. One possible solution is that the player faces two doors without any preference for either door, thus the probability is 50%. Another possible solution is that first the player had a 33% chance of correctly guessing the door with the car. If for example door c is opened, door b remains with a conditional probability of 67% of having the car. Which solution is correct?

Let $\mathcal{X} = \{a, b, c\}$ be the set of doors and U the random variable denoting the location of the car. Conditional probability then states

$$\mathbb{P}[U = b \mid U \in \{a, b\}] = \frac{\mathbb{P}[U = b, U \in \{a, b\}]}{\mathbb{P}[U \in \{a, b\}]} = \frac{\frac{1}{3}}{\frac{2}{3}} = \frac{1}{2},$$

supporting the claim that a probability of 50% is the correct answer. However, the space conditioned on door c is opened is $\{a, b\}$ and the space conditioned on door b is opened is $\{a, c\}$. If door a has the car, the game master can open either door and the set of events we must condition on is $\{\{a, b\}, \{a, c\}\}$. These two events do not form a partition, preventing us from using traditional conditional probability.

Thus in addition to Kolmogorov's statement, we must not only be wary for conditioning on events with measure 0, we cannot condition on arbitrary events at all. I propose that when using conditional probability, one must provide a pair of a sub-σ-algebra and the event from that σ-algebra to condition on. Regarding Monty Hall's problem there is no σ-algebra on $\{a, b, c\}$ containing the set $\{\{a, b\}, \{a, c\}\}$. In the case of the Borel-Kolmogorov paradox providing sub-σ-algebras immediately make clear why the conditional distribution on a great circle is not unique, as both calculations are supported by different sub-σ-algebras.

The crux of the Monty Hall problem is that the initial distribution of the car is unknown and the player does not know with which probability door b is opened given the car is behind door a. Thus there is a set of different possible probability distributions where one is the correct distribution. Using the theory of *safe probability* we can obtain a strategy that give equal results for all distributions in a specified model. This theory is introduced by Grünwald [Gr 18b] and

summarized here in Chapter 3. Safe probability is then applied in Chapter 4 to problems as the Monty Hall problem and to the two envelope problem in Chapter 5.

In Chapter 4 the results of the analysis of Monty Hall's problem are generalized to a theorem, which can be used to provide safe distributions for other problems like the boy or girl problem.

Theorem 4.1.1. *Let $\mathcal{X}$ be countable and $\mathcal{Y}$ be finite. Let U be an $\mathcal{X}$-valued random variable and V be a $\mathcal{Y}$-valued random variable. Let $p_u \in [0,1]$ with $\sum_{u \in \mathcal{X}} p_u = 1$. Let*

$$\mathcal{P}^* \subseteq \{\mathbb{P} \mid \forall u \in \mathcal{X} : \mathbb{P}[U = u] = p_u\} \tag{1.1}$$

be our model of probability distributions on $\mathcal{X} \times \mathcal{Y}$ where $|\mathcal{Y}|$ distributions $\mathbb{P}_1, \ldots, \mathbb{P}_{|\mathcal{Y}|} \in \mathcal{P}^*$ impose $|\mathcal{Y}|$ linearly independent vectors $(\mathbb{P}_i[V = v])_{v \in \mathcal{Y}}$ with $i \in \{1, \ldots, |\mathcal{Y}|\}$. Let $u \in \mathcal{X}$ be arbitrary and let $\tilde{\mathbb{P}}$ be a distribution on $\mathcal{X} \times \mathcal{Y}$ with full support on V, then the following are equivalent:

1. *For all $v \in \mathcal{Y}$ we have $\tilde{\mathbb{P}}[U = u|V = v] = p_u$.*

2. *$\tilde{\mathbb{P}}$ is safe for $\mathbb{1}_{\{U=u\}} |[V]$.*

3. *$\tilde{\mathbb{P}}$ is safe for $\langle \mathbb{1}_{\{U=u\}} \rangle | \langle V \rangle$.*

This theorem and its notation will be explained, proven and applied in Chapter 4, but it essentially states the following: in the case of the Monty Hall problem if one assumes the car is initially distributed evenly between the doors, the probability of the car being behind the originally chosen door a can be assumed to be $\frac{1}{3}$. Using this assumption one must always switch to the other door as the probability of the car being behind that door can be assumed to be $\frac{2}{3}$, resulting in a 67% chance of winning the car. It is not at all clear whether this is the correct distribution. However, this assumption always yields to a 67% chance of winning the car independent on the probability of opening door b when the car is behind door a.

Another paradox we will treat is the two envelope problem. There are two envelopes, where one is filled with value a and the other with value b. The only thing the player knows is that either $a = 2b$ or $b = 2a$. An envelope is given to the player, with the player receiving either envelope with equal probability. Call this envelope A and suppose the player observes value a. Should he switch to envelope B or keep the contents of A?

At first glance, he should. Either $b = 2a$ or $b = \frac{a}{2}$ holds with equal probability, thus

$$\mathbb{E}[B|A = a] = \frac{1}{2} \cdot 2a + \frac{1}{2} \cdot \frac{a}{2} = \frac{5}{4}a.$$

However, the player has received an envelope at random and only knows the contents of that particular envelope. So he could also have been given envelope B with value b, for which $\mathbb{E}[A|B = b] = \frac{5}{4}b$ holds. Therefore no matter what envelope the player receives, he should switch. This solution must clearly be wrong and it is. When conditioning on $A = a$, then either $b = 2a$ or $a = 2b$ holds with probability 1 in the conditioned probability space. This renders the previous calculation of $\mathbb{E}[B|A = a]$ to be incorrect. However, we do not know which event takes place.

Now let $x = \min\{a, b\}$ be the lowest value, then either $a = x$ or $a = 2x$ holds with equal probability. For both envelopes we now have

$$\mathbb{E}[A|B = b] = \mathbb{E}[B|A = a] = \frac{1}{2}x + \frac{1}{2} \cdot 2x = \frac{3}{2}x.$$

Both envelopes thus have on average the same contents, as is expected when the envelopes are handed out uniformly. Unfortunately, we do not know the value of x. Furthermore, if x is picked from an unbounded set and we assume that given an observation the other envelope must hold twice as much or half the value with equal probability, then x must be taken from a uniform probability distribution on the unbounded set. Such distribution does not exist. We conclude that the value of $\mathbb{E}[B|A = a]$ does depend on a prior distribution on the chosen value x.

There are two ways to address this problem. Using safe probability, like in the case of Monty Hall, we will devise a strategy where the player wins $\frac{3}{2}\mathbb{E}[X]$ on average, where X the actual unknown distribution for the lowest value in the envelopes on the positive number line with finite expectation. This strategy is in fact that the player must flip a fair coin and switch when it lands heads, which is quite intuitive.

Another method is using a switching strategy introduced by McDonell and Abbott [MA09, ADP10, MGL$^+$11]. Let $f\colon (0, \infty) \to [0, 1]$ be a function that takes an observation a and equips it with a probability. The player must switch with probability $f(a)$ when observing value a. When f is a decreasing function and strictly decreases on an interval where distribution X has positive measure, then switching using f will always return on average a higher value than $\frac{3}{2}\mathbb{E}[X]$. If f is a threshold switch, for example $f = \mathbb{1}_{(0,a^*]}$, then for some distributions X the strategy f is actually the best strategy available. However, when playing the game, the player does not know how much strategy f outperforms switching using safe probability. By any means, distribution X gives f an arbitrarily small advantage over safe probability, making it an unnecessarily complicated method to marginally increase the average value won.

The two envelope problem will be discussed further in Chapter 5.

For every problem there is a different reason for getting paradoxical results. There is one recurring theme with all paradoxes, namely when analysing these problems most of the times the underlying σ-algebra is not taken into account. This yields to various conflicting results, as conditioning on events that cannot be conditioned on to not recognizing that multiple probability distributions are possible. The theme of is thesis this therefore that when doing probability theory, the underlying probability space and σ-algebra must never be ignored and not providing a sub-σ-algebra with a conditional distribution must become a bad habit instead of an accepted practice.

Chapter 2

The Borel-Kolmogorov paradox

The first paradox we will study in more detail is the Borel-Kolmogorov paradox. Many accredit the origin of the problem to Borel in 1909 [Bor09], but the problem is originally conceived by Bertrand in 1889 [Ber89]. Kolmogorov [Kol33] first provided a meaningful and discussion answer to the problem in 1933, which is why the problem is now called the Borel-Kolmogorov paradox. We will take a look at the following version of the paradox.

Suppose a random variable has a uniform probability distribution on the unit sphere. If one conditions on a great circle, what will the resulting conditional distribution be? If the great circle is viewed as a latitude, the conditional distribution will be uniform as well. However, if the great circle is modelled as a meridian, the conditional distribution has a cosine as density function. We have two different conditional distributions on the same set, thus which one is correct?

This problem seems a bit specific, however it can occur in many other more realistic cases. One example is given by Proschan and Presnell [PP98]. There two random variables X and Y are considered, both distributed according to the standard normal distribution. A professor once asked the students on a test to describe the distribution of Y given $X = Y$. In short, there are three different ways to attack this problem. Some students calculated the density of Y given $Y - X = 0$, others calculated the density of Y given $\frac{Y}{X} = 1$ and another student calculated the density of Y given $\mathbb{1}_{\{Y=X\}} = 1$. All three methods create different density functions, but the methods used by the students are all correct. The crux is that the original question posed by the professor is flawed as the event $\{X = Y\}$ has zero probability if X and Y are continuous random variables. This is why this problem is related to the Borel-Kolmogorov paradox.

The basic solution to the paradox is to use measure-theoretic conditional probability, so that one is given not just a set like a great circle to condition on but rather a set and an accompanying sub-σ-algebra containing this set. The definition of measure-theoretic conditional probability now provides an answer that is unique and meaningful up to a set of zero measure.
When conditional probability is defined traditionally based on density functions, it is defined on every point even though each such point has measure zero.

One may now ask the following question: can we generally extend the measure-theoretic conditional probability to be defined on all points once a sub-σ-algebra is provided, by treating the conditional probability at that point as a limit of traditional conditional probabilities where one conditions on smaller and smaller sets? In Section 2.3.4 we provide Conjecture 2.3.5 that would imply this, but we will then show this conjecture is wrong. If we condition on both a given meridian and a suitably rotated latitude, then two phenomena occur:

1. the only sub-σ-algebra that we can provide is the full Borel σ-algebra and conditioning on the full σ-algebra leads to no effect and

2. the required limit is undefined as the limit of smaller and smaller meridians takes on a different value than the limit of smaller and smaller rotated latitudes.

Kolmogorov [Kol33] pointed out that the great circle has zero measure and according to him conditioning on a great circle must not be allowed. One will be tempted to disregard conditioning on zero sets at all, but this brings many problems as argued by De Finetti [dF72] and Howson [How14]. One example from De Finetti is the field of statistical applications where it is realistic to regard the probability of the event of observations E as 0. It is only the probabilities conditioned on these E's that determine the optimal actions in these cases. Therefore, we cannot just disregard conditioning on probabilities with measure 0, as those conditional probabilities are needed for applying statistics.

Our solution, supported by [Rao88, Bil95, PP98, Eas08, Myr15, GHSR17], is that when performing conditional probability, the σ-algebra of the sets conditioned on must always be taken into account. Otherwise the event E with zero measure has no meaning as there are many methods to approach E by events with positive measures, giving different conditional probabilities.

Our analysis concerns conditioning on zero sets using measure-theoretic conditional expectations as such conditioning is defined there. The article of Gyenis, Hofer-Szabó and Rédei [GHSR17] is followed to compute the conditional distributions on the great circle, viewing that circle both as latitude and as meridian. We will take it one step further by also considering the σ-algebra of both the meridians and latitudes and by considering conditional distributions on rotated latitudes making them coincide with meridians. We will then conclude that once again conditioning on only measure zero sets must not be admissible, as it will lead to contradicting results.

After drawing these conclusions, we will rename the Borel-Kolmogorov paradox to the *Borel-Kolmogorov phenomenon*, as it is more an example on why you should not condition on zero sets than it is a paradox.

2.1 Conditional expectation

First we need to define conditional expectations. We use the standard measure-theoretic definitions as given by David Williams [Wil91].

Let Ω be an arbitrary sample space and let $\mathcal{F}$ be a σ-algebra on Ω. The indicator function $\mathbb{1}_A$ on a set A is defined as

$$\mathbb{1}_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases} \tag{2.1}$$

Let $\mathcal{B}(A)$ be the σ-algebra of all Borel-measurable sets on $A \subseteq \mathbb{R}^n$. We will start with the definition of a random variable.

Definition 2.1.1 (Random variable). Let $\mathcal{X}$ be a measurable space. A function $X \colon \Omega \to \mathcal{X}$ is an $\mathcal{X}$-*valued random variable* if it is $\mathcal{F}$-measurable, thus if $X^{-1}(A) \in \mathcal{F}$ holds for all $A \in \mathcal{G}$ where $\mathcal{G}$ is a σ-algebra on $\mathcal{X}$.

We can now define the conditional expectation as definition 9.2 of [Wil91].

Definition 2.1.2 (Conditional expectation). Let X be an $\mathcal{F}$-measurable random variable with finite $\mathbb{E}[|X|]$. Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. There exists a random variable Y such that

1. Y is $\mathcal{G}$-measurable,

2. $\mathbb{E}[|Y|]$ is finite,

3. for every $G \in \mathcal{G}$ we have

$$\int_G Y\, d\mathbb{P} = \int_G X\, d\mathbb{P}. \tag{2.2}$$

Y is called a *version of the conditional expectation* $\mathbb{E}[X|\mathcal{G}]$ of X given $\mathcal{G}$, written as $Y = \mathbb{E}[X|\mathcal{G}]$ almost surely. Moreover, if $\tilde{Y}$ is another random variable with these properties, then $\tilde{Y} = Y$ equal almost surely. Since two versions of $\mathbb{E}[X|\mathcal{G}]$ coincide almost surely, Y is also called *the* conditional expectation $\mathbb{E}[X|\mathcal{G}]$.

Note that when $\mathcal{G} = \sigma(Z)$ is the smallest σ-algebra generated by a set $Z \in \mathcal{F}$, we also write $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X|Z]$. This generalizes to $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X|Z_1, Z_2, \dots]$ when $\mathcal{G} = \sigma(Z_1, Z_2, \dots)$.

Definition 2.1.2 extends very nicely to the definition of conditional probability. Note that $\mathbb{E}[\mathbb{1}_F] = \mathbb{P}[F]$ holds for all $F \in \mathcal{F}$ and this carries over to conditional probabilities.

Definition 2.1.3 (Conditional probability). If $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{F}$ and $F \in \mathcal{F}$, then the *conditional probability* $\mathbb{P}[F|\mathcal{G}]$ is a version of $\mathbb{E}[\mathbb{1}_F|\mathcal{G}]$.

2.1.1 Comparison to traditional definitions

The traditional definition of traditional density-based conditional expectation and probability is the following: when X and Y are two random variables on $\mathbb{R}$, $f_{X,Y}$ is the joint density function of X and Y and f_Y is the density function of X, the conditional expectation $\mathbb{E}[X|Y = y]$ is defined as

$$\mathbb{E}[X|Y = y] = \int_{\mathbb{R}} x \frac{f_{X,Y}(x,y)}{f_Y(y)}\, dx. \tag{2.3}$$

Definition 2.1.2 agrees on this traditional usage although it can be modified on a set of measure zero, as we will see now. Take $\mathcal{G} = \sigma(\{Y = y\})$. The smallest σ-algebra generated by all $\omega \in \mathbb{R}$ such that $Y(\omega) = y$ and define

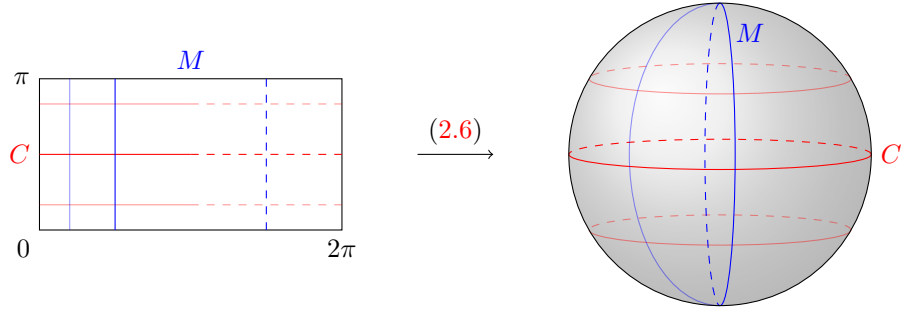$$g(y) = \int_{\mathbb{R}} x \frac{f_{X,Y}(x,y)}{f_Y(y)}\, dx, \tag{2.4}$$

Figure 2.1: A visualization of coordinate transformation (2.6) from Euclidean to spherical coordinates. The rectangle represents the set S. The horizontal lines in the rectangle like C parametrize the latitudes. The vertical lines in the rectangle like M parametrize the meridians. Note that the two blue vertical lines of M are distance π apart, this is needed to fully parametrize a meridian.

then $g(Y)$ is a version of $\mathbb{E}[X|Y]$. The proof and a more general statement can be found in section 9.6 of [Wil91].

The traditional conditional expectation can also be extended to traditional conditional probability.

Definition 2.1.4 (Traditional conditional probability)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $F \in \mathcal{F}$ be with positive measure and let $E \in \mathcal{F}$ be measurable. The *traditional conditional probability* of E given F is given by

$$\mathbb{P}[E|F] = \frac{\mathbb{P}[E \cap F]}{\mathbb{P}[F]}. \tag{2.5}$$

If F has measure zero, the traditional conditional probability of E given F is undefined.

2.2 Formal description of the probability space

To analyse the Borel-Kolmogorov paradox, we need to formalize our probability space. The following construction is visualized in Figure 2.1. Let S be the unit sphere. For easier calculations, we equip S with polar coordinates. Define $S = [0, 2\pi) \times [0, \pi]$, then S is described in the Euclidean space with the function

$$S \to \mathbb{R}^3 : (\phi, \psi) \mapsto (\cos \phi \sin \psi, \sin \phi \sin \psi, \cos \psi). \tag{2.6}$$

Let $\mathcal{B} = \mathcal{B}(S)$ be the Borel-σ-algebra on S. The uniform distribution on S is defined as

$$\mathbb{P}[B] = \frac{1}{4\pi} \iint_B \sin \psi \, d\psi \, d\phi \tag{2.7}$$

for a $B \in \mathcal{B}$. The triple $(S, \mathcal{B}, \mathbb{P})$ form a probability space.

The set of latitudes is described by

$$\mathcal{C} = \{[0, 2\pi) \times \{\psi\} \mid \psi \in [0, \pi]\} \tag{2.8}$$

8

and will be horizontal lines in the rectangle of Figure 2.1. The set of meridians is described by

$$\mathcal{M} = \{\{\phi, \phi + \pi\} \times [0, \pi] \mid \phi \in [0, \pi)\} \tag{2.9}$$

and will be the vertical lines in the rectangle of Figure 2.1.

Note that the function in (2.6) is not a bijection. The image of S is not one-to-one on the north and south pole, but this is a null set and will not cause any problems in our case. The reason why (2.6) is not made a formal bijection is that it eases notation.

2.3 Conditional distributions

We will now explore various ways to describe the conditional distribution on a great circle.

2.3.1 Traditional conditional probability

The first method is the naive method using traditional conditional probability. The probability space here is $(S, \mathcal{B}, \mathbb{P})$. Let $F \in \mathcal{B}$ be a great circle.

Note that $\mathbb{P}[F] = 0$, as a great circle always has zero measure. To be precise, let $f \colon S \to \mathbb{R}^3$ be the coordinate transformation of (2.6). All great circles F' have a rotation $O \colon \mathbb{R}^3 \to \mathbb{R}^3$ such that $(f^{-1} \circ O \circ f)(F') = [0, 2\pi) \times \left\{\frac{\pi}{2}\right\}$. Rotations are orthogonal and have determinant 1, thus

$$\mathbb{P}[F] = \frac{1}{4\pi} \iint_F \sin\psi d\psi d\phi \tag{2.10}$$

$$= \frac{1}{4\pi} \int_0^{2\pi} \int_{\frac{1}{2}\pi}^{\frac{1}{2}\pi} \sin\psi \det(f^{-1} \circ O \circ f) d\psi d\phi \tag{2.11}$$

$$= \frac{1}{4\pi} \int_0^{2\pi} \int_{\frac{1}{2}\pi}^{\frac{1}{2}\pi} \sin\psi d\psi d\phi = 0. \tag{2.12}$$

The inverse f^{-1} is well-defined here since f is a bijection locally around the circle $[0, 2\pi) \times \left\{\frac{\pi}{2}\right\}$. Let $E \subset F$ be a measurable subset of F, then the traditional conditional probability of E given F equals

$$\mathbb{P}[E|F] = \frac{\mathbb{P}[E \cap F]}{\mathbb{P}[F]}, \tag{2.13}$$

which is not defined as $\mathbb{P}[F] = 0$.

Thus the traditional interpretation of conditional probability will not give an answer. Furthermore, a single great circle does not yield enough information to compute any conditional probability. In terms of Definition 2.1.2 the sub-σ-algebra considered here is $\mathcal{G} = \{\emptyset, S, B, S \setminus B\}$, which turns out to be too small.

2.3.2 Conditional probability on latitudes

Since four-element sub-σ-algebras are too small, we need to condition on larger sub-σ-algebras. One option is the σ-algebra of latitudes. Let

$$\mathfrak{C} = \sigma\left(\{[0, 2\pi) \times A \mid A \in \mathcal{B}([0, \pi])\}\right) \tag{2.14}$$

be the σ-algebra of all measurable subsets of latitudes. We can then calculate the conditional expectation of $\mathbb{E}[X|\mathfrak{C}]$.

Proposition 2.3.1. *Let X be $\mathcal{B}$-measurable. The conditional expectation of $\mathbb{E}[X|\mathfrak{C}]$ is given by*

$$\mathbb{E}[X|\mathfrak{C}](\phi, \psi) = \frac{1}{2\pi} \int_0^{2\pi} X(\phi', \psi) d\phi' \tag{2.15}$$

with $(\phi, \psi) \in S$.

Proof. The proof is taken from [GHSR17]. Take $A \in \mathcal{B}([0, 2\pi))$ and consider $C = [0, 2\pi) \times A \in \mathfrak{C}$. Since $\mathbb{P}$ is the uniform measure on the surface of the unit sphere, we have

$$\int_C X d\mathbb{P} = \frac{1}{4\pi} \int_A \int_0^{2\pi} X(\phi, \psi) \sin \psi d\phi d\psi \tag{2.16}$$

by the standard spherical to Euclidean coordinate transformation. We can now apply the same coordinate transformation on the integral of $\mathbb{E}[X|\mathfrak{C}]$:

$$\int_C \mathbb{E}[X|\mathfrak{C}] d\mathbb{P} = \frac{1}{4\pi} \int_A \int_0^{2\pi} \mathbb{E}[X|\mathfrak{C}](\phi, \psi) \sin \psi d\phi d\psi. \tag{2.17}$$

Filling in $\mathbb{E}[X|\mathfrak{C}]$ and rewriting yields

$$\int_C \mathbb{E}[X|\mathfrak{C}] d\mathbb{P} = \frac{1}{4\pi} \int_A \int_0^{2\pi} \mathbb{E}[X|\mathfrak{C}](\phi, \psi) \sin \psi d\phi d\psi \tag{2.18}$$

$$= \frac{1}{4\pi} \int_A \int_0^{2\pi} \frac{1}{2\pi} \left(\int_0^{2\pi} X(\phi', \psi) d\phi' \right) \sin \psi d\phi d\psi \tag{2.19}$$

$$= \frac{1}{4\pi} \int_A \left(\int_0^{2\pi} \frac{1}{2\pi} d\phi \right) \int_0^{2\pi} X(\phi', \psi) \sin \psi d\phi' d\psi \tag{2.20}$$

$$= \frac{1}{4\pi} \int_A \int_0^{2\pi} X(\phi, \psi) \sin \psi d\phi d\psi = \int_C X d\mathbb{P}. \tag{2.21}$$

Note that $\mathfrak{C}$ is generated by sets like C, thus $\mathbb{E}[X|\mathfrak{C}]$ is a well-defined version of the $\mathfrak{C}$-conditional expectation of X. $\qquad \square$

This derivation is slightly different from the one in [GHSR17]. The corrections on [GHSR17] are given in Appendix A.1.

As we now have a $\mathfrak{C}$-conditional expectation, we can look at the measure space $(S, \mathfrak{C})$. Let $\psi' \in (0, \pi)$ be arbitrary, then from $\mathfrak{C}$ we can take a latitude $C = [0, 2\pi) \times \{\psi'\}$ and a measurable arc $A = [\phi_1, \phi_2] \times \{\psi'\} \subset C$. Let $\mathbb{P}'$ be the probability measure taking 1 on C and 0 on C^c, then the conditional probability $\bar{\mathbb{P}}$ of points being on A given there is a point on C is

$$\bar{\mathbb{P}}[A] = \int_S \mathbb{P}[A|\mathfrak{C}] d\mathbb{P}' = \frac{1}{2\pi} \int_S \int_0^{2\pi} \mathbb{1}_A(\phi', \psi) d\phi' d\mathbb{P}'(\phi, \psi) \tag{2.22}$$

$$= \frac{1}{2\pi} \left(\int_{S \setminus C} \int_{\phi_1}^{\phi_2} 0 d\phi' d\mathbb{P}' + \int_C \int_{\phi_1}^{\phi_2} 1 d\phi' d\mathbb{P}' \right) \tag{2.23}$$

$$= \frac{\phi_2 - \phi_1}{2\pi}. \tag{2.24}$$

Thus the conditional probability distribution on C is uniform, resulting in all latitudes having uniform conditional probability measure. Since the unit sphere itself has a uniform probability measure, this result is as expected.

2.3.3 Conditional probability on meridians

A great circle cannot only be described as a latitude, but also as a meridian. Therefore it is interesting whether describing great circles as meridians yield to the same conditional distribution. Let

$$\mathfrak{M} = \sigma(\{A \times [0,\pi] \mid A \in \mathcal{B}([0,2\pi))\}) \tag{2.25}$$

be the σ-algebra of measurable subsets of meridians. We can now calculate the conditional expectation $\mathbb{E}[X|\mathfrak{M}]$.

Proposition 2.3.2. *Let X be $\mathcal{B}$-measurable. The conditional expectation of $\mathbb{E}[X|\mathfrak{M}]$ is given by*

$$\mathbb{E}[X|\mathfrak{M}](\phi,\psi) = \frac{1}{2} \int_0^\pi X(\phi,\psi') \sin\psi' d\psi' \tag{2.26}$$

with $(\phi,\psi) \in S$.

Proof. The proof is largely taken from [GHSR17]. Let $A \in \mathcal{B}([0,2\pi))$ be measurable and consider $M = A \times [0,\pi] \in \mathfrak{M}$. Coordinate transformation between polar coordinates and the uniform measure on the circle $\mathbb{P}$ and further rewrites yield

$$\int_M \mathbb{E}[X|\mathfrak{M}]d\mathbb{P} = \frac{1}{4\pi} \int_0^\pi \int_A \mathbb{E}[X|\mathfrak{M}](\phi,\psi) \sin\psi d\phi d\psi \tag{2.27}$$

$$= \frac{1}{4\pi} \int_0^\pi \int_A \left(\frac{1}{2} \int_0^\pi X(\phi,\psi') \sin\psi' d\psi' \right) \sin\psi d\phi d\psi \tag{2.28}$$

$$= \frac{1}{4\pi} \int_A \int_0^\pi X(\phi,\psi') \sin\psi' d\psi' d\phi \tag{2.29}$$

$$= \int_M X d\mathbb{P}. \tag{2.30}$$

Note that $\mathfrak{M}$ is generated by sets like M, thus $\mathbb{E}[X|\mathfrak{M}]$ is a well-defined version of the $\mathfrak{M}$-conditional expectation of X. □

Note that this representation is different from equations (81) and (112) of [GHSR17], where equation 81 is

$$\mathbb{E}[X|\mathfrak{M}](\phi,\psi) = \frac{1}{2} \int_0^{2\pi} X(\phi,\psi') |\sin\psi'| d\psi' \tag{2.31}$$

and equation 112 is

$$\mathbb{E}[X|\mathfrak{M}](\phi,\psi) = \int_0^\pi X(\phi,\psi') \sin\psi' d\psi'. \tag{2.32}$$

The corrections on [GHSR17] are given in Appendix A.2. We can verify that our version is the correct one.

Take first the meridian $M = \{\phi', \phi' + \pi\} \times [0, \pi] \in \mathfrak{M}$ with $\phi' \in [0, \pi)$. Let $\psi_1^*, \psi_2^* \in \mathbb{R}$ be arbitrary with $\psi_2^* - 2\pi \leq \psi_1^* \leq \psi_2^*$, define the angles $\psi_1 = \psi_1^*$ mod 2π and $\psi_2 = \psi_2^*$ mod 2π and define arc $A \subseteq M$ as

$$A = \begin{cases} \{\phi'\} \times [\psi_1, \psi_2], & \psi_1, \psi_2 \leq \pi, \\ \{\phi'\} \times [\psi_1, \pi] \cup \{\phi' + \pi\} \times [\psi_2 - \pi, \pi], & \psi_1 \leq \pi, \psi_2 > \pi, \\ \{\phi' + \pi\} \times [\psi_1 - \pi, \psi_2 - \pi], & \psi_1, \psi_2 > \pi, \\ \{\phi'\} \times [0, \psi_1] \cup \{\phi' + \pi\} \times [0, \psi_2 - \pi], & \psi_2 \leq \pi, \psi_1 > \pi. \end{cases} \tag{2.33}$$

This definition is exhaustive, yet it provides all possible arcs on a meridian while restricting ourselves to the domain $[0, 2\pi) \times [0, \pi]$. Now, analogous to the latitudes, let $\mathbb{P}'$ be the uniform measure taking 1 on meridian M and 0 on M^c. The conditional probability of $\hat{\mathbb{P}}$ of points being on A with $\psi_1, \psi_2 \leq \pi$ given there is a point on M is given by

$$\hat{\mathbb{P}}[A] = \int_S \mathbb{P}[A|\mathfrak{M}] d\mathbb{P}' = \frac{1}{2} \int_S \int_0^\pi \mathbb{1}_A(\phi, \psi') \sin \psi' d\psi' d\mathbb{P}'(\phi, \psi) \tag{2.34}$$

$$= \frac{1}{2} \left(\int_{S \setminus M} \int_{\phi_1}^{\phi_2} 0 d\psi' d\mathbb{P}' + \int_M \mathbb{1}_{\{\phi'\} \times [0, \pi]} \int_{\phi_1}^{\phi_2} \sin \psi' d\psi' d\mathbb{P}' \right) \tag{2.35}$$

$$= \frac{1}{4} \int_{\phi_1}^{\phi_2} \sin \psi' d\psi' d\mathbb{P}' = \frac{\cos \psi_1 - \cos \psi_2}{4} \tag{2.36}$$

since $\int_{\{\phi'\} \times [0, \pi]} d\mathbb{P}' = \frac{1}{2}$. On the other possible arcs of M the probability $\hat{\mathbb{P}}[A]$ with ψ_1, ψ_2 as in the definition of A becomes

$$\hat{\mathbb{P}}[A] = \begin{cases} \frac{1}{4} (\cos \psi_1 - \cos \psi_2), & \psi_1, \psi_2 \leq \pi, \\ \frac{1}{4} (2 + \cos \psi_1 - \cos \psi_2), & \psi_1 \leq \pi, \psi_2 > \pi, \\ \frac{1}{4} (\cos \psi_2 - \cos \psi_1), & \psi_1, \psi_2 > \pi, \\ \frac{1}{4} (2 - \cos \psi_1 + \cos \psi_2), & \psi_2 \leq \pi, \psi_1 > \pi. \end{cases} \tag{2.37}$$

Now one can immediately check that $\hat{\mathbb{P}}$ is well-defined on M as

$$\hat{\mathbb{P}}[M] = \int_S \mathbb{P}[A|\mathfrak{M}] d\mathbb{P}' = \frac{1}{2} \int_M \int_0^\pi \sin \psi' d\psi' d\mathbb{P}' = -\frac{1}{2} (\cos \pi - \cos 0) = 1. \tag{2.38}$$

Clearly this conditional distribution on meridians is not uniform. However, one should expect they are, as a meridian is just a rotation of a latitude and the points on the sphere are spread uniformly. An explanation of this difference is given in Section 2.4.

2.3.4 Combining latitudes and meridians

Another question one could ask is the following: if I define $\Sigma = \sigma(\mathfrak{C}, \mathfrak{M})$ as the smallest σ-algebra containing both measurable subsets of meridians and latitudes, what is then the conditional probability distribution on a great circle given Σ? The answer is that in this approach the distributions in sections 2.3.2 and 2.3.3 can be recovered as limiting distributions of a sequence of traditional conditional probabilities defined in Section 2.3.1.

First we'll further analyse the new σ-algebra Σ. Consider an arbitrary rectangle $(a, b) \times (c, d) \subset [0, 2\pi) \times [0, \pi]$. Since by definition $(a, b) \times [0, \pi] \in \mathfrak{M}$ and $[0, 2\pi) \times (c, d) \in \mathfrak{C}$, we have

$$(a, b) \times (c, d) = ((a, b) \times [0, \pi]) \cap ([0, 2\pi) \times (c, d)) \in \Sigma. \tag{2.39}$$

Thus all Borel-measurable sets on our sphere are contained in Σ. Furthermore, as $\mathfrak{C}$ and $\mathfrak{M}$ contain only Borel-measurable sets, $\Sigma = \sigma(\mathfrak{C}, \mathfrak{M})$ can only have Borel-measurable sets. Therefore $\Sigma = \mathcal{B}$ is the σ-algebra of Borel-measurable sets on the sphere.

Now take the following approach. Pick $x \in (0, \pi)$. Define the two rectangles $C_\epsilon = [0, 2\pi) \times [x - \epsilon, x + \epsilon]$ and $R_\epsilon = [a, b] \times [x - \epsilon, x + \epsilon] \subseteq C_\epsilon$ for all $\epsilon \in (0, \min\{\pi - x, x\})$, then what is the limiting conditional probability of R_0 given C_0? The conditional probability of R_0 given C_0 can be modeled as the limit of the sequence $\mathbb{P}[R_\epsilon | C_\epsilon]$ with $\epsilon \to 0$, which will take on the uniform distribution on C_0 as calculated in Equation 2.22.

Proposition 2.3.3. *Define the two rectangles $C_\epsilon = [0, 2\pi) \times [x - \epsilon, x + \epsilon]$ and $R_\epsilon = [a, b] \times [x - \epsilon, x + \epsilon] \subset C_\epsilon$ for all $\epsilon \in (0, \min\{\pi - x, x\})$ with $x \in (0, \pi)$. The limiting conditional probability of R_0 given C_0 equals*

$$\mathbb{P}[R_0 | C_0] = \lim_{\epsilon \downarrow 0} \mathbb{P}[R_\epsilon | C_\epsilon] = \frac{b - a}{2\pi}. \tag{2.40}$$

Proof. Let $\epsilon \in (0, \min\{\pi - x, x\})$. Let $C_\epsilon = [0, 2\pi) \times [x - \epsilon, x + \epsilon] \in \mathfrak{C}$. The probability $\mathbb{P}[R_\epsilon | C_\epsilon]$ equals

$$\mathbb{P}[R_\epsilon | C_\epsilon] = \frac{\mathbb{P}[R_\epsilon \cap C_\epsilon]}{\mathbb{P}[C_\epsilon]} = \frac{\int_a^b \int_{x-\epsilon}^{x+\epsilon} \sin \psi \, d\psi \, d\phi}{\int_0^{2\pi} \int_{x-\epsilon}^{x+\epsilon} \sin \psi \, d\psi \, d\phi} = \frac{b - a}{2\pi}. \tag{2.41}$$

Thus $\mathbb{P}[R_\epsilon | C_\epsilon] = \frac{b-a}{2\pi}$ holds for all possible ϵ, resulting in the limiting value

$$\mathbb{P}[R_0 | C_0] = \lim_{\epsilon \downarrow 0} \mathbb{P}[R_\epsilon | C_\epsilon] = \frac{b - a}{2\pi}. \tag{2.42}$$

$\square$

A same proposition can be found for the meridians.

Proposition 2.3.4. *Let $x \in (0, 2\pi)$ and $\epsilon \in (0, \min\{2\pi - x, x\})$. Consider $M_\epsilon = [x - \epsilon, x + \epsilon] \times [0, \pi]$ and $R_\epsilon = [x - \epsilon, x + \epsilon] \times [a, b] \subseteq M_\epsilon$, then the limiting conditional probability of R_0 given M_0 is*

$$\mathbb{P}[R_0 | M_0] = \lim_{\epsilon \downarrow 0} \mathbb{P}[R_\epsilon | M_\epsilon] = \frac{1}{2}(\cos a - \cos b) \tag{2.43}$$

Proof. The conditional probability $\mathbb{P}[R_\epsilon | M_\epsilon]$ equals

$$\mathbb{P}[R_\epsilon | M_\epsilon] = \frac{\mathbb{P}[R_\epsilon \cap M_\epsilon]}{\mathbb{P}[M_\epsilon]} = \frac{\int_{x-\epsilon}^{x+\epsilon} \int_a^b \sin \psi \, d\psi \, d\phi}{\int_{x-\epsilon}^{x+\epsilon} \int_0^\pi \sin \psi \, d\psi \, d\phi} = \frac{1}{2}(\cos a - \cos b). \tag{2.44}$$

The limit of $\epsilon \downarrow 0$ remains $\mathbb{P}[R_0 | M_0] = \frac{1}{2}(\cos a - \cos b)$. $\square$

Note that the probability in Proposition 2.3.4 has $\frac{1}{2}$ as normalization factor instead of $\frac{1}{4}$, as M_ϵ converges to only a half meridian. By symmetry the distribution on the whole meridian follows the absolute cosine function.

The conditional probabilities on $\mathfrak{C}$ and $\mathfrak{M}$ therefore can be approached by traditional conditional probabilities on $\mathcal{B}$. This gives rise to the following conjecture.

Conjecture 2.3.5. *Let $(X, \mathcal{F}, \mathbb{P})$ be a probability space. Let $F \in \mathcal{F}$ have zero measure and let $E \subseteq F$ be measurable. Let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra containing E and F with $\mathcal{G} \neq \mathcal{F}$. Let $\{F_n\}_{n \in \mathbb{N}} \subset \mathcal{G}$ be a sequence converging to F and $\{E_n\}_{n \in \mathbb{N}} \subset \mathcal{G}$ be a sequence converging to E with for all $n \in \mathbb{N}$ the sets E_n and F_n have positive measure, $E_n \subseteq F_n$ holds and the limit $\lim_{n \to \infty} \mathbb{P}[E_n | F_n]$ exists. Then*

$$\mathbb{P}[E|F] := \lim_{n \to \infty} \mathbb{P}[E_n | F_n] = \int_X \mathbb{E}[\mathbb{1}_F \,|\, \mathcal{G}] d\, \mathbb{P}'(E) \qquad (2.45)$$

where $\mathbb{P}'$ is the probability measure taking 1 on F and $\mathbb{E}[\cdot | \mathcal{G}]$ is a version of the conditional expectation on $\mathcal{G}$.

If Conjecture 2.3.5 is true, then all conditional probabilities on null sets can be computed by limiting traditional conditional probabilities. It seems to hold as for the Borel-Kolmogorov paradox when considering latitudes we can take $\mathcal{F} = \mathcal{B}$ and $\mathcal{G} = \mathfrak{C}$ and when considering meridians we can take $\mathcal{F} = \mathcal{B}$ and $\mathcal{G} = \mathfrak{M}$. It however is false as the following proposition will point out.

Proposition 2.3.6. *Conjecture 2.3.5 is false.*

Proof. In sections 2.3.2 and 2.3.3 we have seen that the parametrization using latitudes gave the uniform distribution and the parametrization using meridians gave a cosine. However, in Section 2.3.2 the set $[0, 2\pi) \times \{\frac{\pi}{2}\}$ is considered, the red horizontal line in Figure 2.1. Section 2.3.3 treats the set $\{0, \pi\} \times [0, \pi]$, like the blue vertical lines in Figure 2.1. Those sets differer, however we can find a set which has two different probabilities using a rotation.

Call f the transformation from polar to Euclidean coordinates from Equation 2.6. Let O be a rotation such that

$$(f^{-1} \circ O \circ f)\left([0, \pi] \times \left\{\frac{\pi}{2}\right\}\right) = \{\pi\} \times [0, \pi], \qquad (2.46)$$

thus where the largest half latitude is rotated to a half meridian. Recall that $\mathfrak{C}$ is the σ-algebra of latitudes. We now rotate those latitudes to get the following σ-algebra:

$$\mathfrak{C}' = \left\{F' \in \mathcal{B} \mid F' = \left(f^{-1} \circ O^{-1} \circ f\right)(F), F \in \mathfrak{C}\right\}. \qquad (2.47)$$

Now $F = \{\pi\} \times [0, \pi]$ is an element of both $\mathfrak{C}'$ and $\mathfrak{M}$. We will prove that different methods of converging to F yield to different $\mathbb{P}[E|F]$ for a measurable E. We will take the set $E = \{\pi\} \times [\frac{1}{4}\pi, \frac{3}{4}\pi]$ as our example.

First take a look at $\mathfrak{C}'$. Let $\epsilon \in \left(0, \frac{\pi}{2}\right)$ be arbitrary and consider the sets $C_\epsilon = [0, \pi] \times [\frac{\pi}{2} - \epsilon, \frac{\pi}{2} + \epsilon]$ and $R_\epsilon = [\frac{1}{4}\pi, \frac{3}{4}\pi] \times [\frac{\pi}{2} - \epsilon, \frac{\pi}{2} + \epsilon]$. Let $\{R'_{n^{-1}}\}_{n \in \mathbb{N}}$ be with $R'_{n^{-1}} = (f^{-1} \circ O^{-1} \circ f)(R_{n^{-1}})$ and let the sequence $\{C'_{n^{-1}}\}_{n \in \mathbb{N}}$ be with $C'_{n^{-1}} = (f^{-1} \circ O^{-1} \circ f)(C_{n^{-1}})$. Since $R_{n^{-1}}, C_{n^{-1}} \in \mathfrak{C}$ holds for all $n \in \mathbb{N}$, we

get $\{R'_{n^{-1}}\}_{n\in\mathbb{N}}, \{C'_{n^{-1}}\}_{n\in\mathbb{N}} \subset \mathfrak{C}'$. Furthermore, $f^{-1} \circ O^{-1} \circ f$ has determinant 1 since O is a rotation and f is almost surely a bijection on F, resulting to

$$\mathbb{P}[R'_\epsilon|C'_\epsilon] = \frac{\mathbb{P}[R'_\epsilon]}{\mathbb{P}[C'_\epsilon]} = \frac{\iint_{R'_\epsilon} \sin\psi d\psi d\phi}{\iint_{C'_\epsilon} \sin\psi d\psi d\phi} = \frac{\iint_{R_\epsilon} g(\sin\psi) \cdot 1 d\psi d\phi}{\iint_{C_\epsilon} g(\sin\psi) \cdot 1 d\psi d\phi} \tag{2.48}$$

$$= \frac{\int_{\frac{1}{4}\pi}^{\frac{3}{4}\pi} \int_{\frac{\pi}{2}-\epsilon}^{\frac{\pi}{2}+\epsilon} g(\sin\psi) \cdot 1 d\psi d\phi}{\int_0^\pi \int_{\frac{\pi}{2}-\epsilon}^{\frac{\pi}{2}+\epsilon} g(\sin\psi) \cdot 1 d\psi d\phi} = \frac{\frac{3}{4}\pi - \frac{1}{4}\pi}{\pi} = \frac{1}{2} \tag{2.49}$$

where $g(x)$ is the result of the transformation $f^{-1} \circ O^{-1} \circ f$ on the integrand. The function g is not written out explicitly, as the integral can be split up as a multiplication of integrals and the resulting integrals in the numerator en denominator with g in the integrand are equal, thus cancel out. The result is $\mathbb{P}[R'_0|C'_0] = \frac{1}{2}$, as $\mathbb{P}[R'_\epsilon|C'_\epsilon]$ is constant in ϵ. The sequence $\{R'_{n^{-1}}\}_{n\in\mathbb{N}}$ converges to E and the sequence $\{C'_{n^{-1}}\}_{n\in\mathbb{N}}$ converges to F, thus we have $\mathbb{P}[E|F] = \frac{1}{2}$ as well.

For $\mathfrak{M}$ we will use Proposition 2.3.4. Let $x = \pi$ and $\epsilon \in (0, \pi)$, then we have $M_\epsilon \in \mathfrak{M}$ and $R_\epsilon = [\pi - \epsilon, \pi + \epsilon] \times \left[\frac{1}{4}\pi, \frac{3}{4}\pi\right] \subset M_\epsilon$. Furthermore, the sequence $\{M_{n^{-1}}\}_{n\in\mathbb{N}}$ converges to F and $\{R_{n^{-1}}\}_{n\in\mathbb{N}}$ converges to E, thus Proposition 2.3.4 states that

$$\mathbb{P}[E|F] = \frac{1}{2}\left(\cos\left(\frac{1}{4}\pi\right) - \cos\left(\frac{3}{4}\pi\right)\right) = \frac{\sqrt{2}}{2}. \tag{2.50}$$

We now have both $\mathbb{P}[E|F] = \frac{1}{2}$ and $\mathbb{P}[E|F] = \frac{1}{2}\sqrt{2}$, which are clearly not equal. Conjecture 2.3.5 is therefore false. $\qquad\square$

A proposition like Proposition 2.3.6 was already proven by Hájek [H03] and Jaynes [Jay03]. Hájek argues that conditional probabilities on sets of measure zero cannot be approached by infinitesimals. This approach needs an entirely new model and the axiom of choice supports the existence of many models. Therefore there is no unique method to calculate conditional probabilities on a set of measure zero. Jaynes adds that a circle cannot be described uniquely and every method of describing a measure zero set with a different limiting approach yields a different result.

This demonstrates that when having an $F \in \mathcal{F}$ with zero measure, one cannot uniquely define $\mathbb{P}[E|F]$. A conditional expectation needs to be accompanied by a sub-σ-algebra, otherwise contradictory results can arise.

The idea underlying Conjecture 2.3.5 is however useful to guess a conditional expectation. The measure-theoretic definition of conditional expectation is only a list of properties, not a constructive definition. In the case of a $\mathfrak{C}$-conditional expectation, looking at Proposition 2.3.3 and its proof we observe that the resulting probability is $\frac{b-a}{2\pi}$, thus uniform. Furthermore, ϕ is the only variable that is integrated, thus one can guess ϕ is the only needed variable. Therefore one can suspect that

$$\mathbb{E}\left[\mathbb{1}_{[a,b]\times\left\{\frac{\pi}{2}\right\}}\Big|\mathfrak{C}\right](\phi, \psi) = \frac{1}{2\pi}\int_0^{2\pi} \mathbb{1}_{[a,b]\times\left\{\frac{\pi}{2}\right\}}(\phi', \psi)d\phi' = \frac{b-a}{2\pi}\mathbb{1}_{\left\{\frac{\pi}{2}\right\}}(\psi) \tag{2.51}$$

holds and the guess can be generalized to

$$\mathbb{E}[X|\mathfrak{C}](\phi, \psi) = \frac{1}{2\pi} \int_0^{2\pi} X(\phi', \psi)d\phi'. \tag{2.52}$$

This is in fact the same $\mathfrak{C}$-conditional expectation of X by Proposition 2.3.3. The same steps can be taken to guess $\mathbb{E}[X|\mathfrak{M}]$ using Proposition 2.3.4.

Thus we can conclude that it is wise to look at a converging sequence of traditional conditional probabilities for a potential conditional expectation when conditioning on zero sets. However, it must be noted that the resulting conditional probability is not uniquely defined, must be checked whether it complies to the actual definition of conditional probability and depends on the chosen sub-σ-algebra.

2.4 The Borel-Kolmogorov paradox explained

Following the arguments of [GHSR17] we will argue that the difference in conditional distribution is no paradox, but a misinterpretation of conditional probability.

A first explanation is quite intuitive. All meridians intersect at the north and south pole of the sphere, whereas latitudes do not intersect. Therefore the σ-algebras $\mathfrak{M}$ and $\mathfrak{C}$ are vastly different. If points are spread uniformly on the meridians, the distribution of mass on the whole sphere will not be uniform and the density of mass will be highest at the poles. This is simulated by [Wei18] and his simulations support this statement.

Let $\bar{\mathbb{P}}$ be the conditional probability measure on a latitude, defined by Equation 2.22 and let $\hat{\mathbb{P}}$ be the conditional probability measure on a meridian, defined by Equation 2.37. If the spaces $(S, \mathfrak{C}, \bar{\mathbb{P}})$ and $(S, \mathfrak{M}, \hat{\mathbb{P}})$ are isomorphic with a measurable bijection, then $\hat{\mathbb{P}}$ and $\bar{\mathbb{P}}$ must have equal distribution and the results of sections 2.3.2 and 2.3.3 must be paradoxical and contradictory. If there is a measurable bijection from a meridian to a latitude, their conditional probability distributions must agree. This is not the case here.

Theorem 2.4.1 ([GHSR17]). *Let $f: S \to S$ be a measurable bijection with measurable inverse. There is no Boolean algebra isomorphism $h_f: \mathfrak{C} \to \mathfrak{M}$ that is determined by f.*

Proof. The proof can be found in [GHSR17], but we will repeat it here. Suppose such isomorphism h_f exists. All latitudes C in $\mathfrak{C}$ are the only atoms of $\mathfrak{C}$, therefore $h_f(C)$ are the only atoms of $\mathfrak{M}$ as well, making $h_f(C)$ meridians. We will now prove that $h_f(C)$ cannot be a meridian.

Let $m_0 = \{(0,0), (0,\pi)\}$ be the set of north and south poles. Let $c_0 \in \mathfrak{C}$ be such that $h_f(c_0) = m_0$. Then c_0 consists of two elements as well, thus we can choose latitude C not on the north or south poles such that $C \cap c_0 = \emptyset$. Note that $h_f(C)$ is a meridian, thus $m_0 \subset h_f(C)$ as all meridians pass through the north and south poles. Since h_f is a Boolean algebra isomorphism, we have $h_f(C \cap c_0) = h_f(C) \cap h_f(c_0)$. This all can be combined in the following line:

$$\emptyset = h_f(\emptyset) = h_f(C \cap c_0) = h_f(C) \cap h_f(c_0) = h_f(C) \cap m_0 = m_0. \tag{2.53}$$

This is clearly a contradiction.

16

Let C be the north or south pole of the sphere. Since h_f is a bijection, $h_f(C)$ is only a single point as well. Thus $h_f(C)$ cannot be an atom of $\mathfrak{M}$. Therefore there is a contradiction here as well.

Thus no single atom of $\mathfrak{C}$ result into an atom of $\mathfrak{M}$ by h_f and therefore function h_f cannot exist. $\qquad\square$

Therefore every measurable bijection between $(S, \mathfrak{C})$ and $(S, \mathfrak{M})$ has no Boolean algebra isomorphism $h_f \colon \mathfrak{C} \to \mathfrak{M}$ such that latitudes in $\mathfrak{C}$ are mapped to meridians on $\mathfrak{M}$. This theorem holds on all subalgebras of $\mathfrak{C}$ and $\mathfrak{M}$ as well, as is proven by [GHSR17].

Now it should be clear why the conditional distribution on the latitudes and meridians of the sphere differ, since we are dealing with two entirely different structured spaces. Therefore, the question 'why is the conditional distribution on the latitudes different from the conditional distribution on the meridians' is easy to answer; difference in spaces. To verify that two different methods of conditioning on the same set yield to the correct conditional probability, one must ask the following question: 'if one models a distribution on the sample space using conditional expectation on a sub-σ-algebra, does the resulting conditional probability distribution extend to the original distribution on the whole sample space?' As earlier mentioned, Weisstein [Wei18] demonstrated that the answer to that question is yes. The uniform distribution on latitudes and the cosine on meridians both extend to the uniform distribution on the whole sphere. Thus when dealing with different conditional distributions on measure zero sets, one should rather ask a question like the last one, as the answer to that question must always be yes. Further reading and a more in-depth analysis can be found in [GHSR17].

2.5 Conclusion

After analysing the Borel-Kolmogorov paradox, the first and most important conclusion we can take is that when conditioning, especially on a set F of measure zero, one needs to give the accompanying sub-σ-algebra of the event that is conditioned on. Otherwise the Borel-Kolmogorov phenomenon appears, where one can give different sub-σ-algebras containing F such that the resulting conditional probability distribution is different.

Secondly, we must accept that conditional probability on sets of measure zero are not uniquely defined. That uniqueness only appears when conditioning on sets of positive measure, as then the classical rule $\mathbb{P}[E|F] = \frac{\mathbb{P}[E \cap F]}{\mathbb{P}[F]}$ gives a version of the conditional probability. If a sub-σ-algebra can be used to model the entire sample space, for example the sub-σ-algebra of latitudes with the sphere as sample space, then the conditional probability must be able to extend to the original probability distribution. For example providing the uniform distribution on all latitudes gives back the uniform distribution on the sphere, as well as applying the cosine distribution on the meridians results into the uniform distribution on the sphere.

Thirdly, a conditional distribution on a zero set must be calculated using the measure-theoretic conditional expectation. Conjecture 2.3.5 can help with finding the correct conditional expectation, however Proposition 2.3.6 proves that the resulting conditional probability does not need to be unique. Therefore

Conjecture 2.3.5 cannot be used to calculate a conditional probability and then call it a day, you need to check whether the distribution you found actually satisfies all conditions of conditional probability.

At last, as Kolmogorov [Kol33] already stated in his work, there is nothing paradoxical going on in the Borel-Kolmogorov paradox. The space of the sphere with the latitudes as σ-algebra is not homeomorphic with the space of the sphere and the meridians as σ-algebra. The not-uniqueness of the conditional probability on a zero set is thus as expected and we must give the sub-σ-algebra when conditioning on sets of measure zero.

Chapter 3

Safe probability

Take a look at the following game of dice, devised by Grünwald [Gr 13]. A player and game master are present and the game master casts a fair die. He peeks at the value of the die and must either state that the value is 4 or lower or state that the value is 3 or higher. The game master is not allowed to lie. Upon hearing the game master's advice, what is the probability the die has landed on a 3?

Suppose an alternative game is considered, where the game master can only say that the value is either from 1 to and including 4 or the value is 5 or 6. The question of the probability of the die having 3 becomes easy to answer. When 5 or 6 is revealed, the probability is 0 and when 1 to 4 is revealed, traditional conditional probability gives us that the probability is 0.25.

However, when either '4 or lower' or '3 or higher' is stated, the game master can choose his statement when the die lands on a 3 or 4. His strategy on choosing his statement heavily influences the probability of the die landing on 3 upon hearing the game master's statement, making that probability not unique.

What can we say about the probability of the die landing on 3? Traditional conditional probability will not suffice here. We want to condition on sets with positive measure and the versions of conditional expectation then almost surely coincides with traditional conditional probability, thus using measure-theoretic conditional probability will not give more insight than conditional probability.

If we loose our notion of conditional probability, we can resort to *safe probability*, introduced by Grünwald in 2018 [Gr 18b]. Suppose there are two random variables, U and V, and we want to use the distribution $U|V$. Suppose however that there is no method of finding this distribution, only you that you know that it is a member of a model of distributions $\mathcal{P}^*$. One can then create a new conditional distribution, not necessarily one from our model nor necessarily reflecting the true distribution at all, which gives a probability of U given V that on average performs equal for all distributions in our model $\mathcal{P}^*$ in predicting U given V. For example, one notion of safe probability is that we have a distribution $\tilde{\mathbb{P}}$ such that $\mathbb{E}_{\mathbb{P}}[U] = \mathbb{E}_{\mathbb{P}}[\mathbb{E}_{\tilde{\mathbb{P}}}[U|V]]$ holds for all $\mathbb{P} \in \mathcal{P}^*$. Under this $\tilde{\mathbb{P}}$ the expectation $\mathbb{E}_{\tilde{\mathbb{P}}}[U|V]$ is an unbiased estimator of U for all distributions $\mathbb{P} \in \mathcal{P}^*$. Thus no matter which strategy the game master has, distribution $\tilde{\mathbb{P}}$ allows the player to estimate U on average correctly.

There are more notions on safety, further restricting the freedom $\tilde{\mathbb{P}}$ has. We will first introduce some notions of safe probability and how all notions

are related to each other. The example of the dice game is then continued to illustrate how safe probability can be applied.

3.1 Definitions

Before we will define safe probability, we need to introduce some notation.

Definition 3.1.1 (Range and support)**.** Let $S \colon \mathcal{X} \to \mathcal{Y}$ be a random variable. The *range* of S is

$$\text{range}(S) = \{s \in \mathcal{Y} \mid s = S(x), x \in \mathcal{X}\}. \tag{3.1}$$

Let $\mathbb{P}$ be a probability distribution on $\mathcal{X}$. The *support* of S under $\mathbb{P}$ is

$$\text{supp}_{\mathbb{P}}(S) = \{s \in \mathcal{Y} \mid \mathbb{P}[S = s] > 0\}. \tag{3.2}$$

Let U and U' be two random variables and suppose there is a function f such that $f(U) = U'$ holds, then we will write $U \overset{f}{\rightsquigarrow} U'$. If a function f exists such that $U \overset{f}{\rightsquigarrow} V$, we will write $U \rightsquigarrow V$. Random variable V is then called a *coarsening* of U or equivalently we state that U determines V.

Let $\mathcal{P}^*$ be a set of probability measures, then $\mathcal{P}^*$ is called a *model*.

Lastly, if we keep U and V as random variables, the distribution of U under $\mathbb{P}$ is written as $\mathbb{P}[U]$. The distribution of U given V under $\mathbb{P}$ is written as $\mathbb{P}[U|V]$.

We are now able to define conditional probability. The following definitions and propositions are all taken from [Gr 18b].

Definition 3.1.2 (Safe probabilities)**.** Let Ω be a sample space and let $\mathcal{P}^*$ be a set of distributions on Ω. Let U be a real-valued random variable with finite expectation and let V be a random variable on Ω. Let $\tilde{\mathbb{P}}$ be a probability distribution on Ω. Then $\tilde{\mathbb{P}}$ is *safe* with respect to $\mathcal{P}^*$ for

- $\langle U \rangle | \langle V \rangle$ if $\mathbb{E}_{\mathbb{P}}[U] = \mathbb{E}_{\mathbb{P}}[\mathbb{E}_{\tilde{\mathbb{P}}}[U|V]]$ holds for all $\mathbb{P} \in \mathcal{P}^*$. This $\tilde{\mathbb{P}}$ is called *unbiased* for $U|V$.

- $\langle U \rangle | [V]$ if $\mathbb{E}_{\mathbb{P}}[U] = \mathbb{E}_{\tilde{\mathbb{P}}}[U|V = v]$ holds for all $v \in \text{supp}_{\tilde{\mathbb{P}}}(V)$. This $\tilde{\mathbb{P}}$ is called *marginally valid* for $U|V$.

Since in almost all situations we only consider a single model $\mathcal{P}^*$, we will from now on omit the statement 'with respect to $\mathcal{P}^*$'.

Definition 3.1.3 (Stronger safety)**.** Let $\Omega, \mathcal{P}^*, U, V$ and $\tilde{\mathbb{P}}$ as above. Then $\tilde{\mathbb{P}}$ is *safe* for

- $U|\langle V \rangle$ if for all real-valued random variables U' on Ω with $U \rightsquigarrow U'$ distribution $\tilde{\mathbb{P}}$ is safe for $\langle U' \rangle | \langle V \rangle$.

- $U|[V]$ if for all real-valued random variables U' on Ω with $U \rightsquigarrow U'$ distribution $\tilde{\mathbb{P}}$ is safe for $\langle U' \rangle | [V]$.

Both definitions introduce a lot of new notation. In practice, $\langle V \rangle$ implies that $\mathbb{E}_{\tilde{\mathbb{P}}}[U|V]$ must be an unbiased estimator for U for all $\mathbb{P} \in \mathcal{P}^*$.
When $[V]$ is considered, the expected value of $U|V = v$ under $\tilde{\mathbb{P}}$ must equal the expected value of U under $\mathbb{P}$ for all $v \in \text{supp}_{\mathbb{P}}(V)$ and $\mathbb{P} \in \mathcal{P}^*$. In other

words, the outcome of V will not influence the conditional expectation value of U under $\tilde{\mathbb{P}}$ and this expectation will equal the expectation of U under all $\mathbb{P} \in \mathcal{P}^*$. Note that when $\mathbb{E}_{\mathbb{P}}[U]$ does not remain constant for all $\mathbb{P} \in \mathcal{P}^*$, safety for $[V]$ is immediately impossible.

Lastly, the difference between $\langle U \rangle$ and U is that for $\langle U \rangle$ the requirement for safety does only have to be met for U. When U is considered, the requirement must be met for all coarsenings U' of U.

When we say that $\tilde{\mathbb{P}}$ is safe for $U|[V]$, we actually mean that 'for all $\mathbb{P} \in \mathcal{P}^*$ we have $\mathrm{supp}_{\mathbb{P}}(V) \subseteq \mathrm{supp}_{\tilde{\mathbb{P}}}(V)$, $\mathbb{E}_{\mathbb{P}}[U]$ is well-defined and finite, $\mathbb{E}_{\tilde{\mathbb{P}}}[U|V = v]$ is well-defined for all $v \in \mathrm{supp}_{\tilde{\mathbb{P}}}(V)$, both expectations are equal and this holds for all U' with $U \leadsto U'$'. This statement is quite lengthy, therefore we will abbreviate it by stating that $\tilde{\mathbb{P}}$ is safe for $U|[V]$. Such abbreviations are performed for other notions of safety as well.

Definition 3.1.3 is not easy to work with as constructing a $\tilde{\mathbb{P}}$ that is safe for $U|[V]$ requires checking $\tilde{\mathbb{P}}$ against all coarsenings U' of U. The following proposition from [Gr 18b] gives us tools to actually create a safe $\tilde{\mathbb{P}}$ for $U|\langle V \rangle$, $U|[V]$ and a few extra notions.

Proposition 3.1.4. *Let Ω, $\mathcal{P}^*$, U, V and $\tilde{\mathbb{P}}$ be as above. Then*

1. *$\tilde{\mathbb{P}}$ is safe for $U|\langle V \rangle$ if and only if for all $\mathbb{P} \in \mathcal{P}^*$ there is a distribution $\mathbb{P}'$ on Ω with $\mathbb{P}'[U = u, V = v] = \tilde{\mathbb{P}}[U = u|V = v]\,\mathbb{P}[V = v]$ for all $(u, v) \in \mathrm{range}((U, V))$ such that $\mathbb{P}'[U] = \mathbb{P}[U]$.*

2. *$\tilde{\mathbb{P}}$ is safe for $\langle U \rangle|V$ if and only if*

$$\mathbb{E}_{\mathbb{P}}[U|V] = \mathbb{E}_{\tilde{\mathbb{P}}}[U|V] \tag{3.3}$$

   holds with probability 1 for all $\mathbb{P} \in \mathcal{P}^*$. This $\tilde{\mathbb{P}}$ is called squared error-optimal *for $U|V$.*

3. *$\tilde{\mathbb{P}}$ is safe for $U|V$ if and only if*

$$\mathbb{P}[U|V] = \tilde{\mathbb{P}}[U|V] \tag{3.4}$$

   holds with probability 1 for all $\mathbb{P} \in \mathcal{P}^*$. This $\tilde{\mathbb{P}}$ is called valid *for $U|V$.*

4. *$\tilde{\mathbb{P}}$ is safe for $U|[V]$ if and only if $\mathbb{P}[U] = \tilde{\mathbb{P}}[U|V = v]$ for all $\mathbb{P} \in \mathcal{P}^*$ and $v \in \mathrm{supp}_{\tilde{\mathbb{P}}}(V)$.*

Proof. The proof is in [Gr 16], section A.2. $\square$

It is easy to see that safety for $\langle U \rangle|[V]$ implies safety for $\langle U \rangle|\langle V \rangle$. If we have $\mathbb{E}_{\tilde{\mathbb{P}}}[U|V = v] = \mathbb{E}_{\mathbb{P}}[U]$ for all $v \in \mathrm{supp}_{\tilde{\mathbb{P}}}(V)$, then

$$\mathbb{E}_{\mathbb{P}}[\mathbb{E}_{\tilde{\mathbb{P}}}[U|V]] = \mathbb{E}_{\mathbb{P}}[\mathbb{E}_{\mathbb{P}}[U]] = \mathbb{E}_{\mathbb{P}}[U] \tag{3.5}$$

must hold. All implications are stated in the following proposition.

Note that safety for $U|V$ will almost never occur. When a $\tilde{\mathbb{P}}$ is safe for $U|V$, then all probability measures in $\mathcal{P}^*$ have an almost surely equal distribution on $U|V$. In that case we should not bother ourselves with safe probability and just use one distribution from $\mathcal{P}^*$ as our probability. Furthermore, if $\mathcal{P}^* = \{\mathbb{P}\}$ consists of only a single probability measure, then $\mathbb{P}$ is automatically safe for $U|V$. Therefore it is wise to make the model $\mathcal{P}^*$ large enough when considering safe probabilities, otherwise the safe distribution is prone to overfitting.
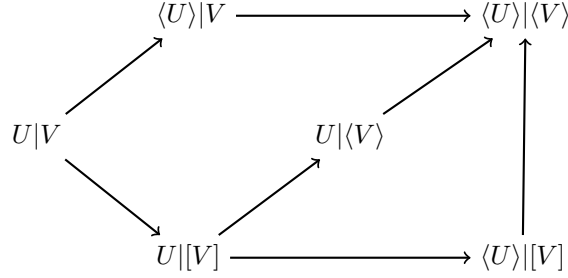
Figure 3.1: A diagram of implications between safe probabilities. If $\tilde{\mathbb{P}}$ is safe for X, then it is safe for all statements that can be reached from X as well.

Proposition 3.1.5. *Let U, V and $\tilde{\mathbb{P}}$ be as above.*

1. *If $\tilde{\mathbb{P}}$ is safe for $U|V$, it is safe for $\langle U\rangle|V$ and $U|[V]$.*

2. *If $\tilde{\mathbb{P}}$ is safe for $U|[V]$, it is safe for $U|\langle V\rangle$ and $\langle U\rangle|[V]$.*

3. *If $\tilde{\mathbb{P}}$ is safe for $\langle U\rangle|V$, $U|\langle V\rangle$ or $\langle U\rangle|[V]$, it is safe for $\langle U\rangle|\langle V\rangle$.*

All relations are visualised in Figure 3.1.

Proof. All implications are found and proven in [Gr 18b]. $\qquad\square$

3.2 The dice game

Recall the dice game from the introduction of this chapter, introduced by [Gr 13]. The game master casts a fair die and observes a value. This value is either in the set $\{1,2,3,4\}$ or in the set $\{3,4,5,6\}$. The game master must reveal one of the two sets to the player and cannot lie. The player must then guess the probability that the die has landed on 3.

It is insufficient to take $\Omega = \{1,2,3,4,5,6\}$ as our sample space. Let U be a random variable denoting the die's value and let V be the statement of the game master, then $\mathbb{P}[U = 3|V = \{1,2,3,4\}]$ cannot be calculated. When the die has rolled to a 3, the game master can choose between $\{1,2,3,4\}$ and $\{3,4,5,6\}$ and this must be taken into account. The smallest σ-algebra containing both $\{1,2,3,4\}$ and $\{3,4,5,6\}$ is

$$\mathcal{G} = \sigma(\{1,2\},\{3,4\},\{5,6\}), \qquad (3.6)$$

thus $\{3\}$ is no member of $\mathcal{G}$ as $\mathcal{G}$ cannot have sets of odd size. We can add $\{3\}$ to $\mathcal{G}$, however we still do need the probability of $\{\{1,2,3,4\},\{3,4,5,6\}\}$. Therefore we cannot condition in Ω to compute $U = 3$ given $V = \{1,2,3,4\}$.

Therefore we need to extend our sample space. We will use the extension introduced in [Gr 13]. Take $\mathcal{X} = \{1,2,3,4,5,6\}$ and $\mathcal{Y} = \{\{1,2,3,4\},\{3,4,5,6\}\}$ and let our sample space be $\Omega = \mathcal{X} \times \mathcal{Y}$. Let U be an $\mathcal{X}$-valued random variable and let V be a $\mathcal{Y}$-valued random variable. In this space conditioning on $\{1,2,3,4\}$ is valid, as $\mathcal{G} = \sigma(\{3\} \times Y)$ is a strict sub-σ-algebra of 2^Ω for all subsets $Y \subseteq \mathcal{Y}$.

3.2.1 Using traditional conditional probability

First we will try using traditional conditional probability. Let $\mathbb{P}$ be a probability measure with the uniform distribution on U. To model the strategy of the game master, we need $p, q \in [0, 1]$ such that

$$\mathbb{P}[V = \{3, 4, 5, 6\} \mid U = 3] = p, \qquad \mathbb{P}[V = \{3, 4, 5, 6\} \mid U = 4] = q. \qquad (3.7)$$

We can now for example calculate the probability of the die landing on 6 after the game master reveals $\{3, 4, 5, 6\}$. This calculation can be found in equation 2 of [Gr 13] and it states that

$$\mathbb{P}[U = 6 \mid V = \{3, 4, 5, 6\}] = \frac{1}{p + q + 2}. \qquad (3.8)$$

Thus the conditional probability of rolling 6 ranges from $\frac{1}{4}$ to $\frac{1}{2}$ depending on the game master's strategy. This should not suffice as answer, as the player does not know the game master's strategy.

3.2.2 Using safe probabilities

Can we create a distribution $\tilde{\mathbb{P}}$ such that the probability of $U = 3$ given V is independent of p and q and behaves on average like the dice? The answer is yes and we need to use safe probability. Consider as model

$$\mathcal{P}^* = \left\{ \mathbb{P} \middle| \forall u \in \mathcal{X} : \mathbb{P}[U = u] = \frac{1}{6}, \forall v \in \mathcal{Y} : \mathbb{P}[U \in v \mid V = v] = 1 \right\}. \qquad (3.9)$$

We do not know the strategy of the game master when 3 or 4 is rolled. However, we do know that the die is fair and that the game master is not able to lie, thus we put this information in our model.

For shorthand purposes, we write $y_1 = \{1, 2, 3, 4\}$ and $y_2 = \{3, 4, 5, 6\}$ such that $\mathcal{Y} = \{y_1, y_2\}$ from now on.

Proposition 3.2.1. *Let $\mathcal{X}$, $\mathcal{Y}$, Ω, U, V and $\mathcal{P}^*$ be as before. Let*

$$\tilde{\mathcal{P}} = \left\{ \mathbb{P} \middle| \begin{array}{l} \mathbb{P}[U = 1 \mid V = y_1] = \mathbb{P}[U = 2 \mid V = y_1], \\ \mathbb{P}[U = 3 \mid V = y_1] = \frac{1}{2} - 5\,\mathbb{P}[U = 1 \mid V = y_1], \\ \mathbb{P}[U = 5 \mid V = y_2] = \mathbb{P}[U = 6 \mid V = y_2], \\ \mathbb{P}[U = 3 \mid V = y_2] = 3\,\mathbb{P}[U = 5 \mid V = y_2] + \frac{1}{2}, \\ \mathbb{P}[U = 1 \mid V = y_1], \mathbb{P}[U = 5 \mid V = y_1] \in \left[0, \frac{1}{10}\right] \end{array} \right\} \qquad (3.10)$$

be a set of probability distributions on Ω. Let $\tilde{\mathbb{P}}$ be a probability distribution on Ω with $\tilde{\mathbb{P}}[U = 1 | V = y_1] = \tilde{\mathbb{P}}[U = 2 | V = y_1]$, $\tilde{\mathbb{P}}[U = 5 | V = y_2] = \tilde{\mathbb{P}}[U = 6 | V = y_2]$ and $\tilde{\mathbb{P}}[U \in y | V = y] = 1$ for all $y \in \mathcal{Y}$. The following are equivalent:

1. *$\tilde{\mathbb{P}}$ is a member of $\tilde{\mathcal{P}}$.*

2. *$\tilde{\mathbb{P}}$ is safe for $\langle U \rangle | [V]$.*

3. *$\tilde{\mathbb{P}}$ is safe for $\langle U \rangle | \langle V \rangle$.*

Distribution $\tilde{\mathbb{P}}$ is not safe for $U | [V]$.

Proof. Here we will abbreviate $\tilde{\mathbb{P}}[U = u | V = v]$ to $\tilde{\mathbb{P}}[u|v]$ for all $(u, v) \in \Omega$. We will start with the implication from 1 to 2. Take an arbitrary $\tilde{\mathbb{P}} \in \tilde{\mathcal{P}}$. We then have

$$\mathbb{E}_{\mathbb{P}}[U] = \sum_{u=1}^{6} u\, \mathbb{P}[U = u] = \frac{7}{2}. \tag{3.11}$$

Consider y_1 first. Writing $\tilde{\mathbb{P}}[u|y_1]$ in terms of $\tilde{\mathbb{P}}[1|y_1]$ for all $u \in \{2, 3, 4\}$ gives

$$\mathbb{E}_{\tilde{\mathbb{P}}}[U|V = y_1] = \tilde{\mathbb{P}}[1|y_1] + 2\tilde{\mathbb{P}}[1|y_1] + 3\left(\frac{1}{2} - 5\tilde{\mathbb{P}}[1|y_1]\right) + 4\left(3\tilde{\mathbb{P}}[1|y_1] + \frac{1}{2}\right) \tag{3.12}$$

$$= \frac{7}{2} + 15\tilde{\mathbb{P}}[1|y_1] - 15\tilde{\mathbb{P}}[1|y_1] = \frac{7}{2}. \tag{3.13}$$

The same calculation holds for $V = y_2$ as well. Therefore we have the equality $\mathbb{E}_{\mathbb{P}}[U] = \mathbb{E}_{\tilde{\mathbb{P}}}[U|V = v]$ for all $\mathbb{P} \in \mathcal{P}^*$ and $v \in \mathrm{supp}_{\tilde{\mathbb{P}}}(V)$, letting $\tilde{\mathbb{P}}$ fulfil the requirement of safety for $\langle U \rangle | [V]$.

The implication from 2 to 3 follows from Proposition 3.1.5.

Lastly, we will prove the implication from 3 to 1. Let $\tilde{\mathbb{P}}$ be safe for $\langle U \rangle | \langle V \rangle$. Let $\mathbb{P} \in \mathcal{P}^*$ be arbitrary and take $p, q \in [0, 1]$ with $\mathbb{P}[V = y_2 | U = 3] = p$ and $\mathbb{P}[V = y_2 | U = 4] = q$.

Definition 3.1.2 states that we need $\mathbb{E}_{\mathbb{P}}[\mathbb{E}_{\tilde{\mathbb{P}}}[U|V]] = \mathbb{E}_{\mathbb{P}}[U]$. The right-hand side is already computed as $\mathbb{E}_{\mathbb{P}}[U] = \frac{7}{2}$. For $\mathbb{E}_{\mathbb{P}}[\mathbb{E}_{\tilde{\mathbb{P}}}[U|V]]$, we first focus on $V = y_1$. Conditioning on U yields

$$\mathbb{P}[V = y_1] = \sum_{u=1}^{6} \mathbb{P}[V = y_1 | U = u]\, \mathbb{P}[U = u] = \frac{1}{6}\sum_{u=1}^{6} \mathbb{P}[V = y_1 | U = u]. \tag{3.14}$$

The game master cannot lie, thus $\mathbb{P}[V = y_1 | U = 1] = \mathbb{P}[V = y_1 | U = 2] = 1$ and $\mathbb{P}[V = y_1 | U = 5] = \mathbb{P}[V = y_1 | U = 6] = 0$ immediately hold. We further have $\mathbb{P}[V = y_2 | U = 3] = p$ and $\mathbb{P}[V = y_2 | U = 4] = q$, thus the value of $\mathbb{P}[V = y_1]$ is

$$\mathbb{P}[V = y_1] = \frac{1}{6}\sum_{u=1}^{6} \mathbb{P}[V = y_1 | U = u] = \frac{2}{3} - \frac{p+q}{6}. \tag{3.15}$$

Since $\mathcal{Y}$ only has two elements, $\mathbb{P}[V = y_2] = \frac{1}{3} + \frac{p+q}{6}$ follows. We can now write out $\mathbb{E}_{\mathbb{P}}[\mathbb{E}_{\tilde{\mathbb{P}}}[U|V]]$ to

$$\mathbb{E}_{\mathbb{P}}[\mathbb{E}_{\tilde{\mathbb{P}}}[U|V]] = \mathbb{E}_{\tilde{\mathbb{P}}}[U|V = y_1]\, \mathbb{P}[V = y_1] + \mathbb{E}_{\tilde{\mathbb{P}}}[U|V = y_2]\, \mathbb{P}[V = y_2] \tag{3.16}$$

$$= \left(\tilde{\mathbb{P}}[1|y_1] + 2\tilde{\mathbb{P}}[2|y_1] + 3\tilde{\mathbb{P}}[3|y_1] + 4\tilde{\mathbb{P}}[4|y_1]\right)\left(\frac{2}{3} - \frac{p+q}{6}\right)$$

$$+ \left(3\tilde{\mathbb{P}}[3|y_2] + 4\tilde{\mathbb{P}}[4|y_2] + 5\tilde{\mathbb{P}}[5|y_2] + 6\tilde{\mathbb{P}}[6|y_2]\right)\left(\frac{1}{3} + \frac{p+q}{6}\right). \tag{3.17}$$

This collapes to

$$\mathbb{E}_{\mathbb{P}}[\mathbb{E}_{\tilde{\mathbb{P}}}[U|V]] = \frac{p+q}{6}\left(\sum_{u=3}^{6} u\tilde{\mathbb{P}}[u|y_2] - \sum_{u=1}^{4} u\tilde{\mathbb{P}}[u|y_1]\right) + \sum_{u=1}^{4} \frac{2u}{3}\tilde{\mathbb{P}}[u|y_1]$$

$$+ \sum_{u=3}^{6} \frac{u}{3}\tilde{\mathbb{P}}[u|y_2]. \tag{3.18}$$

As we have $\mathbb{E}_{\mathbb{P}}[U] = \frac{7}{2}$, we need to get rid of the dependence on $p + q$. This results into the condition

$$\sum_{u=3}^{6} u\tilde{\mathbb{P}}[u|y_2] = \sum_{u=1}^{4} u\tilde{\mathbb{P}}[u|y_1]. \tag{3.19}$$

When this condition is satisfied, we need

$$\sum_{u=1}^{4} \frac{2u}{3}\tilde{\mathbb{P}}[u|y_1] + \sum_{u=3}^{6} \frac{u}{3}\tilde{\mathbb{P}}[u|y_2] = \frac{7}{2} \tag{3.20}$$

for $\tilde{\mathbb{P}}$ to be safe for $\langle U \rangle | \langle V \rangle$. We impose one further assumption on our safe probability to reduce the amount of possible safe probabilities. Suppose the die shows 5 or 6, the game master has to tell y_2. Thus it is logical to assume that $\tilde{\mathbb{P}}[5|y_2] = \tilde{\mathbb{P}}[6|y_2]$ and $\tilde{\mathbb{P}}[1|y_1] = \tilde{\mathbb{P}}[2|y_1]$, as when an y_2 is revealed there is no reason why 5 is preferred more than 6. As the game master cannot lie,

$$\tilde{\mathbb{P}}[5|y_1] = \tilde{\mathbb{P}}[6|y_1] = \tilde{\mathbb{P}}[1|y_2] = \tilde{\mathbb{P}}[2|y_2] = 0 \tag{3.21}$$

can be assumed. Luckily, these assumptions are made mandatory by the proposition. This all results into the following system of equations:

$$\begin{pmatrix} \frac{2}{3} & \frac{4}{3} & 2 & \frac{8}{3} & 1 & \frac{4}{3} & \frac{5}{3} & 2 \\ -1 & -2 & -3 & -4 & 3 & 4 & 5 & 6 \\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \tilde{\mathbb{P}}[1|y_1] \\ \tilde{\mathbb{P}}[2|y_1] \\ \tilde{\mathbb{P}}[3|y_1] \\ \tilde{\mathbb{P}}[4|y_1] \\ \tilde{\mathbb{P}}[3|y_2] \\ \tilde{\mathbb{P}}[4|y_2] \\ \tilde{\mathbb{P}}[5|y_2] \\ \tilde{\mathbb{P}}[6|y_2] \end{pmatrix} = \begin{pmatrix} \frac{7}{2} \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}. \tag{3.22}$$

The set $\tilde{\mathcal{P}}$ is the set of all probability distributions that are solution to this system. As $\tilde{\mathbb{P}} \in \tilde{\mathcal{P}}$ needs to be a valid probability distribution, we need to have $\tilde{\mathbb{P}}[u|v] \in [0,1]$ for all $(u,v) \in \Omega$. Solving (3.22) gives

$$\tilde{\mathbb{P}}[3|y_1] = 3\tilde{\mathbb{P}}[1|y_1] + \frac{1}{2}, \tag{3.23}$$

$$\tilde{\mathbb{P}}[4|y_1] = 5\tilde{\mathbb{P}}[1|y_1] - \frac{1}{2}, \tag{3.24}$$

thus $\tilde{\mathbb{P}}[3|y_1] \in [0,1]$ holds when $\tilde{\mathbb{P}}[1|y_1] \in \left[0, \frac{1}{6}\right]$ and $\tilde{\mathbb{P}}[4|y_1] \in [0,1]$ holds when $\tilde{\mathbb{P}}[1|y_1] \in \left[0, \frac{1}{10}\right]$. Without loss of generality $\tilde{\mathbb{P}}[5|y_2] \in \left[0, \frac{1}{10}\right]$ must hold as well. This proves the last requirement $\tilde{\mathbb{P}}[1|y_2], \tilde{\mathbb{P}}[5|y_2] \in \left[0, \frac{1}{10}\right]$ from $\tilde{\mathcal{P}}$ and completes our proof that if a distribution $\tilde{\mathbb{P}}$ is safe for $\langle U \rangle | \langle V \rangle$ with the requirements in the proposition, it is a member of $\tilde{\mathcal{P}}$.

Lastly we turn to safety for $U | [V]$. Pick probability measure $\tilde{\mathbb{P}}$ on Ω arbitrarily with $\tilde{\mathbb{P}}[1|y_1] = \tilde{\mathbb{P}}[2|y_1]$, $\tilde{\mathbb{P}}[5|y_2] = \tilde{\mathbb{P}}[6|y_2]$ and $\tilde{\mathbb{P}}[U \in y | V = y] = 1$ for all $y \in \mathcal{Y}$. Proposition 3.1.4 implies $\tilde{\mathbb{P}}[u|y_1] = \frac{1}{6}$ for all $u \in \mathcal{X}$, which is a clear contradiction to $\tilde{\mathbb{P}}[U \in y | V = y] = 1$. Thus there is no probability measure $\tilde{\mathbb{P}}$ with the requirements in the proposition that is safe for $U | [V]$. $\qquad \square$

The original question of the dice game did not concern an overall safe probability distribution for the whole die, but rather the probability of rolling 3. Proposition 4.2.1 states that a distribution $\tilde{\mathbb{P}}$ is safe for $\langle \mathbb{1}_{\{U=3\}} \rangle | \langle V \rangle$ as well as $\mathbb{1}_{\{U=3\}} | [V]$ if and only if $\tilde{\mathbb{P}}[U = 3 | V = v] = \frac{1}{6}$ holds for all $v \in \mathcal{Y}$. It is therefore safe to ignore the statement of the game master when guessing the probability of the die rolling to a 3 and safe to state that the probability of rolling to a 3 remains $\frac{1}{6}$.

A safe distribution for a statement of $\langle \mathbb{1}_{\{U=1\}} \rangle | \langle V \rangle$ cannot exist, as the following proposition will prove.

Proposition 3.2.2. *There is no probability distribution $\tilde{\mathbb{P}}$ on Ω that is safe for $\langle \mathbb{1}_{\{U=u\}} \rangle | \langle V \rangle$ with $u \in \{1, 2, 5, 6\}$ when $\tilde{\mathbb{P}}[U \in v | V = v] = 1$ is required for all $v \in \mathcal{Y}$.*

Proof. Pick $u = 1$ and suppose $\tilde{\mathbb{P}}$ is safe for $\langle \mathbb{1}_{\{U=1\}} \rangle | \langle V \rangle$. Let $p, q \in [0, 1]$ be as in the proof of Proposition 3.2.1. As $V = y_1$ is the only possible option when rolling 1, we have

$$\mathbb{E}_{\mathbb{P}}[\mathbb{E}_{\tilde{\mathbb{P}}}[\mathbb{1}_{\{U=1\}} | V]] = \tilde{\mathbb{P}}[U = 1 | V = y_1] \, \mathbb{P}[V = y_1] \qquad (3.25)$$

$$= \tilde{\mathbb{P}}[U = 1 | V = y_1] \left(\frac{2}{3} - \frac{p+q}{6} \right). \qquad (3.26)$$

We need $\mathbb{E}_{\mathbb{P}}[\mathbb{E}_{\tilde{\mathbb{P}}}[\mathbb{1}_{\{U=1\}} | V]] = \mathbb{P}[U = 1] = \frac{1}{6}$, but there is no possible value for $\tilde{\mathbb{P}}[U = 1 | V = y_1]$ such that

$$\tilde{\mathbb{P}}[U = 1 | V = y_1] \left(\frac{2}{3} - \frac{p+q}{6} \right) = \frac{1}{6} \qquad (3.27)$$

holds for all $p, q \in [0, 1]$. This only happens when $p + q = 1$, so there is a contradiction. Therefore $\tilde{\mathbb{P}}$ cannot be safe for $\langle \mathbb{1}_{\{U=1\}} \rangle | \langle V \rangle$.

This proof applies to all $v \in \{1, 2, 5, 6\}$, where in the case of $v \in \{5, 6\}$ we need to condition on $V = y_2$ instead of $V = y_1$. $\qquad \square$

3.2.3 Conclusion

To conclude, in the example of the dice game, we cannot compute a unique conditional distribution for $U | V$. However, we were able to create a set of probability distributions that are all unbiased for $U | V$. When only considering the probability of the die rolling to a specific value, we can create a safe distribution for the values 3 or 4, which will be done in Proposition 4.2.1. A safe distribution for the values 1, 2, 5 and 6 does not exist.

Chapter 4

Discrete conditional paradoxes

Consider the following three problems:

The Monty Hall problem   There are three doors. Two doors have a goat behind and one door a car. The player initially chooses a door, say he chooses door a. The game master then opens either door b or door c, however he never opens the door with the car. After opening a door, the player is asked if he should switch. He ultimately wins the contents behind the door he has chosen. Should the player switch to the other door?

The boy or girl problem   Suppose you encounter a stranger in a bar and you two get in conversation. Suddenly, the stranger states that he has two children and at least one of them is a girl. What is the probability he has two daughters?

The dice game   This game is already introduced in Section 3.2, but here is a repeat. A die is rolled. Only the game master is able to observe the die's value. He then tells you whether the value of the die is in the set $\{1, 2, 3, 4\}$ or in the set $\{3, 4, 5, 6\}$. What is the probability of the die landing on a 3 given the game master's statement?

All three problems essentially have the same structure. Firstly, there is a finite set of outcomes in each problem. Monty Hall's problem has three doors, the boy or girl problem has three distinct family compositions and there are six values on a die. Secondly, the game master must reveal some information where it is essential that at least one outcome exists remaining possible no matter what the game master reveals. For that outcome the game master is free to choose his revelation. If in Monty Hall's problem door a has the car, the game master can choose to either open door a or door b. If for the boy or girl problem the stranger has one son and one daughter, he is free to choose whether he tells at least one child is a boy or at least one is a girl. If the die rolled 3 in the dice game, the game master is free to choose to reveal $\{1, 2, 3, 4\}$ or $\{3, 4, 5, 6\}$.
Since there is at least one outcome remaining possible no matter which revelation the game master makes, traditional conditional probability does not suffice. This

is seen for example in the dice game in Section 3.2, where Equation 3.8 states that the probability of rolling 6 given $\{3,4,5,6\}$ dilates from $\frac{1}{4}$ to $\frac{1}{2}$. Measure-theoretic conditional probability is not sufficient as well, as the smallest sub-σ-algebra of $\{1,2,3,4,5,6\}$ containing the subset $\{\{1,2,3,4\},\{3,4,5,6\}\}$ does not contain the set $\{3\}$. When a larger sub-σ-algebra is chosen that does contain $\{3\}$, a version of the measure-theoretic conditional expectation cannot be found as the problem does not state the probability of for example revealing $\{1,2,3,4\}$ when a 3 is rolled.

All three problems can be analysed using the same methods; they are essentially equal. The game master has a choice and therefore is able to employ a strategy. To counteract such strategies we need to first model all possible probability distributions. We can then create a pragmatic probability distribution using safe probability where the decision maker can act on. This distribution performs for predicting the distribution on the outcome space on average as well as all probability distributions in the model.

In this chapter we first introduce Theorem 4.1.1 to create safe probability distributions in the setting of the three introduced problems. Then each problem is treated individually where Theorem 4.1.1 is applied and we will discuss the results. The dice game will be treated in Section 4.2, the Monty Hall problem will be treated in Section 4.3 and the boy or girl problem will be treated in Section 4.4. We will conclude with Section 4.5 by stating that there is nothing paradoxical going on with the three problems. The paradoxical statements are just misapplications of conditional probability.

4.1 The main theorem

In the general setting we consider a set $\mathcal{X}$ of all possible outcomes called the *outcome space* and a set $\mathcal{Y}$ of observations the player can make called the *observation space*. Our sample space then becomes $\Omega = \mathcal{X} \times \mathcal{Y}$. Let U be an $\mathcal{X}$-valued random variable denoting the outcome of the game, e.g. the door containing the car in the Monty Hall problem or the family composition in the boy or girl problem. Let V be a $\mathcal{Y}$-valued random variable denoting the statement of the game master, e.g. the opening of door b in the Monty Hall problem. Using safe probability we will for all $u \in \mathcal{X}$ create a safe distribution for $\mathbb{1}_{\{U=u\}} \,|[V]$.

Theorem 4.1.1. *Let $\mathcal{X}$ be countable and $\mathcal{Y}$ be finite. Let U be an $\mathcal{X}$-valued random variable and V be a $\mathcal{Y}$-valued random variable. Let $\{p_u\}_{u \in \mathcal{X}} \subset [0,1]$ with $\sum_{u \in \mathcal{X}} p_u = 1$. Let*

$$\mathcal{P}^* \subseteq \{\mathbb{P} \mid \forall u \in \mathcal{X} : \mathbb{P}[U = u] = p_u\} \tag{4.1}$$

be our model of probability distributions on $\mathcal{X} \times \mathcal{Y}$ where $|\mathcal{Y}|$ distributions $\mathbb{P}_1, \ldots, \mathbb{P}_{|\mathcal{Y}|} \in \mathcal{P}^*$ impose $|\mathcal{Y}|$ linearly independent vectors $(\mathbb{P}_i[V = v])_{v \in \mathcal{Y}}$ with $i \in \{1, \ldots, |\mathcal{Y}|\}$. Let $u \in \mathcal{X}$ be arbitrary and let $\tilde{\mathbb{P}}$ be a distribution on $\mathcal{X} \times \mathcal{Y}$ with full support on V, then the following are equivalent:

1. *For all $v \in \mathcal{Y}$ we have $\tilde{\mathbb{P}}[U = u | V = v] = p_u$.*

2. *$\tilde{\mathbb{P}}$ is safe for $\mathbb{1}_{\{U=u\}} \,|[V]$.*

3. *$\tilde{\mathbb{P}}$ is safe for $\langle \mathbb{1}_{\{U=u\}} \rangle | \langle V \rangle$.*

Proof. The implication from 1 to 2 is satisfied by Proposition 3.1.4. Let $v \in \mathcal{Y}$ and $\mathbb{P} \in \mathcal{P}^*$ be arbitrary, then we have

$$\tilde{\mathbb{P}}[U = u | V = v] = p_u = \mathbb{P}[U = u]. \tag{4.2}$$

Proposition 3.1.4 states that $\tilde{\mathbb{P}}$ is safe for $\mathbb{1}_{\{U=u\}} | [V]$.

The implication from 2 to 3 is by Proposition 3.1.5.

Consider lastly the implication from 3 to 1. Safety for $\langle \mathbb{1}_{\{U=u\}} \rangle | \langle V \rangle$ implies

$$\mathbb{E}_{\mathbb{P}}[\mathbb{1}_{\{U=u\}}] = \mathbb{E}_{\mathbb{P}}[\mathbb{E}_{\tilde{\mathbb{P}}}[\mathbb{1}_{\{U=u\}} | V]] \tag{4.3}$$

for all $\mathbb{P} \in \mathcal{P}^*$. We will construct a sufficient $\tilde{\mathbb{P}}$ from this requirement and prove that $\tilde{\mathbb{P}}[U = u | V = v] = p_u$ for all $v \in \mathcal{Y}$ is necessary.

Let $\mathbb{P} \in \mathcal{P}^*$ be arbitrary. Take a first look at $\mathbb{E}_{\mathbb{P}}[\mathbb{1}_{\{U=u\}}]$, then

$$\mathbb{E}_{\mathbb{P}}[\mathbb{1}_{\{U=u\}}] = \mathbb{P}[U = u] = p_u. \tag{4.4}$$

Let $\tilde{\mathbb{P}}$ be a arbitrary distribution on $\mathcal{X} \times \mathcal{Y}$ that is safe for $\langle \mathbb{1}_{\{U=u\}} \rangle | \langle V \rangle$ and let $v \in \mathcal{Y}$ be arbitrary, then we can write out

$$\mathbb{E}_{\tilde{\mathbb{P}}}[\mathbb{1}_{\{U=u\}} | V = v] = \tilde{\mathbb{P}}[U = u | V = v]. \tag{4.5}$$

The value $\tilde{\mathbb{P}}[U = u | V = v]$ exists since $\tilde{\mathbb{P}}$ has full support on V. The expectation $\mathbb{E}_{\mathbb{P}}[\mathbb{E}_{\tilde{\mathbb{P}}}[\mathbb{1}_{\{U=u\}} | V]]$ can now be expanded to

$$\mathbb{E}_{\mathbb{P}}[\mathbb{E}_{\tilde{\mathbb{P}}}[\mathbb{1}_{\{U=u\}} | V]] = \sum_{v \in \mathrm{supp}_{\mathbb{P}}(V)} \mathbb{E}_{\tilde{\mathbb{P}}}[\mathbb{1}_{\{U=u\}} | V = v] \, \mathbb{P}[V = v] \tag{4.6}$$

$$= \sum_{v \in \mathrm{supp}_{\mathbb{P}}(V)} \tilde{\mathbb{P}}[U = u | V = v] \, \mathbb{P}[V = v]. \tag{4.7}$$

We now need to abbreviate our notation. From now on we write $\tilde{\mathbb{P}}[u|v]$ instead of $\tilde{\mathbb{P}}[U = u | V = v]$ for all $(u, v) \in \mathcal{X} \times \mathcal{Y}$. For safety for $\langle \mathbb{1}_{\{U=u\}} \rangle | \langle V \rangle$ we require

$$p_u = \mathbb{E}_{\mathbb{P}}[\mathbb{E}_{\tilde{\mathbb{P}}}[\mathbb{1}_{\{U=u\}} | V]] = \sum_{v \in \mathrm{supp}_{\mathbb{P}}(V)} \tilde{\mathbb{P}}[u|v] \, \mathbb{P}[V = v]. \tag{4.8}$$

We will also abbreviate the index of the sum to $v \in \mathcal{Y}$ instead of $v \in \mathrm{supp}_{\mathbb{P}}(V)$, as $\mathbb{P}[V = v] = 0$ will be zero for every $v \in \mathcal{Y} \setminus \mathrm{supp}_{\mathbb{P}}(V)$ and this will have no impact on the summation.

Note that equation 4.8 is a linear combination of the vectors $(\tilde{\mathbb{P}}[u|v])_{v \in \mathcal{Y}}$ and $(\mathbb{P}[V = v])_{v \in \mathcal{Y}}$ and the values of the latter vector are known, thus the equation $\sum_{v \in \mathcal{Y}} \tilde{\mathbb{P}}[u|v] \, \mathbb{P}[V = v] = p_u$ is linear. Since $\mathcal{P}^*$ has $|\mathcal{Y}|$ distributions $\mathbb{P}_1, \dots, \mathbb{P}_{|\mathcal{Y}|}$ that have linearly independent vectors $(\mathbb{P}_i[V = v])_{v \in \mathcal{Y}}$ with $i \in \{1, \dots, |\mathcal{Y}|\}$, we can create a system of $|\mathcal{Y}|$ linearly independent equations

$$\sum_{v \in \mathcal{Y}} \tilde{\mathbb{P}}[u|v] \, \mathbb{P}_i[V = v] = p_u, \qquad i \in \{1, \dots, |\mathcal{Y}|\}. \tag{4.9}$$

Since we have a system of $|\mathcal{Y}|$ linearly independent equations with $|\mathcal{Y}|$ unknowns $\tilde{\mathbb{P}}[u|v_1], \dots, \tilde{\mathbb{P}}[u|v_{|\mathcal{Y}|}]$, there is a unique solution. The coefficients $\mathbb{P}[V = v]$ with sum up to 1, thus $\tilde{\mathbb{P}}[u|v] = p_u$ is the unique solution to this system.

Let now $\mathbb{P} \in \mathcal{P}^*$ again be arbitrary, then the requirement in Equation 4.8 still holds as

$$\mathbb{E}_{\mathbb{P}}[\mathbb{E}_{\tilde{\mathbb{P}}}[\mathbb{1}_{\{U=u\}} \,|V]] = \sum_{v \in \operatorname{supp}_{\mathbb{P}}(V)} \tilde{\mathbb{P}}[u|v] \, \mathbb{P}[V=v] = p_u \sum_{v \in \operatorname{supp}_{\mathbb{P}}(V)} \mathbb{P}[V=v] = p_u.$$
(4.10)

Therefore $\tilde{\mathbb{P}}[u|v] = p_u$ for all $v \in \mathcal{Y}$ is required for $\tilde{\mathbb{P}}$ when $\tilde{\mathbb{P}}$ is safe for $\langle\mathbb{1}_{\{U=u\}}\rangle|\langle V\rangle$, thus 3 implies 1. $\qquad\square$

It looks like Theorem 4.1.1 can only be applied in a few circumstances. However, the restriction on model $\mathcal{P}^*$ is in many cases quickly satisfied, as in most cases $\mathcal{P}^*$ is not a strict subset of the set in Equation 4.1. If $\mathcal{P}^*$ is equal to the set in Equation 4.1, then the distributions $\mathbb{P}_i$ with $\mathbb{P}_i[V = v_j] = \delta_{ij}$ for all $i, j \in \{1, \ldots, |\mathcal{Y}|\}$, with δ the Kronecker delta, form a sequence of $|\mathcal{Y}|$ linearly independent unit vectors imposed from $|\mathcal{Y}|$ distributions of $\mathcal{P}^*$. In other cases when excluding the unit vectors, if for all $v \in \mathcal{Y}$ there is a $\mathbb{P} \in \mathcal{P}^*$ with $v \in \operatorname{supp}_{\mathbb{P}}(V)$, then finding $|\mathcal{Y}|$ different distributions that impose $|\mathcal{Y}|$ linearly independent marginal probabilities on V is not a hard task; otherwise the model $\mathcal{P}^*$ is simply too small. This leads to the following lemma that is helpful when checking whether the model is large enough to apply Theorem 4.1.1.

Lemma 4.1.2. *Let $\mathcal{X}$, $\mathcal{Y}$, U, V, $\{p_u\}_{u \in \mathcal{X}}$ be as in Theorem 4.1.1. Let*

$$\mathcal{P}^* \subseteq \{\mathbb{P} \mid \forall u \in \mathcal{X} : \mathbb{P}[U = u] = p_u\}.$$
(4.11)

It is necessary that $\bigcup_{\mathbb{P} \in \mathcal{P}^*} \operatorname{supp}_{\mathbb{P}}(V) = \mathcal{Y}$ holds for $\mathcal{P}^*$ to fulfil the requirements of Theorem 4.1.1.

Proof. Suppose $\bigcup_{\mathbb{P} \in \mathcal{P}^*} \operatorname{supp}_{\mathbb{P}}(V) \neq \mathcal{Y}$, then there is a $v' \in \mathcal{Y}$ with $\mathbb{P}[V = v'] = 0$ for all $\mathbb{P} \in \mathcal{P}^*$. Take this v' and let $\mathbb{P}_1, \ldots, \mathbb{P}_{|\mathcal{Y}|} \in \mathcal{P}^*$ be an arbitrary sequence of distributions. Then the vectors $(\mathbb{P}_1[V = v])_{v \in \mathcal{Y}}, \ldots, (\mathbb{P}_{|\mathcal{Y}|}[V = v])_{v \in \mathcal{Y}}$ are not linearly independent as the vectors are of size $|\mathcal{Y}|$ as well and the entry of v' is always zero. $\qquad\square$

Furthermore, Theorem 4.1.1 implies that a safe probability for $\mathbb{1}_{\{U=u\}} |[V]$ can be created for all $u \in \mathcal{X}$. However, earlier in the introduction of Section 4.1 we specifically discussed a special $u' \in \mathcal{X}$ that is supported by $U|V = v$ for all $v \in \operatorname{supp}_{\mathbb{P}}(V)$ and $\mathbb{P} \in \mathcal{P}^*$.

Take for example the dice game. There either 3 or 4 can be picked as u' as those values are still possible after the game master's statement that the value of the die is either in $\{1, 2, 3, 4\}$ or in $\{3, 4, 5, 6\}$. According to Theorem 4.1.1 we can pick $u' = 1$ as well to create a safe distribution $\tilde{\mathbb{P}}$ for $\mathbb{1}_{\{U=1\}} |[V]$, however this implies

$$\tilde{\mathbb{P}}[U = 1|V = \{3, 4, 5, 6\}] = \frac{1}{6}.$$
(4.12)

Thus according to safe probability we need to assume that the game master lies with probability $\frac{1}{6}$ when he tells the die has rolled three or higher, while in the problem we specifically assume the game master is not able to lie. This solution or distribution should therefore not be regarded as admissible, while it is still mathematically valid. When calculating a safe distribution $\tilde{\mathbb{P}}$ for $\mathbb{1}_{\{U=3\}} |[V]$, as is done in Proposition 4.2.1, we obtain $\tilde{\mathbb{P}}[U = 3|V = v] = \frac{1}{6}$ for all $v \in \mathcal{Y}$. Since for all $v \in \mathcal{Y}$ a $\mathbb{P} \in \mathcal{P}^*$ exists with $\mathbb{P}[U = 3|V = v] > 0$, putting a positive

probability on $U = 3$ given any value of V is a wise thing to do.

Therefore, while Theorem 4.1.1 can be applied in many cases, it is wise to only apply it on $\mathbb{1}_{\{U=u\}} | V$ when firstly $\mathbb{P}[U = u] = p_u > 0$ holds for all $\mathbb{P} \in \mathcal{P}^*$ and secondly when this u is a member of the set

$$\bigcap_{\mathbb{P} \in \mathcal{P}^*} \bigcap_{v \in \mathrm{supp}_{\mathbb{P}}(V)} \mathrm{range}(U|V = v). \tag{4.13}$$

Explaining the second requirement, since the event $\{U = u\}$ has positive and equal probability for all $\mathbb{P} \in \mathcal{P}^*$ and since we have $\bigcup_{\mathbb{P} \in \mathcal{P}^*} \mathrm{supp}_{\mathbb{P}}(V) = \mathcal{Y}$ by Lemma 4.1.2, for each $v \in \mathcal{Y}$ a $\mathbb{P} \in \mathcal{P}^*$ must exist with $\mathbb{P}[U = u|V = v] > 0$ when u is a member of the set in Equation 4.13, making it sensible to put $\tilde{\mathbb{P}}[U = u|V = v] = \mathbb{P}[U = u] > 0$.

The requirement $\mathcal{Y} = \bigcup_{\mathbb{P} \in \mathcal{P}^*} \mathrm{supp}_{\mathbb{P}}(V)$ is however restrictive. The following proposition drops this requirement, but the equivalence in Theorem 4.1.1 disappears as well leaving only the implications from 1 to 2 and from 2 to 3.

Proposition 4.1.3. *Let $\mathcal{X}$, $\mathcal{Y}$, U, V and $\{p_u\}_{u \in \mathcal{X}}$ be as before. Let*

$$\mathcal{P}^* \subseteq \{\mathbb{P} \,|\, \forall u \in \mathcal{X} : \mathbb{P}[U = u] = p_u\} \tag{4.14}$$

be non-empty and let $\tilde{\mathbb{P}}$ be a distribution on $\mathcal{X} \times \mathcal{Y}$. Let $u \in \mathcal{X}$, then $\tilde{\mathbb{P}}$ is safe for $\mathbb{1}_{\{U=u\}} | [V]$ if $\tilde{\mathbb{P}}[U = u|V = v] = p_u$ holds for all $v \in \mathcal{Y}$. This $\tilde{\mathbb{P}}$ is safe for $\langle \mathbb{1}_{\{U=u\}} \rangle | \langle V \rangle$ as well.

Proof. Let $\mathbb{P} \in \mathcal{P}^*$ and $v \in \mathcal{Y}$ be arbitrary, then we have

$$\tilde{\mathbb{P}}[U = u|V = v] = p_u = \mathbb{P}[U = u]. \tag{4.15}$$

Proposition 3.1.4 states that this $\tilde{\mathbb{P}}$ is safe for $\mathbb{1}_{\{U=u\}} | [V]$.

Safety for $\langle \mathbb{1}_{\{U=u\}} \rangle | \langle V \rangle$ now immediately follows for $\tilde{\mathbb{P}}$ by Proposition 3.1.5. $\square$

Note that from the proof of Proposition 4.1.3 one can deduce that the elaborate set-up in Theorem 4.1.1 is not necessarily needed and that Proposition 4.1.3 is nothing more than an applied restatement of the fourth property of Proposition 3.1.4. However, in the case of Proposition 4.1.3, you do not know whether distributions exist that are safe for $\langle \mathbb{1}_{\{U=u\}} \rangle | \langle V \rangle$ but not for $\mathbb{1}_{\{U=u\}} | [V]$. In the case of Theorem 4.1.1, safety for $\langle \mathbb{1}_{\{U=u\}} \rangle | \langle V \rangle$ and $\mathbb{1}_{\{U=u\}} | [V]$ are equivalent. Finding a safe distribution for $\mathbb{1}_{\{U=u\}} | [V]$ is easy when the distribution on U in $\mathcal{P}^*$ is known, however for knowing that only these distributions are unbiased for U we need Theorem 4.1.1.

Theorem 4.1.1 only considers single $u \in \mathcal{X}$. We can construct safe probability distributions for $\mathbb{1}_{\{U \in \mathcal{X}'\}} | V$ for larger $\mathcal{X}' \subset \mathcal{X}$ as well in the following corollary.

Corollary 4.1.4. *Let $\mathcal{X}$, $\mathcal{Y}$, U, V, $\{p_u\}_{u \in \mathcal{Y}}$ and $\mathcal{P}^*$ be as in Theorem 4.1.1. Let $\mathcal{X}' \subseteq \mathcal{X}$ be non-empty. Let $\tilde{\mathbb{P}}$ be a distribution on $\mathcal{X} \times \mathcal{Y}$ with full support on V, then the following are equivalent:*

1. *For all $u \in \mathcal{X}'$ and $v \in \mathcal{Y}$ we have $\tilde{\mathbb{P}}[U = u|V = v] = p_u$.*

2. *$\tilde{\mathbb{P}}$ is safe for $\mathbb{1}_{\{U \in \mathcal{X}'\}} | [V]$.*

3. *$\tilde{\mathbb{P}}$ is safe for $\langle \mathbb{1}_{\{U \in \mathcal{X}'\}} \rangle | \langle V \rangle$.*

Proof. The proof is in essence a repetition of the proof of Theorem 4.1.1 where $\mathbb{1}_{\{U \in \mathcal{X}'\}} = \max_{u \in \mathcal{X}'} \mathbb{1}_{\{U=u\}} = \sum_{u \in \mathcal{X}'} \mathbb{1}_{\{U=u\}}$ must be used. $\square$

4.1.1 Accuracy

Suppose you are in a situation where you need to guess $\mathbb{1}_{\{U \in \mathcal{X}'\}}$ for an $\mathcal{X}' \subset \mathcal{X}$. One example is the Monty Hall game where on live television you need to guess whether your door a has the car behind, thus whether $\mathbb{1}_{\{U=a\}} = 1$ holds given the opened door V. To probability of guessing correctly is called the *accuracy*.

Definition 4.1.5 (Accuracy)**.** Let $\mathcal{Z}$ be a countable set. Let X and Y be $\mathcal{Z}$-valued random variables. Let $\mathcal{P}^*$ be a set of probability measures on $\mathcal{Z}$ and let $\mathbb{P}'$ be an arbitrary probability measure on $\mathcal{Z}$ denoting the distribution of X. The *accuracy of X for guessing Y distributed by* $\mathbb{P}$ is defined by

$$\operatorname{acc}_Y^{\mathbb{P}}(X) = \sum_{k \in \mathcal{Z}} \mathbb{P}'[X = k] \, \mathbb{P}[Y = k]. \tag{4.16}$$

When $\operatorname{acc}_Y^{\mathbb{P}}(X)$ is constant for all $\mathbb{P} \in \mathcal{P}^*$, this value is called the *accuracy of X for guessing Y* and is denoted by $\operatorname{acc}_Y(X)$. If from context it is clear we want to guess Y, we write *the accuracy of X* and denote this by $\operatorname{acc}(X)$.

The next theorem optimizes the accuracy of a random variable $\mathbb{1}_{\{U \in \mathcal{X}'\}}^*$ for guessing $\mathbb{1}_{\{U \in \mathcal{X}'\}}$ against all $\mathbb{P} \in \mathcal{P}^*$.

Theorem 4.1.6. *Let $\mathcal{X}$, $\mathcal{Y}$, U, V, $\{p_u\}_{u \in \mathcal{Y}}$ and $\mathcal{P}^*$ be as in Proposition 4.1.3 and let $\mathcal{X}' \subseteq \mathcal{X}$ be non-empty. Let $\mathbb{1}_{\{U \in \mathcal{X}'\}}^*$ be a binary-valued random variable on $\mathcal{X} \times \mathcal{Y}$ with probability measure $\mathbb{P}'$ such that $\mathbb{1}_{\{U \in \mathcal{X}'\}}^*$ is distributed as*

$$\mathbb{1}_{\{U \in \mathcal{X}'\}}^* \sim \begin{cases} 1, & \sum_{u \in \mathcal{X}'} p_u > \frac{1}{2}, \\ 0, & \sum_{u \in \mathcal{X}'} p_u < \frac{1}{2}, \\ \operatorname{Ber}(p), & \sum_{u \in \mathcal{X}'} p_u = \frac{1}{2}, p \in [0,1] \end{cases} \tag{4.17}$$

where $\operatorname{Ber}(p)$ is the Bernoulli distribution with parameter $p \in [0,1]$. This distribution guesses $\mathbb{1}_{\{U \in \mathcal{X}'\}}$ correctly with probability

$$\operatorname{acc}^{\mathbb{P}}\left(\mathbb{1}_{\{U \in \mathcal{X}'\}}^*\right) = \max\left\{\sum_{u \in \mathcal{X}'} p_u, 1 - \sum_{u \in \mathcal{X}'} p_u\right\}. \tag{4.18}$$

Any other binary random variable W has a lower accuracy than $\mathbb{1}_{\{U \in \mathcal{X}'\}}^*$.

Proof. Let $\mathbb{P} \in \mathcal{P}^*$ be arbitrary and let W be an arbitrary $\{0,1\}$-valued random variable on $\mathcal{X} \times \mathcal{Y}$. Let $\hat{\mathbb{P}}$ be an arbitrary probability measure on $\mathcal{X} \times \mathcal{Y}$ which we will apply on W. We will first prove that $\operatorname{acc}^{\mathbb{P}}(W)$ is constant for all $\mathbb{P} \in \mathcal{P}^*$ and then prove that $W = \mathbb{1}_{\{U \in \mathcal{X}'\}}^*$ maximizes the accuracy of W.

By Definition 4.1.5 and the fact that $\mathbb{P} \in \mathcal{P}^*$ we have

$$\operatorname{acc}_{\mathbb{P}}(W) = \hat{\mathbb{P}}[W = 0] \sum_{u \in \mathcal{X} \setminus \mathcal{X}'} \mathbb{P}[U = u] + \hat{\mathbb{P}}[W = 1] \sum_{u \in \mathcal{X}'} \mathbb{P}[U = u] \tag{4.19}$$

$$= \hat{\mathbb{P}}[W = 0] \left(1 - \sum_{u \in \mathcal{X}'} p_u\right) + \hat{\mathbb{P}}[W = 1] \sum_{u \in \mathcal{X}'} p_u, \tag{4.20}$$

thus $\operatorname{acc}^{\mathbb{P}}(W)$ is constant for all $\mathbb{P} \in \mathcal{P}^*$. Abbreviate the terms $\sum_{u \in \mathcal{X}'} p_u = x$ and $\hat{\mathbb{P}}[W = 1] = p$. We now need to maximize

$$f(p) := px + (1 - p)(1 - x) = p(2x - 1) + 1 - x. \tag{4.21}$$

Note that both $x, p \in [0,1]$. Consider the following different cases for x.

$x > \frac{1}{2}$: When $x > \frac{1}{2}$ holds, the function f is strictly increasing in p. Its maximum is then achieved at $p = 1$, which is p. Thus always guessing $W = 1$ maximizes its accuracy.

$x < \frac{1}{2}$: When $x < \frac{1}{2}$ holds, the function f is strictly decreasing in p. Its maximum is then achieved at $p = 0$, which is $1 - x$. Thus always guessing $W = 0$ maximizes its accuracy.

$x = \frac{1}{2}$: When $x = \frac{1}{2}$ holds, the function f is constant. Its value is $f(p) = \frac{1}{2} = x$ for all $p \in [0, 1]$. Therefore if W is made the Bernoulli distribution with parameter p, the accuracy of W is maximized.

Therefore W has maximal accuracy when $W = \mathbb{1}^*_{\{U \in \mathcal{X}'\}}$ and the accuracy of $\mathbb{1}^*_{\{U \in \mathcal{X}'\}}$ is

$$\mathrm{acc}\left(\mathbb{1}^*_{\{U \in \mathcal{X}'\}}\right) = \max\left\{\sum_{u \in \mathcal{X}'} p_u, 1 - \sum_{u \in \mathcal{X}'} p_u\right\}. \tag{4.22}$$

$\square$

Suppose the player applies a safe distribution $\tilde{\mathbb{P}}$ on his guess $\mathbb{1}^*_{\{U \in \mathcal{X}'\}}$ for $\mathbb{1}_{\{U \in \mathcal{X}'\}}$, what will the accuracy be? The following proposition answers this question.

Proposition 4.1.7. *Let $\mathcal{X}$, $\mathcal{X}'$, $\mathcal{Y}$, U, V, $\{p_u\}_{u \in \mathcal{Y}}$ and $\mathcal{P}^*$ be as in Corollary 4.1.4. Let $\tilde{\mathbb{P}}$ be a safe distribution for $\mathbb{1}_{\{U \in \mathcal{X}'\}} | \langle V \rangle$. Let $\tilde{\mathbb{1}}_{\{U \in \mathcal{X}'\}}$ be a $\{0, 1\}$-valued random variable and employ distribution $\mathbb{P}'$ on $\tilde{\mathbb{1}}_{\{U \in \mathcal{X}'\}}$ such that*

$$\mathbb{P}'\left[\tilde{\mathbb{1}}_{\{U \in \mathcal{X}'\}} = 1\right] = \tilde{\mathbb{P}}[\mathbb{1}_{\{U \in \mathcal{X}'\}} = 1 | V = v] = \sum_{u \in \mathcal{X}'} p_u \tag{4.23}$$

holds for all $v \in \mathcal{Y}$. The accuracy of $\tilde{\mathbb{1}}_{\{U \in \mathcal{X}'\}}$ is

$$\mathrm{acc}^{\mathbb{P}}\left(\tilde{\mathbb{1}}_{\{U \in \mathcal{X}'\}}\right) = \left(\sum_{u \in \mathcal{X}'} p_u\right)^2 + \left(1 - \sum_{u \in \mathcal{X}'} p_u\right)^2 \tag{4.24}$$

for all $\mathbb{P} \in \mathcal{P}^*$.

Proof. Let $\mathbb{P} \in \mathcal{P}^*$ be arbitrary. By Corollary 4.1.4 and Definition 4.1.5 we have

$$\mathrm{acc}^{\mathbb{P}}\left(\tilde{\mathbb{1}}_{\{U \in \mathcal{X}'\}}\right) = \sum_{k=0}^{1} \mathbb{P}'\left[\tilde{\mathbb{1}}_{\{U \in \mathcal{X}'\}} = k\right] \mathbb{P}[\mathbb{1}_{\{U \in \mathcal{X}'\}} = k] \tag{4.25}$$

$$= \mathbb{P}'[\mathbb{1}_{\{U \in \mathcal{X}'\}} = 0]\left(1 - \sum_{u \in \mathcal{X}'} p_u\right) + \mathbb{P}'[\mathbb{1}_{\{U \in \mathcal{X}'\}} = 1]\sum_{u \in \mathcal{X}'} p_u. \tag{4.26}$$

Since $\mathbb{P}[\mathbb{1}_{\{U \in \mathcal{X}'\}} = 1] = \sum_{u \in \mathcal{X}'} p_u$ holds for all $\mathbb{P} \in \mathcal{P}^*$, we can conclude that

$$\mathrm{acc}\left(\tilde{\mathbb{1}}_{\{U \in \mathcal{X}'\}}\right) = \left(\sum_{u \in \mathcal{X}'} p_u\right)^2 + \left(1 - \sum_{u \in \mathcal{X}'} p_u\right)^2. \tag{4.27}$$

$\square$

Now we know when a safe distribution $\tilde{\mathbb{P}}$ for $\mathbb{1}_{\{U\in\mathcal{X}'\}}|[V]$ results into being the most accurate guess for $\mathbb{1}_{\{U\in\mathcal{X}'\}}$ as well.

Corollary 4.1.8. *Let $\mathcal{X}$, $\mathcal{X}'$, $\mathcal{Y}$, U, V, $\{p_u\}_{u\in\mathcal{Y}}$ and $\mathcal{P}^*$ be as in Corollary 4.1.4. Let $\mathbb{1}^*_{\{U\in\mathcal{X}'\}}$ be as in Theorem 4.1.6. Let $\tilde{\mathbb{P}}$ be safe for $\mathbb{1}_{\{U\in\mathcal{X}'\}}|[V]$ and let $\tilde{\mathbb{1}}_{\{U\in\mathcal{X}'\}}$ be as in Proposition 4.1.7. The accuracy of $\tilde{\mathbb{1}}_{\{U\in\mathcal{X}'\}}$ for guessing $\mathbb{1}_{\{U\in\mathcal{X}'\}}$ is only optimal when*

$$\sum_{u\in\mathcal{X}'} p_u \in \left\{0, \frac{1}{2}, 1\right\}. \tag{4.28}$$

Proof. Let $\mathbb{P}\in\mathcal{P}^*$ be arbitrary. According to Proposition 4.1.7 the accuracy of $\tilde{\mathbb{1}}_{\{U\in\mathcal{X}'\}}$ is

$$\mathrm{acc}\left(\tilde{\mathbb{1}}_{\{U\in\mathcal{X}'\}}\right) = \left(\sum_{u\in\mathcal{X}'} p_u\right)^2 + \left(1 - \sum_{u\in\mathcal{X}'} p_u\right)^2. \tag{4.29}$$

Theorem 4.1.6 states that the optimal accuracy for predicting $\mathbb{1}_{\{U\in\mathcal{X}'\}}$ is

$$\mathrm{acc}\left(\mathbb{1}^*_{\{U\in\mathcal{X}'\}}\right) = \max\left\{\sum_{u\in\mathcal{X}'} p_u, 1 - \sum_{u\in\mathcal{X}'} p_u\right\}. \tag{4.30}$$

Note that $\tilde{\mathbb{1}}_{\{U\in\mathcal{X}'\}}$ and $\mathbb{1}^*_{\{U\in\mathcal{X}'\}}$ are only equal when $\sum_{u\in\mathcal{X}'} p_u \in \{0,1\}$ holds or when $\sum_{u\in\mathcal{X}'} p_u = \frac{1}{2}$ and $\mathbb{1}^*_{\{U\in\mathcal{X}'\}} = \mathrm{Ber}\left(\frac{1}{2}\right)$ hold. The first case is quite trivial as when $\sum_{u\in\mathcal{X}'} p_u \in \{0,1\}$ we have $\mathrm{acc}\left(\tilde{\mathbb{1}}_{\{U\in\mathcal{X}'\}}\right) = 1$ as result. When $\sum_{u\in\mathcal{X}'} p_u = \frac{1}{2}$ holds, we have $\mathbb{P}[\tilde{\mathbb{1}}_{\{U\in\mathcal{X}'\}} = 1] = \frac{1}{2}$, resulting into $\tilde{\mathbb{1}}_{\{U\in\mathcal{X}'\}} = \mathrm{Ber}\left(\frac{1}{2}\right)$. Since $\mathbb{1}^*_{\{U\in\mathcal{X}'\}} = \mathrm{Ber}\left(\frac{1}{2}\right)$ gives optimal accuracy for $\sum_{u\in\mathcal{X}'} p_u = \frac{1}{2}$ by Theorem 4.1.6, the accuracy of $\tilde{\mathbb{1}}_{\{U\in\mathcal{X}'\}}$ is in this case optimal with $\mathrm{acc}\left(\tilde{\mathbb{1}}_{\{U\in\mathcal{X}'\}}\right) = \frac{1}{2}$. $\square$

Knowing this, one must keep in mind that safe probability for $\mathbb{1}_{\{U\in\mathcal{X}'\}}|[V]$ gives in almost no case a prediction for $\mathbb{1}_{\{U\in\mathcal{X}'\}}|V$ with optimal accuracy. Safe probability merely simulates the distribution of $\mathbb{1}_{\{U\in\mathcal{X}'\}}$, where in the setting of Theorem 4.1.1 the revelation of the game master is in no cases of influence. A safe probability distribution must therefore in this setting not directly be used to make predictions for $\mathbb{1}_{\{U\in\mathcal{X}'\}}$, but it can be used to create a different random variable that does predict $\mathbb{1}_{\{U\in\mathcal{X}'\}}$ with optimal accuracy. In the Monty Hall game, you should not switch with probability $\frac{2}{3}$, as then the probability of winning the car is $\frac{5}{9}$. If the player observes that the safe distribution in the Monty Hall game puts a probability of $\frac{1}{3}$ door a, then never choosing door a yields a probability of winning the car of $\frac{2}{3}$, which is optimal by Theorem 4.1.6.

4.2 Dice game

With the main theorems stated, we can apply them to some well-known paradoxes. We will first continue our example of the dice game, started in Section 3.2. Recall that a die is cast with values in $\mathcal{X} = \{1, 2, 3, 4, 5, 6\}$ and the game master is, after observing the die's value, only able to reveal a member of the set $\mathcal{Y} = \{\{1, 2, 3, 4\}, \{3, 4, 5, 6\}\} =: \{y_1, y_2\}$. Our sample space is $\Omega = \mathcal{X}\times\mathcal{Y}$ with

its power set $\Sigma = 2^\Omega$ as σ-algebra. Let U be an $\mathcal{X}$-valued random variable and V be a $\mathcal{Y}$-valued random variable. The game master is not allowed to lie and the die is fair, thus any probability distribution for this game must be a member of

$$\mathcal{P}^* = \left\{ \mathbb{P} \middle| \forall u \in \mathcal{X} : \mathbb{P}[U = u] = \frac{1}{6}, \forall y \in \mathcal{Y} : \mathbb{P}[U \in y | Y = y] = 1 \right\}. \qquad (4.31)$$

Safe probability

Proposition 3.2.1 gives a set $\tilde{\mathcal{P}}$ of distributions that are all safe for $\langle U \rangle | [V]$. However, in the original problem we are not interested in the whole distribution of U given V, but only want to know the probability of rolling 3 after the game master's statement. Therefore finding safe distributions for $\mathbb{1}_{\{U=3\}} | [V]$ is sufficient.

Proposition 4.2.1. *Let $\mathcal{X}$, $\mathcal{Y}$, Ω, U, V and $\mathcal{P}^*$ be as in the dice game. Let $\tilde{\mathbb{P}}$ be a distribution on Ω with full support on V. Let $u \in \mathcal{X}$, then the following statements are equivalent:*

1. For all $v \in \mathcal{Y}$ we have $\tilde{\mathbb{P}}[U = u | V = v] = \frac{1}{6}$.

2. $\tilde{\mathbb{P}}$ is safe for $\langle \mathbb{1}_{\{U=u\}} \rangle | \langle V \rangle$.

3. $\tilde{\mathbb{P}}$ is safe for $\mathbb{1}_{\{U=u\}} | [V]$.

Proof. This proposition is a direct application of Theorem 4.1.1. The set $\mathcal{X}$ is of size 6 and therefore countable and the set $\mathcal{Y}$ is of size 2, which makes it finite. U is a random variable on $\mathcal{X}$ and V is a random variable on $\mathcal{Y}$. In this dice game we have $p_u = \frac{1}{6}$ as $\mathbb{P} \in \mathcal{P}^*$ implies $\mathbb{P}[U = u] = \frac{1}{6}$ for all $u \in \mathcal{X}$. Lastly, the set $\mathcal{P}^*$ is a subset of $\{\mathbb{P} | \forall u \in \mathcal{X} : \mathbb{P}[U = u] = p_u\}$.

Take $\mathbb{P}_1, \mathbb{P}_2 \in \mathcal{P}^*$ with $\mathbb{P}_i[V = y_i | U = 3] = \mathbb{P}_i[V = y_i | U = 4] = 1$. Let $i \in \{1, 2\}$ be arbitrary, then

$$\mathbb{P}_i[V = y_i] = \sum_{u=1}^{6} \mathbb{P}_i[V = y_i | U = u] \, \mathbb{P}_i[U = u] \qquad (4.32)$$

$$= \frac{1}{6} \left(4 \cdot 1 + 2 \cdot 0 \right) = \frac{2}{3}. \qquad (4.33)$$

Thus the vectors $(\mathbb{P}_i[V = v])_{v \in \mathcal{Y}}$ become $\left(\frac{1}{3}, \frac{2}{3} \right)$ and $\left(\frac{2}{3}, \frac{1}{3} \right)$, which are linearly independent.

Now we can directly apply Theorem 4.1.1 to conclude the proof. $\qquad \square$

Accuracy

Suppose you want to bet on this game and want to know whether you must put money on rolling 3. As the following proposition will state, after observing the game master's statement, always stating that the die has not rolled to a 3 has a $\frac{5}{6}$ chance of being correct independently on the game master's strategy and is the optimal probability.

Proposition 4.2.2. *Let $\mathcal{X}$, $\mathcal{Y}$, U, V and $\mathcal{P}^*$ be as in the dice game. Consider the random variable $\mathbb{1}^*_{\{U=3\}} \sim 0$ distributed as the singular distribution on 0, then this $\mathbb{1}^*_{\{U=3\}}$ has accuracy*

$$\mathrm{acc}^{\mathbb{P}}\left(\mathbb{1}^*_{\{U=3\}}\right) = \frac{5}{6} \tag{4.34}$$

for all $\mathbb{P} \in \mathcal{P}^*$. There is no random variable W with $\mathrm{acc}^{\mathbb{P}}(W) > \mathrm{acc}^{\mathbb{P}}(\mathbb{1}^*_{\{U=3\}})$ for a $\mathbb{P} \in \mathcal{P}^*$.

Proof. We want to apply Theorem 4.1.6. Firstly in this case $\mathcal{X}' = \{3\} \subset \mathcal{X}$ is non-empty. Secondly, we have $\sum_{u \in \mathcal{X}'} p_u = p_3 = \frac{1}{6} < \frac{1}{2}$. Then Theorem 4.1.6 states that $\mathbb{1}^*_{\{U=3\}} \sim 0$ optimizes the accuracy for guessing $\mathbb{1}_{\{U=3\}}$ with

$$\mathrm{acc}\left(\mathbb{1}^*_{\{U=3\}}\right) = \max\left\{\frac{1}{6}, \frac{5}{6}\right\} = \frac{5}{6}. \tag{4.35}$$

$\square$

We now observe that using the information given by the game master does not lead to a higher probability of guessing $\mathbb{1}_{\{U=3\}}$ correct. When always stating that the die has not rolled 3 you maximize the probability of having a correct guess, which is 5 out of 6. The information revealed by the game master can therefore be ignored as it will not yield to a higher probability of guessing correctly.

Final remarks

We have seen that the conditional probability of rolling 3 given the game master's statement is not uniquely defined; it dilates from 0 to $\frac{1}{4}$ depending on the to the player unknown strategy of the game master. However, when playing like the probability of rolling 3 is always $\frac{1}{6}$ disregarding the game master's statement, you are able to simulate the probability distribution of $\mathbb{1}_{\{U=3\}}$. Furthermore, when always stating that the die has not rolled 3, you are correct in five out of six games, which is the maximal fraction of games you can be correct on.

The paradoxical nature of this problem is that most people treat this problem with sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$ and condition on for example the information $\{3, 4, 5, 6\}$. However, as we have seen in Section 3.2, we also need to condition on the possibility $\{1, 2, 3, 4\}$ and those two sets cannot form a partition of Ω. Therefore the sample space Ω is insufficient for solving this problem, resulting in the paradoxical statements.

This underlines the argument that when conditioning the accompanying sub-σ-algebra must always be given, as then it is easily seen that conditioning on a sub-σ-algebra containing $\sigma(\{\{1, 2, 3, 4\}, \{3, 4, 5, 6\}\}) \subset 2^{\Omega}$ is not possible.

4.3 Monty Hall

One of the most classic paradoxes in probability theory is Monty Hall's three door problem. The problem is already explained in the introduction of this chapter, but we will repeat it here.[1]

---

[1] Geschiedenis en bronnenonderzoek toevoegen.

There are three doors called a, b and c. Two doors have a goat behind and one door has a car. The player chooses one door, where after the game master opens one of the remaining doors. The game master does not open the door with the car. The player is then asked whether he wants to switch to the remaining door. The player wins the contents of the door he ultimately chooses.

A logical question, and the centre of this paradox, is whether the player should switch when a door is opened. This question can be answered with multiple viewpoints. Throughout the whole analysis we assume that initially the car is placed behind a door with equal probability. Without loss of generality we now assume the player has chosen door a, as the initial choice of door has become independent of the question whether the player should switch.

The most naive viewpoint is the following. Say door c is opened by the game master, then either door a or door b has a car with equal probability $\frac{1}{2}$. Since there is a 50-50 chance, switching does not increase your chances. This viewpoint is widely disputed.

Another viewpoint using classic conditional probability states the following. In the beginning there is a $\frac{1}{3}$-chance that door a has a car. Assume after choosing door a the game master reveals door c. There was a chance of $\frac{2}{3}$ that the car was behind door b or c, leaving a $\frac{2}{3}$-chance that the car is behind door b. Therefore switching is advised, doubling your chances of winning the car. This viewpoint is however disputed as well.

Grünwald and Halpern [GH03] pointed out that the probability of the car being behind door a is anything between 0 and $\frac{1}{2}$. When the car is behind door a, the game master is free to choose between doors b and c. The game master's strategy influences the probability of the car being behind the other door.

Let us start with a formal statement of the problem. Let $\mathcal{X} = \{a, b, c\}$ be our outcome space and $\mathcal{Y} = \{b, c\}$ be our observation space. Let $\Omega = \mathcal{X} \times \mathcal{Y}$ be our sample space with $\Sigma = 2^{\Omega}$ as σ-algebra. Let U be an $\mathcal{X}$-valued random variable denoting the placement of the car and let V be a $\mathcal{Y}$-valued random variable on $\mathcal{Y}$ denoting the opening of a door.

Note that the doors are called a, b and c and not 1, 2 and 3, as in the latter case the expectation $\mathbb{E}[U]$ suddenly can be given a meaning. There is no such thing as the 'average door', as the labels on the doors are nothing more than names, not aspects indicating any value. Therefore we choose to name the doors a, b and c to not give rise to any confusion.

We assume that the car is initially distributed with equal probability and the game master cannot open a door with a car. A probability distribution must therefore be a member of the set

$$\mathcal{P}^* = \left\{ \mathbb{P} \middle| \forall u \in \mathcal{X} : \mathbb{P}[U = u] = \frac{1}{3}, \forall v \in \mathcal{Y} : \mathbb{P}[V = v | U = v] = 0 \right\}. \quad (4.36)$$

In this section we first give a repetition of the arguments of [Gr 13] as they view the Monty Hall problem in the viewpoint of traditional probability. Then we apply safe probability and Theorem 4.1.1 to the Monty Hall problem, with as result that it is safe to state that the car is behind door a with probability $\frac{1}{3}$.

Traditional conditional probability

Let us start with the viewpoint that result in a probability of $\frac{2}{3}$ of obtaining the car after switching. This viewpoint takes $\Omega' = \{a, b, c\}$ as sample space. Let U' be an Ω'-valued random variable denoting the cars location. Equip Ω' with the uniform distribution $\mathbb{P}$ such that $\mathbb{P}[U' = u] = \frac{1}{3}$. Assume lastly that $\mathbb{P}[U' \in \{a, b\}|U' = a] = \frac{1}{2}$, thus that the game master chooses a door with equal probability when he has free choice. Traditional conditional probability and Bayes' rule then give

$$\mathbb{P}[U' = a|U' \in \{a, b\}] = \frac{\mathbb{P}[U' \in \{a, b\}|U' = a]\,\mathbb{P}[U' = a]}{\sum_{u \in \Omega} \mathbb{P}[U' \in \{a, b\}|U' = u]\,\mathbb{P}[U' = u]} \tag{4.37}$$

$$= \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{3}\left(\frac{1}{2} + 0 + 1\right)} = \frac{1}{3}. \tag{4.38}$$

It is clear that the argument above needs to make assumptions on $\mathbb{P}$ in order to be able to perform the calculation, as following the problem's statement we cannot put a direct value on $\mathbb{P}[U' \in \{a, b\}|U' = a]$. We thus need to make a reasonable assumption for this argument to hold.

Let now $p \in [0, 1]$ be such that $\mathbb{P}[U' \in \{a, b\}|U' = a] = p$, then we have

$$\mathbb{P}[U' = a|U' \in \{a, b\}] = \frac{\mathbb{P}[U' \in \{a, b\}|U' = a]\,\mathbb{P}[U' = a]}{\sum_{u \in \Omega} \mathbb{P}[U' \in \{a, b\}|U' = u]\,\mathbb{P}[U' = u]} = \frac{p}{p + 1}. \tag{4.39}$$

Thus the probability of door a having the car after opening door b dilates from 0 to $\frac{1}{2}$, as [Gr 13] also pointed out in their work.

To create a probability of $\frac{1}{2}$ of our initial door having the car, we need to pick $p = 1$. We thus need to assume that when we initially chose the correct door, the game master *always* opens door b. This also debunks the argument that there are two doors left, thus there is a probability of $\frac{1}{2}$ that door a has the car. The information that door b has no car heavily influences the conditional probability.

Even more is wrong with the above reasoning. Suppose door a has the car. The game master is now able to choose between doors b and c. If door b is opened, the resulting set of possibilities is $\{a, b\}$ and when door c is opened, the resulting set of possibilities is $\{a, c\}$. We need to condition on both the possibilities $\{a, b\}$ and $\{a, c\}$, however both sets do not form a partition. Furthermore, the smallest σ-algebra containing both sets is

$$\mathcal{G} = \sigma(\{a, b\}, \{a, c\}) = 2^{\{a,b,c\}} = 2^{\Omega'}, \tag{4.40}$$

our original σ-algebra. As stated by [Wil91] and also easy to prove, we have $\mathbb{P}[F|\mathcal{G}] = \mathbb{P}[F]$ for all $F \in \mathcal{G}$ when $\mathcal{G} = 2^{\Omega'}$ is our original σ-algebra. We are therefore simply not able to use measure-theoretic conditional probability, let alone traditional conditional probability.

Can we say nothing sensible at all about the probability of winning the car after switching? Consider the Ω, U and V from the formal statement in the introduction of this section. Now we have $\Omega = \mathcal{X} \times \mathcal{Y}$, thus Ω is split up in outcome space $\mathcal{X} = \{a, b, c\}$ and observation space $\mathcal{Y} = \{b, c\}$. Now we can condition on all possibilities of open doors, as $\sigma(\{a\} \times \mathcal{Y})$ is a strict sub-σ-algebra of 2^{Ω}.

The resulting conditional probabilities come in the form of Equation 4.39. Let $\mathbb{P} \in \mathcal{P}^*$, thus such that $\mathbb{P}$ is uniform distributed on U and a door with a car cannot be opened. Then a $p \in [0,1]$ exists such that $\mathbb{P}[V = b | U = a] = p$. This results into

$$\mathbb{P}[U = a | V = b] = \frac{\mathbb{P}[V = b | U = a] \, \mathbb{P}[U = a]}{\sum_{u \in \Omega} \mathbb{P}[V = b | U = u] \, \mathbb{P}[U = u]} = \frac{p}{p+1}, \tag{4.41}$$

$$\mathbb{P}[U = a | V = c] = \frac{\mathbb{P}[V = c | U = a] \, \mathbb{P}[U = a]}{\sum_{u \in \Omega} \mathbb{P}[V = c | U = u] \, \mathbb{P}[U = u]} = \frac{1-p}{2-p}. \tag{4.42}$$

In both cases the probability of door a having the car dilates from 0 to $\frac{1}{2}$. Therefore, you are advised to switch no matter the strategy of the game master. However, we cannot pin down a single probability of door a having the car given the game master's statement. We can put a prior on $\mathbb{P}[V = b | U = a]$, however we do need more information to apply such prior wisely.

This subsection shows another example of why when doing conditional probability one always needs to provide the accompanying sub-σ-algebra. Unlike the Borel-Kolmogorov paradox in Chapter 2, here the measure-theoretic conditional expectation defines a unique probability measure. However, most use $\Omega = \{a, b, c\}$ as sample space and without taking the accompanying σ-algebra in account when conditioning, one should find no reason not to do this. Only when correctly applying conditional probability and noticing that both $\{a, b\}$ and $\{a, c\}$ are possible options, one finds out that $\Omega = \{a, b, c\}$ is insufficient as sample space and paradoxical results will rise.

Safe probability

Using safe probability it is possible to find a pragmatic distribution $\tilde{\mathbb{P}}$ on Ω that is safe against all $\mathbb{P} \in \mathcal{P}^*$. We only need to apply Theorem 4.1.1.

Proposition 4.3.1. *Let $\mathcal{X}$, $\mathcal{Y}$, Ω, U, V and $\mathcal{P}^*$ be as in the Monty Hall problem. Let $\tilde{\mathbb{P}}$ be a distribution on Ω with full support on V. Let $u \in \mathcal{X}$, then the following statements are equivalent:*

1. For all $v \in \mathcal{Y}$ we have $\tilde{\mathbb{P}}[U = u | V = v] = \frac{1}{3}$.

2. $\tilde{\mathbb{P}}$ is safe for $\langle \mathbb{1}_{\{U=u\}} \rangle | \langle V \rangle$.

3. $\tilde{\mathbb{P}}$ is safe for $\mathbb{1}_{\{U=u\}} | [V]$.

Proof. This proposition is a direct application of Theorem 4.1.1 and its proof is equal to the proof of Proposition 4.2.1. The set $\mathcal{X}$ is of size 3 and therefore countable and the set $\mathcal{Y}$ is of size 2, which makes it finite. U is a random variable on $\mathcal{X}$ and V is a random variable on $\mathcal{Y}$. In this Monty Hall problem we have $p_u = \frac{1}{3}$ as $\mathbb{P} \in \mathcal{P}^*$ implies $\mathbb{P}[U = u] = \frac{1}{3}$ for all $u \in \mathcal{X}$. Lastly, the set $\mathcal{P}^*$ is a subset of $\{\mathbb{P} \, | \, \forall u \in \mathcal{X} : \mathbb{P}[U = u] = p_u\}$.

Take $\mathbb{P}_b, \mathbb{P}_c \in \mathcal{P}^*$ with $\mathbb{P}_i[V = i | U = a] = 1$. Let $i \in \{b, c\}$ be arbitrary, then

$$\mathbb{P}_i[V = i] = \sum_{u=1}^{3} \mathbb{P}_i[V = i | U = u] \, \mathbb{P}_i[U = u] \tag{4.43}$$

$$= \frac{1}{3} (2 \cdot 1 + 0) = \frac{2}{3}. \tag{4.44}$$

Thus the vectors $(\mathbb{P}_i[V = v])_{v \in \mathcal{Y}}$ become $\left(\frac{1}{3}, \frac{2}{3}\right)$ and $\left(\frac{2}{3}, \frac{1}{3}\right)$, which are linearly independent.

Now we can directly apply Theorem 4.1.1 to conclude the proof. □

Thus the distribution stating that door a has probability $\frac{1}{3}$ of having the car is safe for $\mathbb{1}_{\{U=a\}} | [V]$, in other words marginally valid. In this case safety for $\mathbb{1}_{\{U=a\}} | [V]$ implies a few other results with *pivotal safety*. Before we can introduce pivotal safety, we first need to introduce the *pivot*. This definition is taken from [Gr 18b].

Definition 4.3.2 (Discrete pivot). Let U and V be random variables on Ω. Assume Ω is countable. A random variable U' is a *discrete pivot* for $U|V$ if

1. $(U, V) \rightsquigarrow U'$ holds, thus there is a function f with $U' = f(U, V)$.

2. For each fixed $v \in \text{range}(V)$ the function

$$f_v \colon \text{range}(U|V = v) \to \text{range}(U'), \qquad u \mapsto f(u, v) \tag{4.45}$$

   is injective.

3. All $\mathbb{P} \in \mathcal{P}^*$ agree on U', thus for all $\mathbb{P}_1, \mathbb{P}_2 \in \mathcal{P}^*$ we have $\mathbb{P}_1(U') = \mathbb{P}_2(U')$.

Pivot U' is *simple* if f_v is a bijection for all $v \in \text{range}(V)$.

Since we will not consider continuous pivots in this thesis, we just state that U' is a *pivot* for $U|V$. Now we can define pivotal safety, where the definition is taken from [Gr 18b].

Definition 4.3.3 (Pivotal safety). Let U and V be as in Definition 4.3.2 and let $\tilde{\mathbb{P}}$ be an arbitrary distribution on Ω. If V has full support under $\tilde{\mathbb{P}}$, thus $\text{supp}_{\tilde{\mathbb{P}}}(V) = \text{range}(V)$, and U' is a discrete pivot such that $\tilde{\mathbb{P}}$ is safe for $U'|[V]$, then $\tilde{\mathbb{P}}$ is *pivotally safe for $U|V$ with pivot U'*.

Now we can prove that $\mathbb{1}_{\{U=a\}}$ is a pivot for $U|V$ and then $\tilde{\mathbb{P}}$ from Proposition 4.3.1 being safe for $U|V$ with pivot $\mathbb{1}_{\{U=a\}}$ immediately follows.

Proposition 4.3.4. *Let $\mathcal{X}$, $\mathcal{Y}$, Ω, U, V, $\mathcal{P}^*$ and $\tilde{\mathbb{P}}$ be as in Proposition 4.3.1. The random variable $\mathbb{1}_{\{U=a\}}$ is a pivot for $U|V$ and $\tilde{\mathbb{P}}$ is pivotally safe for $U|V$ with pivot $\mathbb{1}_{\{U=a\}}$.*

Proof. The proof is a mere check of Definitions 4.3.2 and 4.3.3. We will start with $\mathbb{1}_{\{U=a\}}$ being a pivot for $U|V$ by checking the three requirements of Definition 4.3.2.

1. Take as function $f(U, V) = \mathbb{1}_{\{U=a\}}$. This is a well-defined function, therefore $(U, V) \rightsquigarrow U'$.

2. Note that $\text{range}(V) = \{b, c\}$. Take $v = b$, then $\text{range}(U|V = b) = \{a, c\}$. We have $f_b(a) = f(a, b) = 1$ and $f_b(c) = f(c, b) = 0$, thus f_b is a bijection. Taking $v = c$ we get $f_c(a) = 1$ and $f_c(b) = 0$, thus f_c is a bijection as well.

40

3. Take $\mathbb{P}_1, \mathbb{P}_2 \in \mathcal{P}^*$ arbitrarily. Note that range$(\mathbb{1}_{\{U=a\}}) = \{0, 1\}$. Therefore

$$\mathbb{P}_1[U \neq a] = \mathbb{P}_1[U \in \{2, 3\}] = \frac{2}{3} = \mathbb{P}_2[U \in \{2, 3\}] = \mathbb{P}_2[U \neq a], \quad (4.46)$$

$$\mathbb{P}_1[U = a] = \mathbb{P}_1[U = 1] = \frac{1}{3} = \mathbb{P}_2[U = 1] = \mathbb{P}_2[U = a], \quad (4.47)$$

thus we have $\mathbb{P}_1[\mathbb{1}_{\{U=a\}}] = \mathbb{P}_2[\mathbb{1}_{\{U=a\}}]$.

Since we have proven that f_v is a bijection for all $v \in \text{range}(V)$, the pivot $\mathbb{1}_{\{U=a\}}$ is a simple pivot for $U|V$.

Theorem 4.1.1 states that $\tilde{\mathbb{P}}$ with $\tilde{\mathbb{P}}[U = a|V = v] = \frac{1}{3}$ for all $v \in \mathcal{Y}$ is safe for $\mathbb{1}_{\{U=a\}}|[V]$. This $\tilde{\mathbb{P}}$ has full support under V, as for all $v \in \mathcal{Y}$ the statement $\tilde{\mathbb{P}}[U = a|V = v] > 0$ implies $\tilde{\mathbb{P}}[V = v] > 0$ by definition of traditional conditional probability. Therefore by Definition 4.3.3 probability measure $\tilde{\mathbb{P}}$ is pivotally safe for $U|V$ with pivot $\mathbb{1}_{\{U=a\}}$. $\qquad\square$

Pivotal safety comes with a nice result. Following [Gr 18b], introduce the random variable $\tilde{p}(U|V)$ by

$$\tilde{p}(U|V)(\omega) = \tilde{\mathbb{P}}[U = U(\omega)|V = V(\omega)] \qquad (4.48)$$

for all $\omega \in \Omega$. This random variable maps $\omega \in \Omega$ to the probability of a single outcome $\tilde{\mathbb{P}}[U = U(\omega)|V = V(\omega)]$. Now [Gr 18b] stated the following theorem on pivotal safety.

Theorem 4.3.5. *Let Ω be countable and U, V be random variables on Ω. Let $\tilde{\mathbb{P}}$ be an arbitrary distribution on Ω. Suppose that for all $v \in \text{range}(V)$ there are no two $u_1, u_2 \in \text{range}(U|V = v)$ with $\tilde{\mathbb{P}}[U = u_1|V = v] = \tilde{\mathbb{P}}[U = u_2|V = v]$. The following statements are equivalent:*

1. *$\tilde{\mathbb{P}}$ is safe for $\tilde{p}(U|V)|[V]$.*

2. *$\tilde{\mathbb{P}}$ is pivotally safe for $U|V$ with simple pivot $U' = \tilde{p}(U|V)|[V]$.*

3. *$\tilde{\mathbb{P}}$ is pivotally safe for $U|V$ for some simple pivot U''.*

Proof. The proof can be found in [Gr 16]. $\qquad\square$

Let in our case $\tilde{\mathbb{P}}$ be as in Proposition 4.3.4. We only defined $\tilde{\mathbb{P}}[U = a|V = v]$ for all $v \in \mathcal{Y}$, not $\tilde{\mathbb{P}}[U = u|V = v]$ with $(u, v) \in \{b, c\} \times \mathcal{Y}$. Thus to actually write down the outcomes of $\tilde{p}(U|V)$ with their respective probabilities, we need to put $p_u^v \in \left[0, \frac{2}{3}\right] \setminus \left\{\frac{1}{3}\right\}$ with $p_b^v + p_c^v = \frac{2}{3}$ for all $(u, v) \in \{b, c\} \times \mathcal{Y}$ and apply

$$\tilde{\mathbb{P}}[U = u|V = v] = p_u^v. \qquad (4.49)$$

Now $\tilde{\mathbb{P}}[U = a|V = v]$ is fixed to be $\frac{1}{3}$. No $u \in \{b, c\}$ has $\tilde{\mathbb{P}}[U = u|V = v] = \frac{1}{3}$ as well, since we excluded $\frac{1}{3}$ from the set p_u^v could be picked from. Since $\tilde{\mathbb{P}}$ is pivotally safe for $U|V$ with pivot $\mathbb{1}_{\{U=a\}}$, Theorem 4.3.5 states that $\tilde{\mathbb{P}}$ is pivotally safe for $U|V$ with simple pivot $\tilde{p}(U|V)$ as well.

Summarizing, since $\tilde{\mathbb{P}}$ is pivotally safe for $\mathbb{1}_{\{U=a\}}$ with simple pivot $\mathbb{1}_{\{U=a\}}$ and since we can create $\tilde{p}(U|V)$ to be a simple pivot for $U|V$ as well, we immediately obtain pivotal safety for $U|V$ with simple pivot $\tilde{p}(U|V)$. We also obtain

an infinite family of simple pivots $\tilde{p}(U|V)$ for which $\tilde{\mathbb{P}}$ is safe, as for example $\tilde{\mathbb{P}}[U = b|V = c]$ is not uniquely defined. We have seen that it is safe to assume that the car is behind door a with a probability of $\frac{1}{3}$ and by some special properties of the Monty Hall game, specifically that $\mathbb{1}_{\{U=a\}}$ is a simple pivot for $U|V$, we see that $\tilde{\mathbb{P}}$ is safe for $U'|[V]$ with $U' = \tilde{p}(U|V)$ being another simple pivot constructed by $\tilde{\mathbb{P}}$.

Accuracy

Suppose you are playing the Monty Hall game, which strategy optimizes the probability of winning the car? The answer is that you must always switch for a probability of $\frac{2}{3}$ to win the car, which follows from applying Theorem 4.1.6.

Proposition 4.3.6. *Let $\mathcal{X}$, $\mathcal{Y}$, U, V and $\mathcal{P}^*$ be as in the dice game. Consider the random variable $\mathbb{1}^*_{\{U=a\}} \sim 0$, then this $\mathbb{1}^*_{\{U=a\}}$ has accuracy*

$$\mathrm{acc}^{\mathbb{P}}\left(\mathbb{1}^*_{\{U=a\}}\right) = \frac{2}{3} \tag{4.50}$$

for all $\mathbb{P} \in \mathcal{P}^*$. There is no random variable W with $\mathrm{acc}^{\mathbb{P}}(W) > \mathrm{acc}^{\mathbb{P}}(\mathbb{1}^*_{\{U=a\}})$ for a $\mathbb{P} \in \mathcal{P}^*$.

Proof. We want to apply Theorem 4.1.6. Firstly in this case $\mathcal{X}' = \{a\} \subset \mathcal{X}$ is non-empty. Secondly, we have $\sum_{u \in \mathcal{X}'} p_u = p_a = \frac{1}{3} < \frac{1}{2}$. Theorem 4.1.6 states that when $\mathbb{1}^*_{\{U=a\}}$ is distributed as $\mathbb{1}^*_{\{U=a\}} \sim 0$, the accuracy for guessing $\mathbb{1}_{\{U=a\}}$ is optimized with value

$$\mathrm{acc}\left(\mathbb{1}^*_{\{U=a\}}\right) = \max\left\{\frac{1}{3}, \frac{2}{3}\right\} = \frac{2}{3}. \tag{4.51}$$

$\square$

As in the dice game, the information of the game master does not lead to a higher probability of winning the car. When always switching doors, the probability of winning the car becomes $\frac{2}{3}$ and this is the optimal probability you can have.

Final remarks

The Monty Hall problem applied Theorems 4.1.1 and 4.1.6 in the same manner as the dice game. In the setting of the Monty Hall problem, it is safe to say that the car has probability $\frac{1}{3}$ to be behind the originally chosen door. Always switching therefore yields to an optimal probability of $\frac{2}{3}$ for winning the car.

Since in the case of the Monty Hall problem the random variable $\mathbb{1}_{\{U=a\}}$ is a simple pivot for $U|V$, we can pick $\tilde{\mathbb{P}}$ to construct a family of other simple pivots U' that are safe for $U'|[V]$. This is not possible in the dice game, as there f_v as in Definition 4.3.2 with $v \in \{\{1, 2, 3, 4\}, \{3, 4, 5, 6\}\}$ is not injective.

When pivots are not considered, the Monty Hall problem is essentially equal to the dice game. Here the paradoxical nature of the Monty Hall problem stems from the fact that most use $\Omega = \{a, b, c\}$ when first trying to solve the problem. However, in that case we cannot create a partition to apply traditional conditional probability or a σ-algebra for measure-theoretic conditional probability.

This problem can only be spotted when performing conditional probability correctly, thus with stating the accompanying sub-σ-algebra. Then it becomes clear that Ω must be extended with an observation space $\mathcal{Y} = \{b, c\}$. This makes the Monty Hall problem yet another argument as in why the σ-algebra must be given when performing conditional probability.

4.4 Boy or girl problem

The last problem we will study in this chapter is the boy or girl problem[2]. Suppose you are approached by a stranger and you two get in conversation. The stranger tells you that he has two children and at least one of them is a girl. You want to know what the probability is that he has two daughters.

The paradox arises when one misinterprets the setting. Suppose you ask the stranger if he has at least one girl. If he says 'yes', the probability of him having two girls is $\frac{1}{3}$, as he can also have a boy and a girl and you do not know whether the earlier mentioned girl is his oldest child. This calculation is not controversial, as his answer 'yes' implies the set of possibilities $\{bg, gb, gg\}$ and is answer 'no' implies the set of possibilities $\{bb\}$. These sets do not overlap, thus traditional probability theory can be applied without any problems.
Suppose now the stranger tells you out of the blue that he has two children and at least one of them is a girl. Now the setting changes, as he could also have told you that at least one of his children is a boy. If he said that he has a girl, the set of possibilities is $\{bg, gb, gg\}$ and if he said that he has a boy, the set of possibilities is $\{bg, gb, bb\}$. Now there is an overlap of $\{bg, gb\}$ and traditional probability theory cannot be applied any-more. The probability of the stranger having two girls given he has at least one is now dilated between $\frac{1}{3}$ and 1, depending on the willingness of the stranger to reveal he has a girl. This dilation can be calculated in the same manner as is done for the Monty Hall problem in Equation 4.41.

We will now construct our model. Let $\mathcal{X} = \{bb, bg, gg\}$ be our outcome space and let $\mathcal{Y} = \{b, g\}$ be our observation space. Then $\Omega = \mathcal{X} \times \mathcal{Y}$ becomes our sample space with 2^Ω as its accompanying σ-algebra. We do not care about whether the girl is older than the boy in the event that the stranger has a boy and a girl, thus we remove the outcome gb. Let U be a random variable on $\mathcal{X}$ denoting the children of the stranger and V be a random variable on $\mathcal{Y}$ denoting whether the stranger told you he as at least one boy, event $\{V = b\}$, or at least one girl, event $\{V = g\}$. All probabilities $\mathbb{P}$ on Ω must be a member of

$$\mathcal{P}^* = \left\{ \mathbb{P} \, \middle| \, \begin{array}{l} \mathbb{P}[U = bb] = \frac{1}{4}, \mathbb{P}[U = bg] = \frac{1}{2}, \\ \mathbb{P}[V = g | U = bb] = \mathbb{P}[V = b | U = gg] = 0 \end{array} \right\}. \qquad (4.52)$$

Traditional conditional probability

The analysis of the problem using traditional conditional probability will be very short here, as it is essentially the same analysis as in sections 4.2 and 4.3. Let $\mathbb{P} \in \mathcal{P}^*$ be arbitrary, then there is a $p \in [0, 1]$ with $\mathbb{P}[V = g | U = bg] = p$.

---

[2]Bronnen invoeren

Using Bayes' rule we get

$$\mathbb{P}[U = bg|V = g] = \frac{\mathbb{P}[V = g|U = bg]\,\mathbb{P}[U = bg]}{\sum_{u \in \mathcal{X} } \mathbb{P}[V = g|U = u]\,\mathbb{P}[U = u]} \tag{4.53}$$

$$= \frac{p \cdot \frac{1}{2}}{0 \cdot \frac{1}{4} + p \cdot \frac{1}{2} + 1 \cdot \frac{1}{4}} = \frac{2p}{2p + 1}, \tag{4.54}$$

thus the probability of him having two girls given the stranger told you he has at least one daughter dilates from $\frac{1}{3}$ to 1. Therefore traditional conditional probability does not give a single answer.

Safe probability

Since U is not real-valued, we cannot calculate $\mathbb{E}_{\mathbb{P}}[U]$ for any $\mathbb{P} \in \mathcal{P}^*$. Thus there is no notion of safety for $U|V$. We can speak of safety for $\mathbb{1}_{\{U=bg\}}|V$, as the event $\{U = bg\}$ arises both in conditioning on $\{V = b\}$ and on $\{V = g\}$.

Proposition 4.4.1. *Let $\mathcal{X}$, $\mathcal{Y}$, Ω, U, V and $\mathcal{P}^*$ be as in the boy or girl problem. Let $\tilde{\mathbb{P}}$ be a distribution on Ω with full support on V. Let $u \in \mathcal{X}$, then the following are equivalent:*

1. *For all $v \in \mathcal{Y}$ we have $\tilde{\mathbb{P}}[U = bg|V = v] = \frac{1}{2}$ in the case of $u = bg$ and $\tilde{\mathbb{P}}[U = u|V = v] = \frac{1}{4}$ when $u \in \{bb, gg\}$.*

2. *$\tilde{\mathbb{P}}$ is safe for $\langle \mathbb{1}_{\{U=u\}}\rangle | \langle V\rangle$.*

3. *$\tilde{\mathbb{P}}$ is safe for $\mathbb{1}_{\{U=u\}}\,|[V]$.*

Proof. The proof is a direct application of Theorem 4.1.1 and an almost direct copy of the proof of Proposition 4.3.1. We will nevertheless state the full proof here.

The set $\mathcal{X}$ is of size 3 and therefore countable and the set $\mathcal{Y}$ is of size 2, which makes it finite. U is a random variable on $\mathcal{X}$ and V is a random variable on $\mathcal{Y}$. In this boy or girl problem we have $p_{bg} = \frac{1}{2}$ and $p_{bb} = p_{gg} = \frac{1}{4}$ by $\mathbb{P} \in \mathcal{P}^*$. Lastly, the set $\mathcal{P}^*$ is a subset of $\{\mathbb{P} \,|\forall u \in \mathcal{X} : \mathbb{P}[U = u] = p_u\}$.

Take $\mathbb{P}_b, \mathbb{P}_g \in \mathcal{P}^*$ with $\mathbb{P}_i[V = i|U = bg] = \frac{3}{4}$. Let $i \in \{b, g\}$ be arbitrary, then

$$\mathbb{P}_i[V = i] = \sum_{u \in \mathcal{X} } \mathbb{P}_i[V = i|U = u]\,\mathbb{P}_i[U = u] \tag{4.55}$$

$$= \frac{1}{2} \cdot \frac{3}{4} + \frac{1}{4} \cdot \frac{1}{4} + 0 \cdot \frac{1}{4} = \frac{7}{16}. \tag{4.56}$$

Thus the vectors $(\mathbb{P}_i[V = v])_{v \in \mathcal{Y}}$ become $\left(\frac{7}{16}, \frac{9}{16}\right)$ and $\left(\frac{9}{16}, \frac{7}{16}\right)$, which are linearly independent.

Now we can directly apply Theorem 4.1.1 to conclude the proof. $\square$

Therefore the distribution, for which the probability is $\frac{1}{2}$ that the stranger has two girls after stating he has at least one girl, is safe for $\mathbb{1}_{\{U=bg\}}\,|[V]$. Moreover, $\tilde{\mathbb{P}}$ from Proposition 4.4.1 is pivotally safe for $U|V$ with pivot $\mathbb{1}_{\{U=bg\}}$.

Proposition 4.4.2. *Let $\mathcal{X}$, $\mathcal{Y}$, Ω, U, V, $\mathcal{P}^*$ and $\tilde{\mathbb{P}}$ be as in Proposition 4.4.1. The random variable $\mathbb{1}_{\{U=bg\}}$ is a pivot for $U|V$ and $\tilde{\mathbb{P}}$ is pivotally safe for $U|V$ with pivot $\mathbb{1}_{\{U=bg\}}$.*

Proof. The proof is equal to the proof of Proposition 4.3.4. $\square$

We will not dive deeper in pivots here, as it will become be a repetition of Section 4.3. The problem is equal to the Monty Hall problem apart from the distribution of U in $\mathcal{P}^*$, thus for a more thorough analysis the reader is referred to Section 4.3.

Accuracy

In this problem the optimal accuracy for guessing $\mathbb{1}_{\{U=bg\}}$ correctly becomes interesting. When playing this game multiple times, it does not matter the frequency of guessing that he has a boy against having a girl. According to Proposition 4.4.1 distribution $\tilde{\mathbb{P}}$ with $\tilde{\mathbb{P}}[U = bg|V = v] = \frac{1}{2}$ for all $v \in \mathcal{Y}$ is safe for $\mathbb{1}_{\{U=bg\}}|[V]$, thus Theorem 4.1.6 gives us that random variable $\mathbb{1}^*_{\{U=bg\}}$ distributed like $\mathbb{1}^*_{\{U=bg\}} \sim \mathrm{Ber}(p)$ optimizes the accuracy of guessing $\mathbb{1}_{\{U=bg\}}$ correctly.

Proposition 4.4.3. *Let $\mathcal{X}$, $\mathcal{Y}$, U, V and $\mathcal{P}^*$ be as in the boy or girl problem. Consider the random variable $\mathbb{1}^*_{\{U=bg\}} \sim \mathrm{Ber}(p)$ distributed as the Bernoulli distribution with parameter $p \in [0,1]$, then this $\mathbb{1}^*_{\{U=bg\}}$ has accuracy*

$$\mathrm{acc}^{\mathbb{P}}\left(\mathbb{1}^*_{\{U=bg\}}\right) = \frac{1}{2} \tag{4.57}$$

for all $\mathbb{P} \in \mathcal{P}^*$. There is no random variable W with $\mathrm{acc}^{\mathbb{P}}(W) > \mathrm{acc}^{\mathbb{P}}(\mathbb{1}^*_{\{U=bg\}})$ for a $\mathbb{P} \in \mathcal{P}^*$.

Proof. We want to apply Theorem 4.1.6. Firstly in this case $\mathcal{X}' = \{bg\} \subset \mathcal{X}$ is non-empty. Secondly, we have $\sum_{u \in \mathcal{X}'} p_u = p_{bg} = \frac{1}{2}$. Then Theorem 4.1.6 states that $\mathbb{1}^*_{\{U=bg\}} \sim \mathrm{Ber}(p)$ with $p \in [0,1]$ arbitrary optimizes the accuracy for guessing $\mathbb{1}_{\{U=bg\}}$ with

$$\mathrm{acc}\left(\mathbb{1}^*_{\{U=bg\}}\right) = \max\left\{\frac{1}{2}, \frac{1}{2}\right\} = \frac{1}{2}. \tag{4.58}$$

$\square$

We can therefore just guess randomly whether the stranger has a boy and girl or not. On average we will be correct 50% of the times we find ourselves in this situation, which is the best we can do.

4.4.1 Boy or girl problem 2.0

There is a second variant of the boy or girl problem stated in a lecture by Peter Grünwald [Gr 18a]. He stated a revised boy or girl problem as such:

1. A person is picked uniformly at random from a population and a conversation is started.

2. You ask if at least one of his children is a boy. Suppose his answer is 'yes', then the probability of two boys is $\frac{1}{3}$. As stated in section 4.4, this is not controversial.

3. You ask if he can give you the name of his son. He can choose which son if he has two. The answer is 'Martin'.

4. You ask if Martin is his youngest child.

   Yes: If he answered 'yes', then his youngest child is a boy and the probability of two boys is now $\frac{1}{2}$.

   No: If he answered 'no', then his oldest child is a boy and the probability of two boys is now $\frac{1}{2}$.

The paradoxical nature of this problem is that after learning the name of a son of the stranger, the probability of him having two boys shifts from $\frac{1}{3}$ to $\frac{1}{2}$. However, as the stranger can state any string as name, just learning the name after knowing he has at least a boy must not influence the probability of the stranger having two boys.

This problem is stated more elaborately, but the underlying foundation is equal to the Monty Hall or the original boy or girl problem. The outcome space after the stranger answering 'yes' is $\{gb, bb\}$ and the outcome space after the stranger answering 'no' is $\{bg, bb\}$. Now there is an overlap of bb. Furthermore, asking a boy's name and then asking if he is the youngest is essentially the same as asking whether his youngest child is a boy.

Therefore we can adapt our model to this situation. Let $\mathcal{X} = \{bg, gb, bb\}$ be the outcome space, let $\mathcal{Y} = \{y, o\}$ be the observation space and take $\Omega = \mathcal{X} \times \mathcal{Y}$ as sample space. Now we have both bg and gb in $\mathcal{X}$ as we now do care about whether the boy is the oldest of the children. The outcome gg is not included in $\mathcal{X}$ as we assume he answered 'yes' on the question whether he has at least one boy. Let U be a random variable on $\mathcal{X}$ denoting the family composition of the stranger, where $\{U = bg\}$ is the event that his oldest child is a boy. Let V be a random variable on $\mathcal{Y}$ denoting which of the children is a boy for sure, where $\{V = y\}$ is the event that the youngest child is a boy and $\{V = o\}$ is the event that the oldest child is a boy. Assume lastly that the stranger does not lie. All probability distributions $\mathbb{P}$ on $\mathcal{X} \times \mathcal{Y}$ must lie in

$$\mathcal{P}^* = \left\{ \mathbb{P} \middle| \forall u \in \mathcal{X} : \mathbb{P}[U = u] = \frac{1}{3}, \mathbb{P}[V = bg|Y = y] = \mathbb{P}[V = gb|Y = o] = 0 \right\}.$$
(4.59)

For safe probability we can now look at safety for $\mathbb{1}_{\{U=bb\}}$ as bb is an element of the intersection of the outcome spaces conditioned on $\{V = y\}$ or $\{V = o\}$.

Proposition 4.4.4. *Let $\mathcal{X}$, $\mathcal{Y}$, Ω, U, V and $\mathcal{P}^*$ be as in the boy or girl problem 2.0. Let $\tilde{\mathbb{P}}$ be a distribution on Ω. Let $u \in \mathcal{X}$, then the following are equivalent:*

1. *For all $v \in \mathcal{Y}$ we have $\tilde{\mathbb{P}}[U = u|V = v] = \frac{1}{3}$.*

2. *$\tilde{\mathbb{P}}$ is safe for $\langle \mathbb{1}_{\{U=u\}} \rangle | \langle V \rangle$.*

3. *$\tilde{\mathbb{P}}$ is safe for $\mathbb{1}_{\{U=u\}} \, |[V]$.*

Proof. This proof is equal to the proof of Proposition 4.3.1, the model in the boy or girl problem 2.0 is equal to the model in the Monty Hall problem. □

For this problem we can talk about pivotal safety and accuracy, but as the proof above already stated, this revised boy or girl problem is in essence equal to the Monty Hall problem and the original boy or girl problem. Therefore if the reader wants to dive further in this subject, he is kindly referred to Section 4.3.

4.4.2 Final remarks on both problems

Both versions of the boy or girl problem are in essence equal to the dice game and the Monty Hall problem. When in both versions of the boy or girl problem $\Omega = \{bb, bg, gb, gg\}$ is chosen as sample space, then in the first version the conditioned sample spaces become $\{bg, gb, bb\}$ and $\{bg, gb, gg\}$ and in the second version the conditioned sample spaces become $\{bg, bb\}$ and $\{gb, bb\}$. In both versions the conditioned sample spaces cannot form a partition as there is an overlap in the sets, thus traditional conditional probability cannot be applied. The sample space must be extended, which can only be noticed when performing conditional probability correctly, thus with taking the accompanying σ-algebras into account.

In both versions there is no probability of the stranger either having two girls or having two boys. In the first version the probability of having two girls dilates from $\frac{1}{3}$ to 1, in the second version the probability of him having two boys dilates from 0 to $\frac{1}{2}$. The latter can be proven by following the reasoning in Equation 4.41.

When Theorem 4.1.1 is applied, we find that the first version has a safe distribution stating the probability of having a boy and a girl is $\frac{1}{2}$. In the second version, we found out that the safe distribution puts a probability of $\frac{1}{3}$ on him having two boys.

Most importantly, the second version of the paradox is practically equal to the Monty Hall problem. The first version can also be solved using the same techniques as the only difference being the model of probability distributions.

4.5 Conclusion

After going through the dice game, the Monty Hall problem and the boy or girl problem, we see that Theorem 4.1.1 can be applied nicely. When a problem is founded upon a sample space $\Omega = \mathcal{X} \times \mathcal{Y}$ where $\mathcal{X}$ is the countable space of all possible outcomes, $\mathcal{Y}$ is the finite space of all statements the game master can make to the player and the distribution on $\mathcal{X}$ is fixed, we can apply Theorem 4.1.1 to find safe distributions for single probabilities. The model of probability distributions must be large enough and fulfil an important requirement, however this requirement is almost always fulfilled when $\mathcal{P}^*$ has $|\mathcal{Y}|$ distinct distributions. When $\mathcal{P}^*$ has fewer than $|\mathcal{Y}|$ distributions, then you should at first ask yourself whether $\mathcal{P}^*$ is not too restrictive, making the safe distribution prone for overfit. Taking $\mathcal{P}^* = \{\mathbb{P}\}$ as a single distribution makes this $\mathbb{P} \in \mathcal{P}^*$ even safe for $U|V$, however this result is not quite helpful.

Furthermore, seemly distinct problems like the boy or girl problem and the Monty Hall problem are in essence equal. Therefore the studies on both of these

problems can be combined, but with care as both problems do have different probability models.

The last and most important observation of this chapter is that the paradoxical results people find in the Monty Hall problem or the boy or girl problem all arise when $\Omega = \mathcal{X}$, thus when the outcome space is taken as sample space. In this case traditional conditional probability cannot be applied as the resulting outcome spaces after conditioning on the observed events do not form a partition.

This can be seen when the σ-algebra 2^Ω is taken into account. Using loose notation, take $y_1, y_2 \in \mathcal{Y}$ and suppose $x \in \mathcal{X}$ is present in both conditioned spaces $\mathcal{X}|y_1$ and $\mathcal{X}|y_2$. Then for computing the conditional probability of x you need to condition on $\{\{y_1\}, \{y_2\}\}$, which is either not a element of 2^Ω or the probability on $\{\{y_1\}, \{y_2\}\}$ is not known. Thus a sub-σ-algebra needed for conditioning cannot be found.

This mistake is easily spotted when taking the σ-algebra into account when performing conditional probability. Therefore these problems are more examples of why when applying conditional probability the accompanying σ-algebras must always be provided.

Chapter 5

The two envelopes problem

The last paradox we will consider is the two envelope problem. This problem is first stated in 1953 by Kraitchik [Kra53] as the necktie problem. In this problem each person claims to have the finer necktie. A judge must make a decision, where the winner has to give his necktie to the loser. The argument of each contestant is: 'I know what my tie is worth. I may lose it, but I may also win a better one, so the game is to my advantage.' However, a game that is favourable for both players cannot exist. Furthermore, the game is symmetrical in terms of the contestants, thus [Kra53] states that the chance of winning is actually fifty-fifty.

However, [Kra53] only provides a discussion and no rigorous proofs. In 1989 Nalebuff [Nal89] states the two envelope problem with many variants and provides a wide discussion from various sources. We will study one of the variants, in my opinion the one with the most interesting results, of the paradox and show how all paradoxical statements are incorrect. The reasoning will stay valid for various extensions to the problem. Those extensions will be treated briefly in Section 5.6.

Pick the following version of the two envelope problem from [Nal89]. There are two envelopes called A and B. Envelope A contains a fixed amount of money and envelope B contains either value $\frac{1}{2}A$ or value $2A$. The player can choose between envelope A and B and let us say without loss of generality that he chooses envelope A. After choosing, the player is given the choice whether he wants to switch. Reasoning that envelope B contains $\frac{1}{2}A$ with probability $\frac{1}{2}$ and contains $2A$ with probability $\frac{1}{2}$, he computes the expected value of envelope B to be

$$\mathbb{E}[B] = \frac{1}{2} \cdot \frac{1}{2}A + \frac{1}{2} \cdot 2A = \frac{5}{4}A. \tag{5.1}$$

Thus the player always wants to switch when initially picking envelope A. However, by symmetry, if the player chose envelope B, then to his knowledge A contains either $\frac{1}{2}B$ or $2B$, thus he wants to switch as well. It is however not possible to have a greater expected value in the other envelope in all cases, as the total amount of money in the whole game is fixed. Therefore this result is called the *two envelope paradox*. Some even reason that if the player possesses no memory, he wants to switch infinitely many times, believing to gain infinite money.

First we will provide a naive solution that corrects the calculation in Equa-

tion 5.1. That solution will however be incomplete, as the formalized problem in Section 5.2 will point out. In Section 5.3 we will investigate the prior on the lowest value of both envelopes, as that value will remain constant during the game. Safe probability will be applied in Section 5.4, which gives a safe probability distribution on whether to switch or not. This will however not lead to satisfactory results. Section 5.5 introduces Cover switching, which provides a switching strategy that does on average improve the amount of money won. Lastly, an alternative version called the Ali Baba problem will be introduced in Section 5.6 and shortly discussed with many other variants of the two envelope problem. Eventually we will wrap up in Section 5.7 giving concluding remarks.

5.1 The naive solution

First we need to find the flaw in the reasoning of the two envelope paradox. The biggest flaw is in Equation 5.1. This calculation is performed using incorrect probability measures. An in-depth discussion is in 2014 given by O'Brien and Mitchell [OM14]. Here we only discuss the main solution of the problem.

Let A and B the random variables denoting the values of respectively envelopes A and B. Let $x = \min\{A, B\}$ be the lowest value of the envelopes. The envelopes can only obtain two values: x and $2x$. The probability of A having x is $\mathbb{P}[A = x] = \frac{1}{2}$, thus the probability space is

$$\left(\Omega = \{(x, 2x), (2x, x)\}, \Omega = 2^{\mathcal{X}}, \mathbb{P}\right) \tag{5.2}$$

with $\mathbb{P}$ a probability measure such that $\mathbb{P}[\{(x, 2x)\}] = \mathbb{P}[\{(2x, x)\}] = \frac{1}{2}$. The calculation of $\mathbb{E}[B]$ in the statement of the two envelope problem is the following:

$$\mathbb{E}[B] = \frac{A}{2} \mathbb{P}\left[B = \frac{A}{2}\right] + 2A \mathbb{P}[B = 2A]. \tag{5.3}$$

The flaw here as discussed by [Wik19] is that symbol A is used with two different meanings. In $\mathbb{P}\left[B = \frac{A}{2}\right]$ the A is actually the expected value of envelope A given that envelope A contains less than envelope B. In $\mathbb{P}[B = 2A]$ the opposite happens, where actually the expected value of A given that envelope A contains more than envelope B is calculated. Equation 5.1 must therefore be rewritten to

$$\mathbb{E}[B] = \mathbb{E}[B|A < B] \mathbb{P}[A < B] + \mathbb{E}[B|A > B] \mathbb{P}[A > B] \tag{5.4}$$

$$= \frac{1}{2}\left(\mathbb{E}[2A|A < B] + \mathbb{E}\left[\frac{1}{2}A\Big|A > B\right]\right) \tag{5.5}$$

$$= \mathbb{E}[A|A < B] + \frac{1}{4}\mathbb{E}[A|A > B]. \tag{5.6}$$

Now we clearly see that the expectation of A in both cases are different and we are therefore unable to relate $\mathbb{E}[B]$ to $\mathbb{E}[A]$. The following calculation inspired by [SD08] and [OM14] is correct:

$$\mathbb{E}[B] = \mathbb{E}[\mathbb{E}[B|A]] = \mathbb{E}[B|A = x] \mathbb{P}[A = x] + \mathbb{E}[B|A = 2x] \mathbb{P}[A = 2x] \tag{5.7}$$

$$= \frac{1}{2}\left(2x \mathbb{P}[B = 2x|A = x]\right) + \frac{1}{2}\left(x \mathbb{P}[B = x|A = 2x]\right) \tag{5.8}$$

$$= \frac{1}{2} \cdot 2x + \frac{1}{2}x = \frac{3}{2}x, \tag{5.9}$$

We can calculate using the same method that $\mathbb{E}[A] = \frac{3}{2}x$, thus the expected amount of money in either envelope is equal. Switching therefore does not increase one's chances of getting more money.

So why is this section called 'The naive solution'? The problem is that we do not know the value of x. When a player observes the contents of his envelope, he does not know whether he has value x or $2x$. He cannot put equal probability on having x or $2x$, as that will result in a uniform distribution on an infinite set as will be pointed out in Section 5.3. We therefore need to find a method that takes the possible values of x into account.

5.2 The problem formalized

The discussion in the previous section does not completely resolve the two envelope paradox. Two immediate flaws can be found in the discussion of Section 5.1:

1. How is the value x picked? If the value x is drawn from a probability distribution $\mathbb{P}$ on an unbounded set U, then it is not possible for all $x \in U$ that $\mathbb{P}[B = 2x|A = x] = \mathbb{P}\left[B = \frac{1}{2}x \big| A = x\right] = \frac{1}{2}$ holds as that would imply $\mathbb{P}$ being the uniform distribution on U [CU92, RCU93, NŠ17, TJ18]. It is clear that uniform distributions on unbounded sets do not exist.

2. If x is drawn from a probability distribution with $\mathbb{E}[B] = \mathbb{E}[A] = \infty$, then $\mathbb{E}[B] = \mathbb{E}[\mathbb{E}[B|A]]$ is in some cases not valid any-more [TJ18].

Thus we need to not only model which envelope has the higher value, we must also model the lower value of the envelopes. In this section we will formally create the probability space describing the two envelope problem fully. This probability space will be used through the rest of this chapter.

Let $\mathcal{X} = (0, \infty)$ be the space containing all possible lowest values x of the envelopes. Equip $\mathcal{X}$ with $\Sigma_{\mathcal{X}} = \mathcal{B}((0, \infty))$, the Borel σ-algebra on $(0, \infty)$. Let $\mathcal{Y} = \left\{(x, 2x), (2x, x) \in (0, \infty)^2 \big| x \in \mathcal{X}\right\}$ be all possible values in the envelopes. The set $\mathcal{Y}$ is equipped with σ-algebra $\Sigma_{\mathcal{Y}} = \mathcal{B}(\mathcal{Y})$, the Borel-$\sigma$-algebra of the lines $x \mapsto 2x$ and $x \mapsto \frac{1}{2}x$. Let $\Omega = \mathcal{X} \times \mathcal{Y}$ be our sample space equipped with $\Sigma = \Sigma_{\mathcal{X}} \times \Sigma_{\mathcal{Y}}$ as σ-algebra.

Let X be an $\mathcal{X}$-valued continuous random variable denoting the lowest value of the envelopes and Y be a $\mathcal{Y}$-valued random variable denoting the actual values. All probability distributions $\mathbb{P}$ on (Ω, Σ) must be a member of

$$\mathcal{P}^* = \left\{\mathbb{P} \,\bigg|\, \mathbb{P}[Y = (x, 2x)|X = x] = \mathbb{P}[Y = (2x, x)|X = x] = \frac{1}{2}, \mathbb{E}_{\mathbb{P}}[X] < \infty\right\}. \tag{5.10}$$

Let $\pi_i \colon \mathbb{R}^2 \to \mathbb{R}$ with $i \in \{1, 2\}$ be the projection on the i-th coordinate. Let $A = \pi_1(Y)$ be the value in envelope A and $B = \pi_2(Y)$ be the value in envelope B. Assume that without loss of generality the player initially chooses envelope A, then we would like to know the distribution of $B|A$.

The reason why we demand finite expectation, is that in the case of infinite expectation the two envelope problem starts to behave like the St. Petersburg problem [BK95, Bro95, TJ18]. This problem concerns a game where the coin is flipped until it gives heads. Suppose the coin is flipped n times, then \$$2^n$ is paid out. The expected value of the amount of money paid out is infinite while there

is a probability of 87,5% of a payout of only \$7, which was viewed as paradoxical. This problem does not concern conditional probability and will therefore not be treated in this thesis. Interested readers are referred to William Feller, who described in 1950 the problem with an accessible discussion and solution [Fel50] and provided in 1971 a general theorem for similar problems [Fel71].

In our case, a prior with infinite expectation value on X will lead to the paradoxical result of always having to switch, as will be shown in Section 5.3.2. This is another reason to leave such priors out of $\mathcal{P}^*$.

Note that all $\mathbb{P} \in \mathcal{P}^*$ are fixed on Y when the value of X is known. Thus $\mathcal{P}^*$ is essentially implied by all the probability distributions on $\mathcal{X}$ where X has finite expectation. In the next section we will dive deeper into the distributions $\mathbb{P}$ on Ω and their influence to the problem.

5.3 Prior on envelope values

Recall that all probability distributions usable in the two envelope problem must be of the set

$$\mathcal{P}^* = \left\{ \mathbb{P} \,\middle|\, \mathbb{P}[Y = (x, 2x)|X = x] = \mathbb{P}[Y = (2x, x)|X = x] = \frac{1}{2}, \mathbb{E}_{\mathbb{P}}[X] < \infty \right\}. \tag{5.11}$$

As we now do not have more information on the problem, any distribution from $\mathcal{P}^*$ can be the correct one. However, a player playing the two envelope game probably has a suspicion on how X is distributed. This suspicion can be modelled as either a single distribution $\mathbb{P} \in \mathcal{P}^*$ or a subset $P \subset \mathcal{P}^*$ of distributions. Therefore, we can introduce the following definition of a prior.

Definition 5.3.1 (Prior). A distribution $\mathbb{P} \in \mathcal{P}^*$ is called a *prior* on the envelope values and set $P \subset \mathcal{P}^*$ is called a set of priors. The density function of prior $\mathbb{P} \in \mathcal{P}^*$ on $\mathcal{X}$, the lowest value of the envelopes, will be denoted by $f \colon \mathcal{X} \to [0, \infty)$.

Note that, as mentioned in Section 5.2, there are no two $\mathbb{P}_1, \mathbb{P}_2 \in \mathcal{P}^*$ that are almost surely equally distributed on $\mathcal{X}$ but differently on $\mathcal{Y}$, as the conditional distribution $Y|X$ is fixed in $\mathcal{P}^*$. Thus the distribution on Ω of two $\mathbb{P}$'s in $\mathcal{P}^*$ are only almost surely equal when those $\mathbb{P}$'s are almost surely equal on $\mathcal{X}$.

5.3.1 Prior with no information

First consider the setting where no information on $\mathcal{P}^*$ is known. An exhaustive analysis of this case is done by Albers, Kooi and Schaafsma [AKS05]. They give three noteworthy examples for trying to solve the two envelope problem while nothing is known about prior f, however they conclude that the paradox is just hopeless to tackle without being giving more information.

Example 5.3.2 (Conditional expectation). Let $\mathbb{P} \in \mathcal{P}^*$ be arbitrary and suppose $\mathbb{P}[A > B|A = a] = p \in [0, 1]$ holds, thus the probability of receiving the higher envelope given you observed $A = a$ equals p. The expected value of B

given $A = a$ becomes

$$\mathbb{E}_{\mathbb{P}}[B|A = a] = \frac{a}{2}\,\mathbb{P}\left[B = \frac{a}{2}\Big|A = a\right] + 2a\,\mathbb{P}[B = 2a|A = a] \qquad (5.12)$$

$$= \frac{a}{2}p + 2a(1-p) = a\left(2 - \frac{3}{2}p\right). \qquad (5.13)$$

Therefore $\mathbb{E}_{\mathbb{P}}[B|A = a] < a$ holds when $p > \frac{2}{3}$. Thus when it is likely that you have the higher envelope, switching is not advised.

However, note that

$$\mathbb{P}[A > B|A = a] = \mathbb{E}\left[\mathbb{1}_{\{A>B\}}\big|A = a\right] \qquad (5.14)$$

holds. Following [AKS05], the estimation of $\mathbb{P}[A > B|A = a]$ is replaced by the prediction of $\mathbb{1}_{\{A>B\}}$, which is not possible without having more information on $\mathbb{P}$.

Example 5.3.3 (Entropy analysis)**.** Another route is to estimate the entropy of the distribution on the other envelope. If we can maximize this entropy, an estimated prior will follow giving us advice on whether we want to switch.

Let $\mu \in (0, \infty)$ be arbitrary and let $P \subset \mathcal{P}^*$ be such that $\mathbb{E}_{\mathbb{P}}[B] = \mu$ holds for all $\mathbb{P} \in P$. This will be the only assumption we will make on the prior. We can now construct an estimate f' for the prior f such that

$$H(f') = \begin{cases} -\sum_{y=1}^{\infty} f'(y)\log f'(y), & X \text{ discrete,} \\ -\int_{y=0}^{\infty} f'(y)\log f'(y), & X \text{ continuous,} \end{cases} \qquad (5.15)$$

is maximized with $\mathbb{E}[B] = \mu$ as restriction.

Following this route [AKS05] found that in the discrete case, the envelope must be swapped for odd $a \in \mathbb{N} \setminus 2\mathbb{N}$ and not swapped for even $a \in 2\mathbb{N}$. In the continuous case, the player must always swap.

This answer raises a lot of questions, such as why the player should in the discrete case favour values 1001 and 1003 and discard value 1002. No one playing the two envelope problem would be convinced that this is a working strategy. Furthermore, as we will see in Section 5.5, in the continuous case switching strategies exist yielding higher gain for all priors f than always switching the envelope. Thus the continuous strategy using entropy can be called insufficient, as strategies exist that outperform 'always switching'.

There is a third game-theoretic example given by [AKS05], however even Albers et alii state that this perspective is unnecessarily complicated and will not be helpful.

The overall conclusion of [AKS05] is that if no information on the prior is known, a mathematician must be satisfied with the fact that he just cannot solve the problem. There is not enough relevant information to make justified decisions or create useful strategies.

5.3.2 Investigating different priors

Take a distribution $\mathbb{P}$ on Ω and let f be its accompanying prior. We will take a look at various variations of f. We will observe that for certain priors switching will always be advised, but the expectation $\mathbb{E}[X]$ will be infinite as well. When $\mathbb{E}[X]$ is finite, then useful switching strategies will arise.

Example 5.3.4 (Navara and Šindelár (2017))**.** The following example is taken from [NŠ17]. Assume that f is a discrete prior and $\mathbb{P}$ its accompanying distribution on Ω. We want to know how much we expect to gain by switching given value a is in envelope A. Define the gain $G = B - A$. It is easy to see that

$$\mathbb{E}_{\mathbb{P}}[G] = \mathbb{E}_{\mathbb{P}}[B] - \mathbb{E}_{\mathbb{P}}[A] = \frac{3}{2}\,\mathbb{E}_{\mathbb{P}}[X] - \frac{3}{2}\,\mathbb{E}_{\mathbb{P}}[X] = 0. \tag{5.16}$$

The possible values of the gain given $A = a$ are $G = a$ or $G = -\frac{a}{2}$. The probability distribution of $G|A = a$ is according to [NŠ17] given by

$$\mathbb{P}[G = g | A = a] = \begin{cases} \frac{f(a)}{f(a) + f\left(\frac{a}{2}\right)}, & g = a, \\ \frac{f\left(\frac{a}{2}\right)}{f(a) + f\left(\frac{a}{2}\right)}, & g = -\frac{a}{2}. \end{cases} \tag{5.17}$$

The expectation value of $G|A = a$ is then given by

$$\mathbb{E}_{\mathbb{P}}[G | A = a] = \frac{a f(a) - \frac{a}{2} f\left(\frac{a}{2}\right)}{f(a) + f\left(\frac{a}{2}\right)}. \tag{5.18}$$

Consider now the prior $f(2^t) = \frac{q^t}{1-q}$, then three different cases can be distinguished:

$q < \frac{1}{2}$: Take $q < \frac{1}{2}$, then $\mathbb{E}_{\mathbb{P}}[X]$ exists and is equal to $\mathbb{E}_{\mathbb{P}}[X] = (1-q)^{-1}(1-2q)^{-1}$. In this case we have

$$\mathbb{E}_{\mathbb{P}}\left[G \big| A = 2^t\right] = \begin{cases} 2^{t-1}\frac{2q-1}{q+1}, & t \in \mathbb{N}_{\geq 1} \\ 1, & t = 0. \end{cases} \tag{5.19}$$

Therefore with this prior switching is only recommended when one observes a value of 1 in his envelope, as $\mathbb{E}_{\mathbb{P}}[G|A = 2^t] < 0$ is negative for all $t \in \mathbb{N}_{\geq 1}$.

$q = \frac{1}{2}$: Take $q = \frac{1}{2}$, then $\mathbb{E}_{\mathbb{P}}[X]$ does not exist any-more. The expected gain now simplifies to $\mathbb{E}[G|A = 2^t] = \mathbb{1}_{\{0\}}(t)$, thus there is only gain on switching when $A = 1$.

$q > \frac{1}{2}$: Take $q > \frac{1}{2}$, then $\mathbb{E}_{\mathbb{P}}[X]$ does not exist as well. The conditional gain $G|A = 2^t$ is still valid and equals

$$\mathbb{E}_{\mathbb{P}}\left[G \big| A = 2^t\right] = \begin{cases} 2^{t-1}\frac{2q-1}{q+1}, & t \in \mathbb{N}_{\geq 1}, \\ 1, & t = 0, \end{cases} \tag{5.20}$$

which is strictly positive for all $t \in \mathbb{N}$. Thus the player is motivated to always switch, which is a paradoxical result.

The q controlling the prior f can lead to various results, from non-paradoxical advise to only switch when obtaining the smallest value 1 knowing the other envelope must contain 2, to the original paradox of always having to switch no matter the contents of A. The difference with the original paradox is that the paradoxical result of this example is supported by a prior on X, not by wrongly performing the calculations involved with the problem.

Example 5.3.5 (Broome (1995))**.** This example is taken from [Bro95] and is given by many as a lighting example of when a prior on X can have infinite expectation. Consider first the discrete case. Let $f(2^n) = \frac{2^n}{3^{n+1}}$ with $n \in \mathbb{N}_{\geq 0}$ be a prior. We now want to know the expected value in envelope B. The trivial case is $\mathbb{E}[B|A = 1] = 2$, as the only possible value in envelope B is 2 when observing 1.

Suppose we observe a value $a \neq 1$. Then $B = 2a$ is possible if and only if $X = A = a$. When observing envelope $A = a$, then either $X = a$ or $X = \frac{a}{2}$ can hold. Therefore following [Bro95]

$$\mathbb{P}[B = 2a|A = a] = \mathbb{P}\left[X = a\Big| X = a \vee X = \frac{a}{2}\right] \tag{5.21}$$

$$= \frac{\frac{2^n}{3^{n+1}}}{\frac{2^n}{3^{n+1}} + \frac{2^{n-1}}{3^n}} = \frac{2}{5}. \tag{5.22}$$

We have $\mathbb{P}\left[B = \frac{a}{2}\big|A = a\right] = 1 - \mathbb{P}[B = 2a|A = a] = \frac{3}{5}$ as well, giving an expected value of envelope B of

$$\mathbb{E}[B|A = a] = \frac{2}{5} \cdot 2a + \frac{3}{5} \cdot \frac{a}{2} = \frac{11}{10}a. \tag{5.23}$$

The paradoxical conclusion is that the player must always switch, no matter the contents of his envelope. The reason why $\mathbb{E}[B|A = a] > a$ for all $a = 2^n$ with $n \in \mathbb{N}$ is possible, is that $\mathbb{E}[X] = \sum_{n=1}^{\infty} \left(\frac{4}{3}\right)^n = \infty$ is infinite.

Until now we have only considered discrete distributions, but paradoxical conclusions can result from continuous distributions as well. Let f be a continuous prior on X, then the conditional expectation on $B|A = a$ is

$$\mathbb{E}[B|A = a] = a\frac{8f(a) + f\left(\frac{a}{2}\right)}{4f(x) + 2f\left(\frac{a}{2}\right)} \tag{5.24}$$

as is derived by [Bro95]. Take density function $f(x) = (x+1)^{-2}$ with $x \in (0, \infty)$, then the conditional expectation of $B|A = a$ becomes

$$\mathbb{E}[B|A = a] = a\frac{2(a + 2)^2 + (a + 1)^2}{(a + 2)^2 + 2(a + 1)^2} > a \tag{5.25}$$

for all $a \in (0, \infty)$. Here the situation that switching is advised arises as well. Note that the expectation value of the prior $\mathbb{E}[X] = \int_0^\infty \frac{x}{(x+1)^2} dx = \infty$ is infinite, explaining why $\mathbb{E}[B|A = a] > a$ for all $a \in (0, \infty)$ is again possible.

Take another look at the gain G given $A = a$, defined in Example 5.3.4. When considering the conditional expectation $\mathbb{E}[G|A = a]$, one can quantify when a distribution yields to paradoxical results. For discrete priors, $\mathbb{E}[G|A = a]$ equals [NŠ17, TJ18]

$$\mathbb{E}[G|A = a] = \frac{af(a) - \frac{a}{2}f\left(\frac{a}{2}\right)}{f(a) + f\left(\frac{a}{2}\right)} \tag{5.26}$$

and where it gives rise to the following definition.

Definition 5.3.6 (Discrete paradoxical distribution)**.** A discrete prior f on X is called *paradoxical* if $f(x) > \frac{1}{2}f\left(\frac{x}{2}\right)$ holds for all $x \in \text{supp}_{\mathbb{P}}(X)$.

Lemma 5.3.7. *Let f be a discrete paradoxical prior on X, then the expectation value $\mathbb{E}[X] = \infty$ is infinite.*

Proof. This proof is taken from [TJ18]. Let x be a possible realization of X, then $2^k x$ with $k \in \mathbb{N}$ are possible realizations of X as well by the setting of the two envelope problem. Note that $2^{k+1} x f(2^{k+1} x) > 2^k x f(2^k x)$ holds for all $x \in \mathbb{N}$, implying an unbounded $\mathbb{E}[X] = \sum_{k=0}^{\infty} k f(k) > \sum_{k=0}^{\infty} 2^k x f(2^k x)$ as all terms $2^k x f(2^k x)$ will increase in k. $\qquad\square$

The requirement for a continuous paradoxical distribution is slightly different. Brams, Kilgour and Blachman [BK95, BCU96] calculated when the conditional expectation $\mathbb{E}[B|A = a] > a$ holds for continuous priors.

Definition 5.3.8 (Continuous paradoxical distribution)**.** A continuous prior f on X is called *paradoxical* if $f(x) > \frac{1}{4} f\left(\frac{x}{2}\right)$ holds for all $x \in \mathrm{supp}_{\mathbb{P}}(X)$.

As mentioned in Section 5.2, we will only consider non-paradoxical distributions from now on. It is now clear that, when having full information, paradoxical distributions still yield unexpected results. This is contributed to the expectation value of the prior being undefined, not to a wrong interpretation of the problem. When having full information non-paradoxical priors yield well-defined and useful switching strategies.

5.3.3 Optimal solution

Suppose a non-paradoxical prior on X is known, does an optimal switching strategy exist? This question is asked and answered by [CU92] and further discussed by [BCU93, RCU93, RCU94, BCU96]. Originally, [CU92] proposed a solution for all general priors, however [BCU96] pointed out this is incorrect for continuous distributions. Here we will combine the discussion and provide a resolution.

Discrete priors

Let f be a discrete prior on X with distribution $\mathbb{P} \in \mathcal{P}^*$, then [CU92] states that

$$\mathbb{P}\left[X = a | A = a\right] = \frac{\mathbb{P}[A = a | X = a] f(a)}{\mathbb{P}[A = a | X = a] f(a) + \mathbb{P}\left[A = a | X = \frac{a}{2}\right] f\left(\frac{a}{2}\right)} \quad (5.27)$$

$$= \frac{f(a)}{f(a) + f\left(\frac{a}{2}\right)}, \quad (5.28)$$

$$\mathbb{P}\left[X = \frac{a}{2} \Big| A = a\right] = \frac{\mathbb{P}\left[A = a | X = \frac{a}{2}\right] f\left(\frac{a}{2}\right)}{\mathbb{P}[A = a | X = a] f(a) + \mathbb{P}[A = a | X = 2a] f(2a)} \quad (5.29)$$

$$= \frac{f\left(\frac{a}{2}\right)}{f(a) + f\left(\frac{a}{2}\right)}. \quad (5.30)$$

We can now compute $\mathbb{E}[B|A = a]$ to be

$$\mathbb{E}[B|A = a] = \frac{a}{2} \mathbb{P}\left[X = \frac{a}{2} \Big| A = a\right] + 2a \, \mathbb{P}\left[X = a | A = a\right] \quad (5.31)$$

$$= \frac{a f\left(\frac{a}{2}\right) + 4a f(a)}{2 f(a) + 2 f\left(\frac{a}{2}\right)}. \quad (5.32)$$

Obviously when $\mathbb{E}[B|A = a] > a$ holds, switching is advised. This happens when $f\left(\frac{a}{2}\right) < 2f(a)$.

Does a discrete prior f exist for which the requirement $\mathbb{E}[B|A = a] \leq a$ holds for all $a \in \operatorname{supp}_{\mathbb{P}}(A)$, thus where the player must never switch? The answer is no, as the following proposition from [BK95] points out.

Proposition 5.3.9 (Brams and Kilgour (1995))**.** *There is no discrete prior f on X for which $\mathbb{E}[B|A = a] \leq a$ holds for all $a \in \operatorname{supp}_{\mathbb{P}}(A)$.*

Proof. The proof is taken from [BK95]. Assume f is a prior with $\mathbb{E}[B|A = a] \leq a$ for all $a \in \operatorname{supp}_{\mathbb{P}}(A)$. In order for $\mathbb{E}[B|A = a] \leq a$ we must have $f\left(\frac{a}{2}\right) \geq 2f(a)$. Let $a \in \operatorname{supp}_{\mathbb{P}}(A)$ be such that $f(a) > 0$, then $f(a) \geq 2^n f(2^n a)$ holds for all $n \in \mathbb{Z}$, leading to $f(a) \to 0$ as $n \to -\infty$. We cannot have $f(a)$ being arbitrarily close to zero and strictly positive at the same time, rendering the assumption of the existence of f false. $\square$

On the contrary, does a discrete prior f exist with $\mathbb{E}[B|A = a] > a$ for all $a \in \operatorname{supp}_{\mathbb{P}}(A)$? Then $f\left(\frac{a}{2}\right) > f(a)$ must hold for all $a \in \operatorname{supp}_{\mathbb{P}}(A)$, which is achieved for all paradoxical distributions since $\operatorname{supp}_{\mathbb{P}}(X) \subset \operatorname{supp}_{\mathbb{P}}(A)$. Various examples can be found in [BCU93, Lin94, Bro95] and previously in Section 5.3.2, with $f(2^n) = \frac{2^n}{3^{n+1}}$ as the most common one. In this case a new paradox arises, where switching is always advised.

Note that Christensen and Utts do not agree with this occurrence as being a paradox as stated in [BCU93], as $\mathbb{E}[B|A = a]$ is still finite and can be compared with a.

Continuous case

Let f be a continuous prior on X with distribution $\mathbb{P} \in \mathcal{P}^*$. First [CU92] states that

$$\mathbb{E}[B|A = a] = \frac{af\left(\frac{a}{2}\right) + 4af(a)}{2f(a) + 2f\left(\frac{a}{2}\right)} \tag{5.33}$$

holds for continuous priors as well, however [BCU96] points out this is not correct.

Take a look at the following analysis from [BK95]. Let $Y = \frac{X}{2}$ and let g be a probability density function on Y. Then we have $\mathbb{P}[X \leq x] = \mathbb{P}\left[Y \leq \frac{x}{2}\right]$, thus $F(x) = G\left(\frac{x}{2}\right)$ where F and G are the cumulative density functions of X and Y respectively. Differentiating for the density functions yields $f(x) = \frac{1}{2}g\left(\frac{x}{2}\right)$. Now, analogous to the discrete case, we have

$$\mathbb{P}[X = a|A = a] = \frac{f(x)}{f(x) + g(x)} = \frac{f(x)}{f(x) + \frac{1}{2}f\left(\frac{x}{2}\right)}, \tag{5.34}$$

$$\mathbb{P}\left[X = \frac{a}{2}\Big|A = a\right] = \frac{g(x)}{f(x) + g(x)} = \frac{\frac{1}{2}f\left(\frac{x}{2}\right)}{f(x) + \frac{1}{2}f\left(\frac{x}{2}\right)}. \tag{5.35}$$

The expectation of $B|A = a$ now becomes

$$\mathbb{E}[B|A = a] = \frac{a}{2}\mathbb{P}\left[X = \frac{a}{2}\Big|A = a\right] + 2a\,\mathbb{P}[X = a|A = a] \tag{5.36}$$

$$= \frac{af\left(\frac{a}{2}\right) + 8af(a)}{4f(a) + 2f\left(\frac{a}{2}\right)}. \tag{5.37}$$

This expectation value is different from the version in [BCU96], as they made a mistake mixing up the probabilities while calculating the expectation value. Nevertheless their conclusion that now $4f(a) > f\left(\frac{a}{2}\right)$ must hold in order to get $\mathbb{E}[B|A = a] > a$ is still correct.

The same proposition as in the discrete case that $\mathbb{E}[B|A = a] > a$ must happen for at least one $a \in \mathrm{supp}_{\mathbb{P}}(A)$ is valid here as well.

Proposition 5.3.10 (Brams and Kilgour (1995)). *There is no continuous prior f on X for which $\mathbb{E}[B|A = a] \leq a$ holds for all $a \in \mathrm{supp}_{\mathbb{P}}(A)$.*

Proof. This proposition is stated and proven as Theorem 2 in [BK95]. $\qquad\square$

Examples for which $\mathbb{E}[B|A = a] > a$ holds for all $a \in \mathrm{supp}_{\mathbb{P}}(A)$ can be found in [BCU96, Bro95, BK95] and previously in Section 5.3.2, with $f(a) = (a+1)^{-2}$ for $a \in (0, \infty)$ being one of the most common example [Bro95].

Summarizing switching strategies

To summarize, let f be a non-paradoxical prior on X with distribution $\mathbb{P} \in \mathcal{P}^*$. Suppose the player observes $a \in \mathrm{supp}_{\mathbb{P}}(A)$. The following strategies optimize the values won by the player:

Discrete: When f is discrete, the player must switch his envelope when

$$f(a) > \frac{1}{2}f\left(\frac{a}{2}\right). \tag{5.38}$$

Continuous: When f is continuous, the player must switch his envelope when

$$f(a) > \frac{1}{4}f\left(\frac{a}{2}\right). \tag{5.39}$$

5.4 Safe probability

Consider now applying safe probability to the two envelope problem. If there is a safe distribution for $\mathbb{1}_{\{B=2A\}}|A$, we can use this distribution to see whether we want to switch and when. The random variable $\mathbb{1}_{\{B=2A\}}$ describes whether the second envelope has twice the observed value. The following proposition gives a safe distribution for $\mathbb{1}_{\{B=2A\}}$.

Proposition 5.4.1. *A distribution $\tilde{\mathbb{P}}$ on Ω is safe for $\langle \mathbb{1}_{\{B=2A\}}\rangle|\langle A\rangle$ with model $\mathcal{P}^*$ if*

$$\int_0^\infty \tilde{\mathbb{P}}[B = 2A|A = a]f_A(a)da = \frac{1}{2} \tag{5.40}$$

holds where f_A is the probability density function of A. Furthermore, the following statements are equivalent:

1. *For all $a \in \mathrm{supp}_{\tilde{\mathbb{P}}}(A)$ we have $\tilde{\mathbb{P}}[B = 2A|A = a] = \frac{1}{2}$.*

2. *$\tilde{\mathbb{P}}$ is safe for $\mathbb{1}_{\{B=2A\}}|[A]$.*

3. *$\tilde{\mathbb{P}}$ is safe for $\langle \mathbb{1}_{\{B=2A\}}\rangle|[A]$.*

Proof. We begin with proving the first statement. Let $\tilde{\mathbb{P}}$ be an arbitrary distribution on Ω that is safe for $\langle \mathbb{1}_{\{B=2A\}} \rangle | \langle A \rangle$. Let $\mathbb{P} \in \mathcal{P}^*$ be arbitrary and let f_X be the accompanying prior on X. Firstly we have

$$\mathbb{E}_{\mathbb{P}}\left[\mathbb{1}_{\{B=2A\}}\right] = \mathbb{P}[B=2A] = \int_0^\infty \mathbb{P}[B=2X|X=x]f_X(x)dx \qquad (5.41)$$

$$= \frac{1}{2} \int_0^\infty f_X(x)dx = \frac{1}{2}, \qquad (5.42)$$

as $B=2A$ only happens when $A=X$ and $\mathbb{P}[B=2X|X=x] = \frac{1}{2}$ is dictated by our model $\mathcal{P}^*$. Let $a \in \mathrm{supp}_{\tilde{\mathbb{P}}}(A)$ be arbitrary, then

$$\mathbb{E}_{\tilde{\mathbb{P}}}\left[\mathbb{1}_{\{B=2A\}}\big|A=a\right] = \tilde{\mathbb{P}}[B=2A|A=a] \qquad (5.43)$$

must hold. Therefore we have

$$\mathbb{E}_{\mathbb{P}}\left[\mathbb{E}_{\tilde{\mathbb{P}}}\left[\mathbb{1}_{\{B=2A\}}\big|A=a\right]\right] = \int_0^\infty \tilde{\mathbb{P}}[B=2A|A=a]f_A(a)da. \qquad (5.44)$$

The density f_A of A is unknown. Since $\tilde{\mathbb{P}}$ must be safe for $\langle \mathbb{1}_{\{B=2A\}} \rangle | \langle A \rangle$, we need $\mathbb{E}_{\mathbb{P}}\left[\mathbb{E}_{\tilde{\mathbb{P}}}\left[\mathbb{1}_{\{B=2A\}}\big|A=a\right]\right] = \mathbb{E}_{\mathbb{P}}\left[\mathbb{1}_{\{B=2A\}}\right]$ by definition 3.1.2. This results to the requirement

$$\int_0^\infty \tilde{\mathbb{P}}[B=2A|A=a]f_A(a)da = \frac{1}{2}. \qquad (5.45)$$

Now we will move our focus to the equivalence. First consider the implication from 1 to 2. Let $\tilde{\mathbb{P}}$ be a distribution on Ω with $\tilde{\mathbb{P}}[B=2A|A=a] = \frac{1}{2}$ for all $a \in \mathrm{supp}_{\tilde{\mathbb{P}}}(A)$. Let $\mathbb{P} \in \mathcal{P}^*$ and $a \in \mathrm{supp}_{\tilde{\mathbb{P}}}(A)$ be arbitrary, then

$$\tilde{\mathbb{P}}[B=2A|A=a] = \frac{1}{2} = \mathbb{P}[B=2A]. \qquad (5.46)$$

Thus $\tilde{\mathbb{P}}$ is safe for $\mathbb{1}_{\{B=2A\}} | [A]$ by Proposition 3.1.4.

The implication from 2 to 3 is stated in Proposition 3.1.5.

Lastly, we will turn to the implication from 3 to 1. Let $\tilde{\mathbb{P}}$ be a distribution on Ω that is safe for $\langle \mathbb{1}_{\{B=2A\}} \rangle | [A]$ for all $\mathbb{P} \in \mathcal{P}^*$. Let $\mathbb{P} \in \mathcal{P}^*$ be arbitrary and let $a \in \mathrm{supp}_{\tilde{\mathbb{P}}}(A)$, then safety for $\langle \mathbb{1}_{\{B=2A\}} \rangle | [A]$ implies

$$\mathbb{E}_{\mathbb{P}}[\mathbb{1}_{\{B=2A\}}] = \mathbb{E}_{\tilde{\mathbb{P}}}[\mathbb{1}_{\{B=2A\}}|A=a]. \qquad (5.47)$$

Expanding this requirement gives

$$\mathbb{E}_{\mathbb{P}}[\mathbb{1}_{\{B=2A\}}] = \frac{1}{2} = \mathbb{E}_{\tilde{\mathbb{P}}}[\mathbb{1}_{\{B=2A\}}|A=a] = \tilde{\mathbb{P}}[B=2A|A=a], \qquad (5.48)$$

which gives the result that $\tilde{\mathbb{P}}[B=2A|A=a] = \frac{1}{2}$ must hold for all $a \in \mathrm{supp}_{\tilde{\mathbb{P}}}(A)$. $\qquad \square$

Note that in this case Theorem 4.1.1 cannot be applied, as this theorem dictates that the sample space Ω is split up in $\Omega = \mathcal{X}' \times \mathcal{Y}'$, where $\mathcal{X}'$ is countable and $\mathcal{Y}'$ is finite. In our case we do have $\Omega = \mathcal{X} \times \mathcal{Y}$, however both $\mathcal{X}$ and $\mathcal{Y}$ are uncountably infinite of size.

Therefore while not knowing how $\mathbb{1}_{\{B=2A\}}|A$ is distributed, we can pick a safe distribution $\tilde{\mathbb{P}}$ from Proposition 5.4.1. When we only need an unbiased distribution for $\mathbb{1}_{\{B=2A\}}$ given A, it is sufficient to pick a $\tilde{\mathbb{P}}$ with

$$\int_0^\infty \tilde{\mathbb{P}}[B=2A|A=a]f_A(a)da = \frac{1}{2}. \tag{5.49}$$

For marginal validity we need $\tilde{\mathbb{P}}[B=2A|A=a] = \frac{1}{2}$ for all $a \in \text{supp}_{\tilde{\mathbb{P}}}(A)$.

Now note that safe probability will not provide more insight. Considering marginal validity first, let $\tilde{\mathbb{P}}$ be marginally valid for $\mathbb{1}_{\{B=2A\}}|A$. Then according to Proposition 5.4.1 $\tilde{\mathbb{P}}[B=2A|A=a] = \frac{1}{2}$ is required for all $a \in \text{supp}_{\tilde{\mathbb{P}}}(A)$. When $\text{supp}_{\tilde{\mathbb{P}}}(A)$ is of infinite size, distribution $\tilde{\mathbb{P}}$ cannot exist as is earlier pointed out in Section 5.2 and in [CU92, RCU93, NŠ17, TJ18].
When $\tilde{\mathbb{P}}$ is an unbiased for $\mathbb{1}_{\{B=2A\}}|A$, we have the requirement

$$\int_0^\infty \tilde{\mathbb{P}}[B=2A|A=a]f_A(a)da = \frac{1}{2}. \tag{5.50}$$

Here the probability density function f_A of A is not known and if it is, then a vast array of distributions $\tilde{\mathbb{P}}$ will satisfy this requirement. Thus unbiased distribution for $\mathbb{1}_{\{B=2A\}}|A$ will not give more insight here.

Suppose $\mathbb{1}_{\{B=2A\}}$ is predicted by a random variable $\mathbb{1}^*_{\{B=2A\}}$, not taking A into account. Let $\mathbb{P} \in \mathcal{P}^*$ be the actual probability distribution on Ω and let $\mathbb{1}^*_{\{B=2A\}} \sim \text{Ber}(p)$ be Bernoulli distributed with parameter $p \in [0,1]$. Let $\mathbb{P}'$ be the probability distribution of $\mathbb{1}^*_{\{B=2A\}}$, then

$$\text{acc}(\mathbb{1}_{\{B=2A\}}) = \sum_{k \in \{0,1\}} \mathbb{P}'[\mathbb{1}^*_{\{B=2A\}} = k]\,\mathbb{P}[\mathbb{1}_{\{B=2A\}} = k] = \frac{1}{2}(p+1-p) = \frac{1}{2}. \tag{5.51}$$

Therefore it looks like no matter in which frequency the person switches after an observed $A = a$, the probability of receiving the higher envelope stays equal. The average value the player wins will be $\frac{3}{2}\mathbb{E}_{\mathbb{P}}[X]$ and this cannot be improved without taking A into account.

Safe probability cannot take into account the probability observing $A = a$. Either a marginally valid distributions arises which erases the influence of A or an unbiased distribution arises where there is too less information for the distribution to be convincing. A probability measure $\tilde{\mathbb{P}}$ from Proposition 5.4.1 must be safe against all $\mathbb{P} \in \mathcal{P}^*$ and those $\mathbb{P}$'s all impose vastly different distributions on A. Therefore we can say that model $\mathcal{P}^*$ is simply too large to apply safe probability. We need to either put ill-founded assumptions on our model or try another strategy. We will do the latter in the next section.

5.5 Cover's switching strategy

If we adapt our probability of switching to the observation $A = a$, we can obtain on average higher results than $\frac{3}{2}\mathbb{E}_{\mathbb{P}}[X]$. The technique of switching smartly is first devised by Cover and in 2003 personly communicated to McDonnell and Abbott [MA09], who published this technique with numerous examples [MA09, ADP10, MGL$^+$11]. The techniques used were already lightly touched

upon by Christensen and Utts in 1992 [CU92] and by Ross in 1994 [RCU94] as response on the article of Christensen and Utts. Since McDonell and Abbott fleshed out the switching strategy and describes it in more detail, we will use their articles as foundation.

In short, Cover's switching strategy devises a function $P\colon (0, \infty) \to [0, 1]$ which picks an observed $a \in (0, \infty)$ and assigns a probability $P(a) \in [0, 1]$ to it. The player then switches his envelope with probability $P(a)$. Note that, in contrary to what many initially assume, the function P does not need to integrate to 1; it is not a probability density function. It implies after observing an $a \in (0, \infty)$ a probability distribution of whether or not to switch the envelopes.

Definition 5.5.1 (Switching strategy)**.** A function $P\colon (0, \infty) \to [0, 1]$ is called a *switching strategy*.

Assume that the prior f on X is *continuous* with finite expectation value and let $\mathbb{P} \in \mathcal{P}^*$ be its accompanying distribution on Ω. The function P can be picked in such a way that, while not knowing the distribution $\mathbb{P} \in \mathcal{P}^*$, we can always on average get a higher result than $\frac{3}{2}\mathbb{E}_{\mathbb{P}}[X]$.

Consider for a short while a more general problem where envelope A receives $2X$ with probability $1 - p$ and receives X with probability $p \in [0, 1]$. The original problem can then be retrieved by putting $p = \frac{1}{2}$. After observing value $A = a$, the player randomly chooses envelope B with probability $P(a)$. Let R be the value returned to the player at the end of the game, then the probability distribution of $R|X = x$ is given by

$$\mathbb{P}[R = x | X = x] = \mathbb{P}[A = x | X = x](1 - P(x)) + \mathbb{P}[A = 2x | X = x]P(2x) \tag{5.52}$$

$$= p(1 - P(x)) + (1 - p)P(2x), \tag{5.53}$$

$$\mathbb{P}[R = 2x | X = x] = \mathbb{P}[A = 2x | X = x](1 - P(2x)) + \mathbb{P}[A = x | X = x]P(x) \tag{5.54}$$

$$= (1 - p)(1 - P(2x)) + pP(x). \tag{5.55}$$

Therefore the conditional expectation of the return given $X = x$ is

$$\mathbb{E}[R | X = x] = x\,\mathbb{P}[R = x | X = x] + 2x\,\mathbb{P}[R = 2x | X = x] \tag{5.56}$$

$$= x(2 - p) + pxP(x) - x(1 - p)P(2x)). \tag{5.57}$$

Following [MA09], we get

$$\mathbb{E}[R] = \mathbb{E}[\mathbb{E}[R | X]] = \int_0^\infty f(x)\,\mathbb{E}[R | X = x]dx \tag{5.58}$$

$$= (2 - p)\,\mathbb{E}[X] + \int_0^\infty xf(x)(pP(x) - (1 - p)P(2x))dx. \tag{5.59}$$

When the player never or always switches, thus $P \equiv 0$ or $P \equiv 1$, his expected return is $(2 - p)\,\mathbb{E}[X]$. Define $G = \int_0^\infty xf(x)(pP(x) - (1 - p)P(2x))dx$ as the players gain of his switching strategy. One can immediately see that for all x the requirement $pP(x) \geq (1 - p)P(2x)$ is a sufficient condition for a non-negative

gain. The gain can be rewritten as

$$G = \int_0^\infty x f(x)(pP(x) - (1-p)P(2x))dx \tag{5.60}$$

$$= \int_0^\infty x P(x) \left(pf(x) - \frac{1-p}{4} f\left(\frac{x}{2}\right) \right) dx, \tag{5.61}$$

where both representations will become useful in many different ways. This will be summarized in the following definition.

Definition 5.5.2. Let f be a continuous prior on X with finite expectation value. Let P be a switching strategy. The *gain* of a player is defined by

$$G = \int_0^\infty x f(x)(pP(x) - (1-p)P(2x))dx \tag{5.62}$$

$$= \int_0^\infty x P(x) \left(pf(x) - \frac{1-p}{4} f\left(\frac{x}{2}\right) \right) dx. \tag{5.63}$$

Equation 5.63 immediately gives us the optimal switching strategy.

Theorem 5.5.3 (McDonnell, Abbott [MA09])**.** *The optimal switching strategy* $P^*(x)$ *is*

$$P^*(x) = \mathbb{1}_{(0,\infty)} \left(pf(x) - \frac{1-p}{4} f\left(\frac{x}{2}\right) \right). \tag{5.64}$$

Proof. We can split the integrand of (5.63) in two functions: $x \mapsto P(x)$ and $x \mapsto x\left(pf(x) - \frac{1-p}{4} f\left(\frac{x}{2}\right)\right)$. Function P can be chosen at will, but the image must lie in $[0,1]$. The integration in (5.63) is therefore maximized when $P(x) = 1$ is chosen if $x\left(pf(x) - \frac{1-p}{4} f\left(\frac{x}{2}\right)\right) > 0$ is positive and $P(x) = 0$ otherwise. Since $x \mapsto x$ is positive for all $x \in (0,\infty)$, it is sufficient to require that the function $x \mapsto pf(x) - \frac{1-p}{4} f\left(\frac{y}{2}\right)$ must be positive when $P(x) = 1$. Switching strategy P^* implements this strategy. $\qquad\square$

From now on we will put $p = \frac{1}{2}$ as in our model an envelope is distributed to the player uniformly. In this case Equation 5.62 leads to $P(x) \geq P(2x)$ for all $x \in (0,\infty)$ being sufficient for non-negative gain. We will now look at two example strategies, namely Cover switching and threshold switching. More examples, comparisons between the examples and simulations of the examples can be found in [MA09, MGL$^+$11].

Example 5.5.4 (Cover switching)**.** This example is taken from [MA09, ADP10]. Let $a \in (0,\infty]$ be arbitrary and put $P(x) = e^{-ax}$. This P is called *Cover switching*. The inequality $P(x) \geq P(2x)$ is automatically satisfied as the requirement $e^{-ax} > e^{-2ax}$ holds for all positive x.

Suppose that $X \sim \mathrm{Exp}(c)$ is exponentially distributed. Let R_C be the random variable of the return when Cover switching is used. Switching using Cover's strategy is maximized for $a = \left(\frac{1}{3}\sqrt[3]{2} + \frac{1}{6}\sqrt[3]{4} - \frac{1}{3}\right)c$, with a return of $\mathbb{E}[R_C] = \frac{3}{2c}\left(2 - 2\sqrt[3]{2} + \sqrt[3]{4}\right) \approx \frac{1.6013}{c}$. The return of this prior for never or always switching strategy is $\mathbb{E}[R] = \frac{3}{2}\mathbb{E}[X] = \frac{3}{2c}$, thus Cover switching does improve on never or always switching.

Example 5.5.5 (Threshold switching)**.** This example is taken from [MA09]. Let $b \in (0, \infty)$ be arbitrary and define $P(a) = \mathbb{1}_{(0,b]}(a)$. This strategy is called *threshold switching*. The player tries to switch all the low values he gets, but always holds on high values. Of course, if the prior on X is unknown, no advice on which b to choose can be given. However, as $P(a) \geq P(2a)$ always holds, this strategy has non-negative gain for all possible values b independent of the prior f.

Let $X \sim \text{Exp}(c)$ be exponentially distributed. Looking at Equation 5.63, we want that $P(x) = 1$ when $f(x) \geq \frac{1}{4} f\left(\frac{x}{2}\right)$ holds, thus when $ce^{-cx} \geq \frac{1}{4}ce^{-\frac{c}{2}x}$ holds. This is satisfied when $x < \frac{4\ln(2)}{c}$ holds, thus take $b = \frac{4\ln(2)}{c}$. Let R_T be the value returned to the player when using threshold switching. The expected return will be $\mathbb{E}[R_T] = \frac{51+4\ln(2)}{32c} \approx \frac{1.6804}{c}$, which is higher than Cover switching's $\mathbb{E}[R_C] = \frac{1.6013}{c}$ and the $\mathbb{E}[R] = \frac{3}{2c}$ of never or always switching. Threshold switching is in fact the optimal switching function here, as $f(x) = \frac{1}{4}f\left(\frac{x}{2}\right)$ is obtained for a unique $x = \frac{4\ln(2)}{c}$ and P^* from Theorem 5.5.3 then becomes threshold switching.

5.5.1 Optimizing Cover's switching strategy

Having two strategies, namely Cover and threshold switching, we would like to know how to optimize them while knowing as little of the distribution on X as possible. McDonnell et alii [MGL$^+$11] considered various cases and here we will take a look at the ones with the least information on the prior of X. Assume for the rest of this section that the prior f on X is absolutely continuous and differentiable on $(0, \infty)$.

Cover switching optimized

This example is taken from [MGL$^+$11]. Let $a \in [0, \infty]$ and consider Cover's switching strategy, $P(x) = e^{-ax}$. The gain as a function of a can be written as

$$G(a) = \int_0^\infty pxf(x)e^{-ax} - (1-p)xf(x)e^{-2ax}dx. \tag{5.65}$$

The optimal value a^* solves the functional equation

$$\int_0^\infty e^{-a^*t}x^2f(x)dx = \frac{2(1-p)}{p}\int_0^\infty e^{-2a^*t}x^2f(x)dx, \tag{5.66}$$

where $\int_0^\infty e^{-at}x^2f(x)dx$ is the Laplace transform of $x \mapsto x^2f(x)$ with parameter a. This leads to the following lemma:

Lemma 5.5.6 (McDonnell et al. (2011))**.** *Let $p \in [0,1]$ be the probability the lowest envelope is given to the player. Let P be Cover's switching strategy with parameter a. The following cases must be distinguished for the optimal value a^* of $G(a^*)$:*

$p \in \left[0, \frac{1}{5}\right)$*: The optimal value is $a^* = \infty$, thus never switching is the optimal strategy. The optimal gain is $G(\infty) = 0$.*

$p \in \left[\frac{1}{5}, \frac{2}{3}\right)$*: If the function $x \mapsto pf(x) - \frac{1-p}{4}f\left(\frac{x}{2}\right)$ has a unique sign change, then $a \mapsto G(a)$ has a unique stationary point $a^* \in [0, \infty)$. This a^* maximizes $a \mapsto G(a)$.*

$p \in \left[\frac{2}{3}, 1\right]$: *The optimal value is $a^* = 0$, thus always switching is the best strategy. The optimal gain is $G(0) = (2p - 1)\mathbb{E}[X]$.*

Proof. The proof can be found in [MGL⁺11]. □

Do note that there is an asymmetry in the previous lemma. When the probability of receiving the envelope with the lowest value is below $\frac{1}{5}$, never switching is the best strategy. However, when the probability is above $\frac{2}{3}$, always switching is the best strategy. One should expect the turning points to lie symmetric around $\frac{1}{2}$ such that they are $\frac{1}{3}$ and $\frac{2}{3}$ or $\frac{1}{5}$ and $\frac{4}{5}$. This phenomenon is widely studied in [MGL⁺11].

Threshold switching optimized

The first part of this example is taken from [MGL⁺11]. Let $b \in [0, \infty]$ and consider threshold switching, $P(x) = \mathbb{1}_{[0,b]}(x)$. The gain as a function of b can be written as

$$G(b) = \int_0^\infty pxf(x)\,\mathbb{1}_{[0,b]}(x) - (1-p)xf(x)\,\mathbb{1}_{\left[0,\frac{b}{2}\right]}(x)dx. \tag{5.67}$$

Taking the derivative yields $G'(b) = b\left(pf(b) - \frac{1-p}{4}f(b)\right)$, resulting to a possible set of optimal values. Let $B = \{b \in [0,\infty) | G'(b) = 0\}$ and take $b^* \in B$, then differentiating G twice yields

$$G''(b^*) = b^*\left(pf'(b^*) - \frac{1-p}{8}f'\left(\frac{b^*}{2}\right)\right), \tag{5.68}$$

which must be negative for b^* to be a local maximum. This results to the condition

$$\frac{8p}{1-p}f'(b^*) < f'\left(\frac{b^*}{2}\right). \tag{5.69}$$

Note that [MGL⁺11] does not have a strict inequality, but the second derivative test is inconclusive when $G''(b^*) = 0$. Thus it is still possible for b^* to be a saddle or a minimum in case of $G''(b^*) = 0$, which is why we write a strict inequality here. This all results into the following lemma:

Lemma 5.5.7. *Let $p \in [0,1]$ be such that the probability of receiving the lower envelope is p. Let P be the threshold switching strategy. The set*

$$B = \left\{b \in [0,\infty]\,\middle|\, pf(b) = \frac{1-p}{4}f\left(\frac{b}{2}\right) \wedge \frac{8p}{1-p}f''(b) < f'\left(\frac{b}{2}\right)\right\} \cup \{0,\infty\} \tag{5.70}$$

is the set of all b that locally maximize $b \mapsto G(b)$.

Proof. The proof can be found in [MGL⁺11]. □

This example however requires full knowledge on f. Can we optimize P with as little knowledge as possible? The answer is yes and is given by [EFG15]. Rephrase the gain as

$$G(b) = \frac{1}{2}\mathbb{E}[XP(X) - XP(2X)] = \frac{1}{2}\mathbb{E}\left[X\,\mathbb{1}_{\left(\frac{b}{2},b\right]}(X)\right], \tag{5.71}$$

then we can derive a lower bound of G when knowing only $\mathbb{E}[X]$ and $\mathbb{V}[X]$.

Theorem 5.5.8 (Egozcue and García (2015)). *Let X be a continuous random variable with $\mathbb{E}[X] = \mu$ and $\mathbb{V}[X] = \sigma^2$. Let $m = \mu^2 + \sigma^2$. Let P be the threshold strategy with parameter $b \in (0, \infty)$.*

1. *The gain has lower bound*

$$G(b) \geq \frac{3 + 2\sqrt{2}}{b} \left(\frac{3}{2} b\mu - \frac{1}{2} b^2 - m \right). \tag{5.72}$$

2. *If $\mu^2 \geq 8\sigma^2$ holds, then $b^* = \sqrt{2m}$ is the optimal threshold strategy with gain*

$$G(b^*) \geq (3 + 2\sqrt{2}) \left(\frac{3}{2}\mu - \sqrt{2m} \right) \geq 0. \tag{5.73}$$

The gain is strictly positive when $\mu^2 > 8\sigma^2$.

Proof. The proof can be found in [EFG15]. $\qquad\qquad\square$

Thus if $\mathbb{E}[X]^2 \geq 8\,\mathbb{V}[X]$ holds, we can find an optimal b^* which always results in non-negative gain. The next question is whether there is a $b \in (0, \infty)$ which guarantees us a strictly positive gain for all priors on X. This is not possible, as the following theorem will point out.

Theorem 5.5.9 (Egozcue and García (2015)). *Let $b \in (0, \infty)$. A random variable X with $\mathbb{E}[X]^2 < 8\,\mathbb{V}[X]$ such that $G(b) = 0$ always exists.*

Proof. The proof can be found in [EFG15]. $\qquad\qquad\square$

5.5.2 Discrete priors and the switching strategy

We only considered switching strategies for continuous priors thus far. If X is distributed discretely by distribution $\mathbb{P} \in \mathcal{P}^*$, then switching strategies are still possible. As in the continuous case, we have

$$\mathbb{E}[R|X = x] = x(2 - p) + pxP(x) - x(1 - p)P(2x). \tag{5.74}$$

The expected return is therefore

$$\mathbb{E}[R] = \sum_x f(x)\,\mathbb{E}[R|X = x] \tag{5.75}$$

$$= (2 - p)\,\mathbb{E}[X] + \sum_x xf(x)\big(pP(x) - (1 - p)P(2x)\big). \tag{5.76}$$

A change in variables like the continuous case now becomes a bit awkward. Let S be support of X, then

$$\sum_{x \in \mathrm{supp}_\mathbb{P}(X)} xf(x)P(2x) = \sum_{x \in 2\,\mathrm{supp}_\mathbb{P}(X)} \frac{x}{2} f\left(\frac{x}{2}\right) P(x) \tag{5.77}$$

holds. Thus we get

$$G = \sum_{x \in \mathrm{supp}_\mathbb{P}(X)} xf(x)(pP(x) - (1 - p)P(2x)) \tag{5.78}$$

$$= p \sum_{x \in \mathrm{supp}_\mathbb{P}(X)} xP(x)f(x) - \frac{1 - p}{2} \sum_{x \in 2\,\mathrm{supp}_\mathbb{P}(X)} xf\left(\frac{x}{2}\right) P(x), \tag{5.79}$$

and since in the discrete case $\text{supp}_{\mathbb{P}}(X) = 2\,\text{supp}_{\mathbb{P}}(X)$ does not necessarily hold the latter formula cannot be easily written in a single sum.

Put $p = \frac{1}{2}$, then it becomes clear that $P(x) \geq P(2x)$ is sufficient for non-negative gain. Thus any decreasing switching strategy will again be better than always keeping the envelope or always switching. Since this is also the case for continuous priors, we can recap both cases in the following theorem.

Theorem 5.5.10. *Let f be any prior on X with finite mean and suppose the envelope is handed to the player with uniform distribution. Let P be a decreasing switching strategy that is strictly decreasing on at least one subset $U \subset (0, \infty)$ that has positive probability measure. Then P will lead to a positive gain and is a better strategy than either always keeping the envelope or always switching the envelope.*

Proof. This theorem has already been proven in parts in the previous sections, so we'll only give a summary here. The gain, the difference between strategy P and always keeping or switching envelopes, is defined as

$$G = \frac{1}{2} \int_0^\infty x f(x)(P(x) - P(2x)) dx \tag{5.80}$$

when X is continuous and as

$$G = \frac{1}{2} \sum_x x f(x)(P(x) - P(2x)) \tag{5.81}$$

when X is discrete. As P is decreasing, $P(x) \geq P(2x)$ will always hold, making G non-negative. If P is also strictly decreasing on a subset $U \subset (0, \infty)$ with positive probability measure, G becomes is strictly positive. $\square$

5.6 The Ali Baba problem

We thus far only studied the two envelope problem where there is only one player. Consider now the following Ali Baba problem mentioned by Nalebuff in 1989 [Nal89]. In the Ali Baba problem, envelope A is filled first and then given to Ali. Then a hidden fair coin is tossed to decide whether envelope B must contain value $2A$ or $\frac{1}{2}A$, where after filling the envelope it is given to Baba. Ali and Baba are allowed to privately look at the contents of their envelope. If they both agree that switching is the better choice, they are able to do so.

Take a look at the following reasoning taken from the introduction of [Nal89]. First consider Ali's choice. Baba has $\frac{1}{2}x$ money with probability $\frac{1}{2}$ and $2x$ money with probability $\frac{1}{2}$. He will reason that on average B has $\frac{5}{4}x$ money, thus Ali always proposes a switch. Consider Baba's viewpoint next. Say he has an amount y, then A has $2y$ or $\frac{1}{2}y$ amount of money. Thus Baba expects an amount of $\frac{5}{4}y$ in Ali's envelope, making him also wanting to switch. They both want to switch, while Ali obtained a fixed value and Baba's value is linked to this value. This is considered paradoxical.

The problem in this paradox again arises from wrongly applying probability theory like in Section 5.1. Call x the amount of money in envelope A. The probability space is

$$\Omega = \left(\mathcal{X} = \left\{ (x, 2x), \left(x, \frac{1}{2}x \right) \right\}, 2^{\mathcal{X}}, \mathbb{P} \right) \tag{5.82}$$

where $\mathbb{P}$ is the probability measure with $\mathbb{P}\left[\left(x, \frac{1}{2}x\right)\right] = \mathbb{P}[(x, 2x)] = \frac{1}{2}$. Since Ali always gets an amount x, the expected value of Baba's envelope calculated from Ali's viewpoint remains the same:

$$\mathbb{E}[B] = \sum_{b \in \left\{\frac{x}{2}, 2x\right\}} b\,\mathbb{P}[B = b | A = x] = \frac{x}{2} \cdot \frac{1}{2} + 2x \cdot \frac{1}{2} = \frac{5}{4}x. \tag{5.83}$$

Therefore Ali will always want to switch. Consider now Baba's viewpoint. He either receives an amount $2x$ or $\frac{1}{2}x$. The corresponding expected values are

$$\mathbb{E}[A | B = 2x] = \sum_{a \in \{x, 4x\}} a\,\mathbb{P}[A = a | B = 2x] = 1 \cdot x + 0, \tag{5.84}$$

$$\mathbb{E}\left[A \middle| B = \frac{1}{2}x\right] = \sum_{a \in \left\{\frac{x}{4}, x\right\}} a\,\mathbb{P}[A = a | B = \frac{1}{2}x] = 0 + 1 \cdot x. \tag{5.85}$$

Thus the expected value of Ali's envelope is

$$\mathbb{E}[A] = \mathbb{E}[\mathbb{E}[A|B]] = \sum_{b \in \left\{\frac{x}{2}, 2x\right\}} \mathbb{P}[B = b]\,\mathbb{E}[A | B = b] \tag{5.86}$$

$$= \frac{1}{2}x + \frac{1}{2}x = x. \tag{5.87}$$

It is unclear for Baba whether he should switch, as A is expected to have on average value x.

Let $p \in [0, 1]$ be the probability that A and B agree to switch. Consider the probability space

$$\Omega' = \left(\mathcal{Y} = \left\{(x, -x), \left(-\frac{1}{2}x, \frac{1}{2}x\right), (0, 0)\right\}, 2^{\mathcal{Y}}, \mathbb{P}\right) \tag{5.88}$$

with $\mathbb{P}[\{(x, -x)\}] = \mathbb{P}\left[\left\{-\frac{1}{2}x, \frac{1}{2}x\right\}\right] = \frac{1}{2}p$ and $\mathbb{P}[\{(0,0)\}] = 1-p$. When there is no switch, A and B gain nothing. The probability of no switch is $1-p$, therefore $\{(0,0)\}$ has mass $1-p$. The probability of switching is p and B has $2x$ with probability $\frac{1}{2}$, therefore $\{(x, -x)\}$ has mass $\frac{1}{2}p$. Let G be the vector of gains. The expected gains are

$$\mathbb{E}[G] = \begin{pmatrix} x \\ -x \end{pmatrix} \mathbb{P}[\{(x, -x)\}] + \begin{pmatrix} -\frac{1}{2}x \\ \frac{1}{2}x \end{pmatrix} \mathbb{P}\left[\left\{\left(-\frac{1}{2}x, \frac{1}{2}x\right)\right\}\right] = \begin{pmatrix} \frac{1}{4}px \\ -\frac{1}{4}px \end{pmatrix}. \tag{5.89}$$

The expected gains sum to 0, thus no money is magically put into the system. Furthermore, A will always profit in the long run for any switching strategy with $p > 0$ as he initially obtains a fixed value x and the expected value obtained by B is $\mathbb{E}[B] = \frac{5}{4}\mathbb{E}[A]$. Thus Baba must always refuse the switching request from Ali in order to lose no money, giving a trivial non-paradoxical game.

5.6.1   Different versions of the Ali Baba problem

This is just one version of the Ali Baba problem; At least four different variations can be put upon the problem. Those problems are discussed by Nickerson and Falk in 2006 [NF06] and we will give a quick overview. For some variations the

Envelope known	Contents known	Ali's choice	Baba's choice
Yes	Yes	Trade	Prior
Yes	No	Trade	Keep
No	Yes	Prior	Prior
No	No	Indifferent	Indifferent

Table 5.1: Table taken from [NF06] giving advice to Ali and Baba whether they should switch their envelopes. 'Envelope known' means that both players know Ali received the first envelope. 'Contents known' means both players are allowed to look at their envelope's contents. When 'Prior' is stated, Ali or Baba should decide on the prior they impose on the value of their envelope.

optimal strategy is independent of a prior on X, for others a prior needs to be taken into account.

Nickerson and Falk considered the following three variants on the Ali Baba problem in [NF06]:

1. is the distribution on X unknown or from the uniform distribution between 0 and 100,

2. do the players know which envelope they received and

3. are the players able to look at the contents?

The first variant is not treated here as we assume that the players do not know the distribution on X. Even if they do know, [NF06] only considers the uniform prior between 0 and 100 which we find too restrictive and not realistic. Here we always assume that the prior on the first envelope is unknown, possibly unbounded, reducing the number of different problems to four.

Table 5.1 gives all optimal strategies. Unfortunately, two out of four cases depend on the prior Ali and Baba impose on their envelope. If both players know the contents of their envelope, the prior on the observed value is needed to give an advice to switching.

Note that Ali must always trade if he knows he received the first envelope. In this case the expected value of Baba's envelope is $\mathbb{E}[B] = \frac{5}{4}\mathbb{E}[A]$. In both cases however Baba must either decide on his prior or always keep the envelope, thus blindly switching for all values is only fortunate when the prior has infinite mean. This debunks the paradox in most cases and when the mean is infinite, the paradox can be classified as St. Petersburg-like.

If both players have no information, the symmetry in the game dictates that there is no information to base a player's strategy on. No paradoxical results can be obtained here.

All variants can be investigated further to obtain optimal strategies. Cover's switching strategy can most likely be applied here. I believe that similar results as in the initial two envelope problem will rise here as well. I leave this open for further research.

5.7 Concluding remarks

The two envelope problem is another example of why performing conditional probability incorrectly can lead to contradicting and paradoxical results. Here

paradoxical results arise by wrongly applying conditional probability. Most calculate the expected value of envelope B as

$$\mathbb{E}[B] = \frac{A}{2}\,\mathbb{P}\left[B = \frac{A}{2}\right] + 2A\,\mathbb{P}[B = 2A], \tag{5.90}$$

however random variable A is used here twice with a different context. As observed in Section 5.1, the actual calculation should be

$$\mathbb{E}[B] = \mathbb{E}[A|A < B] + \frac{1}{4}\,\mathbb{E}[A|A > B], \tag{5.91}$$

where $\mathbb{E}[A|A < B]$ and $\mathbb{E}[A|A > B]$ are impossible to calculate without having more information. When the lowest value of both envelopes $X = \min\{A, B\}$ is considered, we observe that

$$\mathbb{E}[A] = \frac{3}{2}\,\mathbb{E}[X] = \mathbb{E}[B] \tag{5.92}$$

which is not considered as paradoxical.

We now only need to know the distribution of X. All probability distributions on $(0, \infty)$ are however possible. In Section 5.4 we saw that safe probability does not lead to new insights, in contrast to the problems in Chapter 4. Proposition 5.4.1 states that it is marginally safe to assume that envelope B has twice the value of A with probability of 50%, however in Section 5.2 we already saw such distribution does not exist when infinitely many values of X are possible. When $\tilde{\mathbb{P}}$ is unbiased for $\mathbb{1}_{\{B=2A\}}\,|A$, then

$$\int_0^\infty \tilde{\mathbb{P}}[B = 2A|A = a] f_A(a) da = \frac{1}{2} \tag{5.93}$$

is needed, where f_A is the probability density function of A. We however need more information on f_A to create a convincing $\tilde{\mathbb{P}}$ satisfying this requirement and this is not possible from the problems statement.

When we state that $B = 2A$ happens with probability $\frac{1}{2}$ independent of A's value and we base our decisions on this assumption, we will on average get a return of $\frac{3}{2}\,\mathbb{E}[X]$. If we drop this assumption, we can improve on this return. Cover switching states that we create a strategy $P\colon (0, \infty) \to [0, 1]$ which assigns a probability of switching to each observed value. The gain G relative to the trivial strategies can then be defined by

$$G = \frac{1}{2} \int_0^\infty x f(x)(P(x) - P(2x)) dx \tag{5.94}$$

with f the density function on X. For the never or always switching strategies we have zero gain. Theorem 5.5.10 states that we only need to make P decreasing in order to perform as well as the 'never switching' strategy. This is quite intuitive, as you are less likely to swap the envelopes when you initially obtain a high value. When a low value is observed, you will lose less if you swap it for an even lower envelope.

Initially it was unknown how much gain a switching strategy yields, as nothing is known of the distribution of X. The gain G can therefore become arbitrarily close to zero. You can for example apply threshold switching with 31 as threshold, while X only takes values higher than 94. Recently in 2015, Egozcue and

Fuentes García [EFG15] proved that in certain conditions only depending on the expectation value and variance of the distribution on X an optimal threshold value for threshold switching can be obtained, resulting in a lower bound for the gain which is strictly positive as described in Theorem 5.5.8. In this case one is able to actually improve on the trivial 'never switching' strategy by at least a certain amount.

The most important observations for this thesis are in Sections 5.1 and 5.2. When explicitly writing down the probability spaces used for studying this problem, one first observes that the calculation performed in Equation 5.1 is wrong. Furthermore, one observes that the naive probability space of Section 5.1 with only $\Omega = \{(x, 2x), (2x, x)\}$ as sample space is too restrictive. Only then you are able to extend this probability space to the one in Section 5.2 including all possible distributions on the lowest value X. This enables you also to observe that $\mathbb{P}[B = 2A|A = a] = \frac{1}{2}$ for all $a \in (0, \infty)$ is simply not possible; an assumption that does follow from safe probability and that everyone starting to study this problem makes. This underlines the message that when performing conditional probability, the probability space with the accompanying σ-algebra should always be taken into account to avoid paradoxical results.

Bibliography

[ADP10]   Derek Abbott, Bruce R. Davis, and Juan M. R. Parrondo. The two-envelope problem revisited. *Fluct. Noise Lett.*, 9(1):1–8, 2010.

[AKS05]   Casper J. Albers, Barteld P. Kooi, and Willem Schaafsma. Trying to resolve the two-envelope problem. *Synthese*, 145(1):89–109, May 2005.

[BCU93]   David A. Binder, Ronald Christensen, and Jessica Utts. Comment and reply to Christensen and Utts (1992). *The American Statistician*, 47(2):160, 1993.

[BCU96]   Nelson M. Blachman, Ronald Christensen, and Jessica M. Utts. Comment and reply to Christensen and Utts (1992). *The American Statistician*, 50(1):98–99, 1996.

[Ber89]   Joseph Bertrand. *Calcul des probabilités.* Gauthier-Villars et Fils, Paris, 1889.

[Bil95]   Patrick Billingsley. *Probability and measure.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, third edition, 1995. A Wiley-Interscience Publication.

[BK95]   Steven J. Brams and D. Marc Kilgour. The box problem: To switch or not to switch. *Mathematics Magazine*, 68(1):27–34, 1995.

[Bor09]   Émile Félix Édouard Justin Borel. *Éléments de la théorie des probabilités : probabilités discontinues, probabilités continues, probabilités des causes.* Hermann, Paris, 1909.

[Bro95]   John Broome. The two-envelope paradox. *Analysis*, 55(1):6–11, 1995.

[CU92]   Ronald Christensen and Jessica Utts. Bayesian resolution of the "exchange paradox". *Amer. Statist.*, 46(4):274–276, 1992.

[dF72]   Bruno de Finetti. *Probability, induction and statistics. The art of guessing.* John Wiley & Sons, London-New York-Sydney, 1972. Wiley Series in Probability and Mathematical Statistics.

[Eas08]   Kenneth Krishnan Easwaran. *The foundations of conditional probability.* ProQuest LLC, Ann Arbor, MI, 2008. Thesis (Ph.D.)– University of California, Berkeley.

[EFG15]   Martin Egozcue and Luis Fuentes García. An optimal threshold strategy in the two-envelope problem with partial information. *J. Appl. Probab.*, 52(1):298–304, 03 2015.

[Fel50]   William Feller. *An Introduction to Probability Theory and Its Applications. Vol. I.* John Wiley & Sons, Inc., New York, N.Y., 1950.

[Fel71]   William Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons, Inc., New York-London-Sydney, 1971.

[GH03]   Peter D. Grünwald and Joseph Y. Halpern. Updating probabilities. *J. Artificial Intelligence Res.*, 19:243–278, 2003.

[GHSR17] Z. Gyenis, G. Hofer-Szabó, and M. Rédei. Conditioning using conditional expectations: the Borel-Kolmogorov paradox. *Synthese*, 194(7):2595–2630, 2017. Corrections are given in appendix A.1.

[Gr 13]   Peter Grünwald. Safe probability: restricted conditioning and extended marginalization. In *Symbolic and quantitative approaches to reasoning with uncertainty*, volume 7958 of *Lecture Notes in Comput. Sci.*, pages 242–253. Springer, Heidelberg, 2013.

[Gr 16]   Peter Grünwald. Safe probability. ArXiv, 2016. ArXiv version of [Gr 18b], containing proofs and an example using Monty Hall's problem.

[Gr 18a]   Peter Grünwald. Private communication, October 2018. Provided me his second version of the boy or girl problem.

[Gr 18b]   Peter Grünwald. Safe probability. *Journal of Statistical Planning and Inference*, 195:47 – 63, 2018. Confidence distributions.

[H03]   Alan Hájek. What conditional probability could not be. *Synthese*, 137(3):273–323, 2003.

[How14]   Colin Howson. Finite additivity, another lottery paradox and conditionalisation. *Synthese*, 191(5):989–1012, 2014.

[Jay03]   E. T. Jaynes. *Probability theory.* Cambridge University Press, Cambridge, 2003. The logic of science, Edited and with a foreword by G. Larry Bretthorst.

[Kol33]   Andrej Nikolaevič Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Ergebnisse der Mathematik und ihrer Grenzgebiete ; 2. Bd, 3. 832253545. Springer, Berlin, 1933.

[Kra53]   Maurice Kraitchik. *Mathematical recreations.* Dover Publications, Inc., New York, N. Y., 1953. 2d ed.

[Lin94]   Elliot Linzer. The two envelope paradox. *The American Mathematical Monthly*, 101(5):417–419, 1994.

[MA09]   Mark D. McDonnell and Derek Abbott. Randomized switching in the two-envelope problem. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 465(2111):3309–3322, 2009.

[MGL⁺11] Mark D. McDonnell, Alex J. Grant, Ingmar Land, Badri N. Vellambi, Derek Abbott, and Ken Lever. Gain from the two-envelope problem via information asymmetry: on the suboptimality of randomized switching. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 467(2134):2825–2851, 2011.

[Myr15] Wayne C. Myrvold. You can't always get what you want some considerations regarding conditional probabilities. *Erkenntnis*, 80(3):573–603, Jun 2015.

[Nal89] Barry Nalebuff. The other person's envelope is always greener. *J. of Economic Perspectives*, 3:171–181, 1989.

[NF06] Raymond S. Nickerson and Ruma Falk. The exchange paradox: Probabilistic and cognitive analysis of a psychological conundrum. *Thinking & Reasoning*, 12(2):181–213, 2006.

[NŠ17] Mirko Navara and Jirí Šindelár. The role of information in the two envelope problem. *ITAT*, pages 112–119, 2017.

[OM14] S. B. G. O'Brien and S. L. Mitchell. The two envelope problem: there is no conundrum. *Teaching Mathematics and its Applications: An International Journal of the IMA*, 33(4):249–262, 07 2014.

[PP98] Michael A. Proschan and Brett Presnell. Expect the unexpected from conditional expectation. *Amer. Statist.*, 52(3):248–252, 1998.

[Rao88] M. M. Rao. Paradoxes in conditional probability. *J. Multivariate Anal.*, 27(2):434–446, 1988.

[RCU93] Terry Ridgway, Ronald Christensen, and Jessica Utts. Comment and reply to Christensen and Utts (1992). *The American Statistician*, 47(4):311, 1993.

[RCU94] Sheldon M. Ross, Ronald Christensen, and Jessica Utts. Comment and reply to Christensen and Utts (1992). *The American Statistician*, 48(3):267–268, 1994.

[SD08] Eric Schwitzgebel and Josh Dever. The two envelope paradox and using variables within the expectation formula. *Sorites*, 20:135–140, 2008.

[TJ18] Joseph Tzur and Arie Jacobi. Decision making under uncertainty and the two-envelope paradox. Available at SSRN: https://ssrn.com/abstract=3141049 or http://dx.doi.org/10.2139/ssrn.3141049, March 2018.

[Wei18] Eric W. Weisstein. Sphere point picking. From MathWorld – A Wolfram Web Resource. http://mathworld.wolfram.com/SpherePointPicking.html, 2018. Visited on 19-12-18.

[Wik19] Wikipedia contributors. Two envelopes problem — Wikipedia, the free encyclopedia. https://en.wikipedia.org/w/index.php?title=Two_envelopes_problem&oldid=899098671, 2019. Visited on 24-06-19.

[Wil91]   David Williams. *Probability with martingales.* Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991.

Appendix A

Corrections to [GHSR17]

A.1   Corrections to appendix 1

In appendix 1, equation 94 of the proof of the conditional expectation on latitudes in [GHSR17], they claim that the volume of the unit sphere is 2π, whereas they should have claimed that the surface area of the unit sphere is 4π. Luckily this constant is not actually used in the computations, thus their mistake has no impact on the rest of their article.

A.2   Corrections to appendix 2

I suggest the following corrections on appendix 2, the proof of conditional expectation on meridians:

1. In the verification [GHSR17] claims that $\int_0^\pi \sin\theta d\theta = 1$ holds between equations (107) and (108) and between equations (115) and (116), while that integral actually has value 2.

2. The conditional distribution of [GHSR17] is verified on a half meridian arc on page 2614, while this distribution must be verified on a full circle in order to be compared with the latitudes. Verification on a full meridian, e.g. computation of $q(C)$ in formula (85), quickly reveals that the conditional distribution of [GHSR17] integrates to 2.

3. The random variable X is integrated on the domain $[0, 2\pi)$, while X is only defined on $[0, 2\pi) \times [0, \pi] \in \mathfrak{M}$. The set $A \times [0, 2\pi)$ is not an element of $\mathfrak{M}$.

4. Between equations (105) and (106) of [GHSR17], they implicitly claim that the identity $X(\phi, \theta) = X(\phi, \theta + \pi)$ holds for all $\mathcal{F}$-measurable X, where now notation of [GHSR17] is used. This is most certainly false without any more restrictions on X.

5. Before equation (85) a measure $q_{\mathcal{M}}$ is defined on a whole meridian. Since the integral is taken from $\psi = 0$ to $\psi = 2\pi$, the integral of $q_{\mathcal{M}}$ over S becomes 1. However, as pointed out earlier, $\psi > \pi$ is not in our domain.

Thus the integral must be split up in two arcs with $q_{\mathcal{M}}$ taking value $\frac{1}{2}$ on each arc.

6. The normalization constant is used in equation (103) of [GHSR17] is 2π, where it must be 4π. This does however not impact further calculations, like the same mistake made from equation (109) to (110).