

# Thesis

Mathijs Kolkhuis Tanke

May 17, 2019

## **Abstract**

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>The Borel-Kolmogorov paradox</b>	<b>6</b>
2.1	Conditional expectation . . . . .	6
2.2	A formal description of the Borel-Kolmogorov paradox . . . . .	8
2.3	Conditional distributions . . . . .	8
2.3.1	Naive conditional probability . . . . .	9
2.3.2	Conditional probability on longitudes . . . . .	9
2.3.3	Conditional probability on meridians . . . . .	10
2.3.4	Combining longitudes and meridians . . . . .	13
2.4	The Borel-Kolmogorov phenomenon explained . . . . .	14
<b>3</b>	<b>Safe probability</b>	<b>15</b>
<b>4</b>	<b>Discrete paradoxes</b>	<b>16</b>
<b>5</b>	<b>The two envelopes problem</b>	<b>17</b>
<b>6</b>	<b>The Sleeping Beauty problem</b>	<b>18</b>

# Chapter 1

## Introduction

Probability theory is one of the most important and most researched fields in mathematics. It is entirely based on 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, the two envelope problem and the Sleeping Beauty problem. We study these problems and deduce ways to resolve their paradoxical statements.

Take a look at the Borel-Kolmogorov paradox first. A sphere is given with unit probability measure. The information that a point exists on a great circle of the sphere, however you do not know where that point is. Is there a probability distribution on the great circle for that point? If one models this problem using longitudes, then 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 looked at 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. (A. Kolmogoroff, [Kol33])

Kolmogorov states that the concept of conditional probability 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 with [GHSR17] giving more insight and depth to the problem, but eventually draws the same conclusion as Kolmogorov. This problem is more elaborately discussed in chapter 2.

Another paradox arising from conditional probability is Monty Hall's problem. In Monty Hall's problem, a player is facing three doors, doors  $a$ ,  $b$  and  $c$ . One door contains a car, 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 if 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 over, 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

$$\mathbb{P}[U = b \mid U \in \{a, b\}] = \frac{\mathbb{P}[U = b, U \in \{a, b\}]}{\mathbb{P}[U \in \{a, b\}]} = \frac{\frac{1}{3}}{\frac{2}{3}} = \frac{1}{2}, \quad (1.1)$$

supporting the claim that 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 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 events at all. When using conditional probability, one must provide a pair of a sub- $\sigma$ -algebra and the event to condition on. That event must be a member of the sub- $\sigma$ -algebra. There is no  $\sigma$ -algebra on  $\{a, b, c\}$  such that equation (1.1) still holds. In the case of the Borel-Kolmogorov paradox we can provide underlying sub- $\sigma$ -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 model of probability distributions that can all be the correct distribution, but we do not know which one is correct. Using the theory of safe probability we can obtain predictions that give equal results for all distributions in the model. This theory is introduced by [Gr 18] and in chapter 3. We will study safe probability more in-depth in chapter 4, using it to make safe predictions on problems as Monty Hall and on the two envelope problem in chapter 5 as well.

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

**Theorem.** *Let  $\mathcal{X}$  and  $\mathcal{Y}$  be countable. Let  $U$  be a random variable on  $\mathcal{Y}$  and  $V$  be a random variable on  $\mathcal{X}$ . Let  $p_u \in [0, 1]$  with  $\sum_{u \in \mathcal{Y}} p_u = 1$  and let  $\mathcal{P}^* \subseteq \{\mathbb{P} \mid \forall u \in \mathcal{Y} : \mathbb{P}[U = u] = p_u\}$  be our model of probability distributions on*

$\mathcal{X} \times \mathcal{Y}$ . Suppose there is a

$$u' \in \bigcap_{\mathbb{P} \in \mathcal{P}^*} \bigcap_{v \in \text{range}_{\mathbb{P}}(V)} \text{range}_{\mathbb{P}}(U|V = v). \quad (1.2)$$

Let  $\tilde{\mathbb{P}}$  be a distribution on  $\mathcal{X} \times \mathcal{Y}$  with

$$\tilde{\mathbb{P}}[U = u'|V = v] = p_{u'} \quad (1.3)$$

for all  $v \in \mathcal{X}$ . This  $\tilde{\mathbb{P}}$  is safe for  $\mathbb{1}_{\{U=u'\}}|V$ .

If for all  $v \in \mathcal{X}$  a  $\mathbb{P} \in \mathcal{P}^*$  exists such that  $v \in \text{range}_{\mathbb{P}}(V)$  holds, then this requirement on  $\tilde{\mathbb{P}}$  is necessary for safety for  $\langle \mathbb{1}_{\{U=u'\}} \rangle | \langle V \rangle$ .

This theorem will be explained, proven and applied in chapter 4, but it essentially states that 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 modelled as  $\frac{1}{3}$ .

A third still unsolved problem is the Sleeping Beauty problem. In this problem Sleeping Beauty is willing to do an experiment. At Sunday she goes to sleep. A fair coin is tossed with the following results: Sleeping Beauty wakes on Monday with heads and wakes on Monday and Tuesday with tails. If she is awake, she must state her credence of the coin giving heads. After the question, she goes to sleep using an amnesia-inducing drug which makes her forget that she has waken before. At Wednesday she wakes and the experiment is over.

A first argument of the probability distribution of the coin is given by the halfers. When Sleeping Beauty wakes, she has no clue what day it is. She knows that the coin is fair and has no reason guess otherwise, thus she must say the the coin is still fair.

Another argument is given by the thirders. Namely that there are three possible events, namely awake on Monday and coin has heads, awake on Monday and coin has tails and as last awake on Tuesday and coin has tails. She does not know which event she attends, thus the events happen with uniform probability. There is one event with the coin giving heads, thus the probability of the coin having heads is 33%.

I propose a solution that the thirders use a different model than the halfers. The question concerns the probability distribution of the coin, which remains 50%. The thirders answer the question ‘what is the probability of guessing correctly if I guess that the coin landed heads’, which is 33%. This solution will be discussed with more depth in chapter 6.

Every problem has 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  $\sigma$ -algebra taken into account. This yields to various results, as conditioning on events that cannot be conditioned on to not recognizing that multiple probability distributions are possible. The theme of is thesis is therefore that when doing probability theory, the underlying probability space and  $\sigma$ -algebra must never be ignored and not providing a  $\sigma$ -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. It is first study by Borel [Bor09] and Kolmogorov [Kol33] and has sparked debate over the centuries afterwards<sup>1</sup>. We will take a look 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 be the resulting conditional distribution? If the great circle is viewed as a longitude, the conditional distribution will be uniform as well. However, if the great circle is viewed 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?

As Kolmogorov [Kol33] points out, the great circle has zero measure and therefore conditioning on a great circle must not be allowed. Our analysis will follow [GHSR17], who comes to the same conclusion as Kolmogorov with a more elaborate explanation.

Before we can start with the analysis, we need to define conditional expectations in the measure-theoretic setting. We will then prove that the conditional distributions on a great circle can differ and conclude from the results that conditioning on zero sets cannot be admissible. The Borel-Kolmogorov paradox will then be renamed 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 will use the definitions given by David Williams [Wil91].

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

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

---

<sup>1</sup>Bronnen lezen en invoeren.

Lastly, let  $\mathcal{B}(A)$  be the  $\sigma$ -algebra of all Borel-measurable sets on  $A \subseteq \mathbb{R}^n$ . We will first define a random variable.

**Definition 2.1.1** (Random variable). A function  $X: \Omega \rightarrow \mathbb{R}$  is a *random variable* if it is  $\mathcal{F}$ -measurable, thus if  $X^{-1}(A) \in \mathcal{F}$  holds for all  $A \in \mathcal{B}$  where  $\mathcal{B}$  is the Borel  $\sigma$ -algebra on  $\mathbb{R}$ .

We can now define the conditional expectation.

**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- $\sigma$ -algebra of  $\mathcal{F}$ . Then 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

$$\int_G Y d\mathbb{P} = \int_G X d\mathbb{P}. \quad (2.2)$$

$Y$  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  $\sigma$ -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  $\mathcal{G} = \sigma(Z_1, Z_2, \dots)$ .

The traditional definition of conditional expectation is that 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  $Y$ , the conditional expectation  $\mathbb{E}[X|Y = y]$  is defined as

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

Definition 2.1.2 agrees on this traditional usage. Take  $\mathcal{G} = \sigma(\{Y = y\})$ , the smallest  $\sigma$ -algebra generated by all  $\omega \in \mathbb{R}$  such that  $Y(\omega) = y$ , and define

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

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].

Definition 2.1.2 extends very nicely to the definition of conditional probability. For an  $F \in \mathcal{F}$  we have  $\mathbb{E}[\mathbb{1}_F] = \mathbb{P}[F]$ , which carries over to conditional expectations as well.

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

We conclude this section with a list of useful properties of conditional expectation, taken from section 9.7 of [Wil91].



**Property 2.1.4.** *Let  $X$  be a random variable,  $\mathcal{G}, \mathcal{H}$  be a sub- $\sigma$ -algebra of  $\mathcal{F}$ , then:*

1. *If  $Y$  is a version of  $\mathbb{E}[X|\mathcal{G}]$ , then  $\mathbb{E}[Y] = \mathbb{E}[X]$ .*
2. *If  $X$  is  $\mathcal{G}$ -measurable, then  $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$  almost surely.*
3.  *$\mathbb{E}[X|\mathcal{G}]$  is almost surely linear in  $X$ .*
4. *If  $X \geq 0$ , then  $\mathbb{E}[X|\mathcal{G}] \geq 0$ .*
5. *If  $\mathcal{H}$  is a sub- $\sigma$ -algebra of  $\mathcal{G}$ , then  $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}]$  almost surely.*
6. *If  $Z$  is  $\mathcal{G}$ -measurable and bounded, then  $\mathbb{E}[ZX|\mathcal{G}] = Z \mathbb{E}[X|\mathcal{G}]$  almost surely.*
7. *If  $\mathcal{H}$  is independent of  $\sigma(\sigma(X), \mathcal{G})$ , then  $\mathbb{E}[X|\sigma(\mathcal{G}, \mathcal{H})] = \mathbb{E}[X|\mathcal{G}]$  almost surely. If  $X$  is independent of  $\mathcal{H}$ , then  $\mathbb{E}[X|\mathcal{H}] = \mathbb{E}[X]$  almost surely.*

## 2.2 A formal description of the Borel-Kolmogorov paradox

To analyse the Borel-Kolmogorov paradox, we need a formal version of the problem. 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

$$S \rightarrow \mathbb{R}^3 : (\phi, \psi) \mapsto (\cos \phi \sin \psi, \sin \phi \sin \psi, \cos \psi). \quad (2.5)$$

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

$$\mathbb{P}[B] = \frac{1}{4\pi} \iint_B \sin \psi d\psi d\phi. \quad (2.6)$$

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

The set of longitudes is described by

$$\mathcal{C} = \{[0, 2\pi] \times \{\psi\} \mid \psi \in [0, \pi]\} \quad (2.7)$$

and the set of meridians is described by

$$\mathcal{M} = \{[\phi, \phi + \pi] \times [0, \pi] \mid \phi \in [0, \pi]\}. \quad (2.8)$$

Note that the function in (2.5) is not a bijection, however the image of  $S$  coincides on a null set which will not cause any problems in our case. The meridians  $\{0, \pi\} \times [0, \pi]$  and  $\{\pi, 2\pi\} \times [0, \pi]$  in  $\mathcal{M}$  also define the same meridian on  $S$ , but this will also not cause any concerns. The reason why for example (2.5) is not a formal bijection is that it eases notation where there is no explicit need to make (2.5) a bijection.

## 2.3 Conditional distributions

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

### 2.3.1 Naive conditional probability

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

Note that  $\mathbb{P}[B] = 0$ , as a great circle always has zero measure. To be precise, let  $f: S \rightarrow \mathbb{R}^3$  be the coordinate transformation of (2.5). All great circles  $B'$  have a rotation  $O: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  such that  $(f^{-1} \circ O \circ f)(B') = [0, 2\pi] \times \{\frac{\pi}{2}\}$ . Rotations are orthogonal and have determinant 1, thus

$$\mathbb{P}[B] = \frac{1}{4\pi} \iint_B \sin \psi d\psi d\phi \quad (2.9)$$

$$= \frac{1}{4\pi} \int_0^{2\pi} \int_{\frac{1}{2}\pi}^{\frac{3}{2}\pi} \sin \psi \det(f^{-1} \circ O \circ f) d\psi d\phi \quad (2.10)$$

$$= \frac{1}{4\pi} \int_0^{2\pi} \int_{\frac{1}{2}\pi}^{\frac{3}{2}\pi} \sin \psi d\psi d\phi = 0. \quad (2.11)$$

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

$$\mathbb{P}[F|B] = \frac{\mathbb{P}[F \cap B]}{\mathbb{P}[B]}, \quad (2.12)$$

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

Thus classical Bayesian 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- $\sigma$ -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 longitudes

Since four-element sub- $\sigma$ -algebras are too small, we need to condition on bigger sub- $\sigma$ -algebras. One option is the  $\sigma$ -algebra of longitudes. Let

$$\mathfrak{C} = \sigma(\{[0, 2\pi] \times A \mid A \in \mathcal{B}([0, \pi])\}) \quad (2.13)$$

be the  $\sigma$ -algebra of all measurable subsets of longitudes. 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*

$$\mathbb{E}[X|\mathfrak{C}](\phi, \psi) = \frac{1}{2\pi} \int_0^{2\pi} X(\phi', \psi) d\phi' \quad (2.14)$$

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

$$\int_C X d\mathbb{P} = \frac{1}{4\pi} \int_A \int_0^{2\pi} X(\phi, \psi) \sin \psi d\phi d\psi \quad (2.15)$$

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

$$\int_C \mathbb{E}[X|\mathfrak{C}] d\mathbb{P} = \frac{1}{4\pi} \int_A \int_0^{2\pi} \mathbb{E}[X|\mathfrak{C}](\phi, \psi) \sin \psi d\phi d\psi. \quad (2.16)$$

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

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

$$= \frac{1}{4\pi} \int_A \int_0^{2\pi} \frac{1}{2\pi} \left( \int_0^{2\pi} X(\phi', \psi) d\phi' \right) \sin \psi d\phi d\psi \quad (2.18)$$

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

$$= \frac{1}{4\pi} \int_A \int_0^{2\pi} X(\phi, \psi) \sin \psi d\phi d\psi = \int_C X d\mathbb{P}. \quad (2.20)$$

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  $X$ .  $\square$

This derivation is slightly different from the one in [GHSR17]. There they claim that the volume of the unit sphere is  $2\pi$ , whereas they actually meant the surface area of the unit sphere which is  $4\pi$ . Luckily this constant is not actually used in the computations, thus their mistake has no impact on the rest of their article.

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 longitude  $C = [0, 2\pi] \times \{\psi'\}$  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

$$\bar{\mathbb{P}}[A] = \int_S \mathbb{P}[A|\mathfrak{C}] d\mathbb{P}' = \frac{1}{2\pi} \int_S \int_0^{2\pi} \mathbb{1}_A(\phi', \psi) d\phi' d\mathbb{P}'(\phi, \psi) \quad (2.21)$$

$$= \frac{1}{2\pi} \left( \int_{S \setminus C} \int_{\phi_1}^{\phi_2} 0 d\phi' d\mathbb{P}' + \int_C \int_{\phi_1}^{\phi_2} 1 d\phi' d\mathbb{P}' \right) \quad (2.22)$$

$$= \frac{\phi_2 - \phi_1}{2\pi}. \quad (2.23)$$

Thus the conditional probability distribution on  $C$  is uniform, resulting in all longitudes having uniform conditional probability measure.

### 2.3.3 Conditional probability on meridians

A great circle can not only be described as a longitude, but as a meridian as well. Therefore we would like to see whether describing great circles as meridians yield to the same conditional distribution. Let

$$\mathfrak{M} = \sigma(\{A \times [0, \pi] \mid A \in \mathcal{B}([0, 2\pi])\}) \quad (2.24)$$

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

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

$$\mathbb{E}[X|\mathfrak{M}](\phi, \psi) = \frac{1}{2} \int_0^\pi X(\phi, \psi') \sin \psi' d\psi' \quad (2.25)$$

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

$$\int_M \mathbb{E}[X|\mathfrak{M}] d\mathbb{P} = \frac{1}{4\pi} \int_0^\pi \int_A \mathbb{E}[X|\mathfrak{M}](\phi, \psi) \sin \psi d\phi d\psi \quad (2.26)$$

$$= \frac{1}{4\pi} \int_0^\pi \int_A \left( \frac{1}{2} \int_0^\pi X(\phi, \psi') \sin \psi' d\psi' \right) \sin \psi d\phi d\psi \quad (2.27)$$

$$= \frac{1}{8\pi} \left( \int_0^\pi \sin \psi d\psi \right) \int_A \int_0^\pi X(\phi, \psi') \sin \psi' d\psi' d\phi \quad (2.28)$$

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

$$= \int_M X d\mathbb{P}. \quad (2.30)$$

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  $X$ .  $\square$

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

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

and

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

as those versions wrong for the following reasons:

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 longitudes. 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. 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  $X$ .

4. 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  $\mathfrak{M}$ .
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  $2\pi$ , where it must be  $4\pi$ . This does however not impact further calculations, like the same mistake made from equation (109) to (110) in the section of longitudes.

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  $\psi_1 = \psi_1^* \bmod 2\pi$  and  $\psi_2 = \psi_2^* \bmod 2\pi$  and define arc  $A \subseteq M$  as

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

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 longitudes, let  $\mathbb{P}'$  be the measure taking 1 on meridian  $M$  and 0 on  $M^c$ , then the conditional probability of  $\mathbb{P}$  of points being on  $A$  with  $\psi_1, \psi_2 \leq \pi$  given there is a point on  $M$  is given by

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

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

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

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  $\psi_1, \psi_2$  as in the definition of  $A$  becomes

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

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

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

Clearly this conditional distribution on meridians is not uniform. However, one should expect they are, as a meridian is just a rotation of a longitude and the points on the sphere where spread uniformly.

### 2.3.4 Combining longitudes and meridians

Another question one could ask is the following: if I define  $\Sigma = \sigma(\mathfrak{C}, \mathfrak{M})$  as the smallest  $\sigma$ -algebra containing both measurable subsets of meridians and longitudes, what would the conditional probability distribution on a great circle be given  $\Sigma$ ? 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 Bayesian conditional probabilities defined in section 2.3.1.

First we'll further analyze the new  $\sigma$ -algebra  $\Sigma$ . 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

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

Thus we have proven that all Borel-measurable sets on our sphere are contained in  $\Sigma$ . Furthermore, as  $\mathfrak{C}$  and  $\mathfrak{M}$  contain only Borel-measurable sets,  $\Sigma = \sigma(\mathfrak{C}, \mathfrak{M})$  can only have Borel-measurable sets. Thus  $\Sigma = \mathcal{B}$  is the set 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  $\epsilon \rightarrow 0$ , which will take on the uniform distribution on  $C_0$  as calculated in equation (2.21).

**Proposition 2.3.3.** *Define the two rectangles  $R_\epsilon = (a, b) \times [x - \epsilon, x + \epsilon]$  and  $C_\epsilon = [0, 2\pi] \times [x - \epsilon, x + \epsilon]$  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*

$$\mathbb{P}[R_0|C_0] = \lim_{\epsilon \downarrow 0} \mathbb{P}[R_\epsilon|C_\epsilon] = \frac{b-a}{2\pi}. \quad (2.39)$$

*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

$$\mathbb{P}[R_\epsilon|C_\epsilon] = \frac{\mathbb{P}[R_\epsilon \cap C_\epsilon]}{\mathbb{P}[C_\epsilon]} = \frac{\int_a^b \int_{x-\epsilon}^{x+\epsilon} \sin \psi d\psi d\phi}{\int_0^{2\pi} \int_{x-\epsilon}^{x+\epsilon} \sin \psi d\psi d\phi} = \frac{b-a}{2\pi}. \quad (2.40)$$

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

$$\mathbb{P}[R_0|C_0] = \lim_{\epsilon \downarrow 0} \mathbb{P}[R_\epsilon|C_\epsilon] = \frac{b-a}{2\pi}. \quad (2.41)$$

□

Now consider 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*

$$\mathbb{P}[R_0|M_0] = \lim_{\epsilon \downarrow 0} \mathbb{P}[R_\epsilon|M_\epsilon] = \frac{1}{2}(\cos a - \cos b) \quad (2.42)$$

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

$$\mathbb{P}[R_\epsilon|M_\epsilon] = \frac{\mathbb{P}[R_\epsilon \cap M_\epsilon]}{\mathbb{P}[M_\epsilon]} = \frac{\int_{x-\epsilon}^{x+\epsilon} \int_a^b \sin \psi d\psi d\phi}{\int_{x-\epsilon}^{x+\epsilon} \int_0^\pi \sin \psi d\psi d\phi} = \frac{1}{2}(\cos a - \cos b). \quad (2.43)$$

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

Therefore the limit is  $\mathbb{P}[R_0|M_0] = \frac{1}{2}(\cos a - \cos b)$ , which is the same probability as in equation 2.37. Note that this probability has a normalization factor of  $\frac{1}{2}$  instead of  $\frac{1}{4}$ , as  $M_\epsilon$  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 Bayesian conditional probabilities on  $\Sigma$ . 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- $\sigma$ -algebra with  $\mathcal{G} \neq \mathcal{F}$  containing  $E$  and  $F$ . 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  $\mathbb{P}[F_n] > 0$ ,  $\mathbb{P}[E_n] > 0$  and  $E_n \subseteq F_n$  for all  $n \in \mathbb{N}$  and with existing limit  $\lim_{n \rightarrow \infty} \mathbb{P}[E_n|F_n]$ . Then*

$$\mathbb{P}[E|F] := \lim_{n \rightarrow \infty} \frac{\mathbb{P}[E_n \cap F_n]}{\mathbb{P}[F_n]} = \lim_{n \rightarrow \infty} \mathbb{P}[E_n|F_n] = \int_X \mathbb{E}[\mathbb{1}_F | \mathcal{G}] d\mathbb{P}'(E)$$

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 Bayesian probabilities. It however is false, but it is interesting to search for a slightly different conjecture that is true.

Note that conjecture 2.3.5 holds in the some cases for the Borel-Kolmogorov paradox. When considering longitudes we have  $\mathcal{F} = \mathcal{B}$  and  $\mathcal{G} = \mathfrak{C}$  and when considering meridians we have  $\mathcal{F} = \mathcal{B}$  and  $\mathcal{G} = \mathfrak{M}$ .

**Proposition 2.3.6.** *Conjecture 2.3.5 is false.*

*Proof.*  $\square$

## 2.4 The Borel-Kolmogorov phenomenon explained

## Chapter 3

# Safe probability



## Chapter 4

# Discrete paradoxes

## Chapter 5

# The two envelopes problem

## Chapter 6

# The Sleeping Beauty problem

# Bibliography

- [Bor09] É. Borel. éléments de la théorie des probabilités. Paris: A. Hermann. VI u. 200 S. 8° (1909)., 1909.
- [GHSR17] Z. Gyenis, G. Hofer-Szabó, and M. Rédei. Conditioning using conditional expectations: the Borel-Kolmogorov paradox. *Synthese*, 194(7):2595–2630, 2017.
- [Gr 18] Peter Grünwald. Safe probability. *Journal of Statistical Planning and Inference*, 195:47 – 63, 2018. Confidence distributions.
- [Kol33] A. Kolmogoroff. Grundbegriffe der wahrscheinlichkeitsrechnung. Ergebnisse der Math. und ihrer Grenzgebiete. 2, No. 3. Berlin: Julius Springer. iv, 62 pp. (1933), 1933.
- [Wil91] David Williams. *Probability with martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991.