Ma 3 — KC Border
Introduction to Probability and Statistics — Winter 2015

Chapter 1: Introduction

Relevant textbook passages:

Pitman [15]: Sections 1.1, 1.2, first part of 1.3, pp. 1–26.

Larsen–Marx [12]: Sections 1.3, 2.1, 2.2, pp. 7–26.

1.1 Uncertainty, randomness, and probability

Play *O Fortuna* before class. Time it to end at 10:00:30

Karl Orff's *O Fortuna* is a musical tribute to **Fortune**. The lyrics are from an irreverent 13th century poem attributed to student monks (http://en.wikipedia.org/wiki/O_Fortuna,http://en.wikipedia.org/wiki/Carmina_Burana). The poem paints a picture of Fortune as "*variabilis, semper crescis aut decrescis* [changeable, ever waxing and waning]." Fortune is associated with "*Sors immanis et inanis, rota tu volubilis, status malus* [Fate—monstrous and empty, you whirling wheel, you are malevolent]."

This view of Fortune, or randomness, or uncertainty, as monstrous and subject to no law save its own malevolence is an ancient view of randomness. See. e.g.. Larsen–Marx [12, §1.4]. Indeed some have gone so far as to suggest that it was this view of luck that kept the ancient Greeks from developing the insurance and financial infrastructure needed to conquer the world. Peter Bernstein [1, p. 1] writes (emphasis mine):

> What is it that distinguishes the thousands of years of history from what we think of as modern times? The answer goes way beyond the progress of science, technology, capitalism, and democracy.
>
> [...]
>
> *The revolutionary idea that defines the boundary between modern times and the past is the mastery of risk: the notion that the future is more than a whim of the gods and that men and women are not passive before nature. Until human beings discovered a way across that boundary, the future was a mirror of the past or the murky domain of oracles and soothsayers who held a monopoly over knowledge of anticipated events.*

But traces of the ancient view remain. It is perhaps this view of randomness as chaos, anarchy, and malevolence, that led Albert Einstein (in a 1926 letter to Max Born) to insist that

Gott würfelt nicht mit dem Universum.
[God does not play dice with the universe.]

One of my colleagues in applied math, , suggested that mixing probability and data analysis in a single course was dangerous because students might "believe that things are probabilistic." (I disagree with him on this point.) This view was also expressed by a Ma 2b student as, "But earthquakes don't happen at random. They happen for a reason."

Our view of luck and fortune began to change in the 17th century when Blaise Pascal and Pierre de Fermat began systematic investigations into games of chance.

We now understand that

Randomness is not simply anarchy.
It obeys mathematical laws.

It is these laws that we shall begin to study in this course.

1.2 Probability and its interpretations

Probability is our way of quantifying or measuring our uncertainty. We normalize it to be a number between 0 and 1 inclusive. The Institute has an entire course, (**HPS/Pl 122**. Probability, Evidence, and Belief) devoted to the interpretation of these numbers, but I shall briefly discuss the major views as I see them.

Pitman [15]: § 1.2

Frequentist school: The frequentist school views probabilities as long-run average frequencies. Joseph Hodges and Erich Lehmann [8, pp. 4, 9–10] put it this way:

> We shall refer to experiments that are not deterministic, and thus do not always yield the same result when repeated under the same conditions, as *random experiments*. Probability theory and statistics are the branches of mathematics that have been developed to deal with random experiments.
>
> [...]
>
> Data ... gathered from many sources over a long period of time, indicate the following *stability property of frequencies:* for sequences of sufficient length the value of [the frequency] f will be practically constant; that is, if we observed f in several such sequences, we would find it to have practically the same value in each of them. ...
>
> It is essential for the stability of long-run frequencies that the conditions of the experiment be kept constant. ... Actually, in reality, it is of course never possible to keep the conditions of the experiment exactly constant. There is in fact a circularity in the argument here:

> we consider that the conditions are *essentially* constant as long as the frequency is observed to be stable. ...
>
> The stability property of frequencies ... is not a consequence of logical deduction. It is quite possible to conceive of a world in which frequencies would not stabilize as the number of repetitions of the experiment become large. That frequencies actually do possess this property is an empirical or observational fact based on literally millions of observations. This fact is the experimental basis for the concept of probability ...

Putative examples:

- Coin tossing: the fraction of heads in repeated tosses tends to 1/2 (or does it?)
- The fraction of times two dice total 7 tends to 1/6.
- Games of chance, such as poker, roulette, and bridge, are full of examples of that frequentists would accepts as probabilities.

There are many conceptual problems with the frequentist approach. We often do not get enough observations to figure out probabilities. Moreover, one of the things we shall prove in this course is that if the probability that a coin toss results in Heads is 1/2, then the probability of getting n Heads in $2n$ tosses of a coin actually tends to zero, as n tends to infinity. So how could we ever figure out the frequentist probability? Do we just have to settle for statements like "the probability that a coin toss results in Heads is probably about 1/2?" [The answer, I believe, is yes.] For a vicious dissection of the frequentist approach see the papers by my former colleague, Alan Hájek [9, 10].

Empirical Probability: Empirical probabilities are observed frequencies in large samples. For example:

- The probability that a child is a boy. **Pitman [15]:** § 1.5

In the U.S. from 2000 through 2008, 51.2% of all live births were boys, so the probability of a child being born a boy is 0.512. (Source: U.S. Census Bureau, *Statistical Abstract of the United States, 2012*, Table 80. http://www.census.gov/compendia/statab/2012/tables/12s0080.pdf)

- Life Tables.

According to the U.S. Centers for Disease Control, National Vital Statistics Report, vol. 61, no. 3 (Sep. 24, 2012), http://www.cdc.gov/nchs/data/nvsr/nvsr61/nvsr61_03.pdf, Table 5, pp. 18–19:[1]

[1] There are two types of life tables: the cohort (or generation) life table and the period (or current) life table. The cohort life table presents the mortality experience of a particular birth cohort—all persons born in a particular year from the moment of birth through consecutive ages in successive calendar years. The drawback of a cohort table is doesn't lend itself to projecting the future mortality of those currently alive. The period life table tries to circumvent this

A U.S. white male has an 86.2% chance of surviving to age 60; and an 80.9% chance of living to age 65. Does that mean that a 60-year old white male has only a 80.9% chance of living to 65?

Ask the class!

No. Since he has already lived to 60, his chance of making to 65 is actually 80.9/86.2=93.9%. This is an example of **conditional probability** that we shall discuss in just a bit.

[You might ask, why did I look at the tables for white males? When I was a 60-year old white male, I had to decide whether to renew my term life insurance policy.]

Physical Probability and Initial Conditions: In this view, the probability of an event is derived from an analysis of the laws of physics. For example, consider coin tossing. We know the physics of rotating and falling objects, so the only uncertainty stems from not observing the initial conditions.

Example: Coin tossing:

- Karl Menger [14] provides a simple model of coin tossing in which the height h from which the coin was dropped and its angular velocity ω determined whether it turns up as Heads or Tails. The key point is the set of initial conditions (h, ω) contains an equal area of conditions that lead to Heads as Tails.

Here are the initial conditions that lead to hitting on edge after k half-turns.

$$h = c\frac{k^2\pi^2}{\omega^2} + 1, \qquad k = 1, 2, \ldots$$

where the coin has radius 1, and c depends on units and the acceleration of gravity. These loci are graphed for various k in Figure 1. The regions between these curves alternately produce Heads and Tails. See Figure 1.1.

- A more sophisticated model of the physics of tossing and catching a coin, due to Persi Diaconis, Susan Holmes, and Richard Montgomery [4] takes into account wobbling and precession, and a claibrated version of their model suggests that the probability a coin comes up in the same position it started is about 51%!

That is why your first assignment will be to toss coins, but more on that later.

- Andrzej Lasota and Michael Mackey's book [13], *Chaos, Fractals, and Noise* (1994), formerly known as *Probabilistic Properties of Deterministic Systems* (1985),[2] make a persuasive case that *chaotic* dynamics are best described in terms of probability.

Pitman [15]: pp. 16–17

Subjective Probability: The subjective school of probability treats all probabilities as statements about the **degree of belief** of an individual decision maker.

problem by looking at a particular reference year, and finding the death rate for each age in that year. (What fraction of those born in that year, died in their first year; what fraction of one year olds in that year died before age two, etc.) It then calculates what would happen to a cohort if the death rate at each age for the cohort is the same as the death rate for that age in the reference year. The table in this report is a period life table.

[2] The new title is a lot sexier and more marketable.

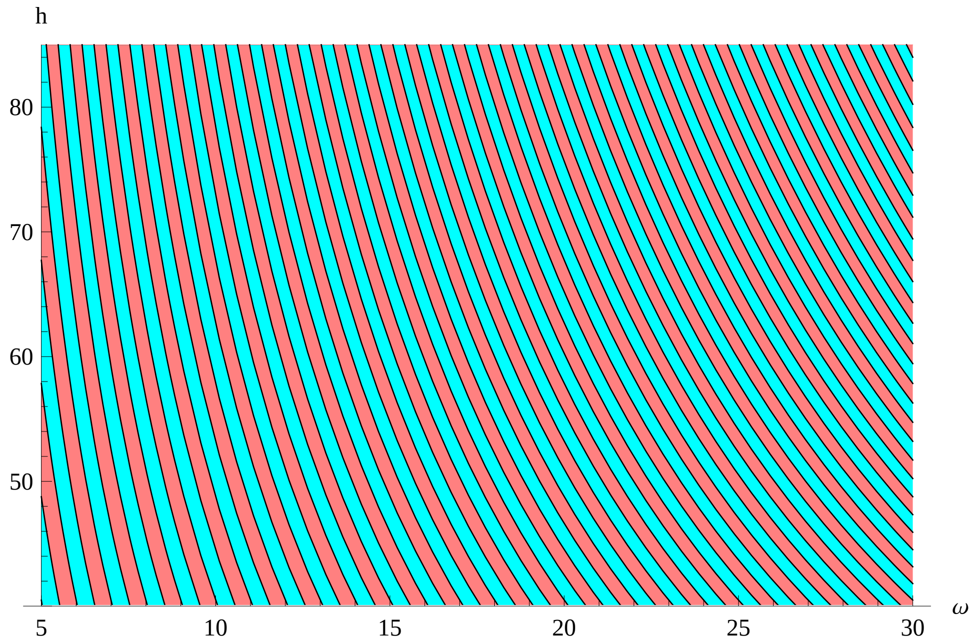

Figure 1.1: Initial conditions (ω, h) that land vertically in the Menger model are indicated by the thin black lines. Shaded areas in between these curves alternate in producing Head and Tails.

Bruno de Finetti, a first rate mathematician, takes the extreme view [6, p. x] that "in order to avoid becoming involved in a philosophical controversy," we should simply agree that "Probability does not exist." By that he means it has no independent existence outside of our minds.

Examples:

- Horse racing. There is an old saying that it takes a difference of opinion to make a horse race. Different bettors have different beliefs about which horse will win. These beliefs may be based on a variety of evidence, but it is unlikely to come from a physical model of the horses.

- Weather forecasting is partially subjective: This is why "skill scoring rules" were invented.

The practice of expressing weather forecasts in terms of rough probabilities was initiated in Western Australia by W. E. Cooke in 1905 [3]. Interestingly, his idea was criticized by E. B. Garriot [7] of the U.S. because "the bewildering complication of uncertainties it involves would confuse even the patient interpolator" and "our public insist upon having our forecasts expressed concisely and in unequivocal terms."

- The "Principle of Insufficient Reason" is often invoked to assign equal probabilities to events, and it is a form of subjective belief.

- The maximum entropy distribution is a more sophisticated version of the principle of insufficient reason for assigning probabilities. See, e.g., Edwin Jaynes [11] for a persuasive argument in this regard. It is usually not considered to be part of a subjectivist point of view, especially by its most ardent practitioners. They would argue that probability can and **must be deduced on logical grounds**.

One might question why purely subjective degrees of belief would obey the rules of probability that we are about to lay out. An answer was given by Bruno de Finetti [5]. He showed that if beliefs are not subject to the laws of probability, then they are *incoherent*. That is, if your subjective beliefs are not probabilistic, then you can be forced to lose money in a gambling situation.

1.2.1 An observation on random sequence generation

But before we go further, indulge me, and let me make the following outrageous claim.

> The following statement represents the opinion of the author, and does not necessarily reflect that of the California Institute of Technology or the Mathematics Department.
>
> The digits of π are a random sequence.

By this I mean that it you cannot predict how a sequence of digits of π will continue. For instance,

- What digit follows the following sequence:

$$3 \quad 1 \quad 4 \quad 1 \quad 5 \quad \dots$$

I hope most of you would say 9, because 9 is the fifth digit after the decimal point in the decimal expansion of π. But that is not necessarily the case. Let's see why, by examining the first Billion digits of π. I asked Mathematica 10 to compute π to a billion places, and it did so in 41 and a half minutes on my five-year old Mac Pro. (By the way, Mathematica 8 would only compute about 200 million digits before crashing.) I then asked it to count the number of occurrences of each digit. This took another 16 minutes plus change. Here are the digit counts:

digit	number	deviation
0	99,997,333	-2,667
1	100,002,411	2,411
2	99,986,912	-13,088
3	100,011,958	11,958
4	99,998,885	-1,115
5	100,010,387	10,387
6	99,996,061	-3,939
7	100,001,839	1,839
8	100,000,272	272
9	99,993,942	-6,058
	1,000,000,000	0

If the digits were evenly distributed you would expect about 100 million of each. You can see that we are close. Is it close enough? We shall learn later on about the marvelous chi-square test for uniformity, and see that if the digits were randomly generated, the distance from perfect uniformity would be greater than this about 26% of the time. That means that the digits of π pass this simple test for randomness.

But now let's get back to the question of what comes after 31415? By my count, the sequence 31415 occurs 10,010 times in the first billion and one digits of π. (When I asked Mathematica to write out the billon digits it actually wrote out about a billion and forty past the decimal point. I don't know why. So I kept the initial 3, threw out the decimal point and took the next billion digits. It took Mathematica 20 minutes to write the file to disk. But it took my Perl script 9 seconds to read the file and count the occurrences of 31415.) There are slightly less than a billion starting points for sequences of five consecutive digits in a billion and one digits. There are 100,000 different 5-digit sequences. If each were equally likely, there would be about 10,000 of each in a billion, so 10,010 is uncannily close. (Note that two sequences of 31415 cannot overlap, so each occurrence wipes out 4 more starting points. But that effect is negligible. The actual distribution is somewhat complicated, and later on I will give you some hints on how to deal with the problem.) More remarkable is what comes next. There are 100 different digit-pairs that can follow 31415. With 10,010 such pairs we would expect about 100 occurrences of each, and trust me, we get about that. We'll revisit this when we get to the chi-square test in about seven weeks.

There are other tests for randomness that we can perform, and as I wrote this, my computer is busy working on some of them. With luck, by the time class starts, I'll have more to say.

The point of all this is that

even though the sequence of digits in the decimal expansion of π is completely deterministic, it still makes a good random number generator, in the sense that if I do not tell you where I start in the sequence, you cannot tell what is coming next—the next digit behaves as if it were random. The digits of π are as random in this sense as a sequence of coin tosses.

(Actually, the digits of π are a terrible random sequence generator, because computing the sequence of digits of π is very time-consuming. It takes 41 minutes to generate a billion digits of π, but only 20 seconds to generate a billion pseudorandom digits using Mathematica's built-in `RandomInteger` function. If you are interested in algorithms to generate the digits of π you might want to start with this nice paper by Borwein, Borwein, and Bailey [2]. At the time it was an impressive accomplishment to generate the first billion digits of π.)

The idea that completely deterministic algorithms can mimic random processes is at the heart of "pseudorandom" number generation. My take is that pseudorandom numbers are as random as coin tosses. But the great John von Neumann opined, "Anyone who attempts to generate random numbers by deterministic means is, of course, living in a state of sin." (brainyquote.com)

1.2.2 The advantage of a formal model

My own view leans toward's de Finetti's, as I really want to avoid becoming embroiled in metaphysical controversies. But I am also impressed by all that empirical evidence that Lehmann and Hodges cite.

We shall take a "formal approach," to probability. That is, we shall introduce "primitive terms" and be careful with our reasoning. The advantage of this is that you don't have to grok the interpretation.

1.3 Administrative Details

Now that everyone has had a chance to arrive and settle in, it is a good time to over administrative details. Almost everything can be found on the course website at

`http://www.math.caltech.edu/~2014-15/2term/ma003/`

- Ma 3/103
- Kim Border, 205 Baxter, x4218, `kcb@caltech.edu`
- Office Hours?
- Lead TA: Marius Lemm, `mlemm@caltech.edu`
- **All questions regarding grading, extensions, late work, etc., should be directed to the Lead TA.**

- Course administrator: Kristy Aubry, 253 Sloan, x 4087, `kaubry@caltech.edu`

- **All requests for changes of section should be addressed to the Course Administrator.**

- Collaboration: See the course web page for details, but collaboration is encouraged on the homework, and not allowed on exams.

- Assignment 0: Coin tossing

1.4 A formal approach to probability

1.4.1 Experiments and sample spaces

A **random experiment** is something we observe that generates an **outcome** that will not be known or precisely predictable in advance. (For the time being I will leave the term random as a primitive.) The set of all possible outcomes is called the **sample space** or the **outcome space** of the experiment. The sample space is sometimes denoted S, or sometimes (as in Pitman) Ω.

Pitman [15]: § 1.3, p. 19

1.4.1 Example (Coin tossing) Consider the results of tossing a coin. The outcome of the toss could be either Heads, denoted H or tails, T, so we could take as our sample space the set:

$$S = \{H, \quad T\}.$$

Or perhaps we are willing to accept the possibility that the coin could land on edge, E. Then the sample space would be

$$S = \{H, \quad T, \quad E\}.$$

Or I might wish to include the possibility that my crazed Labrador Retriever might see this as an opportunity to demonstrate her talent for retrieving flying objects and snatch the coin out of the air, outcome L, so maybe the sample space should be

$$S = \{H, \quad T, \quad E, \quad L\}.$$

Or maybe the FBI would confiscate the coin in a counterfeiting investigation. (This is actually rather unlikely since the Secret Service investigates counterfeiting.)

The point is, the sample space is a *mathematical model* chosen by you the analyst, to represent the outcomes worthy of consideration. And for most uses, that means the sample space for a coin toss has two points,

$$S = \{H, \quad T\}.$$

□

1.4.2 Example (Repeated coin tossing) Now consider the results of tossing a coin three times. The outcome of each toss could be either Heads, denoted H or tails, T. (We won't consider Labradors or coins on edge, or intervention by aliens or the FBI.) With three tosses there are eight possible outcomes to the experiment, so we take as our sample space the set:

$$S = \{HHH, \quad HHT, \quad HTH, \quad HTT, \quad THH, \quad THT, \quad TTH, \quad TTT\}.$$

Clearly, if we toss a coin n times the sample space will contain 2^n outcomes. □

1.4.3 Example (Repeated coin tossing with a stopping rule) In this experiment, we toss a coin repeatedly until it comes up heads. The sample space for this experiment is quite large. In fact it is infinite, but denumerably infinite. It includes every finite sequence of n Tails followed by a single Head, for $n = 0, 1, 2, \ldots$, and it includes the infinite sequence of only Tails.

$$S = \{H, \quad TH, \quad TTH, \quad \cdots, \quad \underbrace{TT\cdots T}_{n}H, \quad \cdots, \quad \overline{TTTT\cdots}\ \}.$$

□

1.4.2 Events

The next element in our formal approach is the notion of an **event**. An event is simply an "observable" subset of the sample space. I use the word observable here as a primitive, and I will come back to that later. If the experiment produces an outcome $s \in S$ and s belongs to the event E, then we say that the event E **occurs** (or has occurred).

Pitman [15]: § 1.3, pp. 19–21
Larsen–Marx [12]: §2.2, pp. 18–27

The set of all events is denoted $\mathcal{E}$, (or sometimes, in keeping with a Greek theme, by Σ). Often, especially when the sample space is finite or denumerably infinite, $\mathcal{E}$ will consist of *all* subsets of S. [As you go on to study more mathematics, you will learn that there are problems with a nondenumerable sample space that force you to work with a smaller set of events.]

We require at a minimum that the set of events be an **algebra** or **field** of sets. That is, $\mathcal{E}$ satisfies:

1. $\varnothing \in \mathcal{E}$, $S \in \mathcal{E}$.

2. If $E \in \mathcal{E}$, then $E^{\mathrm{c}} \in \mathcal{E}$.

3. If E and F belong to $\mathcal{E}$, then EF and $E \cup F$ belong to $\mathcal{E}$.

Most probabilists assume further that $\mathcal{E}$ is a **σ-algebra** or **σ-field**, which requires in addition that

3′. If $E_1, E_2, \ldots$ belong to $\mathcal{E}$, then $\bigcap_{i=1}^{\infty} E_i$ and $\bigcup_{i=1}^{\infty} E_i$ belong to $\mathcal{E}$.

Note that if S is finite and $\mathcal{E}$ is an algebra then, it is automatically a σ-algebra. Why?

The reason for these properties is that we think of events as having a description on some language. Then we can think of the descriptions being joined by *or* or *and* or *not*. The correspond to union, intersection, and complementation.

1.4.4 Example (Coin tossing events) For the sample space in Example 1.4.3, Coin Tossing until Heads, let $\mathcal{E}$ be the set of all subsets of S. We can consider events such as

$$E = \text{the first Head occurs on an odd-numbered toss} = \{H, TTH, TTTTH, \cdots\}$$
$$F = \text{the first Head occurs on an even-numbered toss} = \{TH, TTTH, TTTTTH, \cdots\}$$
$$G = \text{Heads never occur} = \{\overline{TTTT\cdots}\}.$$

Note that $E \cup F \neq S$, but $(E \cup F)^{\mathrm{c}} = G$, and $EF = \varnothing$. □

Aside: The notion of the set of events as a set of subsets of S my seem unwieldy. You may be used to thinking of sets of points, not of sets. But you have used such collections for years. Think of the set of intervals on a line, or the set of triangles in a plane. These are all sets of sets.

In every real application I can think of, for each outcome $s \in S$, the singleton set $\{s\}$ is an event, that is, $\{s\} \in \mathcal{E}$. [Note that $s \notin \mathcal{E}$!]

1.4.3 Probability measures

A **probability measure** or **probability distribution** (as in Pitman) or simply a **probability** (although this usage can be confusing) is a **set function** $P \colon \mathcal{E} \to [0, 1]$ that satisfies:

Larsen–Marx [12]: §2.3, pp. 27–32
Pitman [15]: §1.3, pp. 19–32

Normalization $P(\varnothing) = 0$; and $P(S) = 1$.

Additivity If $E \cap F = \varnothing$, then $P(E \cup F) = P(E) + P(F)$.

Most probabilists require the following stronger property, called **countable additivity**:

Countable additivity $P\Big(\bigcup_{i=1}^{\infty} E_i\Big) = \sum_{i=1}^{\infty} P(E_i)$ provided $E_i \cap E_j = \varnothing$ for $i \neq j$.

Aside: You need to take an advanced analysis course to understand that there can be probability measures that are additive, but not countably additive. So don't worry too much about it.

Note that while the domain of P is technically $\mathcal{E}$, the set of events, we may also refer to P as a probability (measure) on S, the set of samples.

To cut down on the number of delimiters in our notation, when a set is delimited with braces or with statistician's notation, we may omit the parentheses surrounding it and simply write something like $P\,(f(s) = 1)$ or $P\{s \in S : f(s) = 1\}$ instead of $P\big(\{s \in S : f(s) = 1\}\big)$ and we may write $P(s)$ instead of $P\big(\{s\}\big)$. You will come to appreciate this.

1.4.5 Definition *For any event E, let $|E|$ denote the number of elements of E.*

1.4.6 Theorem (Uniform probability) *Consider the case where S is finite and $\mathcal{E}$ contains all subsets of S. Enumerate S as $S = \{s_1, \dots, s_n\}$. Then $1 = P(S) = P(s_1) + \cdots + P(s_n)$. (Why?) If each outcome is equally likely (has the same probability), then $P(s_1) = \cdots = P(s_n) = 1/n$, and*

$$P(E) = \frac{|E|}{n}.$$

1.4.7 Example (Coin Tossing) We usually think of a coin as being equally likely to come up H as T. That is, $P\{H\} = P\{T\}$. If our sample space is the simple $S = \{H, T\}$ and $\mathcal{E}$ is all four subsets of S, $\mathcal{E} = \big\{\varnothing, S, \{H\}, \{T\}\big\}$, then

$$\{H\}\{T\} = \varnothing \text{ and } \{H\} \cup \{T\} = S$$

so $P\{H\} = P\{T\}$ implies

$$1 = P(S) = P\big(\{H\} \cup \{T\}\big) = P\{H\} + P\{T\},$$

which implies

$$P\{H\} = P\{T\} = 1/2.$$

□

Bibliography

[1] P. L. Bernstein. 1996. *Against the gods: The remarkable story of risk*. New York: Wiley.

[2] J. M. Borwein, P. B. Borwein, and D. H. Bailey. 1989. Ramanujan, modular equations, and approximations to pi or how to compute one billion digits of pi. *American Mathematical Monthly* 96(3):201–219.
http://www.jstor.org/stable/2325206

[3] W. E. Cooke. 1906. Forecasts and verifications in Western Australia. *Monthly Weather Review* 34(1):23–24.
http://docs.lib.noaa.gov/rescue/mwr/034/mwr-034-01-0023.pdf

[4] P. Diaconis, S. Holmes, and R. Montgomery. 2007. Dynamical bias in the coin toss. *SIAM Review* 49(2):211–235. DOI: 10.1137/S0036144504446436

[5] B. de Finetti. 1937. La prevision: Ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré* 7(1):1–68. Translated as "Foresight: Its Logical Laws, Its Subjective Sources" in *Studies in Subjective Probability*, H. E. Kyburg, Jr. and H. E. Smokler, eds., Robert E. Krieger Publishing, Huntington, New York, 1980, pages 53–118.
http://www.numdam.org/item?id=AIHP_1937__7_1_1_0

[6] ——— . 1974. *Theory of probability*, volume 1. London: Wiley.

[7] E. B. Garriott. 1906. Note by Prof. E. B. Garriott. *Monthly Weather Review* 34(1):24.
http://docs.lib.noaa.gov/rescue/mwr/034/mwr-034-01-0023.pdf

[8] J. L. Hodges, Jr. and E. L. Lehmann. 2005. *Basic concepts of probability and statistics*, 2d. ed. Number 48 in Classics in Applied Mathematics. Philadelphia: SIAM.

[9] A. Hájek. 1996. “Mises redux”—redux: Fifteen arguments against finite frequentism. *Erkenntnis* 45(2–3):209–227. DOI: 10.1007/BF00276791

[10] ——— . 2009. Fifteen arguments against hypothetical frequentism. *Erkenntnis* 70(2):211–235. DOI: 10.1007/s10670-009-9154-1

[11] E. T. Jaynes. 1968. Prior probabilities. *IEEE Transactions on Systems Science and Cybernetics* 4(3):227–241. DOI: 10.1109/TSSC.1968.300117

[12] R. J. Larsen and M. L. Marx. 2012. *An introduction to mathematical statistics and its applications*, fifth ed. Boston: Prentice Hall.

[13] A. Lasota and M. C. Mackey. 1994. *Chaos, fractals, and noise*, 2d. ed. New York: Springer–Verlag. Second edition of *Probabilistic Properties of Deterministic Systems*, published by Cambridge University Press, 1985.

[14] K. Menger. 1954. Tossing a coin. *American Mathematical Monthly* 61(9):634–636. http://www.jstor.org/stable/2307682

[15] J. Pitman. 1993. *Probability*. Springer Texts in Statistics. New York, Berlin, and Heidelberg: Springer.