Modeling Mindsets

The Many Cultures of Learning From Data

Christoph Molnar

Modeling Mindsets

The Many Cultures of Learning From Data

Christoph Molnar

2023, First Edition

ISBN 9798358729339 (paperback)

Christoph Molnar
c/o MUCBOOK, Heidi Seibold
Elsenheimerstraße 48
80687 München, Germany

commit id: f9ada8d

Contents

Preface

Is a linear regression model statistical modeling or machine learning? This is a question that I have heard more than once, but it starts from the false premise that models somehow belong to one approach or the other. Rather, we should ask: What's the mindset of the modeler?

Differences in mindsets can be subtle yet substantial. It takes years to absorb a modeling mindset because, usually, the focus is on the methods and math rather than the mindset. Sometimes modelers aren't even aware of their modeling limitations due to their mindset. I studied statistics, which gave me an almost pure frequentist mindset. Because of that narrow lens on modeling, I often hit a wall in my modeling projects, ranging from a lack of causal reasoning to crappy predictive models. I couldn't solve these problems by doubling down on my current mindset. Instead, I made the most progress when I embraced new modeling mindsets. Modeling Mindsets is the book I wish I had read earlier to save myself time and headaches.

The inspiration for Modeling Mindsets was the article "Statistical Modeling: The Two Cultures" by the statistician Leo Breiman. His article was the first to show me that modeling is not only about math and methods but it's about the mindset through which you see the world. Leo Breiman's paper is over 21 years old and hasn't lost any of its relevance even today. Modeling Mindsets builds on the same principle of making mindsets explicit and embraces the multiplicity of mindsets. I kept Modeling Mindsets short, (mostly) math-free and poured over a decade of modeling experiences into it. My hope is that this book will still be relevant 21 years from now and a great investment for you.

Who This Book is For

This book is for everyone who builds models from data: data scientists, statisticians, machine learners, and quantitative researchers.

To get the most out of this book:

- You should already have experience with modeling and working with data.
- You should feel comfortable with **at least one of the mindsets** in this book.

Don't read this book if:

- You are completely new to working with data and models.
- You cling to the mindset you already know and aren't open to other mindsets.

You will get the most out of Modeling Mindsets if you keep an open mind. You have to challenge the rigid assumptions of the mindset that feels natural to you.

1 Introduction

Every day, people use data for prediction, automation, science, and making decisions. For example:

- Identifying patients prone to drug side effects.
- Finding out how climate change affects bee hives.
- Predicting which products will be out-of-stock.

Each data point has details to contribute:

- Patient with ID 124 got acne.
- Bee colony 27 shrank during the drought in 2018.
- On that one Tuesday, the flour was sold out.

Data alone aren't enough to solve the above tasks. Data are noisy and high-dimensional – most of the information will be irrelevant to the task. No matter how long a human analyzes the data, it's difficult to gain insights just by sheer human willpower.

Therefore, people rely on models to interpret and use data. A model simplifies and represents an aspect of the world. Models learned from data – the focus of this book – glue together the raw data and the world. With a model, the modeler can make predictions, learn about the world, test theories, make decisions and communicate results to others.

Models Have Variables And Learnable Functions

There is no philosophical consensus on the definition of a model. For our purpose, we'll go with this definition: **a**

mathematical model that consists of variables and functions.

Variables represent aspects of the data and the model:

- A numerical variable can contain the blood pressure of a patient.
- A 3-dimensional variable may represent the colors of a pixel.
- Variables can also represent abstract aspects like happiness.
- A cluster variable represents data grouping.

Variables have different names in different mindsets: Random variables, covariates, predictors, latent variables, features, targets, and outcomes. The names can reveal the role of a variable in a model: For example, the "target" or "dependent variable" is the variable the modeler wants to predict.

Functions relate the variables to each other (Weisberg 2012), see also Figure 1.1.

- A linear regression model expresses one variable as a weighted sum of the other variables.
- In deep Q-learning – a form of reinforcement learning – the value of an action (variable) is a function of the state of the environment (variables).
- A clustering model can be a function that takes, as input, the variables of the data point and returns its cluster (variable).
- The joint distribution of two variables describes the co-occurrence of certain variable values in terms of probability.
- A causal model represents the causal relationship between variables.

Functions range from simple, like $Y = 5 \cdot X$, to complex, like a deep neural network with millions of parameters.

So far, we have discussed variables and functions, the ingredients of a model, but not how they are connected to data. Modelers use data to find the best[1] function to relate the

[1] What the best model is depends on the mindset.

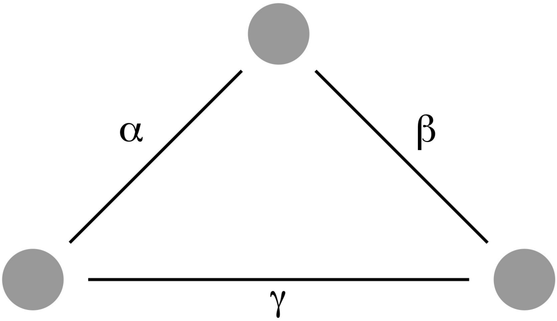

Figure 1.1: A mathematical model sets variables (dots) into relation (lines) using parameterized functions.

variables. Depending on the mindset of the modeler, the process of adapting a model's function is called estimation, training, fitting or learning. In this process, the model function is optimized using data:

- In a linear regression model, the coefficients (also called weights) are optimized to minimize the squared difference between the weighted sum and the target variable.
- K-means clustering cycles between assigning data points to cluster centers and optimizing the centers to minimize distances.
- A decision tree is grown by finding split points in variables using data.

Models Are Embedded In Mindsets

The interpretation and use of the model can't be derived from the model itself. Two mathematically identical models might be used in different ways by different modelers. The use of a model depends on the mindset.

A model is a mathematical construct that doesn't contain the purpose of the model. The purpose of the model – how to use and interpret it – depends on the modeling mindset. To derive knowledge about the world from the model, modelers need to make further assumptions.

Consider a linear regression model that predicts regional rice yield as a function of rainfall, temperature, and fertilizer use. It's a model, but interpretation needs a mindset:

- Can the modeler interpret the effect of fertilizer as causal to rice yield? *Yes, if based on a causal model.*
- Is the model good enough for predicting rice yields? *Depends if the modeler had a supervised learning mindset and has evaluated the generalization error properly.*
- Is the effect of fertilizer on yield significant? *This requires a frequentist mindset.*

A Mindset Is A Perspective Of The World

A modeling mindset provides the framework for modeling the world with data (Figure 1.2). Modeling means investigating a real-world phenomenon indirectly using a model (Weisberg 2007). Modeling mindsets are like different lenses. All lenses show us the world, but with a different focus. Some lenses magnify things that are close, and other lenses detect things that are far away. Also, some glasses are tinted so you can see in bright environments.

Mindsets differ in how they interpret probabilities – or whether probabilities are central to the mindset at all. While mindsets cover many different modeling tasks, they have some specific tasks where they really shine. Each mindset allows different questions. Hence, it shapes how to view the world through the model. In supervised machine learning, for example, everything becomes a prediction problem, while in Bayesian inference, the goal is to update beliefs about the world using probability theory.

A modeling mindset limits the questions that can be asked. Some tasks are out of scope because they don't make sense

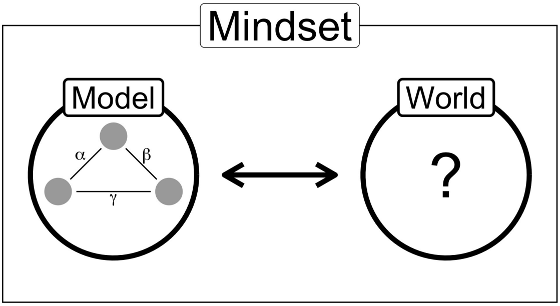

Figure 1.2: The real-world purpose of the model depends on the modeling mindset.

in a particular modeling mindset. Supervised machine learners formulate tasks as prediction problems. Questions about probability distributions are out of scope since the mindset is: to choose the model with the lowest generalization error given new data. So the best model could be any function, even an ensemble of many models (random forest, a neural network, linear regression model). If the best model can be any function, questions that a statistician would normally ask (hypothesis tests, parameter estimates) become irrelevant, as the modeler can't guarantee that the best model is a statistical model. Choosing a suboptimal model would violate the purely supervised learning mindset.

Each mindset has a different set of permissible models and ways to evaluate how good a model is. These sets may, however, overlap – for example, linear regression models are both used in frequentist inference and supervised learning. But whether the linear regression is a good model for a given task is evaluated in different ways, as we will discuss.

Mindsets Are Cultural

Modeling mindsets are not just theories; they shape communities and are shaped by people who apply the mindset. In many scientific communities, the frequentist mindset is very common. I once consulted a medical student for his doctoral thesis, and I helped him visualize some data. A few days later, he came back, "I need p-values with this visualization." His advisor had told him that any data visualization needed p-values. His advisor's advice was a bit extreme and not an advice that an experienced statistician would have given. However, it serves as a good example of how dominant a mindset can be and perpetuated by the community. Likewise, if you were trying to publish a machine learning model in a journal that publishes only Bayesian analysis, I would wish you good luck.

The people within a shared mindset also accept the assumptions of that mindset. These assumptions are usually not challenged but mutually agreed upon. At least implicitly. In a team of Bayesians, whether or not to use priors won't be challenged for every model. In machine learning competitions, the model with the lowest prediction error on new data wins. You will have a hard time arguing that your model should have won because it's the only causal model. Modelers that find causality important wouldn't participate in such a challenge. Only modelers that have accepted the supervised learning mindset will thrive in such machine learning competitions.

Mindsets Are Archetypes

The modeling mindsets, as I present them in this book, are archetypes: pure and extreme forms of these mindsets. In reality, the boundaries between mindsets are much more fluid:

- A data scientist who primarily builds machine learning models might also use regression models with hypothesis tests – without cross-validating the models' generalization error.
- A research community could accept analyses that use both frequentist and Bayesian models.

- A machine learning competition could include a human jury that awards additional points if the model is interpretable and causal.
- Ideas from one mindset might be applied in another: A modeler might use statistical models but evaluate them with the generalization error as typical in supervised learning.

Have you ever met anyone who is really into supervised learning? The first question they ask is, “Where are the labels?” The supervised machine learner turns every problem into a prediction problem. Or perhaps you’ve worked with a statistician who always wants to run hypothesis tests and regression models? Or have you had an intense discussion with a hardcore Bayesian about probability? Some people are walking archetypes of singular mindsets. But most people learned more than one mindset and embraced bits of other mindsets. Most people’s mindset is already a mixture of multiple modeling mindsets, and that’s a good thing. Having an open mind about modeling ultimately makes you a better modeler.

Mindsets Covered In This Book

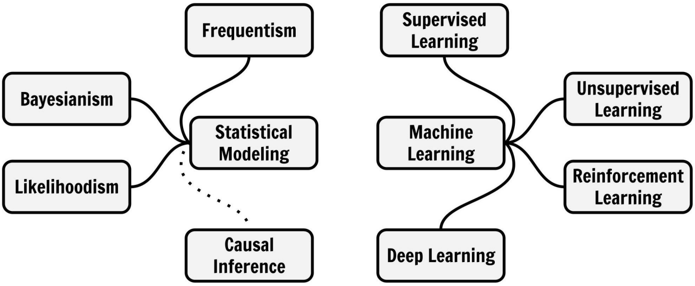

Figure 1.3: Overview of mindsets

- Statistical modeling and machine learning are “parent” mindsets.

- Frequentism, likelihoodism, and Bayesianism are flavors of statistical modeling.
- Supervised, unsupervised, reinforcement, and deep learning are flavors of machine learning.
- Causal inference can work with machine learning, but is culturally closer to statistical modeling.

For example, to understand Bayesian inference, you need to read the chapters on statistical modeling and Bayesianism.

The book is not exhaustive: some mindsets are *not covered*, although they are also data-driven. For example, the book doesn't cover mindsets based on sampling theory, experiments, or visualization. Also, some flavors such as self-supervised learning or non-parametric statistics didn't make it into this first edition of the book. If the book is popular, I'll write a second edition with more mindsets. Nevertheless, Modeling Mindsets already covers many of the most popular mindsets.

2 Statistical Modeling – Reason Under Uncertainty

Premise: The world is best approached through probability distributions.

Consequence: Estimate aspects of these distributions using statistical models to reason under uncertainty.

The statistician had come to duel the monster of randomness. The fight was fierce, and the statistician was driven back, step by step. But with every attack, the statistician learned more about the monster and suddenly realized how to win: figure out the data-generating process of the monster. With one final punch, the statistician separated signal from randomness.

Do you become more productive when you drink a lot of water? Your productivity will vary naturally from day to day – independent of hydration. This uncertainty obscures the effect of water intake on productivity. Uncertainty so often stands between data and clear answers.

Statistical modeling offers mathematical tools to handle uncertainty.[1] The rough idea is that the data are generated by a process that involves probability distributions or at least can be represented by distributions. Statistical models approximate the process by relating variables and specifying their distribution (in full or in part). The models are estimated with data, and modelers interpret them to reason under uncertainty.

[1] This chapter looks at the parts of statistical modeling that Bayesian inference, frequentist inference and likelihoodism share.

Every "Thing" Has A Distribution

Statistical modelers think in random variables. These variables are mathematical objects that encode uncertainty in the form of probability distributions. In statistical modeling, data are realizations of these variables (see Figure 2.1). Someone drinking 1.7 liters of water is a realization of the variable "daily water intake," but also a sign that this person might not be drinking enough. Other examples of variables:

- Outcome of a dice roll.
- Number of ducks in a park.
- Whether a customer canceled their contract last month.
- Daily number of duck attacks.
- The color of a t-shirt.
- Pain on a scale from 1 to 10 due to duck bites.

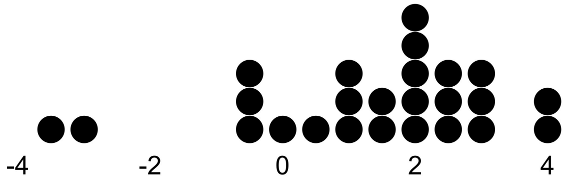

Figure 2.1: Each dot is a realization of a variable. The x-axis shows the variable's values. Dots are stacked by frequency

Probability distributions describe how variables behave. A distribution is a function that assigns a probability to each possible outcome of a variable. Value in, probability out. For the outcome of a fair dice, there are six possible outcomes, each with a probability of 1/6. For continuous outcomes such as temperature, the Normal distribution is a common choice. See Figure 2.2 on the upper left.[2] Part of statistical modeling is to pick the best distributions that match the nature of the variables. Each distribution has parameters that modify it, such as mean or variance. That's an important property as it

[2] For continuous probability distributions the probability is given by an integral over a range of values.

allows fitting distributions to data. However, distributions alone aren't expressive enough for modeling – the modeler creates statistical models.

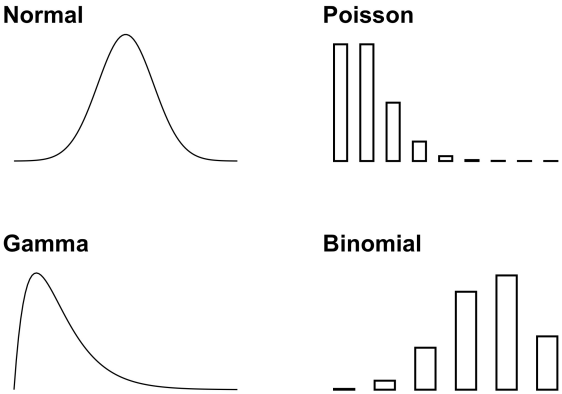

Figure 2.2: Distributions

Models Encode The Data-Generating Process

The modeler translates the data-generating process into a statistical model. All kinds of assumptions go into the model:

- Choice of relevant variables.
- How variables are distributed.
- If data are independent and identically distributed (IID).
- How variables are related to each other (such as linear, smooth effects, coding of categorical variables).

Let's say the modeler has data from an experiment where participants recorded both their daily water intake and the number of completed tasks. The modeler might assume that productivity, given water intake and day of the week, follows a Poisson distribution. More specifically, the modeler expresses the mean of the productivity distribution as a function of water intake and day of the week. Each participant appears multiple times in the data (multiple days), so the modeler makes assumptions that there is an additional "participant" effect that has to be accounted for. The modeler thinks through all these nuances of the data-generating process and encodes them in the statistical model.[3] This focus stands in contrast with machine learning, where solving the task at hand (like prediction) is more important than replicating the data-generating process.

You can see statistical models as having a fixed part, which are all these assumptions. In the next step, the model has to be combined with the data. That's where the flexible part of the model comes into play. Modelers "fit" models by adapting the parameters[4], such as regression coefficients, so that the data seem likely under the statistical model – aka having a high likelihood (see Figure 2.3). This fit is measured and optimized via the likelihood function, which takes as input the parameters and returns how likely the data are given these parameters. This makes model fitting an optimization problem (maximum likelihood estimation). While likelihoodists and frequentists directly maximize the likelihood function, Bayesians estimate the posterior parameter distribution, which is more complex than likelihood maximization alone.

There is another angle to model fitting: estimating parameters. Usually, the parameters within the model describe aspects of the distribution that the modeler is interested in. This can be a coefficient targeting the (conditional) mean, the correlation between variables, or even the full distribution.[5] In the productivity example, the modeler might be interested in the coefficients (parameters) that relate water intake and day of the week to the mean of the productivity distribution. The process of modeling can also be seen

[3] For the connoisseurs among you: The modeler might end up with something like a mixed-effects Poisson regression model.

[4] Statistical modeling has also distribution-free and non-parametric methods, which both relax some assumptions.

[5] Why not always model the full data distribution? Because it's complex or requires strong assumptions. Many questions can be answered with, for example, conditional distributions (like treatment effects).

as estimating model parameters such as these coefficients. Statistical modelers often draw conclusions via model parameters about real-world phenomena. Interpretation is often supported by element-wise representation (Freiesleben et al. 2022): All relevant parts of the data-generating process are represented within the model as variables, parameters or functions. Frequentists and Bayesians focus more on the estimation aspect, and likelihoodists focus more on the fitting aspect: likelihoodists compare models based on the likelihood ratio, which can tell which model fits better.

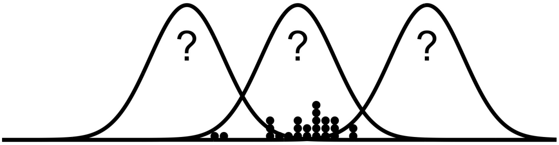

Figure 2.3: Fitting distributions to data

Good Models Satisfy Assumptions And Fit Data

The evaluation of statistical models consists of two parts: model diagnostics and goodness-of-fit evaluation.

The role of model diagnostics is to check whether the modeling assumptions are reasonable. This check is often visual: If a modeler assumes a variable follows a Normal distribution, they can check this assumption with a plot. Another assumption is homoscedasticity: The variance of the target is independent of other variables. Homoscedasticity can be checked with a residual plot, which shows on the x-axis the values of a variable and on the y-axis the residuals (actual target minus its predicted value) against each of the other variables. A Bayesian can verify with a posterior predictive check that the observed distribution of the target variable matches what the posterior distributions simulate.

A model that passes the diagnostic tests is not automatically a good model because it might not fit the data well enough. Statistical modelers use goodness-of-fit measures to compare different models and evaluate modeling choices, such as which variables to integrate into the model. The fit can be evaluated with goodness-of-fit measures bearing names such as R-squared, Akaikes Information Criterion, and the Bayes factor.

Goodness-of-fit is often computed with the same data used for fitting the statistical models. This choice may look like a minor detail, but it says a lot about the statistical modeling mindset.[6] The critical factor here is overfitting: The more flexible a model is, the more it might copy the randomness in the data. Many goodness-of-fit metrics, therefore, account for model complexity. In comparison, supervised learning relies on evaluation schemes that use unseen data to avoid overfitting and obtain honest estimates of the generalization error.

Models Enable Conclusions

Statistical models can help make decisions, understand the world, and make predictions. But using the model as a representation of the world isn't for free. The modeler must consider the representativeness of the data and make further (philosophical) assumptions.

Are the data representative of the population studied? Let's say modelers analyze data on whether a sepsis screening tool reduced the incidence of sepsis in a hospital. They concluded the tool reduced sepsis-related deaths at that hospital and want to make a recommendation to use it in all hospitals in the country. But are the data representative of all hospitals in the country? Or is there a reason why the patients of the

[6] A reminder that these mindsets are archetypes: some modelers only use statistical models but evaluate them with cross-validation as common in supervised learning.

studied hospital differ? A good modeler defines the population to be studied and discusses whether the data represent this population well.

When it comes to the modeler's attitude toward the nature of probability and what counts as evidence, the matter becomes philosophical:

> "It is unanimously agreed that statistics depends somehow on probability. But, as to what probability is and how it is connected with statistics, there has seldom been such complete disagreement and breakdown of communication since the Tower of Babel."

– Leonard Savage, 1972[7]

Different interpretations of probability and evidence lead to different flavors of the statistical modeling mindset:

- Frequentist inference sees probability as relative frequencies in the long run.
- Bayesian inference is based on an interpretation of probability as a degree of belief about the world.
- Likelihoodism equates the likelihood of a model as evidence for a hypothesis.

All these flavors of statistical modeling have a probabilistic view of the world in common.

Strengths & Limitations

+ Distributions and statistical models provide a language to describe the world and its uncertainties. The same language is even used to describe and understand machine learning algorithms under the label of statistical learning.

[7] Savage, Leonard J. The foundations of statistics. Courier Corporation, 1972.

+ Statistical modeling has an extensive theoretical foundation: From measurement theory as the basis of probability to thousands of papers for specific statistical models.

+ The data-generating process is a powerful mental model that encourages asking questions about the data.

+ Statistical models provide the means to reason under uncertainty, such as making decisions, understanding the world, evaluating hypotheses, and making predictions.

– Modeling the data-generating process can be quite manual and tedious, as many assumptions have to be made. More automatable mindsets, such as supervised learning, are more convenient.

– The statistical modeling mindset struggles with complex distributions, like image and text data. This is where machine learning and especially deep learning shine.

– Goodness-of-fit doesn't guarantee high predictive performance on new data. If the goal is to make predictions, a supervised learning approach is more suitable.

3 Frequentism – Infer "True" Parameters

Premise: The world is best approached through probability distributions with fixed but unknown parameters.

Consequence: Estimate and interpret the parameters using long-run frequencies to make decisions under uncertainty.

Once upon a time, there was a frequentist who dealt with p-values. She was the best in the business, and people would flock to her from all over to get their hands on her wares. One day, a young scientist came to the frequentist's shop. The scientist said: "I'm looking for the p-value that will allow me to get my paper published." The frequentist smiled and reached under the counter. She pulled out a dusty old book and thumbed through the pages until she found what she was looking for. "The p-value you're looking for is 0.05," she said. The scientist's eyes lit up. "That's exactly what I need! How much will it cost me?" The frequentist leaned in close and whispered, "It will cost your soul."

Drinking alcohol is associated with a 1.81 higher risk of diabetes in middle-aged men. At least, this is what a study claims (Kao et al. 2001). The researchers modeled type II diabetes as a function of variables such as alcohol intake. The researchers used frequentist inference to draw this conclusion from the data. There is no particular reason why I chose this study other than it's a typical frequentist analysis. Modelers[1] that think in significance levels, p-values, hypothesis tests, and confidence intervals are likely frequentists.

[1] Why I am not using the term "statistician" here: Statisticians do more than just statistical modeling. They visualize data, plan experiments, collect data, complain about machine learning, design surveys, and much more.

In many scientific fields, such as medicine and psychology, frequentist inference is the dominant modeling mindset. Frequentist papers follow similar patterns, make similar assumptions, and contain similar tables and figures. Understanding frequentist concepts such as confidence intervals and hypothesis tests is, therefore, one of the keys to understanding scientific progress. Frequentism has a firm foothold in the industry as well: Statisticians, data scientists, and whatever the future name of the role will be, often use frequentist inference, from analyzing A/B tests for a website to calculating portfolio risk to monitoring quality on production lines.

Probability Is A Long-Run Frequency

Frequentist inference is a statistical modeling mindset relying on variables, distributions, and statistical models. But it comes with a specific interpretation of probability: Probability is seen as the relative frequency of an event in infinitely repeated trials. But how do these long-run frequencies help gain insights from the model?

Let's go back to the 1.81 increase in diabetes risk among men who drink a lot of alcohol. 1.81 is larger than 1, so there seems to be a difference between men who drink alcohol and the ones who don't. But how can the researchers be sure that the 1.81 is not a random result? If you flip a coin 100 times, and it comes up with tails 51 times, our gut-feeling would say that the coin is fair. But where is the threshold? Is it 55 tails? 60? Viewing probability as long-run frequencies allows the modeler to define a threshold to decide when the coin is unfair. In addition, frequentists assume that each parameter has a fixed, true value. Repeated observations reveal the true values in the long run. The coin has a fixed probability for tails, and the modeling goal is to estimate (=uncover) it and quantify how certain the estimate is.

The researchers of the study applied the same type of frequentist thinking to decide whether the effect of alcohol is random or not. In this study, the parameter of interest was a

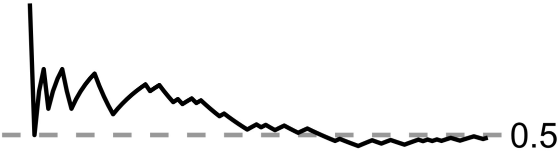

Figure 3.1: The line shows how the relative frequency of tails changes as the number of coin tosses increases from 1 to 100 (left to right).

coefficient in a logistic regression model. The diabetes study reported a 95% confidence interval for the alcohol coefficient ranging from 1.14 to 2.92. The interval doesn't contain 1, so the researchers concluded that alcohol is significantly associated with diabetes in men. We will discuss confidence intervals in more detail later in this chapter, as they are typically frequentist.

Imagined Experiments Underpin Inference

The best way to understand frequentism is to slowly work our way from the data and statistical model to the estimator to imagined experiments and finally to a decision.

1) Data Are Random Variables

Assume a modeler has observed 10 independent draws of a variable. Let's say we are talking about the weights of 10 apples. The modeler first assumes a statistical model for the weight variable by saying that it follows a Gaussian distribution. The modeler wants to know A) the average weight of an apple, including uncertainty estimates, and B) whether the average apple weighs less than 90 grams.

2) Estimators Are Random Variables

To answer the questions, the modeler first estimates the mean of the apple weight distribution. The estimation, in this case, is simple: add all observed values and divide by 10. Let's say the result is 79 grams. As a function of a variable, this mean estimator is a random variable itself. And the estimated mean of 79 is just one observation of that variable. The mean estimator is a rather simple estimation for a simple statistical model. We can imagine more complex models for the apples:

- A statistical model that relates the weight of an apple with its "redness".
- The weight distribution might be estimated conditional on variables such as apple type and season.
- Another model might target the 90% quantile of apple weights.

3) Long-Run Frequencies Interpretation

79 is smaller than 90 (question B). But the estimator is a random variable that comes with uncertainty. Assuming that the true average weight is 90, couldn't a sample of 10 apples weigh 79 grams on average, just by chance?

Bad news: The modeler has only one observation of the mean estimate and no motivation to buy more apples. Who on earth should eat all those apples? Good news: The modeler can derive the distribution of the mean estimate from the data distribution. Since the apple weights follow a Gaussian distribution, the modeler concludes that the mean follows a t-distribution.[2] The distribution of the mean has to be interpreted in terms of long-run frequency: It tells the modeler what to expect of the mean estimates in future experiments. Now the modeler has an estimate of the average weight of the apple and even knows the distribution, which is enough

[2] The average of a Gaussian distributed variable also follows a Gaussian distribution. However, if the variance is unknown and is estimated from data, the average follows a Student's t-distribution.

to quantify the uncertainty of the estimator. The modeler could visualize the distribution of the mean estimate, which would help with question A) about uncertainty. The modeler might also eyeball whether 90 grams seems likely (question B). But that's not good enough for the frequentist.

4) Making Decision Based On Imaginary Experiments

To draw conclusions about the population, the frequentist elaborates on how the mean estimate would behave if the experiment were repeated. The experiment was: Sample 10 apples, weigh them, and calculate the average weight. The frequentist has two tools available to make probabilistic statements about the average weight of the apples. One tool is hypothesis tests, and the other is confidence intervals. Both tools make clear, binary statements about which values are likely and which are not, allowing clear-cut decisions.

Decide With Tests And Intervals

A **hypothesis test** is a method to decide whether the data support a hypothesis. In the apple case: Does the observed average of 79 grams still support the hypothesis of 90 grams average apple weight? Ninety grams is also called the null hypothesis. Based on the assumed distributions, the corresponding test would be a one-sample, one-sided Student t-test. The t-test calculates a p-value that indicates how likely a result of 79 (or more extreme) is if 90 grams is the true average apple weight. If the p-value is below a certain confidence level (often 5%), then 79 is so far away from 90 that the modeler rejects the 90-gram-hypothesis. If the null hypothesis is really true, the test would falsely reject the hypothesis in 5% of the cases (1 out of 20). This so-called