

BAYESIAN THINKING

ETHAN LEVIEN

CONTENTS

1. Motivation	1
2. Bayes Theorem	1

1. MOTIVATION

We revisit the basic Linear regression model:

$$(1) \quad Y \sim \text{Normal}(\alpha + \beta x, \sigma).$$

Sometimes we know the values of α and β before seeing any data, for example, the ideal gas law. In this case, there is no need to carry out statistical inference. In the previous section, we discussed how statistical inference of the parameters α , β and σ works, assuming no knowledge of these parameters. There is an intermediate between these two regimes: We might have a rough idea of what the parameters are from physical considerations or previous experiments. In this case, it is appropriate to incorporate this knowledge into our statistical inference. This can be done via *priors*. In order to understand priors, let's work with the following example.

Exercise 1. *Consider the election example, suppose we know (e.g. from polling data that). What does this tell us about α ?*

The important thing to realize is that priors are simply an extension of our statistical model, in these sense that they are variables which we make assumptions about. As such, we can express our model and proceed to make predictions as usual.

Exercise 2. *Consider the model*

$$(2) \quad Y \sim \text{Normal}(\alpha + \beta x, 0.4)$$

$$(3) \quad \alpha \sim \text{Normal}(0, 2)$$

$$(4) \quad \beta \sim \text{Normal}(1, 2)$$

What is the chance that $Y > 0.3$.

In the above examples, we fixed σ , but it is reasonable to place priors on this as well. The question is: How do we do this in a way that ensures $\sigma > 0$?

Exercise 3. *Consider the election example, suppose we know (e.g. from polling data that). What does this tell us about α ?*

2. BAYES THEOREM

How do we incorporate priors into our statistical inference?

$$(5) \quad P(D|\theta)P(\theta) = P(\theta|D)P(D)$$

Rearranging this yields

$$(6) \quad P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

Exercise 4. *Can you explain why Bayes theorem holds for both probability distributions and probability densities?*

The distributions appearing in Equation (6) have the following names and interpretations:

- $P(\theta|D)$ is the posterior.
- $P(D|\theta)$ is the likelihood
- $P(\theta)$
- $P(D)$

Bayes theorem provides us with the mathematical formalism to think about statistical inference probabilistically. In some cases, this mathematics can yield analytical formula for the

Exercise 5.

In general, even for simple models, the posterior is not available to us

3. POSTERIOR PREDICTION

4. BAYESIAN p -VALUES