Notational conventions
Class 13, 18.05
Jeremy Orloff and Jonathan Bloom

1 Learning Goals

1. Be able to work with the various notations and terms we use to describe probabilities and likelihood.

2 Introduction

We've introduced a number of different notations for probability, hypotheses and data. We collect them here, to have them in one place.

3 Notation and terminology for data and hypotheses

The problem of labeling data and hypotheses is a tricky one. When we started the course we talked about outcomes, e.g. heads or tails. Then when we introduced random variables we gave outcomes numerical values, e.g. 1 for heads and 0 for tails. This allowed us to do things like compute means and variances. We need to do something similar now. Recall our notational conventions:

- Events are labeled with capital letters, e.g. A, B, C.
- A random variable is capital X and takes values small x.
- The connection between values and events: '$X = x$' is the event that X takes the value x.
- The probability of an event is capital $P(A)$.
- A discrete random variable has a probability mass function small $p(x)$ The connection between P and p is that $P(X = x) = p(x)$.
- A continuous random variable has a probability density function $f(x)$ The connection between P and f is that $P(a \leq X \leq b) = \int_a^b f(x)\,dx$.
- For a continuous random variable X the probability that X is in an infinitesimal interval of width dx around x is $f(x)\,dx$.

In the context of Bayesian updating we have similar conventions.

- We use capital letters, especially $\mathcal{H}$, to indicate a hypothesis, e.g. $\mathcal{H}$ = 'the coin is fair'.

- We use lower case letters, especially θ, to indicate the hypothesized value of a model parameter, e.g. the probability the coin lands heads is $\theta = 0.5$.
- We use upper case letters, especially $\mathcal{D}$, when talking about data as events. For example, $\mathcal{D}$ = 'the sequence of tosses was HTH.
- We use lower case letters, especially x, when talking about data as values. For example, the sequence of data was $x_1, x_2, x_3 = 1, 0, 1$.
- When the set of hypotheses is discrete we can use the probability of individual hypotheses, e.g. $p(\theta)$. When the set is continuous we need to use the probability for an infinitesimal range of hypotheses, e.g. $f(\theta)\,d\theta$.

The following table summarizes this for discrete θ and continuous θ. In both cases we are assuming a discrete set of possible outcomes (data) x. Tomorrow we will deal with a continuous set of outcomes.

	hypothesis	prior	likelihood	Bayes numerator	posterior
	$\mathcal{H}$	$P(\mathcal{H})$	$P(\mathcal{D}\|\mathcal{H})$	$P(\mathcal{D}\|\mathcal{H})P(\mathcal{H})$	$P(\mathcal{H}\|\mathcal{D})$
Discrete θ:	θ	$p(\theta)$	$p(x\|\theta)$	$p(x\|\theta)p(\theta)$	$p(\theta\|x)$
Continuous θ:	θ	$f(\theta)\,d\theta$	$p(x\|\theta)$	$p(x\|\theta)f(\theta)\,d\theta$	$f(\theta\|x)\,d\theta$

Remember the continuous hypothesis θ is really a shorthand for 'the parameter θ is in an interval of width $d\theta$ around θ'.