



Bookmarks

- Introduction
- 1. Probability and Inference

Introduction to Probability (Week 1)  
Exercises due Sep 21, 2016 at 21:00 UTC

Probability Spaces and Events (Week 1)  
Exercises due Sep 21, 2016 at 21:00 UTC

Random Variables (Week 1)  
Exercises due Sep 21, 2016 at 21:00 UTC

Jointly Distributed Random Variables (Week 2)  
Exercises due Sep 28, 2016 at 21:00 UTC

Conditioning on Events (Week 2)  
Exercises due Sep 28, 2016 at 21:00 UTC

Homework 1 (Week 2)  
Homework due Sep 28, 2016 at 21:00 UTC

Inference with Bayes' Theorem for Random Variables (Week 3)  
Exercises due Oct 05, 2016 at 21:00 UTC

Independence Structure (Week 3)  
Exercises due Oct 05, 2016 at 21:00 UTC

Homework 2 (Week 3)  
Homework due Oct 05, 2016 at 21:00 UTC

Notation Summary (Up Through Week 3)

Mini-project 1: Movie Recommendations (Week 3)  
Mini-projects due Oct 12, 2016 at 21:00 UTC

1. Probability and Inference > Inference with Bayes' Theorem for Random Variables (Week 3) > Bayes' Theorem for Random Variables

Bookmark

# Bayes' Rule for Random Variables (Also Called Bayes' Theorem for Random Variables)

6.008.1x - Bayes' Rule for Random Variables (Also Called Bayes' Theorem for Rand



0:00 / 0:00

1.25x

Speaker icon

Fullscreen icon

CC icon

Quote icon

Video  
Download video file

Transcripts  
Download SubRip (.srt) file  
Download Text (.txt) file

These notes cover roughly the same content as the video:

## BAYES' THEOREM FOR RANDOM VARIABLES (COURSE NOTES)

In inference, what we want to reason about is some unknown random variable  $X$ , where we get to observe some other random variable  $Y$ , and we have some model for how  $X$  and  $Y$  relate. Specifically, suppose that we have some “prior” distribution  $p_X$  for  $X$ ; this prior distribution encodes what we believe to be likely or unlikely values that  $X$  takes on, *before we actually have any observations*. We also suppose we have a “likelihood” distribution  $p_{Y|X}$ .

After observing that  $Y$  takes on a specific value  $y$ , our “belief” of what  $X$  given  $Y = y$  is now given by what’s called the “posterior” distribution  $p_{X|Y}(\cdot | y)$ . Put another way, we keep track of a probability distribution that tells us how plausible we think different values  $X$  can take on are. When we observe data  $Y$  that can help us reason about  $X$ , we proceed

to either upweight or downweight how plausible we think different values  $\mathbf{X}$  can take on are, making sure that we end up with a probability distribution giving us our updated belief of what  $\mathbf{X}$  can be.

Thus, once we have observed  $\mathbf{Y} = \mathbf{y}$ , our belief of what  $\mathbf{X}$  is changes from the prior  $p_{\mathbf{X}}$  to the posterior  $p_{\mathbf{X}|\mathbf{Y}}(\cdot | \mathbf{y})$ .

Bayes' theorem (also called Bayes' rule or Bayes' law) for random variables explicitly tells us how to compute the posterior distribution  $p_{\mathbf{X}|\mathbf{Y}}(\cdot | \mathbf{y})$ , i.e., how to weight each possible value that random variable  $\mathbf{X}$  can take on, once we've observed  $\mathbf{Y} = \mathbf{y}$ . Bayes' theorem is the main workhorse of numerous inference algorithms and will show up many times throughout the course.

**Bayes' theorem:** Suppose that  $\mathbf{y}$  is a value that random variable  $\mathbf{Y}$  can take on, and  $p_{\mathbf{Y}}(\mathbf{y}) > 0$ . Then

$$p_{\mathbf{X}|\mathbf{Y}}(\mathbf{x} | \mathbf{y}) = \frac{p_{\mathbf{X}}(\mathbf{x})p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x})}{\sum_{\mathbf{x}'} p_{\mathbf{X}}(\mathbf{x}')p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x}')}$$

for all values  $\mathbf{x}$  that random variable  $\mathbf{X}$  can take on.

**Important:** Remember that  $p_{\mathbf{Y}|\mathbf{X}}(\cdot | \mathbf{x})$  could be undefined but this isn't an issue since this happens precisely when  $p_{\mathbf{X}}(\mathbf{x}) = 0$ , and we know that  $p_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y}) = 0$  (for every  $\mathbf{y}$ ) whenever  $p_{\mathbf{X}}(\mathbf{x}) = 0$ .

**Proof:** We have

$$p_{\mathbf{X}|\mathbf{Y}}(\mathbf{x} | \mathbf{y}) \stackrel{(a)}{=} \frac{p_{\mathbf{X},\mathbf{Y}}(\mathbf{x}, \mathbf{y})}{p_{\mathbf{Y}}(\mathbf{y})} \stackrel{(b)}{=} \frac{p_{\mathbf{X}}(\mathbf{x})p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x})}{p_{\mathbf{Y}}(\mathbf{y})} \stackrel{(c)}{=} \frac{p_{\mathbf{X}}(\mathbf{x})p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x})}{\sum_{\mathbf{x}'} p_{\mathbf{X},\mathbf{Y}}(\mathbf{x}', \mathbf{y})} \stackrel{(d)}{=} \frac{p_{\mathbf{X}}(\mathbf{x})p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x})}{\sum_{\mathbf{x}'} p_{\mathbf{X}}(\mathbf{x}')p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y} | \mathbf{x}')}.$$

where step (a) uses the definition of conditional probability (this step requires  $p_{\mathbf{Y}}(\mathbf{y}) > 0$ ), step (b) uses the product rule (recall that for notational convenience we're not separately writing out the case when  $p_{\mathbf{X}}(\mathbf{x}) = 0$ ), step (c) uses the formula for marginalization, and step (d) uses the product rule (again, for notational convenience, we're not separately writing out the case when  $p_{\mathbf{X}}(\mathbf{x}') = 0$ ).  $\square$

© All Rights Reserved



© 2016 edX Inc. All rights reserved except where noted. EdX, Open edX and the edX and Open EdX logos are registered trademarks or trademarks of edX Inc.

POWERED BY  
OPENedX



