EPSE 581C: Bayesian Methods

Ed Kroc

University of British Columbia

ed.kroc@ubc.ca

September 16, 2019

Last time

- Intro:
  - Frequentist vs. Bayesian methodology, basics
  - Bayes' Theorem (discrete set of hypotheses)
  - Critical Bayesian quantities: posterior probabilities, priors, likelihoods, normalizing factor

Today

- Mathematical foundations of Bayesian methodology:
  - Discrete vs. continuous random variables
  - Bayes' Theorem (continuous set of hypotheses)

Bayes' Theorem

All of Bayesian methodology is predicated upon a simple mathematical relation: Bayes' Theorem (proven independently by Rev. Thomas Bayes circa 1760 and Pierre-Simon Laplace circa 1774).

Bayes' Theorem

Let the sets $F_1, F_2, \ldots, F_n$ be disjoint (no overlap). Further, suppose that they partition the entire universe of events: $S = \{F_1 \text{ or } F_2 \text{ or } \ldots \text{ or } F_n\}$. Then:

$$\Pr(F_i \mid E) = \frac{\Pr(E \mid F_i)\Pr(F_i)}{\sum_{j=1}^{n} \Pr(E \mid F_j)\Pr(F_j)}$$

- Jeffreys (1973): "Bayes' theorem is to the theory of probability what the Pythagorean theorem is to geometry."
- To understand this theorem, we need to understand *conditional probability*.

Bayes' theorem to Bayesian methodology

$$\Pr(H \mid \text{data}) = \frac{\Pr(\text{data} \mid H) \cdot \Pr(H)}{\Pr(\text{data})}$$

$$\Pr(H_1 \mid \text{data}) = \frac{\Pr(\text{data} \mid H_1) \cdot \Pr(H_1)}{\sum_{i=1}^{n} \Pr(\text{data} \mid H_i) \Pr(H_i)}$$

Critical terminology:

- Likelihood
- Prior probability
- Posterior probability
- Normalizing factor

Bayes' theorem to Bayesian methodology

$$\Pr(H \mid \text{data}) = \frac{\Pr(\text{data} \mid H) \cdot \Pr(H)}{\Pr(\text{data})}$$

$$\Pr(H_1 \mid \text{data}) = \frac{\Pr(\text{data} \mid H_1) \cdot \Pr(H_1)}{\sum_{i=1}^{n} \Pr(\text{data} \mid H_i) \Pr(H_i)}$$

- The posterior probability has a natural interpretation: it gives an explicit measure of certainty to a hypothesis given some evidence for or against that hypothesis.
- This is the quantity we are always most interested in in practice for scientific inquiry.

Bayes' Theorem

- But what about when there are infinitely many disjoint sets (hypotheses) that partition the universe of events?
- Think of the simple t-test: the mean μ of a random variable can be *any* real number: an infinite *continuum* of possibilities!
- Simple summation is not going to work....
- *Integration* is required: calculus-based probability.

Calculus in Statistics

There are two main uses for calculus in statistics:

(1) The concept of the *derivative* allows us to *minimize* or *maximize* functions

- (MLE) maximum likelihood estimation (equivalent to OLS regression in simple cases)
- common frequentist approach: a model determines a *likelihood function*, and we then take derivatives of this function to find the specific model parameters (regression coefficients) that *maximize the likelihood*; then use the observed data to *estimate* these parameters.

Example: MLE in simple regression

- Consider the simple regression model:

$$Y = \beta_0 + \beta_X X + \varepsilon, \tag{1}$$

  with the usual assumption that $\varepsilon \sim N(0, \sigma^2)$ for some fixed σ^2.

- Note: if X is *binary*, then this model is equivalent to the ordinary t-test: i.e. β_X is equivalent to the ordinary two group t-statistic.
- Regardless: for any given value of X, equation (1) implies $Y \sim N(\beta_0 + \beta_X X, \sigma^2)$. Thus, this model is described by the likelihood function:

$$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\beta_0-\beta_X X)^2}{2\sigma^2}}$$

- Now take derivatives to find what values of β_0 and β_X would *maximize* this likelihood.

Example: OLS solution in simple regression

- Consider the simple regression model:

$$Y = \beta_0 + \beta_X X + \varepsilon,$$

with the usual assumption that $\varepsilon \sim N(0, \sigma^2)$ for some fixed σ^2.

- Could also compute the *ordinary least squares* solution for this regression equation by *minimizing* the sum of the squared errors:

$$\text{minimize} \quad \sum_{i=1}^{n} (y_i - \beta_0 - \beta_x x_i)^2$$

- Take derivatives to find what values of β_0 and β_X would *minimize* this expression.

- Note: can show that these OLS solutions are the *same* as the MLE solutions for any simple linear regression model.

Calculus in Statistics

There are two main uses for calculus in statistics:

(2) The concept of the *integral* allows us to *average* (or *sum*) infinitely many numbers in a coherent way, even if those numbers form a *continuum*.

- allows us to generalize Bayes' Theorem so that we can apply it to a *continuum of hypotheses*

Discrete vs. Continuous Space

A *discrete* space is (basically) one that assumes finitely many or countably infinitely many distinct values: e.g.

- N-point space: $\{1, 2, \ldots, N\}$
- All positive integers: $\{1, 2, 3, \ldots\}$
- All pairs of integer coordinates: $\{(x, y) : x, y \in \mathbb{Z}\}$.
- All rational numbers (fractions of integers)

We can always make sense of *sums* over discrete spaces, so can accumulate discrete probabilities, because the possibilities can be *counted/listed*: e.g.

- $\sum_{i=1}^{n} a_i = a_1 + \cdots + a_n$
- $\sum_{I=1}^{\infty} a_i = a_1 + a_2 + \cdots$

Discrete vs. Continuous Space

A *continuous* space is (basically) one that assumes uncountably many distinct values along a continuum: e.g.

- All real numbers, $\mathbb{R}$
- All positive real numbers, $\mathbb{R}^+$
- All pairs of real-numbered coordinates: $\{(x, y) : x, y \in \mathbb{R}\}$
- All real numbers in the interval $[0, 1]$ or $(0, 1)$
- All real numbers in *any* interval $[a, b]$ or (a, b)

But how do we count/list all values in a continuum? Answer: we can't!

The real numbers are not countable

Here's a proof that the real numbers in the interval $[0, 1]$ *cannot* be counted/listed:

- Proof by contradiction: start by supposing all numbers in $[0, 1]$ *can* be listed.
- Then that means we can write $[0, 1] = \{x_1, x_2, x_3, \ldots\}$.
- Since these numbers can all be expressed as (infinite) decimals, let's write them out:

$$
\begin{aligned}
x_1 &= 0.x_{11}x_{12}x_{13}x_{14}x_{15}\ldots \\
x_2 &= 0.x_{21}x_{22}x_{23}x_{24}x_{25}\ldots \\
x_3 &= 0.x_{31}x_{32}x_{33}x_{34}x_{35}\ldots \\
x_4 &= 0.x_{41}x_{42}x_{43}x_{44}x_{45}\ldots \\
x_5 &= 0.x_{51}x_{52}x_{53}x_{54}x_{55}\ldots \\
\vdots &= \vdots \quad \vdots
\end{aligned}
$$

The real numbers are not countable

- Consider only the decimal digits along the *diagonal* of this list:

$$
\begin{aligned}
x_1 &= 0.x_{11}x_{12}x_{13}x_{14}x_{15}\ldots \\
x_2 &= 0.x_{21}x_{22}x_{23}x_{24}x_{25}\ldots \\
x_3 &= 0.x_{31}x_{32}x_{33}x_{34}x_{35}\ldots \\
x_4 &= 0.x_{41}x_{42}x_{43}x_{44}x_{45}\ldots \\
x_5 &= 0.x_{51}x_{52}x_{53}x_{54}x_{55}\ldots \\
\vdots &= \vdots \quad \vdots \qquad \ddots
\end{aligned}
$$

- Now we will define a new real number $r = 0.r_1 r_2 r_3 \ldots$ in $[0, 1]$ that is *not* in this list:
  - If $x_{ii} = 0$, then define $r_i = 1$,
  - If $x_{ii} \neq 0$, then define $r_i = 0$.

The real numbers are not countable

- Notice that r *cannot* be contained in our supposed list of all numbers in [0,1]. Why not?
- If r was in our list, then $r = x_k$ for some k.
- So, in particular, must have $r_k = x_{kk}$.
- But we defined r so that its kth decimal is *always different* from the kth decimal of the kth real number in our list: if $x_{kk} = 0$, then $r_k = 1$ and if $x_{kk} \neq 0$, then $r_k = 0$.
- Thus, we have a contradiction: we assumed we could list all the real numbers in [0,1], but ended up constructing one that can *never* be on the list.
- Logically then, *the initial assumption was wrong*; hence, there is no way to list/count the real numbers.

Bayes' theorem

- Recall: Bayes' Theorem gives us an explicit way to decompose the probability of a hypothesis given some data:

$$\Pr(H_k \mid \text{data}) = \frac{\Pr(\text{data} \mid H_k) \cdot \Pr(H_k)}{\sum_{i=1}^{n} \Pr(\text{data} \mid H_i)\Pr(H_i)}$$

- Can generalize the proof we gave last time to show Bayes' Theorem over *countably* many hypotheses:

$$\Pr(H_k \mid \text{data}) = \frac{\Pr(\text{data} \mid H_k) \cdot \Pr(H_k)}{\sum_{i=1}^{\infty} \Pr(\text{data} \mid H_i)\Pr(H_i)}$$

- But since we can't *count* the real numbers, our proof isn't going to work when we have a *continuum* of hypotheses.

Integration

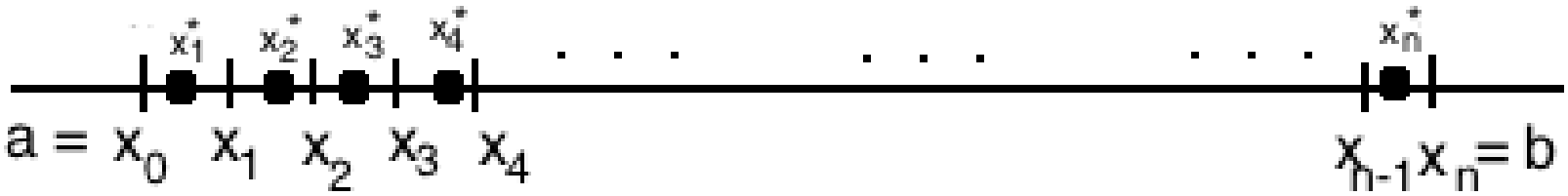
- To solve this problem, we need to understand how *integration* works.
- There are many different kinds of integration, but all aim to allow us to coherently generalize the concept of a *sum* to a continuum of numbers:
  - Riemann integration (traditional, what we will use)
  - Riemann-Steltjes integration (more general)
  - Lebesgue integration (*much* more general)
  - Lebesgue-Steltjes integration (ultimate generalization)

Integration

- For a given function f, we define the *integral* (Riemann integral) of f over an interval $[a, b]$ as:

$$\int_a^b f(x)\ dx := \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*)(x_i - x_{i-1}),$$

where the interval $[a, b]$ is split into n equally-sized pieces $[x_0, x_1]$, $[x_1, x_2], \ldots, [x_{n-1}, x_n]$, and x_i^* is any point inside the ith one of these intervals.

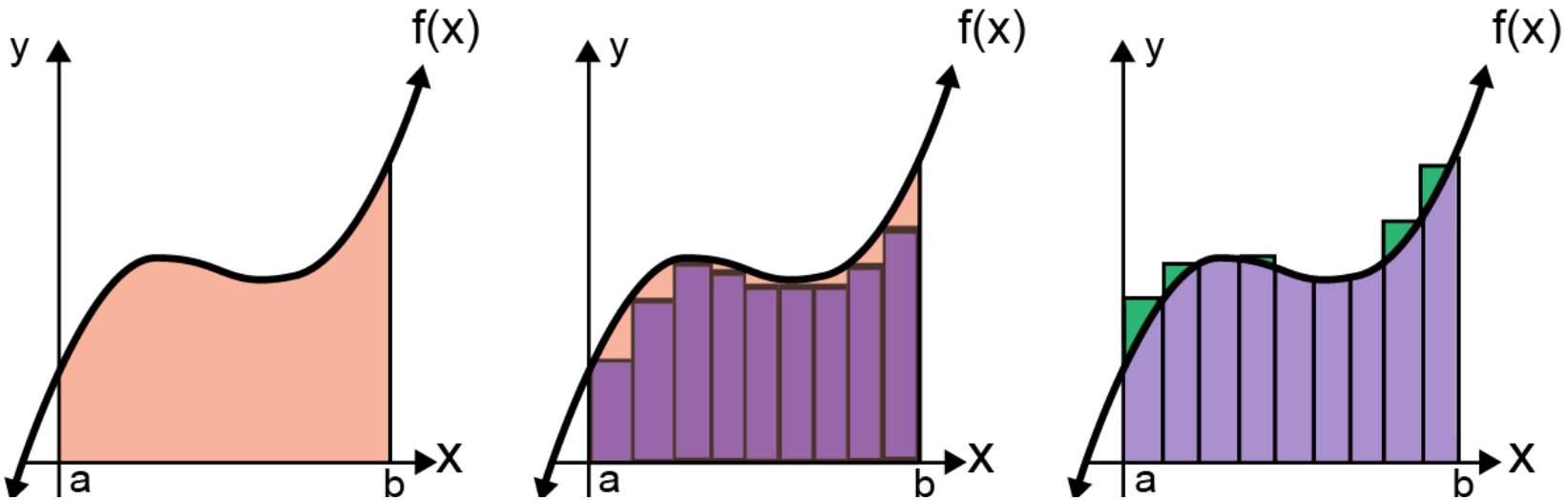


Integration

- Notice that the "Riemann sum" is just a sum of a bunch of rectangles: all have the same width $x_i - x_{i-1} = \frac{b-a}{n}$ and each have height $f(x_i^*)$.

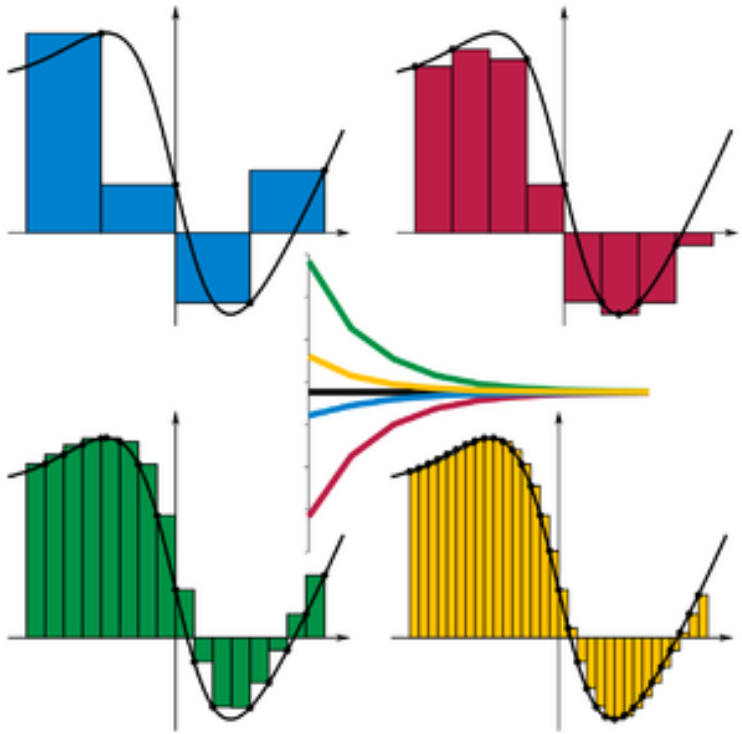
$$\int_a^b f(x)\ dx := \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*)(x_i - x_{i-1}),$$

- We say "the integral of the function f is the limit of its Riemann sums."



Integration

- Now when we take the *limit*, we make these approximating rectangles finer and finer, thus recovering the *area under the graph* of f on $[a, b]$.



See: `https://www.desmos.com/calculator/tgyr42ezjq`

Integration

Let's recap:

- In general, suppose we want to compute the area under $f(x)$ on an interval $[a, b]$. Then:

  (1) Split the interval $[a, b]$ into n pieces of the same size: this creates n subintervals $I_1, \ldots, I_n$ of the same length. These form the bases of the approximating rectangles.

  (2) Pick any value of the function on each of these subintervals to represent the heights of the bars: $f(x_1^*), \ldots, f(x_n^*)$.

  (3) Then the area under the curve is approximated by the area under the histogram given by

$$\sum_{i=1}^{n} f(x_i^*) \cdot |I_i|,$$

  where $|I_i| = \frac{b-a}{n}$.

Integration

- As we *take the limit* of this process, we define the (Riemann) integral of the curve, $f(x)$. This is the actual area under the curve:

$$\int_a^b f(x)\ dx := \lim_{n\to\infty} \sum_{i=1}^{n} f(x_i^*) \cdot \frac{b-a}{n}$$

- We read this as "the integral of f on $[a, b]$."
- The integral sign $\int$ imitates the summation sign $\sum$.
- The bounds of the integral $\int_a^b$ tell you where you are calculating the area under the curve.
- The *integrand* $f(x)$ imitates the "height" of the approximating rectangles.
- The *differential* dx imitates the "width" of the approximating rectangles. In the limit, these widths go to zero, but the *limit of the Riemann sums* approaches the area under the curve.

Integration

- Important to realize that integrals make use of *dummy variables*
- That is,

$$\int_a^b f(x)\ dx = \int_a^b f(t)\ dt = \int_a^b f(\checkmark)\ d\checkmark$$

- The "variable" inside the integral and the differential is a *dummy variable*; it doesn't matter what we call it.
- But *be careful* with your notation otherwise this can get confusing, e.g.

$$\int_x^y f(x)\ dx$$

is meaningless. You can't integrate a function of x from the fixed point x to the fixed point y. Fix this by writing instead

$$\int_x^y f(t)\ dt$$

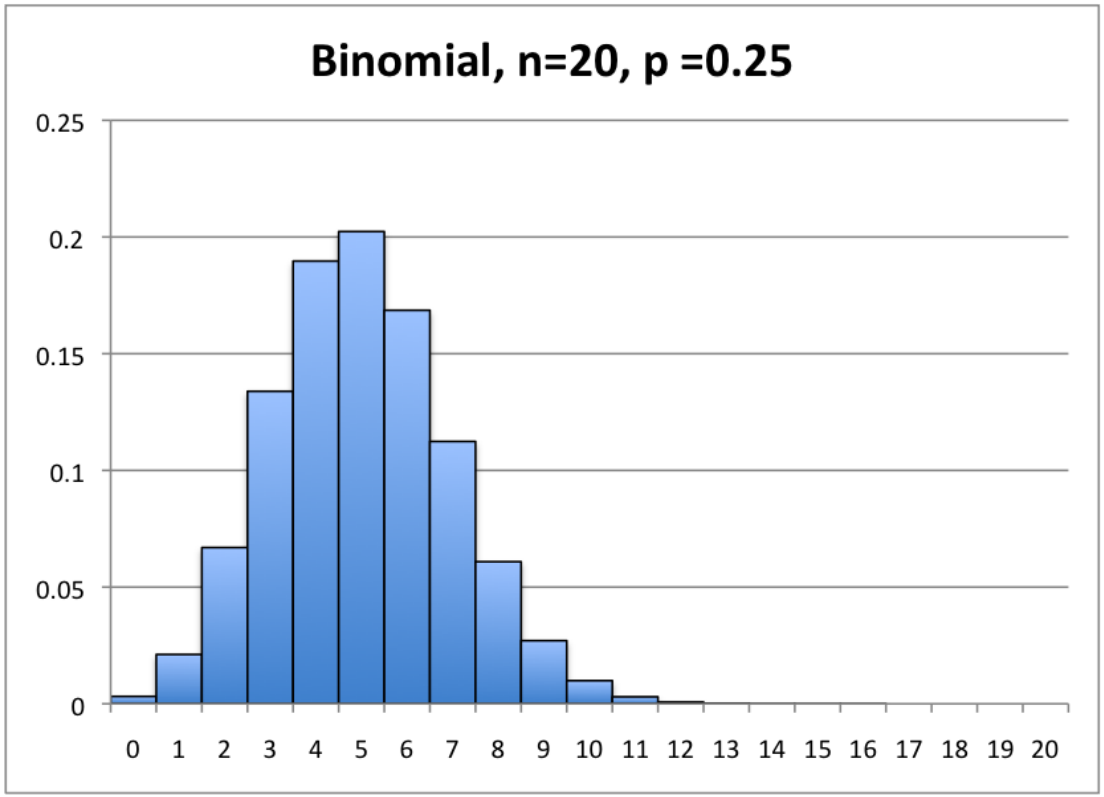
Integration

- With the tool of integration, we can now make precise sense of probabilities of events generated from *continuous* random variables.

- To contextualize this, let's recall what defines a random variable and what distinguishes a discrete r.v. from a continuous one.

Probability Mass Functions

A discrete random variable X is totally defined by its *probability mass function* (PMF): $\Pr(X = x)$, often visualized as a histogram:



E.g. $\Pr(1 \leqslant X \leqslant 3) = \Pr(X = 1) + \Pr(X = 2) + \Pr(X = 3) = 0.222$

Probability Mass Functions

Recall the fundamental property that

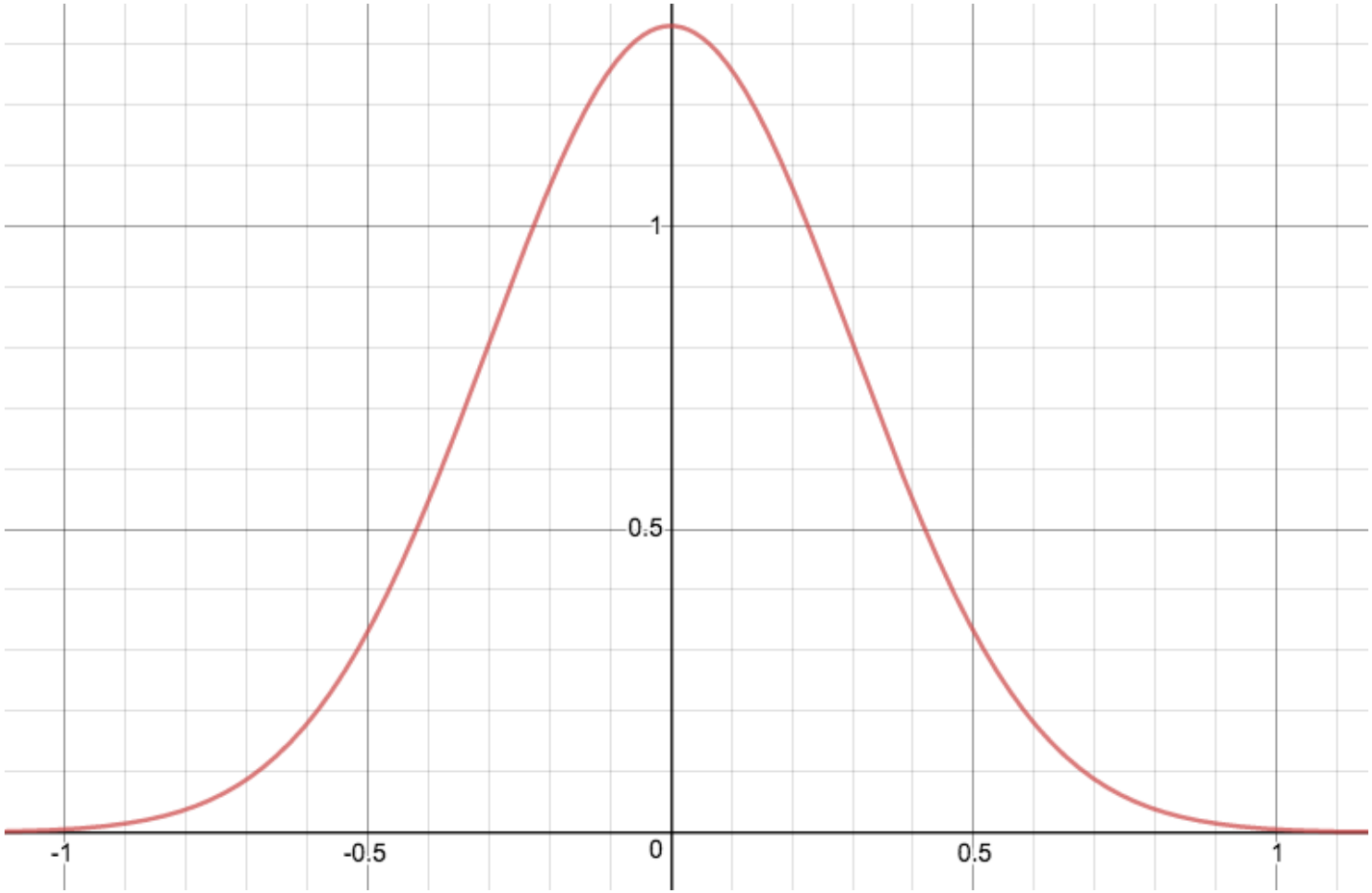
$$\sum_x \Pr(X = x) = 1.$$

That is, if we sum up the probabilities of all possible outcomes, these have to sum to 1.

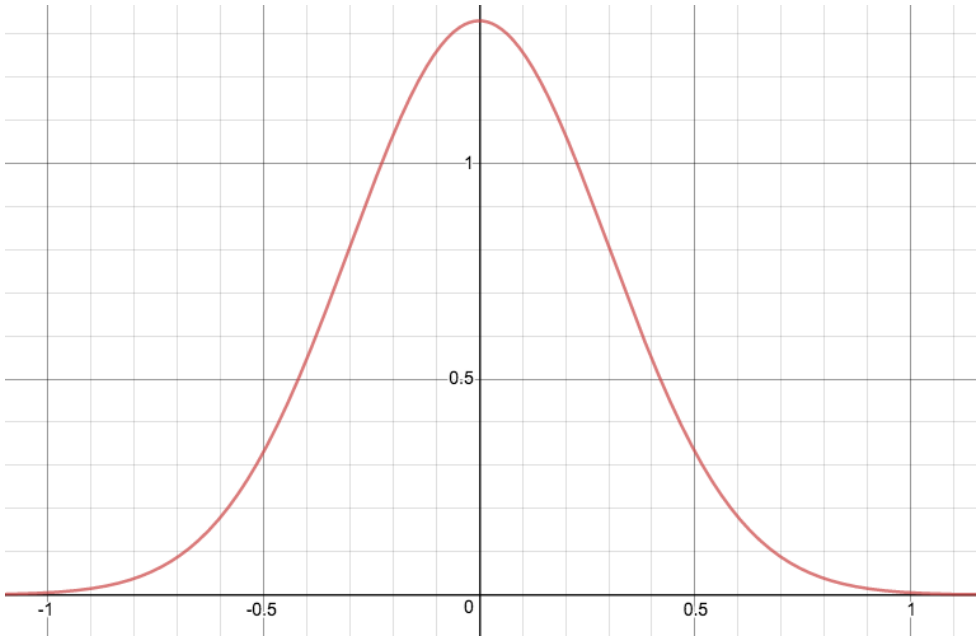
Certainly, the same should hold for continuous random variables too, but we can't just sum up all probabilities of a continuous outcome; instead, we *integrate*.

Probability Density Functions

- Consider the classic bell curve, representing the *probability density function* (PDF) of a normal (continuous) random variable: $N(0, 0.3)$



Probability Density Functions



- For any real numbers a, b, we have

$$\Pr(a \leqslant X \leqslant b) = \int_a^b f(x)\ dx$$

- Note: this is the area under the graph of the PDF from a to b, which exactly mirrors how we would calculate such a probability for discrete r.v.s via a PMF.

Probability Density Functions

Just as the PMF characterizes the probability distribution of a discrete r.v., the PDF, $f(x)$, characterizes the probability distribution of a continuous r.v. X:

(1) $f(x)$ is always nonnegative.

(2) The total area under the PDF (and above the x-axis) equals 1.

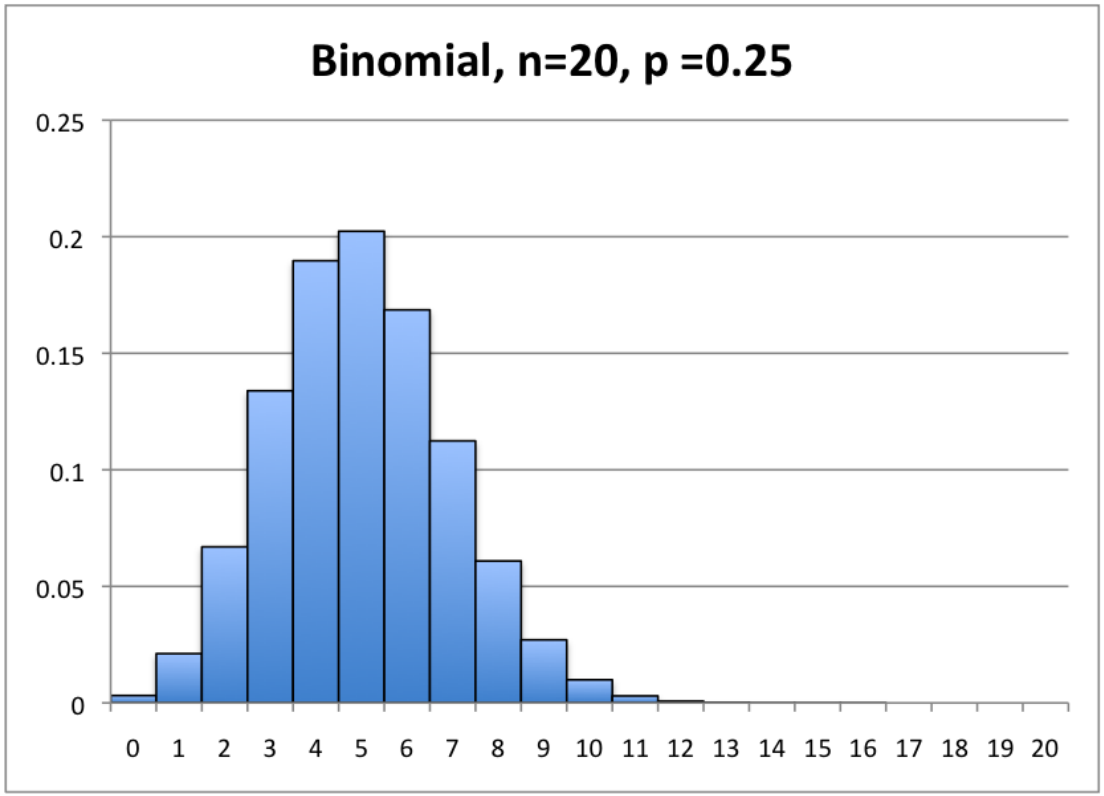
(3) In general, the probability of an event $\{a \leqslant X \leqslant b\}$ is given by the area under the PDF on the interval $[a, b]$; i.e.

$$\Pr(a \leqslant X \leqslant b) = \int_a^b f(x)\ dx$$

However, for continuous r.v., it makes *no sense* to write $\Pr(X = x)$.

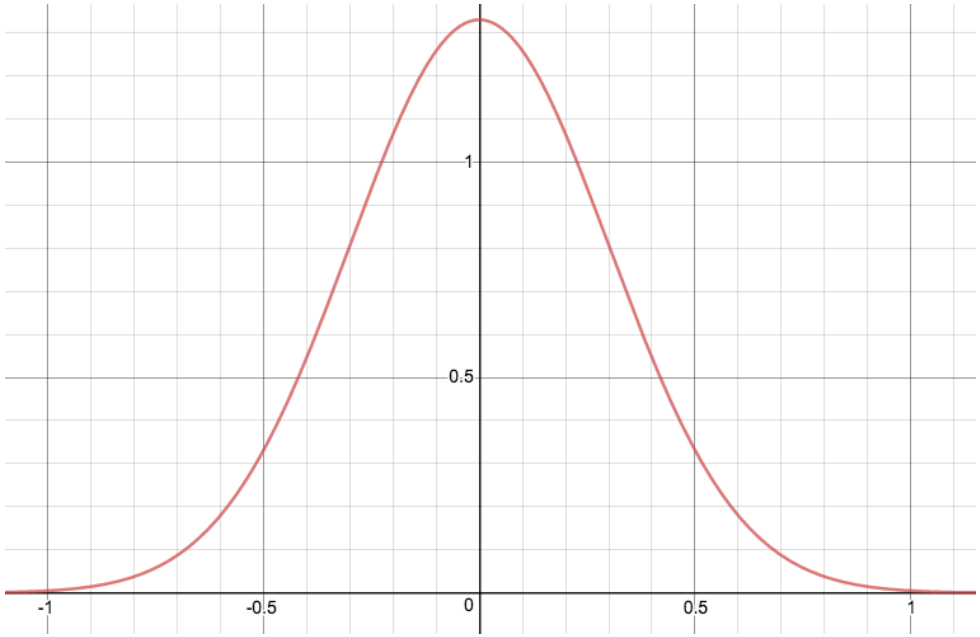
Discrete probabilities

Recall the PMF of a discrete r.v.:



Here, $\Pr(X = x)$ is exactly equal to the area of the rectangle at $X = x$.

Continuous probabilities



- Here, $\Pr(X = x) = \int_x^x f(t)\ dt = 0$; i.e. there is *no area* under the graph at a single point.
- Moreover, notice that for this particular PDF, $f(0) > 1$. So the PDF $f(x)$ does *not* encode probabilities in its values (unlike a PMF).

Continuous probabilities

So for a continuous r.v. X:

- $\Pr(X = x) = 0$ always
- $\Pr(a \leqslant X \leqslant b) = \int_a^b f(x)\ dx$

Thus, it is the *integral of the PDF* that encodes probabilities for continuous r.v.s, *not* simply the PDF.

Also, we must always consider a *range* of possible values for X, otherwise we will always be considering events that *cannot happen*.

Note, because the area under the graph at a single point is always zero:

$$\Pr(a \leqslant X \leqslant b) = \Pr(a < X \leqslant b) = \Pr(a \leqslant X < b) = \Pr(a < X < b)$$

Towards Bayes' Theorem

- Remember: we are trying to get to a point where we can make sense of a continuous version of Bayes' Theorem.
- We have PDFs for continuous random variables.
- We know how to calculate probabilities for these random variables via integration.
- Finally, we need to understand how *conditional probability* works for continuous random variables. To do this, we need to talk about *joint distributions* of more than one random variable at the same time.

Joint discrete random variables

Let X and Y be two discrete random variables.

- The joint probability mass function of X and Y is:

$$p(x, y) = \Pr(X = x, Y = y).$$

Note, we require $0 \leqslant p(x, y) \leqslant 1$ for all x, y, and $\sum_x \sum_y p(x, y) = 1$.

- Example: X denotes flipping a fair coin and Y denotes rolling a fair 6-sided die. Then

$$\begin{aligned}\Pr(X = H,\ Y \in \{1, 2\}) &= \Pr(X = H,\ Y = 1) + \Pr(X = H,\ Y = 2) \\ &= \frac{1}{2} \cdot \frac{1}{6} + \frac{1}{2} \cdot \frac{1}{6} \\ &= \frac{1}{6}\end{aligned}$$

Joint discrete random variables

Let X and Y be two discrete random variables.

- The marginal probability mass functions of X and Y are:

$$\text{For } X: \Pr(X = x) = \sum_{\text{all } y} p(x, y)$$

$$\text{For } Y: \Pr(Y = y) = \sum_{\text{all } x} p(x, y)$$

- Example: X denotes flipping a fair coin and Y denotes rolling a fair 6-sided die. Then

$$\begin{aligned}\Pr(X = H) &= \sum_{\text{all } y} p(H, y) \\ &= \sum_{i=1}^{6} \Pr(X = H,\ Y = i) \\ &= \frac{1}{2} \cdot \frac{1}{6} + \frac{1}{2} \cdot \frac{1}{6} + \cdots + \frac{1}{2} \cdot \frac{1}{6} = \frac{1}{2}\end{aligned}$$

Joint continuous random variables

Let X and Y be two continuous random variables.

- Let $C \subseteq \mathbb{R}^2$. Then $C = A \times B$ for some sets $A, B \subseteq \mathbb{R}$. Then the probability that (X, Y) lies in the set C is:

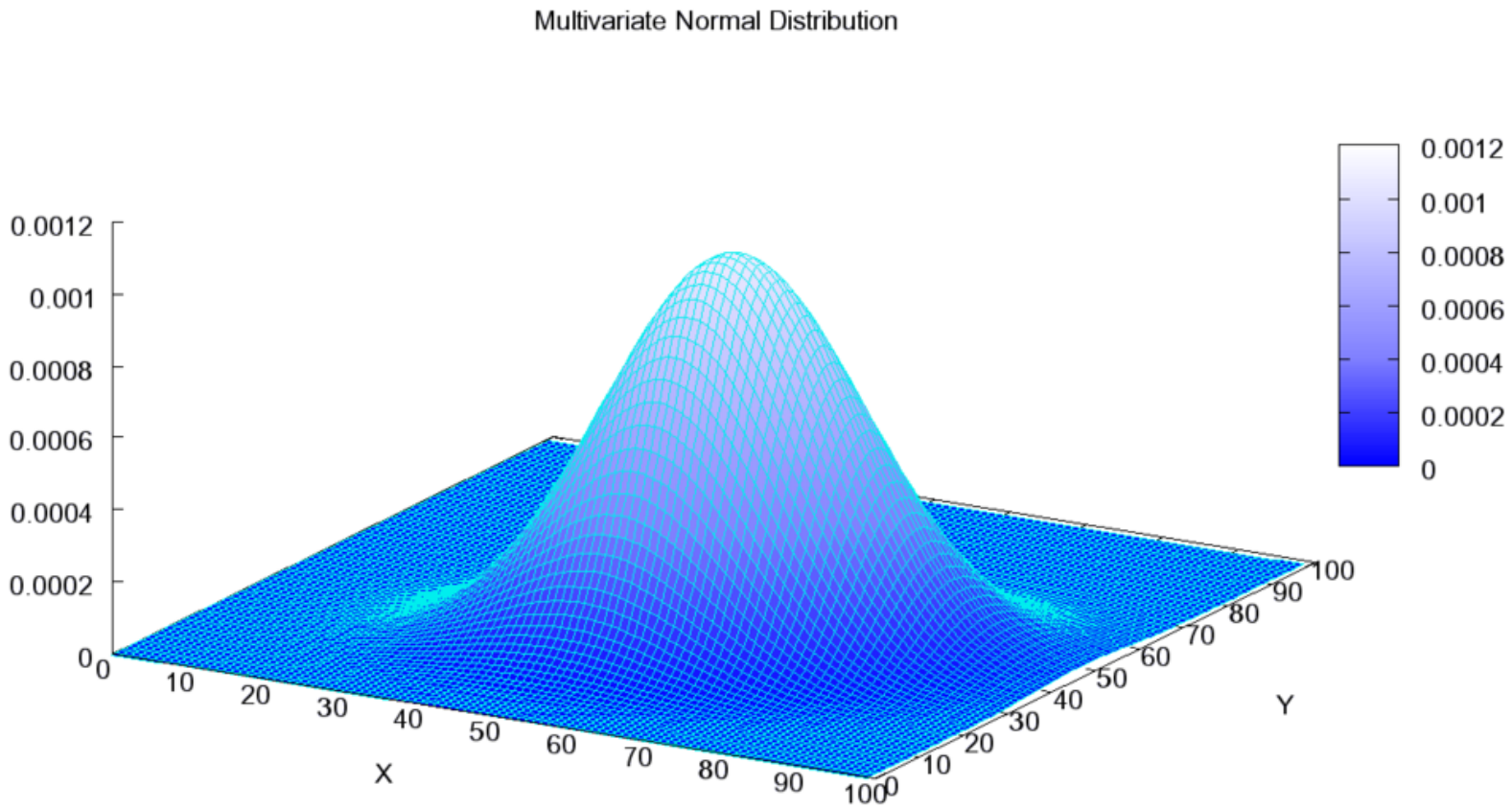
$$\Pr((X, Y) \in C) = \Pr(X \in A, Y \in B) = \int_A \int_B f(x, y) dy dx.$$

  Here, $f(x, y)$ is the joint probability density function of X and Y.

- Because the joint PDF defines how the calculate probabilities for X and Y simultaneously, we have:
  - $f(x, y) \geqslant 0$
  - $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dy dx = 1$

Joint continuous random variables

If X and Y are both normally distributed, then their joint density $f(x, y)$ is given by a *multivariate bell curve* and defines a *multivariate normal distribution*:



Joint continuous random variables

Let X and Y be two continuous random variables.

- The marginal probability density functions of X and Y are:

$$\text{For } X : f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy$$

$$\text{For } Y : f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx$$

- Notice how this definition directly mirrors the definition for discrete r.v.s
- Since we aren't actually learning how to integrate functions in this class, it is only the concept and the notation that is important for us. You will *not* have to compute integrals!

Independence of random variables

Two random variables X and Y are said to be independent if and only if:

- $\Pr(X = x, Y = y) = \Pr(X = x)\Pr(Y = y)$ for discrete X, Y, or,
- $f(x, y) = f_X(x)f_Y(y)$ for continuous X, Y.

Example: recall the example of the fair coin and the fair die. Those random phenomena (discrete random variables) were *independent*, e.g.:

$$\begin{aligned}
\Pr(X = H,\ Y \in \{1, 2\}) &= \Pr(X = H,\ Y = 1) + \Pr(X = H,\ Y = 2) \\
&= \frac{1}{2} \cdot \frac{1}{6} + \frac{1}{2} \cdot \frac{1}{6} \\
&= \frac{1}{6} \\
&= \Pr(X = H)\Pr(Y \in \{1, 2\}) \\
&= \frac{1}{2} \cdot \frac{2}{6} = \frac{1}{6}
\end{aligned}$$

Conditional distributions

Let X and Y be discrete random variables.

- The conditional probability mass function of X given $Y = y$, for $\Pr(Y = y) > 0$, is:

$$\Pr(X = x \mid Y = y) = \frac{\Pr(X = x, Y = y)}{\Pr(Y = y)}$$

- Notice that this is just the ordinary definition of conditional probability from last time!
- Note: if X and Y are independent, then

$$\Pr(X = x \mid Y = y) = \frac{\Pr(X = x)\Pr(Y = y)}{\Pr(Y = y)} = \Pr(X = x)$$

Conditional distributions

Example: let Y denote the outcome of a toss of a fair 6-sided die, and let X denote the number of heads that comes up after tossing a fair coin Y many times.

- Then the *conditional random variable* $X \mid Y \sim \text{Bin}(Y, 0.5)$.
- For example,

$$\begin{aligned}
\Pr(X = 1 \mid Y = 2) &= \frac{\Pr(X = 1, Y = 2)}{\Pr(Y = 2)} \\
&= \frac{\Pr(HT,\ Y = 2) + \Pr(TH,\ Y = 2)}{\Pr(Y = 2)} \\
&= \frac{\Pr(HT)\Pr(Y = 2) + \Pr(TH)\Pr(Y = 2)}{\Pr(Y = 2)} \\
&= \frac{\frac{1}{4} \cdot \frac{1}{6} + \frac{1}{4} \cdot \frac{1}{6}}{\frac{1}{6}} = \frac{1}{2}
\end{aligned}$$

Conditional distributions

Let X and Y be continuous random variables.

- The conditional probability density function of X given $Y = y$, for $f_Y(y) > 0$, is $f_{X|Y}(x|y)$, also denoted $f_{X|Y=y}(x)$:

$$f_{X|Y=y}(x) = \frac{f(x,y)}{f_Y(y)}$$

- Notice that this is equivalent to

$$f(x,y) = f_{X|Y=y}(x) \cdot f_Y(y)$$

- If X and Y are independent, then

$$f_{X|Y=y}(x) = \frac{f_X(x) f_Y(y)}{f_Y(y)} = f_X(x)$$

Bayes' Theorem

Finally! We are now prepared to state and prove Bayes' Theorem for continuous random variables.

Bayes' Theorem

Let X be a continuous random variable; i.e. can take on any real number. Let Y be any random variable (discrete or continuous). Then the distribution of X given Y is determined by the conditional PDF:

$$f_{X|Y=y}(x \mid Y = y) = \frac{f_{Y|X=x}(y \mid X = x) f_X(x)}{\int_{-\infty}^{\infty} f_{Y|X=t}(y \mid X = t) f_X(t)\ dt}$$

or, more simply:

$$f(x \mid y) = \frac{f(y \mid x) f(x)}{\int_{-\infty}^{\infty} f(y \mid t) f(t)\ dt}$$

Bayes' Theorem

Proof of Bayes:

$$
\begin{aligned}
f(x \mid y) &= \frac{f(x,y)}{f(y)} \quad \text{(defn. cond. PDF)} \\
&= \frac{f(y \mid x) f(x)}{f(y)} \quad \text{(defn. cond. PDF)} \\
&= \frac{f(y \mid x) f(x)}{\int_{-\infty}^{\infty} f(t,y)\ dt} \quad \text{(defn. marg. PDF)} \\
&= \frac{f(y \mid x) f(x)}{\int_{-\infty}^{\infty} f(y \mid t) f(t)\ dt} \quad \text{(defn. cond. PDF)}
\end{aligned}
$$

Recall: a PDF totally characterizes the probability distribution of a random variable/phenomenon.

Thus: we can now "partition" a continuum of events and flip the order of conditioning!

Bayes' theorem to Bayesian methodology

$$f(\theta \mid y) = \frac{f(y \mid \theta) f(\theta)}{f(y)}$$

$$f(\theta \mid y) = \frac{f(y \mid \theta) f(\theta)}{\int_{-\infty}^{\infty} f(y \mid \psi) f(\psi)\, d\psi}$$

Critical terminology:

- Likelihood
- Prior probability
- Posterior probability
- Normalizing factor

Bayes' theorem to Bayesian methodology

- The posterior probability reflects the *updated* belief in a hypothesis/event, given the data and the initial prior.
- The likelihood is easy to calculate (determined by the *model* and the *data*).
- The prior is *not* determined by the data; the researcher must set its value using prior information.
- The normalizing factor only acts as a scaling factor (so that the conditional PDF integrates to 1); it depends on the data, the model, and the prior, but it does *not* depend on the particular hypothesis/event of interest.
- All of Bayesian inference is based on *properties of the posterior distribution/density*.