Week 2

Introduction to Probability

Paul Laskowski and Alex Hughes

July 19, 2022

UC Berkeley, School of Information

Introduction

Introduction

Motivating Examples

Plan for the Week

Two sections:

1. Defining random variables
2. Joint distributions

Plan for the Week (cont.)

At the end of this week, you will be able to:

- Understand how random variables fit as a part of statistical modeling
- Apply the mathematical tools we have for characterizing random variables
- Solve problems by applying properties of expectation and variance

Part 1: Introducing Random Variables

Part 1: Introducing Random Variables

Introduction Random Variables

What Is a Random Variable?

Intuition:

- A model for where the numbers in data come from.
- Like a regular variable but with uncertainty about its value.

Random Variable

Given probability space (Ω, S, P), a *random variable* is a measurable function from Ω to $\mathbb{R}$.

What Is a Random Variable (Formally)?

160

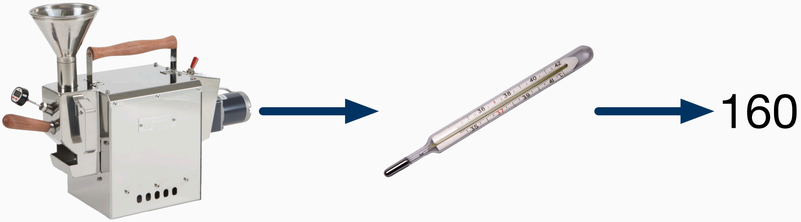
What Is a Random Variable (Formally)?

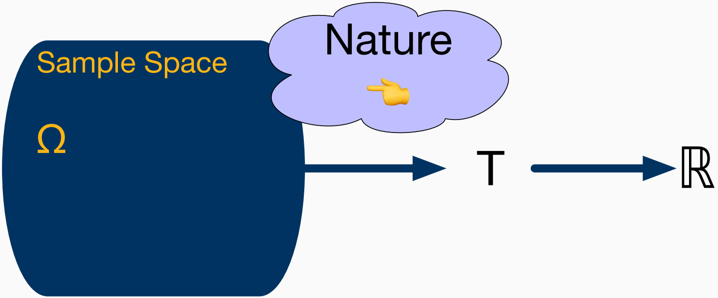
160

Sample Space

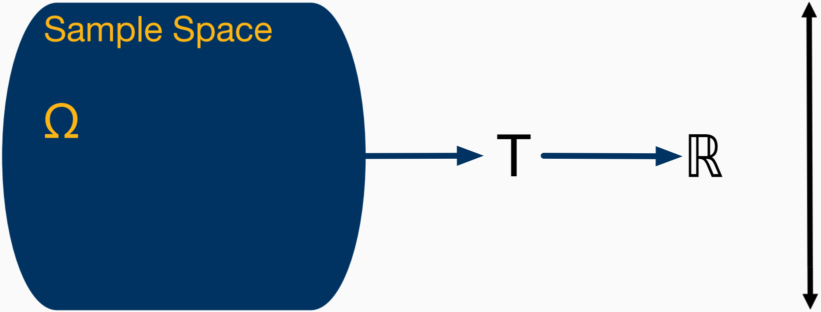
$\Omega \longrightarrow T \longrightarrow \mathbb{R}$

What Is a Random Variable (Formally)?





Probabilities of Real Numbers



Part 1: Introducing Random Variables

Reading: Random Variables

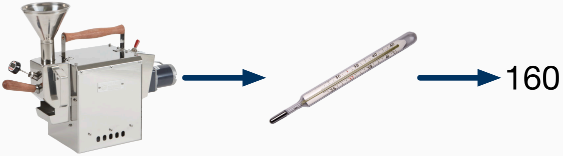
Reading Assignment

Read *Foundations of Agnostic Statistics*, section 1.2.0–1.2.1.

Part 1: Introducing Random Variables

Functions, Events, and Operators

Functions of a Random Variable

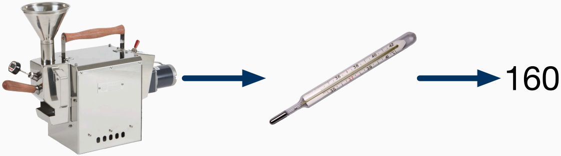


- Random variable T represents temperature in Fahrenheit, but you need Celsius.

Creating Events from a Random Variable

You want to know whether the temperature is below 150.

Operators Applied to a Random Variable

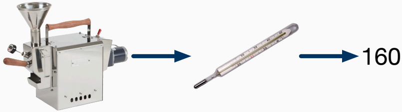


You want to know the minimum possible temperature.

Part 1: Introducing Random Variables

Learnosity: Roaster Temperature

Learnosity: Roaster Temperature



Suppose T is a random variable representing the temperature of a roaster.

What kind of object is each of the following?

1. $g(T)$ where $g(t) = \begin{cases} 1, & t < 150 \\ 0, & \text{otherwise} \end{cases}$. Answer: random variable
2. R such that $R[T] = P(T > 150)$. Answer: operator
3. $T < (5/9)(T - 32)$. Answer: event

Part 1: Introducing Random Variables

Reading Call

Reading Assignment

Read *Foundations of Agnostic Statistics*, section 1.2.2.

Part 1: Introducing Random Variables

Discrete Random Variables

Introducing Discrete Random Variables

A random variable X is *discrete* if its range, $X(\Omega) \in \mathbb{R}$, is a countable set.

- $X(\Omega)$ can be finite.
  - A roll of a die: $\{1, 2, 3, 4, 5, 6\}$
- $X(\Omega)$ can be countably infinite.
  - Edits to an article: $\{0, 1, 2, 3 \ldots\}$

The Bernoulli Distribution

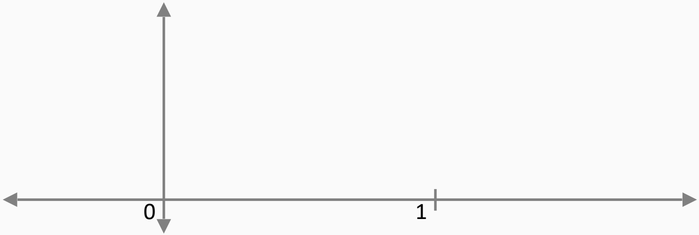
A Bernoulli trial is a simple (but important) discrete random variable.

- It can represent:
  - a coin flip
  - a decision to buy or not to buy
  - success or failure of experiment
- Convention: 1 represents success, 0 failure

The Probability Mass Function (PMF)

A function that gives the probability that a discrete random variable equals each number in $\mathbb{R}$

- $f_X(x) = P(X = x)$



Key Properties of the PMF

Theorem

For a discrete random variable X with pmf f,

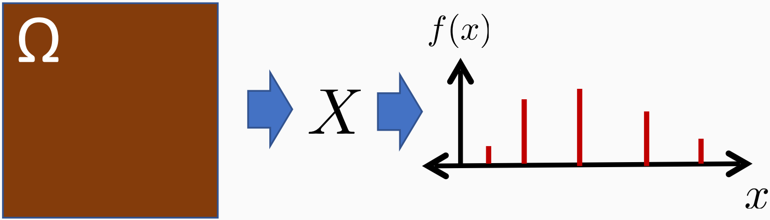
1. For all $x \in \mathbb{R}$, $f(x) \geq 0$.
2. $\sum_{x \in X(\Omega)} f(x) = 1$.

Computing Other Event Probabilities from the PMF

Theorem: The PMF fully characterizes a discrete RV

For a discrete random variable X with pmf f, if $D \subseteq \mathbb{R}$,

$$P(X \in D) = \sum_{x \in X(\Omega) \cap D} f(x).$$



Take-away: The pmf contains *all* the information about the probability distribution of X.

Part 1: Introducing Random Variables

Reading: Cumulative Distribution Functions

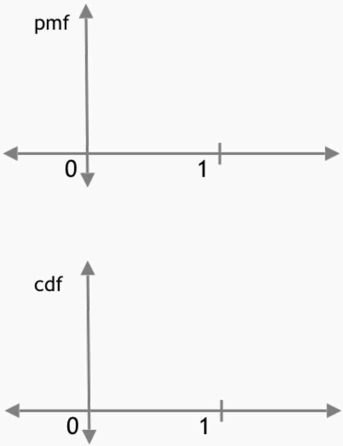
Reading Assignment

Read *Foundations of Agnostic Statistics*, section 1.2.3.

Part 1: Introducing Random Variables

Lightboard: The Cumulative Distribution Function

The Cumulative Distribution Function

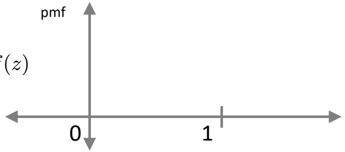


The Cumulative Distribution Function (Solution)

For R.V. X, $F(x) = P(X \leq x)$ for all $x \in \mathbb{R}$

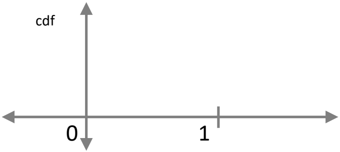
For discrete X,

$$P(X \leq x) = P\big(X \in (-\infty, x]\big) = \sum_{z \in X(\Omega), z \leq x} f(z)$$



Ex: X ~ Bernoulli (p)

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - p, & 0 \leq x < 1 \\ 1, & 1 \leq x \end{cases}$$



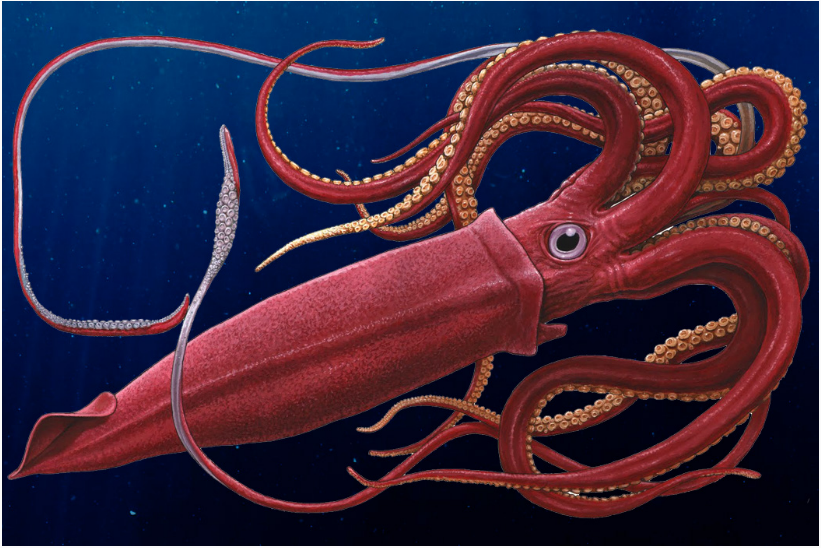
Key properties

1. F is Nondecreasing
2. $lim_{x \to -\infty} F(x) = 0$
3. $lim_{x \to \infty} F(x) = 1$

Part 1: Introducing Random Variables

Continuous Random Variables

Motivation for Continuous Random Variables



Motivation for Continuous Random Variables

Motivation for Continuous Random Variables

Motivation for Continuous Random Variables

The Probability Density Function

Part 1: Introducing Random Variables

Reading: Continuous Random Variables

Reading: Continuous Random Variables

Read *Foundations of Agnostic Statistics*, section 1.2.4.

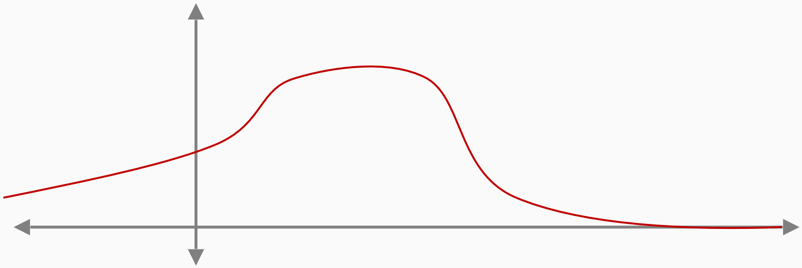
Part 1: Introducing Random Variables

Properties of Continuous Random Variables

Definition

A random variable X is *continuous* if there exists a non-negative function $f : \mathbb{R} \to \mathbb{R}$ such that the cdf of X is

$$F(x) = \int_{-\infty}^{x} f(u)du, \text{ for all } x \in \mathbb{R}$$

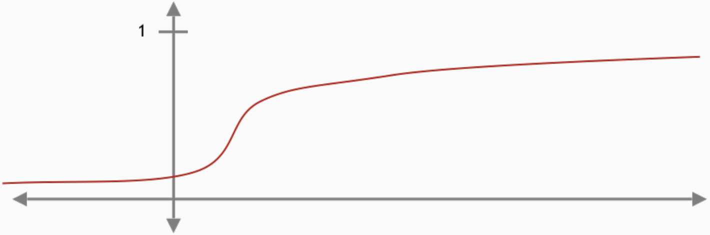


Characterizing Continuous Random Variables

Deriving the PDF from the CDF

For a continuous random variable with cdf F, the probability density function of X is

$$f(x) = \left.\frac{dF(u)}{du}\right|_{u=x}, \text{ for all } x \in \mathbb{R}$$



Properties of Probability Density Functions

For a continuous random variable X with pdf f,

1. For all $x \in \mathbb{R}$, $f(x) \geq 0$.
2. $\int_{-\infty}^{\infty} f(x)dx = 1$.

Support

Definition

For a random variable X with pmf or pdf f, the *support* of X is

$$\text{Supp}[X] = \{x \in \mathbb{R} : f(x) > 0\}$$

Part 1: Introducing Random Variables

Uniform Random Variables

Uniform Random Variables

Note: This is a placeholder for a video that we're going to re-use from: Use the old video 3.9 Uniform Random Variables - only the start through 6.12, before I start talking about expectation.

Part 1: Introducing Random Variables

Normal Random Variables

Use the old video 3.10 The Normal Distribution

Part 1: Introducing Random Variables

Normal Random Variables

The Normal Distribution

The most important distribution in statistics

- The variables measured don't always come directly from a normal distribution
    - Human height has a close-to-normal distribution–so do anthropometric measurements on fossils, reaction times in psychological experiments, etc.
    - Most variables deviate from a normal distribution
- We'll see how the normal distribution pops up as a consequence of basic mathematical laws
- A famous theorem is the **Central Limit Theorem**. Many of our statistical tests depend on this theorem, and it tells us when certain distributions will approach a normal distribution

The Normal Distribution equation

A continuous rv X is said to have a **normal distribution** with parameters μ and σ (or μ and σ^2), where $-\infty < \mu < \infty$ and $0 < \sigma$, if the pdf of X is:

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}, -\infty < x < \infty \qquad (1)$$

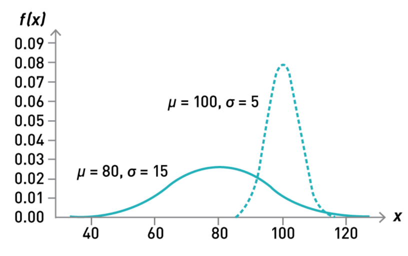
This is the equation for the density function of a normal distribution

The Normal Distribution equation

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}, -\infty < x < \infty \qquad (2)$$

- There are two parameters, μ and σ
    - There is an entire family of normal distributions
- It can be shown that $E(x) = \mu$ and $V(X) = \sigma^2$, so the parameters are the mean and the standard deviation of X
- The statement that X is normally distributed with parameters μ and σ^2 is often abbreviated $X \sim N(\mu, \sigma^2)$

Normal Curve Examples



- These are normal curves with different μ and σ
- Each curve is symmetric around its mean
- μ is both the mean and the median, changing it moves the curve left or right
- σ is also called the scale parameter, since it stretches the curve horizontally

The Standard Normal Distribution

- Often, we need to select one representative distribution out of the normal family
- Typically, we choose $\mu = 0$ and $\sigma = 1$
    - This is called the *standard normal distribution*, also called a *z-distribution*
    - It is also denoted by a capital Z
    - We can write the pdf of Z:

$$f(z; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, -\infty < X < \infty \tag{3}$$

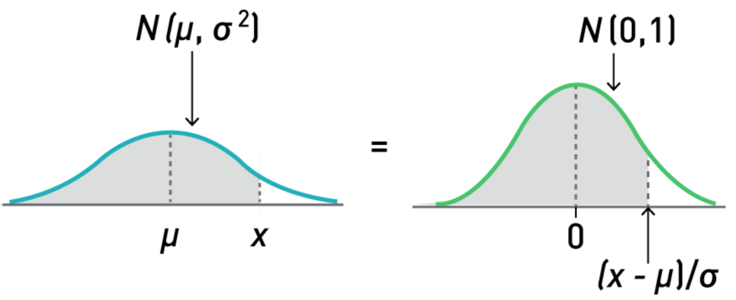
- The cumulative distribution of Z is often written as $\phi(z)$

Standardizing a Normal Variable

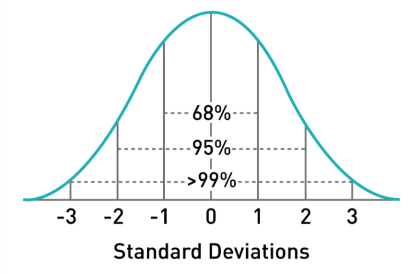
- Suppose we have a normal variable that is not standard normal $X \sim N(\mu, \sigma^2)$
  - We want to simplify our computations
- We can compute the random variable $(X - \mu)/\sigma$
  - This new variable has a standard normal distribution
  - This can make it easier to perform tests using the variable
- Any probabilities involving X can be expressed in terms of areas under the z-distribution

Standardizing a Normal Variable (cont.)

These are easy to compute in statistical software



Properties of the Normal Distribution



This picture shows certain areas under a normal curve

- 68% of the area is within 1 standard deviation of the mean
- 95% is within 2 standard deviations
- 99.7% is within 3 standard deviations

Properties of the Normal Distribution (cont.)

Important Rules of Thumb (helpful to memorize)

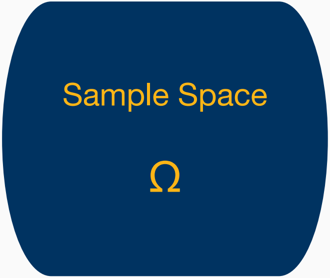
1. It can be helpful when you have a variable that is approximately normally distributed
2. You would expect about 68% of values to be within 1 standard deviation
3. Very few values are more than 2 standard deviation away from the mean (this will be important when we look at hypothesis testing)

Part 2: Joint Distributions

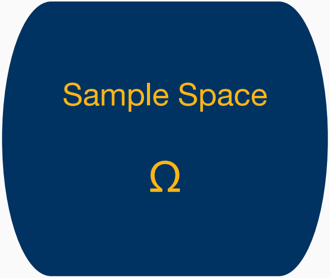
Part 2: Joint Distributions

Joint Distributions

Introducing Joint Distributions



Introducing Joint Distributions



Part 2: Joint Distributions

Reading Assignment

Reading Assignment

Read *Foundations of Agnostic Statistics*, section 1.3–1.3.1.

Part 2: Joint Distributions

Discrete Bivariate Distributions

Equality of Random Variables

Equality of random variables

Let X and Y be random variables. Then, $X = Y$ if, for all $\omega \in \Omega, X(\omega) = Y(\omega)$.

Equality of Random Variables

Equality of random variables

Let X and Y be random variables. Then, $X = Y$ if, for all $\omega \in \Omega, X(\omega) = Y(\omega)$.

Don't assume $X = Y$ just because:

1. X and Y have the same distribution
    - $X = 1$ if Paul's coin lands heads, 0 otherwise
    - $Y = 1$ if Alex's coin lands heads, 0 otherwise
    - $f_X = f_Y$ but $X \neq Y$

Equality of Random Variables

Equality of random variables

Let X and Y be random variables. Then, $X = Y$ if, for all $\omega \in \Omega, X(\omega) = Y(\omega)$.

Don't assume $X = Y$ just because:

1. X and Y have the same distribution
    - $X = 1$ if Paul's coin lands heads, 0 otherwise
    - $Y = 1$ if Alex's coin lands heads, 0 otherwise
    - $f_X = f_Y$ but $X \neq Y$
2. $P(X = Y) = 1$
    - $X = 1$ if Paul climbs Half Dome, 0 otherwise
    - $Y = 1$ if Alex climbs El Capitan, 0 otherwise
    - $P(X = Y) = 1$ but $X \neq Y$

The Joint PMF and Car Dealers

$$
\begin{aligned}
B: &\quad \Omega \to \{0,1\}, && \text{Does the customer buy?} \\
D: &\quad \Omega \to \{0,1,2\}, && \text{Number of test drives}
\end{aligned}
$$

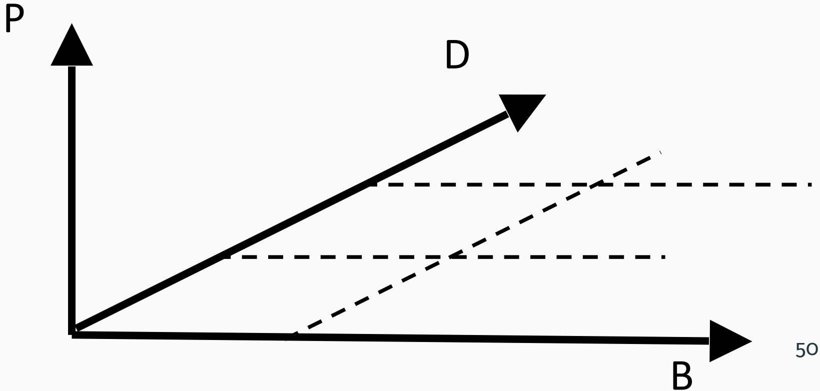
$$f(b,d) = P(B = b, D = d)$$

The Joint PMF and Car Dealers (cont.)

	$D = 0$	$D = 1$	$D = 2$
$B = 0$	0.1	0.4	0.1
$B = 1$	0.1	0.1	0.2

The Joint PMF and Car Dealers (cont.)

	$D = 0$	$D = 1$	$D = 2$
$B = 0$	0.1	0.4	0.1
$B = 1$	0.1	0.1	0.2



The Joint Cumulative Probability Function

$$F(b, d) = P(B \leq b, D \leq d)$$

- Recall that $F_X(x) = \int_{-\infty}^{x} f(x)$ relates the CDF and PDF for the single variable case.
- In a similar way:

$$F_{Y,X} = \int_{-\infty}^{x} \int_{-\infty}^{y} f(y, x) dydx$$

- This can be hard to work with but can describe *any* random variables.

Part 2: Joint Distributions

Learnosity: The Probability Mass Function

Learnosity: The Probability Mass Function

Suppose that X_1 is a Bernoulli random variable representing a fair coin. Suppose that X_2 is a Bernoulli random variable representing a second fair coin. Let $Y = X_1 + X_2$.

Fill in the following table for the PMF of X_1 and Y.

	$X_1 = 0$	$X_1 = 1$
$Y = 0$		
$Y = 1$		
$Y = 2$		

Part 2: Joint Distributions

Discrete Marginal and Conditional Distributions

Part 2: Joint Distributions

Reading Assignment: Discrete Marginal and Conditional Distributions

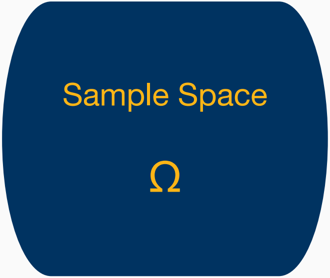
Reading Assignment

Read *Foundations of Agnostic Statistics*, section 1.3.2.

Part 2: Joint Distributions

Discrete Marginal Distributions

Marginal vs. Joint Distributions



From Joint to Marginal Distributions

$$f_B(b) = P(B = b) = \sum_{d \in \text{Supp}[D]} f(b, d) \text{ for all } b \in \mathbb{R}$$

	$B = 0$	$B = 1$
$D = 0$	.1	.1
$D = 1$	.4	.1
$D = 2$	.1	.2

From Marginal to Joint Distributions?

Is it possible to compute the joint distribution, f, given the marginals f_B and f_D?

	$B = 0$	$B = 1$	F_D
$D = 0$			0.2
$D = 1$			0.5
$D = 2$			0.3
F_B	0.6	0.4	

From Marginal to Joint Distributions? (cont.)

Two joint distributions with the same marginals:

	$B = 0$	$B = 1$	F_D
$D = 0$	0.1	0.1	0.2
$D = 1$	0.4	0.1	0.5
$D = 2$	0.1	0.2	0.3
F_B	0.6	0.4	

	$B = 0$	$B = 1$	F_D
$D = 0$			0.2
$D = 1$			0.5
$D = 2$			0.3
F_B	0.6	0.4	

Part 2: Joint Distributions

Discrete Conditional Distributions

The Conditional PMF

- Given two test drives ($D = 2$), what does the model say about buying behavior (B)?
- What is the conditional pmf, $f_{B|D}$?

	$B = 0$	$B = 1$	F_D
$D = 0$	0.1	0.1	0.2
$D = 1$	0.4	0.1	0.5
$D = 2$	0.1	0.2	0.3
F_B	0.6	0.4	

The Conditional PMF (cont.)

Conditional expectation, car dealers

$$\begin{aligned} f_{B|D}(b|d) &= P(B = b|D = d) \\ &= \frac{P(B = b, D = d)}{P(D = d)} = \frac{f(b, d)}{f_D(d)} \end{aligned}$$

The Conditional PMF (cont.)

Conditional expectation, car dealers

$$
\begin{aligned}
f_{B|D}(b|d) &= P(B = b|D = d) \\
&= \frac{P(B = b, D = d)}{P(D = d)} = \frac{f(b, d)}{f_D(d)}
\end{aligned}
$$

$$\text{conditional} = \frac{\text{joint}}{\text{marginal}} \Leftrightarrow$$

The Conditional PMF (cont.)

Conditional expectation, car dealers

$$\begin{aligned} f_{B|D}(b|d) &= P(B = b|D = d) \\ &= \frac{P(B = b, D = d)}{P(D = d)} = \frac{f(b, d)}{f_D(d)} \end{aligned}$$

$$\text{conditional} = \frac{\text{joint}}{\text{marginal}} \Leftrightarrow \text{joint} = \text{marginal} \cdot \text{conditional}$$

Part 2: Joint Distributions

Reading: Jointly Continuous Random Variables

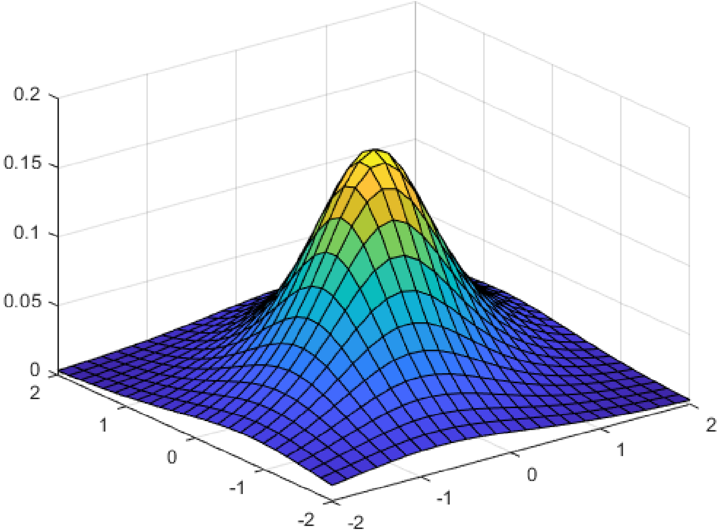
Reading Assignment

Read *Foundations of Agnostic Statistics*, section 1.3.3.

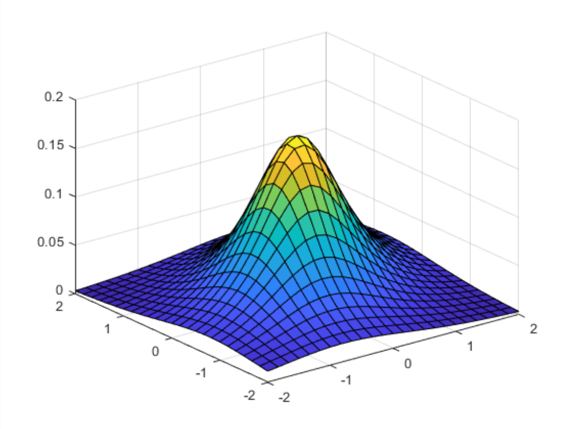
Part 2: Joint Distributions

Jointly Continuous Random Variables

The Joint Density Function

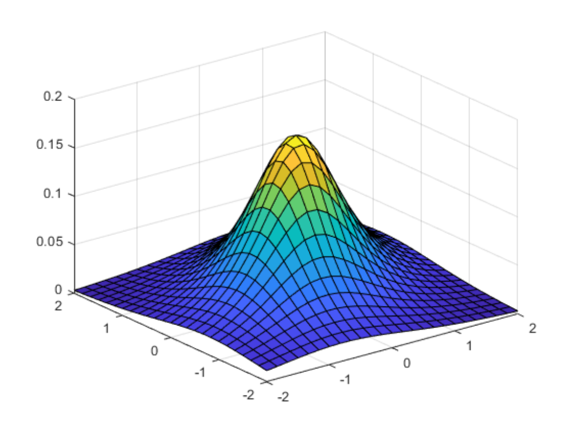


Probabilities of Events



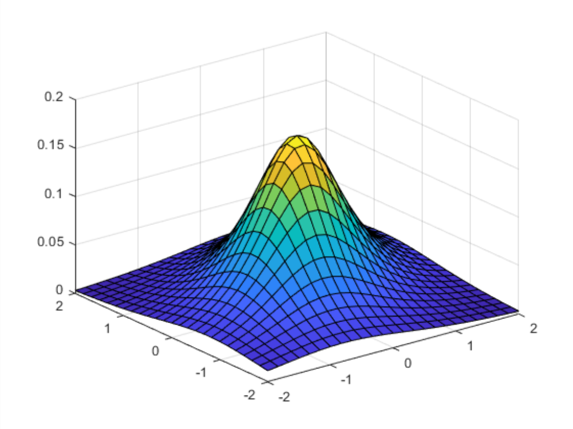
How to compute:
$P((X, Y) \in A)$

Continuous Marginal Distributions



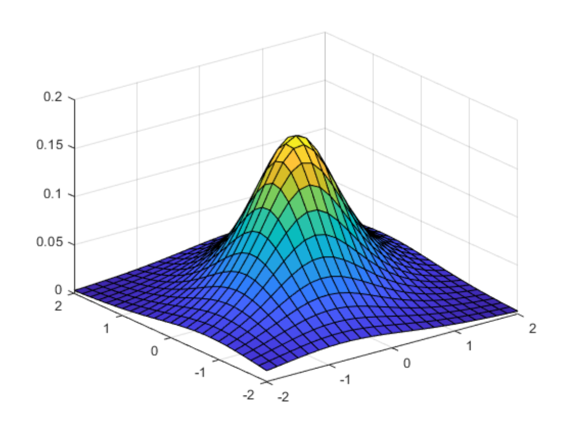
How to compute:
$f_X(x)$

Continuous Conditional Distributions



How to compute:
$f_{B|D}(b|d)$

Continuous Conditional Distributions



How to compute:
$f_{B|D}(b|d)$

$$\text{conditional} = \frac{\text{joint}}{\text{marginal}}$$

Part 2: Joint Distributions

Reading: Independent Random Variables

Reading Assignment

Read *Foundations of Agnostic Statistics*, section 1.3.4.

Part 2: Joint Distributions

Independent Random Variables

Independence

If X and Y are independent, $X \perp\!\!\!\perp Y$.

Independence

If X and Y are independent, $X \perp\!\!\!\perp Y$.

1. The marginals fully describe the distribution: $f(x, y) = f_X(x) f_Y(y)$

Independence

If X and Y are independent, $X \perp\!\!\!\perp Y$.

1. The marginals fully describe the distribution: $f(x, y) = f_X(x)f_Y(y)$
2. Knowing X tells you nothing about Y: $f_{Y|X}(y|x) =$

Drawing Independent Continuous RV

Drawing Non-Independent Continuous RV

Importance of Independence

B_1 represents blood pressure of patient 1.

B_2 represents blood pressure of patient 2.

$\vdots$

Without simplifying assumptions, the joint distribution of $B_1, B_2, \ldots$ is too complex to estimate with limited data.

Independence $\implies$ only need to estimate the marginals

Assessing Independence in Practice

Independence is rarely (if ever) perfectly met.

What if:

- Patient 1 and patient 2 are related
- Technicians trade off every 10 patients
- After seeing an unusually high blood pressure, the technician adjusts their mannerisms

These are potential *dependencies*.