Some problems
Some problems

A means to capture uncertainty
Some problems

A means to capture uncertainty
You have data from two sources, are they different?
Some problems

A means to capture uncertainty
You have data from two sources, are they different?
How can you generate data or fill in missing data?
Some problems

A means to capture uncertainty
You have data from two sources, are they different?
How can you generate data or fill in missing data?
What explains the observed data?
Every event has a probability of occurring, some event always occurs, and combining separate events adds their probabilities. Why these axioms? Many other choices are possible: Possibility theory, probability intervals, belief functions, upper and lower probabilities.
Every event has a probability of occurring, some event always occurs, and combining separate events adds their probabilities.
Every event has a probability of occurring, some event always occurs, and combining separate events adds their probabilities.

Why these axioms? Many other choices are possible:
Every event has a probability of occurring, some event always occurs, and combining separate events adds their probabilities.

Why these axioms? Many other choices are possible:
Possibility theory, probability intervals
Every event has a probability of occurring, some event always occurs, and combining separate events adds their probabilities.

Why these axioms? Many other choices are possible:
Possibility theory, probability intervals
Every event has a probability of occurring, some event always occurs, and combining separate events adds their probabilities.

Why these axioms? Many other choices are possible:
- Possibility theory, probability intervals
- Belief functions, upper and lower probabilities
I offer that if $X$ happens I pay out $R$, otherwise I keep your money.
I offer that if $X$ happens I pay out $R$, otherwise I keep your money. Why should I always value my bet at $pR$, where $p$ is the probability of $X$?
I offer that if $X$ happens I pay out $R$, otherwise I keep your money. Why should I always value my bet at $pR$, where $p$ is the probability of $X$?

Negative $p$
I offer that if $X$ happens I pay out $R$, otherwise I keep your money. Why should I always value my bet at $pR$, where $p$ is the probability of $X$?

Negative $p$

I’m offering to pay $R$ for $-pR$ dollars.
I offer that if $X$ happens I pay out $R$, otherwise I keep your money. Why should I always value my bet at $pR$, where $p$ is the probability of $X$?

Negative $p$

I’m offering to pay $R$ for $-pR$ dollars.

Probabilities don’t sum to one
Probability is special: Dutch books

I offer that if $X$ happens I pay out $R$, otherwise I keep your money. Why should I always value my bet at $pR$, where $p$ is the probability of $X$?

Negative $p$
I’m offering to pay $R$ for $-pR$ dollars.

Probabilities don’t sum to one
Take out a bet that always pays off.
Probability is special: Dutch books

I offer that if $X$ happens I pay out $R$, otherwise I keep your money.

Why should I always value my bet at $pR$, where $p$ is the probability of $X$?

Negative $p$

I’m offering to pay $R$ for $-pR$ dollars.

Probabilities don’t sum to one

Take out a bet that always pays off.

If the sum is below 1, I pay $R$ for less than $R$ dollars.
I offer that if $X$ happens I pay out $R$, otherwise I keep your money.

Why should I always value my bet at $pR$, where $p$ is the probability of $X$?

Negative $p$

I’m offering to pay $R$ for $-pR$ dollars.

Probabilities don’t sum to one

Take out a bet that always pays off.

If the sum is below 1, I pay $R$ for less than $R$ dollars.

If the sum is above 1, buy the bet and sell it to me for more.
I offer that if $X$ happens I pay out $R$, otherwise I keep your money. Why should I always value my bet at $pR$, where $p$ is the probability of $X$?

Negative $p$

I’m offering to pay $R$ for $-pR$ dollars.

Probabilities don’t sum to one

Take out a bet that always pays off.

If the sum is below 1, I pay $R$ for less than $R$ dollars.

If the sum is above 1, buy the bet and sell it to me for more.

When $X$ and $Y$ are incompatible the value isn’t the sum
I offer that if $X$ happens I pay out $R$, otherwise I keep your money.

Why should I always value my bet at $pR$, where $p$ is the probability of $X$?

Negative $p$

I’m offering to pay $R$ for $-pR$ dollars.

Probabilities don’t sum to one

Take out a bet that always pays off.

If the sum is below 1, I pay $R$ for less than $R$ dollars.
If the sum is above 1, buy the bet and sell it to me for more.

When $X$ and $Y$ are incompatible the value isn’t the sum

If the value is bigger, I still pay out more.
I offer that if $X$ happens I pay out $R$, otherwise I keep your money. Why should I always value my bet at $pR$, where $p$ is the probability of $X$?

Negative $p$
I’m offering to pay $R$ for $-pR$ dollars.

Probabilities don’t sum to one
Take out a bet that always pays off.
If the sum is below 1, I pay $R$ for less than $R$ dollars.
If the sum is above 1, buy the bet and sell it to me for more.

When $X$ and $Y$ are incompatible the value isn’t the sum
If the value is bigger, I still pay out more.
If the value is smaller, sell me my own bets.
Probability is special: Dutch books

I offer that if $X$ happens I pay out $R$, otherwise I keep your money. Why should I always value my bet at $pR$, where $p$ is the probability of $X$?

Negative $p$

I’m offering to pay $R$ for $-pR$ dollars.

Probabilities don’t sum to one

Take out a bet that always pays off.

If the sum is below 1, I pay $R$ for less than $R$ dollars.

If the sum is above 1, buy the bet and sell it to me for more.

When $X$ and $Y$ are incompatible the value isn’t the sum

If the value is bigger, I still pay out more.

If the value is smaller, sell me my own bets.

Decisions under probability are “rational”.
Experiments, theory, and funding
Experiments, theory, and funding
Data

Mean

\[ \mu_X = \mathbb{E}[X] = \sum x_p(x) \]

Variance

\[ \sigma^2_X = \text{var}(X) = \mathbb{E}[(X - \mu_X)^2] \]

Covariance

\[ \text{cov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] \]

Correlation

\[ \rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} \]
Mean

\[ \mu_X = E[X] = \sum_x x p(x) \]
Data

Mean \( \mu_X = E[X] = \sum_x x p(x) \)

Variance \( \sigma_X^2 = \text{var}(X) = E[(X - \mu)^2] \)
Data

Mean \[ \mu_X = E[X] = \sum_x x p(x) \]

Variance \[ \sigma_X^2 = \text{var}(X) = E[(X - \mu)^2] \]

Covariance \[ \text{cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] \]
Data

Mean \( \mu_X = E[X] = \sum_x x p(x) \)

Variance \( \sigma_X^2 = \text{var}(X) = E[(X - \mu)^2] \)

Covariance \( \text{cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] \)

Correlation \( \rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} \)
Often implicitly assume that our data comes from a normal distribution.
Mean and variance

Often implicitly assume that our data comes from a normal distribution.
That our samples are i.i.d. (independent identically distributed).
Often implicitly assume that our data comes from a normal distribution.

That our samples are i.i.d. (independent identically distributed).

Generally these don’t capture enough about the underlying data.
Mean and variance

Often implicitly assume that our data comes from a normal distribution.
That our samples are i.i.d. (independent identically distributed).
Generally these don’t capture enough about the underlying data.
Uncorrelated does not mean independent!
Correlation vs independence

\[ V = N(0, 1), \quad X = \sin(V), \quad Y = \cos(V) \]

Correlation only measures linear relationships.

Andrei Barbu (MIT)

Probability

August 2018
Correlation vs independence

\[ V = N(0, 1), \ X = \sin(V), \ Y = \cos(V) \]
Correlation vs independence

\[ V = N(0, 1), \quad X = \sin(V), \quad Y = \cos(V) \]

Correlation only measures linear relationships.
Data dinosaurs

X Mean: 54.2659224
Y Mean: 47.8313999
X SD : 16.7649829
Y SD : 26.9342120
Corr. : -0.0642526
Data

Mean
Variance
Covariance
Correlation

Are two players the same?
How do you know and how certain are you?
What about two players is different?
How do you quantify which differences matter?
Here’s a player, how good will they be?
What is the best information to ask for?
What is the best test to run?
If I change the size of the board, how might the results change?

Andrei Barbu (MIT)
Data

Mean
Variance
Covariance
Correlation

Are two players the same? How do you know and how certain are you? What about two players is different? How do you quantify which differences matter?

Here's a player, how good will they be? What is the best information to ask for? What is the best test to run?

If I change the size of the board, how might the results change?
Are two players the same?

Mean
Variance
Covariance
Correlation
Are two players the same?
How do you know and how certain are you?

Mean
Variance
Covariance
Correlation
Are two players the same?
How do you know and how certain are you?
What about two players is different?

Mean
Variance
Covariance
Correlation
Are two players the same? How do you know and how certain are you? What about two players is different? How do you quantify which differences matter?
Data

Are two players the same?
How do you know and how certain are you?
What about two players is different?
How do you quantify which differences matter?
Here’s a player, how good will they be?

Mean
Variance
Covariance
Correlation
Are two players the same?
How do you know and how certain are you?
What about two players is different?
How do you quantify which differences matter?
Here’s a player, how good will they be?
What is the best information to ask for?
Are two players the same?
How do you know and how certain are you?
What about two players is different?
How do you quantify which differences matter?
Here’s a player, how good will they be?
What is the best information to ask for?
What is the best test to run?

Mean
Variance
Covariance
Correlation
Are two players the same?
How do you know and how certain are you?
What about two players is different?
How do you quantify which differences matter?
Here’s a player, how good will they be?
What is the best information to ask for?
What is the best test to run?
If I change the size of the board, how might the results change?
Are two players the same?

How do you know and how certain are you?

What about two players is different?

How do you quantify which differences matter?

Here’s a player, how good will they be?

What is the best information to ask for?

What is the best test to run?

If I change the size of the board, how might the results change?

...
A machine enters a random state, its current state is an event. Events, \( x \), have probabilities associated, \( p(x) \) (shorthand).

Sets of events, \( A \), Random variables, \( X \), are a function of the event.

The probability of two events

\[
p(A \cup B) = p(A) + p(B) - p(A \cap B)
\]

The probability of either event

\[
p(\neg x) = 1 - p(x)
\]

Joint probabilities

\[
P(x, y) = P(x)P(y)
\]

Independence

\[
P(x, y) = P(x)P(y)
\]

Conditional probabilities

\[
P(x | y) = \frac{P(x, y)}{P(y)}
\]

Law of total probability

\[
\sum_a P(a) = 1 \text{ when events } A \text{ are a disjoint cover}
\]
A machine enters a random state, its current state is an event.
Probability as an experiment

A machine enters a random state, its current state is an event. Events, $x$, have probabilities associated, $p(X = x)$ ($p(x)$ shorthand)
A machine enters a random state, its current state is an event. Events, $x$, have probabilities associated, $p(X = x)$ ($p(x)$ shorthand)

Sets of events, $A$
A machine enters a random state, its current state is an event. Events, $x$, have probabilities associated, $p(X = x)$ ($p(x)$ shorthand). Sets of events, $A$. Random variables, $X$, are a function of the event.
A machine enters a random state, its current state is an event. Events, $x$, have probabilities associated, $p(X = x)$ ($p(x)$ shorthand). Sets of events, $A$. Random variables, $X$, are a function of the event. The probability of two events $p(A \cup B)$. 
Probability as an experiment

A machine enters a random state, its current state is an event. Events, $x$, have probabilities associated, $p(X = x)$ ($p(x)$ shorthand).

Sets of events, $A$

Random variables, $X$, are a function of the event

The probability of two events $p(A \cup B)$

The probability of either event $p(A \cap B)$
A machine enters a random state, its current state is an event. Events, $x$, have probabilities associated, $p(X = x)$ ($p(x)$ shorthand)

Sets of events, $A$

Random variables, $X$, are a function of the event

The probability of two events $p(A \cup B)$

The probability of either event $p(A \cap B)$

$p(\neg x) = 1 - p(x)$
Probability as an experiment

A machine enters a random state, its current state is an event. Events, $x$, have probabilities associated, $p(X = x)$ ($p(x)$ shorthand). Sets of events, $A$

Random variables, $X$, are a function of the event

The probability of two events $p(A \cup B)$
The probability of either event $p(A \cap B)$

$p(\neg x) = 1 - p(x)$

Joint probabilities $P(x, y)$
A machine enters a random state, its current state is an event. Events, $x$, have probabilities associated, $p(X = x)$ ($p(x)$ shorthand)

Sets of events, $A$

Random variables, $X$, are a function of the event

The probability of two events $p(A \cup B)$

The probability of either event $p(A \cap B)$

$p(\neg x) = 1 - p(x)$

Joint probabilities $P(x, y)$

Independence $P(x, y) = P(x)P(y)$
Probability as an experiment

A machine enters a random state, its current state is an event. Events, $x$, have probabilities associated, $p(X = x)$ ($p(x)$ shorthand)

Sets of events, $A$
Random variables, $X$, are a function of the event

The probability of two events $p(A \cup B)$
The probability of either event $p(A \cap B)$

$p(\neg x) = 1 - p(x)$

Joint probabilities $P(x, y)$
Independence $P(x, y) = P(x)P(y)$

Conditional probabilities $P(x|y) = \frac{P(x,y)}{P(y)}$
Probability as an experiment

A machine enters a random state, its current state is an event. Events, $x$, have probabilities associated, $p(X = x)$ ($p(x)$ shorthand)

Sets of events, $A$

Random variables, $X$, are a function of the event

The probability of two events $p(A \cup B)$

The probability of either event $p(A \cap B)$

$p(\neg x) = 1 - p(x)$

Joint probabilities $P(x, y)$

Independence $P(x, y) = P(x)P(y)$

Conditional probabilities $P(x | y) = \frac{P(x, y)}{p(y)}$

Law of total probability $\sum_A a = 1$ when events $A$ are a disjoint cover
Beating the lottery

Andrei Barbu (MIT) Probability August 2018
Beating the lottery

Andrei Barbu  (MIT)
Probability
August 2018
You want to play the lottery, and have a method to win. 0.5% of tickets are winners, and you have a test to verify this. You are 85% accurate (5% false positives, 10% false negatives). Is this test useful? How useful? Should you be betting?

Andrei Barbu (MIT) 
Probability
August 2018
Analyzing a test

You want to play the lottery, and have a method to win.
Analyzing a test

You want to play the lottery, and have a method to win. 0.5% of tickets are winners, and you have a test to verify this.
You want to play the lottery, and have a method to win. 0.5% of tickets are winners, and you have a test to verify this. You are 85% accurate (5% false positives, 10% false negatives)
Analyzing a test

You want to play the lottery, and have a method to win. 0.5% of tickets are winners, and you have a test to verify this. You are 85% accurate (5% false positives, 10% false negatives)
You want to play the lottery, and have a method to win. 0.5% of tickets are winners, and you have a test to verify this. You are 85% accurate (5% false positives, 10% false negatives). Is this test useful? How useful? Should you be betting?
You want to play the lottery, and have a method to win. 0.5% of tickets are winners, and you have a test to verify this. You are 85% accurate (5% false positives, 10% false negatives). Is this test useful? How useful? Should you be betting?
Is this test useful?

Andrei Barbu (MIT)  Probability  August 2018
Is this test useful?

\[(D-, T+) (D-, T-) (D+, T+) (D+, T-)\]

\[
\begin{align*}
D- & \quad .995 \\
T+ & \quad .05 \\
T- & \quad .95 \\
D+ & \quad .005 \\
T+ & \quad .9 \\
T- & \quad .1
\end{align*}
\]
Is this test useful?

\[(D-, T+) \ (D-, T-) \ (D+, T+) \ (D+, T-)]

What percent of the time when my test comes up true am I winner?
Is this test useful?

\[
\begin{align*}
(D-, T+) & \quad (D-, T-) \quad (D+, T+) \quad (D+, T-) \\
D- & \quad .995 & \quad D+ & \quad .005 \\
T+ & \quad .05 & \quad T- & \quad .95 \\
T+ & \quad .9 & \quad T- & \quad .1
\end{align*}
\]

What percent of the time when my test comes up true am I winner?

\[
\frac{D+ \cap T+}{T+} \approx 0.995 \times 0.05 + 0.995 \times 0.95 = 0.833 = P(T+ | D+) P(D+) P(T+) \]

Andrei Barbu (MIT) Probability August 2018
Is this test useful?

\[(D-, T+) (D-, T-) (D+, T+) (D+, T-)\]

What percent of the time when my test comes up true am I winner?

\[
\frac{D+ \cap T+}{T+} = \frac{0.9 \times 0.005}{0.9 \times 0.005 + 0.995 \times 0.05}
\]
Is this test useful?

\[(D-, T+) (D-, T-) (D+, T+) (D+, T-)\]

What percent of the time when my test comes up true am I winner?

\[
\frac{D + \cap T+}{T+} = \frac{0.9 \times 0.005}{0.9 \times 0.005 + 0.995 \times 0.05} = 8.3\%
\]
Is this test useful?

What percent of the time when my test comes up true am I winner?

\[
\frac{D^+ \cap T^+}{T^+} = \frac{0.9 \times 0.005}{0.9 \times 0.005 + 0.995 \times 0.05} = 8.3\% = \frac{P(T^+ | D^+) P(D^+)}{P(T^+)}
\]
Is this test useful?

\[(D-, T+) \quad (D-, T-) \quad (D+, T+) \quad (D+, T-)\]

What percent of the time when my test comes up true am I winner?

\[
\frac{D^+ \cap T^+}{T^+} = \frac{0.9 \times 0.005}{0.9 \times 0.005 + 0.995 \times 0.05} = 8.3\% = \frac{P(T^+ | D^+) P(D^+)}{P(T^+)}
\]

\[
P(A | B) = \frac{P(B | A) P(A)}{P(B)} \quad \text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{probability of data}}
\]
**Bernoulli Distribution**

- Given a **Bernoulli experiment**, that is, a **yes/no experiment** with outcomes 0 (“failure”) or 1 (“success”)
- The Bernoulli distribution is a **discrete probability distribution**, which takes value 1 with success probability $\lambda$ and value 0 with failure probability $1 - \lambda$

**Probability mass function**

$$
\begin{align*}
    p(x = 0) &= 1 - \lambda \\
    p(x = 1) &= \lambda
\end{align*}
$$

- **Notation**

$$
\text{Bern}_x(\lambda) = \lambda^x(1 - \lambda)^{1-x}
$$

**Parameters**
- $\lambda$: probability of observing a success

**Expectation**
- $E[x] = \lambda$

**Variance**
- $\text{Var}[x] = \lambda(1 - \lambda)$
Binomial Distribution

- Given a sequence of Bernoulli experiments
- The binomial distribution is the discrete probability distribution of the number of successes \( m \) in a sequence of \( N \) independent yes/no experiments, each of which yields success with probability \( \lambda \)
- **Probability mass function**

\[
p(m) = \binom{N}{m} \lambda^m (1 - \lambda)^{N-m}
\]

- **Notation**

\[
Bin_m(N, \lambda) = \binom{N}{m} \lambda^m (1 - \lambda)^{N-m}
\]
Gaussian Distribution

- **Most widely** used distribution for continuous variables
- Reasons: (i) simplicity (fully represented by only two moments, mean and variance) and (ii) the central limit theorem (CLT)
- The CLT states that, under mild conditions, the mean (or sum) of many independently drawn random variables is distributed approximately normally, irrespective of the form of the original distribution
- **Probability density function**

\[ p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]

**Parameters**
- \( \mu \): mean
- \( \sigma^2 \): variance

**Expectation**
- \( E[x] = \mu \)

**Variance**
- \( \text{Var}[x] = \sigma^2 \)
Gaussian Distribution

• Notation

\[ \mathcal{N}_x(\mu, \sigma^2) = p(x) \]

• Called standard normal distribution for \( \mu = 0 \) and \( \sigma = 1 \)

• About 68% (~two third) of values drawn from a normal distribution are within a range of ±1 standard deviations around the mean

• About 95% of the values lie within a range of ±2 standard deviations around the mean

• Important e.g. for hypothesis testing

Parameters

- \( \mu \): mean
- \( \sigma^2 \): variance

Expectation

- \( E[x] = \mu \)

Variance

- \( \text{Var}[x] = \sigma^2 \)
Multivariate Gaussian Distribution

- For $d$-dimensional random vectors, the **multivariate Gaussian distribution** is governed by a $d$-dimensional **mean vector $\mu$** and a $D \times D$ **covariance matrix $\Sigma$** that must be symmetric and positive semi-definite.

- **Probability density function**
  \[
p(x) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right\}
  \]

- **Notation**
  \[\mathcal{N}_x(\mu, \Sigma) = p(x)\]

**Parameters**
- $\mu$: mean vector
- $\Sigma$: covariance matrix

**Expectation**
- $E[x] = \mu$

**Variance**
- $\text{Var}[x] = \Sigma$
Multivariate Gaussian Distribution

- For \( d = 2 \), we have the **bivariate** Gaussian distribution
- The covariance matrix \( \Sigma \) (often \( C \)) determines the **shape of the distribution** (video)

**Parameters**
- \( \mu \): mean vector
- \( \Sigma \): covariance matrix

**Expectation**
- \( E[x] = \mu \)

**Variance**
- \( \text{Var}[x] = \Sigma \)
Multivariate Gaussian Distribution

- For $d = 2$, we have the **bivariate** Gaussian distribution
- The covariance matrix $\Sigma$ (often $C$) determines the **shape of the distribution** (video)

\[
C = \begin{bmatrix}
0.020 & -0.012 \\
-0.012 & 0.020
\end{bmatrix}
\]

- $\lambda_1 = 0.008$
- $\lambda_2 = 0.032$
- $\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = -0.618$

**Parameters**
- $\mu$: mean vector
- $\Sigma$: covariance matrix

**Expectation**
- $E[x] = \mu$

**Variance**
- $\text{Var}[x] = \Sigma$
Multivariate Gaussian Distribution

- For $d = 2$, we have the **bivariate** Gaussian distribution
- The covariance matrix $\Sigma$ (often $C$) determines the **shape of the distribution** (video)

### Parameters
- $\mu$: mean vector
- $\Sigma$: covariance matrix

### Expectation
- $E[x] = \mu$

### Variance
- $\text{Var}[x] = \Sigma$
Poisson Distribution

- Consider independent events that happen with an average rate of $\lambda$ over time.
- The Poisson distribution is a discrete distribution that describes the probability of a given number of events occurring in a fixed interval of time.
- Can also be defined over other intervals such as distance, area or volume.

Probability mass function

$$p(x) = \frac{\lambda^k e^{-\lambda}}{k!}$$

- Notation

$$\text{Pois}_x(\lambda) = p(x)$$

Parameters

- $\lambda$ : average rate of events over time or space

Expectation

- $E[x] = \lambda$

Variance

- $\text{Var}[x] = \lambda$
Bayesian updates

\[
P(\theta | X) = \frac{P(X | \theta) P(\theta)}{P(X)}
\]

I have some prior over how good a player is: informative vs uninformative.

\[
P(X | \theta) \text{: I think dart throwing is a stochastic process, every player has an unknown mean.}
\]

\[
X : I \text{ observe them throwing darts.}
\]

\[
P(X) \text{: Across all parameters this is how likely the data is.}
\]

Normalization is usually hard to compute, but it's often not needed.

Say \( P(\theta) \) is a normal distribution with mean 0 and high variance. And \( P(X | \theta) \) is also a normal distribution.

What's the best estimate for this player's performance?

\[
\frac{\partial}{\partial \theta} \log P(\theta | X) = 0
\]
Bayesian updates

\[ P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)} \]

I have some prior over how good a player is: informative vs uninformative.

I think dart throwing is a stochastic process, every player has an unknown mean.

I observe them throwing darts.

Across all parameters this is how likely the data is.

Normalization is usually hard to compute, but it's often not needed.

Say \( P(\theta) \) is a normal distribution with mean 0 and high variance.

And \( P(X|\theta) \) is also a normal distribution.

What's the best estimate for this player's performance?
Bayesian updates

\[ P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)} \]

\( P(\theta) \): I have some prior over how good a player is: informative vs uninformative.
Bayesian updates

\[ P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)} \]

\(P(\theta)\): I have some prior over how good a player is: informative vs uninformative.

\(P(X|\theta)\): I think dart throwing is a stochastic process, every player has an unknown mean.
Bayesian updates

\[ P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)} \]

\( P(\theta) \): I have some prior over how good a player is: informative vs uninformative.

\( P(X|\theta) \): I think dart throwing is a stochastic process, every player has an unknown mean.

\( X \): I observe them throwing darts.
Bayesian updates

\[ P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)} \]

- **P(\theta)**: I have some prior over how good a player is: informative vs uninformative.
- **P(X|\theta)**: I think dart throwing is a stochastic process, every player has an unknown mean.
- **X**: I observe them throwing darts.
- **P(X)**: Across all parameters this is how likely the data is.

Normalization is usually hard to compute, but it's often not needed.

Say **P(\theta)** is a normal distribution with mean 0 and high variance. And **P(X|\theta)** is also a normal distribution.

What's the best estimate for this player's performance?

\[ \frac{\partial}{\partial \theta} \log P(\theta|X) = 0 \]
Bayesian updates

\[ P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)} \]

\( P(\theta) \): I have some prior over how good a player is: informative vs uninformative.

\( P(X|\theta) \): I think dart throwing is a stochastic process, every player has an unknown mean.

\( X \): I observe them throwing darts.

\( P(X) \): Across all parameters this is how likely the data is.

Normalization is usually hard to compute, but it’s often not needed.
Bayesian updates

\[ P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)} \]

\( P(\theta) \): I have some prior over how good a player is: informative vs uninformative.

\( P(X|\theta) \): I think dart throwing is a stochastic process, every player has an unknown mean.

\( X \): I observe them throwing darts.

\( P(X) \): Across all parameters this is how likely the data is.

Normalization is usually hard to compute, but it’s often not needed.

Say \( P(\theta) \) is a normal distribution with mean 0 and high variance.
Bayesian updates

\[ P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)} \]

\(P(\theta)\): I have some prior over how good a player is: informative vs uninformative.

\(P(X|\theta)\): I think dart throwing is a stochastic process, every player has an unknown mean.

\(X\): I observe them throwing darts.

\(P(X)\): Across all parameters this is how likely the data is. Normalization is usually hard to compute, but it’s often not needed.

Say \(P(\theta)\) is a normal distribution with mean 0 and high variance.
And \(P(X|\theta)\) is also a normal distribution.
Bayesian updates

\[ P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)} \]

\( P(\theta) \): I have some prior over how good a player is: informative vs uninformative.

\( P(X|\theta) \): I think dart throwing is a stochastic process, every player has an unknown mean.

\( X \): I observe them throwing darts.

\( P(X) \): Across all parameters this is how likely the data is.

Normalization is usually hard to compute, but it’s often not needed.

Say \( P(\theta) \) is a normal distribution with mean 0 and high variance.
And \( P(X|\theta) \) is also a normal distribution.

What’s the best estimate for this player’s performance?
Bayesian updates

\[ P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)} \]

\( P(\theta) \): I have some prior over how good a player is: informative vs uninformative.

\( P(X|\theta) \): I think dart throwing is a stochastic process, every player has an unknown mean.

\( X \): I observe them throwing darts.

\( P(X) \): Across all parameters this is how likely the data is.

Normalization is usually hard to compute, but it’s often not needed.

Say \( P(\theta) \) is a normal distribution with mean 0 and high variance.

And \( P(X|\theta) \) is also a normal distribution.

What’s the best estimate for this player’s performance?

\[ \frac{\partial}{\partial \theta} \log P(\theta|X) = 0 \]
Graphical models

So far we've talked about independence, conditioning, and observation. A toolkit to discuss these at a higher level of abstraction.
So far we’ve talked about independence, conditioning, and observation.
Graphical models

So far we’ve talked about independence, conditioning, and observation.

A toolkit to discuss these at a higher level of abstraction.
So far we’ve talked about independence, conditioning, and observation.

A toolkit to discuss these at a higher level of abstraction.
Break down a speech signal into parts.
Speech recognition: Naïve bayes

Break down a speech signal into parts.
Speech recognition: Naïve bayes

Break down a speech signal into parts.

Recover the original speech
Create a set of features, each sound is composed of combinations of features.
Speech recognition: Naïve bayes

Create a set of features, each sound is composed of combinations of features.
Speech recognition: Naïve bayes

Create a set of features, each sound is composed of combinations of features.

\[ P(c|X) \propto \prod_{k} P(X_k|c)P(c) \]
Speech recognition: Gaussian mixture model
Speech recognition: Gaussian mixture model
Speech recognition: Gaussian mixture model
Speech recognition: Hidden Markov model
Speech recognition: Hidden Markov model

\[ x(t-1) \rightarrow x(t) \rightarrow x(t+1) \]

\[ y(t-1) \rightarrow y(t) \rightarrow y(t+1) \]
Summary

Probabilities defined in terms of events
Random variables and their distributions
Reasoning with probabilities and Bayes' rule
Updating our knowledge over time
Graphical models to reason abstractly

A quick tour of how we would build a more complex model

Andrei Barbu (MIT)
Probabilities defined in terms of events
Probabilities defined in terms of events
Random variables and their distributions
Probabilities defined in terms of events
Random variables and their distributions
Reasoning with probabilities and Bayes’ rule
Probabilities defined in terms of events
Random variables and their distributions
Reasoning with probabilities and Bayes’ rule
Updating our knowledge over time
Probabilities defined in terms of events
Random variables and their distributions
Reasoning with probabilities and Bayes’ rule
Updating our knowledge over time
Graphical models to reason abstractly
Probabilities defined in terms of events
Random variables and their distributions
Reasoning with probabilities and Bayes’ rule
Updating our knowledge over time
Graphical models to reason abstractly
A quick tour of how we would build a more complex model