Intro to Probability

Andrei Barbu

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?

Uncertainty

Uncertainty

Every event has a probability of occurring, some event always occurs, and combining separate events adds their probabilities.

Uncertainty

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:

Uncertainty

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

Uncertainty

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

Uncertainty

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

Probability is special: Dutch books

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

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?

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

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.

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

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.

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.

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

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.

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.

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.

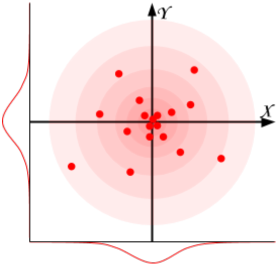
If the value is smaller, sell me my own bets.

Decisions under probability are "rational".

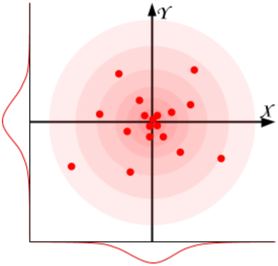
Experiments, theory, and funding

Experiments, theory, and funding

Data

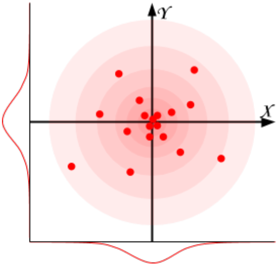


Data



Mean $\mu_X = E[X] = \sum_x x\mathrm{p}(x)$

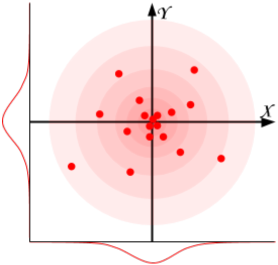
Data



Mean $\mu_X = E[X] = \sum_x x \mathsf{p}(x)$

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

Data

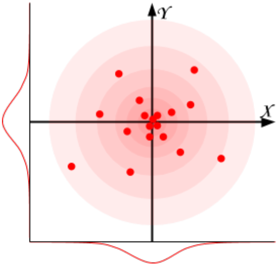


Mean $$\mu_X = E[X] = \sum_x x\mathrm{p}(x)$$

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

Covariance $$\mathrm{cov}(X,Y) = E[(X-\mu_x)(Y-\mu_Y)]$$

Data



Mean $$\mu_X = E[X] = \sum_x x\mathrm{p}(x)$$

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

Covariance $$\mathrm{cov}(X,Y) = E[(X-\mu_x)(Y-\mu_Y)]$$

Correlation $$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$

Mean and variance

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 indentically distributed).

Mean and variance

Often implicitly assume that our data comes from a normal distribution.

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

Generally these don't capture enough about the underlying data.

Uncorrelated does not mean independent!

Correlation vs independence

Correlation vs independence

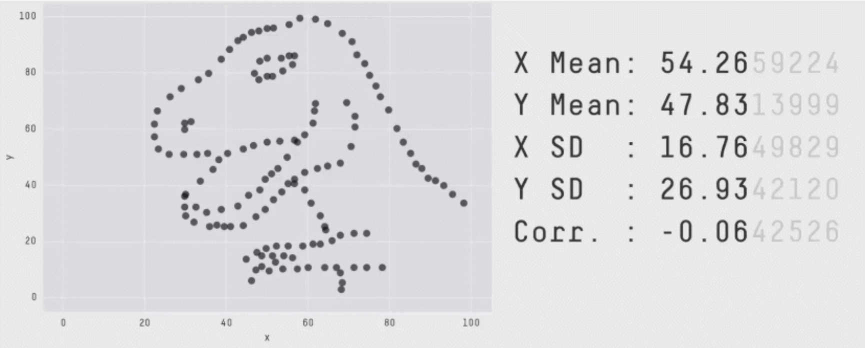
$$V = N(0,1),\ X = sin(V),\ Y = cos(V)$$

Correlation vs independence

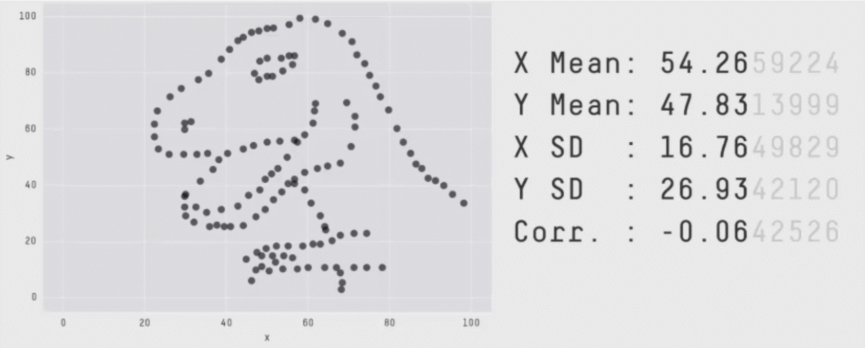
$$V = N(0,1),\ X = sin(V),\ Y = cos(V)$$

Correlation only measures linear relationships.

Data dinosaurs

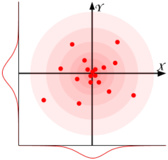


Data dinosaurs



Data

Data



Mean

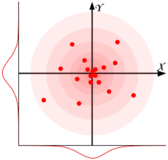
Variance

Covariance

Correlation

Data

Are two players the same?



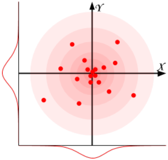
Mean

Variance

Covariance

Correlation

Data



Mean

Variance

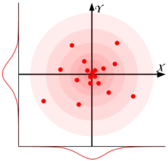
Covariance

Correlation

Are two players the same?

How do you know and how certain are you?

Data



Are two players the same?

How do you know and how certain are you?

What about two players is different?

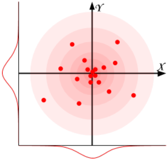
Mean

Variance

Covariance

Correlation

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?

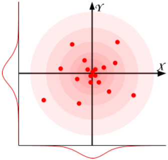
Mean

Variance

Covariance

Correlation

Data



Mean

Variance

Covariance

Correlation

Are two players the same?

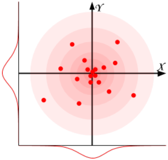
How do you know and how certain are you?

What about two players is different?

How do you quantify which differences matter?

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

Data



Mean

Variance

Covariance

Correlation

Are two players the same?

How do you know and how certain are you?

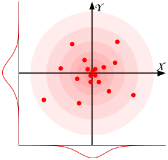
What about two players is different?

How do you quantify which differences matter?

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

What is the best information to ask for?

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?

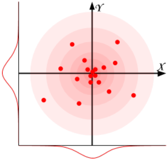
How do you quantify which differences matter?

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

What is the best information to ask for?

What is the best test to run?

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?

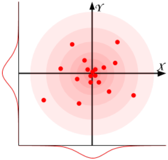
Here's a player, how good will they be?

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?

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?

...

Probability as an experiment

Probability as an experiment

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)

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

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

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 $p(A \cup B)$

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 $p(A \cap B)$

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 $p(A \cap B)$

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

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 $P(x, y)$

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

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

Beating the lottery

Analyzing a test

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.

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)

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)

Analyzing a test

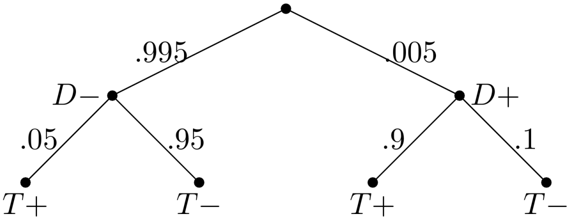
You want to play the lottery, and have a method to win.

0.5% of tickets are winners, and you have a test to verify this.

You are 85% accurate (5% false positives, 10% false negatives)

Is this test useful? How useful? Should you be betting?

Analyzing a test



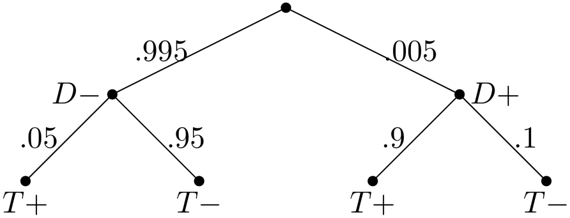
You want to play the lottery, and have a method to win.

0.5% of tickets are winners, and you have a test to verify this.

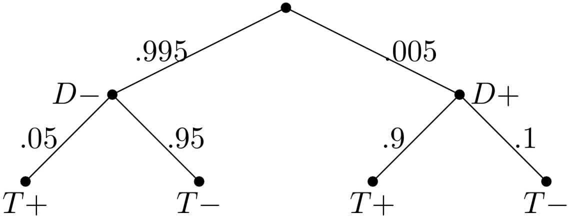
You are 85% accurate (5% false positives, 10% false negatives)

Is this test useful? How useful? Should you be betting?

Is this test useful?

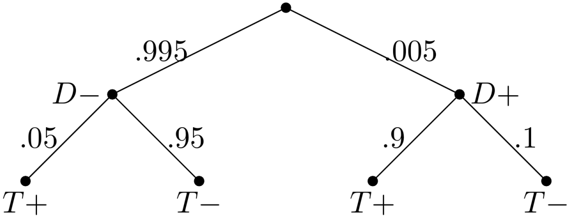


Is this test useful?



$(D-, T+)$ $(D-, T-)$ $(D+, T+)$ $(D+, T-)$

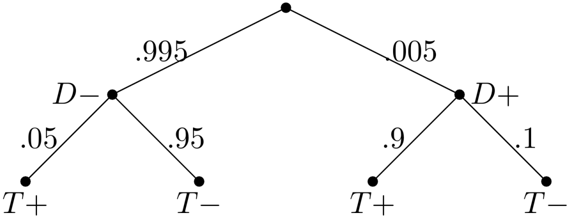
Is this test useful?



$(D-, T+)$ $(D-, T-)$ $(D+, T+)$ $(D+, T-)$

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

Is this test useful?

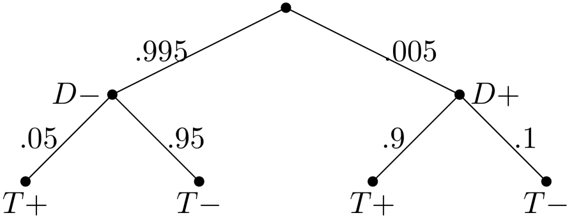


$(D-, T+)$ $(D-, T-)$ $(D+, T+)$ $(D+, T-)$

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

$$\frac{D + \cap T+}{T+}$$

Is this test useful?

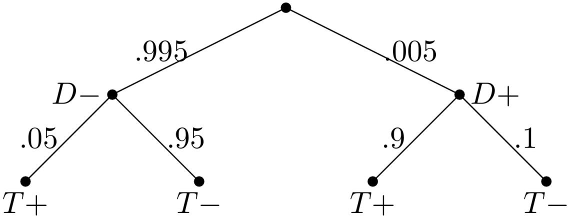


$(D-, T+)$ $(D-, T-)$ $(D+, T+)$ $(D+, T-)$

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?

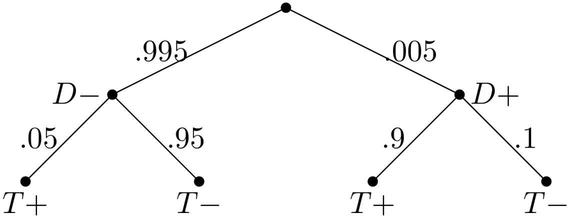


$(D-, T+)$ $(D-, T-)$ $(D+, T+)$ $(D+, T-)$

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?

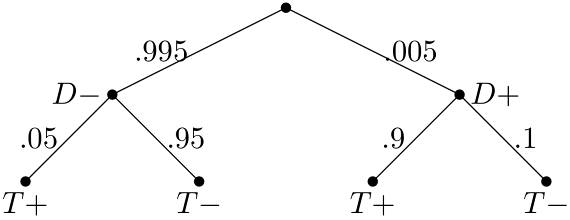


$(D-, T+)$ $(D-, T-)$ $(D+, T+)$ $(D+, T-)$

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

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

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?

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

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \qquad \text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{probability of data}}$$

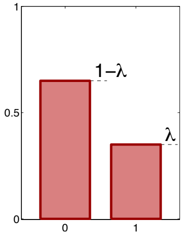
Bernoulli Distribution

- 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 λ and value 0 with failure probability 1 – λ
- **Probability mass function**

$$\left.\begin{array}{l} p(x=0) = 1-\lambda \\ p(x=1) = \lambda \end{array}\right\} \quad p(x) = \lambda^x (1-\lambda)^{1-x}$$

- Notation

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



Parameters
- λ : probability of observing a success

Expectation
- $\mathrm{E}[x] = \lambda$

Variance
- $\mathrm{Var}[x] = \lambda(1-\lambda)$

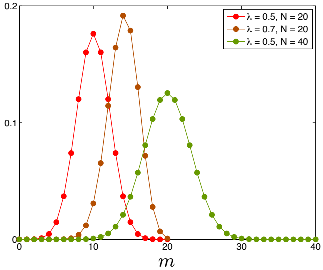
Binomial Distribution

- 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 λ
- **Probability mass function**

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

- Notation

$$\text{Bin}_m(N, \lambda) = \binom{N}{m} \lambda^m (1-\lambda)^{N-m}$$



Parameters
- N : number of trials
- λ : success probability

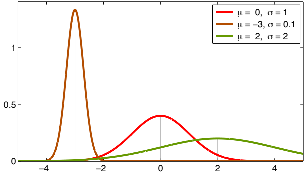
Expectation
- $\text{E}[m] = N\,\lambda$

Variance
- $\text{Var}[m] = N\,\lambda(1-\lambda)$

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

- μ : mean
- σ^2: variance

Expectation

- $\mathrm{E}[x] = \mu$

Variance

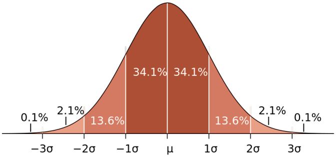
- $\mathrm{Var}[x] = \sigma^2$

Gaussian Distribution

- Notation

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

- Called **standard normal distribution** for μ = 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

- μ : mean
- σ^2: variance

Expectation

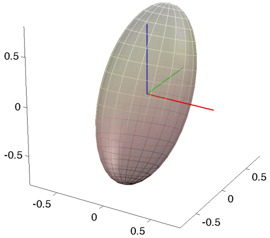
- $\mathrm{E}[x] = \mu$

Variance

- $\mathrm{Var}[x] = \sigma^2$

Multivariate Gaussian Distribution

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



- **Probability density function**

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x}-\boldsymbol{\mu})\right\}$$

- Notation

$$\mathcal{N}_x(\boldsymbol{\mu}, \Sigma) = p(\mathbf{x})$$

Parameters

- $\boldsymbol{\mu}$: mean vector
- Σ: covariance matrix

Expectation

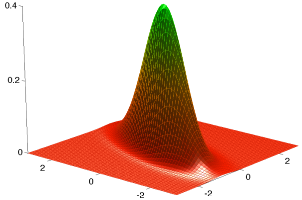
- $\mathrm{E}[\mathbf{x}] = \boldsymbol{\mu}$

Variance

- $\mathrm{Var}[\mathbf{x}] = \Sigma$

Multivariate Gaussian Distribution

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



Parameters

- $\boldsymbol{\mu}$: mean vector
- Σ: covariance matrix

Expectation

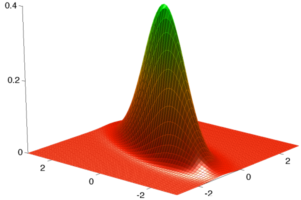
- $\mathrm{E}[\mathbf{x}] = \boldsymbol{\mu}$

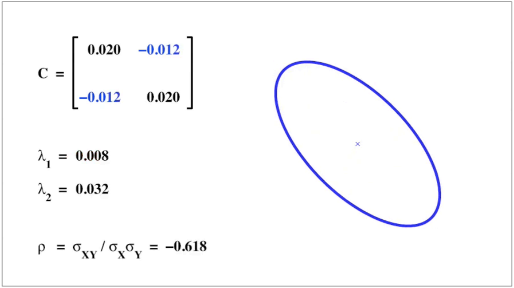
Variance

- $\mathrm{Var}[\mathbf{x}] = \Sigma$

Multivariate Gaussian Distribution

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





Parameters

- $\boldsymbol{\mu}$: mean vector
- Σ: covariance matrix

Expectation

- $\mathrm{E}[\mathbf{x}] = \boldsymbol{\mu}$

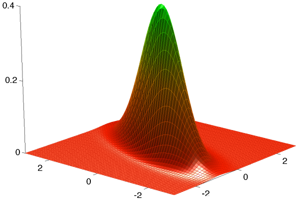
Variance

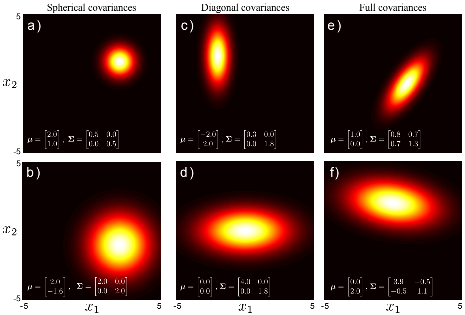
- $\mathrm{Var}[\mathbf{x}] = \Sigma$

Human-Oriented Robotics
Prof. Kai Arras
Social Robotics Lab

Multivariate Gaussian Distribution

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





Source [1]

Parameters

- $\boldsymbol{\mu}$: mean vector
- Σ: covariance matrix

Expectation

- $\mathrm{E}[\mathbf{x}] = \boldsymbol{\mu}$

Variance

- $\mathrm{Var}[\mathbf{x}] = \Sigma$

Common Probability Distributions

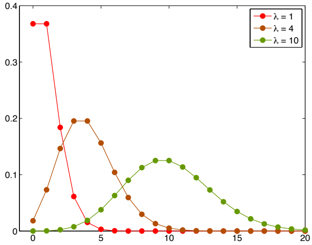
Poisson Distribution

- Consider independent **events** that **happen with an average rate** of λ 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

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



Parameters

- λ : average rate of events over time or space

Expectation

- $\mathrm{E}[x] = \lambda$

Variance

- $\mathrm{Var}[x] = \lambda$

Bayesian updates

Bayesian updates

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)}$$

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.

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

Graphical models

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

Graphical models

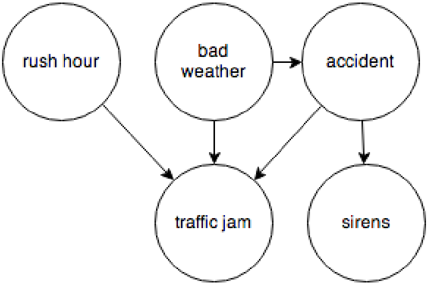
So far we've talked about independence, conditioning, and observation.

A toolkit to discuss these at a higher level of abstraction.

Graphical models

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

A toolkit to discuss these at a higher level of abstraction.

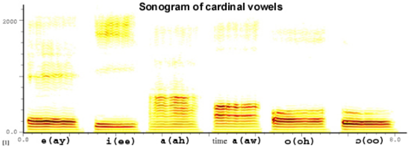


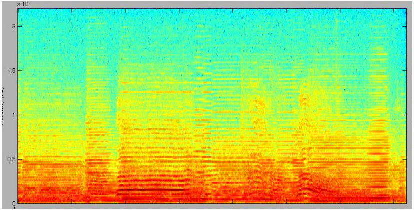
Speech recognition: Naïve bayes

Break down a speech signal into parts.

Speech recognition: Naïve bayes

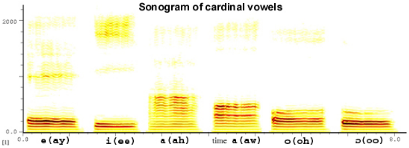
Break down a speech signal into parts.

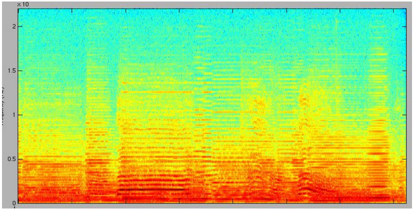




Speech recognition: Naïve bayes

Break down a speech signal into parts.





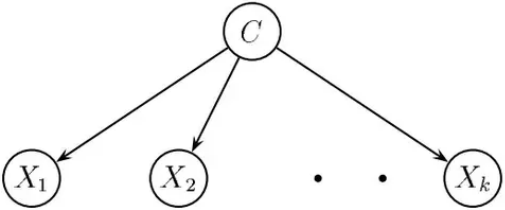
Recover the original speech

Speech recognition: Naïve bayes

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

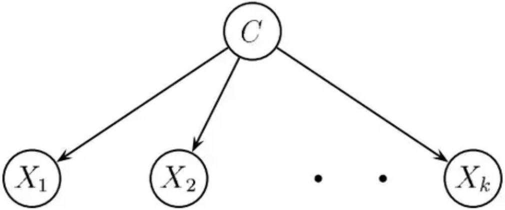
Speech recognition: Naïve bayes

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



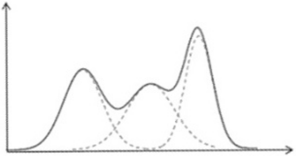
Speech recognition: Naïve bayes

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

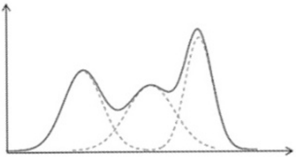


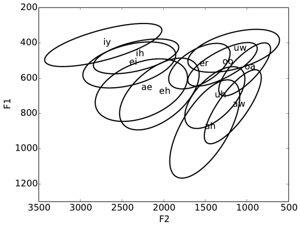
$$P(c|X) \propto \prod_K P(X_k|c)P(c)$$

Speech recognition: Gaussian mixture model

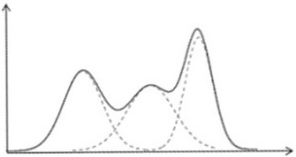


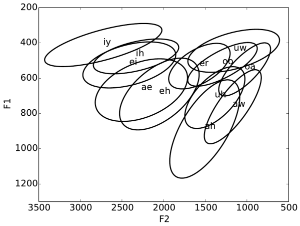
Speech recognition: Gaussian mixture model

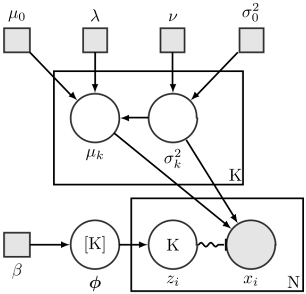




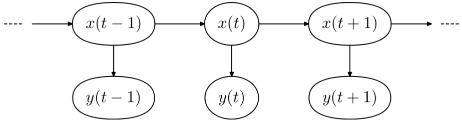
Speech recognition: Gaussian mixture model



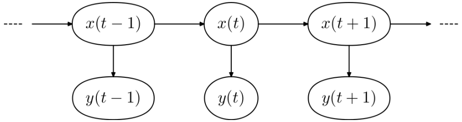


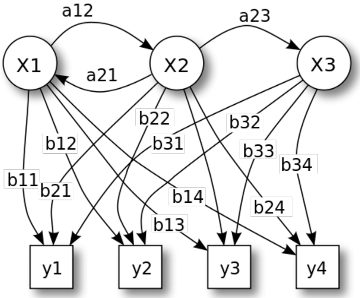


Speech recognition: Hidden Markov model



Speech recognition: Hidden Markov model





Summary

Summary

Probabilities defined in terms of events

Summary

Probabilities defined in terms of events

Random variables and their distributions

Summary

Probabilities defined in terms of events

Random variables and their distributions

Reasoning with probabilities and Bayes' rule

Summary

Probabilities defined in terms of events

Random variables and their distributions

Reasoning with probabilities and Bayes' rule

Updating our knowledge over time

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

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