CSCI 4022 Spring 2021
Clustering and Likelihoods

Example: Suppose we have a set of data that we *know* is Gaussian. What's the best way to estimate the true population mean μ and σ^2?

estimate μ with $\overline{X}$

pop   sample

σ^2 with $S^2 = \dfrac{\sum\limits_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$

Announcements and To-Dos

Announcements:

1. HW 2 posted

Clustering Recap

Two methods so far

Hierarchical

costly

- No initialization
- At each step, we compute *all* distances between pairs of clusters. Then merge the nearest two clusters.
- Once a cluster is formed, it is represented only by its **centroid** (average) or **clustroid** (median/representative point)
- Stop at k clusters.

kmeans

- Random initialization of k clusters
- Loop over points; compute *all* point-to-cluster distances. Then assign each point to its nearest two.
- Loop over clusters; represent each only by its **centroid** (average)
- Stop at "things aren't changing"

Each method required a notion of distance where it was easy to compare dimension-to-dimension. We can relax this!
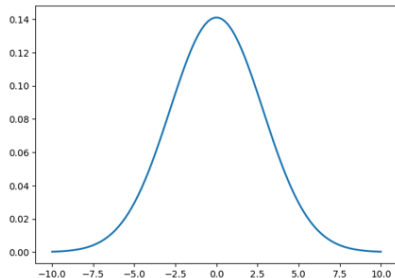
The Gaussian (normal) distribution

You should recall the *normal distribution*m *Gaussian distribution*, or *bell curve*. In one dimension, it's a beautiful little function. (squared negative exponential).

The normal distribution has two arguments or *parameters*:

- μ: The mean or center of the curve. The most important *location*. A single scalar.
- σ^2: The variance/width of the curve. The most common measure of *dispersion* or spread. A single positive scalar.

$$N\left(\mu, \sigma^2\right)$$

A normal

The Gaussian (normal) distribution
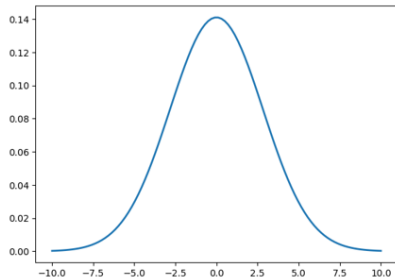**Formalism**

pdf: The *probability density function* of the normal distribution is

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

cdf: The **integral** of the pdf lets us answer probability questions like

$$P(a < X < b) = \int_a^b f(x)\, dx$$

- The normal is known for assigning *low probabilities* to outliers, or points *far from* μ

A normal

The 2-D Gaussian

$M = [240, 266]$

In multiple dimensions, everything gets a bit more complex.

parameters:

μ: The mean or center of the surface. This is a 2-D (x, y) tuple, so we may write $\mu = (\mu_1, \mu_2) \in \mathbb{R}^2$
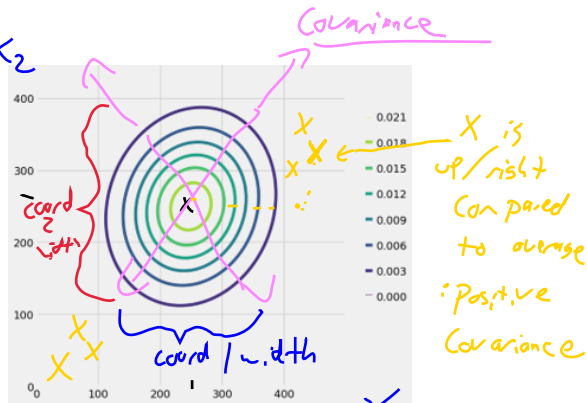
Σ: A 2×2 *covraiance* matrix whose entries are

$$\begin{bmatrix} \sigma_1^2 & cov(x_1, x_2) \\ cov(x_2, x_1) & \sigma_2^2 \end{bmatrix}$$

where σ_1^2 is the variance in the **first axis direction**, σ_2^2 is the variance in the **second axis direction**, and

$$cov(x_1, x_2) = E\left[(X_1 - E[X_1])(X_2 - E[X_2])\right]$$

Covariance

X_2

Coord 2

coord / width

A 2-D normal's contour plot

X is up/right compared to average : Positive Covariance

is the point it avg > or < in the left-right direction

up-down compared to average

X_1

0.021
0.018
0.015
0.012
0.009
0.006
0.003
0.000

The 2-D Gaussian

$3-D: \mu = [m_1, m_2, m_3]$ +3 means

$$\begin{bmatrix} \sigma_1^2 & cov12 & cov13 \\ & \sigma_2^2 & cov2,1 \\ & & \sigma_3^2 \end{bmatrix} = 6 \text{ things}$$

$= 9$

This is actually **5** unique parameters.
parameters:

μ_1: The location of the center in the axis-1 dimension.

μ_2: The location of the center in the axis-2 dimension.
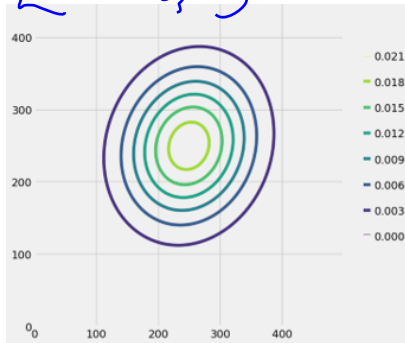
σ_1^2: The variance in the axis-1 direction.

σ_2^2: The variance in the axis-2 direction.

cov: $cov(x_1, x_2)$ The covariance of the data set (dimension one covariance with dimension 2).

Usage: NP.COV(X1, Y1)

A 2-D normal's contour plot

The 2-D Gaussian: variance and covariance

1. The off-diagonal arguments of Σ lead to *rotations*.

   **Results:** if $cov(x_1, x_2) = 0$, then the Gaussian is an ellipse that's oriented **vertically/horizontally** (i.e. the semi-major and semi-minor axis are parallel to the coordinate axes.).

2. σ_1^2 and σ_2^2 determine the widths in their respective coordinate directions

   Results: if $cov(x_1, x_2) = 0$ **and** $\sigma_1 = \sigma_2$, the contours would be circles... like k-means.

   Allowing σ_1 and σ_2 to vary stretches the circle in the corresponding direction. This gives a new method of clustering: **Gaussian Mixture Models**

$$\begin{bmatrix} \sigma_1^2 & cov(x_1, x_2) \\ cov(x_2, x_1) & \sigma_2^2 \end{bmatrix}$$
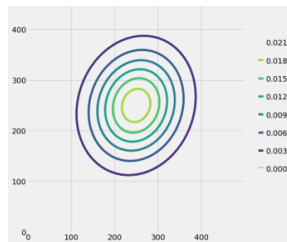
The Gaussian Mixture Model (GMM): 1D Example

Motivating example/cautionary tale: Suppose you go to Chuck E. Cheese's for your niece's birthday party. As you look around, you start to feel rather self-conscious because there seem to be very few people around your age. Needing to pass the time, because you feel so, so awkward, you collect data on everyone's ages.

The Gaussian Mixture Model (GMM): 1D Example

Motivating example/cautionary tale: Suppose you go to Chuck E. Cheese's for your niece's birthday party. As you look around, you start to feel rather self-conscious because there seem to be very few people around your age. Needing to pass the time, because you feel so, so awkward, you collect data on everyone's ages.
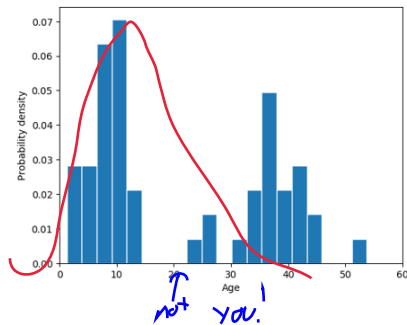
You are then kicked out of and permanently banned form Chuck E. Cheese's because you were accosting children and asking about their ages...

... Fair enough.

Now that your afternoon is freed up, you plot up a histograme of the age data you so creepily and painstakingly collected. It looks like this:
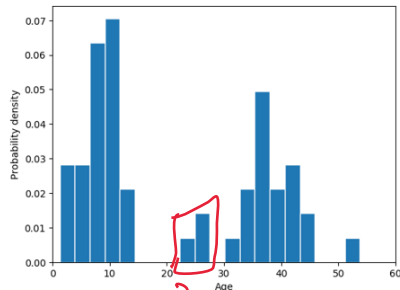
The Gaussian Mixture Model (GMM): 1D Example

The task: How can we best model this distribution?

It's clearly **bimodal**, so a simple normal Gaussian is not appropriate.
Instead, we view it as possibly the combination of **two**
distributions:

1. The **kids'** ages seem like they might be reasonably
   Gaussian, aside from the pesky fact of non-negativity on
   ages.
2. The **parents'** ages might also be modeled by a *different*
   Gaussian distribution
3. This would make the overall distribution of patrons' ages
   a *Gaussian mixture model*

The Gaussian Mixture Model (GMM): 1D Example

The whole model:
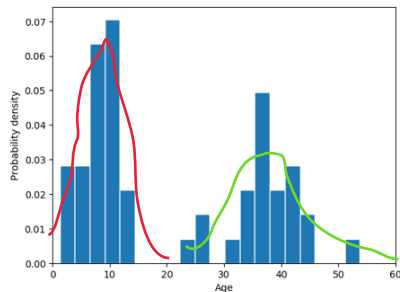**Kids' ages**: a normal $X_1 \sim N(\mu_1, \sigma_1^2)$

Parents' ages: a normal $X_2 \sim N(\mu_2, \sigma_2^2)$

Who's who: a patron could be either of these at some *mixing proportion*. Define the *Bernoulli* random variable Δ where the probability that a patron is an adult ($\Delta = 1$) by π. Then:

Δ has pdf: $P(\Delta = 1) = \pi$

$P(\Delta = 0) = 1 - \pi$

All patrons' ages:

$$X = (1 - \Delta) \cdot X_1 + \Delta \cdot X_2$$

The Gaussian Mixture Model (GMM): 1D Example

The whole model:

$$X = (1 - \Delta) \cdot X_1 + \Delta \cdot X_2$$

Let's unpack, since this is secretly just adding up all the possible ways we can observe a specific age:

$$\underbrace{X}_{\text{Prob of specific age}} = \overbrace{(1 - \Delta)}^{\text{Prob person is adult}} \cdot \underbrace{X_1}_{\text{Prob an "adult" is that age}} + \overbrace{\Delta}^{\text{Prob person is child}} \cdot \underbrace{X_2}_{\text{Prob a "child" is that age}}$$

$$P(adult) \cdot P(age \mid adult)$$
$$= P(age \ AND \ adult).$$

The Gaussian Mixture Model (GMM): 1D Example

The whole model: $X = (1 - \Delta) \cdot X_1 + \Delta \cdot X_2$

Definition: The GMM is a *generative* model, since it specifies the probabilities for new data points.

1. Sample or simulate a Δ with a coin flip or NP.RANDOM.CHOICE

2. Based on Δ, sample a random normal from:

    2.1 X_1 as a $N(\mu_1, \sigma_1^2)$ if $\Delta = 0$ **OR**

    2.2 X_2 as a $N(\mu_2, \sigma_2^2)$ if $\Delta = 1$

Our task is sometimes to *generate*, but first we have to *estimate* the underlying parameters used in the model. To use the model, we have **5** things to estimate or choose.

$$\Theta = \left(\pi, \mu_1, \sigma_1^2, \mu_2, \sigma_2^2 \right)$$

The Gaussian Mixture Model (GMM): 1D Example

**The whole model:** $X = (1 - \Delta) \cdot X_1 + \Delta \cdot X_2$ We need to estimate:

$$\Theta = \left(\pi, \mu_1, \sigma_1^2, \mu_2, \sigma_2^2\right)$$

Assuming we actually *know* all 5 parameters, we can simply write down the full probability density function for our process. If we denote $\phi(x|\mu, \sigma^2)$ as the normal with mean μ and variance σ^2, the model is now

$$f(x|\Theta) = (1 - \pi)\phi(x|\mu_1, \sigma_1^2) + \pi\phi(x|\mu_2, \sigma_2^2)$$

GMM Example: Varying Theta

Here are some pdfs, depending on different choices of the parameter set Θ.

$\mu_1=10, \mu_2=38$ — mean ages

$\sigma_1^2=4, \sigma_2^2=36$ — spreads.

$\pi=0.35$

% adults

GMM Example: Varying Theta

Here are some pdfs, depending on different choices of the parameter set Θ.
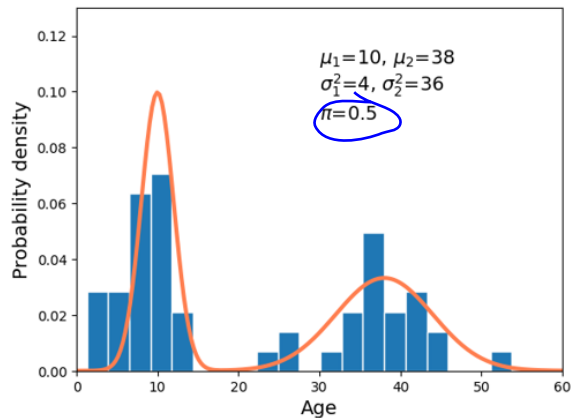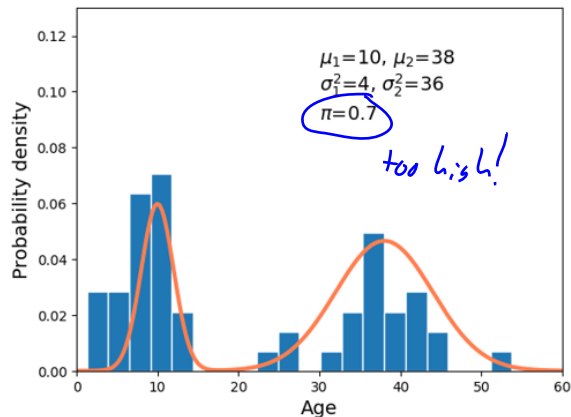
GMM Example: Varying Theta

Here are some pdfs, depending on different choices of the parameter set Θ.

GMM Example: Varying Theta

Here are some pdfs, depending on different choices of the parameter set Θ.

$\mu_1 = 10, \mu_2 = 38$
$\sigma_1^2 = 4, \sigma_2^2 = 36$
$\pi = 0.9$
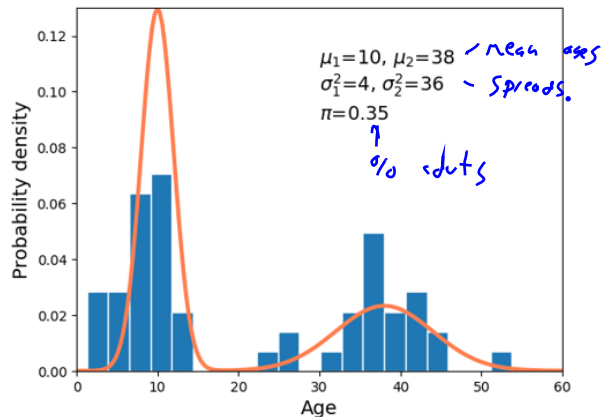
This is too many adults: let's go back to $\pi = .5$...

GMM Example: Varying Theta

Here are some pdfs, depending on different choices of the parameter set Θ.

$\mu_1=10, \mu_2=38$
$\sigma_1^2=7, \sigma_2^2=36$
$\pi=0.5$

Changing σ_1

GMM Example: Varying Theta

Here are some pdfs, depending on different choices of the parameter set Θ.

Changing σ_2

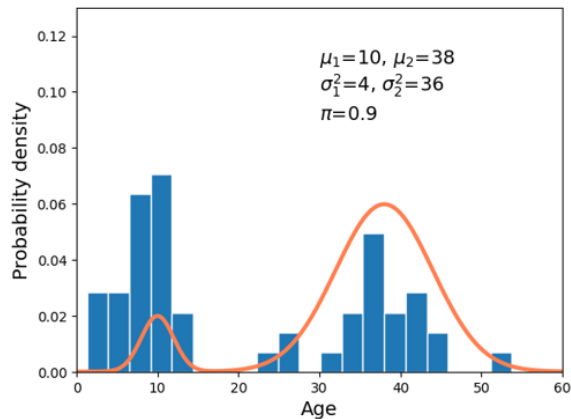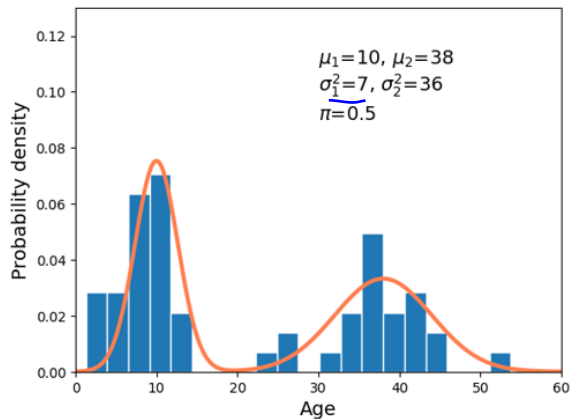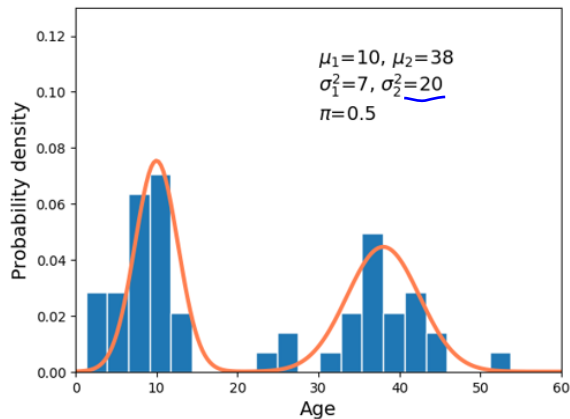# GMM Example: Underview

Curve: population
histogram: data gathered.

The whole model: $X = (1 - \Delta) \cdot X_1 + \Delta \cdot X_2$ We need to estimate:

$$\Theta = \left(\pi, \mu_1, \sigma_1^2, \mu_2, \sigma_2^2\right)$$

1. It would be **tedious** and hand-wavy to manually try to estimate those parameters, even in one dimension.

2. Note that in 2D, each variance was a 2×2 covariance matrix.

   This problem grows in size *quickly*: 2 Gaussians in 2D is now 9 unknowns (4 per Gaussian plus a mixture probability).

Optimization

We're going to zoom out for a bit from Gaussians and talk more generally about *estimating parameters*.

(0,0,4)
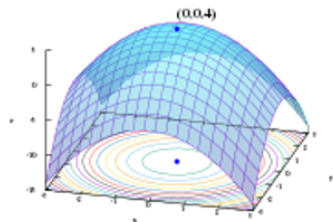
▶ We are often tasked with asking what the "best" values for a set of parameters is.
  ▶ In regression: the β values.
  ▶ In GMMs: the mixture probabilities and component variables
▶ The data scientist has to define **best**!
  ▶ Least Squares?
  ▶ Calculus for local extrema?
  ▶ Minimized Loss? How do we define loss? Can we use cross-validation?
  ▶ Fastest algorithm that's "almost" best?

Probability Models

In the Gaussian mixture model, we're able to write down a **model**. We could answer exact questions like:

1. What are the exact probabilities associated with new data points? For example, suppose I got a new observation. If I knew the *parameters* (Θ), what would be the $P(a < Y < b)$ for a new observation Y? [pdf]; integrate for   prob

2. What is the *joint distribution* of multiple new data points? Assume we get multiple new *independent* actualizations (draws/observations) from the model. They should **all** follow the exact probabilities $P(a < Y_i < b)$!

This is the crux of likelihood theory, which describes a method for estimating terms when we have a probability density function.

pdfs

Definition: The *Probability density function* (pdf) of a random variable X is the function that describes the probability distribution of its outcomes. For (1D) continuous distributions, we add up *intervals* of outcomes on f to get probabilities, which turns into an integral:

$$P(a \le X \le b) = \int_a^b f(x)\,dx$$

tells us the probability of all outcomes from a to b of a continuous RV with pdf $f(x)$.

A couple of examples:

1. *Normal:* $f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$

2. *Exponential:* $f(x; \lambda) = \lambda e^{-\lambda x}; \quad x \ge 0$
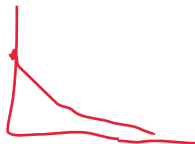
On Parameters

Our goal in the GMM problem is to estimate *parameters* - for the Gaussian that's a mean, variance, and possibly covariances - but it may be other terms for other distributions.

The *Poisson* and *Exponential* distributions were based on using *rates* (λ), and the *binomial* and *geometric* distributions were based on estimating probabilities (p).

In the general notation we denote those parameters as the vector Θ, and think of Θ as the list of "knobs" or "dials" can adjust to change the shape of a distribution in a given model.

The problem with pdfs is that they're backwards of what we want. A pdf gives the "probability of *data* values given *parameter* values." But the data scientist already has data, we need "probability of *parameter* values given *data* values."

Likelihoods

Notationally, we have $f(x|\theta)$: the pdf is the probabilities of data values x if we already know θ. If we want to "flip" those things, Bayes' theorem comes to the rescue!

$$P(\theta|x) = \frac{f(x|\theta)f(\theta)}{f(x)}$$

throw this out:
we have no prob.
assumption on θ.

Definition: The *likelihood function* of random variable with pdf $f(x; \Theta)$ and parameters Θ with a set of observations $\vec{x} = [x_1, x_2, \ldots x_n]$ is the probability of the parameters given the data; $f(\theta|\vec{x})$. If $f(\theta)$ is a constant (an "uninformed prior"), it is *proportional* to the joint density $f(\vec{x}|\theta)$

Idea: There's a lot of theory there and notation here that we won't cover in great detail, but the summary: to estimate parameters for a data set $\vec{x}$ and a model/pdf f, we can ask "which parameter values would have led to the largest value of the pdf at exactly $\vec{x}$?".

Likelihood Example

$$f(x) = \lambda \, e^{-\lambda x} \qquad \lambda > 0.$$

Time to get to the real payoff: how do we estimate parameters if our data is **Gaussian**?
Example: Suppose our *model* is that the data is an exponential random variable. Our goal is to estimate λ. We generate a sample of 4 values, and get $x = [1, 3, 10, 5].$

1. What is $P(\vec{x}|\lambda)$?

"like" $P(\text{bus} \, x = 3)$

$$f(3) = \lambda \, e^{-\lambda \cdot 3}$$

$$f(10) = \lambda \cdot e^{-\lambda 10}$$

$x_1 = 1$ AND
$x_2 = 3$ AND
$x_3 = 10$ AND
$x_4 = 5$

each is ind. identical, exponential

2. How would we "choose λ to maximize this?"

Likelihood Example

Time to get to the real payoff: how do we estimate parameters if our data is **Gaussian**?
Example: Suppose our *model* is that the data is an exponential random variable. Our goal is to estimate λ. We generate a sample of 4 values, and get $x = [1, 3, 10, 5]$.

1. What is $P(\vec{x}|\lambda)$?

   This is secretly an *and* statement: we have 4 *independent* x values.

   $$P(\vec{x}|\lambda) = P(\vec{x} = [x_1, x_2, x_3, x_4]|\lambda) = P(x_1 = 1 \cap x_2 = 3 \cap x_3 = 10 \cap x_4 = 5|\lambda)$$

2. How would we "choose λ to maximize this?"

Likelihood Example

Time to get to the real payoff: how do we estimate parameters if our data is **Gaussian**?
Example: Suppose our *model* is that the data is an exponential random variable. Our goal is to estimate λ. We generate a sample of 4 values, and get $x = [1, 3, 10, 5]$.

1. What is $P(\vec{x}|\lambda)$?

   This is secretly an *and* statement: we have 4 *independent* x values.

   $$P(\vec{x}|\lambda) = P(\vec{x} = [x_1, x_2, x_3, x_4]|\lambda) = P(x_1 = 1 \cap x_2 = 3 \cap x_3 = 10 \cap x_4 = 5|\lambda)$$
   $$= P(x_1 = 1|\lambda) \bullet P(x_2 = 3|\lambda) \bullet P(x_3 = 10|\lambda) \bullet P(x_4 = 5|\lambda)$$
   $$= \lambda e^{-\lambda 1} \cdot \lambda e^{-\lambda 3} \cdot \lambda e^{-\lambda 10} \cdot \lambda e^{-\lambda 5}$$

2. How would we "choose λ to maximize this?"

Likelihood Example

Time to get to the real payoff: how do we estimate parameters if our data is **Gaussian**?
Example: Suppose our *model* is that the data is an exponential random variable. Our goal is to estimate λ. We generate a sample of 4 values, and get $x = [1, 3, 10, 5]$.

1. What is $P(\vec{x}|\lambda)$?

   This is secretly an *and* statement: we have 4 *independent* x values.

$$P(\vec{x}|\lambda) = P(\vec{x} = [x_1, x_2, x_3, x_4]|\lambda) = P(x_1 = 1 \cap x_2 = 3 \cap x_3 = 10 \cap x_4 = 5|\lambda)$$
$$= P(x_1 = 1|\lambda) \cdot P(x_2 = 3|\lambda) \cdot P(x_3 = 10|\lambda) \cdot P(x_4 = 5|\lambda)$$
$$= \lambda e^{-\lambda 1} \cdot \lambda e^{-\lambda 3} \cdot \lambda e^{-\lambda 10} \cdot \lambda e^{-\lambda 5}$$

Find λ that gives max of: $\lambda^4 e^{-\lambda[1+3+10+5]}$.

2. How would we "choose λ to maximize this?" **Hit it with the Calculus stick!**

Likelihood Example

$P(\vec{x}|\lambda) = \lambda e^{-\lambda 1} \cdot \lambda e^{-\lambda 3} \cdot \lambda e^{-\lambda 10} \cdot \lambda e^{-\lambda 5}$

How would we "choose λ to maximize this?"

1. We'd probably collect the λs at the start into λ^4.

2. We'd probably combine exponentials into $e^{-\lambda(1+3+10+5)}$

3. We'd possibly realize that our goal is to pick the λ value that returns **maximum**. But products and exponentials are harder to deal with than sums, so we could take the *logarithm* of our function and find its maximum instead.

$\lambda^4 \cdot e^{-\lambda[1+3+10+5]}$

prod rule $\ddot{x}$.

ℓa_5:

Likelihood Example

$P(\vec{x}|\lambda) = \lambda e^{-\lambda 1} \cdot \lambda e^{-\lambda 3} \cdot \lambda e^{-\lambda 10} \cdot \lambda e^{-\lambda 5}$

How would we "choose λ to maximize this?"

1. We'd probably collect the λs at the start into λ^4.

2. We'd probably combine exponentials into $e^{-\lambda(1+3+10+5)}$

3. We'd possibly realize that our goal is to pick the λ value that returns **maximum**. But products and exponentials are harder to deal with than sums, so we could take the *logarithm* of our function and find its maximum instead.

The **log-likelihood** of this problem is

$$\log\left(\lambda^4 e^{-\lambda(1+3+10+5)}\right) = \log \prod_{i=1}^{4} f(x_i, \lambda)$$

Likelihood Example

The **log-likelihood** of this problem is

$$\log\left(\lambda^4 e^{-\lambda(1+3+10+5)}\right) = \log\prod_{i=1}^{4} f(x_i, \lambda)$$

To maximize with respect to λ, we differentiate and set equal to zero!

Likelihood Example

The **log-likelihood** of this problem is

$$\log\left(\lambda^4 e^{-\lambda(1+3+10+5)}\right) = \log\prod_{i=1}^{4} f(x_i, \lambda)$$

To maximize with respect to λ, we differentiate and set equal to zero!

$$LL(x, \theta) = \log\left(\lambda^4 e^{-\lambda(1+3+10+5)}\right);$$

$$\frac{d}{d\lambda} LL(x, \theta) = \frac{d}{d\lambda}\log\lambda^4 + \frac{d}{d\lambda}\left(-\lambda\left(1+3+10+5\right)\right)$$

$$= \frac{4}{\lambda} - 19; \qquad \text{now set to zero;}$$

$$0 = \frac{4}{\lambda} - 19;$$

$$\lambda = \frac{4}{19}$$

So our estimate of the rate of these things is 4 events per 19 units of time.

The Gaussian Likelihood

Time to get to the real payoff: how do we estimate parameters if our data is **Gaussian**?

The setup to maximum likelihood is always the same: we look at all n of our data points $x_1, x_2, \ldots x_n$ and ask about the probability of $f(x = x_i | \mu, \sigma^2)$ for each one, then multiply them all together as an "and" or *joint* probability. Then we hit it with a logarithm to make maximization easier.

$$f(\vec{x} | \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x_i - \mu)^2}{2\sigma^2}}$$

The Gaussian Likelihood

Time to get to the real payoff: how do we estimate parameters if our data is **Gaussian**?

The setup to maximum likelihood is always the same: we look at all n of our data points $x_1, x_2, \ldots x_n$ and ask about the probability of $f(x = x_i | \mu, \sigma^2)$ for each one, then multiply them all together as an "and" or *joint* probability. Then we hit it with a logarithm to make maximization easier.

$$f(\vec{x} | \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x_i - \mu)^2}{2\sigma^2}}$$

$$LL(\mu, \sigma^2 | x) = \log \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x_i - \mu)^2}{2\sigma^2}}$$

$$= \frac{-n}{2} \log \left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$$

The Gaussian Likelihood

Finally, we have to take two derivatives and set both $\frac{d}{d\mu}LL$ and $\frac{d}{d\sigma^2}LL$ equal to zero.

$$LL(\mu, \sigma^2|x) = \frac{-n}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

$$\frac{d}{d\mu}LL = \frac{d}{d\mu}\left(\frac{-n}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right)$$

The Gaussian Likelihood

Finally, we have to take two derivatives and set both $\frac{d}{d\mu} LL$ and $\frac{d}{d\sigma^2} LL$ equal to zero.

$$LL(\mu, \sigma^2 | x) = \frac{-n}{2} \log \left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$$

$$\frac{d}{d\mu} LL = \frac{d}{d\mu} \left(\frac{-n}{2} \log \left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right)$$

$$= -\frac{1}{2\sigma^2} \sum_{i=1}^{n} -2(x_i - \mu)$$

$$0 = \sum_{i=1}^{n} (x_i - \mu) = \sum_{i=1}^{n} (x_i) - \sum_{i=1}^{n} \mu$$

$$\implies n\mu = \sum_{i=1}^{n} (x_i) \implies \mu = \bar{X}$$

The Gaussian Likelihood

What does that even mean? It means that the "best guess" of the mean μ of the *probability density function* giving rise to our data was the sample mean $\bar{x}$. This is one of the measures as to why the sample mean is a **great measure** of population mean.

There are others, including the central limit theorem and the result that the sample mean is closest point in terms of sums of squared deviations (Euclidean distance) to each other point.

What about σ^2?

The Gaussian Likelihood

What does that even mean? It means that the "best guess" of the mean μ of the *probability density function* giving rise to our data was the sample mean $\bar{x}$. This is one of the measures as to why the sample mean is a **great measure** of population mean.

There are others, including the central limit theorem and the result that the sample mean is closest point in terms of sums of squared deviations (Euclidean distance) to each other point.

What about σ^2?

The **maximum likelihood estimate** (MLEs) for μ, σ^2 of a Gaussian is $\bar{x}, \frac{\sum (x_i - \bar{x})}{n}$... but it turns out MLEs are often slightly worse - especially for small samples - than the estimates you learned in your intro class. They're just a way to arrive at similar results: if you want to estimate the mean and variance of a normal, you use *sample* mean and *sample* variance.

And you can use *sample* covariance for covariances!

Acknowledgments

Some material is adapted/adopted from Mining of Massive Data Sets, by Jure Leskovec, Anand Rajaraman, Jeff Ullman (Stanford University) http://www.mmds.org
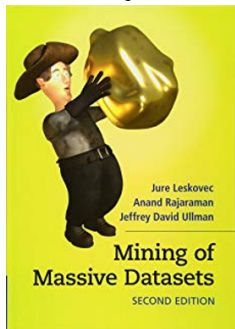
Special thanks to Tony Wong for sharing his original adaptation and adoption of slide material.