Lecture 10

Inference of Parameters and Models
A source emits particles; they decay at locations exponentially distributed with lengthscale $\lambda$. Decays can be observed only if the location falls in a window $(x_{\text{min}}, x_{\text{max}}) = (a, b) = (1, 20)$.

$N$ decays are observed at locations $\{x_1, \ldots, x_N\}$. What is $\lambda$?
How would you have solved this problem before you heard about likelihood functions and Bayes' theorem?

A source emits particles; they decay at locations exponentially distributed with lengthscale $\lambda$. Decays can be observed only if the location falls in a window $(x_{\text{min}}, x_{\text{max}}) = (a, b) = (1, 20)$.

$N$ decays are observed at locations $\{x_1, \ldots, x_N\}$. What is $\lambda$?
PUT DATA IN BINS
sistency between

Had to be chosen
PUT DATA IN BINS

Count

\[ \langle F_b \rangle, F_1, F_2, F_3, \ldots, F_B \]
Goodness of fit 1 = \sum_{b} \left( F_b - \langle F_b \rangle \right)^2

Goodness of fit 2 = \sum_{b} F_b 

\langle F_b \rangle
Goodness of fit \( Q_1 \) = \( \sum \) \( b \) 

Goodness of fit \( Q_2 \) = \( \)
\[ P(x_n | \lambda, a, b) = \begin{cases} e^{-\frac{x_n}{\lambda}} & \text{if } x \in (a, b) \\ 0 & \text{otherwise} \end{cases} \]

\[ Z = \int_a^b e^{-\frac{x}{\lambda}} \, dx \]
\[ P(x | \lambda, a, b) = \frac{e^{-x/\lambda}}{\sum e^{-(a+b)x/\lambda}} \]

\[ NP(x | \lambda, a, b) \Delta x \]

- expected number in bin
\[ \log \langle F \rangle = \log \frac{N}{Z} - \frac{x}{\lambda} \]
Put data in bins, make histogram
Put data in bins, make histogram, and take log

then 'fit a straight line'
(how do we handle zero counts?)
$\alpha$ Had to be Chosen

$\sum_{i=1}^{N}(x_i - a) \frac{1}{N}$

$P(x|\lambda) = e^{-x/\lambda}$

$\langle x \rangle = \lambda$
\[ x \, dx = N(a, b, x) \]

\[ N(a, b, x) \]

\[ \bar{x} = \frac{\sum x_n}{N} \]
another thing
Other ideas

- Chi-squared
  - minimize with respect to lambda
  - also depends on bins

- Other measure of 'error' between theory and data

- Find how the data-mean is expected to vary with lambda

- Other 'estimators', depending on other data-properties
The answer using Bayes's theorem

\[
P(\lambda \mid \{x_n\}_{n=1}^N) = \frac{P(\{x_n\}_{n=1}^N \mid \lambda) P(\lambda)}{P(\{x_n\}_{n=1}^N)}
\]

Or, making explicit the underlying assumptions -

\[
P(\lambda \mid \{x_n\}_{n=1}^N, \mathcal{H}_1) = \frac{P(\{x_n\}_{n=1}^N \mid \lambda, \mathcal{H}_1) P(\lambda \mid \mathcal{H}_1)}{P(\{x_n\}_{n=1}^N \mid \mathcal{H}_1)}
\]

\[
P(\text{Parameters} \mid \text{Data}, \mathcal{H}_1) = \frac{P(\text{Data} \mid \text{Parameters}, \mathcal{H}_1) P(\text{Parameters} \mid \mathcal{H}_1)}{P(\text{Data} \mid \mathcal{H}_1)}
\]

- **POSTERIOR**
- **LIKELIHOOD**
- **PRIOR**
- **EVIDENCE FOR \( \mathcal{H}_1 \)**
\( P(x_n \mid x, H) = \)
\[ Z = \frac{\int_a^b e^{-x/\alpha} \, dx}{\int_0^x e^{-x/\alpha} \, dx} \]
\[ P(x\mid 3 \times 5, \mathcal{H}) = \prod_{i=1}^{N} \]
\[ P(3 \times 5 | \lambda, \mathbf{H}) = \prod_{x} \frac{1}{Z(\lambda, a, b)} e^{-\frac{\lambda}{a} x} \]
\[
P(3 \times 3 | \lambda, H) = \frac{1}{N} \prod_{i=1}^{N} \frac{1}{Z(\lambda, a, b)} e^{-\frac{x_i}{2}} \times P(\chi | H)
\]
The answer using Bayes's theorem

```plaintext
# gnuplot commands to illustrate how simple Bayesian inference is
# Define the density P(x|lambda)
p(x,1) = exp (-x/1) / Z(1)/1
Z(1) = exp (-a/1) - exp(-b/1)
a=1
b=20

# The first plot illustrates the predictive
distributions for different values of lambda
plot [1:20][0:0.25] p(x,2) t "P(x|lambda=2)", \ 
p(x,5) t "P(x|lambda=5)", \ 
p(x,10) t "P(x|lambda=10)"

# Now imagine we receive data \{x\} = \{3, 5, 12\}
# This is the definition of the likelihood function
like(l) = p(3,l) * p(5,l) * p(12,l)

# The next plot shows the three factors, P(x_n|lambda),
in the likelihood function
set logscale x
set xlabel "lambda"
plot [0.3:140] p(3,x) t "P(x=3|lambda)", \ 
p(5,x) t "P(x=5|lambda)", \ 
p(12,x) t "P(x=12|lambda)"

# Finally, plot the likelihood
plot [0.3:140] like(x) t 'P(\{x_n|lambda\})' w 1 5
```
\[ P(x \mid \lambda) \]
The answer using Bayes's theorem

\[
P(x|\lambda=2) \quad P(x|\lambda=5) \quad P(x|\lambda=10)
\]

```
# gnuplot commands to illustrate how simple Bayesian inference is
# Define the density P(x|\lambda)
p(x, l) = \exp \left( -x/l \right) / Z(l)/l
Z(l) = \exp \left( -a/l \right) - \exp \left( -b/l \right)
a=1
b=20

# The first plot illustrates the predictive distributions for different values of lambda
plot [1:20][0:0.25] p(x, 2) t "P(x|\lambda=2)",
    p(x, 5) t "P(x|\lambda=5)",
    p(x, 10) t "P(x|\lambda=10)"

# Now imagine we receive data \{ x \} = \{3, 5, 12\}
# This is the definition of the likelihood function
like(l) = p(3, l) * p(5, l) * p(12, l)

# The next plot shows the three factors, P(x_n|\lambda), in the likelihood function
set logscale x
set xlabel "\lambda"
plot [0.3:140] p(3, x) t "P(x=3|\lambda)",
    p(5, x) t "P(x=5|\lambda)",
    p(12, x) t "P(x=12|\lambda)"

# Finally, plot the likelihood
plot [0.3:140] like(x) t 'P(\{x\}|\lambda)' w 1 5
```

\[
\begin{bmatrix}
\begin{array}{c}
3 \\
5 \\
12
\end{array}
\end{bmatrix}
\]

\[
a=1 \quad b=20
\]
The answer using Bayes's theorem

\[
P(x|\lambda=2) \quad \quad P(x|\lambda=5) \quad \quad P(x|\lambda=10)
\]

The plot illustrates the predictive distributions for different values of \(\lambda\) when we receive data \(\{x\} = \{3, 5, 12\}\).
The answer using Bayes's theorem

The plot illustrates the predictive distributions for different values of \( \lambda \).

Given the data \( \{x\} = \{3, 5, 12\} \), we receive data. The definition of the likelihood function is:

\[
p(3,1) \times p(5,1) \times p(12,1)
\]

The plot shows the three factors, \( P(x | \lambda) \), the likelihood function.

```
plot(x, lambda)
:1401 p(3,x) t "P(x=3|lambda)",
    p(5,x) t "P(x=5|lambda)",
    p(12,x) t "P(x=12|lambda)"

plot the likelihood
:1401 like(x) t 'P(x|lambda)' w 1 5
```

\( a = 1 \) \( \quad b = 20 \)
\[ P(x=3|\lambda) \]
The answer using Bayes's theorem

```
# gnuplot commands to illustrate how simple Bayesian inference is
# Define the density P(x|\lambda)
  p(x,1) = exp(-x/1) / Z(1)/1
  Z(1) = exp(-a/1) - exp(-b/1)
a=1
b=20

# The first plot illustrates the predictive distributions for different values of \lambda
plot [1:20][0:0.25] p(x,2) t "P(x|\lambda=2)", \
    p(x,5) t "P(x|\lambda=5)", \
    p(x,10) t "P(x|\lambda=10)"

# Now imagine we receive data \{ x \} = \{3, 5, 12\}
# This is the definition of the likelihood function
like(1) = p(3,1) * p(5,1) * p(12,1)

# The next plot shows the three factors, P(x|\lambda), in the likelihood function
set logscale x
set xlabel "\lambda"
plot [0.3:140] p(3,x) t "P(x=3|\lambda)", \
    p(5,x) t "P(x=5|\lambda)", \
    p(12,x) t "P(x=12|\lambda)"

# Finally, plot the likelihood
plot [0.3:140] like(x) t 'P(x|\lambda)' w 1 5
```

```
3
5
12
a=1
b=20
```
\log_{a} n > 3
\[ P(\mathcal{X}^3 | \mathcal{H}) \]

\[ \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \]

\[ \times \quad P(\mathcal{X} | \mathcal{H}) \]

\[ \text{approx. likelihood by a Gaussian?} \]
Bayesian answer, alternate data set

\[ \{x\} = \{1.7, 1.5, 2.0\} \]

\[ P(\{x\} | \lambda) \]

\[ \text{set title "\{x\} = \{1.7, 1.5, 2.0\}"} \]

\[ \text{# This is the definition of the likelihood function} \]

\[ \text{like}(l) = p(1.7, l) \times p(1.5, l) \times p(2, l) \]

\[ a = 1 \]  \hspace{1cm}  \[ b = 20 \]
A source emits particles; they decay at locations exponentially distributed with lengthscale $\lambda$. Decays can be observed only if the location falls in a window $(x_{\text{min}}, x_{\text{max}}) = (a, b) = (1, 20)$.

$N$ decays are observed at locations $\{x_1, \ldots, x_N\}$. What is $\lambda$?

+ What about the possibility that the distribution is a mixture of two exponentials?
\[ P(X | \lambda_1, \lambda_2, \pi_1, \pi_2, H_2) \]

\[ = \begin{cases} \pi_1 \frac{e^{-\lambda_1} x^{\lambda_1 - 1}}{\lambda_1} & \text{if } x \in (a, b) \\ \pi_2 \frac{e^{-\lambda_2} x^{\lambda_2 - 1}}{\lambda_2} & \text{otherwise} \end{cases} \]

\[ a = 0 \quad \pi_1 > 0 \quad \pi_1 + \pi_2 = 1 \]

\[ \pi_1 = \frac{1}{2} \]
\[ P(\lambda_1, \lambda_2 \mid 3 \times 5, H) = \frac{P(3 \times 5 \mid \lambda_1, \lambda_2, H_2) P(\lambda_1, \lambda_2 \mid H_2)}{P(3 \times 5 \mid H_2)} \]

Model Comparison:

\[ P(H_1 \mid 3 \times 5, I) = \frac{P(3 \times 5 \mid H_1) P(H_1 \mid I)}{P(3 \times 5 \mid H_2) P(H_2 \mid I)} \]

\[ P(H_2 \mid 3 \times 5, I) = \]

\[ = \]
\[
\frac{1}{Z(\lambda, a, b)} \times P(\lambda | H_1)
\]

\[
e^{-\frac{1}{\lambda} \left( \sum_{i=1}^{n} x_i \right)}
\]
\[ P \left( \sum_{n=1}^{N} 3 k_n s_{n=1} \right) \]
\[ P(3 \times \xi_i \mid \lambda, \lambda, \frac{1}{\lambda}, \sum_{k=1}^{\infty} e^{-\lambda} n^{-1}) \]
maximum likelihood distribution
maximum likelihood distribution
maximum likelihood distribution
maximum likelihood distribution
P(\text{Data} \mid \lambda_1, \lambda_2, k_1..k_N = 00000111)
\[ P(\text{Data} \mid \lambda_1, \lambda_2, k_1..k_N = 11111111) \]
Marginal likelihood of lambda1 and lambda2 - \( P(\text{Data} \mid \text{lambda1, lambda2}) \)
Posterior distribution of the labels

$P(k_1..kN | Data)$

$k_1..kN$
Data set 3 - 8 points

Data = 0.64 0.016 0.100 0.050 3.09 2.14 0.200 7.10

Model 1

optimized lambda = 1.6637
ML_model1 = 5.6254e-06
Evidence1 = 5.4555e-07

Model 2

Evidence2 = 1.3994e-05
ML = 2.5026e-04
tau1ml = 0.080025; tau2ml = 2.8289
k1..kN optimized (optimal labels) = 01110010
EvidenceRatio = 25.651
PosteriorProbOfH2 = 0.96248
PosteriorProbOfH1 = 0.037522
Clustering

K means clustering
Recommended homework

- Noisy channels - Chapters 8, 9, 10 (10.1-10.4 only)
  - Exercises 9.17 (p155); 10.12 (172); 15.12 (235)
  - and (if you want more practice) 15.11, 15.13, 15.15

- Invent a channel to pose to your colleagues:
  - 'what's the capacity of _this_?'

The reading associated with the current lectures is Chapters 3, 21 (especially sec 21.2), and 22 (especially sec 22.1), and 27.

Other recommended exercises are listed on handout 2.

www.inference.phy.cam.ac.uk/itprnn/