Lecture 4: Maximum Likelihood
Review – Common Distributions

**Continuous**
- Uniform
- Normal
- Lognormal
- Beta
- Exponential/Laplace
- Gamma

**Discrete**
- Binomial
- Bernoulli
- Poisson
- Negative Binomial
- Geometric
What are we trying to do?

• “Confronting models with data”
• How is the data modeled?
  – What type of data?
  – What process generated this data?
  – What distributions are an appropriate description of the data?
• How is the process modeled?
• How are the parameters modeled?
Why are we trying to do this?

• Quantify states & relationships
  – What is Y?
  – How is Y related to X?

• Test Hypotheses

• Prediction

• Decision making
How do we do this?
Likelihood

\[ L = P(X = x | \theta) = P(data | model) \]

- Probability of observing a given data point \( x \) conditional on parameter value \( \theta \)
- Likelihood principle: a parameter value is more likely than another if the one for which the data are more probably
Example – Mortality Rate

- Assume mortality rate is constant – $\rho$ – but is an UNKNOWN we want to estimate
- $a_i$ is a KNOWN time of death

$$Pr(a < a_i < a + \Delta a) = Pr\left(die\ now\ given\ that\ right) \cdot Pr\left(plant\ is\ still\ alive\ die\ now\ given\ that\ right) \cdot Pr\left(plant\ is\ still\ alive\ right)$$

$$\approx \rho \Delta a \times e^{-\rho a}$$

$$= \text{Exp}(a|\rho) \Delta a$$
\[ L = Pr(a | \rho) \propto \text{Exp}(a | \rho) \]

- Exact probability is this area.
- Approx probability is shaded box, \( f(a) \, da \).

- Density \( f(a) \).
- Age \( (a) \).
An Observation

- A plant is observed to die on day 10
- From this observation, what is the best estimate for $\rho$?
A few things to note

- A likelihood surface is NOT a PDF
- $Pr(X \mid \theta) \neq Pr(\theta \mid X)$
- Does not integrate to 1
- No, you can't just normalize it
- The model parameter is being varied, not the random variable
  - i.e. the x-axis is fixed, not random
- Cannot interpret surface in terms of it's mean, variance, quantiles
Maximum Likelihood

• Step 1: Write a likelihood function describing the likelihood of the observation
• Step 2: Find the value of the model parameter that maximized the likelihood

\[
\frac{dL}{d\rho} = 0
\]
\[ L = \rho e^{-\rho a} \]

\[ \ln L = \ln \rho - \rho a \]

\[ \frac{\partial \ln L}{\partial \rho} = \frac{1}{\rho} - a = 0 \]

\[ \rho_{ML} = \frac{1}{a} = 0.1 \text{ day}^{-1} \]
\[
\begin{align*}
L &= \rho e^{-\rho a} \\
\ln L &= \ln \rho - \rho a \\
\frac{\partial \ln L}{\partial \rho} &= \frac{1}{\rho} - a = 0 \\
\rho_{ML} &= \frac{1}{a} = 0.1 \text{ day}^{-1}
\end{align*}
\]
\[
\begin{align*}
\ln L &= \ln \rho - \rho a \\
\frac{\partial \ln L}{\partial \rho} &= \frac{1}{\rho} - a = 0 \\
\rho_{ML} &= \frac{1}{a} = 0.1 \text{ day}^{-1}
\end{align*}
\]
\[ L = \rho e^{-\rho a} \]
\[ \ln L = \ln \rho - \rho a \]
\[ \frac{\partial \ln L}{\partial \rho} = \frac{1}{\rho} - a = 0 \]
\[ \rho_{ML} = \frac{1}{a} = 0.1 \text{ day}^{-1} \]
\[ \frac{1}{\rho} = a \]
\[ 1 = \rho a \]
\[ \frac{1}{a} = \rho \]
a = 10
A second data point

• Suppose a second plant dies at day 14
• Step 1: Define the likelihood

\[ L = Pr(a_1, a_2 | \rho) \]

\[ = Pr(a_2 | a_1, \rho) Pr(a_1 | \rho) \]

\[ = Pr(a_2 | \rho) Pr(a_1 | \rho) \]

\[ = \text{Exp}(a_2 | \rho) \text{Exp}(a_1 | \rho) \]

Assume measurements are independent
• Step 2: Find the maximum

\[ L = \rho e^{-\rho a_1} \cdot \rho e^{-\rho a_2} \]

\[ \ln L = 2 \ln \rho - \rho a_1 - \rho a_2 \]

\[ \frac{\partial \ln L}{\partial \rho} = \frac{2}{\rho} - (a_1 + a_2) = 0 \]

\[ \rho_{ML} = \frac{2}{a_1 + a_2} = 0.0833 \text{ day}^{-1} \]
A whole data set

- Step 1: Define Likelihood

\[ L = \text{Pr}(a_1, a_2, \cdots, a_n \mid \rho) \]

Assume measurements are independent

\[ = \prod_{i=1}^{n} \text{Pr}(a_i \mid \rho) \]

\[ = \prod_{i=1}^{n} \text{Exp}(a_i \mid \rho) \]
Step 2:
Find the maximum

\[ L = \prod_{i=1}^{n} \rho e^{-\rho a_i} \]

\[ \ln L = \sum_{i=1}^{n} \left( \ln \rho - \rho a_i \right) \]
\[ = n \ln \rho - \rho \sum_{i=1}^{n} a_i \]

\[ \frac{\partial \ln L}{\partial \rho} = \frac{n}{\rho} - \sum_{i=1}^{n} a_i = 0 \]

\[ \rho_{ML} = \frac{n}{\sum_{i=1}^{n} a_i} = 1/\bar{a} \]
FIGURE 3.2. Likelihood functions for the exponential model with three different sample sizes. Note the different scales on the vertical axes.