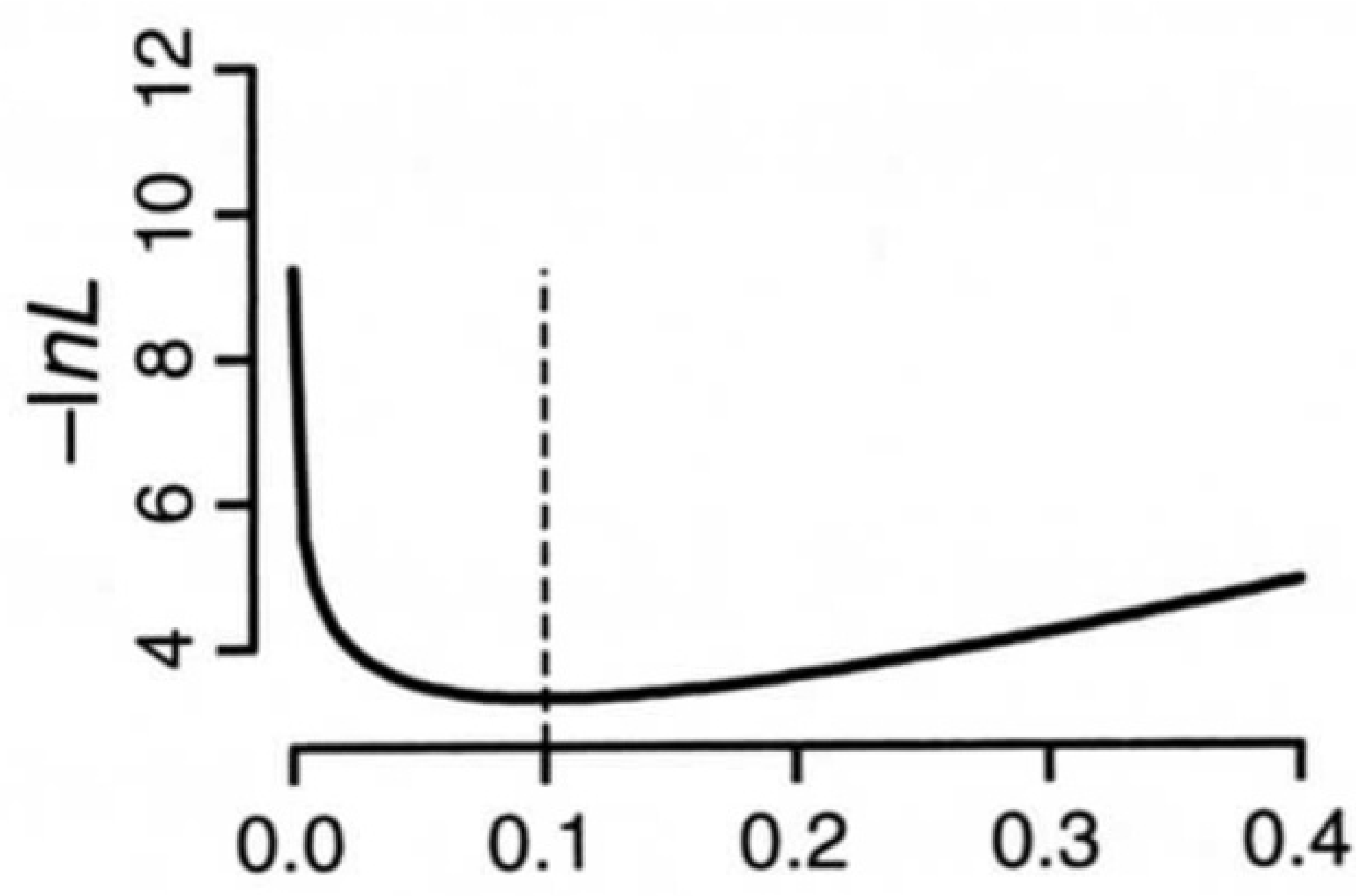
Lecture 4: Maximum Likelihood



Review – Common Distributions

Continuous

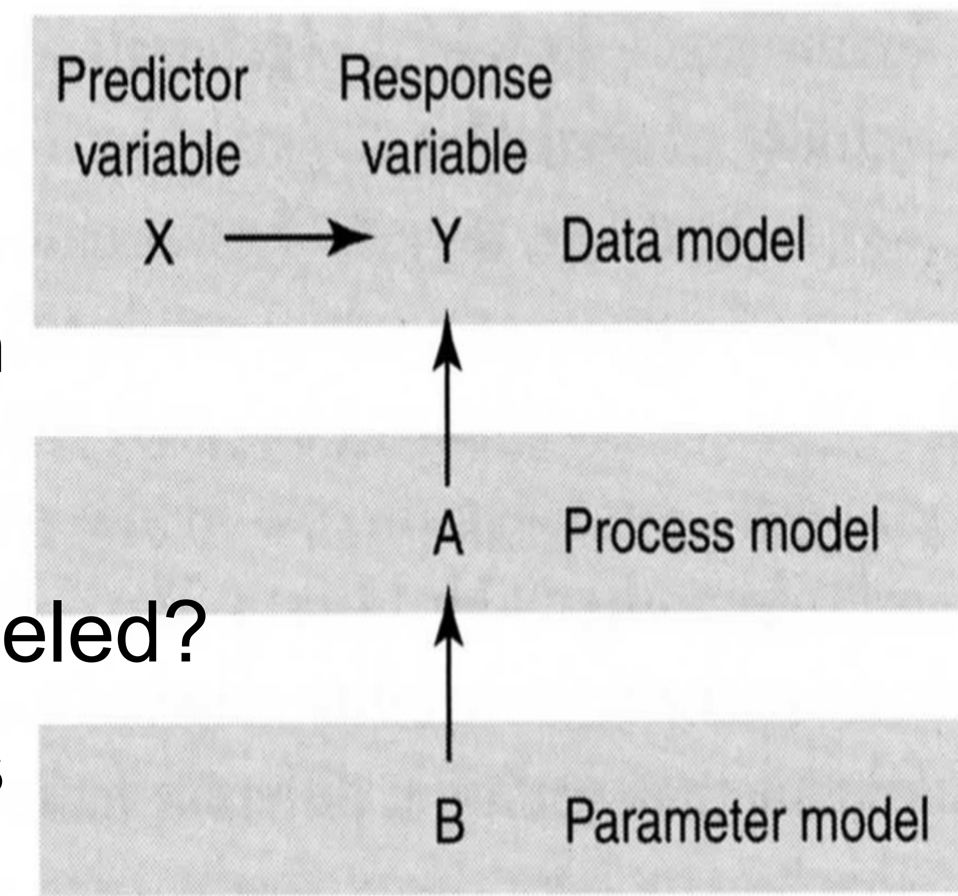
- Uniform
- Normal
- Lognormal
- Beta
- Exponental/Laplace
- Gamma

Discrete

- Binomial
- Bernoulli
- Poisson
- Negative Binomial
- Geometric

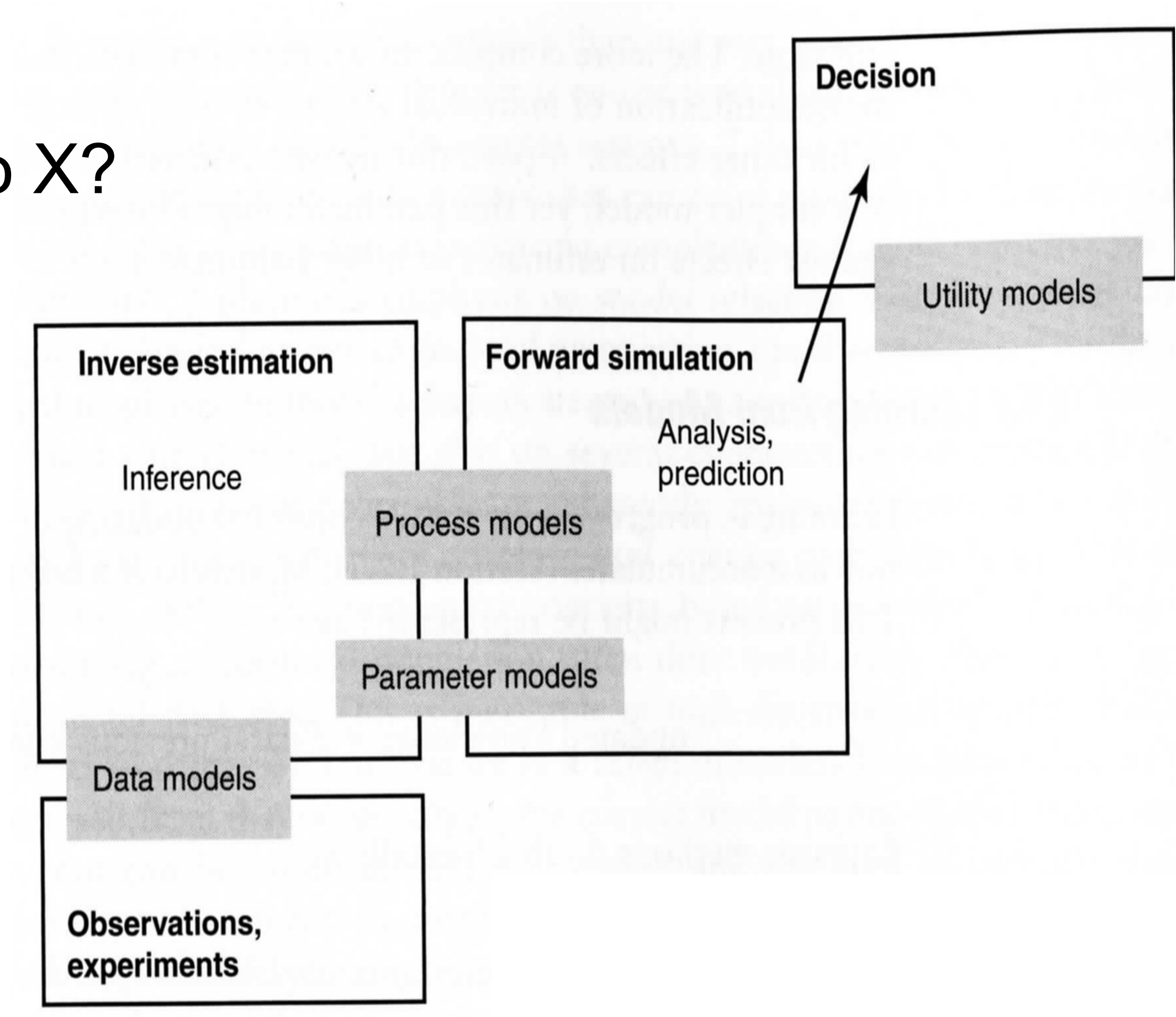
What are we trying to do?

- "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?

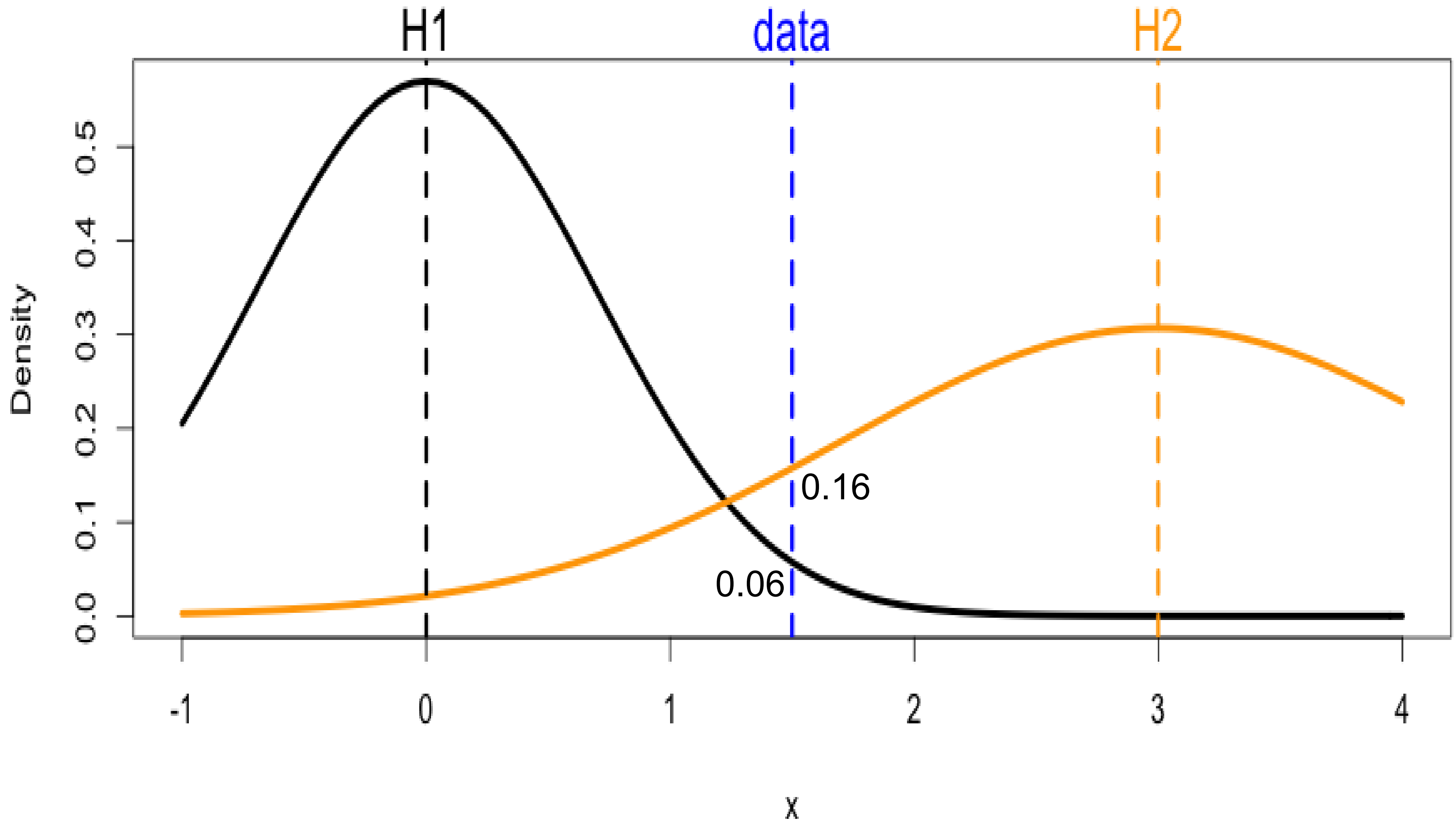


Why are we trying to do this?

- Quantify states & relationships
  - What is Y?
  - How is Y related to X?
- Test Hypotheses
- Prediction
- Decision making



How do we do this?


H1
data
H2
0.16
0.06
Density
0.0
0.1
0.2
0.3
0.4
0.5
-1
0
1
2
3
4
x

Likelihood

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

- Probability of observing a given data point x conditional on parameter value θ
- Likelihood principle: a parameter value is more likely than another if it is the one for which the data are more probable

Maximum Likelihood

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

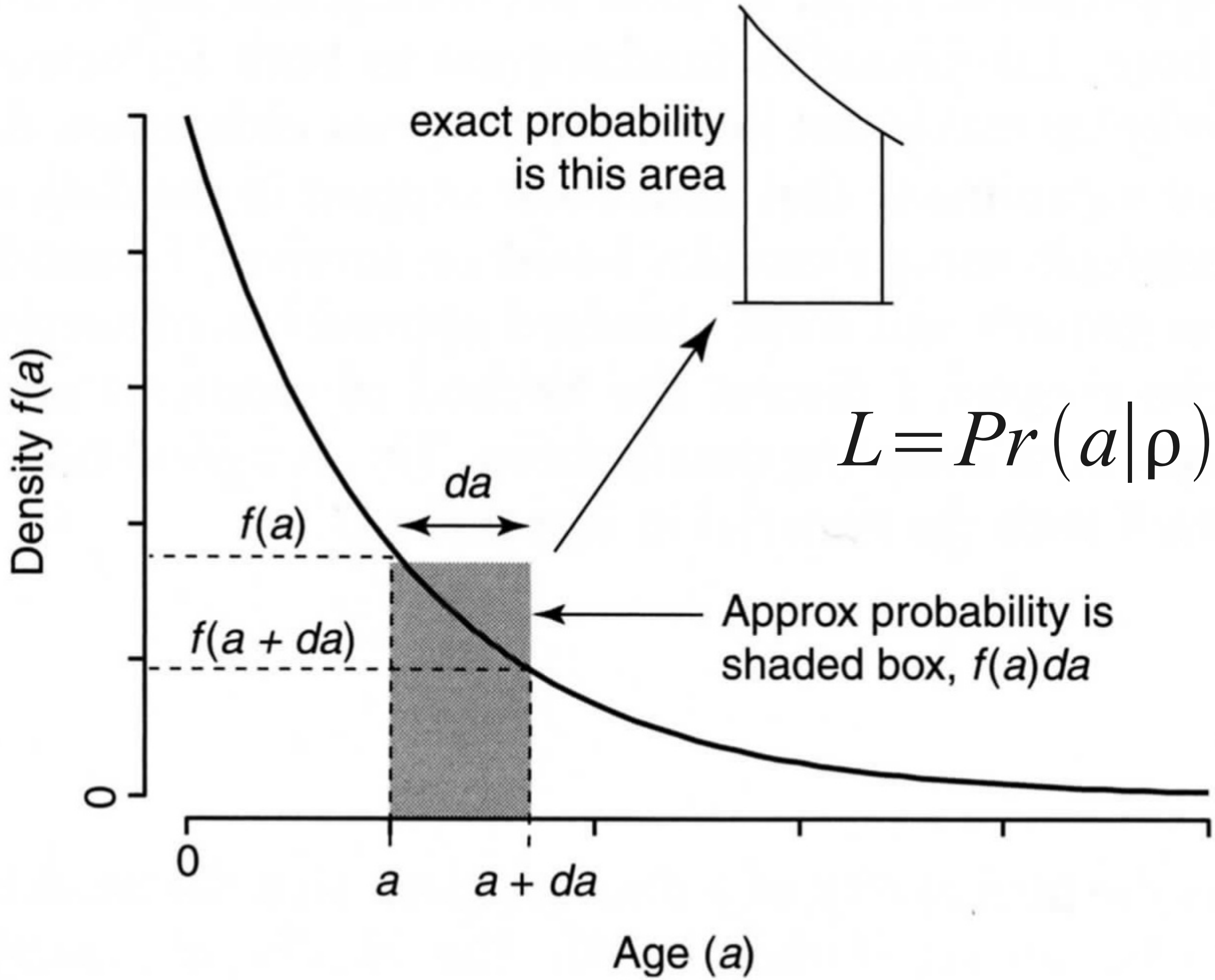
Goal: Find the θ that maximizes L

- Step 1: Construct Likelihood
- Step 2: Maximize function
  - Take Log of likelihood function
  - Take derivative of function
  - Set derivative = 0
  - Solve for parameter

Example – Mortality Rate

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

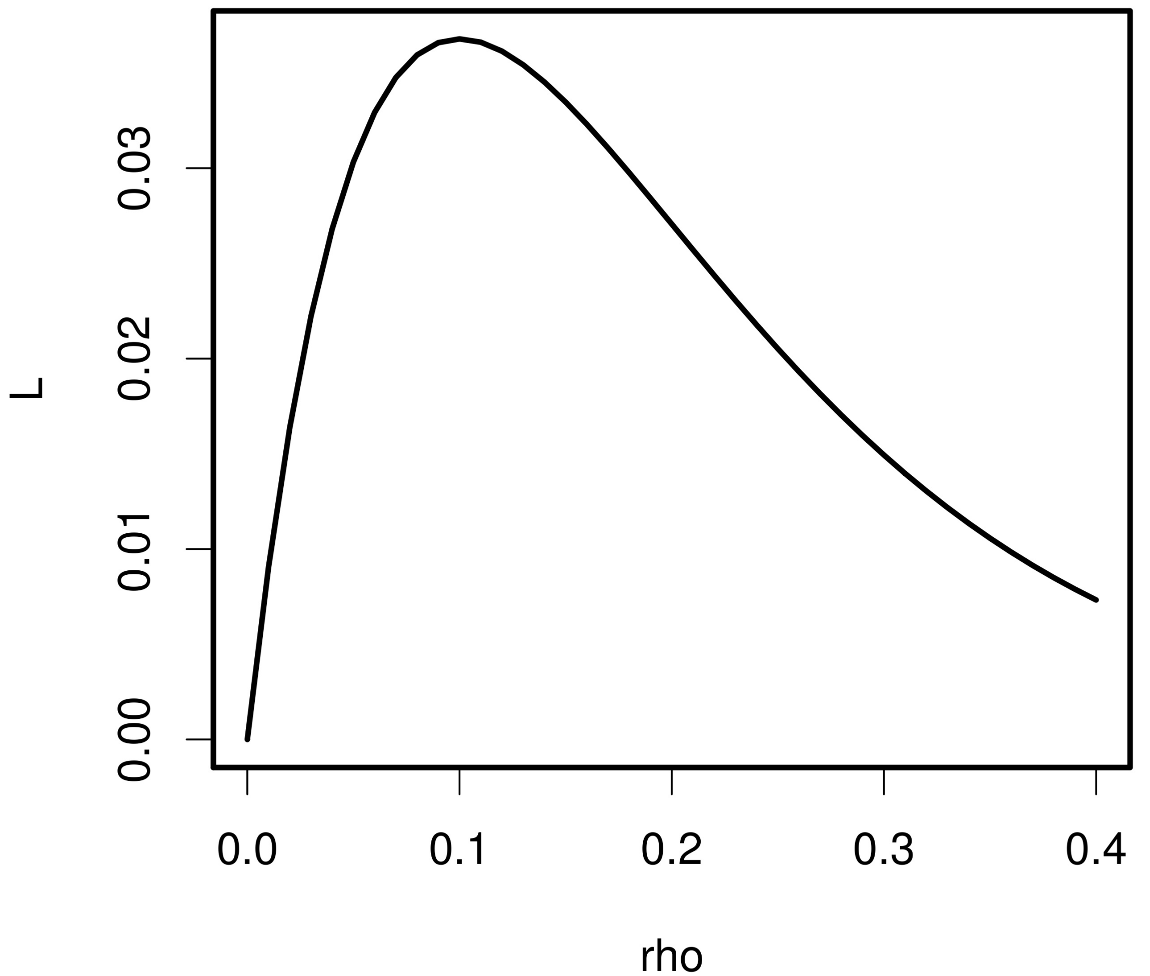
$$
\begin{aligned}
Pr(a < a_i < a + \Delta a) &= Pr\left(\begin{matrix}\text{die now given that}\\ \text{plant is still alive}\end{matrix}\right) \cdot Pr\left(\begin{matrix}\text{plant is}\\ \text{still alive}\end{matrix}\right) \\
&\approx \rho \Delta a \times e^{-\rho a} \\
&= \mathrm{Exp}(a|\rho) \Delta a
\end{aligned}
$$



$$L = Pr(a|\rho) \propto \mathrm{Exp}(a|\rho)$$

An Observation

- A plant is observed to die on day 10
- From this observation, what is the best estimate for ρ?



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
- <u>Cannot</u> 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

$$\frac{dL}{d\rho} = 0$$

$$
\begin{aligned}
L &= \rho\, e^{-\rho a} \\
\ln L &= \ln \rho - \rho a \\
\frac{\partial \ln L}{\partial \rho} &= \frac{1}{\rho} - a = 0 \\
\rho_{ML} &= \frac{1}{a} = 0.1\, day^{-1}
\end{aligned}
$$

$$L = \rho e^{-\rho a}$$

$$\ln L = \ln \rho - \rho a$$

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

$$\rho_{ML} = \frac{1}{a} = 0.1\, day^{-1}$$

$$\ln\left(\rho e^{-\rho a}\right)$$

$$\ln(\rho) + \ln\left(e^{-\rho a}\right)$$

$$\ln(\rho) - \rho a$$

$$L = \rho e^{-\rho a}$$

$$\ln L = \ln \rho - \rho a$$

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

$$\rho_{ML} = \frac{1}{a} = 0.1\, day^{-1}$$

$$\frac{\partial [\ln(\rho) - \rho a]}{\partial \rho}$$

$$\frac{\partial \ln(\rho)}{\partial \rho} - \frac{\partial \rho a}{\partial \rho}$$

$$\frac{1}{\rho} - a$$

$$L = \rho e^{-\rho a}$$

$$\ln L = \ln \rho - \rho a$$

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

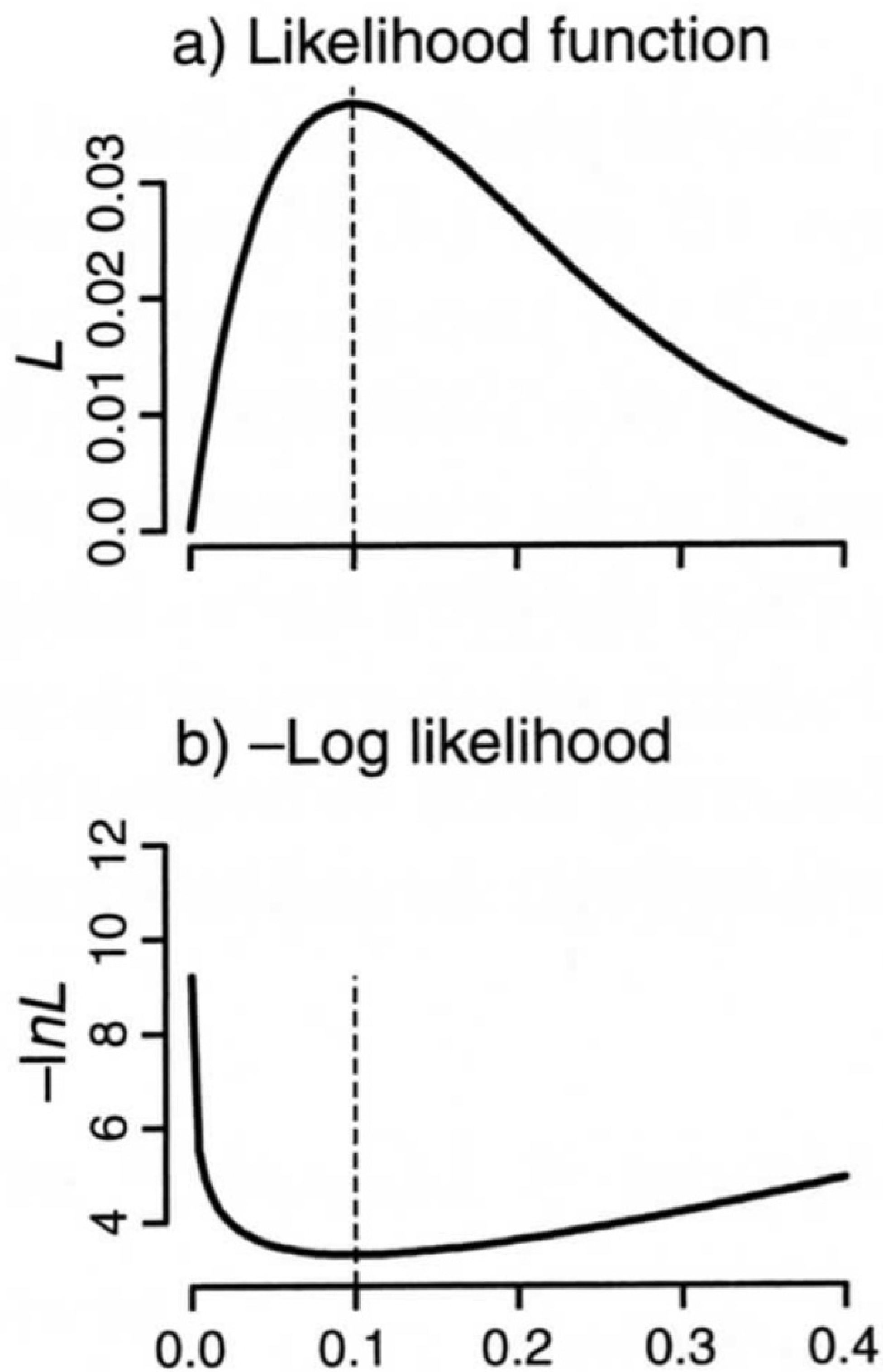
$$\frac{1}{\rho} = a$$

$$1 = \rho a$$

$$\frac{1}{a} = \rho$$

$$\rho_{ML} = \frac{1}{a} = 0.1\, day^{-1}$$

a = 10



a) Likelihood function
L
0.0
0.01
0.02
0.03
b) –Log likelihood
–lnL
4
6
8
10
12
0.0
0.1
0.2
0.3
0.4

A second data point

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

$$
\begin{aligned}
L &= Pr(a_{1,} a_2 | \rho) \\
&= Pr(a_2 | a_{1,} \rho) Pr(a_1 | \rho) \\
&= Pr(a_2 | \rho) Pr(a_1 | \rho) \\
&\propto \mathrm{Exp}(a_2 | \rho) \mathrm{Exp}(a_1 | \rho)
\end{aligned}
$$

Assume measurements are independent

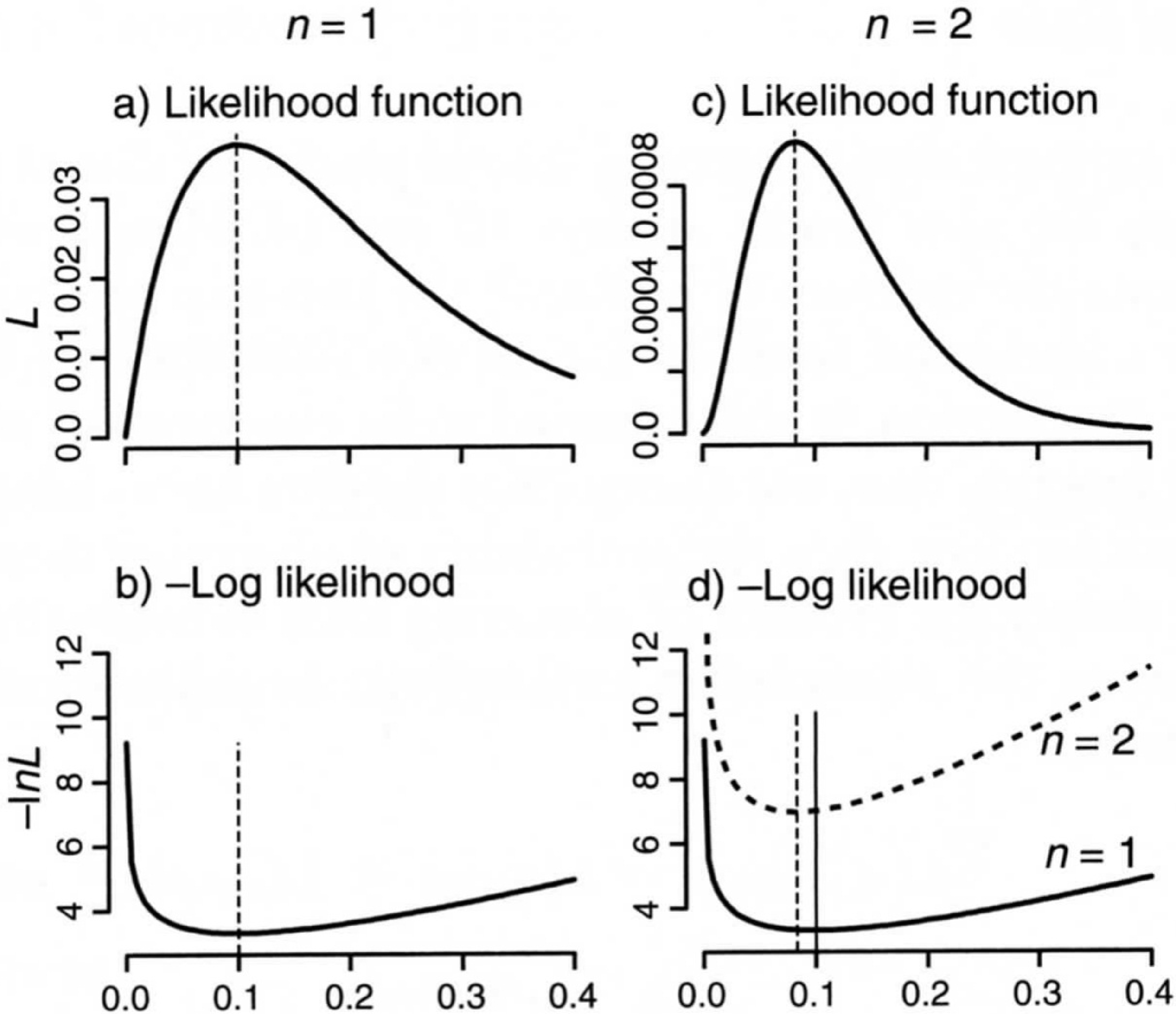
- Step 2: Find the maximum

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

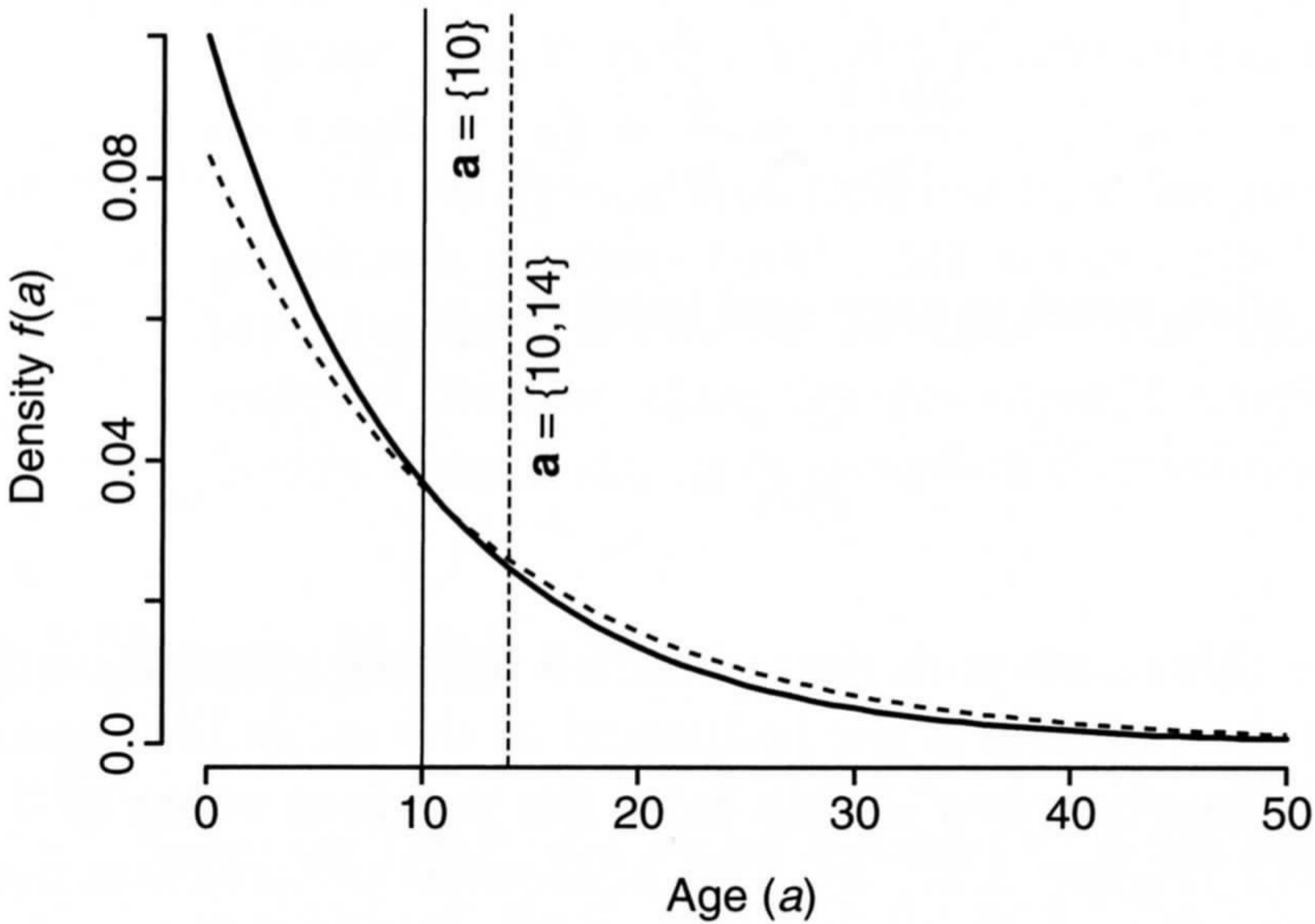
$$\ln L = 2 \ln \rho - \rho a_1 - \rho a_2$$

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

$$\rho_{ML} = \frac{2}{a_1 + a_2} = 0.0833\, day^{-1}$$



n = 1
n = 2
a) Likelihood function
L
0.0 0.01 0.02 0.03
c) Likelihood function
0.0 0.0004 0.0008
b) –Log likelihood
–lnL
4 6 8 10 12
0.0 0.1 0.2 0.3 0.4
d) –Log likelihood
n = 2
n = 1
4 6 8 10 12
0.0 0.1 0.2 0.3 0.4



Density f(a)
0.0
0.04
0.08
a = {10}
a = {10,14}
0
10
20
30
40
50
Age (a)

A whole data set

- Step 1: Define Likelihood

$$L = Pr(a_{1,} a_{2,} \cdots, a_n | \rho)$$

$$= \prod_{i=1}^{n} Pr(a_i | \rho)$$

Assume measurements are independent

$$= \prod_{i=1}^{n} \mathrm{Exp}(a_i | \rho)$$

- Step 2:
  Find the maximum

$$L = \prod_{i=1}^{n} \rho\, e^{-\rho\, a_i}$$

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

$$= n \ln \rho - \rho \sum_{i=1}^{n} a_i$$

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

$$\rho_{ML} = \frac{n}{\sum_{i=1}^{n} a_i} = 1/\bar{a}$$

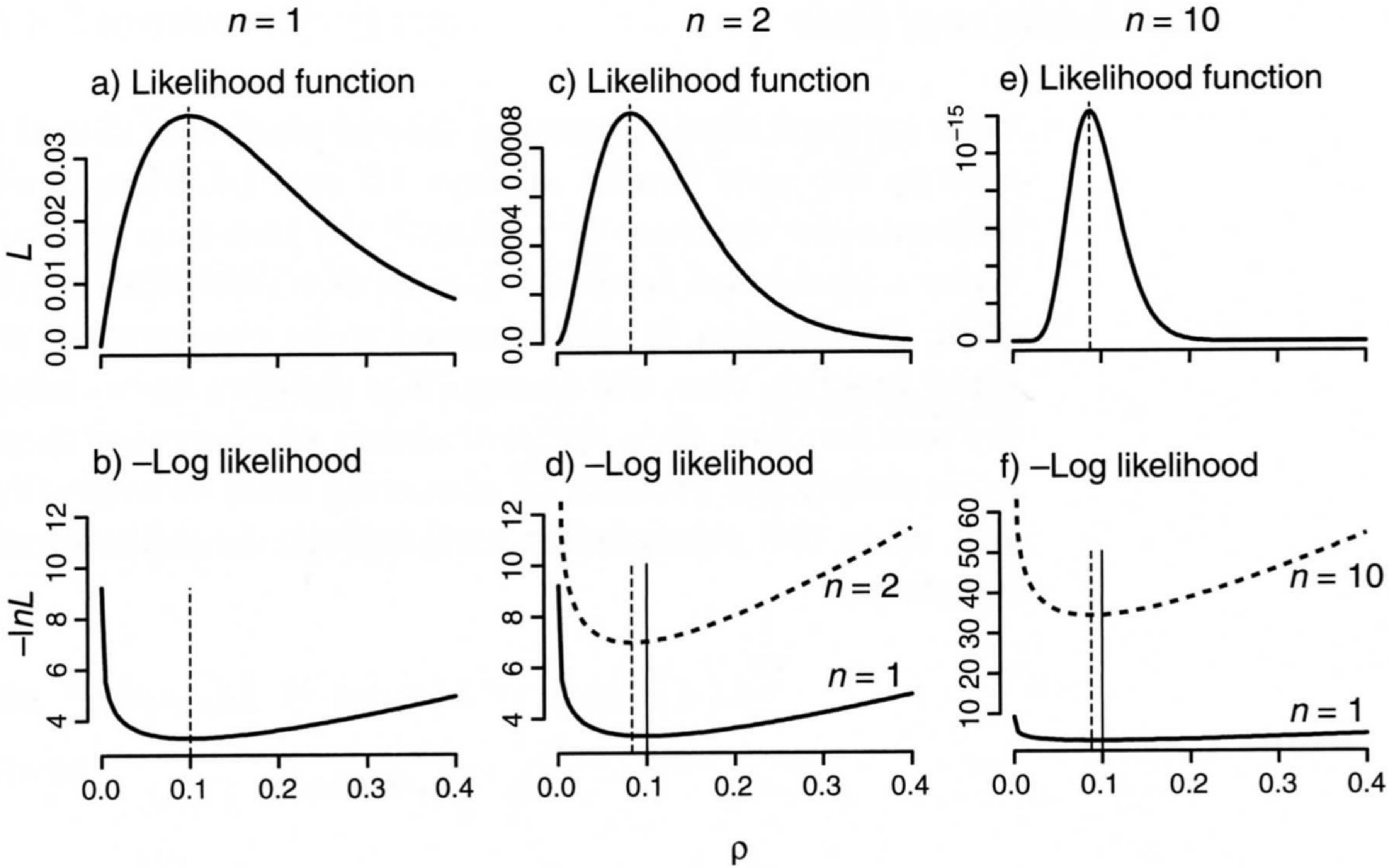


FIGURE 3.2. Likelihood functions for the exponential model with three different sample sizes. Note the different scales on the vertical axes.