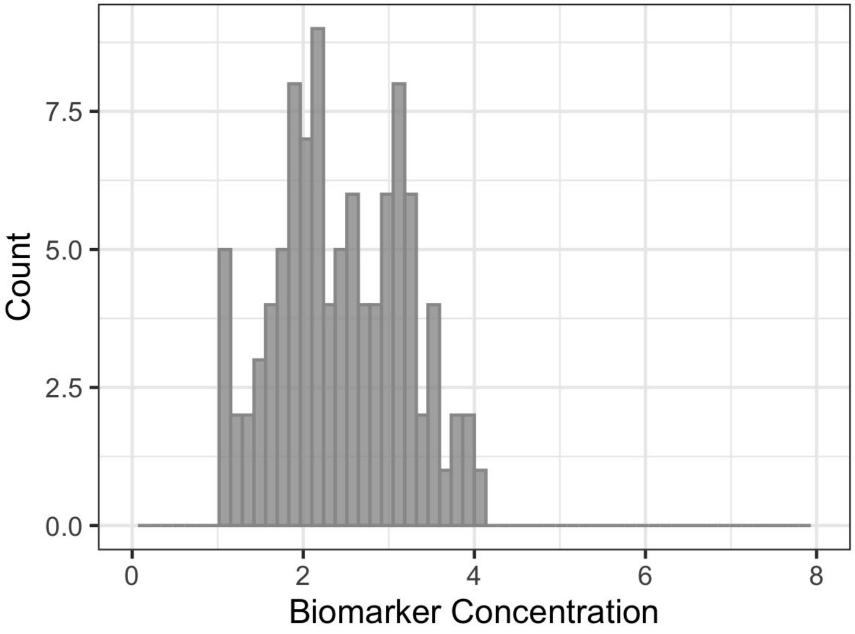
Chapter 5: The Basics of Maximum Likelihood Estimation

Modern Clinical Data Science
Chapter Guides
Bethany Percha, Instructor

Icahn School of Medicine at Mount Sinai

How to Use this Guide

- Read the corresponding notes chapter first
- Try to answer the discussion questions on your own
- Listen to the chapter guide (should be 15 min, max) while following along in the notes

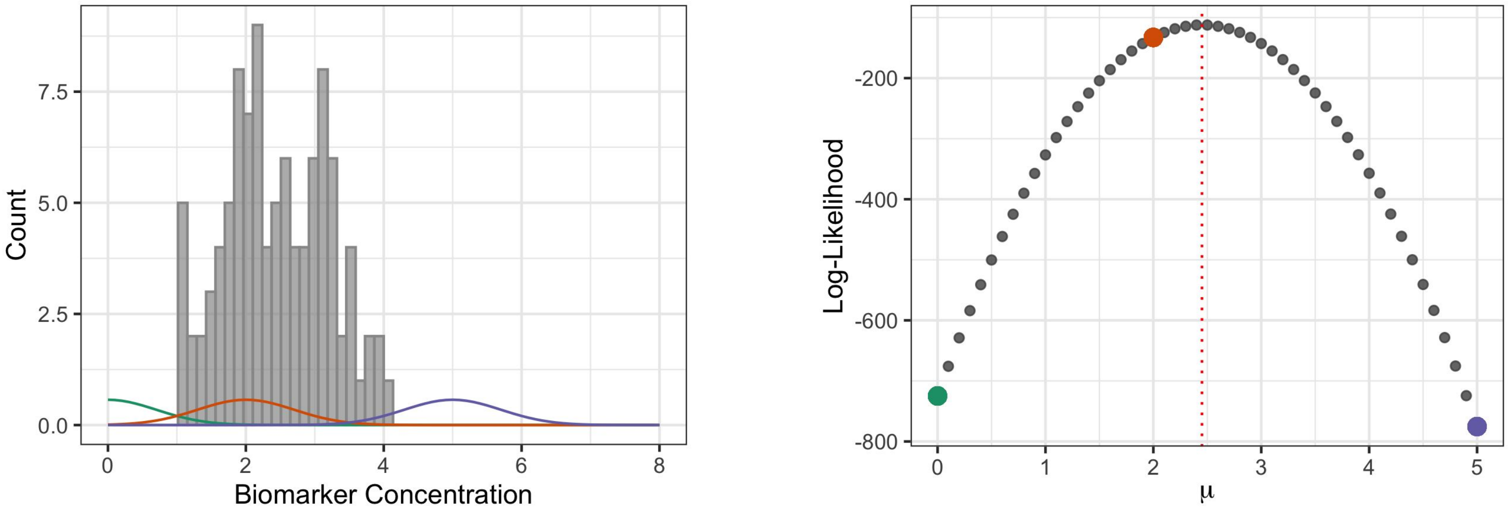


7.5
5.0
2.5
0.0
Count
0
2
4
6
8
Biomarker Concentration

Question 5.1

What are some reasons why we might want to fit data to a probability distribution?

- Data summarization (don't need to keep entire dataset around)
- Sampling - ability to simulate data that look like ours
- Many models and hypothesis tests rely on assumptions about underlying probability distributions
- Detecting outliers or faulty assumptions



Count
7.5
5.0
2.5
0.0
0
2
4
6
8
Biomarker Concentration
Log-Likelihood
-200
-400
-600
-800
0
1
2
3
4
5
μ

i	$x^{(i)}$
1	1
2	0
3	0
4	0
5	1
6	1
7	1
8	0
9	0
10	0

$$\begin{aligned}\log \mathcal{L}(\mu) &= \sum_{i=1}^{n} \log p(x^{(i)}|\mu) \\ &= \sum_{i=1}^{n} \log\left(\mu^{x^{(i)}}(1-\mu)^{1-x^{(i)}}\right) \\ &= \sum_{i=1}^{n}\left[x^{(i)}\log(\mu) + (1-x^{(i)})\log(1-\mu)\right]\end{aligned}$$

$$\frac{d}{d\mu}\log\mathcal{L}(\mu) = \sum_{i=1}^{n}\left[\frac{x^{(i)}}{\mu} - \frac{1-x^{(i)}}{1-\mu}\right]$$

$$\begin{aligned}\sum_{i=1}^{n}\left[\frac{x^{(i)}}{\hat{\mu}} - \frac{1-x^{(i)}}{1-\hat{\mu}}\right] = 0 \implies (1-\hat{\mu})\sum_{i=1}^{n}x^{(i)} &= \hat{\mu}\sum_{i=1}^{n}(1-x^{(i)}) \\ \implies \boxed{\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}x^{(i)}}\end{aligned}$$

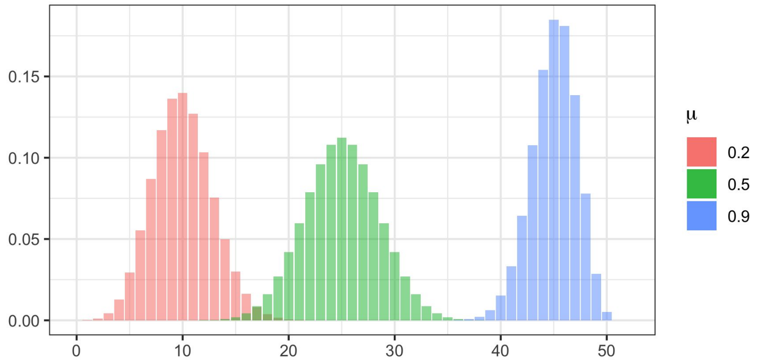
Question 5.2

Interpret the MLE for the parameter, μ, of a binomial distribution, assuming fixed m (number of trials). Does the MLE for μ make intuitive sense to you? Think through a few of your examples from Question 4.3.

Every year, 10 scientists go to the same geographic area (same Lyme prevalence) and they each collect 40 ticks. They test each tick for Lyme disease and record the number of ticks that have Lyme. Let $x^{(i)}$ be the number of ticks with Lyme in the ith scientist's bunch.

$$\hat{\mu} = \frac{1}{nm} \sum_{i=1}^{n} x^{(i)}$$

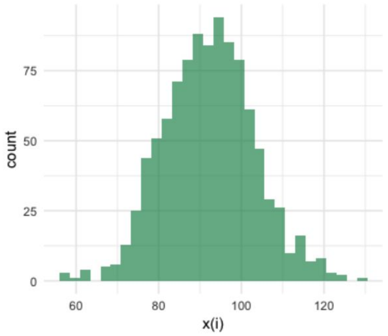
Scientist ID (i)	$x^{(i)}$
1	8
2	9
3	14
4	15
5	12
6	7
7	6
8	8
9	8
10	14



Question 5.3

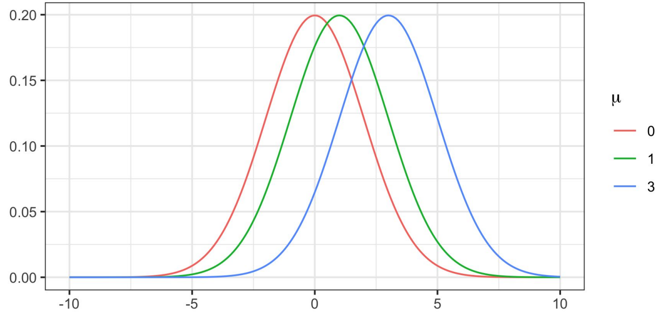
Interpret the MLEs for the parameters, μ and σ, of a normal distribution. Do these results make intuitive sense to you? Think through a few of your examples from Question 4.1.

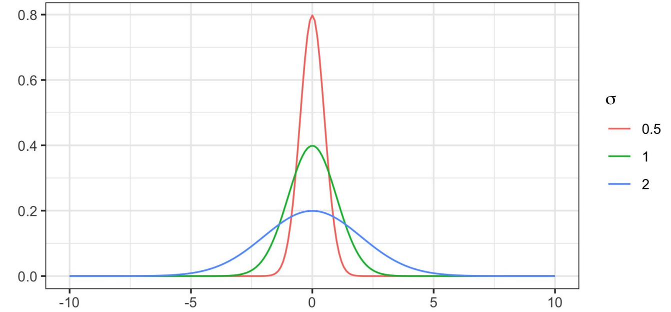
You have waist circumference data on 1045 men aged 70 and above (see Dey's 2002 paper in the Journal of the American Geriatric Society). It looks like this:



$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x^{(i)}$$

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x^{(i)} - \hat{\mu}\right)^2}$$





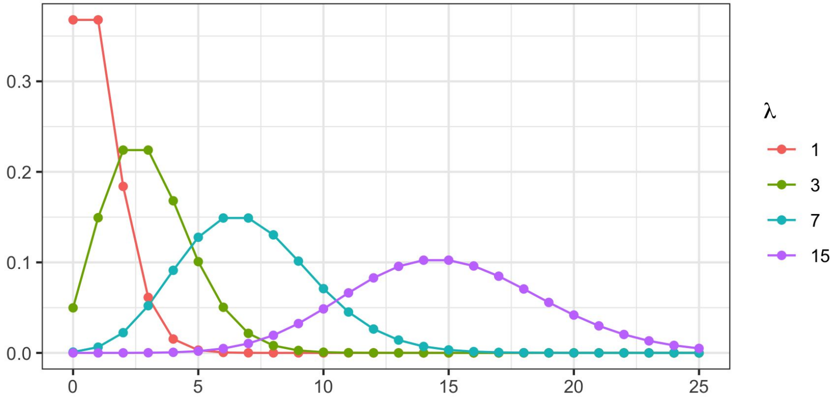
Question 5.4

Interpret the MLE for the parameter, λ, of a Poisson distribution. Does this result make intuitive sense to you? Think through a few of your examples from Question 4.4.

Imagine you are Ladislaus Bortkiewicz, and you are modeling the number of persons killed by mule or horse kicks in the Prussian army per year. You have data from the late 1800s over the course of 20 years. Let $x^{(i)}$ be the number of people killed in year i.

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} x^{(i)}$$

Year (i)	$x^{(i)}$	Year (i)	$x^{(i)}$
1	8	11	9
2	10	12	7
3	5	13	10
4	3	14	12
5	10	15	8
6	8	16	7
7	7	17	8
8	2	18	8
9	6	19	10
10	11	20	7



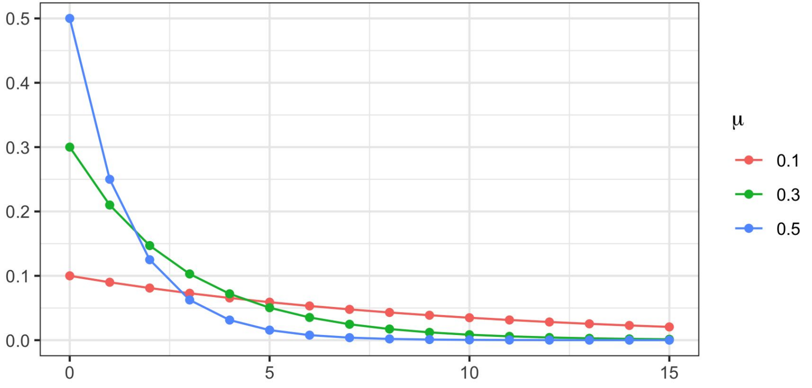
Question 5.5

Interpret the MLE for the parameter, μ, of a geometric distribution. Does this result make intuitive sense to you? Think through a few of your examples from Question 4.5.

You are observing several patients' skin in a clinical study to see how long it takes them to develop a rash. You take a picture each day. Let $x^{(i)}$ be the number of days of *no rash* before the rash occurs.

$$\hat{\mu} = \frac{n}{\sum_{i=1}^{n}(x^{(i)} + 1)}$$

Patient ID (i)	$x^{(i)}$
1	4
2	1
3	0
4	2
5	2
6	4
7	3
8	1
9	0
10	1



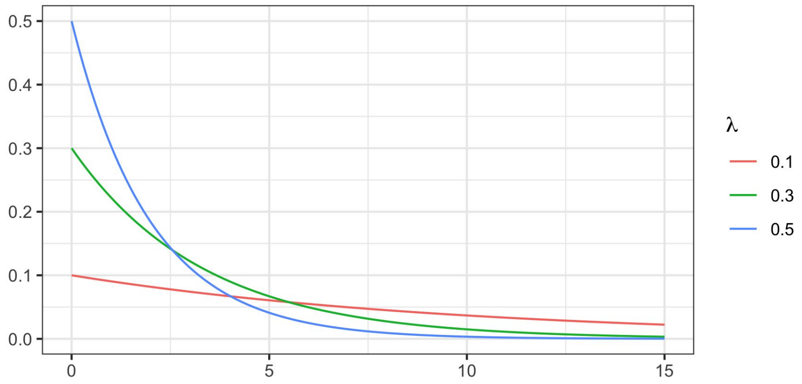
Question 5.6

Interpret the MLE for the parameter, λ, of an exponential distribution. Does this result make intuitive sense to you? Think through a few of your examples from Question 4.6.

Same situation as above except that instead of taking a picture each day, the patient texts you at the moment he/she observes a rash. The data look like this, where $x^{(i)}$ is the time (in days) at which patient i develops a rash:

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x^{(i)}}$$

Patient ID (i)	$x^{(i)}$
1	2.25
2	3.43
3	0.68
4	0.04
5	3.78
6	5.65
7	2.88
8	3.88
9	2.83
10	1.87



Distribution	Parameter	ML Estimate	Domain of $x^{(i)}$
Univariate Normal	μ	$\frac{1}{n}\sum_{i=1}^{n} x^{(i)}$	$\mathbb{R}$
	σ	$\frac{1}{n}\sum_{i=1}^{n} \left(x^{(i)} - \hat{\mu}\right)^2$	$\mathbb{R}$
Multivariate Normal	μ	$\frac{1}{n}\sum_{i=1}^{n} x^{(i)}$	$\mathbb{R}^m$
	$\mathbf{\Sigma}$	$\frac{1}{n}\sum_{i=1}^{n}(x^{(i)} - \hat{\mu})(x^{(i)} - \hat{\mu})^T$	$\mathbb{R}^m$
Bernoulli	μ	$\frac{1}{n}\sum_{i=1}^{n} x^{(i)}$	$\{0, 1\}$
Binomial (fixed m)	μ	$\frac{1}{nm}\sum_{i=1}^{n} x^{(i)}$	$\{0, 1, \dots, m\}$
Poisson	λ	$\frac{1}{n}\sum_{i=1}^{n} x^{(i)}$	$\{0, 1, \dots\}$
Geometric	μ	$\frac{n}{\sum_{i=1}^{n}(x^{(i)} + 1)}$	$\{0, 1, \dots\}$
Exponential	λ	$\frac{n}{\sum_{i=1}^{n} x^{(i)}}$	$\mathbb{R}^+$

Question 5.7

In Question 4.7, we examined several examples of experimental conditions and datasets and discussed which probability distribution best modeled each one. Using the formulas above and the actual datasets from Question 4.7, calculate the MLEs for the parameter(s) of your chosen probability distributions.

Question 5.2 (Binomial)	MLE of mu = 0.2525
Question 5.3 (Normal)	Need to estimate from picture. Mean approx. 90, sigma approx. 15.
Question 5.4 (Poisson)	MLE of lambda = 7.8
Question 5.5 (Geometric)	MLE of mu = 0.3571
Question 5.6 (Exponential)	MLE of lambda = 0.3664