Using R in Undergraduate Probability and Mathematical Statistics Courses

Amy G. Froelich
Department of Statistics
Iowa State University

Undergraduate Probability and Mathematical Statistics at Iowa State

- Introduction to Probability (Calculus-Based)
  - Probability
  - Discrete Random Variables
  - Continuous Random Variables
  - Multivariate Distributions
- Introductory to Mathematical Statistics (Calculus-Based)
  - Transformations of Random Variables and Sampling Distributions
  - Mathematical Statistics and Inference
  - From Theory to Practice and Back
- Textbook: Mathematical Statistics with Applications by Wackerly, Mendenhall & Scheaffer

Why Use R in Prob/Math Stat?

- Curriculum: Exposure to R
- Content: Model nature of statistics
- Pedagogy: Ground theoretical concepts in understanding obtained from earlier applied statistics courses.

Gamma Distribution

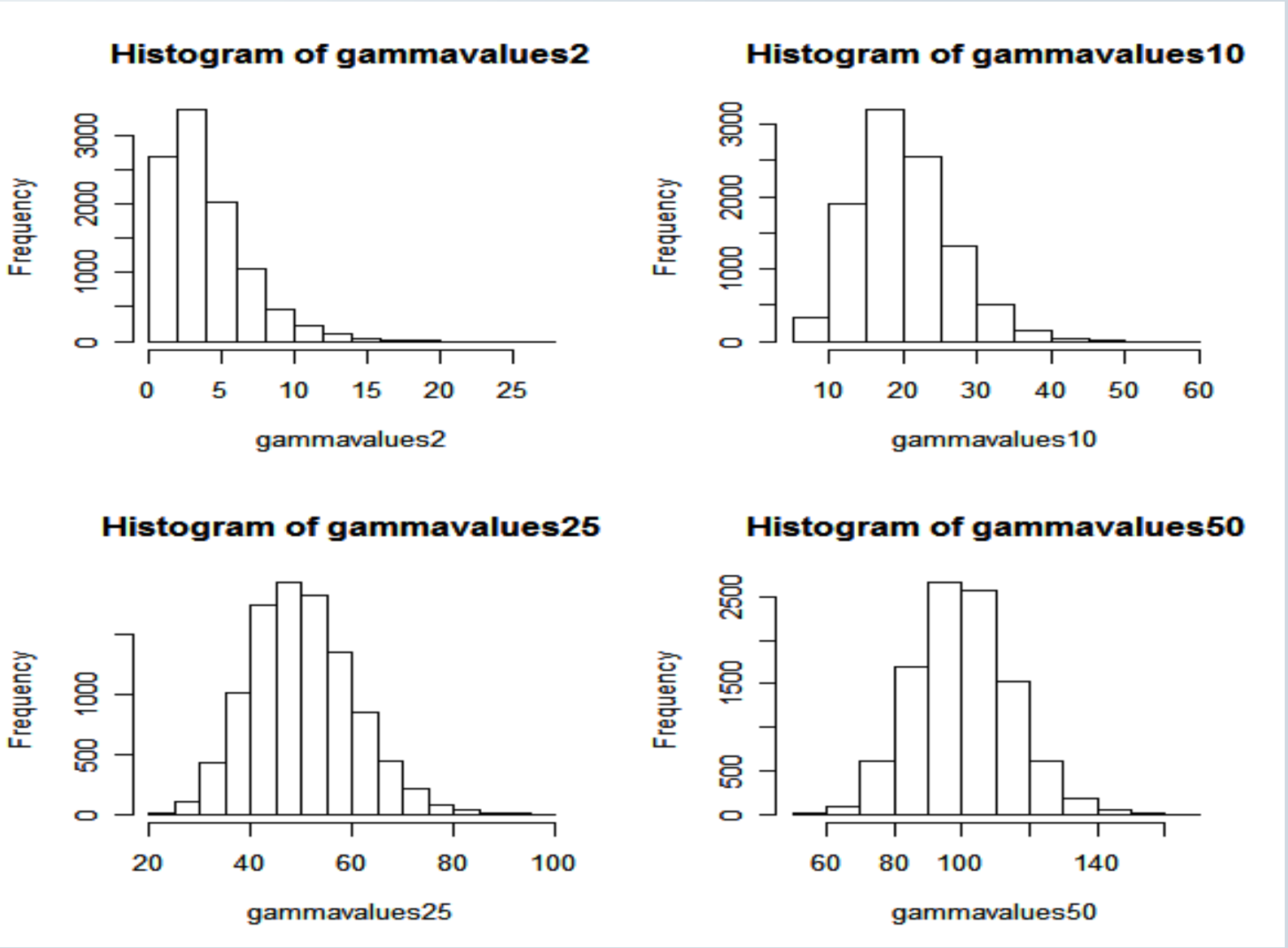
- Distribution Characteristics
- Distribution of Sample Mean
- Estimation of Parameters

Distribution Characteristics

- Study influence of parameters on
  - Shape
  - Mean
  - Variance
  - Percentiles
- Study through
  - Observed distributions through simulation
  - Theoretical distributions

Distribution Characteristics

- Simulation Study for Shape
  - Parameters (alpha = 2, 10, 25, 50; beta = 2)
  - 10,000 observations

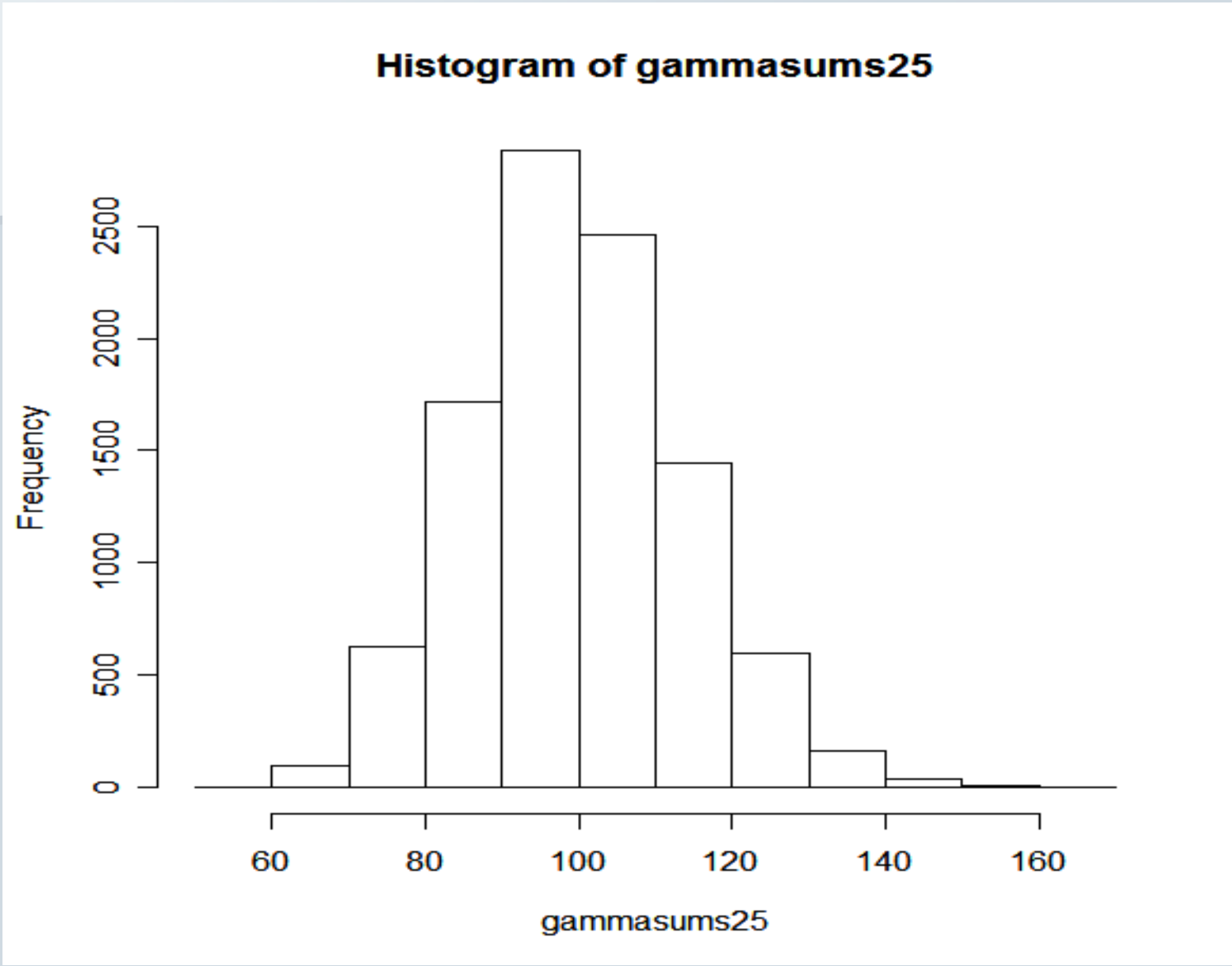

Histogram of gammavalues2
Frequency
0
1000
2000
3000
0
5
10
15
20
25
gammavalues2
Histogram of gammavalues10
Frequency
0
1000
2000
3000
10
20
30
40
50
60
gammavalues10
Histogram of gammavalues25
Frequency
0
500
1000
20
40
60
80
100
gammavalues25
Histogram of gammavalues50
Frequency
0
500
1500
2500
60
80
100
140
gammavalues50

Distribution of Sample Mean

- What is the distribution of the sample mean from a random sample of gamma random variables?
  - Distribution of sum of gamma random variables.
    - Introduce concept through simulation.
    - Characteristics of observed simulation.
    - Intuition about distribution.

Distribution of Sum of Gamma R.V.s

- Simulation Example
  - Parameters (alpha = 2, beta = 2)
  - Sample Size (n = 25)


Histogram of gammasums25
Frequency
0
500
1000
1500
2000
2500
60
80
100
120
140
160
gammasums25

Distribution of Sum of Gamma R.V.s

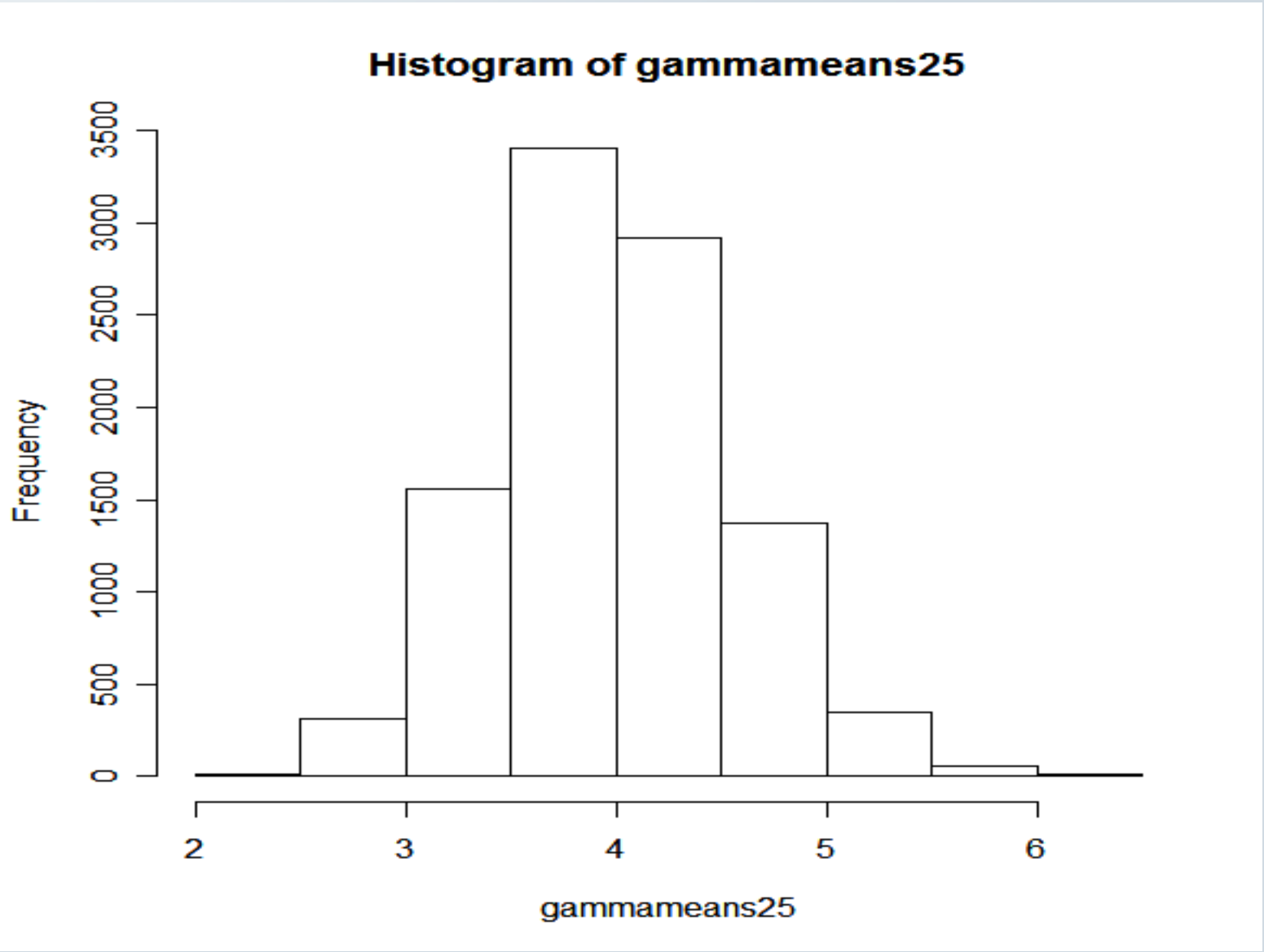
- Observed Distribution
  - Very Slightly Skewed Right
  - Observed mean = 99.77758
  - Observed variance = 196.796
- Intuition about Distribution
  - Gamma distribution?
  - Parameters? alpha = 50, beta = 2

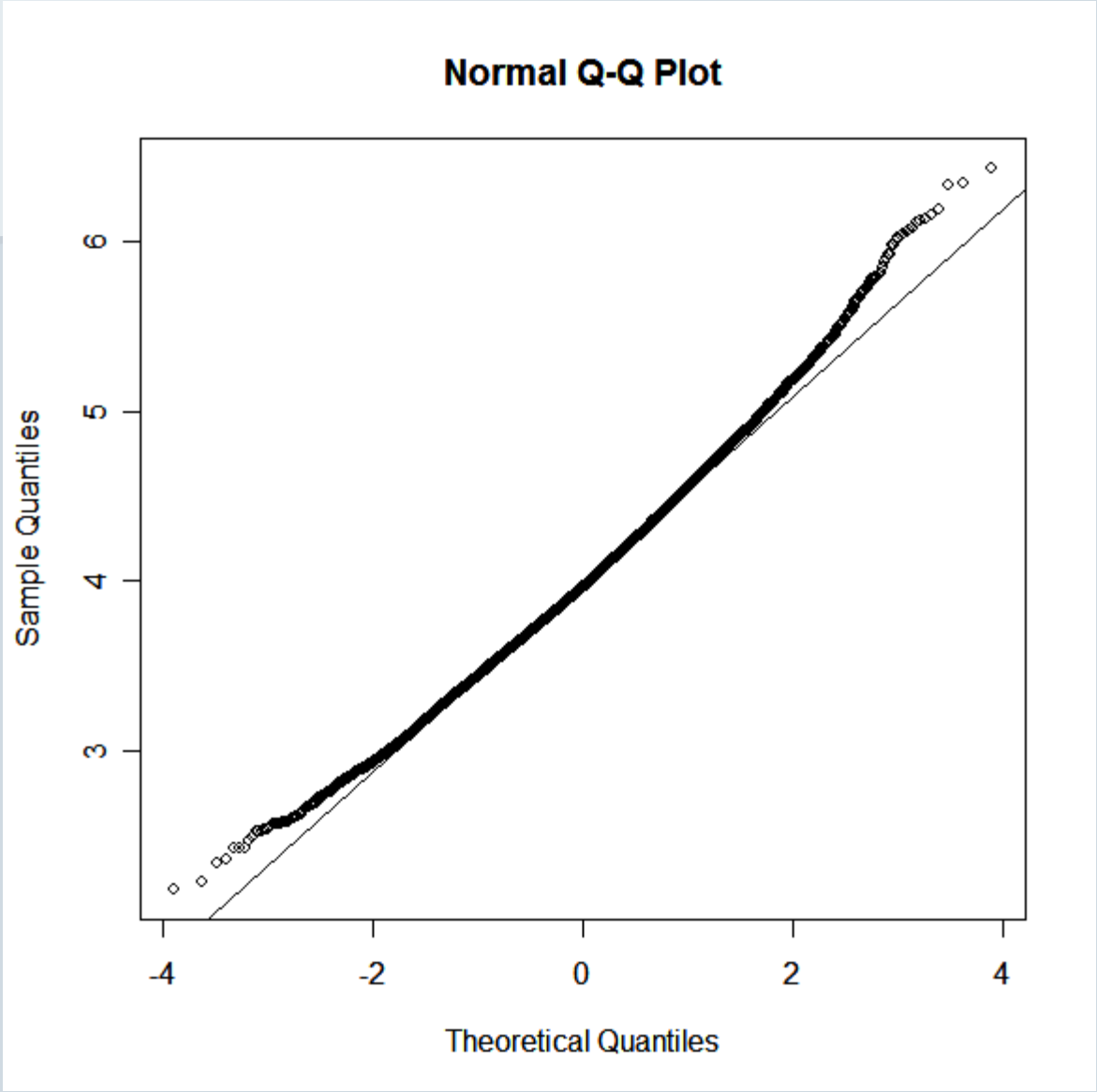
Distribution of Sample Mean

- Distribution of sample mean
  - Connection to distribution of sum.
  - Introduce concept through simulation.
  - Characteristics of observed distribution.
  - Intuition about distribution.
  - Introduce Central Limit Theorem.

Distribution of Sample Mean

- Simulation Example
  - Parameters (alpha = 2, beta = 2)
  - Sample Size (n = 25)


Histogram of gammameans25
Frequency
0
500
1000
1500
2000
2500
3000
3500
2
3
4
5
6
gammameans25


Normal Q-Q Plot
Sample Quantiles
Theoretical Quantiles
3
4
5
6
-4
-2
0
2
4

Distribution of Sample Mean

- Observed Distribution
  - Slightly Skewed Right Distribution
  - Observed Mean = 3.991103
  - Observed Variance = 0.3148736
- Intuition about distribution
  - Gamma Distribution?
  - Parameters? alpha = 50 beta = 2/25
  - Normal Distribution?

Parameter Estimation

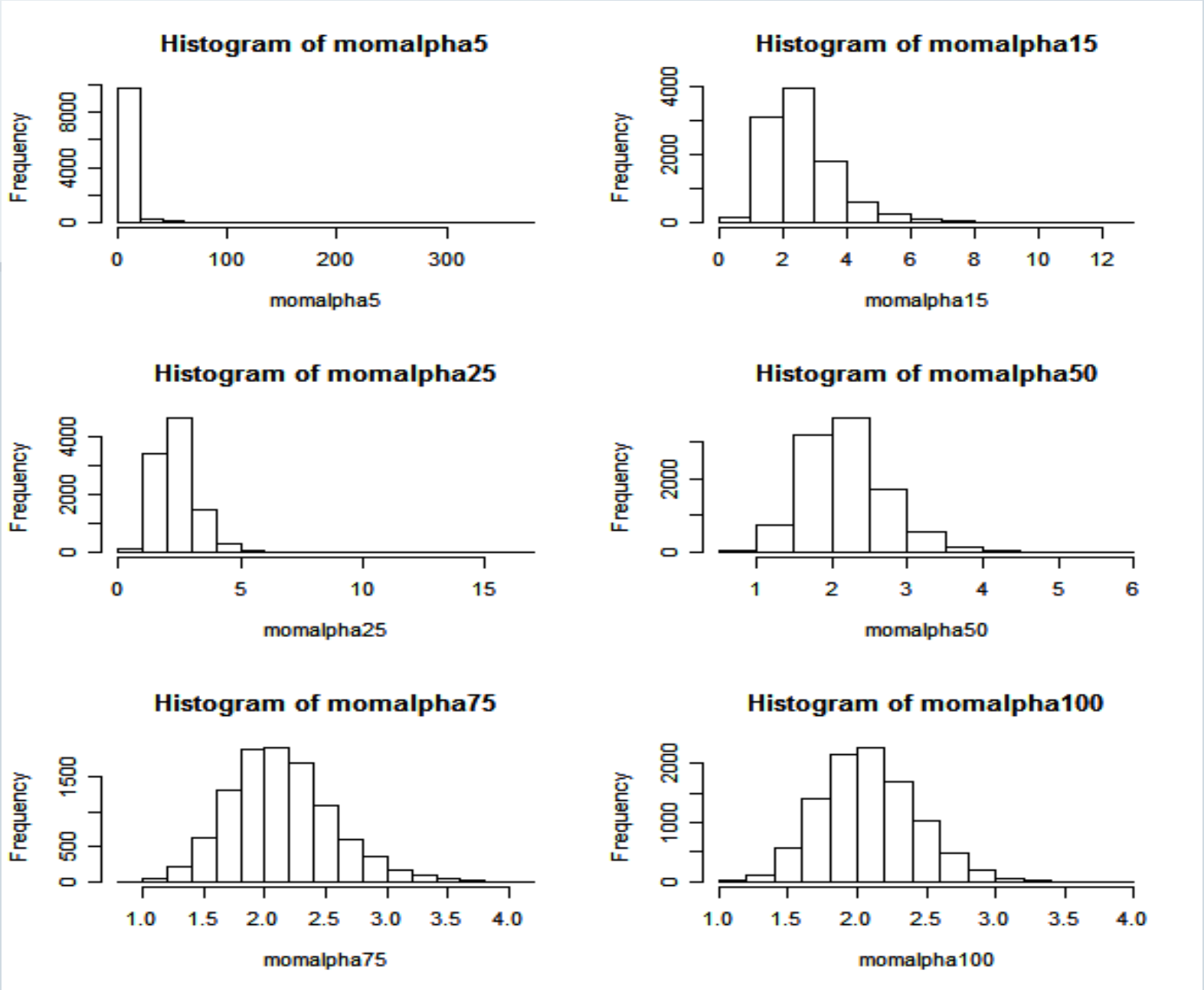
- Two Methods
  - Method of Moments
  - Maximum Likelihood
- Properties of Estimators
  - Distribution (Shape, Center, Spread)
  - Bias
  - Mean Square Error (MSE)
  - Mean Absolute Deviation (MAD)

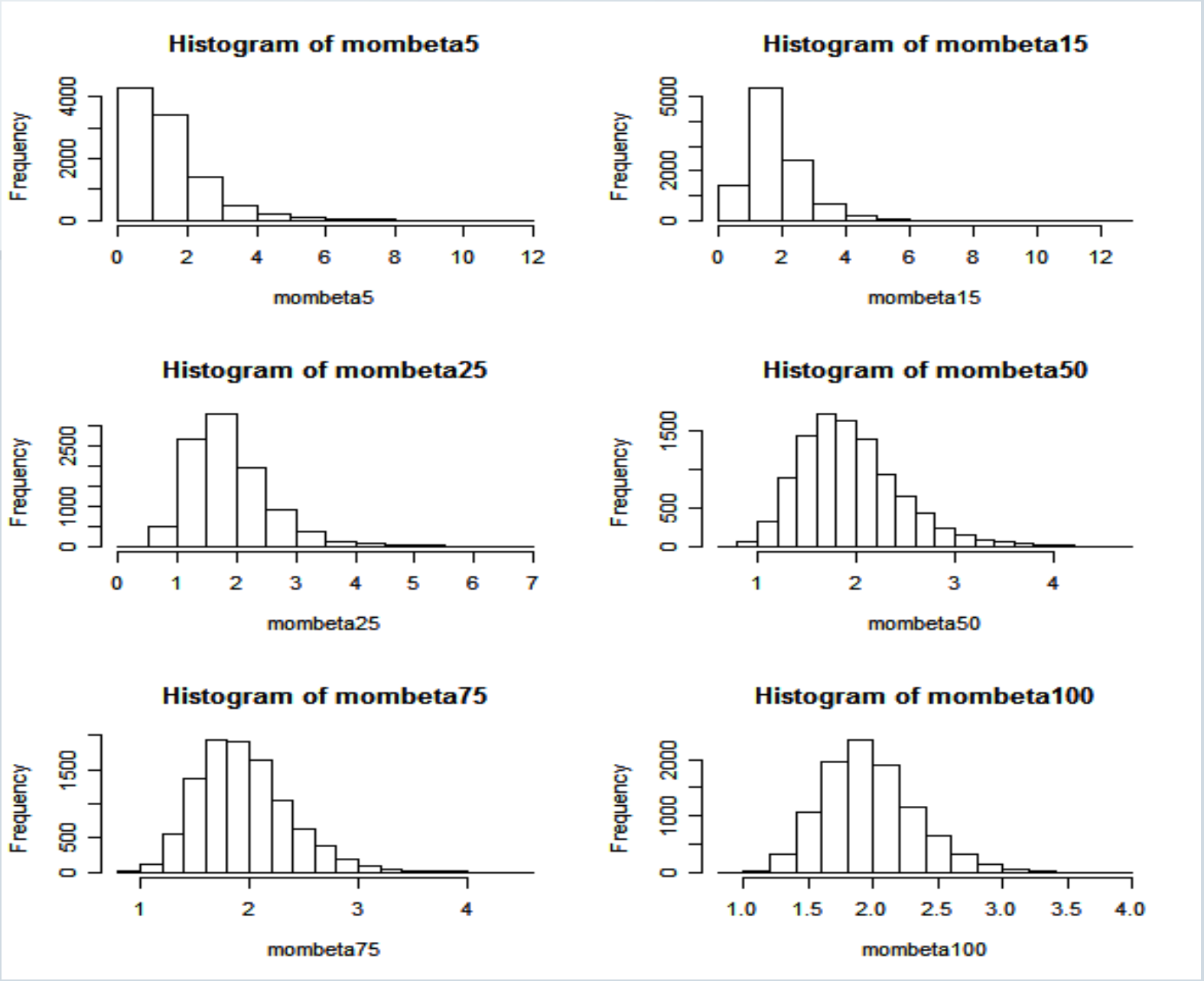
Method of Moments

- Introduce concept through previous work.
- Theoretically derive formulas for MOM estimators.
- Study properties of MOM estimators through simulation.
  - Characteristics of observed distributions.
  - Introduce concepts of observed bias, MSE, MAD.

Method of Moments

- Simulation Example
  - Parameter Values (alpha = 2, beta = 2)
  - Sample Size (n = 5, 15, 25, 50, 75, 100)


Histogram of momalpha5
Frequency
0
4000
8000
0
100
200
300
momalpha5
Histogram of momalpha15
Frequency
0
2000
4000
0
2
4
6
8
10
12
momalpha15
Histogram of momalpha25
Frequency
0
2000
4000
0
5
10
15
momalpha25
Histogram of momalpha50
Frequency
0
2000
1
2
3
4
5
6
momalpha50
Histogram of momalpha75
Frequency
0
500
1500
1.0
1.5
2.0
2.5
3.0
3.5
4.0
momalpha75
Histogram of momalpha100
Frequency
0
1000
2000
1.0
1.5
2.0
2.5
3.0
3.5
4.0
momalpha100


Histogram of mombeta5
Frequency
0
2000
4000
0
2
4
6
8
10
12
mombeta5
Histogram of mombeta15
Frequency
0
2000
5000
0
2
4
6
8
10
12
mombeta15
Histogram of mombeta25
Frequency
0
1000
2500
0
1
2
3
4
5
6
7
mombeta25
Histogram of mombeta50
Frequency
0
500
1500
1
2
3
4
mombeta50
Histogram of mombeta75
Frequency
0
500
1500
1
2
3
4
mombeta75
Histogram of mombeta100
Frequency
0
1000
2000
1.0
1.5
2.0
2.5
3.0
3.5
4.0
mombeta100

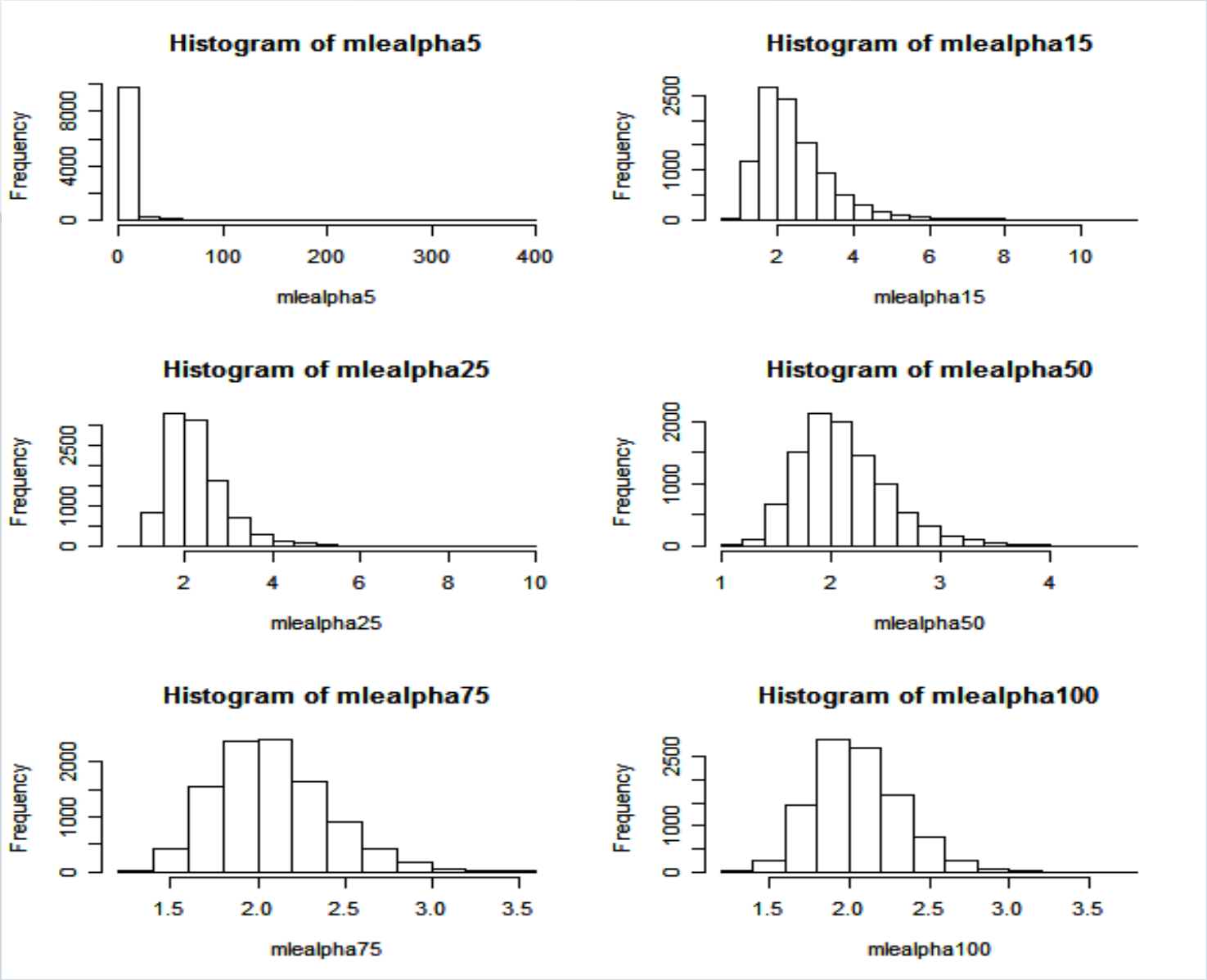
	α		β	
Sample Size	**Mean**	**Variance**	**Mean**	**Variance**
5	5.072479	61.87657	1.443441	1.278346
15	2.618371	1.35053	1.803246	0.7176061
25	2.365998	0.6300843	1.877318	0.474249
50	2.178479	0.2695496	1.938572	0.2601572
75	2.127446	0.1724364	1.953107	0.1791825
100	2.086805	0.1228160	1.971858	0.1325997

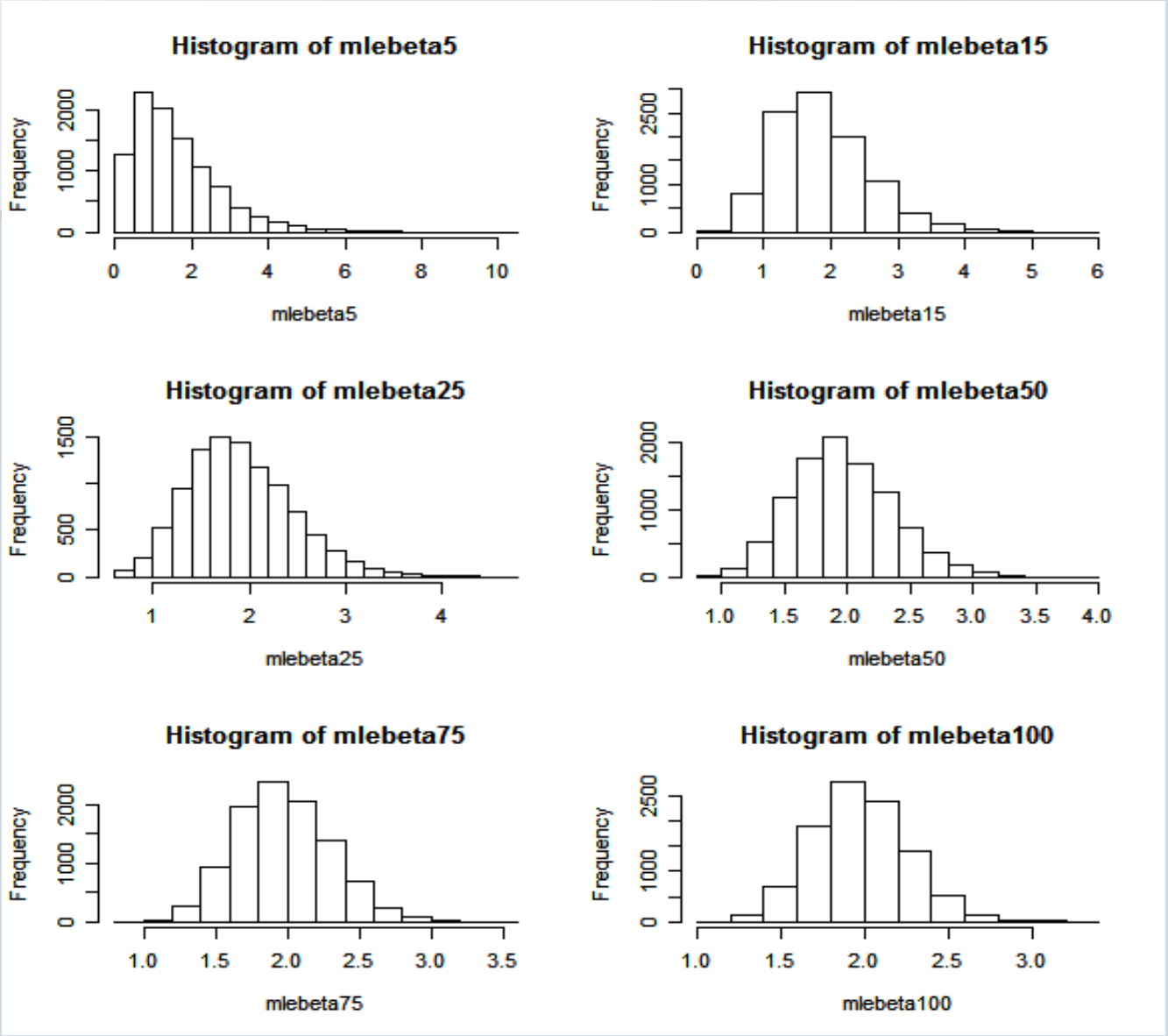
Maximum Likelihood

- Introduce concept through simulated data.
- Theoretically derive the ML Estimators.
  - Closed form solution not possible.
  - Maximization scheme needed (ex. Newton's Method.)
- Study properties of ML Estimators through simulation.
  - Characteristics of observed distribution.
  - Introduce concepts of observed bias, MSE, MAD.

Maximum Likelihood

- Simulation Example
  - Parameter Values (alpha = 2, beta = 2)
  - Sample Size (n = 5, 15, 25, 50, 75, 100)


Histogram of mlealpha5
Frequency
0
4000
8000
0
100
200
300
400
mlealpha5
Histogram of mlealpha15
Frequency
0
1000
2500
2
4
6
8
10
mlealpha15
Histogram of mlealpha25
Frequency
0
1000
2500
2
4
6
8
10
mlealpha25
Histogram of mlealpha50
Frequency
0
1000
2000
1
2
3
4
mlealpha50
Histogram of mlealpha75
Frequency
0
1000
2000
1.5
2.0
2.5
3.0
3.5
mlealpha75
Histogram of mlealpha100
Frequency
0
1000
2500
1.5
2.0
2.5
3.0
3.5
mlealpha100


Histogram of mlebeta5
Frequency
0
1000
2000
0
2
4
6
8
10
mlebeta5
Histogram of mlebeta15
Frequency
0
1000
2500
0
1
2
3
4
5
6
mlebeta15
Histogram of mlebeta25
Frequency
0
500
1500
1
2
3
4
mlebeta25
Histogram of mlebeta50
Frequency
0
1000
2000
1.0
1.5
2.0
2.5
3.0
3.5
4.0
mlebeta50
Histogram of mlebeta75
Frequency
0
1000
2000
1.0
1.5
2.0
2.5
3.0
3.5
mlebeta75
Histogram of mlebeta100
Frequency
0
1000
2500
1.0
1.5
2.0
2.5
3.0
mlebeta100

	α		β	
Sample Size	**Mean**	**Variance**	**Mean**	**Variance**
5	4.729084	62.63001	1.60065	1.447492
15	2.448018	1.040122	1.864778	0.5549284
25	2.238958	0.4268584	1.924708	0.3371712
50	2.112406	0.1702163	1.959108	0.1746927
75	2.079821	0.1055121	1.968891	0.1168963
100	2.056419	0.07579958	1.979734	0.08670047

Comparison of MOM and ML Estimators

- Simulation Example
  - Parameter Values (alpha = 2, beta = 2)
  - Sample Sizes (n = 10, 50, 100)
- Characteristics of observed distribution.
  - Mean (bias), Var, MSE, MAD.

α

	Obs. Mean		Obs. Var.	
Sample Size	**MOM**	**MLE**	**MOM**	**MLE**
n = 10	3.029221	2.786177	3.266518	2.795095
n = 50	2.172688	2.105872	0.257089	0.1637300
n = 100	2.097151	2.060918	0.129137	0.0784782

α

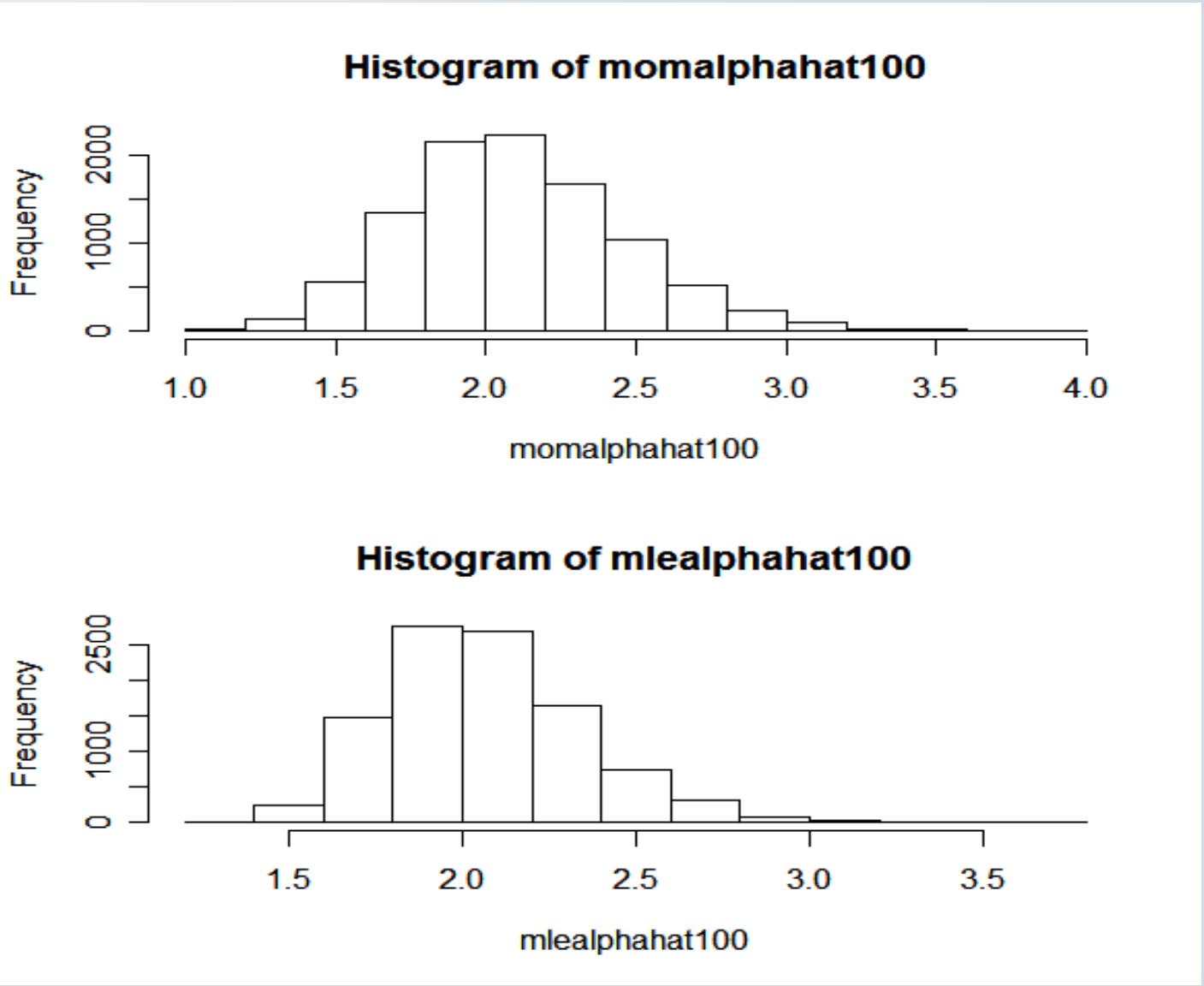
	Obs. MSE		Obs. MAD	
Sample Size	**MOM**	**MLE**	**MOM**	**MLE**
n = 10	4.32549	3.41289	1.295606	1.087239
n = 50	0.286885	0.174923	0.409188	0.3190955
n = 100	0.138562	0.082181	0.289018	0.221859

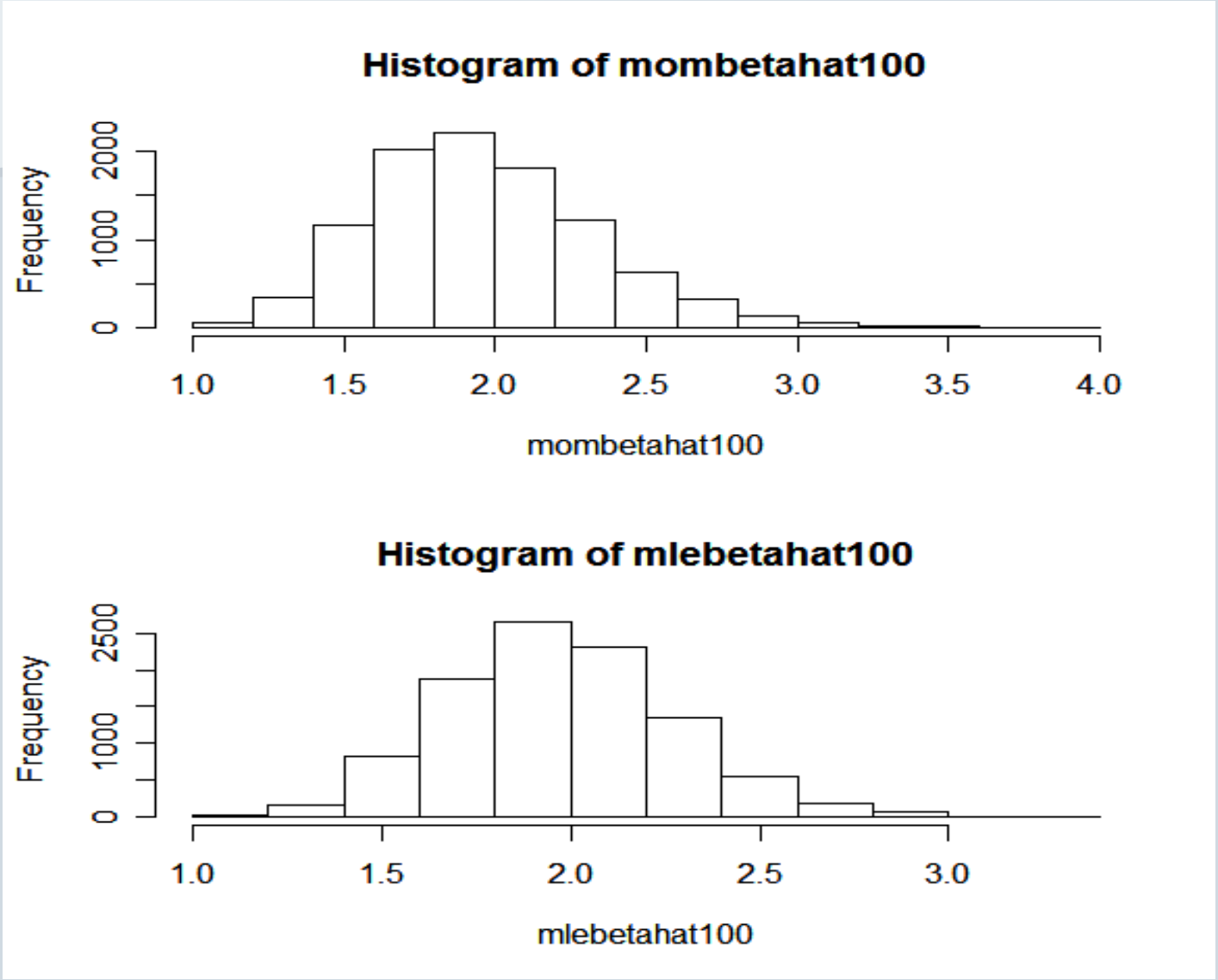
β

	Obs. Mean		Obs. Var.	
Sample Size	**MOM**	**MLE**	**MOM**	**MLE**
n = 10	1.695161	1.791985	0.893835	0.786961
n = 50	1.94285	1.965812	0.260478	0.173177
n = 100	1.962924	1.974595	0.137522	0.089519

β

	Obs. MSE		Obs. MAD	
Sample Size	**MOM**	**MLE**	**MOM**	**MLE**
n = 10	0.986672	0.830152	0.801810	0.739329
n = 50	0.263718	0.174328	0.402426	0.333662
n = 100	0.138883	0.090155	0.295451	0.2393529


Histogram of momalphahat100
Frequency
0
1000
2000
1.0
1.5
2.0
2.5
3.0
3.5
4.0
momalphahat100
Histogram of mlealphahat100
Frequency
0
1000
2500
1.5
2.0
2.5
3.0
3.5
mlealphahat100


Histogram of mombetahat100
Frequency
0
1000
2000
1.0
1.5
2.0
2.5
3.0
3.5
4.0
mombetahat100
Histogram of mlebetahat100
Frequency
0
1000
2500
1.0
1.5
2.0
2.5
3.0
mlebetahat100

Comparison of MOM and ML Estimators

- Emphasizing observed nature of comparison.
- Emphasizing limitations due to
  - Sample size
  - Parameter Values
- Emphasizing asymptotic properties of MOM and ML Estimators.

Possible Implementation Issues

- Textbook
  - No textbook available matching this approach, yet!
- Using R
  - Need lots of supplements for students.
- Student expectations from course
  - Ease them in early and often
- Departmental Support
  - Not a problem at Iowa State.

Student Reactions to R

- Course evaluations – Fall 2005.
- Did you like using R? Was it helpful to you? Did you find it easy or difficult to use?
  - 34 responses –32 positive; 1 indifferent; 1 negative comments to using R.

Selected Student Comments

- “I liked using R. I found it easy with the examples given in class and the handouts clearly explaining how to use it.”
- “I liked using R, it was helpful in doing the difficult distribution problems. I also liked it because we could construct histograms to visual(ize) the distributions.”
- “I liked using R, easy to use, somewhat difficult to know what to type in. Continue using it in 341 since it’s used in the real world.”
- “I thought R was very helpful and easy to use… In most cases, the professor’s instructions made the program very understandable.”
- “It was nice to use R because it seems really versatile and I like the easy to understand interface.”
- “I think R was very helpful, because we didn’t have to calculate everything by hand. It was easy to use too, because we got a good explanation about it.”