

Capture Recapture estimation using finite mixtures of arbitrary dimension

Richard Arnold
Yu Hayakawa
Paul Yip

International Biometrics Society
Taupo, 30 Nov-3 Dec 2009

Outline

- Capture-Recapture experiments and models
- Reversible Jump MCMC
- Application: Software reliability

Capture-Recapture

- k repeated samples taken from a population
- Population size N is unknown
- D distinct individuals are seen
- Some seen on multiple occasions, some only once
- **Goal:** estimate population size N

Capture-Recapture: Applications

- **Size of animal populations:**

Samples are occasions on which animals are trapped, marked (for re-identification) and released.

N is the number of animals in the population.

Capture-Recapture: Applications

- **Size of animal populations:**

Samples are occasions on which animals are trapped, marked (for re-identification) and released.

N is the number of animals in the population.

- **Software testing:**

Samples are independent software testers detecting errors.

N is the number of errors in the piece of software.

Observations: Capture Matrix X

		Samples					
		1	2	3	...	k	
Individuals	1	1	0	1	...	0	Observed
	2	1	0	1	...	0	
	3	0	0	0	...	0	
	4	0	1	1	...	0	
	5	0	0	1	...	0	
	⋮						
	D	0	1	1	...	0	
$D+1$	0	0	0	...	0	Unobserved	
$D+2$	0	0	0	...	0		
⋮							
N	0	0	0	...	0		

An unknown number ($N - D$) of zero rows: to be estimated.

Capture-Recapture

Observations form the $D \times k$ **capture matrix**

$$X_{ij} = \begin{cases} 1 & \text{if individual } i \text{ appears in sample } j \\ 0 & \text{otherwise} \end{cases}$$

Probability of capture

$$\begin{aligned} p_{ij} &= \text{Probability individual } i \text{ is captured in sample } j \\ &= p(X_{ij} = 1) \end{aligned}$$

$$X_{ij} \sim \text{Bernoulli}(p_{ij})$$

$$p(X|P, N) = \frac{N!}{\prod_{\mathbf{x}} N_{\mathbf{x}}!} \prod_{i=1}^N \prod_{j=1}^k p_{ij}^{x_{ij}} (1 - p_{ij})^{1-x_{ij}}$$

Some key assumptions

1. All individuals are catchable
2. **Closed population:** no births/deaths or migration
3. No loss of marks (identifiable individuals)
4. No impact of sampling on capture probabilities (i.e. no behavioural responses)

(Models do exist for violations of 2-4)

What makes this a difficult problem?

What makes this a difficult problem?

1. Heterogeneity

What makes this a difficult problem?

1. Heterogeneity

- **Individuals differ** – some are more easily caught: model M_h

What makes this a difficult problem?

1. Heterogeneity

- **Individuals differ** – some are more easily caught: model M_h
- **Samples differ** – some samples are more successful: model M_t

What makes this a difficult problem?

1. Heterogeneity

- **Individuals differ** – some are more easily caught: model M_h
- **Samples differ** – some samples are more successful: model M_t
- **Both sources** – model M_{th}

What makes this a difficult problem?

1. Heterogeneity

- **Individuals differ** – some are more easily caught: model M_h
- **Samples differ** – some samples are more successful: model M_t
- **Both sources** – model M_{th}

2. Identifiability

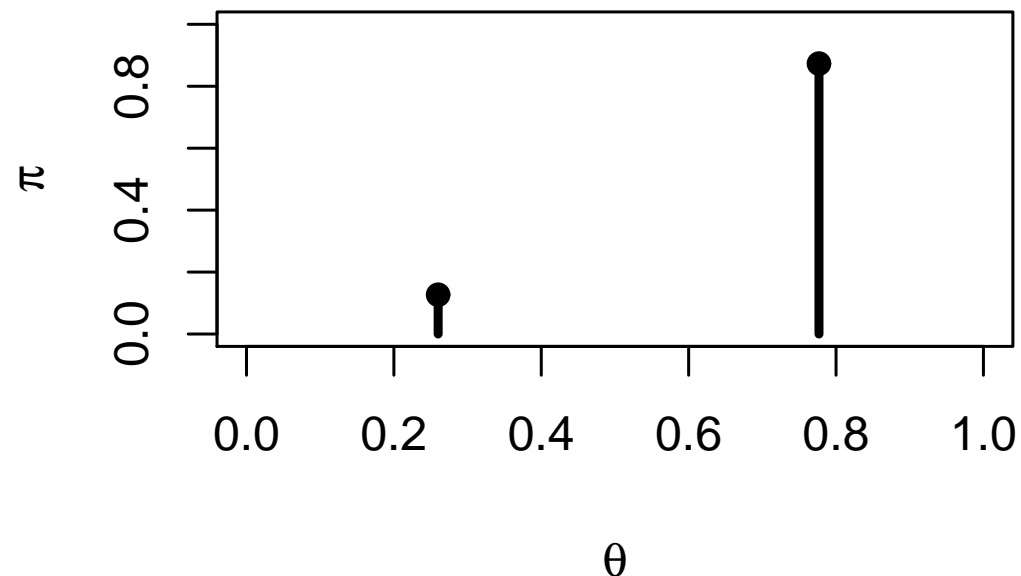
- N is identifiable **within** but not **between** model classes.
e.g. finite vs. beta (infinite) mixture models for P

Modelling Heterogeneity by Finite Mixtures

1. **Individuals** – animals/errors differ in their catchability

Individuals belong to A latent classes:

- Membership probabilities $\{\pi_a\}$
- Capture probabilities $\{\phi_a\}$



$$f(p_i) = \sum_{a=1}^A \pi_a \delta(p_i - \phi_a) \qquad \sum_{a=1}^A \pi_a = 1$$

Modelling Heterogeneity by Finite Mixtures

2. **Samples** – probability of capture varies between samples

- **Model** $M_{t[f]h}$: k fixed effects

$$p_{ij} = \theta_{aj}$$

$$\text{logit}(\theta_{aj}) = \text{logit}(\phi_a) + \beta_j + \gamma_{aj}$$

Individual i is in latent class a

Modelling Heterogeneity by Finite Mixtures

2. **Samples** – probability of capture varies between samples

- **Model** $M_{t[f]h}$: k fixed effects

$$p_{ij} = \theta_{aj}$$

$$\text{logit}(\theta_{aj}) = \text{logit}(\phi_a) + \beta_j + \gamma_{aj}$$

Individual i is in latent class a

OR

Modelling Heterogeneity by Finite Mixtures

2. **Samples** – probability of capture varies between samples

- **Model** $M_{t[f]h}$: k fixed effects

$$p_{ij} = \theta_{aj}$$
$$\text{logit}(\theta_{aj}) = \text{logit}(\phi_a) + \beta_j + \gamma_{aj}$$

Individual i is in latent class a

OR

- **Model** M_{th} B latent classes: membership probabilities $\{\lambda_b\}$
capture probabilities $\{\psi_b\}$

$$p_{ij} = \theta_{ab}$$
$$\text{logit}(\theta_{ab}) = \text{logit}(\phi_a) + \text{logit}(\psi_b) - \text{logit}(\phi_1) + \gamma_{ab}$$

Individual i is in latent class a and

Sample j is in latent class b

Identifiability

- Ordering constraints on mixture component probabilities

$$0 < \phi_1 < \dots < \phi_A < 1 \quad \text{and} \quad \phi_1 = \psi_1 < \psi_2 < \dots < \psi_B < 1$$

Implemented by ‘repulsive’ priors (following Green 1995)

$$\underline{\phi}|A \sim \text{EOS}(\text{Uniform}(0, 1)^{2A+1})$$

$\{\phi_a\}$ are the even order statistics of $(2A + 1)$ draws from $U(0, 1)$

- Sum to zero constraints on fixed effects

$$\sum_j \beta_j = \sum_a \gamma_{aj} = \sum_j \gamma_{aj} = 0 \quad \text{or}$$

$$\sum_a \gamma_{ab} = \sum_b \gamma_{ab} = 0$$

Ensured by degenerate Normal priors, e.g.

$$\beta_j \sim \text{Normal}(0, \sigma_\beta'^2) \quad \text{with} \quad \sum_j \beta_j = 0$$

Priors

Parameter	Distribution	Constants	Models	
			$M_{t[f]h}$	M_{th}
N	Geometric($1 - \eta$)	$\eta = 0.999$	★	★
A	Geometric($1 - \rho_h$)	$\rho_h = 0.8$	★	★
B	Geometric($1 - \rho_h$)	$\rho_h = 0.8$		★
$\underline{\pi} A$	Dirichlet($\underline{1}\alpha_h$)	$\alpha_h = \frac{3}{2}$	★	★
$\underline{\phi} A$	EOS(Uniform($0,1$) $^{2A+1}$)		★	★
$\underline{\sigma}_\beta^2$	InverseGamma(ν_β, κ_β)	$\nu_\beta = 3, \kappa_\beta = 40$	★	
$\underline{\beta}$	DegenNormal($k;0,\sigma_\beta^2$)		★	
$\underline{\lambda} B$	Dirichlet($\underline{1}\alpha_t$)	$\alpha_t = \frac{3}{2}$		★
$\underline{\psi} B, \phi_1$	EOS(Uniform($\phi_1,1$) $^{2B-1}$)			★
$\underline{\zeta} A, (B)$	Bernoulli(p_γ)	$p_\gamma = \frac{1}{2}$	★	★
$\gamma \zeta, A, (B)$	DegenNormal($A,B;0,\sigma_\gamma^2$)	$\sigma_\gamma = 5$	★	★
$\underline{b} B, \underline{\lambda}$	Discrete($\{1, \dots, B\}; \underline{\lambda}$)			★

Bayesian Estimation – RJMCMC

Estimation by Reversible Jump MCMC

- Draw from posterior

$$p(\{m, \Theta_m\} | X)$$

- Reversible Jump MCMC allows dimension switching moves:
Addition/deletion of mixture components.
- 4 types of MCMC move in model M_h .

RJMCMC Updates

MCMC Step	1				
	$\theta_1^{(1)}$				
	$\theta_2^{(1)}$				
	$\theta_3^{(1)}$				

RJMCMC Updates

MCMC Step	1	2			
	$\theta_1^{(1)}$	$\theta_1^{(2)}$			
	$\theta_2^{(1)}$	$\theta_2^{(2)}$			
	$\theta_3^{(1)}$	$\theta_3^{(2)}$			

RJMCMC Updates

MCMC Step	1	2	3		
	$\theta_1^{(1)}$	$\theta_1^{(2)}$	$\theta_1^{(3)}$		
	$\theta_2^{(1)}$	$\theta_2^{(2)}$	$\theta_2^{(3)}$		
	$\theta_3^{(1)}$	$\theta_3^{(2)}$	$\theta_3^{(3)}$		

RJMCMC Updates

MCMC Step	1	2	3	4	
	$\theta_1^{(1)}$	$\theta_1^{(2)}$	$\theta_1^{(3)}$	$\theta_1^{(4)}$	
	$\theta_2^{(1)}$	$\theta_2^{(2)}$	$\theta_2^{(3)}$	$\theta_2^{(4)}$	
	$\theta_3^{(1)}$	$\theta_3^{(2)}$	$\theta_3^{(3)}$	$\theta_3^{(4)}$	

RJMCMC Updates

MCMC Step

	1	2	3	4	...
θ_1	$\theta_1^{(1)}$	$\theta_1^{(2)}$	$\theta_1^{(3)}$	$\theta_1^{(4)}$...
θ_2	$\theta_2^{(1)}$	$\theta_2^{(2)}$	$\theta_2^{(3)}$	$\theta_2^{(4)}$...
θ_3	$\theta_3^{(1)}$	$\theta_3^{(2)}$	$\theta_3^{(3)}$	$\theta_3^{(4)}$...

RJMCMC Updates

MCMC Step

1	2	3	4	...
$\theta_1^{(1)}$	$\theta_1^{(2)}$	$\theta_1^{(3)}$	$\theta_1^{(4)}$...
$\theta_2^{(1)}$	$\theta_2^{(2)}$	$\theta_2^{(3)}$	$\theta_2^{(4)}$...
$\theta_3^{(1)}$	$\theta_3^{(2)}$	$\theta_3^{(3)}$	$\theta_3^{(4)}$...

RJMCMC Step

1				
$m^{(1)}$				
$\theta_{m^{(1)}:1}^{(1)}$				
$\theta_{m^{(1)}:2}^{(1)}$				

RJMCMC Updates

MCMC Step

	1	2	3	4	...
$\theta_1^{(1)}$	$\theta_1^{(2)}$	$\theta_1^{(3)}$	$\theta_1^{(4)}$...	
$\theta_2^{(1)}$	$\theta_2^{(2)}$	$\theta_2^{(3)}$	$\theta_2^{(4)}$...	
$\theta_3^{(1)}$	$\theta_3^{(2)}$	$\theta_3^{(3)}$	$\theta_3^{(4)}$...	

RJMCMC Step

	1	2			
$m^{(1)}$	$m^{(2)}$				
$\theta_{m^{(1)}:1}^{(1)}$	$\theta_{m^{(2)}:1}^{(2)}$				
$\theta_{m^{(1)}:2}^{(1)}$	$\theta_{m^{(2)}:2}^{(2)}$				
	$\theta_{m^{(2)}:3}^{(2)}$				
	$\theta_{m^{(2)}:4}^{(2)}$				

RJMCMC Updates

MCMC Step

	1	2	3	4	...
$\theta_1^{(1)}$	$\theta_1^{(2)}$	$\theta_1^{(3)}$	$\theta_1^{(4)}$...	
$\theta_2^{(1)}$	$\theta_2^{(2)}$	$\theta_2^{(3)}$	$\theta_2^{(4)}$...	
$\theta_3^{(1)}$	$\theta_3^{(2)}$	$\theta_3^{(3)}$	$\theta_3^{(4)}$...	

RJMCMC Step

	1	2	3		
$m^{(1)}$	$m^{(2)}$	$m^{(3)}$			
$\theta_{m^{(1)}:1}^{(1)}$	$\theta_{m^{(2)}:1}^{(2)}$	$\theta_{m^{(3)}:1}^{(3)}$			
$\theta_{m^{(1)}:2}^{(1)}$	$\theta_{m^{(2)}:2}^{(2)}$	$\theta_{m^{(3)}:2}^{(3)}$			
	$\theta_{m^{(2)}:3}^{(2)}$	$\theta_{m^{(3)}:3}^{(3)}$			
	$\theta_{m^{(2)}:4}^{(2)}$	$\theta_{m^{(3)}:4}^{(3)}$			

RJMCMC Updates

MCMC Step

	1	2	3	4	...
$\theta_1^{(1)}$	$\theta_1^{(2)}$	$\theta_1^{(3)}$	$\theta_1^{(4)}$...	
$\theta_2^{(1)}$	$\theta_2^{(2)}$	$\theta_2^{(3)}$	$\theta_2^{(4)}$...	
$\theta_3^{(1)}$	$\theta_3^{(2)}$	$\theta_3^{(3)}$	$\theta_3^{(4)}$...	

RJMCMC Step

	1	2	3	4	
$m^{(1)}$	$m^{(2)}$	$m^{(3)}$	$m^{(4)}$		
$\theta_{m^{(1)}:1}^{(1)}$	$\theta_{m^{(2)}:1}^{(2)}$	$\theta_{m^{(3)}:1}^{(3)}$	$\theta_{m^{(4)}:1}^{(4)}$		
$\theta_{m^{(1)}:2}^{(1)}$	$\theta_{m^{(2)}:2}^{(2)}$	$\theta_{m^{(3)}:2}^{(3)}$	$\theta_{m^{(4)}:2}^{(4)}$		
	$\theta_{m^{(2)}:3}^{(2)}$	$\theta_{m^{(3)}:3}^{(3)}$			
	$\theta_{m^{(2)}:4}^{(2)}$	$\theta_{m^{(3)}:4}^{(3)}$			

RJMCMC Updates

MCMC Step

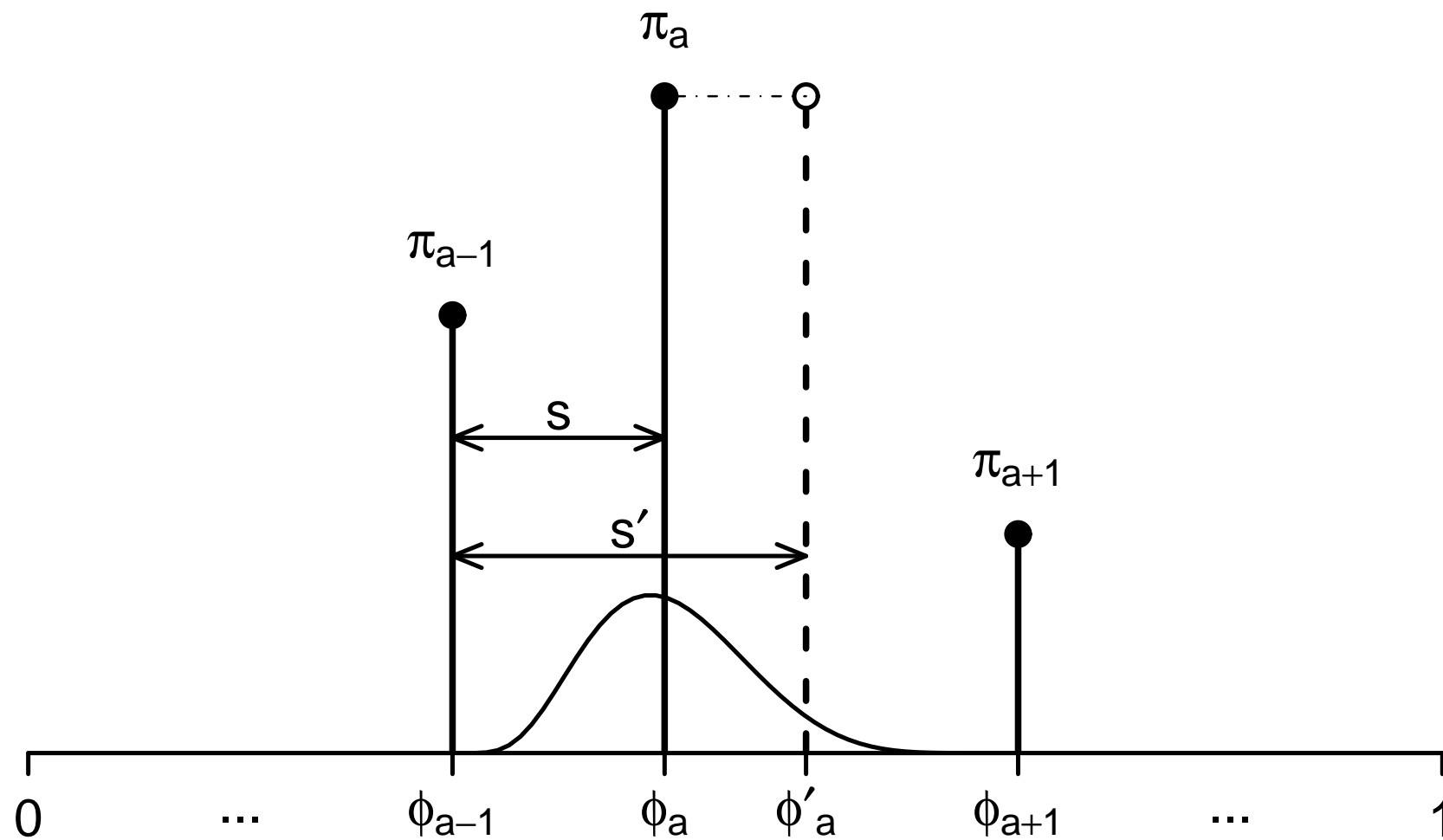
	1	2	3	4	...
$\theta_1^{(1)}$	$\theta_1^{(2)}$	$\theta_1^{(3)}$	$\theta_1^{(4)}$...	
$\theta_2^{(1)}$	$\theta_2^{(2)}$	$\theta_2^{(3)}$	$\theta_2^{(4)}$...	
$\theta_3^{(1)}$	$\theta_3^{(2)}$	$\theta_3^{(3)}$	$\theta_3^{(4)}$...	

RJMCMC Step

	1	2	3	4	...
$m^{(1)}$	$m^{(2)}$	$m^{(3)}$	$m^{(4)}$...	
$\theta_{m^{(1)}:1}^{(1)}$	$\theta_{m^{(2)}:1}^{(2)}$	$\theta_{m^{(3)}:1}^{(3)}$	$\theta_{m^{(4)}:1}^{(4)}$...	
$\theta_{m^{(1)}:2}^{(1)}$	$\theta_{m^{(2)}:2}^{(2)}$	$\theta_{m^{(3)}:2}^{(3)}$	$\theta_{m^{(4)}:2}^{(4)}$...	
	$\theta_{m^{(2)}:3}^{(2)}$	$\theta_{m^{(3)}:3}^{(3)}$...	
	$\theta_{m^{(2)}:4}^{(2)}$	$\theta_{m^{(3)}:4}^{(3)}$...	

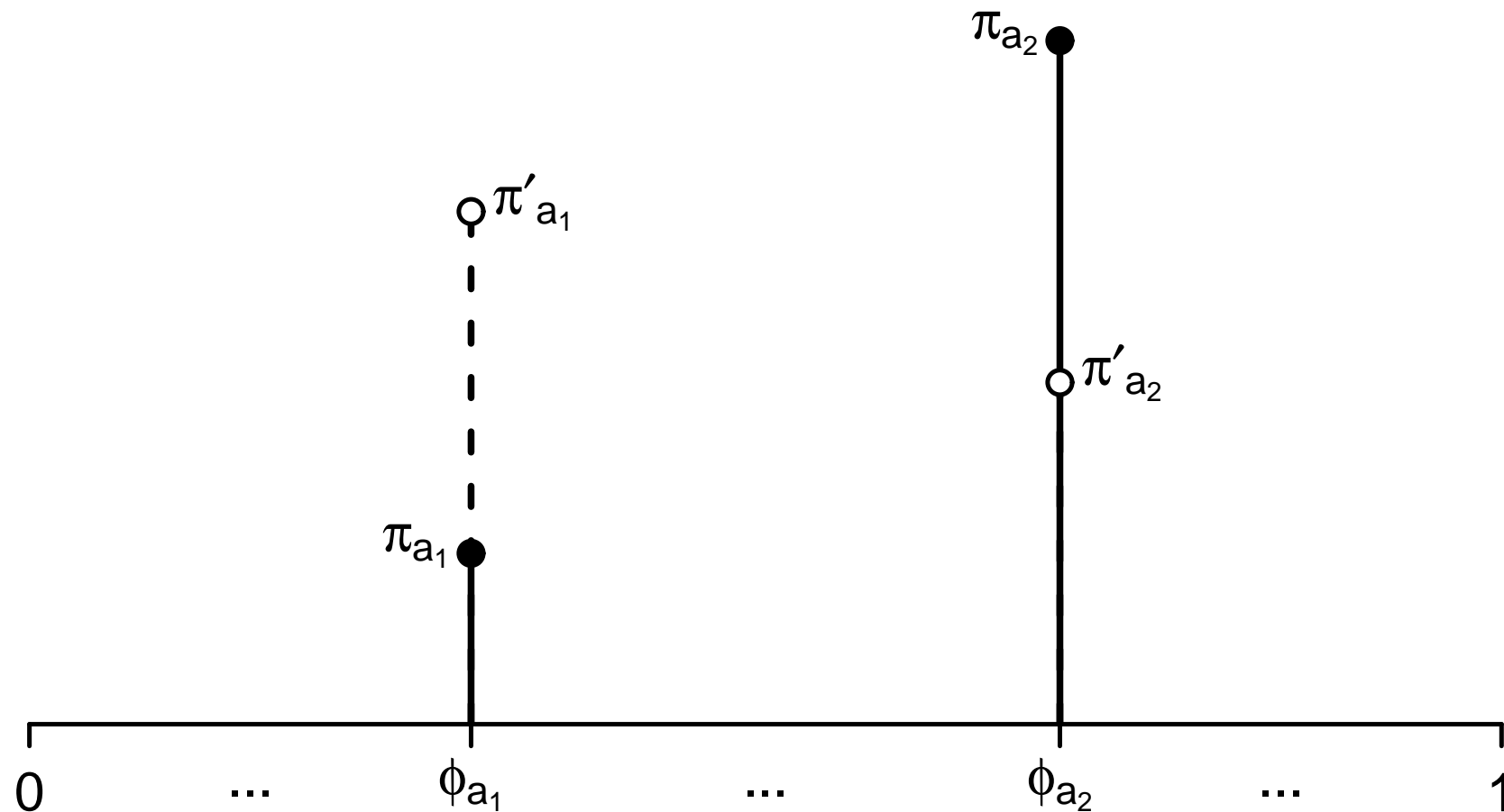
RJMCMC Updates

(1) Shift a support point



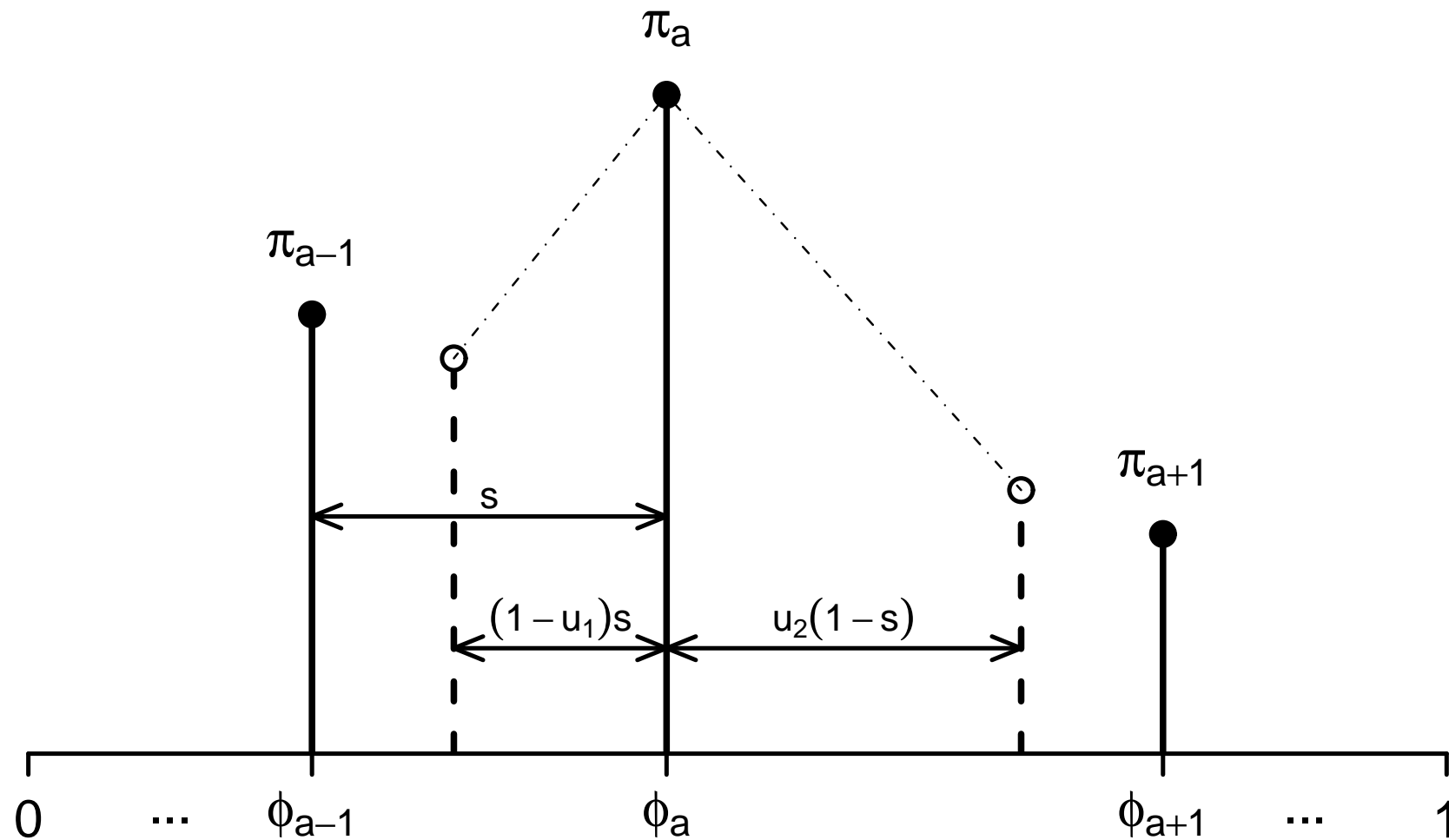
RJMCMC Updates

(2) Exchange probability between points



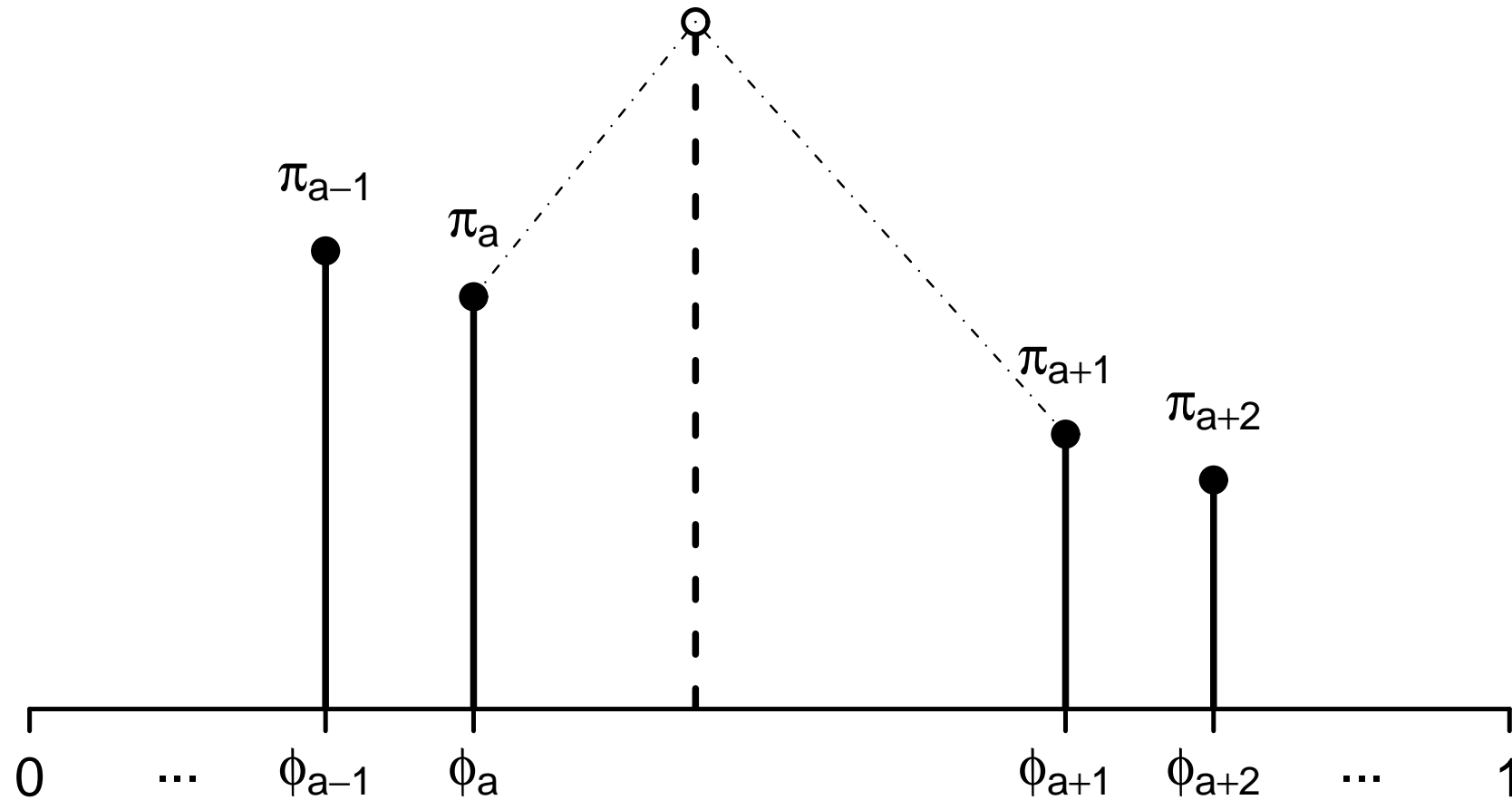
RJMCMC Updates

(3) Split a support point



RJMCMC Updates

(4) Merge two support points



Label switching and poor mixing

Estimation of finite mixture models affected if the MCMC sampler **can't mix** because of

- artificial identifiability constraints, (e.g. so that individuals/samples become persistently misallocated) or
- updating protocols making one mode inaccessible from another.

Label switching and poor mixing

Estimation of finite mixture models affected if the MCMC sampler **can't mix** because of

- artificial identifiability constraints, (e.g. so that individuals/samples become persistently misallocated) or
- updating protocols making one mode inaccessible from another.

In our case:

- Mixture components are labelled by a **single parameter**: have a natural ordering (e.g. preventing misallocations);
- Dimension-switching RJMCMC allows components to be deleted and reappear – the chain moves rapidly around model space.

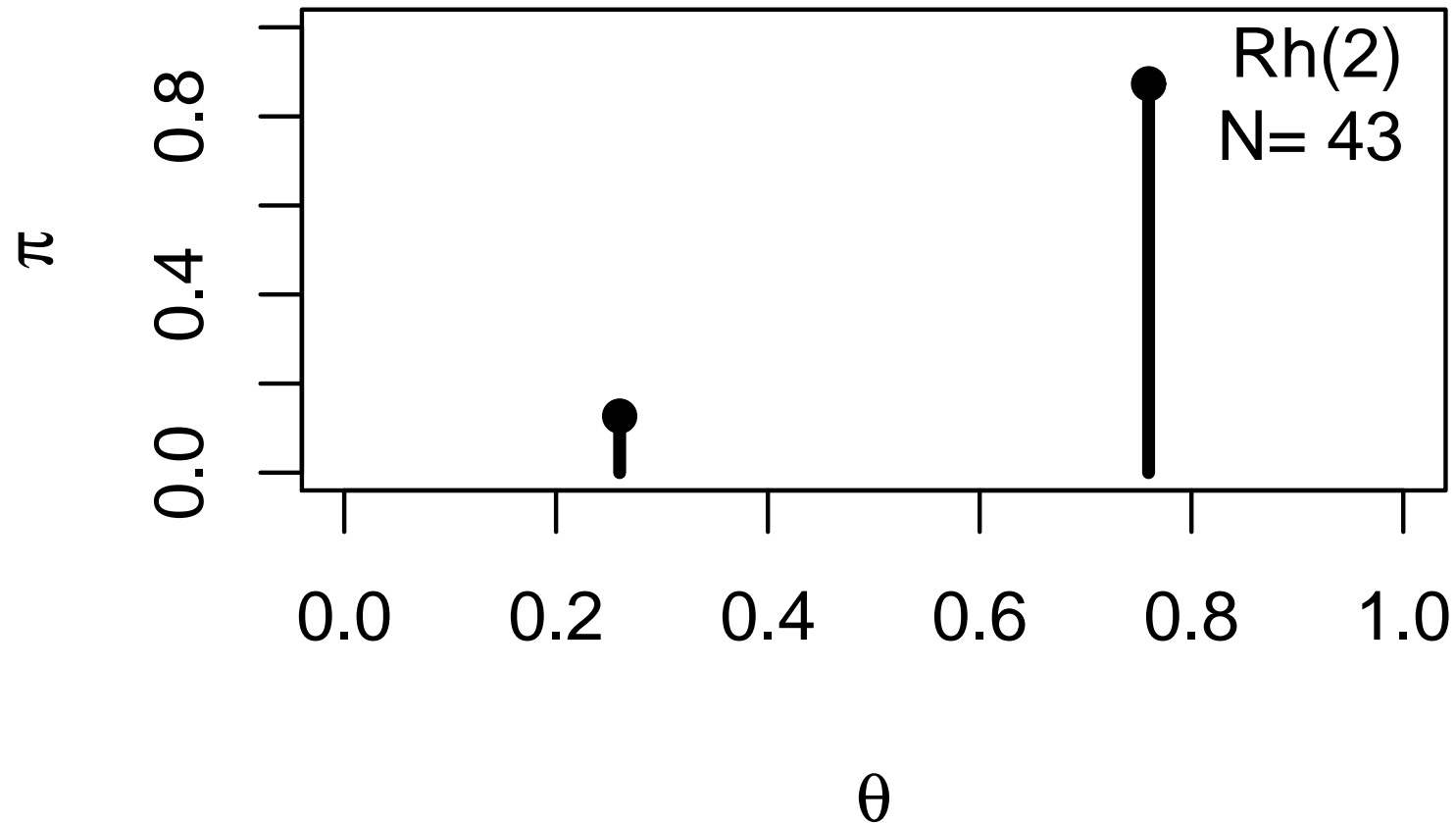
Application: AT&T Switch Testing

$k = 6$ reviewers tested switches and found $D = 43$ errors (Basu 1998)

Error i	Reviewer j						x_i	Error i	Reviewer j						x_i
	1	2	3	4	5	6			1	2	3	4	5	6	
1	1	0	0	0	0	0	1	23	0	0	0	0	0	1	1
2	0	0	0	1	0	1	2	24	0	0	0	0	1	0	1
3	0	0	0	0	1	0	1	25	0	0	0	1	1	0	2
4	1	0	0	0	0	0	1	26	1	0	0	0	0	0	1
5	0	0	0	1	0	0	1	27	0	0	0	0	1	0	1
6	0	0	0	1	0	0	1	28	0	0	1	0	0	0	1
7	0	0	0	1	0	0	1	29	1	0	0	0	0	0	1
8	1	0	0	0	0	0	1	30	1	0	0	1	1	0	3
9	1	0	0	0	0	0	1	31	1	0	0	1	0	0	2
10	0	0	0	1	0	0	1	32	1	0	0	0	0	0	1
11	1	0	0	1	0	0	2	33	1	0	0	0	0	0	1
12	1	0	0	0	0	0	1	34	0	0	1	0	0	0	1
13	1	0	0	1	0	0	2	35	0	0	1	0	0	0	1
14	1	0	0	0	0	0	1	36	0	0	0	0	0	1	1
15	1	0	0	1	0	0	2	37	0	0	1	0	0	0	1
16	1	0	0	1	0	0	2	38	1	0	0	0	0	1	2
17	1	1	0	1	1	1	5	39	1	0	0	0	0	0	1
18	1	0	0	0	0	1	2	40	1	0	0	0	0	0	1
19	0	1	0	0	0	0	1	41	0	0	0	0	1	0	1
20	0	1	0	0	0	0	1	42	1	0	0	0	0	0	1
21	1	0	0	1	0	0	2	43	1	0	0	0	0	0	1
22	1	0	0	1	0	0	2	n_j	25	3	4	15	7	6	60

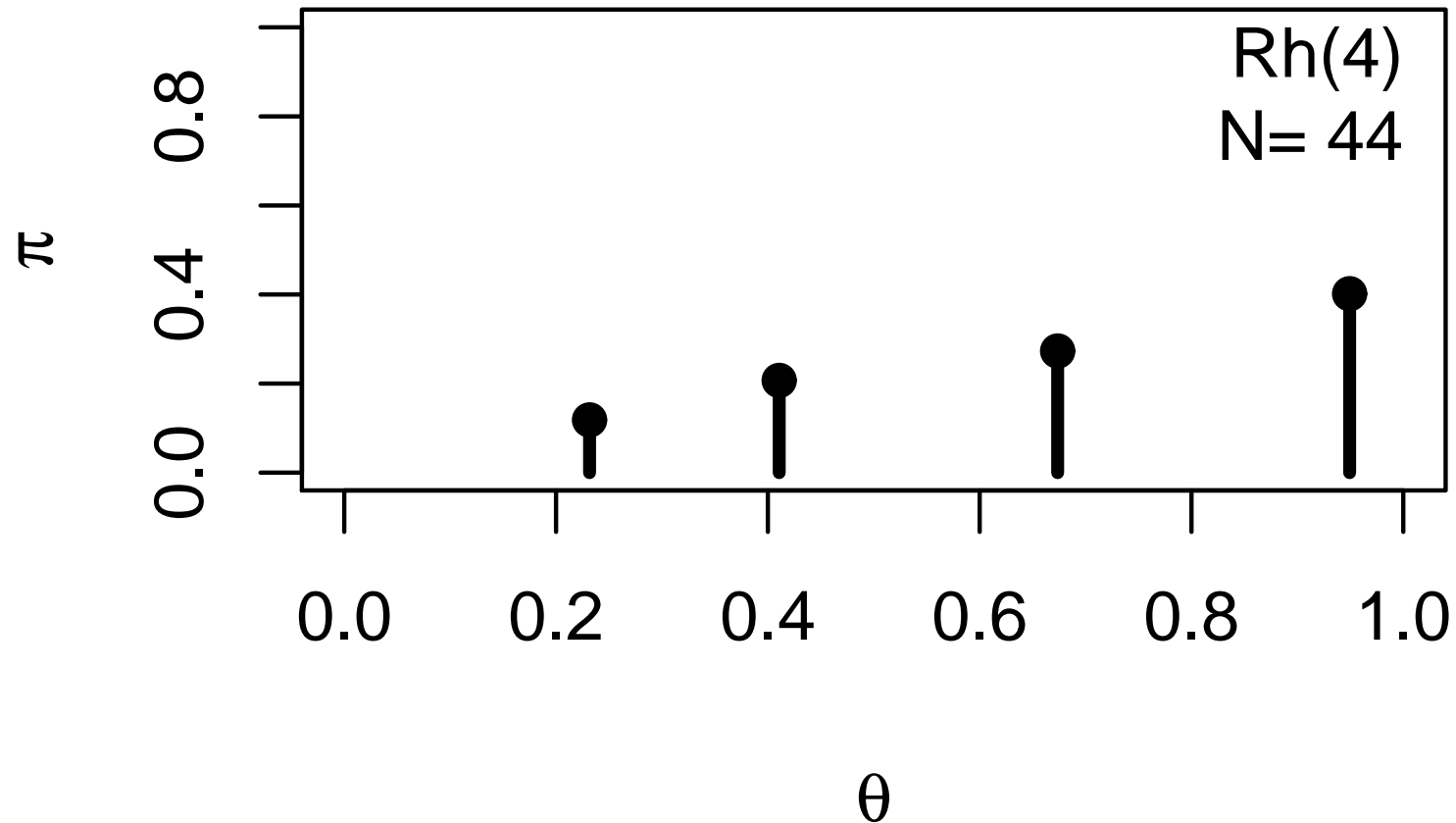
Step 1

$$\log L = -91.9596431614535$$



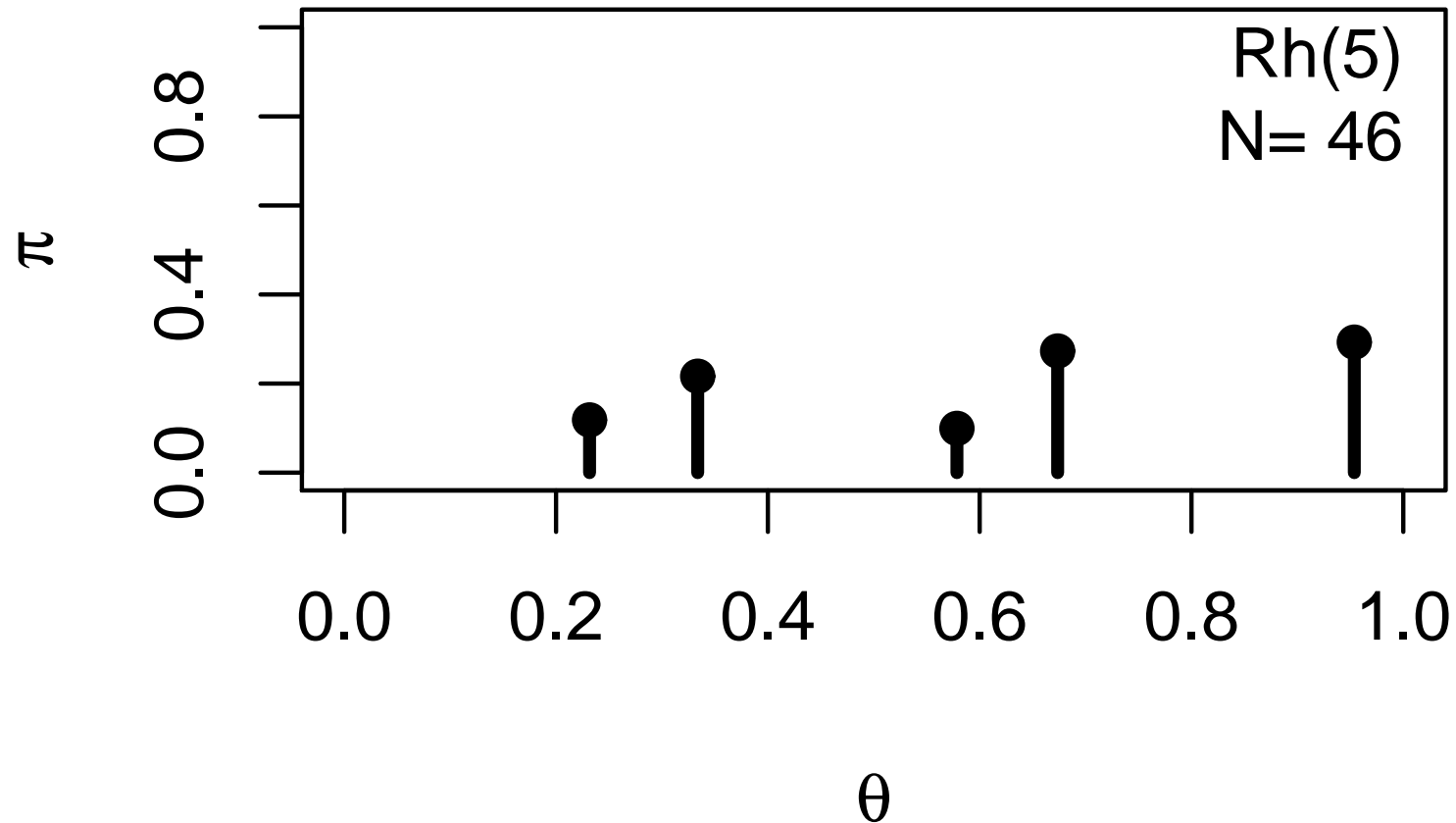
Step 2

$$\log L = -67.9229646689119$$



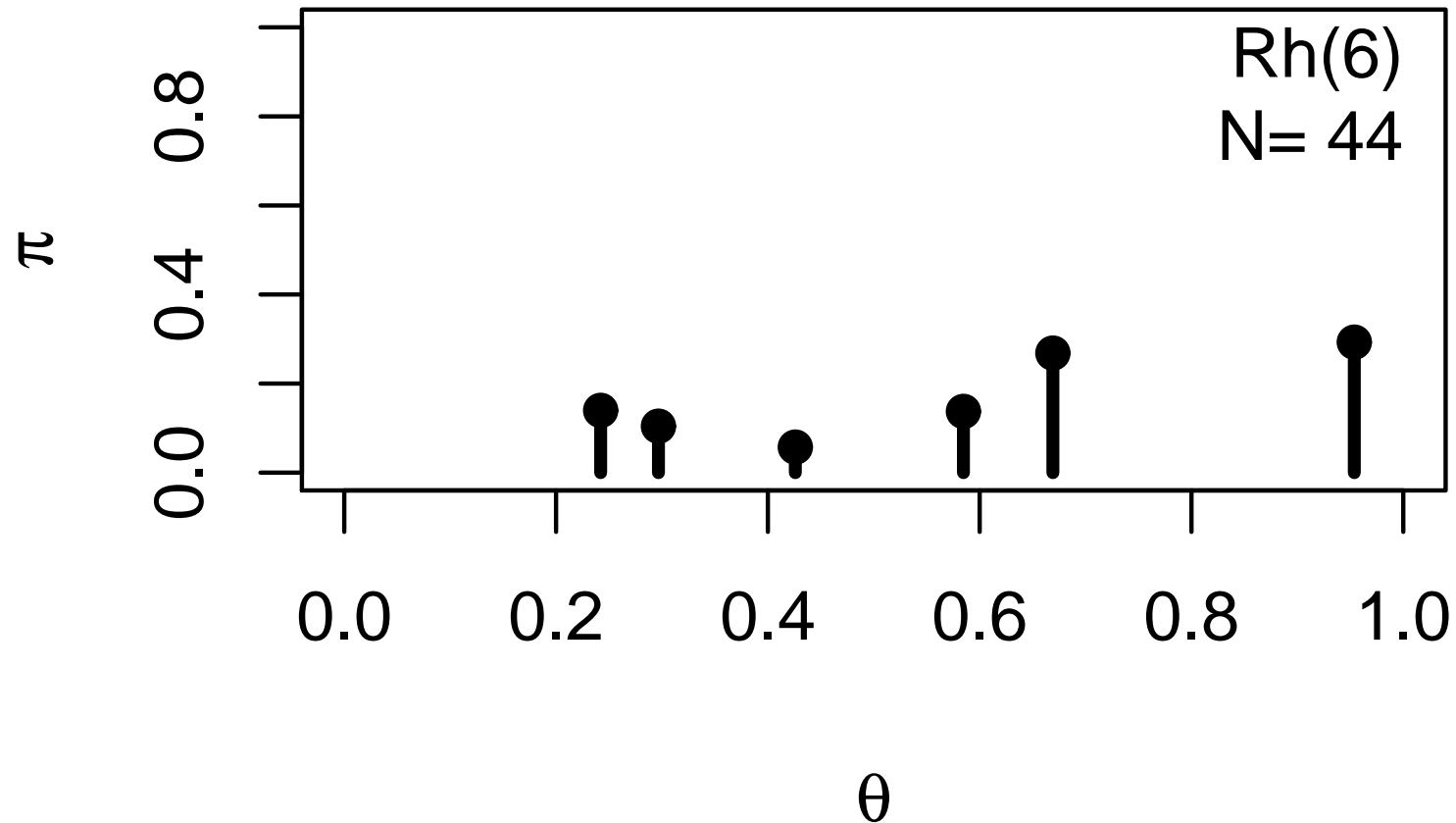
Step 3

$$\log L = -58.1418394830594$$



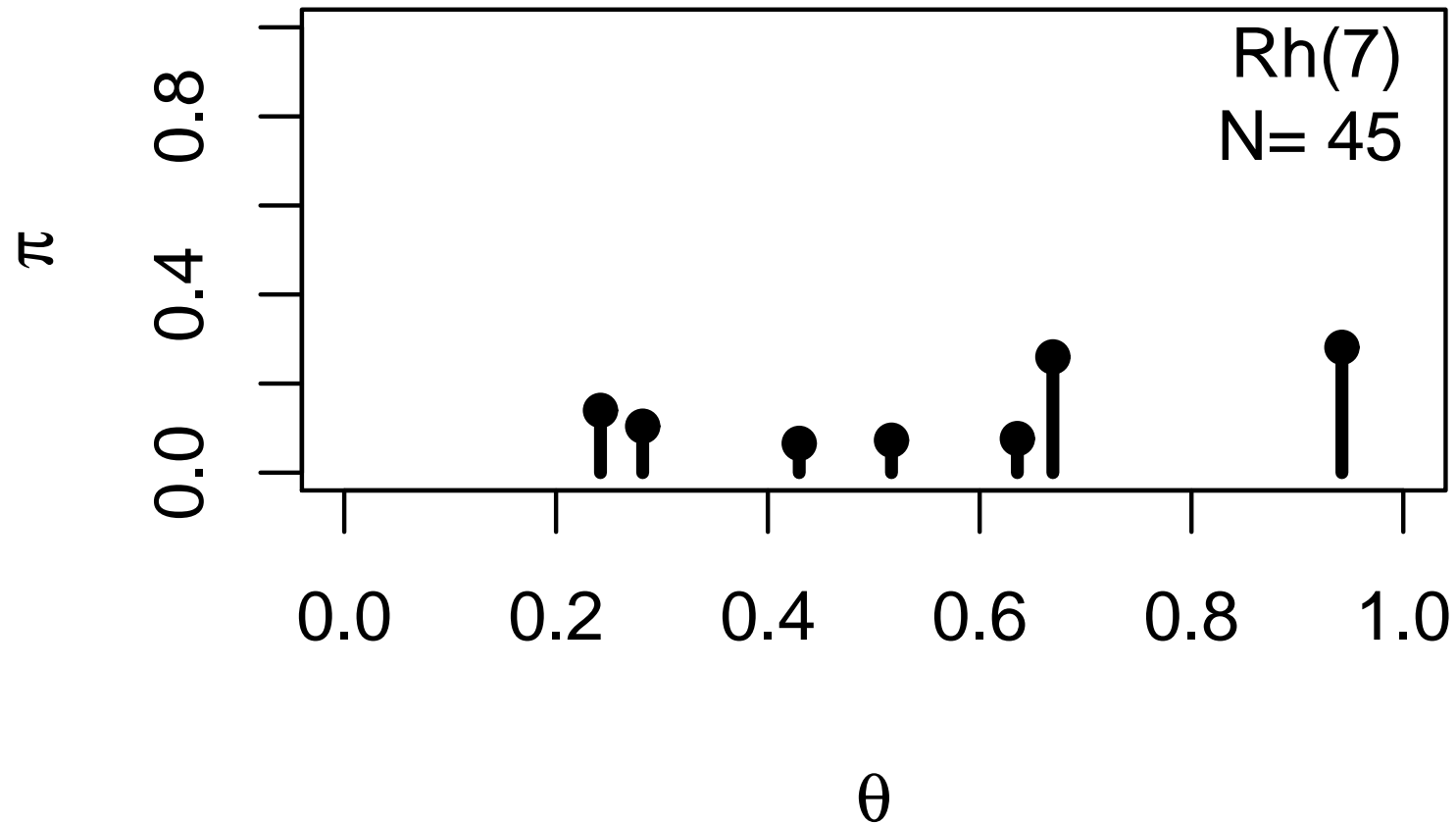
Step 4

$$\log L = -60.4784899550032$$



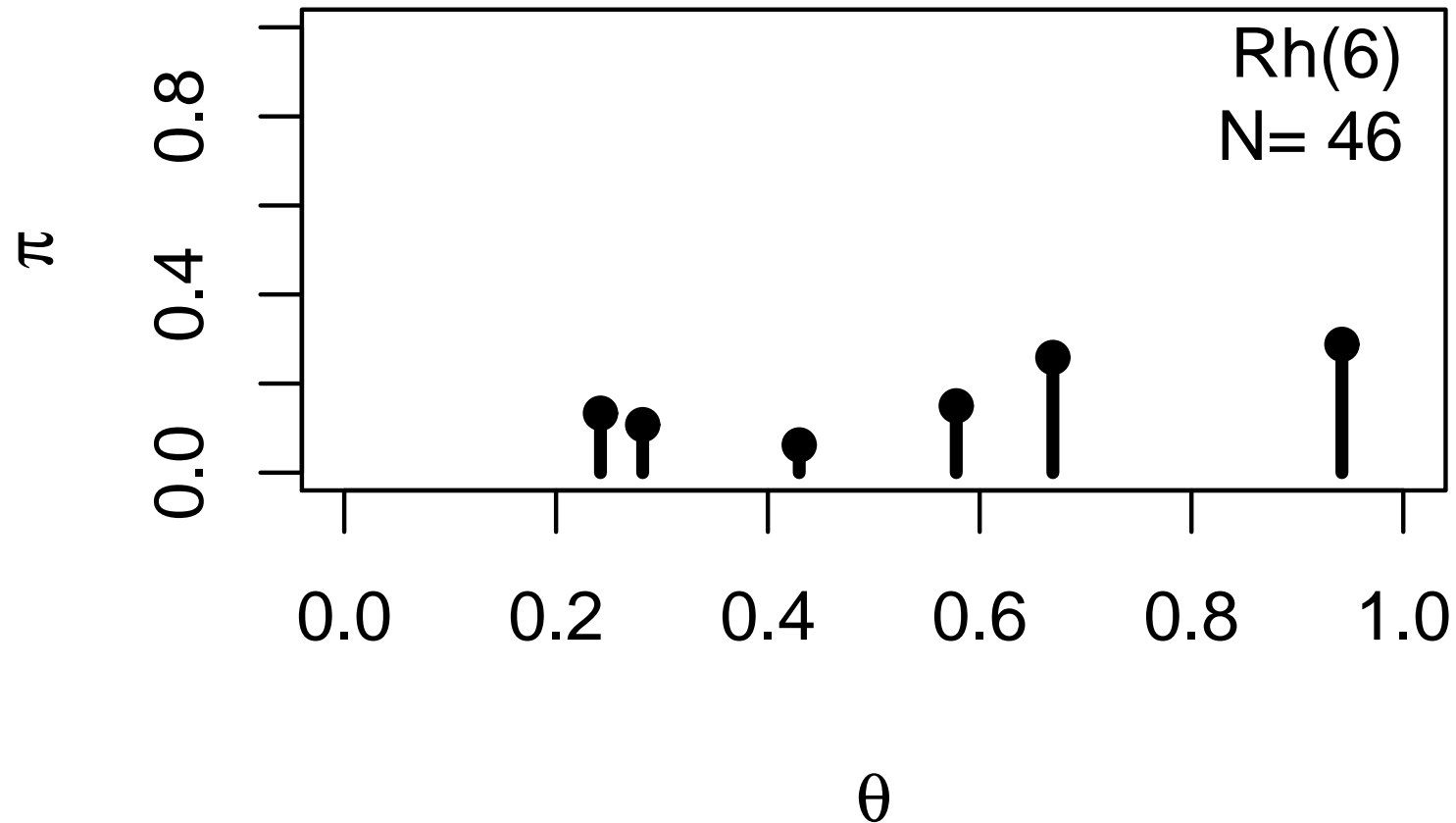
Step 5

$$\log L = -58.7074675287983$$



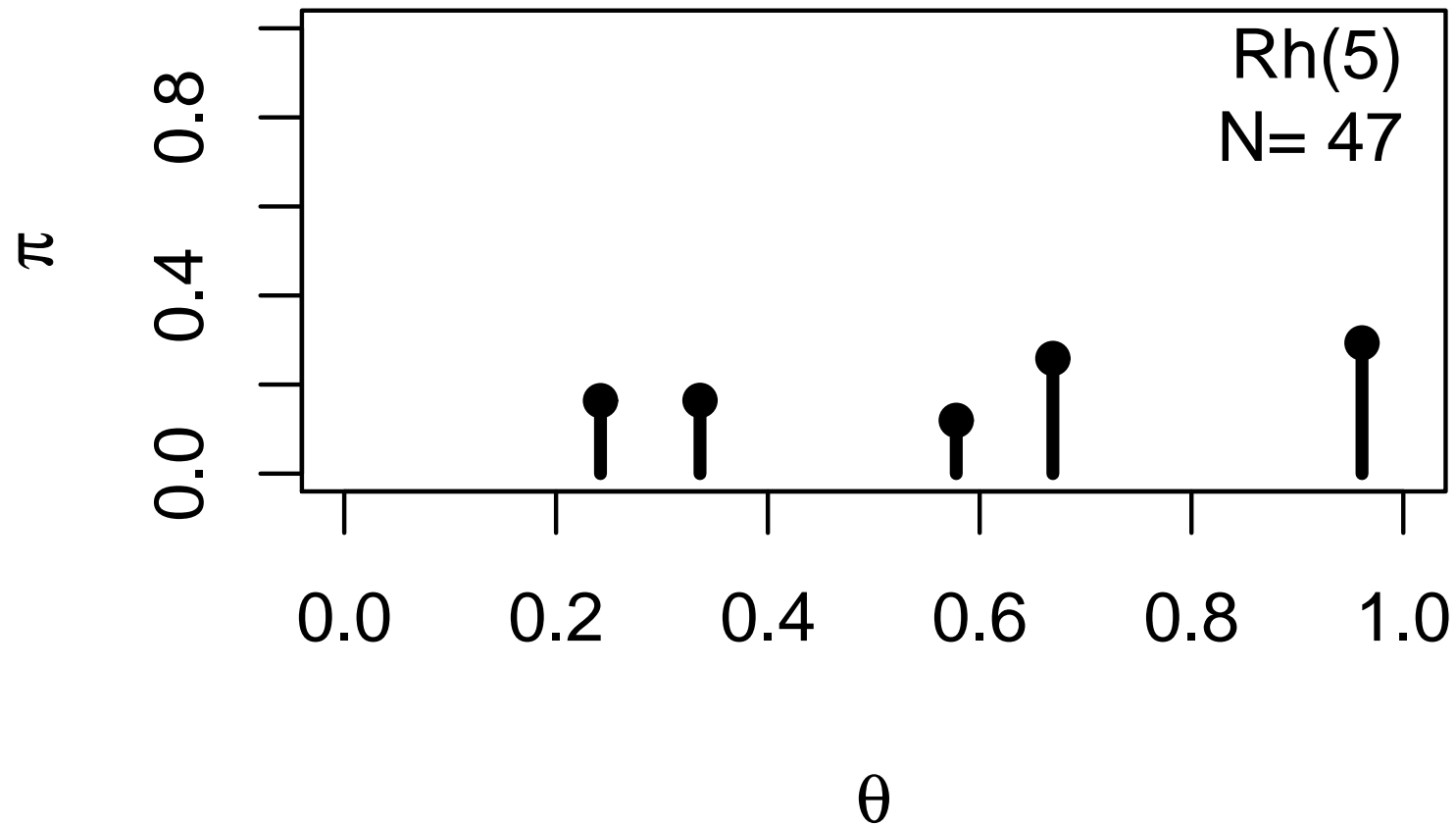
Step 6

$$\log L = -60.0611555129325$$



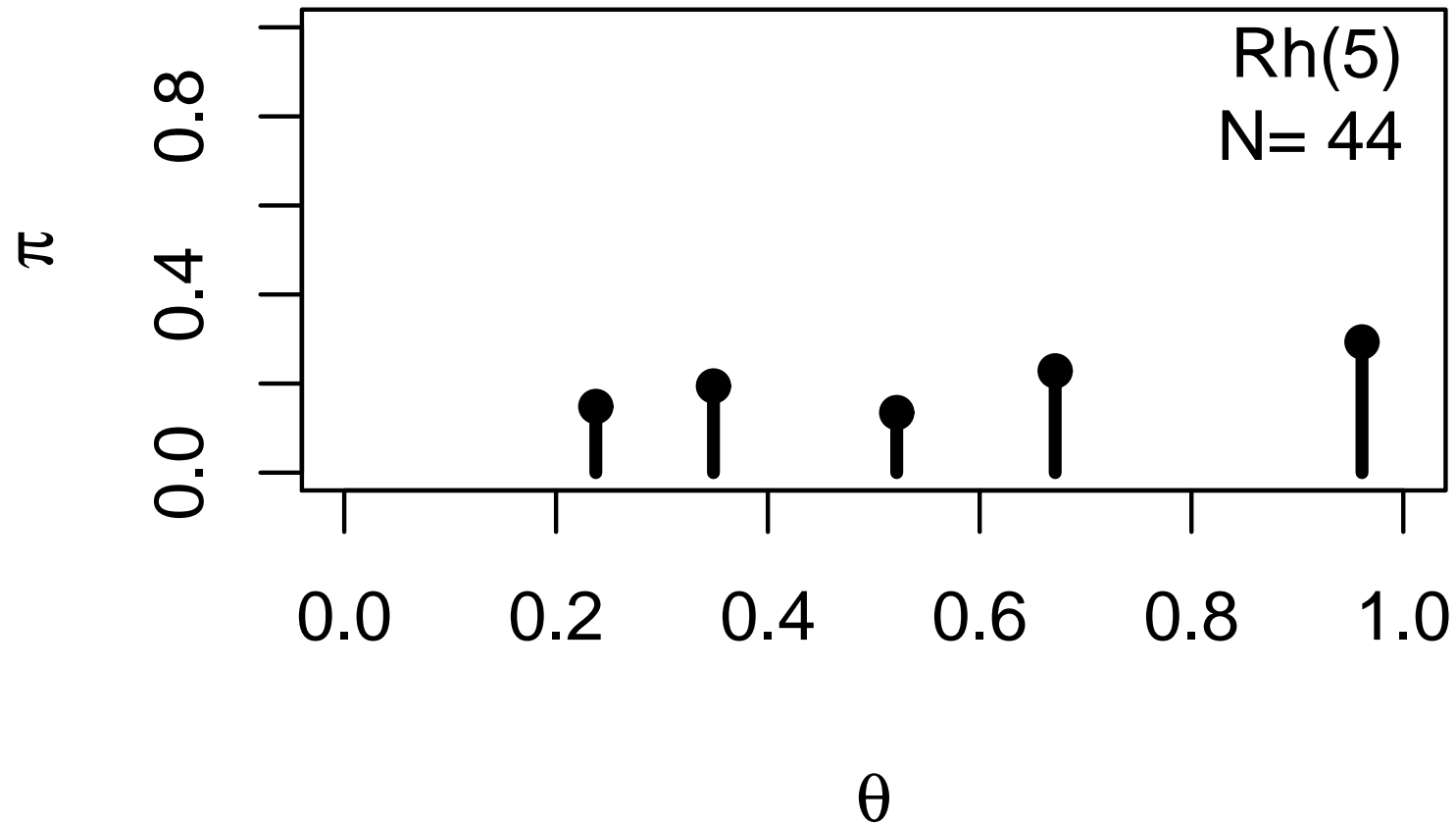
Step 7

$$\log L = -58.0130422680327$$



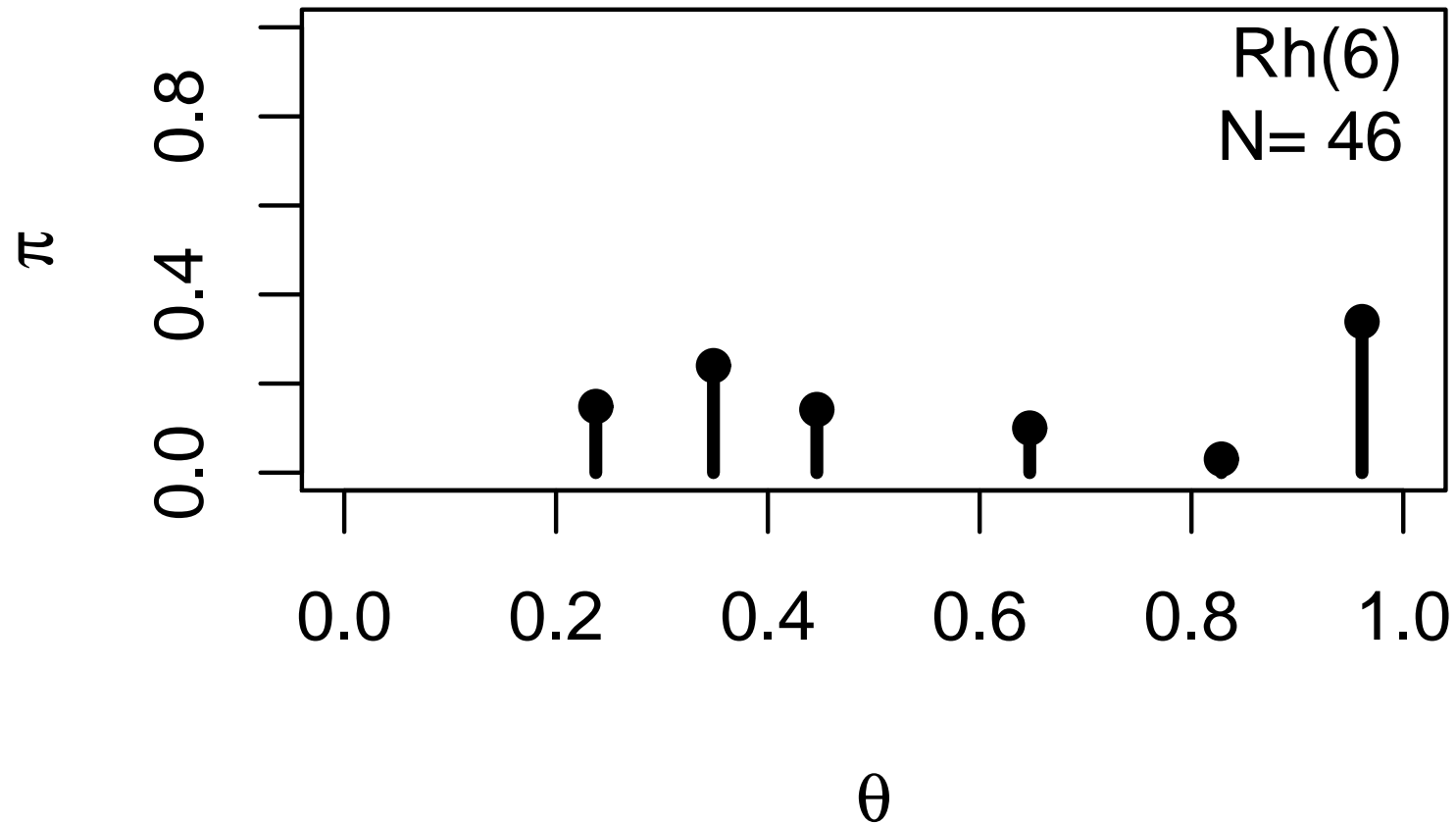
Step 8

$$\log L = -55.1097509120916$$



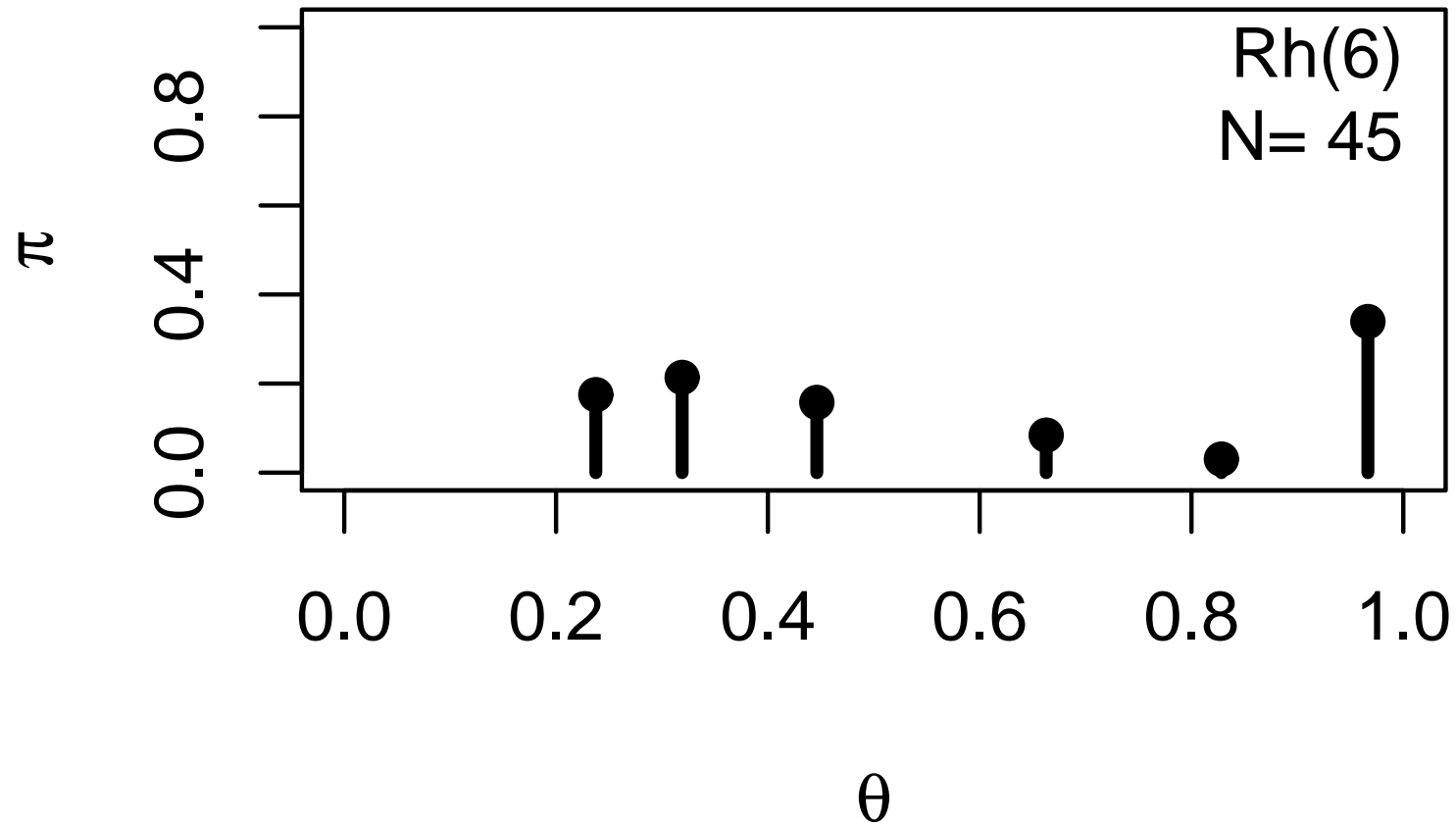
Step 9

$$\log L = -49.5727659877251$$



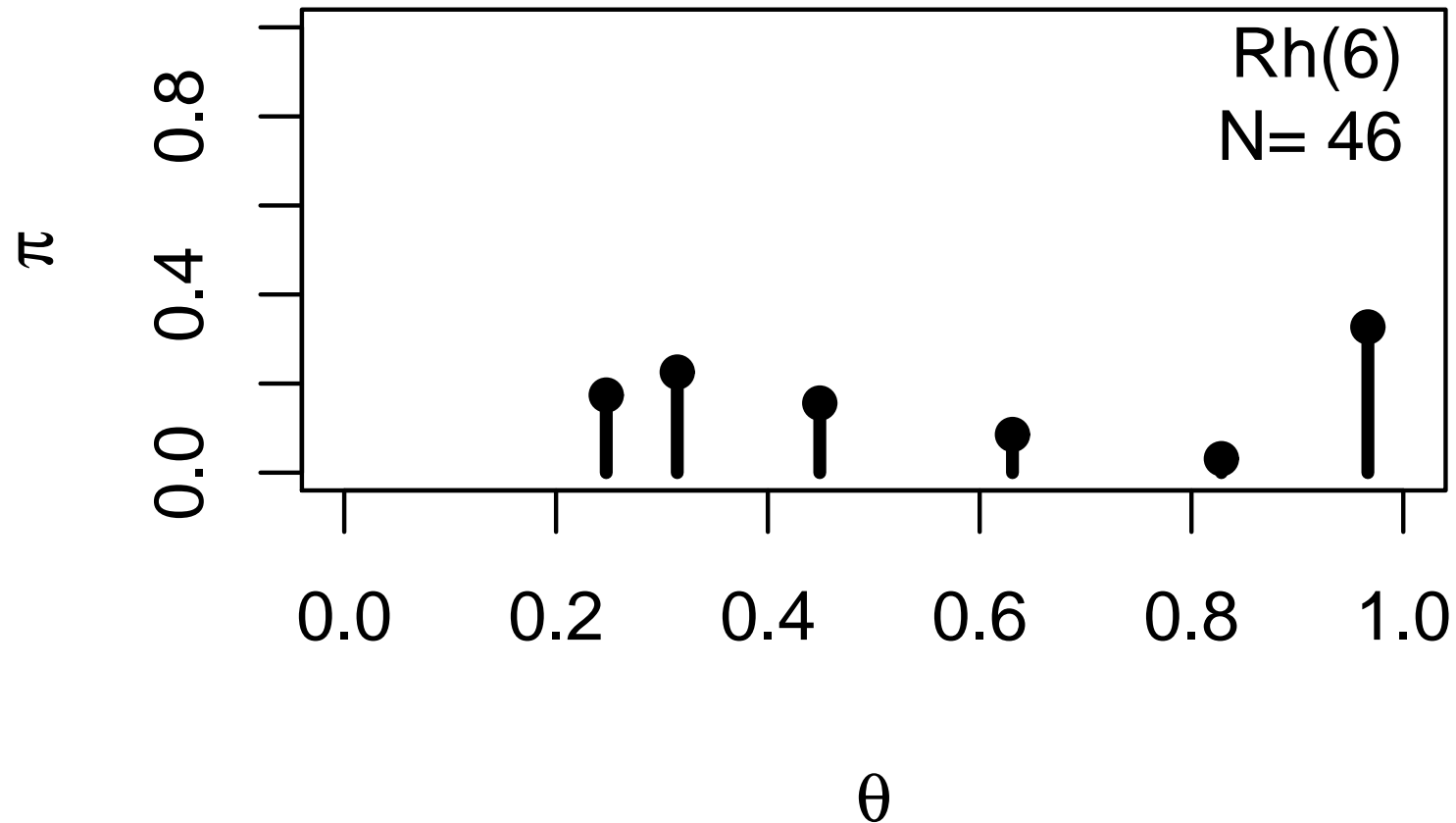
Step 10

$$\log L = -46.5488986935539$$



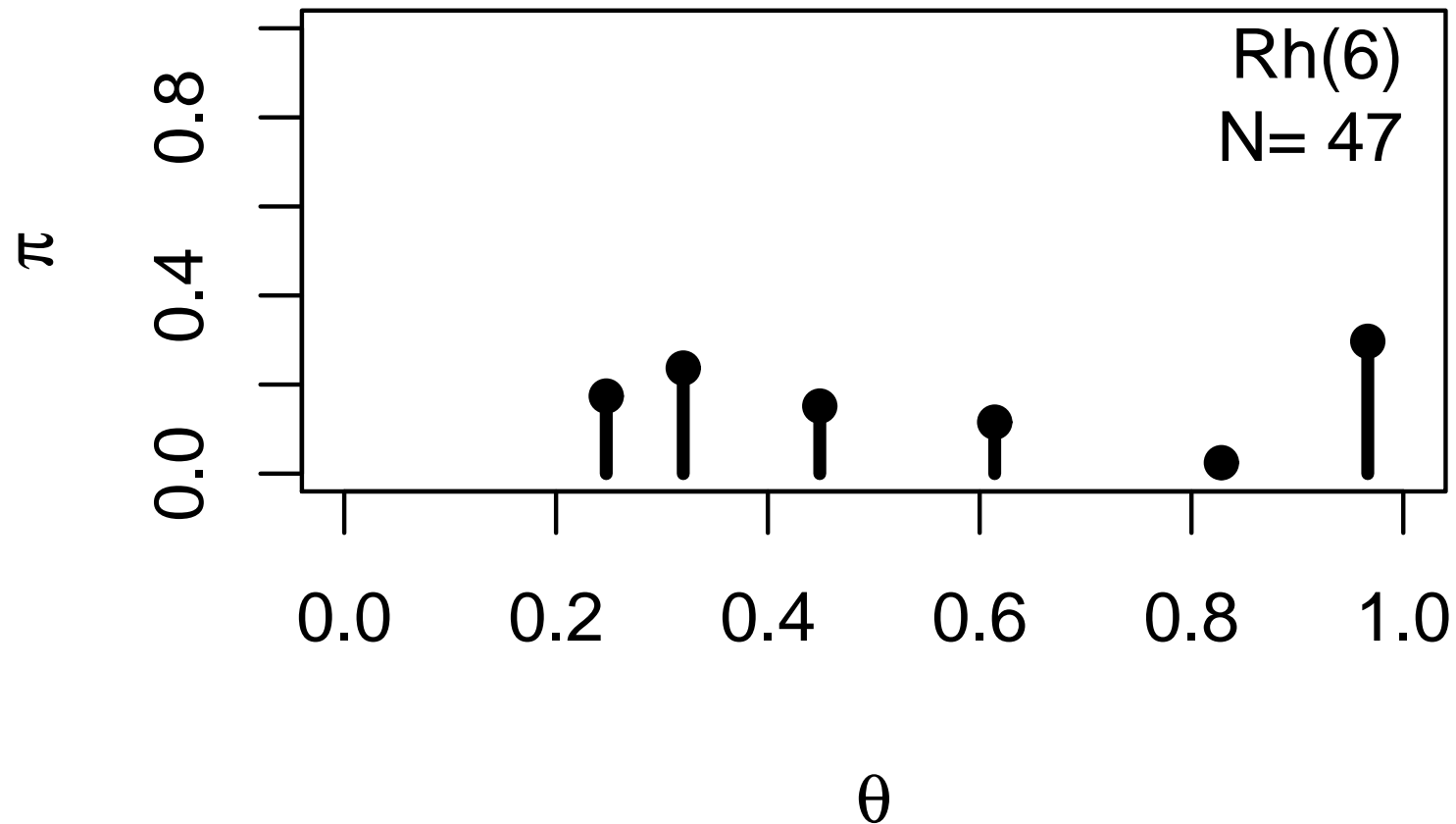
Step 11

$$\log L = -45.664424798022$$



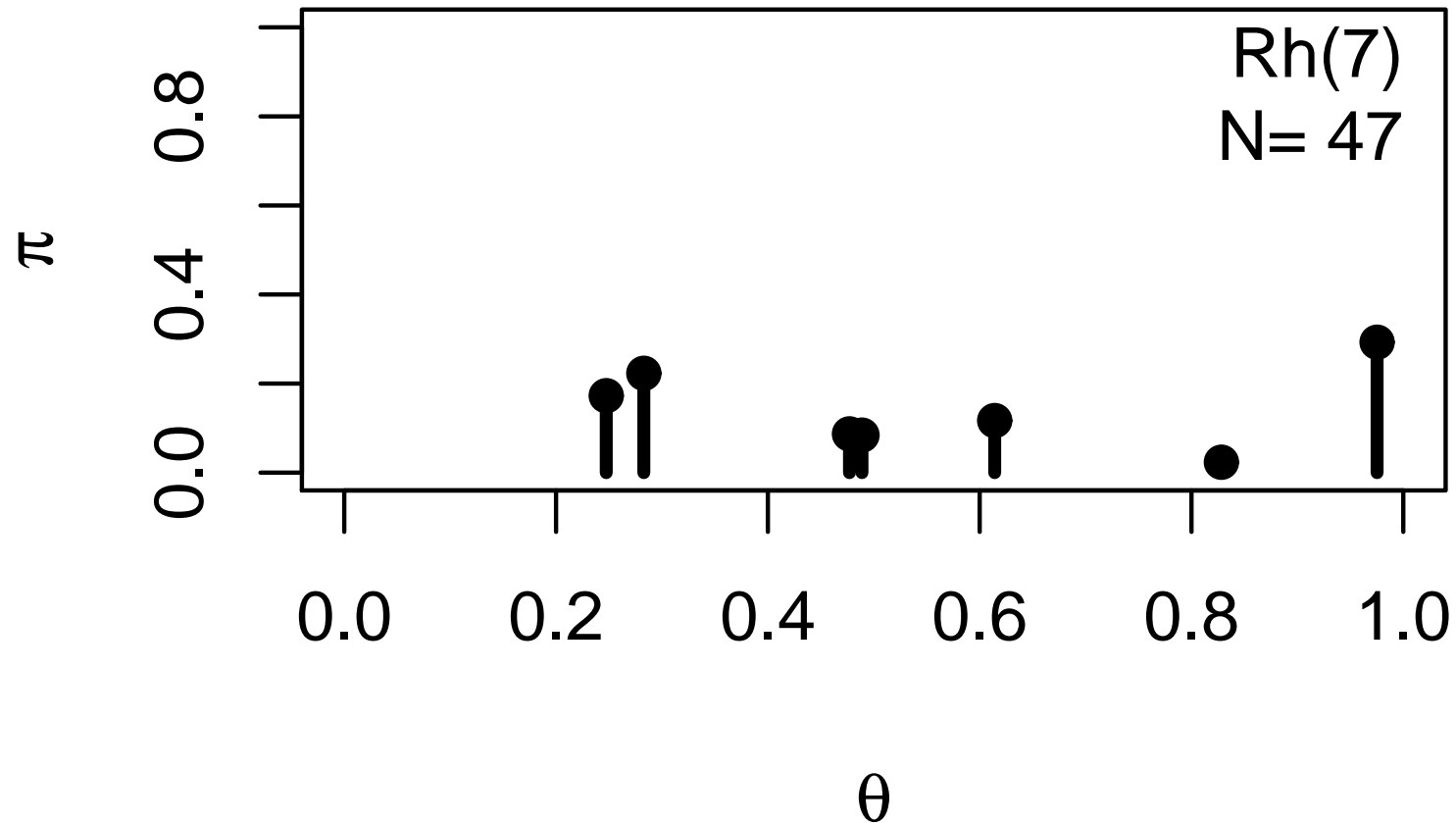
Step 12

$$\log L = -45.0363060198825$$



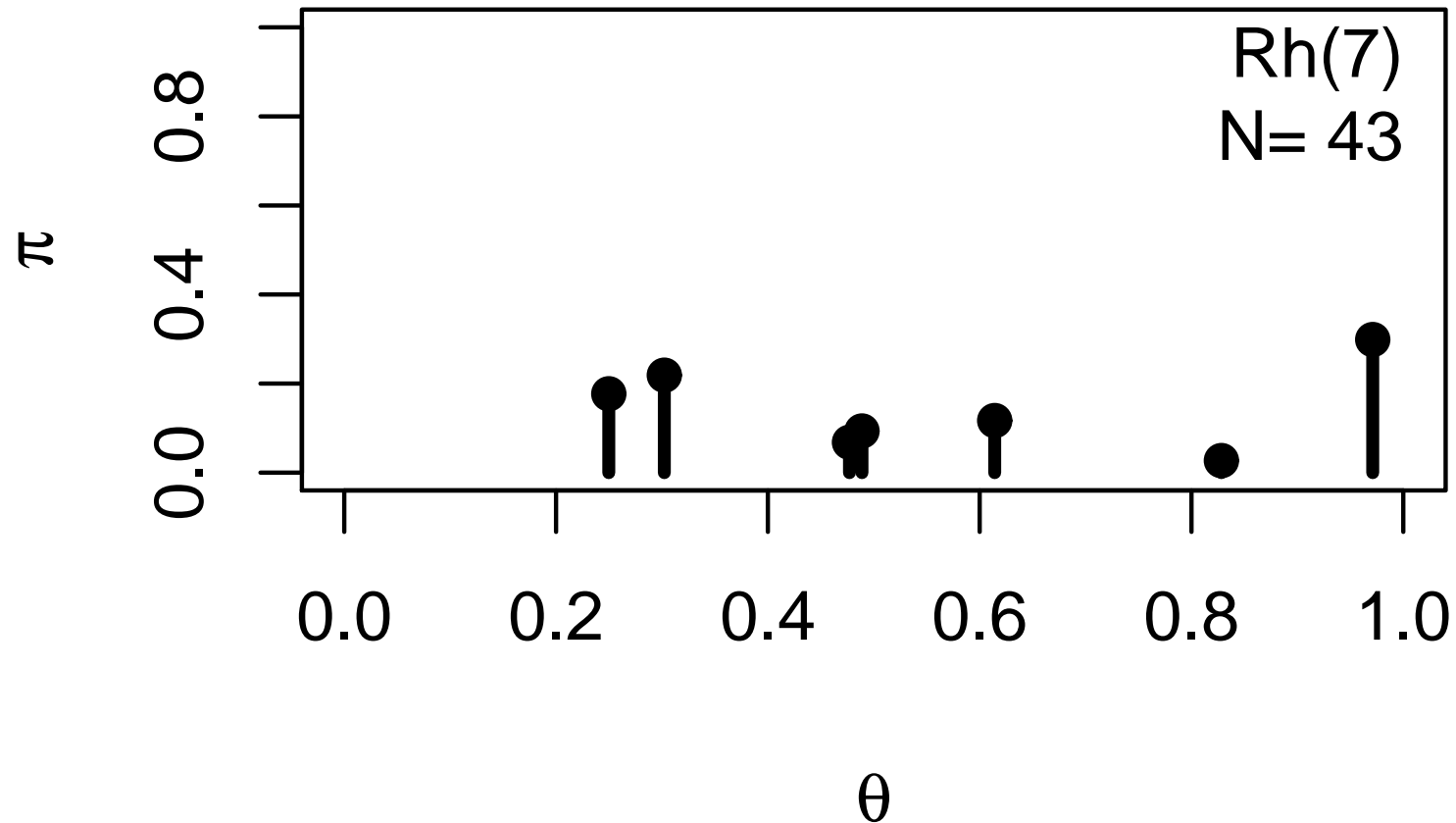
Step 13

$$\log L = -44.6636665499824$$



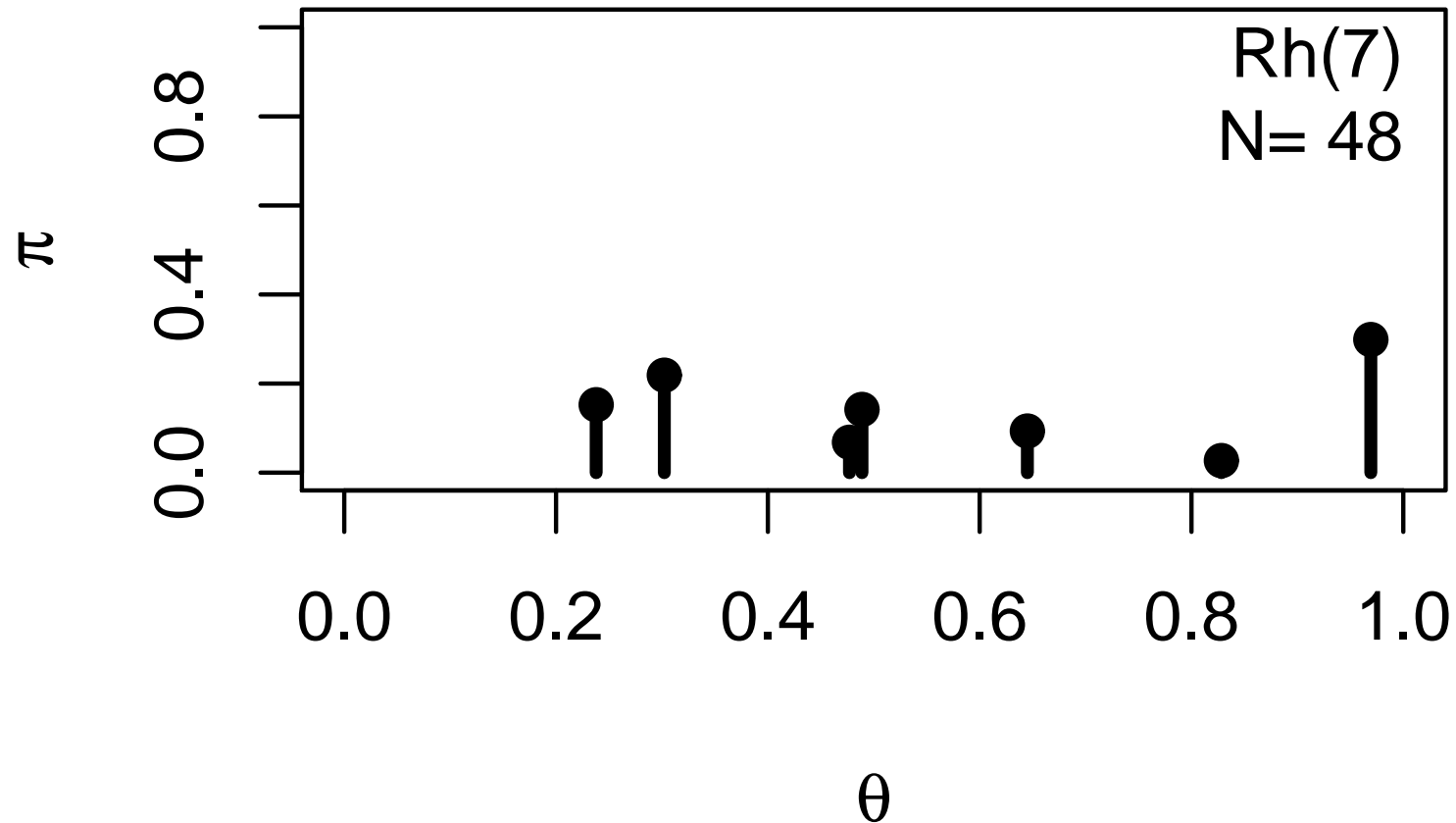
Step 14

$$\log L = -47.0076330888468$$



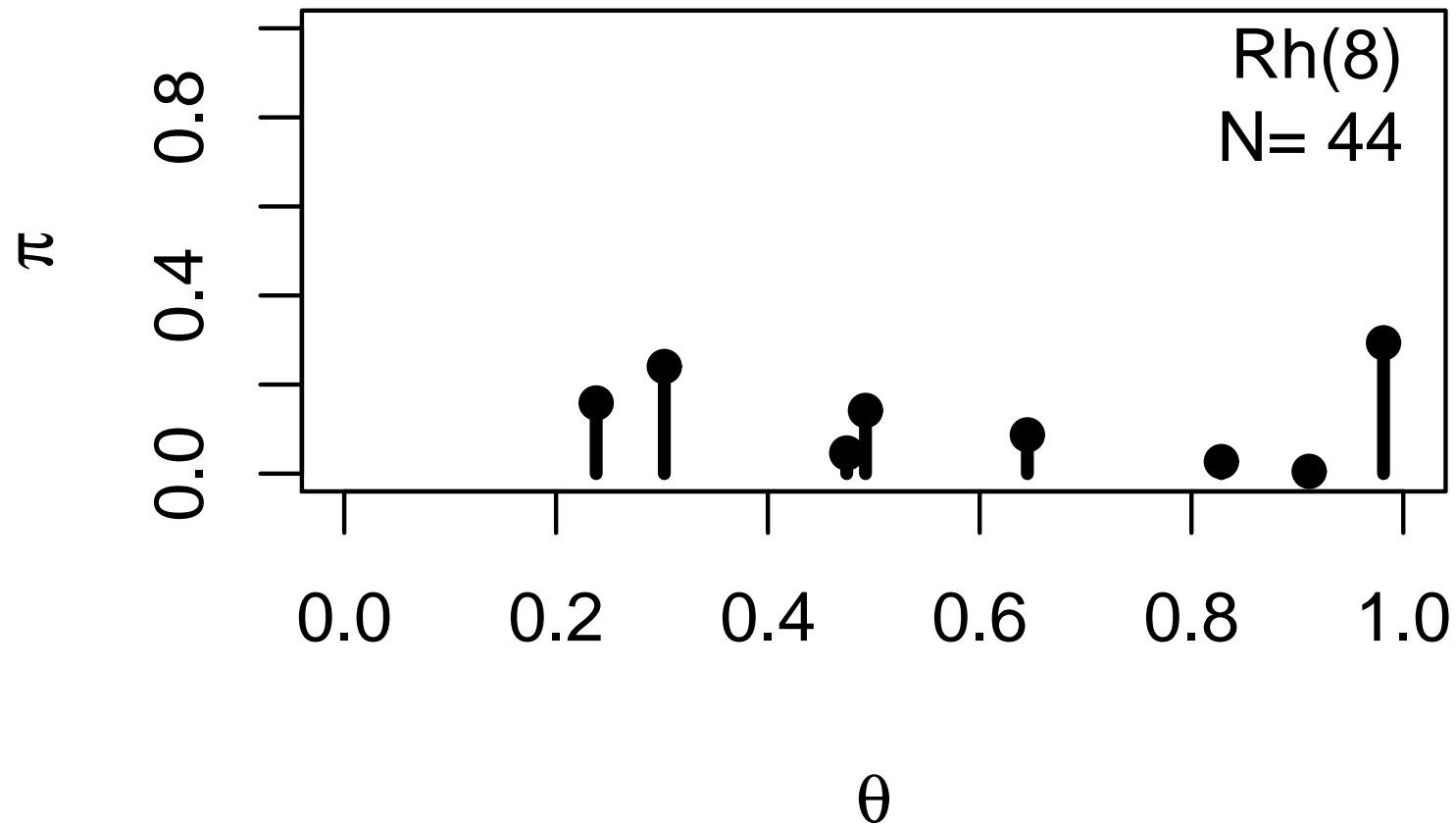
Step 15

$$\log L = -47.6636985662387$$



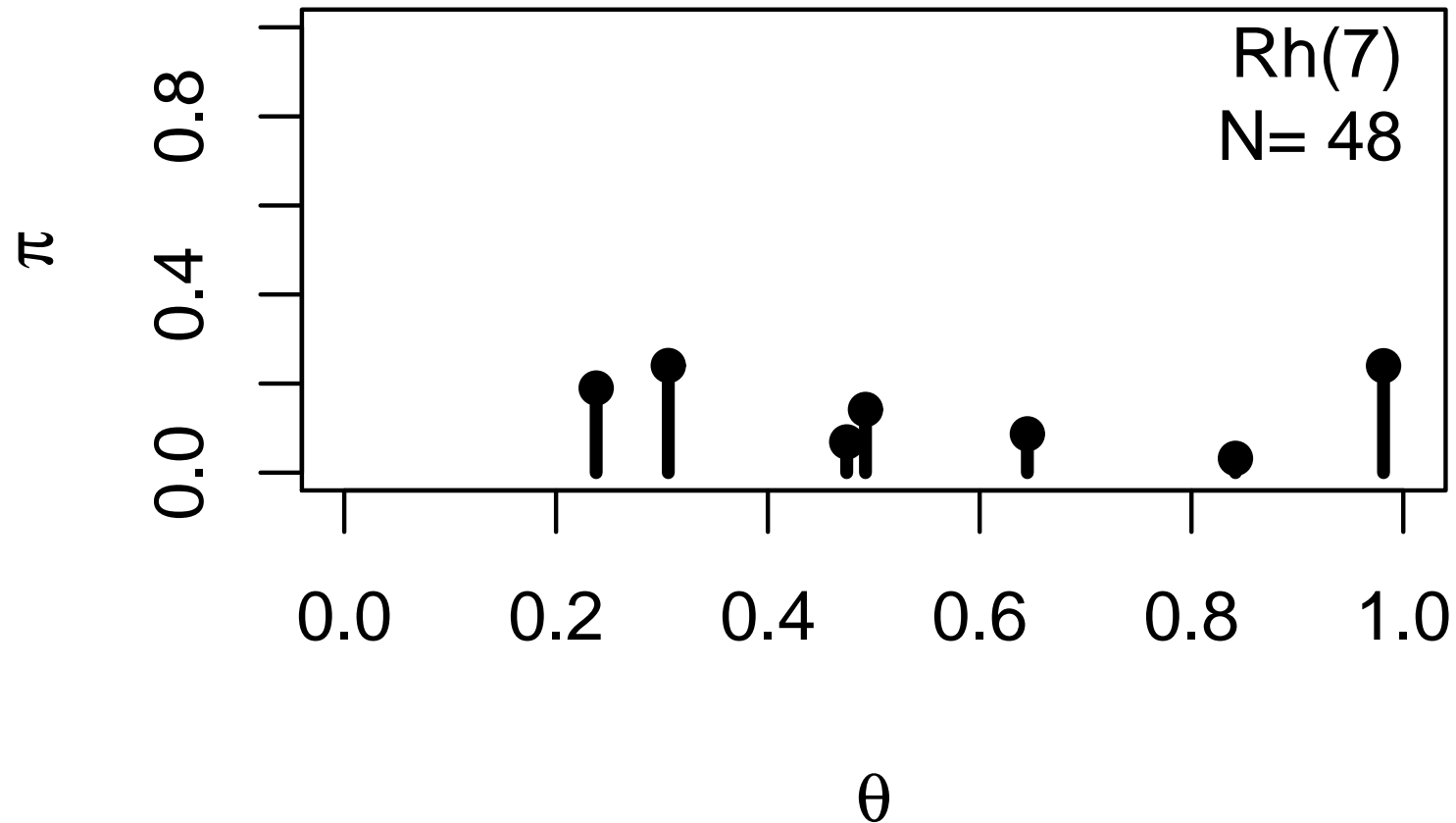
Step 16

$$\log L = -45.737454851479$$



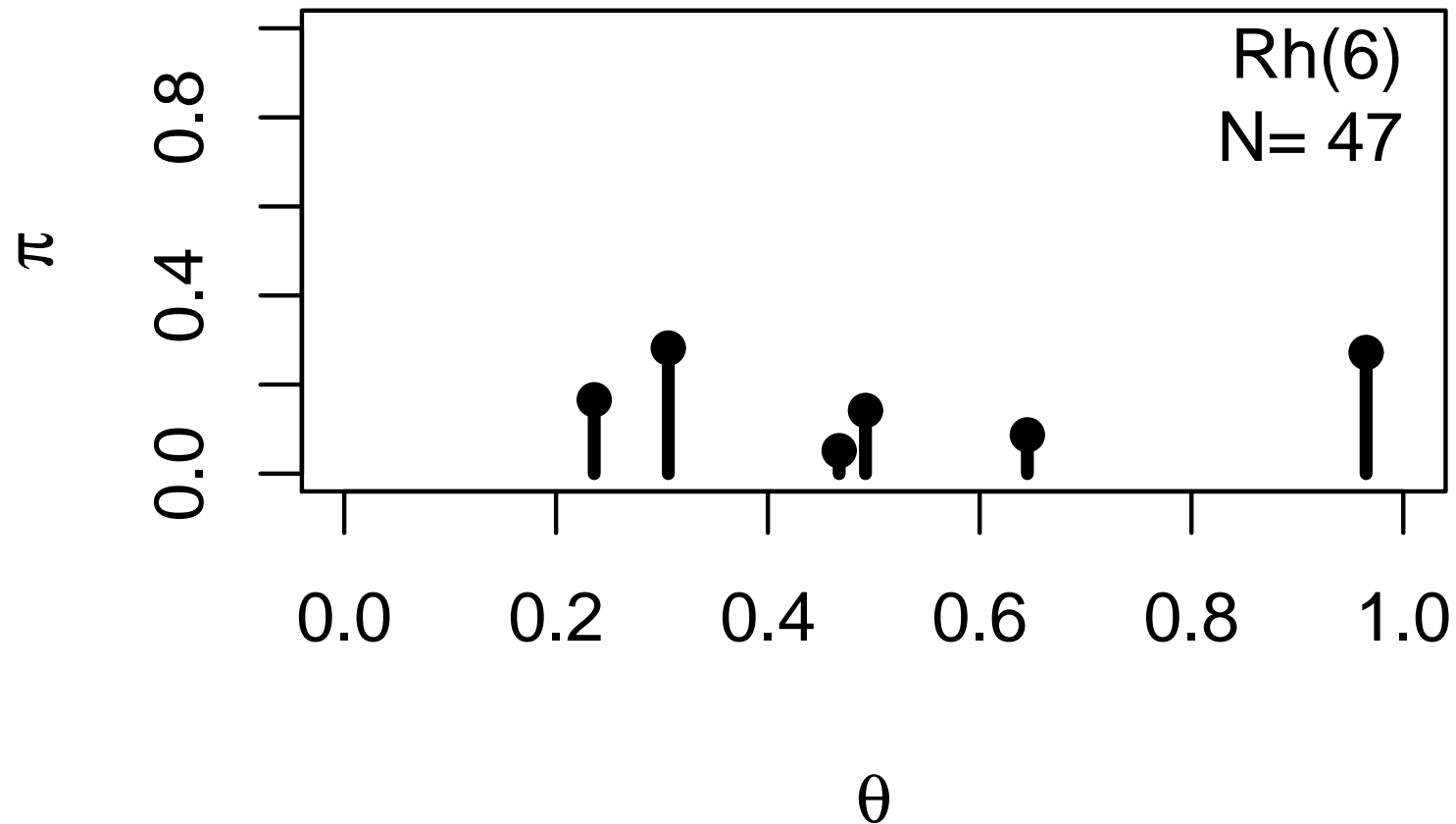
Step 17

$$\log L = -42.4371645695473$$



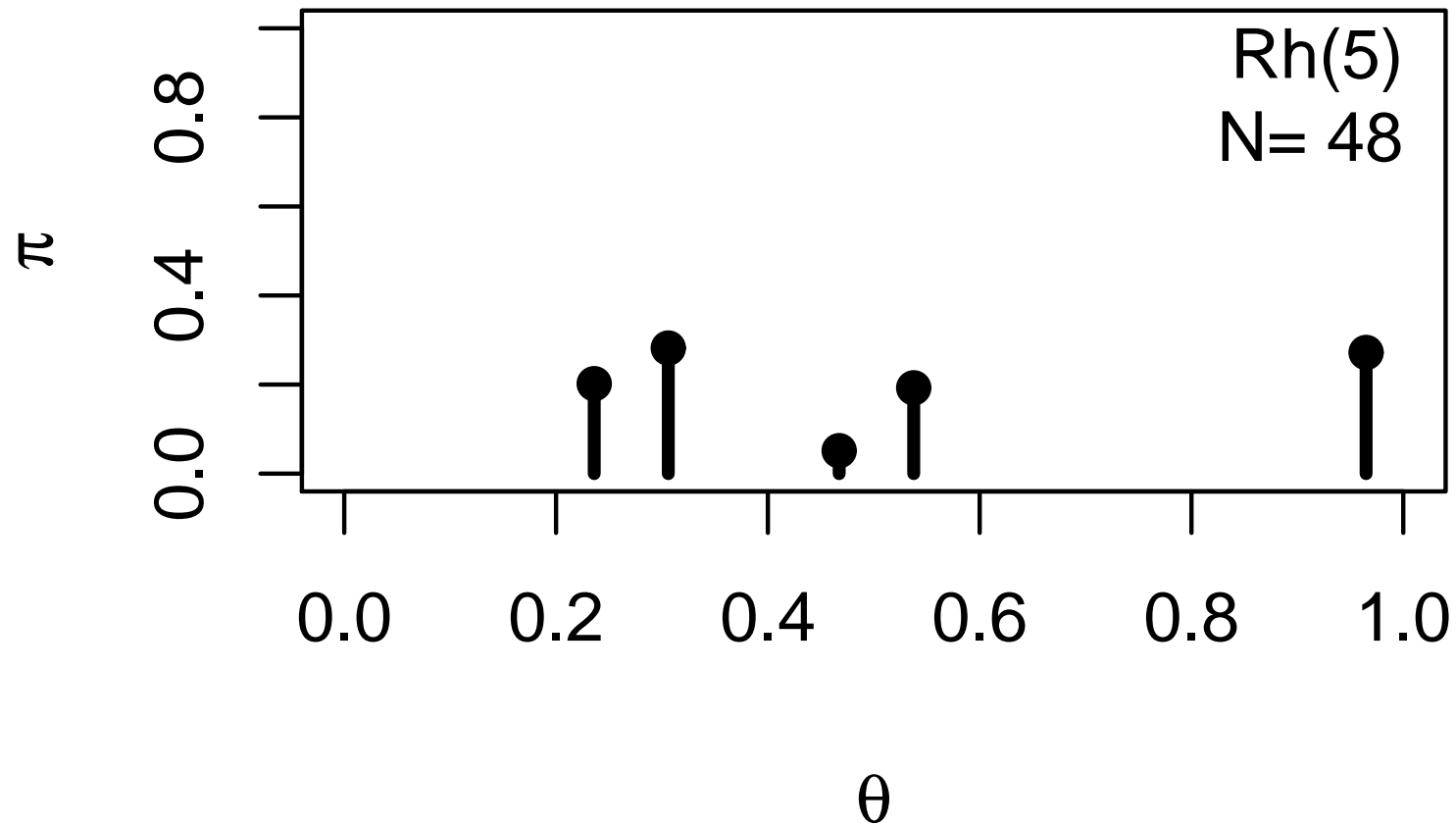
Step 18

$$\log L = -41.4345231830313$$



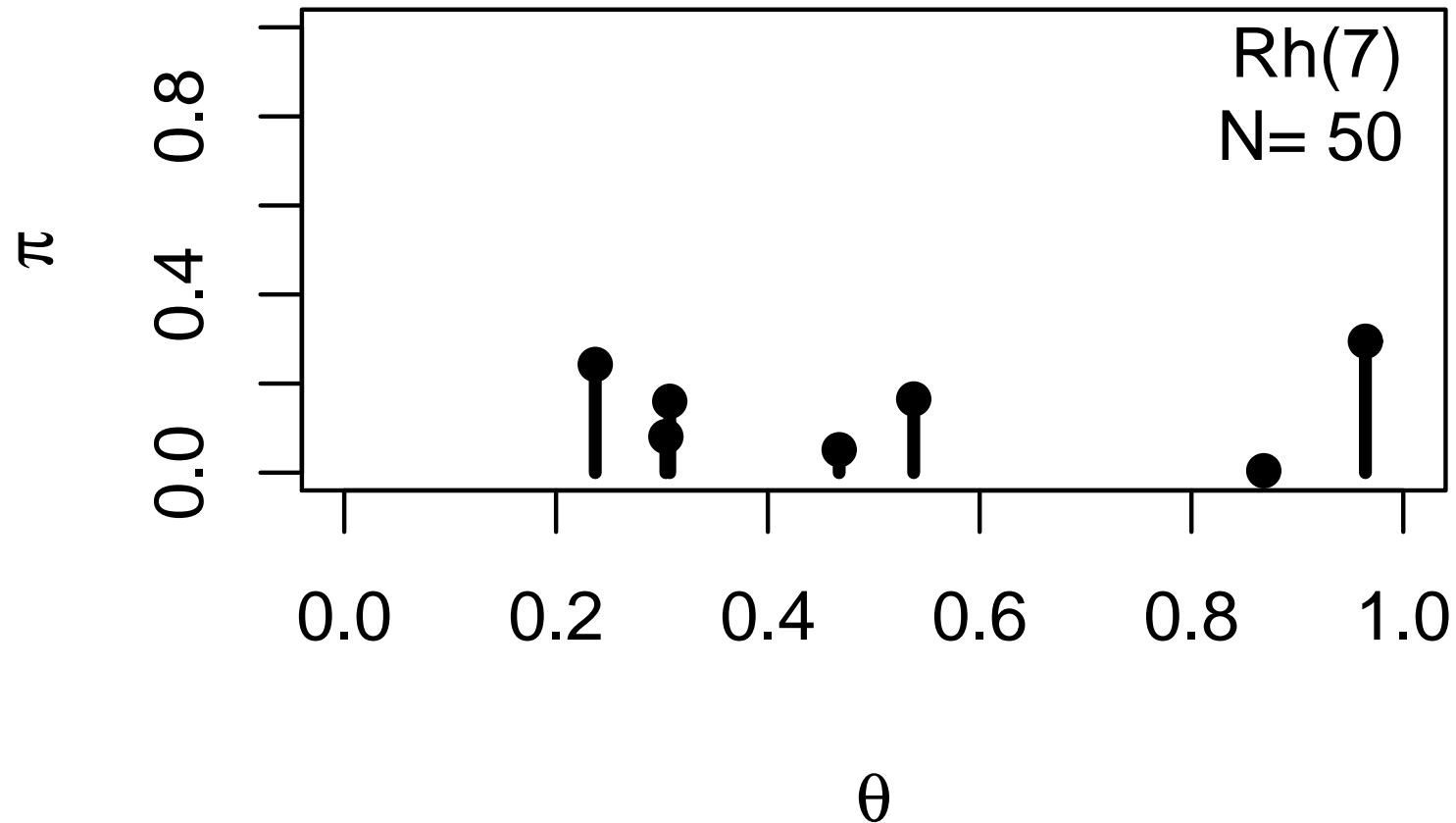
Step 19

$$\log L = -39.3970709062431$$



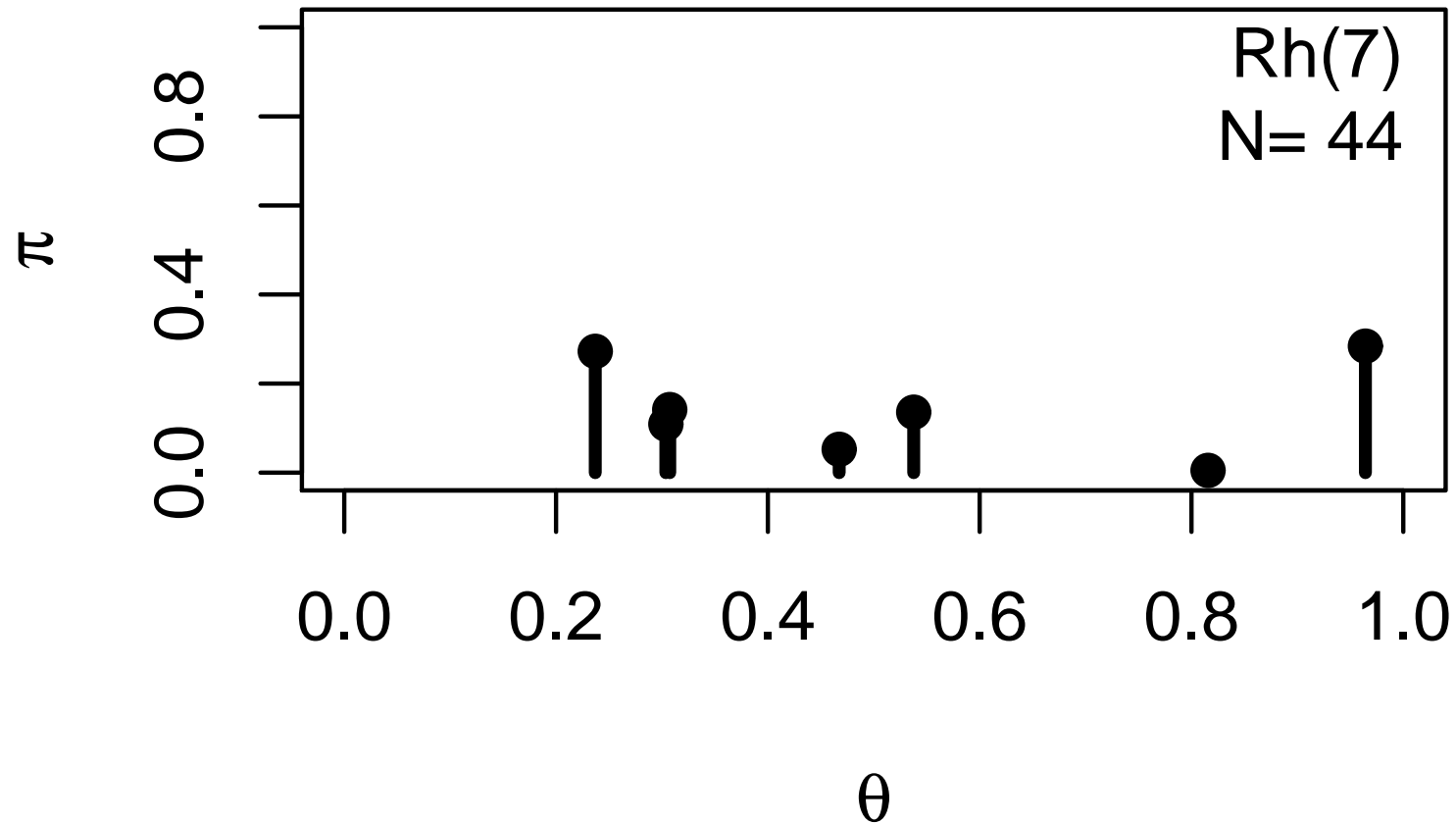
Step 20

$\log L = -40.5227716139258$



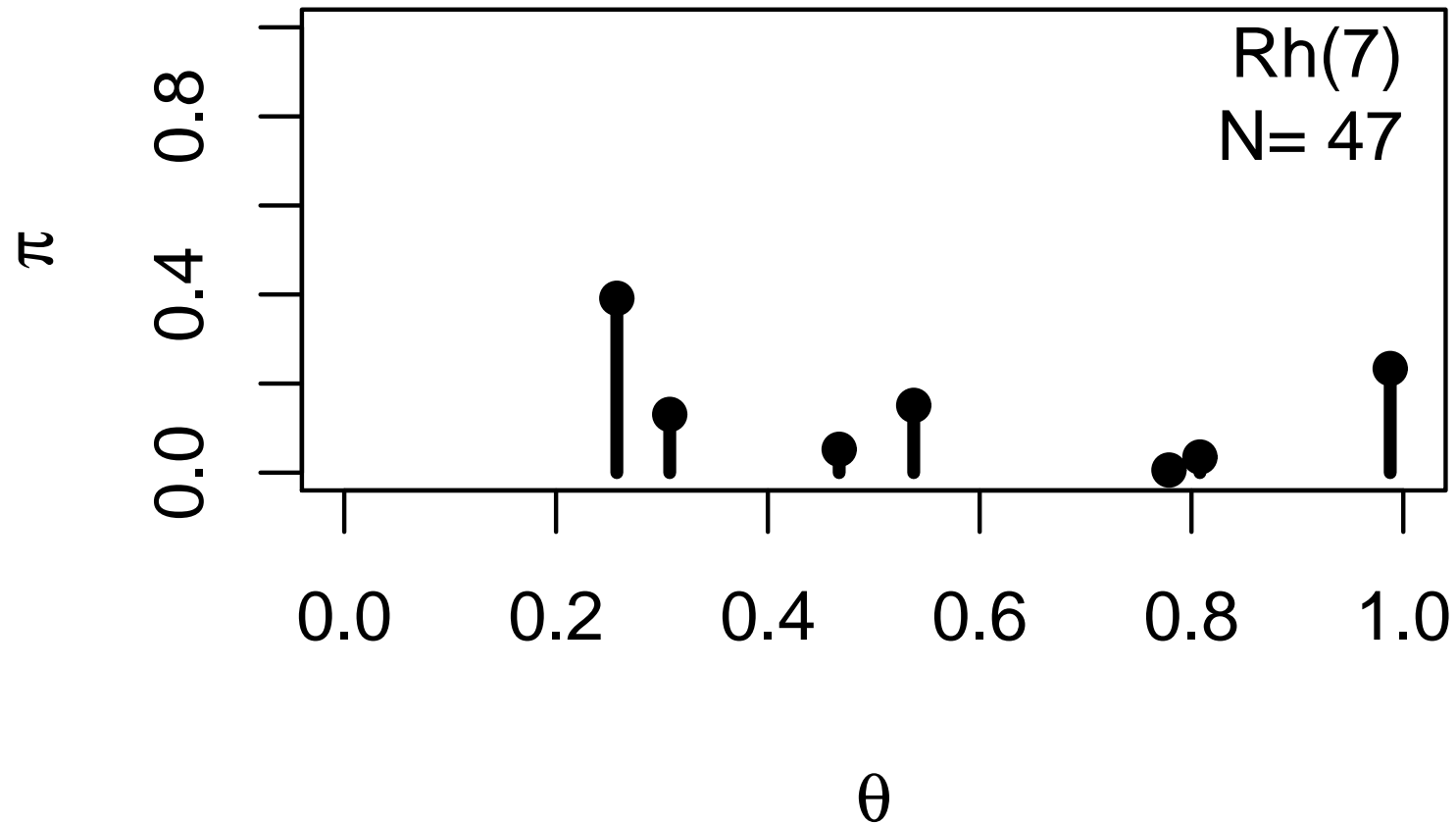
Step 21

$$\log L = -37.5693882100671$$



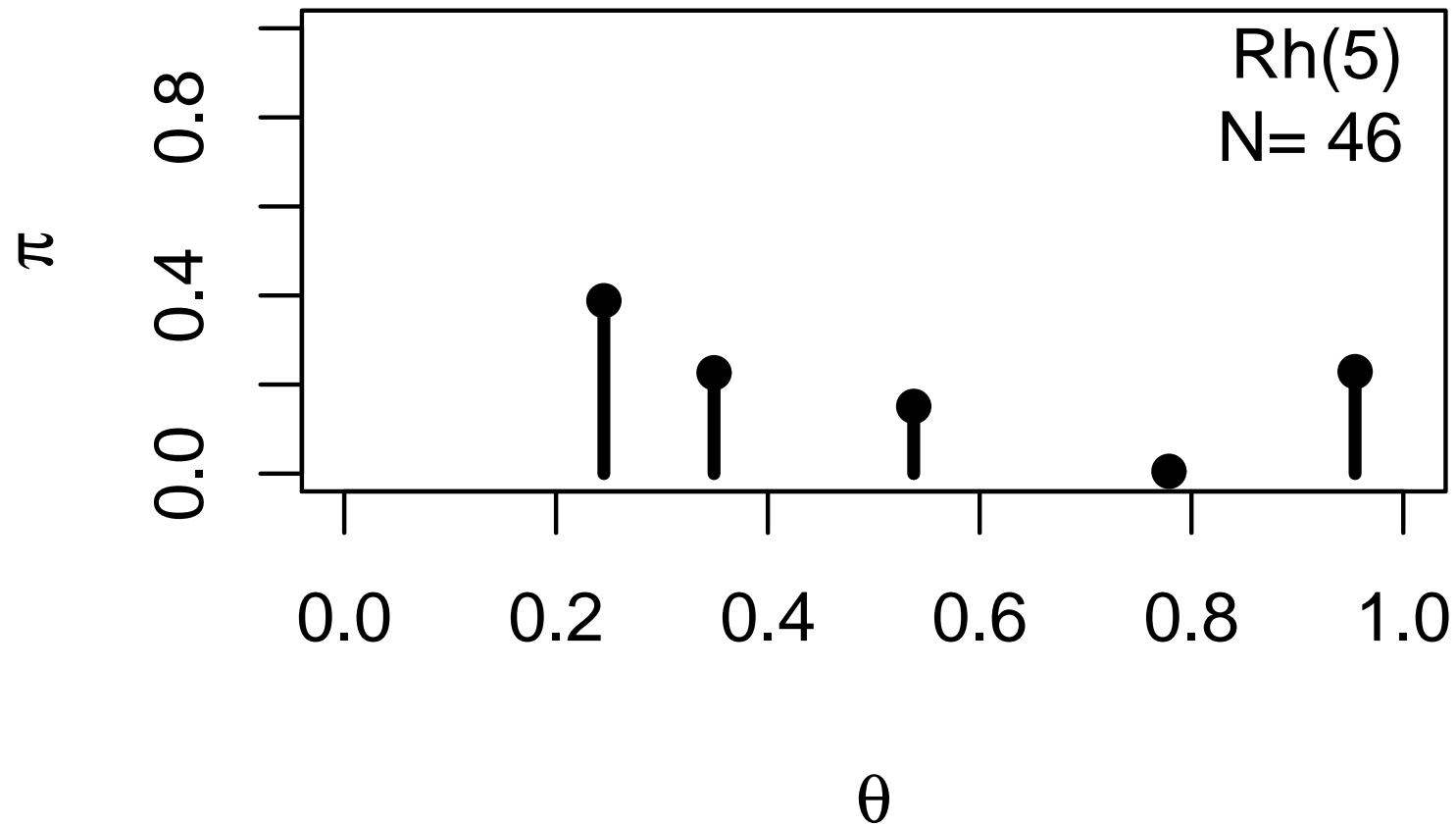
Step 22

$$\log L = -36.4306162111048$$



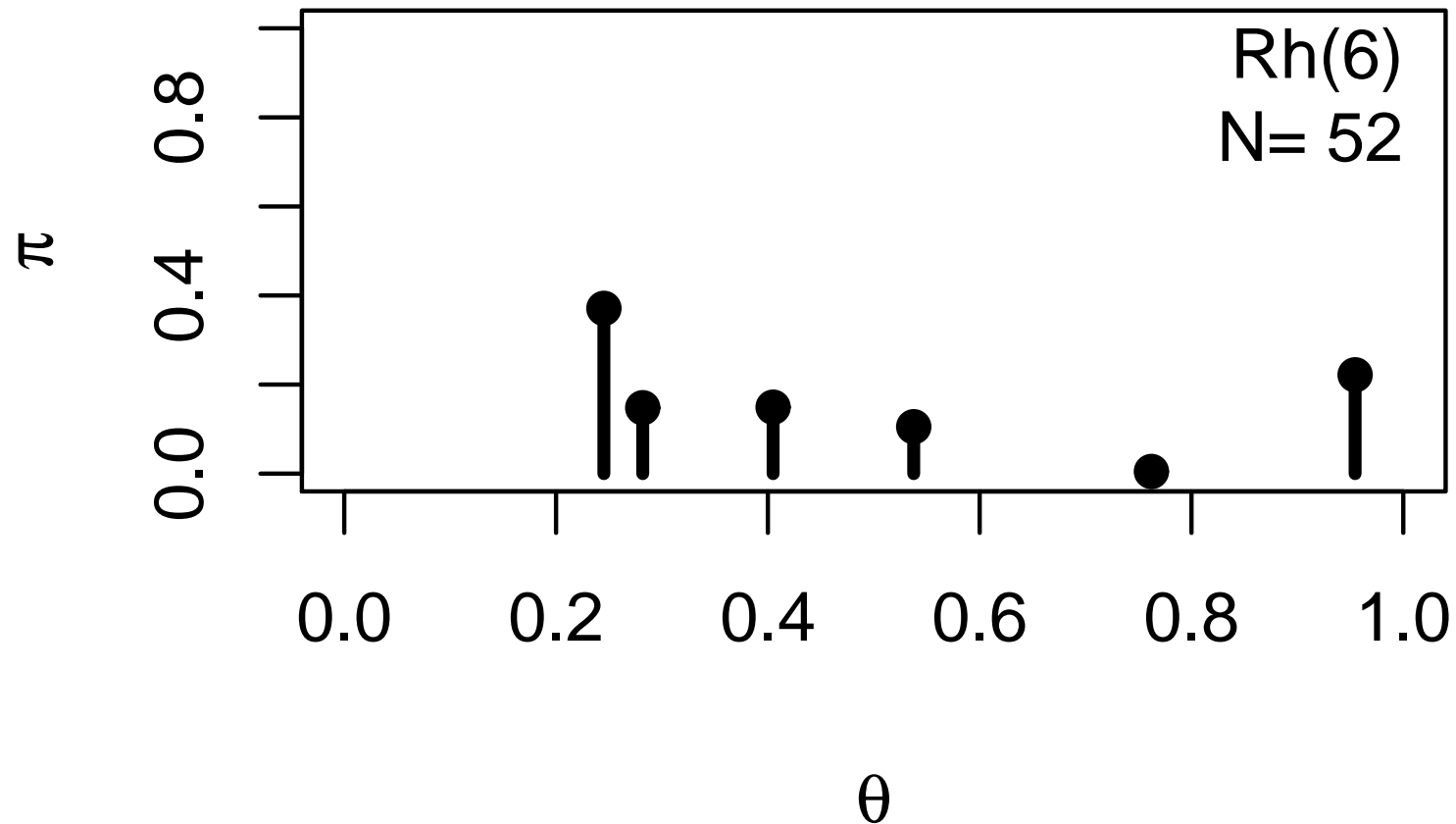
Step 23

$$\log L = -33.0389452678949$$



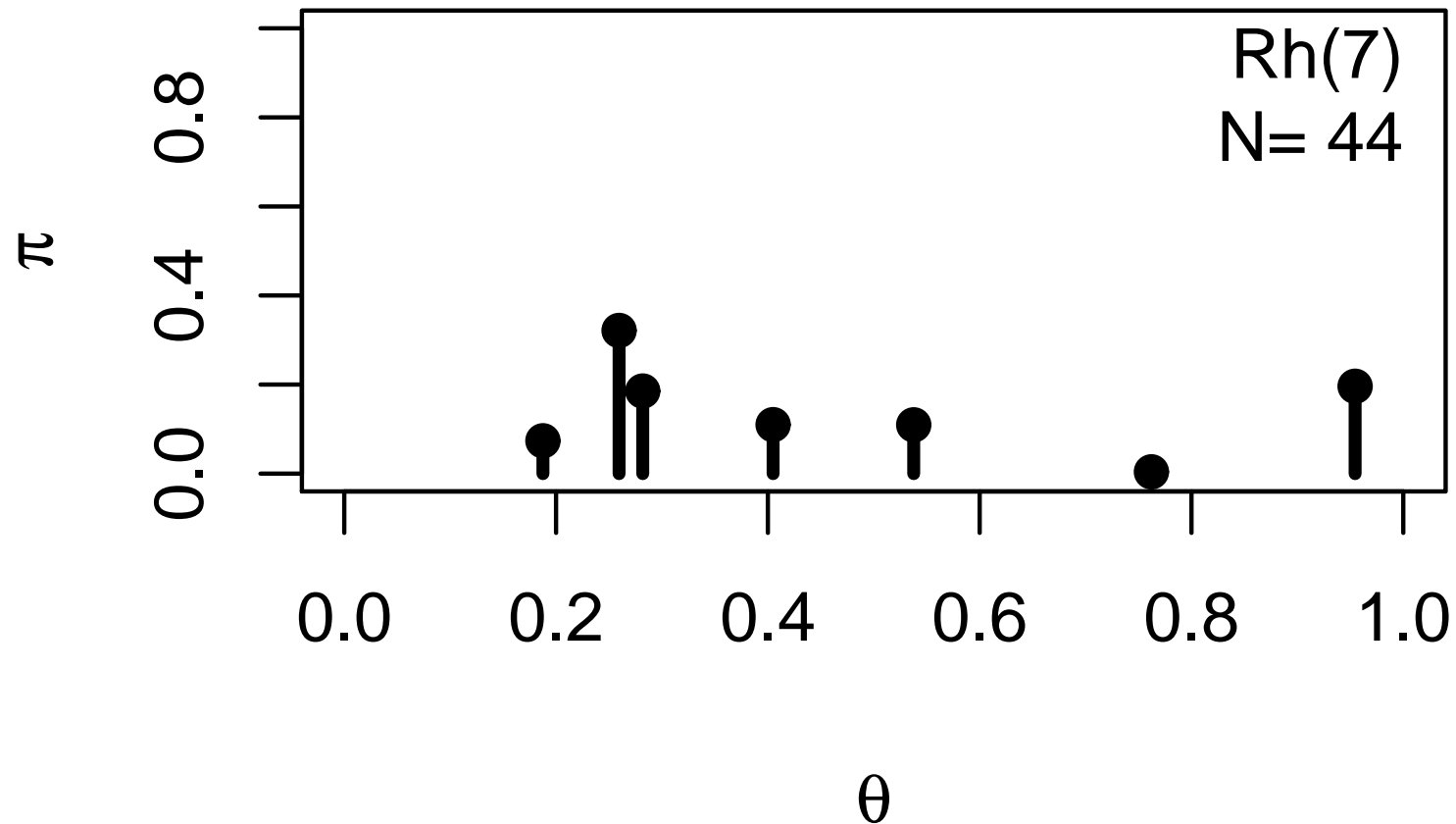
Step 24

$$\log L = -32.9077337310434$$



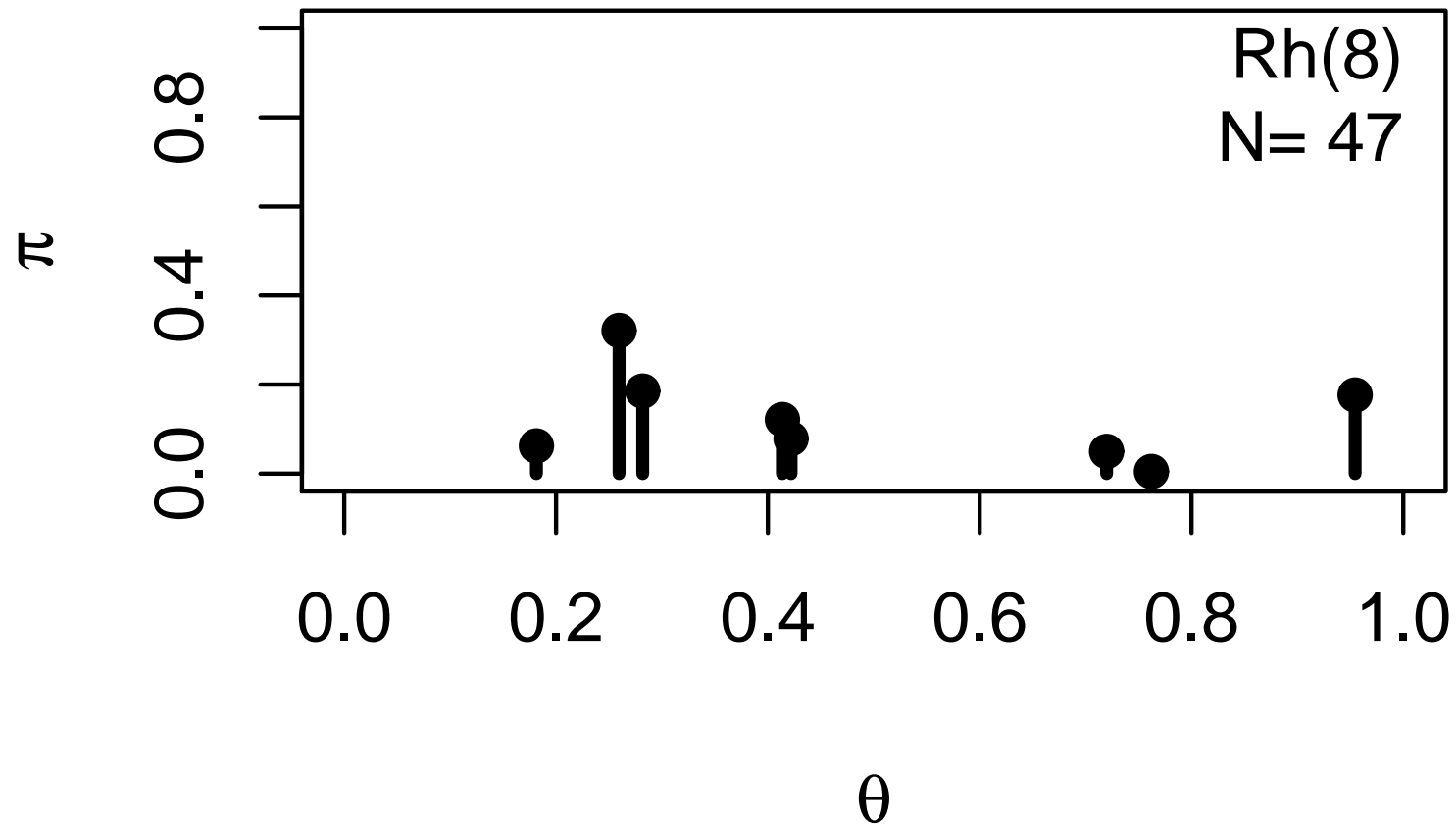
Step 25

$$\log L = -30.4732050323583$$



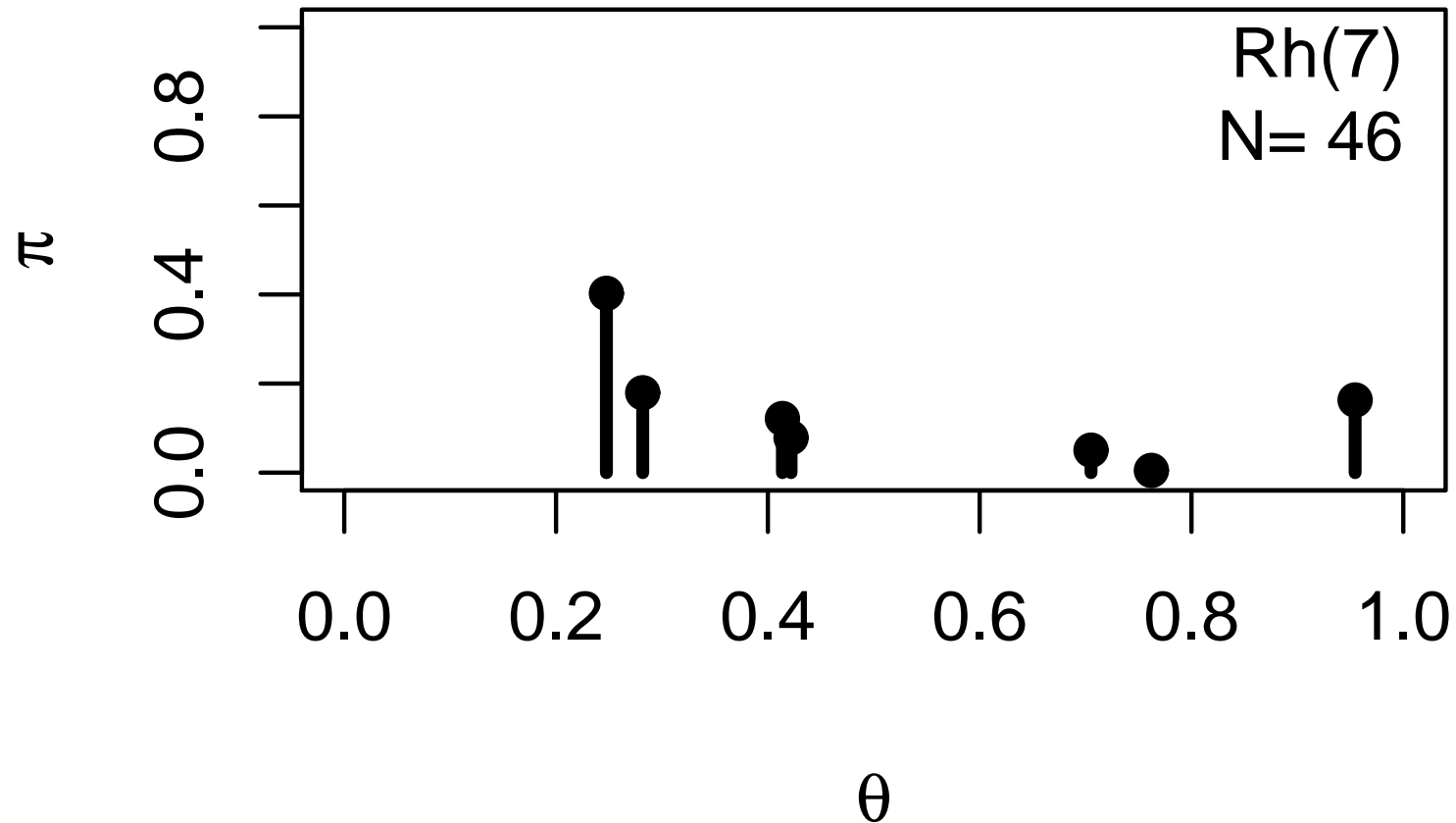
Step 26

$\log L = -28.4096812545929$



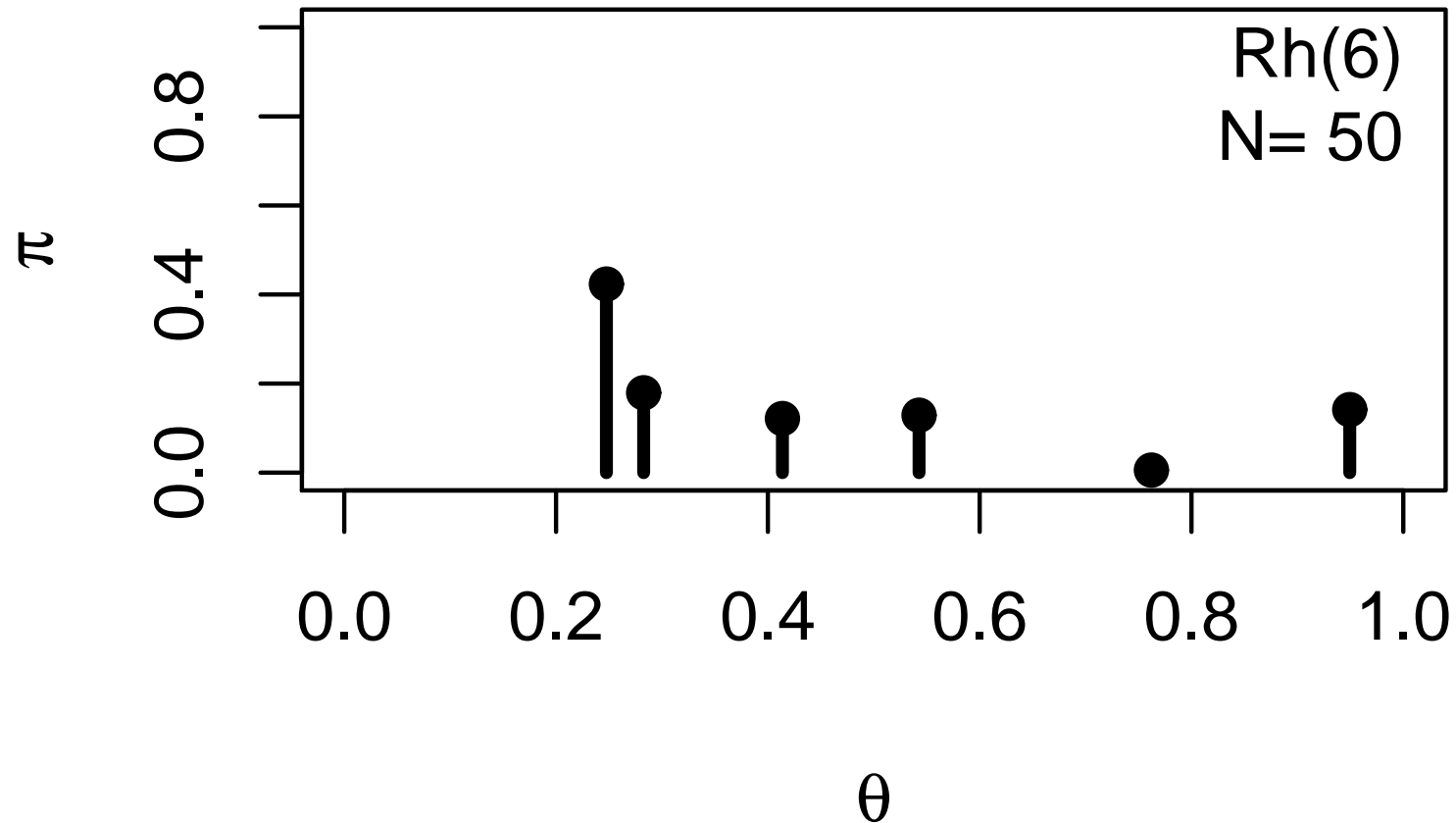
Step 27

$$\log L = -27.59453640333$$



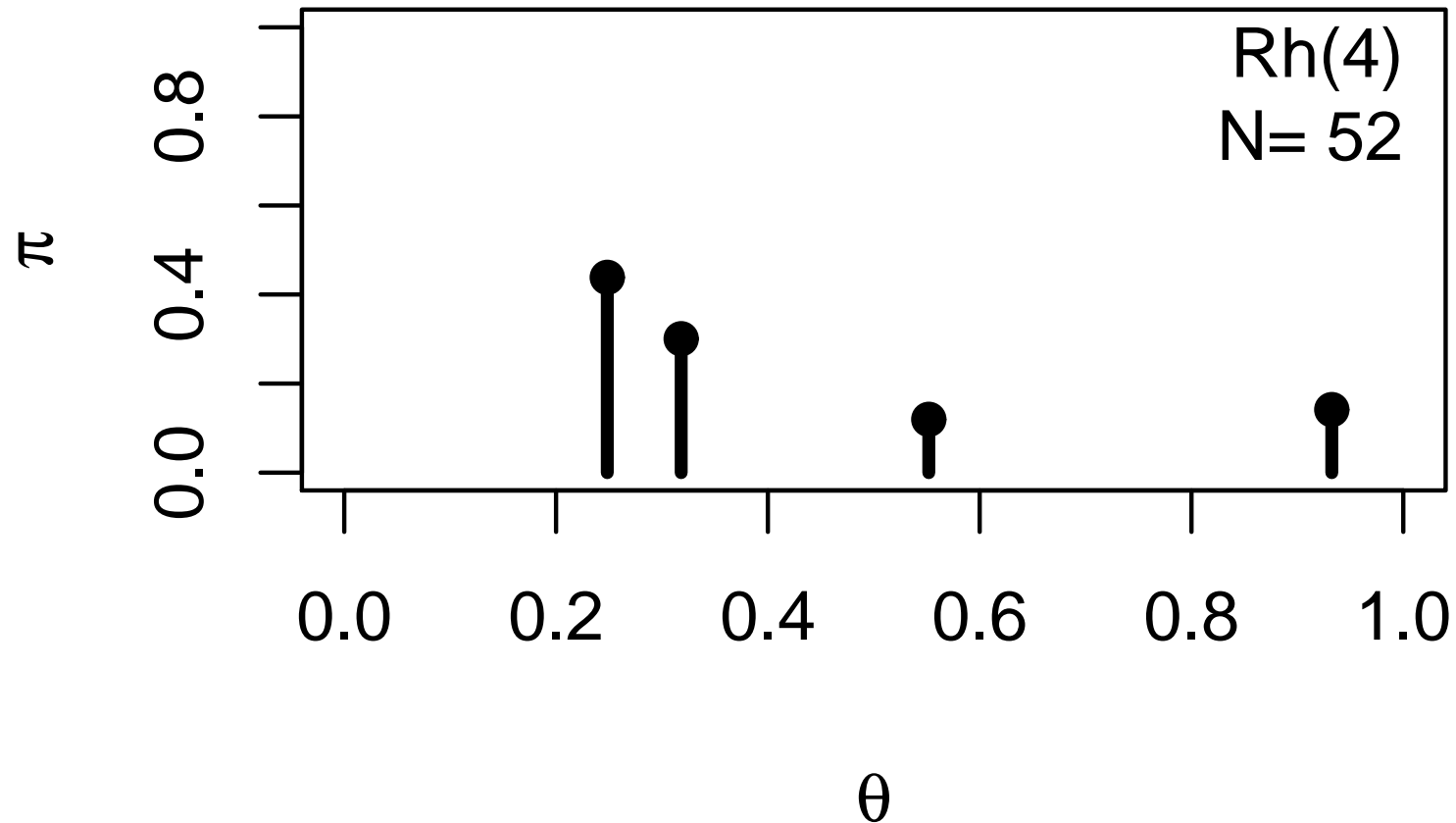
Step 28

$$\log L = -27.1528912375878$$



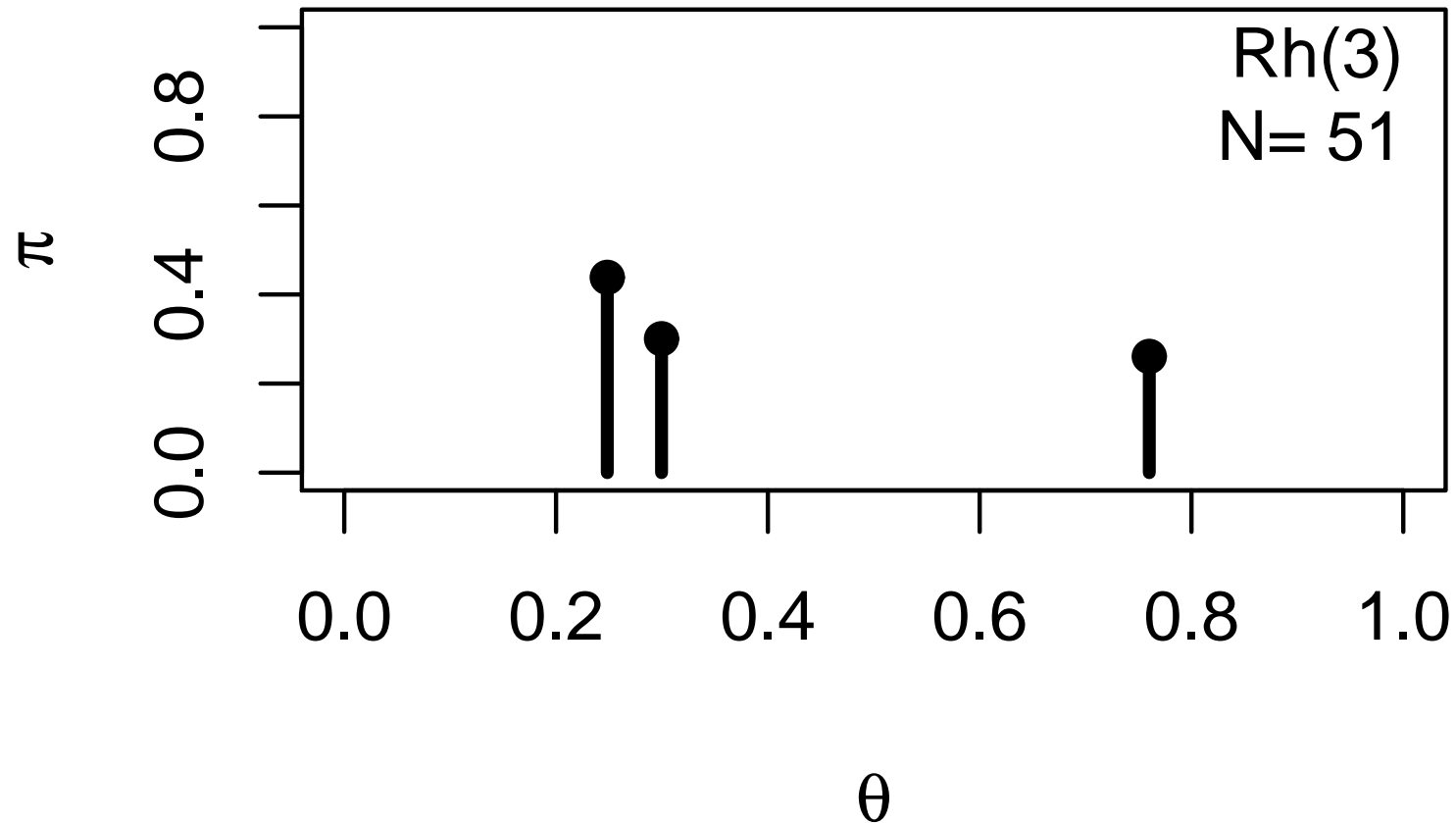
Step 29

$$\log L = -26.2582030338909$$



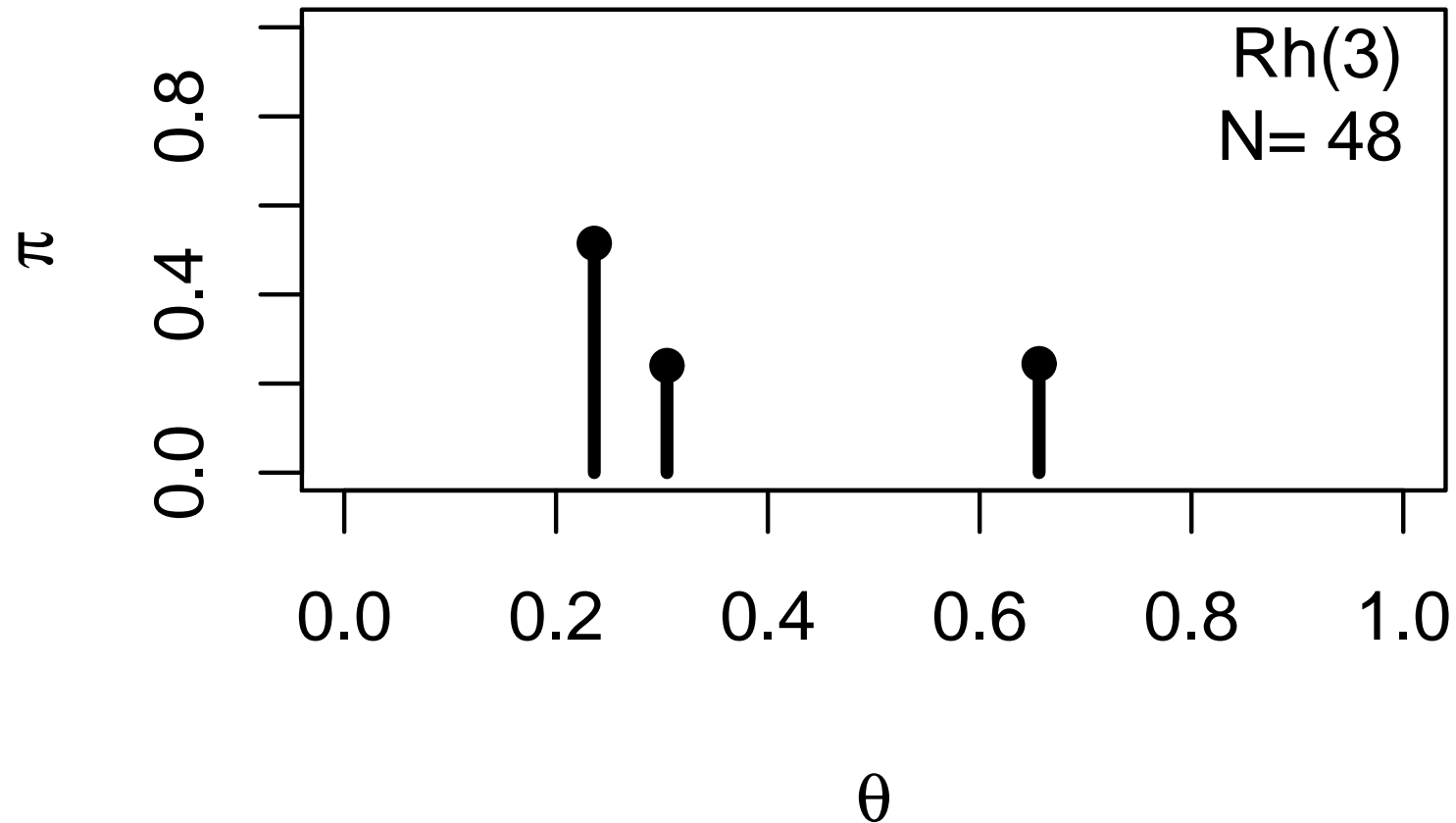
Step 30

$$\log L = -25.7864676274672$$



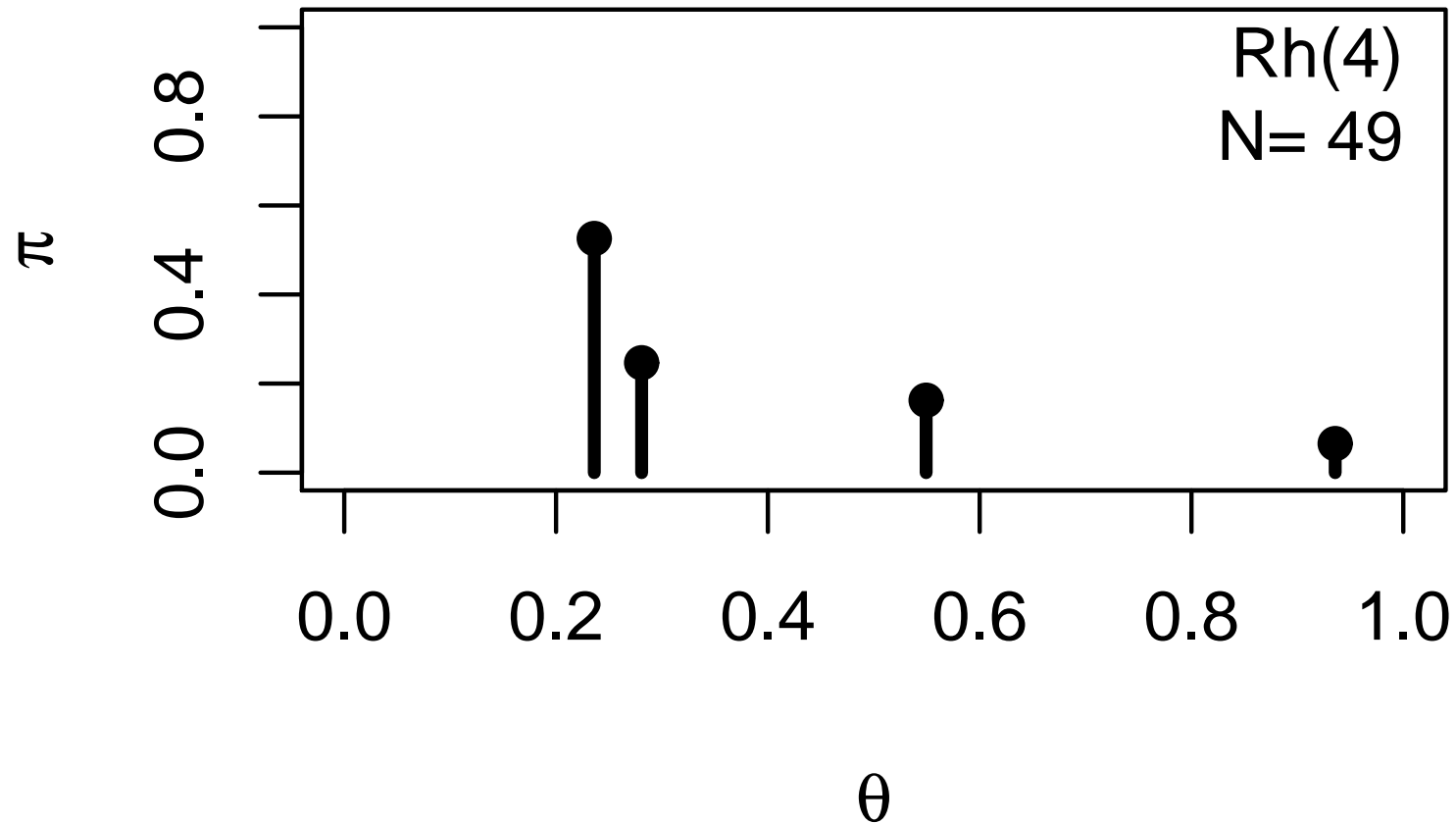
Step 31

$\log L = -22.3220481167324$



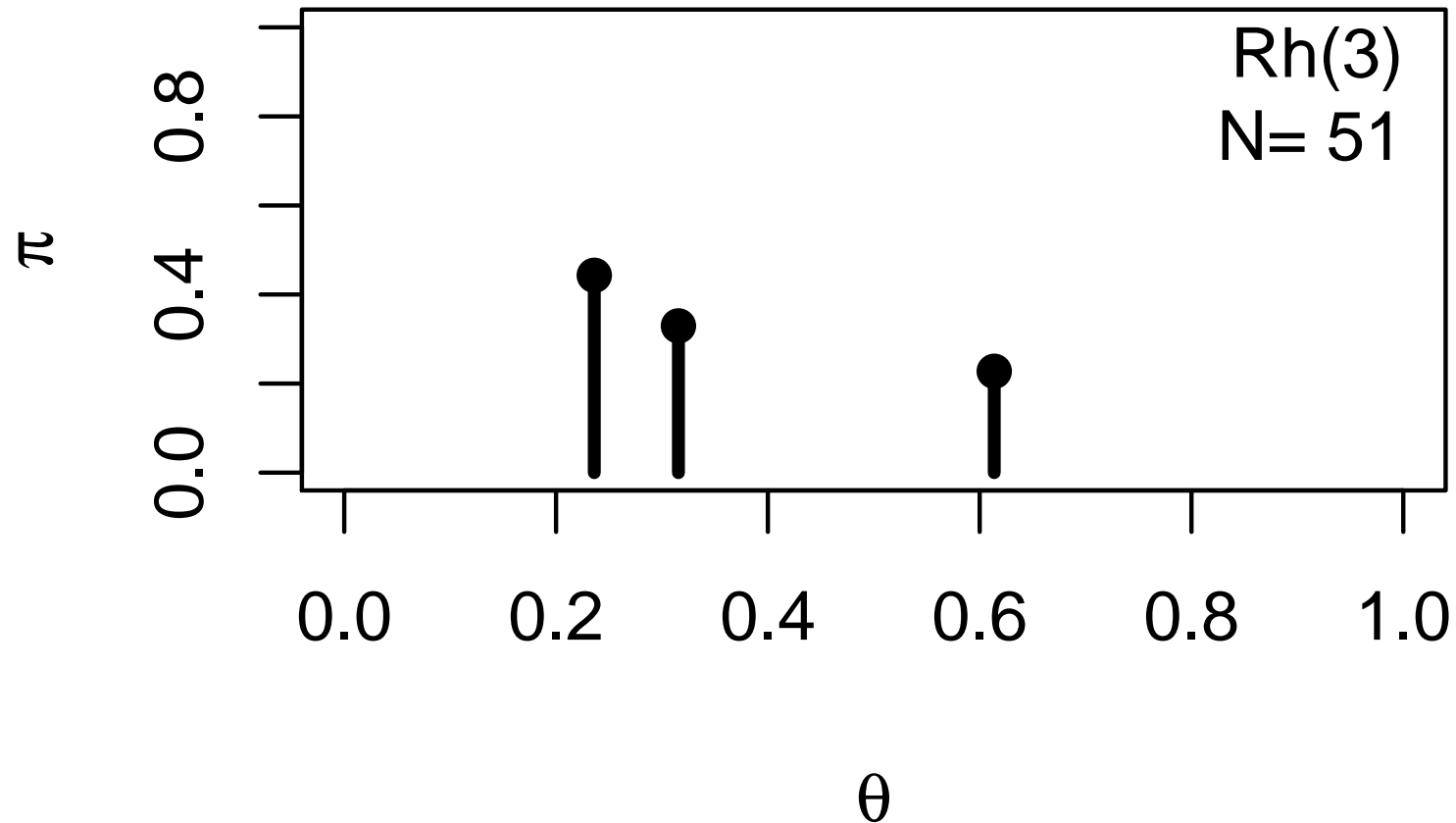
Step 32

$\log L = -19.8536269710666$



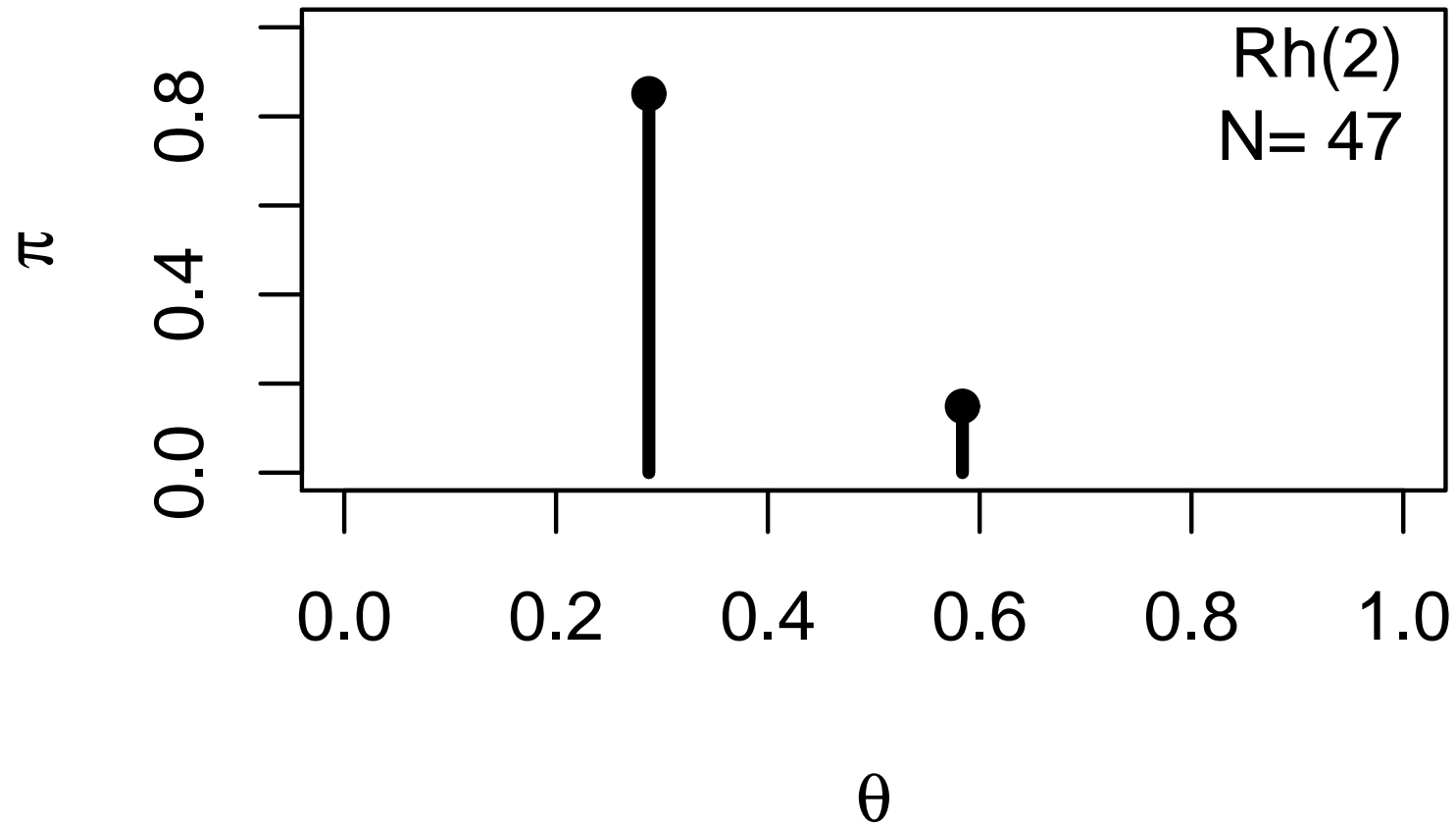
Step 33

$$\log L = -22.6319674062755$$



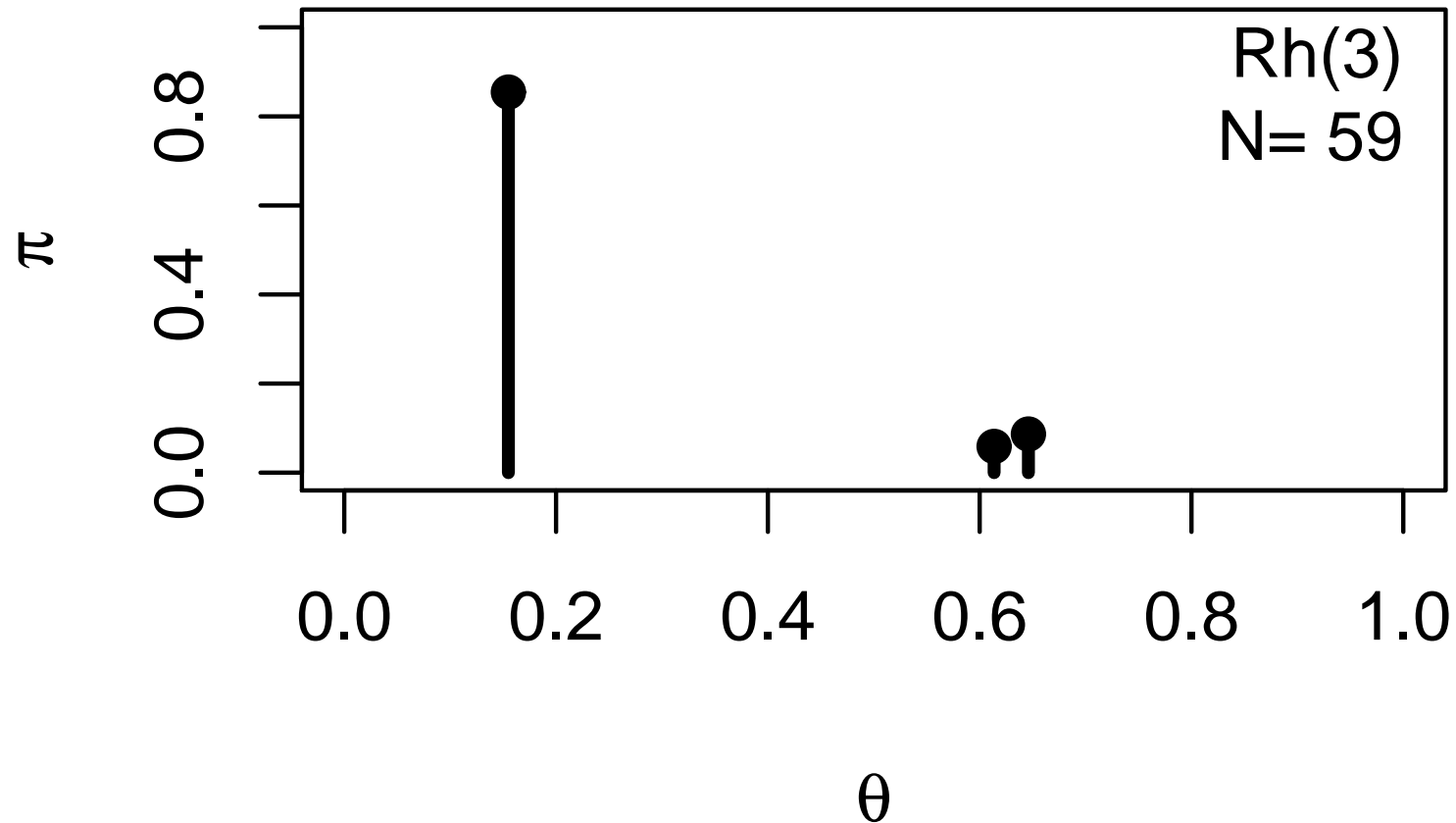
Step 34

$$\log L = -20.6214301093238$$



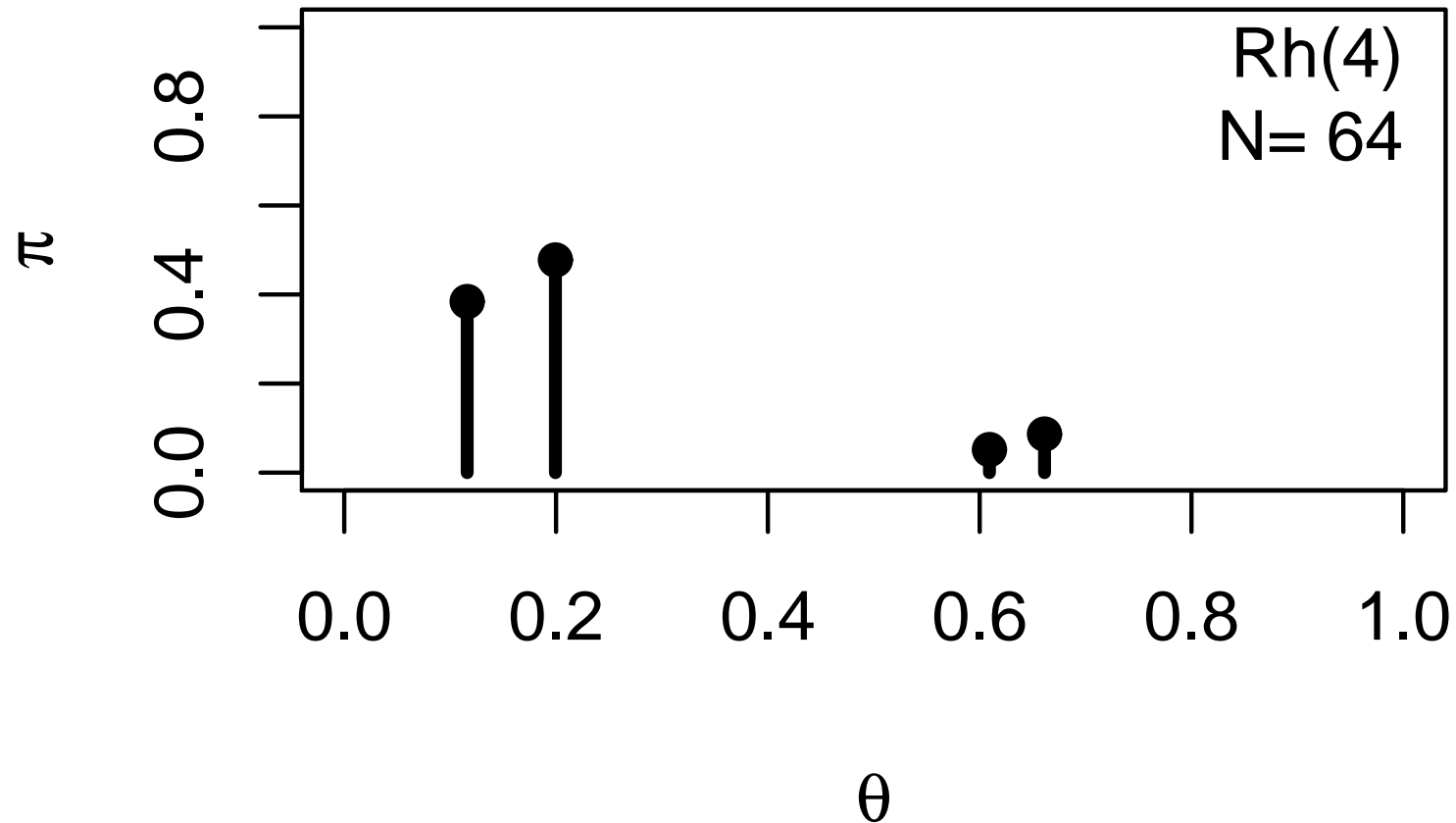
Step 35

$$\log L = -10.2673835002677$$



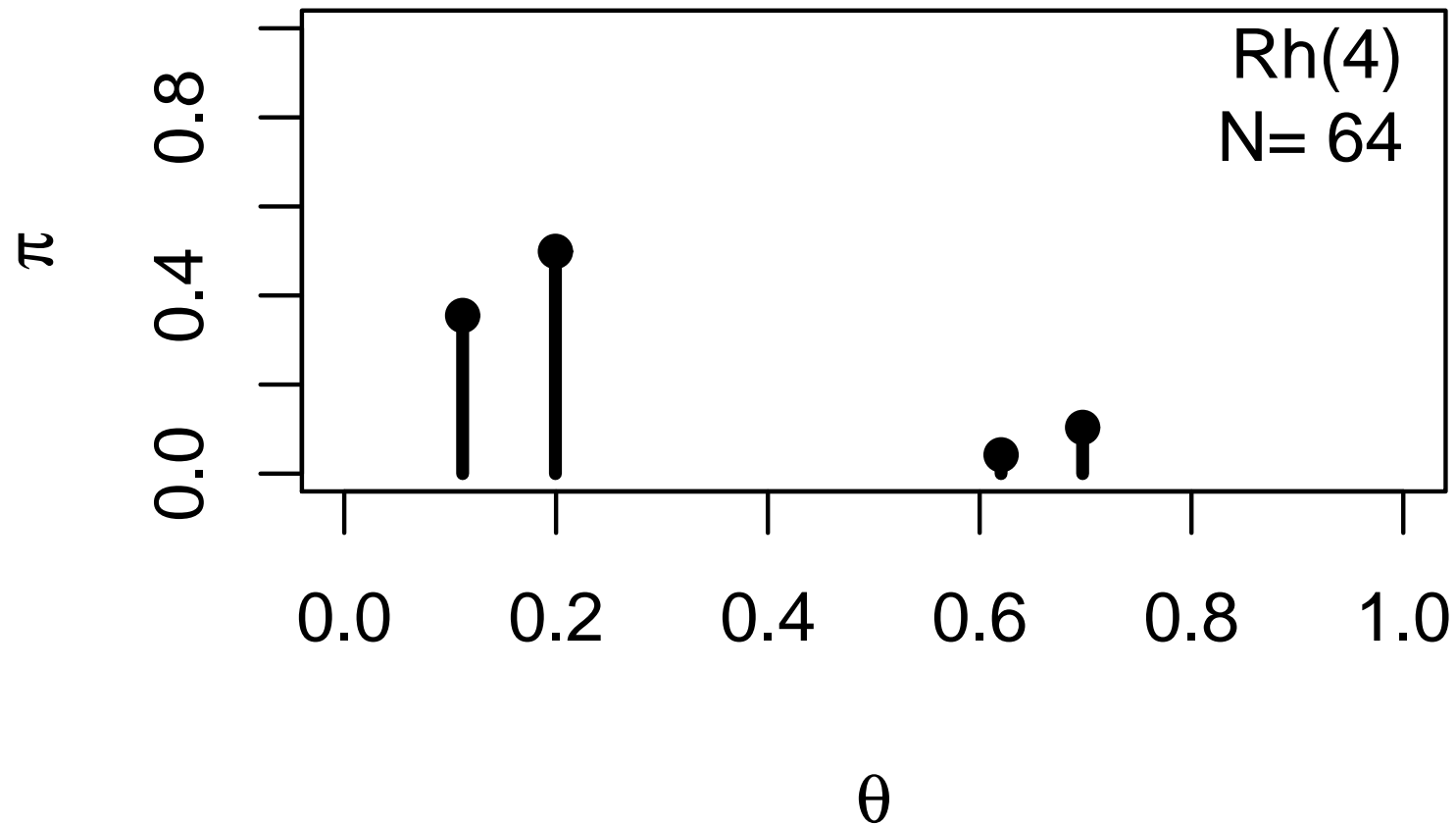
Step 36

$\log L = -10.819486250473$



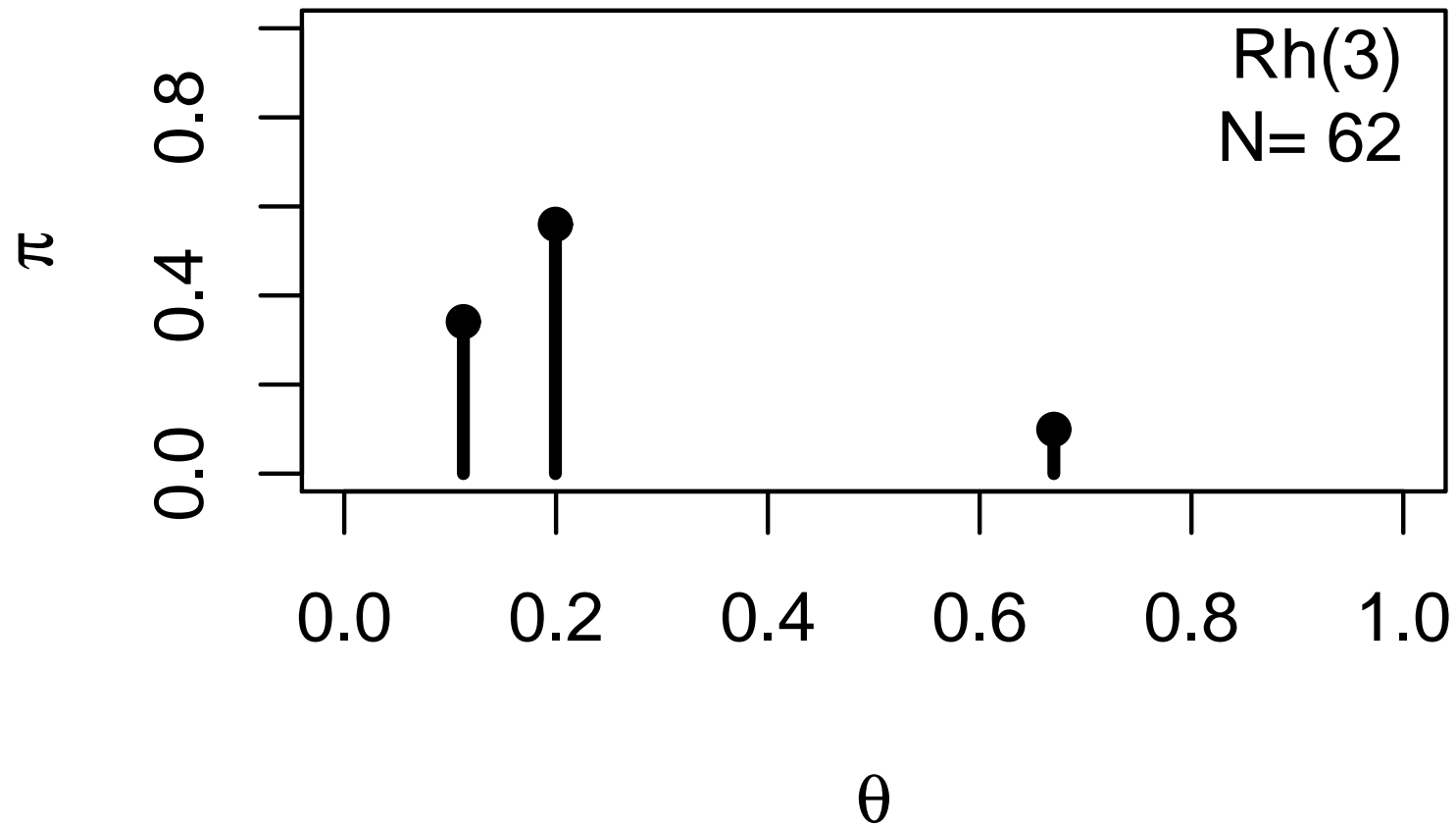
Step 37

$\log L = -11.5513213071656$



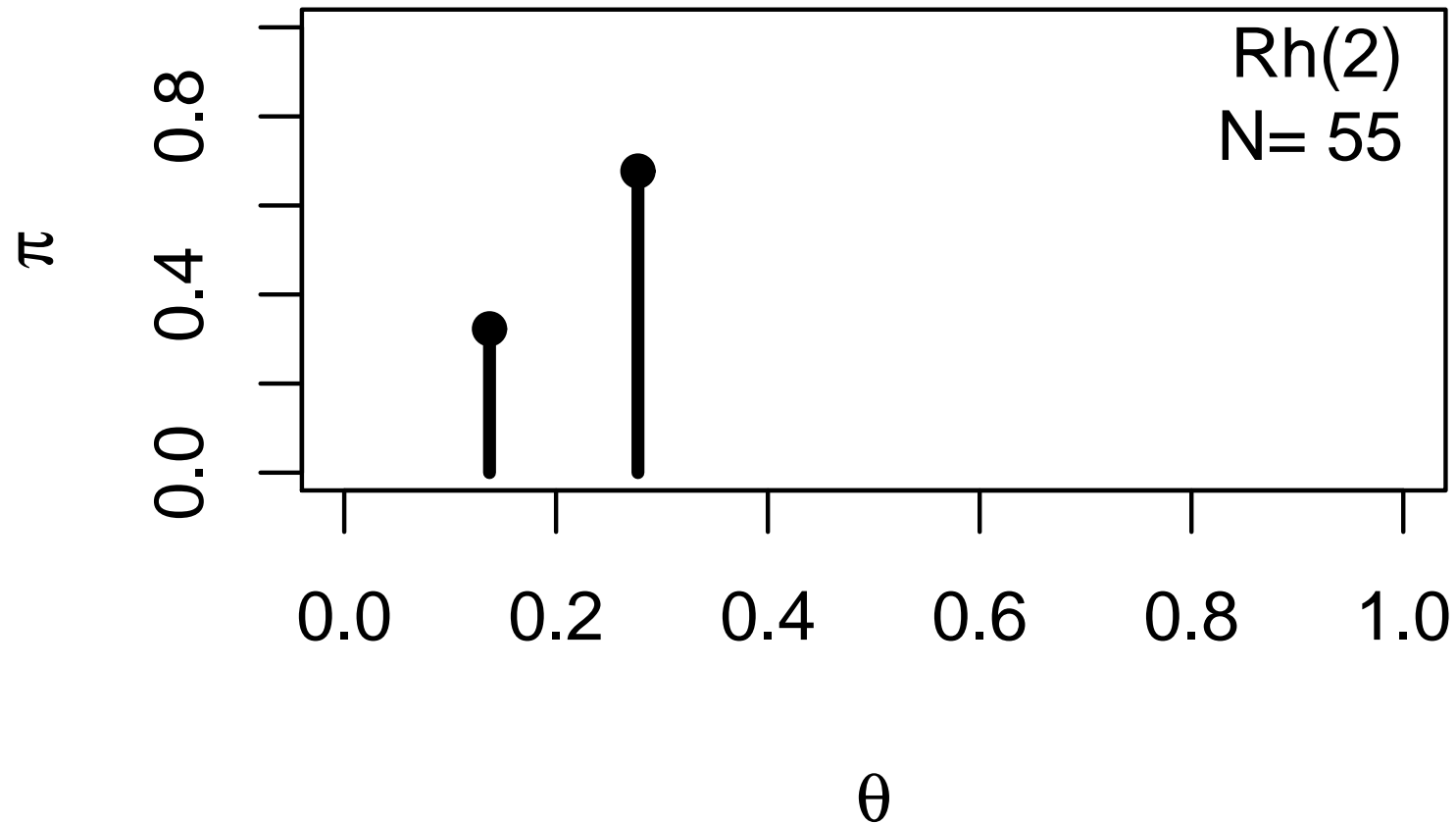
Step 38

$$\log L = -9.16791808262374$$



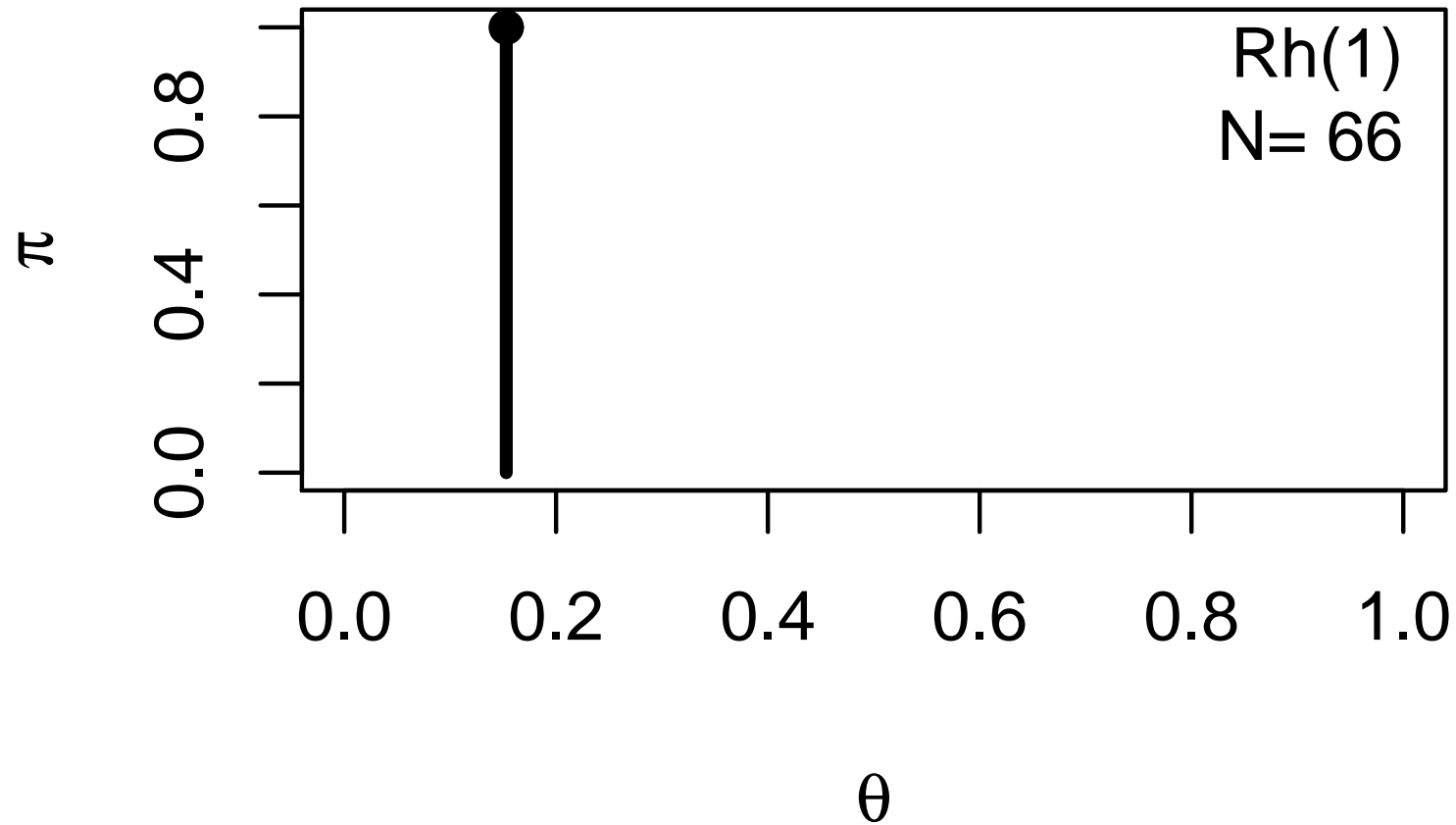
Step 39

$$\log L = -11.3187838098980$$



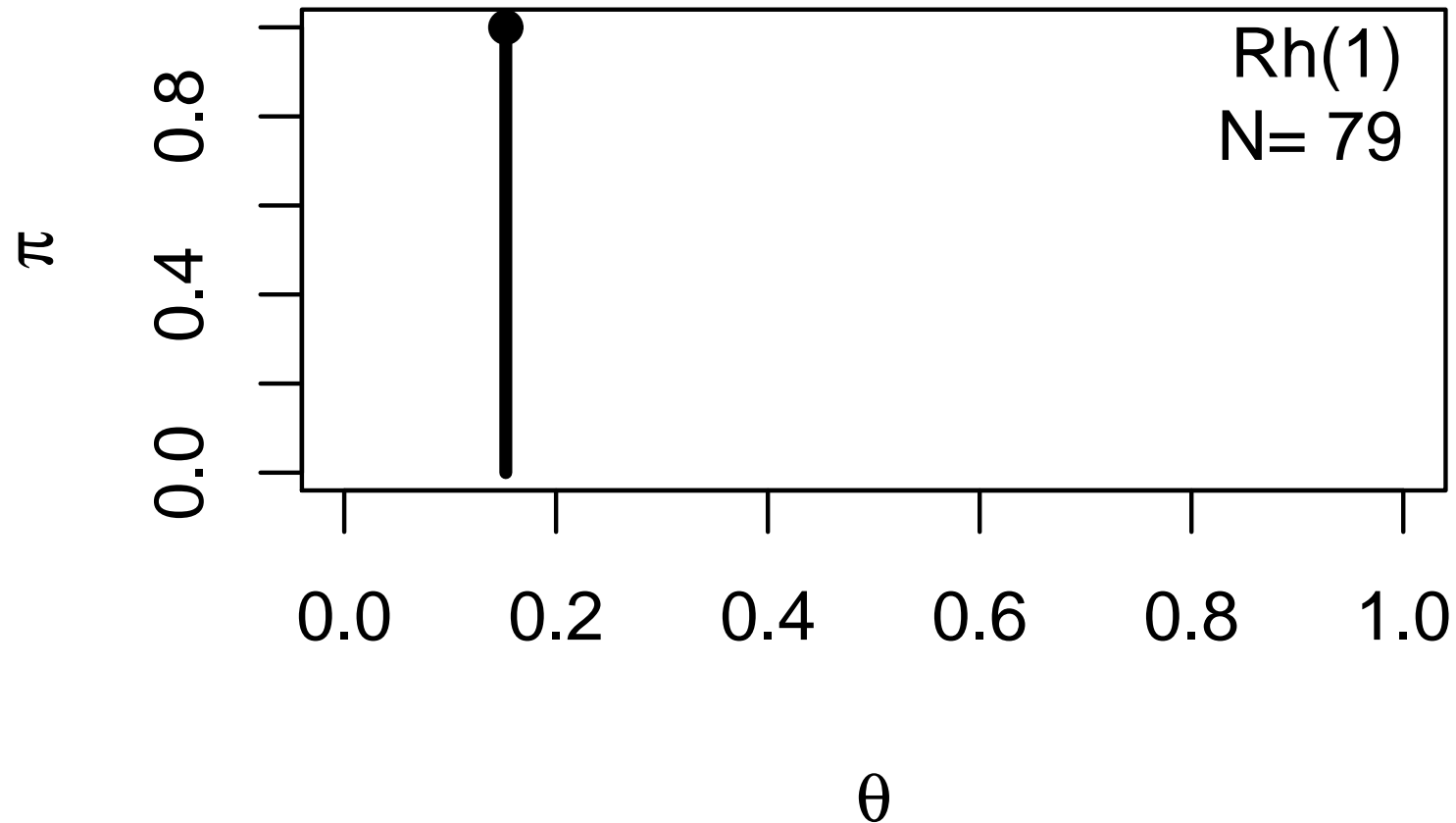
Step 40

$$\log L = -6.50771712565023$$



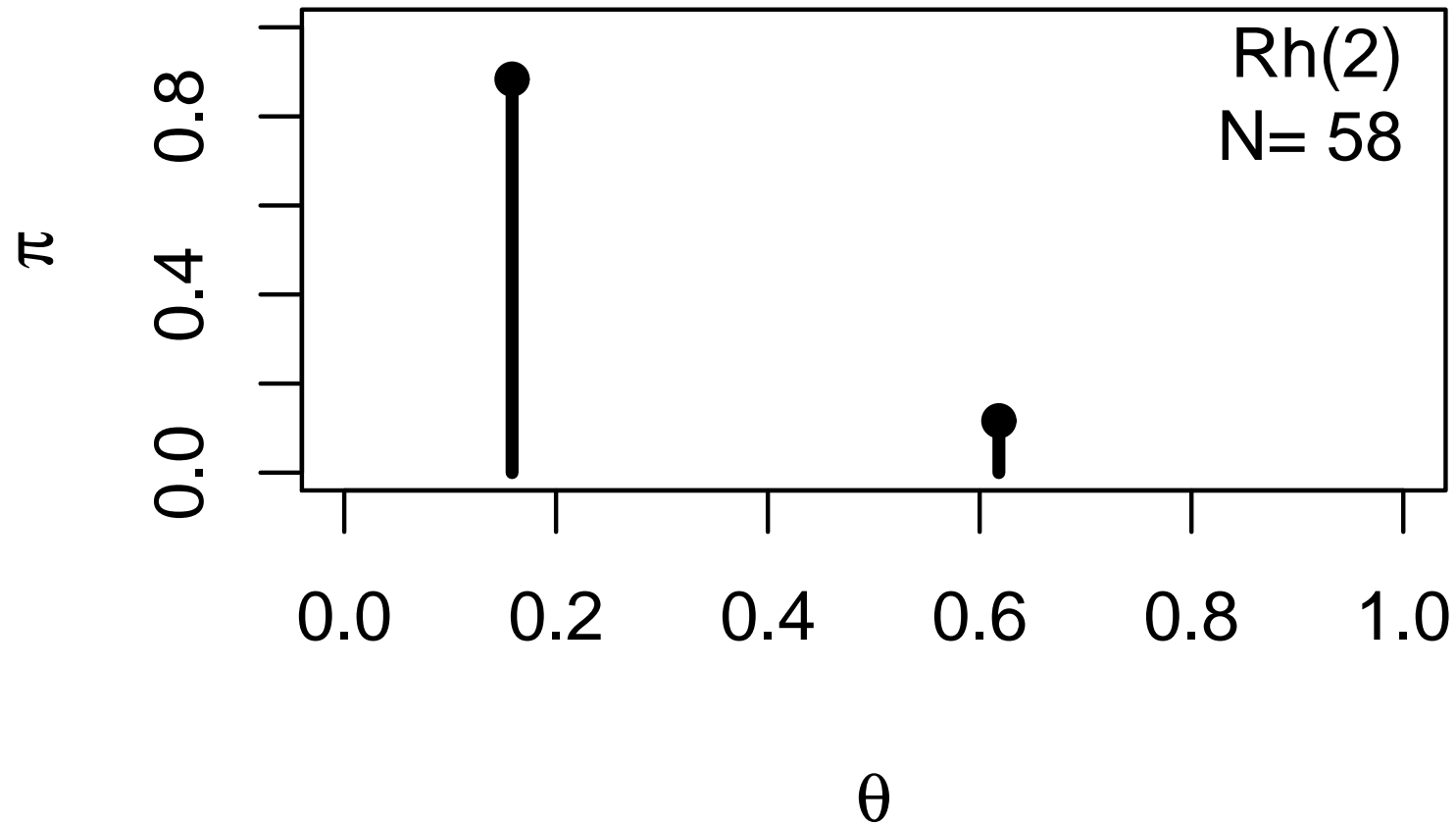
Step 41

$$\log L = -7.77292212428978$$



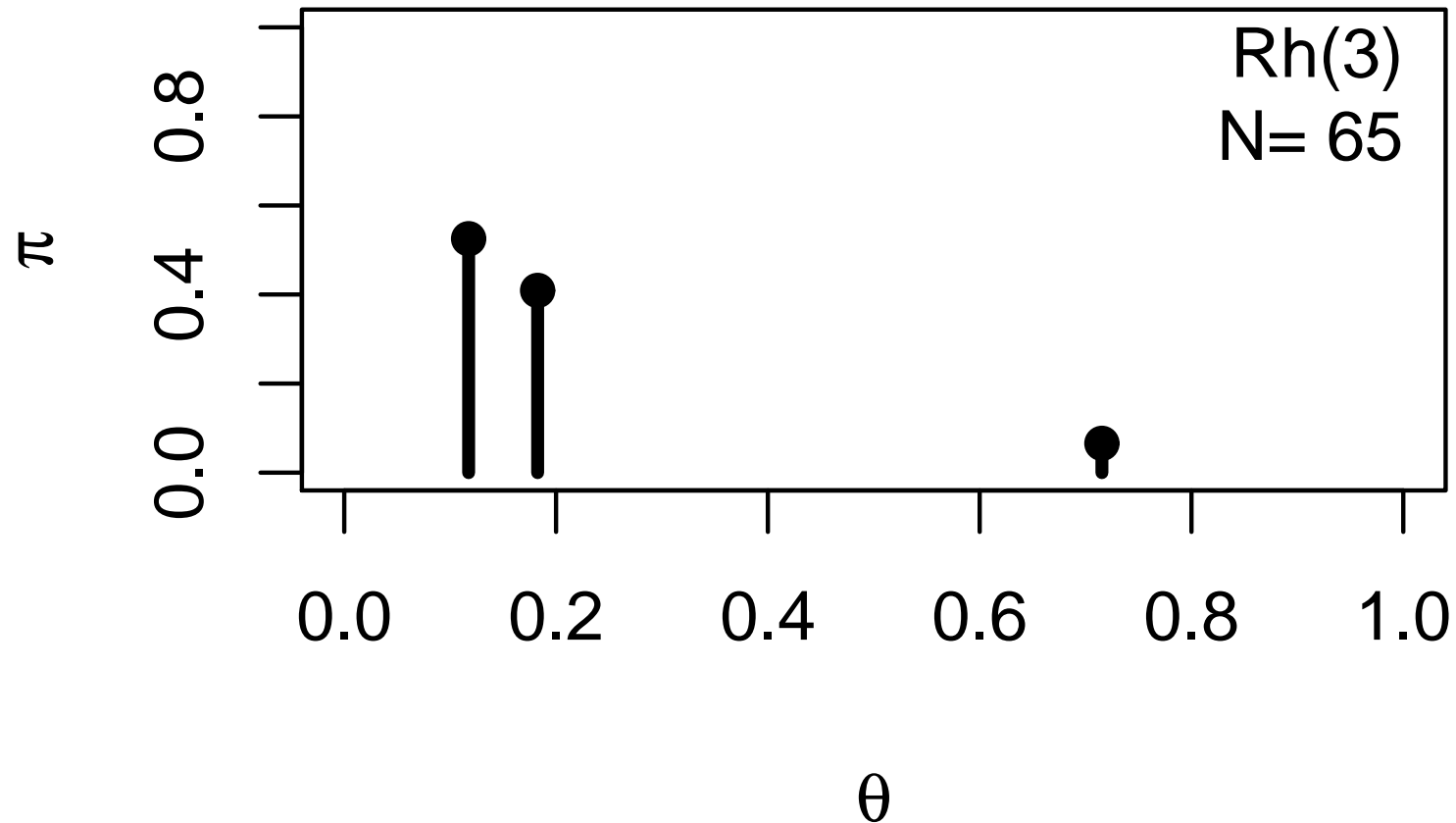
Step 42

$$\log L = -9.07442912991769$$



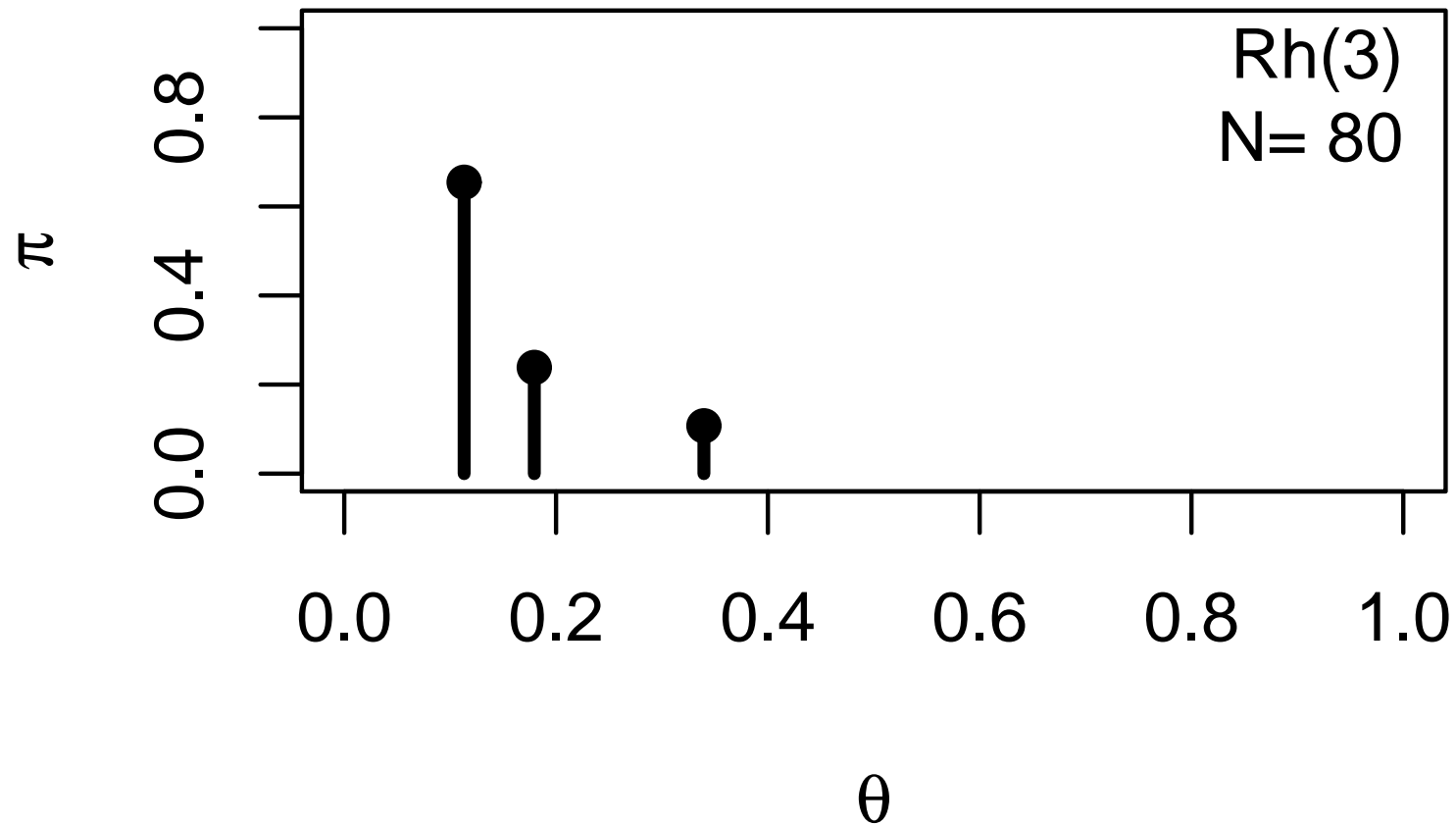
Step 43

$$\log L = -6.7347692978596$$



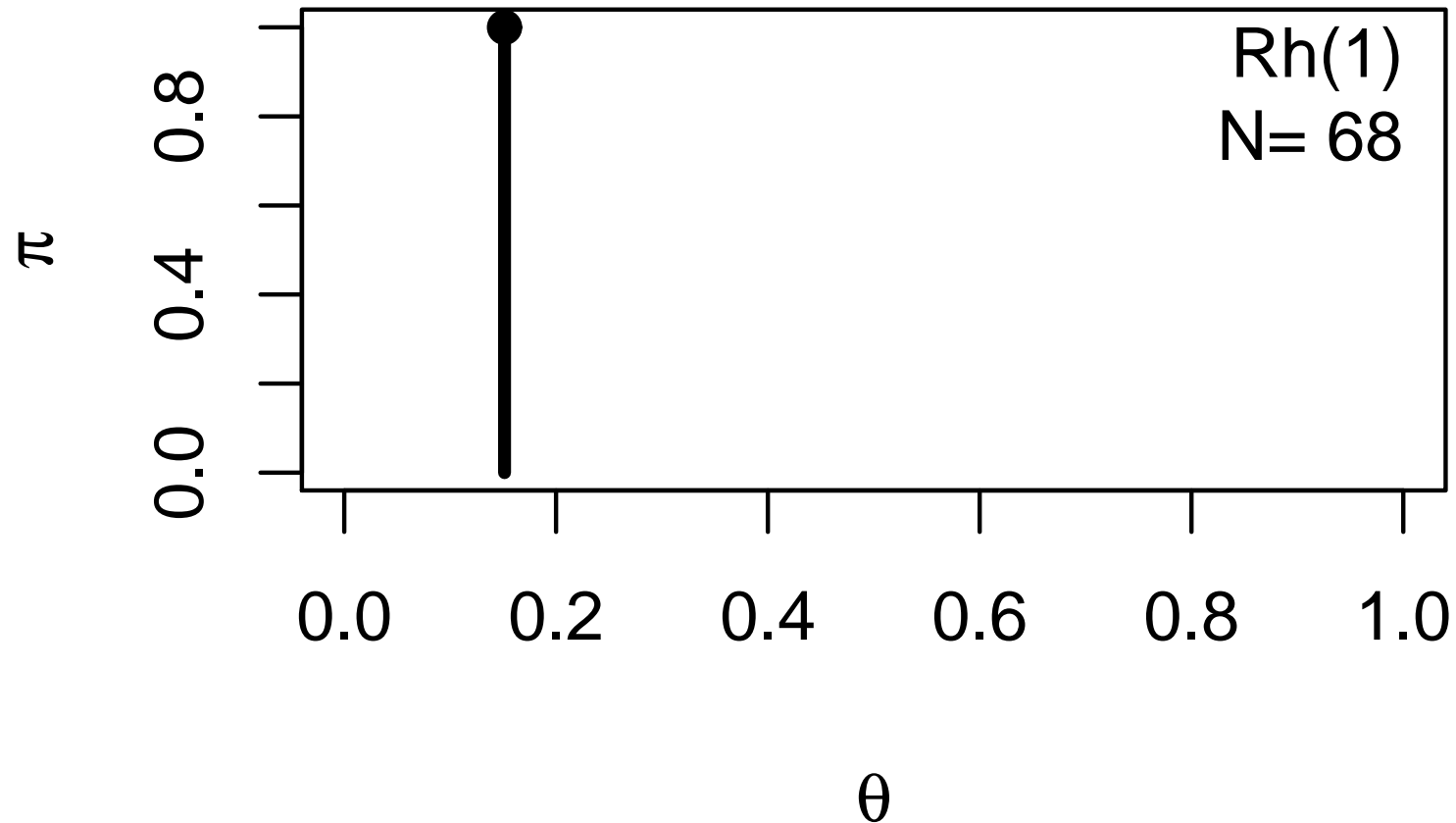
Step 44

$$\log L = -6.85775153090216$$



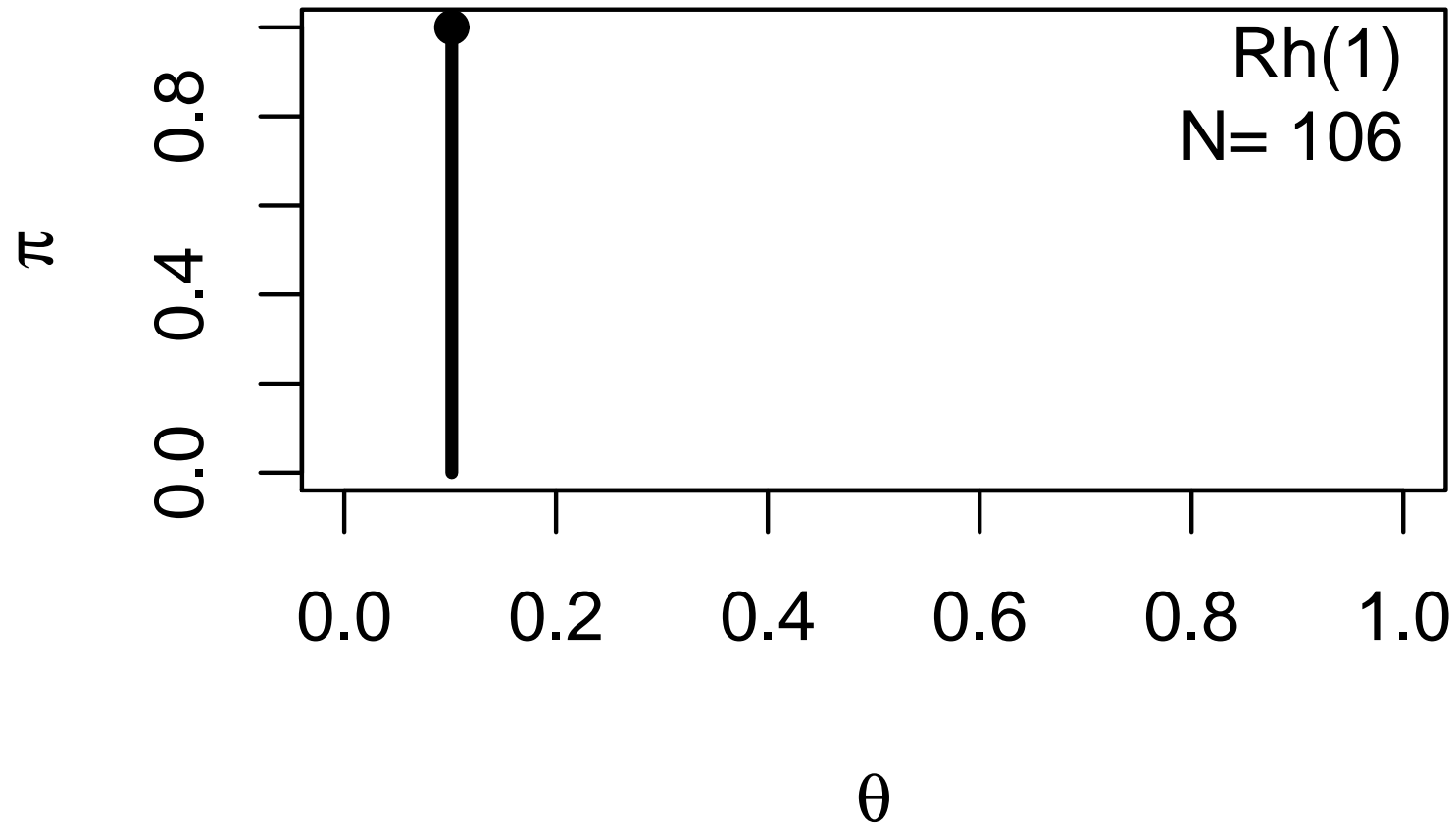
Step 45

$$\log L = -6.44686591772984$$



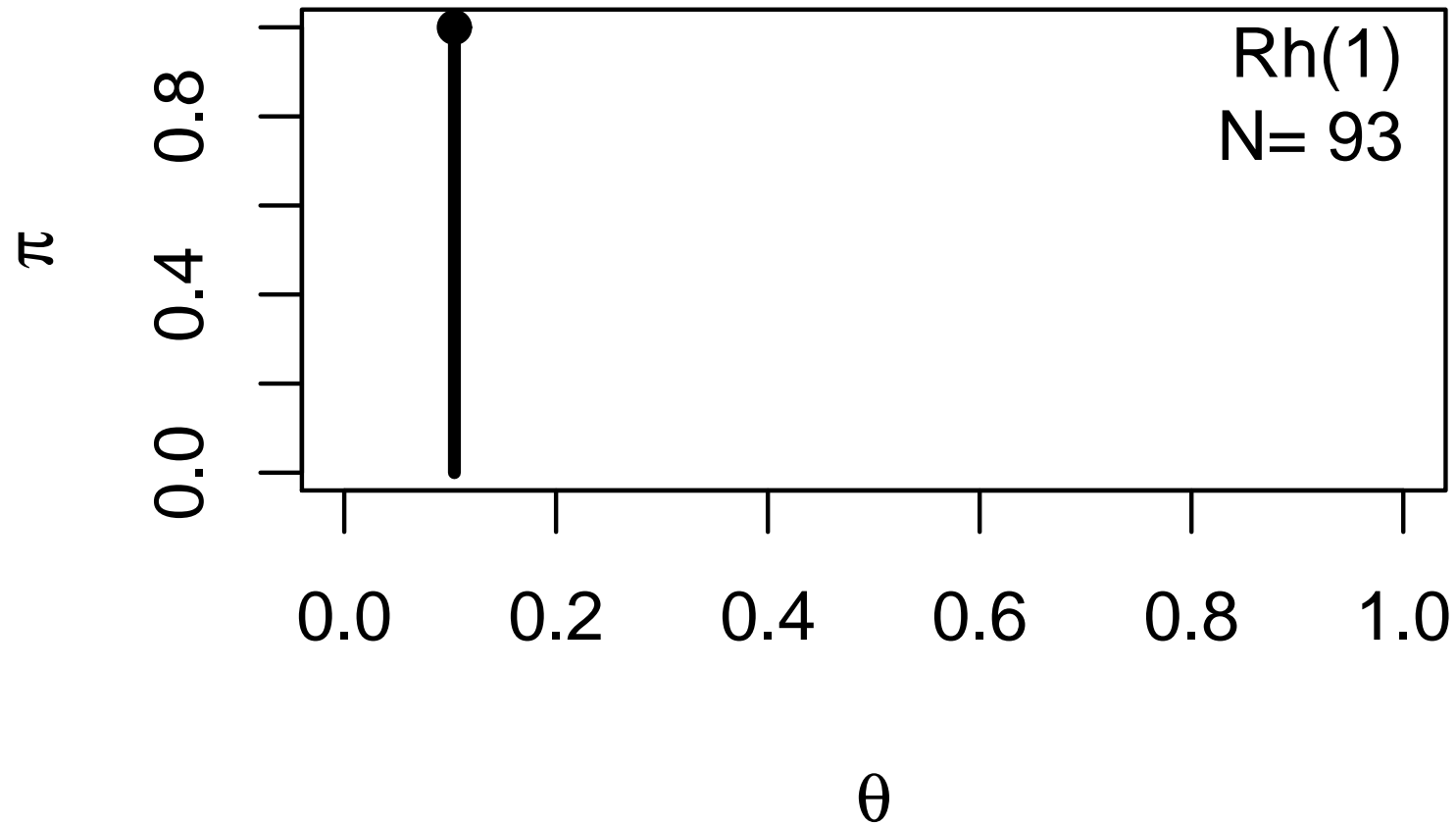
Step 46

$$\log L = -8.34608316367388$$



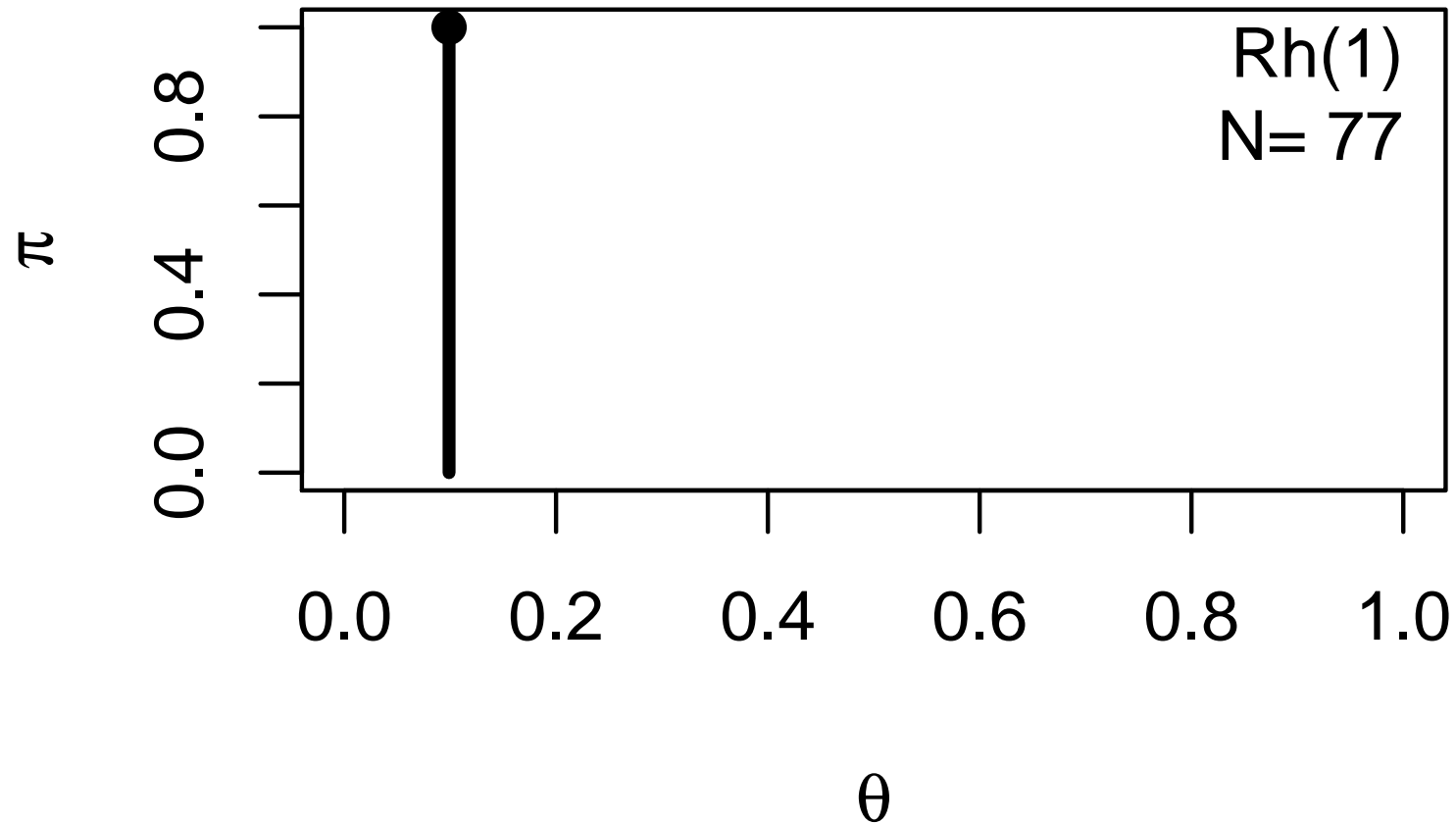
Step 47

$$\log L = -7.24947491612883$$



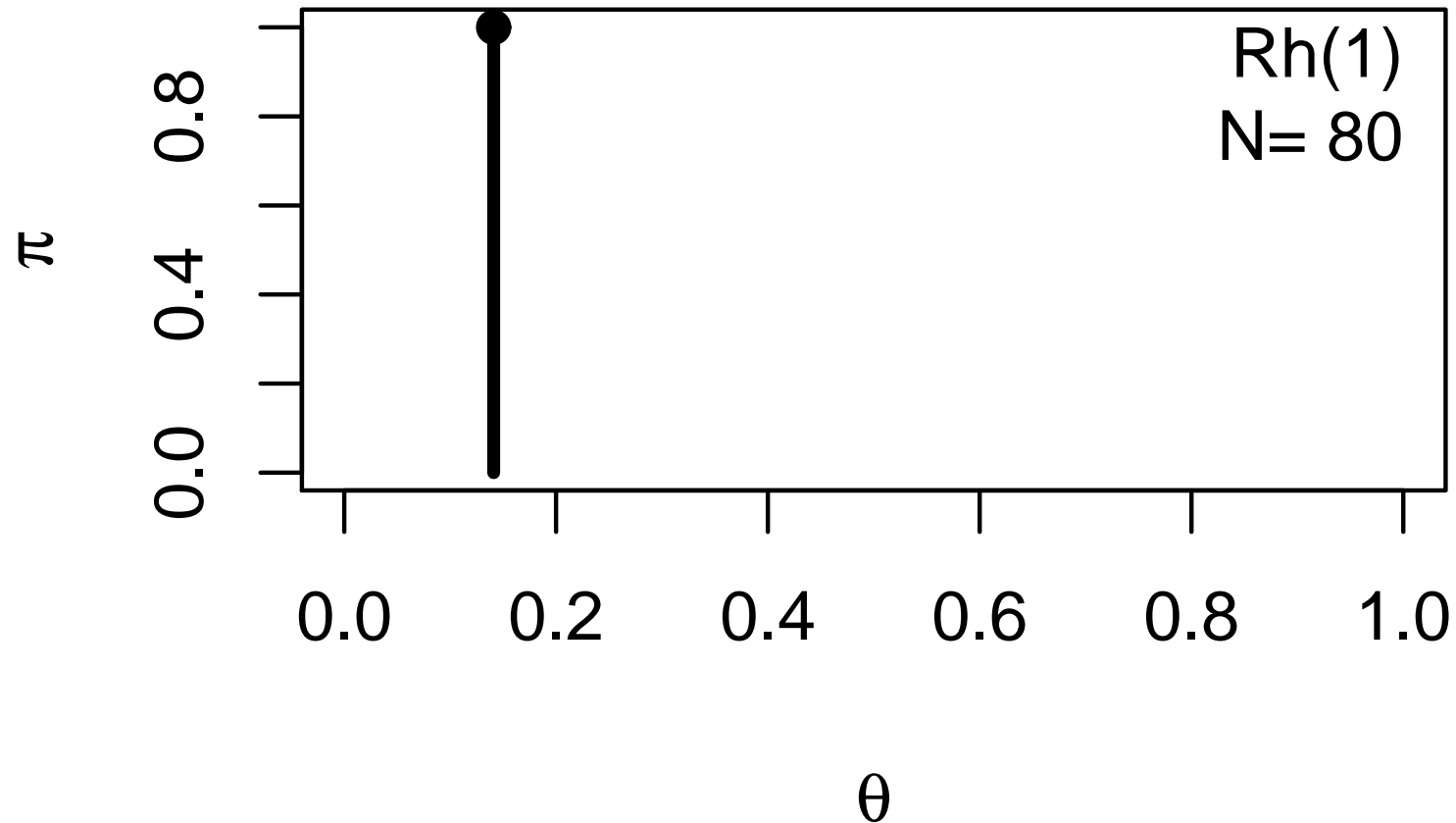
Step 48

$$\log L = -8.68619170427243$$



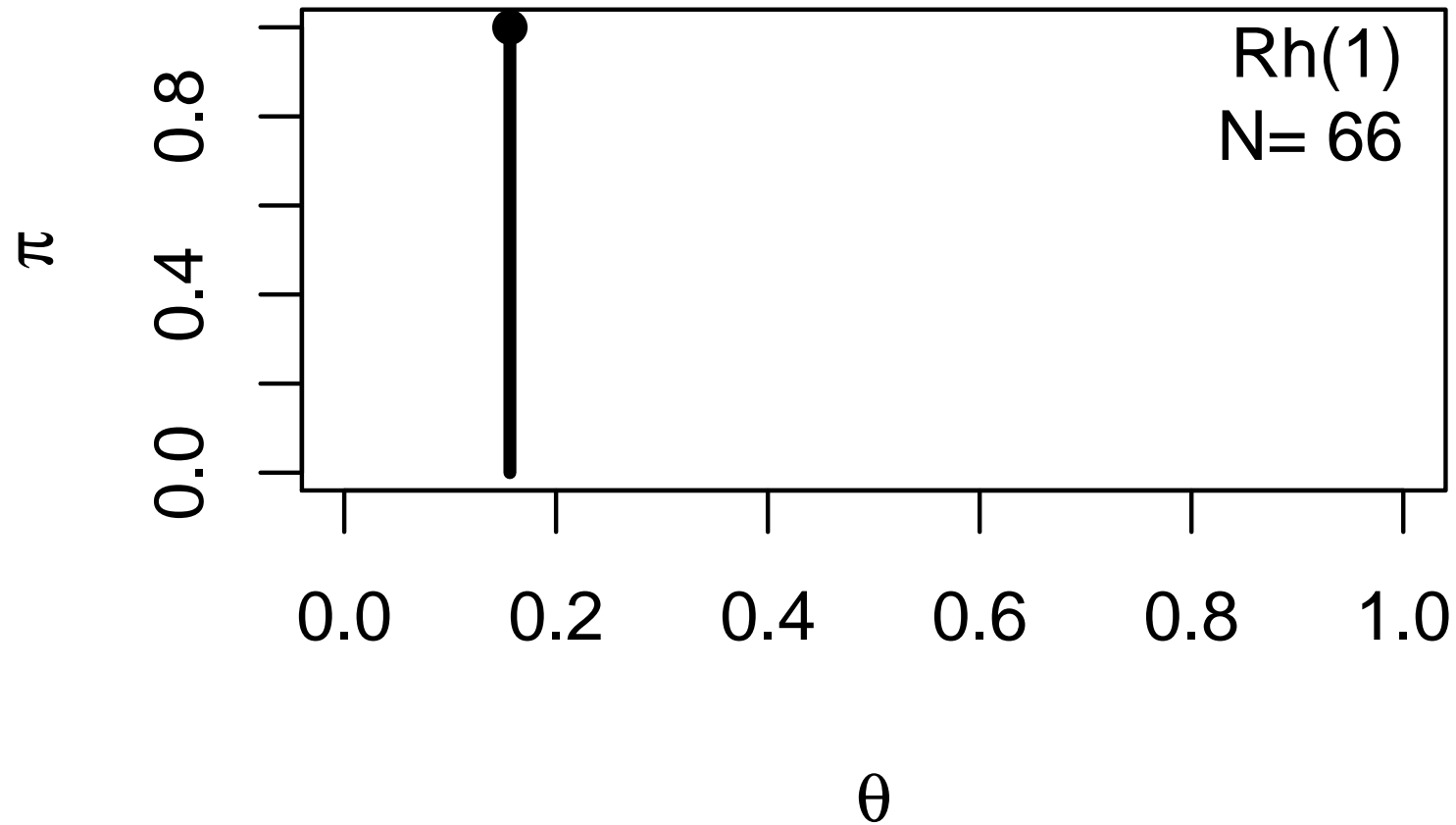
Step 49

$$\log L = -7.03769307699849$$



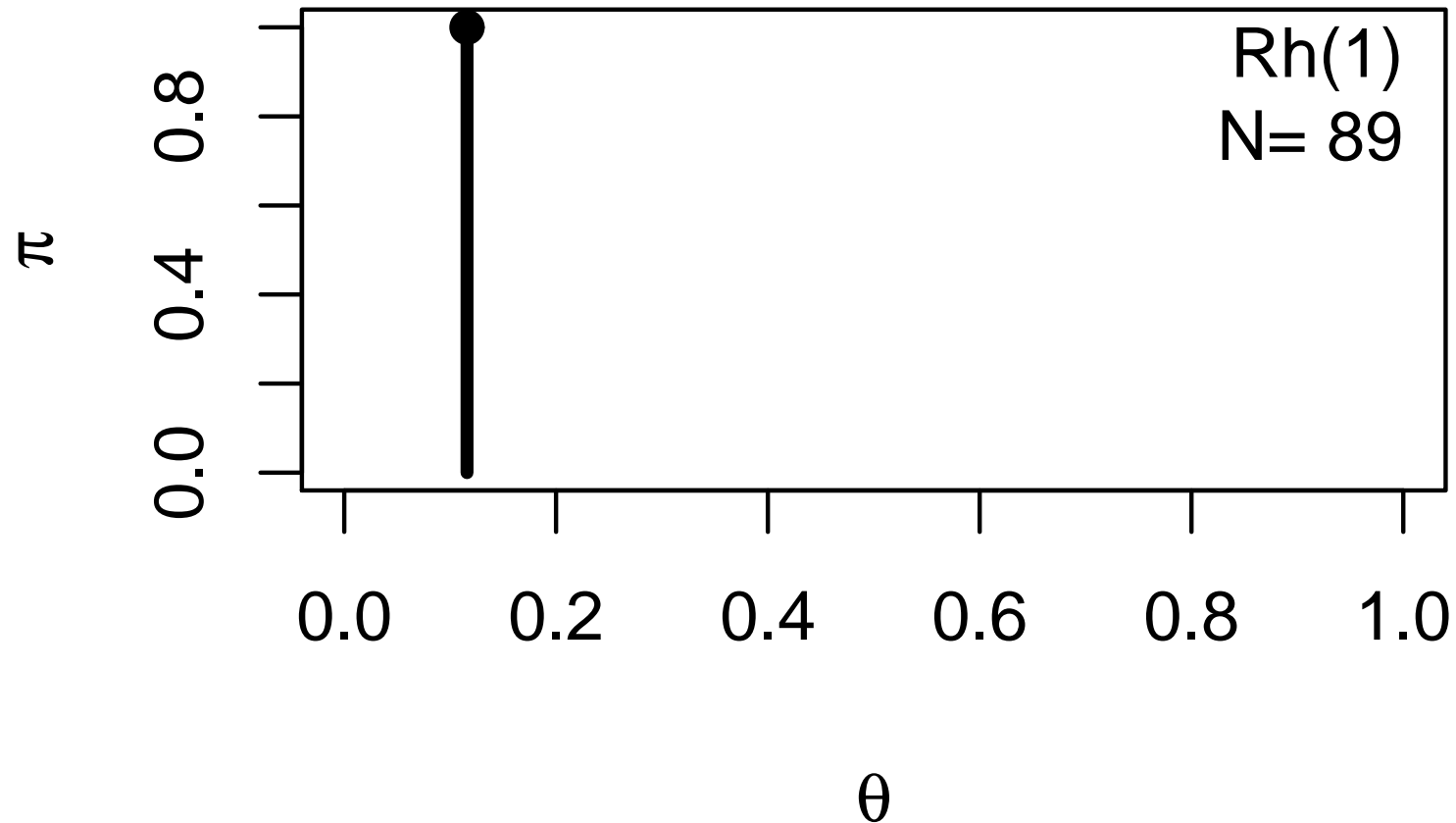
Step 50

$$\log L = -6.54071271014857$$



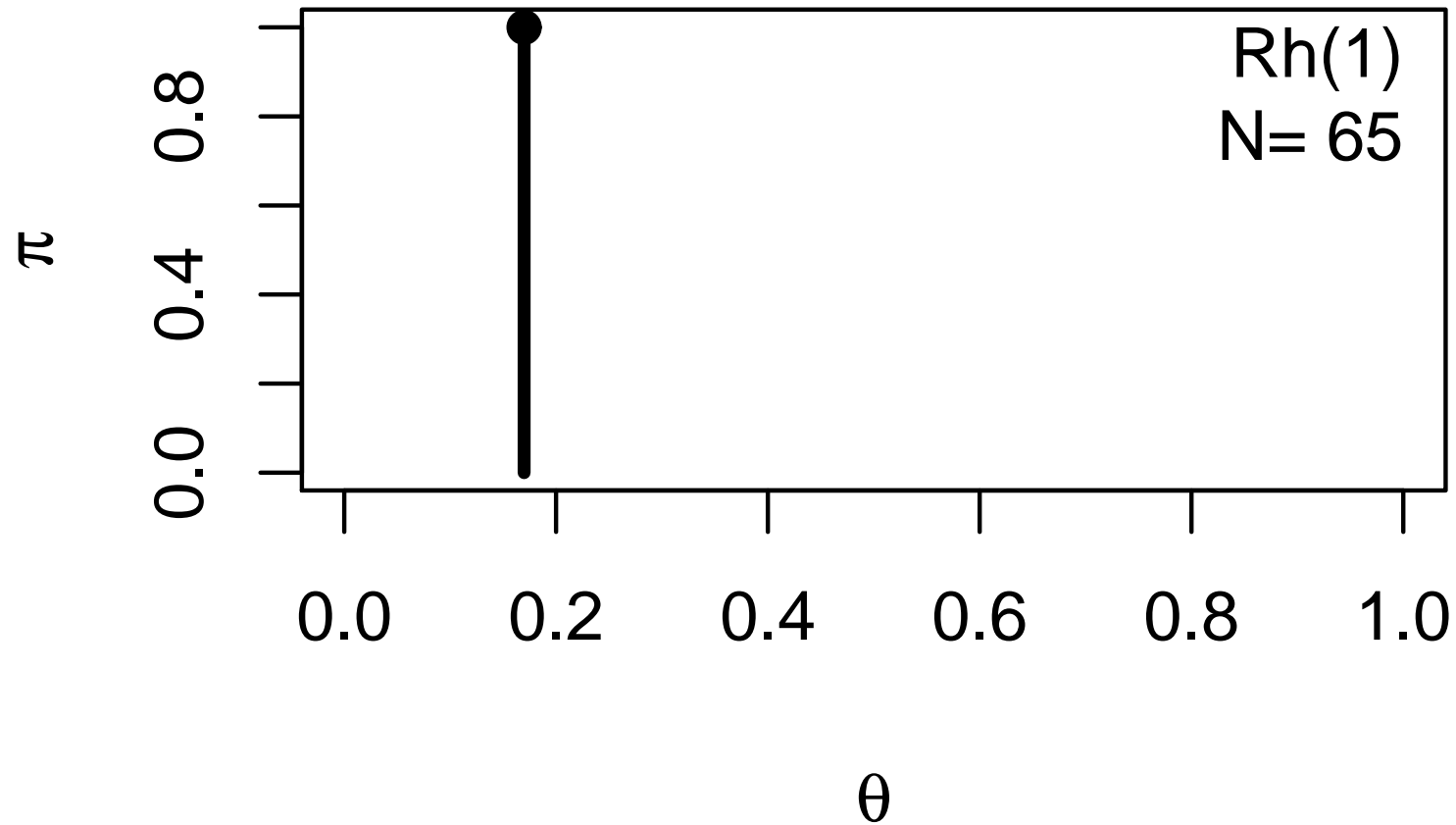
Step 51

$$\log L = -6.99200544598219$$



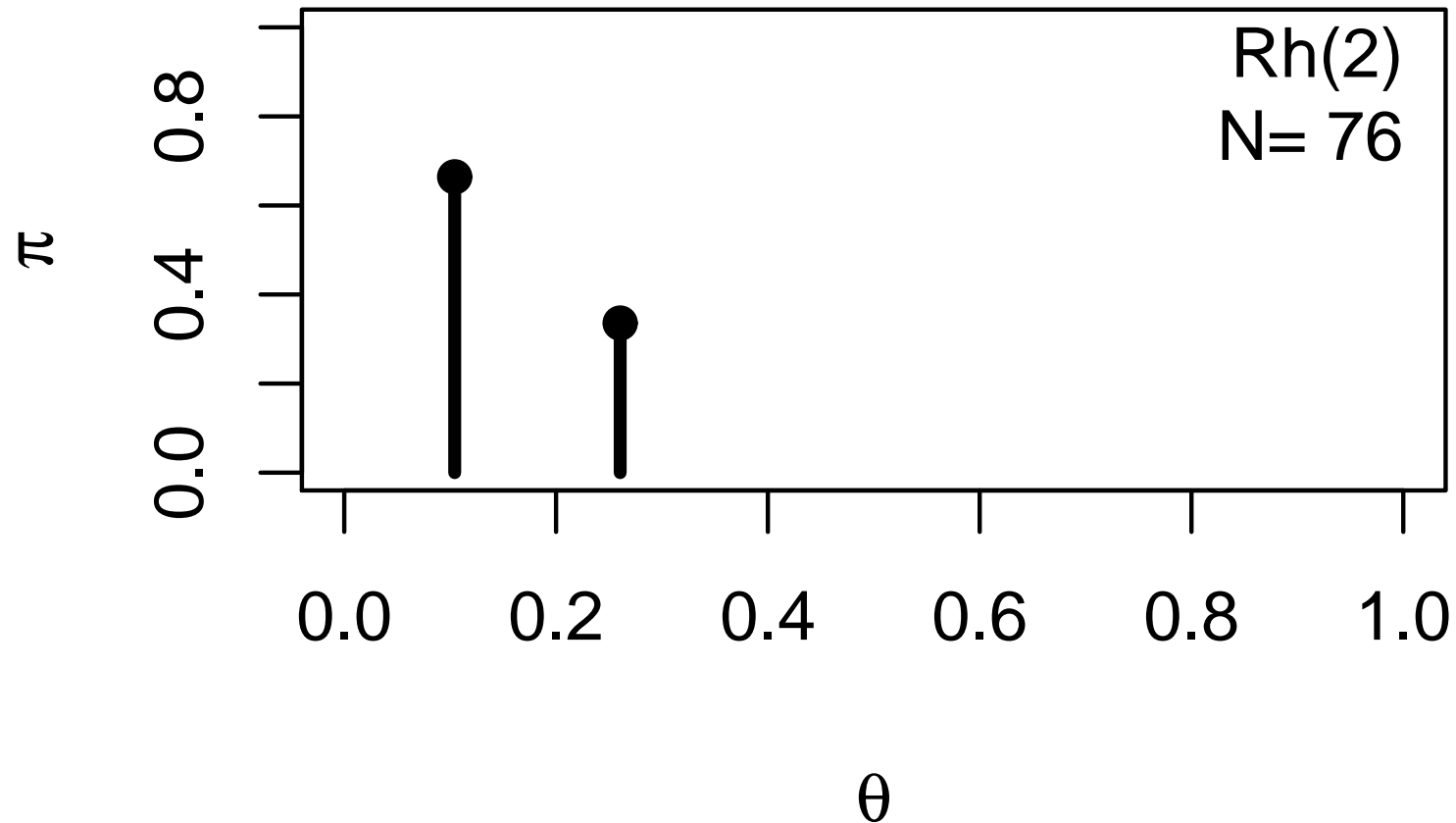
Step 52

$$\log L = -6.92828015307384$$



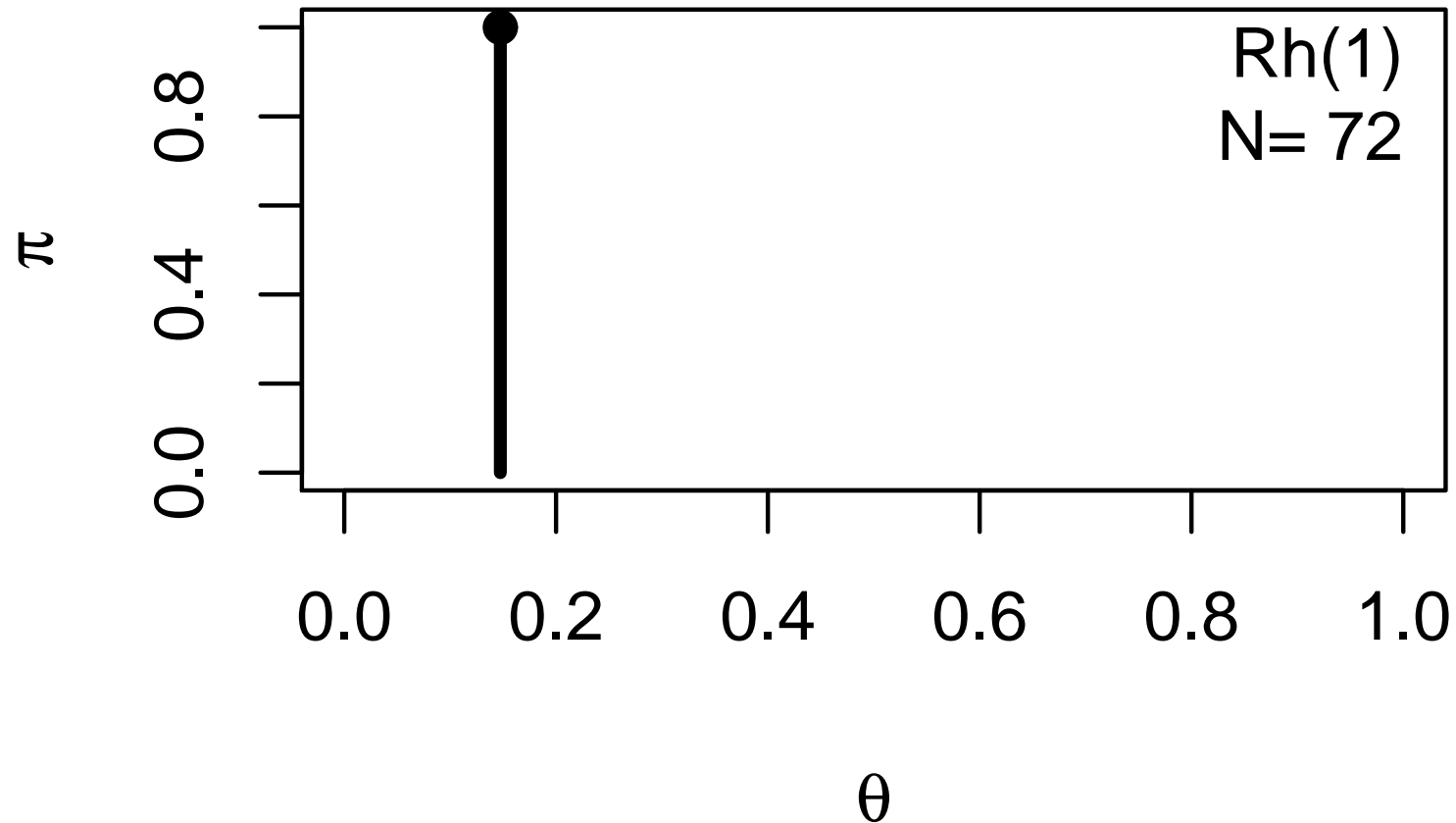
Step 53

$$\log L = -7.0832078891851$$



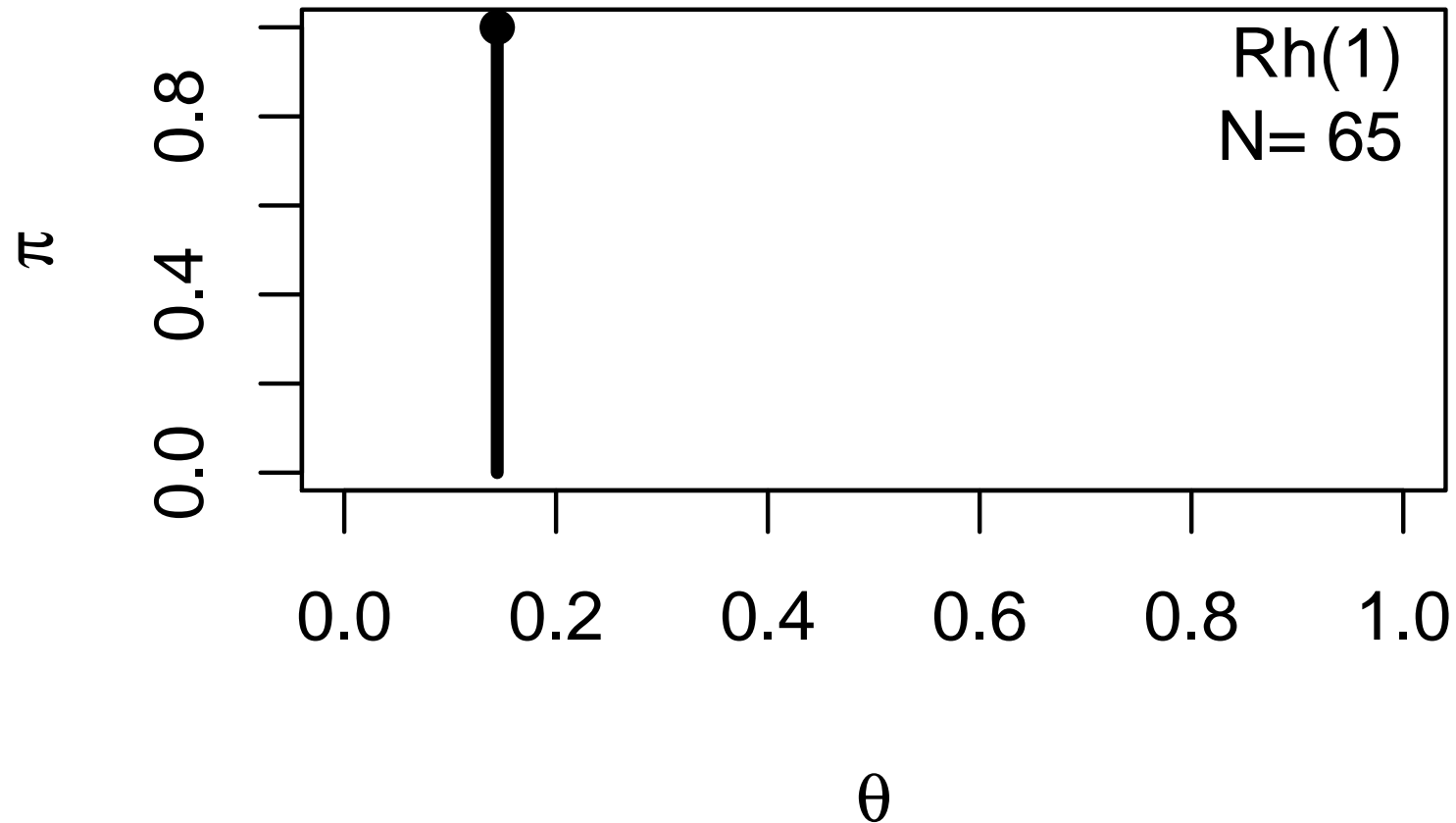
Step 54

$$\log L = -6.47554471024264$$



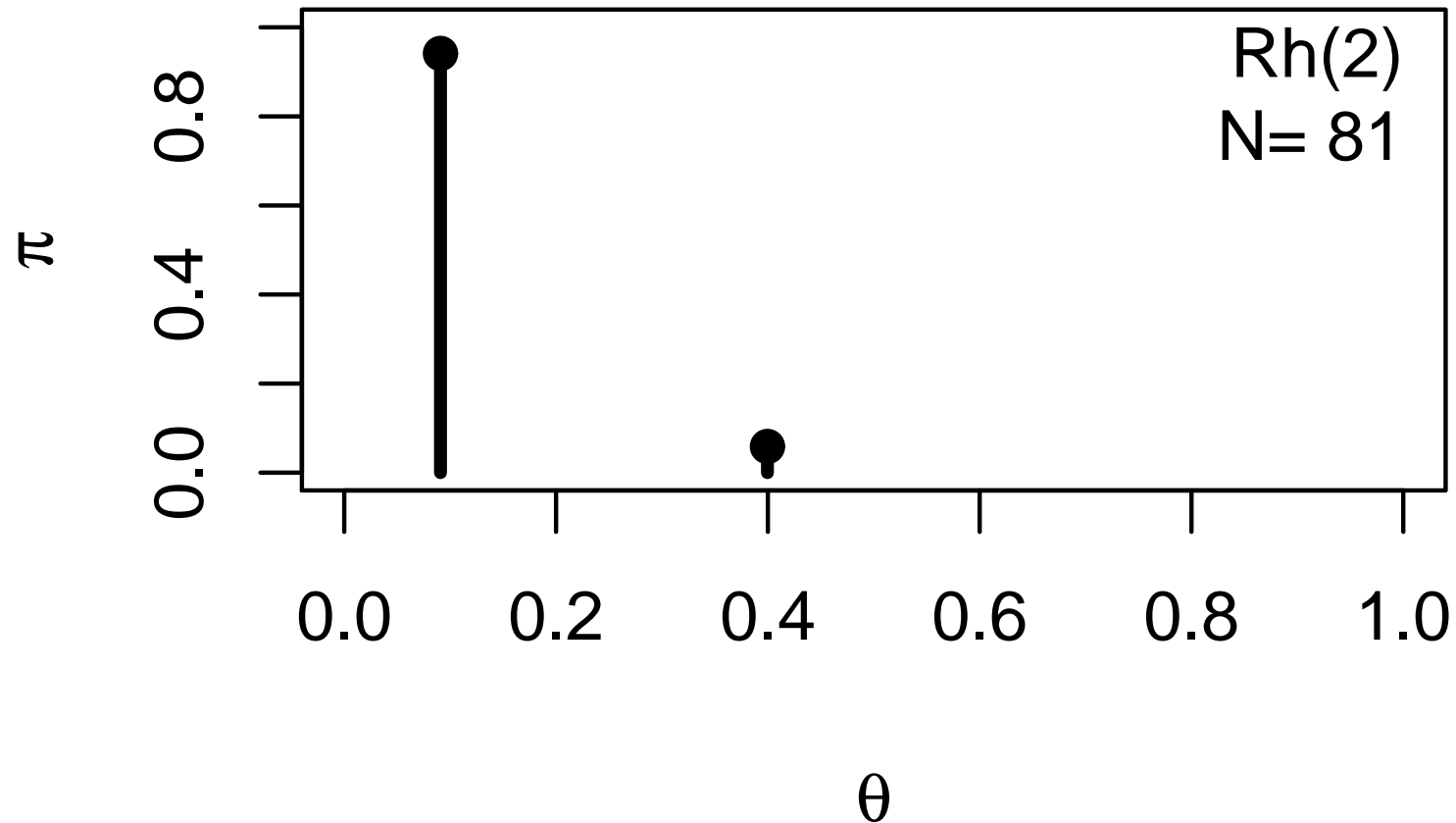
Step 55

$$\log L = -6.7014630905897$$



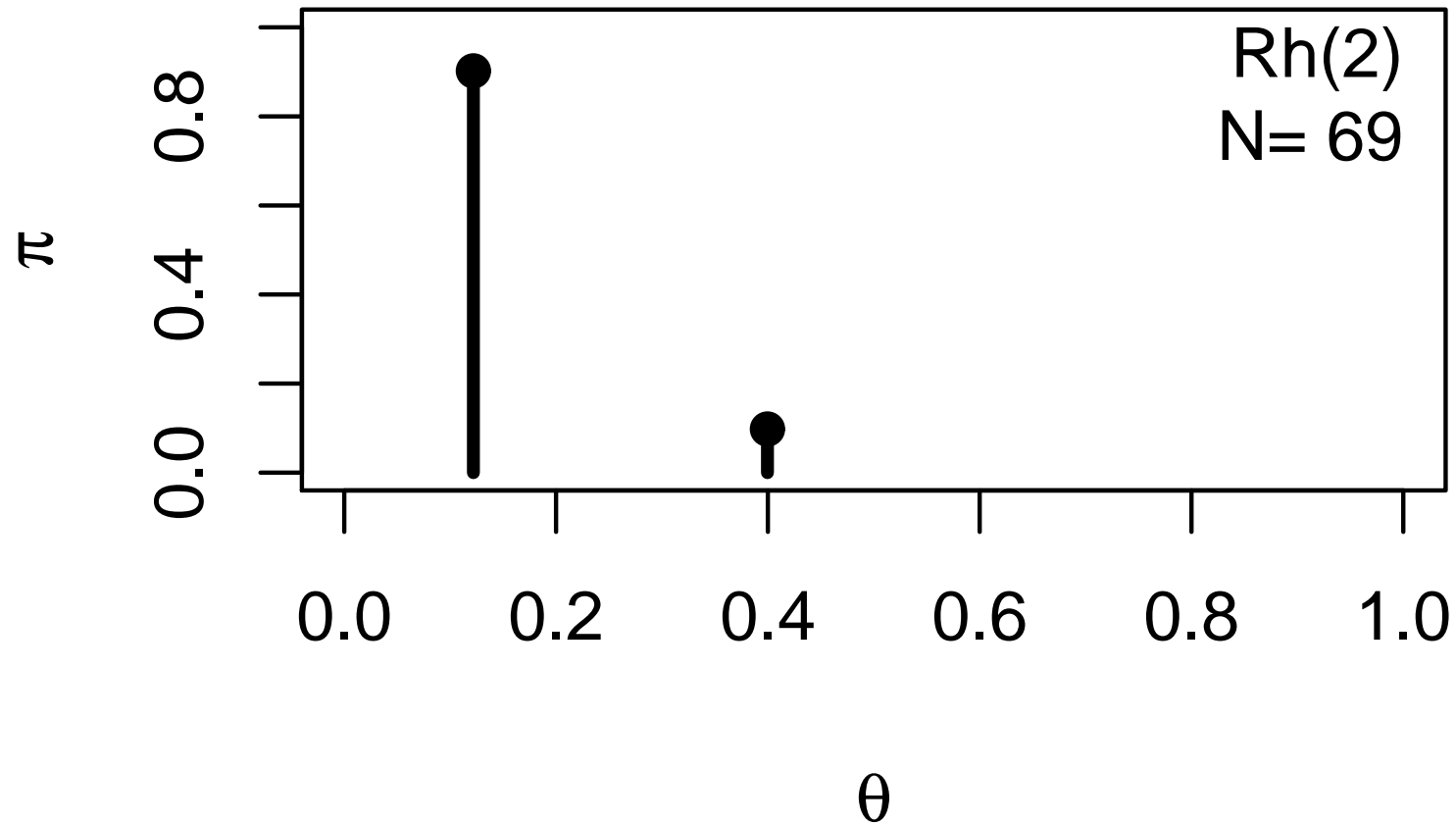
Step 56

$$\log L = -5.64197053930326$$



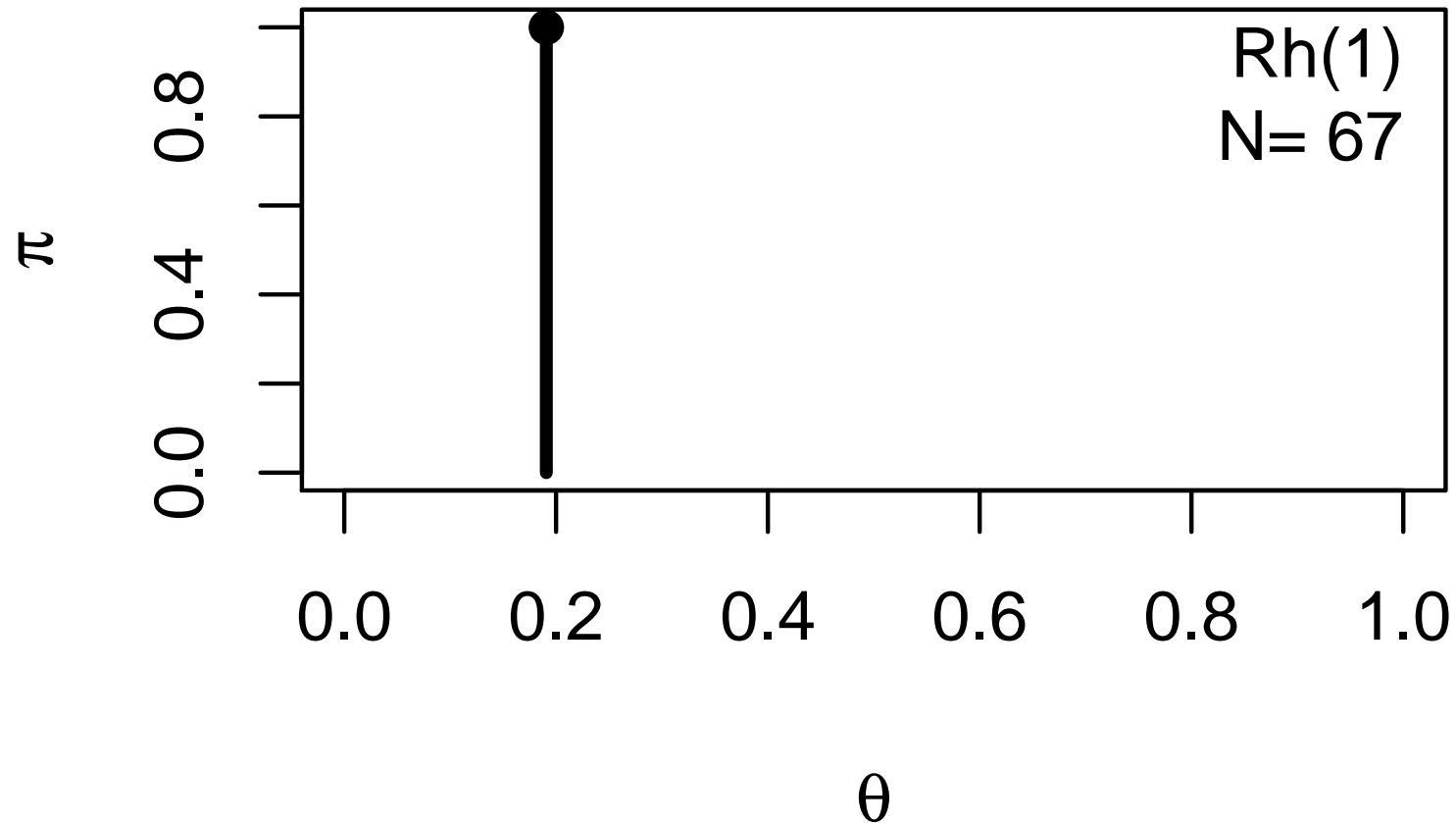
Step 57

$$\log L = -5.96373888768323$$



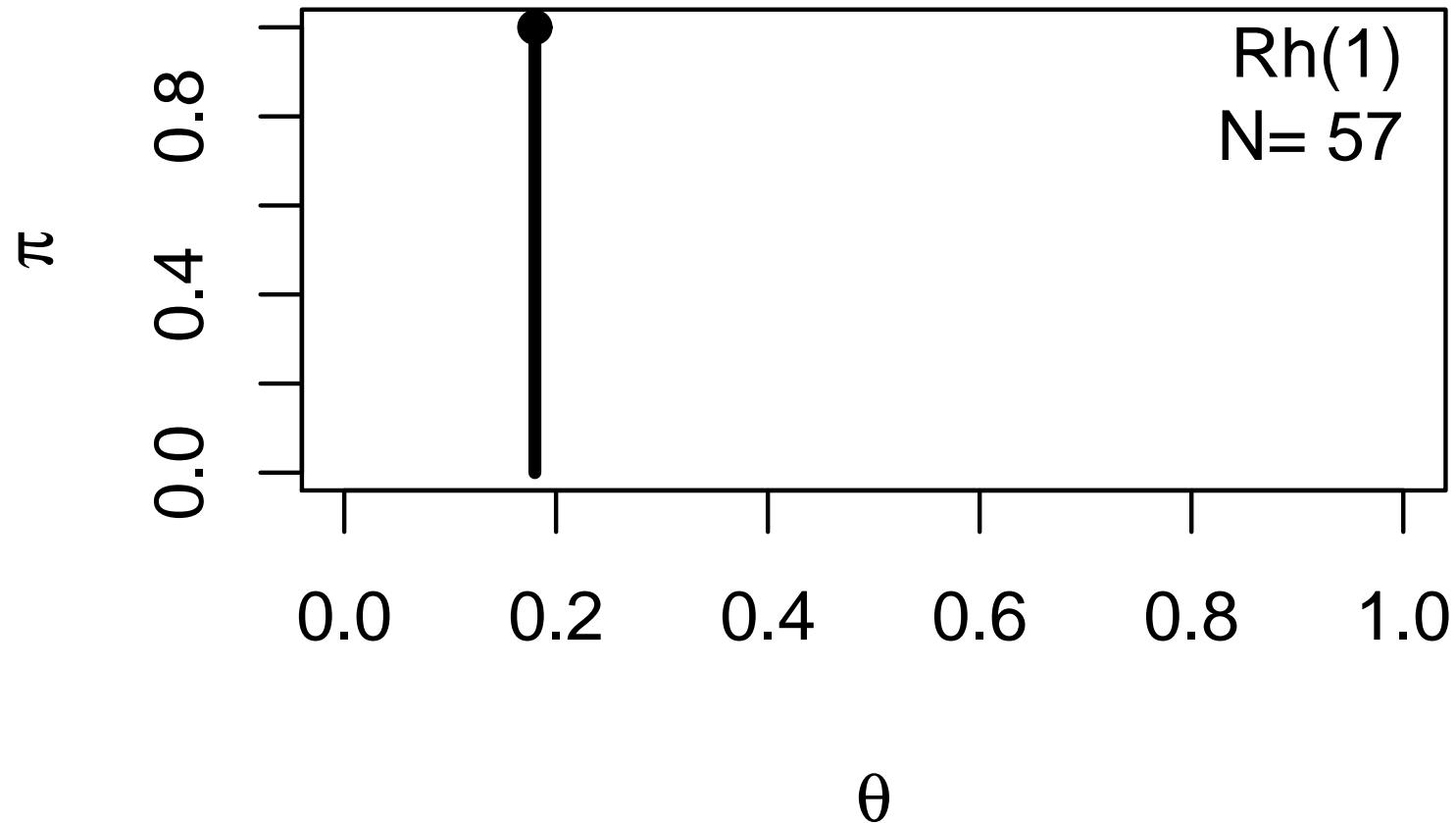
Step 58

$$\log L = -8.84591665324345$$



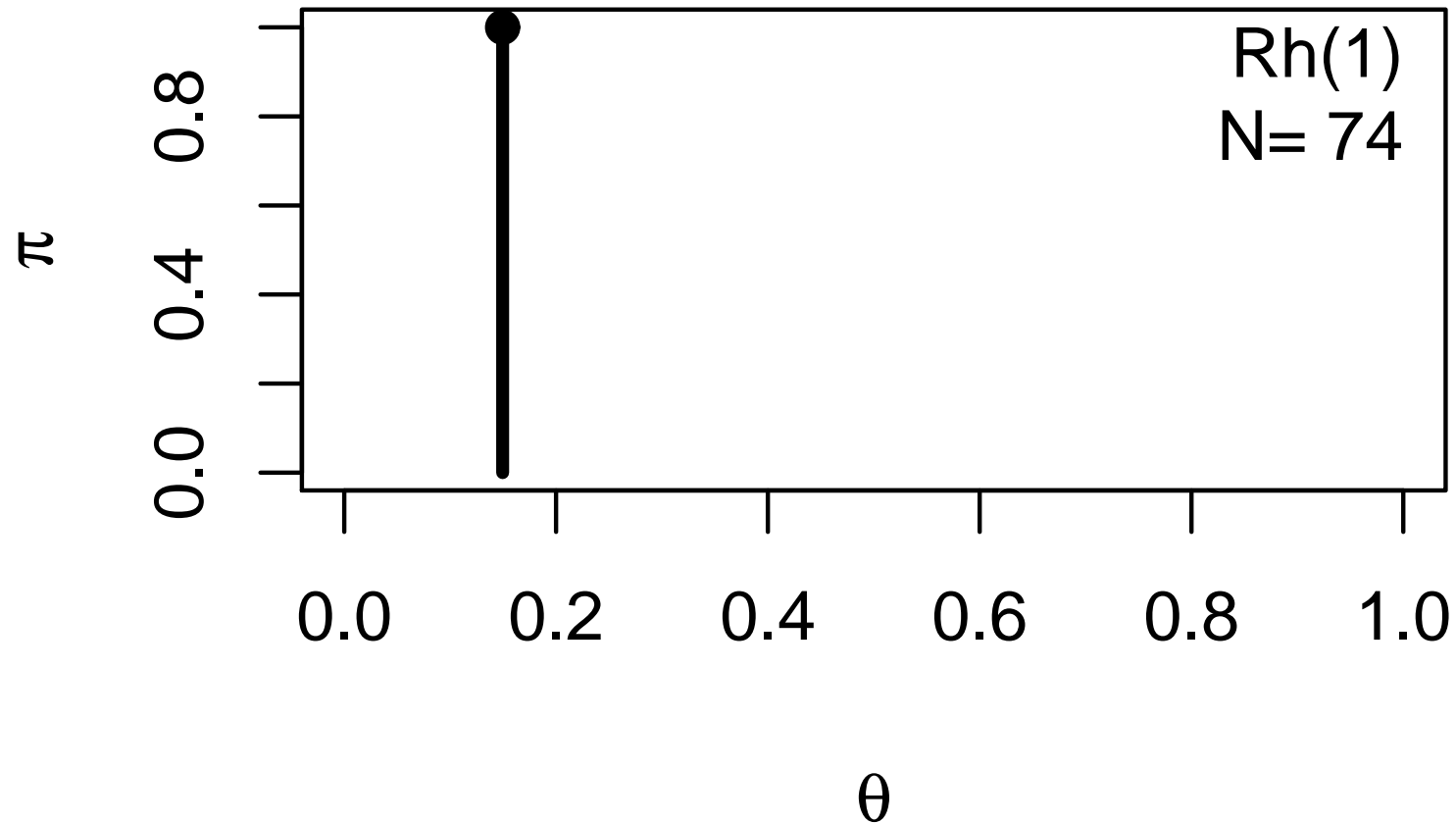
Step 59

$$\log L = -7.64672016924314$$



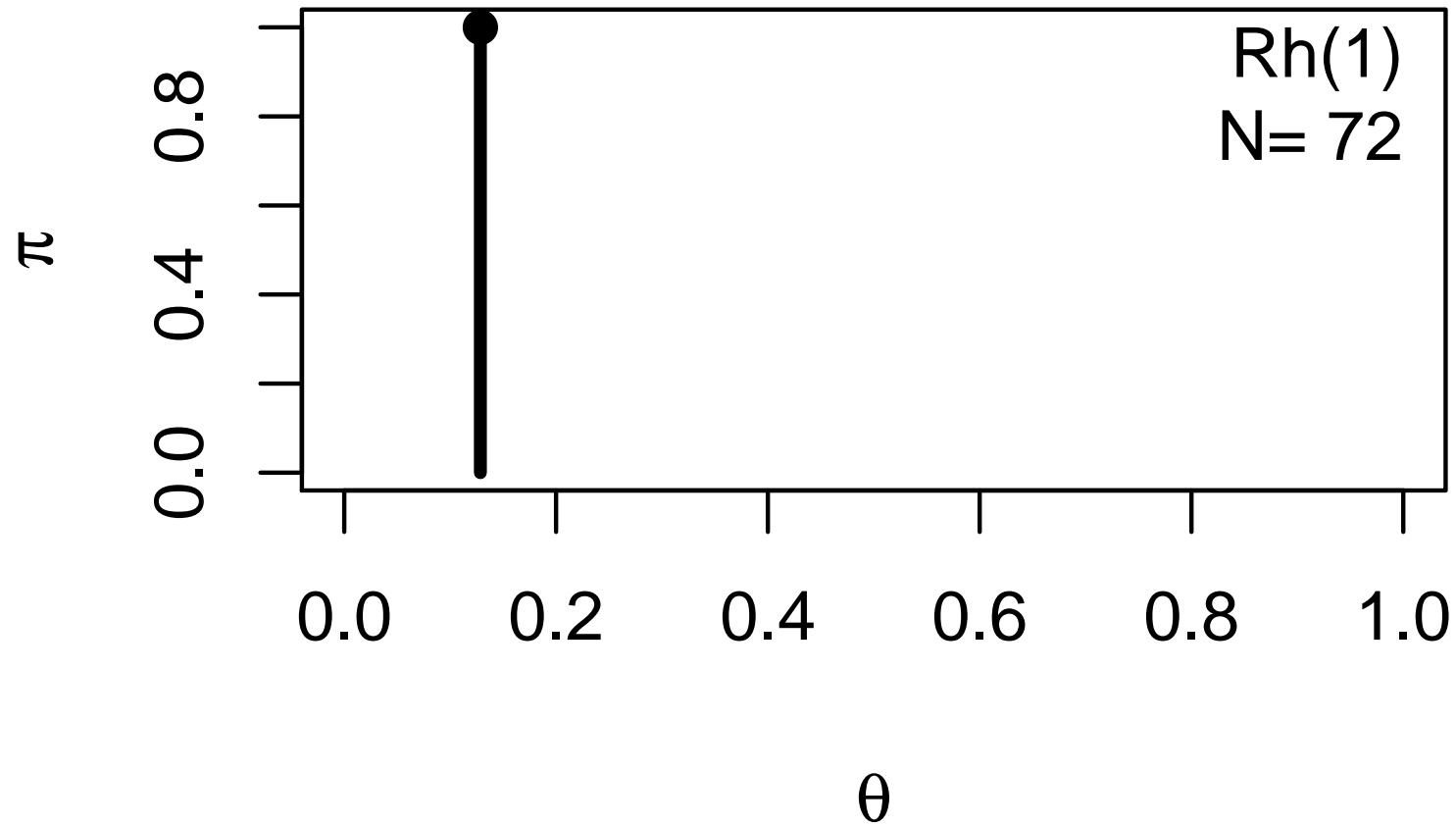
Step 60

$$\log L = -6.73174374308653$$



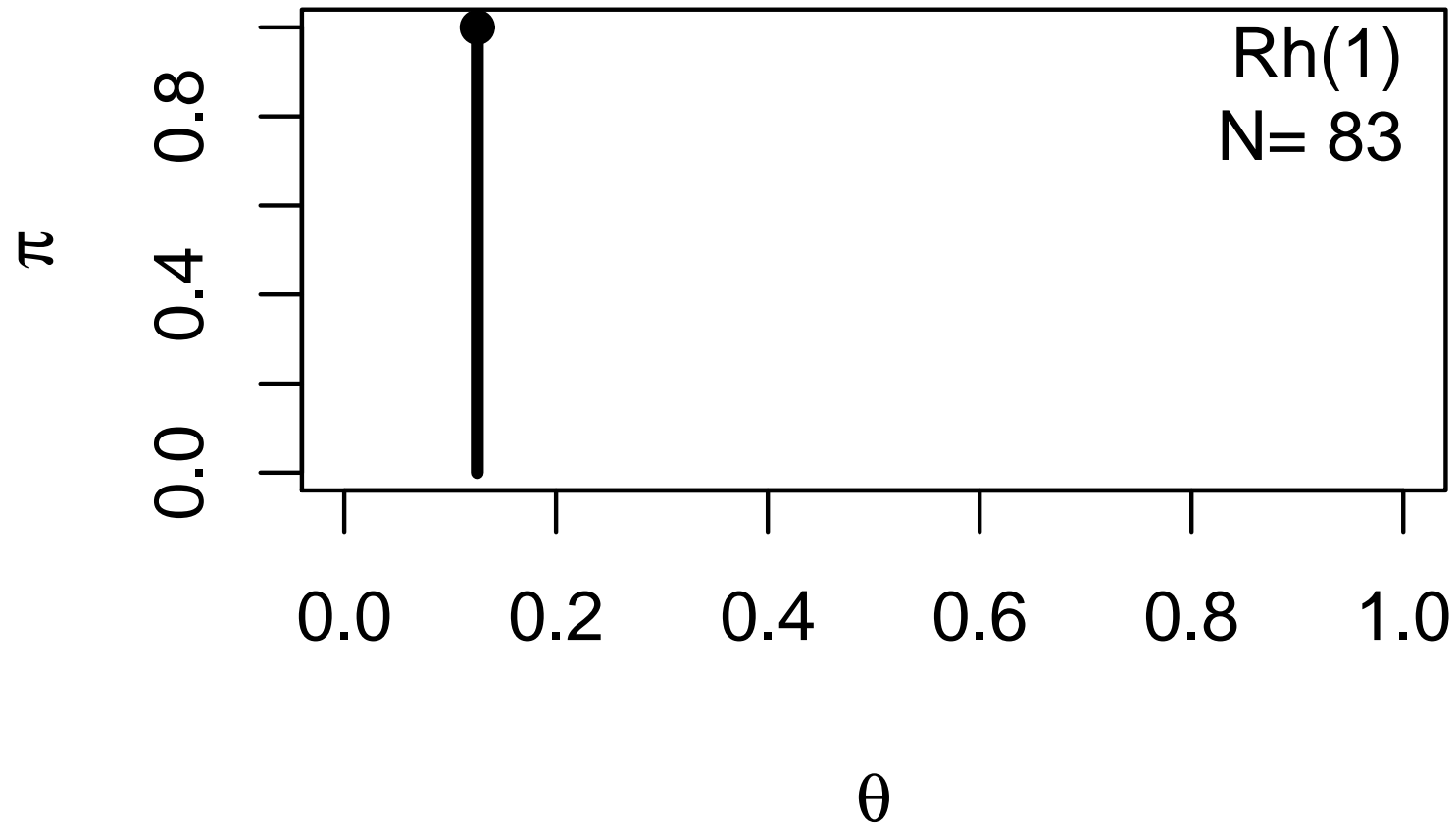
Step 61

$$\log L = -6.55142574603127$$



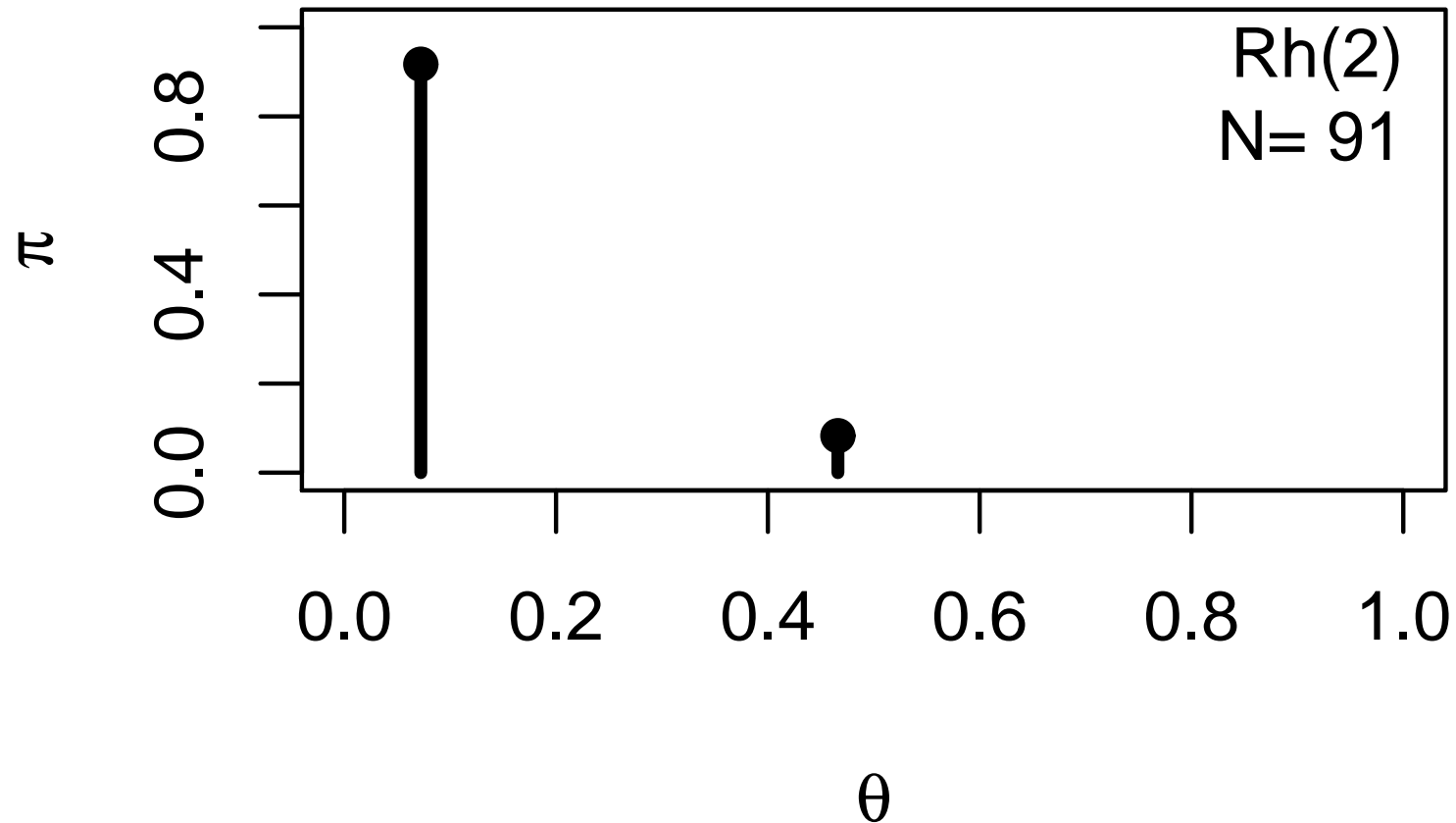
Step 62

$$\log L = -6.69541020421036$$



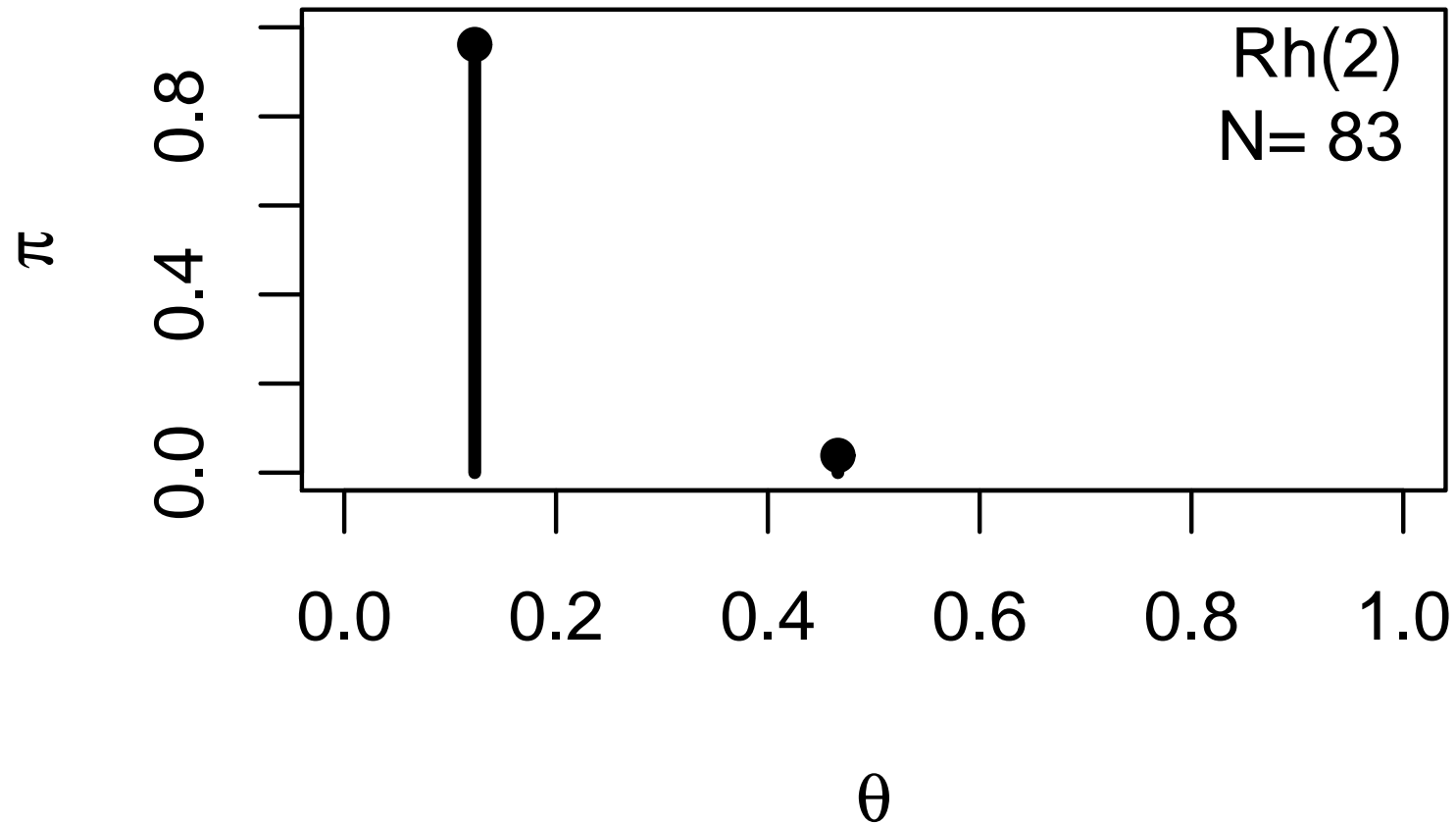
Step 63

$$\log L = -6.8475622252277$$



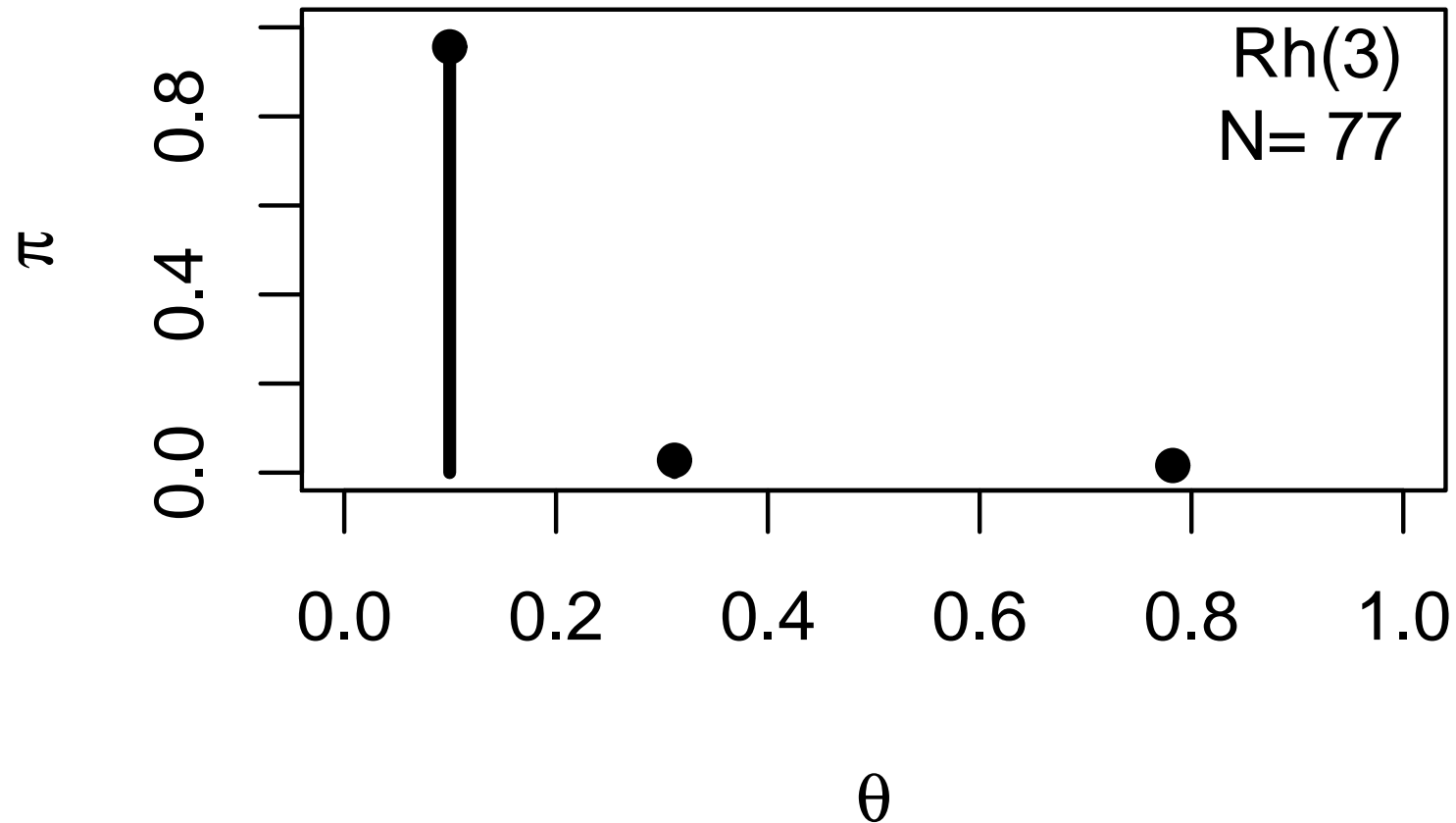
Step 64

$$\log L = -5.22857102139116$$



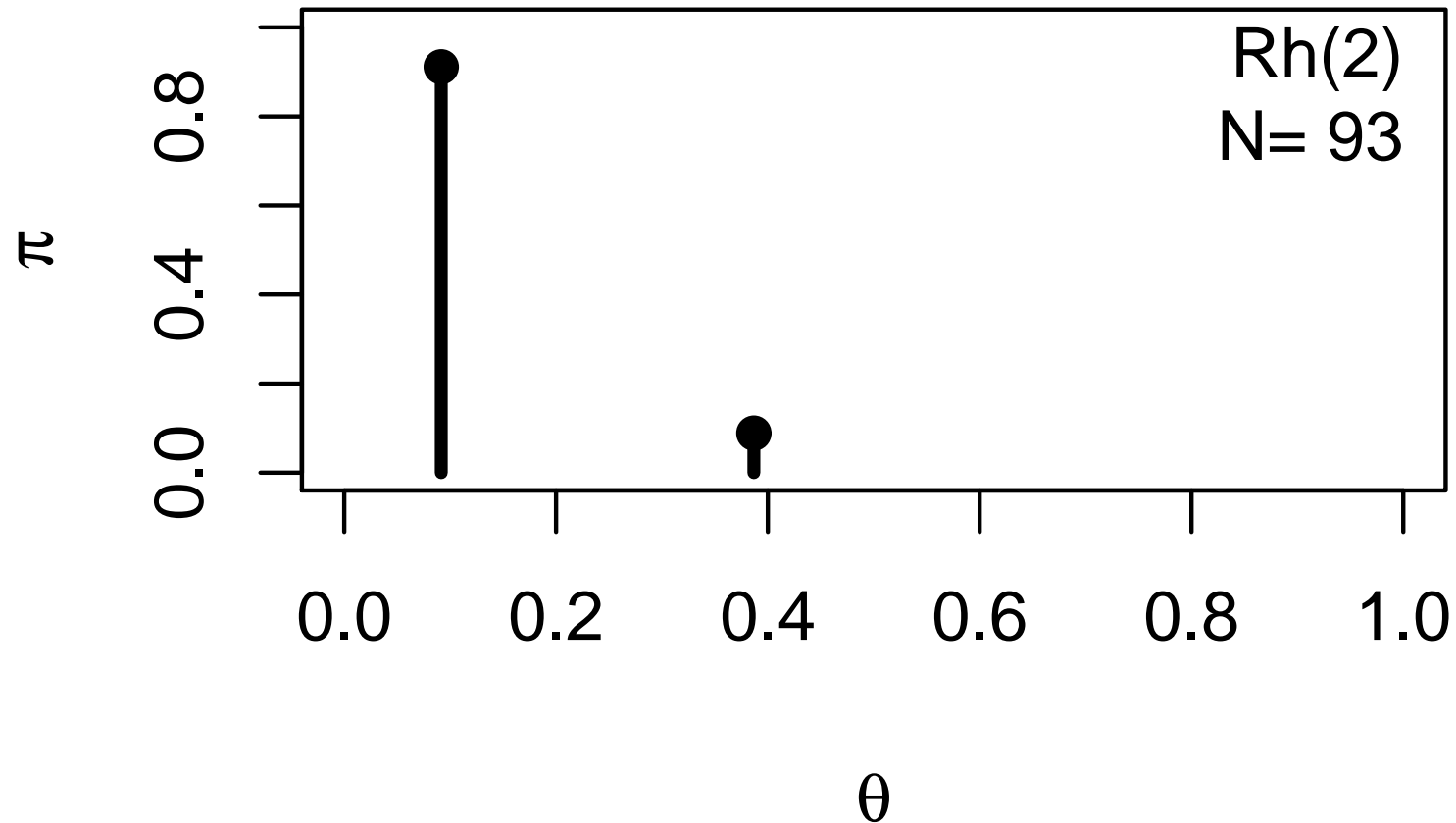
Step 65

$$\log L = -4.81063005314064$$



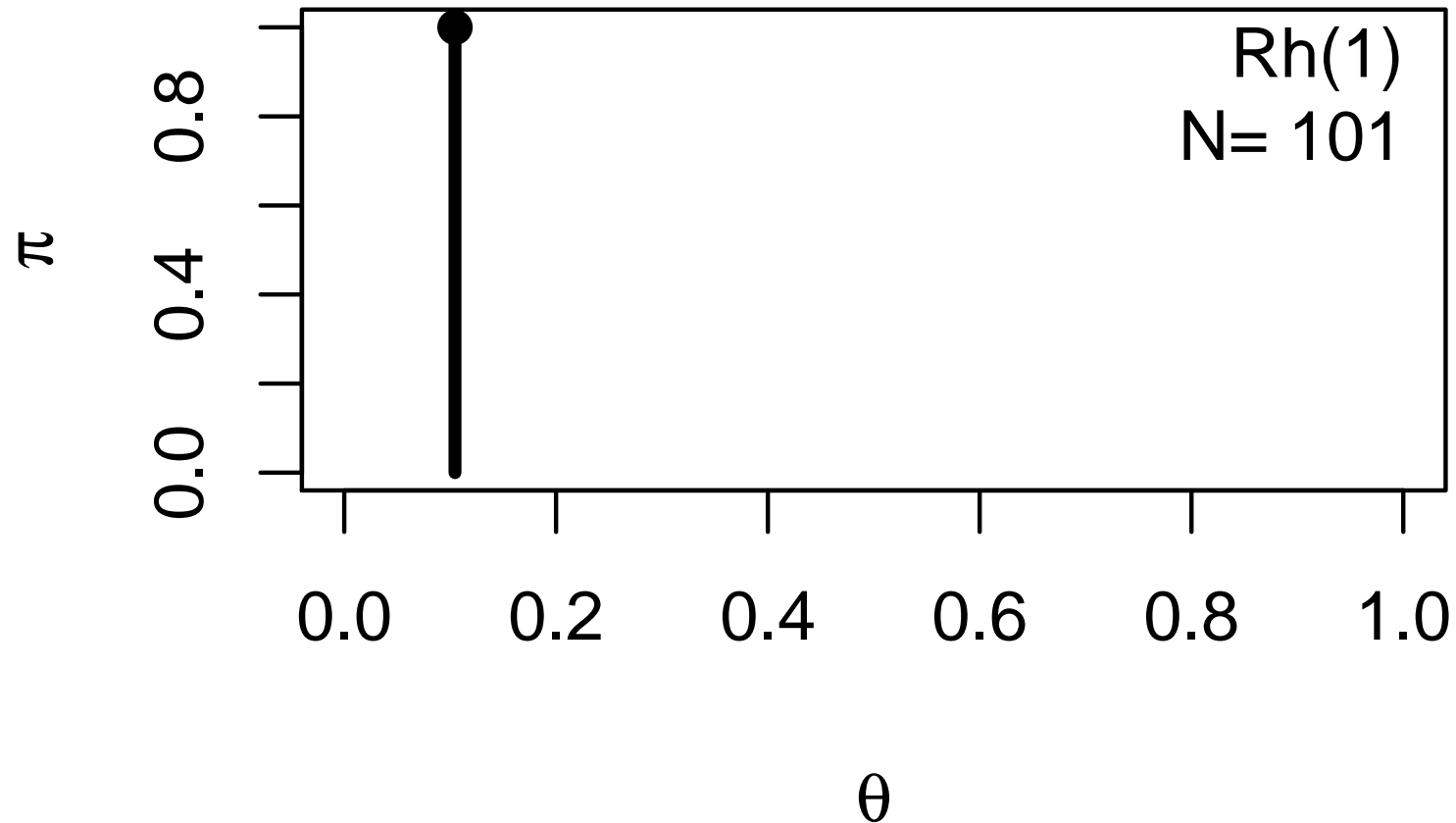
Step 66

$$\log L = -5.46413707072475$$



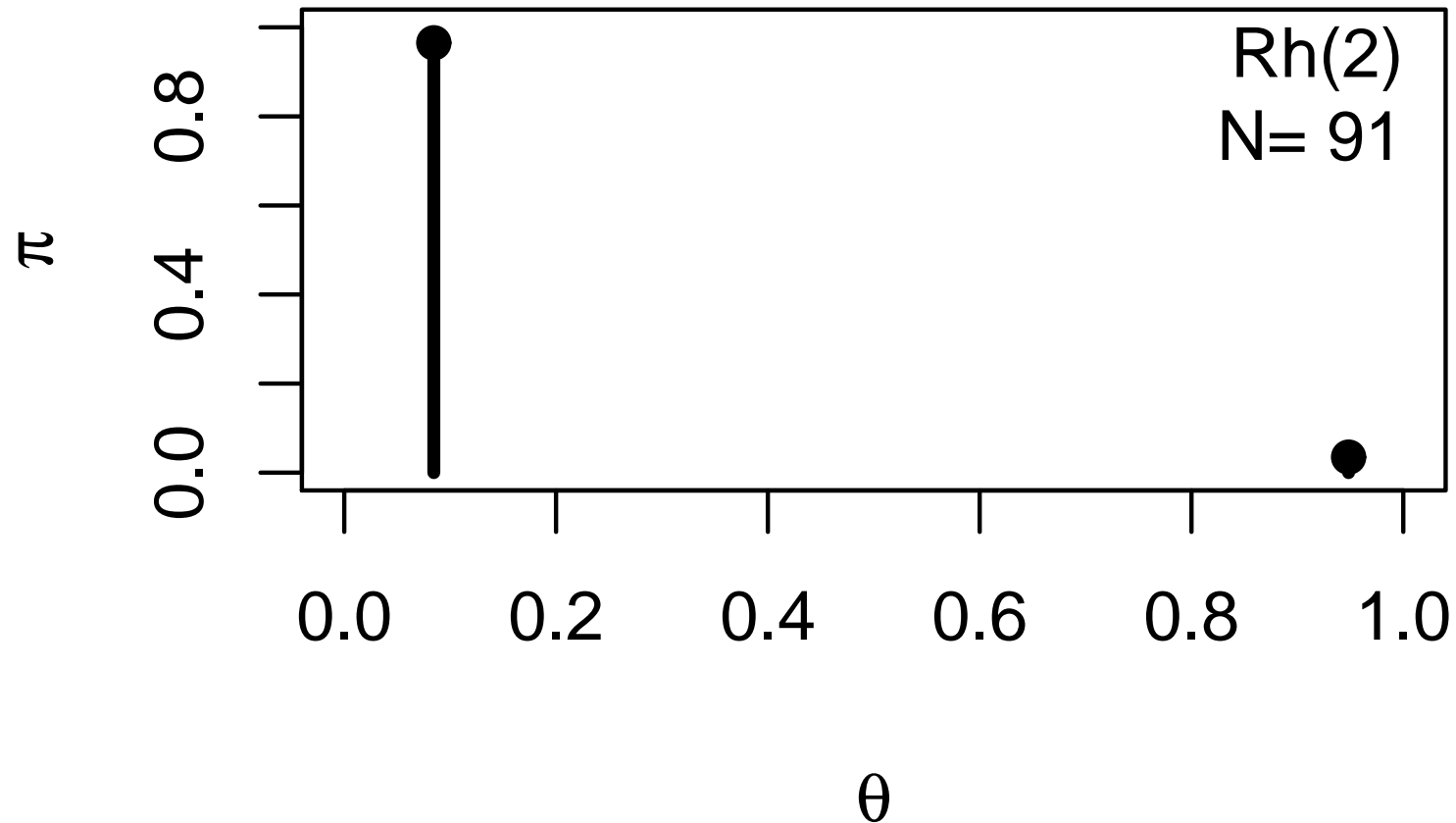
Step 67

$$\log L = -7.88242747736024$$



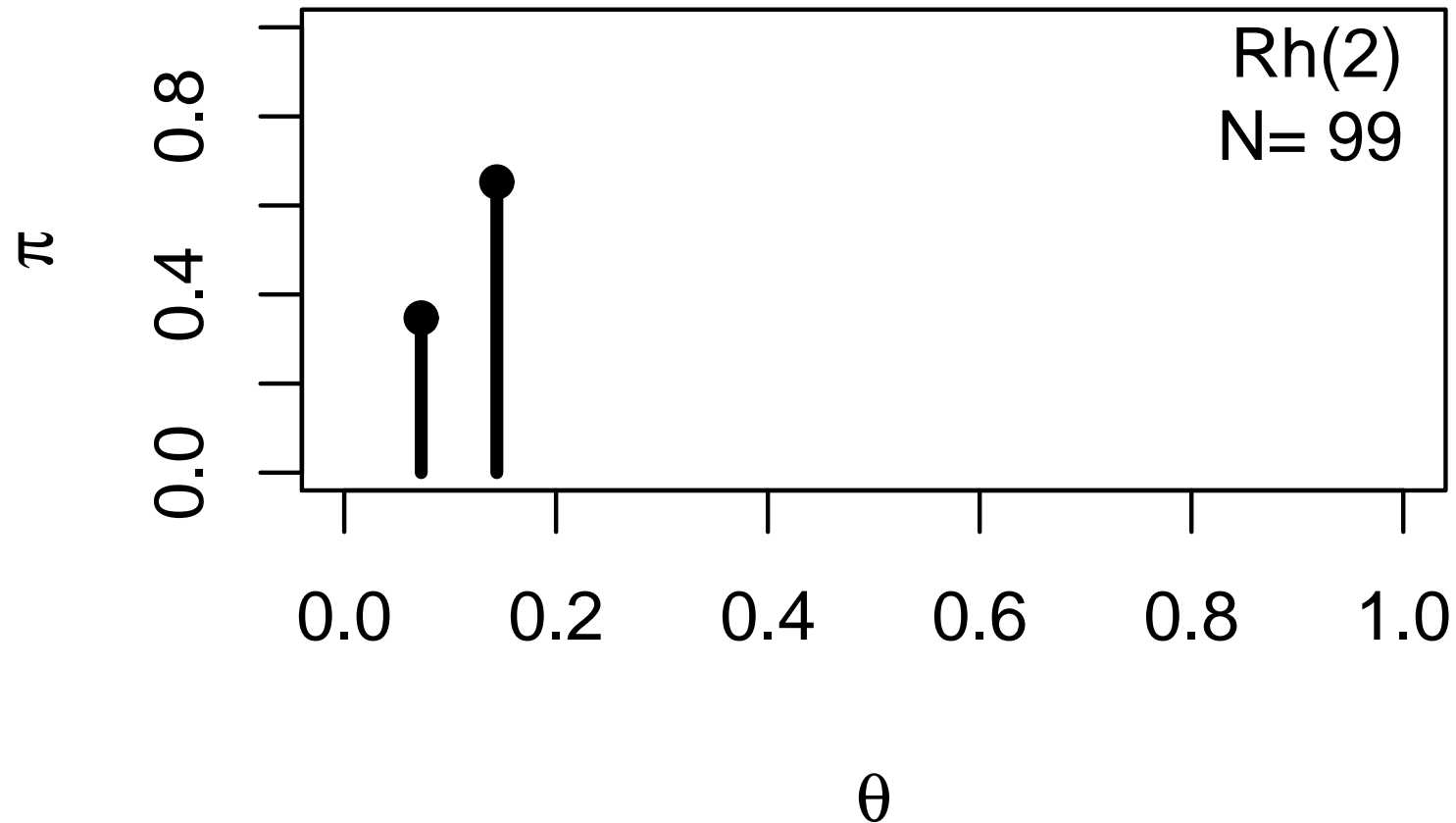
Step 68

$$\log L = -6.48263347003243$$



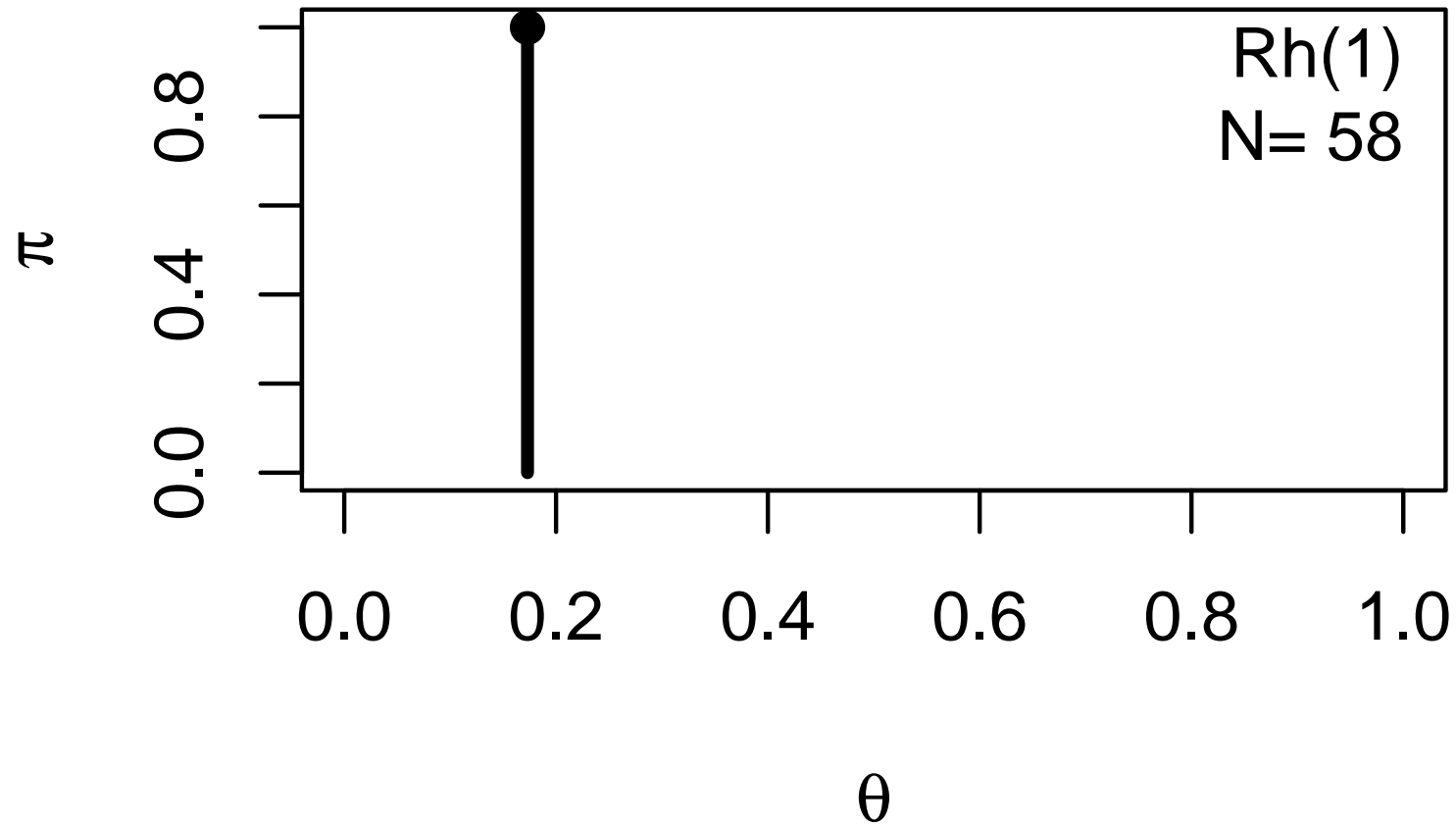
Step 69

$$\log L = -7.91704677251013$$



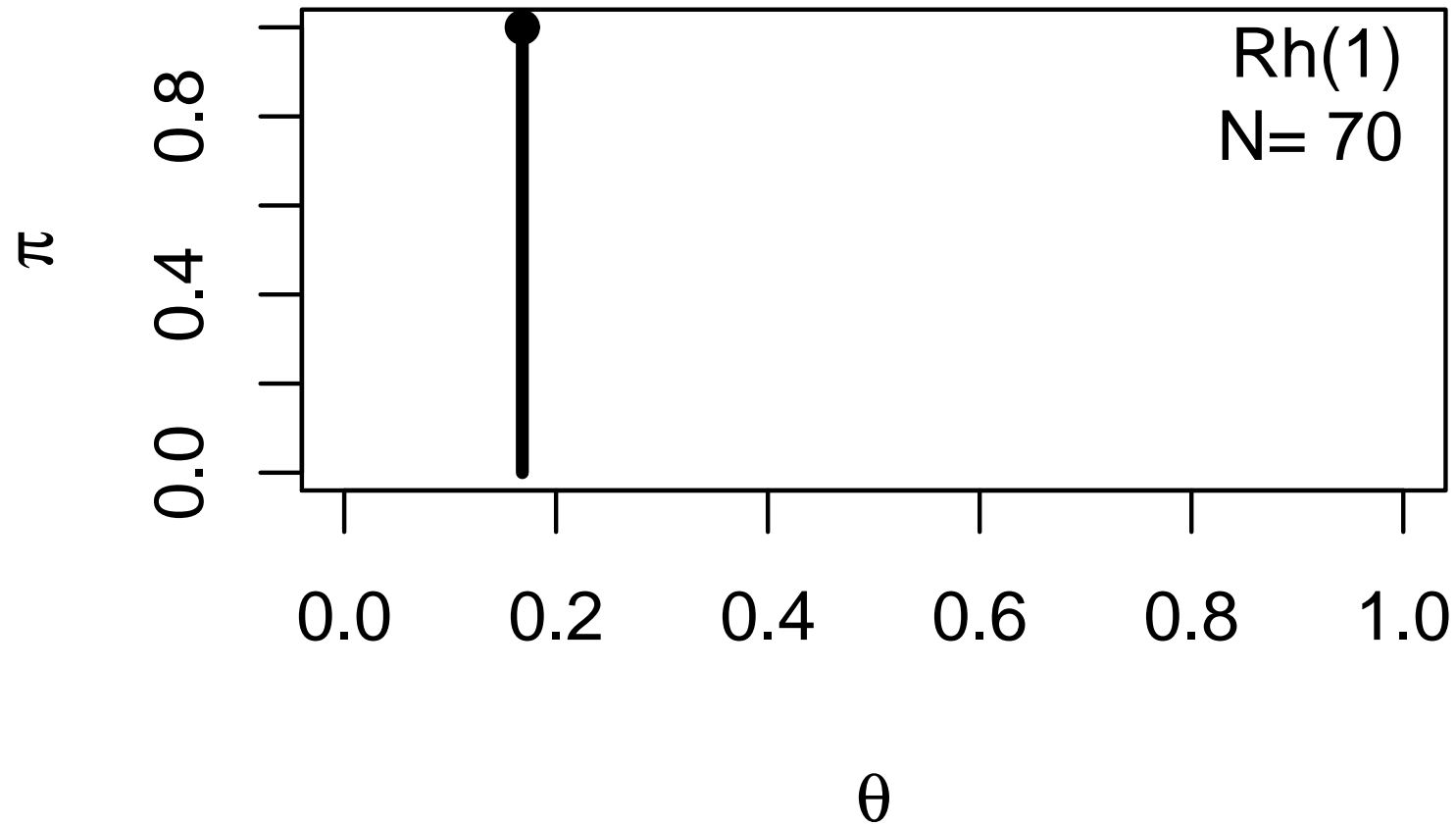
Step 70

$$\log L = -7.41664781432527$$



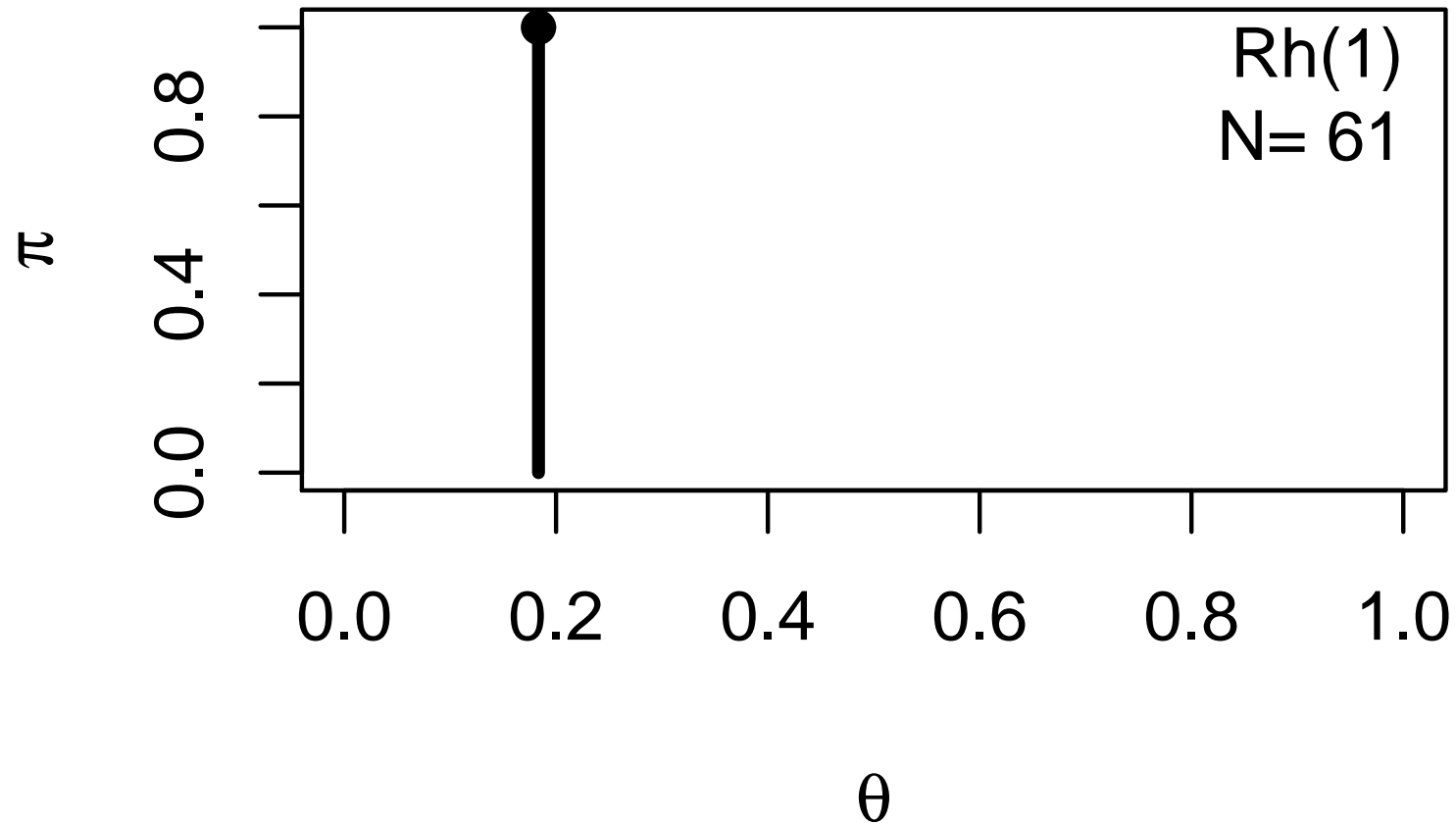
Step 71

$$\log L = -7.36978157074631$$



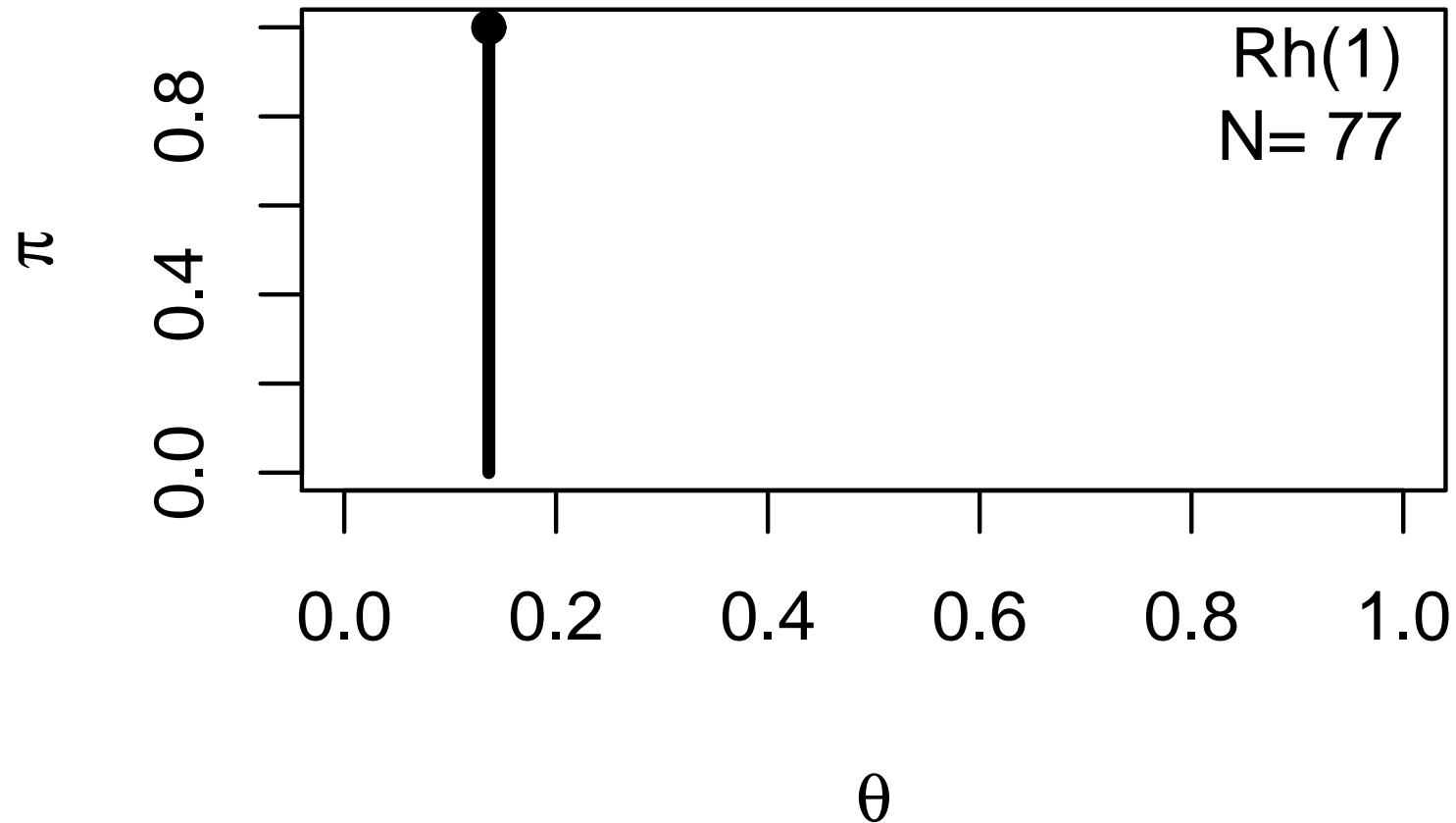
Step 72

$$\log L = -7.42029372486004$$



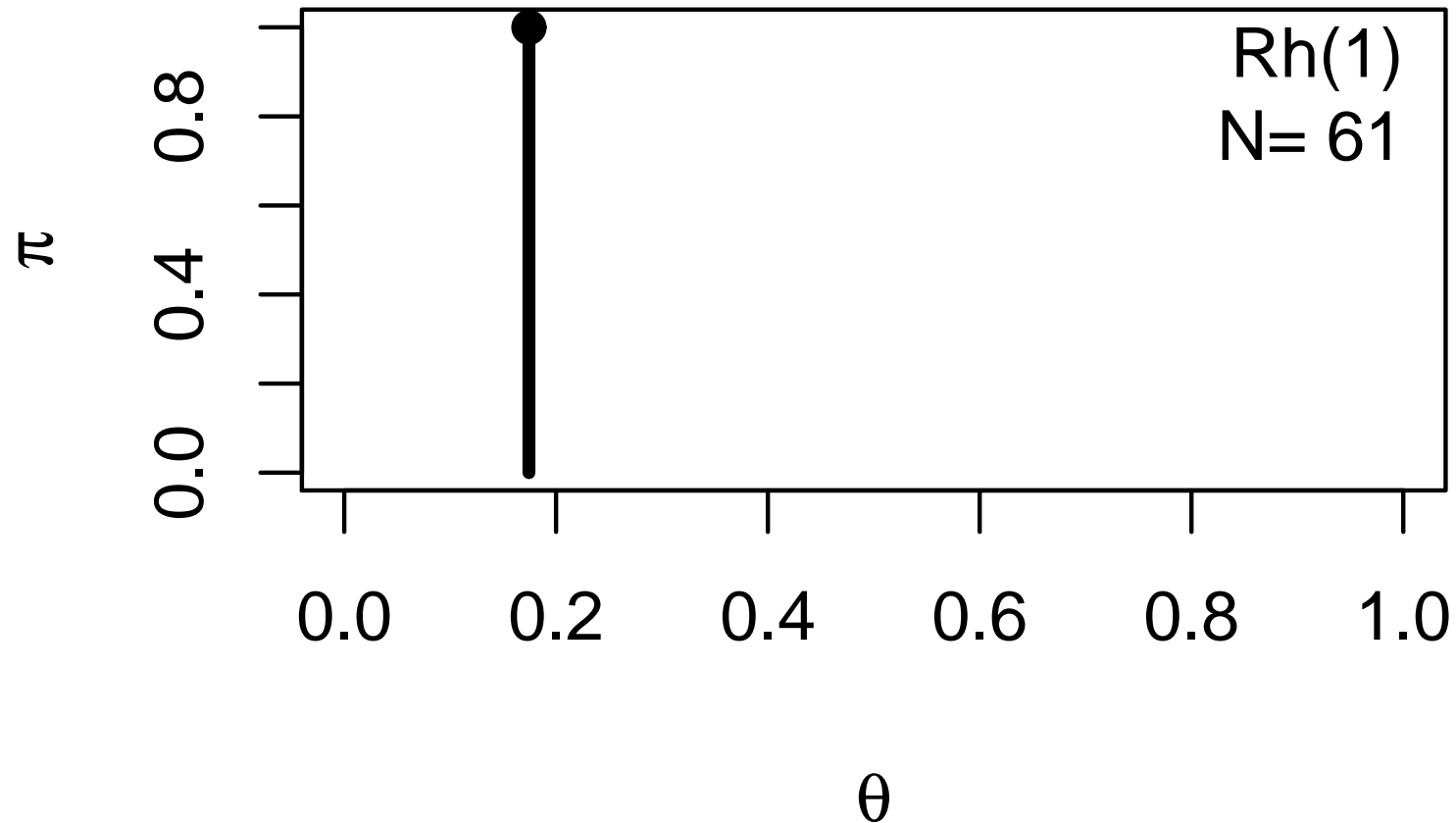
Step 73

$$\log L = -6.50030323887847$$



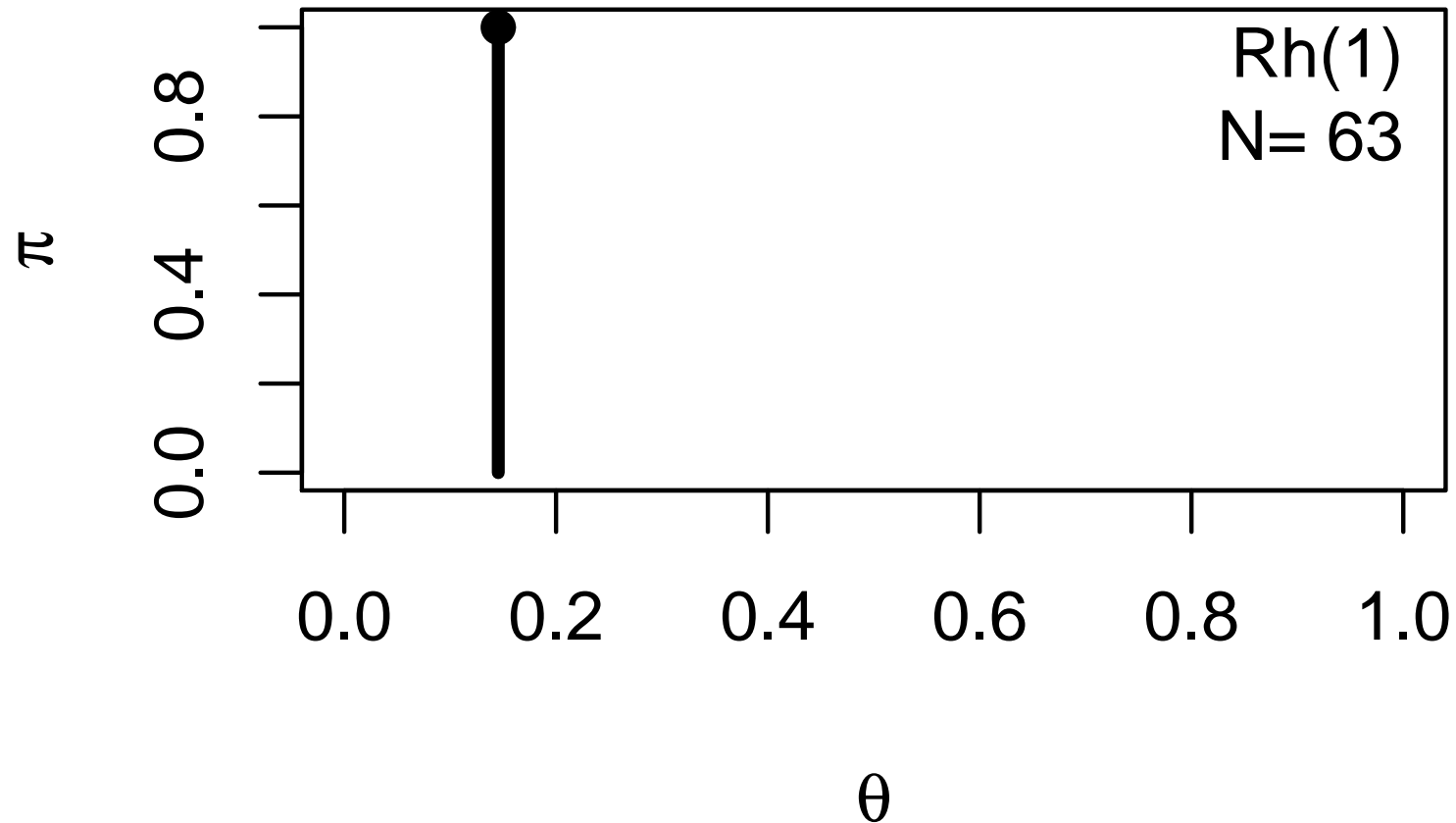
Step 74

$$\log L = -7.08379337128636$$



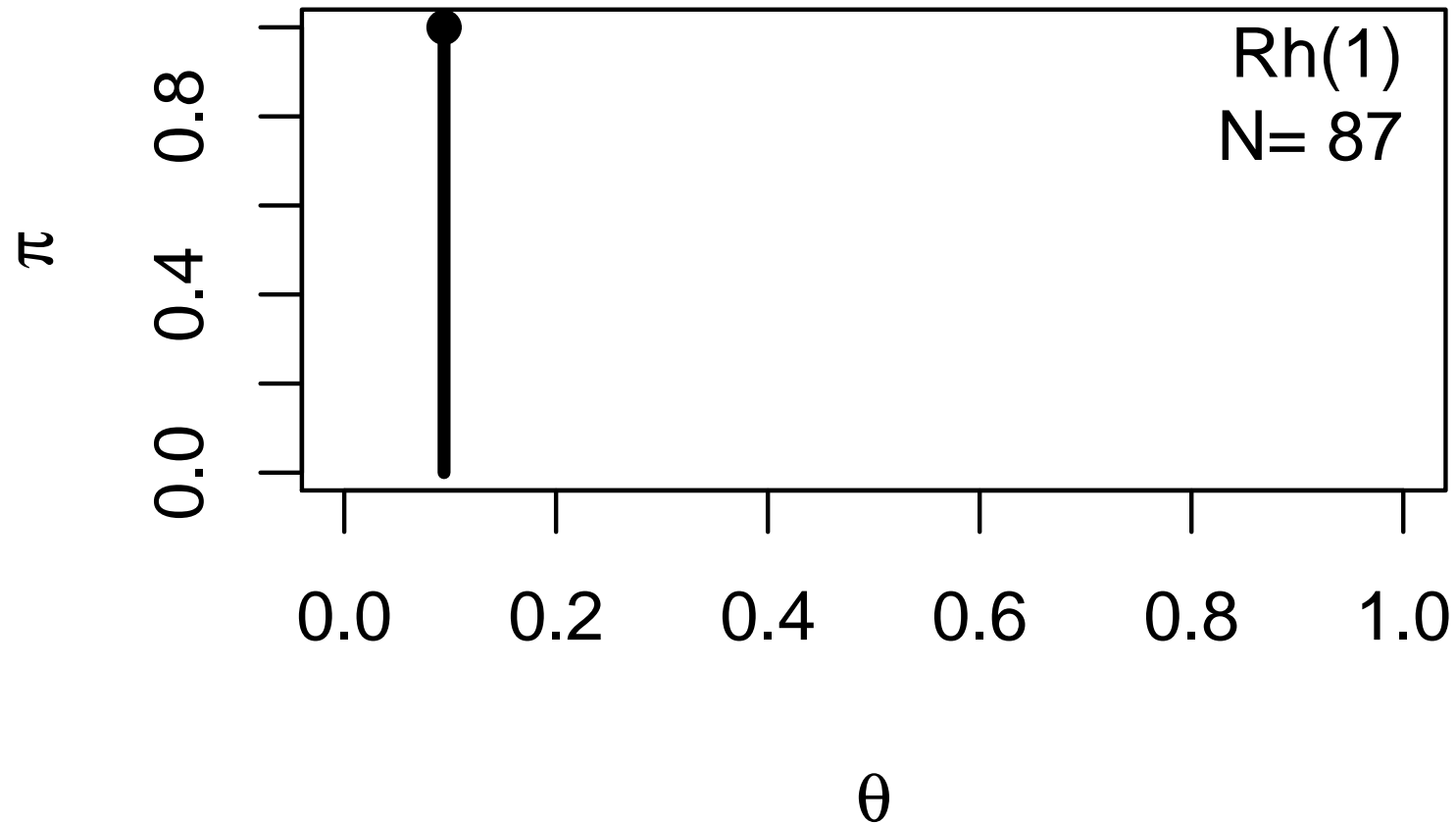
Step 75

$$\log L = -6.98594579627081$$



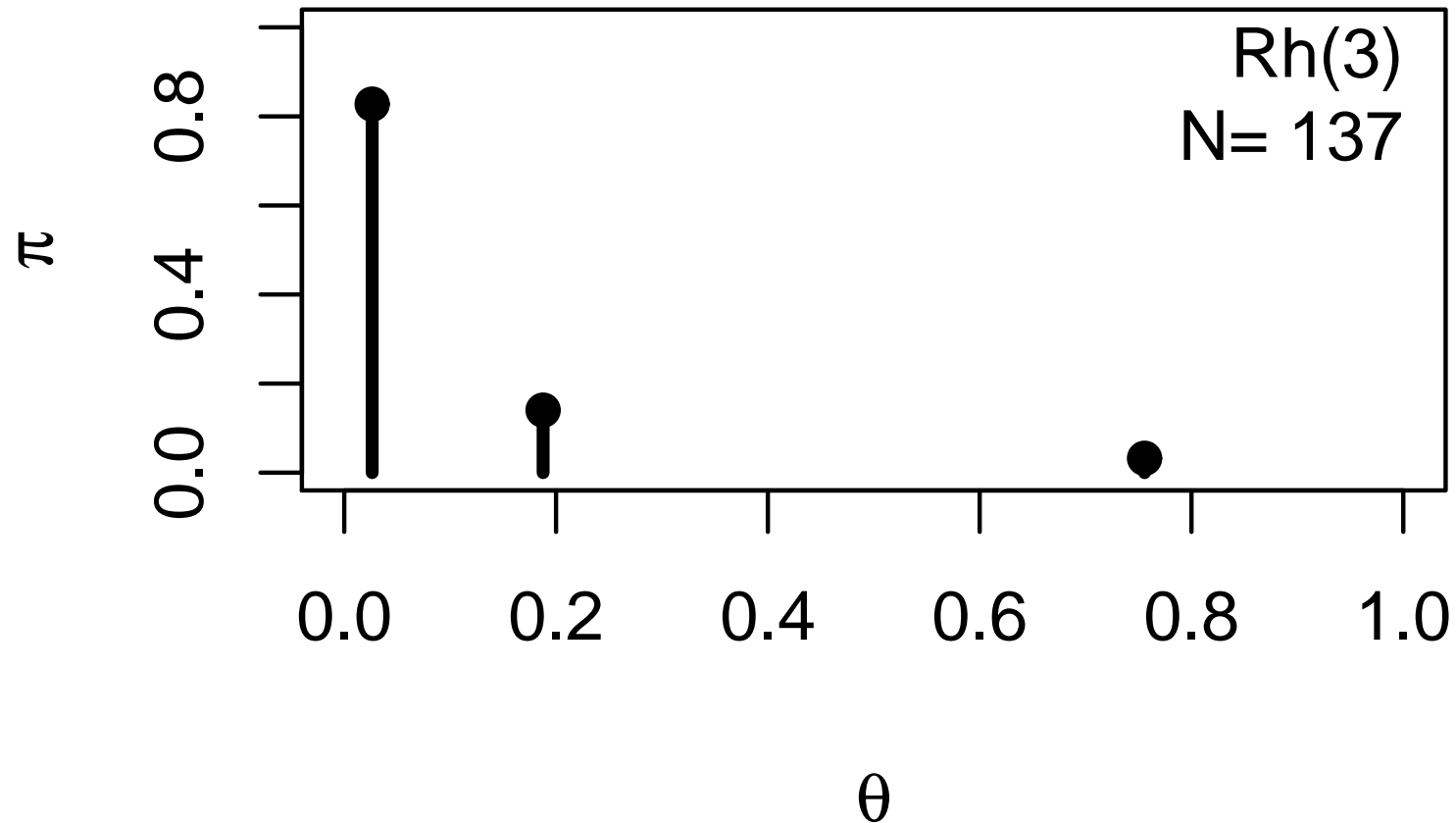
Step 76

$$\log L = -8.07461582671763$$



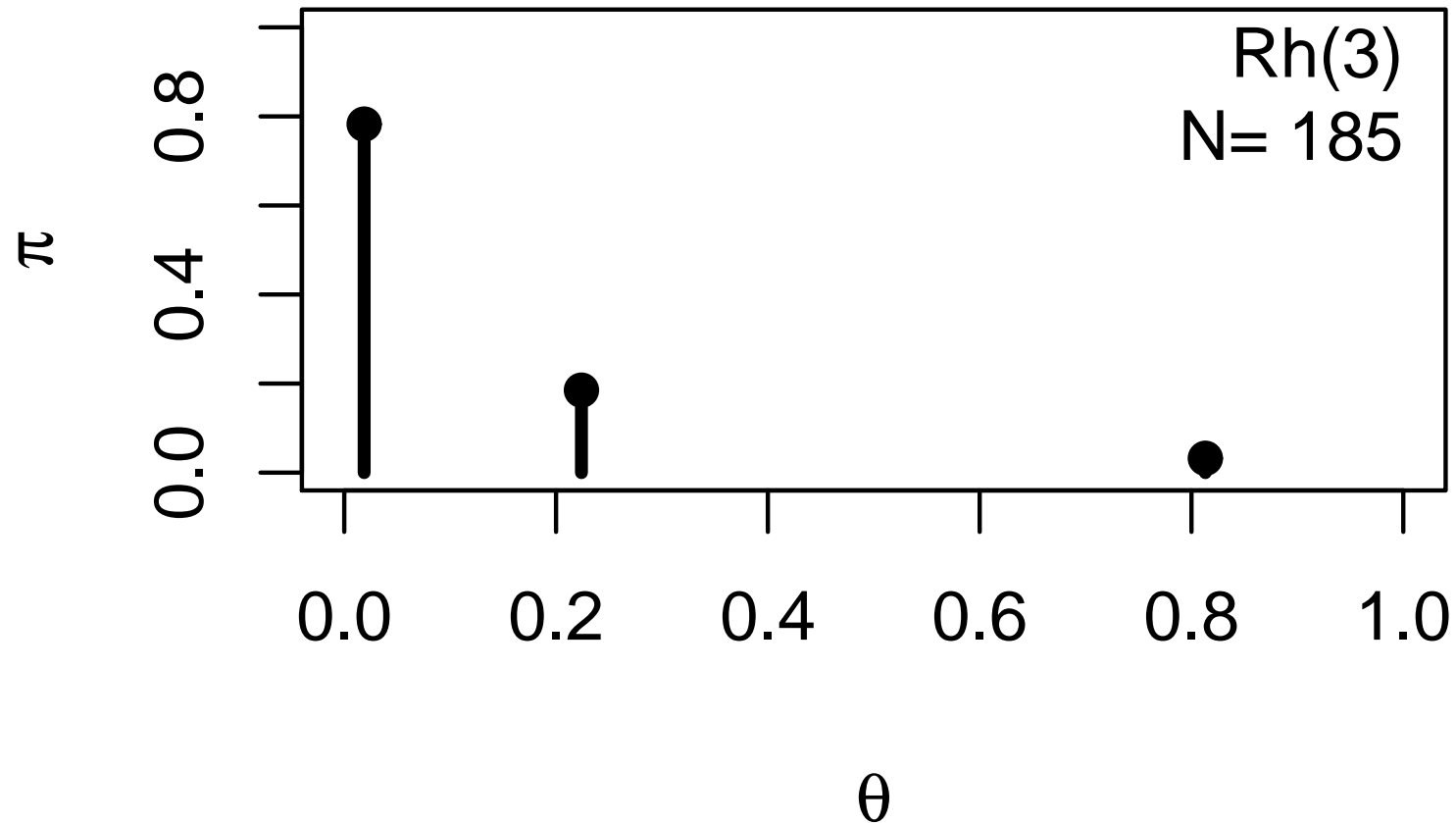
Step 77

$\log L = -8.69088322869459$



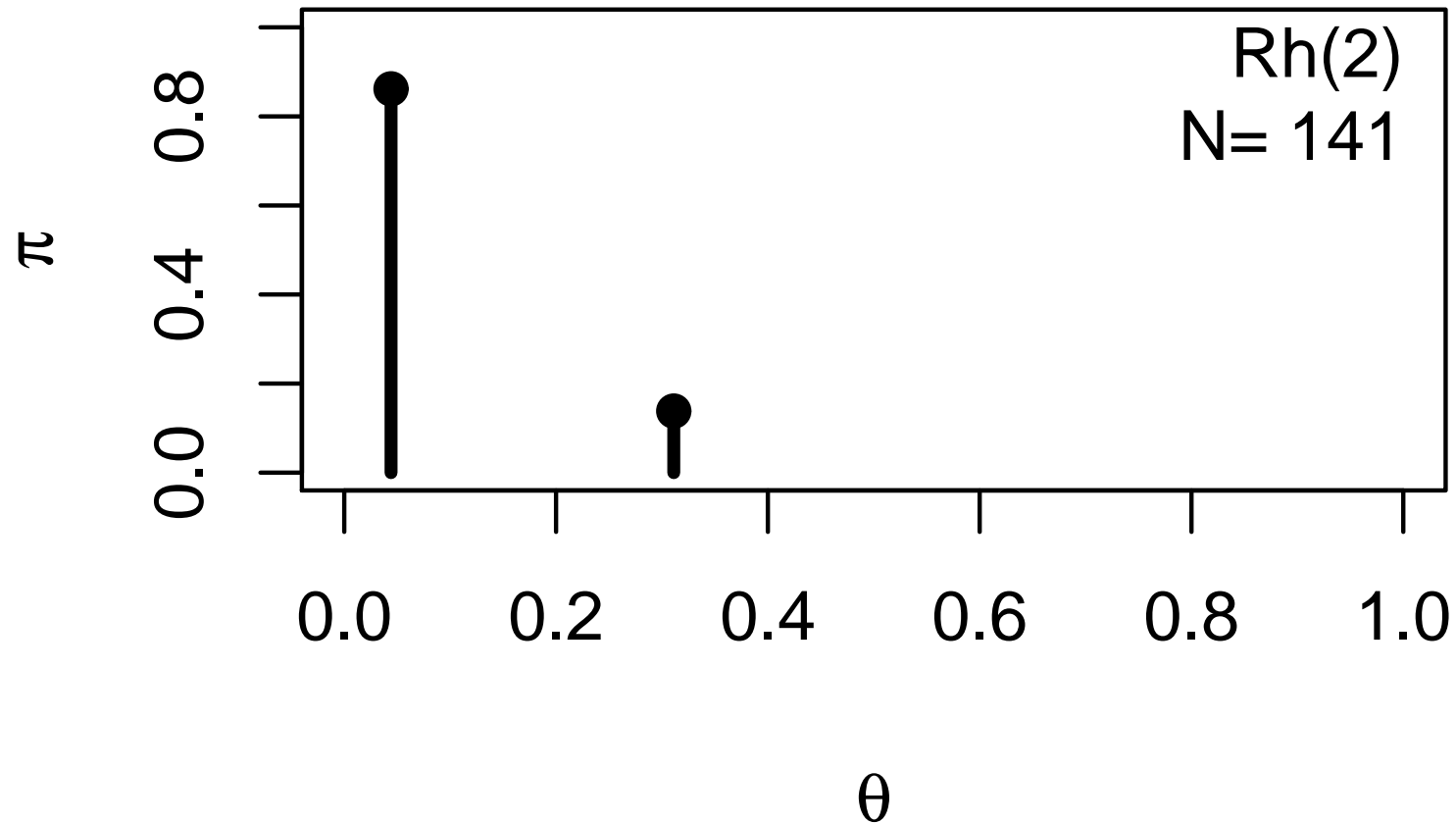
Step 78

$\log L = -9.06991579820229$



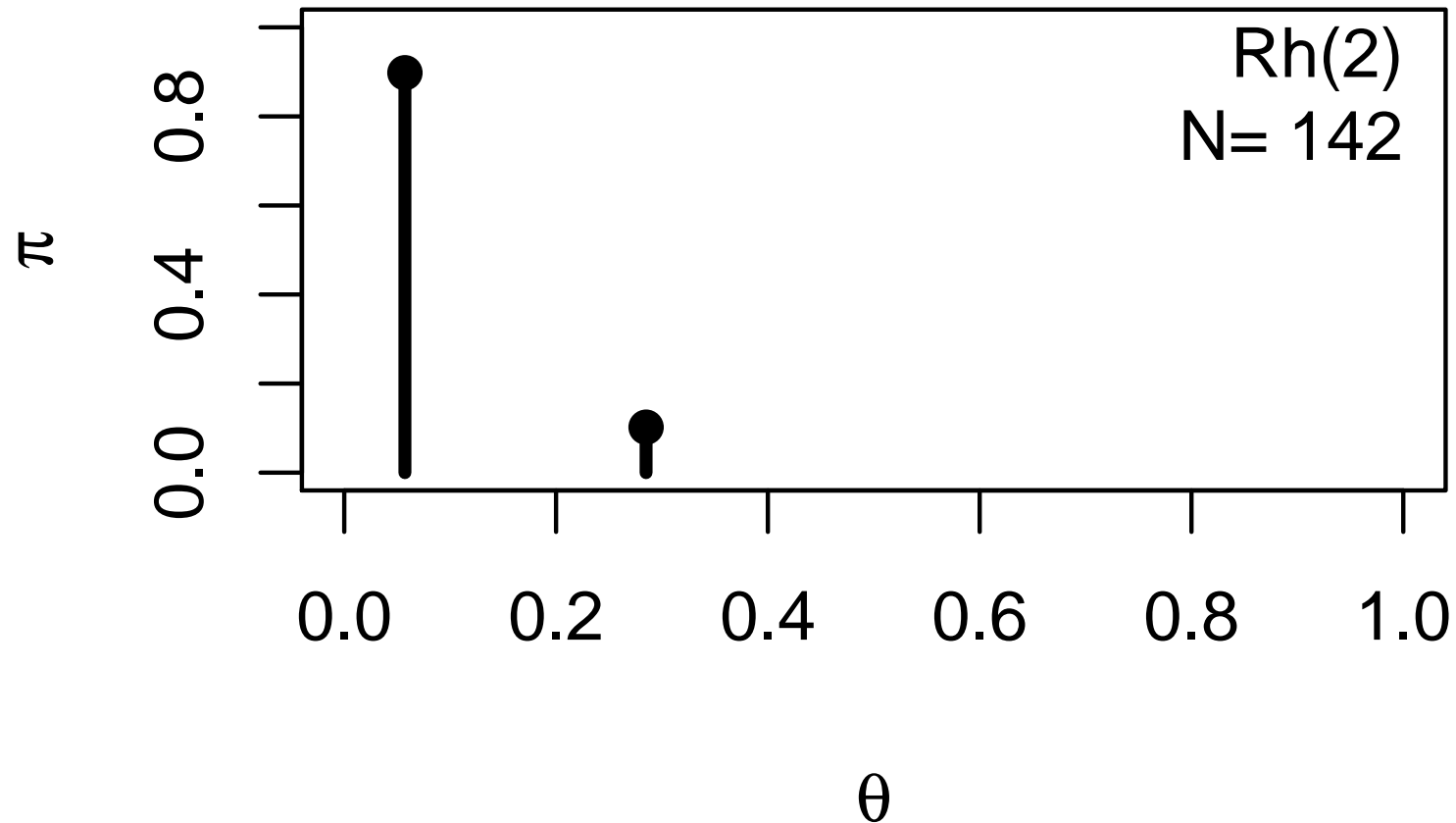
Step 79

$$\log L = -6.59579706710454$$



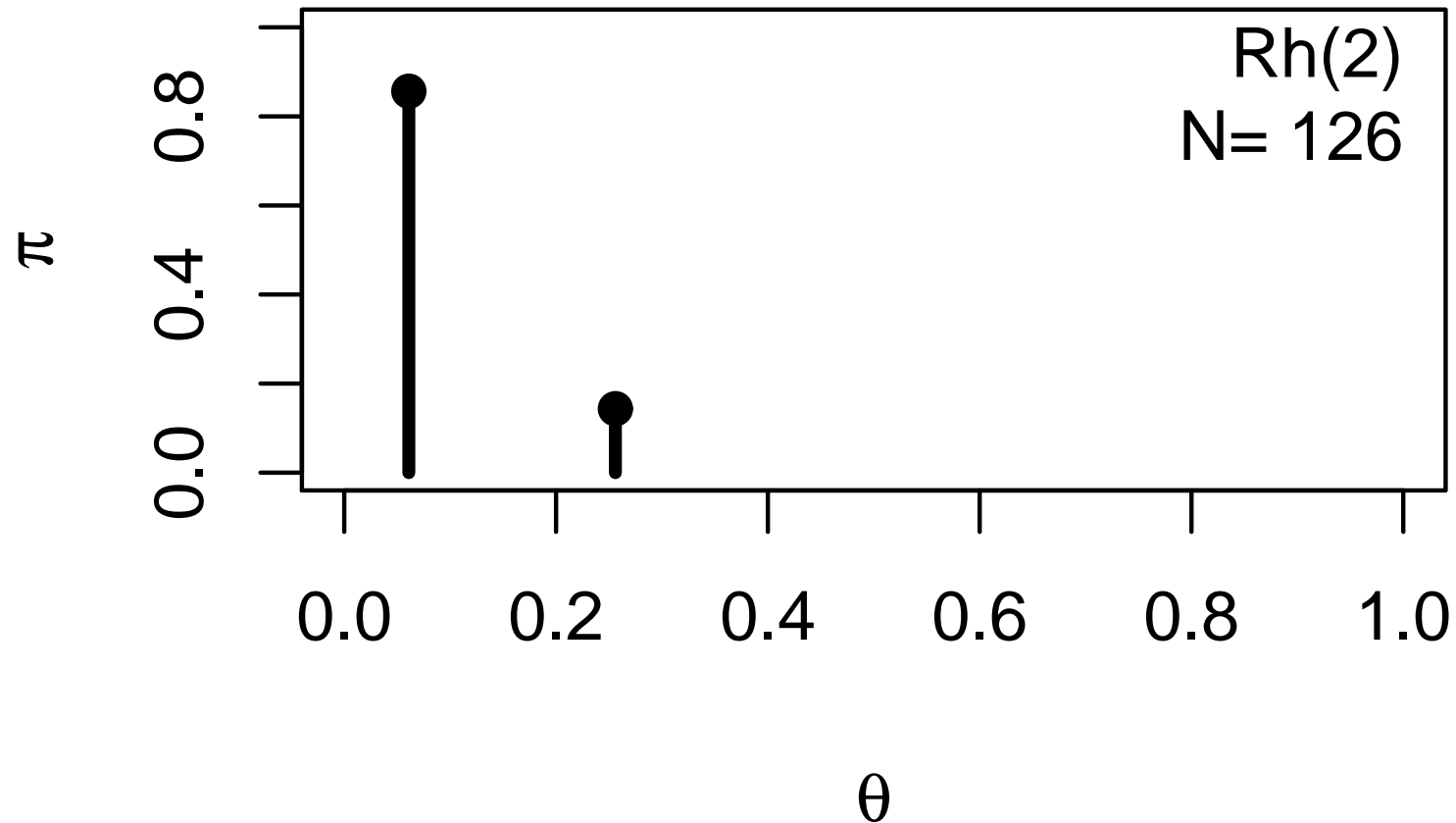
Step 80

$$\log L = -6.72727099921829$$



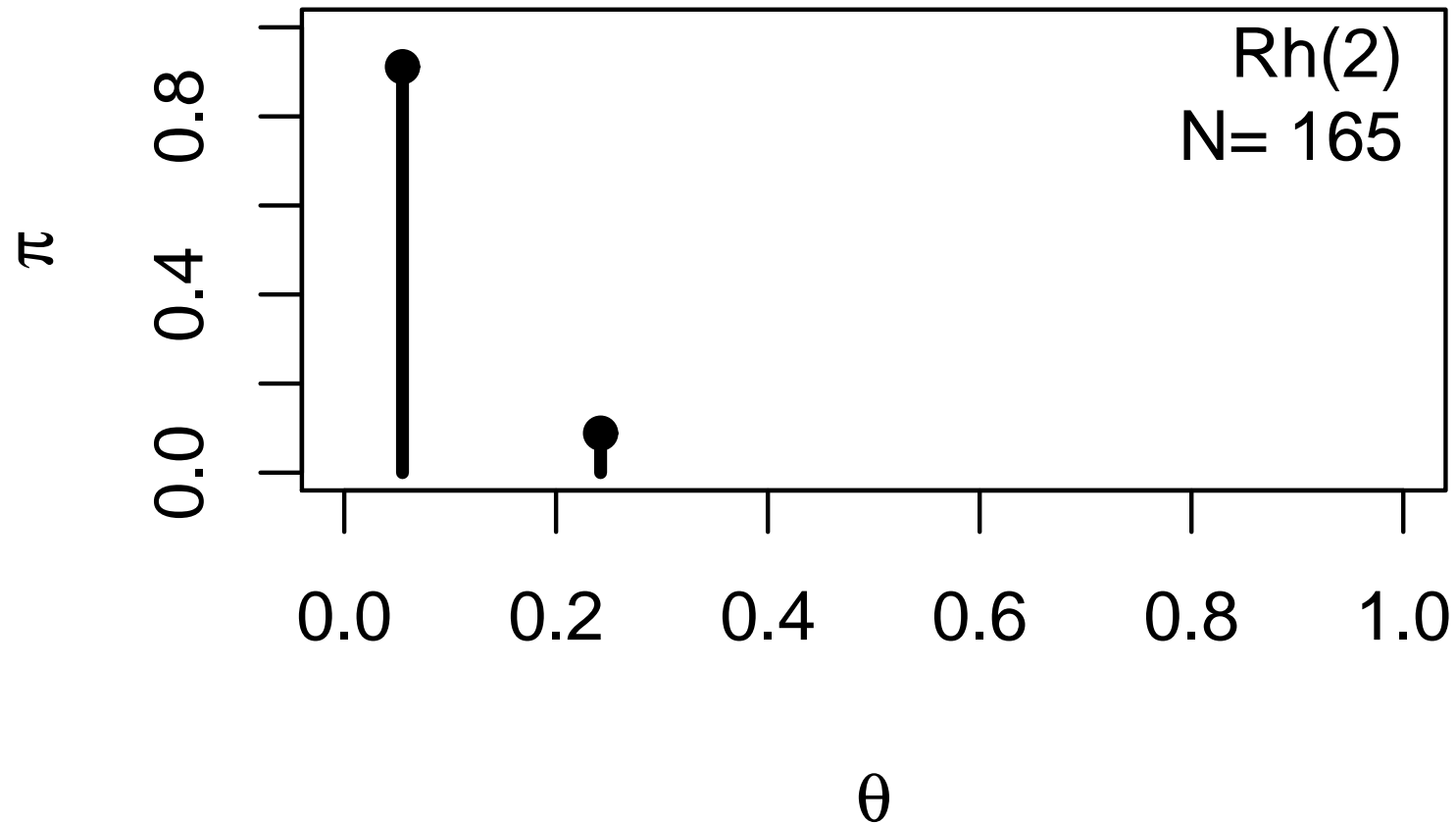
Step 81

$$\log L = -6.37796903283765$$



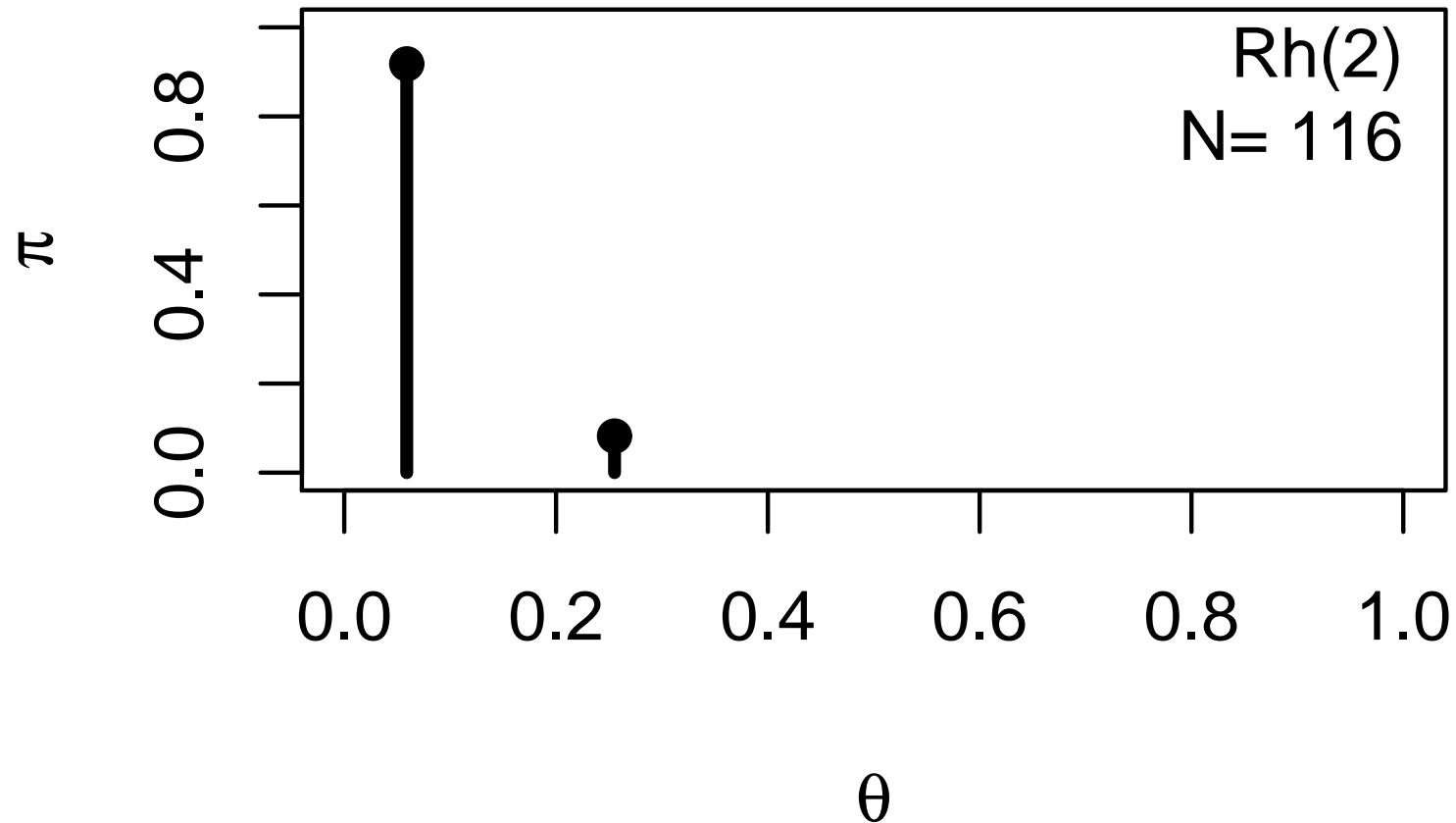
Step 82

$$\log L = -8.43700588880688$$



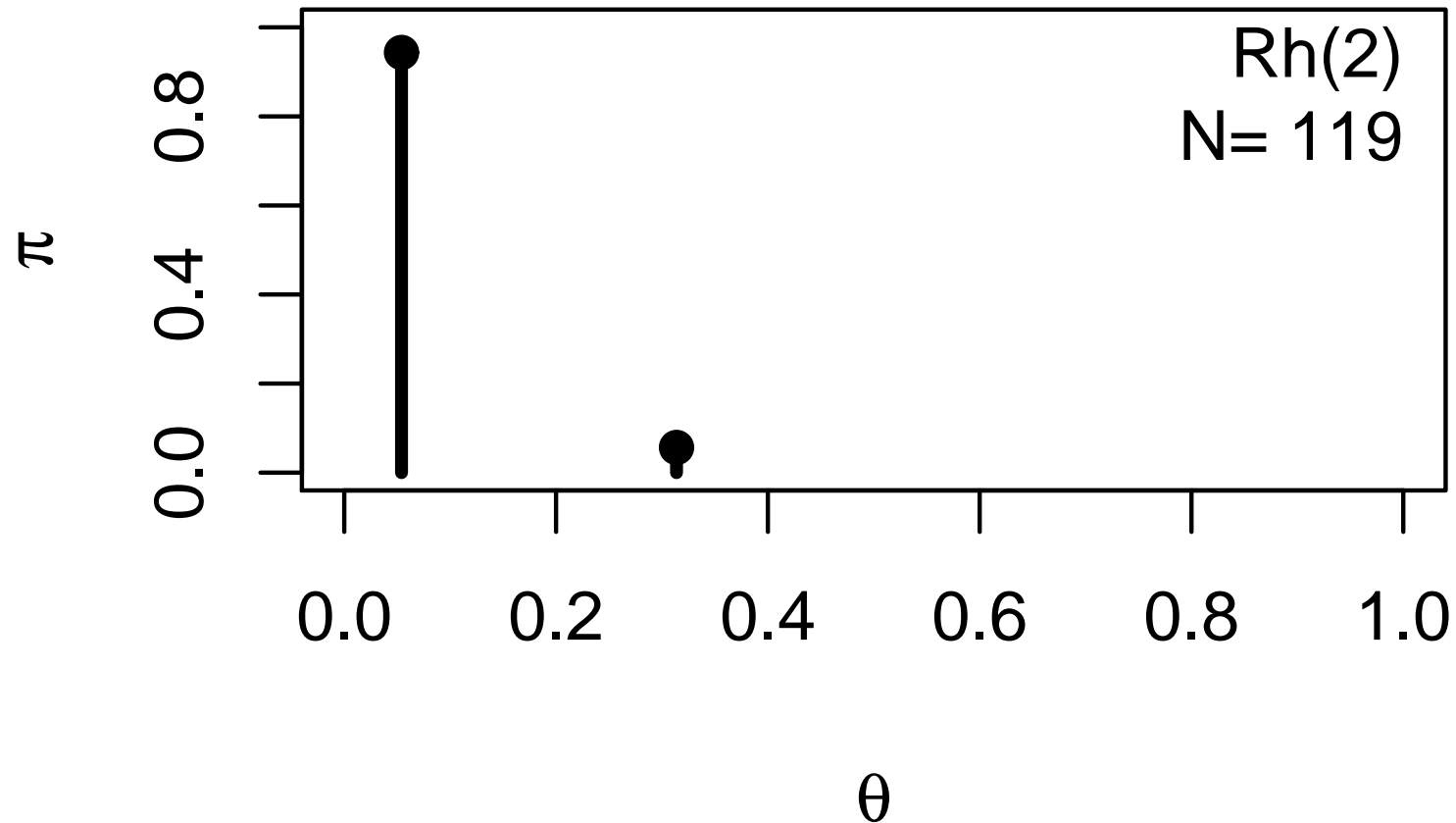
Step 83

$$\log L = -6.2297006532936$$



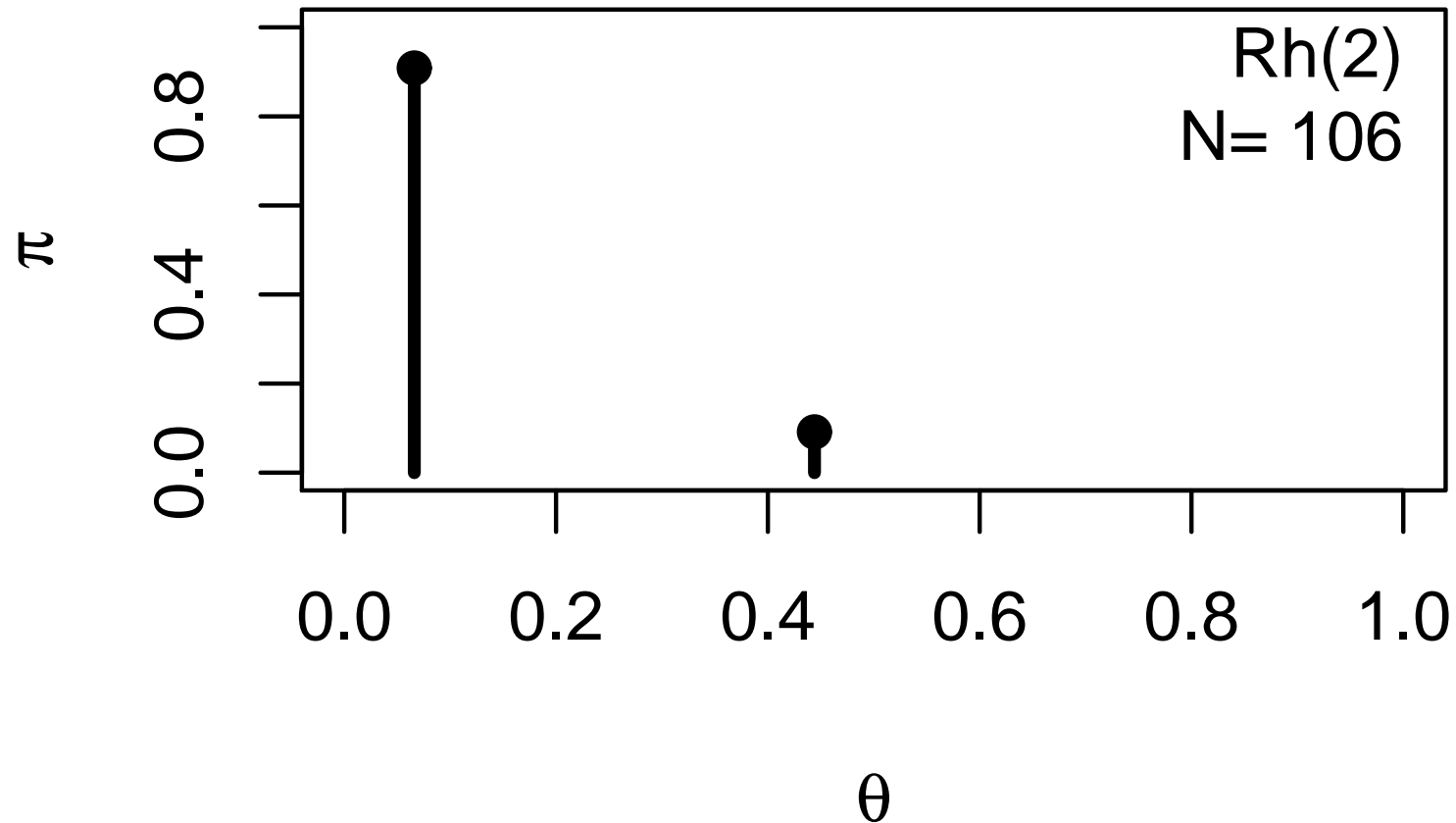
Step 84

$$\log L = -6.59598781624027$$



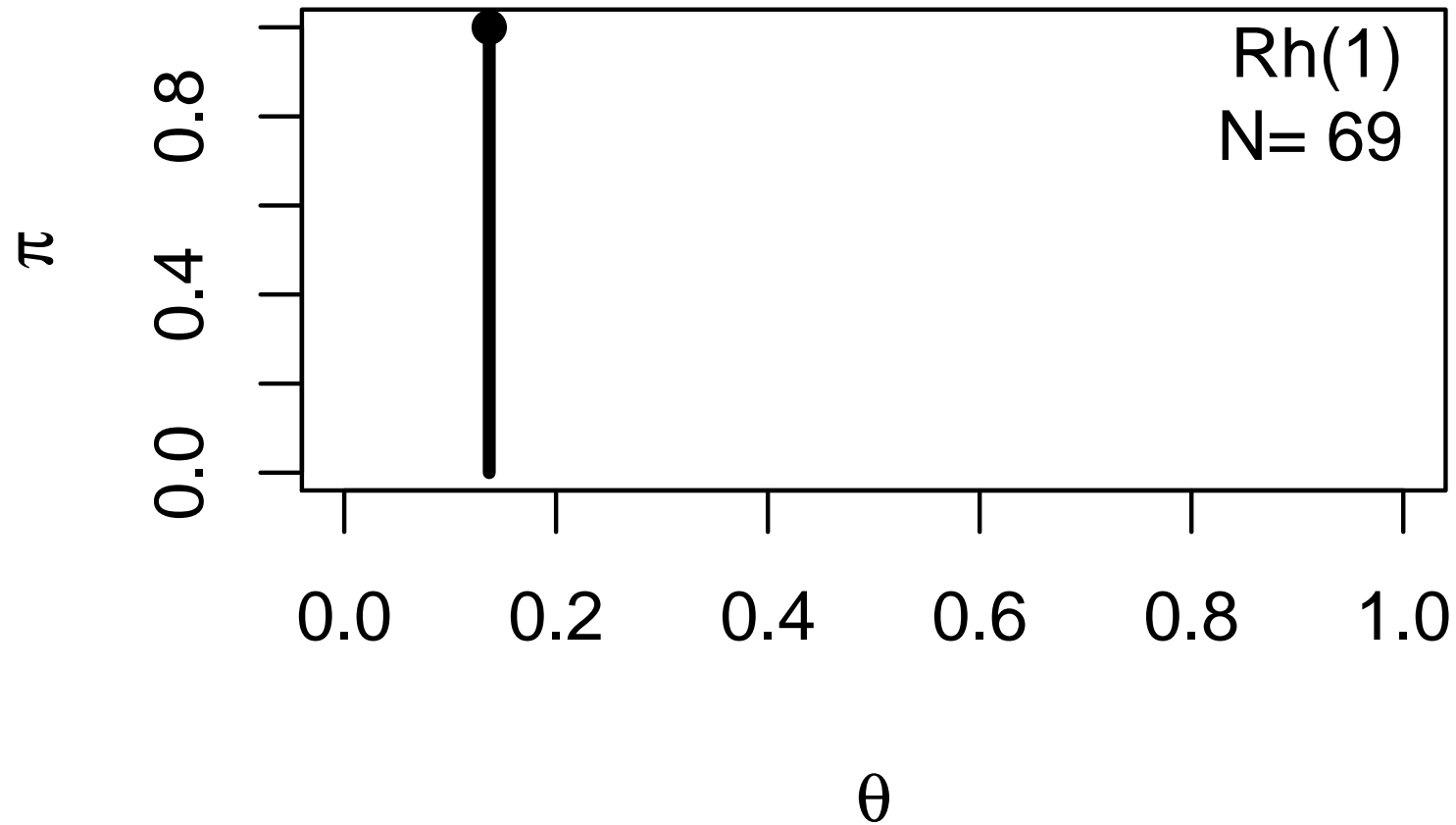
Step 85

$$\log L = -6.53657057131011$$



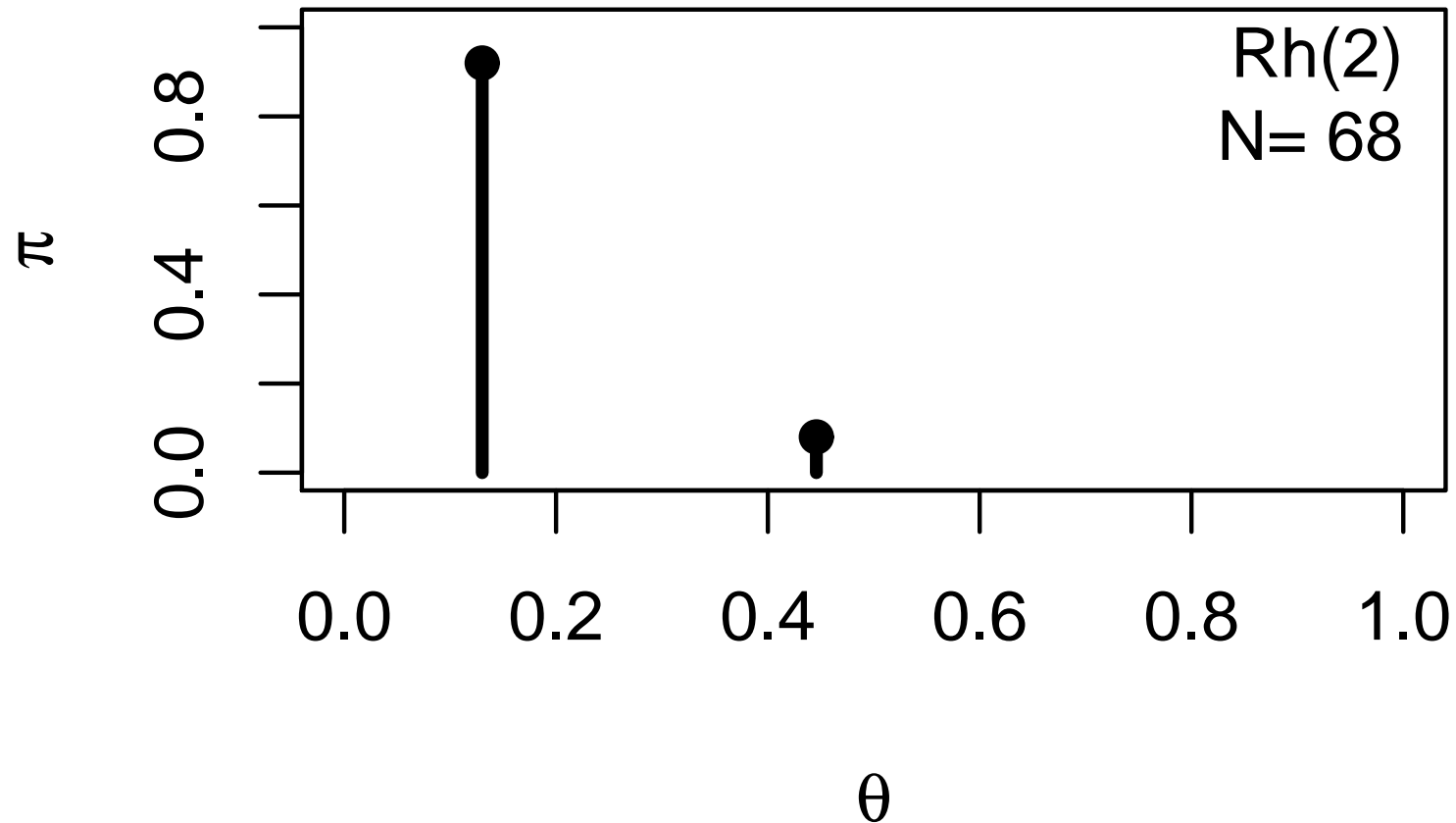
Step 86

$$\log L = -6.49950133785143$$



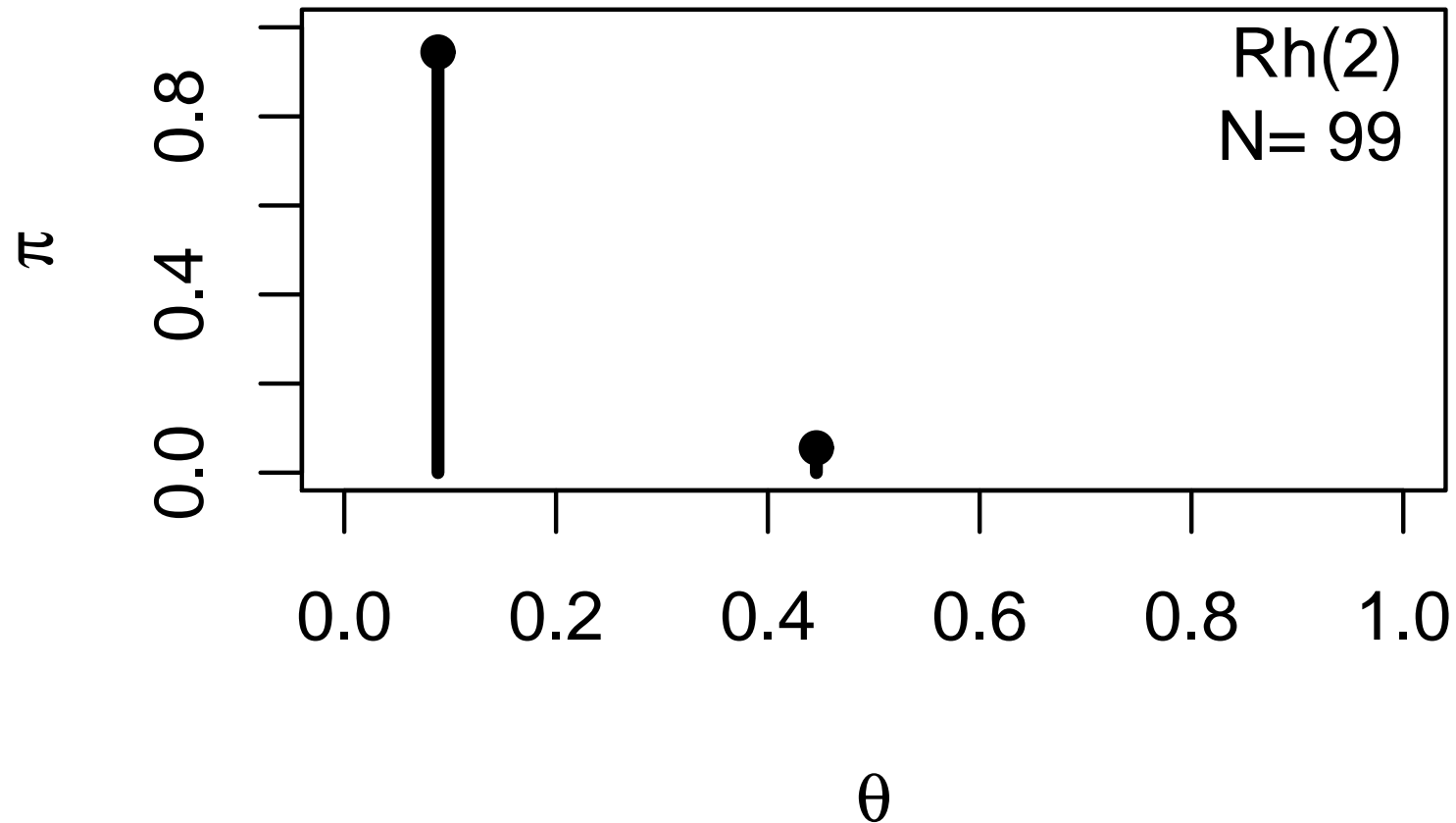
Step 87

$$\log L = -5.86507082741511$$



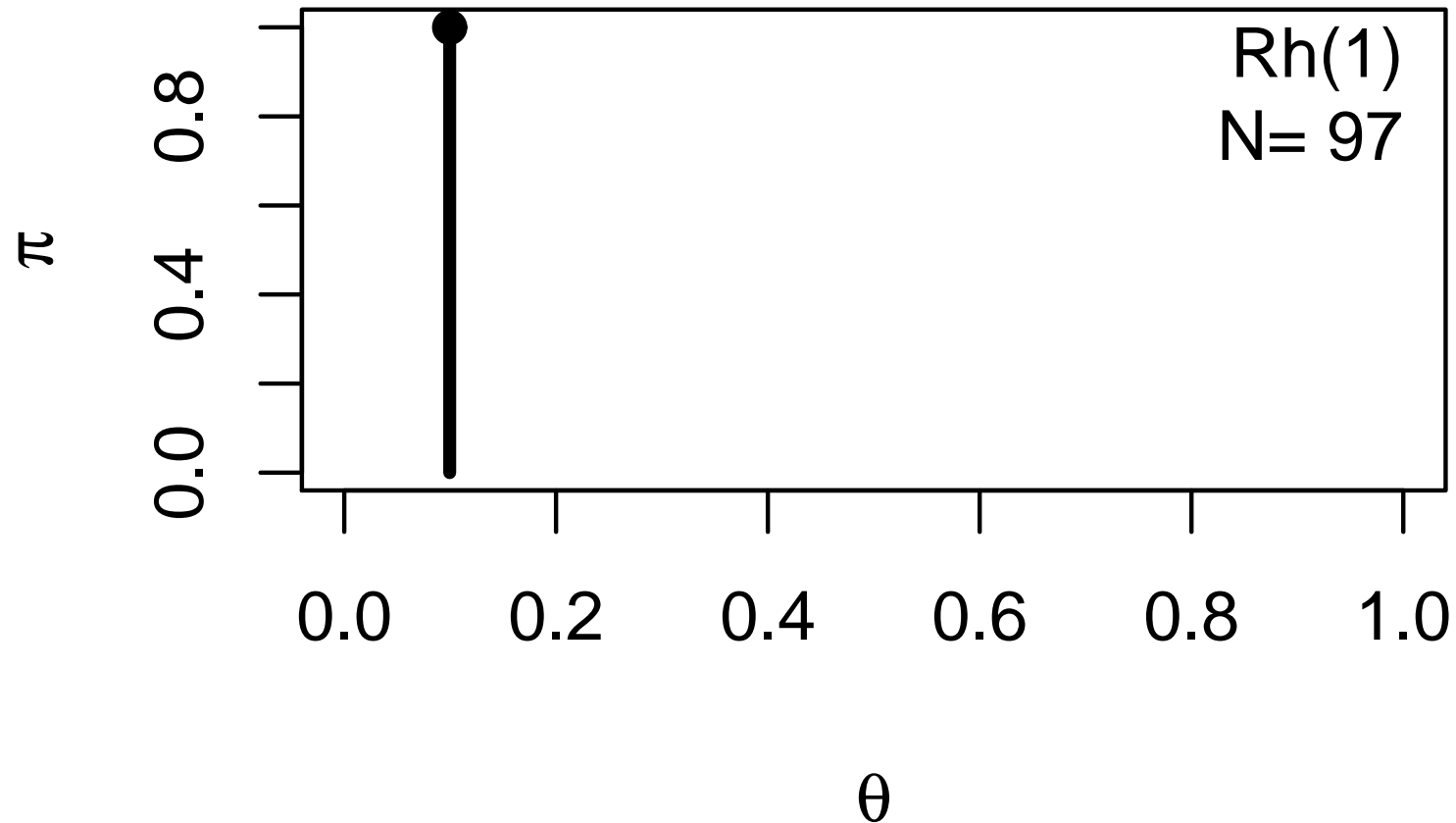
Step 88

$$\log L = -5.15535623326514$$



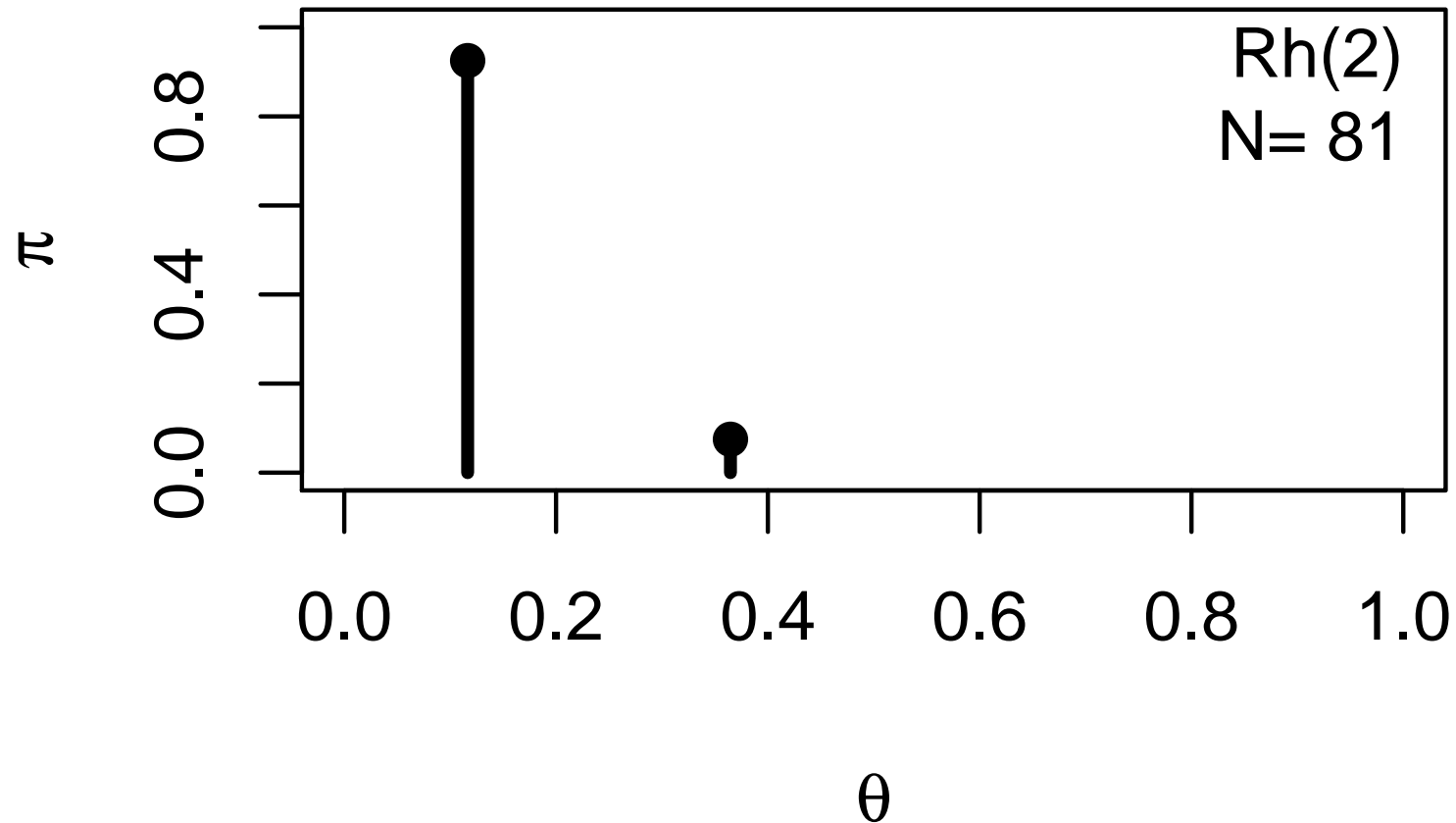
Step 89

$$\log L = -7.52915403726294$$



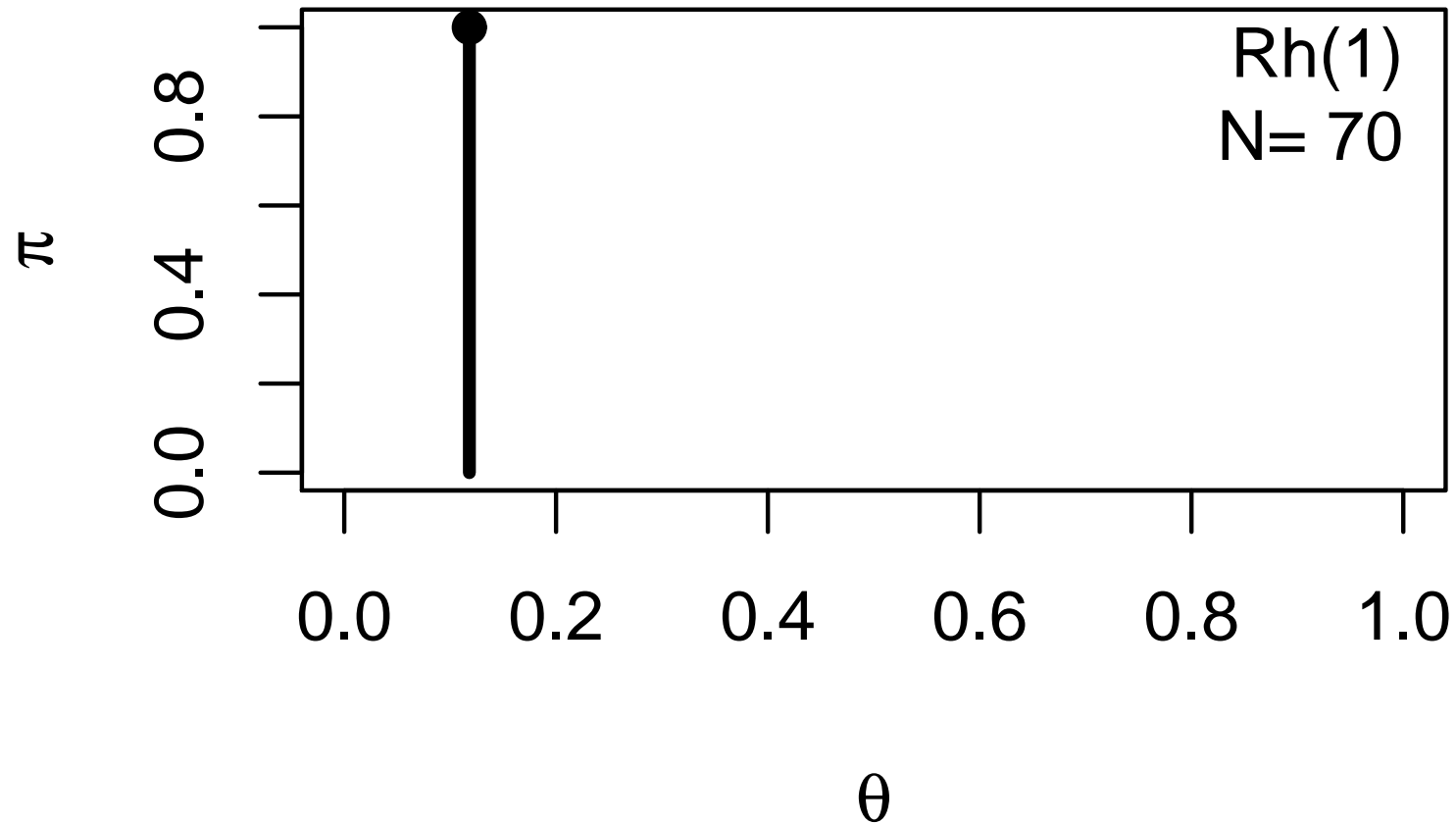
Step 90

$$\log L = -5.42621622335881$$



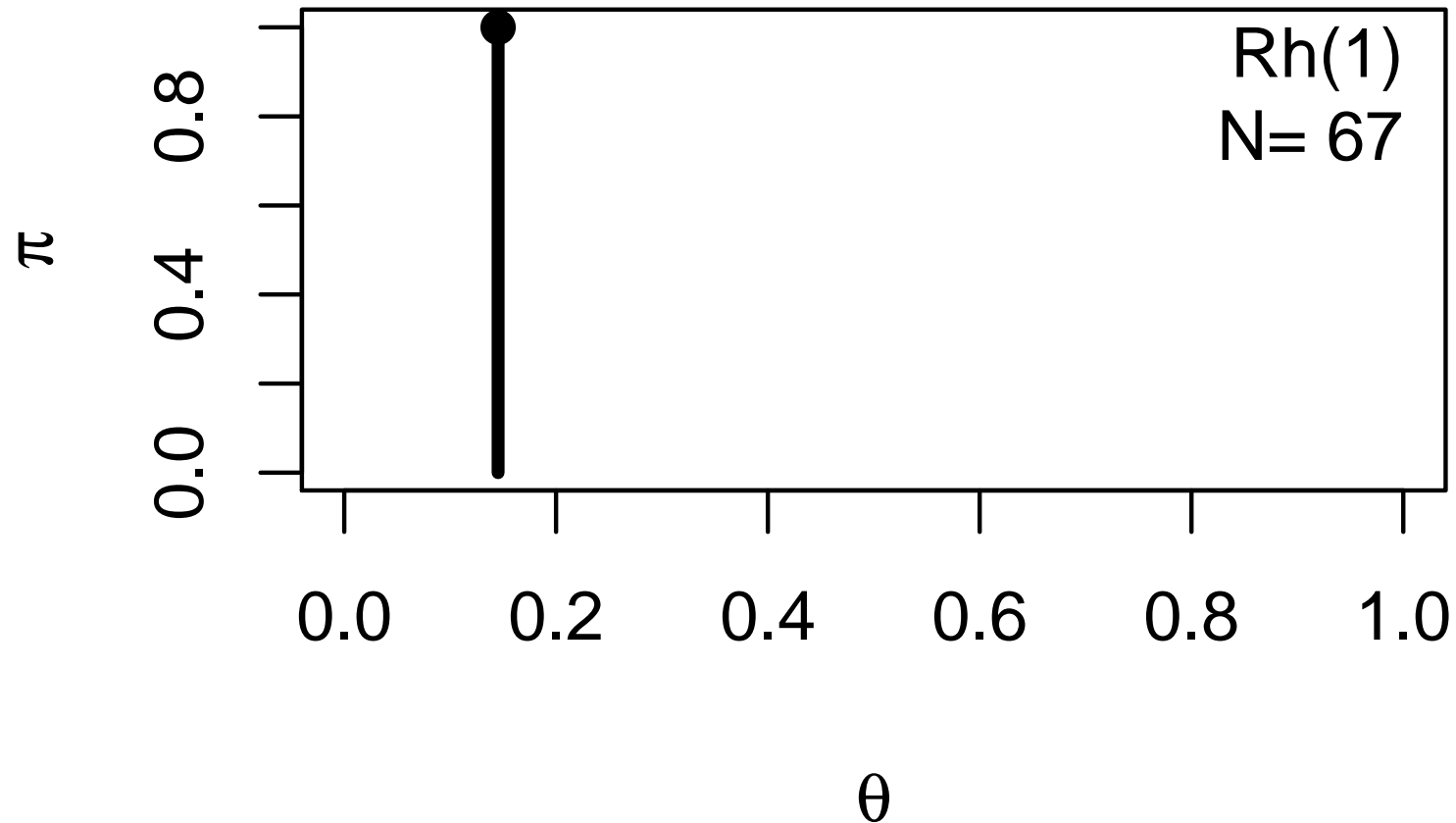
Step 91

$$\log L = -7.53051945806166$$



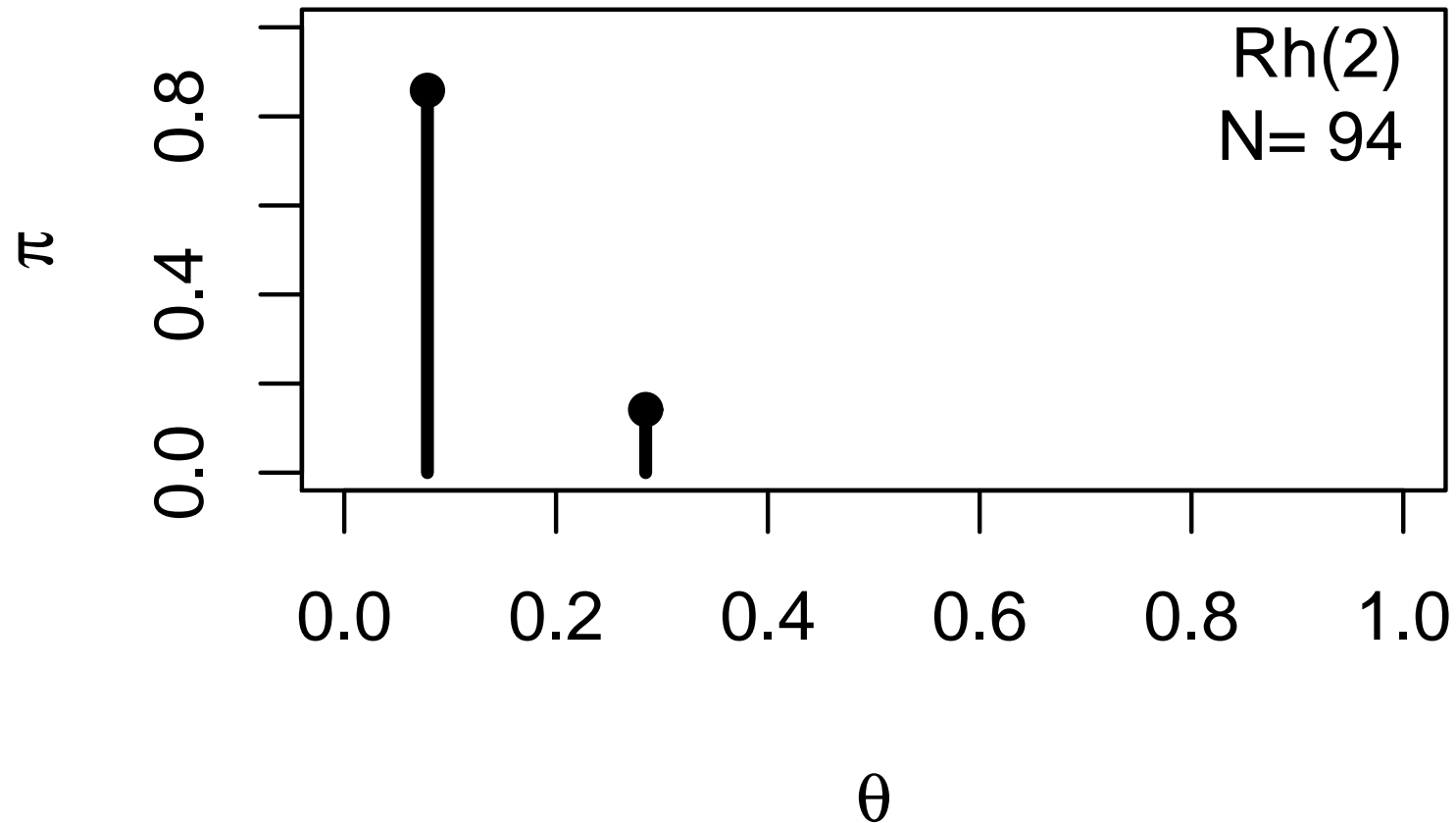
Step 92

$$\log L = -6.48125416952934$$



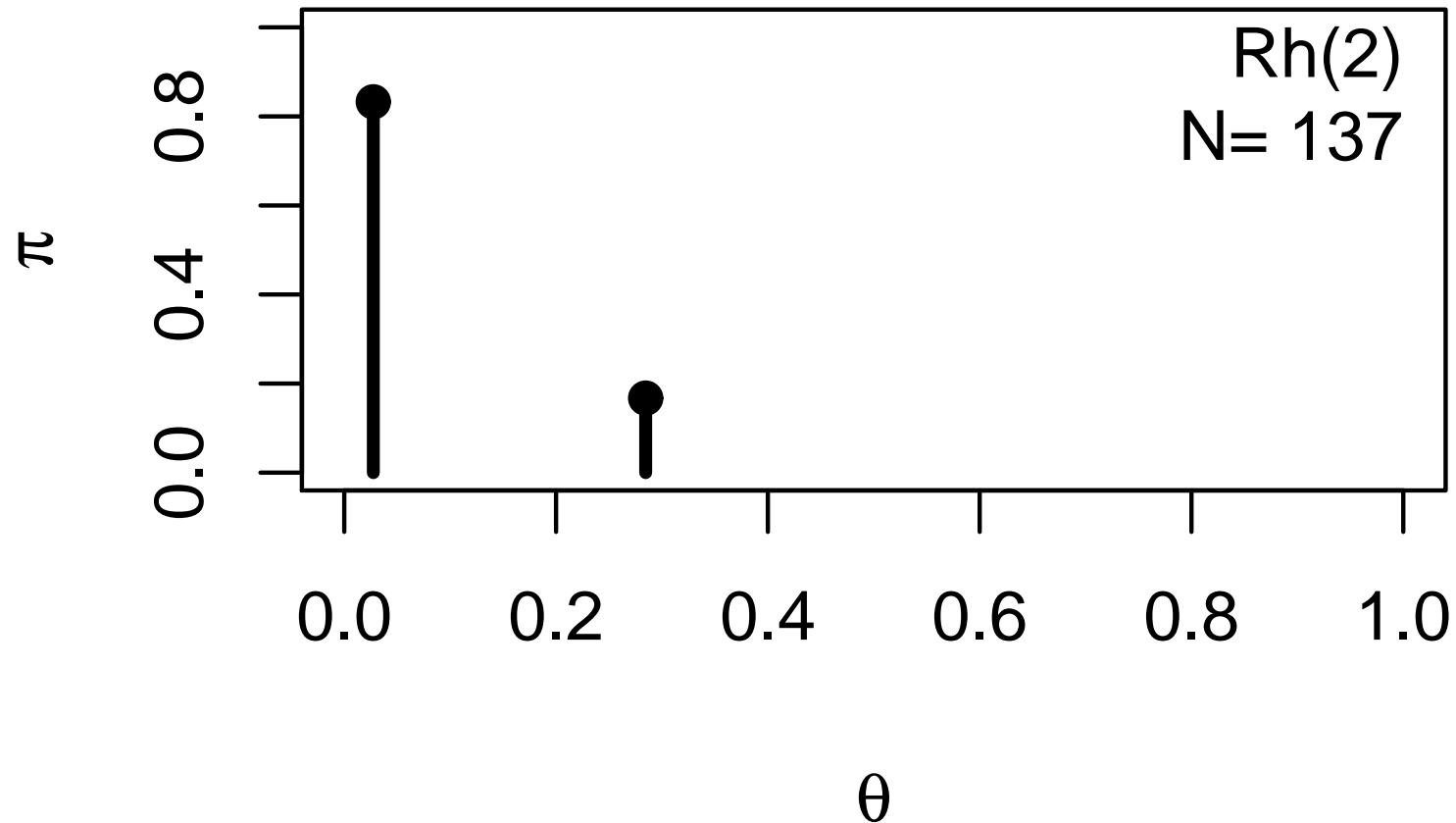
Step 93

$$\log L = -5.56786282597483$$



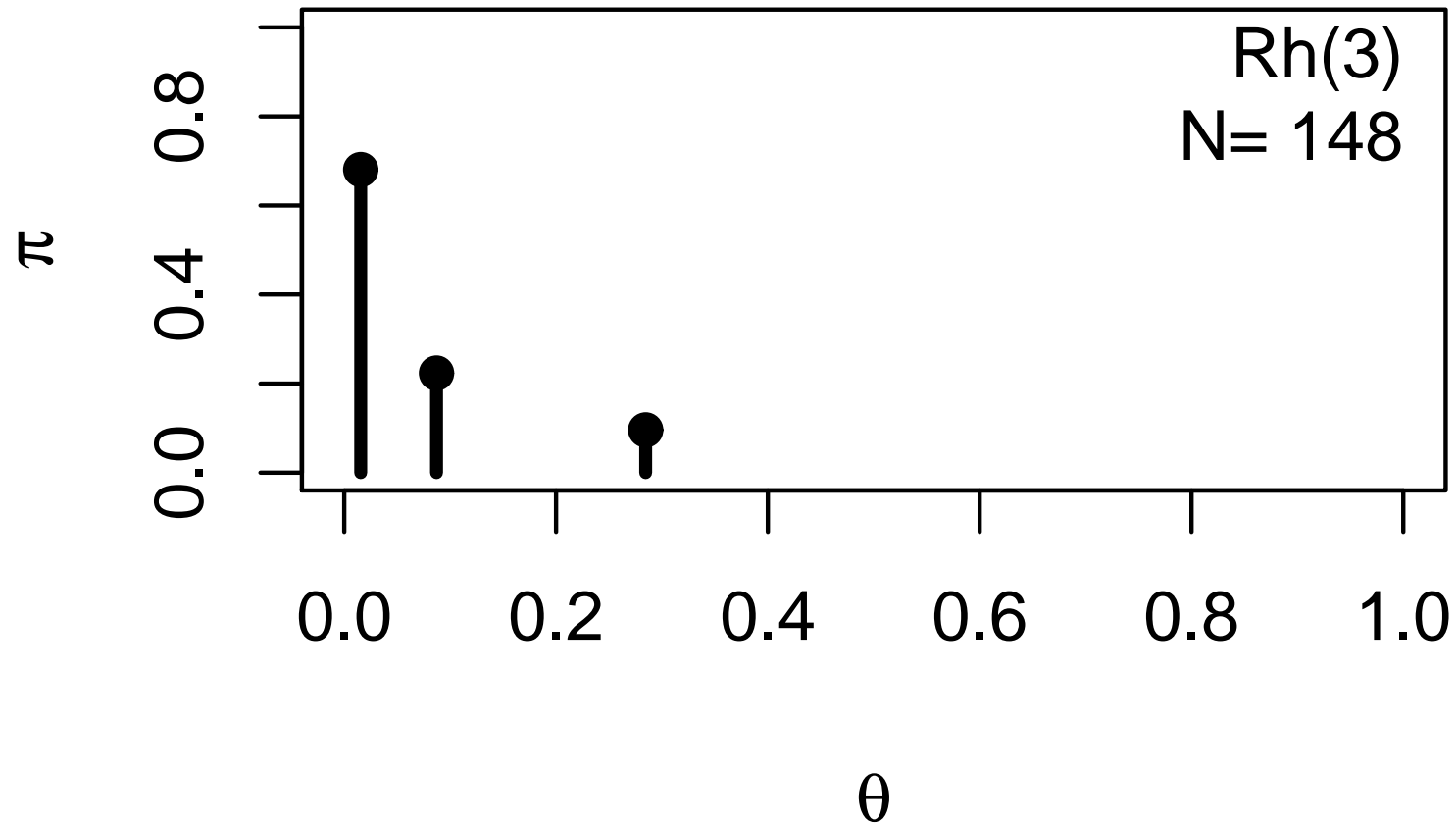
Step 94

$$\log L = -7.61592304565889$$



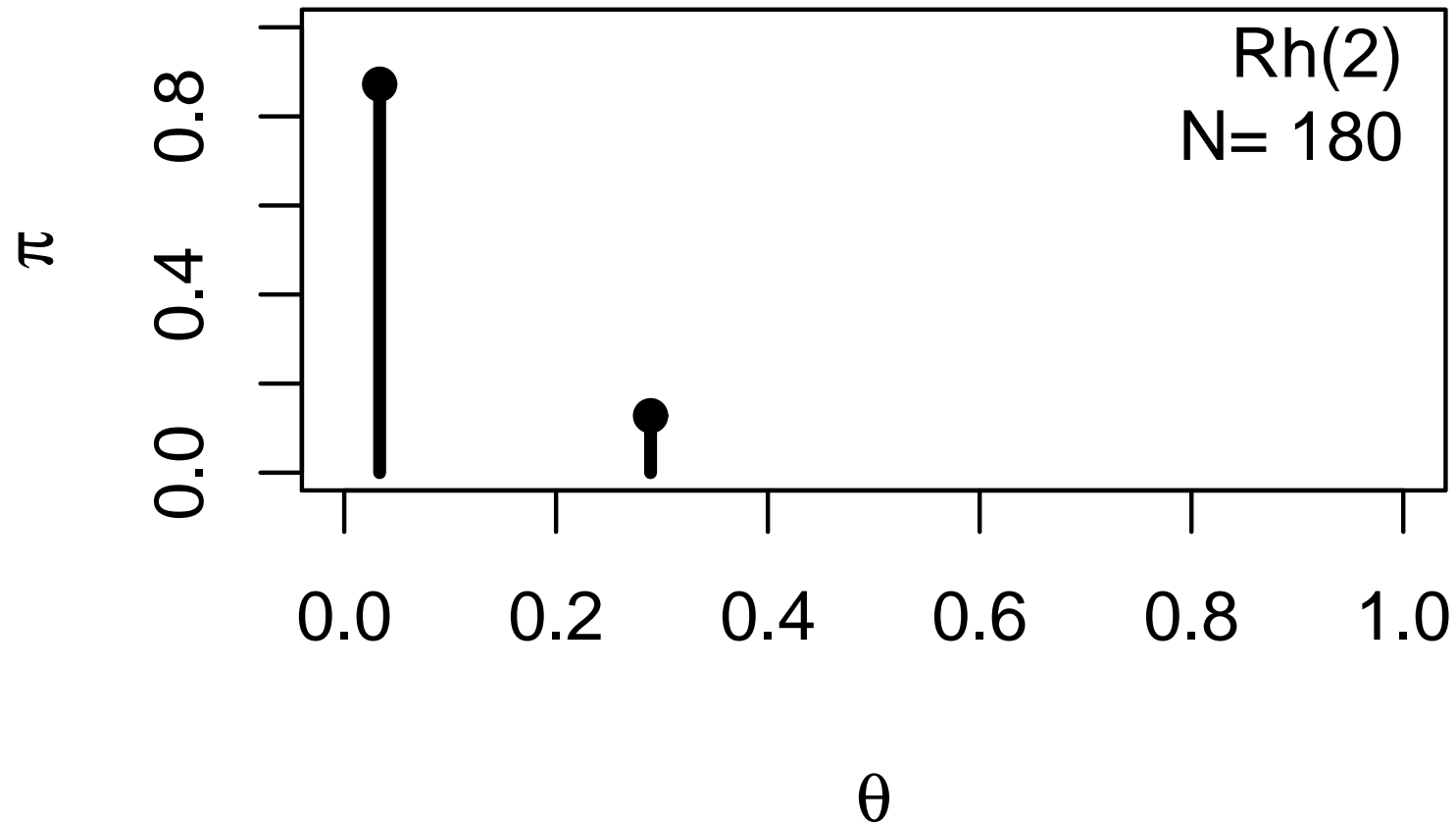
Step 95

$$\log L = -7.14847973079398$$



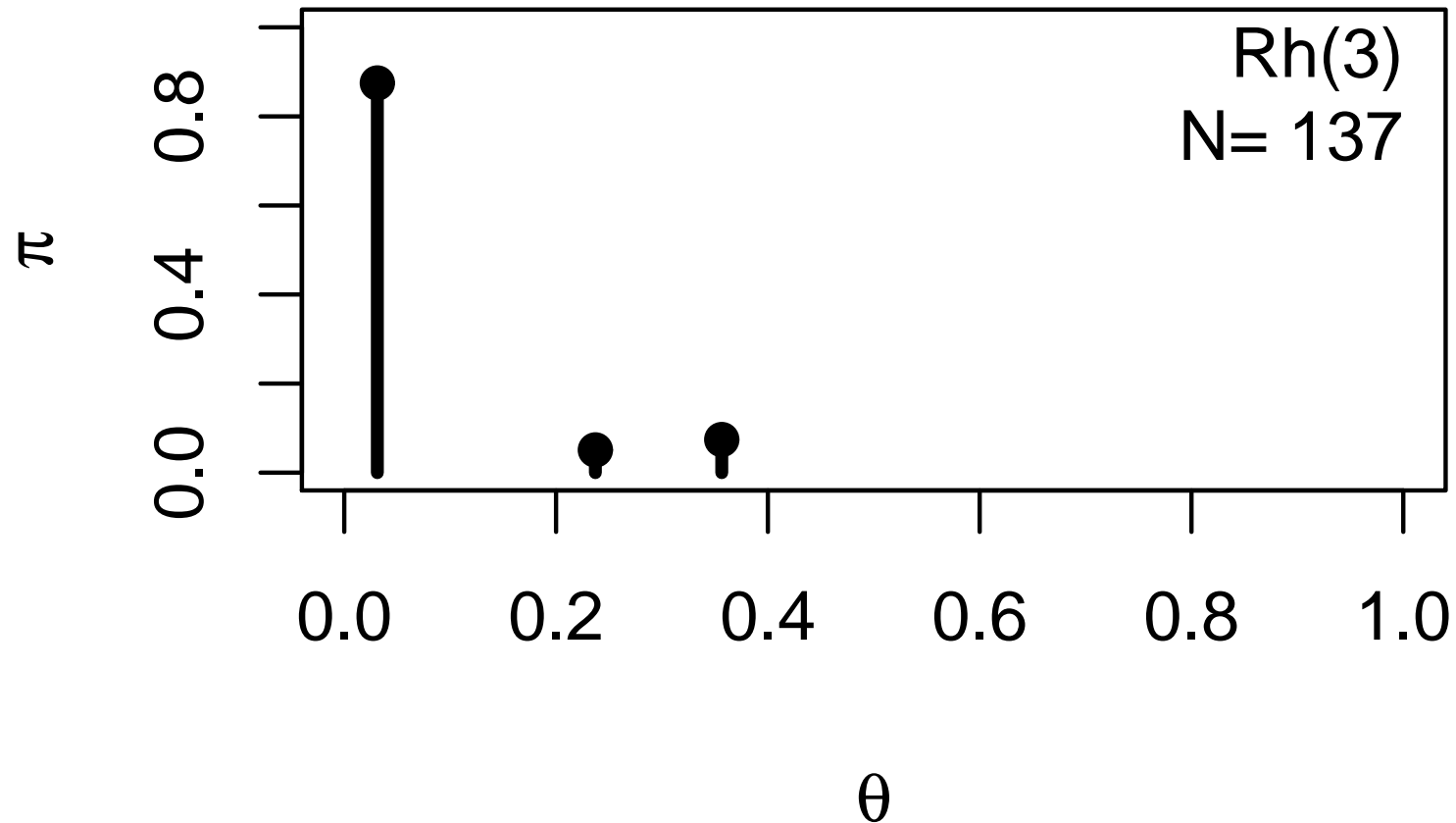
Step 96

$\log L = -6.96778206285407$



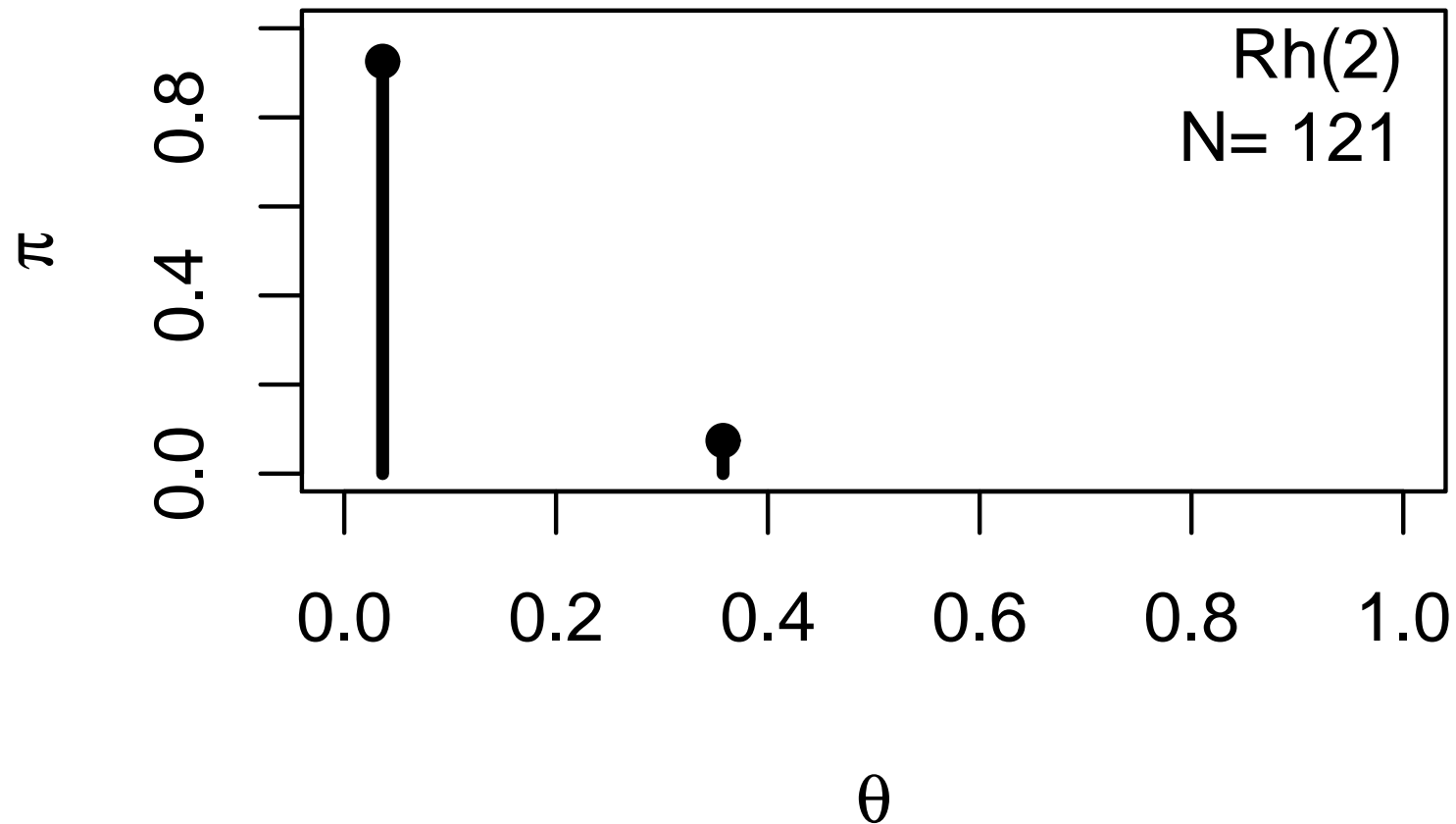
Step 97

$$\log L = -7.71918326510837$$



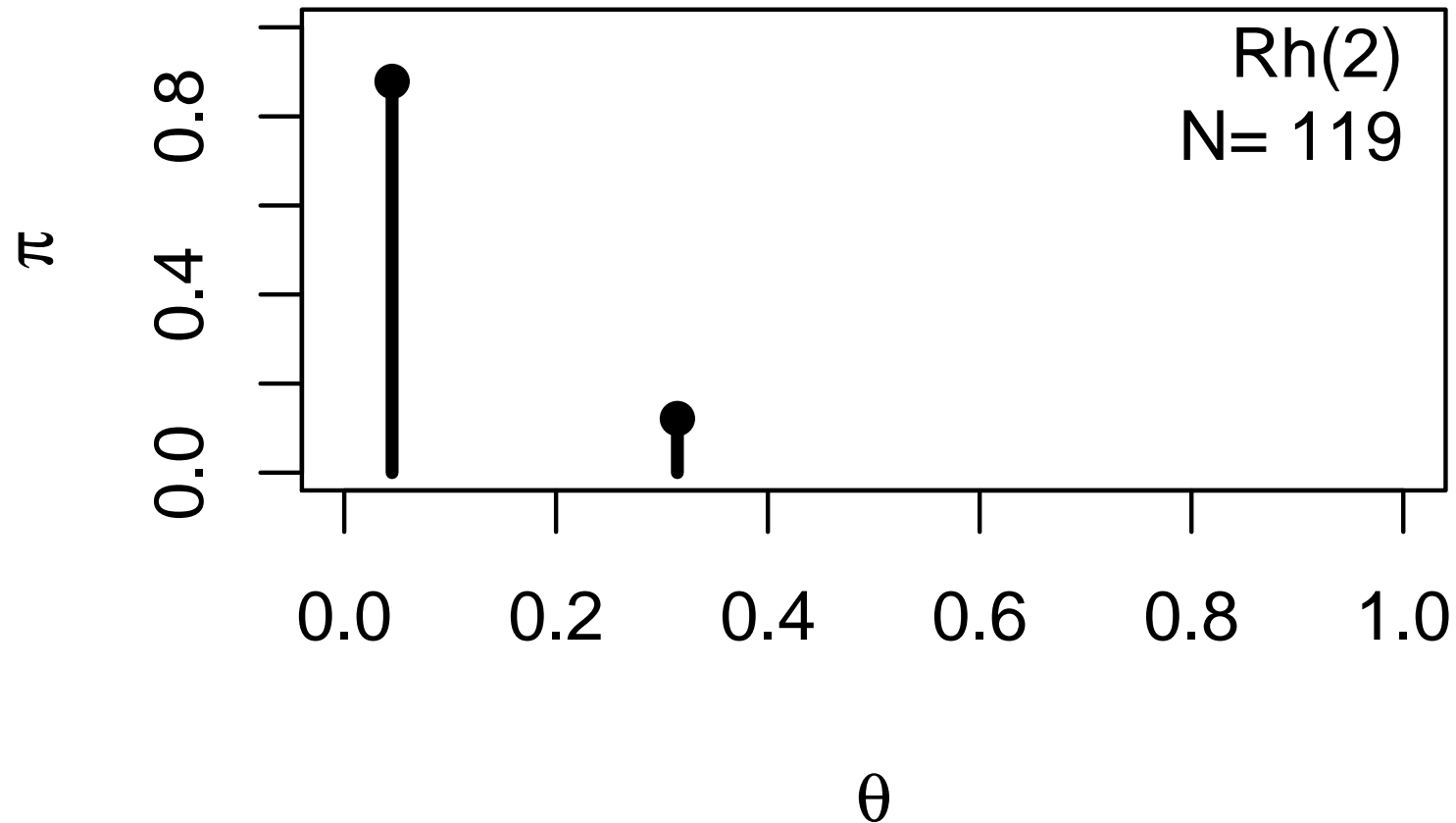
Step 98

$\log L = -9.8181414985707$



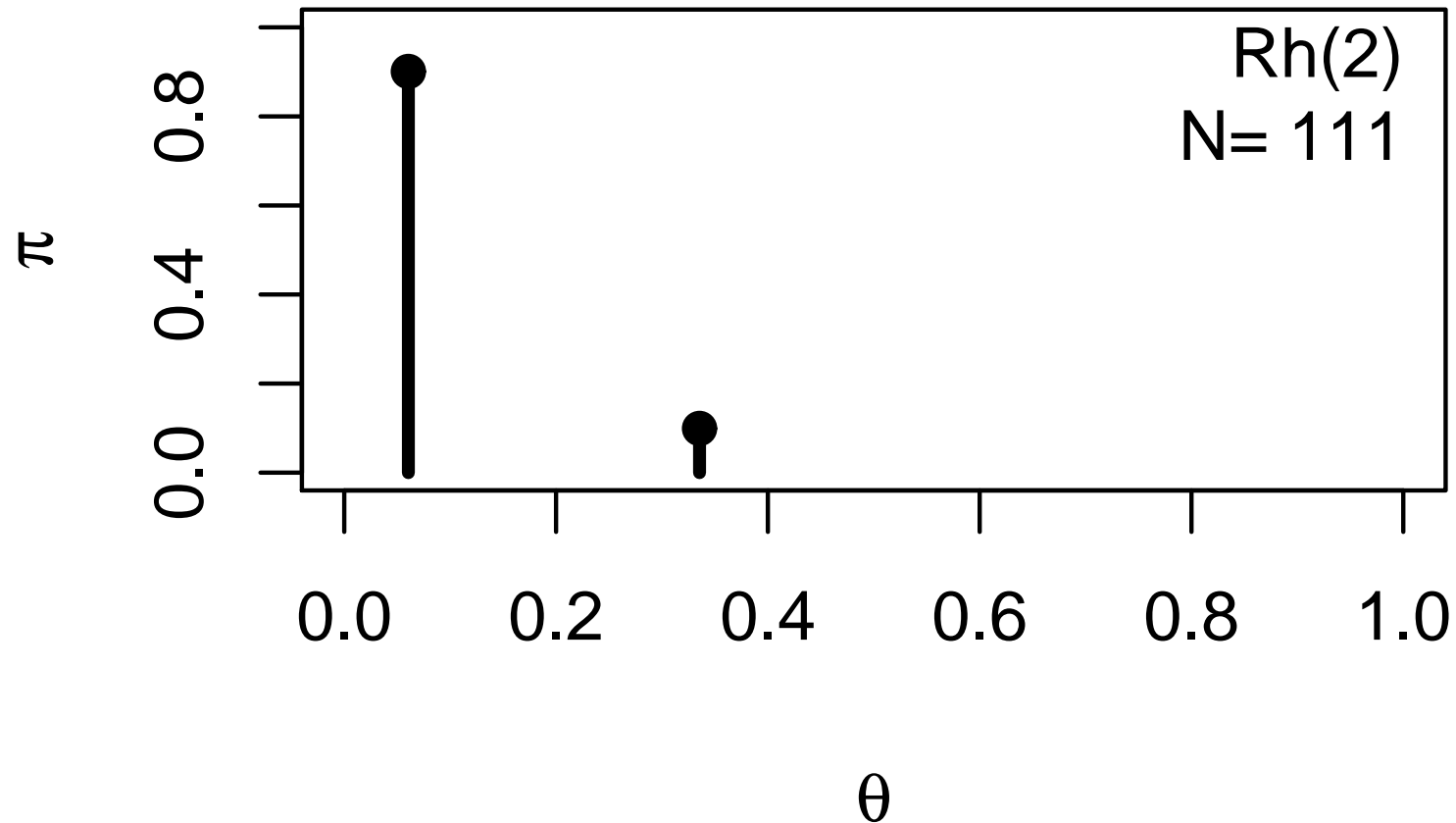
Step 99

$$\log L = -6.64868771633988$$



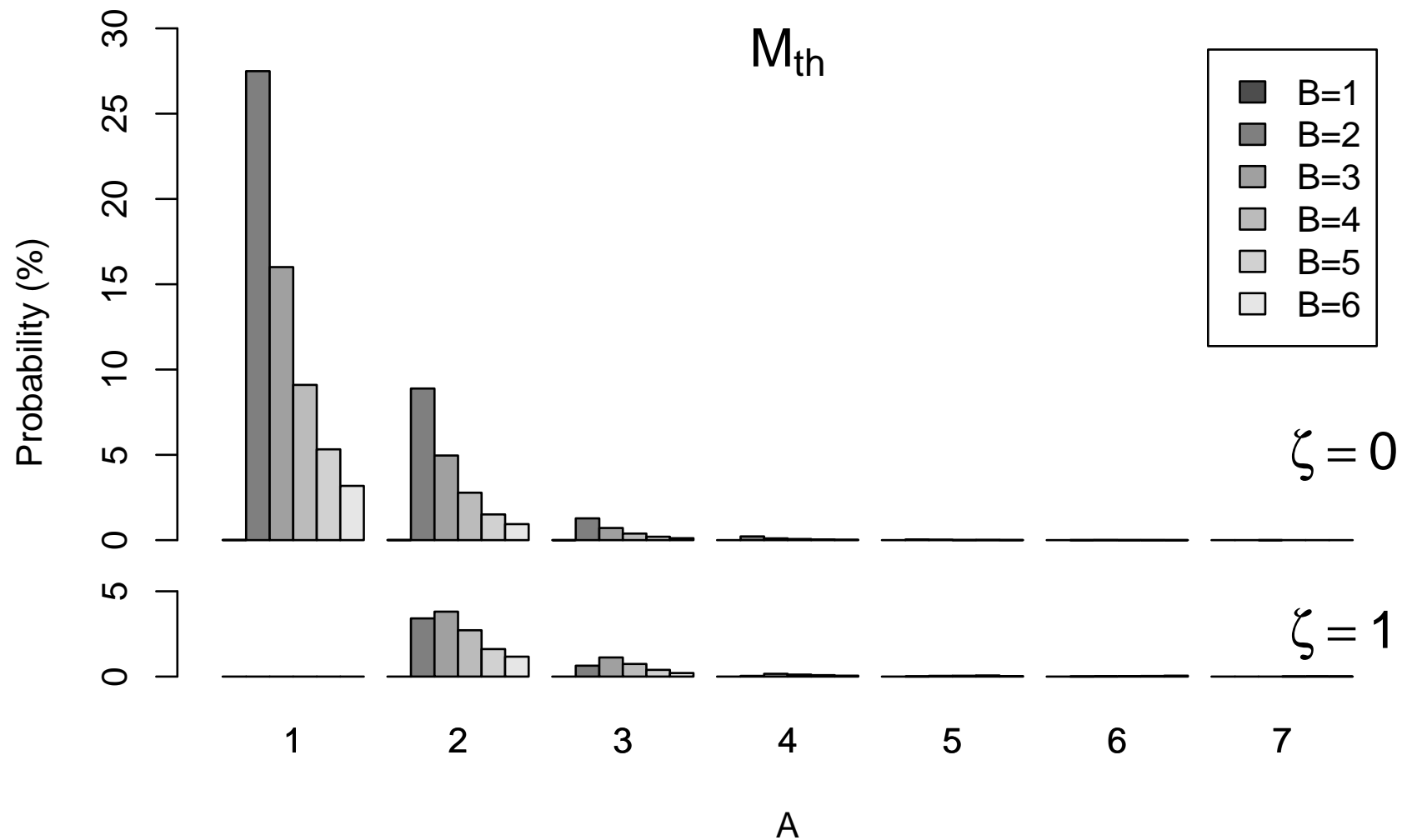
Step 100

$\log L = -5.83175113202486$



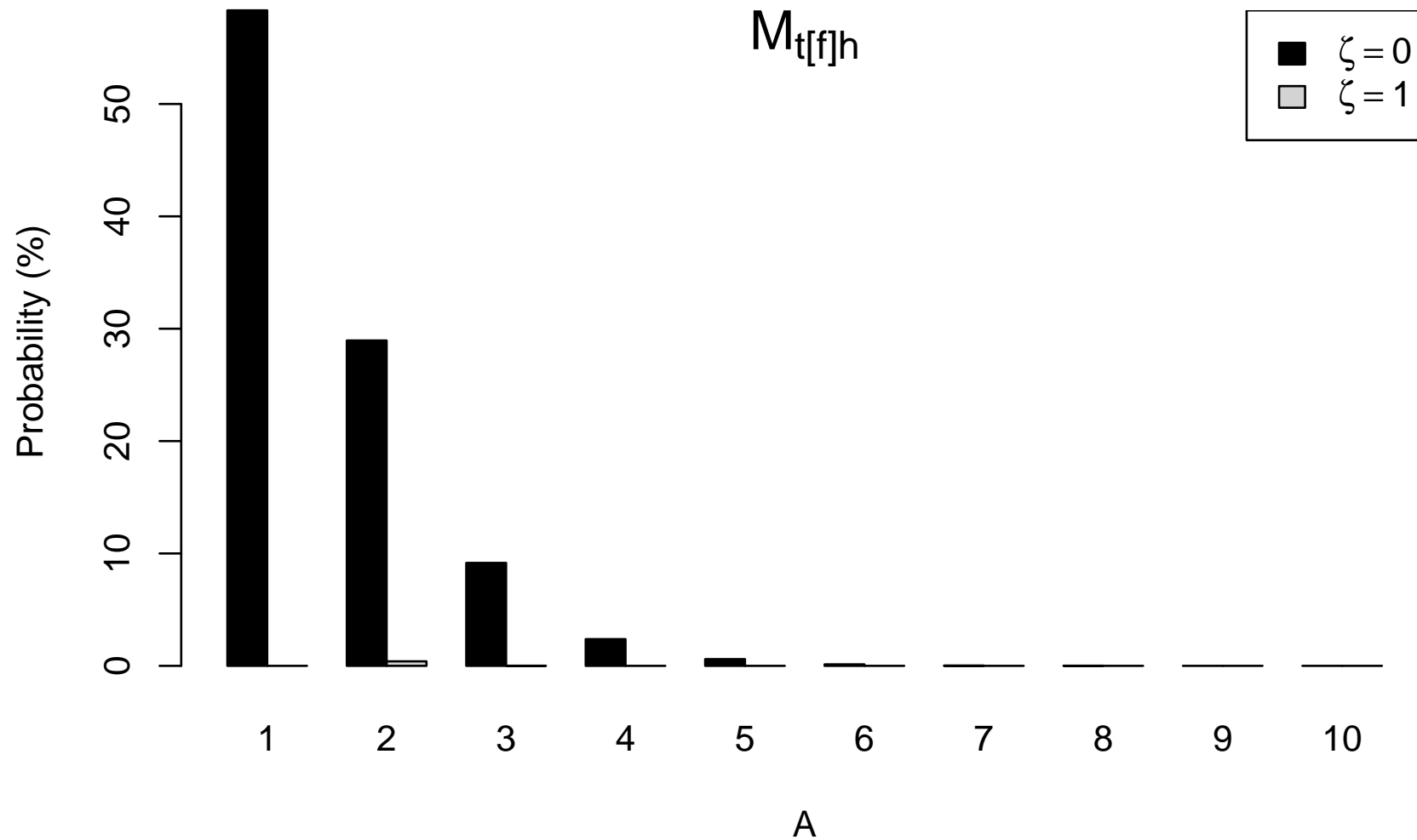
Application: AT&T Switch Testing

M_{th} : 2-way finite mixture



Application: AT&T Switch Testing

$M_{t[f]h}$: 1-way finite mixture with fixed sample effects



Posterior Model Probabilities

A	$p(A, \zeta X, M_{t[f h]})$		$p(A, B, \zeta X, M_{th})$										
	$\zeta = 0$	$\zeta = 1$	$\zeta = 0$					$\zeta = 1$					
B =			1	2	3	4	5	6	2	3	4	5	6
1	58.3			27.5	16.0	9.1	5.3	3.2					
2	29.0	0.4		8.9	5.0	2.8	1.5	0.9	3.4	3.8	2.7	1.6	1.2
3	9.2			1.3	0.7	0.4	0.2	0.1	0.6	1.1	0.7	0.4	0.2
4	2.4			0.2	0.1	0.1				0.2	0.1	0.1	0.1
5	0.6											0.1	
6	0.1												0.1
7													
8													
9													
10													

Posterior Means

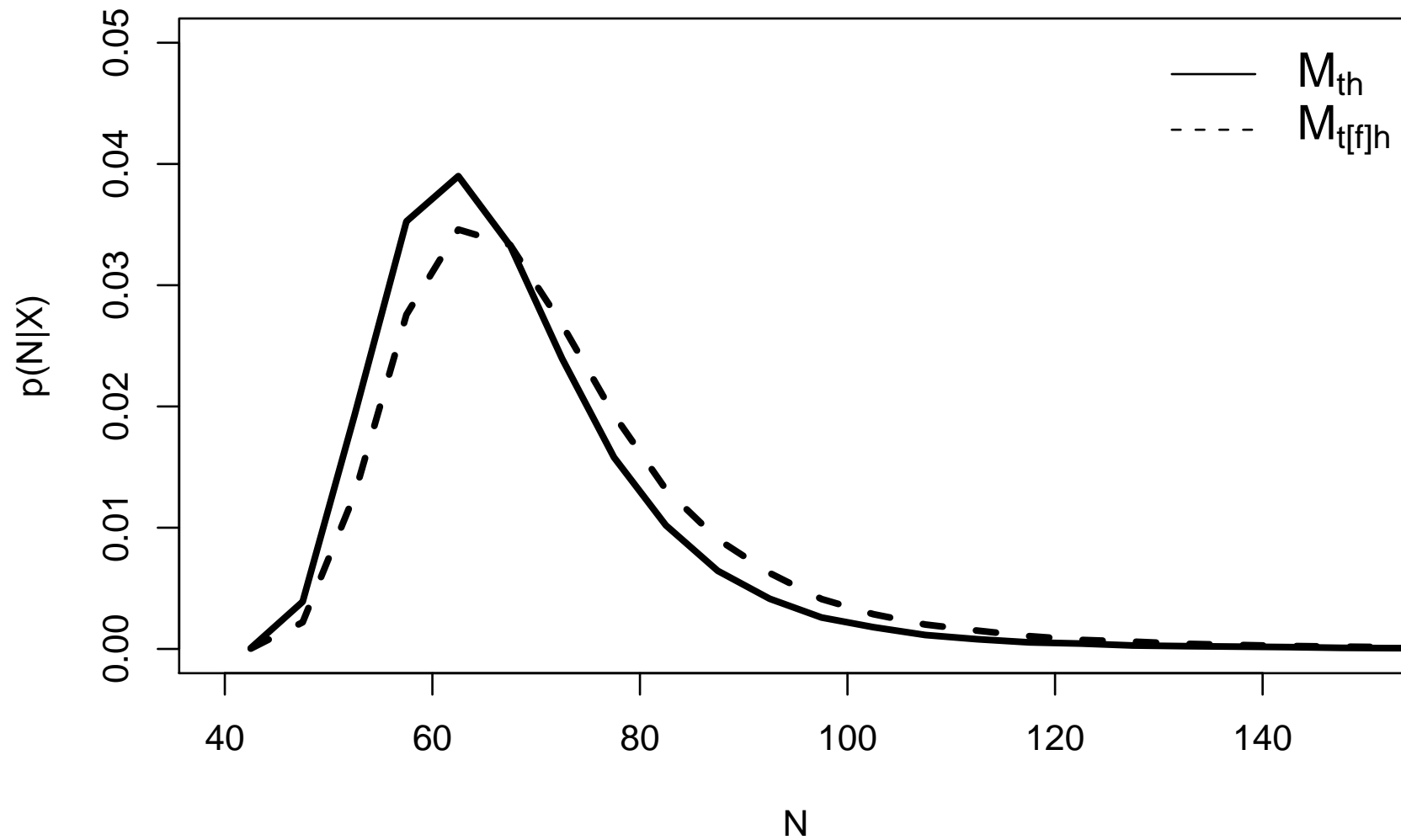
A	$E[N X, M_{t[f]h}]$		$E[N X, M_{th}]$											
	$\zeta = 0$	$\zeta = 1$	$\zeta = 0$					$\zeta = 1$						
1	67.8		78.5	65.9	64.7	64.3	64.2	64.2						
2	78.6	81.8	103.6	72.1	70.5	70.1	70.0	70.0	79.4	79.2	76.8	76.9	75.9	
3	79.6	65.2	100.0	71.2	69.8	69.2	68.3	70.4	76.1	81.4	82.9	83.4	80.0	
4	78.1			71.0	70.8	67.7	67.5	68.9	75.6	85.8	77.8	73.6	72.8	
5	75.5			68.4	67.1	69.5	64.7	64.0	67.5	78.8	75.6	67.0	69.7	
6	75.0			70.2	64.9	62.0	67.8	58.0	65.0	91.3	78.2	64.5	66.1	
7	76.4				62.0						74.0	66.3	67.7	
8	84.0													
9	83.0													
10	64.0													

Application: AT&T Switch Testing

Population size estimates \hat{N} in fixed and variable dimension models.

Model	Mean	Med.	95% CI	Npar
M_0	78.5	76	(63,101)	2
M_{t_2}	65.9	66	(54,97)	10
M_{t_3}	64.7	64	(55,77)	12
M_{t_4}	64.3	62	(60,64)	14
$M_{t_2+h_2}$	72.1	70	(63,77)	12
$M_{t_2 \times h_2}$	79.4	73	(55,122)	13
M_{th}	68.7	66	(51,103)	
$M_{t[f]}$	67.8	66	(52,94)	8
$M_{t[f]+h_2}$	78.6	74	(54,129)	10
$M_{t[f] \times h_2}$				15
$M_{t[f]h}$	72.4	69	(52,114)	

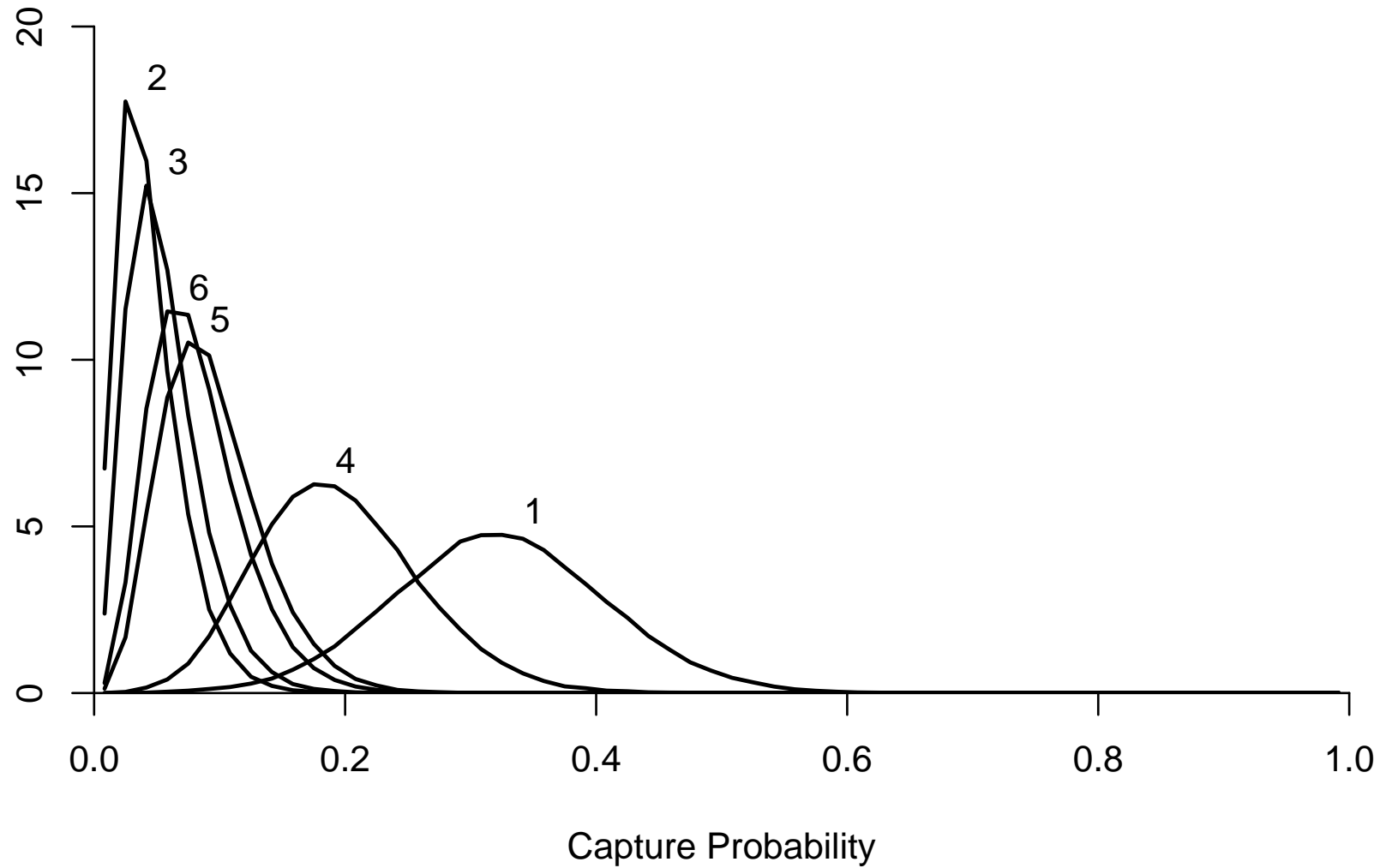
Posterior distribution of N



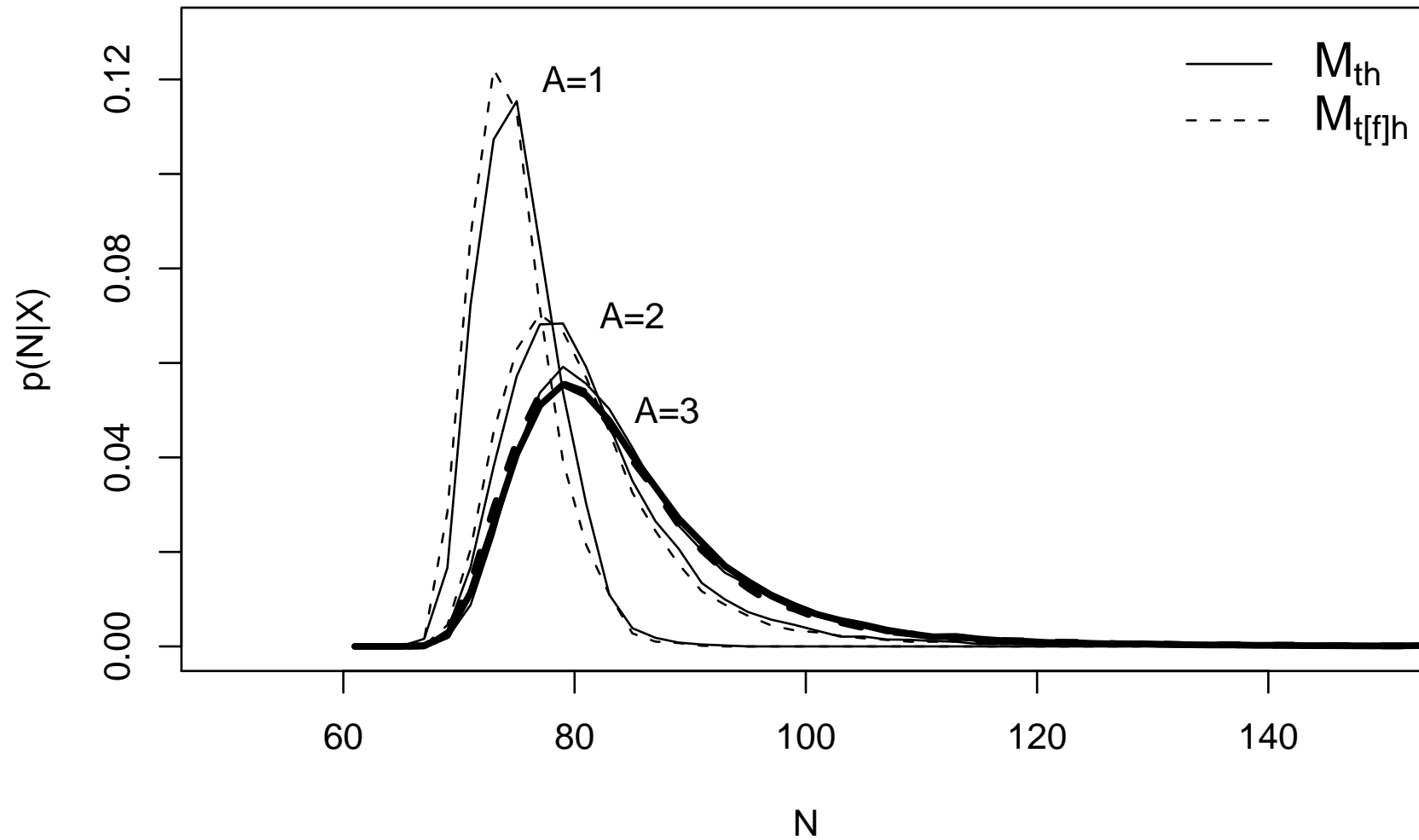
Application: AT&T Switch Testing

- Models M_{th} and $M_{\text{t[f]h}}$ both predict of the order 70 errors in total (95% credible intervals [51,103] and [52,114])
- ~ 30 errors not so far detected
- Evidence for heterogeneity among **reviewers**, but not amongst faults
- Bayes Factor $\text{BF}(M_{\text{th}} : M_{\text{t[f]h}}) = 0.42$: neither model class convincingly favoured.

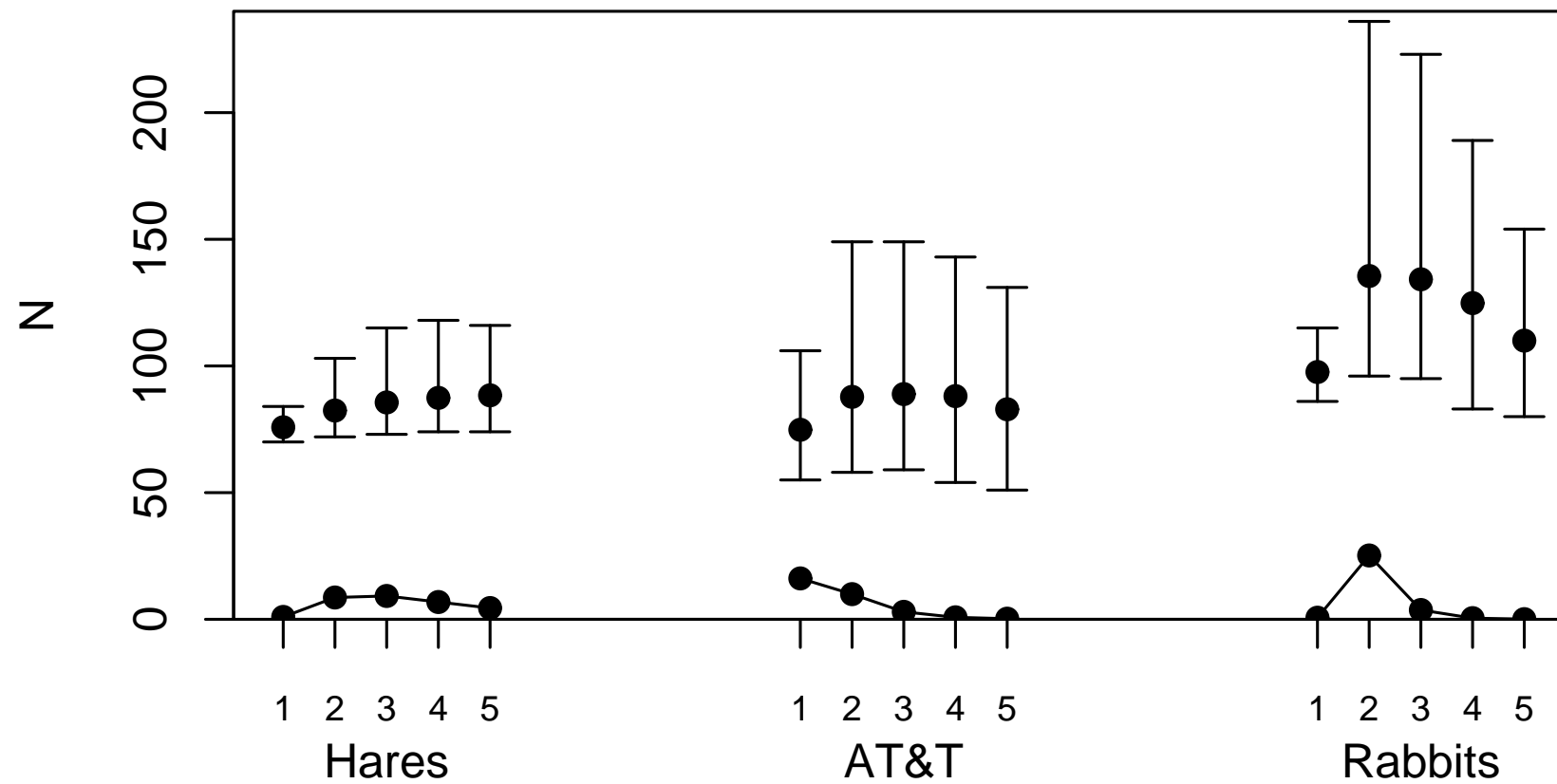
Skills of the 6 reviewers



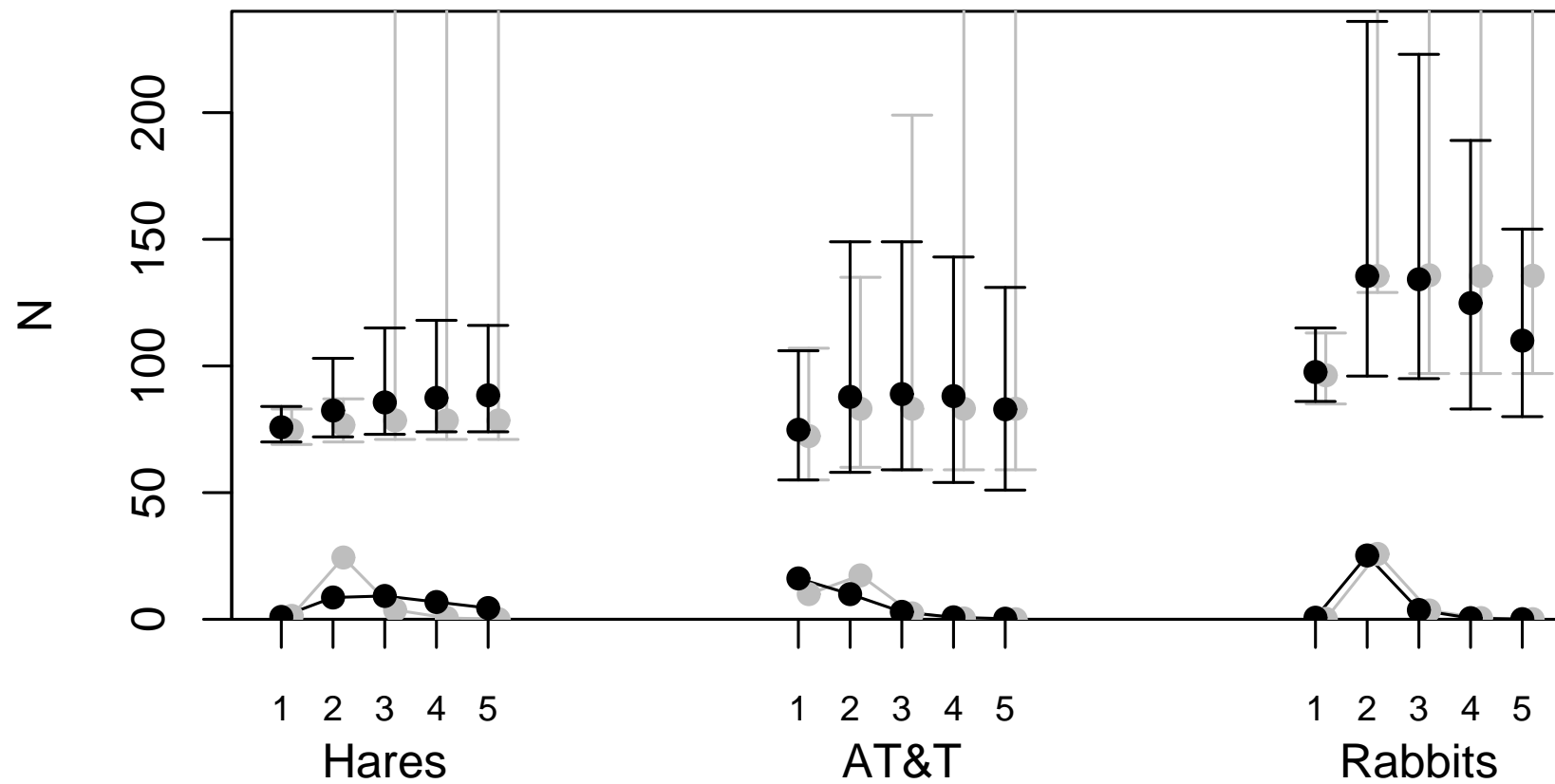
Application: Snowshoe Hares



Interval estimates – 3 examples



Interval estimates – 3 examples



Summary

- RJMCMC is a practicable means for model selection/averaging in Capture-Recapture with finite mixtures.
- Doesn't solve non-identifiability between model classes
- Priors regularise the likelihood – adding extra components does not affect estimates much.