Interval Estimation
Part II

Frequentist Confidence Interval

- Def'n: The fraction of intervals calculated from a large number of data sets generated by the same process that would include the true parameter value

Bayesian Credible Interval

- Def'n: Posterior probability that the parameter lies within the interval

Bayesian Credible Intervals

$$\int_{-\infty}^{A} p(\theta|Y)\,d\theta = \int_{B}^{\infty} p(\theta|Y)\,d\theta = \alpha/2$$

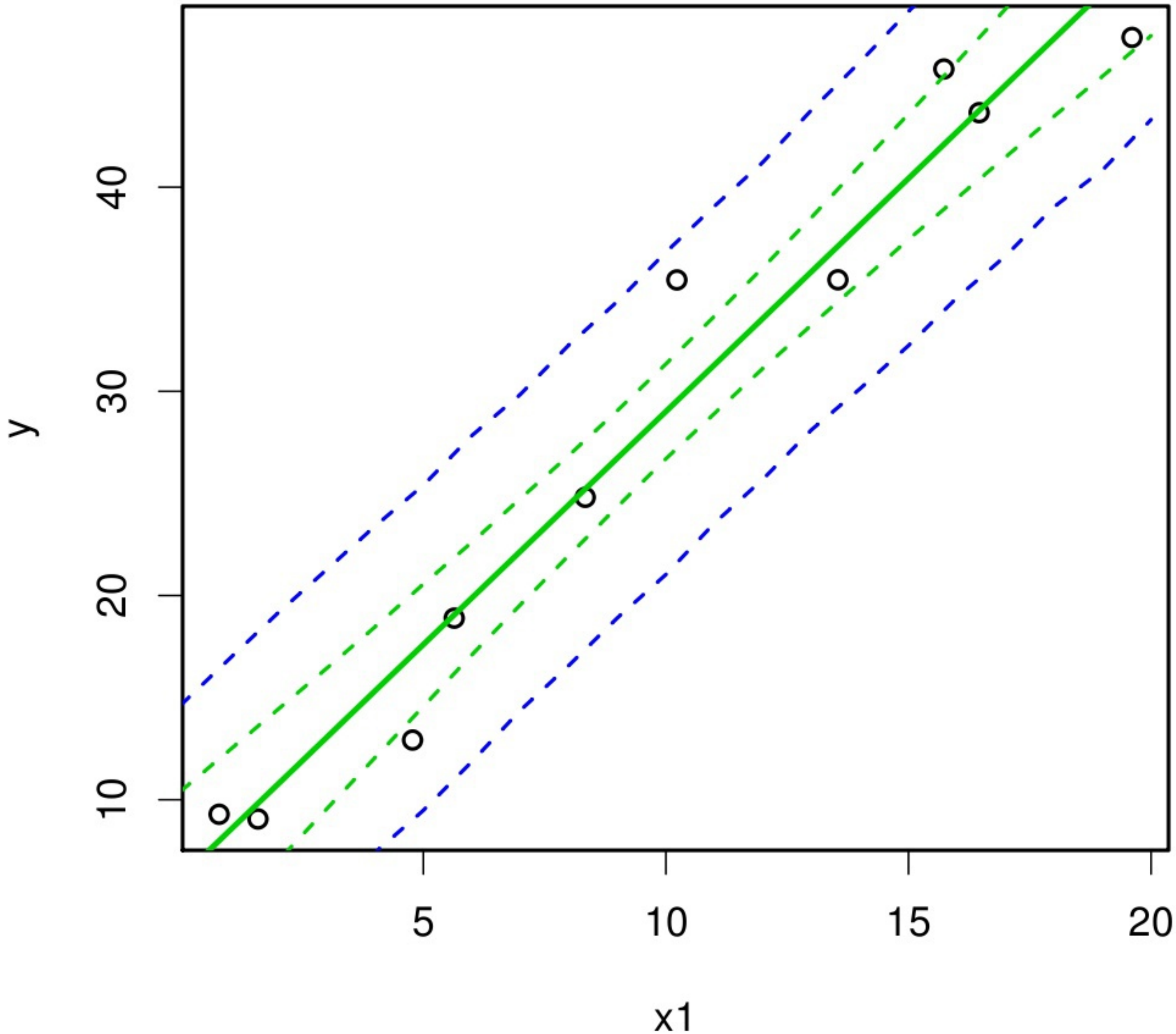
- Parameter CI:
  - Analytical from inverse CDF (e.g. qnorm)
  - Numerically from quantiles (R: quantile)
- **NOT estimated based on standard deviation**
- Not necessarily symmetric
- Requires no additional assumptions

Bayesian Prediction

- Consider an observed data set Y and a model with parameters θ
- Want to calculate the posterior PDF of some <u>*new* data point y'</u>
- Need to integrate over all values θ can take on for ALL the model parameters (including variances)

$$p(y'|y) = \int \underbrace{p(y'|\theta)}_{Likelihood\ of\ new\ data} \underbrace{p(\theta|Y)}_{Posterior} d\theta$$

Bayesian Prediction Intervals

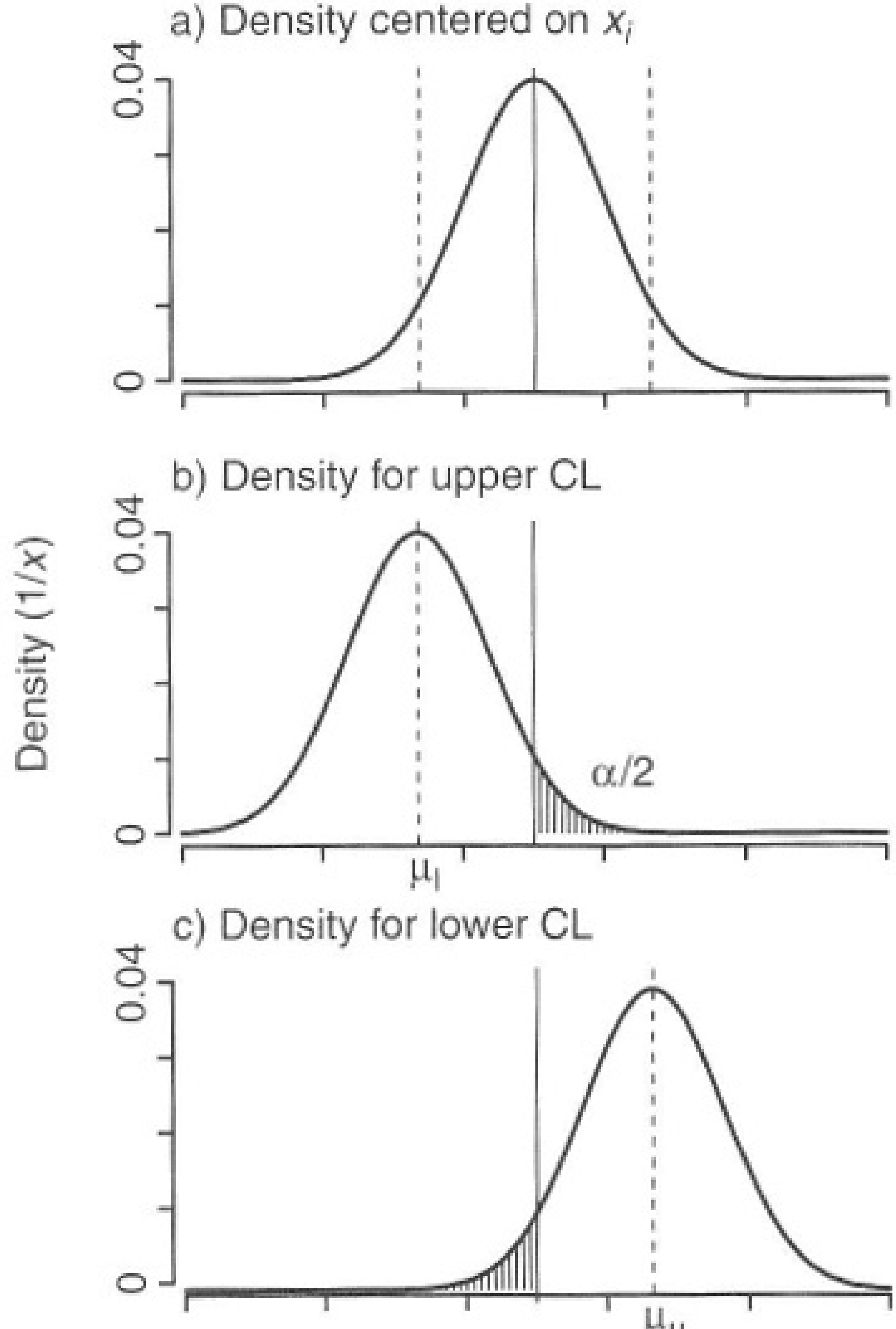


- CI of p(y'|y) for each x
- Includes both data and parameter uncertainty

Frequentist Confidence Interval

- Will consider four approaches to estimating confidence interval
  - Standard Error
  - Likelihood Profile
  - Fisher Information
  - Bootstrap
- All require additional assumptions

- Frequentist CI does not assume a density centered on the MLE
- Cannot integrate likelihood profile
- Assumes density **centered on upper/lower bound** and calc. tail probabilities
- Equivalent if symmetric (e.g. Normal)



Normal CI

- Goal:
  - Find μ_u and μ_l the locations where the distributions should be centered so that they have the desired tail probability

$$\int_{\mu_{MLE}}^{\infty} N(x|\mu_u, \sigma^2) = \alpha/2$$

$$\int_{-\infty}^{\mu_{MLE}} N(x|\mu_l, \sigma^2) = \alpha/2$$

- As we know at $\alpha = 0.05$ (95% CI) these are located at +/- 1.96 σ^2
- Approx 1.96 SE

Std Error Estimator

- Only approximate if not Normal
- Always symmetric
- Can lead to non-sensible estimates for other distributions
- **Choice of likelihood not equivalent to distribution of parameter estimator**
- Other methods of estimating frequentist CI based on likelihood surface

Likelihood Profile

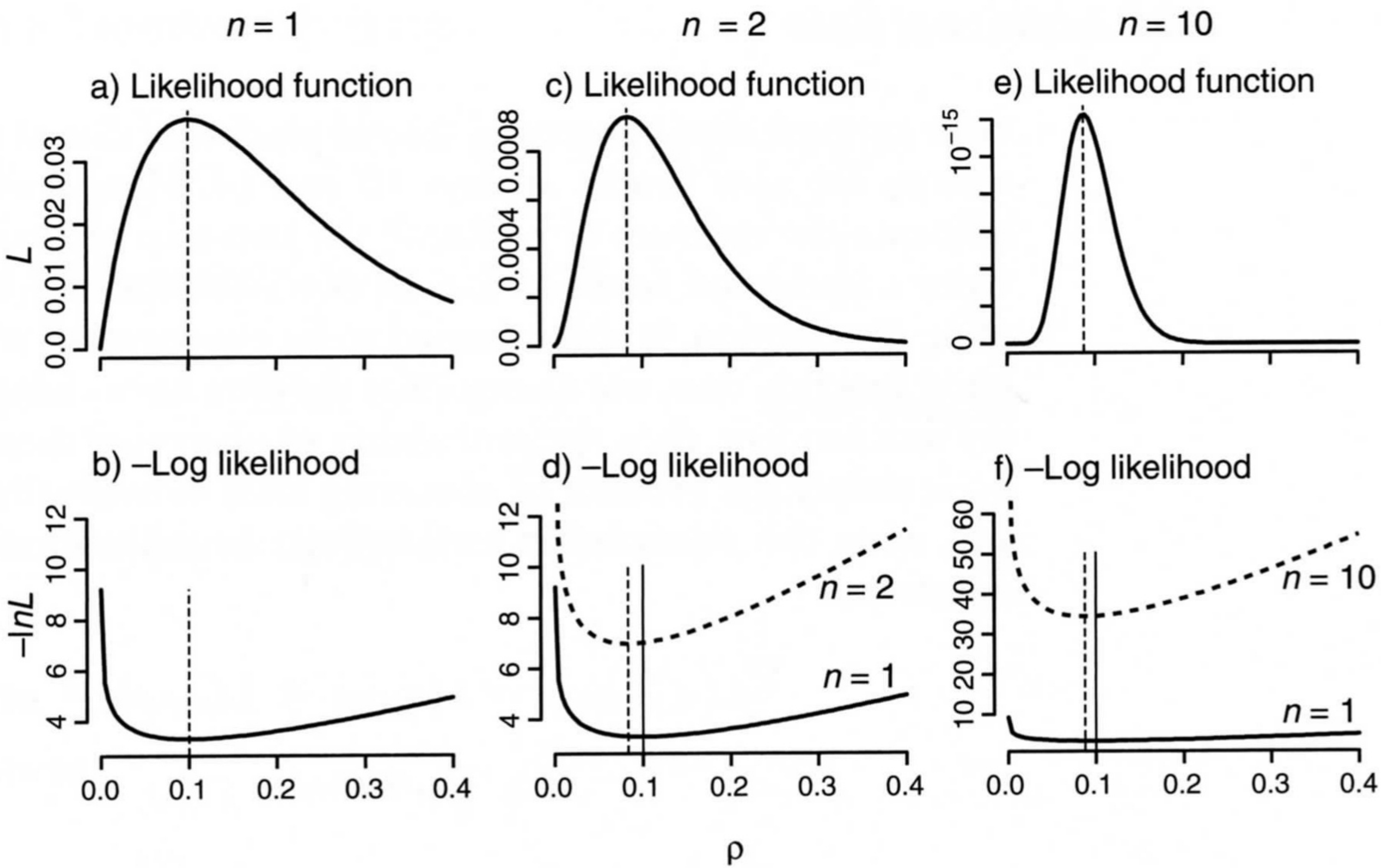


FIGURE 3.2. Likelihood functions for the exponential model with three different sample sizes. Note the different scales on the vertical axes.

Likelihood Profile

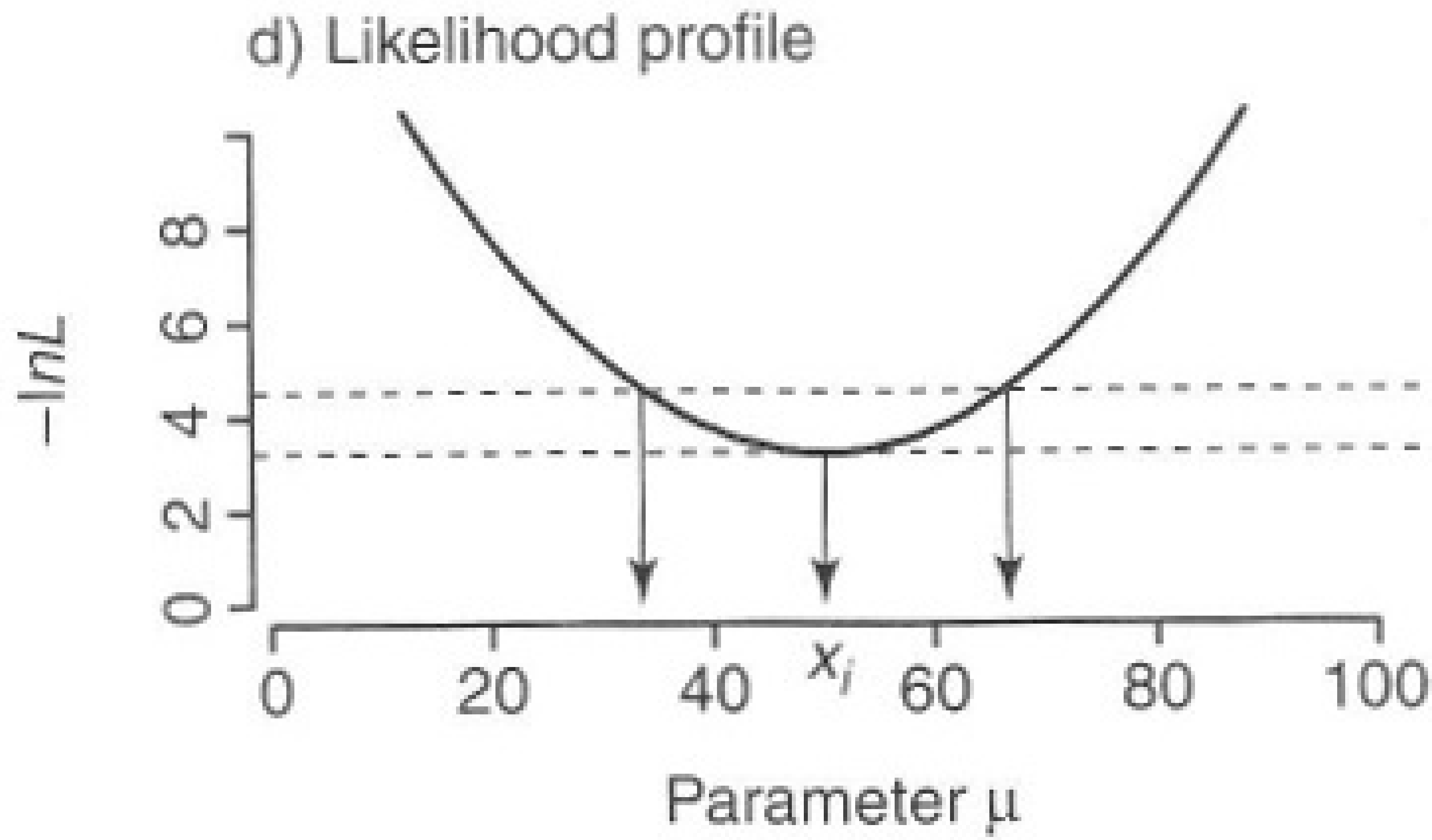
- Narrows as the sample size increases
- More constrained estimate of the parameter
- Can't interpret as a PDF
- Can we use to estimate uncertainty in the parameter, and if so HOW?

Likelihood Ratio Test

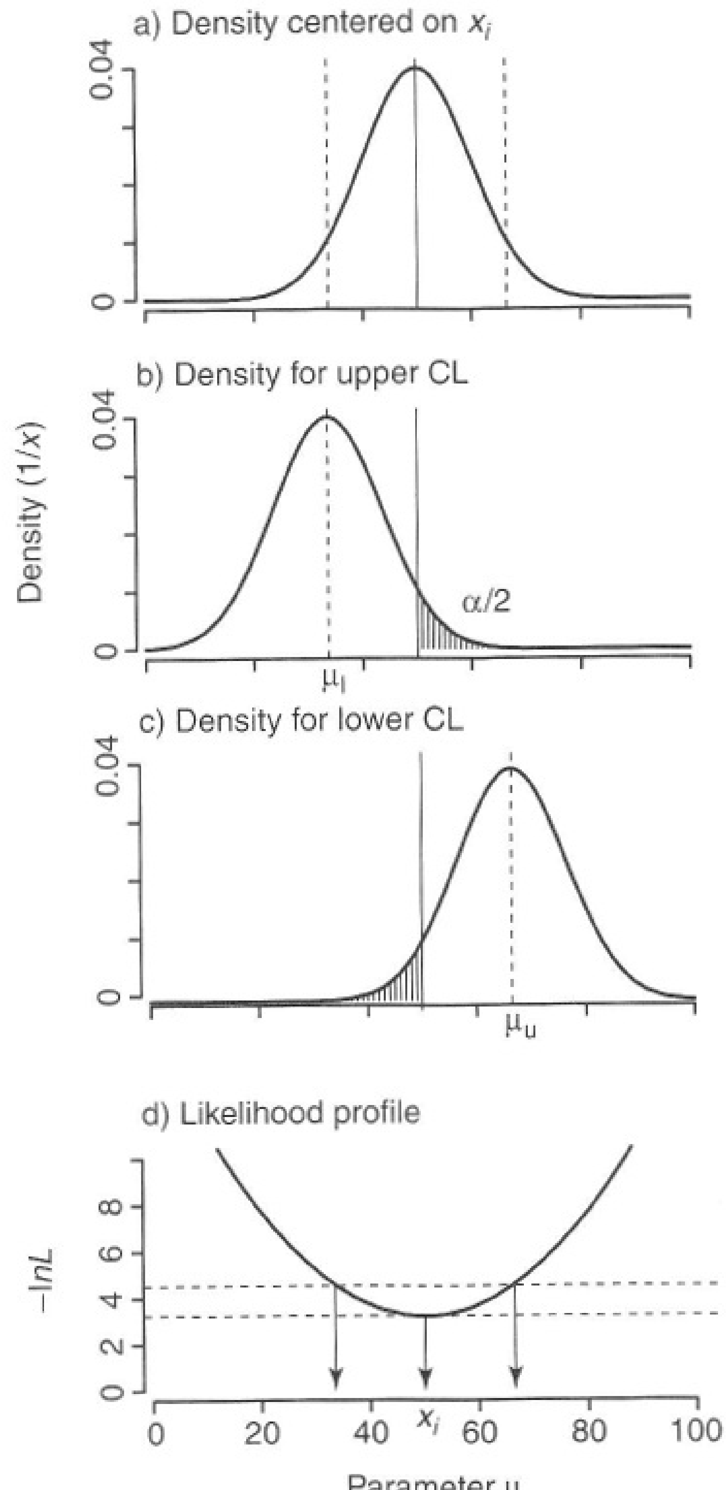
- Recall from Lab 4 that we compared two nested models using the Likelihood Ratio test, which was based on the ratio of the likelihood (equivalently the differences in log likelihood)
- Can similarly construct a CI based on the LRT

Example

- Consider a likelihood function for dataset x that has a parameter mu: $L(x|\mu)$
- Compare the likelihood of some arbitrary μ_0 as compared to the MLE: μ_{MLE}.
- $LR = L(x|\mu_0)/L(x|\mu_{MLE})$
- $D = -2\ln L(x|\mu_0) - -2\ln L(x|\mu_{MLE})$
- The test statistic D is known to be distributed with a χ^2 distribution with df 1 (1 parameter)



- CI constructed based on difference in Deviance from MLE
- χ^2(0.95,1 d.f.) --> ΔD ~ 3.84, ΔlnL ~ 1.92
  - qchisq(0.95,1)
- p-value = 1-pchisq(D,1)



a) Density centered on x_i
b) Density for upper CL
c) Density for lower CL
d) Likelihood profile
Density (1/x)
0.04
0
α/2
μl
μu
−lnL
0
2
4
6
8
0
20
40
x_i
60
80
100
Parameter μ

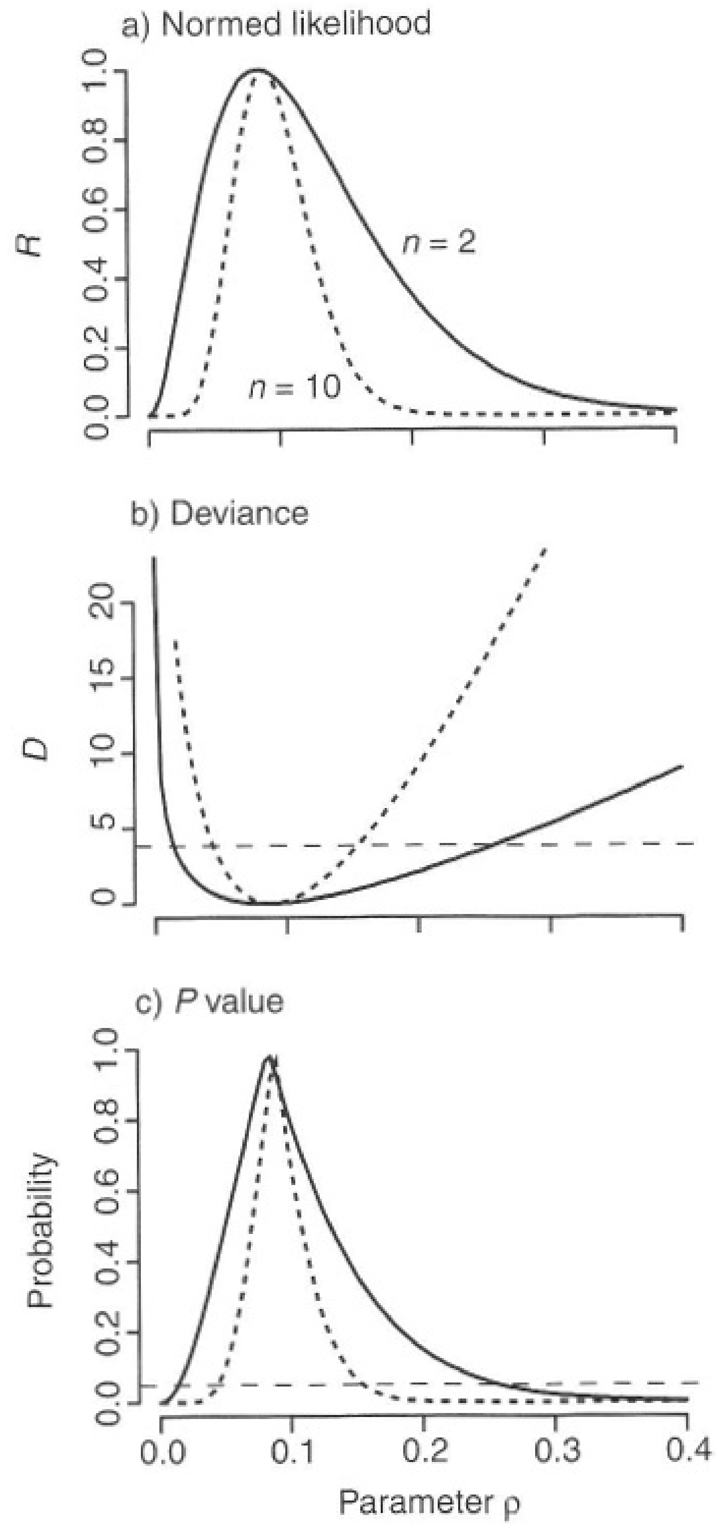
Exponential

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

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

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

- Produces asymmetric CI



Frequentist Confidence Interval

- Will consider four approaches to estimating confidence interval
  - Standard Error
  - Likelihood Profile
  - **Fisher Information**
  - Bootstrap
- All require additional assumptions

Fisher's Information

- Uses the curvature of the ln likelihood to estimate variance of parameter error dist'n

$$I = \left. \frac{-d^2 \ln L(\theta)}{d\theta^2} \right|_{\theta_{ML}}$$

$$se_\theta = \frac{1}{\sqrt{(I)}}$$

- Quadratic approximation of lnL (exact for N)
- C.I. *is* based on standard error

Example: Normal mean

- Assume L = N(x|μ,σ^2)

$$L = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{\sum (x_i - \mu)^2}{2\sigma^2} \right)$$

$$lnL = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{\sum (x_i - \mu)^2}{2\sigma^2}$$

$$\frac{\partial\, lnL}{\partial \mu} = \frac{1}{\sigma^2}\sum (x_i - \mu) = \frac{\sum x_i}{\sigma^2} - \frac{n\mu}{\sigma^2}$$

$$\frac{\partial^2 lnL}{\partial \mu^2} = \frac{-n}{\sigma^2} \longrightarrow I = \frac{n}{\sigma^2} \longrightarrow se = \frac{\sigma}{\sqrt{n}}$$

Example: Exponential

$$L = \prod \rho \exp\left(-\rho x_i\right)$$

$$lnL = n \ln(\rho) - \rho \sum x_i$$

$$\frac{\partial\, lnL}{\partial \rho} = \frac{n}{\rho} - \sum x_i$$

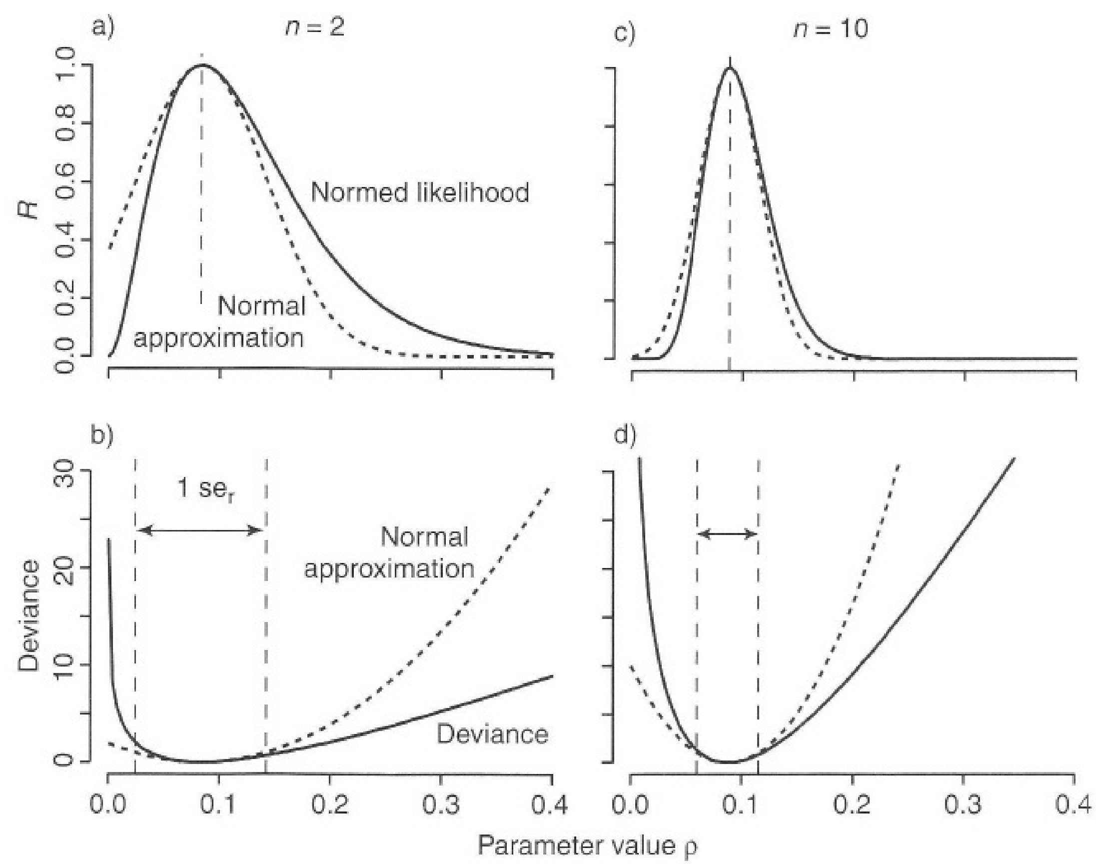
$$\frac{\partial^2 lnL}{\partial \rho^2} = -\frac{n}{\rho^2}$$

$$I = \frac{n}{\rho^2}$$

$$se = \frac{\rho_{ML}}{\sqrt{(n)}}$$

Fisher's Pro/Con

- Analytical solution
  - Generalization
  - Requires Math
- Approximation
  - Can be biased, especially at small sample size
- Asymptotic
  - CLT assumption that parameter dist'n is asymptotically Normal



Frequentist Confidence Interval

- Will consider four approaches to estimating confidence interval
  - Standard Error
  - Likelihood Profile
  - Fisher Information
  - **Bootstrap**
- All require additional assumptions

Bootstrap

- Monte Carlo method (numerical)
- Based on idea of generating parameter distribution based on large number of **replicate data sets** that are the *same size* as original (data random)
- Two variants
  - Parametric: pseudodata
  - Nonparametric: resample data

Non-parametric bootstrap

- Draw a replicate data set by resampling from the original data
- Fit parameters to resample
- Repeat procedure n times
- Estimate parameter CI based on sample quantiles
- Estimate parameter std error as sample s.d.

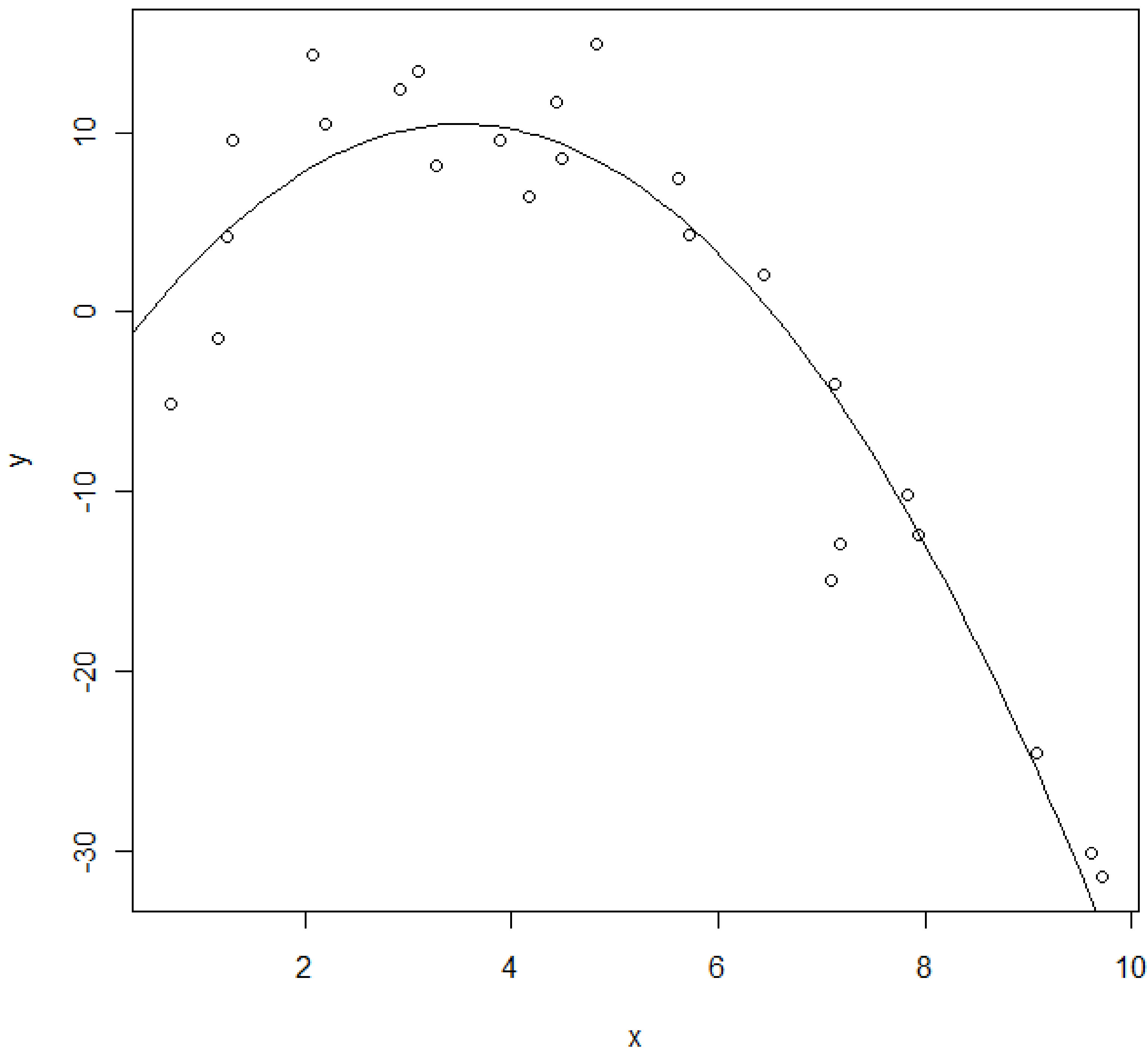
Resampling

Original	1	2	3	4	5	6	7	8	9	10
Sample 1	5	3	5	1	7	9	10	8	2	4
Sample 2	4	9	5	6	5	9	3	10	10	2
...										
sample N	4	10	3	9	2	9	6	5	2	6

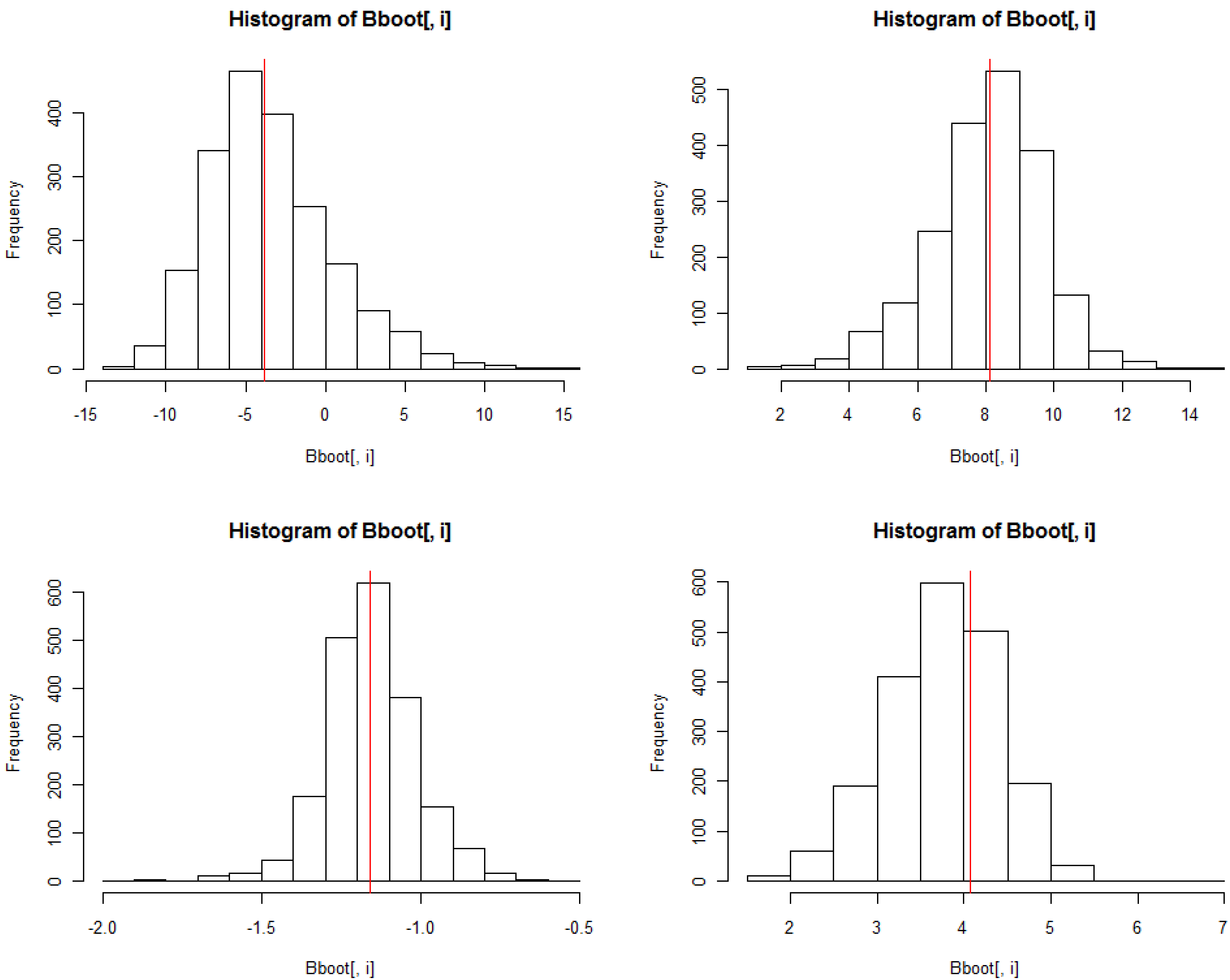
- For CI, resample all covariates simultaneously
- For null model, resample response variable independent of covariates
- Difficult for highly structured data
- R: sample(x,length(x),replace=TRUE)
  - Hint: For simult. sample the *indices* not the data

R example: Quadratic

```
lnL <- function(beta,x,y){                    ### - ln likelihood
-sum(dnorm(y,beta[1] + beta[2]*x + beta[3]*x^2,beta[4],log=TRUE))
}
ic <- c(mean(y),0,0,sd(y))                    ###  initial condition

outMLE <- optim(ic,lnL,x=x,y=y)               ### MLE fit

### general code for non-parameteric bootstrap
nboot <- 2000
Bboot <- matrix(NA,nboot,4)                   ### storage
for(i in 1:nboot){
  samp <- sample(1:length(y),length(y),replace=TRUE)  ### sample
  out <- optim(ic,lnL,x=x[samp],y=y[samp])            ### fit sample
  Bboot[i,] <- out$par
}
```



10
0
-10
-20
-30
y
2
4
6
8
10
x

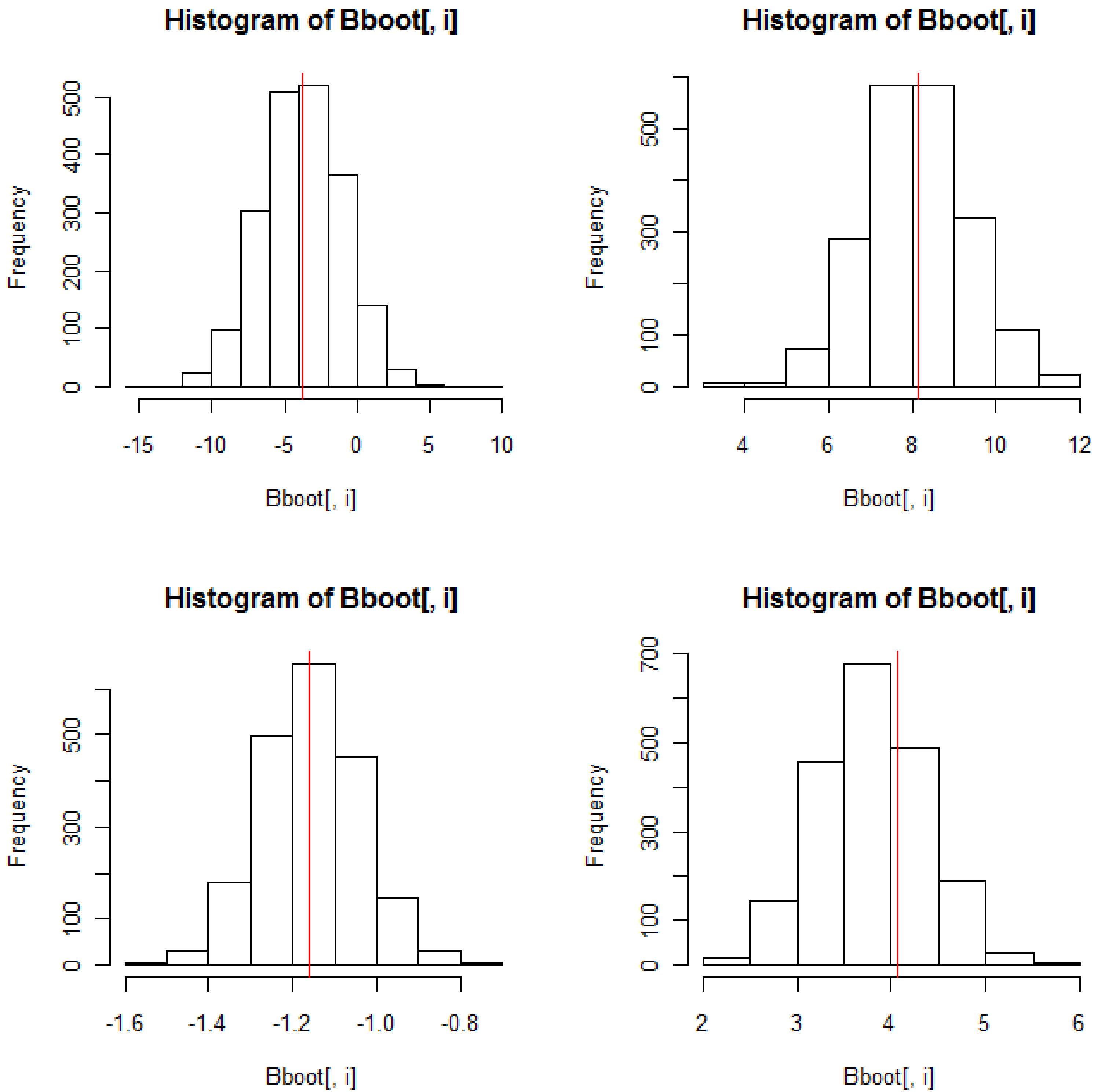

Histogram of Bboot[, i]
Frequency
0
100
200
300
400
-15
-10
-5
0
5
10
15
Bboot[, i]
Histogram of Bboot[, i]
Frequency
0
100
200
300
400
500
2
4
6
8
10
12
14
Bboot[, i]
Histogram of Bboot[, i]
Frequency
0
100
200
300
400
500
600
-2.0
-1.5
-1.0
-0.5
Bboot[, i]
Histogram of Bboot[, i]
Frequency
0
100
200
300
400
500
600
2
3
4
5
6
7
Bboot[, i]

Parametric bootstrap

- Based on parameters fit to original data set generate pseudodata with same dist'n
- Fit parameters to resample
- Repeat procedure n times
- Estimate parameter CI based on sample quantiles
- Estimate parameter std error as sample s.d.

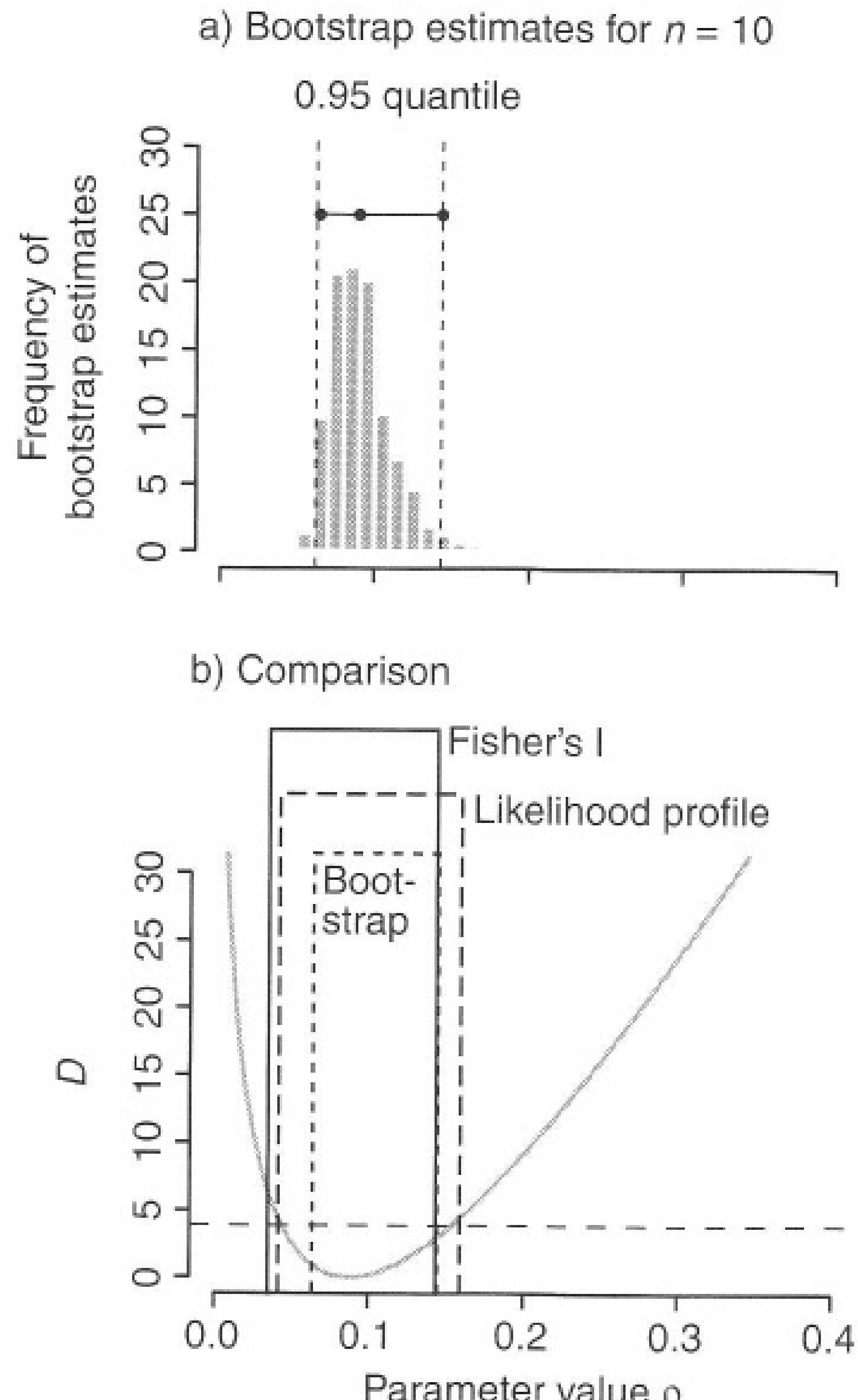
R example: Quadratic

```
lnL <- function(beta,x,y){                        ### - ln likelihood
-sum(dnorm(y,beta[1] + beta[2]*x + beta[3]*x^2,beta[4],log=TRUE))
}
ic <- c(mean(y),0,0,sd(y))                        ###  initial condition

outMLE <- optim(ic,lnL,x=x,y=y)                   ### MLE fit
beta <- outMLE$par

### general code for non-parameteric bootstrap
nboot <- 2000
Bboot <- matrix(NA,nboot,4)                       ### storage
for(i in 1:nboot){
  yboot <- rnorm(n,beta[1] + beta[2]*x + beta[3]*x^2,beta[4]) ##pseudo
  out <- optim(ic,lnL,x=x,y=yboot)                ### fit pseudo
  Bboot[i,] <- out$par
}
```


Histogram of Bboot[, i]
Frequency
0
100
200
300
400
500
-15
-10
-5
0
5
10
Bboot[, i]
Histogram of Bboot[, i]
Frequency
0
100
300
500
4
6
8
10
12
Bboot[, i]
Histogram of Bboot[, i]
Frequency
0
100
300
500
-1.6
-1.4
-1.2
-1.0
-0.8
Bboot[, i]
Histogram of Bboot[, i]
Frequency
0
100
300
500
700
2
3
4
5
6
Bboot[, i]

Bootstrap Pro/Con

- No fancy math
- Code/computation (though less than MCMC)
- Easy to extend to
  - Multiple parameter models
  - Estimation of covariance
  - Prediction
- Nonparameteric:
  - Inference limited to sample
  - Not for small sample size (var sample < var pop'n)
- Parameteric: assumes model is true



a) Bootstrap estimates for n = 10
0.95 quantile
Frequency of bootstrap estimates
0
5
10
15
20
25
30
b) Comparison
Fisher's I
Likelihood profile
Boot-strap
D
0
5
10
15
20
25
30
0.0
0.1
0.2
0.3
0.4
Parameter value ρ

Classic Error Propagation

- Taylor series approximation:

$$var[f(x)] \approx \sum \left(\frac{\partial f}{\partial \theta_i} \right)^2 var[\theta_i] + \sum_{i \neq j} \left(\frac{\partial f}{\partial \theta_i} \right) \left(\frac{\partial f}{\partial \theta_j} \right) cov[\theta_i, \theta_j]$$

- Individual parameter variances generated from methods discussed earlier (Fisher's I, bootstrap, MCMC)

Example: Regression CI

$$f(x) = \beta_0 + \beta_1 x$$

$$\frac{\partial f}{\partial \beta_0} = 1 \qquad \frac{\partial f}{\partial \beta_1} = x$$

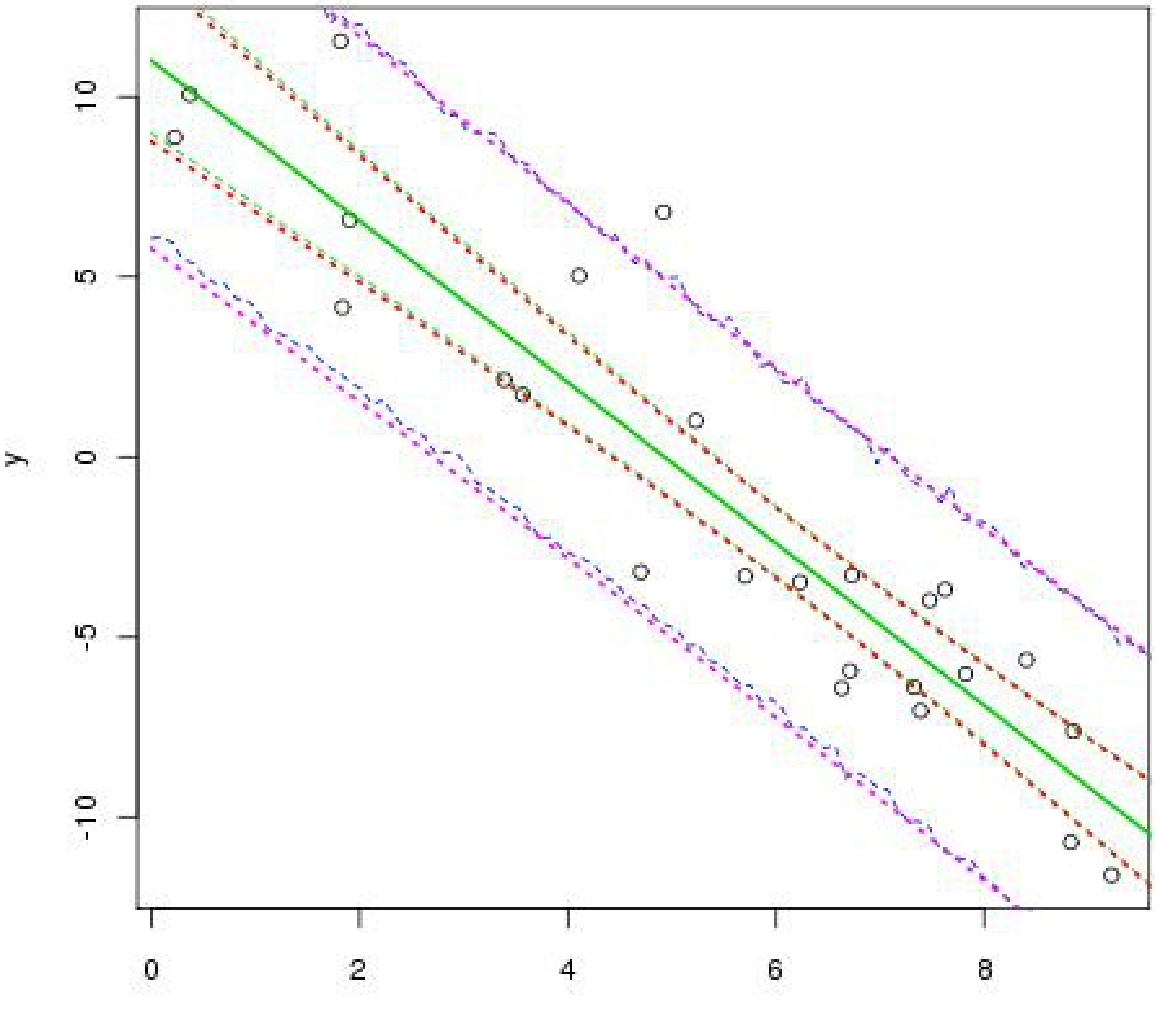
$$Var[f(x)] \approx 1 \cdot Var[\beta_0] + x^2 Var[\beta_1] + x\, Cov[\beta_{0,} \beta_1]$$

Example: Regression PI

$$f(x) = \beta_0 + \beta_1 x + \epsilon$$

$$\frac{\partial f}{\partial \beta_0} = 1 \qquad \frac{\partial f}{\partial \beta_1} = x \qquad \frac{\partial f}{\partial \epsilon} = 1$$

$$Var[f(x)] \approx Var[\beta_0] + x^2 Var[\beta_1] + \sigma^2 + x\, Cov[\beta_{0,} \beta_1]$$



y
10
5
0
-5
-10
0
2
4
6
8
x

Frequentist Confidence Interval

- Will consider four approaches to estimating confidence interval
  - Standard Error (+/- 1.96 se)
  - Likelihood Profile (+3.84 Deviance)
  - Fisher Information $se_\theta = \frac{1}{\sqrt{(I)}}$
  - Bootstrap (simulation)
- Model CI and PI
  - Bootstrap → Monte Carlo
  - Taylor Series approximation