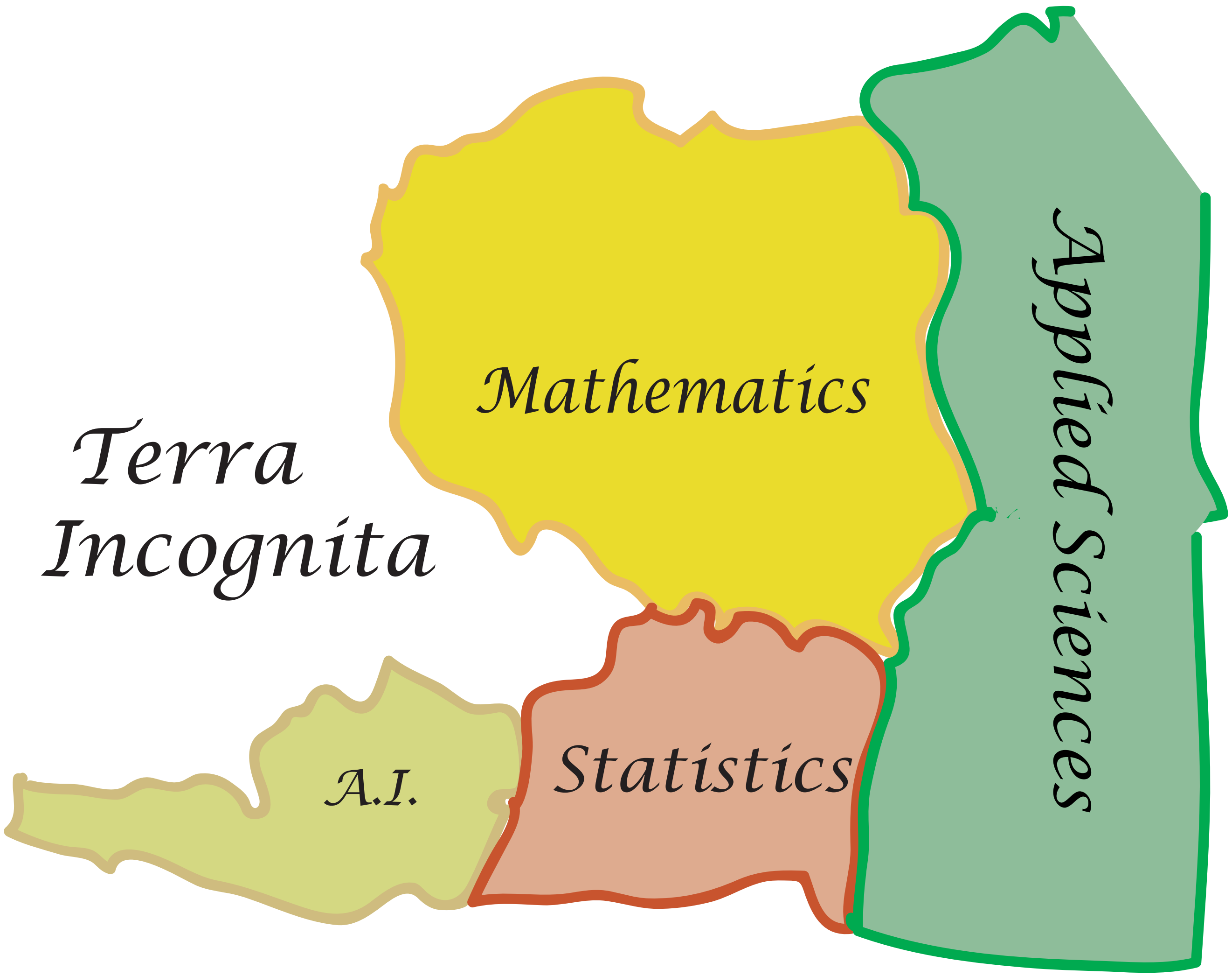
A 250-Year Argument

Belief, Behavior, and the Bootstrap

Bradley Efron

Stanford University

The Greater World of Mathematics and Science



The Physicist's Twins

- Sonogram "Twin boys on the way!"
- Physicist "What's the probability my twins will be *identical* rather than *fraternal*?"
- Doctor "One third of twins are identical."

Bayes Rule for the Twins

- Prior odds: $\dfrac{\Pr\{\text{identical}\}}{\Pr\{\text{fraternal}\}} = \dfrac{1/3}{2/3} = \dfrac{1}{2}$ (past experience)
- Likelihood ratio: $\dfrac{\Pr\{\text{same sex}|\text{identical}\}}{\Pr\{\text{same sex}|\text{fraternal}\}} = \dfrac{1}{1/2} = 2$ (current evidence)
- Posterior odds: $\dfrac{\Pr\{\text{identical}|\text{same sex}\}}{\Pr\{\text{fraternal}|\text{same sex}\}} = ?$ (updated beliefs)
- Bayes rule:

$$\boxed{\text{Posterior odds} = \text{Prior odds} \cdot \text{Likelihood ratio}} = \frac{1}{2} \cdot 2 = 1$$

- *My answer*: "50/50"

If All Twins Were Sonogrammed:

Sonogram shows:

Twins are:

	Same sex	*Different*	
Identical	*a* **1/3**	*b* **0**	1/3
Fraternal	*c* **1/3**	*d* **1/3**	2/3

Doctor

Physicist

Belief and Inference

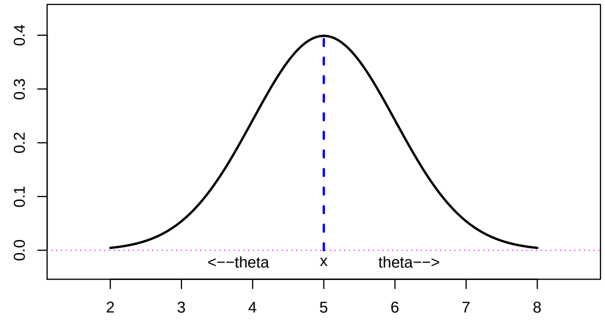
- θ: unknown state of nature (identical or fraternal?)
- $\pi(\theta)$: prior beliefs for θ (1/3, 2/3)
- x: current evidence (sonogram)
- $f_\theta(x)$: probability model for x given θ
- Question **What is $\pi(\theta|x)$?** (posterior beliefs given x)

Bayes Rule (1763)

- $$\pi(\theta|x) = \quad c\pi(\theta) \quad \cdot \quad f_\theta(x)$$

↑	↑	↑
posterior	prior	likelihood
beliefs	beliefs	function

- "c" makes $\pi(\theta|x)$ sum to 1
- Likelihood function $f_\theta(x)$ with x fixed, θ varying, e.g., $f_\theta(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\theta-x)^2}$:



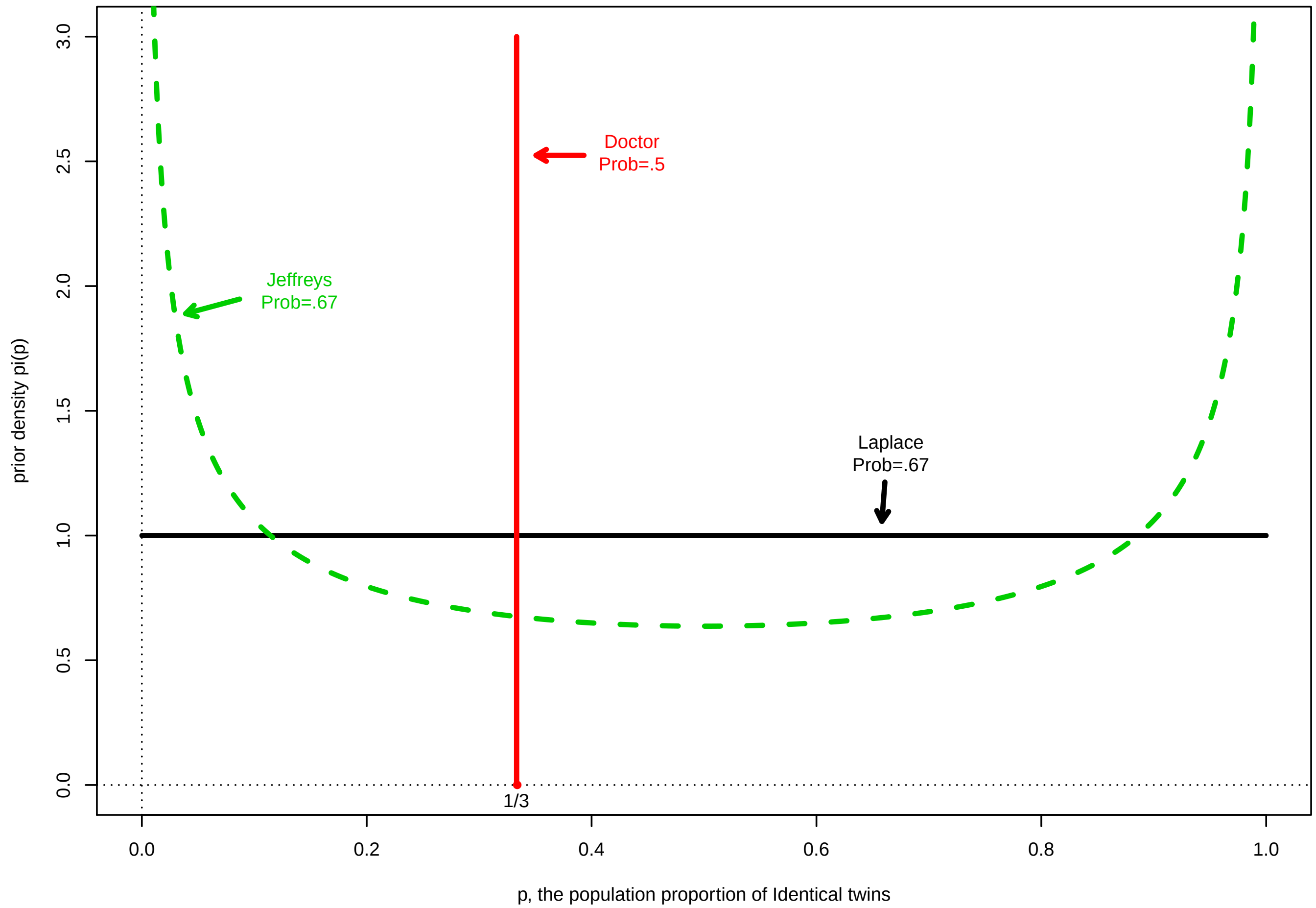
Bayes Inference without Prior Experience

"Objective Bayes"

- "p" population proportion of identical twins [*Doctor*: $p = \frac{1}{3}$]
- Principle of insufficient reason (Laplace, Bernoulli) "In the absence of prior experience, assume p equally likely to have any value between 0 and 1." [*opposed* Venn, Keynes, Fisher]
- Invariant prior (Harold Jeffreys, 1930s):

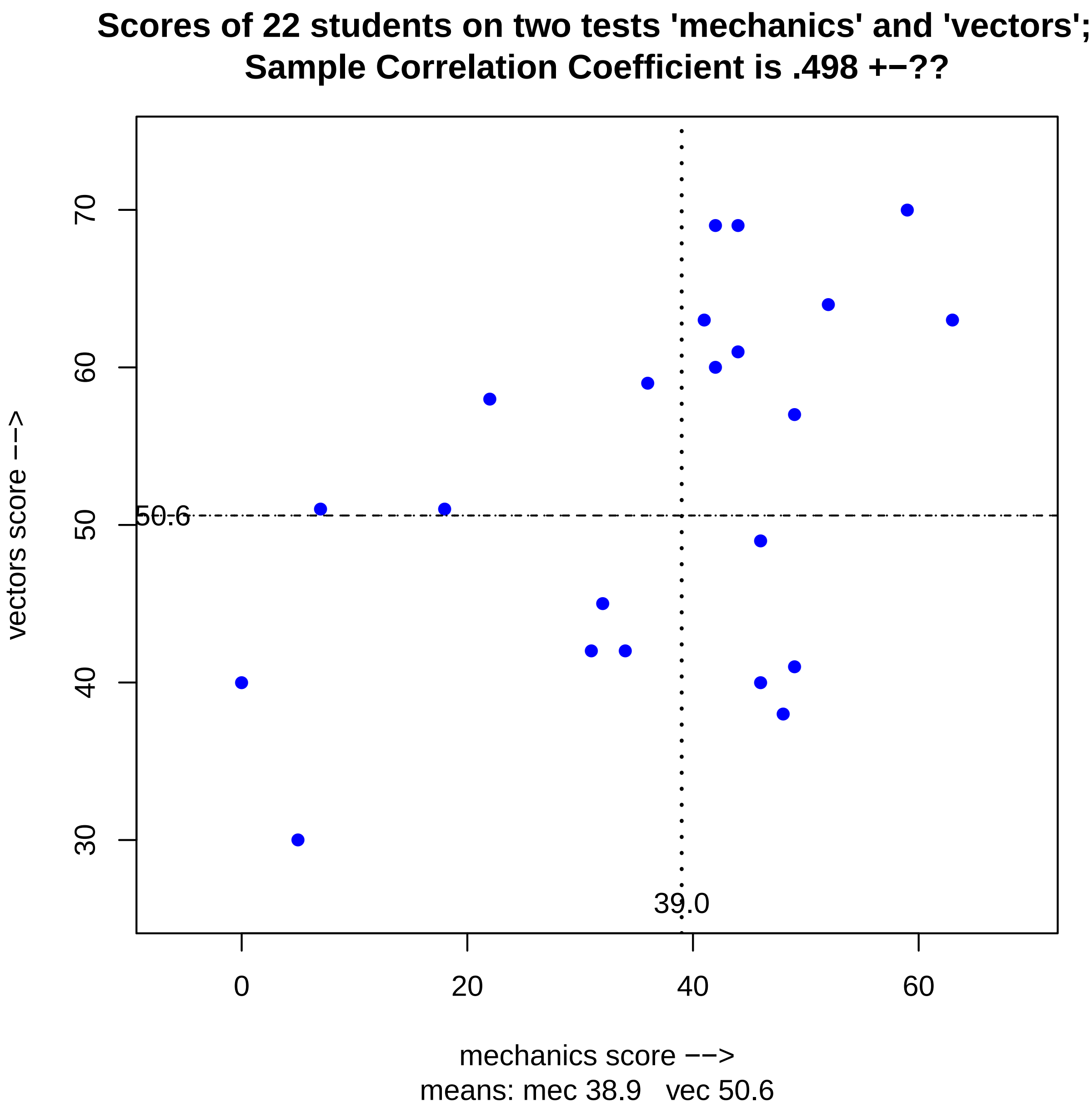
$$\boxed{\pi(p) = cp^{-\frac{1}{2}}(1-p)^{-\frac{1}{2}}}$$

Possible prior densities for p, the Population prop Identical,
And the corresponding predictions for the Physicist



Frequentist Statistics (Behaviorism)

- θ = unknown parameter, x = observed data,
  $f_\theta(x)$ probability model (but no prior beliefs $\pi(\theta)$)
- "$t(x)$" some statistical procedure
  (test, estimate, confidence interval, ...)
- Inference based on *behavior* of $t(x)$ in repeated use
- Optimality   find *best* $t(x)$
  (R.A. Fisher, 1920s; J. Neyman, 1930s)


Scores of 22 students on two tests 'mechanics' and 'vectors';
Sample Correlation Coefficient is .498 +−??
vectors score −−>
70
60
50
40
30
50.6
39.0
0
20
40
60
mechanics score −−>
means: mec 38.9   vec 50.6

Student Score Data

- $n = 22$ students' scores on two tests: *mechanics*, *vectors*
- Data $\boldsymbol{y} = (y_1, y_2, \ldots, y_{22})$ with $y_i = (\text{mec}_i, \text{vec}_i)$
- *Parameter of interest* $\theta =$ correlation (mec, vec)
- *Sample correlation coefficient* $\hat{\theta} = 0.498\pm$??

R. A. Fisher

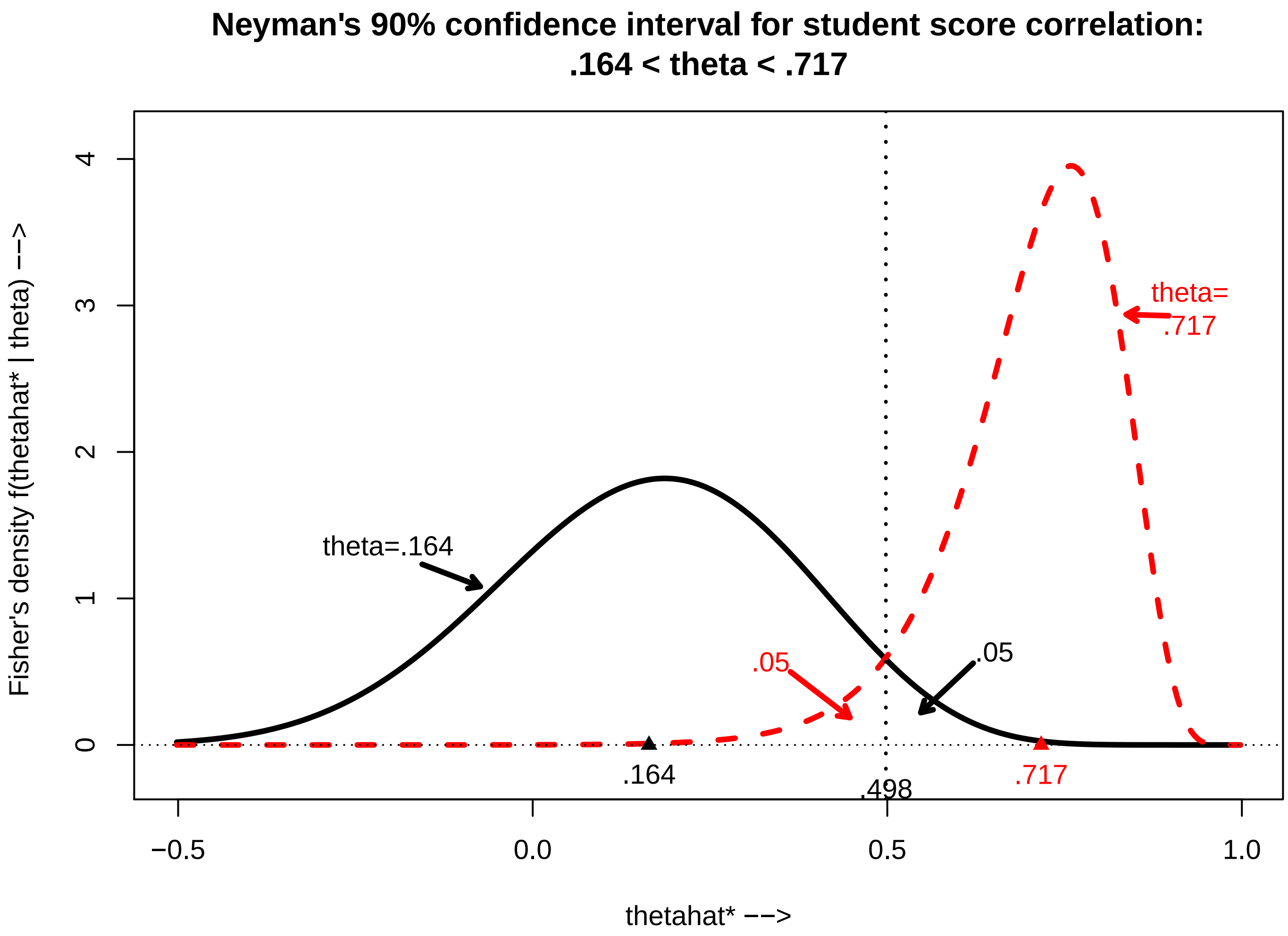
- *1915*: probability density $f_\theta\left(\hat{\theta}\right)$ (hypergeometric series)
- *1922–30*: $\hat{\theta}$ is *maximum likelihood estimate* (MLE)
- Frequentist optimality of MLE minimize expected squared error $E\left\{\left(\hat{\theta}-\theta\right)^2\right\}$
- *Bivariate normal models*

Jerzy Neyman (1930s)

- Optimal frequentist tests and confidence intervals
- 90% confidence interval for θ:

$$\boxed{\theta \in [0.164, 0.717]}$$

- Neyman's construction covers true θ 90% of the time, in repeated use


Neyman's 90% confidence interval for student score correlation:
.164 < theta < .717
Fisher's density f(thetahat* | theta) -->
thetahat* -->
0
1
2
3
4
−0.5
0.0
0.5
1.0
theta=.164
theta=
.717
.05
.05
.164
.498
.717

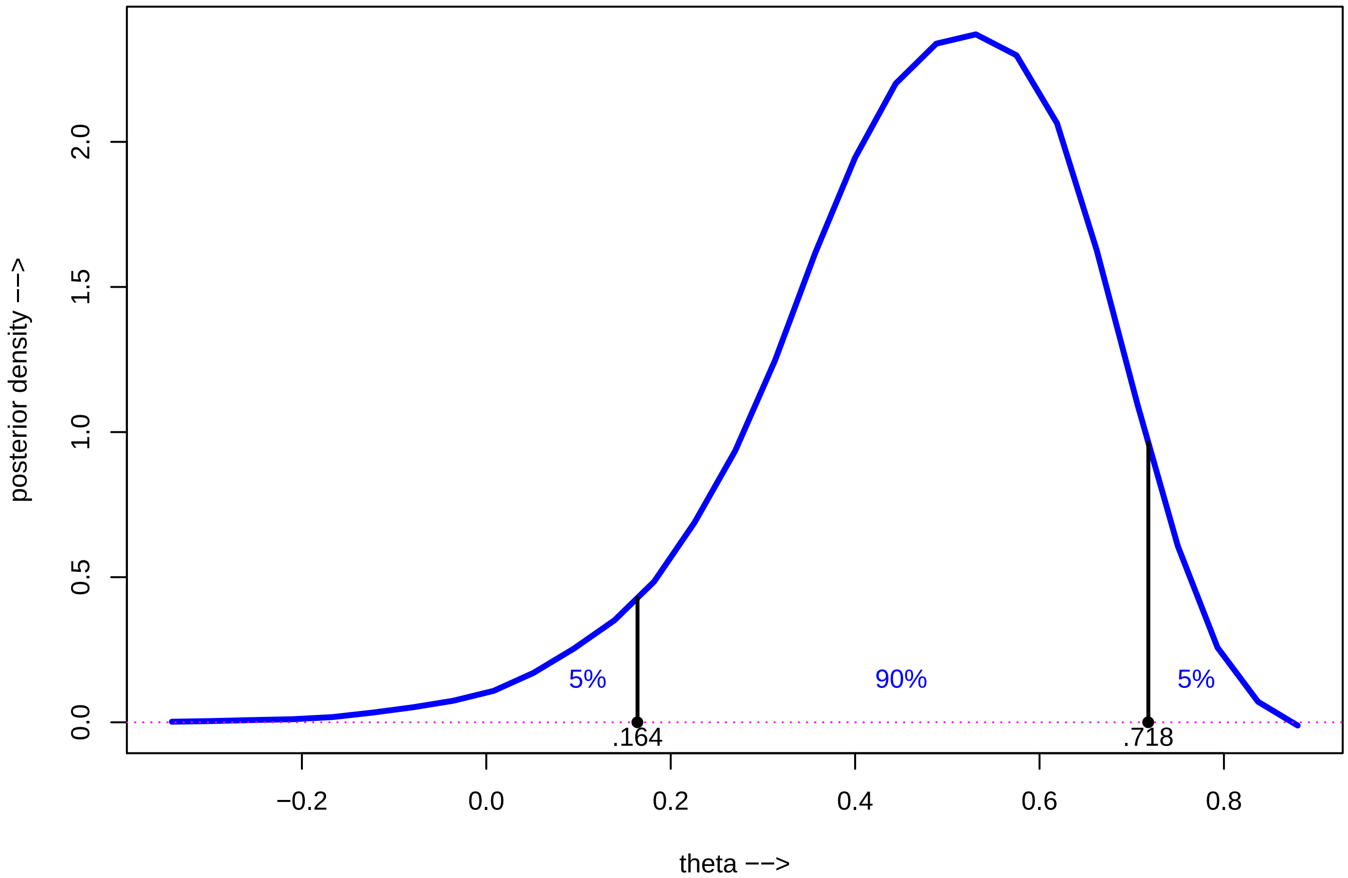
Jeffreys' Invariant Prior

- Jeffreys' objective (or "uninformative") prior for correlation:

$$\boxed{\pi(\theta) = 1/(1-\theta^2)}$$

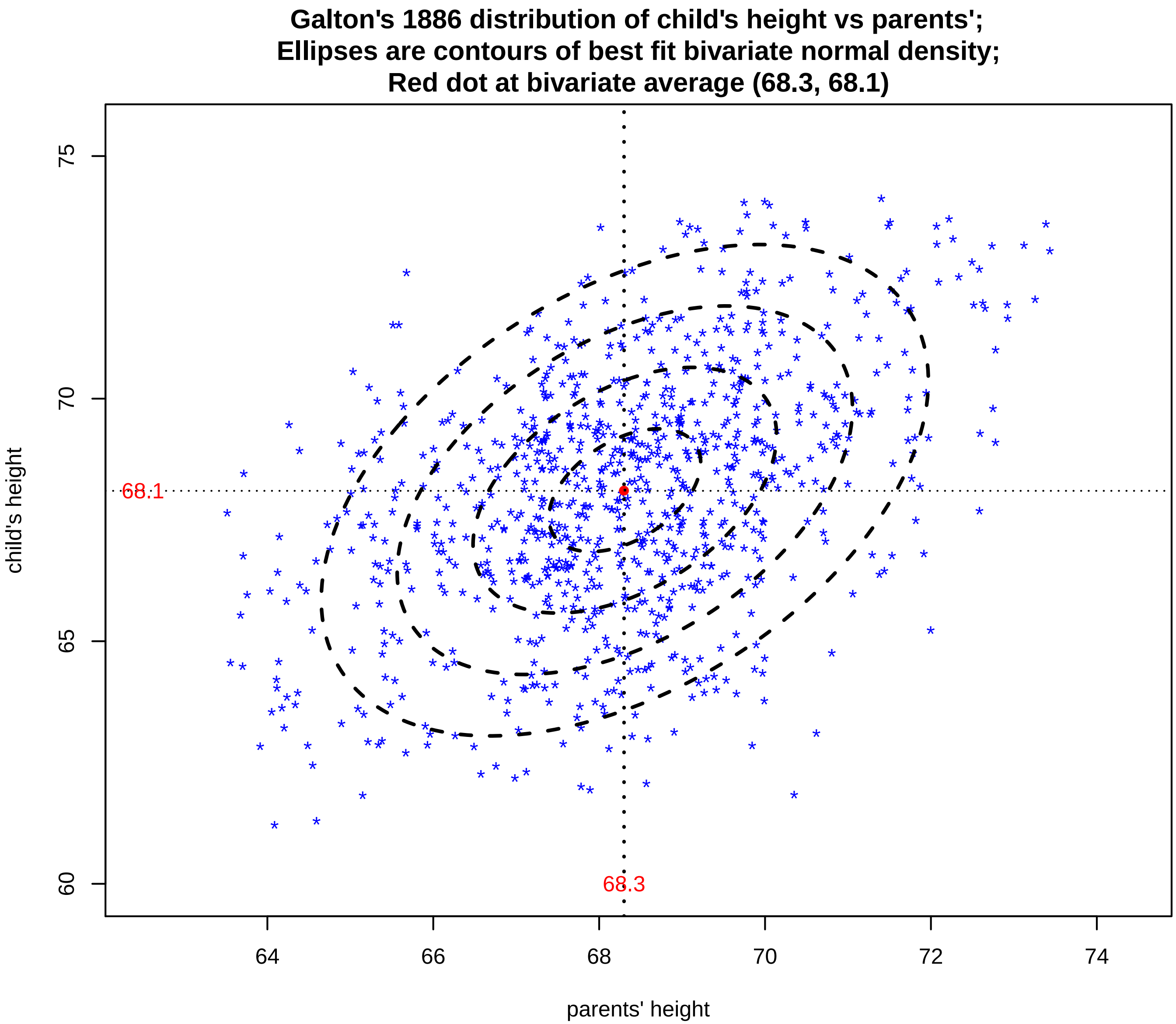
- *General formula* one over square root of Fisher's information bound for the variance of the MLE (transforms correctly under change of variables)

Bayes posterior density pi(theta | thetahat) for the 22 students;
90% Credible Limits = [.164,.718]; Neyman Limits [.164,.717]



More Students

n	$\hat{\theta}$
22	.498
44	.663
66	.621
88	.553
∞	[.415, .662]


Galton's 1886 distribution of child's height vs parents';
Ellipses are contours of best fit bivariate normal density;
Red dot at bivariate average (68.3, 68.1)
child's height
parents' height
60
65
70
75
64
66
68
70
72
74
68.1
68.3

Bivariate Normal Distribution

- "$y \sim \mathcal{N}_2(\mu, \Sigma)$" $\quad (y, \mu \in \mathbb{R}^2,\ \Sigma\ 2 \times 2$ pos def$)$:

$$f_{\mu,\Sigma}(y) = \frac{1}{2\pi}|\Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2}(y-\mu)^t \Sigma^{-1}(y-\mu)}$$

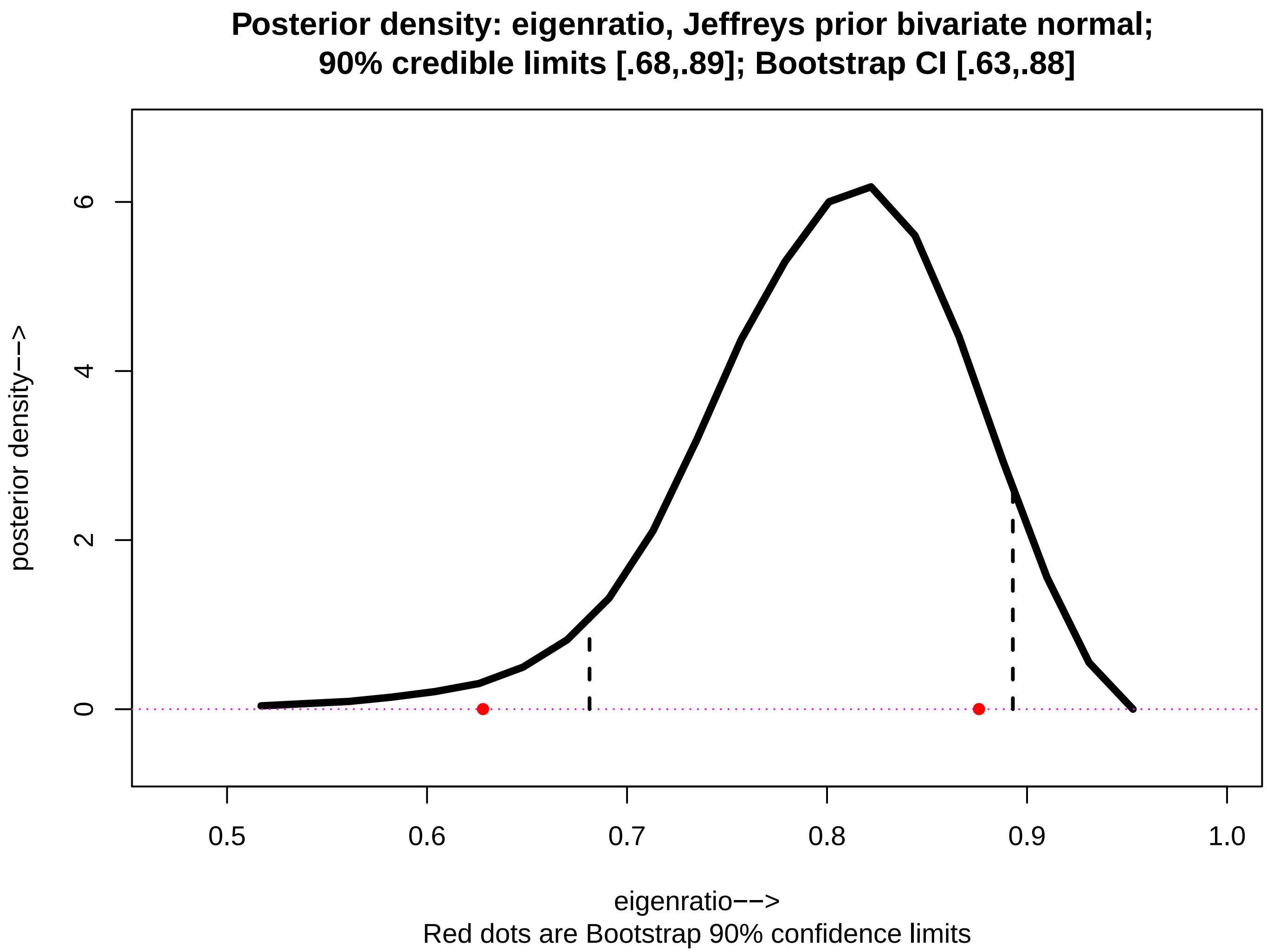
- μ center of ellipse, Σ their shape
- 5 parameters: 2 means, 2 variances, 1 correlation

A More Difficult Problem

- θ = "eigenratio" $= \dfrac{\lambda_1}{\lambda_1 + \lambda_2}$ $\quad (\lambda_1 > \lambda_2$ eigenvalues $\boldsymbol{\Sigma})$
- Student score data $\quad \boldsymbol{y}$ (22×2) gives MLEs $\hat{\mu}$, $\hat{\boldsymbol{\Sigma}}$, and

$$\boxed{\hat{\theta} = 0.793 \pm ?}$$

- Not true: $f_{\mu,\Sigma}\left(\hat{\theta}\right)$ depends only on θ
- There are 4 "*nuisance parameters*"


Posterior density: eigenratio, Jeffreys prior bivariate normal;
90% credible limits [.68,.89]; Bootstrap CI [.63,.88]
posterior density-->
0
2
4
6
0.5
0.6
0.7
0.8
0.9
1.0
eigenratio-->
Red dots are Bootstrap 90% confidence limits

Bootstrap Methods (Automatic Frequentist Inference)

- *Original data* $y_i \sim \mathcal{N}_2(\mu, \Sigma), \quad i = 1, 2, \ldots, 22$
  - gives MLEs $\hat{\mu}, \hat{\Sigma}$, and $\hat{\theta} = 0.793$
- *Bootstrap data* $y_i^* \sim \mathcal{N}_2(\hat{\mu}, \hat{\Sigma}), \quad i = 1, 2, \ldots, 22$
  - gives $\hat{\theta}^*$ = bootstrap eigenratio
- 10,000 $\hat{\theta}^*$s
- 58% exceed $\hat{\theta}$ (upward bias)
- *Reweighting formula* puts bigger weights on smaller $\hat{\theta}^*$s
- *Confidence limits* are the weighted bootstrap percentiles


10000 bootstrap eigenratio values from student score data
(bivariate normal model); Red line shows confidence weights
Frequency
0
100
200
300
400
500
600
MLE=.793
0.5
0.6
0.7
0.8
0.9
1.0
bootstrap eigenratios -->
58% of the bootstrap values exceed .793

Gibbs Sampling (Automatic Bayes Inference)

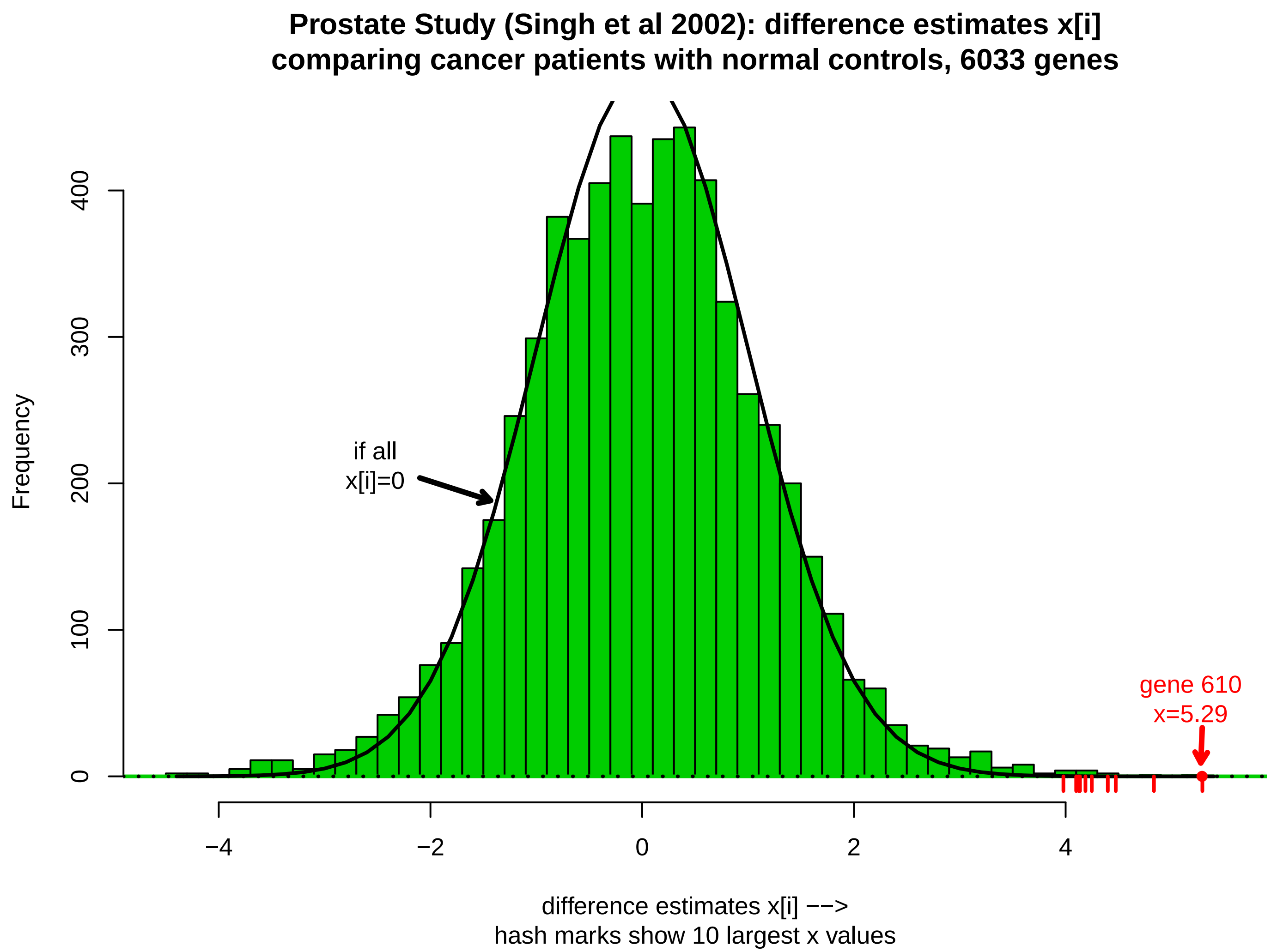
- *Given*: prior $\pi(\theta)$, data x, model $f_\theta(x)$
- *Approximates*: $\pi(\theta|x)$ by Markov chain random walk
- "MCMC", "Metropolis-Hastings", . . . (A-Bomb?)
- Most often used with convenient " uninformative" priors

Prostate Cancer Study
(Singh et al 2002)

- *102 men*: 52 prostate cancer, 50 healthy controls
- Each man assessed for activity of 6033 genes
- *Statistic* x_i measures differences in activity, patients minus controls, for gene$_i$, $i = 1, 2, \ldots, 6033$.
- Probability model

$$\boxed{x_i \sim \mathcal{N}(\delta_i, 1)} \quad \text{(normal, mean } \delta_i\text{, variance 1)}$$

δ_i the true difference or *effect size*


Prostate Study (Singh et al 2002): difference estimates x[i]
comparing cancer patients with normal controls, 6033 genes
Frequency
0
100
200
300
400
if all
x[i]=0
gene 610
x=5.29
−4
−2
0
2
4
difference estimates x[i] −−>
hash marks show 10 largest x values

Bayesian Analysis (for one gene)

- Assume δ has prior density $\pi(\delta)$
- *Prob model* $f_\delta(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\delta)^2}$
- Marginal density $m(x) = \displaystyle\int_{-\infty}^{\infty} f_\delta(x)\pi(\delta)\, d\delta$

  (overall density of x taking account of randomness in δ)
- *Bayes posterior expectation* ("Tweedie's formula")

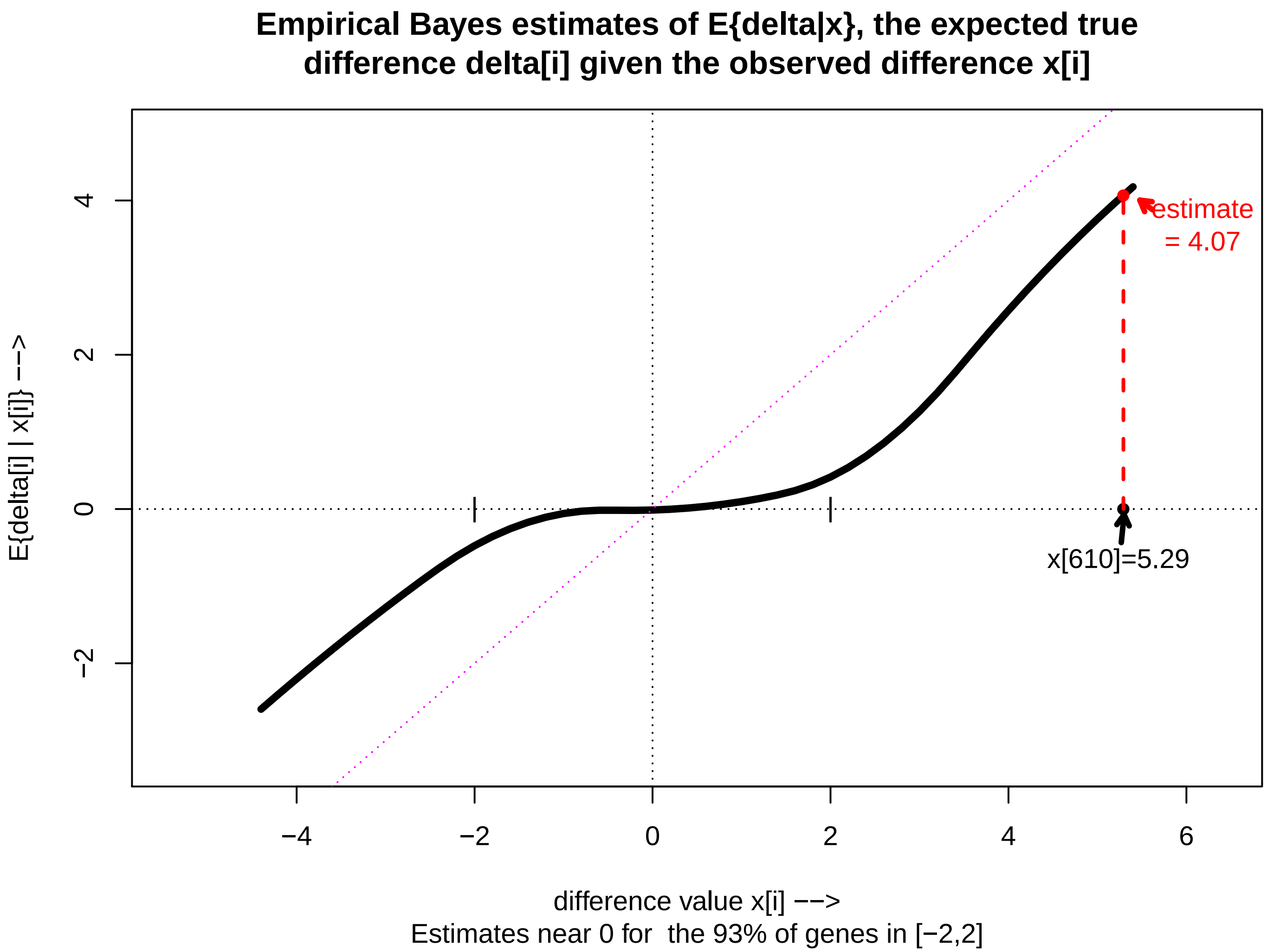
$$\boxed{E\{\delta|x\} = x + \frac{d}{dx}\log m(x)}$$

Empirical Bayes Analysis

- We don't know prior $\pi(\delta)$, but histogram provides a smooth estimate $\hat{m}(x)$ for $m(x)$
- *Empirical Bayes estimate*:

$$\hat{E}\{\delta_i|x_i\} = x_i + \frac{d}{dx}\log \hat{m}(x)\Big|_{x_i}$$

- Frequentist estimation of a Bayesian inference


Empirical Bayes estimates of E{delta|x}, the expected true
difference delta[i] given the observed difference x[i]
E{delta[i] | x[i]} −−>
estimate
= 4.07
x[610]=5.29
−4
−2
0
2
4
6
difference value x[i] −−>
Estimates near 0 for the 93% of genes in [−2,2]

Score Sheet

Bayes	Frequentist
1. Belief (prior)	1. Behavior (method)
2. Principled	2. Opportunistic
3. One distribution	3. Many distributions (bootstrap?)
4. Dynamic	4. Static
5. Individual (subjective)	5. Community (objective)
6. Aggressive	6. Defensive