Bootstrap Confidence Intervals

Bradley Efron

Efron and Tibshirani "Introduction to the Bootstrap"
Chapman and Hall, Chapters 12–14 and 22

• EXACT CONFIDENCE INTERVALS

Binomial, Poisson, Normal correlation, ratio of normal means

• APPROXIMATE CONFIDENCE INTERVALS

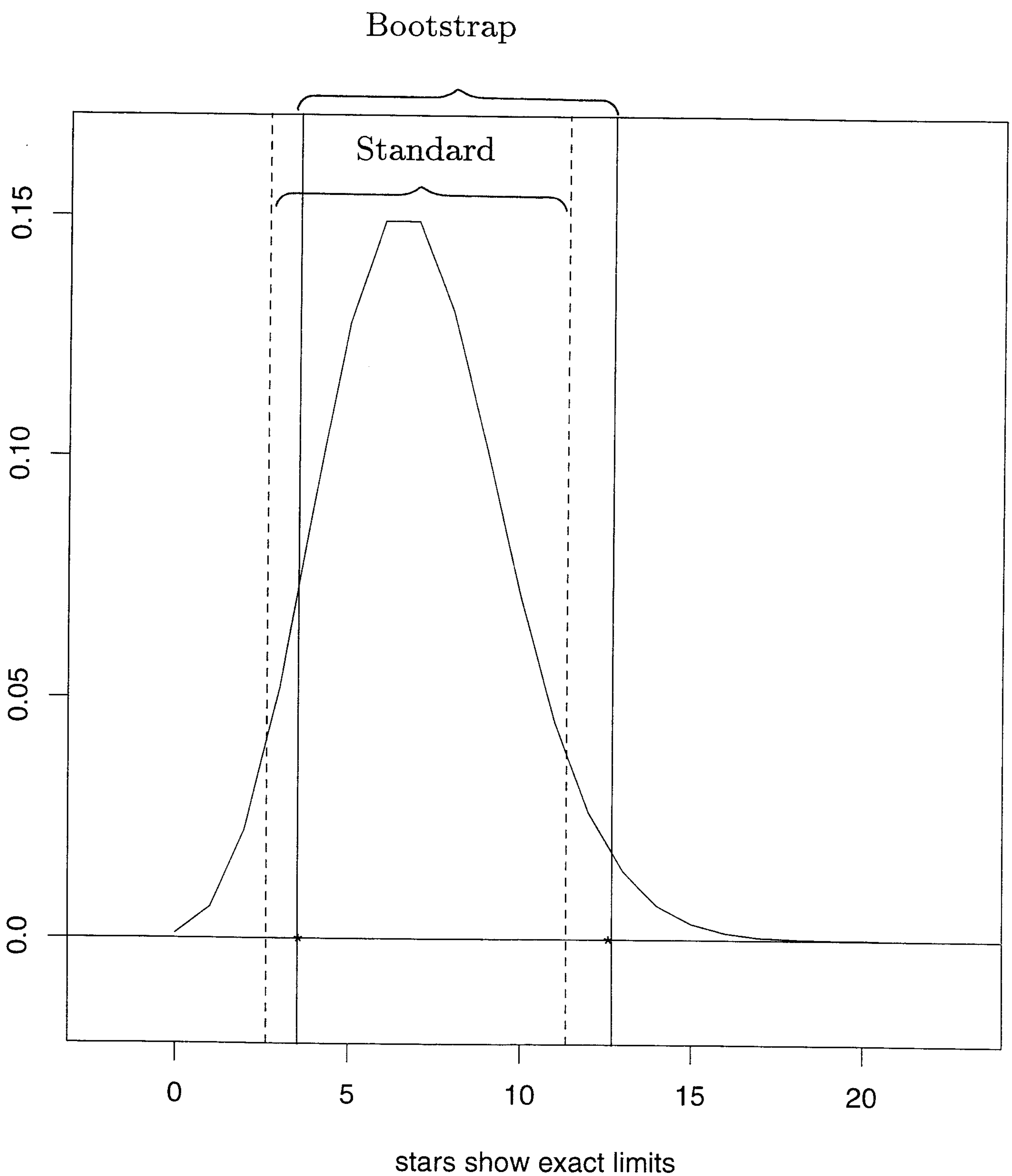
"Standard Interval"

$$\widehat{\theta} \pm 1.645 \; \widehat{\sigma}$$

MLE   For 90% Coverage   Estimated Standard Error

• BOOTSTRAP APPROXIMATE CONFIDENCE INTERVALS

Poisson (.05, .95) Confidence Limits, $X = 7$



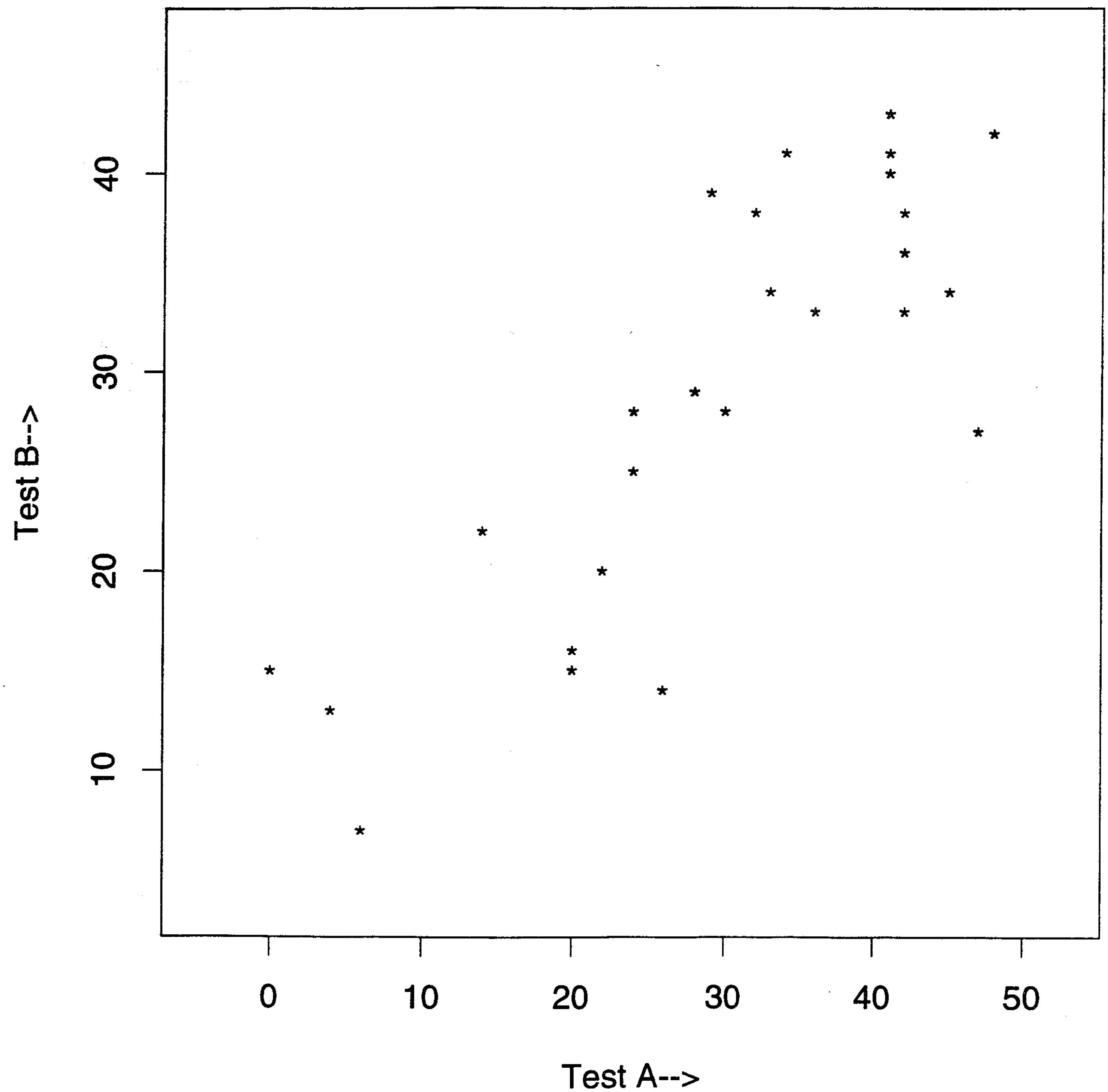
$$\underbrace{\widehat{\theta}_{\mathrm{lo}} \longmapsto \widehat{\theta}}_{L}\ \underbrace{\longmapsto\!\!\longrightarrow \widehat{\theta}_{\mathrm{up}}}_{R}$$

- Length $= L + R$
- Shape $= R/L$

	Length	Shape	Actual Probabilities $\widehat{\theta}_{\mathrm{lo}}$	$\widehat{\theta}_{\mathrm{up}}$
Standard:	8.7	1.00	.012	.091
Bootstrap:	9.1	1.64	.048	.047
Exact:	9.0	1.63	.05	.05

$\mathrm{Prob}_{\hat{\theta}_{\mathrm{lo}}}\{X \geq 7\}$ $\qquad$ $\mathrm{Prob}_{\hat{\theta}_{\mathrm{up}}}\{X \leq 7\}$

Spatial Test Data,
n=26 children



A	48	36	20	29	42	42	20	42	22	41	45	14	6	0	33	28	34	4	32	24	47	41	24	26	30	41
B	42	33	16	39	38	36	15	33	20	43	34	22	7	15	34	29	41	13	38	25	27	41	28	14	28	40

CORRELATION COEFFICIENT

$$\widehat{\theta} = \text{Sample correlation between A and B}$$
$$= .821$$

- Normal Theory Confidence Limits

	$\widehat{\theta}_{\text{lo}}$	$\widehat{\theta}_{\text{up}}$	Length	Shape
Standard:	.716	.926	.21	1.00
Bootstrap:	.668	.901	.23	.52
Exact:	.665	.902	.24	.52

- Actual probabilities for Standard Endpoints:

$\text{Prob}_{.716}\{\widehat{\theta} > .821\} = .090$ $\text{Prob}_{.926}\{\widehat{\theta} < .821\} = .012$

One-Sample Problems

- $$F \longrightarrow \mathbf{x} = (x_1, x_2, \cdots, x_n) \longrightarrow \widehat{\theta} = t(\mathbf{x})$$

Unknown probability distribution	data: random sample of size n from F	Statistic of interest

- <u>Spatial Data:</u> $n = 26$, $x_i = (A_i, B_i)$, $\mathbf{x}$ = data matrix (26×2)

$$\widehat{\theta} = \text{sample corr} \quad t(\mathbf{x}) = \frac{\Sigma(A_i - \bar{A})(B_i - \bar{B})}{[\Sigma(A_i - \bar{A})^2 \Sigma(B_i - \bar{B})^2]^{1/2}}$$

- <u>Bootstrap</u>

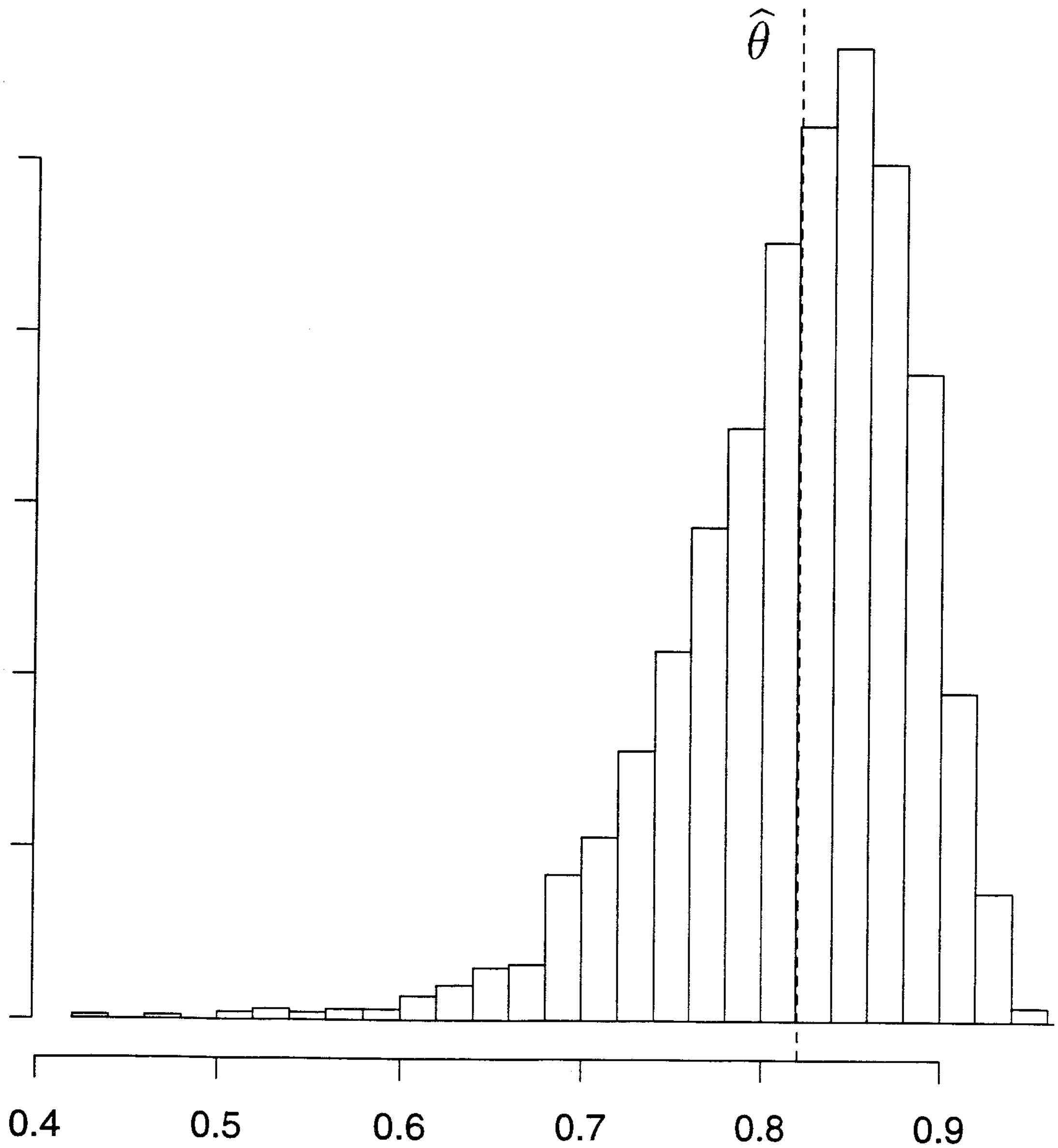
$$\widehat{F} \longrightarrow \mathbf{x}^* = (x_1^*, x_2^*, \cdots, x_n^*) \longrightarrow \widehat{\theta}^* = t(\mathbf{x}^*)$$

estimate of F	bootstrap data: random sample of size n from $\widehat{F}$	bootstrap replication of $\widehat{\theta}$

Nonparametric: $\widehat{F}$ = empirical distribution of the data

($x_1^*, x_2^*, \cdots, x_n^*$ are sampled WITH replacement from $\{x_1, x_2, \cdots, x_n\}$)

2000 nonparametric bootstrap replications
of correlation coeff for Spatial data

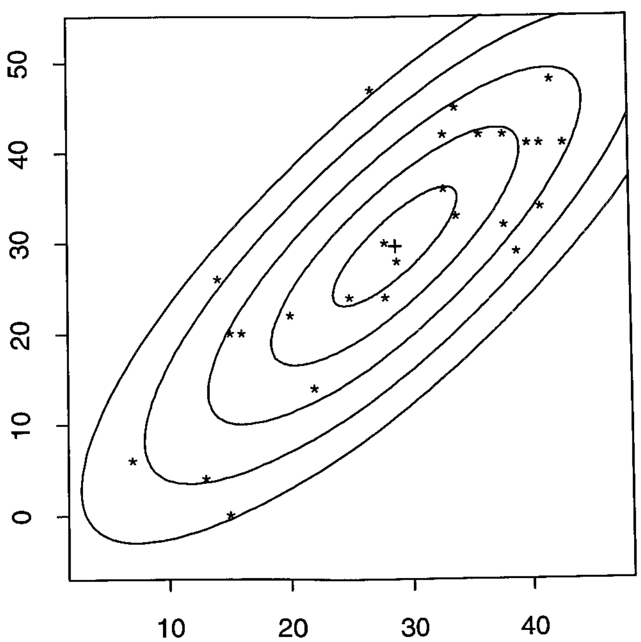


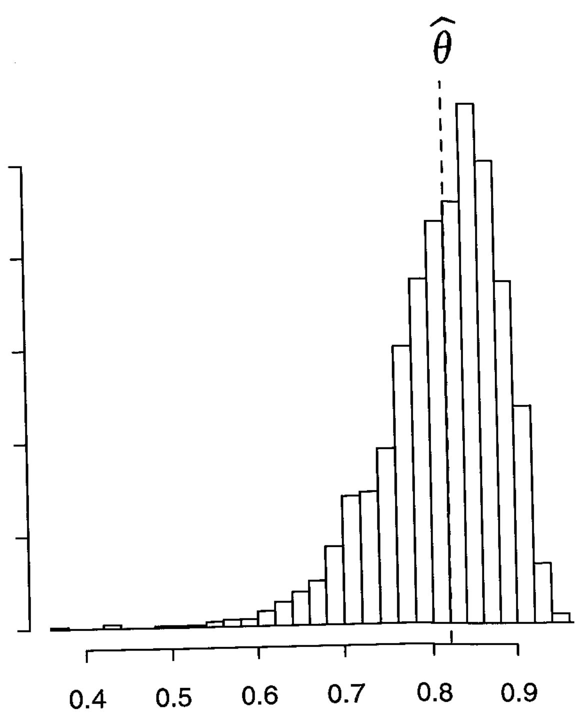
- $\hat{\sigma} = .066$
- 45% of the $\widehat{\theta}^*$ values $< \widehat{\theta}$

NORMAL THEORY BOOTSTRAP

• Take $\widehat{F}$ to be the bivariate normal distribution that best fits the data (MLE).

• $\widehat{F} \to \mathbf{x}^* = (x_1^*, x_2^*, \cdots, x_n^*)$ is random sample of size n from $\widehat{F}$.

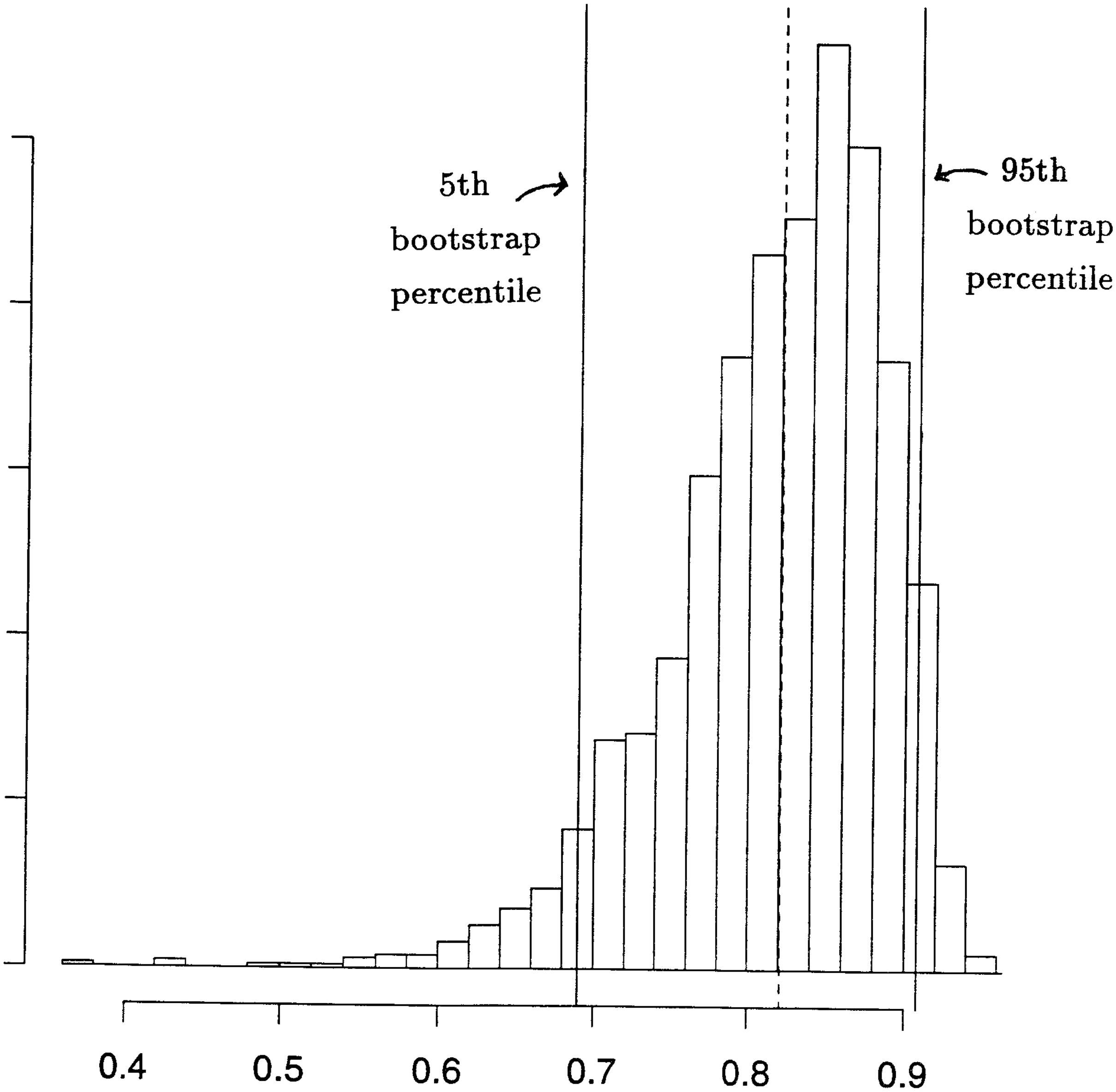




• $\hat{\sigma} = .070$ (Compared to .066 nonparametrically)

• 46% of the 2000 $\widehat{\theta}^*$'s $< \widehat{\theta}$.

2000 normal-theory bootstrap replications of correlation, spatial data



<u>Percentile Interval</u>: $(.690, .908)$

Exact: $(.665, .902)$

BC_a Method

- Instead of .05 and .95 percentiles of the bootstrap distribution, use $\alpha_{.05}$ and $\alpha_{.95}$ percentiles, where

$$\alpha_{.95} = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + 1.645}{1 - \hat{a}(z_0 + 1.645)}\right)$$

- $\Phi(z)$ is standard normal cdf $\int_0^z \exp\{-t^2/2\}dt/\sqrt{2\pi}$

- $\hat{z}_0 =$ "bias-correction" $= \Phi^{-1}$ (proportion of $\widehat{\theta}^*$'s $< \widehat{\theta}$)

$$= \Phi^{-1}\left(\tfrac{914}{2000}\right) = -.108 \text{ for normal bootstraps.}$$

.457

- $\hat{a} =$ "acceleration" $= .000$ for normal bootstraps

($= .035$ nonparametric)

- If $\hat{z}_0 = \hat{a} = 0$ then $BC_a =$ percentile method

- If also bootstrap histogram normal, then $BC_a =$ Standard

- For normal theory spatial data $\alpha_{.05} = 032$ $\quad\alpha_{.95} = .924$

Acceleration $\hat{a}$

- Standard interval is based on asymptotic approximation

$$\widehat{\theta} \dot{\sim} N\,(\theta,\, \sigma^2)$$

Normal unbiased Constant Variance

- BC_a allows for non-normal distributions, biased estimates, and non-constant variance.

- "$\hat{a}$" is a measure of how quickly variance is changing

- Let $\widehat{\theta}_{(i)} = t(\mathbf{x}_{(i)})$ where

  $\mathbf{x}_{(i)} = (x_1, x_2, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n)$

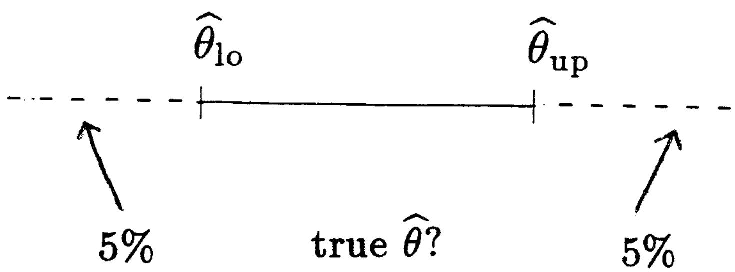
Then

$$\hat{a} = \frac{\Sigma[\widehat{\theta}_{(\cdot)} - \widehat{\theta}(i)]^3}{6[\Sigma(\widehat{\theta}_{(\cdot)} - \widehat{\theta}_{(i)})^2]^{3/2}} \qquad \text{where} \qquad \theta_{(\cdot)} = \frac{\Sigma\widehat{\theta}_i}{n}$$

(nonparametric)

Second-Order Accuracy

• **Each side of exact interval has .05 probability of not covering the true θ**



• Standard interval has non-coverage probabilities

$$.05 + \frac{c}{\sqrt{n}}$$

"first order accurate"

• BC_a interval has non-coverage probabilities

$$.05 + \frac{c}{n}$$

"second order accurate"

(both parametric and nonparametric)

BOOTSTRAP-T Intervals

- Suppose

$$F \to \mathbf{x} \begin{array}{l} \nearrow \widehat{\theta} = t(\mathbf{x}) \quad \text{estimate of} \quad \theta \\ \searrow \hat{\sigma} = s(\mathbf{x}) \quad \text{estimate of} \quad \widehat{\theta} \quad \text{sterr} \end{array}$$

- Define $T = \frac{\widehat{\theta}-\theta}{\hat{\sigma}}$

- Let $T^{(.05)}$ and $T^{(.95)}$ be percentiles of T

- Then $(.05, .95)$ confidence interval for θ is $(\widehat{\theta}_{\text{lo}}, \widehat{\theta}_{\text{up}})$,

$$\boxed{\widehat{\theta}_{\text{lo}} = \widehat{\theta} - T^{(.95)}\hat{\sigma} \qquad\qquad \widehat{\theta}_{\text{up}} = \widehat{\theta} - T^{(.05)}\hat{\sigma}}$$

- If $\widehat{\theta} = \bar{x}$, $\hat{\sigma} = [\Sigma(x_i - \bar{x})^2/n(n-1)]^{\frac{1}{2}}$ then get Student's t.

Bootstrap-T

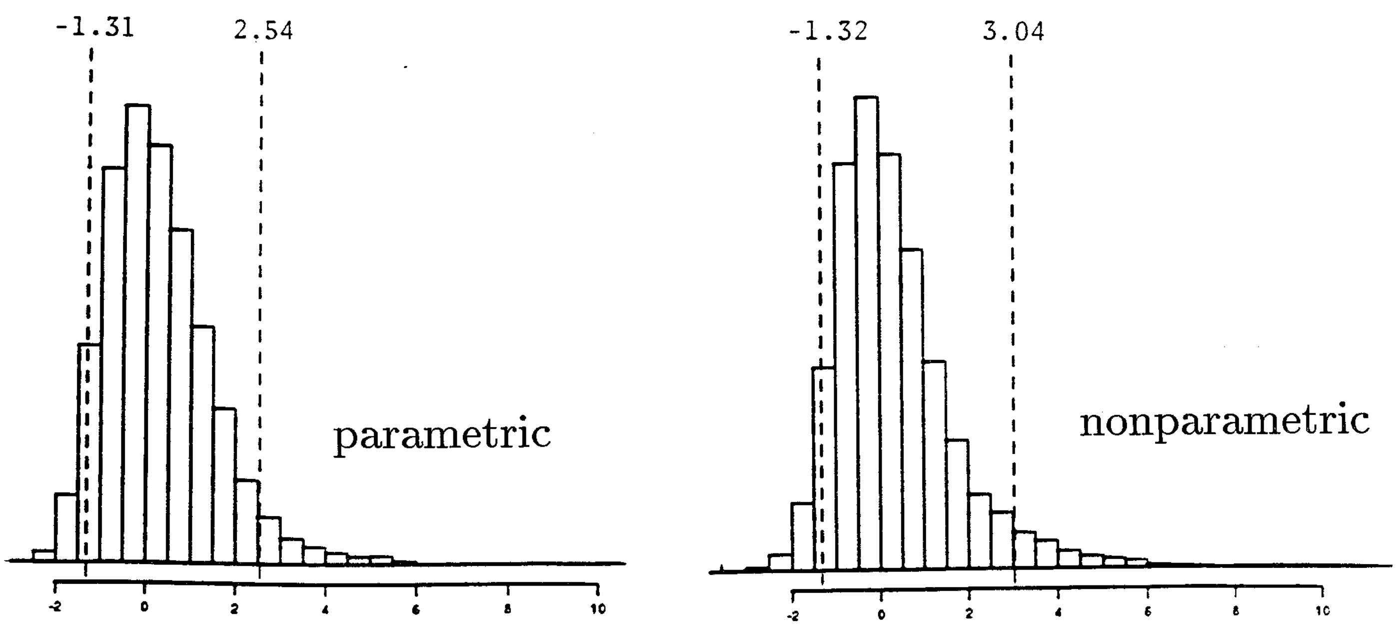
$$\widehat{F} \longrightarrow \mathbf{x}^* \begin{array}{l} \nearrow \ \widehat{\theta}^* = t(\mathbf{x}^*) \ \searrow \\ \searrow \ \hat{\sigma}^* = s(\mathbf{x}^*) \ \nearrow \end{array} \quad T^* = \frac{\widehat{\theta}^* - \widehat{\theta}}{\hat{\sigma}^*}$$

- Distribution of T^* values gives bootstrap percentiles $T^{*(.05)}$, $T^{*(.95)}$

- Use

$$\boxed{\widehat{\theta}_{\text{lo}} = \widehat{\theta} - T^{*(.95)}\hat{\sigma} \qquad\qquad \widehat{\theta}_{\text{up}} = \widehat{\theta} - T^{*(.05)}\hat{\sigma}}$$

- Spatial Correlation Example, $\hat{\sigma} = (1 - \widehat{\theta^2})/\sqrt{26}$:



	$\widehat{\theta}_{\text{lo}}$	$\widehat{\theta}_{\text{up}}$	$\widehat{\theta}_{\text{lo}}$	$\widehat{\theta}_{\text{up}}$
Boot-T	.653	.905	.627	.905
BC_a(ABC)	.668	.901	.675	.892
Exact	.665	.902	?	?
Standard	.716	.926	.726	.916
	Normal Theory		Nonparametric	

- Boot-T is 2nd order accurate.

- Disadvantages:
  - Need expression for $\hat{\sigma}$ (or 2nd level bootstrap)
  - Not trustworthy in nonparametric settings
  - Not transformation invariant

Transformation Invariance

- Suppose we change parameter of interest from $\theta =$ correlation coefficient to $R = \sqrt{1-\theta^2}$.

- Then BC_a confidence interval changes in the obvious way:

$$\widehat{R}_{\mathrm{lo}} = \sqrt{1-\widehat{\theta}^2_{\mathrm{lo}}} \qquad \widehat{R}_{\mathrm{up}} = \sqrt{1-\widehat{\theta}^2_{\mathrm{up}}}$$

- Works for any monotone transformation.

- Exact intervals have same property.

- But standard intervals, Boot-T intervals don't.

Student Score Data

- $n = 22$ students have each taken 5 tests, but some of the scores are missing:

Observed Data **o**

student	A	B	C	D	E
1	?	63	65	70	63
2	53	61	72	64	73
3	51	67	65	65	?
4	?	69	53	53	53
5	?	69	61	55	45
6	?	49	62	63	62
7	44	61	52	62	?
8	49	41	61	49	?
9	30	69	50	52	45
10	?	59	51	45	51
11	?	40	56	54	?
12	42	60	54	49	?
13	?	63	53	54	?
14	?	55	59	53	?
15	?	49	45	48	?
16	17	53	57	43	51
17	39	46	46	32	?
18	48	38	41	44	33
19	46	40	47	29	?
20	30	34	43	46	18
21	?	30	32	35	21
22	?	26	15	20	?

- Parameter of interest: $\theta =$ maximum eigenvalue of covariance matrix of the 5 scores.

Estimate θ by $\widehat{\theta}$:

(a) Impute missing values in 22×5 data matrix $\mathbf{x}$ by two-way additive model $\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j$

fit from non-missing values

(b) Compute usual sample covariance matrix for the imputed data matrix.

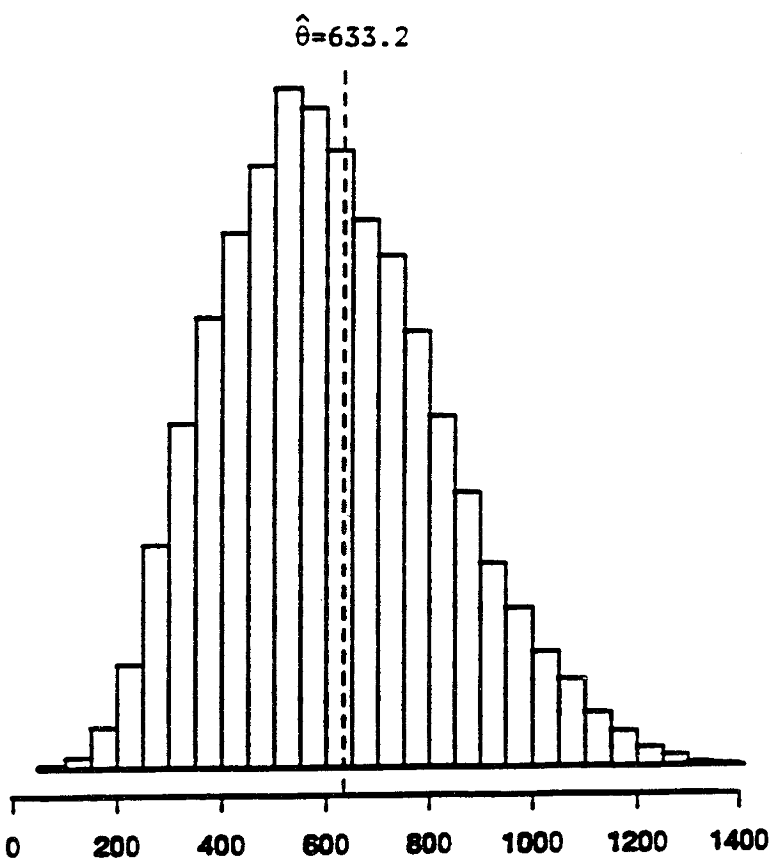
(c) Take its maximum eigenvalue.

- $\widehat{\theta} = 633.2 \pm ?$

• **Nonparametric bootstrap analysis of $\widehat{\theta}$**

• Bootstrap data matrix $\mathbf{x}^*$ is obtained by resampling the rows of $\mathbf{x}$ (including the question marks).

• Then $\widehat{\theta}^*$ is obtained from $\mathbf{x}^*$ in same way that $\widehat{\theta}$ was obtained from $\mathbf{x}$.



- 2200 bootstraps of $\widehat{\theta}$ gave

$$\hat{\sigma} = 212.0$$

- $(.05, .95)$ Confidence Limits for θ:

	$\widehat{\theta}_{\text{lo}}$	$\widehat{\theta}_{\text{up}}$
BC_a	379	1164
ABC	379	1172
Standard	284	982

- No gold standard

ABC Method

- Approximates the endpoints of BC_a interval analytically.

- Uses numerical 2nd derivatives in place of Monte Carlo.

- Needs $2n + 4$ recomputations of $\widehat{\theta}$, rather than 2000.

- Works for "smooth" statistics $\widehat{\theta} = s(\mathbf{x})$.

- Standard interval requires $(\widehat{\theta}, \hat{\sigma})$. • ABC also needs

$$(\hat{a}, \hat{z}_0, \hat{c})$$

"nonlinearity"

- Also an approximation to Bootstrap-T

- Need to write $\widehat{\theta} = t(\mathbf{x})$ as function of bootstrap weights on sample points $x_1, x_2, \cdots, x_n$. [e.g. $\bar{x}^* = \sum_i \frac{N_i}{n} x_i$].

Nonparametric ABC Program in "S"

```
"abcnon" <-
function(tt, n, epsi = 0.001, alpha = c(.025,.05,.1,.16,.84,.9,.95,.975))
{
#abc for nonparametric problems,  sample size n
#tt(P)is statistic in resampling form, where P[i] is weight on x[i]
        ep <- epsi/n; I<- diag(n); P0<- rep(1/n,n)
        t0 <- tt(P0)
#calculate t. and t.. .........................................
        t. <- t.. <- numeric(n)
        for(i in 1:n) {  di <- I[i,  ] - P0
                         tp <- tt(P0 + ep * di)
                         tm <- tt(P0 - ep * di)
                         t.[i] <- (tp - tm)/(2 * ep)
                         t..[i] <- (tp - 2 * t0 + tm)/ep^2}
#calculate sighat,a,z0,and cq ....................................
        sighat <- sqrt(sum(t.^2))/n
        a <- (sum(t.^3))/(6 * n^3 * sighat^3)
        delta <- t./(n^2 * sighat)
        cq <- (tt(P0+ep*delta) -2*t0 + tt(P0-ep*delta))/(2*sighat*ep^2)
        bhat <- sum(t..)/(2 * n^2)
        curv <- bhat/sighat - cq
        z0 <- qnorm(2 * pnorm(a) * pnorm( - curv))
#calculate interval endpoints.....................................
        Z <- z0 + qnorm(alpha)
        za <- Z/(1 - a * Z)^2
        stan <- t0 + sighat * qnorm(alpha)
        abc <- seq(alpha)
        for(i in seq(alpha)) abc[i] <- tt(P0 + za[i] * delta)
        lims <- cbind(alpha, abc, stan)
#output in list form..............................................
        list(lims=lims, stats=c(t0,sighat,bhat), cons=(c(a,z0,cq)), t.=t.)
}
```

- $n = 8$ subjects each given 3 hormone patches: Placebo, Approved, New.

- Blood levels of hormone:

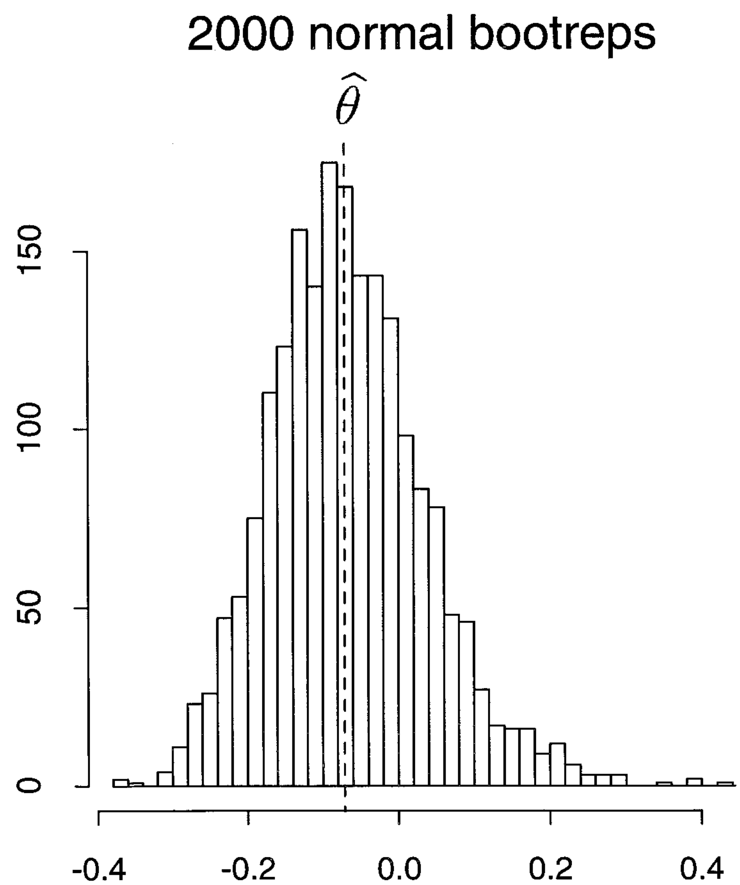
				y	z
Patient	Placebo	Approved	New	App-Pla	New-App.
1.	9243	17649	16449	8406	-1200
2.	9671	12013	14614	2342	2601
3.	11792	19979	17274	8187	-2705
4.	13357	21816	23798	8459	1982
5.	9055	13850	12560	4795	-1290
6.	6290	9806	10157	3516	351
7.	12412	17208	16570	4796	-638
8.	18806	29044	26325	10238	-2719
mean	11328	17671	17218	6342	-452

- $y =$ Approved-Placebo $\qquad z =$ New-Approved

- Parameter of interest $\theta = E\{z\}/E\{y\}$

- $\widehat{\theta} = \frac{-452}{6342} = -.071 \pm ?$

• **Normal Theory:** Assume $x_i = (y_i, z_i)$ bivariate normal vectors.



- $\hat{\sigma} = .103.$

- 51.1% of $\widehat{\theta}^*$ values $< \widehat{\theta}$

Confidence Limits for θ

	$\widehat{\theta}_{\text{lo}}$	$\widehat{\theta}_{\text{up}}$	Length	Shape
Exact	-.249	.170	.42	1.36
BC_a	-.212	.115	.33	1.32
ABC	-.215	.111	.33	1.27
ABC_{CAL}	-.257	.175	.43	1.33
Standard	-.232	.089	.32	1.00

- "Third-order errors" in length of BC_a, ABC. $(O(n^{-\frac{3}{2}}))$

CALIBRATION

• Let $\widehat{\theta}[\alpha]$ be endpoint of level-α approximate confidence interval

$$\widehat{\theta}[.05] = \widehat{\theta}_{\text{lo}}, \ \widehat{\theta}[.95] = \widehat{\theta}_{\text{up}}$$

• $$\underbrace{\beta(\alpha)}_{\substack{\text{true coverage}\\ \text{probability}}} = \text{Prob}_F\{\theta < \underbrace{\widehat{\theta}[\alpha]}_{\substack{\text{nominal coverage}\\ \text{probability}}}\}$$

• If we knew for example that $\beta = .95$ corresponded to $\alpha = .98$ then we could set

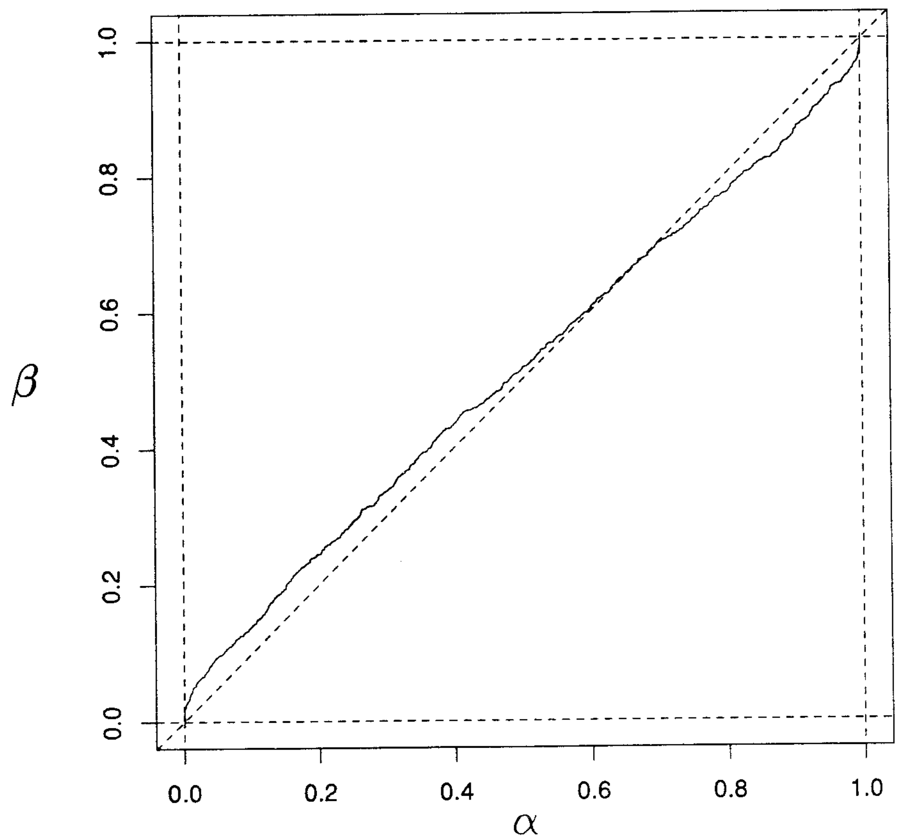
$$\widehat{\theta}_{\text{up}} = \widehat{\theta}[.98].$$

• Bootstrap Calibration

$$\hat{\beta}(\alpha) = \text{Prob}_{\hat{F}}\{\widehat{\theta} < \widehat{\theta}[\alpha]^*\}$$

(Proportion that bootstrap limit $\widehat{\theta}[\alpha^*]$ exceeds $\widehat{\theta}$)

Normal Theory Bootstrap Calibration of ABC



- Says that

$$.05 = \hat{\beta}[.0137] \qquad .95 = \hat{\beta}[.9834]$$

- $\mathrm{ABC}_{\mathrm{CAL}} = (\widehat{\theta}_{\mathrm{ABC}}[.0137],\ \widehat{\theta}_{\mathrm{ABC}}[.9834])$