

THE AUTOMATIC CONSTRUCTION OF BOOTSTRAP CONFIDENCE INTERVALS

Bradley Efron

Stanford University

THE STANDARD INTERVALS

$$\boxed{\hat{\theta} \pm z^{(\alpha)} \hat{\sigma}} \quad (z^{(0.975)} = 1.96)$$

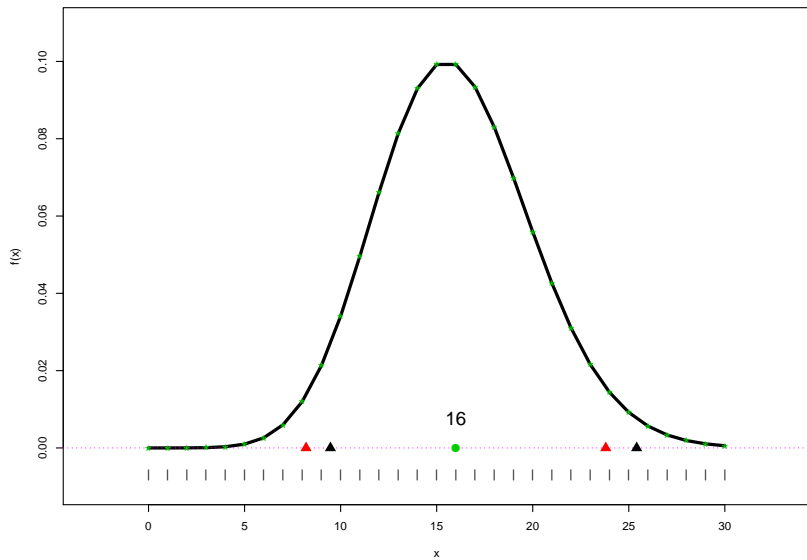
- $\hat{\theta}$ a point estimate of θ
- $\hat{\sigma}$ an estimate of its standard error
(Taylor series, jackknife, bootstrap)
- Automatic!

POISSON EXAMPLE

- $x \sim \text{Poisson}(\theta)$
- Observe $x = 16$
- 95% central intervals:

	Lo	Up	Right/Left
<i>Standard</i>	8.16	23.84	1
<i>Exact</i>	9.47	25.41	1.44
<i>Bootstrap</i>	9.42	25.53	1.45

Poisson example $x \sim \text{Poisson}(\theta)$, observe $x=16$;
exact (black) and standard (red) 95% endpoints



STUDENT- t CORRECTIONS

- $x_i \stackrel{\text{ind}}{\sim} \mathcal{N}(\theta, \sigma^2) \quad i = 1, 2, \dots, n$

- $\hat{\theta} = \bar{x}, \quad \hat{\sigma}^2 = \sum (x_i - \bar{x})^2 / (n - 1)$

$$\theta \in \hat{\theta} \pm t_{n-1}^{(\alpha)} \hat{\sigma}$$

- $t_{15}^{(0.975)} = 2.13 \quad \text{cf } z^{(0.975)} = 1.96$

- t correction is order $O(1/n)$

BOOTSTRAP CONFIDENCE INTERVALS

- `bca` algorithm makes 3 corrections to standard endpoints
- Corrections of order $O(1/\sqrt{n})$
- *Nonparametric* `bcajack` (automatic)
- *Parametric* `bcapar` (almost automatic)

THE DIABETES DATA

- $n = 442$ patients
- $x_i = (p_i, y_i)$ $\left\{ \begin{array}{l} p_i \text{ vector of 10 baseline predictors} \\ y_i \text{ measure of disease progression, 1 year} \end{array} \right.$
- \mathbf{x} a 442×11 matrix
- $\hat{\theta} =$ adjusted R^2 of y regressed on p
- $\mathbf{x} = (\text{lm}(y \sim p)\$ \text{adj.r.squared} = 0.507)$

DIABETES DATA

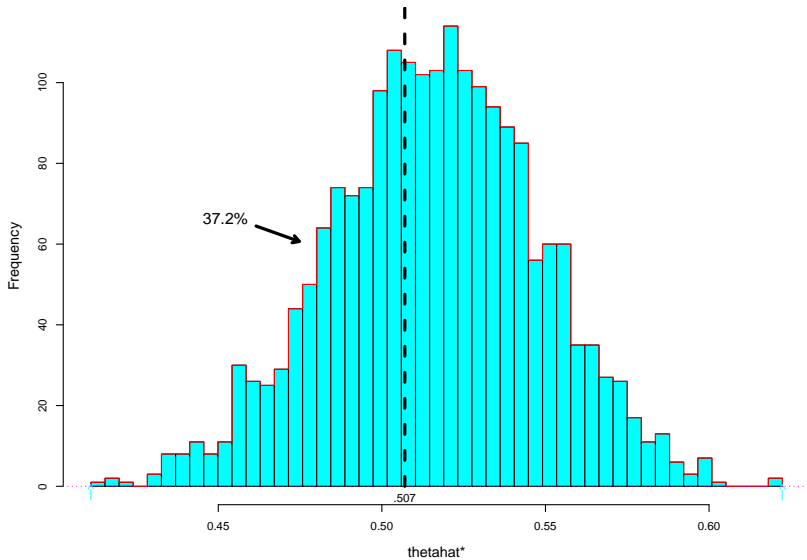
$n = 442$ SUBJECTS, 10 PREDICTORS, RESPONSE y "PROGRESSION".

age	sex	bmi	map	tc	ldl	hdl	tch	ltg	glu	y
.04	.05	.06	.02	-.04	-.03	-.04	.00	.02	-.02	-1.13
.00	-.04	-.05	-.03	-.01	-.02	.07	-.04	-.07	-.09	-77.13
.09	.05	.04	-.01	-.05	-.03	-.03	.00	.00	-.03	-11.13
-.09	-.04	-.01	-.04	.01	.02	-.04	.03	.02	-.01	53.87
.01	-.04	-.04	.02	.00	.02	.01	.00	-.03	-.05	-17.13
-.09	-.04	-.04	-.02	-.07	-.08	.04	-.08	-.04	-.10	-55.13
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

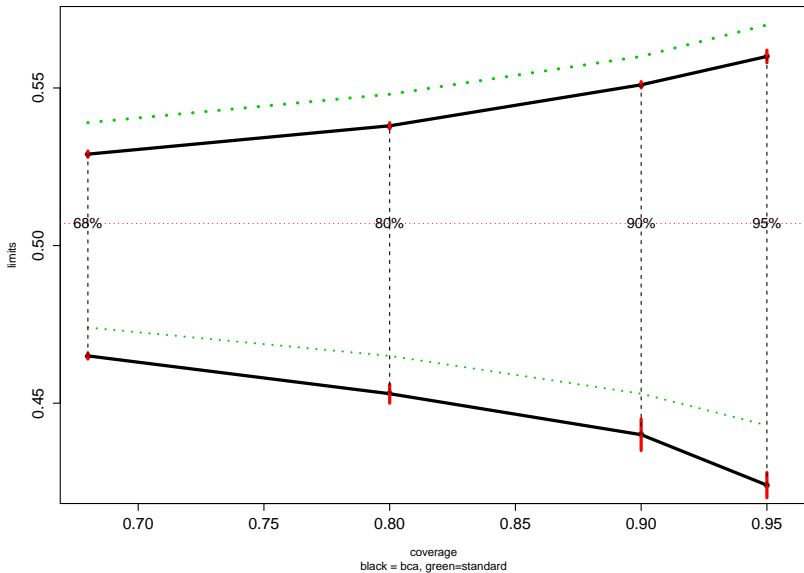
NONPARAMETRIC BOOTSTRAPPING

- Data matrix \mathbf{x} : rows are p -vectors
 $n \times p$
- $\hat{\theta} = t(\mathbf{x})$ estimates parameter of interest θ
- Bootstrap data matrix \mathbf{x}^* : randomly choose n rows from \mathbf{x}
with replacement
- Gives $\hat{\theta}^* = t(\mathbf{x}^*)$
- $\{\hat{\theta}^*(1), \hat{\theta}^*(2), \dots, \hat{\theta}^*(B)\}$ provides inferences for θ
($\hat{\sigma}_{\text{boot}}$ = standard deviation of $\hat{\theta}^*$)

B=2000 nonparametric bootstrap reps for adj.r.squared;
Only 37.2% of the boots are less than .507



Bca and standard two-sided intervals: adjusted R2 statistic
(from bcajack; red bars indicate monte carlo error)



CORRECTING THE STANDARD INTERVALS

- Standard CIs assume $\hat{\theta} \sim \mathcal{N}(\theta, \sigma^2)$, σ^2 constant
- bca makes 3 corrections:
 - 1 for non-normality: using boot cdf \widehat{G}
 - 2 for bias: using $\hat{z}_0 = \Phi^{-1}\widehat{G}(\hat{\theta})$
(diabetes: $\hat{z}_0 = \Phi^{-1}(0.372) = -0.327$)
 - 3 for “acceleration”: or nonconstant σ^2 , using jackknife

THE BCA LEVEL α ENDPOINT

$$\hat{\theta}_{\text{bca}}(\alpha) = \widehat{G}^{-1} \Phi \left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(\alpha)}}{1 - \hat{a}(\hat{z}_0 + z^{(\alpha)})} \right)$$

- \widehat{G} bootstrap cdf
- Φ standard normal cdf
- $\hat{z}_0 = \Phi^{-1} \widehat{G}(\hat{\theta})$
- \hat{a} acceleration estimate
- **Second-order accuracy** Claimed coverage accurate to $O(1/n)$, compared to $O(1/\sqrt{n})$ for standard CIs



JACKKNIFE CALCULATIONS

- $\mathbf{x}_{(i)}$ is $(n - 1) \times p$ matrix having x_i removed from \mathbf{x}
- $\hat{\theta}_{(i)} = t(\mathbf{x}_{(i)})$
- $\hat{\theta}_{(\cdot)} = \sum_1^n \hat{\theta}_{(i)} / n$

$$D_i = \hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)}$$

- Jackknife standard error

$$\hat{\sigma}_{\text{jack}} = \left(\frac{n-1}{n} \sum_1^n D_i^2 \right)^{1/2}$$

- Acceleration estimate

$$\hat{a} = \frac{1}{6} \sum_1^n D_i^3 / \left(\sum_1^n D_i^2 \right)^{3/2}$$



GROUPED JACKKNIFE

- n can be enormous these days
- partition \mathbf{x} into m groups of size $g = n/m$:

$$X_k = \{x_{i_1}, x_{i_2}, \dots, x_{i_g}\} \quad (k = 1, 2, \dots, m)$$

- Now $\mathbf{X} = (X_1, X_2, \dots, X_m)$ has just m “points”
- **Diabetes** $m = 40, g = 11$ (2 left over)

```
bcajack(x, 2000, rfun, m = 40)
```

PARTIAL OUTPUT OF BCAJACK(x , 2000, RFUN, $M=40$)

$X =$ DIABETES DATA, RFUN = ADJUSTED R^2

α	bcalims	jacksd	standard
.025	.424	.004	.443
.16	.465	.001	.474
.5	.497	.001	.507
.84	.529	.001	.539
.975	.560	.002	.570

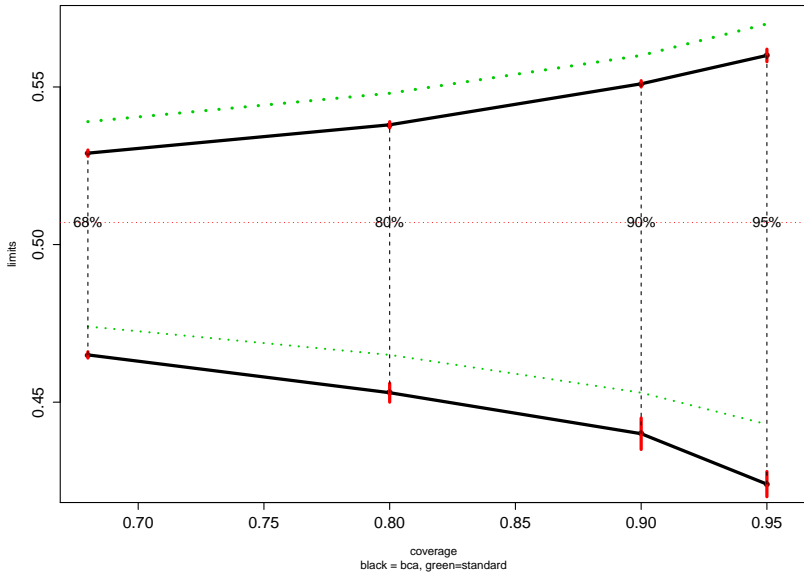
Monte Carlo Error

- *Two types of error* “sampling” and “Monte Carlo”
- $\hat{\sigma}_{\text{boot}}$ concerns sampling error of $\hat{\theta}$
- Was $B = 2000$ enough bootstrap replications?
- **Answer** Jackknife the bootstrap replications.

JACKKNIFE ESTIMATES OF MONTE CARLO ERROR

- Randomly partition bootstrap replications $\{\hat{\theta}_{(1)}^*, \hat{\theta}_{(2)}^*, \dots, \hat{\theta}_{(2000)}^*\}$ into 10 groups of 200 each
- Remove one group at a time, and rerun `bcajack`
- Use jackknife formula to get “jacksd”
- No new bootstrapping required

Monte Carlo error is small for upper limits, not so small for lower limits



MORE OUTPUT OF BCAJACK(x , 2000, RFUN, M=40)

	θ	sdboot	z_0	a	sdjack
estimate	.507	.033	-.327	-.004	.034
jacksd	.000	.001	.027	.000	.000

PARAMETRIC EXAMPLE: THE PEDIATRIC DEATH DATA

- 800 cases African facility for very sick babies
- 600 survived, 200 died
- 11 predictors $\mathbf{x} = (\text{respiration, heart rate, weight, } \dots)$
- *Logistic regression model*

π_i death probability ($i = 1, 2, \dots, 800$)

$$y_i = \begin{cases} 1 & \text{probability } \pi_i \\ 0 & \text{probability } 1 - \pi_i \end{cases} \quad \text{where } \pi_i = \frac{1}{1 + e^{-\mathbf{x}'_i \boldsymbol{\alpha}}}$$

$\theta = \text{"resp" coefficient } \alpha_1$

PARAMETRIC BOOTSTRAPPING

■ $\text{glm}(\mathbf{y} \sim_{800 \times 11} \mathbf{X}, \text{binomial})$ gives MLE $\hat{\alpha}$, $\hat{\theta} = \hat{\alpha}_1 = 0.943$

■ $\hat{\pi}_i = 1 / (1 + e^{-x_i' \hat{\alpha}})$, $i = 1, 2, \dots, 800$

■ *Bootstrap sample* $y_i^* = \begin{cases} 1 & \text{prob } \hat{\pi}_i \\ 0 & \text{prob } 1 - \hat{\pi}_i \end{cases}$

$\text{glm}(\mathbf{y}^* \sim \mathbf{X}, \text{binomial})$ gives $\hat{\alpha}^*$ and $\hat{\theta}^* = \hat{\alpha}_1^*$

■ Also need $b^* = \mathbf{X}' \mathbf{y}^*$ for bca calculations

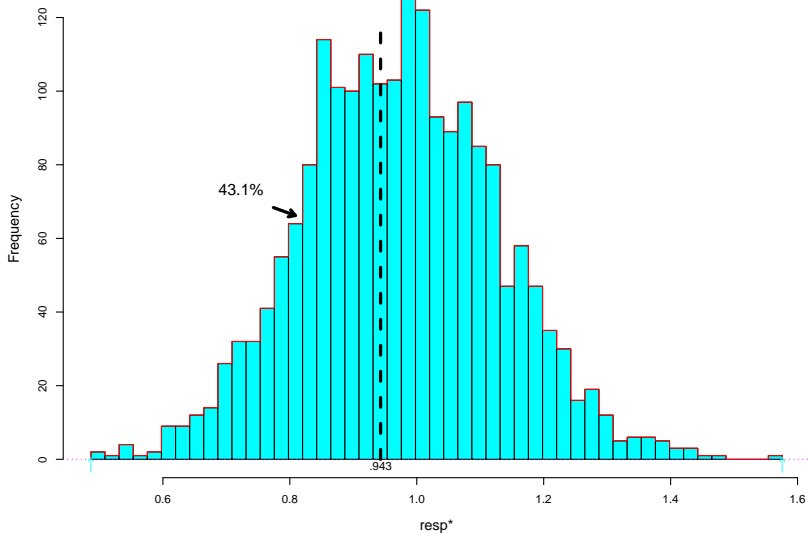
PROGRAM BCAPAR

- B bootstrap samples give

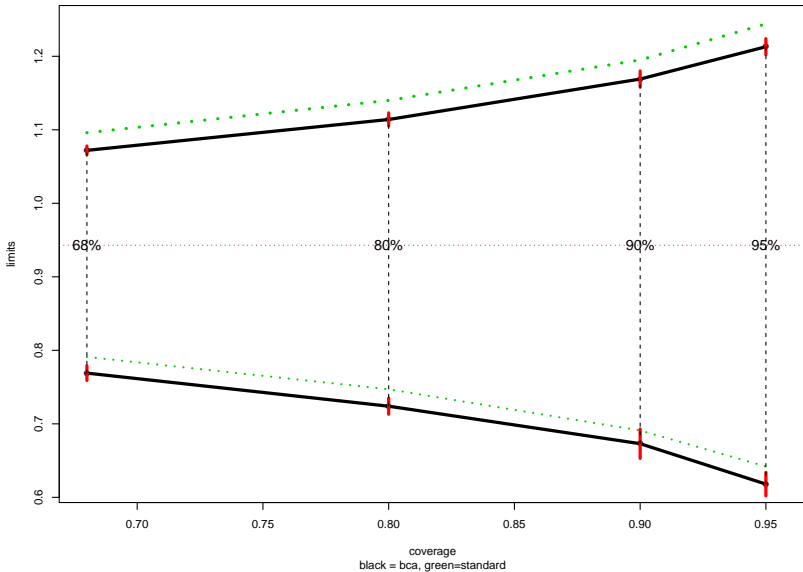
$$\hat{\theta}^* = (\hat{\theta}^*(1), \hat{\theta}^*(2), \dots, \hat{\theta}^*(B)) \quad \text{and} \quad \mathbf{b}^*_{1 \times B} = (b^*(1), b^*(2), \dots, b^*(B))$$

- $\text{bcapar}(\hat{\theta}, \hat{\theta}^*, \mathbf{b}^*)$ yields bca limits
- $\hat{\theta}$ and $\hat{\theta}^*$ give \widehat{G} and \hat{z}_0
- \mathbf{b}^* gives acceleration \hat{a}

2000 parametric bootstrap replications, ped death data;
coef 'resp', logistic regression, 11 covariates



confidence intervals for resp, using bcapar
(resphat = .943)

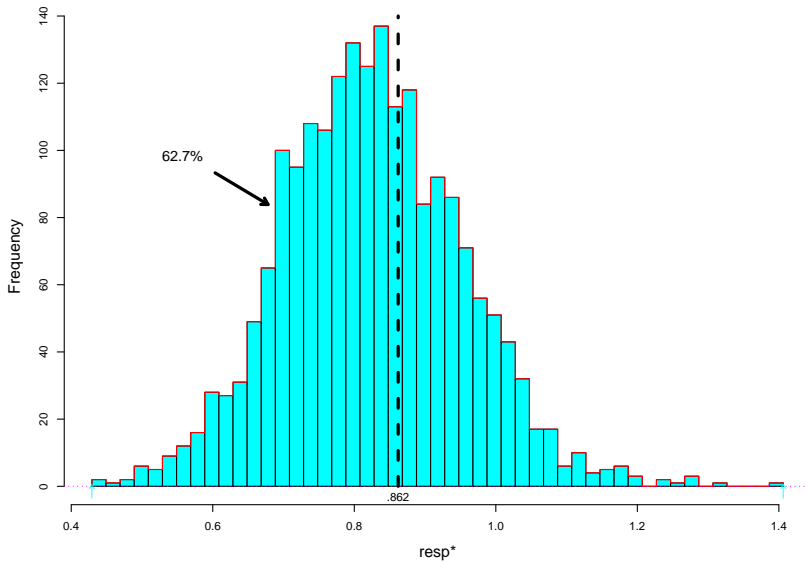


GLMNET ESTIMATE OF α

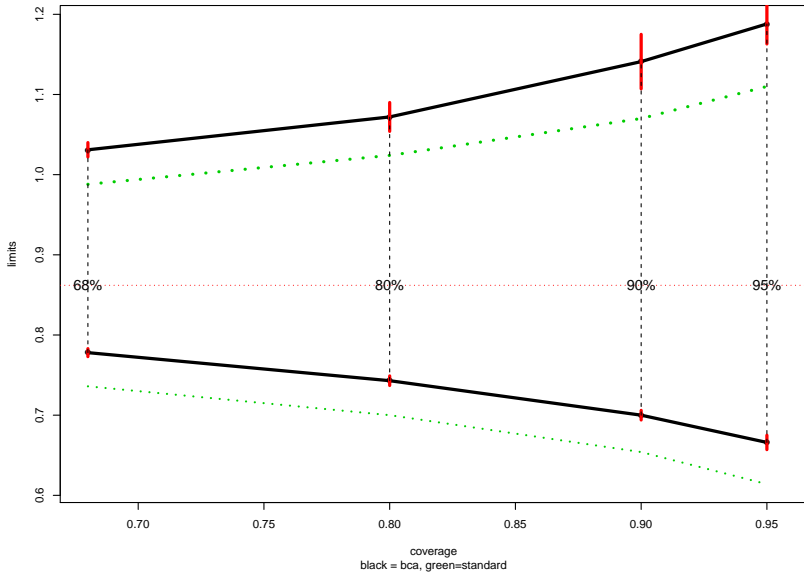
(INSTEAD OF MLE)

- `glmnet(\mathbf{X}, \mathbf{y})` fits increasing succession of logistic regressions; cross-validation used to choose “best” one (*shrinkage*)
- `glmnet` gave $(\hat{\alpha})$ and $\hat{\theta} = \hat{\alpha}_1 = 0.862$ (cf 0.943 for MLE)
- **Bootstrapping** $\hat{\alpha} \rightarrow \hat{\pi} \rightarrow \mathbf{y}^*$ and then `glmnet(\mathbf{X}, \mathbf{y}^*)` gives $\hat{\theta}^*$ and $\hat{\beta}^*$
- Confidence limits from `bcapar($\hat{\theta}, \hat{\theta}^*, \mathbf{b}^*$)`

2000 boot estimates of 'resp' using glmnet estimation
(point estimate .862)



bcapar confidence limits for 'resp' based on glmnet ;
(glmnet point estimate = .862)



COMPARING MLE AND GLMNET ESTIMATES OF “RESP”

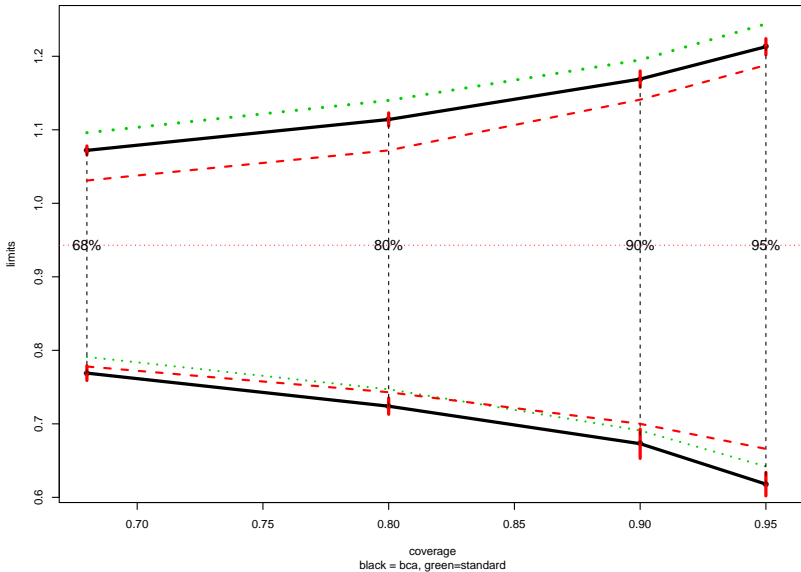
- glmnet estimate $\hat{\theta}$ is smaller than MLE:

glmnet .86 ± .13

MLE .94 ± .15

- However bcapar confidence limits have almost the same centering (but glmnet intervals shorter)

Compare logistic regression limits (black) with glmnet limits (red)



REFERENCES

- DiCiccio, T. J. and Efron, B. (1996). Bootstrap confidence intervals. *Statist. Sci.* 11: 189–228.
- Efron, B. (2018). R program `bcajack`. Available from the author's web site `efron.web.stanford.edu` under “Talks”.
- Efron, B. and Hastie, T. (2016). *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*. Cambridge: Cambridge University Press, Institute of Mathematical Statistics Monographs (Book 5).