*The Annals of Statistics*
2016, Vol. 44, No. 5, 1821–1836
DOI: 10.1214/16-AOS1489


# PETER HALL'S CONTRIBUTIONS TO THE BOOTSTRAP

BY SONG XI CHEN

*Peking University and Iowa State University*

**1. Introduction.** Professor Peter Hall's first publication appeared in 1977, one year after he was awarded DPhil from the University of Oxford and MSc from the Australian National University (ANU). His first paper on the bootstrap appeared in 1985 [Hall (1985)] on the block bootstrap of spatial point patterns, six years after the seminal paper [Efron (1979)] that introduced the bootstrap method. Hall (1985) was his 103rd publication and proposed resampling "tiles" of spatial data ("coverage patterns" in his words) to preserve the spatial dependence. The proposal was one year earlier than Carlstein's (1986) proposal of resampling nonoverlapping blocks of data sub-series for time series data.

The early publications of Peter were predominantly in probability theory: strong approximation and limiting distributions of various stochastic quantities. In the first ten years (1976–1985) as a publishing academic, he started as a probabilist and made a transition to Statistics. An inspection on the publications reveals that among the 102 papers published prior to his first bootstrap paper [Hall (1985)], 66 papers were published in the probability journals and 36 in the statistics journals: the Annals, Biometrika, JRSS-B, the Annals of the Institute of Statistical Mathematics (AISM), Journal of Multivariate Analysis and Australian Journal of Statistics. The first statistics paper was Hall (1980) appeared in the Japanese Annals AISM, which studied estimating a probability density function on the half real line with the orthogonal series. Hall (1981) was his first *The Annals of Statistics* (AoS) paper and his second statistics paper, which was on the density estimation with trigonometric series. The two works indicated that Peter's transition to Statistics started from nonparametric curve estimation, a field to which he has made enormous contributions which is surveyed by Cheng and Fan (2016) in this memorial issue.

By a search of four key words "bootstrap,""bagging," "resampling" and "percentile" on titles of Peter's more than 650 publications, we find 80 papers in the 30 year span from 1985 to 2015, with 50 of these papers published in the first ten years 1985–1994. See Table 1 for yearly counts on the bootstrap publications, which indicates a high intensity of output by Peter on the subject in this period. Figure 1

Received June 2016.
*MSC2010 subject classifications.* 62G09.
*Key words and phrases.* Block bootstrap, confidence intervals, Edgeworth expansion, hypothesis testing, iterated bootstrap, percentile method, percentile-t method.

TABLE 1
*Yearly publication counts of Professor Hall on the bootstrap with four key words: "bootstrap," "bagging," "resampling" and "percentile" in the titles of publications*

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Year | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 |
| Publications | 1 | 2 | 2 | 5 | 8 | 8 | 5 | 9 | 6 | 4 | 1 | 3 |
| Year | 98 | 99 | 00 | 01 | 03 | 06 | 07 | 08 | 09 | 10 | 13 | 15 |
| Publications | 3 | 3 | 5 | 3 | 2 | 1 | 2 | 1 | 3 | 1 | 1 | 1 |

presents a word-cloud based on the titles of the 80 bootstrap papers, which provides a summary on the topics of these works. It is clear from the word-cloud that "confidence intervals/regions" is a center-piece of Peter's work on the bootstrap. However, it is not clear which would come next. Hence, I exercise my personal selection that is biased toward the first 10 years or so from 1985 to mid-1990s as it was the most exciting period for the development of the bootstrap technology. My selection is by no means comprehensive even to the mentioned period, and only reflects a personal view on the contributions of Peter on the development of the bootstrap method.

**2. An edgeworth view.** After Efron's (1979) pioneering paper, there was "an almost bewildering array of bootstrap methods" [Hall (1988a)] that firmly injects computation to the statistical inference in general and resampling in particular. Using a comment from Peter's monograph on the bootstrap [Hall (1992a)], "Efron's contribution was to marry the bootstrap to modern computational power."

One basic appeal of the bootstrap method is in its providing an alternative computational approach to approximate the distribution of a statistic. Before the coming of the bootstrap, the approximation had been predominantly carried out via

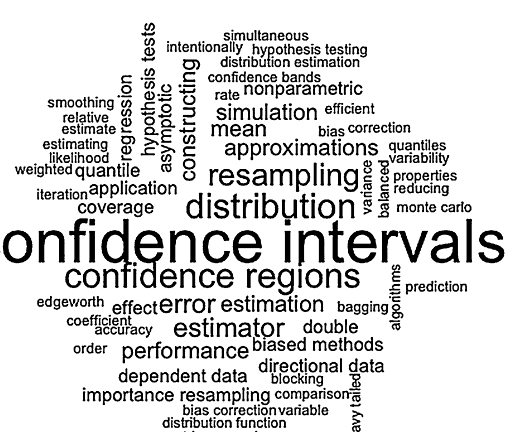


FIG. 1. *Word-cloud from the* 80 *bootstrap publications after removing "bootstrap."*

asymptotic analysis that replied on delicate mathematical derivations and large sample sizes.

Efron introduced in his seminar paper in 1979, and more elaborately in his monograph [Efron (1982)], the percentile confidence intervals for a parameter $\theta$. Let $\theta = \theta(F)$ where $F$ is the distribution from which a random sample (IID observations) $\mathcal{X}_n = \{X_1, \dots, X_n\}$ is drawn, and $\hat{\theta} = \theta(F_n)$ be the plug-in estimator by replacing $F$ with $F_n$, the empirical distribution based on the sample. Let $X_1^*, \dots, X_n^*$ be an IID resample drawn from $F_n$, and $\hat{\theta}^* = \theta(F_n^*)$ be the bootstrap estimate based on the resample, where $F_n^*$ denotes the empirical distribution based on the resample. Let $F_{\hat{\theta}^*}(y)$ denote the distribution of $\hat{\theta}^*$. A virtue of the bootstrap is in using the Monte Carlo simulation to approximate $F_{\hat{\theta}^*}(y)$ and the quantile function $F_{\hat{\theta}^*}^{-1}(p)$ for $p \in (0, 1)$. Efron's percentile confidence interval (one-sided) with a nominal $1 - \alpha$ coverage is

$$I_{\text{pt}} = \big(-\infty, F_{\hat{\theta}^*}^{-1}(1 - \alpha)\big). \tag{2.1}$$

The Monte Carlo approximation is carried out by repeatedly generating independent resamples and mass producing copies of the bootstrap estimates $\hat{\theta}^{*,1}, \dots, \hat{\theta}^{*,B}$ for a large integer $B$. The empirical distribution based on the $B$ bootstrap estimates, denoted as $F_{\hat{\theta},B}^*(y) = B^{-1} \sum_{b=1}^{B} I(\hat{\theta}^{*,b} \le y)$ is used to estimate $F_{\hat{\theta}}^*(y)$, with the estimation error solely controlled by $B$, which leads to the bootstrap estimate of the quantile $F^{*-1}_{\hat{\theta}}(p)$.

There were a set of theoretical analyzes on the properties of various versions of the bootstrap in the early to middle 1980s, trying to find theoretical justification to Efron's proposal. Singh (1981) studied the bootstrap approximation to the distribution of the standardized sample mean $\sqrt{n}(\bar{X}_n - \mu)$, where $\mu = E(X)$ and $\sigma^2 = \mathrm{Var}(X)$ are both univariate. He found that $\sqrt{n}(\bar{X}_n^* - \bar{X}_n)$ has the same one-term Edgeworth expansion as that of $\sqrt{n}(\bar{X}_n - \mu)$.

In the same issue of the Annals where Singh (1981) appeared, Bickel and Freedman (1981) presented the bootstrap distributional approximation of $\sqrt{n}(\bar{X}_n - \mu)$, and extended to other statistics (the U-statistics, von Mises functionals, and empirical processes). They showed that for the above mentioned statistics, the bootstrap provides accurate approximation in the leading order. Beran (1982) gave minimax results regarding the bootstrap estimates and its one-term Edgeworth expansion. In 1983, Singh published a paper on the Annals with his PhD adviser, G. J. Babu [Babu and Singh (1983), Issue 3], that extended Singh (1981) to a weighted average of multi-population means.

On the previous issue of the Annals, Peter [Hall (1983b), Issue 2] published his third Annals paper titled "Inverting an Edgeworth Expansion" for a general parameter $\theta = f(\mu)$ for a smoothed function $f$. This was the setting for many of his studies on the bootstrap and a related computer intensive statistical method called the empirical likelihood [Hall and La Scala (1990); Diccico, Hall and Romano

(1991)]. Peter had just published a paper [Hall (1983a)] on Edgeworth expansions of Stein's statistics in the *Math. Proceeding of Cambridge Philosophical Society*. We note that closed related key words "rate of convergence" and "central limit theorem" had simultaneously appeared in several titles of Peter's publications prior to 1983.

Hall (1983b) considered an estimator $\hat{\theta}_n$ of $\theta = f(\mu)$, which can be usually obtained by the plug-in estimator $f(\bar{X}_n)$. Suppose that $\hat{\theta}_n$ admits an Edgeworth expansion

$$P\{\sqrt{n}(\hat{\theta}_n - \theta) \le x\} = \Phi(x) + n^{-1/2}\xi_{11}(x)\phi(x) + n^{-1}\xi_{12}(x)\phi(x) + \cdots \tag{2.2}$$

for some functions $\xi_{11}, \xi_{12}, \ldots$, where $\Phi$ denoted the standard normal distribution. The paper first established an one-term inversion of the Edgeworth expansion

$$P\{\sqrt{n}(\hat{\theta}_n - \theta) \le x - n^{-1/2}\hat{\xi}_{11}(x)\} = \Phi(x) + n^{-1}\xi_{22}(x)\phi(x) + \cdots \tag{2.3}$$

uniformly in $x$ over compact intervals, where $\hat{\xi}_{11}$ is a $\sqrt{n}$-consistent estimator of $\xi_{11}$. And soon he laid out a general iteration of the inversion such that

$$\begin{aligned} &P\{\sqrt{n}(\hat{\theta}_n - \theta) \le x - n^{-1/2}\hat{\eta}_{11}(x) - \cdots - n^{(k-1)/2}\hat{\eta}_{k-1}(x)\} \\ &\quad = \Phi(x) + O(n^{-k/2}). \end{aligned} \tag{2.4}$$

Peter proved the validity of the inversion for $k = 3$. And it turned out that this framework of Edgeworth expansions and their inversions were the very technique Peter later used to analyze the bootstrap method. The iterative inversion of the Edgeworth expansions is the tool behind the iterative bootstrap proposed in Hall (1986a) and Hall and Martin (1988).

Noting Hall's (1983b) results on inverting the Edgeworth expansion, Abramovitch and Singh (1985) suggested an explicit corrections by estimating $\xi_{11}$ to attain the second correctness of the one-sided confidence interval for the bootstrap. The authors utilized (2.3) and carried out explicit correction to $\hat{\theta}$ that involved estimating $\xi_{11}$. Peter did not like the approach as this was too mechanical as clearly expressed in Hall (1986a). At this time, the magic power of the bootstrap was still to be discovered.

Hall [(1986a), AoS] was an important paper that made critical discoveries which paved the way for further developments. On the same issue of the Annals, there was an invited paper by Jeff Wu [Wu (1986)] on the jackknife and bootstrap for regressions. Hall (1986a) first uncovered the advantage of bootstrapping a Studentized statistic for, again, a $\theta = f(\mu)$. Let $\sigma^2(\hat{\theta})$ be the asymptotic variance of $\hat{\theta}_n = f(\bar{X}_n)$, and $\hat{\sigma} =: \hat{\sigma}(\hat{\theta}_n)$ be the plug-in estimator that replaces all the population moments by the sample ones. Peter defined a Studentized quantity $r_2(\bar{X}_n, \hat{\sigma}) = n^{1/2}\{f(\bar{X}_n) - f(\mu)\}/\hat{\sigma}$ which is asymptotically pivotal due to the Studentizing. Let $r_2^*(\bar{X}_n, \hat{\sigma}) = n^{1/2}\{f(\bar{X}_n^*) - f(\bar{X}_n)\}/\hat{\sigma}^*$ be the bootstrap version of $r_2(\bar{X}_n, \hat{\sigma})$, and $t^*_\alpha$ be the $\alpha$th conditional quantile of $r^*_2(\bar{X}_n, \hat{\sigma})$ given the original sample. Peter proved that

$$P\bigl(n^{1/2}\{f(\bar{X}_n) - f(\mu)\}/\hat{\sigma} \le t^*_\alpha\bigr) = \alpha + O\bigl(n^{-1}\bigr), \tag{2.5}$$

where the $n^{-1/2}$-order term such as that in (2.2) vanished. In addition, he found that if the unknown $\sigma$ was used instead of $\hat{\sigma}$ in $r^*_2(\bar{X}_n, \hat{\sigma})$, the $n^{-1/2}$-term would reappear in the corresponding Edgeworth expansion to (2.5). This means that the bootstrap when properly Studentized can automatically perform a one-term Edgeworth expansion inversion, making the one-sided confidence intervals second-order accurate. Later in the paper, Peter outlined the foundation of the bootstrap iteration, namely if one do bootstrap on the bootstrap resamples of the Studentized quantity $r^*_2(\bar{X}_n, \hat{\sigma})$, namely the double bootstrap, the $n^{-1}$-term in (2.5) would disappear making the bootstrap the third-order accurate. This would go on with any extra layer of the bootstrapping, provided that there are enough moments existed and $f$ is sufficiently smooth. These were the most elegant results for the bootstrap, established by Peter's Edgeworth view. The results in Hall (1986a) motivated a set of other findings on the bootstrap. Peter's significant contribution to the understanding of the bootstrap method was largely conceived by this paper, which is further enhanced by two papers two years later in Hall (1988a, 1988b).

In the same issue of the Annals as Hall (1986a) Hall (1986b) quantified the role of the number of bootstrap repetitions $B$ on the coverage probability of the percentile-t bootstrap confidence intervals for the general smooth function of mean estimator $\hat{\theta}_n$. Using the results in Hall (1986a), Peter proved that if $B$ was chosen such that the nominal coverage $1 - \alpha = b/(B + 1)$ for an integer $b \le B$, which could be easily done for integer percentage for $1 - \alpha$ like 0.95 or 0.99, the coverage error due to using any value of $B$ including for finite $B$ does not exceed the worse coverage error under $B = \infty$. He later outlined that the impact of using a small $B$ was largely on the length of the intervals rather than the coverage. He stressed that the above results may not be true for confidence intervals not based on Studentization, for instance the percentile intervals. This further contributed to his belief for Studentizing and using the percentile-t confidence intervals.

**3. Studentizing and percentile-t bootstrap.** At the early years of the bootstrap revolution, a range of methods had been proposed for constructing confidence intervals for a parameter. The percentile method and the hybrid method were the early favorite, along with a bias-corrected one [Efron (1982)]. And yet, practical evaluations of these methods [Buckland (1984)] and Schenker (1985) led to results with mixed fortune. These prompted Efron to proposed another version of the BC, the accelerated BC (ABC) [Efron (1987)] based on the idea of transformation and variance stabilizing.

There was an urgent need to provide in-depth analyses on the properties of these confidence intervals, that could provide guidelines to practitioners. As a master of the Edgeworth expansions, and with the preparatory works in Hall (1986a, 1986b), Hall (1988a) settled the problem in a special invited paper of the AoS.

Let $H(x) = P\{n^{1/2}(\hat{\theta}_n - \theta)/\sigma \le x\}$ and $K(x) = P\{n^{1/2}(\hat{\theta}_n - \theta)/\hat{\sigma} \le x\}$ be the distributions of two versions of the standardizations, which were termed as "ordinary" and "Studentized" by Peter. Let $\hat{H}(x) = P\{n^{1/2}(\hat{\theta}_n^* - \hat{\theta}_n)/\hat{\sigma} \le x|\mathcal{X}\}$ and $\hat{K}(x) = P\{n^{1/2}(\hat{\theta}_n^* - \hat{\theta}_n)/\hat{\sigma}^* \le x|\mathcal{X}\}$ be the bootstrap versions of $H(x)$ and $K(x)$. Bootstrap estimates of the $\alpha$th quantiles are $\hat{x}_\alpha = \hat{H}^{-1}(\alpha)$ and $\hat{y}_\alpha = \hat{K}^{-1}(\alpha)$.

Hall (1988a) first evaluated four types of upper confidence intervals with nominal coverage $1 - \alpha$, constructed using the following upper critical points:

$$\begin{aligned} &\text{Z-interval:} \quad \hat{\theta}_{\mathrm{ORD}}(\alpha) = \hat{\theta}_n - n^{-1/2}\sigma\hat{x}_{1-\alpha}, \\ &\text{Percentile-t:} \quad \hat{\theta}_{\mathrm{STUD}}(\alpha) = \hat{\theta}_n - n^{-1/2}\hat{\sigma}\hat{y}_{1-\alpha}, \\ &\text{Hybrid:} \quad \hat{\theta}_{\mathrm{HYB}}(\alpha) = \hat{\theta}_n - n^{-1/2}\hat{\sigma}\hat{x}_{1-\alpha} \quad \text{and} \\ &\text{Percentile:} \quad \hat{\theta}_{\mathrm{BACK}}(\alpha) = \hat{\theta}_n + n^{-1/2}\hat{\sigma}\hat{x}_{\alpha}. \end{aligned} \tag{3.1}$$

Peter pointed out that while $\hat{\theta}_{\mathrm{ORD}}$ and $\hat{\theta}_{\mathrm{STUD}}$ were sensible as they both "look at the right tables" (the quantiles were obtained from the correct bootstrap distributions), the hybrid $\hat{\theta}_{\mathrm{HYB}}$ "looks at the wrong table" as it mixes up $\hat{\sigma}$ with $\hat{x}_{1-\alpha}$, and $\hat{\theta}_{\mathrm{BACK}}$ (which is equivalent to the percentile method) "amounts to looking up the wrong table backward."

Realizing the issue with the percentile confidence interval, Efron (1982) proposed the Bias-corrected (BC) confidence interval with an adjustment to the level $\alpha$ used in $\hat{\theta}_{\mathrm{BACK}}(\alpha)$. Let $\hat{G}(x) = P(\hat{\theta}^* \le x|\mathcal{X})$ and $\hat{m} = \Phi^{-1}\{\hat{G}(\hat{\theta})\}$. As $\hat{G}(\hat{\theta})$ may not be 0.5 so that $\hat{m}$ may not be zero, Efron suggested shifting $\alpha$ to $\hat{\beta} = \Phi(z_\alpha + 2\hat{m})$ in the bootstrap critical point, namely using $\hat{\theta}_{\mathrm{BACK}}(\hat{\beta})$ instead of $\hat{\theta}_{\mathrm{BACK}}(\alpha)$. However, mixed numerical results on the percentile and the BC prompted Efron to propose another version of the BC, the accelerated BC (ABC) [Efron (1987)], which amounted to adjust $\hat{\beta}$ by adding a $n^{-1/2}$-order correction term, and had the critical point $\hat{\theta}_{\mathrm{BACK}}(\hat{\beta}_{\mathrm{ABC}})$ where $\hat{\beta}_{\mathrm{ABC}} = \Phi[\hat{m} + (z_\alpha + \hat{m})\{1 - \hat{a}(z_\alpha + \hat{m})\}^{-1}]$.

Applying the insightful "Edgeworth view," Hall (1988a) presented thorough analyzes on the four bootstrap confidence intervals that used the four critical points given in (3.1) plus the BC and ABC confidence intervals, and delivered the following assessments:

(i) Both the z- and percentile-t intervals have coverage errors of order $n^{-1}$, which are the second-order correct.

(ii) The percentile and the hybrid confidence intervals both incur coverage errors of order $n^{-1/2}$, which do not improve upon the conventional normal approximation based confidence interval.

(iii) The BC confidence interval cannot completely rescue the percentile method as it only removes the constant in the coefficient polynomial (quadratic) to the $n^{-1/2}$ term in the coverage error.

(iv) The ABC removes the entire $n^{-1/2}$ terms in the coverage error, and hence is at a par with the percentile-t method.

(v) Between the percentile-t and the ABC, the percentile-t is preferred due to its simplicity and having more attractive third-order coverage property for two-sided confidence intervals.

These together with his study on the symmetric bootstrap confidence intervals in Hall (1988b) had led Peter to advocate for the percentile-t method and Studentization in general. This message was much more emphasized in his monograph on the bootstrap [Hall (1992a)]. Indeed, the power of the bootstrap is only realized based with the version that Studentizes.

While Peter was a strong believer of the Studentization, he was fully aware of its reliance on a quality and "stable estimate of $\sigma^2$." It is particularly refreshing for me to read this passage 20 years after my own work [Chen (1996)] on the confidence intervals for a probability density function, which was a case where a stable variance estimation is not easy to come by. Peter had foreseen this in Hall (1988a).

**4. Iterative bootstrap.** Although the idea of bootstrap iteration already appeared in Hall (1986a) [also Beran (1987) and Loh (1987)], which was framed by applying successive Edgeworth expansion inversions, a full spelling of the method was given only in Hall and Martin (1988). Unlike Hall (1986a), Hall and Martin's bootstrap iteration was rooted in "a measure of quality or accuracy, expressed in the form of an equation whose solution is sought," which facilitates a bootstrap iteration via "iterating the empirical solution to this equation so as to improve accuracy."

Let $F_0$ be the underlying distribution and $F_1$ be the empirical distribution based on a random sample drawn from $F_0$ and $f(F_0, F_1)$ be an inferential equation that can manifest various forms of statistical inference problems. Specifically, the target of the inference was defined via

$$E\{f(F_0, F_1)|F_0\} = 0. \tag{4.1}$$

An array of inference tasks (bias and variance estimation, quantile estimation or coverage of confidence intervals) can be expressed via (4.1) as shown in Hall and Martin (1988) explicitly. Let $F_i$ be the empirical distribution on an $i$th level resample drawn conditionally from $F_{i-1}$, for $i = 2, 3, \ldots$. Specifically for $i = 2$, the inference equation (4.1) can be mirrored by an one-level higher replica

$$E\{f(F_1, F_2)|F_1\} = 0, \tag{4.2}$$

which can be simulated numerically whereas its parent (4.1) cannot. These processes can be continued to the next level, which become the essence of the bootstrap iteration.

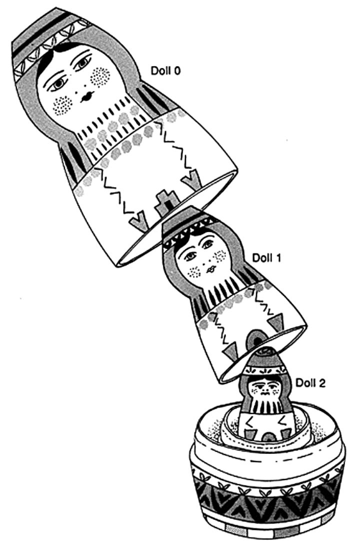


FIG. 2. *Illustrating the idea of bootstrap iteration via a Russian matryoshka doll. The figure was hand-drawn by Peter and is reproduced from a figure in Hall (1983a) with permission from Springer.*

Figure 2 is a drawing of a Russian matryoshka doll that appeared in the monograph Hall (1992a) (with permission). Peter drew the figure himself from a matryoshka doll he had bought in Moscow in September 1974 when he traveled on the Trans-Siberian Railway on his way to Oxford to start his DPhil. The drawing was a lively illustration of the ideas of the bootstrap iteration, namely the inference at a level $i$ based on a version of (4.1) can be mimicked by that at the level $i + 1$, utilizing an analogue with respect to the number of dots on the faces of the dolls.

Hall, Martin and Schucany (1989) demonstrated the ability of the bootstrap iteration for the notorious problem of constructing confidence intervals for the correlation coefficient $\rho$. The percentile-t was known [Efron (1982)] for not working well for $\rho$ especially when $|\rho|$ is close to 1, largely due to the instability in the variance estimation as well as the excessive skewness. Hall, Martin and Schucany implemented the generally inferior percentile method (which is more stable than the percentile-t due to not having to Studentize) with one level of bootstrap iteration, that is, the double bootstrap, which produced substantial improvement in the coverage. In one of Peter's last papers, Chang and Hall (2015) showed that a single double bootstrap resampling is enough for achieving bias reduction, but not so for coverage improvement.

**5. Block bootstrap.** As mentioned at the beginning of this article, Peter's first published paper on the bootstrap was on "resampling a coverage pattern" [Hall (1985)]. The "coverage pattern" that Peter referred to was a spatial point pattern, a form of two-dimensional dependent data. Peter suggested two schemes of "resampling" units for the spatial point patterns. The first one (termed as "fixed" tiles) partitions the entire observation region $\mathcal{R}$ "into $m$ congruent and nonoverlapping tiles" denoted as $\{t_i + \mathcal{A}\}$, where $t_i$ are the centers of the tiles and $\mathcal{A}$ is a standard title that is congruent to $\mathcal{R}$. The second scheme (the "moving" tiles) resamples from $t + \mathcal{A}$, where $t$ is uniformly distributed over $\mathcal{B}$, the set of $t$ such that $t + \mathcal{A} \subset \mathcal{R}$.

The "fixed" tile scheme is just the "nonoverlapping data blocks" for time series proposed by Carlstein (1986), while the "moving" tile scheme is equivalent to the "moving data block" proposed by Künsch (1989) and Liu and Singh (1992). Hall (1985) demonstrated that both resampling schemes led to consistent bootstrap variance estimation of general statistics $\hat{\theta}$ on the "coverage pattern" with much similar properties. He quantified the role of the number of the blocks $m$ and the total spatial area $\mathcal{R}$ on the bias and the variance of the bootstrap variance estimators: the bias can be reduced by increasing $m$ and $\mathcal{R}$ while the variability of the bootstrap variance estimation can be reduced by the same increase in $m$ and $\mathcal{R}$. It appeared that neither Carlstein (1986) nor Künsch (1989) were aware of Hall (1985). Although the contexts of these latter two papers are on the time series, where the dependence has a one-dimensional source, the idea and the purpose were essentially the same, namely in consistently acquiring variance estimation of an estimator based on dependent data. Hall (2003) gave an interesting historic coverage on the use of resampling methods for the crop yield surveys by P. C. Mahalanobis.

Davison and Hall [(1993), Austral. J. Statist.] considered the accuracy of the distributional approximation offered by using the block bootstrap method to that of a Studentized statistic with dependent data. The statistic considered was the sample mean, and the dependent data were generated according to a linear process with IID innovations. The study was made in the context of "Götze and Künsch (1990) announced that in the case of dependent data, a particular bootstrap method can be used to improve on the normal approximation to the distribution of the Studentized mean. In this note, we show that this result depends fundamentally on the method of Studentization, in respect of which there are important differences between the approaches which should be taken for dependent and independent data" [Davison and Hall (1993)].

Götze and Künsch (1990) was an abstract that claimed that a Studentized bootstrap approximation could offer a better distributional approximation than the conventional asymptotic normality, which was analogue to the percentile-t method established for the IID data. Götze and Künsch (1996) appeared to be the paper abstracted by Götze and Künsch in 1990. The contribution of Davison and Hall (1993) was to warn and demonstrate specifically that in order to attain the claimed distributional approximation by a version of the percentile-t bootstrap care had to be exercised on how to formulate the variance estimate for the Studentizing.

Let $\{\bar{Y}_i\}_{i=1}^{b}$ be the averaged obtained on each data block. While treating these statistics from the data blocks as independent for bootstrap variance estimation may produce consistent variance estimation to the underlying variance of the statistics, the percentile-t bootstrap using the variance estimate to Studentize failed to produce a more accurate distributional approximation to the distribution of

$$T = n^{1/2} b^{-1} \hat{\sigma}^{-1} \sum_{i=1}^{b} (Y_i - \mu)$$

than the conventional normal approximation. In the above expression, $\hat{\sigma}^2 = lb^{-1} \sum_{i=1}^{b} (\bar{Y}_i - \bar{Y})^2$ where $n = lb$ and $\bar{Y}$ is the average of the entire data. Specifically, the $o(n^{-1/2})$ in (2.5) cannot be attained. Davison and Hall (1993) showed instead if $\hat{\sigma}^2$ is modified to the long-run variance form to include the covariance terms, the $o(n^{-1/2})$ accuracy can be attained by the percentile-t bootstrap.

A critical issue for the block bootstrap method which had not been addressed in 1995 is the choice of the block length. Hall, Hororwitz and Jing (1995) provided a set of timely results on the block length for inference on a parameter that is a smooth function of means. “It is shown that optimal block size depends significantly on context, being equal to $n^{1/3}$, $n^{1/4}$ and $n^{1/5}$ in the case of the variance or bias estimation, estimation of a one-sized distribution functions and the estimation of a two-sided distribution functions, respectively” [Hall, Hororwitz and Jing (1995)]. The paper showed that the rules for the block lengths for the nonoverlapping block and the overlapping (moving blocks) were largely the same. The paper provided an empirical cross-validation method for practically selecting the block length. This paper was a highly influential paper that motivates further analysis in this direction; see Politis and Romano (1994) for an alternative block bootstrap formulation, and Lahiri (1999) and (2003) for further developments.

**6. Bootstrap hypothesis testing.** Much of the attention on the bootstrap had been around the estimation and confidence intervals; there was far less work on issues related to the hypothesis testing. A key aspect with the hypothesis testing was respecting the null hypothesis by the bootstrap resampling and the associated test statistics, on which Young (1986), Beran (1988), Hinkley (1988) and Fisher and Hall (1990) had investigated.

Hall and Wilson (1991) published in the Consultant’s Forum of *Biometrics*, which offered “two guidelines for bootstrap hypothesis testing.” The paper was written out of an urgency to let practitioners to use the sound bootstrap method, which was clearly shown in the author’s writing toward the end of the third paragraph:

“However, these relatively technical contributions are not easily accessible to biometricians, and so their theoretical recommendations can go unheeded in practice. Indeed, the present note was partly motivated by an article of...” [Hall and Wilson (1991)].

The two guidelines were written specifically for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, which can be easily modified for one-sided $H_1$. Let $\hat{\theta}$ be an estimator of $\theta$ based on a sample $X_1, \ldots, X_n$, and $\hat{\theta}^*$ be the estimator based a bootstrap resample.

The first guideline offered was "Resample $\hat{\theta}^* - \hat{\theta}$, not $\hat{\theta}^* - \theta_0$" in order to respect $H_0$. Resampling of $\hat{\theta}^* - \theta_0$ was seen in publications which commanded certain influence and would distort the power of the testing. The guideline would help the test to uphold its intrinsic power of the test.

The second guideline was to "based on the bootstrap distribution of $(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}^*$, not on the bootstrap distribution of $(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}$ or of $\hat{\theta}^* - \theta$." This was advocated based on the more accurate bootstrap distribution approximation offered by the percentile-t method as shown by Hall (1988a, 1988b), which would translate to more accurate approximation to the nominal significant level of the test. The Studentization that produced the pivotal quantity also made the task of respecting the first guideline easier.

**7. Other contributions.** In additional to the fundamental contributions outlined in the previous sections, Peter has made significant contributions on using the bootstrap for a set of important areas of statistical inference. These includes the nonparametric curve estimation with representative works Hall (1992b, 1992c, 1993) that validated the percentile-t method for constructing confidence intervals and bands for density and nonparametric regression functions, as well as the advocate for the smooth bootstrap. See the review by Professors Cheng and Fan [Cheng and Fan (2016)] in this issue for Peter's work in the nonparametric curve estimation. Peter has also skillfully applied the bootstrap methods for high dimensional statistical inference and classification, which includes Hall and Samworth (2005) for applying the bagging (bootstrap aggregating) method with the nearest neighbor classifier, Fan, Hall and Yao (2007) for the bootstrap calibration in high dimensional multiple hypothesis testing, and Hall, Lee and Park (2009) for a bootstrap subsampling penalty based LASSO that achieved the oracle performance for high dimensional variable selection. Readers should read the review article in this issue by Professor Samworth [Samworth (2016)] on Peter's other contribution in high dimensional statistics and classification. For Peter's contributions in the deconvolution method and in the functional data analysis, readers should see Delaigle (2016) and Müller (2016) in this memorial issue.

**8. Personal reminisces.** One day in August of 1989, while I walked past a bulletin board on the campus of the Victoria University of Wellington (VUW), I saw an advertisement from the ANU for PhD student recruitment with scholarships. The contact person was Professor Hall, as the Head of the Statistical Research Section at the School of Mathematical Science. After consulting with Professor David Vere-Jones, the Chair Professor at VUW and Dr Xiaogu Zheng, my teacher from my Beijing Normal University (BNU) days (Dr Zheng worked at the VUW at the time), I decided to apply. The economic situation in New Zealand at the time was not that promising, as reflected in a much tighter scholarship scene for overseas students.

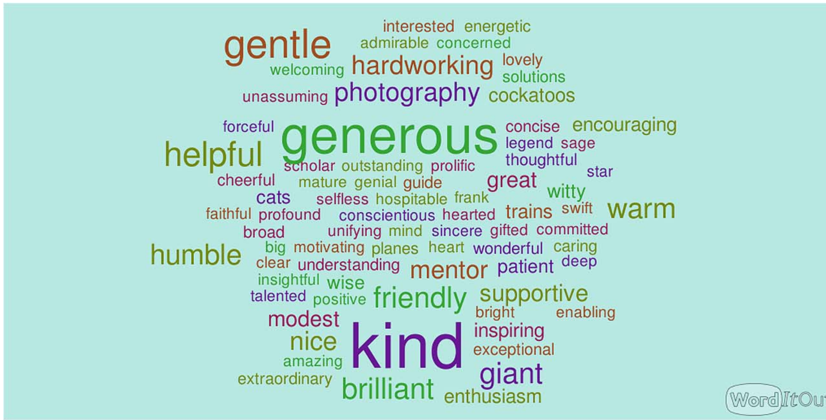


FIG. 3. *Word-cloud of tributes left at the memorial website* (*http://www.stat-center.pku.edu.cn/Peter_Gavin_Hall/*) *at the Center for Statistical Science*, *Peking University*, *up to February* 08*th*, 2016. *Composed by Terry Speed*.

FIG. 4. *Group photo of a short course by Professor Hall in March* 2012 *at Peking University. The short course consisted* 16 *lectures on Methodology and Theory of the Bootstrap. Front row*: *Bin Guo*, *a student*, *Weixiang Zhang*, *Dayue Chen*, *Song Xi Chen*, *Peter Hall*, *Gang Tian*, *Zhi Geng*, *Jinzhu Jia and a student*.

FIG. 5. *On the Great Wall at Mutianyu after a snow/rain mix in March* 2012. *From Left*: *Song Xi Chen*, *Peter Hall with his camera*, *Jinyuan Chang*, *Jing He and Jiani Song*.

I received a full ANU PhD scholarship in January 1990. However, my visa to Australia was not easy. I remember one day in the office of David, he phoned Peter in Canberra to inform my ordeal. Peter's immediate response was "let me talk to Chris and get back to you." A few weeks later, I received a phone call from Professor Heyde telling me that "the thing had been refreshed" and "it should be ok." With the tremendous help of Professors Vere-Jones, Hall, Heyde and Zheng (and his family), we finally flew to Canberra in late September 1990.

My first meeting Peter was at the arrival hall of the Canberra airport, which was quite late at night. Peter came to pick us up. After a brief greeting, to my great astonishment, he picked up two pieces of heavy luggage (with one hand each) and walked quickly to his white Subaru. He had already collected the key to an apartment in the Graduate Court. After we got there, he lifted the luggage to the unit on the third level of the building, and said good night. Whenever I recall the first meeting with Peter, I feel the warmth and generosity of him as a very kind and gentle person. This has been surely echoed by many people in their tributes registered at a memorial website established at the Center for Statistical Science, Peking University. Figure 3 displays a word-cloud based on these tributes, which has been kindly composed by Professor Terry Speed at the University of Melbourne. The sentiment of these tributes is highly consistent with that of my first meeting with Peter and those in my subsequent 26 years as his student, mentee and colleague.

Peter visited the Center for Statistical Science at the Peking University in the spring of 2012 and 2013. He was a member of the International Advisory Committee of the Center. In March 2012, Peter gave a short course on the bootstrap method, which consisted of 16 lectures; see Figures 4 and 5. He returned in March 2013 with six lectures on the functional data analysis. On both occasions, more than 120 graduate students and faculties from all over China attended his two short courses. These lectures have had long lasting impacts on the participants not only on the scientific content but also on the scholarship and professionalism displayed by Peter during his very dedicated teaching.

**Acknowledgments.** I thank Jeannie Hall for providing the source file and the background information regarding Figure 2, Professors Richard Samworth and Bingyi Jing for comments and suggestions and Xiaojun Mao and Jing He for editorial assistance. The author was partially supported by Natural Science Foundation of China grants 11131002, 71371016 and 71532001, National Key Basic Research Program of China Grant 2015CB856000,and NSF grant DSM-1309210.

## REFERENCES

ABRAMOVITCH, L. and SINGH, K. (1985). Edgeworth corrected pivotal statistics and the bootstrap. *Ann. Statist.* **13** 116–132. MR0773156

BABU, G. J. and SINGH, K. (1983). Inference on means using the bootstrap. *Ann. Statist.* **11** 999–1003. MR0707951

BERAN, R. (1982). Estimated sampling distributions: The bootstrap and competitors. *Ann. Statist.* **10** 212–225. MR0642733

BERAN, R. (1987). Prepivoting to reduce level error of confidence sets. *Biometrika* **74** 457–468. MR0909351

BERAN, R. (1988). Prepivoting test statistics: A bootstrap view of asymptotic refinements. *J. Amer. Statist. Assoc.* **83** 687–697. MR0963796

BICKEL, P. J. and FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217. MR0630103

BUCKLAND, S. T. (1984). Monte Carlo confidence intervals. *Biometrics* **40** 811–817.

CARLSTEIN, E. (1986). Simultaneous confidence regions for predictions. *Amer. Statist.* **40** 277–279. MR0866907

CHANG, J. and HALL, P. (2015). Double-bootstrap methods that use a single double-bootstrap simulation. *Biometrika* **102** 203–214. MR3335106

CHEN, S. X. (1996). Empirical likelihood confidence intervals for nonparametric density estimation. *Biometrika* **83** 329–341. MR1439787

CHENG, M.-Y. and FAN, J. (2016). Peter Hall's contributions to nonparametric function estimation and modeling. *Ann. Statist.* **44** 1837–1853.

DAVISON, A. C. and HALL, P. (1993). On Studentizing and blocking methods for implementing the bootstrap with dependent data. *Austral. J. Statist.* **35** 215–224. MR1244850

DELAIGLE, A. (2016). Peter Hall's main contributions to deconvolution. *Ann. Statist.* **44** 1854–1866.

DICICCIO, T., HALL, P. and ROMANO, J. (1991). Empirical likelihood is Bartlett-correctable. *Ann. Statist.* **19** 1053–1061. MR1105861

EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26. MR0515681

EFRON, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans. CBMS-NSF Regional Conference Series in Applied Mathematics* **38**. SIAM, Philadelphia, PA. MR0659849

EFRON, B. (1987). Better bootstrap confidence intervals. *J. Amer. Statist. Assoc.* **82** 171–200. MR0883345

FAN, J., HALL, P. and YAO, Q. (2007). To how many simultaneous hypothesis tests can normal, Student's $t$ or bootstrap calibration be applied? *J. Amer. Statist. Assoc.* **102** 1282–1288. MR2372536

FISHER, N. I. and HALL, P. (1990). On bootstrap hypothesis testing. *Austral. J. Statist.* **32** 177–190. MR1079387

GÖTZE, F. and KÜNSCH, H. R. (1990). Blockwise bootstrap for dependent observation: Higher order approximations for Studentized statistics (abstract). *Inst. Math. Statist. Bull.* **19** 443.

GÖTZE, F. and KÜNSCH, H. R. (1996). Second-order correctness of the blockwise bootstrap for stationary observations. *Ann. Statist.* **24** 1914–1933. MR1421154

HALL, P. (1980). Estimating a density on the positive half line by the method of orthogonal series. *Ann. Inst. Statist. Math.* **32** 351–362. MR0609028

HALL, P. (1981). On trigonometric series estimates of densities. *Ann. Statist.* **9** 683–685. MR0615446

HALL, P. (1983a). Edgeworth expansion of the distribution of Stein's statistic. *Math. Proc. Cambridge Philos. Soc.* **93** 163–175. MR0684286

HALL, P. (1983b). Inverting an Edgeworth expansion. *Ann. Statist.* **11** 569–576. MR0696068

HALL, P. (1985). Resampling a coverage pattern. *Stochastic Process. Appl.* **20** 231–246. MR0808159

HALL, P. (1986a). On the bootstrap and confidence intervals. *Ann. Statist.* **14** 1431–1452. MR0868310

HALL, P. (1986b). On the number of bootstrap simulations required to construct a confidence interval. *Ann. Statist.* **14** 1453–1462. MR0868311

HALL, P. (1988a). Theoretical comparison of bootstrap confidence intervals. *Ann. Statist.* **16** 927–985. MR0959185

HALL, P. (1988b). On symmetric bootstrap confidence intervals. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **50** 35–45. MR0954730

HALL, P. (1992a). *The Bootstrap and Edgeworth Expansion*. Springer, New York. MR1145237

HALL, P. (1992b). Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *Ann. Statist.* **20** 675–694. MR1165587

HALL, P. (1992c). On bootstrap confidence intervals in nonparametric regression. *Ann. Statist.* **20** 695–711. MR1165588

HALL, P. (1993). On Edgeworth expansion and bootstrap confidence bands in nonparametric curve estimation. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **55** 291–304.

HALL, P. (2003). A short prehistory of the bootstrap. *Statist. Sci.* **18** 158–167. MR2026077

HALL, P., HOROWITZ, J. L. and JING, B.-Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika* **82** 561–574. MR1366282

HALL, P. and LA SCALA, B. (1990). Methodology and algorithms of empirical likelihood. *Intl. Statist. Review* **58** 1039–1049.

HALL, P., LEE, E. R. and PARK, B. U. (2009). Bootstrap-based penalty choice for the lasso, achieving oracle performance. *Statist. Sinica* **19** 449–471. MR2514171

HALL, P. and MARTIN, M. A. (1988). On bootstrap resampling and iteration. *Biometrika* **75** 661–671. MR0995110

HALL, P., MARTIN, M. A. and SCHUCANY, W. R. (1989). Better nonparametric bootstrap confidence intervals for the correlation coefficient. *J. Stat. Comput. Simul.* **33** 161–172. MR1029439

HALL, P. and SAMWORTH, R. J. (2005). Properties of bagged nearest neighbour classifiers. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **67** 363–379. MR2155343

HALL, P. and WILSON, S. R. (1991). Two guidelines for bootstrap hypothesis testing. *Biometrics* **47** 757–762. MR1132543

HINKLEY, D. V. (1988). Bootstrap methods. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **50** 321–337, 355–370. MR0970971

KÜNSCH, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17** 1217–1241. MR1015147

LAHIRI, S. N. (1999). Theoretical comparisons of block bootstrap methods. *Ann. Statist.* **27** 386–404. MR1701117

LAHIRI, S. N. (2003). *Resampling Methods for Dependent Data*. Springer, New York. MR2001447

LIU, R. Y. and SINGH, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the Limits of Bootstrap* (*East Lansing*, *MI*, 1990) (R. Lepage and L. Billard, eds.) 225–248. Wiley, New York. MR1197787

LOH, W.-Y. (1987). Calibrating confidence coefficients. *J. Amer. Statist. Assoc.* **82** 155–162. MR0883343

MÜLLER, H.-G. (2016). Peter Hall, functional data analysis and random objects. *Ann. Statist.* **44** 1867–1887.

POLITIS, D. N. and ROMANO, J. P. (1994). The stationary bootstrap. *J. Amer. Statist. Assoc.* **89** 1303–1313. MR1310224

SAMWORTH, R. J. (2016). Peter Hall's work on high-dimensional data and classification. *Ann. Statist.* **44** 1888–1895.

SCHENKER, N. (1985). Qualms about bootstrap confidence intervals. *J. Amer. Statist. Assoc.* **80** 360–361. MR0792734

SINGH, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9** 1187–1195. MR0630102

WU, C.-F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. Statist.* **14** 1261–1350. MR0868303

YOUNG, A. (1986). Conditioned data-based simulations: Some examples from geometrical statistics. *Int. Stat. Rev.* **54** 1–13. MR0959648

GUANGHUA SCHOOL OF MANAGEMENT AND
CENTER FOR STATISTICAL SCIENCE
PEKING UNIVERSITY
BEIJING 100871
CHINA
AND
DEPARTMENT OF STATISTICS
IOWA STATE UNIVERSITY
AMES, IOWA 50011-1210
USA
E-MAIL: songchen@iastate.edu