Department Copy

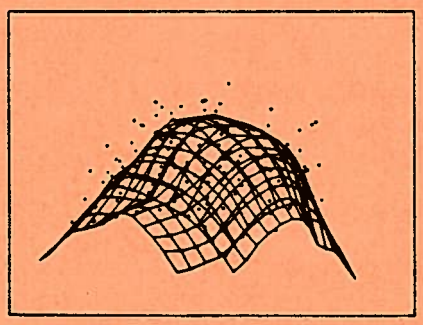
BAYESIAN NONPARAMETRIC BOOTSTRAP CONFIDENCE INTERVALS

Nils Lid Hjort

Technical Report No. 20

November 1985

Laboratory for Computational Statistics



Department of Statistics
Stanford University

BAYESIAN NONPARAMETRIC BOOTSTRAP CONFIDENCE INTERVALS*

Nils Lid Hjort

Department of Statistics, Stanford University

and

Norwegian Computing Centre

LCS Technical Report No. 20

and

Department of Statistics Report No. 240

November 1985

ABSTRACT

Let $X_1, \ldots, X_n$ be a random sample from an unknown probability distribution P on the sample space X, and let $\theta = \theta(P)$ be a parameter of interest. The present paper gives a "Bayesian bootstrap" method of obtaining Bayes estimates and Bayesian confidence limits for θ, using a (non-degenerate) Dirichlet process prior for P. This extends methods and results of Rubin (1981) and Efron (1982), in that they assume the sample space to be finite and use only a particular degenerate Dirichlet prior. An asymptotic justification of the Bayesian bootstrap is given, parallelling results of Bickel and Freedman (1981).

---

* Work supported by a National Science Foundation Grant MCS80-24649, Office of Naval Research contract N00014-83-K-0472.

1. Exact Bayesian intervals.

Let $X_1,\ldots,X_n$ be independent and identically distributed (i.i.d.) according to an unknown distribution P. For convenience take the sample space to be $\mathcal{X} = \mathbb{R}$, so that P can be identified with its distribution function (c.d.f.) F. Most of the results in this report can be generalised to any complete, separable metric space $\mathcal{X}$.

Let $\theta = \theta(F)$ be a parameter functional of interest. We shall be concerned with Bayesian nonparametric confidence statements about θ, and need to start out with a prior distribution on the space of all c.d.f.s. A natural class from which to choose is provided by Ferguson's (1973, 1974) Dirichlet processes; the class is rich, each member has large support, and basic posterior calculations are feasible. Thus let

$$F \sim \mathrm{Dir}(aF_o), \tag{1.1}$$

i.e. F is a Dirichlet process with parameter aF_o. $F_o(.) = E_B F(.)$ is the prior guess c.d.f. whereas $a > 0$ has interpretation as prior sample size.

Identify the observed sample $x_1,\ldots,x_n$ with the empirical c.d.f.

$$\hat{F}(t) = \frac{1}{n}\sum_{i=1}^{n} I\{x_i \le t\}. \tag{1.2}$$

The posterior distribution of F is

$$F|\hat{F} \underset{B}{\sim} \mathrm{Dir}(aF_o + n\hat{F}). \tag{1.3}$$

(E_B, $\underset{B}{\sim}$ etc. indicate statements relative to the chosen Bayesian framework.) Thus the function

$$G(t) = \Pr_B\{\theta(F) \le t \,|\, \hat{F}\} \tag{1.4}$$

is in principle known. We wish to calculate θ_{LOW}, θ_{UP} from the data, satisfying

$$\Pr_B\{\theta_{LOW} \le \theta(F) \le \theta_{UP} \,|\, \hat{F}\} \doteq 1 - 2\alpha, \tag{1.5}$$

say. Thus

$$\theta_{LOW} = G^{-1}(\alpha), \quad \theta_{UP} = G^{-1}(1-\alpha) \tag{1.6}$$

are the natural choices; $G^{-1}(p) = \inf\{t: G(t) \ge p\}$.

The fact that G above is only rarely explicitly available, however, necessitates devising computational approximations. The problem becomes much simpler for the particular case $a \to 0$, which is the "non-informative" case Rubin (1981) and Efron (1982, Ch. 10) consider. (Actually, they consider only finite sample spaces, but the extension to the present generality is easily made via the theory of Dirichlet processes.) Then $F|\hat{F}$ is concentrated on the observed data values,

$$F(.) = \sum_{i=1}^{n} d_i \, \delta(x_i), \tag{1.7}$$

with weights $(d_1,\ldots,d_n)$ following a Dirichlet $(1,\ldots,1)$ (uniform on the simplex of non-negative weights summing to one). It follows that values of $\theta(F)$ can be simulated according to (1.4), i.e. G may be closely approximated by Monte Carlo. (The d_i's may be simulated as $d_i = e_i/(e_1+\ldots+e_n)$, where the e_i's are i.i.d. unit exponential. If $\theta(F) = \int x dF(x)$ is the mean, for example, then a large number of realisations of $\theta(F) = \Sigma_{i=1}^{n} d_i x_i = \Sigma_{i=1}^{n} e_i x_i / \Sigma_{i=1}^{n} e_i$ can be generated, the histogram of these values would approximate G, enabling one to get good numerical approximations to the interval (1.6).) Rubin (1981) discusses this point and notes that the resulting approach, though different in interpretation, agrees well, operationally and inferentially, with the ordinary bootstrap procedure. This may be taken as but another example where Bayesian inference, starting with a non-informative ("neutral", "objective") prior distribution, resembles classical frequentist inference.

The rest of the report is concerned with the informative case $a > 0$.

2. Approximating the posterior distribution of θ.

For a few parameter functionals the posterior distribution (1.4) can be evaluated exactly. Section 5 provides calculations for $\theta = F\{A\}$, A a set of interest, and $\theta = F^{-1}(p)$, the p-quantile. In some other instances can $\theta(F)|\hat{F}$ be simulated directly, i.e. a sequence $Y_1, Y_2, \ldots$ being i.i.d. with a distribution equal to or very close to G can be generated, thus enabling one to obtain a close approximation to G and to the sought-after θ_{LOW}, θ_{UP}.

Example. Let $\theta = \int x dF(x)$ be the unknown mean of F. The exact distribution of θ given data can be obtained for some choices of the prior guess F_o, but the resulting expressions are complicated, making "exact simulation" difficult. (This approach would have to use results of the type reached by Hannum, Hollander, and Langberg (1981) and Yamato (1984), but with Dirichlet process parameter $aF_o + n\hat{F}$.) However, the posterior distribution of θ can be approximated with that of $\theta' = \Sigma_{i=1}^{n} x_i F\{x_i\} + \Sigma_{j=1}^{m} y_j F\{A_j\}$, say, where $A_1, \ldots, A_m$ is a fine partition of $R - \{x_1, \ldots, x_n\}$, and $y_j \in A_j$. θ' can then be simulated.

Hjort (1985) shows that $\alpha_m \xrightarrow{d} \alpha$ in X implies $Dir(\alpha_m) \xrightarrow{d} Dir(\alpha)$ in the space of probability measures on X, and that $\int x dF_m(x) \xrightarrow{d} \int x dF(x)$ under a mild extra condition on $\{\alpha_m\}$. This justifies $\theta \doteq \theta'$ above.

The example illustrates that (1.4) in general will be difficult to obtain from (1.3) via direct simulation of $\theta(F)|\hat{F}$. A simpler method, which we call the Bayesian bootstrap (BB) method, is however possible, and is now described.

Note first that

$$\hat{F}_B(t) = E_B\{F(t)|\hat{F}\} = \frac{a}{a+n} F_o(t) + \frac{n}{a+n} \hat{F}(t) \qquad (2.1)$$

is the natural Bayes estimate of F(t). Generate a BB sample $X_1^*, \ldots, X_{n+a+1}^*$ from $\hat{F}_B$. (This is easy, provided it is feasible to sample from F_o: X_i^* is from F_o with probability $a/(a+n)$ and is equal to x_j with probability $1/(a+n)$, $j = 1, \ldots, n$.) Define

$$\hat{F}_B^*(t) = \frac{1}{n+a+1} \sum_{i=1}^{n+a+1} I\{X_i^* \leq t\}, \tag{2.2}$$

and evaluate

$$\hat{\theta}_B^* = \theta(\hat{F}_B^*). \tag{2.3}$$

The proposed approximation to G is

$$\hat{G}(t) = \Pr_{*,B}\{\hat{\theta}_B^* \leq t\}, \tag{2.4}$$

which in practice would have to be evaluated as

$$\hat{G}(t) \doteq \frac{1}{BOOT} \sum_{b=1}^{BOOT} I\{\hat{\theta}_B^{*b} \leq t\} \tag{2.5}$$

for a large number BOOT of independent drawings $\hat{\theta}_B^{*b}$ of the type described.

The resulting confidence interval

$$\hat{G}^{-1}(\alpha) \leq \theta(F) \leq \hat{G}^{-1}(1-\alpha) \tag{2.6}$$

could be termed the BB percentile interval.

The description above assumed a to be an integer. If $a = m + \beta$, say, $0 < \beta < 1$ and m an integer, generate $n + m + 2$ X_i^*'s from $\hat{F}_B$ instead, and use

$$\hat{F}_B^*(t) = \frac{1}{n+a+1}\left[\sum_{i=1}^{n+m+1} I\{X_i^* \leq t\} + \beta\, I\{X_{n+m+2}^* \leq t\}\right].$$

The motivation for the BB method is as follows. The two conditional distributions $F|\hat{F}$ and $\hat{F}_B^*|\hat{F}$ are reasonably similar. In fact, judicious calculations give

$$E_B\{F(t)|\hat{F}\} = \hat{F}_B(t),$$

$$E_{*,B}\{\hat{F}_B^*(t)|\hat{F}\} = \hat{F}_B(t),$$

$$\mathrm{Var}_B\{F(t)|\hat{F}\} = \frac{1}{n+a+1}\hat{F}_B(t)\{1 - \hat{F}_B(t)\},$$

$$\mathrm{Var}_{*,B}\{\hat{F}_B^*(t)|\hat{F}\} = \frac{1}{n+a+1}\hat{F}_B(t)\{1 - \hat{F}_B(t)\}.$$

Hence, for well-behaved functionals $\theta = \theta(F)$ we would expect

$$\theta(F)|\hat{F} \doteq \theta(\hat{F}_B^*)|\hat{F}, \tag{2.7}$$

i.e. $G \doteq \hat{G}$.

As a point of further comparison, it may be considered a bit annoying that the skewness of $F(t)|\hat{F}$ is about twice that of $\hat{F}_B^*|\hat{F}$, but they are both small;

$$E_B\{F(t) - \hat{F}_B(t)\}^3|\hat{F} = \frac{2}{(n+a+1)(n+a+2)}\,\hat{F}_B(t)\,\{1 - \hat{F}_B(t)\}\,\{1 - 2\hat{F}_B(t)\},$$

$$E_{*,B}\,\{\hat{F}_B^*(t) - \hat{F}_B(t)\}^3|\hat{F} = \frac{1}{(n+a+1)^2}\,\hat{F}_B(t)\,\{1 - \hat{F}_B(t)\}\,\{1 - 2\hat{F}_B(t)\}.$$

The next section provides an asymptotic justification for the BB.

Remark 1. Consider once more the non-informative case a close to zero (or, rather, a/n close to zero). Then the BB procedure advocates taking bootstrap samples of size n + 1 from the usual $\hat{F}$, as opposed to the traditional size n. This points to the fact that the BB sample size n + a + 1 was chosen merely to make also the second moments of $F|\hat{F}$, $\hat{F}_B^*|\hat{F}$ agree.

Remark 2. Even disregarding the small n versus n + 1 controversy, Rubin's (1981) "simple BB" does not come out of letting $a \to 0$ in the proposed BB of the present paper. Rubin's method smooths the weights, but rigidly sticks to the observed sample points (as does the ordinary bootstrap), cf. (1.7), whereas the more universally applicable method proposed here smooths also outside the data points, using $\hat{F}_B$.

One may call this paper's BB the informative Bayesian bootstrap and Rubin's BB the non-informative Bayesian bootstrap, in order to distinguish them. The remarks above indicate that the present informative version comes much closer to being a proper Bayesian generalisation of Efron's bootstrap.

3. Asymptotic justification.

Assume first, and mostly for illustrational purposes, that the sample space is finite, say $X = \{1,\ldots,L\}$. Let $f_\ell = Pr_F\{X_i = \ell\}$, $\hat{f}_\ell = \#\{x_i = \ell\}/n$, $\hat{f}_{B,\ell} = (af_{o,\ell} + n\hat{f}_\ell)/(a + n)$. Efron (1982, Ch. 5.6) observes

$$\sqrt{n}\,(\hat{f} - f) \xrightarrow{d} N_L(0, \Sigma_f), \tag{3.1}$$

$$\sqrt{n}\,(\hat{f}^* - \hat{f})|\hat{f} \xrightarrow[*]{d} N_L(0, \Sigma_f) \text{ a.s.}, \tag{3.2}$$

where $\hat{f}^*$ stems from the ordinary bootstrap and where Σ_f has elements $f_\ell \delta_{\ell m} - f_\ell f_m$, and discusses why this may be taken as an asymptotic justification for a class of inferential procedures based on the bootstrap. The above results rely only on asymptotic theory for the multinomial distribution.

(3.1), (3.2) can now be accompanied by results for the exact and BB approximated posterior distributions $f|\hat{f}$, $\hat{f}_B^*|\hat{f}$:

$$(n+a+1)^{\frac{1}{2}}\,(f - \hat{f}_B)|\hat{f} \xrightarrow[B]{d} N_L(0, \Sigma_{f_\infty}) \text{ a.s.}, \tag{3.3}$$

$$(n+a+1)^{\frac{1}{2}}\,(\hat{f}_B^* - \hat{f}_B)|\hat{f} \xrightarrow[*,B]{d} N_L(0, \Sigma_{f_\infty}) \text{ a.s.} \tag{3.4}$$

An explanation is needed here: we prefer on this occasion to study the limiting behaviour of $f|\hat{f}$, $\hat{f}_B^*|\hat{f}$ in the ordinary frequentist framework, where the observed frequencies $\hat{f}_\ell$ converge, on a set Ω_o having probability one, to the true ones, say $f_{true,\ell}$. The parameter a may be fixed in (3.3), (3.4), but can also go to infinity with n, as long as

$$\hat{f}_{B,\ell} = \frac{a}{a+n} f_{o,\ell} + \frac{n}{a+n} \hat{f}_\ell \to f_{\infty,\ell} \text{ on } \Omega_o ;$$

f_∞ will be just f_{true} provided $a/n \to 0$.

Now the framework for (3.3), (3.4) is explained. (3.3) follows from asymptotic properties of the Dirichlet distribution, whereas (3.4) is essentially the central limit theorem. Note that exactly the same a.s. set Ω_o is at work in (3.2), (3.3), (3.4).

Efron's discussion of the consequences of (3.1), (3.2) (1979, p. 23; 1982, Ch. 5.6) can now be applied to (3.3), (3.4), and provides the asymptotic justification for the BB procedure.

Remark 3. It is interesting to note that if only $a/\sqrt{n} \to 0$, then $\sqrt{n}\ (f_\ell - \hat{f}_\ell)$ $-\sqrt{n}\ (f_\ell - \hat{f}_{B,\ell})$ goes to zero, and

$$\sqrt{n}\ (f - \hat{f})|\hat{f} \xrightarrow[B]{d} N_L(0, \Sigma_{f_{true}}) \text{ a.s.}, \tag{3.5}$$

$$\sqrt{n}\ (\hat{f}_B^* - \hat{f})|\hat{f} \xrightarrow[*,B]{d} N_L(0, \Sigma_{f_{true}}) \text{ a.s.} \tag{3.6}$$

Accordingly, four different approaches will lead to the same inferential statements, up to first order asymptotics: the classical based on $\hat{f}$; the ordinary Efron bootstrap; the proper posterior Bayes; and the BB.

Now consider the extension of the preceding results and conclusions to $X = R$. The degree to which (3.1), (3.2) and its consequences have analogues for $X = R$ is investigated in Bickel and Freedman (1981) and Singh (1981). The canonical parallel to (3.1) is

$$\sqrt{n}\ \{\hat{F}(.) - F(.)\} \xrightarrow{d} W^o\{F(.)\} \text{ in } D[-\infty, \infty] \tag{3.7}$$

where W^o is a Brownian bridge, see for example Billingsley (1968). Bickel and Freedman (1981) prove the bootstrap companion

$$\sqrt{n}\ \{\hat{F}^*(.) - \hat{F}(.)\}|\hat{F} \xrightarrow{d} W^o\{F(.)\} \text{ a.s.}, \tag{3.8}$$

and conclude that the bootstrap works for well-behaved functionals $\theta = \theta(F)$.

These results can be parallelled in the present Bayesian posterior context. Again, we look at limiting properties in an ordinary framework in which $\hat{F}$ according to the Glivenko-Cantelli theorem converges uniformly to $F = F_{true}$ on a set Ω_o of probability one.

Theorem. Let a vary with n in such a way that $\hat{F}_B = (aF_o + n\hat{F})/(a + n) \to$ some F_∞ on Ω_o. (Most often, F_∞ is just F_{true}.) Then

$$(n+a+1)^{\frac{1}{2}}\ \{F(.) - \hat{F}_B(.)\}|\hat{F} \xrightarrow[B]{d} W^o\{F_\infty(.)\}, \tag{3.9}$$

$$(n+a+1)^{\frac{1}{2}}\ \{\hat{F}_B^*(.) - \hat{F}_B(.)\}|\hat{F} \xrightarrow[*,B]{d} W^o\{F_\infty(.)\}, \tag{3.10}$$

along every sequence in Ω_o.

Proof: The second statement is within reach of (the triangular version of) the classical invariance theorem for i.i.d. random variables. The first statement involves showing finite-dimensional convergence, by looking at Dirichlet distributions, and proving tightness, which follows by the proof of Billingsley's (1968) Theorem 15.6, upon noticing that

$$(n+a+1)^2 \, E\,\{F(s,\,t] - \hat{F}_B(s,\,t]\}^2\,\{F(t,\,u] - \hat{F}_B(t,\,u]\}^2 \,|\, \hat{F}$$
$$\leq \hat{F}_B(s,\,u]^2, \quad s \leq t \leq u. \;\square$$

Thus the conditional distributions $\theta(F)\,|\,\hat{F}$ and $\theta(\hat{F}_B^{\,*})\,|\,\hat{F}$ will be close to each other for well-behaved functionals, justifying the BB method. Particular examples can be worked through, as in Bickel and Freedman (1981). Their tentative description of well-behavedness (p. 1209) can also be subscribed to here. Sufficient conditions for

$$(n+a+1)^{\frac{1}{2}}\,\{\theta(F) - \theta(\hat{F}_B)\}\,|\,\hat{F} \xrightarrow[B]{d} N(0,\,\sigma^2) \text{ a.s.},$$
$$(n+a+1)^{\frac{1}{2}}\,\{\theta(\hat{F}_B^{\,*}) - \theta(\hat{F}_B)\}\,|\,\hat{F} \xrightarrow[*,B]{d} N(0,\,\sigma^2) \text{ a.s.}$$

to hold, for appropriate variance σ^2, can be written down, using von Mises methods, for example as in Boos and Serfling (1980), Parr (1985), who use Fréchet differentiability, or the more universally applicable machinery of Hadamard or compact differentiability, as in Reeds (1976) and Fernholz (1983).

Remark 4. If a is fixed, or only $a/\sqrt{n} \to 0$, then

$$\sqrt{n}\,\{F(.) - \hat{F}(.)\}\,|\,\hat{F} \xrightarrow{d} W^o\{F_{true}(.)\} \text{ a.s.}, \tag{3.11}$$
$$\sqrt{n}\,\{\hat{F}_B^{\,*}(.) - \hat{F}(.)\}\,|\,\hat{F} \xrightarrow[*,B]{d} W^o\{F_{true}(.)\} \text{ a.s.} \tag{3.12}$$

A conclusion concerning the approximate agreement among the four statisticians referred to in the previous remark thus can be reached also for $X = R$ (and for more general spaces).

Remark 5. The functional $\theta = \theta(F)$ can depend upon n; the described BB procedure works specifically for the given n. θ is also allowed to depend upon the actual data sample, say $\theta = \theta(F, X_1,\ldots,X_n)$ (in contrast to simpler ones like $\theta(F) = \int x dF(x)$, $\sigma(F) = [\int\{x - \theta(F)\}^2 dF(x)]^{\frac{1}{2}}$). An example is $\theta = \sup_t |F(t) - \hat{F}(t)|$, of importance in connection with confidence bands. Our BB method is general enough to handle such cases too.

Let us illustrate this comment with a description of how a nonparametric Bayesian might construct a simultaneous confidence band for F. Consider $\theta = \sup_{a\le t\le b} |F(t) - \hat{F}_B(t)|/[\hat{F}_B(t)\{1 - \hat{F}_B(t)\}]^{\frac{1}{2}}$. The natural band is

$$\hat{F}_B(t) - t_{1-\alpha}[\hat{F}_B(t)\{1 - \hat{F}_B(t)\}]^{\frac{1}{2}} \le F(t) \le \hat{F}_B(t) + t_{1-\alpha}[\hat{F}_B(t)\{1 - \hat{F}_B(t)\}]^{\frac{1}{2}}$$

for $a \le t \le b$, where $t_{1-\alpha}$ ideally would be determined by $Pr_B\{\theta(F, X_1,\ldots,X_n) \le t_{1-\alpha}|\hat{F}\} = 1 - \alpha$. This $t_{1-\alpha}$ would be almost impossible to find. The BB method consists of generating perhaps 1000 values $\hat{\theta}_B^{*b}$ of $\hat{\theta}_B^* = \sup_{a\le t\le b} |\hat{F}_B^*(t) - \hat{F}_B(t)| /[\hat{F}_B(t)\{1 - \hat{F}_B(t)\}]^{\frac{1}{2}}$, and using $\hat{t}_{1-\alpha}$ = empirical upper α-point for these 1000 realisations instead of $t_{1-\alpha}$. One may prove that $(n+a+1)^{\frac{1}{2}}(\hat{t}_{1-\alpha} - t_{1-\alpha}) \to 0$ a.s. in this case. (Strictly speaking, this is true provided $BOOT_n$ realisations are generated instead of 1000 and $BOOT_n/(n \log n)$ grows to infinity.)

Remark 6. The bootstrap sample size BOOT in (2.5) should of course be large in order for $\hat{G}_{BOOT}^{-1}(1-\alpha)$, $\hat{G}_{BOOT}^{-1}(\alpha)$ to come close to $\hat{G}^{-1}(1-\alpha)$, $\hat{G}^{-1}(\alpha)$. The investigation of Efron (1985, Section 8) indicates that BOOT = 1000 may be a rough minimum.

Remark 7. The exposition of the BB method has so far emphasized its use to construct confidence intervals. There are other uses for (an approximation to) the posterior distribution, however, a major example being the evaluation of the usual Bayes estimate (under quadratic loss), $\bar{\theta}_B = \int t dG(t)$. Closed form

solutions are only available for special cases. An approximation is now possible,

$$\theta_B \doteq \int t d\hat{G}(t) \doteq \frac{1}{BOOT} \sum_{b=1}^{BOOT} \hat{\theta}_B^{*b}.$$

BOOT need in this case not be as large as 1000 to make the second approximation a good one, BOOT = 100 may be sufficient, cf. Efron (1985, Section 8).

Remark 8. The starting point for our quest for the construction of confidence intervals has been (1.6). Sometimes highest posterior regions are advocated instead, see for example Box and Tiao (1973). In the present case this would involve approximating the posterior distribution G with one with a density g(t), and then letting $\{t: g(t) \geq g_o\}$ be the confidence region, for appropriate level g_o. This approach makes most sense when g is unimodal, which it would not be in many important cases here, due to the fact that the posterior distribution of F places extra weight on the observed data points. This is illustrated in Section 5 for the case of the median.

Remark 9. The BB method is easily generalised to for example two sample situations. To illustrate, let $X_1,\ldots,X_n$ and $Y_1,\ldots,Y_m$ be samples from respectively F_1 and F_2, and assume $\theta(F_1, F_2) = F_1^{-1}(\frac{1}{2}) - F_2^{-1}(\frac{1}{2})$ is of interest. A Bayes estimate and a confidence interval for this difference of population medians can be obtained by generating perhaps 1000 realisations of $\hat{\theta}_B^* = \text{median}\{X_1^*,\ldots, X_{n+a+1}^*\} - \text{median}\{Y_1^*,\ldots,Y_{m+b+1}^*\}$, where the X_i^*'s are drawn from $(aF_{1,o} + n\hat{F}_1)/(a+n)$ and the Y_i^*'s from $(bF_{2,o} + m\hat{F}_2)/(b+m)$, and then treating the resulting histogram as the posterior distribution of θ.

Remark 10. It is perhaps surprising that a simple method like the BB, constructed merely to make the first and second moments of the exact and the approximate distributions of $F\{A\}$ agree for each $A \subset X$ can work well for the vast majority of parameter functionals. As indicated in Section 3, this is at least

partly the work and the magic of the central limit theorem. This also points to the possibility of using "small-sample asymptotics" machinery to arrive at other approximations to the posterior distribution G, for example Edgeworth-Cramér expansions combined with Taylor expansions. Such an approach would be functional-dependent, however; a primary virtue of the BB is that it is both simple and versatile. A similar remark of course applies to the usual bootstrap.

Remark 11. The BB has been constructed for situations where the statistician is willing to approximate the prior uncertainty about the unknown F with a Dirichlet process, involving only the prior guess F_o and the "prior sample size" parameter a. One can conceivably construct similar bootstrap-like devices for more complex prior distributions, like mixtures of Dirichlet processes and neutral to the right processes. This would be valuable, considering the difficulty with which even Bayes point estimates are evaluated in such situations.

Remark 12. A critisism sometimes voiced against the ordinary nonparametric bootstrap is that it too rigidly sticks to the observed data points. The Bayesian bootstrap proposed in this paper provides a generalisation of the ordinary bootstrap, as indicated in Remarks 1 and 2, towards having the possibility of smoothing also outside the data, in a reasonable and non-ad hoc way (unless one discards Bayesian statistics in general as being too ad hoc). Let us point out that hybrids can be invented. One may take the "strength of belief" constant a, which in principle is user-defined, to be just an unknown parameter instead, and estimate it based on the data. A large estimated a should result if the data fits F_o to a high degree; if the data seriously contradicts F_o then the estimated a should be close to zero. A further de-Bayesification of the Bayesian bootstrap could allow unknown parameters in F_o too, placing matters in an empirical Bayes framework, and estimate these too from the data.

4. A bias corrected BB percentile interval.

The BB percentile interval cannot be bias-corrected the way this is done for the ordinary bootstrap case, as presented e.g. in Efron (1982, Ch. 10). In fact, one can argue that the "bias" is already taken care of. If there exists a smooth increasing transformation h such that

$$h\{\theta(F)\} - h(\hat{\theta}_{obs}) \overset{\cdot}{\sim}_{B} N(z_o\sigma, \sigma^2), \tag{4.1}$$

$$h\{\theta(\hat{F}_B^*)\} - h(\hat{\theta}_{obs}) \overset{\cdot}{\sim}_{*,B} N(z_o\sigma, \sigma^2) \tag{4.2}$$

for some constants z_o, σ, putting $\hat{\theta}_{obs} = \theta(\hat{F}_B)$ for short (these assumptions appear reasonable in view of the preceding section), then one may deduce $z_o \doteq -\Phi^{-1}\{\hat{G}(\hat{\theta}_{obs})\}$ as in the cited reference, but the bias corrected interval on the $h(\theta)$ scale,

$$h(\hat{\theta}_{obs}) + z_o\sigma - z^{(1-\alpha)}\sigma \leq h(\theta) \leq h(\hat{\theta}_{obs}) + z_o\sigma + z^{(1-\alpha)}\sigma,$$

where $z^{(1-\alpha)} = \Phi^{-1}(1-\alpha)$ is the upper α-point for the standard normal, happens to tranform back again to (2.6) again on the θ scale:

$$\begin{aligned}\hat{G}[h^{-1}\{h(\hat{\theta}_{obs}) + z_o\sigma + z^{(1-\alpha)}\sigma\}] &= Pr_{*,B}\{h(\hat{\theta}_B^*) \leq h(\hat{\theta}_{obs}) + z_o\sigma + z^{(1-\alpha)}\sigma\} \\ &\doteq Pr\{N(z_o, 1) \leq z_o + z^{(1-\alpha)}\} = 1 - \alpha.\end{aligned}$$

Even if the above approach had led to something non-trivial the result would not have been as trustworthy as Efron's bias corrected percentile interval is, comparatively speaking. The comments about skewness following (2.7) would imply that the implicit assumption in (4.1), (4.2), namely that z_o can be taken to be the same in the two situations, hardly could be trusted, this in contrast to the ordinary bootstrap framework in which it is known that the bootstrap approximation usually is good also to the "next order". It may however be possible to deduce a simple and likely relationship between the two z_o's in (the rephrased) (4.1), (4.2), and then correct the

BB percentile interval based on this.

There is another possibility of detecting and repairing a bias, however. For each in a respectable catalogue of examples there is a known transformation h, perhaps the identity, such that the posterior expectation of $h\{\theta(F)\}$ is explicitly calculable by some published formula, i.e.

$$\nu_o = E_B[h\{\theta(F)\}|\hat{F}]$$

is known. The BB procedure considers

$$\hat{H}(t) = Pr_{*,B}[h\{\theta(\hat{F}_B^*)\} \leq t|\hat{F}] = \hat{G}\{h^{-1}(t)\}$$

and approximates ν_o with

$$\hat{\nu}_o = \int t d\hat{H}(t) = \frac{1}{BOOT}\sum_{b=1}^{BOOT} h\{\theta(\hat{F}_B^{*b})\} = \nu_o + \varepsilon, \tag{4.3}$$

say. Accordingly, if $\varepsilon \neq 0$, then $\hat{H}$ is not a perfect estimate of H, the c.d.f. of $h\{\theta(F)\}|\hat{F}$. $\hat{H}_\varepsilon(t) = \hat{H}(t + \varepsilon)$ is a new estimate, this time getting the mean right. Hence

$$\hat{H}_\varepsilon^{-1}(\alpha) = \hat{H}^{-1}(\alpha) - \varepsilon \leq h\{\theta(F)\} \leq \hat{H}^{-1}(1 - \alpha) - \varepsilon = \hat{H}_\varepsilon^{-1}(1 - \alpha)$$

would be a natural corrected confidence interval for $h\{\theta(F)\}$. Transforming back we obtain

$$h^{-1}[h\{\hat{G}^{-1}(\alpha)\} - \varepsilon] \leq \theta(F) \leq h^{-1}[h\{\hat{G}^{-1}(1 - \alpha)\} - \varepsilon] \tag{4.4}$$

as the bias corrected BB percentile interval for $\theta(F)$. Of course this interval is just (2.6) if $\varepsilon = 0$ above.

One can also write down a slightly more general bias and variance corrected BB percentile interval which also takes into account the value of $\tau_o^2 = Var_B[h\{\theta(F)\}|\hat{F}]$ if it is available. Assume that, in addition to (4.3),

$$\hat{\tau}_o^2 = \int (t - \hat{\nu}_o)^2 d\hat{H}(t) = \frac{1}{BOOT}\sum_{b=1}^{BOOT} \{h(\hat{\theta}_B^{*b}) - \hat{\nu}_o\}^2 = \tau_o^2(1 + \delta)^2.$$

A perhaps better estimate of $G\{h^{-1}(t)\}$ is then $\hat{H}_{\varepsilon,\delta}(t) = \hat{H}\{t(1+\delta) + \varepsilon - \nu_o\delta\}$, since it gets both the mean and the variance right. Using $\hat{H}_{\varepsilon;\delta}^{-1}(p) =$

$\{\hat{H}^{-1}(p) + \nu_o\delta - \varepsilon\}/(1+\delta)$ one ends up with

$$h^{-1}[\frac{h\{\hat{G}^{-1}(\alpha)\}}{1+\delta} + \frac{\nu_o\delta - \varepsilon}{1+\delta}] \leq \theta(F) \leq h^{-1}[\frac{h\{\hat{G}^{-1}(1-\alpha)\}}{1+\delta} + \frac{\nu_o\delta - \varepsilon}{1+\delta}]. \qquad (4.5)$$

5. Some examples.

This section looks briefly into the nature of the BB approximation method, and compares confidence intervals arising from different prior distributions in an artificial example.

5.1. A probability.

If $\theta(F) = P(A)$ for some set A of interest, then

$$\theta(F)\,|\,\hat{F} \underset{B}{\sim} \text{Beta}\{aF_o(A) + \#(x_i \in A),\ aF_o(A^c) + \#(x_i \notin A)\}.$$

Thus (1.5) and (1.6) can be obtained from tables of the incomplete Beta function. In this case the BB method amounts to approximating the Beta distribution G above with that of $Y/(n+a+1)$, Y being binomial $[n+a+1,\ \{aF_o(A) + \#(x_i \in A)\}/(a+n)]$.

If U is Beta $\{mp,\ m(1-p)\}$ and V is Bin $\{m+1,p\}/(m+1)$, then $EU = EV = p$, $\text{Var } U = \text{Var } V = \frac{1}{m+1}\,p(1-p)$. U and V differ in skewness and kurtosis, but not to any dramatic extent:

$$\text{skew (U)} = 2\,\frac{(m+1)^{\frac{1}{2}}}{m+2}\,\frac{1-2p}{\{p(1-p)\}^{\frac{1}{2}}},$$

$$\text{skew (V)} = \frac{1}{(m+1)^{\frac{1}{2}}}\,\frac{1-2p}{\{p(1-p)\}^{\frac{1}{2}}};$$

$$\text{kurtosis (U)} = \frac{6m}{(m+2)(m+3)}\{1 - \frac{1+1/m}{p(1-p)}\},$$

$$\text{kurtosis (V)} = \frac{1}{m+1}\,\frac{1-4p(1-p)}{p(1-p)}.$$

Brief investigations have shown the distributions of U and V, and therefore confidence intervals based on either the exact or BB approximated distributions, to be remarkably similar even for moderate m, provided p is not too

close to zero or one, provided α is not too close to zero, and finally provided the discrete distribution of V is interpolated. Rather than using $\hat{G}(t) = \Pr[\text{Bin}\{m+1,p\}/(m+1) \le t]$, which jumps at the points $j/(m+1)$, we use $\tilde{G}\{j/(m+1)\} = \frac{1}{2}\Pr\{V \le j/(m+1)\} + \frac{1}{2}\Pr\{V \le (j-1)/(m+1)\}$ (which is what one gets if one interprets $\Pr\{V \le j/(m+1)\}$ as $\Pr\{V + \varepsilon \le j/(m+1)\}$ where ε is independent of V and say normal $(0, 10^{-6})$) and interpolate linearly in-between. Similar modifications to $\hat{G}$ are advocated in other cases where $\hat{G}$ increases in sharp jumps only.

5.2. The median.

The p-quantile functional is another example where it is possible to compute the posterior distribution explicitly, but the resulting expressions are complex, and the BB would be much easier to carry out in practice. For simplicity only the median $\theta(F) = F^{-1}(\frac{1}{2}) = \inf\{t: F(t) \ge \frac{1}{2}\}$ is considered below.

Assume for concreteness that the data points are distinct, say $x_1 < \ldots < x_n$. We shall find $G(t) = \Pr_B\{\theta(F) \le t | \hat{F}\}$.

First look at a data point x_j. Then

$$\begin{aligned} G\{x_j\} &= \Pr_B\{\theta(F) = x_j | \hat{F}\} \\ &= \Pr_B\{F(-\infty,x_j) < \tfrac{1}{2},\ F(-\infty,x_j] \ge \tfrac{1}{2} | \hat{F}\} \\ &= \Pr\{U < \tfrac{1}{2},\ U + V \ge \tfrac{1}{2}\} = \Pr\{U < \tfrac{1}{2},\ W < \tfrac{1}{2}\}, \end{aligned}$$

where (U, V, W) is Dirichlet (α, β, γ); $\alpha = aF_o(x_j-) + j - 1$, $\beta = aF_o\{x_j\} + 1$, $\gamma = aF_o(x_j,\infty) + n - j$. Assuming the prior guess c.d.f. to be continuous one gets

$$\begin{aligned} G\{x_j\} &= \frac{\Gamma(\alpha+\beta+\gamma)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\gamma)} \frac{1}{\alpha} (\tfrac{1}{2})^{\alpha} \frac{1}{\gamma} (\tfrac{1}{2})^{\gamma} \\ &= \frac{\Gamma(a+n)}{\Gamma(aF_o(x_j)+j)\Gamma(a\{1-F_o(x_j)\}+n-j+1)} (\tfrac{1}{2})^{a+n-1}. \end{aligned} \tag{5.1}$$

Next consider $G[t, t+dt]$ for some t outside $\{x_1,\ldots,x_n\}$, and let for further convenience F_o be the integral of a prior guess density f_o. Following the

reasoning above one may show that G has a density at $t \in (x_j, x_{j+1})$ equal to

$$g(t) = \frac{\Gamma(a+n)}{\Gamma(aF_o(t)+j)\Gamma(a\{1-F_o(t)\}+n-j)} \, af_o(t) \, J[aF_o(t)+j, \, a\{1-F_o(t)\}+n-j], \tag{5.2}$$

where

$$J[\alpha, \gamma] = \int_0^{\frac{1}{2}} \int_0^{\frac{1}{2}} u^{\alpha-1} w^{\gamma-1} (1-u-w)^{-1} \, du \, dw$$

$$= (\tfrac{1}{2})^{\alpha+\gamma} \sum_{m=0}^{\infty} \sum_{i=0}^{m} \binom{m}{i} (\tfrac{1}{2})^m \frac{1}{\alpha+i} \frac{1}{\gamma+m-i}.$$

It is (in principle) possible to compute for example the posterior expectation and the lower and upper 5 percent points for the distribution G based on the results above.

Now consider the BB approximation method in this situation. It approximates the complicated G above with

$$\hat{G}(t) = \Pr_{*,B}\{\hat{\theta}_B^* = \theta(\hat{F}_B^*) \le t\}$$

$$= \Pr_{*,B}\{\text{median}\{X_1^*, \ldots, X_{n+a+1}^*\} \le t\}, \tag{5.3}$$

where the X_i^*'s are i.i.d. from $\hat{F}_B = (aF_o + n\hat{F})/(a+n)$. Assume for simplicity that n+a+1 is an odd integer, say 2m + 1. Then

$$\hat{G}(t) = \Pr_{*,B}\{X_{(m+1)}^* \le t\}$$

$$= \Pr[\text{Bin}\{2m+1, \, \hat{F}_B(t)\} \ge m+1]. \tag{5.4}$$

Expressions for $\hat{G}\{x_j\}$ and for the density $\hat{g}$ that $\hat{G}$ has between data points can be worked out based on this, and they can be compared with G and g obtained above. Such a study is not pursued here. Notice that the endpoints of the BB confidence interval can be found using binomial tables.

Note also that in the non-informative case, where a tends to zero, both G and $\hat{G}$ are supported on the data points, with

$$G\{x_j\} = \binom{n-1}{j-1} (\tfrac{1}{2})^{n-1},$$

$$\hat{G}\{x_j\} = \frac{(n+1)!}{(\frac{1}{2}n)!(\frac{1}{2}n)!} \left(\frac{j-1}{n}\right)^{\frac{1}{2}n} \frac{1}{n} \left(\frac{n-j}{n}\right)^{\frac{1}{2}n}.$$

5.3. The endpoint of a distribution.

As the final explicit example, consider

$$\theta(F) = \sup \{t: F(t) < 1\}.$$

The exact posterior distribution of θ based on the Dirichlet process assumption is particularly simple in this case. One has $\theta(F) \leq t$ if and only if $F(t) = 1$, and $F(t)$ is Beta $\{aF_o(t) + n\hat{F}(t), a\{1 - F_o(t)\} + n\hat{F}(t,\infty)\}$. The probability of a Beta $\{\alpha, \beta\}$ variable being 1 is zero unless $\beta = 0$, in which case the probability suddenly is one. Thus $G(t)$ is zero if F_o or $\hat{F}$ have some mass left for (t,∞), and is one only if $F_o(t) = 1$ and no x_i's are greater than t, i.e. G is simply concentrated at the single point $\max\{\theta(F_o), \max_{i\leq n} x_i\}$.

Now let us watch BB at work. $\hat{\theta}_B^* = \theta(\hat{F}_B^*) = \max\{X_1^*,\ldots,X_{n+a+1}^*\}$ has distribution

$$\begin{aligned}\hat{G}(t) &= \Pr_{*,B}\{\text{every } X_i^* \leq t\} \\ &= \hat{F}_B(t)^{n+a+1}.\end{aligned}$$

If the right endpoint of the prior guess is ∞, then $G\{\infty\} = 1$, i.e. the a posteriori opinion is that the right endpoint of the unknown F is indeed ∞. This can be compared to $\hat{G}(t) = \{\frac{a}{a+n} F_o(t) + \frac{n}{a+n}\}^{n+a+1}$ for $t \geq \max_{i\leq n} x_i$. If on the other hand the right endpoint of the prior guess is finite, say $\theta(F_o) = 100$, then the exact a posteriori distribution is concentrated at the sensible point $\max\{100, \max_{i\leq n} x_i\}$, and the BB approximation does not seriously disagree:

$$\begin{aligned}\hat{G}(t) &= \{\tfrac{a}{a+n} F_o(t) + \tfrac{n}{a+n}\}^{n+a+1} && \text{, if } \max_{i\leq n} x_i \leq t < 100, \\ &= 1 && \text{, if } t \geq 100 \text{ and } \max_{i\leq n} x_i \leq 100, \\ &= \{\tfrac{a}{a+n} + \tfrac{n}{a+n} \hat{F}(t)\}^{n+a+1} && \text{, if } 100 \leq t < \max_{i\leq n} x_i, \\ &= 1 && \text{, if } 100 \leq \max_{i\leq n} x_i \leq t.\end{aligned}$$

5.4 An artificial example.

The following example is indeed constructed but hopefully illustrative. Assume that the abilities and intelligence of a certain interesting minority population is studied, and assume that 20 individuals from this population are given a standard IQ test. Let us suppose that the resulting $X_1,\ldots,X_{20}$ really come from a normal $(125, 10^2)$ distribution; the data points $x_1,\ldots,x_{20}$ used below were simulated from this distribution. Some parameters of interest could be

$$\theta(F) = \text{mean of } F = \int x dF(x),$$

$$\sigma(F) = \text{standard deviation of } F = [\int\{x - \theta(F)\}^2 dF(x)]^{\frac{1}{2}},$$

$$\gamma(F) = \text{upper 25 percent point of } F = F^{-1}(3/4).$$

We include four different prior distributions in this modest experiment, namely (i) F_o = normal $(100, 15^2)$ (from which our own IQs presumably once were drawn), a = 2; (ii) F_o = normal $(100, 15^2)$, a = 6; (iii) F_o = normal $(100, 15^2)$, a = 10; (iv) F_o = uniform on [100, 150], a = 6. The data points turned out to be 94.8, 114.7, 115.5, 117.5, 119.1, 121.2, 122.6, 124.0, 124.6, 128.7, 130.0, 130.5, 130.8, 131.4, 132.2, 133.5, 134.4, 135.3, 136.4, 144.6. The empirical mean is 126.07 and the empirical standard deviation is 10.70.

The mean. Histograms of BOOT = 1000 BB values of $\hat{\theta}_B^*$ are shown in Figure 1 (i), (ii), (iii), (iv), corresponding to the four combinations of (a, F_o) listed above. For example, the 1000 values leading to Figure 1 (i) are of the type $\hat{\theta}_B^* = \theta(\hat{F}_B^*)$ = mean $\{X_1^*,\ldots,X_{23}^*\}$ where $X_1^*,\ldots,X_{23}^*$ are i.i.d. from $\frac{2}{22} F_o + \frac{20}{22} \hat{F}$. The distributions are unimodal and fairly symmetric, and for all practical purposes continuous ($\hat{\theta}_B^*$ has microscopic point masses $1.33\ 10^{-31}$ in each of the 20 data points in Experiment (i), for instance).

Table 1 lists 95 percent, 90 percent, and 80 percent confidence intervals for the unknown mean $\theta(F)$, reached by Bayesians (i), (ii), (iii), (iv). Also listed is the median of the BB posterior distribution, which is also a good approximation to the Bayesian point estimate using absolute loss. Note that Bayesian (i), with a = 2, behaves almost like the standard bootstrap user.

TABLE 1

Confidence intervals for the mean, reached by Bayesians (i), (ii), (iii), (iv).

	Bayesian (i)	(ii)	(iii)	(iv)
95 percent	[117.74, 129.02]	[113.33, 125.60]	[110.53, 123.10]	[121.41, 130.01]
90 percent	[118.71, 128.19]	[114.22, 124.93]	[111.55, 122.13]	[121.94, 129.33]
80 percent	[119.98, 127.12]	[115.69, 123.90]	[113.10, 121.24]	[122.83, 128.69]
median	123.68	120.07	117.37	125.74

The mean functional is simple enough to compute moments of its true posterior distribution. Ferguson (1973) has shown

$$E_B\{\theta(F)\,|\,\text{data}\} = \frac{a}{a+n}\,\theta(F_o) + \frac{n}{a+n}\,\theta(\hat{F}), \tag{5.5}$$

and methods from the same paper can be used to obtain

$$\text{Var}_B\{\theta(F)\,|\,\text{data}\} = \frac{1}{n+a+1}\left[\frac{a}{a+n}\,\sigma^2(F_o) + \frac{n}{a+n}\,\sigma^2(\hat{F}) + \frac{a}{a+n}\,\frac{n}{a+n}\,\{\theta(F_o) - \theta(\hat{F})\}^2\right]. \tag{5.6}$$

(Here $\theta(\hat{F}) = \bar{x} = \Sigma_{i=1}^{n} x_i/n$ and $\sigma^2(\hat{F}) = \frac{1}{n}\,\Sigma_{i=1}^{n}(x_i - \bar{x})^2$.) Hence we are in a position to correct the BB intervals of Table 1 for bias and variance, using (4.4) and (4.5).

These changes turn out to be small, and perhaps insignificant compared to the variation resulting from differences in prior opinion from Bayesian to Bayesian. As a random example, consider 90 percent intervals for θ in Experiment (ii).

The uncorrected BB percentile interval is $[\hat{G}^{-1}(.05), \hat{G}^{-1}(.95)] = [114.22, 124.93]$ as in Table 1. (5.5) above gives $\nu_o = E_B\{\theta(F)\,|\,\text{data}\} = \frac{6}{26}\,100 + \frac{20}{26}\,\bar{x} = 120.0573$, whereas the BB approximation gives $\hat{\nu}_o = E_{*,B}\{\hat{\theta}_B^*\} \doteq (1/1000)\,\Sigma_{b=1}^{1000}\,\hat{\theta}_B^{*b} = 119.8952$, i.e. $\varepsilon = -.1621$, cf. (4.3). The bias corrected BB interval is accordingly [114.06, 124.77]. Next, $\tau_o^2 = \text{Var}_B\{\theta(F)\,|\,\text{data}\} = 9.6501$ using (5.6), whereas $\hat{\tau}_o^2 = 10.2493$, i.e. $1 + \delta = 1.0306$ in the notation of Section 4. The bias and variance corrected interval (4.5) becomes [114.55, 124.94].

The standard deviation. Histograms of BOOT = 1000 values of $\hat{\sigma}_B^*$ = standard deviation of $X_1^*,\ldots,X_{20+a+1}^*$ are shown in Figure 2 (i), (ii), (iii), (iv), corresponding again to the four experiments. The distributions are again unimodal and practically continuous, but not symmetric.

The bias correction procedure of Section 4 turns out to be more important for this functional than it was for the mean. One may prove, again using methods of Ferguson (1973), that

$$E_B\{\sigma^2(F)\,|\,\text{data}\} = \frac{n+a}{n+a+1}\left[\frac{a}{n+a}\,\sigma^2(F_o) + \frac{n}{a+n}\,\sigma^2(\hat{F}) + \frac{a}{n+a}\,\frac{n}{n+a}\,\{\theta(F_o) - \theta(\hat{F})\}^2\right],$$

cf. (5.6). The $1 - 2\alpha$ bias corrected confidence interval for $\sigma(F)$ is therefore

$$\{\hat{G}^{-1}(\alpha)^2 - \varepsilon\}^{\frac{1}{2}} \le \sigma(F) \le \{\hat{G}^{-1}(1-\alpha)^2 - \varepsilon\}^{\frac{1}{2}}, \tag{5.7}$$

where ε is the difference between the average value of the observed $(\hat{\sigma}_B^*)^2$'s and $E_B\{\sigma^2(F)\,|\,\text{data}\}$, as in (4.3). Notice the similarity between $E_B\{\sigma^2(F)\,|\,\text{data}\}$ and $\text{Var}_B\{\theta(F)\,|\,\text{data}\}$.

Table 2 gives confidence intervals for $\sigma(F)$ in the four experiments, both the uncorrected BB and the bias corrected BB interval (5.7). Also listed are the posterior medians, to be thought of as Bayes point estimates of $\sigma(F)$ (under absolute loss).

TABLE 2

Confidence intervals for the standard deviation in Experiments (i), (ii), (iii), (iv); usual BB interval (upper line) and bias corrected BB interval (lower line).

	Bayesian (i)	(ii)	(iii)	(iv)
95 percent	[7.82, 19.18]	[10.94, 21.00]	[13.02, 21.44]	[7.57, 14.85]
	[7.10, 18.90]	[10.20, 20.62]	[12.43, 21.09]	[7.02, 14.58]
90 percent	[8.33, 17.90]	[11.83, 20.31]	[13.71, 20.77]	[8.20, 14.47]
	[7.65, 17.60]	[11.15, 19.92]	[13.15, 20.40]	[7.69, 14.19]
80 percent	[9.45, 16.77]	[12.61, 19.27]	[14.48, 20.20]	[9.03, 13.79]
	[8.86, 16.44]	[11.96, 18.86]	[13.95, 19.82]	[8.57, 13.49]
median	13.02	15.92	17.34	11.57

It is also possible to correct the intervals further for possible inaccuracy of the BB approximation to the variance of $\sigma^2(F)$ given data; this would involve a quite lengthy formula for $E_B\{\sigma^4(F)|\text{data}\} - [E_B\{\sigma^2(F)|\text{data}\}]^2$, however, and is not pursued here.

The upper quartile. The histograms in Figure 3 (i), (ii), (iii), (iv) come from BOOT = 1000 values of $\hat{\gamma}_B^* = \gamma(\hat{F}_B^*)^{-1}(3/4)$. The sample sizes of the BB samples in question are 23, 27, 31, 27, so we take $\hat{\gamma}_B^*$ to be respectively order statistic 18, order statistic 21, order statistic 24, and order statistic 21. The distribution of $\hat{\gamma}_B^*$ is different from that of $\hat{\theta}_B^*$ and $\hat{\sigma}_B^*$ in that it has most of its mass concentrated on the observed data points, cf. theoretical calculations for the median in 5.2.

Table 3 lists 95, 90, and 80 percent BB confidence intervals for $\gamma(F)$ for the four choices of prior distribution of F in its space of all distributions.

TABLE 3

Confidence intervals for the upper quartile in Experiments (i), (ii), (iii), (iv).

95 percent	[130.0, 135.3]	[126.0, 135.3]	[124.0, 134.4]	[130.5, 136.4]
90 percent	[130.5, 135.3]	[128.7, 134.4]	[124.9, 134.4]	[130.8, 136.4]
80 percent	[130.8, 134.4]	[130.0, 134.4]	[128.7, 133.5]	[131.4, 135.4]
median	132.2	131.4	130.8	133.5

Acknowledgements. This paper has been written during my stay at Stanford University with grants from the Norwegian Computer Centre and the Royal Norwegian Council for Scientific and Industrial Research. I have had the privilege to get a detailed list of comments to an earlier draft from Persi Diaconis. Thanks are also due to Mark Knowles for computer programming assistance and to Bradley Efron and Paul Switzer for comments and encouragement.

References.

Bickel, P. and Freedman, D. (1981). Some asymptotic theory for the bootstrap. Ann. Statist. 9, 1196-1217.

Billingsley, P. (1968). Convergence of probability measures. Wiley, New York.

Boos, D.D. and Serfling, R.J. (1980). A note on differentials and the CLT and LIL for statistical functions with application to M-estimates. Ann. Statist. 8, 618-624.

Box, G.E.P. and Tiao, G.C. (1973). Bayesian inference in statistical analysis. Reading, Mass. Addison-Wesley.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. Ann. Statist. 7, 1-26.

Efron, B. (1982). The jackknife, the bootstrap and other resampling plans. SIAM, Philadelphia.

Efron, B. (1985). Better bootstrap confidence intervals. Tech. Report #226, Stanford University, Department of Statistics.

Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. Ann. Statist. 1, 209-230.

Ferguson, T.S. (1974). Prior distributions on spaces of probability measures. Ann. Statist. 2, 615-629.

Fernholz, L.T. (1983). von Mises calculus for statistical functionals. Lecture notes in statistics, Springer, New York.

Hannum, R., Hollander, M., and Langberg, N. (1981). Distributional results for random functionals of a Dirichlet process. Ann. Probability 9, 665-670.

Hjort, N.L. (1985). An invariance theorem for Dirichlet processes. Manuscript.

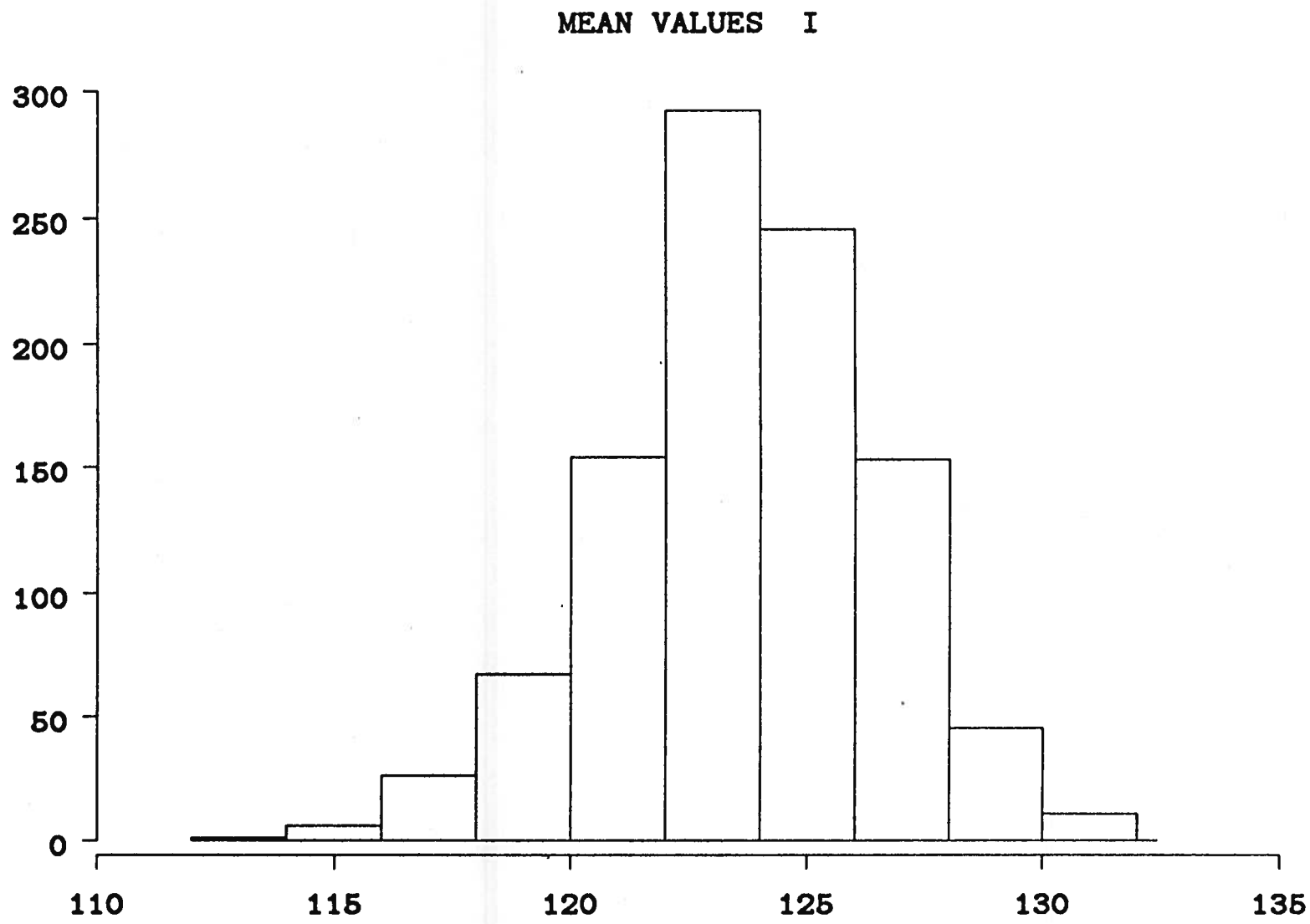
Parr, W.C. (1985). The bootstrap: some large sample theory, and connections with robustness. Statistics & Probability Letters 3, 97-100.

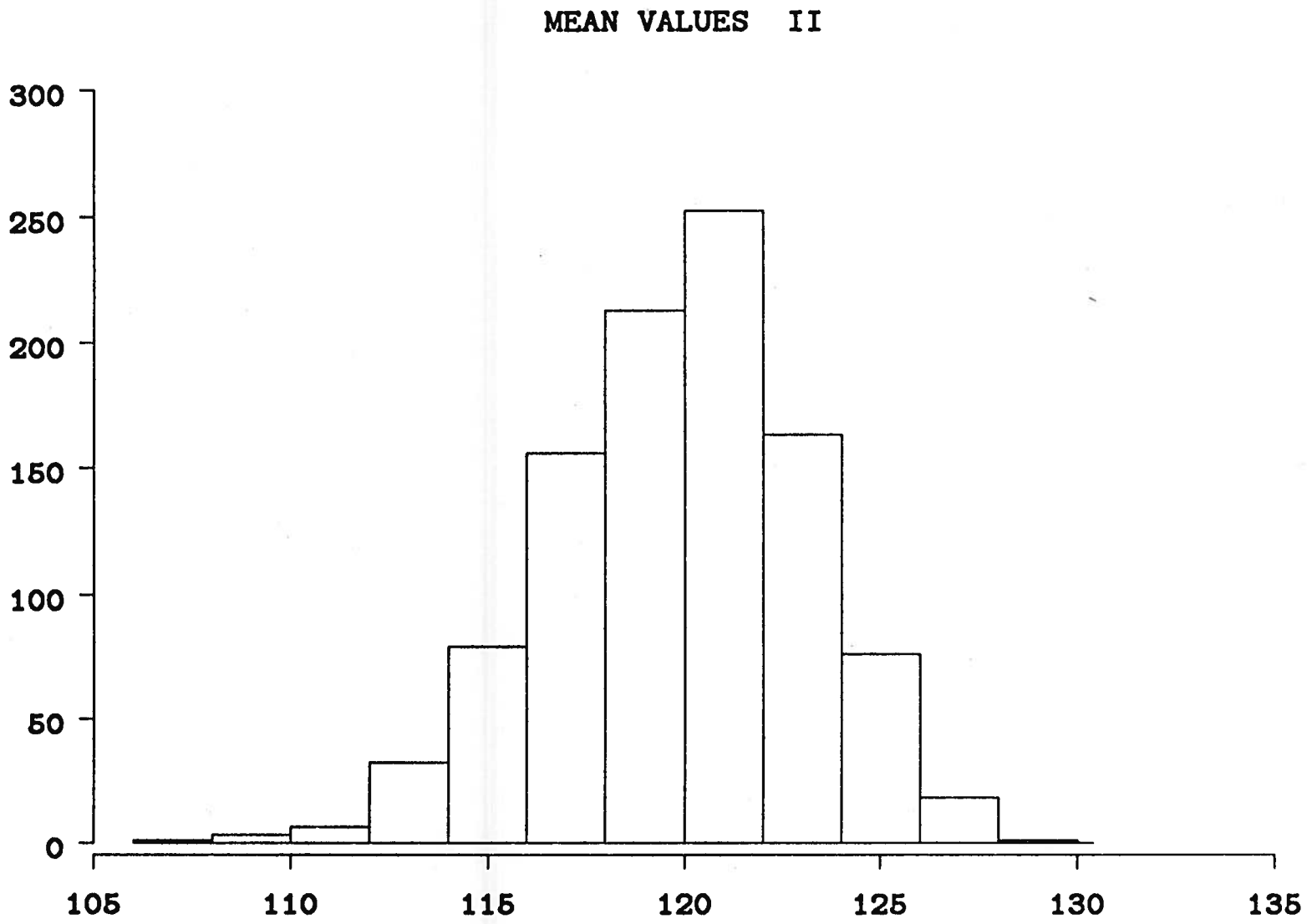
Reeds, J. (1976). On the definition of a von Mises functional. Thesis, Harvard University.

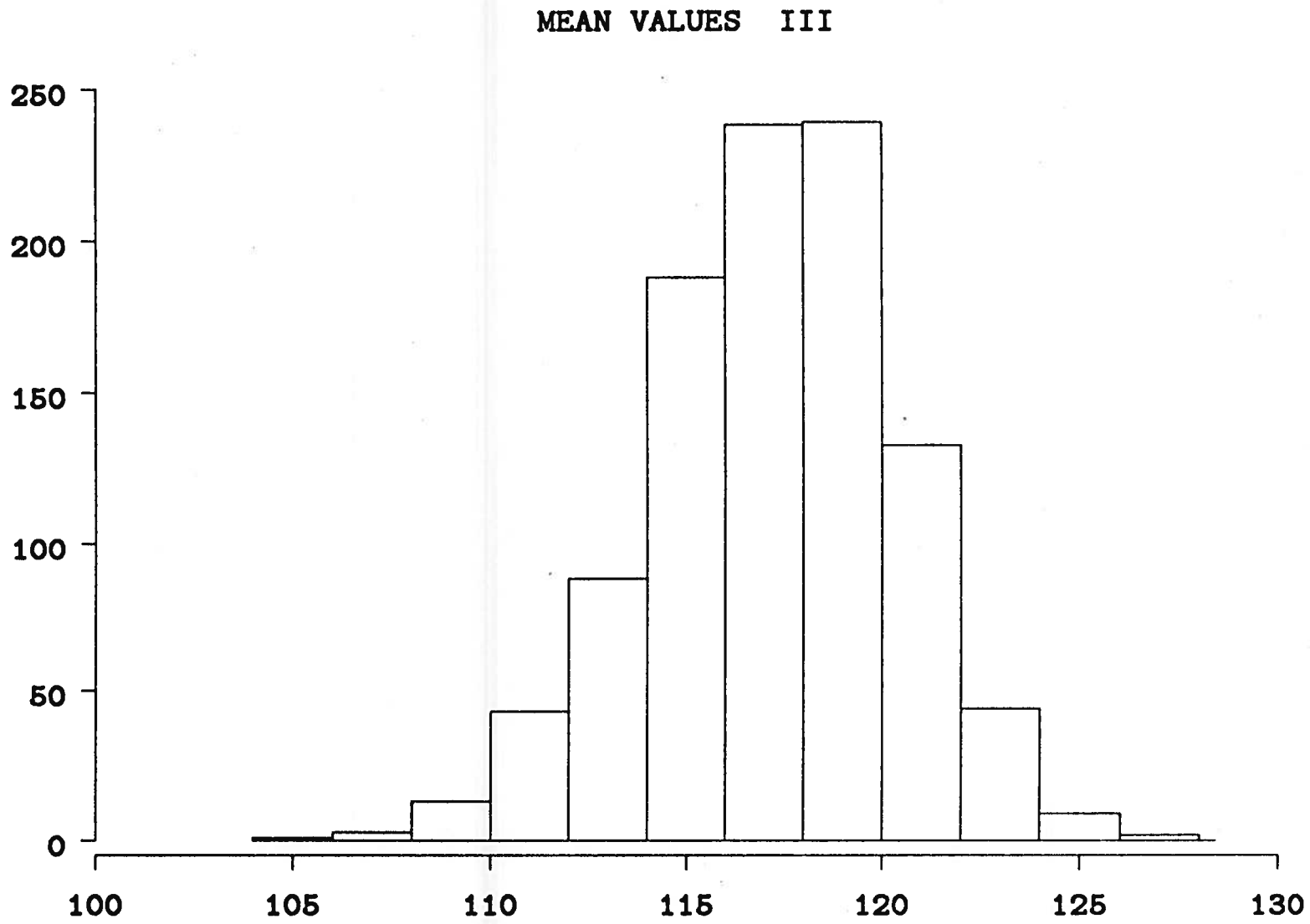
Rubin, D.B. (1981). The Bayesian bootstrap. Ann. Statist. 9, 130-134.

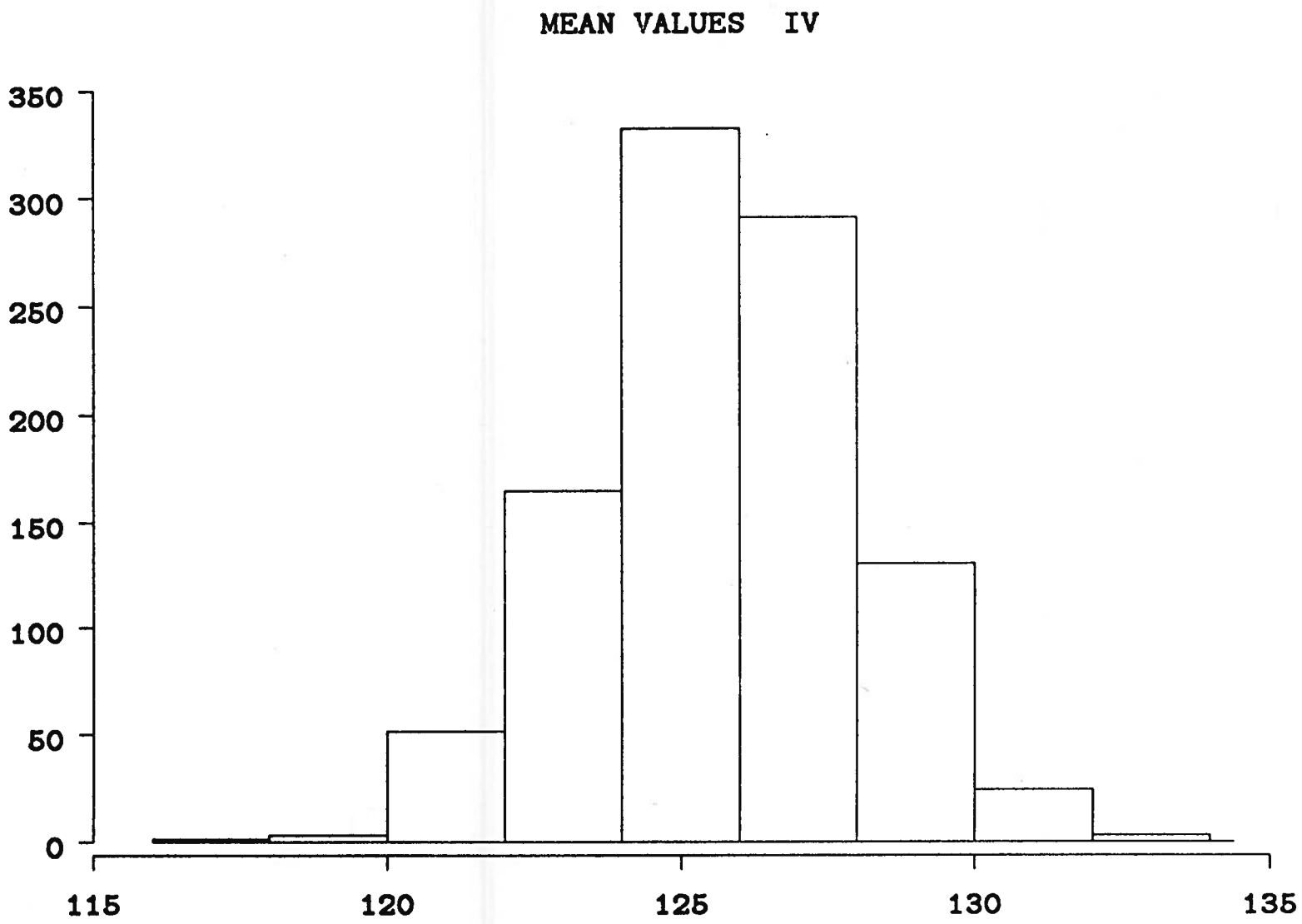
Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap. Ann. Statist. 9, 1187-1195.

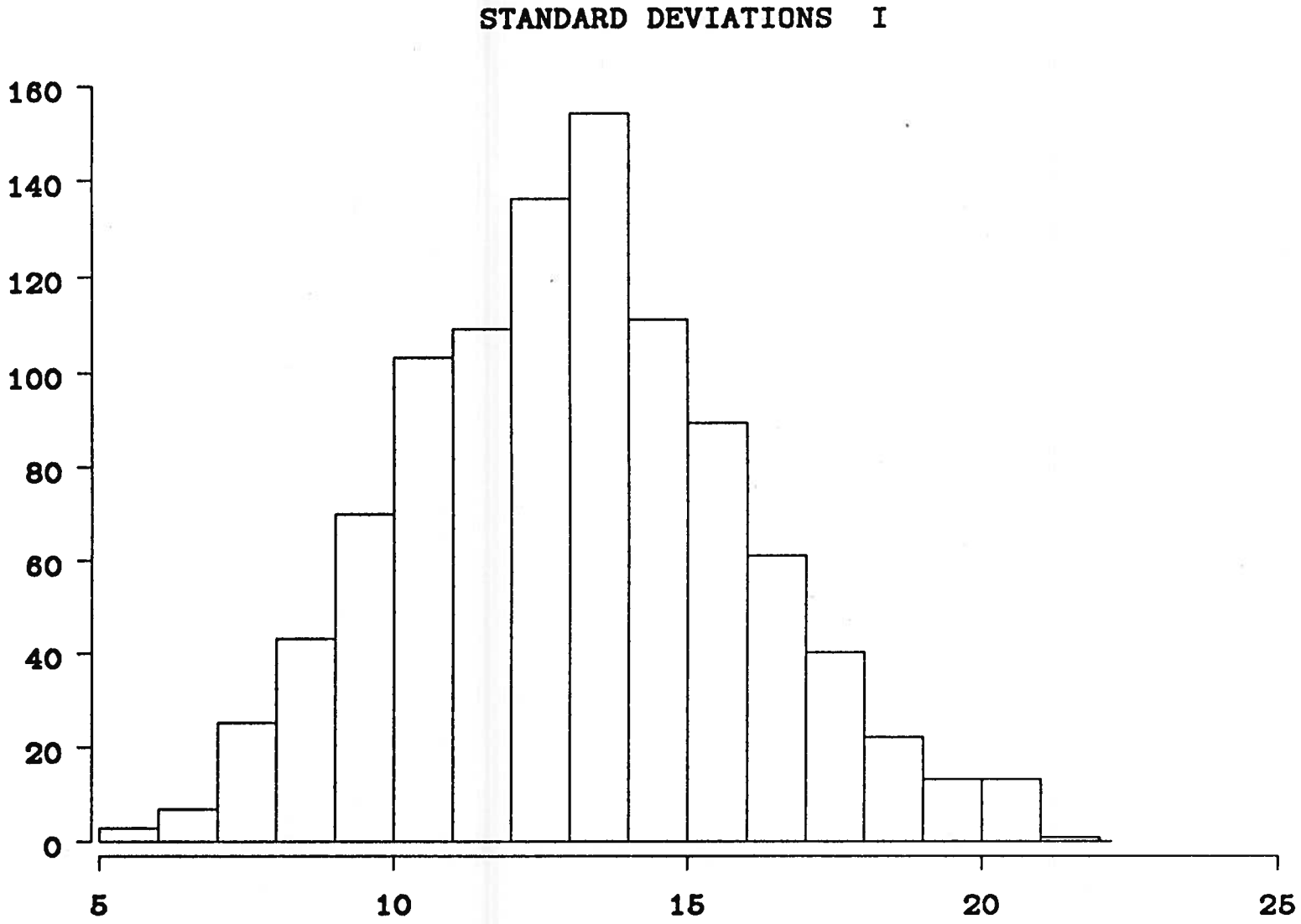
Yamato, H. (1984). Characteristic functions of means of distributions chosen from a Dirichlet process. Ann. Probability 12, 262-267.

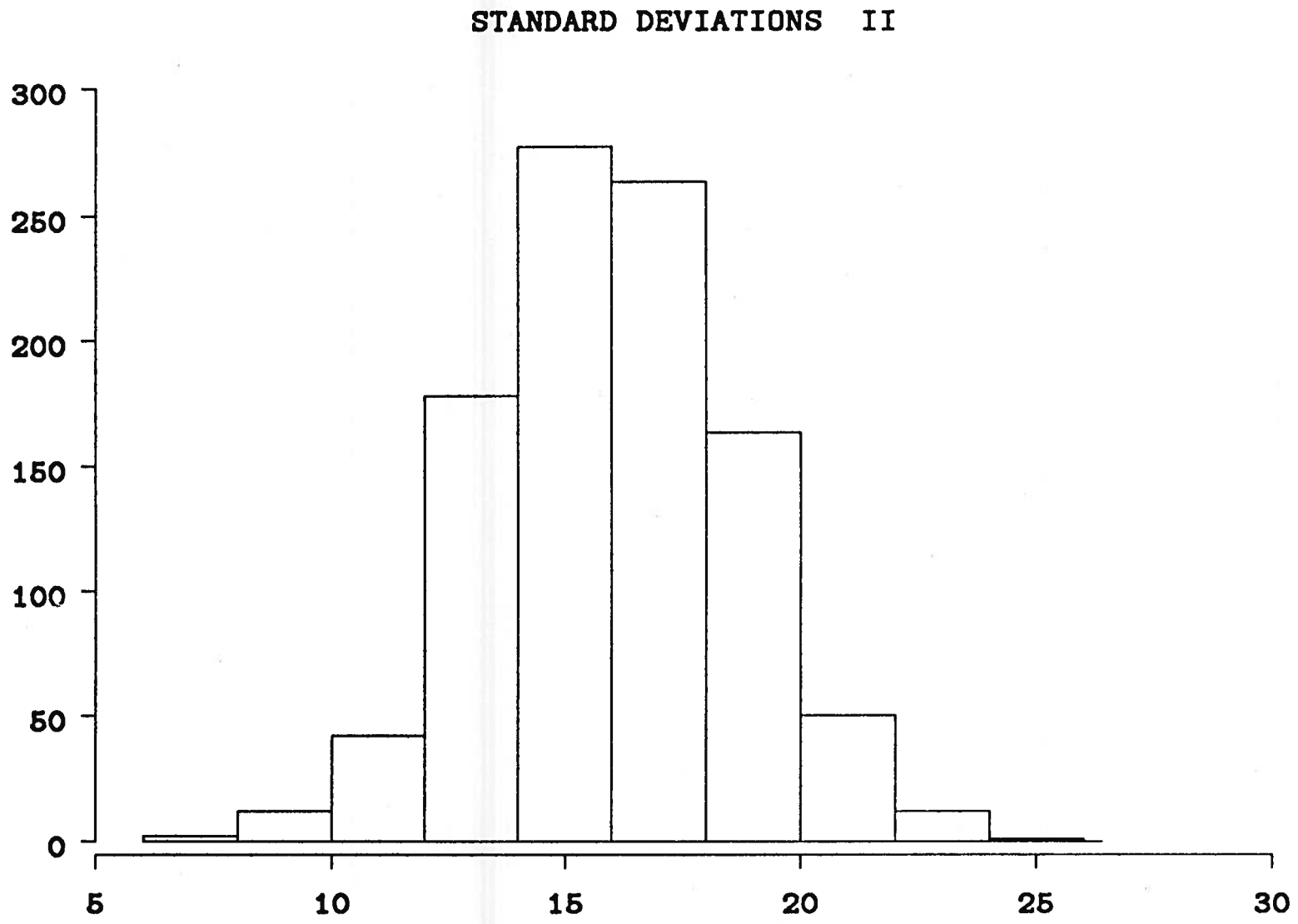

MEAN VALUES I
300
250
200
150
100
50
0
110
115
120
125
130
135

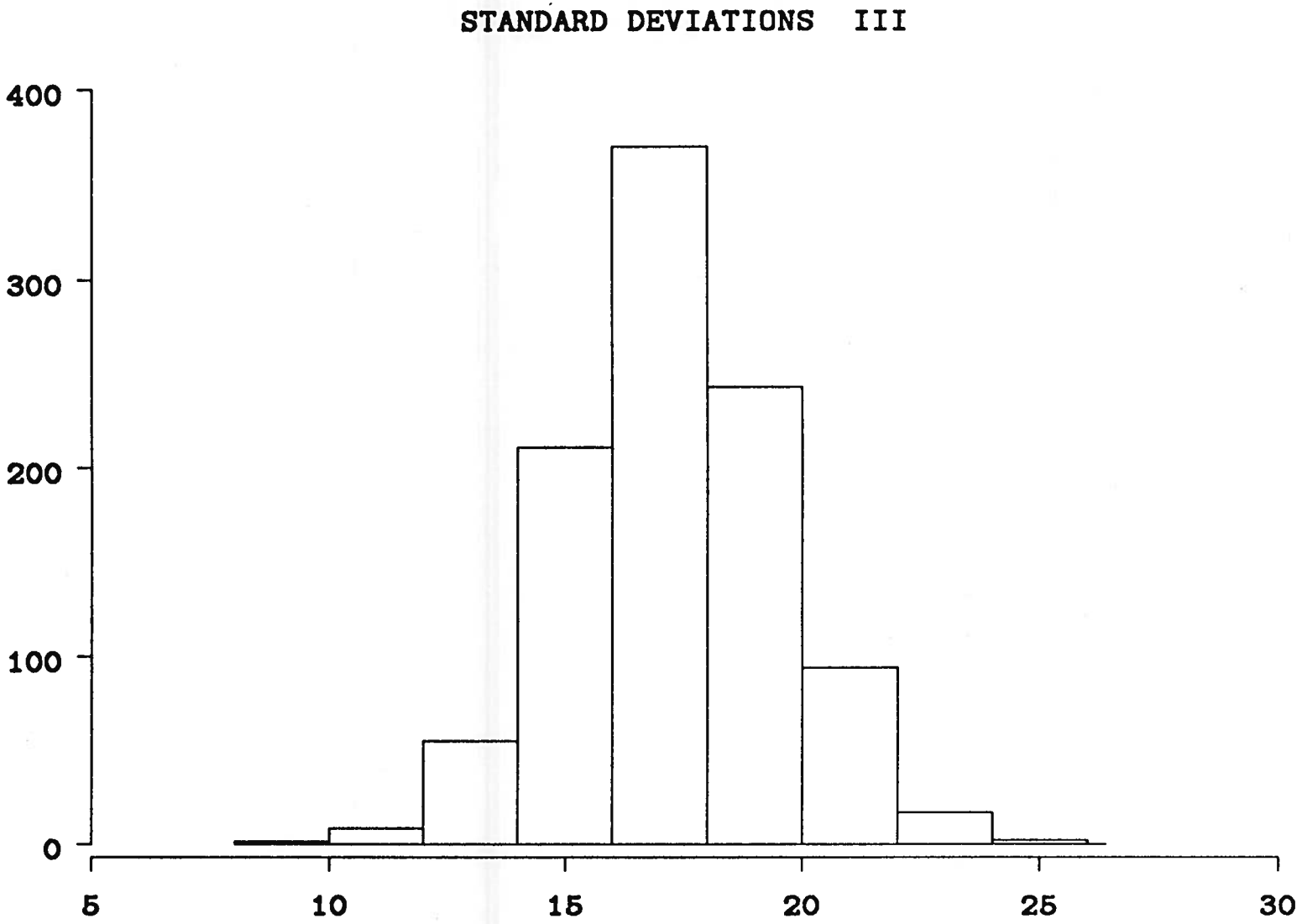

MEAN VALUES II
300
250
200
150
100
50
0
105
110
115
120
125
130
135

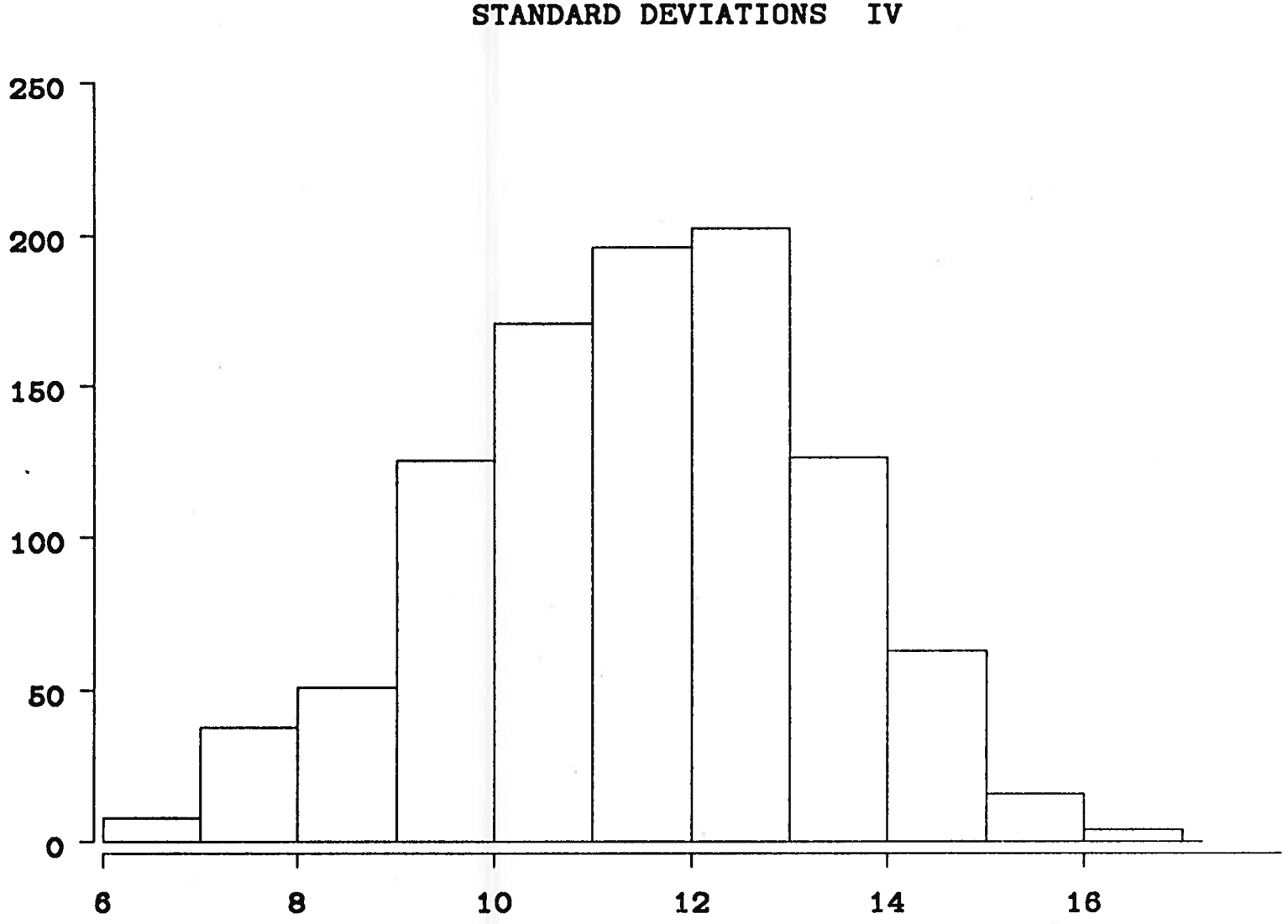

MEAN VALUES III
250
200
150
100
50
0
100
105
110
115
120
125
130

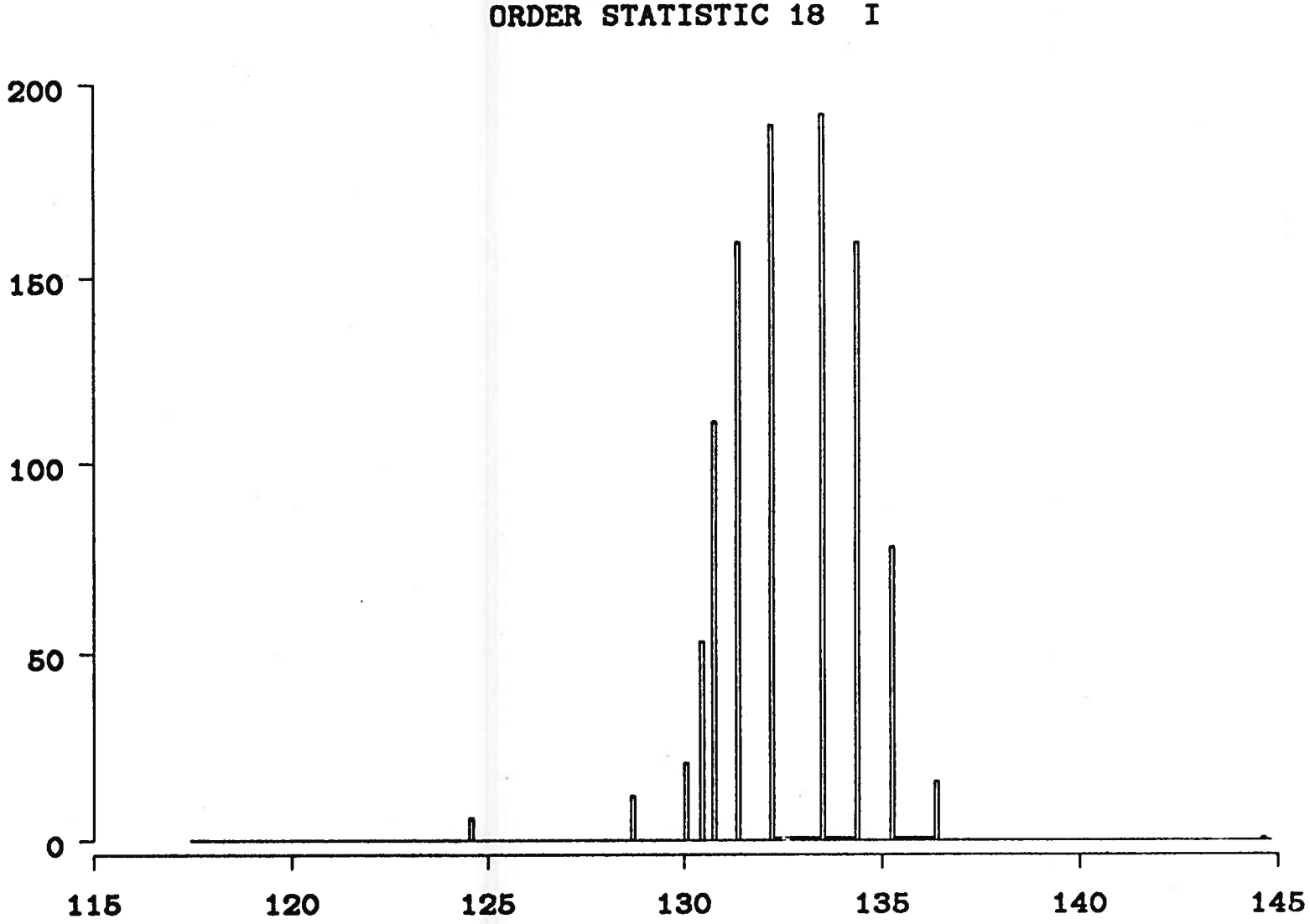

MEAN VALUES IV
350
300
250
200
150
100
50
0
115
120
125
130
135

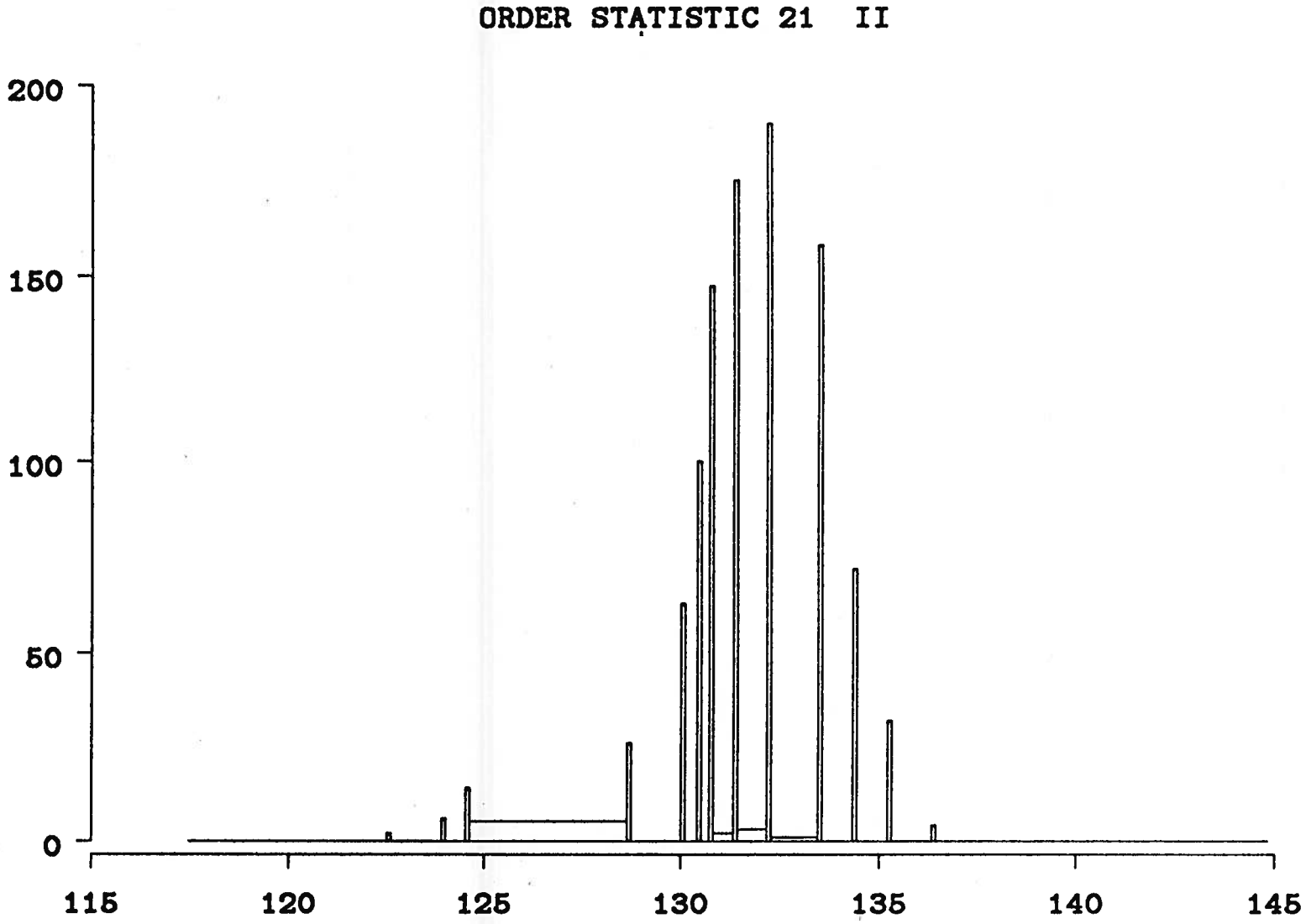

STANDARD DEVIATIONS I
160
140
120
100
80
60
40
20
0
5
10
15
20
25

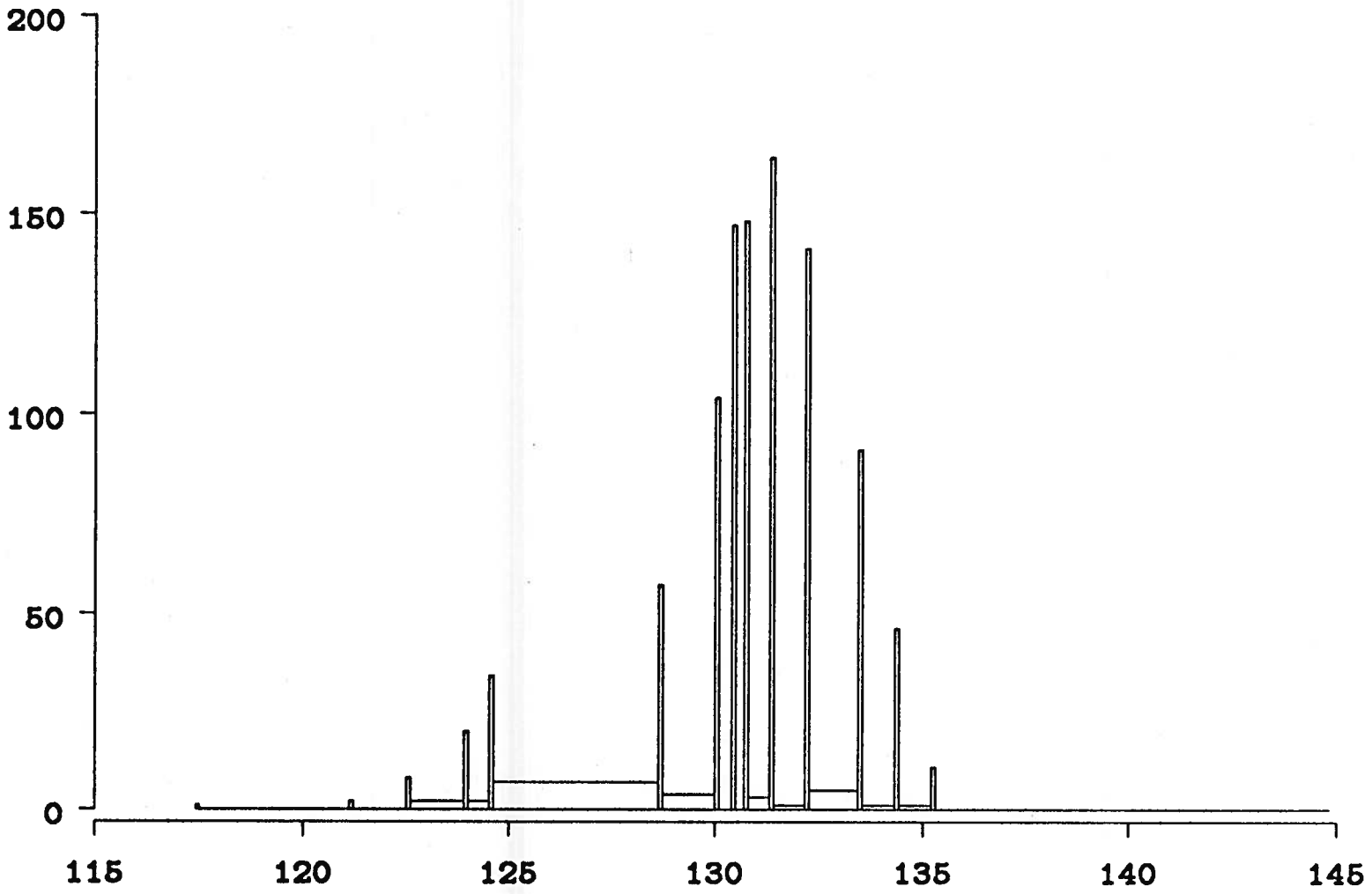

STANDARD DEVIATIONS II
300
250
200
150
100
50
0
5
10
15
20
25
30

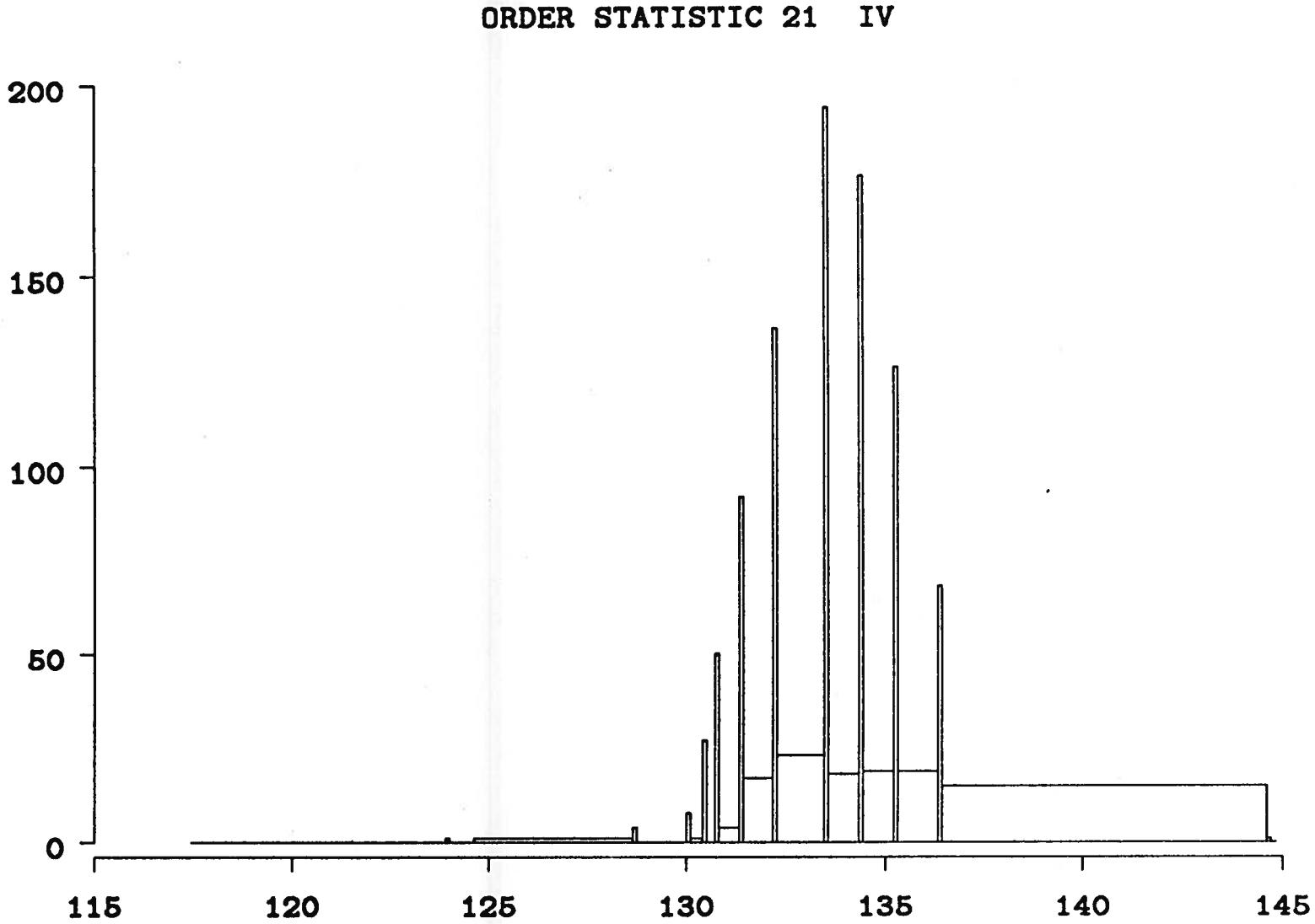

STANDARD DEVIATIONS III
400
300
200
100
0
5
10
15
20
25
30


STANDARD DEVIATIONS IV
250
200
150
100
50
0
6
8
10
12
14
16


ORDER STATISTIC 18 I
200
150
100
50
0
115
120
125
130
135
140
145


ORDER STATISTIC 21 II
200
150
100
50
0
115
120
125
130
135
140
145


ORDER STATISTIC 24 III
200
150
100
50
0
115
120
125
130
135
140
145


ORDER STATISTIC 21 IV
200
150
100
50
0
115
120
125
130
135
140
145