# Selecting the Number of Bins in a Histogram: A Decision Theoretic Approach

Kun He*
Department of Mathematics
University of Kansas
Lawrence, KS 66045

Glen Meeden†
School of Statistics
University of Minnesota
Minneapolis, MN 55455

Appeared in *Journal of Statistical Planning and Inference*, Vol 61 (1997), 59-59.

*Research supported in part by University of Kansas General Research Fund
†Research supported in part by NSF Grant SES 9201718

# Selecting the Number of Bins in a Histogram: A Decision Theoretic Approach

Short Running Title
Number of Bins in a Histogram


## ABSTRACT

In this note we consider the problem of, given a sample, selecting the number of bins in a histogram. A loss function is introduced which reflects the idea that smooth distributions should have fewer bins than rough distributions. A stepwise Bayes rule, based on the Bayesian bootstrap, is found and is shown to be admissible. Some simulation results are presented to show how the rule works in practice.



Key Words: histogram, Bayesian bootstrap, stepwise Bayes, admissibility, non-informative Bayes and entropy.

AMS 1991 Subject Classification: Primary 62C15; Secondary 62F15, 62G07.

# 1 Introduction

The histogram is a statistical technique with a long history. Unfortunately there exist only a few explicit guidelines, which are based on statistical theory, for choosing the number of bins that appear in the histogram. Scott [8] gave a formula for the optimal histogram bin width which asymptotically minimizes the integrated mean squared error. Since the underlying density is usually unknown, it is not immediately clear how one should apply this in practice. Scott suggested using the Gaussian density as a reference standard, which leads to the data-based choice for the bin width of $a \times s \times n^{-1/3}$, where $a = 3.49$ and $s$ is an estimate of the standard deviation. (See also Terrell and Scott [10] and Terrell [9].) As Scott noted many authors advise that for real data sets histograms based on 5-20 bins usually suffice. Rudemo [7] suggested a cross-validation technique for selecting the number of bins. But such methods seem to have large sampling variation.

In this note we will give a decision theoretic approach to the problem of choosing the number of bins in a histogram. We will introduce a loss function which incorporates the idea that smoother densities require less bins in their histogram estimates than rougher densities. A non-informative Bayesian approach, based on the Bayesian bootstrap of Rubin [6], will yield a data dependent decision rule for selecting the number of bins. We will then give a stepwise Bayes argument which proves the admissibility of this rule and shows the close connection of the rule to the notion of maximum likelihood, which also underlies the idea of a histogram. Finally we give some simulation results which show how our rule works in practice and compares to Scott's rule. In section 2 we describe the rule and give the simulation results, while the proof of admissibility is deferred to section 3.

# 2 Selecting the Number of Bins

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random sample from some unknown continuous distribution on the known interval $[a, b]$. The parameter space is $\Theta$, the class of all probability density functions $f$ on $[a, b]$. We are considering the decision problem where a histogram is an estimate of the unknown density $f$. Since we are only considering histograms where the bins are all of the same size, a typical decision $d$ consists of two components, the number of bins and the mass assigned to each bin. Hence we will write $d = (\, k, \mathbf{m}_k \,)$, where $k$ is a

positive integer such that $5 \leq k \leq 20$ and $\mathbf{m}_k = (\, m_{k,1}, \ldots, m_{k,k}\,)$ where the components of $\mathbf{m}_k$ are nonnegative and sum to one.

Next we must specify a loss function when $d = (\, k, \mathbf{m}_k\,)$ is an estimate of $f$. We will assume that the loss depends on $f$ only through the mass it assigns to the $k$ equal subintervals of $[a, b]$, i.e. on $\mathbf{p}_k(f) = \mathbf{p}_k = (\, p_{k,1}, \ldots, p_{k,k}\,)$. This assumption is not only mathematically convenient for what follows but seems to be in the spirit underlying the idea of estimating $f$ with a histogram. In addition we will assume that our loss function depends on the data $\mathbf{X} = \mathbf{x}$ and it will be denoted by

$$L(f, d, \mathbf{x}) = L(\mathbf{p}_k, k, \mathbf{m}_k, \mathbf{x})$$

where

$$L(\mathbf{p}_k, k, \mathbf{m}_k, \mathbf{x}) = \begin{cases} c(k, \mathbf{x}) \sum_1^k (p_{k,i} - m_{k,i})^2 & \text{if no } m_{k,i} = 1 \\ \\ 1/k & \text{if } \exists\, m_{k,i} = 1 \end{cases} \tag{2.1}$$

Before defining $c(k, \mathbf{x})$ we will explain the necessity for the term $1/k$ in the second part of the loss function. To consider a very simple case suppose that we are on the unit interval $[0, 1]$ and we have just four possible bins. Furthermore suppose the unknown $f$ puts all its mass in the first bin, i.e. on $[0, .25]$. Now consider two different decisions $d_1 = (2, (1, 0))$ and $d_2 = (4, (1, 0, 0, 0))$. Note for such an $f$ we have that for both of these decisions the sum in the first part of the loss function is zero. But clearly $d_2$ should be preferred to $d_1$ and this in fact is what is done by the second part of the loss function.

It remains to define $c(k, \mathbf{x})$. In principle there could be many reasonable choices of $c(k, \mathbf{x})$. At first sight our choice, given in equation 2.4, might seem quite mysterious and since it is really the key to what follows, we shall now give a detailed explanation of our choice.

The basic difficulty in selecting the number of bins is how to trade off the increased 'accuracy' that comes from adding more bins with the increased 'cost' of the additional bins. Or in decision theoretic terms, how can we calibrate the risk functions of the various problems, so that they could be sensibly compared. One possibility would be to assign a cost function to the number of bins, which increases as the number of bins increase. Unfortunately there seems to be no obvious choice for such a function. Furthermore we could never find any such function which seemed to work in a reasonable

fashion in the examples we considered. Perhaps there is such a function, but it is unknown to us.

We were interested in attacking the problem from a non-informative Bayesian prospective. Often non-parametric problems such as these are quite difficult from this point of view. However Rubin [6] proposed the Bayesian bootstrap, which can be quite useful for such problems. Meeden et al [5] proved the admissibility of various non-parametric procedures by showing that they were stepwise Bayes procedures against the Bayesian bootstrap. Our idea was to apply a similar approach to the problem of selecting the number of bins.

To this end, for a given k, let

$$\Theta(k) = \{\, \mathbf{p}_k = \mathbf{p}_k(f) \,:\, f \in \Theta \,\}$$

Given that $\mathbf{X} = \mathbf{x}$ let $\mathbf{V}_k(\mathbf{x}) = \mathbf{v}_k$ be the count vector for the number of observations that fall in each bin. That is if $\mathbf{v}_k = (\, v_{k,1}, \ldots, v_{k,k} \,)$ then $v_{k,i}$ is the number of observations that fell into bin $i$. So if $f$ is the true but unknown distribution generating $\mathbf{X}$ then $\mathbf{V}_k$ is multinomial$(n, \mathbf{p}_k)$. After $\mathbf{X} = \mathbf{x}$ has been observed we will take as our 'posterior distribution' over $\Theta(k)$ the Dirichlet distribution with parameter vector $\mathbf{v}_k$. This 'posterior' is the Bayesian bootstrap, but we believe that it cannot be a true posterior distribution on $\Theta(k)$, i.e. that it arises in the usual way from a prior distribution. In the next section we will give the underlying stepwise Bayes justification for this 'posterior', but for now we will see out it leads to a sensible loss function.

Suppose $\mathbf{X} = \mathbf{x}$ has been observed, $k$, the number of bins, is fixed and $\mathbf{v}_k$ has been computed. Then the choice of $\mathbf{m}_k$ which minimizes the 'posterior' expectation, under the Bayesian bootstrap, of $\sum_1^k (p_{k,i} - m_{k,i})^2$ is just $\mathbf{v}_k/n$ where $n = \sum_{i=1}^k v_{k,i}$, the sample size. The corresponding 'posterior' risk is

$$\sum_{i=1}^{k} v_{k,i}(n - v_{k,i})/(n^2(n+1)).$$

Now it is easy to check that for a fixed $\mathbf{x}$ this posterior risk increases as $k$ increases and so by itself can never be used to choose the number of bins. Since $(n+1)^{-1} \sum_{i=1}^k m_{k,i}(1 - m_{k,i})$ is maximized by taking $m_{k,i} = 1/k$ for each $i$ the 'posterior' risk is always at least as large as $(1 - 1/k)/(n+1)$. From this it follows that

$$\frac{\sum_{i=1}^k v_{k,i}(n - v_i)/(n^2(n+1))}{(1-1/k)/(n+1)} \leq 1 \tag{2.2}$$

This ratio, as a function of the number of bins $k$, compares the actual 'posterior' risk under the Bayesian bootstrap to its maximum possible value. This can be viewed as an attempt to calibrate the loss over the different problems. However it takes no account of the roughness or smoothness of the underlying distribution. To do this we will modify the denominator further. Now it is clear that when sampling from a rough distribution we want our histogram to have more bins than when sampling from a smooth distribution. Hence histograms with fewer bins should be penalized when the data becomes rougher. A convenient measure of the smoothness or uncertainty of a probability distribution is its entropy. Given $\mathbf{v}_k$ we can think of $\mathcal{E}_{\mathbf{v}_k} = -\sum_{i=1}^{k}(v_{k,i}/n)\log(v_{k,i}/n)$ as an estimate of the smoothness of the unknown $f$. Then the ratio

$$r(\mathbf{x}, k) = \frac{\mathcal{E}_{\mathbf{v}_k}}{\log k},$$

which is the estimated entropy of $f$ over $k$ bins divided by the entropy of the uniform distribution over $k$ bins, gives a standardized estimate of the measure of the smoothness of the distribution. Now $r(\mathbf{x}, k)$ always lies between 0 and 1 and for a fixed $k$ decreases as the data becomes rougher. Hence if the denominator of equation 2.2 is raised to the power $1 + (1 - r(\mathbf{x}, k))$ then it is decreased as $r(\mathbf{x}, k)$ decreases. Furthermore, because of the factor $1 - 1/k$, this decrease is proportionally greater for smaller values of $k$. This means that histograms with fewer bins are penalized by a greater amount than histograms with more bins when the data are rough.

Formally, our proposal is to consider the ratio

$$\frac{\sum_{i=1}^{k} v_{k,i}(n - v_{k,i})/(n^2(n+1))}{\{(1-1/k)/(n+1)\}^{\{1+(1-r(\mathbf{X},k))\}}} \tag{2.3}$$

and to select the histogram with $k_0$ bins where $k_0$ is the value of $k$ which minimizes the above ratio when $5 \leq k \leq 20$. Under this proposal, when sampling from a smooth distribution, one would expect to see on the average histograms with fewer bins than when sampling from a rougher distribution. To implement this suggestion we define $c(k, \mathbf{x})$ as follows

$$c(k, \mathbf{x})^{-1} = \{(1-1/k)/(n+1)\}^{\{1+(1-r(\mathbf{X},k))\}} \tag{2.4}$$

In the above we have given a justification of the loss function defined in equation 2.1 with $c(k, \mathbf{x})$ defined in equation 2.4. The general form of

the loss function is quite reasonable and the particular choice of $c(k, \mathbf{x})$ was closely tied to our desire to exploit the Bayesian bootstrap to give a sensible solution for this problem. To see how it works in practice we considered eight different densities on the unit interval. The first was the Uniform distribution, the second was the Beta(.9,10) distribution, the third was the Beta(.9,2) distribution, the fourth was the Beta(2,4) distribution and the fifth was the Beta(3,3) distribution. The final three were each a mixture of two Betas and the graphs of these final three densities are given in Figure 1. Together they represent most of the common types of distributions one would expect to see in practice.

put figure 1 about here

For each density we took 500 samples of size 50, 100 and 200 and found the number of bins selected by our method and the number of bins found by the rule proposed by Scott. When applying our method we only considered histograms with five to twenty bins and selected the one that minimized equation 2.3. Rather than considering subintervals that run over [0,1] we let them run over the interval $[min(\mathbf{x}), max(\mathbf{x})]$. This will matter little in practice and would be the sensible thing to do when the range of the unknown distribution, $[a, b]$, is also unknown. For each of the 500 random samples we found the sample mean and sample standard deviation of the number of bins for the two methods. The results are given in Table 1.

put table 1 about here

The results for Scott's rule are what we would expect. The average number of bins depends very little on the underlying distribution, especially for smaller sample sizes, and increases as the sample size increases. On the other hand for the method given here, based on the Bayesian bootstrap, the average number of bins varies a good deal as the underlying population changes. In particular smooth densities generate histograms with fewer bins than the rougher densities. Moreover, for smooth populations the average number of bins tends to decrease as the sample size increases, while for the rougher densities just the opposite is true.

The formula for the optimal histogram bin width which asymptotically minimizes the integrated mean squared error, as derived in Scott [8], is given by

$$\{6/\int_a^b f'(x)^2\,dx\}^{1/3}\,n^{-1/3}$$

for a given density $f$, when the integral exists. He also gives two other methods for increasing the number of bins based on skewness and kurtosis. For density 5 with samples of size 200 the above rule yields 8.3 bins which is very close to the average given by Scott's rule in the Table. Our method yields almost twice as many bins, but this just reflects the fact that we are using a different loss structure.

Terrell and Scott [10] proposed the following rule of thumb: For a histogram of data from a density believed to be moderately smooth on the real line, use $(2n)^{1/3}$ equal-width bins or a convenient slightly larger number over the sample range. For $n = 50$ this is 4.6 bins and for $n = 4000$ this is 20 bins. This is consistent with the suggestion noted earlier that for most practical problems only considering histograms with 5 to 20 bins is not unreasonable. Scott gives an asymptotic justification of his rule. Although the sample size should play some role in the selecting the number of bins our approach is essentially a non-asymptotic one. To see this suppose $n$ is large enough so that $v_{k,i}/n$ is a very good estimate of $p_{k,i}$. Let $f$ be the density to be estimated and let $\mathcal{E}_f(k)$ just be $\mathcal{E}_{\mathbf{v}_k}$ where $\mathbf{v}_k$ has been replaced with $\mathbf{p}_k(f)$. Then $r(f,k) = \mathcal{E}_f(k)/\log k$ is just the true entropy of $f$ over $k$ bins divided by the entropy of the Uniform distribution over $k$ bins. Hence if we replace $\mathbf{v}_k$ with $\mathbf{p}_k(f)$ every where it appears in equation 2.3 we get the following equation

$$(n+1)^{r(f,k)-1}(1-1/k)^{r(f,k)-2}\sum_{i=1}^{k} p_{k,i}(1-p_{k,i}). \tag{2.5}$$

Using our loss structure this equation measures how well a $k$ bin histogram approximates $f$ where for each bin we use the true bin probability under $f$. Hence, for us, an optimal number bins for fitting the true $f$ is the value of $k$ which minimizes the above equation. Note if $f$ is the Uniform density then the above function is constant in $k$. But for any other density and any value of $n$ we can find the minimizing value of $k$ as $k$ ranges from five to twenty. Our investigations found that the value of $n$ has little affect on this minimizing value. For $n = 500$ for each of the other 7 densities we found the value of $k$ which minimizes the above equation. For densities 2, 3, 4 and 5 the minimizing value is 20. For densities 6, 7 and 8 the minimizing value is 5. Note that this is very consistent with the results given in the Table expect for density 7. For density 7 we took an additional 500 samples of size 500 and another 500 samples of size 1,000. The average number of bins found by our method was 11.36 and 6.05 respectively. Because of the shape of density 7 it

just takes a larger sample before our method approaches its true optimal bin size. For density 4 we took an additional 500 random samples of size 1,000. In this case the average number of bins found by our method was 19.77. We believe that the number of bins in a histogram should depend both on the sample size and the shape of the unknown density to be estimated. Our method places more importance on the unknown shape in contrast to the more traditional asymptotic methods. For most sample sizes encountered in practice this does not seem unreasonable.

The above just compares the two methods in terms of the number of bins each of them produces. There are two more natural comparisons that can be considered. The first is integrated squared error loss which is the criterion Scott used to derive his rule and the second is just the loss function given in equation 2.3 which gives our procedure. Table 2 shows these comparison for 500 random samples at the three sample sizes for densities 3 and 5.

put table 2 about here

To help get a feeling for how the histograms of our method and Scott's method would differ on particular samples we selected a sample of size 50 from density 8 for which our method yielded 14 bins and Scott's rule 4 bins. Note 4 is the average number of bins for this problem for Scott's method, while the average for our method is 14.55. (We only selected a few samples and took the first one with 14 and 4 bins.) The two histograms are plotted in Figure 2 along with the density. We see that the histogram with 4 bins completely misses the shape of the density. Even though the histogram with 14 bins is probably too rough it does a much better job in capturing the variability of the density. Further investigation indicates that this example is fairly typical. That is, Scott's rule gives fewer bins and will give better pictures for smooth densities but at the cost of increasing the probability of obscuring important structure for rougher densities. In short, our method seems to work in an intuitive and sensible way and should yield a useful method for determining the number of bins in practice.

put figure 2 about here

In the next section we will show that the usual theoretical justification of the Bayesian bootstrap, based on a stepwise Bayes argument, can be adopted to the problem of this note and the admissibility of the proposed method will be demonstrated.

# 3 The stepwise Bayes justification

The stepwise Bayes technique was introduced in Johnson [3] (see also Wald and Wolfowitz [11]) and named in Hsuan [2]. He showed that the class of unique stepwise procedures is the minimal complete class when the parameter space is finite and the loss function is strictly convex. Meeden and Ghosh [4] gave an alternative formulation of this result. Brown [1] gave a quite general complete class theorem for estimation problems having a finite sample space, again using the stepwise Bayes idea. In these arguments a finite sequence of disjoint subsets of the parameter space is selected, where the order of the specified subsets is important. A different prior distribution is defined on each of the subsets. Then the Bayes procedure is found for each sample point that receives positive probability under the first prior. Next the Bayes procedure is found for the second prior for each sample point which receives positive probability under the second prior and which was not taken care of under the first prior. Next the Bayes procedure is found for the sample points with positive probability under the third prior and which had not been considered in the first two stages. This process is continued over each subset of the sequence in the order given. If the sequences of subsets and priors are such that a procedure is defined at every sample point of the sample space then the resulting procedure is admissible. To prove the admissibility of a given procedure one must select the sequence of subsets, their order, and the sequence of priors appropriately.

To see how this will work for the problem of selecting a histogram, recall that we are assuming that we have a random sample of size $n$ from an unknown population with some density function on the known interval $[a, b]$. Given the sample we want to select a histogram with $k$ bins to represent the data. In each case the bins are always of the same size and $k$ must satisfy the conditions $k_1 \leq k \leq k_2$ where $k_1$ and $k_2$ are known positive integers. For the simulations in the last section $k_1 = 5$ and $k_2 = 20$.

Let $M$ be the least common multiple of the integers $k_1, k_1 + 1, \ldots, k_2$. Typically $M$ will be quite large and we will assume for notational convenience that $n < M$, but this plays no essential role in what follows. We will take as our parameter space $\Theta(M)$. The first step is to define a sequence of disjoint subsets of $\Theta(M)$. This will be done using a certain class of vectors $\mathbf{u}$ of lengths $2, 3, \ldots, n + 1$. For such a $\mathbf{u}$ the possible set of values for $u_1$ is $\{ 1, 2, \ldots, n \}$ and the length of $\mathbf{u}$ is always $u_1 + 1$. The remaining entries of $\mathbf{u}$ are just $u_1$ distinct members of the set $\{ 1, 2, \ldots, M \}$ arranged in increasing

order. We can associate with each such $\mathbf{u}$ the subset of $\Theta(M)$

$$\Theta(M, \mathbf{u}) = \{ \mathbf{p}_M \, : \, p_{M,i} > 0 \text{ for } i = u_2, u_3, \ldots, u_{u_1+1} \text{ and } \sum_{i=2}^{u_1+1} p_{M,u_i} = 1 \}$$

So if $\mathbf{u} = (1, 3)$ then $\Theta(M, \mathbf{u})$ is just the one vector that puts mass one on the third subinterval of $[a, b]$ and if $\mathbf{u} = (3, 1, 4, 7)$ then $\Theta(M, \mathbf{u})$ is the set of all those vectors which place all their mass on the first, fourth and seventh subintervals and where each of these three subintervals receive positive mass. More generally a typical $\mathbf{u}$ selects $u_1$ subintervals which are given by $u_2, u_3, \ldots, u_{u_1+1}$. Note that if $\mathbf{u}_1$ and $\mathbf{u}_2$ are distinct then $\Theta(M, \mathbf{u}_1)$ and $\Theta(M, \mathbf{u}_2)$ are disjoint. We will order the $\mathbf{u}'s$, and hence their corresponding induced subsets, by the usual lexicographical ordering. It is this ordering that we will use in the stepwise Bayes argument.

Recall that with the parameter space $\Theta(M)$, the random variable we observe is just $\mathbf{V}_M$ the count vector for the number of observations which fall into each subinterval. The distribution of $\mathbf{V}_M$ is just multinomial$(n, \mathbf{p})$. Let $\mathbf{u}$ be given and suppose we have already taken care of all the sample points which get positive probability under all those vectors $\mathbf{u}'$ which precede $\mathbf{u}$ in the lexicographic ordering. For the given $\mathbf{u}$ the points in the sample space of $\mathbf{V}_M$ which get assigned positive probability under $\mathbf{p}_M \in \Theta(M, \mathbf{u})$ are just those with $v_i \geq 0$ for $i = u_2, u_3, \ldots, u_{u_1+1}$. If any of these $v_i$'s are zero they would have been taken care of in an earlier stage of the argument. So at this stage we need only consider those sample points where each of these $v_i$'s are strictly greater than zero. The prior chosen at this stage is essentially

$$1/ \prod_{i=2}^{u_1+1} p_{M,u_i}$$

(We are ignoring here a function of $\mathbf{p}_M$ which plays no role since it also appears in the renormalized probability function for this stage and cancels.) Then for those sample points with each of these $v_i$'s strictly greater than zero the 'posterior' at this stage under this prior is just Dirichlet$(\mathbf{v})$, i.e. the Bayesian bootstrap. Then for such a $\mathbf{v}$ selecting the $k$ which minimizes the posterior expected loss is equivalent to selecting the $k$ which minimizes equation 2.3.

This shows that the procedure given in the previous section is indeed a stepwise Bayes procedure. If it were the unique stepwise Bayes procedure then it would be admissible by the standard arguments. However it need not

be unique since for a given $\mathbf{v}$ there can be more than one $k$ which minimizes equation 2.3. However it is easy to see that any rule which is admissible within this family of stepwise Bayes rules must be admissible for the original problem as well. One way to select an admissible rule from this family is to specify a continuous prior density function, say $g$, over $\Theta(M)$, which is strictly positive at every point and from this restricted family select a rule which is admissible, for this restricted problem, against the prior $g$. In practice there will usually be few samples for which more than one value of $k$ minimizes equation 2.3 and the choice of $g$ will not matter much. In the simulations we did not actually specify a $g$ but always took the smallest value of $k$ which minimized equation 2.3.

This concludes the theoretical justification for the method proposed here for selecting the number of bins in a histogram. We have seen that it is an admissible stepwise Bayes procedure which is based on the Bayesian bootstrap. As in some other non-parametric problems it can be thought of as a non-informative Bayesian solution to this problem. Moreover the simulation study indicates that in practice it should yield useful and sensible answers to a problem of some interest.

# References


[1] Lawerence D. Brown. A complete class theorem for statistical problems with finite sample space. *Annals of Statistics*, 9:1289–1300, 1981.

[2] Francis C. Hsuan. A stepwise Bayes procedure. *Annals of Statistics*, 7:860–868, 1979.

[3] Bruce McK. Johnson. On admissible estimators for certain fixed sample binomial problems. *Annals of Mathematical Statistics*, 42:1579–1587, 1971.

[4] Glen Meeden and Malay Ghosh. Admissibility in finite problems. *Annals of Statistics*, 9:296–305, 1981.

[5] Glen Meeden, Malay Ghosh, and Stephen Vardeman. Some admissible nonparametric and related finite population sampling estimators. *Annals of Statistics*, 13:811–817, 1985.

[6] Donald B. Rubin. The Bayesian bootstrap. *Annals of Statistics*, 9:130–134, 1981.

[7] Mats Rudemo. Empirical choice of histograms and kernel density estimators. *Scandinavian J. of Stat.*, 9:65–78, 1982.

[8] David W. Scott. On optimal and data-based histograms. *Biometrika*, 66:605–610, 1979.

[9] George R. Terrell. The maximal smoothing principle in density estimation. *J. of the Amer. Stat'l. Assn.*, 85:470–477, 1990.

[10] George R. Terrell and David W. Scott. Oversmooth nonparametric density estimates. *J. of the Amer. Stat'l. Assn.*, 80:209–214, 1985.

[11] A. Wald and J. Wolfowitz. Characterization of the minimal complete class of decision functions when the number of distributions and decisions is finite. In *Proc. Second Berk. Symp. Math. Statist. and Prob.*, pages 149–157, 1951.

Table 1: The mean and standard deviation , for 500 random samples, of the number of bins of the histogram found by Scott's rule, denoted by S, and the method based on the Bayesian Bootstrap, denoted by BBS

| | | Sample Size 50 | | Sample Size 100 | | Sample Size 200 | |
|---|---|---|---|---|---|---|---|
| Density | Type | Mean | Stdev | Mean | Stdev | Mean | Stdev |
| 1 | S | 4.02 | 0.14 | 5.03 | 0.17 | 6.09 | 0.29 |
| | BBS | 7.25 | 3.28 | 6.41 | 1.75 | 6.28 | 1.57 |
| 2 | S | 5.02 | 0.70 | 7.01 | 0.84 | 9.51 | 1.17 |
| | BBS | 18.49 | 1.97 | 18.97 | 1.35 | 19.21 | 1.06 |
| 3 | S | 4.32 | 0.47 | 5.69 | 0.48 | 7.13 | 0.40 |
| | BBS | 15.11 | 4.81 | 16.24 | 3.81 | 17.98 | 2.21 |
| 4 | S | 4.82 | 0.48 | 6.23 | 0.46 | 8.16 | 0.54 |
| | BBS | 13.22 | 5.56 | 15.09 | 4.63 | 17.30 | 3.06 |
| 5 | S | 4.79 | 0.45 | 6.21 | 0.44 | 8.05 | 0.46 |
| | BBS | 11.26 | 5.50 | 12.42 | 5.35 | 15.95 | 4.14 |
| 6 | S | 4.03 | 0.18 | 5.05 | 0.23 | 6.44 | 0.50 |
| | BBS | 6.94 | 3.07 | 6.16 | 1.69 | 5.86 | 1.40 |
| 7 | S | 3.50 | 0.50 | 4.31 | 0.46 | 5.71 | 0.46 |
| | BBS | 12.62 | 5.71 | 12.85 | 6.16 | 13.37 | 5.86 |
| 8 | S | 4.00 | 0.09 | 5.00 | 0.06 | 6.13 | 0.34 |
| | BBS | 14.55 | 5.08 | 11.85 | 4.78 | 7.75 | 2.27 |

Table 2: The average integrated mean squared error, denoted by Isqer, and the average value of the loss function in equation 2.3, denoted by eq2.3, for 500 random samples, of the number of the histogram found by Scott's rule, denoted by S, and the method based on the Bayesian Bootstrap, denoted by BBS

| | | Sample Size 50 | | Sample Size 100 | | Sample Size 200 | |
|---|---|---|---|---|---|---|---|
| Density | Type | Isqer | eq2.3 | Isqer | eq2.3 | Isqer | eq2.3 |
| 3 | S | 0.166 | 1.41 | 0.101 | 1.51 | 0.068 | 1.59 |
| | BBS | 0.327 | 1.21 | 0.176 | 1.31 | 0.101 | 1.41 |
| 5 | S | 0.140 | 1.23 | 0.085 | 1.31 | 0.057 | 1.43 |
| | BBS | 0.222 | 1.14 | 0.119 | 1.22 | 0.074 | 1.33 |

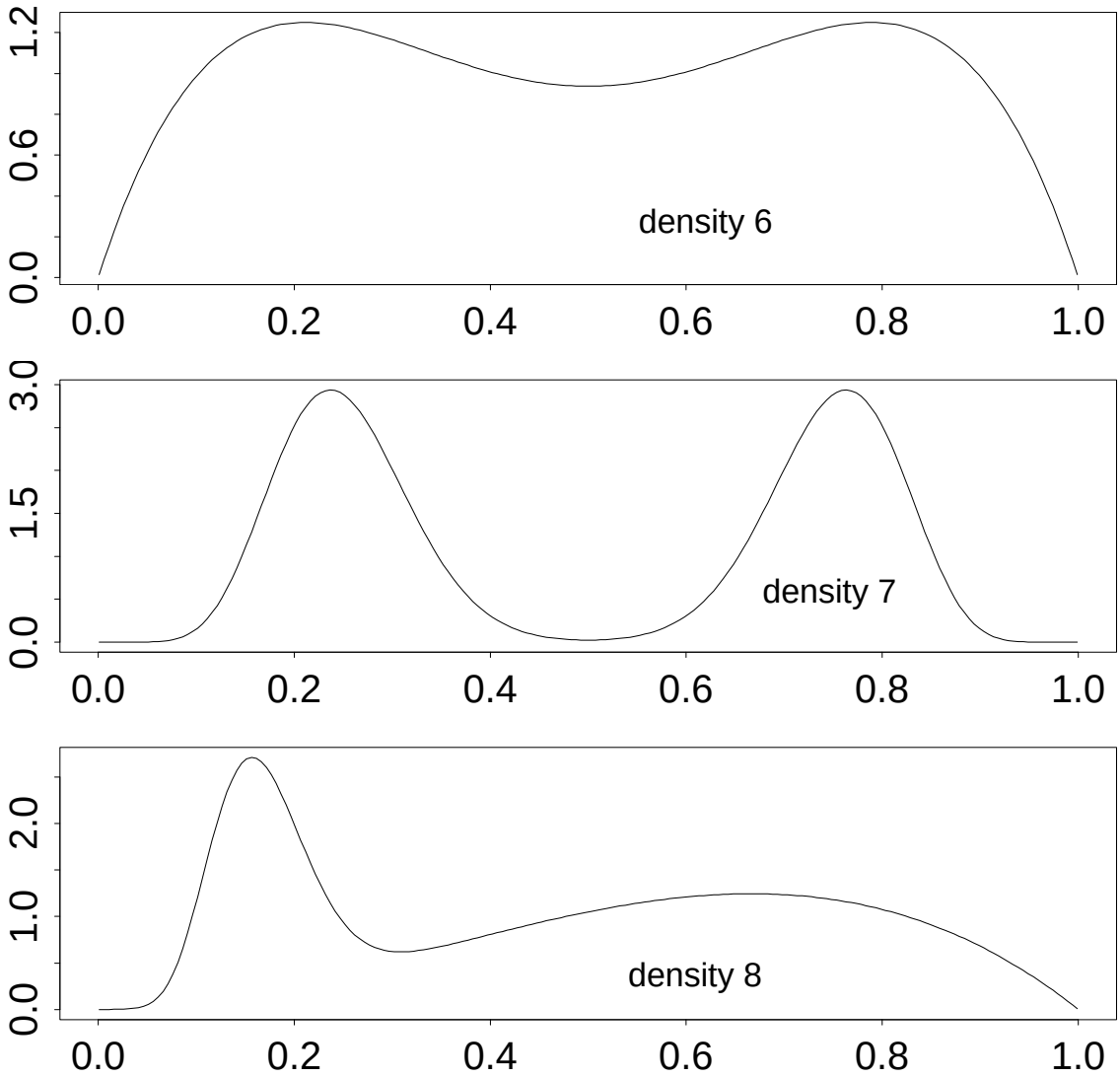


Figure 1: The graphs of densities 6, 7 and 8

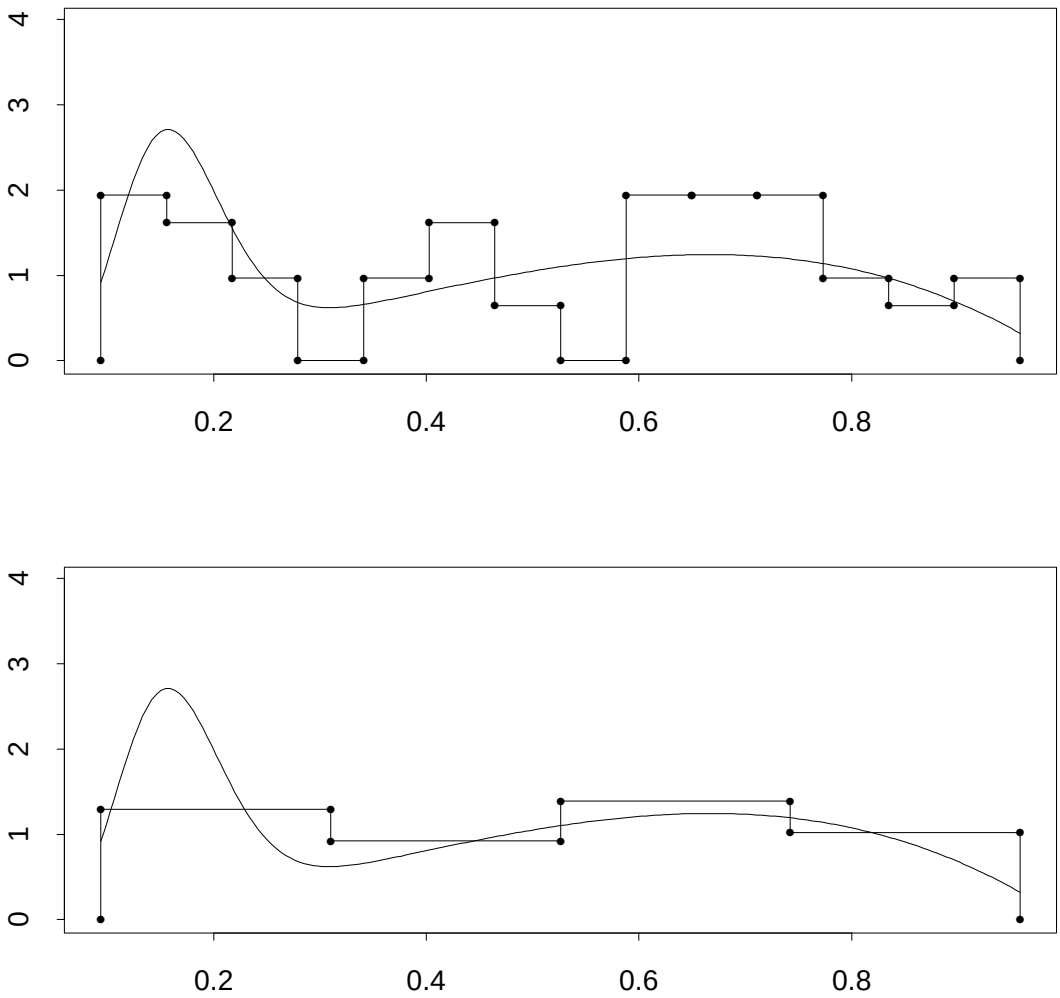


Figure 2: For a sample of size 50 from density 8 the BBS histogram, with 14 bins, and the histogram based on Scott's rule, with 5 bins