Computing and Graphing Highest Density Regions

Rob J. Hyndman*

Abstract

Many statistical methods involve summarizing a probability distribution by a region of the sample space covering a specified probability. One method of selecting such a region is to require it to contain points of relatively high density. Highest density regions are particularly useful for displaying multimodal distributions and, in such cases, may consist of several disjoint subsets — one for each local mode. In this paper, I propose a simple method for computing a highest density region from any given (possibly multivariate) density $f(\boldsymbol{x})$ which is bounded and continuous in $\boldsymbol{x}$. Several examples of the use of highest density regions in statistical graphics are also given. A new form of boxplot is proposed based on highest density regions; versions in one and two dimensions are given. Highest density regions in higher dimensions are also discussed and plotted.

KEY WORDS: Highest density regions, graphical summary, density estimation, boxplots.

1 Introduction

Many statistical methods involve summarizing a probability distribution by a region of the sample space covering a specified probability. For example, reporting a prediction interval for a future observation involves an implicit summary of the underlying distribution. A boxplot is an alternative summary where the various components mark regions of the empirical distribution function corresponding approximately to regions of the underlying probability distribution.

In summarizing a probability distribution by a region, it is not always clear which region should be used. Suppose you wish to give a two-sided 95% prediction interval

*Rob Hyndman is Lecturer, Department of Mathematics, Monash University, Clayton, Victoria, Australia, 3168.
The author thanks Gary Grunwald (University of Melbourne) for several constructive discussions in developing the ideas contained here, a referee for several suggestions which led to improvements in the paper, and David Scott (Rice University, Houston, Texas) for generously providing the software to produce Figure 6.

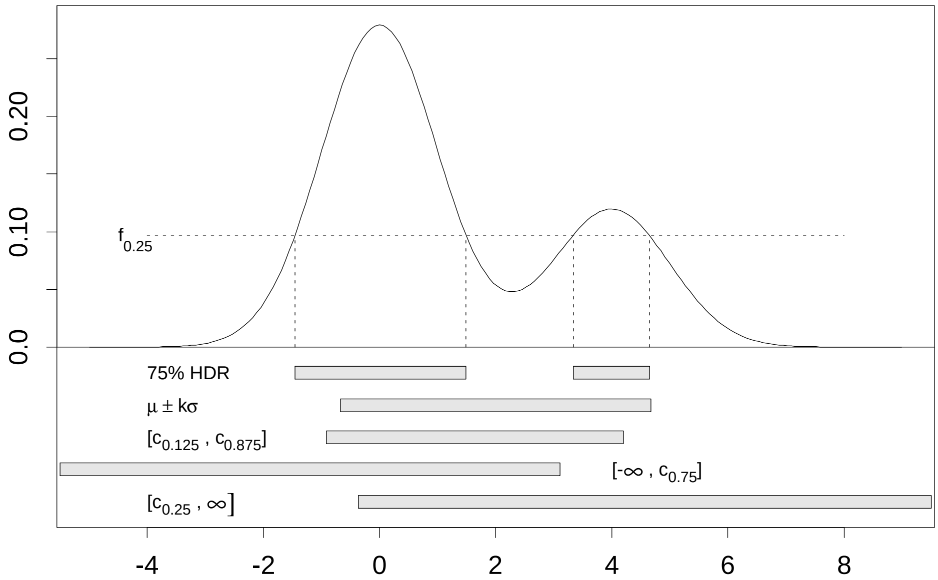


Figure 1: *Five different 75% probability regions from a normal mixture density. Here, c_q denotes the qth quantile, μ denotes the mean and σ denotes the standard deviation of the density.*

from a given distribution. Should you use the interval symmetric about the mean, the interval symmetric about the median, the interval defined between the 2.5% and 97.5% quantiles, the interval of shortest length, the interval which minimizes the probability of covering a given set? Each of these intervals has 95% coverage and they may all be different. In higher dimensions, the difficulty of selecting an appropriate region is even greater.

This article investigates one approach to this problem by suggesting that highest density regions are often the most appropriate subset to use to summarize a probability distribution. As a motivating example, Figure 1 shows five possible 75% regions for a mixture density comprising a N(0,1) density and a N(4,1) density with weights 0.7 and 0.3 respectively. Only the highest density region shows the bimodality.

The usual purpose in summarizing a probability distribution by a region of the sample space is to delineate a comparatively small set which contains most of the probability although the density may be non-zero over infinite regions of the sample space. This is the underlying philosophy of prediction regions and boxplots. Clearly, there are an infinite number of ways to choose a region with given coverage probability. In some cases, the nature of the problem will suggest a specific region. (For instance, 'one-sided' regions of given coverage are uniquely defined.) Often, though, which region to use is not clear and it is necessary to decide what properties we wish the region to possess. The following criteria seem intuitively sensible:

1 The region should occupy the smallest possible volume in the sample space;

2 Every point inside the region should have probability density at least as large as every point outside the region.

In fact, the criteria are equivalent (Box and Tiao, 1973) and lead to regions called *highest density regions* or HDRs. I shall define them more precisely using the second criterion.

DEFINITION *Let $f(x)$ be the density function of a random variable X. Then the $100(1-\alpha)\%$ HDR is the subset $R(f_\alpha)$ of the sample space of X such that*

$$R(f_\alpha) = \{x : f(x) \geq f_\alpha\}$$

where f_α is the largest constant such that $\Pr(X \in R(f_\alpha)) \geq 1-\alpha$.

The same definition also applies for discrete variables with the density function replaced by the probability mass function.

Figure 1 shows the particular value of $f_{0.25}$ which was used to construct the 75% region shown.

It follows immediately from the definition that of all regions of probability coverage $1-\alpha$, the HDR has the smallest possible volume in the sample space of X. Furthermore, the mode is contained in every HDR. Unlike other probability regions, HDRs are easily summarized, even for high dimensions, as they are defined by a single number f_α.

Such regions are common in Bayesian analysis where they are applied to a posterior distribution (e.g., Box and Tiao, 1973). In that context, they are also called "credible sets", "plausible sets", "highest posterior density regions" or "Bayesian confidence sets".

In the case of a normal distribution, an HDR coincides with the usual probability region symmetric about the mean, spanning the $\alpha/2$ and $1-\alpha/2$ quantiles. The same is true for any unimodal, symmetric distribution. However, in the case of a multimodal distribution, an HDR often consists of several disjoint sub-regions. This provides useful information which is 'masked' by other probability regions.

We shall first demonstrate the value of highest density regions by considering several graphical display methods in Section 2. The computational algorithms which were used to construct these graphs will be discussed in Section 3.

2 Graphical display

2.1 HDR boxplots

Boxplots, introduced by Tukey (1977), are a common method for summarizing univariate samples. Several variations of boxplots are discussed by McGill et al. (1978), Benjamini (1988) and Esty and Banfield (1992). All of these boxplot variants include a central box bounded by Q_1 and Q_3 where Q_1 and Q_3 denote sample quartiles or approximate sample quartiles such as fourths. Hence, the box contains approximately 50% of the observations. The most common form of boxplot contains 'whiskers' extending to $1.5(Q_3 - Q_1)$ beyond the ends of the central box. For a standard normal distribution the quartiles are ± 0.6745 so for a large data set, the whiskers lie at approximately $\pm[0.6745 + 1.5(1.349)] = \pm 2.698$. Hence the probability of an observation falling inside the whiskers is approximately $1 - 2[1 - \Phi(2.698)] = 99.30\%$ for large samples. (Hoaglin et al. (1986) show that for small to moderate samples the probability of observations inside the whiskers is smaller than this asymptotic rate.)

Here, I propose a new form of boxplot based on HDRs which summarizes the distribution in a similar way but allows the display of multimodality. An HDR boxplot replaces the box bounded by the interquartile range with the 50% HDR. In both cases, the coverage probability of the region is 50% but only the HDR will display multimodality. Similarly, the region bounded by the whiskers is replaced by the 99% HDR. This is chosen to roughly reflect the probability coverage of the whiskers on a standard boxplot for a normal distribution. Data beyond the 99% HDR are displayed as points. To emphasize the different densities of the two regions, shaded boxes are used with higher density shading in the 50% HDR. Finally, in keeping with the emphasis on highest density, the mode rather than the median is marked by a horizontal line.

Example: Figure 2 shows both standard and HDR boxplots for the daily maximum temperature in Melbourne, Australia between 1981 and 1990. The data are divided according to the temperature of the previous day and the density for each group was estimated using a kernel density estimator (e.g. Scott, 1992). The HDR boxplots were calculated using the density quantile approach of Section 3. Both displays demonstrate that the mean and variance of tomorrow's temperature increase as today's temperature increases, except on very hot days (over 40°C) which tend to be followed by cooler days. However, only the HDR boxplots reveal the bimodality of the distribution of the temperature following a warm to hot day. The 50% HDRs consist of two disjoint intervals showing that days of 30–39°C tend to be followed by days of similar temperature or of much lower temperature; they are not generally followed by days with maximum temperature in the high 20s. This occurs because temperatures slowly increase as high pressure systems pass over the city from west to east. At the tail end of a high pressure system, a strong north wind often blows (from off the Australian mainland) bringing high temperatures. A high pressure system is often followed by a cold front causing a rapid drop in temperature. Hence, hot days are generally followed by days of similar or greater temperature or by much cooler days. The hotter the day, the more likely it is to be followed by a cool day.

The bimodality of these distributions are, perhaps, the most interesting feature of these data. Yet standard boxplots give no hint that such a feature exists.

The vaseplots of Benjamini (1988) can show some forms of multimodality in that they allow the shape of the central box to be proportional to a density estimate of the data. However, there are several drawbacks to vaseplots compared with HDR boxplots. If the modes are close to or outside the quartiles, as in the 30–34°C group of temperatures in Figure 2, the vaseplots will not adequately display the multimodality. Also, for distributions with regions of very low density between modes, HDR boxplots allow outliers to be present between the modes unlike vaseplots or any other boxplot variant. Finally, unlike HDR boxplots, vaseplots do not extend naturally to allow continuous conditioning or to higher dimensions.

2.2 Conditional HDRs

The preceding example is easily extended to allow a more continuous display of the density of the maximum daily temperature conditional on the previous day's temperature. The conditional densities were calculated using a kernel approach as described in Hyndman et al. (1994) and the HDRs conditional on the previous day's temperature being 7,8,…,43 were computed. The results are displayed in Figure 3. This display avoids the artificial grouping of Figure 2 and allows the gradual change in the shape of the conditional densities to be seen more clearly.

In forecasting, it is common to compute the density of future values conditional on observed values of the series and the time horizon of the forecast. In this case, the HDRs may be plotted against time.

EXAMPLE: A famous data set in non-linear time series analysis is A.J. Nicholson's blowfly data. These consist of the number of Australian blowflies recorded every two days in a caged population on a strictly controlled diet (Nicholson, 1957). The data are also given in Brillinger et al. (1980). Of particular interest are the aperiodic population cycles. Tong (1988) fitted a threshold autoregressive model to the log of the population. We are interested in forecasting the process for the next 20 observation times. Hyndman (1993) used a Monte-Carlo technique to obtain forecast densities of the logged population at each time. These densities were then transformed back to the original scale. The highest density forecast regions were obtained from the densities on the original scale.

Figure 4 shows the 50% and 95% HDRs for these distributions. The means of the forecast densities are plotted as a solid line. The HDRs clearly show positive skewness and bimodality from time 203. The left hump becomes narrower until time 207 and then widens again so that by time 210 the humps corresponding to each mode have become sufficiently small that the 50% HDR contains both modes. The bimodality indicates uncertainty in the period of the next population cycle: the lower mode corresponds to the possibility of a minimum in the cycle and the larger mode corresponds to a possible maximum in the cycle. Standard forecast regions, usually obtained from quantiles, are unable to show this bimodality.

2.3 HDRs for multivariate densities

It follows immediately from the definition that the boundary of an HDR consists of those values of the sample space with equal density. Hence a plot of a bivariate HDR is simply a form of contour plot. A bivariate HDR boxplot may be constructed using the 50% HDR and 99% HDR with points lying outside the 99% HDR displayed as in a scatterplot. The mode is marked by a circle (O).

Example: Azzalini and Bowman (1990) examine duration and waiting times for eruptions from the Old Faithful geyser in Yellowstone National Park, Wyoming. The data were collected continuously from August 1st until August 15th, 1985. Several similar data sets have also been analysed: Weisberg (1985) and Denby and Pregibon (1987) consider data collected in 1978 and 1979 and Cook and Weisberg (1982) consider data collected in 1980. These authors mainly consider the regression relationship between the waiting time between eruptions and the previous eruption duration.

The focus in this example is the bivariate density of the duration of each eruption and the duration of the previous eruption. There are 299 observations; the times are measured in minutes. Figure 5 shows a bivariate HDR boxplot of the data based on a kernel estimate of the density. This is similar to a plot in Scott (1991) showing the contours of an estimate of the analogous density for the 1978 data.

Figure 5 shows that eruptions tend to be either long (around 4 minutes) or short (around 2 minutes) but rarely of medium length (around 3 minutes). Furthermore, there are never consecutive short eruptions. These observations are not new; previous studies of Old Faithful data have noted similar phenomena and Azzalini and Bowman present a tentative physical model for these eruption patterns.

The striking features of the Old Faithful data are much more obvious from the HDR boxplot than from other bivariate boxplots which have been proposed. Becketti and Gould (1987) proposed a form of bivariate boxplot which simply superimposes the boxplots of the marginal variables on the scatterplot. As such, it fails to capture any correlation, let alone the unusual structure seen in Figure 5. Goldberg and Iglewicz (1992) propose bivariate boxplots in which both the inner region and outer region are constructed using four quarter-ellipses. While their boxplot will show the correlation and possibly the lack of data in the lower left of the plot, it cannot reveal the presence of the three modes since both regions must be convex. If the data are from a unimodal density, then the boxplots of Goldberg and Iglewicz (1992) will yield very similar results to HDR boxplots. Hence, HDR bivariate boxplots are suitable for a greater variety of data than the alternative bivariate boxplots.

In three dimensions, a highest density region is a shell in three dimensional space. If the density is trivariate normal, the HDRs are nested hyperellipsoids. David Scott (Scott, 1991; Scott, 1992) has proposed a similar concept which he calls an α-shell defined by the surfaces $S_\alpha(x, y, z) := \{(x, y, z) : f(x, y, z) = \alpha f(\text{mode})\}$. In fact, these are HDRs under a different parameterization to that used in this article.

For d-dimensional densities where $d > 3$, Scott proposes plotting a three dimensional shell after conditioning on $d-3$ of the variables. Computer animation is also possible where the probability coverage is changed through time or, for $d > 3$, one of the conditioned variables is changed through time.

Example: Figure 6 shows the 70% HDR from the trivariate density computed from the duration of the Old Faithful geyser data and its first two lagged variables. If L denotes a long duration and S a short duration, then the five shells correspond to the situations SLS, SLL, LSL, LLS and LLL. The situations SSL, LSS and SSS do not occur as two short eruptions are never consecutive.

This example does not reveal any new information about the geyser data that was not apparent in Figure 5. However, it shows clearly the value of HDR plots in identifying clusters in three dimensional data. Using a three-dimensional scatterplot, even with spinning, it is very difficult to spot the five clusters since they tend to overlap when projected onto a two-dimensional plane.

3 Calculation of HDRs

For discrete valued distributions, HDRs simply consist of those elements of the sample space with highest probability. Therefore computation is simple.

3.1 Numerical integration approach

For continuous distributions, there have been several suggestions for constructing an HDR from a general univariate density $f(x)$, which is a bounded, continuous function of x. Wright (1986) proposed an algorithm involving numerical integration of $f(x)$ but he assumed the density was unimodal and so restricted the HDR to a single interval. Hyndman (1990) developed a more general algorithm which computed an HDR for any given density where $f(x)$ is a bounded, continuous function of x and the inverse of $f(x)$ is uniquely defined in the neighborhood of the boundary of the HDR. Both Wright and Hyndman use various numerical methods to improve the speed of computation. However, neither of these algorithms is easily generalized to multivariate densities. One major problem is the computational difficulty in numerically integrating over a general region in high dimensional space. An alternative approach is required that avoids explicit integration.

3.2 Density quantile approach

Let $\boldsymbol{X}$ be a (possibly multivariate) random variable with density $f(\boldsymbol{x})$. Rather than use numerical integration under $f(\boldsymbol{x})$, we may obtain information about the probability coverage of a given region of the sample space using a Monte-Carlo technique.

Consider the random variable, $Y = f(\boldsymbol{X})$, obtained by the transforming $\boldsymbol{X}$ by its own density function. Now f_α is such that $\Pr(f(\boldsymbol{X}) \geq f_\alpha) = 1 - \alpha$. So f_α is the α-quantile of Y. Therefore f_α can be estimated as a sample quantile from a set of iid random variables with the same distribution as Y.

Let $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ denote a set of independent observations from the density $f(\boldsymbol{x})$. Then $\{f(\boldsymbol{x}_1), \ldots, f(\boldsymbol{x}_n)\}$ is a set of independent observations from the distribution of Y. Let $f_{(j)}$ be the jth largest of $\{f(\boldsymbol{x}_i)\}$ so that $f_{(j)}$ is the (j/n) sample quantile of Y. We shall use $f_{(j)}$ as an estimate of f_α. Specifically, we choose $\hat{f}_\alpha = f_{(j)}$ where $j = \lfloor \alpha n \rfloor$. Then $\hat{f}_\alpha \to f_\alpha$ as $n \to \infty$ and so $R(\hat{f}_\alpha) \to R(f_\alpha)$ as $n \to \infty$.

If $f(\boldsymbol{x})$ is a known function, the observations will usually be generated pseudo-randomly. The approximation can be made arbitrarily accurate by increasing n.

Often, however, we will not know the density $f(\boldsymbol{x})$ but will have a set of i.i.d. observations $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_m\}$ from an unknown density. In this case, $f(\boldsymbol{x})$ may be estimated empirically from $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_m\}$ (see, for example, Scott, 1992). If m is large, there is no need to generate the sample $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ as the actual observations $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_m\}$ may be used directly by setting $n = m$ and $\boldsymbol{x}_i = \boldsymbol{y}_i$. If m is moderate, it may be preferable to generate "observations" $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ pseudo-randomly from the density estimate. For small m, it may not be possible to get a reasonable density estimate. Besides, with few observations and no prior knowledge of the underlying density, there seems little point attempting to summarise the sample space. See Wand and Jones (1995) for some discussion on the number of observations needed for a reasonable density estimate. Note that the problem here is not with the density quantile algorithm (which will give results to an arbitrary degree of accuracy given a density), but with estimating the density from insufficient data.

A crude form of the density quantile algorithm can also be found in Tanner (1993, pp.70–71) in the context of a posterior density from a data-augmentation algorithm.

The density quantile algorithm was used to estimate the HDRs displayed in Section 2.

3.3 Confidence regions for estimated HDRs

Because the density quantile algorithm involves a statistical approximation, it is helpful to compute some uncertainty limits on the estimated regions. Here I only consider the case where $\boldsymbol{X}$ is univariate.

We first obtain the distribution of $Y = f(X)$. Let $\{z_i\}$ denote those points in the sample space of X such that $f(z_i) = y$, $i = 1, 2, \dots, n(y)$. That is, $\{z_i\}$ denote the endpoints of the sub-intervals which make up $R(y)$. Also define $A(y) = \int_{R(y)} f(u)du$ so that $A(f_\alpha) = 1 - \alpha$. Then for small δ,

$$A(y + \delta) = A(y) - \delta y \sum_{i=1}^{n(y)} |f'(z_i)|^{-1} + O(\delta^2)$$

and so

$$\frac{dA(y)}{dy} = -y \sum_{i=1}^{n(y)} |f'(z_i)|^{-1} .$$

Now $\Pr(Y \le y) = 1 - A(y)$. Therefore the density of Y is simply

$$g(y) = -\frac{dA(y)}{dy} = y \sum_{i=1}^{n(y)} |f'(z_i)|^{-1} .$$

Then, using the standard asymptotic results for a sample quantile (e.g., Cox and Hinkley, 1974) we have that $\hat{f}_\alpha$ is asymptotically normally distributed with mean f_α and variance $\frac{\alpha(1-\alpha)}{n[g(f_\alpha)]^2}$. We can use this result to obtain a confidence interval for f_α, say $[f_L, f_U]$. Then $R(f_L)$ and $R(f_U)$ represent lower and upper confidence regions for $R(f_\alpha)$.

The top two plots of Figure 7 show 50% and 90% HDRs for the mixture density of Figure 1. The estimate of f_α and a 95% confidence interval for $\hat{f}_\alpha$ are shown as horizontal lines. The HDR estimate is shown as the dark shaded region and the values of $R(f_L)$ and $R(f_U)$ are shown with lighter shading. Here, $n = 200$ has been used to compute the HDR.

To assess the effect of using an empirical density estimate rather than the actual density, $f(x)$, analogous plots are shown in the bottom half of Figure 7. Here the density was estimated using a kernel estimator from the 200 observations. The additional variability in the HDR estimate due to the density estimation seems to be relatively small.

3.4 Computational details and timing

The density quantile algorithm and the computation of confidence regions has been implemented in Splus. For example, Figure 8 gives code for producing a bivariate boxplot as displayed in Figure 5. Here, the data are used directly in the algorithm by setting $n = m$ and $\boldsymbol{x}_i = \boldsymbol{y}_i$. This function can be easily modified to produce other HDRs or to compute the density from a known function rather than use an empirical estimate.

Note that two calls must be made to the function `density2d`, one to compute the density at each of the observation points and one to compute the density on a 20×20 grid. Linear interpolation is used between the grid points when computing the contour.

Most of the computational work in this function is in estimating the density using `density2d`. Hence, the speed of the function depends largely on the speed of `density2d`. Using a relatively slow implementation of bivariate kernel density estimation, Figure 5 was produced using this function in 6.1 seconds of processor time on Splus 3.1 running on a DECstation 5000/25. With a faster implementation of density estimation, using the ideas of Fan and Marron (1994), this time should be able to be reduced substantially.

Ironically, it is more difficult to plot a univariate HDR than a bivariate HDR because there is no equivalent of the `contour` function in one dimension. Therefore, a new function has been written for this purpose. Again, the density is computed (or estimated) on a grid of values; between the grid points, spline interpolation is used. Table 1 shows the processor times (in seconds) to compute an HDR and its associated 95% confidence regions using a DECstation 5000/25 running Splus 3.1. The mixture density of Figure 1 was used and measurements were made for different values of n

and α. Because the density is known in this case, kernel density estimation was not necessary.

n	α			
	0.5	0.25	0.05	0.01
10,000	1.38	1.64	1.45	1.36
1,000	1.01	1.28	1.19	1.02
200	1.09	1.25	0.85	0.76

Table 1: *Processor times (in seconds) for Splus to compute an HDR and its associated 95% confidence regions for the mixture density of Figure 1.*

The Splus code used to find HDRs and produce HDR boxplots in one and two dimensions can be obtained from the S archive of statlib@lib.stat.cmu.edu.

References

AZZALINI, A. and BOWMAN, A.W. (1990), A look at some data on the Old Faithful geyser, *Appl. Statist.*, **39**, 357–365.

BECKETTI, S and GOULD, W. (1987), Rangefinder box plots: a note, *Amer. Statist.*, **41**, 149.

BENJAMINI, YOAV (1988), Opening the box of a boxplot, *Amer. Statist.*, **42**, 257–262.

BOX, G.E.P. and TIAO, G.C. (1973), *Bayesian inference in statistical analysis*, (Addison-Wesley: Reading, Massachusetts).

BRILLINGER, D.R., GUCKENHEIMER, J., GUTTORP, P.E., and OSTER, G. (1980), Empirical modelling of population time series data: the case of age and density dependent vital rates, volume 13 of *Lectures on Mathematics in the Life Sciences*, pages 65–90. (Amer. Math. Soc.).

COOK, R.D. and WEISBERG, S. (1982), *Residuals and influence in regression*, (Chapman and Hall: London).

COX, D.R. and HINKLEY, C.V. (1974), *Theoretical Statistics*, (Chapman and Hall: London).

DENBY, L. and PREGIBON, D. (1987), An example of the use of graphics in regression, *Amer. Statist.*, **41**, 33–38.

ESTY, W.W. and BANFIELD, J.D. (1992), The box-percentile plot, Technical report, Department of Mathematical Sciences, Montana State University.

FAN, J. and MARRON, J.S. (1994), Fast implementations of nonparametric curve estimators, *J. Comp. Graph. Statist.*, **3**, 35–56.

GOLDBERG, K.M. and IGLEWICZ, B. (1992), Bivariate extensions of the boxplot, *Technometrics*, **34**, 307–320.

HOAGLIN, D.C., IGLEWICZ, B., and TUKEY, J.W. (1986), Performance of some resistance rules for outlier labeling, *J. Amer. Statist. Assoc.*, **81**, 991–999.

HYNDMAN, R.J., BASHTANNYK, D.M., and GRUNWALD, G.K. (1994), Estimating and visualizing conditional densities, Research report, The University of Melbourne. Submitted to *J. Comp. Grap. Statist.*

HYNDMAN, R.J. (1990), An algorithm for constructing highest density regions. Unpublished manuscript available from the author.

Hyndman, R.J. (1993), Highest density forecast regions for non-linear and non-normal time series models, Research report, The University of Melbourne. To appear in *J. Forecasting*.

McGill, R., Tukey, J.W., and Larsen, W.A. (1978), Variations of box plots, *Amer. Statist.*, **32**, 12–16.

Nicholson, A.J. (1957), The self-adjustment of populations to change, *Cold Spring Harbour Symp. Quant. Biol.*, **22**, 153–173.

Scott, D.W. (1991), On estimation and visualization of higher dimensional surfaces, in Buja, A. and Tukey, P.A., editors, *Computing and graphics in statistics*, IMA vol. 36 in mathematics and its applications. (Springer–Verlag).

Scott, D.W. (1992), *Multivariate density estimation: theory, practice, and visualization*, (Wiley: New York).

Tanner, M.A. (1993), *Tools for statistical inference: methods for the exploration of posterior distributions and likelihood functions*, (Springer-Verlag: New York), second edition.

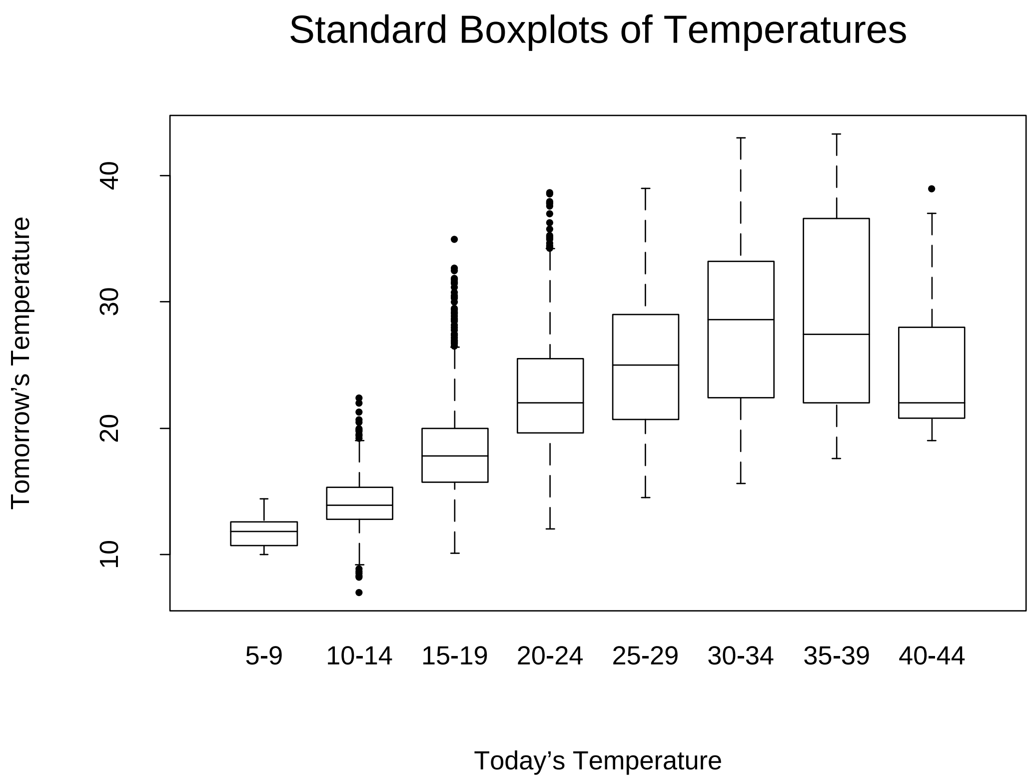
Tong, H. (1988), Non-linear time series modelling in population biology: a preliminary case study, in *Nonlinear time series and signal processing*, volume 106 of *Lecture Notes in Control and Information Sciences*, pages 75–87. (Springer–Verlag).

Tukey, J.W. (1977), *Exploratory data analysis*, (Addison-Wesley: Reading).

Wand, M.P. and Jones, M.C. (1995), *Kernel smoothing*, (Chapman and Hall: London).

Weisberg, S. (1985), *Applied linear regression*, (John Wiley & Sons: New York), second edition.

Wright, D.E. (1986), A note on the construction of highest posterior density intervals, *Appl. Statist.*, **35**, 49–53.



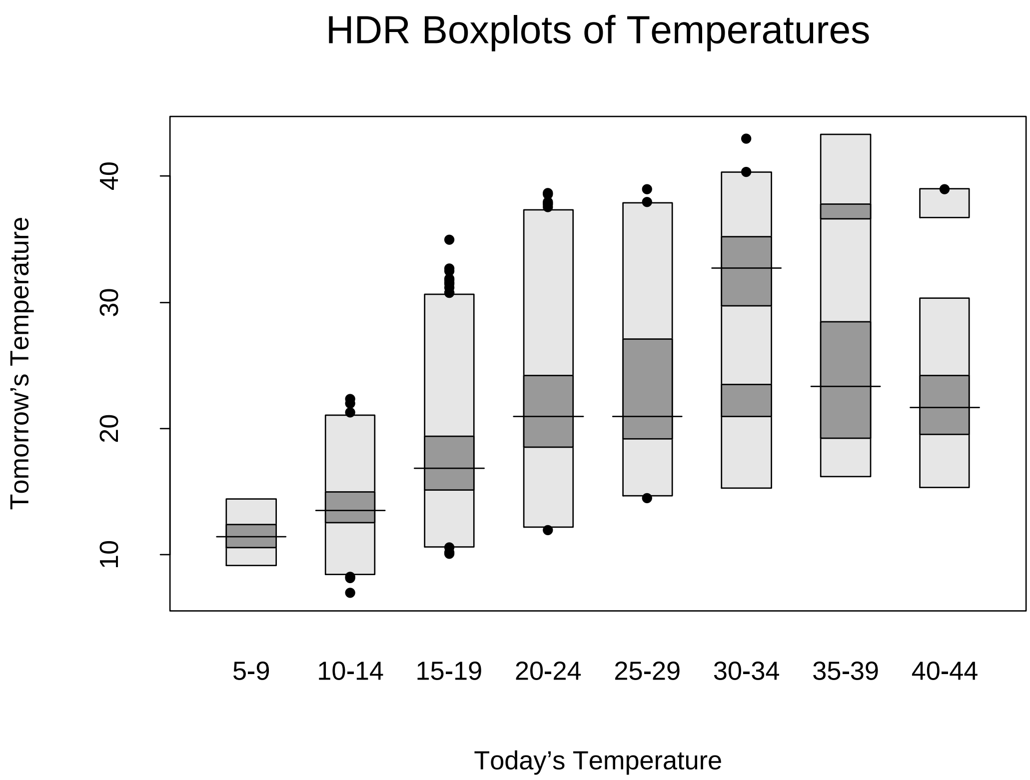


Figure 2: *Boxplots for the daily maximum temperature in Melbourne Australia between 1981 and 1990. The bimodality of the distribution following a hot day, which is due to the possible onset of a cool change, is only visible with the HDR boxplots.*

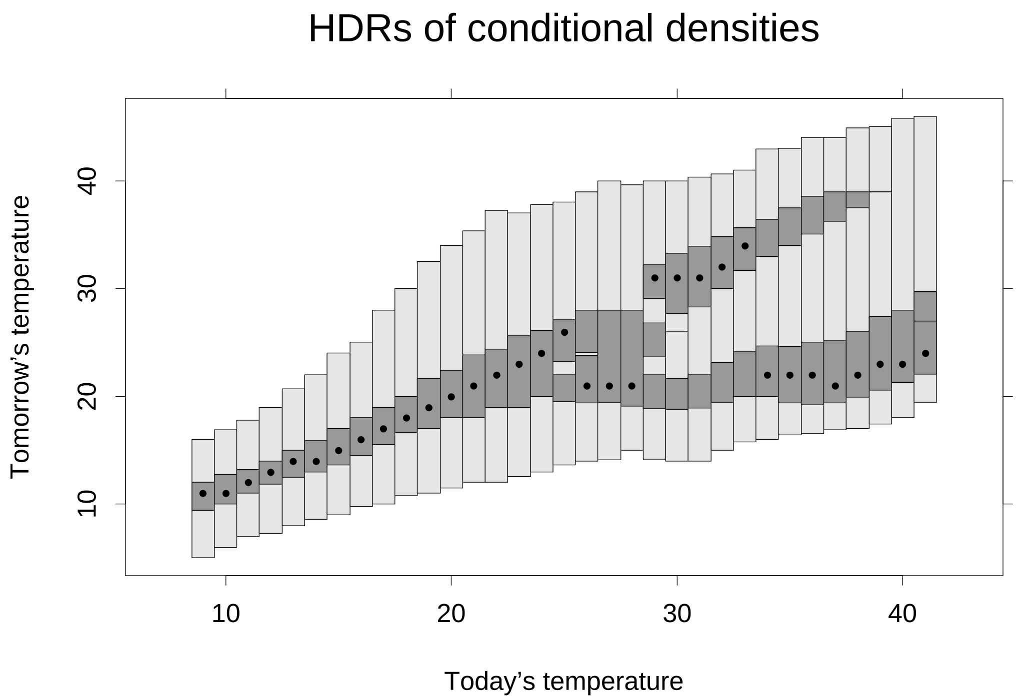
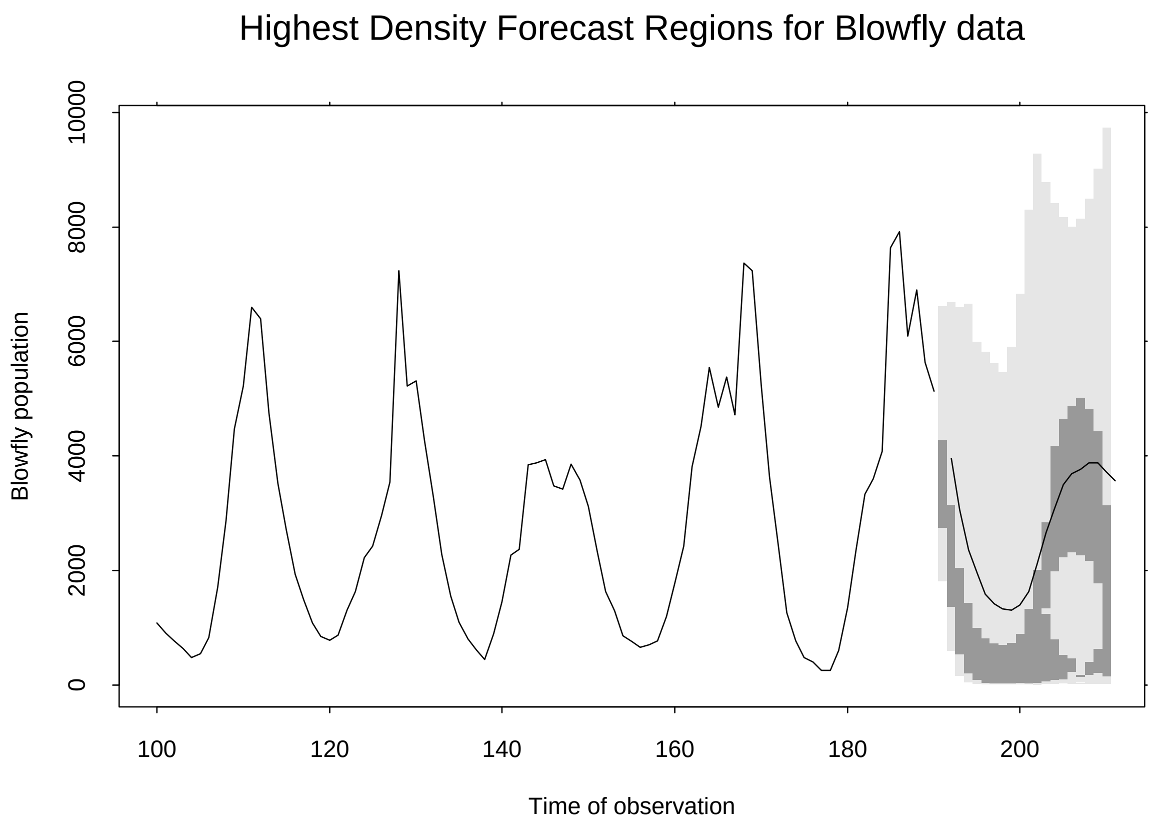


Figure 3: *An extension of Figure 2. Highest density regions (50% and 99%) for tomorrow's temperature conditional on today's temperature are displayed. The conditional modes are marked by • for each conditioning value.*


Highest Density Forecast Regions for Blowfly data
Blowfly population
0
2000
4000
6000
8000
10000
100
120
140
160
180
200
Time of observation

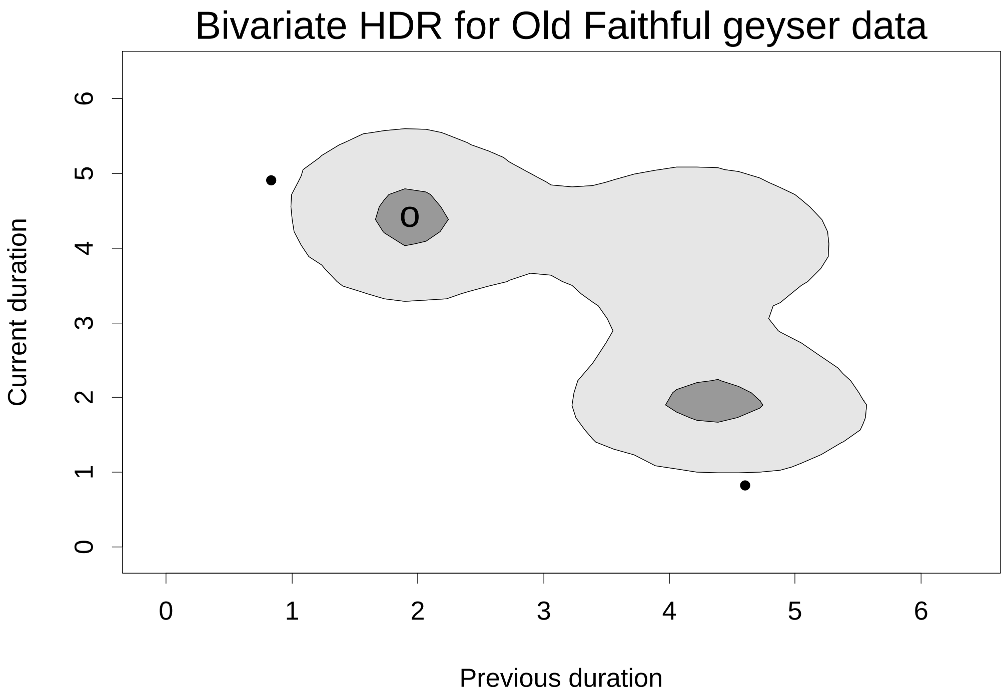
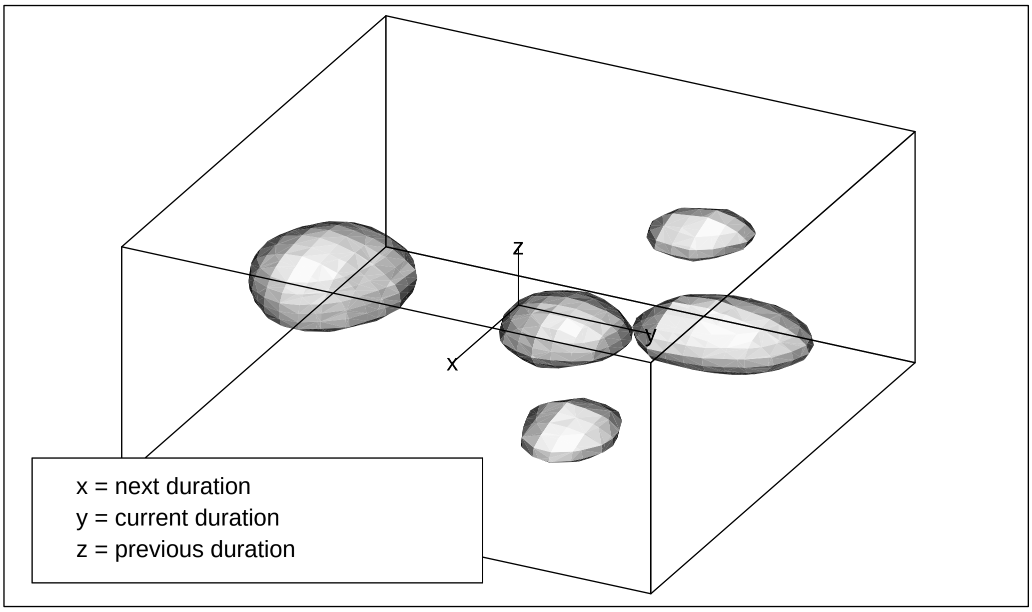


Figure 5: *A bivariate HDR boxplot for the Old Faithful geyser data. The duration of each eruption is one the vertical axis, the duration of the previous eruption is on the horizontal axis. Times are measured in minutes.*


z
x
y
x = next duration
y = current duration
z = previous duration

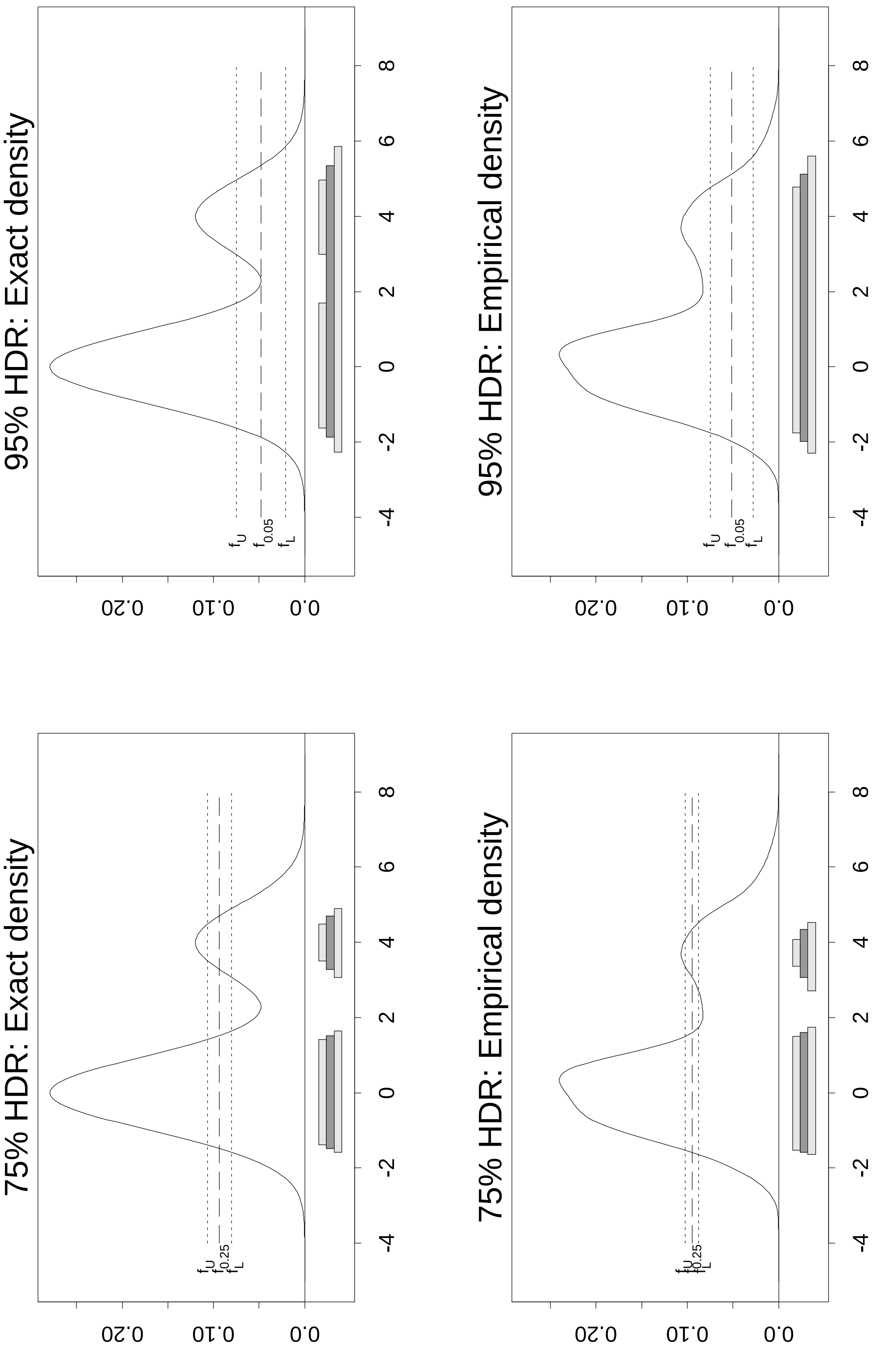


Figure 7: *50% and 75% HDRs for the mixture density of Figure 1 with 95% confidence regions.*

```
hdrboxplot <- function(x, y, a, b, ...)
# Input: x and y are independent observations on a density f(x,y).
# Call: density2d(x,y,x0,y0,a,b) returns a vector containing the
#       estimated density of (x,y) at the points x0,y0 using the
#       smoothing parameters a and b.
# Output: bivariate HDR boxplot
{
    fxy <- density2d(x, y, x, y, a, b)
    falpha <- quantile(fxy, c(0.01, 0.5))
    range.x <- diff(range(x))
    range.y <- diff(range(x))
    grid <- expand.grid(list(
           x = seq(min(x)-0.2*range.x, max(x)+0.2*range.x, length=20),
           y = seq(min(y)-0.2*range.y, max(y)+0.2*range.y, length=20)))
    fxy.grid <- density2d(x, y, grid$x, grid$y, a, b)
    junk <- contour(interp(grid$x,grid$y,fxy.grid),levels=falpha,
              xlab=deparse(substitute(x)), ylab=deparse(substitute(y)),
              labex=0, save=T, plotit=T, ...)
    polygon(junk[[1]]$x, junk[[1]]$y, col=4)
    polygon(junk[[2]]$x, junk[[2]]$y, col=2)
    index <- fxy < 0.999*falpha[1]
    points(x[index], y[index])
    index <- (1:length(x))[fxy==max(fxy)]
    points(x[index],y[index],pch="o")
}
```

Figure 8: *Splus code for bivariate data: computes* $\hat{f}_\alpha$ *using the density quantile algorithm and plots a HDR boxplot.*