Details on the standard error of a special density estimate

C.J. Albers* W. Schaafsma*

Abstract

De Bruin *et al.*[2] provides a unique nonparametric method to estimate the probability density f from a sample, given an initial guess ψ of f. This report provides an approximation to the standard error of this estimate $f_n(x)$ of $f(x)$.

The smoothed analogues $f_n^{(m)}$ of f_n introduced in Albers *et al.*[1] perform better with respect to the rate of convergence. Numerical experiences are favorable, but a satisfactory theoretical analysis seems to be impossible.

The notation of Albers *et al.*[1] is used; this report can be seen as an 'appendix'.

1 Introduction

For a large variety of reasons, it is of interest to pay attention to nonparametric density estimation methods other than the familiar kernel methods. See, e.g., Silverman [4] for motivation and, also, De Bruin *et al.*[2] and Albers *et al.*[1] for more specific motivation.

This report discusses the asymptotic behaviour of the estimates introduced in De Bruin *et al.*[2] and adapted in Albers *et al.*[1].

In Section 2 it is established that the quantile density estimates $b_n(p)$ used in De Bruin *et al.*[2] can be equipped with the standard error

$$\frac{b_n(p)}{\sqrt[4]{4\pi np(1-p)}}$$

A similar result for the probability density estimate $f_n(x)$ is not (completely) attainable. It is of interest to remark that the estimates f_n do display a regular behavior, but perhaps less regular than suggested by De Bruin *et al.*: the error

This report will be published as part of the first author's forthcoming thesis.

*Department of Mathematics and Computing Science, University of Groningen, P.O. Box 800, 9700 AV Groningen, Netherlands. E-mail: c.j.albers@math.rug.nl

in the derivative $f_n' - f'$ does not vanish; it is of order $O(n^{1/4})$ (Section 3). In this regard, kernel estimates provide better results because (under some weak conditions) their derivatives do converge to the real f' (Silverman[4]).

In spite of the divergence of the derivative, f_n does converge to f if $\psi = f$. For the case $\psi \neq f$ we have not been able to complete the proof.

Unfortunately, the rate of convergence $O(n^{-1/4})$ for b_n (and, approximately, for f_n) is too small to be satisfactory. That is why Albers *et al.*[1] replace $f_n(x)$ by a smoothed version $f_n^{(m)}(x)$. Numerical results indicate better performance, and so do some 'theoretical remarks'. The rate of convergence will be improved but an asymptotically correct expression for the standard error is not available at the moment, though a practical approximation is presented. An extensive numerical analysis of this new estimator is discussed in Albers *et al.*[1].

2 The asymptotic distribution of $\mathbf{b_n(p)}$

In this section the asymptotic distribution of b_n will be considered. It will be established that $n^{1/4}(b_n(p) - b(p))$ has a limiting normal distribution with zero-mean, and a variance that does not depend on Ψ. To prove this result, we will use two lemmas involving the asymptotic behaviour of b_n.

The values $y_{[1]}, \ldots, y_{[n]}$ are ordered outcomes of an independent random sample from the distribution on $[0,1]$ with distribution function $F \circ \Psi^{-1} = B^{-1}$. If we define $u_{[i]} = F(\Psi^{-1}(y_{[i]})) = B^{-1}(y_{[i]})$, then we have that $(u_{[1]}, \ldots, u_{[n]})$ is the ordered outcome of an independent random sample from the uniform distribution on $[0,1]$. Hence, replacing $y_{[i]}$ by $B(u_{[i]})$, we obtain, by applying the Mean Value Theorem, that $v_i \in [u_{[i]}, u_{[i+1]}]$ $(i = 0, \ldots, n)$ exist such that

$$\begin{aligned} b_n(p) &= \sum_{i=0}^{n} \left(B(u_{[i+1]}) - B(u_{[i]})\right)(n+1)\binom{n}{i} p^i (1-p)^{n-i} \\ &= \sum_{i=0}^{n} b(v_i)\left(u_{[i+1]} - u_{[i]}\right)(n+1)\binom{n}{i} p^i (1-p)^{n-i}. \end{aligned}$$

The expectations $b_{n,i} = \binom{n}{i} p^i (1-p)^{n-i}$ of the weights of the $b(v_i)$ are largest if $i \approx np$ and, hence, $v_i \approx p$. This suggests to replace $b(v_i)$ by $b(p)$ Using the approximation $\tilde{b}_n(p)$ of $b_n(p)$ thus obtained, we shall have to study the error $\tilde{b}_n(p) - b_n(p)$.

This will be done in Lemma 2.2. We start out with the following result.

Lemma 2.1 *With respect to $\tilde{b}_n(p)$ just defined, we have:*

$$\mathcal{L}\, n^{1/4}(\tilde{b}_n(p) - b(p)) \to \mathcal{N}\left(0, \frac{b(p)^2}{\sqrt{4\pi p(1-p)}}\right).$$

Proof. Note that the coverages $u_{[i+1]} - u_{[i]}$ are outcomes of random variables $U_{[i+1]} - U_{[i]}$ which have the same joint distribution as the r.v.'s $E_i/(E_0 + \ldots + E_n)$ where $E_0, \ldots, E_n$ constitute an independent random sample from the standard negative-exponential distribution. Hence

$$\mathcal{L}\, \tilde{b}_n(p) = \mathcal{L}\left(\frac{b(p) \sum_{i=0}^n E_i \binom{n}{i} p^i (1-p)^{n-i}}{\sum_{i=0}^n E_i/(n+1)} \right).$$

The denominator on the right-hand side has a Gamma$(n+1, n+1)$-distribution. As it converges to 1 (in probability), it suffices to show that

$$\mathcal{L}\left(n^{1/4} b(p) \left(\sum_{i=0}^n E_i b_{n,i} - 1 \right) \right) \to \mathcal{N}\left(0, \frac{b(p)^2}{\sqrt{4\pi p(1-p)}}\right),$$

where $b_{n,i} = \binom{n}{i} p^i (1-p)^{n-i} = \mathrm{P}(S = s)$ if S has the binomial distribution $\mathrm{B}(n,p)$. The expectation of $\sum E_i b_{n,i}$ is obvious (that of E_i is equal to 1 and $\sum b_{n,i} = 1$). The variance of it is equal to

$$\mathrm{Var}\left(\sum_{i=0}^n E_i b_{n,i} \right) = \sum_{i=0}^n b_{n,i}^2 = \mathrm{P}(S_1 - S_2 = 0),$$

where $S_1, S_2 \sim \mathrm{B}(n,p)$ are independent, and $S_1 - S_2$ has a distribution on $\{-n, -n+1, \ldots, n\}$ which is asymptotically $\mathcal{N}(0, 2np(1-p))$. Using a local form of the CLT, we have that

$$\mathrm{P}(S_1 - S_2 = 0) \approx \frac{1}{\sqrt{2\pi \cdot 2np(1-p)}}$$

What remains will follow from the Lyapunov Central Limit Theorem. This theorem implies

$$\mathcal{L}\left(\frac{\sum_{i=0}^n E_i b_{n,i} - 1}{\sqrt{\mathrm{Var}(\sum_{i=0}^n E_i b_{n,i})}} \right) \to \mathcal{N}(0,1)$$

if the condition

$$\frac{\mathbf{E} \sum_{i=0}^n |(E_i - 1) b_{n,i}|^3}{\{\mathrm{Var}(\sum_{i=0}^n E_i b_{n,i})\}^{3/2}} \to 0 \quad \text{as} \quad n \to \infty$$

is satisfied. The asymptotic behaviour of the denominator has been obtained in the above as $(\sqrt{4\pi n p(1-p)})^{-3/2}$, it suffices to establish that the numerator is $o(n^{-3/4})$ or, equivalently, that $\sum b_{n,i}^3$ behaves that way. Note that, using notations similar to before,

$$\sum_{i=0}^n b_{n,i}^3 = \mathrm{P}(S_1 = S_2 = S_3) = \mathrm{P}(S_1 - S_2 = S_1 - S_3 = 0)$$

where $(S_1-S_2, S_1-S_3)'$ has a distribution on $\{(i,j) \mid i,j \in \{-n,-n+1,\ldots,n\}\}$, asymptotically normal with vector of expectations $(0,0)'$, variances $2np(1-p)$ and covariance $np(1-p)$. Finally, using a local form of the multivariate CLT we obtain that

$$\begin{aligned} \mathrm{P}(S_1 - S_2 = S_1 - S_3 = 0) &\approx \left[2\pi(np(1-p))^2 \det \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\right]^{-1/2} \\ &= (2\pi np(1-p))^{-1} 3^{-1/2}. \end{aligned}$$

Hence the numerator behaves like n^{-1} which is obviously $o(n^{-3/4})$. □

Lemma 2.2 *Under weak regularity assumptions*

$$n^{1/4}(b_n(p) - \tilde{b}_n(p)) \to 0 \qquad \textit{in probability.}$$

Proof. Reformulate

$$n^{1/4}(b_n(p) - \tilde{b}_n(p)) = n^{1/4} \sum_{i=0}^{n} (b(V_i) - b(p))(U_{[i+1]} - U_{[i]})(n+1) b_{n,i},$$

where V_i is a random variable satisfying $U_{[i]} \leq V_i \leq U_{[i+1]}$. About f we know that it is smooth and strictly positive on a given interval (a,b). Almost equivalently, $b(p)$ will be assumed to be smooth and strictly positive. For the sake of convenience, the additional assumption is made that b is differentiable with $|b'(p)| \leq M \quad \forall p$ for some finite M. The Mean Value Theorem states that $n^{1/4}(b_n(p) - \tilde{b_n}(p))$ can be rewritten as

$$n^{1/4} \sum_{i=0}^{n} b'(W_i)(V_i - p)(U_{[i+1]} - U_{[i]})(n+1) b_{n,i}$$

with W_i between p and V_i. Hence

$$\begin{aligned} |n^{1/4}(b_n(p) - \tilde{b_n}(p))| &\leq M n^{1/4} \sum_{i=0}^{n} |V_i - p|(U_{[i+1]} - U_{[i]})(n+1) b_{n,i} \\ &\leq M n^{1/4} \sum_{i=0}^{n} (|U_{[i+1]} - p| + |U_{[i]} - p|)(U_{[i+1]} - U_{[i]})(n+1) b_{n,i} \end{aligned}$$

because $V_i \in [U_{[i]}, U_{[i+1]}]$. To establish that

$$n^{1/4} \sum_{i=0}^{n} |U_{[i+1]} - p|(U_{[i+1]} - U_{[i]})(n+1) b_{n,i} \to 0$$

in probability, it suffices to prove that the expectation of the left-hand side tends to 0. The other contribution follows similarly. The Cauchy-Schwartz inequality provides

$$\mathbf{E}\, |U_{[i+1]} - p|(U_{[i+1]} + U_{[i]}) \leq \sqrt{\mathbf{E}\, (U_{[i+1]} - p)^2} \frac{\sqrt{2}}{n+1},$$

because the 'coverage' $(U_{[i+1]} - U_{[i]}) \sim \text{Beta}(1, n)$ has expectation $\mathbf{E}\,(U_{[i+1]} - p)^2 = \frac{2}{(n+1)(n+2)} < \frac{2}{(n+1)^2}$. Hence it suffices to establish that

$$n^{1/4} \sum_{i=0}^{n} \sqrt{\mathbf{E}\,(U_{[i+1]} - p)^2}\, b_{n,i} \to 0,$$

where $U_{[i+1]} \sim \text{Beta}(i+1, n-i)$ is such that

$$\mathbf{E}\,(U_{[i+1]} - p)^2 = \frac{(i+1)(n-i)}{(n+2)(n+1)^2} \left(\frac{i+1}{n+1} - p \right)^2 .$$

The last step, as $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for nonnegative a and b, it suffices that both

$$n^{1/4} \sum_{i=0}^{n} \sqrt{\frac{(i+1)(n-i)}{(n+2)(n+1)^2}}\, b_{n,i} = n^{1/4} \mathbf{E} \sqrt{\frac{(S+1)(n-S)}{(n+2)(n+1)^2}} \to 0,$$

and

$$n^{1/4} \sum_{i=0}^{n} \left| \frac{i+1}{n+1} - p \right| b_{n,i} = n^{1/4} \mathbf{E} \left| \frac{S+1}{n+1} - p \right| \to 0.$$

Both statements are trivial, and, hence, the lemma holds. □

Theorem 2.1 *The quantile-density estimate defined in De Bruin* et al.*[2] has the following limiting density*

$$\mathcal{L}\; n^{1/4}(b_n(p) - b(p)) \to \mathcal{N}\left(0, \frac{b(p)^2}{\sqrt{4\pi p(1-p)}}\right).$$

Proof. This follows immediately from Lemma 2.1 and Lemma 2.2. □

3 The asymptotic distribution of $\mathbf{f_n(x)}$

Applications in discriminant analysis and pattern recognition require estimates of $f(x)$, not of $b(p)$. The relationship between these population concepts is as follows. Starting from the distribution function F of X_1 and the corresponding quantile function $H = F^{-1}$, we have that $Y_1 = \Psi(X_1)$ has distribution function $F \circ \Psi^{-1}$ and quantile function $B = \Psi \circ F^{-1}$. Hence $H = \Psi^{-1} \circ B$. For the densities $f = F'$, $h = H'$ and $b = B'$ we have that $f(x) = 1/h(F(x))$ and $h(p) = 1/f(H(p))$ while

$$h(p) = \frac{b(p)}{\psi(\Psi^{-1}(B(p)))}.$$

Hence

$$f(H(p)) = \frac{\psi(\Psi^{-1}(B(p)))}{b(p)}$$

and

$$f(x) = \frac{\psi(\Psi^{-1}(B(F(x))))}{b(F(x))} = \frac{\psi(x)}{b(F(x))}.$$

In Albers *et al.*[1] the sampling analogues are defined. Sampling analogues are derived accordingly, starting from $b_n = B_n'$.

Theorem 3.1 suggests that $n^{1/4}(f_n(x) - f(x))$ has a normal limiting distribution similar to that of $n^{1/4}(b_n(p) - b(p))$. Our proof contains a suggestion which cannot be made in general. However, if $\Psi = F$, then the required result can be established as follows.

Theorem 3.1 *If $\Psi = F$ then the difference between the density estimate defined in De Bruin* et al.*[2] and the true density $f = \psi$ satisfies*

$$\mathcal{L}\, n^{1/4}(f_n(x) - f(x)) \to \mathcal{N}\left(0, \frac{f(x)^2}{\sqrt{4\pi F(x)(1-F(x))}}\right).$$

Proof. Concentrate the attention on $f_n(x) - f(x)$:

$$f_n(x) - f(x) = \frac{\psi(x)\,[b(F(x)) - b_n(F_n(x))]}{b_n(F_n(x))\; b(F(x))},$$

which equals

$$\frac{\psi(x)\,[b_n(F(x)) - b_n(F_n(x))] - \psi(x)\,[b_n(F(x)) - b(F(x))]}{b_n(F_n(x))\; b(F(x))}.$$

Theorem 2.1 implies that

$$\mathcal{L}\, n^{1/4}(b_n(F(x)) - b(F(x))) \to \mathcal{N}\left(0, \frac{b(F(x))^2}{\sqrt{4\pi F(x)(1-F(x))}}\right).$$

Furthermore, not spelling out the details, we have

$$\frac{\psi(x)^2}{b_n(F_n(x))^2 b(F(x))^2}\,\frac{b(F(x))^2}{\sqrt{4\pi F(x)(1-F(x))}} \approx \frac{f(x)^2}{\sqrt{4\pi F(x)(1-F(x))}}.$$

and complete the proof by establishing, in the following lemma, that $b_n(F(x)) - b_n(F_n(x))$ is of smaller order of magnitude than $b_n(F(x)) - b(F(x))$, which is $O(n^{-1/4})$. □

Lemma 3.1 *If $\Psi = F$ then $b_n(F_n(x)) - b_n(F(x))$ is $o(n^{-1/4})$.*

Proof. The proof is, in large parts, similar to that of Lemma 2.1. A difficulty is that $b'_n(p)$ does *not* tend to $b'(p)$, the difference being $O(n^{1/4})$. Compare $b_n(p)$ with $b_n(r)$ $(0 \le p, r \le 1)$.

$$b_n(p) - b_n(r) = \sum_{i=0}^{n} (u_{[i+1]} - u_{[i]})(n+1)(\mathrm{P}(B_{n,p} = i) - \mathrm{P}(B_{n,r} = i))$$

where $B_{n,p} \sim \mathrm{Bin}(n, p)$ and the $u_{[i]}$ are ordered outcomes of the standard uniform distribution. This equation can be considered as the outcome of

$$\frac{\sum_{i=0}^{n} E_i(\mathrm{P}(B_{n,p} = i) - \mathrm{P}(B_{n,r} = i))}{\sum_{i=0}^{n} E_i/(n+1)}$$

with $E_0, \ldots, E_n$ a random sample from the standard negative exponential distribution. Remember from the previous section that the denominator is practically 1. The expectation of the numerator is

$$\sum_{i=0}^{n} \mathrm{P}(B_{n,p} = i) - \sum_{i=0}^{n} \mathrm{P}(B_{n,r} = i) = 0,$$

its variance is

$$\begin{aligned}
& \sum_{i=0}^{n} (\mathrm{P}(B_{n,p} = i) - \mathrm{P}(B_{n,r} = i))^2 \\
= \; & \mathrm{P}(B^{(1)}_{n,p} = B^{(2)}_{n,p}) + \mathrm{P}(B^{(1)}_{n,p} = B^{(2)}_{n,r}) - 2\mathrm{P}(B^{(1)}_{n,p} = B^{(1)}_{n,r}) \\
\approx \; & \frac{1}{\sqrt{4\pi np(1-p)}} + \frac{1}{\sqrt{4\pi nr(1-r)}} - 2\frac{\exp\left[-\frac{1}{2}\left(\frac{p-r}{p(1-p)+r(1-r)}\right)^2\right]}{\sqrt{2\pi n(p(1-p)+r(1-r))}} \\
= \; & \frac{1}{\sqrt{4\pi n}}\left(\frac{1}{\sqrt{p(1-p)}} + \frac{1}{\sqrt{r(1-r)}} - \frac{2\sqrt{2}}{\sqrt{p(1-p)+r(1-r)}}\right) \\
& + \frac{2}{\sqrt{4\pi n}} \frac{\exp\left[-\frac{1}{2}\left(\frac{p-r}{p(1-p)+r(1-r)}\right)^2\right] - 1}{\sqrt{p(1-p)+r(1-r)}}.
\end{aligned}$$

For the first terms, see the proof of Lemma 2.1. The last term follows from $B^{(1)}_{n,p} - B^{(1)}_{n,r} \sim \mathcal{N}(n(p-r),\ n(p(1-p)+r(1-r)))$.

As $p = F_n(x)$ tends to $r = F(x)$, the entire expression for the variance is, indeed, $o(n^{-1/2})$ □

Now, let's see what happens when the restriction $\Psi = F$ is abandoned. Theorem 3.1 and Lemma 3.1 no longer apply, since they depend on the restriction in the step where $b_n(p) - b_n(r)$ is seen as outcome of the random variable specified.

Under certain regularity conditions for ψ (and f), it might be possible to prove a similar theorem for the general case. At the moment, we are in doubt whether the theorem holds in general.

However, our theory is developed for situations where the practical researcher has some good intuition about f. Numerical analysis [1] suggests that in those situations our method provides a useful contribution to existing density estimation theory. In these cases, and especially when Ψ and F are not too irregular, the behaviour of our density estimate will not be much different from that for the ideal situation, and the, now approximate, standard-error

$$\frac{f(x)}{\sqrt[4]{4\pi n F(x)(1-F(x))}}$$

will usually suffice.

4 The asymptotic distribution of $\mathbf{f_n^{(m)}}(\mathbf{x})$

When the first author tried to compute B_n, b_n and f_n for large values of n, he encountered numerical difficulties following from the large sample sizes. To overcome them he splitted the sample into two sub-samples of size $\frac{1}{2}n$ and took the arithmic average of the estimates. This turned out to be smoother and more accurate than the estimate based on the entire sample. A natural way of smoothing can be achieved by partitioning the sample into $k = m^{-1}n$ sub-samples of size m each, and taking the arithmic average of the resulting estimates.

As $m \to \infty$, $\mathcal{L}\, m^{1/4}(f_{m,h}(x) - f(x)) \approx \mathcal{N}(0, \sigma^2(x))$ implies that $\mathcal{L}\, m^{1/4}(\bar{f}_{n,m}(x) - f(x)) \approx \mathcal{N}(0, k^{-1}\sigma^2(x))$ or, equivalently, that the standard error of $\bar{f}_{n,m}(x)$ is $m^{-1/4}k^{-1/2}\sigma(x) = k^{-1/4}n^{-1/4}\sigma(x)$. (The status of these formulas is that they are useful suggestions. Note that $\sigma(x) = \frac{f(x)}{\sqrt[4]{4\pi F(x)(1-F(x))}}$).

To remove the permutation dependence, U-statistic symmetrization is used. The smoothed analogue

$$B_n^{(m)}(p) = \binom{n}{m}^{-1} \sum_{1 \le \alpha_1 < \ldots < \alpha_m \le n} B_m(p | y_{\alpha_1}, \ldots, y_{\alpha_m}),$$

$B_n(p)$ can be rewritten as the L-statistic

$$p^{m+1} + \sum_{j=1}^{m} \binom{m+1}{j} p^j (1-p)^{m+1-j} \sum_{i=j}^{n-m+j} \frac{\binom{i-1}{j-1}\binom{n-i}{m-j}}{\binom{n}{m}} y_{[i]}.$$

From this we get $F_n^{(m)} = {B_n^{(m)}}^{-1} \circ \Psi$, $f_n^{(m)} = {F_n^{(m)}}'$, etc. For $m \to \infty$, the $B_m(p|y_{\alpha_1}, \ldots, y_{\alpha_m})$ are consistent estimators of $(\Psi^{-1} \circ H)(p)$. For m fixed,

$B_n^{(m)}(p)$ is an unbiased estimate of $\mathbf{E}\ (B_m(p|X_1, \ldots, X_m))$ which, of course, depends on the underlying distribution function of X_1. In principle, the (exact) theory of Hoeffding[3] about U-statistics is applicable but the complexity of the formulas and the dependence on f, F, H, etc. precludes application in practice. The formula $k^{-1/4}n^{-1/4}\sigma(x)$ suggested for the standard error for the estimate $\bar{f}_{n,m}(x)$ suggests that the U-statistic symmetrization $f_{n,m}(x)$ will have an even better asymptotic performance. This suggests that the rate of convergence of $f_{n,m}$ is better than $n^{-3/8}$, which arises when $k = n^{1/2}$ is taken.

However, the bias of this estimator will be larger than of the original estimator $f_n(x)$. For a good comparison between this and other methods, one has to look at e.g. the Mean (Integrated) Squared Error. We prefer to rely on computational experiences because the theoretical analyses are both (too) complicated and too much dependent on irrelevant asymptotics (in practice, m is chosen to be less than 10 or 20 and asymptotic theory becomes questionable in such cases.

Remark. This Report, as well as the (submitted) article [1], can be downloaded from `http://www.math.rug.nl/∼casper`. Another way to retrieve the work is by contacting the first author (`casper@math.rug.nl`).

References

[1] C.J. Albers and W. Schaafsma. Estimating a density by adapting an initial guess. *submitted for publication*, 2001. Available as IWI preprint 2001-5-02.

[2] R. De Bruin, D. Salomé, and W. Schaafsma. A semi-Bayesian method for nonparametric density estimation. *Computational Statistics & Data Analysis*, 30:19–30, 1999.

[3] W. Hoeffding. A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics*, 19:293–325, 1948.

[4] B.W. Silverman. *Denstiy estimation for statistics and data analysis.* Chapman and Hall, London, 1986.