# AN APPROXIMATION USEFUL IN UNIVARIATE STRATIFICATION

By Gunnar Ekman

*Institute of Mathematical Statistics, University of Stockholm*

**1. Introduction.** The problem of minimizing a sum $\sum_1^n P_h \sigma_h$, where $P_h$ and $\sigma_h^2$ denote the area and conditional variance in the interval $(x_{h-1}, x_h)$ of a density $f(x)$, arises in the theory of optimum univariate stratification (see Dalenius, [1]). In [1] Dalenius shows that the sum $\sum_1^n P_h\sigma_h$ is minimized when the conditions

$$\frac{\sigma_h^2 + (x_h - \mu_h)^2}{\sigma_h} = \frac{\sigma_{h+1}^2 + (x_h - \mu_{h+1})^2}{\sigma_{h+1}} \tag{1}$$

are fulfilled, $(h = 1, \cdots (n - 1), x_0 = -\infty, x_n = +\infty)$, where $\mu_h$ denotes the conditional mean in the interval $(x_{h-1}, x_h)$.

In order to avoid the computational difficulties presented by determining $\{x_h\}$ such that the conditions (1) are satisfied, various approximations to (1) have been proposed. A brief summary of these results is given in [2]. In this article a new approximation will be derived and numerical examples will be given.

We shall show, that under certain conditions[1] and for a density over a finite range, points $\{x_h\}$ satisfying the equalities

$$(1)^{(a)} \qquad (x_h - x_{h-1})P_h = C_n, \qquad h = 1, 2, \cdots, n.$$

where $C_n$ is a constant dependent on $n$, approximately satisfy the minimal conditions (1). For a density over an infinite range the above is obviously not applicable, in which case, however, a certain modification can be made, substituting the conditions (12), (13) below for $(1)^{(a)}$. The basic result will be derived under the assumption of a large $n$, i.e. correspondingly small intervals $(x_h - x_{h-1})$; asymptotically as $n$ approaches infinity the conditions (1) and $(1)^{(a)}$ will be proven equivalent. In practice, of course, $n$ is often rather small, hardly greater than 4 or 5, which certainly does not fulfill this requirement of a large $n$. It is still possible that even for $n = 2, 3$, etc., the points satisfying $(1)^{(a)}$ provide a good approximation to (1) in the sense of a $\sum_1^n P_h\sigma_h$ near the minimal value. In order to ascertain whether this may be the case or not, actual numerical computation has been carried out for three densities and for $n = 2, 3, 4$, and 5. In the table under "Numerical Examples" the points $\{x_h\}$ obtained by applying $(1)^{(a)}$ or the substitute conditions (12), (13), may be

Received February 28, 1958; revised July 3, 1958.

[1] These conditions, which concern primarily the regularity of the density function $f(x)$, are imposed in order to facilitate the mathematical derivation of the final result, and should be of no interest or concern to the reader desirous only of applying this result in practice.

compared with the points satisfying the actual minimal conditions (1); below is given a comparison of the respective $[\sum P_h \sigma_h]^2$:

| $n$ | $f(x) = 2(1 - x)$ | | $f(x) = e^{-x}$ | | $f(x) = xe^{-x}$ | |
|---|---|---|---|---|---|---|
| | $(1)^{(a)}$ | $(1)$ | $(1)^{(a)}$ | $(1)$ | $(1)^{(a)}$ | $(1)$ |
| 2 | 0.01517 | 0.01505 | 0.2856 | 0.2855 | 0.6420 | 0.6370 |
| 3 | 0.00693 | 0.00688 | 0.1333 | 0.1332 | 0.3090 | 0.3075 |
| 4 | 0.00395 | 0.00393 | 0.0769 | 0.0768 | 0.1811 | 0.1804 |
| 5 | 0.00255 | 0.00254 | 0.0500 | 0.0500 | 0.1189 | 0.1185 |

It therefore seems not unlikely that the conditions $(1)^{(a)}$ are suitable substitutes for the exact conditions (1); the differences in the above table are seen to be comparatively small and decrease both absolutely and percentagewise as $n$ increases. It is nevertheless appropriate to mention a general type of density for which $(1)^{(a)}$ may be expected to give rather poor results, namely such $f(x)$ with a long (finite) "tail" (for example a $\chi^2$-distribution with a large number of d.f.) or with $f(x) = 0$ in an end point. In this case it is advisable to use (12), (13) unless these are difficult analytical expressions; in the case of an infinite "tail" this is imperative.

The determination of points $\{x_h\}$ satisfying $(1)^{(a)}$ or (12), (13) is by no means entirely free of computational difficulties, although these points are of course considerably easier to find than those satisfying (1). Trial and error or various iterative methods may suffice in some cases, but the need for a more systematic approach arises even for comparatively convenient expressions for $f(x)$. It is apparent that with a knowledge of the constant $C_n$ the determination of the points becomes trivial, as they may then be found simply by "fitting" with the aid of a table over the distribution function. An iterative method of finding $C_n$ is described in the paragraphs "First approximation to the constant $C_n$" and "Adaption to Numerical Computation". The mathematical arguments leading to $(1)^{(a)}$ following in the next paragraph are rather simple and straightforward, involving only the use of Taylor series and elementary algebraic manipulation. An outline of a method of finding the points satisfying the exact minimal conditions (1) concludes the paper.

**2. Approximation over a finite range.** Consider a density $f(x)$ over a finite range $(x_0, x_n)$. We assume that the derivatives $f'(x)$ and $f''(x)$ exist and are continuous over the whole range. We introduce a function $H(x)$ by defining

$$\begin{aligned} H'''(x) &= f(x), \\ H''(x) &= \int_{-\infty}^{x} H'''(t)\,dt, \\ H'(x) &= \int_{-\infty}^{x} H''(t)\,dt, \\ H(x) &= \int_{-\infty}^{x} H'(t)\,dt. \end{aligned}$$

$H'(x) \cdots H^{(5)}(x)$ exist therefore and are continuous. One finds by partial integration:

$$\int_a^b x^2 f(x)\,dx = 2[H(b) - H(a)] - 2[bH'(b) - aH'(a)] + [b^2H''(b) - a^2H''(a)],$$

$$\int_a^b xf(x)\,dx = -[H'(b) - H'(a)] + [bH''(b) - aH''(a)],$$

$$\int_a^b f(x)\,dx = [H''(b) - H''(a)].$$

By substituting the expressions in the right hand members of these equalities in

$$P_h\,\mu_h = \int_{x_{h-1}}^{x_h} xf(x)\,dx,$$

$$P_h[\sigma_h^2 + \mu_h^2] = \int_{x_{h-1}}^{x_h} x^2 f(x)\,dx,$$

where $P_h$, $\mu_h$, and $\sigma_h$ are the area, conditional mean and conditional variance of $f(x)$ in the interval $(x_{h-1}, x_h)$, we obtain by some calculations:[2]

$$P_h[\sigma_h^2 + (x_h - \mu_h)^2] = 2\left\{H(x_h) - \left[H(x_{h-1}) + (x_h - x_{h-1})H'(x_{h-1}) + \frac{(x_h - x_{h-1})^2}{2!}H''(x_{h-1})\right]\right\},$$

$$P_h[x_h - \mu_h] = \{H'(x_h) - [H'(x_{h-1}) + (x_h - x_{h-1})H''(x_{h-1})]\},$$

$$P_h = [H''(x_h) - H''(x_{h-1})].$$

We note that the right hand members consist of $H(x_h)$, $H'(x_h)$ and $H''(x_h)$ minus the first terms in their respective Taylor expansions about the point $x = x_{h-1}$. Continuing these expansions, the following results are obtained:

$$P_h[\sigma_h^2 + (x_h - \mu_h)^2] = 2\left[\frac{(x_h - x_{h-1})^3}{3!}H'''(x_{h-1}) + \frac{(x_h - x_{h-1})^4}{4!}H^{(4)}(x_{h-1}) + \frac{(x_h - x_{h-1})^5}{5!}H^{(5)}(\xi_1)\right], \tag{2}$$

$$P_h[x_h - \mu_h] = \left[\frac{(x_h - x_{h-1})^2}{2!}H'''(x_{h-1}) + \frac{(x_h - x_{h-1})^3}{3!}H^{(4)}(x_{h-1}) + \frac{(x_h - x_{h-1})^4}{4!}H^{(5)}(\xi_2)\right], \tag{3}$$

$$P_h = \left[\frac{(x_h - x_{h-1})}{1!}H'''(x_{h-1}) + \frac{(x_h - x_{h-1})^2}{2!}H^{(4)}(x_{h-1}) + \frac{(x_h - x_{h-1})^3}{3!}H^{(5)}(\xi_3)\right], \tag{4}$$

[2] These identities may also be obtained by partially integrating $\int_{x_{h-1}}^{x_h} (x_h - x)^k f(x)\,dx$, $k = 0, 1, 2$.

where $\xi_i$ are points in the interval $(x_{h-1}, x_h)$. By multiplying (2) by (4) and subtracting the square of (3), the following identity is obtained:

$$(5)\quad [P_h \sigma_h]^2 = \frac{(x_h - x_{h-1})^4}{12}\{H'''(x_{h-1})[H'''(x_{h-1}) + (x_h - x_{h-1})H^{(4)}(x_{h-1})] + R_1\}.$$

Squaring both members of (2) we obtain:

$$(6)\quad P_h^2[\sigma_h^2 + (x_h - \mu_h)^2]^2 = \frac{(x_h - x_{h-1})^6}{9}\Big\{H'''(x_{h-1})[H'''(x_{h-1}) + \frac{H^{(4)}(x_{h-1})}{2!}(x_h - x_{h-1})] + R_2\Big\}.$$

$R_1$ and $R_2$ are terms of the second order or higher in $(x_h - x_{h-1})$. For large $n$, that is for small intervals $(x_h - x_{h-1})$, these terms may be neglected, introducing thereby only a slight error. Substituting $f(x)$ and $f'(x)$ for $H'''(x)$ and $H^{(4)}(x)$, and dividing (6) by (5), the following approximative identity is obtained:

$$(7)\quad \left[\frac{\sigma_h^2 + (x_h - \mu_h)^2}{\sigma_h}\right]^2 \sim \frac{4(x_h - x_{h-1})}{3} \cdot \frac{f(x_{h-1})(x_h - x_{h-1}) + \frac{f'(x_{h-1})}{2!}(x_h - x_{h-1})^2}{f(x_{h-1}) + f'(x_{h-1})(x_h - x_{h-1})}.$$

In the numerator of the second factor of the right hand member of (7) we have the first two terms of the Taylor expansion of $P_h$ about the point $x = x_{h-1}$, whereas the denominator is the partial derivative of the numerator with respect to $x_h$, i.e. the first two terms of the expansion of $f(x_h)$ in the same point. Approximating once again by neglecting terms of greater order in the expansions of $P_h$ and $f(x_h)$, we obtain finally

$$(8)\quad \left[\frac{\sigma_h^2 + (x_h - \mu_h)^2}{\sigma_h}\right]^2 \sim \frac{4(x_h - x_{h-1})P_h}{3f(x_h)}.$$

Proceeding in entirely the same fashion the analogous expression

$$(9)\quad \left[\frac{\sigma_{h+1}^2 + (x_h - \mu_{h+1})^2}{\sigma_{h+1}}\right]^2 \sim \frac{4(x_{h+1} - x_h)P_{h+1}}{3f(x_h)}$$

may be derived.

Applying these results to the identity (1), one obtains

$$(10)\quad (x_h - x_{h-1})P_h \sim (x_{h+1} - x_h)P_{h+1}.$$

Applying (10) finally with $h = 1, \cdots, (n-1)$, we find an approximation to (1), namely

$$(11)\quad (x_h - x_{h-1})P_h = C_n,$$

where $C_n$ is a constant.

We have tacitly assumed $f(x) \neq 0$; the case $f(x) = 0$ will usually present itself only for $x = x_0$ or $x = x_n$, in which case (12), (13) below may be used instead of (11). When $f(x) \neq 0$, and assuming without loss of generality a range of unit length, we see from (7) that the first neglected term in the numerator is $O(x_h - x_{h-1})^4$, which implies the same degree of approximation in (8), (9). Accordingly the square roots of the members of (8), (9) differ by terms of the third order. This in turn implies both that the partial derivatives of $\sum P_h\sigma_h$ (being in fact $(f(x_h)/2)\{[\sigma_h^2 + (x_h - \mu_h)^2]/\sigma_h - [\sigma_{h+1}^2 + (x_h - \mu_{h+1})^2]/\sigma_{h+1}\})$ are $O(x_h - x_{h-1})^3 + O(x_{h+1} - x_h)^3$ in the points satisfying (11), and that the approximate $\{x_h\}$ must be adjusted by $\sum_1^n O(x_h - x_{h-1})^3 = O(x_j - x_{j-1})^2 = K_h$ in order to satisfy (1), since the members of (1) are of the first order. We deduce from these results and from the expansion to $\sum (a_h K_h^2 + b_h K_h K_{h+1})$ of $\sum P_h \sigma_h$ about the points satisfying (11) that the approximate and true minimal values of the sum differ by a sum of $n$ terms each $= O(x_j - x_{j-1})^4$, that is by $O(x_j - x_{j-1})^3$. The conditions (11) thus generate points $\{x_h\}$ differing from the true minimal points by terms of greater order than the interval lengths and result in an approximate minimal $\sum P_h\sigma_h$ differing from the true value by a term of greater order than the sum itself, i.e., $O(x_j - x_{j-1})$; these differences should furthermore decrease monotonically as $n$ increases, that is as $(x_j - x_{j-1})$ decreases. These conclusions are immediately extended to comprise even the case $f(x) = 0$ in the end points, and are therefore valid for any reasonable density over a finite range. By a truncation argument we may immediately ascertain the asymptotic equivalence of the sums, when (12), (13) below are used to approximate (1), even over an infinite range, whereas in this case the two sets of points do not necessarily converge, as may be seen by taking $f(x) = e^{-x}$.

**3. Approximation over an infinite range.** If $x_0 = -\infty$, $x_n = +\infty$, the approximation (11) can still be applied for $h = 2, \cdots, (n-1)$. The identities (8) and (9) with $h = 1$ and $(n-1)$ respectively suggest putting

$$(12) \qquad (x_h - x_{h-1})P_h = C_n, \qquad h = 2, \cdots, (n-1),$$

$$(13) \qquad \frac{3f(x_1)}{4}\left[\frac{\sigma_1^2 + (x_1 - \mu_1)^2}{\sigma_1}\right]^2 = \frac{3f(x_{n-1})}{4}\left[\frac{\sigma_n^2 + (x_{n-1} - \mu_n)^2}{\sigma_n}\right]^2 = C_n,$$

whereby an approximation over the infinite range is obtained. The functions in the left hand members of (13) depend only on the variables $x_1$ and $x_{n-1}$ respectively and are often convenient analytical expressions, e.g. for $f(x) = e^{-x}$,

$$\frac{3f(x_{n-1})}{4}\left[\frac{\sigma_n^2 + (x_{n-1} - \mu_n)^2}{\sigma_n}\right]^2 = 3e^{-x_{n-1}}.$$

The result (11) may be compared to some analogous results discussed in [2]. In [5] it is proven that $P_h\sigma_h = C_n$ gives an approximate solution to (1). Now under the assumption of small intervals $(x_h - x_{h-1})$, $f(x)$ may be approximated by a constant $f(\xi)$ in this interval $(x_{h-1} \leqq \xi \leqq x_h)$, in which case

$$\sigma_h = (x_h - x_{h-1})/\sqrt{12}.$$

Therefore this result [5] and (11) are substantially equivalent, each one follows from the other. (11) might seem advantageous from a computational point of view. From a theoretical point of view we have at the same time derived (8) and (9), which permits (12) and (13) to be used in some cases mentioned in the introduction, where both (11) and the result in [5] give comparatively poor results.

In [4] it is shown that $[x_h - x_{h-1}] = C_n$ also gives $\{x_n\}$ approximating (1). This result follows from (7) above, if there in the series expansions we neglect even the second terms, obtaining then

$$\frac{\sigma_h^2 + (x_h - \mu_h)^2}{\sigma_h} \sim \frac{2}{\sqrt{3}} (x_h - x_{h-1}),$$

and the analogous result for the right hand member of (1). It is therefore reasonable to assume that Aoyama's result gives a poorer approximation (at least this is so asymptotically) than (11) or the result in [5], although computationally, of course, $(x_h - x_{h-1}) = C_n$ is better than both of these; it has nevertheless the disadvantage of being restricted to a finite range.

The approximation $P_h\mu_h = C_n$ has been proposed. We see from (11) that asymptotically this would imply $(x_h - x_{h-1})/\mu_h = $ a constant, so that the interval lengths would increase with $h$, whereas the $P_h$ necessarily decrease. This so-called principle of equipartition might therefore be of use when dealing with decreasing $f(x)$, e.g. of exponential type.

**4. First approximation of the constant $C_n$.** With a knowledge of $C_n$ the set $\{x_h\}$ satisfying (11) or (12), (13) may easily be found, which set then approximately satisfies (1). The necessity of at least an approximation to $C_n$ arises.

We shall derive an approximation of $C_n$ under the assumption that $C_{n-r}$ has already been obtained; this will be done by the following heuristic argument. Suppose that for a density with a finite range $(x_n - x_0)$ a set of points $x_1, \cdots, x_{n-1}$ has been found such that the relations (11) are satisfied (there is always a unique such solution, as can be seen immediately). The left side of the identity

$$\sum_{h=1}^{n} \left(\frac{x_h - x_{h-1}}{x_n - x_0}\right) P_h = \frac{nC_n}{(x_n - x_0)} \tag{14}$$

may be considered as a weighted mean either of the $(x_h - x_{h-1})/(x_n - x_0)$ weighted by the $P_h$ or vice versa. We noticed above that Aoyama's approximation $(x_h - x_{h-1}) = C = $ a constant is asymptotically correct when all terms but the first are neglected in (5) and (6), so that for large $n$

$$[(x_h - x_{h-1})/(x_n - x_0)] \sim 1/n,$$

and (14) becomes

$$\frac{1}{n}\sum_{h=1}^{n} P_h = \frac{1}{n} \sim \frac{nC_n}{(x_n - x_0)},$$

that is:

$$C_n \sim \frac{(x_n - x_0)}{n^2}. \tag{15}$$

As a first approximation $C'_n$ to $C_n$ the expression

$$C'_n = \frac{(n-r)^2 C_{n-r}}{n^2} \tag{16}$$

may therefore be used.

Of course other first approximations $C'_n$ might be conceived of. There is for example some logic in adjusting $C'_n$ as given by (16) by the factor $C_{n-r}/C'_{n-r}$, thereby obtaining

$$C'_n = \frac{(n-r)^2 C^2_{n-r}}{n^2 C'_{n-r}}. \tag{17}$$

**5. Adaption to numerical computation.** The above results (11) and (12), (13), together with (16), (17) suggest a reasonably simple method of finding a set of points $\{x_h\}$ which approximatively satisfy the identities (1), that is, which approximately minimize the sum $\sum_1^n P_h\sigma_h$ for a given density. The method to be described is applicable to the case of a finite range; in the general case the identities (13) are used instead of $(x_1 - x_0)P_1 = (x_n - x_{n-1})P_n = C_n$.

Assuming that the set $\{x_h\}$ satisfying (11) has been found for some $n = n - r$, (e.g. for $n = 2$, $x_1$ may be found by trial and error), a first approximation to $C_n$ is obtained by (16) or (17). A set $\{x'_h\}$ may be found such that (in general) all $(x'_h - x'_{h-1})P'_h$ but one are equal to $C'_n$. Let us assume that $(x'_j - x'_{j-1})P'_j \neq C'_n$ (e.g. $j = n$). Then a second approximation $C''_n$ to $C_n$ may be obtained by putting

$$C''_n = \frac{(n-1)C'_n + (x'_j - x'_{j-1})P'_j}{n}. \tag{18}$$

Proceeding in the same manner a set $\{x''_h\}$ such that all $(x''_h - x''_{h-1})P''_h$ but one are equal to $C''_n$ may be found, and a new approximation $C'''_n$ to $C_n$ is obtained by an analogous formula to (18), etc. The $C^{(i)}_n$ thus obtained converge to $C_n$, and the sets $\{x^{(i)}_h\}$ correspondingly to the set $\{x_h\}$ satisfying (11).

**6. Numerical examples.** The method described above has been applied to three densities, whereof two are over a semi-finite range. For $n = 2$ the point $x_1$ has been found by trial and error. As first approximations $C'_n$, (16) with $r = 1$, has been used in all cases but one ($f(x) = xe^{-x}$, $n = 4$), where (17) with $r = 1$ was used. The points $\{x_h\}$ have been found to three decimals; as a rule a comparable degree of accuracy will not be necessary, at least for the first approximation.

The results are summarized in table I. The results under Min. are the points satisfying the exact minimal conditions (1), with reservation for the last decimal, which may differ by a unit from the true value. The exact minimal variances are given in the last column.

We note that both $f(x) = 2(1 - x)$ and $f(x) = xe^{-x}$ are densities of the type mentioned in the introduction as being not quite suitable for application of equations $(1)^{(a)}$; here better approximations can be obtained by using (13) for $h = n$ in the first, $h = 1$ (and of course $h = n$) in the second case.

**7. Note on the method used in obtaining the exact minimal points.** The method described above gives in many cases a fairly good approximate solution

TABLE I

| Number of intervals | Approximation | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $C$ | $(\Sigma P_h \sigma_h)^2$ | |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 1 | 0.382 | 1.000 | | | | | 0.01517 | |
| | min. | 0.354 | 1.000 | | | | $C_2 = 0.2361$ | 0.01505 | |
| 3 | 1 | 0.244 | 0.528 | 1.000 | | | $C_3' = 0.1049$ | 0.00692 | |
| | 2 | 0.242 | 0.531 | 1.000 | | | $C_3'' = 0.1030$ | 0.00693 | |
| | min. | 0.230 | 0.503 | 1.000 | | | $C_3''' = 0.1029$ | 0.00688 | $f(x)$ |
| 4 | 1 | 0.178 | 0.379 | 0.613 | 1.000 | | $C_4' = 0.0579$ | 0.00395 | $= 2(1$ |
| | 2 | 0.177 | 0.376 | 0.615 | 1.000 | | $C_4'' = 0.0573$ | 0.00395 | $- x)$ |
| | min. | 0.171 | 0.361 | 0.588 | 1.000 | | $C_4''' = 0.0573$ | 0.00393 | |
| 5 | 1 | 0.141 | 0.294 | 0.466 | 0.668 | 1.000 | $C_5' = 0.0367$ | 0.00255 | |
| | 2 | 0.140 | 0.292 | 0.463 | 0.669 | 1.000 | $C_5'' = 0.0364$ | 0.00255 | |
| | min. | 0.136 | 0.283 | 0.448 | 0.644 | 1.000 | $C_5''' = 0.0365$ | 0.00254 | |
| 2 | 1 | 1.233 | $\infty$ | | | | | 0.2856 | |
| | min. | 1.262 | $\infty$ | | | | $C_2 = 0.8737$ | 0.2855 | |
| 3 | 1 | 0.742 | 2.045 | $\infty$ | | | $C_3' = 0.3883$ | 0.1334 | |
| | 2 | 0.766 | 1.991 | $\infty$ | | | $C_3'' = 0.4095$ | 0.1333 | |
| | 3 | 0.763 | 1.997 | $\infty$ | | | $C_3''' = 0.4071$ | 0.1333 | |
| | min. | 0.764 | 2.026 | $\infty$ | | | $C_3'''' = 0.4071$ | 0.1332 | |
| 4 | 1 | 0.545 | 1.295 | 2.572 | $\infty$ | | $C_4' = 0.2292$ | 0.0769 | $f(x) =$ |
| | 2 | 0.553 | 1.317 | 2.547 | $\infty$ | | $C_4'' = 0.2349$ | 0.0769 | $e^{-x}$ |
| | min. | 0.551 | 1.315 | 2.577 | $\infty$ | | $C_4''' = 0.2345$ | 0.0768 | |
| 5 | 1 | 0.430 | 0.977 | 1.730 | 2.995 | $\infty$ | $C_5' = 0.1501$ | 0.0500 | |
| | 2 | 0.433 | 0.985 | 1.748 | 2.982 | $\infty$ | $C_5'' = 0.1522$ | 0.0500 | |
| | min. | 0.431 | 0.982 | 1.746 | 3.008 | $\infty$ | $C_5''' = 0.1521$ | 0.0500 | |
| 2 | 1 | 2.125 | $\infty$ | | | | | 0.6420 | |
| | min. | 2.291 | $\infty$ | | | | $C_2 = 1.3319$ | 0.6370 | |
| 3 | 1 | 1.423 | 2.976 | $\infty$ | | | $C_3' = 0.5920$ | 0.3125 | |
| | 2 | 1.467 | 3.109 | $\infty$ | | | $C_3'' = 0.6328$ | 0.3092 | |
| | 3 | 1.472 | 3.123 | $\infty$ | | | $C_3''' = 0.6368$ | 0.3090 | |
| | min. | 1.571 | 3.252 | $\infty$ | | | $C_3'''' = 0.6371$ | 0.3075 | |
| 4 | 1 | 1.175 | 2.304 | 3.949 | $\infty$ | | $C_4' = 0.3859$ | 0.1809 | $f(x) =$ |
| | 2 | 1.157 | 2.259 | 3.835 | $\infty$ | | $C_4'' = 0.3720$ | 0.1811 | $xe^{-x}$ |
| | min. | 1.234 | 2.324 | 3.915 | $\infty$ | | $C_4''' = 0.3696$ | 0.1804 | |
| 5 | 1 | 0.955 | 1.785 | 2.792 | 4.277 | $\infty$ | $C_5' = 0.2365$ | 0.1191 | |
| | 2 | 0.960 | 1.796 | 2.814 | 4.327 | $\infty$ | $C_5'' = 0.2396$ | 0.1189 | |
| | 3 | 0.961 | 1.798 | 2.818 | 4.336 | $\infty$ | $C_5''' = 0.2401$ | 0.1189 | |
| | min. | 1.032 | 1.859 | 2.876 | 4.425 | $\infty$ | $C_5'''' = 0.2401$ | 0.1185 | |

to the problem of minimizing $\sum P_h\sigma_h$, from the point of view of generating points $\{x_h\}$ fairly close to the true minimal values. When this is the case, the "usual" iterative procedure of finding successively better approximations converging to the true values may be employed, which method will be briefly reviewed for completeness.

To this end denote by $A_h$ the left hand member of (1), $B_{h+1}$ the right hand member, that is:

$$A_h = \frac{\sigma_h^2 + (x_h - \mu_h)^2}{\sigma_h},$$

$$B_{h+1} = \frac{\sigma_{h+1}^2 + (x_h - \mu_{h+1})^2}{\sigma_{h+1}}.$$

The minimal conditions (1) then take on the form

$$A_h - B_{h+1} = 0, \qquad h = 1, \cdots, (n-1).$$

The conditions $(1)^{(a)}$ or (12), (13) have resulted in approximative values of these expressions, which may be denoted by $A_h^{(0)}$, $B_{h+1}^{(0)}$.

Taylor expansions of $A_h$ and $B_{h+1}$ about $x_{h-1}^{(0)}$, $x_h^{(0)}$, and $x_h^{(0)}$, $x_{h+1}^{(0)}$, respectively, have then the form

$$A_h\,(x_{h-1}^{(0)} + K_{h-1}, x_h^{(0)} + K_h) = A_h^{(0)} + \left(\frac{\partial A_h}{\partial x_{h-1}}\right)_0 K_{h-1} + \left(\frac{\partial A_h}{\partial x_h}\right)_0 K_h,$$

$$B_{h+1}(x_h^{(0)} + K_h, x_{h+1}^{(0)} + K_{h+1}) = B_{h+1}^{(0)} + \left(\frac{\partial B_{h+1}}{\partial x_h}\right)_0 K_h + \left(\frac{\partial B_{h+1}}{\partial x_{h+1}}\right)_0 K_{h+1},$$

if terms of second order or higher are neglected, and where the subscript 0 denotes the value of the partial derivatives in $x_h^{(0)}$, etc. We should now attempt to find $\{K_h\}$ so that these expressions equalize, that is, we should solve the system of linear equations:

$$\text{(a)} \quad \left(\frac{\partial A_h}{\partial x_{h-1}}\right)_0 K_{h-1} + \left[\left(\frac{\partial A_h}{\partial x_h}\right)_0 - \left(\frac{\partial B_{h+1}}{\partial x_h}\right)_0\right] K_h - \left(\frac{\partial B_{h+1}}{\partial x_{h+1}}\right)_0 K_{h+1}$$
$$= B_{h+1}^{(0)} - A_h^{(0)}, \qquad h = 1, \cdots, (n-1).$$

whereafter $K_h$ is added to $x_h^{(0)}$ to obtain the new approximative $x_h$. The matrix of this equation will be $>0$ if the first approximation is good, since this matrix is definite positive in the exact minimal points. The system may be solved in a relatively simple manner by first finding for example $K_1$ by Cramer's formula, and by then successively determining $K_2$, $K_3$, $\cdots$ from the equations for $h = 1$, $2, \cdots$. It may be noted that the matrix is a so called continuant matrix, and has a simple recursive form discussed in the literature. Equations (a) are of course used iteratively, (if necessary), once the first set of $\{K_h\}$ has been found, etc.

It remains to find the expressions for the partial derivatives appearing in the equations (a). We first multiply both numerator and denominator of $A_h$ by $P_h$. From the expressions for $P_h[\sigma_h^2 + (x_h - \mu_h)^2]$ and $P_h[x_h - \mu_h]$ in the function $H$ and its derivatives, we find immediately

$$\frac{\partial}{\partial x_h} [P_h [\sigma_h^2 + (x_h - \mu_h)^2]] = 2P_h [x_h - \mu_h],$$

and by simple derivation

$$\frac{\partial}{\partial x_{h-1}} [P_h [\sigma_h^2 + (x_h - \mu_h)^2]] = -(x_h - x_{h-1})^2 f(x_{h-1}).$$

Keeping in mind that the original minimal conditions were derived from

$$\frac{\partial [P_h \sigma_h]}{\partial x_{h-1}} = -\frac{f(x_{h-1})}{2} B_h,$$

$$\frac{\partial [P_h \sigma_h]}{\partial x_h} = \frac{f(x_h)}{2} A_h,$$

we then find by simple computation (the procedure is analogous for $B_{h+1}$):

$$\text{(b)} \quad \begin{aligned} \frac{\partial A_h}{\partial x_{h-1}} &= \frac{f(x_{h-1})}{P_h \sigma_h} \left[ \frac{A_h B_h}{2} - (x_h - x_{h-1})^2 \right], \\ \frac{\partial A_h}{\partial x_h} &= \frac{1}{P_h \sigma_h} \left[ 2P_h(x_h - \mu_h) - \frac{f(x_h)}{2} A_h^2 \right], \\ \frac{\partial B_{h+1}}{\partial x_h} &= \frac{1}{P_{h+1} \sigma_{h+1}} \left[ 2 P_{h+1}(x_h - \mu_{h+1}) + \frac{f(x_h)}{2} B_{h+1}^2 \right], \\ \frac{\partial B_{h+1}}{\partial x_{h+1}} &= \frac{f(x_{h+1})}{P_{h+1} \sigma_{h+1}} \left[ (x_{h+1} - x_h)^2 - \frac{A_{h+1} B_{h+1}}{2} \right] = -\frac{f(x_{h+1})}{f(x_h)} \cdot \frac{\partial A_{h+1}}{\partial x_h}. \end{aligned}$$

These expressions involve only the functions $A_h$, etc. themselves, and $\mu_h$, $\sigma_h$, $f(x_h)$, etc., which have already been calculated to obtain $A_h$, etc., and are therefore not difficult to compute.

Using equations (a) on the approximate values found by $(1)^{(a)}$ for the three densities above, the exact values (with reservation for the third decimal in a few cases) were found with only one application of (a). This would seem to imply, that no greater error is introduced by neglecting second order terms and higher in the expansions of $A_h$, $B_{h+1}$, and that application of the above method on the points satisfying conditions $(1)^{(a)}$ very often provides a reasonably simple and systematic method of finding points satisfying the minimal conditions (1).

**Acknowledgements.** I stand in appreciation to Docent Tore Dalenius of the Institute of Statistics, University of Stockholm for stating the problem and allowing me to examine his article [2] before it appeared in print.

I am also indebted to the referee and to Mr. J. M. Hammersley for helpful suggestions in the revision of the original paper.

## REFERENCES

[1] T. Dalenius, "The problem of optimum stratification," *Skand. Aktuarietids.*, (1950), pp. 203–213.

[2] T. Dalenius and J. L. Hodges, Jr. "Minimum variance stratification," *J. Amer. Stat. Assn.*, (1958), (submitted in December 1957).

[3] T. Dalenius and M. Gurney, "The problem of optimum stratification II," *Skand. Aktuarietids.*, (1951), pp. 133–148.

[4] H. Aoyama, "A study of the stratified random sampling," *Ann. Inst. Stat. Math.*, (1954), pp. 1–36.

[5] T. Dalenius and J. L. Hodges, Jr. "The choice of stratification points," (1958), (Soon to appear in *Skand. Aktuarietids.*).