Recursive Integration Methodologies with Statistical Applications

A. J. Hayter
Industrial & Systems Engineering
Georgia Tech, Atlanta, USA
ajh@isye.gatech.edu

Summary

This paper shows how recursive integration methodologies can be used to evaluate high dimensional integral expressions. This has applications to many areas of statistical inference where probability calculations and critical point evaluations often require such high dimensional integral evaluations. Recursive integration can allow an integral expression of a given dimension to be evaluated by a series of calculations of a smaller dimension. This significantly reduces the computation time. The application of the recursive integration methodology is illustrated with several examples.

Key words: Recursive integration; Numerical integration; Computational intensity; Tree structure; Markov property; Critical point calculation; Multiple comparisons; Sequential analysis.

1 Introduction

Let $\boldsymbol{x} = (x_1, \ldots, x_n) \in \Re^n$ and consider the integral

$$A = \int \cdots \int_{\boldsymbol{x} \in S} f(\boldsymbol{x})\, d\boldsymbol{x}$$

where for some d with $1 \leq d \leq n-1$, the function $f(\boldsymbol{x})$ can be expressed as

$$f(\boldsymbol{x}) = \prod_{i=1}^{n-d+1} f_i(x_i, \ldots, x_{d+i-1}) \tag{1}$$

and the set $S \subseteq \Re^n$ can be expressed as

$$S = \bigcap_{i=1}^{n-d} S_i$$

where the set S_i places restrictions only on $(x_i, \ldots, x_{d+i})$. The n-dimensional integral A is thus defined in terms of variables x_i with an inherent order, where the integrand

$f(\boldsymbol{x})$ factors into terms consisting of at most d adjacent x_i, and where the integration is over a set S which consists of restrictions involving at most $d+1$ adjacent x_i.

The purpose of this paper is to demonstrate that while the evaluation of A ostensibly requires an n-dimensional numerical integration, it can in fact be evaluated by $n - d + 1$ calculations which each have at most d-dimensional computational intensity. In other words, using this approach the computational intensity of evaluating A is at most equivalent to evaluating $n - d + 1$ integrals each having dimension d. Consequently, the approximation of the integral expression using a numerical integration method with a grid of N points in each dimension for a bounded set S results in $O(N^d)$ mathematical operations when the recursive integration method is applied versus $O(N^n)$ operations when the numerical integration method is directly applied.

If the random variables $\boldsymbol{X} = (X_1, \ldots, X_n)$ have a probability density function $f(\boldsymbol{x})$ then $A = P(\boldsymbol{X} \in S)$, and so this problem has many important applications to statistical inference procedures. For example, probability calculations and critical point calculations for multiple comparisons and sequential procedures require the evaluation of high dimensional integral expressions of the form A. More generally, many important probabilities such as orthant probabilities for some multivariate normal distributions can be evaluated using the recursive integration methodology.

Section 2 contains some illustrations of the statistical applications of the methodology, and the recursive integration methodology itself is explained in section 3. The recursive integration methodology can also be applied to variables arranged in a tree structure, and this is discussed in section 4.

2 Statistical applications of recursive integration

There are several important probability expressions that can be evaluated using the recursive integration methodology with $d = 1$ and $d = 2$. Notice that if the random variables $\boldsymbol{X} = (X_1, \ldots, X_n)$ are independent then their probability density function has the form (1) with $d = 1$. In addition, if the X_i have a first order Markov property then their probability density function has the form (1) with $d = 2$ since

$$f(\boldsymbol{x}) = f_1(x_1, x_2) \prod_{i=2}^{n-1} f(x_{i+1}|x_1, \ldots, x_i) = f_1(x_1, x_2) \prod_{i=2}^{n-1} f(x_{i+1}|x_i) \tag{2}$$

where $f_1(x_1, x_2)$ is the marginal probability density function of (X_1, X_2). This first order Markov property arises in sequential analysis where $X_i = Y_1 + \ldots + Y_i$ for some independent random variables Y_i.

The recursive integration methodology can be applied, for example, to evaluate the probability

$$P(X_1 < X_2 < \ldots < X_n). \tag{3}$$

In isotonic regression this probability is needed to find level probabilities (see Robertson *et al.* (1988) pages 74-77) and in ranking and selection procedures this probability represents the probability of a correct ranking (see Bechhofer (1954)). This probability is a special case of the more general expression

$$P(X_1 \in S_0, X_2 \in S_1(X_1), \ldots, X_n \in S_{n-1}(X_{n-1})) \tag{4}$$

where the set S_i places restrictions on the random variables X_i and X_{i+1}. If the X_i are independent then the probability (4) can be evaluated by the recursive integration methodology with $d = 1$ (this was first shown for (3) in Hayter & Liu (1996)).

Another application concerns the evaluation of the cumulative distribution function of the sum of independent random variables

$$P(Y_1 + \ldots + Y_n \leq c). \tag{5}$$

This is useful when the distributions of the Y_i are known but the distribution of their sum is not easily obtained. If $X_i = Y_1 + \ldots + Y_i$, $1 \leq i \leq n$, then the X_i have a first order Markov property. Moreover, the terms $f_i(x_i, x_{i+1}) = f(x_{i+1}|x_i)$ in the factorization of $f(\boldsymbol{x})$ given in (2) can be obtained from the distributions of the Y_i. The probability (5) is $P(X_n \leq c)$ which is a special case of the more general probability

$$P((X_1, X_2, X_3) \in S_1, (X_2, X_3, X_4) \in S_2, \ldots, (X_{n-2}, X_{n-1}, X_n) \in S_{n-2}) \tag{6}$$

which can be evaluated by the recursive integration methodology with $d = 2$. In sequential analysis X_i may be the test statistic after i stages, and the properties of the sequential procedure may be obtained from expression (6) with the set S_i only placing restrictions on X_i. The implementation of recursive methods to sequential analysis has been shown in Armitage *et al.* (1969) and Jennison & Turnbull (1991).

An additional application is given in Miwa *et al.* (2003) where the problem of evaluating general non-centered orthant probabilites for a multivariate normal distribution is discussed. This fundamental problem has many statistical applications, and it is shown how any orthant probability can be expressed in terms of orthant probabilities for a multivariate normal distribution with a tri-diagonal covariance matrix, which can each be evaluated by the recursive integration methodology with $d = 1$.

Liu *et al.* (2000) provide an application to the multiple comparison procedure concerning successive comparisons of treatment effects in a one way layout. One-sided and two-sided sets of simultaneous confidence intervals are constructed for the differences $\mu_{i+1} - \mu_i$, $2 \leq i \leq k$, which can be useful when the treatment effects μ_i have a known intrinsic ordering such as for a sequence of increasing dose levels of a drug. In this case the recursive integration methodology is employed with $d = 1$ together with an additional integration over the conditioned value of the variance estimate s^2 in order to find the necessary critical points.

Finally, Kuriki *et al.* (2002) contains an implementation of the recursive integration methodology with $d = 2$ applied to critical point calculations for the isotonic range statistic in order restricted inference. For testing the null hypothesis $H_0 : \mu_1 = \ldots = \mu_k$ against $H_A : \mu_1 \leq \ldots \leq \mu_k$ with at least one strict inequality in a one way layout, the isotonic range statistic is the range $\hat{\mu}_k - \hat{\mu}_1$ of the isotonic estimates of the treatment effects. The critical points of this test procedure also allow the construction of simultaneous confidence intervals for all monotone contrasts of the treatment effects.

3 The general recursive integration methodology

3.1 General methodology

For $1 \leq i \leq n-d$ define

$$h_i(x_{i+1}, \ldots, x_{d+i}) = \int_{B_i(x_{i+1}, \ldots, x_{d+i})} \cdots \int \prod_{j=1}^{i} f_j(x_j, \ldots, x_{d+j-1}) \, dx_1 \ldots dx_i \qquad (7)$$

where

$$B_i(x_{i+1}, \ldots, x_{d+i}) = \{(x_1, \ldots, x_i) : (x_1, \ldots, x_{d+1}) \in S_1, \ldots, (x_i, \ldots, x_{d+i}) \in S_i\} \subseteq \Re^i.$$

Also, define

$$g_1(y_1, x_2, \ldots, x_d) = \int_{x_1=-\infty}^{y_1} f_1(x_1, \ldots, x_d) \, dx_1 \qquad (8)$$

and

$$g_i(y_i, x_{i+1}, \ldots, x_{d+i-1}) = \int_{x_i=-\infty}^{y_i} f_i(x_i, \ldots, x_{d+i-1}) \; h_{i-1}(x_i, \ldots, x_{d+i-1}) \, dx_i \qquad (9)$$

for $2 \leq i \leq n-d$.

The set S_i places restrictions on the values of $(x_i, \ldots, x_{d+i})$, and so for given values of $(x_{i+1}, \ldots, x_{d+i})$ the set S_i restricts x_i to lie in a series of disjoint intervals (l_{ij}, u_{ij}), $1 \leq j \leq m_i$, say. The lower endpoints l_{ij} and the upper endpoints u_{ij} of these intervals may depend on $(x_{i+1}, \ldots, x_{d+i})$. Therefore, if

$$\begin{aligned} S_i^*(x_{i+1}, \ldots, x_{d+i}) &= \{x_i : (x_i, \ldots, x_{d+i}) \in S_i\} \\ &= \bigcup_{j=1}^{m_i} (l_{ij}(x_{i+1}, \ldots, x_{d+i}), u_{ij}(x_{i+1}, \ldots, x_{d+i})) \end{aligned} \qquad (10)$$

for $1 \leq i \leq n-d$ where the union is of disjoint intervals, it follows from the definition of h_i in (7) that (taking $h_0 = 1$)

$$h_i(x_{i+1}, \ldots, x_{d+i}) = \int_{x_i \in S_i^*} f_i(x_i, \ldots, x_{d+i-1}) \; h_{i-1}(x_i, \ldots, x_{d+i-1}) \, dx_i$$

$$= \sum_{j=1}^{m_i} [g_i(u_{ij}(x_{i+1}, \ldots, x_{d+i}), x_{i+1}, \ldots, x_{d+i-1}) - g_i(l_{ij}(x_{i+1}, \ldots, x_{d+i}), x_{i+1}, \ldots, x_{d+i-1})] \qquad (11)$$

for $1 \leq i \leq n-d$.

The functions g_i, $1 \leq i \leq n-d$, can be calculated recursively. First, g_1 is calculated from (8). Next g_2 can be calculated from (9) with $i = 2$ where the values of h_1 can be obtained from g_1 using expression (11). The subsequent g_i are similarly calculated from (9) with the values of h_{i-1} being obtained from g_{i-1} using expression (11). Once the function g_{n-d} has been calculated, the required probability can be calculated as

$$A = \int_{x_{n-d+1}=-\infty}^{\infty} \ldots \int_{x_n=-\infty}^{\infty} f_{n-d+1}(x_{n-d+1}, \ldots, x_n) \; h_{n-d}(x_{n-d+1}, \ldots, x_n) \, dx_{n-d+1} \ldots dx_n \qquad (12)$$

where the values of h_{n-d} are obtained from g_{n-d} using expression (11).

3.2 Computational intensity

The functions g_i, $1 \leq i \leq n-d$, are each functions of d variables, and they can be evaluated and stored over a grid of points in d dimensions. If the grid values of y_i are y_{ij}, $1 \leq j \leq N$, then

$$\begin{aligned} g_i(y_{i,j+1}, x_{i+1}, \ldots, x_{d+i-1}) &= g_i(y_{ij}, x_{i+1}, \ldots, x_{d+i-1}) \\ &\quad + \int_{x_i=y_{ij}}^{y_{i,j+1}} f_i(x_i, \ldots, x_{d+i-1})\ h_{i-1}(x_i, \ldots, x_{d+i-1})\ dx_i \end{aligned}$$

so that the values of g_i can be obtained for the d-dimensional grid points once the quantities $g_i(y_{11}, x_{i+1}, \ldots, x_{d+i-1})$ and

$$\int_{x_i=y_{ij}}^{y_{i,j+1}} f_i(x_i, \ldots, x_{d+i-1})\ h_{i-1}(x_i, \ldots, x_{d+i-1})\ dx_i, \quad 1 \leq j \leq N-1, \tag{13}$$

have been calculated for each of the $(d-1)$-dimensional grid points $(x_{i+1}, \ldots, x_{d+i-1})$. The adjacent grid points y_{ij} and $y_{i,j+1}$ can be chosen to be close together, and so the integrals in (13) can be approximated by standard Newton-Cotes methods. The values of h_{i-1} required in the evaluation of the integrals in (13) can be obtained from (11) using interpolation between the stored values of g_{i-1} if necessary.

The evaluation of all of the integrals in (13) has a computational intensity proportional to N^d if each of the variables of g_i is calculated over a grid with N points. This process must be repeated for each g_i, $1 \leq i \leq n-d$. In addition, the final calculation of A from g_{n-d} using (12) requires an additional d-dimensional numerical integration. Therefore, the computational intensity of evaluating A using this method is equivalent to $n-d+1$ calculations which each have at most computational intensity proportional to N^d. This is preferable to evaluating A directly as an n-dimensional numerical integration because the direct computation has computational intensity proportional to N^n.

3.3 Illustrations of the methodology

The recursive integration methodology is illustrated for some simple cases.

An integral with $n = 2$ and $d = 1$

If $n = 2$ and $d = 1$ then

$$A = \int_{S_1} \int f_1(x_1) f_2(x_2)\ dx_1 dx_2$$

where $S_1 \subseteq \Re^2$ is any set. The first step in the evaluation of A is to calculate and store the function

$$g_1(y_1) = \int_{x_1=-\infty}^{y_1} f_1(x_1)\ dx_1$$

over a grid of y_1 values. Then A can be evaluated as

$$A = \int_{x_2=-\infty}^{\infty} f_2(x_2) h_1(x_2)\ dx_2$$

where

$$h_1(x_2) = \sum_{j=1}^{m_1} [g_1(u_{1j}(x_2)) - g_1(l_{1j}(x_2))]$$

for

$$S_1^*(x_2) = \{x_1 : (x_1, x_2) \in S_1\} = \bigcup_{j=1}^{m_1} (l_{1j}(x_2), u_{1j}(x_2))$$

where the union is of disjoint intervals. The evaluation requires two calculations which each have one-dimensional computational intensity.

An integral with $n = 3$ and $d = 1$

If $n = 3$ and $d = 1$ then

$$A = \int_S \dots \int f_1(x_1) f_2(x_2) f_3(x_3)\ dx_1 dx_2 dx_3$$

where

$$S = \{(x_1, x_2, x_3) : (x_1, x_2) \in S_1, (x_2, x_3) \in S_2\}.$$

Again, the first step in the evaluation of A is to calculate and store the function

$$g_1(y_1) = \int_{x_1=-\infty}^{y_1} f_1(x_1)\ dx_1$$

over a grid of y_1 values. Next, the function

$$g_2(y_2) = \int_{x_2=-\infty}^{y_2} f_2(x_2) h_1(x_2)\ dx_2$$

is calculated and stored over a grid of y_2 values where

$$h_1(x_2) = \sum_{j=1}^{m_1} [g_1(u_{1j}(x_2)) - g_1(l_{1j}(x_2))]$$

for

$$S_1^*(x_2) = \{x_1 : (x_1, x_2) \in S_1\} = \bigcup_{j=1}^{m_1} (l_{1j}(x_2), u_{1j}(x_2))$$

where the union is of disjoint intervals. Finally, A can be evaluated as

$$A = \int_{x_3=-\infty}^{\infty} f_3(x_3) h_2(x_3)\ dx_3$$

where

$$h_2(x_3) = \sum_{j=1}^{m_2} [g_2(u_{2j}(x_3)) - g_2(l_{2j}(x_3))]$$

for

$$S_2^*(x_3) = \{x_2 : (x_2, x_3) \in S_2\} = \bigcup_{j=1}^{m_2} (l_{2j}(x_3), u_{2j}(x_3))$$

where the union is of disjoint intervals. The evaluation requires three calculations which each have one-dimensional computational intensity.

An integral with $n = 3$ and $d = 2$

If $n = 3$ and $d = 2$ then

$$A = \int_{S_1} \dots \int f_1(x_1, x_2) f_2(x_2, x_3)\ dx_1 dx_2 dx_3$$

where $S_1 \subseteq \Re^3$ is any set. The first step in the evaluation of A is to calculate and store the function

$$g_1(y_1, x_2) = \int_{x_1=-\infty}^{y_1} f_1(x_1, x_2)\ dx_1$$

over a two-dimensional grid of y_1 and x_2 values. Then A can be evaluated as

$$A = \int_{x_2=-\infty}^{\infty} \int_{x_3=-\infty}^{\infty} f_2(x_2, x_3) h_1(x_2, x_3)\ dx_2 dx_3$$

where

$$h_1(x_2, x_3) = \sum_{j=1}^{m_1} [g_1(u_{1j}(x_2, x_3), x_2) - g_1(l_{1j}(x_2, x_3), x_2)]$$

for

$$S_1^*(x_2, x_3) = \{x_1 : (x_1, x_2, x_3) \in S_1\} = \bigcup_{j=1}^{m_1} (l_{1j}(x_2, x_3), u_{1j}(x_2, x_3))$$

where the union is of disjoint intervals. The evaluation requires two calculations which each have two-dimensional computational intensity.

3.4 Practical aspects

The shape of the set S has implications to the practical implementation of the recursive integration methodology. First of all, if S is infinite it should be replaced by a bounded set so that the grids over which the functions are stored are finite. It can be assumed that the required answer A is finite so that it can be closely approximated by replacing S by a suitable bounded set. This is equivalent to defining the functions f_i to be zero outside of bounded sets.

In addition, as usual for numerical integration methods it is necessary to have the set S piecewise smooth so that the required functions can be well represented by storing them at close grid points. Furthermore, if the calculation of l_{ij} and u_{ij} in (10) is not trivial, such as in situations where the solution of non-linear equations is required, then this will slow down the implementation of the procedure. However, the statistical examples discussed in section 2 all have closed form expressions for l_{ij} and u_{ij} (with the m_i generally being equal to one).

Another important aspect of the implementation of the methodology is the choice of the grid size for storing the functions. Three sources of error can be identified in the evaluation of the terms in equation (13): (I) the error in the function h_{i-1} at the grid points due to errors in the function g_{i-1}, (II) the errors in interpolating the x_i

variable in the function g_{i-1} in order to obtain the function h_{i-1}, and (III) the error in the approximation of the integral in equation (13).

If the distance between grid points is h, then the implementation of the two-point Newton-Cotes method for the evaluation of the integrals results in errors of magnitude $O(h^2)$ for source (III), while these errors are $O(h^4)$ if the three-point Newton-Cotes method is used (see Thisted (1988)). Also, if an r^{th} degree polynomial is used for the interpolation then the error for source (II) is $O(h^{r+1})$ (see Shampine *et al.* (1997)). It is sensible to have the three errors of similar magnitudes. For example, if the trapezoidal rule is used for the integrations and linear interpolation is used, then the overall error will be $O(h^2)$. If Simpson's rule is used for the integrations then cubic polynomial interpolation affords an overall error of $O(h^4)$.

In principal, any method can be used to evaluate the final answer from the d dimensional integral in equation (12). For example, a Newton-Cotes method may be used, or simulation and Monte-Carlo techniques may be employed.

While the advantage of the dimensional reduction from n to d is generally obvious, the actual advantage of the recursive integration methodology relative to other methodologies will depend upon the actual problem under consideration. Similarly, the best choice of grid points will depend upon the forms of the functions f_i and the sets S_i. For example, in Miwa *et al.* (2000) the functions f_i are normal probability density functions and the use of grid intervals with lengths inversely proportional to the 1/4 power of the density function are considered, and some indications of the accuracy of the procedure are provided.

4 Recursive integration methodologies for tree structures

In section 3 the recursive integration methodology was described for a set of ordered random variables. In this section extensions of the methodology to random variables arranged in a tree structure are discussed.

Consider a set of random variables arranged in a tree structure as follows. Let X_1 be the root, with branches $X_{11}, \ldots, X_{1b_1}$, and in general let the random variable $X_{i_1 \ldots i_r}$ have branches $X_{i_1 \ldots i_r j}$, $1 \leq j \leq b_{i_1 \ldots i_r}$. Suppose that the probability density function $f(\boldsymbol{x})$ of all the random variables can be factored as

$$f(\boldsymbol{x}) = \prod f_{i_1 \ldots i_r}(x_{i_1 \ldots i_{r-d+1}}, x_{i_1 \ldots i_{r-d+2}}, \ldots, x_{i_1 \ldots i_r})$$

so that each term in the factorization depends upon at most d random variables which are connected in a line from a point in the tree heading back towards the root. Also, suppose that the set S can be expressed as

$$S = \bigcap S_{i_1 \ldots i_r}$$

where the set $S_{i_1 \ldots i_r}$ places restrictions only on $(x_{i_1 \ldots i_{r-d}}, x_{i_1 \ldots i_{r-d+1}}, \ldots, x_{i_1 \ldots i_r})$, which are the values of at most $d+1$ random variables connected in a line from a point in the tree heading back towards the root.

A recursive integration approach can be employed to evaluate the probability

$$A = \int_{\boldsymbol{x} \in S} \ldots \int f(\boldsymbol{x}) \, d\boldsymbol{x}$$

as a series of d-dimensional calculations. The recursion begins at the periphery of the tree and works back towards the root random variable. The special case discussed in section 3 corresponds to a tree with a root X_k connected to one branch X_{k-1}, which itself has one branch X_{k-2}, and so one towards the single periphery random variable X_1. The general methodology for a tree structure uses equations similar to those given in section 3, but with each of the branches at each stage being taken into consideration. The methodology is demonstrated in the following examples for general trees with $d = 1$ and for a tree with three levels with $d = 2$.

A tree with $d = 1$

When $d = 1$ the random variables are all independent and S is defined by

$$X_1 \in I_1 \subseteq \Re$$

and

$$X_{i_1 \ldots i_r} \in I_{i_1 \ldots i_r}(X_{i_1 \ldots i_{r-1}}) \tag{14}$$

for all of the other random variables. In other words, the set S can be specified by a region for the root random variable X_1, together with a region for every other random variable which respectively depends only on the value of the random variable from which it is branched.

The evaluation of A can be performed as a series of one-dimensional integral calculations. In order to show how this can be accomplished, it is helpful to define the function $G_{i_1 \ldots i_r}(x)$ to be the probability that the random variable $X_{i_1 \ldots i_r}$ lies in the interval $(-\infty, x)$, and that all of the random variables $X_{i_1 \ldots i_r \ldots}$ which are derived from $X_{i_1 \ldots i_r}$ through one or more branches (these are all of the random variables in the subtree which has $X_{i_1 \ldots i_r}$ as its root) satisfy the restrictions placed upon them by the set S.

Suppose that a random variable which has no subsequent branches is described as a terminating random variable. Consider then a random variable $X_{i_1 \ldots i_r}$ which is not a terminating random variable, and which has branches $X_{i_1 \ldots i_r j}$, $j \in T_{i_1 \ldots i_r}$, which are terminating branches, and branches $X_{i_1 \ldots i_r j}$, $j \in T^*_{i_1 \ldots i_r}$, which are not terminating branches. Here $T_{i_1 \ldots i_r} \cup T^*_{i_1 \ldots i_r} = \{1, \ldots, b_{i_1 \ldots i_r}\}$, with $b_{i_1 \ldots i_r} \geq 1$, and either the set $T_{i_1 \ldots i_r}$ or the set $T^*_{i_1 \ldots i_r}$ may be empty.

If the set $T^*_{i_1 \ldots i_r}$ is empty, then the random variable $X_{i_1 \ldots i_r}$ has branches $X_{i_1 \ldots i_r j}$, $1 \leq j \leq b_{i_1 \ldots i_r}$, none of which have any subsequent branches. In this case the first step is to calculate and store the cumulative distribution functions $F_{i_1 \ldots i_r j}$, $1 \leq j \leq b_{i_1 \ldots i_r}$, of each of these terminating random variables. Next, the function

$$G_{i_1 \ldots i_r}(x) = \int_{y=-\infty}^{x} f_{i_1 \ldots i_r}(y) \left(\prod_{j=1}^{b_{i_1 \ldots i_r}} P(X_{i_1 \ldots i_r j} \in I_{i_1 \ldots i_r j}(y)) \right) dy \tag{15}$$

can be calculated and stored, where $f_{i_1 \ldots i_r}(x)$ is the probability density function of the random variable $X_{i_1 \ldots i_r}$. In the evaluation of this function $G_{i_1 \ldots i_r}(x)$, the cumulative distribution functions $F_{i_1 \ldots i_r j}$, $1 \leq j \leq b_{i_1 \ldots i_r}$, can be referred to for the calculations of $P(X_{i_1 \ldots i_r j} \in I_{i_1 \ldots i_r j}(y))$, $1 \leq j \leq b_{i_1 \ldots i_r}$. Notice that, as required, the function $G_{i_1 \ldots i_r}(x)$ represents the probability that the random variable $X_{i_1 \ldots i_r}$ lies in the interval $(-\infty, x)$,

and that all of the random variables $X_{i_1\ldots i_r j}$, $1 \leq j \leq b_{i_1\ldots i_r}$, satisfy the restrictions placed upon them by the set S.

Now more generally, consider a random variable $X_{i_1\ldots i_r}$ for which the set $T^*_{i_1\ldots i_r}$ is not empty, and for which the functions $G_{i_1\ldots i_r j}(x)$, $j \in T^*_{i_1\ldots i_r}$, have been calculated and stored. In this case the first step is again to calculate and store the cumulative distribution functions $F_{i_1\ldots i_r j}$, $j \in T_{i_1\ldots i_r}$, of the terminating random variables. Next, the function

$$G_{i_1\ldots i_r}(x) = \int_{y=-\infty}^{x} f_{i_1\ldots i_r}(y) \left(\prod_{j \in T_{i_1\ldots i_r}} P(X_{i_1\ldots i_r j} \in I_{i_1\ldots i_r j}(y)) \right) \left(\prod_{j \in T^*_{i_1\ldots i_r}} H_{i_1\ldots i_r j}(y) \right) dy \qquad (16)$$

can be calculated and stored.

In the calculation of expression (16) the cumulative distribution functions $F_{i_1\ldots i_r j}$, $j \in T_{i_1\ldots i_r}$, can be referred to in order to calculate the probabilities $P(X_{i_1\ldots i_r j} \in I_{i_1\ldots i_r j}(y))$, $j \in T_{i_1\ldots i_r}$. In addition, the function $H_{i_1\ldots i_r j}(y)$ is defined to be the probability that the random variable $X_{i_1\ldots i_r j}$ lies in the interval $I_{i_1\ldots i_r j}(y)$ and that all of the random variables $X_{i_1\ldots i_r j\ldots}$ which are derived from $X_{i_1\ldots i_r j}$ through one or more branches satisfy the restrictions placed on them by the set S. However, $H_{i_1\ldots i_r j}(y)$ can be found from the interval $I_{i_1\ldots i_r j}(y)$ in conjunction with reference to the function $G_{i_1\ldots i_r j}(x)$, since if

$$I_{i_1\ldots i_r j}(y) = \bigcup_{i=1}^{m(y)} (l_i(y), u_i(y))$$

say, where the intervals $(l_i(y), u_i(y))$ are disjoint, then

$$H_{i_1\ldots i_r j}(y) = \sum_{i=1}^{m(y)} \left[G_{i_1\ldots i_r j}(u_i(y)) - G_{i_1\ldots i_r j}(l_i(y)) \right].$$

Again it can be seen that the function $G_{i_1\ldots i_r}(x)$ calculated from expression (16) meets its required definition, which is that it represents the probability that the random variable $X_{i_1\ldots i_r}$ lies in the interval $(-\infty, x)$, and that all of the random variables $X_{i_1\ldots i_r\ldots}$ which are derived from $X_{i_1\ldots i_r}$ through one or more branches satisfy the restrictions placed upon them by the set S.

In this manner the functions $G_{i_1\ldots i_r}(x)$ can be calculated recursively for each of the non-terminating random variables, moving backwards from the periphery of the tree towards the root. Thus, if the tree is considered to grow away from the root, the recursion is applied in an opposite direction to the growth. Eventually, it will be possible to calculate and store all of the functions $G_{1j}(x)$, $j \in T^*_1$. If the cumulative distribution functions $F_{1j}(x)$ of each of the terminating random variables X_{1j}, $j \in T_1$, are then calculated and stored, the required probability can be calculated as

$$A = \int_{y \in I_1} f_1(y) \left(\prod_{j \in T_1} P(X_{1j} \in I_{1j}(y)) \right) \left(\prod_{j \in T^*_1} H_{1j}(y) \right) dy \qquad (17)$$

where the cumulative distribution functions $F_{1j}(x)$ are referred to for the evaluation of $P(X_{1j} \in I_{1j}(y))$, $j \in T_1$, and the functions $G_{1j}(x)$ and regions $I_{1j}(y)$ are referred to for the evaluation of $H_{1j}(y)$, $j \in T_1^*$.

Each step in this calculation has the computational intensity of a one-dimensional integral evaluation. The last step using expression (17) is clearly a one-dimensional integral computation, and every other step can be considered to be the calculation and storage of a function $F(x)$ from a known function $f(x)$ with

$$F(x) = \int_{y=-\infty}^{x} f(y)dy.$$

For terminating random variables this is the calculation of the random variable's cumulative distribution function from its probability density function, which may be unnecessary if the cumulative distribution function is already stored or is otherwise available. For non-terminating random variables $F(x)$ is the function $G(x)$ where the function $f(y)$ is the integrand of expression (15) or (16) which can be obtained from the probability density function of the random variable and the previously stored functions.

A well known example (see Tong (1990), page 192, for example) of a tree structure with two levels and with $d = 1$ concerns the evaluation of

$$P(l_i \leq Z_i \leq u_i; 1 \leq i \leq k) \tag{18}$$

for some (possibly infinite) lower and upper bounds l_i and u_i, where the random variables Z_i have a k-dimensional multivariate normal distribution with means μ_i, variances σ_i^2 and covariances $\sigma_{ij} = \lambda_i\lambda_j$, for some $\lambda_i \in \Re$. Let X_1 be a standard normal random variable, and let $X_{1i} \sim N(\mu_i, \tau_i^2)$, $1 \leq i \leq k$, be independent normal random variables which are also independent of X_1, with $\tau_i^2 = \sigma_i^2 - \lambda_i^2$. This forms a tree of independent random variables consisting of the root X_1 with branches X_{1i}, $1 \leq i \leq k$. Moreover, the random variables Z_i can be expressed as $Z_i = \lambda_i X_1 + X_{1i}$, $1 \leq i \leq k$, so that

$$S_{1i} = \{(x_1, x_{1i}) : l_i \leq \lambda_i x_1 + x_{1i} \leq u_i\}, \quad 1 \leq i \leq k.$$

The probability (18) can be evaluated as

$$\int_{x_1=-\infty}^{\infty} \phi(x_1) \left(\prod_{i=1}^{k} \left[\Phi\left(\frac{u_i - \mu_i - \lambda_i x_1}{\tau_i} \right) - \Phi\left(\frac{l_i - \mu_i - \lambda_i x_1}{\tau_i} \right) \right] \right) dx_1.$$

The cumulative distribution functions of the branch variables x_{1i} can each be expressed in terms of the standard normal cumulative distribution function $\Phi(x)$, and these are incorporated into the single one-dimensional integration which is performed for the root variable x_1 with the standard normal probability density function $\phi(x_1)$.

A tree with $d = 2$ and three levels

Suppose that a root random variable X_1 has branches $X_{11}, \ldots, X_{1b_1}$, and that the random variables X_{1i} have subsequent branches X_{1ij}, $1 \leq j \leq b_{1i}$, although some of the b_{1i} may be zero. Also, suppose that the probability density function of these random variables can be expressed as

$$f(\boldsymbol{x}) = \prod_{i=1}^{b_1} \left[f_{1i}(x_1, x_{1i}) \prod_{j=1}^{b_{1i}} f_{1ij}(x_{1i}, x_{1ij}) \right]$$

and that S is the intersection of sets S_{1ij}, $1 \leq i \leq b_1$, $1 \leq j \leq b_{1i}$, where the set S_{1ij} places restrictions only on (X_1, X_{1i}, X_{1ij}), so that $d = 2$.

For $1 \leq i \leq b_1$, $1 \leq j \leq b_{1i}$, define

$$g_{1ij}(x_{1i}, y_{1ij}) = \int_{x_{1ij}=-\infty}^{y_{1ij}} f_{1ij}(x_{1i}, x_{1ij})\ dx_{1ij}$$

and then for $1 \leq i \leq m_1$ define

$$g_{1i}(x_1) = \int_{x_{1i}=-\infty}^{\infty} f_{1i}(x_1, x_{1i}) \prod_{j=1}^{b_{1i}} \left[\sum_{k=1}^{m_{1ij}} (g_{1ij}(x_{1i}, u_{1ijk}(x_1, x_{1i})) - g_{1ij}(x_{1i}, l_{1ijk}(x_1, x_{1i}))) \right] dx_{1i}$$

where for given values of x_1 and x_{1i}

$$\{x_{1ij} : (x_1, x_{1i}, x_{1ij}) \in S_{1ij}\} = \bigcup_{k=1}^{m_{1ij}} (l_{1ijk}(x_1, x_{1i}), u_{1ijk}(x_1, x_{1i}))$$

where the union is of disjoint intervals. The required probability can be calculated as

$$A = \int_{x_1=-\infty}^{\infty} \prod_{i=1}^{b_1} g_{1i}(x_1)\ dx_1.$$

The recursive integration methodology is implemented by first storing each function g_{1ij} for a two-dimensional grid of values. Each of these functions requires a two-dimensional calculation. Using these stored values, the functions g_{1i} can then be calculated and stored for a one-dimensional grid of values, and each of these functions requires an additional two-dimensional calculation. The final calculation of A from the stored values of the g_{1i} only requires a one-dimensional calculation. The evaluation therefore requires $b_1 + \sum_{i=1}^{b_1} b_{1i}$ calculations which each have two-dimensional computational intensity.

Acknowledgement

The author would like to gratefully acknowledge the assistance of the special issue editor and the helpful comments of three referees which have led to improvements in this paper.

References

Armitage, P., McPherson, C. K. and Rowe, B. C. (1969), "Repeated significance tests on accumulating data," *Journal of the Royal Statistical Society, Series A*, 132, 235-244.

Bechhofer, R.E. (1954), "A single-sample multiple decision procedure for ranking means of normal populations with known variances," *The Annals of Mathematical Statistics*, 25, 16-39.

Hayter, A. J. and Liu, W. (1996), "A note on the calculation of $P(X_1 < X_2 < \ldots < X_k)$," *The American Statistician*, 85, 778-785.

Jennison, C. and Turnbull, B. W. (1991), "Exact calculations for sequential t, χ^2 and F tests," *Biometrika*, 78, 133-141.

Kuriki, S., Shimodaira, H. and Hayter, A. J. (2002), "On the isotonic range statistic for testing against an ordered alternative," *Journal of Statistical Planning and Inference*, 105, 347-362.

Liu, W., Miwa, T. and Hayter, A. J. (2000), "Simultaneous confidence interval estimation for successive comparisons of ordered treatment effects," *Journal of Statistical Planning and Inference*, 88, 75-86.

Miwa, T., Hayter, A. J., and Liu, W. (2000), "Calculations of level probabilities for normal random variables with unequal variances with applications to Bartholomew's test in unbalanced one-way models," *Computational Statistics and Data Analysis*, 34, 17-32.

Miwa, T., Hayter, A. J., and Kuriki, S. (2003), "The evaluation of general non-centred orthant probabilities," *Journal of the Royal Statistical Society, Series B*, 65, 223-234.

Robertson, T., Wright, F. T. and Dykstra, R.L. (1988), "Order Restricted Statistical Inference," Wiley.

Shampine, L. F., Allen, R. C., and Pruess, S. (1997), "Fundamentals of numerical computing," Wiley.

Thisted, R. A. (1988), "Elements of statistical computing," Chapman and Hall.

Tong, Y. L. (1990), "The multivariate normal distribution," Springer-Verlag.