CONDITIONAL PROBABILITY AND CONDITIONAL EXPECTATION OF A MULTIVARIATE NORMAL RANDOM VARIABLE OVER A RECTANGLE

Horand I. Gassmann

December 1984
WP-84-100

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
2361 Laxenburg, Austria

PREFACE

In this paper, the author surveys techniques that could be used to find the conditional probability and conditional mean of a normally distributed random vector ξ over a rectangle of the form $A = \prod_{i=1}^{n}[a_i, b_i]$. This information is sometimes useful in stochastic optimization, for instance, when establishing upper and lower bounds on the expected value of the random vector ξ. Numerical results illustrate the performance of the various methods.

This paper was written while the author was visiting the Adaptation and Optimization Project in the System and Decision Sciences Program.

A.B. Kurzhanski
Chairman
System and Decision Sciences
Program

ABSTRACT

The idea of orthonormal variates, designed by Deák to find conditional probabilities of a multivariate normal random variable, is extended to compute the conditional expectation given that the observation falls inside an n-dimensional rectangle. Another possible technique is presented in the form of a transformation to independence. Numerical results are provided to illustrate the performance of the methods.

CONDITIONAL PROBABILITY AND CONDITIONAL EXPECTATION OF A MULTIVARIATE NORMAL RANDOM VARIABLE OVER A RECTANGLE

Horand I. Gassmann

Faculty of Commerce and Business Administration,
The University of British Columbia, Vancouver, Canada V6T 1Y8

1. INTRODUCTION

This paper surveys techniques for finding the conditional probability and conditional mean of a normally distributed random vector ξ over a rectangle of the form

$$A = \prod_{i=1}^{n} [a_i, b_i] \quad . \tag{1}$$

This information is sometimes of interest in stochastic optimization, for instance when one wants to establish lower and upper bounds on the expected value of a convex function of the random vector ξ [1,10]. The bounds are known in the literature as the inequalities of Jensen and Edmundson–Madansky, respectively (see [8] for an extension of the classical result which appeared in [11]). They are used in the following way.

Given a convex function $\varphi : \mathbb{R}^n \to \mathbb{R}$ and a partition $\Pi = \{A_1, \ldots, A_m\}$ of the sample space $\mathbb{R}^n$, we have

$$\sum_{j=1}^{m} p_j \, \varphi(\bar{\xi}_j) \leq \mathrm{E}\varphi \leq \sum_{j=1}^{m} p_j \, u^j \quad , \tag{2}$$

where p_j is the mass of the subset A_j, $\bar{\xi}_j$ its conditional mean and u^j is the Edmundson–Madansky upper bound on A_j.

By partitioning the sample space $\mathbb{R}^n$ in a suitable manner, it is possible to achieve arbitrary accuracy with these bounds, i.e., for every $\varepsilon > 0$ there is a partition Π_ε of $\mathbb{R}^n$ such that the upper and lower bounds in (2) differ by at most ε.

This report was prepared during a visit to the International Institute for Applied Systems Analysis in Laxenburg, Austria. The author gratefully acknowledges the hospitality extended to him as well as the free use of the computing facilities.

It seems natural to construct such a partition by considering only cuts generated by hyperplanes perpendicular to the coordinate axes. Every subset of Π then takes the form of an n-dimensional rectangle as in (1), with obvious modifications if a_i or b_i is infinite for some i (the case $a_i = -\infty$, $b_i = +\infty$ has to be excluded for technical reasons). Hence we are led to study the quantities [9]

$$p = \int_A \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\xi-\mu)^T\Sigma^{-1}(\xi-\mu)} d\xi \quad , \tag{3.1}$$

$$q_i = \int_A \frac{\xi_i}{(2\pi)^{n/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\xi-\mu)^T\Sigma^{-1}(\xi-\mu)} d\xi \quad , \tag{3.2}$$

where ξ has mean μ and covariance matrix Σ, which is assumed to be symmetric and positive definite.

The conditional probability over the rectangle A is then $\Pr\{\xi \in A\} = p$, while the conditional expectation is given by $E\{\xi | \xi \in A\} = (1/p)(q_1,\ldots,q_n)$. For notational simplicity we shall assume that $\mu = 0$ and Σ is a correlation matrix such that $\mathrm{diag}(\Sigma) = (1,\ldots,1)$. This does not constitute a loss of generality since the transformation Z defined by $Z_i(x) = (x_i - \mu_i)/\sqrt{\sigma_{ii}}$ is one-to-one and readily inverted.

2. CONDITIONAL PROBABILITIES

Let us first consider some easy cases.

If $n = 1$, then

$$p = \int_{a_1}^{b_1} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt \quad .$$

This integral can be expanded into a power series which permits fast and accurate evaluation on a computer. The quantity q_i is even easier to obtain by analytic integration:

$$q_i = \int_{a_1}^{b_1} \frac{t}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt = \frac{1}{\sqrt{2\pi}}\left(e^{-a_1^2/2} - e^{-b_1^2/2}\right) \quad ,$$

and accurate computation poses no problem at all. Hence the conditional expectation q_1/p is also known to within machine precision.

For $n = 2$, an efficient method for computing

$$\int_{a_1}^{b_1}\int_{a_2}^{b_2} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}[t_1^2+t_2^2+2\rho t_1 t_2]} \, dt_2 \, dt_1$$

is described in [14, Chap. 8] (see also [5]). Commercial software which calculates p to five-digit accuracy is also available, e.g., the routine MDBNOR in the IMSL library.

The numerators q_1, q_2 can be reduced to four one-dimensional integrals of type (3.1) after completing the square in the exponent of the integrand, and hence can be computed, at least in principle, to arbitrary precision. Details are given in [7].

Thus the conditional mean is available with reasonable accuracy if p is not too small (three decimal places if $p \geq 0.001$). This is sufficient for most practical purposes.

The last trivial case occurs for independent components. If that happens, then Σ is the identity matrix,

$$\Pr\{\xi \in A\} = \prod_{i=1}^{n} \Pr\{\xi_i \in [a_i, b_i]\}$$

and

$$E\{\xi_i \mid \xi \in A\} = E\{\xi_i \mid \xi_i \in [a_i, b_i]\} .$$

Conditional means and conditional expectations can thus be computed by appealing to the one-dimensional case for each component separately.

Let us now turn to the case $n \geq 3$ with dependent components. There are theoretical methods for computing p exactly in three and four dimensions [6,12], but they are not very efficient, and the only viable option seems to be to employ a Monte-Carlo method. This approach has been studied extensively by Deák [2,3,4], whose results are summarized briefly because they also provide the basis for the treatment of the q_i.

A first crude estimator for p is given by

$$\vartheta_0 = \frac{1}{N} \sum_{i=1}^{N} f_0(x^i) \quad . \tag{4}$$

where f_0 is the indicator function of A, that is,

$$f_0(t) = \begin{cases} 1 \text{ if } t \notin A \\ 1 \text{ if } t \in A \end{cases}$$

and $\{x^i, i=1,\ldots,N\}$ is an independent sample of size N from the distribution of the random vector ξ. It can be shown that ϑ_0 is an unbiased estimator for p, but unfortunately its variance only decreases linearly with the sample size, and hence a rather large sample is necessary if any significant accuracy requirement is imposed.

Deák [2] introduced the following elegant decomposition of ξ to take advantage of fairly sophisticated variance reduction schemes. It is not hard to see that ξ can be written as $\xi = \lambda Tz$, where λ is chi-distributed with n degrees of freedom, z is distributed uniformly on S^n, the unit sphere in $\mathbb{R}^n$, λ and z are independent, and T is any matrix such that $T T^{\mathrm{T}} = \Sigma$. Here λ can be regarded as the length of the random vector ξ, while z can be thought of as its direction. An attractive choice for T is the Cholesky decomposition of Σ because efficient methods exist for its generation and because its triangular structure reduces the computational effort necessary to calculate Tz.

Writing F, χ_n, V for the distribution functions of ξ, λ, z, respectively, we obtain

$$p = \int_A dF(\xi) = \int_{\mathbb{R}^n} f_0(\xi) dF(\xi) = \int_{S^n} \int_0^\infty f_0(\lambda Tz) d\chi_n(\lambda)\, dV(z) = \int_{S^n} g(z)\, dV(z) \quad .$$

where

$$g(z) = \int_0^\infty f_0(\lambda Tz) d\chi_n(\lambda) = \chi_n(r_2(z)) - \chi_n(r_1(z))$$

and

$$r_2(z) = \sup\{r \geq 0 \mid rTz \in A\}$$

$$r_1(z) = \inf\{r \geq 0 \mid rTz \in A\} \quad .$$

Sampling then proceeds in the following manner. For each $j=1,\ldots,N$ a randomized system of orthonormalized directions $\{z_1^j,\ldots,z_n^j\}$ is obtained by first generating a sample of n observations from the uniform distribution on S^n and then orthonormalizing them by the Gram–Schmidt procedure. This defines

estimators

$$\vartheta_k = \frac{1}{N}\sum_{j=1}^{N} f_k(z_1^j,\ldots,z_n^j) = \frac{1}{2^k\binom{n}{k}}\sum_{s,i} g(b^j(s,i)) \quad , \tag{5}$$

where

$$s = (s_1,\ldots,s_n) \quad \text{is a sign vector,} \qquad s_l = +1 \text{ or } -1, \quad l=1,\ldots,k$$
$$i = (i_1,\ldots,i_n) \quad \text{is an index vector,} \quad i_l \in \{1,\ldots,n\}, \quad l=1,\ldots,k$$

and

$$b_j(s,i) = \frac{1}{\sqrt{k}}\sum_{j=1}^{k} s_l \, Tz_{i_l}^j \quad .$$

In other words, all directions are considered which can be formed by adding any k (linearly independent) vectors from $\{Tz_1^j,\ldots,Tz_n^j\} \cup \{-Tz_1^j,\ldots,-Tz_n^j\}$ and renormalizing them so that the new vector will still fall onto TS^n. It is straightforward to verify that for each k, ϑ_k is an unbiased estimator for p. In principle, k can be chosen arbitrarily from $\{1,\ldots,n\}$, but the computational effort increases rapidly.

The reduction in the variance of the ϑ_k can be traced primarily to two causes. Firstly, the sample points $b^j(s,i)$ are uniformly and quite regularly distributed on TS^n, and secondly, their number is quite large. From (5) it is obvious that there are $2^k\binom{n}{k}$ terms in the series which could be expected to lead to an approximate reduction in the variance by the same factor. Numerical results show that this is indeed the case.

Thus, if we wish to minimize the variance of the ϑ_k for a given sample size N without regard to computational complexity, the best estimator would be determined by the $\hat{k}$ which maximizes the expression $2^k\binom{n}{k}$. It is not hard to show that $\hat{k} = (2n+1)/3$.

3. CONDITIONAL MEANS

Similar considerations also hold for the q_i.

As a first step we obtain crude estimators

$$\varphi_0^i = \frac{1}{N}\sum_{j=1}^{N} x_j^i f_0(x_j) \quad . \tag{6}$$

These estimators are again unbiased, and the variance decreases linearly in N, just as in (4).

In order to apply Deák's variance reduction scheme, we can again use the decomposition $\xi = \lambda T z$. This gives

$$q_i = \int_A \xi_i \, dF(\xi) = \int_{S^n} \int_{r_1(z)}^{r_2(z)} \lambda T_i z \, d\chi_n(\lambda) \, dV(z) = \tag{7}$$

$$= \int_{S^n} T_i z \int_{r_1(z)}^{r_2(z)} \lambda \, d\chi_n(\lambda) \, dV(z) = \int_{S^n} T_i z \, \tilde{g}(z) \, dV(z) \quad ,$$

where

$$\tilde{g}(z) = \int_{r_1(z)}^{r_2(z)} \lambda \, d\chi_n(\lambda) = \frac{\Gamma(\frac{n+1}{2})\sqrt{2}}{\Gamma(n)} \int_{r_1(z)}^{r_2(z)} d\chi_{n+1}(\lambda) =$$

$$= \frac{\Gamma(\frac{n+1}{2})\sqrt{2}}{\Gamma(n)} \left[\chi_{n+1}(r_2(z)) - \chi_{n+1}(r_1(z)) \right] \quad .$$

An easy way to prove this identity is to note that $\lambda \leq r$ if and only if $\lambda^2 \leq r^2$ and use the density of a chi-squared distribution, which can be written down explicitly.

Thus one arrives at the unbiased estimators

$$\varphi_k^i = \frac{1}{N} \sum_{j=1}^{N} \tilde{f}_k(z_1^j, \ldots, z_n^j), \quad \tilde{f}_k(z_1^j, \ldots, z_n^j) = \frac{1}{2^k \binom{n}{k}} \sum_{s,i} \tilde{g}(b^j(s,i)) \tag{8}$$

with all other notation as in (5).

In particular, it should be clear that the same sample can be used to extract all the information of interest and that $r_1(z), r_2(z)$ have to be computed just once for each sample point. Deák [4] gives an efficient algorithm for doing this.

The difficulty with this approach lies in the fact that the quotient $\varphi_k^i / \vartheta_k$ is not an unbiased estimator for q_i / p [13]. Using a Taylor expansion about the point q_i / p, it is not hard to verify that

$$E\left[\frac{\varphi_k^i}{\vartheta_k} \right] = \frac{q_i}{p} + \frac{1}{Np^2} \left(\sigma_{00} \frac{q_i}{p} - \sigma_{i0} \right) + O(N^{-2}) \quad , \tag{9}$$

where $\sigma_{00} = \mathrm{var}(\vartheta_k)$, $\sigma_{i0} = \mathrm{cov}(\vartheta_k, \varphi_k^i)$.

The bias can be reduced (but not eliminated) by forming the estimators

$$\psi_k^i = \frac{\varphi_k^i}{\vartheta_k} + \frac{1}{N\vartheta_k^2}\left[\mathrm{Cov}(\vartheta_k, \varphi_k^i) - \mathrm{Var}(\vartheta_k)\frac{\varphi_k^i}{\vartheta_k}\right] \tag{10}$$

Note that in formula (9) the true variance and covariance are used, while the quantities appearing in (10) are the sample variance and covariance. Expansion into a Taylor series shows that

$$\mathrm{E}(\psi_k^i) = \frac{q_i}{p} + O(N^{-2}) \quad .$$

It is also possible to use this technique to obtain approximate variances for different estimators. Thus

$$\mathrm{Var}\left[\frac{\varphi_k^i}{\vartheta_k}\right] = \frac{1}{Np^2}\left[\sigma_{ii} - 2\sigma_{i0}\frac{q_i}{p} + \sigma_{00}\left[\frac{q_i}{p}\right]^2\right] + O(N^{-2}) \quad , \tag{11.1}$$

$$\mathrm{Var}(\psi_k^i) = \frac{1}{Np^2}\left[\sigma_{ii} - \sigma_{00}\left[\frac{q_i}{p}\right]^2\right] + O(N^{-2}) \quad . \tag{11.2}$$

Sample output from five small problems is supplied in the Appendix. Several conclusions can be drawn: If p is of moderate size, all methods perform reasonably well, the variance depending on the sample size in the predicted manner. The reduction in the variance due to orthonormal variates is appreciable, but even the crude estimator gives acceptable results. For small p, the performance of the crude estimator grows continually worse, and the orthonormal variates fare only marginally better.

After some reflection it should not be too hard to see why this might be the case. Assume for the moment that p is known with certainty, $p = 0.01$, let us say. In this case dividing by p will result in a variance inflation of the numerator q_i by a factor of 10^4, i.e., if three-digit accuracy is required for the conditional expectation, q_i has to be sampled with a variance of at most 10^{-10}. This requires a rather large N, and uncertainty in the value of p can only make things worse.

The size of the variance obviously tells us little about the accuracy of the estimator. Even for rather large variances, the estimate may be quite close to

the true value. Conversely, a small variance does not mean that the estimate is necessarily good. Of the 520 estimates listed in Tables 1–5, 145 are more than one standard deviation away from the true value. This proportion of 27.9% outliers compares rather well with the number predicted by the law of large numbers, namely $1 - 2\Phi(1) = .317$.

Furthermore, the variances of the quotients ψ_k^i are often seriously underestimated, underscoring the fact that there is also some bias inherent in formula (11). Of course the estimates could be improved by considering higher-order information in the Taylor expansion as well as higher-order sample moments. This would, however, increase the storage and computational requirements, and the possible gain is not easily assessed since the principal problem of variance inflation cannot be addressed in this way. But all may not be lost as far as the envisaged application is concerned, for rectangles of little mass contribute little to the overall bounds in formula (2), and perhaps the accuracy requirements need not be too stringently adhered to in this case.

The next section describes a way to avoid the problem altogether.

4. INDEPENDENCE TRANSFORMATIONS

It is known from elementary matrix algebra that every positive definite, symmetric matrix Σ can be factored into

$$\Sigma = PDP^{\mathrm{T}}$$

where P is orthogonal and D is diagonal or, equivalently,

$$D = P^{\mathrm{T}}\Sigma P \quad .$$

In addition, the random vector $\eta = P^{\mathrm{T}}\xi$ is normally distributed with mean $P^{\mathrm{T}}\mu = 0$ and covariance matrix $\Sigma_\eta = P^{\mathrm{T}}\Sigma_\xi P = D$, such that the components of η are seen to be independent.

Now

$$\Pr\{\xi \in A\} = \Pr\{P^{\mathrm{T}}\xi \in P^{\mathrm{T}}A\} = \Pr\{\eta \in P^{\mathrm{T}}A\} = \int_{P^{\mathrm{T}}A} \mathrm{d}F_\eta(\eta) \qquad (12.1)$$

$$\mathrm{E}\{\xi \mid \xi \in A\} = \mathrm{E}\{P\eta \mid P\eta \in A\} = P\,\mathrm{E}\{\eta \mid \eta \in P^{\mathrm{T}}A\} \quad . \qquad (12.2)$$

The complexion of the problem has changed, but not the complexity. In the space of the random vector η the integrand is separable, but unfortunately

the region of integration has lost its simple appearance. On the other hand, $P^{\mathrm{T}}A$ is still a rectangle, whose coordinates are readily computable, and a partitioning scheme seems to suggest itself. The situation in two dimensions is depicted in Figure 1, and the generalization to higher dimensions is obvious.

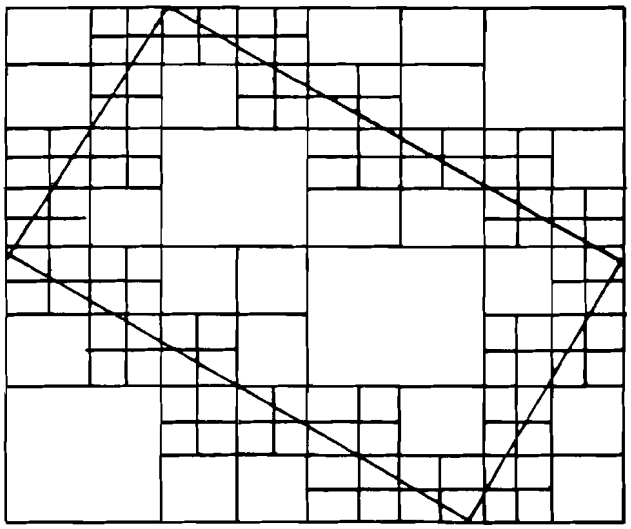

Figure 1.

In contrast to the Monte-Carlo methods of Section 3, it is clear that one can obtain lower bounds on p by considering only subsets which lie entirely within $P^{\mathrm{T}}A$, and upper bounds by taking all subsets whose intersection with $P^{\mathrm{T}}A$ is nonempty. The bounds can be made arbitrarily accurate by suitably refining the partition.

Inner and outer approximations for $\tau_i = P^{\mathrm{T}}q_i$ can be obtained in an analogous way, and error bounds are straightforward if the plane $\eta_i = 0$ is used in the partitions whenever η_i changes sign over $P^{\mathrm{T}}A$.

A provisional version of the algorithm was implemented on the computer – some results are reported in the Appendix. The method works quite well if the components of ξ are independent from the start, but seems extremely slow otherwise. An increase in speed might be effected by Romberg-type extrapolations on data obtained from coarse partitions, but it is clear that even if dramatic improvement can be made on computing times, there are serious limitations to the approach and its use cannot be recommended for $n > 3$ if the components of the random vector ξ are correlated.

On the other hand, independence transformations can be successfully applied to formula (2). It may be recalled that the partitions there were constructed in what seemed a convenient way at the time. In retrospect it turns out that cuts perpendicular to the coordinate axes are more of a hindrance than a help and the partitions should really be guided by the directions of the *eigenvectors* of Σ in the space of the random vector ξ (which coincide with the coordinate directions in the η-space). In other words, it might be easier to exploit independence and find bounds on $E_\eta \tilde{\varphi}(\eta)$, which is easily recognized as being equal in value to $E_\xi \varphi(\xi)$ if $\tilde{\varphi}(\eta) = \varphi(P\eta)$.

Numerical work is currently under way at the University of British Columbia to apply the preceding ideas to a problem of portfolio revision.

APPENDIX

The following five simple problems were used to illustrate the performance of the various methods described in the text.

Problem 1:

$$n = 2,\ \Sigma = I,\ A = [0,2] \times [0,2].$$

Problem 2:

$$n = 3,\ \Sigma = I,\ A = [0,0.5] \times [0,1.0] \times [0,1.5]$$

Problem 3:

$$n = 5,\ \Sigma = I,\ A = [0,0.5] \times [0,1.0] \times [0,1.5] \times [0,2.0] \times [0,2.5]$$

Problem 4:

$$n = 2,\ \Sigma = \begin{bmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{bmatrix},\ A = [0,3.5] \times [-0.5,+0.5]$$

Problem 5:

$$n = 3,\ \Sigma = \begin{bmatrix} 1.0 & 0.5 & 0.5 \\ 0.5 & 1.0 & 0.5 \\ 0.5 & 0.5 & 1.0 \end{bmatrix},\ A = [-0.5,1.0] \times [-0.5,1.0] \times [-0.5,1.0]$$

Tables 1–5 give the results for the Monte-Carlo methods. The notation is as in the text, the second figure in each table entry showing the approximate

standard deviation of the estimator, calculated from the sample variance. The true values of problems 4 and 5 were determined by a numerical integration routine and are accurate to five decimal places. Due to time constraints and heavy use of the computer it was impossible to obtain accurate running times for the different problems. Problems 1 and 4 took about 90 seconds total computer time on a VAX 11/750, the other three slightly under three minutes. These times are not very meaningful, however, as they include total CPU time as well as the time taken to swap the program in and out of the core. Precise figures would probably be much smaller.

Results obtained using the partitioning methods are given in Tables 6 and 7. Only problems 4 and 5 are considered, since the others have independent components, and so the correct values could be found without any partitioning. The following additional notation is used in Tables 6 and 7.

A superscript "o" (p^o, q_1^o, etc.) denotes outer approximations, while superscript "i" stands for inner approximations. The average of the two values is denoted by an overbar. The estimate $\hat{\Psi}$ is defined by $\hat{\Psi} = \frac{1}{\bar{p}} P\bar{\mathbf{q}}$, where $\bar{\mathbf{q}} = (\bar{q}_1, \ldots, \bar{q}_n)^T$. The number of divisions refers to the number of cuts made in each direction, e.g., for the three-dimensional problem 5, #div = 2 indicates that a total of 8 subsets were used. Problem 4 took a little over 4 minutes of total computer time, and on Problem 5 the global time limit of 5 minutes was exceeded, although, as before, these figures have limited information content due to the congestion on the computer.

Table 1. Monte-Carlo results for Problem 1.

	N	p	q^1	q^2	ψ^1	ψ^2
k=0	100	.270±.045	.200±.042	.190±.043	.739±.096	.702±.106
	200	.255±.031	.180±.029	.182±.029	.706±.075	.712±.074
	500	.238±.019	.180±.018	.169±.018	.756±.046	.709±.047
	1000	.215±.013	.161±.012	.152±.012	.749±.034	.706±.035
k=1	100	.2275±.0009	.1766±.0075	.1520±.0078	.782±.033	.675±.034
	200	.2284±.0007	.1693±.0054	.1638±.0053	.744±.024	.720±.023
	500	.2277±.0004	.1654±.0035	.1634±.0034	.727±.015	.719±.015
	1000	.2278±.0003	.1654±.0025	.1643±.0025	.727±.011	.722±.011
k=2	100	.2277±.0009	.1580±.0074	.1720±.0079	.699±.033	.761±.035
	200	.2271±.0007	.1620±.0055	.1634±.0058	.716±.024	.722±.025
	500	.2278±.0004	.1643±.0034	.1655±.0035	.722±.015	.727±.015
	1000	.2277±.0003	.1640±.0025	.1649±.0025	.721±.011	.725±.011
true value		.22777	.16463	.16463	.72279	.72279

Table 2. Monte-Carlo results for Problem 2.

	N	p	q^1	q^2	q^3	ψ^1	ψ^2	ψ^3
$k=0$	100	.0400±.0197	.0131±.0067	.0090±.0064	.0254±.0164	.328±.041	.225±.114	.635±.266
	200	.0400±.0139	.0103±.0040	.0146±.0066	.0224±.0102	.257±.044	.364±.106	.560±.165
	500	.0340±.0081	.0084±.0023	.0160±.0048	.0197±.0062	.246±.032	.472±.087	.579±.119
	1000	.0340±.0057	.0090±.0017	.0144±.0030	.0204±.0044	.264±.024	.424±.053	.601±.082
$k=1$	100	.0289±.0031	.0063±.0007	.0135±.0018	.0183±.0030	.2206±.0078	.4701±.0351	.6430±.0768
	200	.0266±.0021	.0061±.0005	.0128±.0012	.0157±.0020	.2269±.0063	.4848±.0256	.5956±.0576
	500	.0286±.0014	.0067±.0003	.0131±.0008	.0181±.0014	.2360±.0033	.4580±.0158	.6344±.0366
	1000	.0289±.0010	.0070±.0002	.0129±.0005	.0182±.0010	.2422±.0026	.4470±.0114	.6304±.0252
$k=2$	100	.0304±.0018	.0080±.0005	.0139±.0010	.0187±.0017	.2836±.0037	.4622±.0179	.6233±.0443
	200	.0298±.0013	.0075±.0003	.0134±.0007	.0184±.0013	.2524±0.	.4509±.0129	.6197±.0322
	500	.0290±.0008	.0070±.0002	.0135±.0005	.0178±.0008	.2411±.0017	.4660±.0080	.6133±.0206
	1000	.0289±.0006	.0070±.0001	.0136±.0003	.0177±.0006	.2454±.0011	.4687±.0058	.6128±.0147
$k=3$	100	.0251±.0025	.0063±.0006	.0115±.0013	.0146±.0023	.2509±0.	.4614±.0282	.5919±.0735
	200	.0269±.0018	.0066±.0004	.0122±.0010	.0172±.0018	.2456±0.	.4568±.0219	.6448±.0493
	500	.0268±.0011	.0067±.0003	.0124±.0006	.0170±.0011	.2490±0.	.4626±.0130	.6346±.0308
	1000	.0273±.0008	.0067±.0002	.0130±.0005	.0168±.0008	.2465±.0010	.4749±.0091	.6150±.0209
true value		.02831	.00693	.01302	.01761	.24479	.45991	.62196

Table 3(a) Monte-Carlo integration for Problem 3.

	N	p	q^1	q^2	q^3	q^4	q^5
$k=0$	100	.0100±.0100	.0047±.0047	.0061±.0061	.0149±.0149	.0051±.0051	.0030±.0030
	200	.0050±.0050	.0023±.0023	.0031±.0031	.0074±.0074	.0026±.0026	.0015±.0015
	500	.0060±.0035	.0021±.0013	.0028±.0019	.0059±.0038	.0049±.0037	.0010±.0007
	1000	.0080±.0028	.0023±.0009	.0041±.0016	.0056±.0025	.0063±.0028	.0043±.0021
$k=1$	100	.00860±.00196	.00211±.00051	.00422±.00110	.00522±.00138	.00650±.00206	.00636±.00217
	200	.00663±.00120	.00173±.00031	.00316±.00064	.00394±.00085	.00483±.00123	.00510±.00140
	500	.00648±.00078	.00171±.00021	.00309±.00040	.00438±.00062	.00464±.00074	.00519±.00091
	1000	.00684±.00055	.00176±.00015	.00322±.00029	.00417±.00041	.00485±.00055	.00544±.00064
$k=2$	100	.00621±.00068	.00164±.00018	.00283±.00033	.00408±.00052	.00425±.00066	.00455±.00077
	200	.00609±.00049	.00156±.00013	.00274±.00024	.00389±.00038	.00418±.00046	.00458±.00059
	500	.00587±.00031	.00148±.00008	.00273±.00016	.00357±.00023	.00432±.00032	.00450±.00037
	1000	.00636±.00024	.00158±.00006	.00290±.00012	.00396±.00016	.00468±.00023	.00492±.00028
$k=3$	100	.00603±.00050	.00145±.00012	.00297±.00026	.00335±.00016	.00476±.00050	.00446±.00057
	200	.00620±.00032	.00156±.00008	.00303±.00017	.00359±.00023	.00461±.00033	.00442±.00036
	500	.00620±.00021	.00156±.00005	.00297±.00011	.00381±.00016	.00434±.00020	.00460±.00024
	1000	.00650±.00016	.00160±.00004	.00305±.00008	.00404±.00012	.00463±.00015	.00500±.00019
true value		.00667	.00163	.00307	.00415	.00482	.00515

Table 3(b) Conditional expectations for Problem 3.

	N	ψ^1	ψ^2	ψ^3	ψ^4	ψ^5
$k=0$	100	.4665±.0001	.611±0.	1.489±.0001	.514±.00004	.295±0.
	200	.4665±.0001	.611±0.	1.489±.0001	.514±.00004	.295±0.
	500	.3546±.0583	.463±.156	.987±.282	.816±.406	.166±.053
	1000	.2850±.0435	.511±.088	.696±.183	.781±.223	.535±.175
$k=1$	100	.2465±.0203	.496±.062	.612±.083	.770±.167	.755±.188
	200	.2666±0.	.477±.045	.596±.069	.738±.130	.779±.159
	500	.2646±.0072	.477±.025	.678±.050	.718±.076	.805±.102
	1000	.2573±.0049	.476±.018	.610±.033	.711±.056	.797±.068
$k=2$	100	.2668±0.	.460±.019	.665±.041	.694±.076	.744±.094
	200	.2567±0.	.452±.015	.643±.034	.691±.050	.757±.074
	500	.2550±.0020	.466±.010	.610±.023	.738±.037	.768±.049
	1000	.2491±.0030	.455±.008	.624±.017	.736±.024	.774±.072
$k=3$	100	.2423±.0007	.497±.013	.561±.036	799±.052	.750±.033
	200	.2522±.0032	.491±.011	.582±.021	.747±.036	.717±.045
	500	.2519±.0007	.480±.008	.616±.015	.702±.022	.744±.030
	1000	.2463±0.	.469±.005	.623±.011	.714±.016	.770±.022
true value		.24479	.45991	.62196	.72279	.77242

Table 4. Monte-Carlo results for Problem 4.

	N	p	q^1	q^2	ψ^1	ψ^2
$k=0$	100	.220±.042	.118±.033	-.0065±.0135	.536±.110	-.0295±.0611
	200	.230±.030	.156±.029	.0010±.0101	.676±.088	.0045±.0440
	500	.204±.018	.143±.017	.0036±.0057	.700±.054	.0174±.0277
	1000	.186±.012	.125±.011	.0029±.0039	.670±.038	.0157±.0208
$k=1$	100	.1802±.0068	.1186±.0098	.0036±.0033	.664±.048	.0199±.0183
	200	.1899±.0047	.1306±.0069	.0040±.0026	.691±.032	.0213±.0136
	500	.1888±.0030	.1302±.0043	.0065±.0016	.691±.020	.0343±.0084
	1000	.1904±.0021	.1326±.0031	.0066±.0011	.697±.014	.0347±.0059
$k=2$	100	.2042±.0068	.1554±.0103	.0112±.0037	.767±.043	.0554±.0179
	200	.1938±.0047	.1389±.0070	.0107±.0026	.720±.032	.0556±.0133
	500	.1942±.0030	.1382±.0044	.0079±.0016	.713±.020	.0409±.0084
	1000	.1925±.0021	.1358±.0031	.0078±.0012	.704±.014	.0405±.0060
true value		.19145	.13402	.00705	.70003	.03681

Table 5. Monte-Carlo results for Problem 5.

	N	p	q^1	q^2	q^3	ψ^1	ψ^2	ψ^3
$k=0$	100	.220±.042	.0198±.0177	.0455±.0201	.0508±.0252	.0900±.0785	.2067±.0824	.2310±.1060
	200	.225±.030	.0435±.0146	.0453±.0146	.0488±.0169	.1954±.0596	.2014±.0593	.2168±.0696
	500	.220±.019	.0490±.0093	.0456±.0095	.0528±.0103	.2226±.0377	.2072±.0393	.2398±.0423
	1000	.206±.013	.0468±.0062	.0428±.0063	.0440±.0069	.2274±.0264	.2077±.0278	.2136±.0308
$k=1$	100	.1942±.0030	.0415±.0021	.0439±.0025	.0497±.0029	.2152±.0102	.2278±.0125	.2583±.0143
	200	.1941±.0021	.0436±.0016	.0442±.0018	.0479±.0019	.2257±.0079	.2288±.0089	.2476±.0096
	500	.1944±.0014	.0446±.0011	.0433±.0011	.0452±.0012	.2295±.0054	.2232±.0055	.2330±.0061
	1000	.1947±.0010	.0440±.0008	.0434±.0008	.0448±.0009	.2263±.0039	.2230±.0039	.2303±.0042
$k=2$	100	.1970±.0015	.0453±.0017	.0450±.0017	.0412±.0017	.2319±.0085	.2308±.0083	.2112±.0085
	200	.1963±.0010	.0448±.0012	.0429±.0012	.0428±.0011	.2293±.0058	.2197±.0059	.2191±.0057
	500	.1959±.0006	.0441±.0007	.0430±.0007	.0445±.0007	.2255±.0036	.2200±.0036	.2274±.0035
	1000	.1957±.0004	.0439±.0005	.0432±.0005	.0449±.0005	.2248±.0025	.2211±.0025	.2296±.0025
$k=3$	100	.1934±.0021	.0413±.0022	.0427±.0018	.0461±.0021	.2154±.0109	.2226±.0092	.2402±.0104
	200	.1955±.0016	.0432±.0015	.0452±.0014	.0455±.0015	.2220±.0076	.2320±.0071	.2339±.0072
	500	.1962±.0011	.0435±.0010	.0449±.0009	0463±.0010	.2223±.0047	.2293±.0045	.2364±.0047
	1000	.1959±.0008	.0441±.0007	.0448±.0007	.0458±.0007	.2251±.0033	.2290±.0033	.2340±.0034
true value		.19716	.04439	.04439	.04439	.22514	.22514	.22514

Table 6. Results from partitioning on Problem 4.

# div	p^0	p^1	$\bar{p}$	q^0	q^1	$\bar{q}$	$\bar{\psi}$
2	.4170	0.	.2085	.1501 .3006	0. 0.	.0750 .1503	.7642 .2552
4	.3358	0.	.1679	.1247 .1693	0. 0.	.0624 .0852	.6191 .0939
8	.2987	.0525	.1756	.1114 .1376	.0272 .0277	.0693 .0827	.6119 .0537
16	.2575	.1223	.1899	.1070 .1263	.0650 .0687	.0860 .0975	.6830 .0429
32	.2311	.1609	.1960	.1035 .1200	.0828 .0903	.0931 .1052	.7153 .0436
64	.2112	.1751	.1931	.0963 .1091	.0858 .0944	.0911 .1017	.7058 .0390
128	.2008	.1825	.1916	.0925 .1036	.0872 .0963	.0899 .0999	.7005 .0371
256	.1954	.1862	.1908	.0906 .1009	.0879 .0972	.0893 .0990	.6977 .0362
512	.1935	.1889	.1912	.0903 .1004	.0889 .0986	.0896 .0995	.6992 .0366
1024	.1926	.1903	.1914	.0901 .1002	.0894 .0993	.0898 .0997	.6999 .0368
∞			.19145			.08978 .09969	.70003 .03681

Table 7. Results from partitioning on Problem 5.

# div	p^0	p^1	$\bar{p}$	q^0	q^1	$\bar{q}$	Φ
2	.5022	0.	.2511	.0000 .0000 .1631	0. 0. 0.	.0000 .0000 .0816	.1875 .1875 .1875
4	.4239	0.	.2120	.4276 -.4399 .1451	0. 0. 0.	.2138 -.2199 .0726	.2050 .2020 .1860
8	.3343	.0775	.2059	.0008 -.0008 .1215	-.0002 -.0002 .0320	.0003 -.0005 .0768	.2163 .2167 .2129
16	.2678	.1286	.1982	.00016 -.00004 .1007	-.00006 .00000 .0517	.00005 -.00002 .0762	.2221 .2219 .2217
32	.2335	.1612	.1974	.00007 -.00002 .0894	-.00004 .00001 .0640	.00002 .00000 .0767	.2244 .2243 .2243
64	.2155	.1789	.1972	.00003 -.00001 .0833	-.00002 .00001 .0704	.00000 .00000 .0768	.2250 .2249 .2249
∞			.19716			0. 0. .07688	.22514 .22514 .22514

REFERENCES

1. J. Birge and R. Wets, "Designing approximation schemes for stochastic optimization problems, in particular for stochastic problems with recourse", *Math. Prog. Study* (to appear).

2. I. Deák, "Computation of multiple normal probabilities", in P. Kall and A. Prékopa (Eds.), *Recent Results in Stochastic Programming*, Springer-Verlag, New York, 1980, pp.107–120.

3. I. Deák, "Three-digit accurate multiple normal probabilities", *Numerische Math.* 35 (1980) 369–380.

4. I. Deák, "Procedures for the computation of normal distribution function values and probabilities of rectangles", Working Paper No. MO/48, Computer and Automation Institute, The Hungarian Academy of Sciences, Budapest, 1983.

5. T.G. Donnelly, "Bivariate normal distribution, algorithm 462", *Comm. ACM* 16 (1973) 638.

6. J.E. Dutt, "A representation of multivariate normal probability integrals by integral transforms", *Biometrika* 60 (1973) 637–645.

7. H.I. Gassmann, "Means of a bivariate normal distribution over a rectangle: an efficient algorithm", Working Paper No. 971, Faculty of Commerce, The University of British Columbia, 1983.

8. H.I. Gassmann and W.T. Ziemba, "A tight upper bound for the expectation of a convex function of a multivariate random variable", *Math Prog. Study* (to appear).

9. N.L. Johnson and S. Kotz, *Distributions in Statistics*, Vol. 4, Wiley, New York, 1972.

10. P. Kall, "Computational methods for solving two-stage stochastic linear programming problems", *Z. Angew. Math. Phys.* 30 (1979) 261–271.

11. A. Madansky, "Bounds on the expectation of a convex function of a multivariate random variable", *Ann. Math. Stat.* 30(1959) 743–746.

12. R.C. Milton, "Computer evaluation of the multivariate normal integral", *Technometrics* 14 (1972) 881–889.

13. A.M. Mood, Greybill and Bose, *Introduction to the Theory of Statistics*, McGraw Hill, New York, 1974.

14. D.B. Owen, *Handbook of Statistical Tables*, Addison-Wesley, Reading, Massachusetts, 1962.