ON ERROR BOUNDS FOR APPROXIMATIONS TO MULTIVARIATE DISTRIBUTIONS II

BY

BJØRN SUNDT[1] AND RALUCA VERNIC[2]

ABSTRACT

In the present paper, we study error bounds for approximations to multivariate distributions. In particular, we discuss some general versions of compound multivariate distributions and look at distributions of dependent random variables constructed by linear transforms of independent random variables or vectors. Special attention is paid to the case when the support of the original distribution is restricted. We also look at some applications with multivariate Bernoulli distributions.

1. INTRODUCTION

Dhaene and De Pril (1994) presented a framework for approximations to univariate aggregate claims distributions, and within this framework they developed general results for error bounds. Some of these results were reformulated in terms of De Pril transforms and discussed by Dhaene and Sundt (1998). Dhaene and Sundt (1997) discussed some error bounds without introducing the De Pril transform. Sundt (2000*b*) extended some of the approximations and error bounds of these papers to the multivariate case, utilising the multivariate De Pril transform introduced in Sundt (2000*a*).

In this paper, we study some error bounds for approximations to multivariate distributions, extending the framework of Sundt (2000*b*) and using the same measure for the distance between the exact distribution and the approximation.

In Section 2, we present some notation and conventions.

In Section 3, we introduce the distance measure as defined in Sundt (2000*b*) and deduce an alternative expression that can be more convenient in some cases when the exact distribution or the approximation has a finite support.

Linear transforms are often used to construct dependent random variables from independent random variables or vectors. In Section 4, we discuss error

[1] Vital Forsikring ASA, Norway.
[2] "Ovidius" University of Constanta, Romania.

ASTIN BULLETIN, Vol. 32, No. 1, 2002, pp. 57-69

bounds for approximations to multivariate distributions when the approximation is constructed by the same linear transform as the exact distribution.

Sundt (2000*b*) discussed multivariate compound distributions in the case with univariate counting distribution and multivariate severity distribution. In this case, he showed that if we approximate the compound distribution by approximating the counting distribution and keeping the severity distribution unchanged, then the distance between the counting distribution and its approximation is an upper bound for the distance between the compound distribution and its approximation. In Section 5, we show that this result holds for a much more general class of compound distributions.

Finally, in Section 6 we apply the theory developed in the earlier sections to some examples with multivariate Bernoulli distributions.

2. Notation and conventions

A column vector will be denoted by a bold letter, and its elements will normally be denoted by the corresponding italic with the number of the element indicated by a subscript; the subscript • indicates the sum of the elements of the vector. For a positive integer m, let $\mathbb{N}_m$ denote the set of $m \times 1$ vectors where all elements are non-negative integers. By $\mathbf{0}$ we denote the $m \times 1$ vector consisting of only zeros and by $\mathbf{e}_j$ the jth $m \times 1$ unit vector; for convenience, we do not indicate the dimension m in the notation for these vectors, hoping that this will not cause confusion. We also introduce $\mathbb{N}_{m+} = \mathbb{N}_m \setminus \{\mathbf{0}\}$. For $\mathbf{x}, \mathbf{y} \in \mathbb{N}_m$, by $\mathbf{y} \le \mathbf{x}$ we shall mean that $\mathbf{x} - \mathbf{y} \in \mathbb{N}_m$.

In this paper, we shall represent probability distributions by their probability functions. Therefore, for convenience, we shall refer to probability functions as distributions.

By $\mathcal{P}_m$ and $\mathcal{F}_m$, we denote the classes of distributions and functions, respectively, on $\mathbb{N}_m$; by $\mathcal{P}_{m0}$ the class of distributions in $\mathcal{P}_m$ with a positive mass at zero, and by $\mathcal{P}_{m+}$ the class of distributions on $\mathbb{N}_{m+}$.

For $f \in \mathcal{F}_m$, we let P_f denote the signed measure induced by f, that is,

$$P_f(A) = \sum_{\mathbf{x} \in A} f(\mathbf{x}). \qquad (A \subset \mathbb{N}_m)$$

Furthermore, we introduce $v(f) = P_f(\mathbb{N}_m)$ and the marginal means $\mu_j(f) = \sum_{\mathbf{x} \in \mathbb{N}_m} x_j f(\mathbf{x})$ for $j = 1, \ldots, m$. When appearing in a formula, we assume that these quantities exist and are finite. If $f \in \mathcal{P}_m$, then $v(f) = 1$.

We use the notation $x_+ = \max(x, 0)$ for any real number x.

3. A distance measure

Like in Sundt (2000*b*), we apply the measure

$$\varepsilon(f, g) = \sum_{\mathbf{x} \in \mathbb{N}_m} |f(\mathbf{x}) - g(\mathbf{x})| \qquad (3.1)$$

https://doi.org/10.2143/AST.32.1.1014 Published online by Cambridge University Press

for the distance between functions f and g in $\mathcal{F}_m$. This is a metric on $\mathcal{F}_m$. The following theorem gives an alternative expression for $\varepsilon(f, g)$. As we shall see later, this new expression will be convenient in some cases when one of the functions has restricted support.

Theorem 3.1. *If $f, g \in \mathcal{F}_m$ satisfy $v(|f|) < \infty$ and $v(|g|) < \infty$, then*

$$\varepsilon(f,g) = 2\sum_{\mathbf{x}\in\mathbb{N}_m}(f(\mathbf{x}) - g(\mathbf{x}))_+ - v(f) + v(g). \tag{3.2}$$

Proof. We have

$$\varepsilon(f,g) = \sum_{\mathbf{x}\in\mathbb{N}_m}|f(\mathbf{x}) - g(\mathbf{x})| = \sum_{\mathbf{x}\in\mathbb{N}_m}(2(f(\mathbf{x}) - g(\mathbf{x}))_+ - (f(\mathbf{x}) - g(\mathbf{x}))),$$

from which we obtain (3.2). □

The following corollary follows easily from Theorem 3.1.

Corollary 3.1. *Under the assumptions of Theorem 3.1, we assume that $g(\mathbf{x}) \geq f(\mathbf{x})$ when $\mathbf{x} \in \mathbb{N}_m \setminus A$ for some $A \subset \mathbb{N}_m$. Then*

$$\varepsilon(f,g) = 2\sum_{\mathbf{x}\in A}(f(\mathbf{x}) - g(\mathbf{x}))_+ - v(f) + v(g).$$

We now turn to the case when $f, g \in \mathcal{P}_m$.

As a distribution is non-negative and sums to one, the two following corollaries follow easily from Theorem 3.1 and Corollary 3.1.

Corollary 3.2. *If $f, g \in \mathcal{P}_m$, then*

$$\varepsilon(f,g) = 2\sum_{\mathbf{x}\in\mathbb{N}_m}(f(\mathbf{x}) - g(\mathbf{x}))_+. \tag{3.3}$$

Corollary 3.3. *If $f, g \in \mathcal{P}_m$ and $f(\mathbf{x}) = 0$ when $\mathbf{x} \in \mathbb{N}_m \setminus A$ for some $A \subset \mathbb{N}_m$, then*

$$\varepsilon(f,g) = 2\sum_{\mathbf{x}\in A}(f(\mathbf{x}) - g(\mathbf{x}))_+.$$

Corollary 3.4. *If $f, g \in \mathcal{P}_m$,* then

$$\varepsilon(f,g) = 2\sup_{A\subset\mathbb{N}_m}(P_f(A) - P_g(A)) = \sup_{A\subset\mathbb{N}_m}|P_f(A) - P_g(A)|.$$

Proof. The first equality follows from (3.3), and as $\varepsilon(f, g) = \varepsilon(g, f)$, we obtain the second equality. □

https://doi.org/10.2143/AST.32.1.1014 Published online by Cambridge University Press

4. CONSTRUCTION OF DEPENDENT RANDOM VARIABLES BY LINEAR TRANSFORMS

4A. In insurance mathematics, one has traditionally assumed that risks within an insurance portfolio are independent. However, there is a growing interest for modelling dependent risks. One way of doing this, is to construct dependent random variables by using a linear transform of a vector of independent random variables, cf. e.g. Ambagaspitiya (1999). In the present section, we shall work in a somewhat more general framework as we consider linear transforms of vectors of independent sub-vectors of random variables. We approximate the distribution of such a vector by another distribution constructed in the same way, using the same linear transform.

4B. We recall that the convolution $f*g$ of two functions $f,\ g \in \mathcal{F}_m$ is defined by

$$(f*g)(\mathbf{x}) = \sum_{\mathbf{0} \le \mathbf{y} \le \mathbf{x}} g(\mathbf{y}) f(\mathbf{x}-\mathbf{y}). \tag{4.1}$$

By Theorem 4.1 in Sundt (2000*b*) we have that if $f_i,\ g_i \in \mathcal{P}_m$ $(i = 1, \ldots, t)$, then

$$\varepsilon\left(\overset{t}{\underset{i=1}{*}} f_i, \overset{t}{\underset{i=1}{*}} g_i\right) \le \sum_{i=1}^{t} \varepsilon(f_i, g_i). \tag{4.2}$$

In Theorem 4.1 we shall give a generalisation of this result. To prove that theorem, we shall need the following lemma.

Lemma 4.1. *Let $\tilde{f},\ \tilde{g} \in \mathcal{F}_r$ and define $f,\ g \in \mathcal{F}_m$ by*

$$f(\mathbf{x}) = \begin{cases} \tilde{f}(\mathbf{u}) & (\mathbf{x} = \mathbf{A}\mathbf{u};\ \mathbf{u} \in \mathbb{N}_r) \\ 0 & (\text{otherwise}) \end{cases}$$

$$g(\mathbf{x}) = \begin{cases} \tilde{g}(\mathbf{u}) & (\mathbf{x} = \mathbf{A}\mathbf{u};\ \mathbf{u} \in \mathbb{N}_r) \\ 0, & (\text{otherwise}) \end{cases}$$

where $\mathbf{A}$ is an $m \times r$ matrix with non-negative integer-valued elements and full rank $r \le m$. Then $\varepsilon(f, g) = \varepsilon(\tilde{f}, \tilde{g})$.

Proof. We have

$$\varepsilon(f,g) = \sum_{\mathbf{x} \in \mathbb{N}_m} |f(\mathbf{x}) - g(\mathbf{x})| = \sum_{\mathbf{u} \in \mathbb{N}_r} |f(\mathbf{A}\mathbf{u}) - g(\mathbf{A}\mathbf{u})| = \sum_{\mathbf{u} \in \mathbb{N}_r} |\tilde{f}(\mathbf{u}) - \tilde{g}(\mathbf{u})| = \varepsilon(\tilde{f}, \tilde{g}),$$

which proves the lemma. □

The impact of Lemma 4.1 is more easily seen in the case of distributions; if the $r \times 1$ random vector $\mathbf{U}$ has distribution $\tilde{f}$, then $\mathbf{A}\mathbf{U}$ has distribution f.

To more easily see the impact of the following result, we express it in terms of random variables.

https://doi.org/10.2143/AST.32.1.1014 Published online by Cambridge University Press

Theorem 4.1. *Let* $\mathbf{U} = (\mathbf{U}^{(1)\prime}, \ldots, \mathbf{U}^{(t)\prime})'$ *and* $\mathbf{V} = (\mathbf{V}^{(1)\prime}, \ldots, \mathbf{V}^{(t)\prime})'$ *be* $r \times 1$ *random vectors of mutually independent sub-vectors* $\mathbf{U}^{(1)}, \ldots, \mathbf{U}^{(t)}$ *and* $\mathbf{V}^{(1)}, \ldots, \mathbf{V}^{(t)}$ *respectively. For* $i = 1, \ldots, t$, $\mathbf{U}^{(i)}$ *and* $\mathbf{V}^{(i)}$ *are of dimension* $r_i \times 1$ *with* $\sum_{i=1}^{t} r_i = r$, *and their distributions are* f_i *and* g_i *respectively, both in* $\mathcal{P}_{r_i}$. *Let* f *and* g *denote the distributions of the random vectors* $\mathbf{X} = \mathbf{AU}$ *and* $\mathbf{Y} = \mathbf{AV}$ *respectively, where* $\mathbf{A} = (\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(t)})$ *is an* $m \times r$ *matrix with non-negative integer-valued elements; for* $i = 1, \ldots, t$, *the sub-matrix* $\mathbf{A}^{(i)}$ *is of dimension* $m \times r_i$ *and full rank* $r_i \leq m$. *Then*

$$\varepsilon(f,g) \leq \sum_{i=1}^{t} \varepsilon(f_i, g_i). \tag{4.3}$$

Proof. For $i = 1, \ldots, t$, let $\tilde{f}_i$ and $\tilde{g}_i$ denote the distributions of $\mathbf{X}^{(i)} = \mathbf{A}^{(i)}\mathbf{U}^{(i)}$ and $\mathbf{Y}^{(i)} = \mathbf{A}^{(i)}\mathbf{V}^{(i)}$ respectively. Lemma 4.1 gives that $\varepsilon(\tilde{f}_i, \tilde{g}_i) = \varepsilon(f_i, g_i)$. However, we also have that $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(t)}$ are mutually independent, and that $\mathbf{Y}^{(1)}, \ldots, \mathbf{Y}^{(t)}$ are mutually independent. Furthermore, $\mathbf{X} = \sum_{i=1}^{t} \mathbf{X}^{(i)}$ and $\mathbf{Y} = \sum_{i=1}^{t} \mathbf{Y}^{(i)}$. Thus, $f = *_{i=1}^{t} \tilde{f}_i$ and $g = *_{i=1}^{t} \tilde{g}_i$, and application of (4.2) gives

$$\varepsilon(f,g) = \varepsilon\left(*_{i=1}^{t} \tilde{f}_i, *_{i=1}^{t} \tilde{g}_i\right) \leq \sum_{i=1}^{t} \varepsilon(\tilde{f}_i, \tilde{g}_i) = \sum_{i=1}^{t} \varepsilon(f_i, g_i),$$

which proves the theorem. □

4C. Let us now look at some special cases:

i) For each i, we let $r_i = m$ and let $\mathbf{A}_i$ be the $m \times m$ identity matrix. Then $f = *_{i=1}^{t} f_i$ and $g = *_{i=1}^{t} g_i$, and (4.3) reduces to (4.2).

ii) We now consider a method that is often applied for obtaining multivariate distributions from univariate distributions. Let $\mathbf{U} = (U_0, U_1, \ldots, U_m)'$ be a vector of independent non-negative integer-valued random variables, and let $\mathbf{X} = (X_1, \ldots, X_m)'$ with $X_i = U_0 + U_i$ for $i = 1, \ldots, m$. The distribution of $\mathbf{X}$ is denoted by f and the distribution of U_i by f_i ($i = 0, 1, \ldots, m$). The random vectors $\mathbf{V}$ and $\mathbf{Y}$ are defined correspondingly, and we denote by g the distribution of $\mathbf{Y}$, and by g_i the distribution of V_i ($i = 0, 1, \ldots, m$). Then $\mathbf{X} = \mathbf{AU}$ and $\mathbf{Y} = \mathbf{AV}$ with $\mathbf{A} = (\sum_{i=1}^{m} \mathbf{e}_i, \mathbf{e}_1, \ldots, \mathbf{e}_m)$, and Theorem 4.1 gives that $\varepsilon(f, g) \leq \sum_{i=0}^{m} \varepsilon(f_i, g_i)$.
In the particular case when the g_is are Poisson distributions, there exist recursions for evaluating g, or, more generally, compound distributions of Type 2 of subsection 5A where g is the counting distribution, cf. Sundt (2000*c*).

iii) We now assume that $r = m$, and that $\mathbf{A}$ is the $m \times$ m identity matrix. Then $\mathbf{X} = \mathbf{U}$ and $\mathbf{Y} = \mathbf{V}$, and we have

$$f(\mathbf{x}) = \prod_{i=1}^{t} f_i(\mathbf{x}^{(i)}); \quad g(\mathbf{x}) = \prod_{i=1}^{t} g_i(\mathbf{x}^{(i)}) \quad (\mathbf{x} \in \mathbb{N}_m) \tag{4.4}$$

https://doi.org/10.2143/AST.32.1.1014 Published online by Cambridge University Press

with $\mathbf{x}^{(i)} = \left(x_{\sum_{j=1}^{i-1} r_j + 1}, \ldots, x_{\sum_{j=1}^{i} r_j}\right)'$ for $i = 1, \ldots, t$. We see that if $\mathbf{X}$ can be split into mutually independent sub-vectors and the corresponding sub-vectors of $\mathbf{Y}$ are also mutually independent, then the distance between the distributions of $\mathbf{X}$ and $\mathbf{Y}$ is bounded by the sum of the distances within each pair of distributions of corresponding sub-vectors. This property becomes even more interesting if all of these sub-vectors are of the same dimension. Then the bound given by Theorem 4.1 for the distance between the distributions of $\mathbf{U}$ and $\mathbf{V}$, is the same as the bound given by (4.2) for the distance between the distributions of $\sum_{i=1}^{t} \mathbf{U}^{(i)}$ and $\sum_{i=1}^{t} \mathbf{V}^{(i)}$. As we have found the same upper bound for these two distances, it is natural to ask what relation is there between them. From the following theorem, we obtain that the distance between the distributions of $\mathbf{U}$ and $\mathbf{V}$ is greater than or equal to the distance between the distributions of $\sum_{i=1}^{t} \mathbf{U}^{(i)}$ and $\sum_{i=1}^{t} \mathbf{V}^{(i)}$.

Theorem 4.2. *For $i = 1, \ldots, t$,* let $f_i, g_i \in \mathcal{F}_r$, *and let $f, g \in \mathcal{F}_{tr}$ be defined by* (4.4) *with $m = tr$. Then*

$$\varepsilon\left(\overset{t}{\underset{i=1}{*}} f_i, \overset{t}{\underset{i=1}{*}} g_i\right) \leq \varepsilon(f,g).$$

Proof. We have

$$\varepsilon\left(\overset{t}{\underset{i=1}{*}} f_i, \overset{t}{\underset{i=1}{*}} g_i\right) = \sum_{\mathbf{x} \in \mathbb{N}_r} \left| \left(\overset{t}{\underset{i=1}{*}} f_i\right)(\mathbf{x}) - \left(\overset{t}{\underset{i=1}{*}} g_i\right)(\mathbf{x}) \right| =$$

$$\sum_{\mathbf{x} \in \mathbb{N}_r} \left| \sum_{\left\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(t)} \in \mathbb{N}_r : \sum_{i=1}^{t} \mathbf{x}^{(i)} = \mathbf{x}\right\}} \left(\left(\prod_{i=1}^{t} f_i(\mathbf{x}^{(i)}) \right) - \left(\prod_{i=1}^{t} g_i(\mathbf{x}^{(i)}) \right) \right) \right| \leq$$

$$\sum_{\mathbf{x} \in \mathbb{N}_r} \sum_{\left\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(t)} \in \mathbb{N}_r : \sum_{i=1}^{t} \mathbf{x}^{(i)} = \mathbf{x}\right\}} \left| \left(\prod_{i=1}^{t} f_i(\mathbf{x}^{(i)}) \right) - \left(\prod_{i=1}^{t} g_i(\mathbf{x}^{(i)}) \right) \right| =$$

$$\sum_{\mathbf{x}^{(1)} \in \mathbb{N}_r} \sum_{\mathbf{x}^{(2)} \in \mathbb{N}_r} \cdots \sum_{\mathbf{x}^{(t)} \in \mathbb{N}_r} \left| \left(\prod_{i=1}^{t} f_i(\mathbf{x}^{(i)}) \right) - \left(\prod_{i=1}^{t} g_i(\mathbf{x}^{(i)}) \right) \right| =$$

$$\sum_{\mathbf{x} \in \mathbb{N}_r} | f(\mathbf{x}) - g(\mathbf{x}) | = \varepsilon(f,g). \qquad \square$$

By combining Theorems 4.1 and 4.2, we see that if $f_i, g_i \in \mathcal{P}_r$ for $i = 1, 2, \ldots, t$, then

$$\varepsilon\left(\overset{t}{\underset{i=1}{*}} f_i, \overset{t}{\underset{i=1}{*}} g_i\right) \leq \varepsilon(f,g) \leq \sum_{i=1}^{t} \varepsilon(f_i, g_i).$$

Thus, as a bound, $\sum_{i=0}^{t} \varepsilon(f_i, g_i)$ is sharper for $\varepsilon(f, g)$ than for $\varepsilon(*_{i=1}^{t} f_i, *_{i=1}^{t} g_i)$.

https://doi.org/10.2143/AST.32.1.1014 Published online by Cambridge University Press

5. Compound distributions

5A. In the univariate case, a compound distribution is the distribution of the sum of independent and identically distributed random variables, where the number of terms is itself a random variable assumed to be independent of the terms. Denoting by $p \in \mathcal{P}_1$ the distribution of the number of terms (the counting distribution) and by $h \in \mathcal{P}_1$ the distribution of the terms (the severity distribution), the compound distribution $p \vee h$ is given by

$$(p \vee h)(x) = \sum_{n \in \mathbb{N}_1} p(n) h^{n^*}(x). \quad (x \in \mathbb{N}_1) \tag{5.1}$$

In the actuarial literature, the concept of compound distributions has been extended to the multivariate case in two different ways:

1. The counting distribution is still univariate, but the severities are m-dimensional. In this setting, we have

$$(p \vee h)(\mathbf{x}) = \sum_{n \in \mathbb{N}_1} p(n) h^{n^*}(\mathbf{x}) \quad (\mathbf{x} \in \mathbb{N}_m) \tag{5.2}$$

with $p \in \mathcal{P}_1$ and $h \in \mathcal{P}_m$. In an actuarial setting, we can interpret the counting variable as the number of claim events in an insurance portfolio, and the ith severity vector as the vector of claim amounts incurred by the ith claim event for the m policies within the portfolio.

2. The counting distribution is m-variate, and there are m univariate severity distributions. In this setting, we have

$$(p \vee \mathbf{h})(\mathbf{x}) = \sum_{\mathbf{n} \in \mathbb{N}_m} p(\mathbf{n}) \prod_{j=1}^{m} h_j^{n_j *}(x_j), \quad (\mathbf{x} \in \mathbb{N}_m) \tag{5.3}$$

where $p \in \mathcal{P}_m$ and $\mathbf{h} = (h_1, \ldots, h_m)$ is the vector of the m severity distributions $h_1, \ldots, h_m \in \mathcal{P}_1$. In an actuarial setting, we can interpret the jth counting variable as the number of claims for the jth policy in an insurance portfolio and the jth severity distribution as the severity distribution of that policy.

Sundt (2000*c*) discusses the second setting and gives references for both settings.

A natural combination of the two models would be to assume that the counting distribution is multivariate like in the second setting, but now assume that the severity distributions could also be multivariate. Keeping m as the dimension of the aggregate severity vector, we denote by k the dimension of the counting vector, so that $p \in \mathcal{P}_k$, and let $\mathbf{h} = (h_1, \ldots, h_k)$ where $h_j \in \mathcal{P}_{m_j}$ for $j = 1, \ldots, k$ with $\sum_{j=1}^{k} m_j = m$. We obtain

$$(p \vee \mathbf{h})(\mathbf{x}) = \sum_{\mathbf{n} \in \mathbb{N}_k} p(\mathbf{n}) \prod_{j=1}^{k} h_j^{n_j *}(\mathbf{x}^{(j)}). \quad (\mathbf{x} \in \mathbb{N}_m) \tag{5.4}$$

Here we have split the $m \times 1$ vector $\mathbf{x}$ into k sub-vectors so that $\mathbf{x} = (\mathbf{x}^{(1)\prime}, \ldots, \mathbf{x}^{(k)\prime})'$ where $\mathbf{x}^{(j)}$ is m_j-dimensional.

https://doi.org/10.2143/AST.32.1.1014 Published online by Cambridge University Press

5B. To allow for approximations to distributions, it is natural to extend the definitions given in the previous subsection so that we obtain compound functions where we allow for the counting distributions and severity distributions to be replaced with functions that are not necessarily distributions themselves. In the following, we shall allow for the counting distributions to be replaced with more general functions, but keep the severity distributions as distributions.

Sundt (2000*b*) showed that if $p, q \in \mathcal{F}_1$ and $h \in \mathcal{P}_{m+}$ in the first setting in the previous subsection, then we have

$$\varepsilon(p \vee h, q \vee h) \leq \varepsilon(p, q); \tag{5.5}$$

in the univariate case ($m = 1$), this result was shown in Dhaene and Sundt (1997).

The inequality (5.5) says that if we keep the severity distribution fixed, but approximate the counting distribution, then the distance between the aggregate claims distribution and its approximation is bounded by the distance between the counting distribution and its approximation. It is natural to believe that this result extends to the generalised setting of (5.4). In the next subsection, we shall prove this conjecture in an even more general setting.

The reason for the constraint $h(\mathbf{0}) = 0$ for (5.5) is to ensure the existence of $(p \vee h)(\mathbf{x})$ in (5.2), considering the infinite summation area; under the constraint, we have $h^{n*}(\mathbf{x}) = 0$ when $n > x$. so that we get

$$(p \vee h)(\mathbf{x}) = \sum_{n=0}^{x.} p(n) h^{n*}(\mathbf{x}), \quad (\mathbf{x} \in \mathbb{N}_m)$$

where the summation area is finite. For proper distributions, the constraint does not represent a real restriction as if it is not satisfied, then we can redefine the severity distribution and counting distribution by

$$\tilde{h}(\mathbf{x}) = \frac{h(\mathbf{x})}{1 - h(\mathbf{0})} \quad (\mathbf{x} \in \mathbb{N}_{m+})$$

$$\tilde{p}(n) = p(n)(1 - h(\mathbf{0}))^n. \quad (n \in \mathbb{N}_1)$$

For the more general (5.4), we can reason analogously. When assuming that $h_j \in \mathcal{P}_{m_j}$ for $j = 1, \ldots, k$, we obtain

$$(p \vee \mathbf{h})(\mathbf{x}) = \sum_{i=1}^{k} \sum_{n_i=0}^{x.^{(i)}} p(\mathbf{n}) \prod_{j=1}^{k} h_j^{n_j *}(\mathbf{x}^{(j)}), \quad (\mathbf{x} \in \mathbb{N}_m)$$

and if the assumption is not satisfied, then we can redefine the severity distributions and the counting distribution by

$$\tilde{h}_j(\mathbf{x}^{(j)}) = \frac{h_j(\mathbf{x}^{(j)})}{1 - h_j(\mathbf{0})} \quad (\mathbf{x}^{(j)} \in \mathbb{N}_{m_j+};\ j = 1, \ldots, k)$$

$$\tilde{p}(\mathbf{n}) = p(\mathbf{n}) \prod_{j=1}^{k} (1 - h_j(\mathbf{0}))^{n_j}. \quad (\mathbf{n} \in \mathbb{N}_k)$$

https://doi.org/10.2143/AST.32.1.1014 Published online by Cambridge University Press

5C. A key issue for compound distributions as discussed above, is that the severities (or severity vectors) for a fixed counting variable are mutually independent and identically distributed and independent of the counting variables. In an insurance context, this assumption excludes situations like when, say, a high number of claims could indicate that the amounts of the individual claims would tend to be low, or when the claims are positively correlated so that if some of the claims are large, then the others would also tend to be large.

Furthermore, in the settings where we assumed that a claim event generated a vector of claims from the individual policies in the portfolio, we considered the number of policies to be a fixed non-random number. It is tempting to allow for a random number of policies that could depend on the number of claim events and the severities.

We shall now extend our model to allow for such effects.

Let $h = \{h_{\mathbf{n}}\}_{\mathbf{n} \in \mathbb{N}_k}$, where $h_{\mathbf{n}}$ is the joint probability function of a non-negative integer-valued random variable and a random vector of dimension equal to this random variable and with non-negative integer-valued elements. For $p \in \mathcal{F}_k$ we define

$$(p \vee h)(m, \mathbf{x}) = \sum_{\mathbf{n} \in \mathbb{N}_k} p(\mathbf{n}) h_{\mathbf{n}}(m, \mathbf{x}), \quad (m \in \mathbb{N}_1; \mathbf{x} \in \mathbb{N}_m) \tag{5.6}$$

assuming that this sum exists. Analogous to what we did at the end of subsection 5B, we can achieve this by assuming that $h_{\mathbf{n}}(m, \mathbf{0}) = 0$ for all $n \in \mathbb{N}_k$ and $m \in \mathbb{N}_1$.

If the dimension of the severity vector is a fixed non-random number $m \geq k$, then we drop this dimension as argument in $h_{\mathbf{n}}$. In this case, we have $h_{\mathbf{n}} \in \mathcal{P}_m$. The case represented by (5.4) is obtained by letting

$$h_{\mathbf{n}}(\mathbf{x}) = \prod_{j=1}^{k} h_j^{n_j *}(\mathbf{x}^{(j)}). \quad (\mathbf{x} \in \mathbb{N}_m) \tag{5.7}$$

To be able to compare two compound functions $p \vee h$ and $q \vee h$ in the general setting with a random number of policies, we have to generalise the distance measure ε to functions of vectors of varying dimension. With a natural extension of the definition (3.1), we let

$$\varepsilon(p \vee h, q \vee h) = \sum_{m \in \mathbb{N}_1} \sum_{\mathbf{x} \in \mathbb{N}_m} |(p \vee h)(m, \mathbf{x}) - (q \vee h)(m, \mathbf{x})|.$$

We are now ready to generalise the bound (5.5).

Theorem 5.1. *For $p, q \in \mathcal{F}_k$ and $h = \{h_{\mathbf{n}}\}_{\mathbf{n} \in \mathbb{N}_k}$ defined as above, we have*

$$\varepsilon(p \vee h, q \vee h) \leq \varepsilon(p, q).$$

https://doi.org/10.2143/AST.32.1.1014 Published online by Cambridge University Press

Proof. We have

$$\varepsilon(p \vee h, q \vee h) = \sum_{m \in \mathbb{N}_1} \sum_{\mathbf{x} \in \mathbb{N}_m} |(p \vee h)(m, \mathbf{x}) - (q \vee h)(m, \mathbf{x})| =$$

$$\sum_{m \in \mathbb{N}_1} \sum_{\mathbf{x} \in \mathbb{N}_m} \left| \sum_{\mathbf{n} \in \mathbb{N}_k} (p(\mathbf{n}) - q(\mathbf{n}))\, h_{\mathbf{n}}(m, \mathbf{x}) \right| \leq$$

$$\sum_{m \in \mathbb{N}_1} \sum_{\mathbf{x} \in \mathbb{N}_m} \sum_{\mathbf{n} \in \mathbb{N}_k} |(p(\mathbf{n}) - q(\mathbf{n}))|\, h_{\mathbf{n}}(m, \mathbf{x}) =$$

$$\sum_{\mathbf{n} \in \mathbb{N}_k} |p(\mathbf{n}) - q(\mathbf{n})| \sum_{m \in \mathbb{N}_1} \sum_{\mathbf{x} \in \mathbb{N}_m} h_{\mathbf{n}}(m, \mathbf{x}) = \sum_{\mathbf{n} \in \mathbb{N}_k} |p(\mathbf{n}) - q(\mathbf{n})| = \varepsilon(p, q),$$

which proves the theorem. □

6. Multivariate Bernoulli distributions

6A. In this section, we shall apply the results of the earlier sections to multivariate Bernoulli distributions.

A distribution $p \in \mathcal{P}_m$ is called an m-variate Bernoulli distribution if $p(\mathbf{x}) = 0$ unless all the elements of $\mathbf{x}$ are equal to zero or one. In this case, Corollary 3.3 gives that

$$\varepsilon(p, q) = 2 \sum_{\mathbf{x} \in \{0,1\}^m} (p(\mathbf{x}) - q(\mathbf{x}))_+ \tag{6.1}$$

for any distribution $q \in \mathcal{P}_m$.

Multivariate Bernoulli distributions can give fairly transparent illustrations to the theory. Furthermore, the theory for error bounds discussed in the present paper consists of several building blocks that can be applied together in various ways, and we believe that the results on multivariate Bernoulli distributions may also be useful in connection with compound distributions and convolutions.

6B. Let $p \in \mathcal{P}_m$ be a multivariate Bernoulli distribution with

$$p(\mathbf{n}) = \begin{cases} \pi_0 & (\mathbf{n} = \mathbf{0}) \\ \pi_j & (\mathbf{n} = \mathbf{e}_j;\ j = 1, \ldots, m) \\ 0 & (\text{otherwise}) \end{cases} \tag{6.2}$$

and $\sum_{j=0}^{m} \pi_j = 1$.

For $p, q \in \mathcal{P}_m$ with p given by (6.2), (6.1) gives

$$\varepsilon(p, q) = 2\left((\pi_0 - q(\mathbf{0}))_+ + \sum_{j=1}^{m} (\pi_j - q(\mathbf{e}_j))_+ \right). \tag{6.3}$$

https://doi.org/10.2143/AST.32.1.1014 Published online by Cambridge University Press

Theorem 6.1. *If $p, q \in \mathcal{P}_m$ with p given by* (6.2) *and such that p and q have the same marginal means, then*

$$\varepsilon(p,q) = 2\sum_{j=1}^{m}\left(\pi_j - q(\mathbf{e}_j)\right) = 2\left(1-\pi_0-\sum_{j=1}^{m} q(\mathbf{e}_j)\right). \tag{6.4}$$

Proof. For $j = 1, \ldots, m$, we have

$$\pi_j = \mu_j(p) = \mu_j(q) = \sum_{\mathbf{n}\in\mathbb{N}_{m+}} n_j\, q(\mathbf{n}) \geq q(\mathbf{e}_j).$$

On the other hand,

$$1-\pi_0 = \sum_{j=1}^{m}\pi_j = \sum_{j=1}^{m}\sum_{\mathbf{n}\in\mathbb{N}_{m+}} n_j\, q(\mathbf{n}) = \sum_{\mathbf{n}\in\mathbb{N}_{m+}} n.q(\mathbf{n}) \geq \sum_{\mathbf{n}\in\mathbb{N}_{m+}} q(\mathbf{n}) = 1-q(\mathbf{0}),$$

that is, $\pi_0 \leq q(\mathbf{0})$. Insertion in (6.3) gives (6.4). □

In the univariate case ($m = 1$), (6.4) gives

$$\varepsilon(p,q) = 2(\pi_1 - q(1)). \tag{6.5}$$

This result was proved by Dhaene and Sundt (1997).

6C. As discussed in Sundt (2000*a*,*b*), any distribution $p \in \mathcal{P}_{m0}$ can be expressed as a compound distribution $p = r \vee h$ of Case 1 of subsection 5A with $r \in \mathcal{P}_{10}$ being a (univariate) Bernoulli distribution given by

$$r(0) = p(\mathbf{0}) = 1 - r(1)$$

and $h \in \mathcal{P}_{m+}$ given by

$$h(\mathbf{n}) = \frac{p(\mathbf{n})}{r(1)}. \qquad (\mathbf{n}\in\mathbb{N}_{m+})$$

If we approximate p with a $q \in \mathcal{P}_m$ in the form $q = s \vee h$ with $s \in \mathcal{P}_1$, then (5.5) gives $\varepsilon(p, q) \leq \varepsilon(r, s)$.

A natural question is now, how sharp is the bound $\varepsilon(r, s)$? In the following, we shall consider this question in the special case when p is a multivariate Bernoulli distribution and s is a Poisson distribution with parameter equal to $\lambda = r(1)$, that is,

$$s(n) = \frac{\lambda^n}{n!}e^{-\lambda}. \qquad (n\in\mathbb{N}_1)$$

Then

$$q(\mathbf{0}) = s(0) = e^{-\lambda} \tag{6.6}$$

https://doi.org/10.2143/AST.32.1.1014 Published online by Cambridge University Press

and for $j = 1, \ldots, m$

$$q(\mathbf{e}_j) = s(1)h(\mathbf{e}_j) = e^{-\lambda} p(\mathbf{e}_j). \tag{6.7}$$

Furthermore, (6.5) gives

$$\varepsilon(r,s) = 2(r(1) - s(1)) = 2(\lambda - \lambda e^{-\lambda}),$$

that is,

$$\varepsilon(r,s) = 2\lambda(1 - e^{-\lambda}). \tag{6.8}$$

Let us first consider the case when p is given by (6.2) with $\pi_0 > 0$. Then, insertion of (6.7) in (6.4) gives

$$\varepsilon(p,q) = 2\left(1 - \pi_0 - e^{-\lambda}\sum_{j=1}^{m} p(\mathbf{e}_j)\right) = 2\left(1 - \pi_0 - e^{-\lambda}(1-\pi_0)\right) = 2\lambda(1-e^{-\lambda}) = \varepsilon(r,s)$$

Let us now turn to the case with $m = 2$. Then

$$q(1,1) = s(1)h(1,1) + 2s(2)h(1,0)h(0,1) = e^{-\lambda}(p(1,1) + p(1,0)p(0,1)),$$

and by inserting this together with (6.6) and (6.7) in (6.1) and utilising (6.8) we obtain after some calculus

$$\varepsilon(r,s) = \varepsilon(p,q) + 2\left(p(1,1)(1-e^{-\lambda}) - (p(1,1)(1-e^{-\lambda}) - p(1,0)p(0,1)e^{-\lambda})_+\right).$$

From this we see that $\varepsilon(r,s) = \varepsilon(p,q)$ if $p(0,1)$, $p(1,0)$, or $p(1,1)$ is equal to zero. The case $p(1,1) = 0$ is the bivariate case of (6.2). The case $p(0,1) = 0$ corresponds to a situation where an insurance policy can have two sorts of claims, and a claim of type one can occur only in connection with a claim of type two, e.g. if type one is a special case of type two.

ACKNOWLEDGEMENT

The authors are grateful to the referee for comments that led to improvements of the paper.

REFERENCES

AMBAGASPITIYA, R.S. (1999) On the distributions of two classes of correlated aggregate claims. *Insurance: Mathematics and Economics* **24**, 301-308.

DHAENE, J. and DE PRIL, N. (1994) On a class of approximative computation methods in the individual model. *Insurance: Mathematics and Economics* **14**, 181-196.

DHAENE, J. and SUNDT, B. (1997) On error bounds for approximations to aggregate claims distributions. *ASTIN Bulletin* **27**, 243-262.

DHAENE, J. and SUNDT, B. (1998) On approximating distributions by approximating their De Pril transforms. *Scandinavian Actuarial Journal*, 1-23.

SUNDT, B. (2000*a*). The multivariate De Pril transform. *Insurance: Mathematics and Economics* **27**, 123-136.

SUNDT, B. (2000*b*). On error bounds for approximations to multivariate distributions. *Insurance: Mathematics and Economics* **27**, 137-144.

SUNDT, B. (2000*c*). On multivariate Vernic recursions. *ASTIN Bulletin* **30**, 111-122.

BJØRN SUNDT
Vital Forsikring ASA
P.O.Box 250
N-1326 Lysaker
NORWAY

RALUCA VERNIC
Department of Mathematics and Informatics
'Ovidius' University of Constanta
124, Bd. Mamaia
8700 Constanta
ROMANIA

https://doi.org/10.2143/AST.32.1.1014 Published online by Cambridge University Press