EUROPEAN JOURNAL OF PURE AND APPLIED MATHEMATICS
Vol. 3, No. 5, 2010, 779-785
ISSN 1307-5543 – www.ejpam.com

Convex Ordering of Random Variables and its Applications in Econometrics and Actuarial Science

Arjun K. Gupta*, Mohammad A. S. Aziz

Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, Ohio 43403, U.S.A

Abstract. It is well known that in economics and finance, the data usually have "fat tail" and in this case the Normal distribution is not a good model to use. The skew normal distributions recently draw considerable attention as an alternative model. Unfortunately, the distribution of the sum of log-skew normal random variables does not have a closed form. In this work, we discuss the use of lower convex order of random variables to approximate this distribution. Further, two application of this approximate distribution are given : first to describe the final wealth of a series of payments, and second to describe the present value of a series of payments.

2000 Mathematics Subject Classifications: Primary 62E17; Secondary 62P05

Key Words and Phrases: Log-skew normal random variable, lower convex order bound, comonotonocity

---

1. Preliminaries

Definition 1. *Consider two random variables X and Y such that $E[\phi(X)] \leq E[\phi(Y)]$, for all convex functions ϕ, provided expectation exist.Then X is said to be smaller than Y in the convex order denoted as $X \leq_{cx} Y$.*

Definition 2 (Convex order definition using stop-loss premium)**.** *Consider two random variables X and Y. Then X is said to precede Y in convex order sense if and only if*

$$E[X] = E[Y]$$

$$E[(X-d)_+] \leq E[(Y-d)_+], I_{(-\propto,\propto)}(d)$$

where

$$(X-d)_+ = max(X-d,0)$$

*Corresponding author.

Email address: gupta@bgsu.edu (Arjun K. Gupta)

An equivalent definition can be derived from the following relation

$$E[(X-d)_+]-E[(d-X)_+]=E(X)-d$$

For the random variable Y the same relation is given by,

$$E[(Y-d)_+]-E[(d-Y)_+]=E(Y)-d$$

Now assume $X \le_{cx} Y$, which implies that

$$E[X]=E[Y]$$

and

$$E[(X-d)_+]\le E[(Y-d)_+], I_{(-\propto,\propto)}(d)$$

Hence

$$E[(d-X)_+]\le E[(d-Y)_+]$$

Therefore, a definition equivalent to the definition here is

$$E[X]=E[Y]$$

$$E[(d-X)_+]\le E[(d-Y)_+]$$

1.1. Properties of Convex Ordering of Random Variables

1. If X precedes Y in convex order sense i.e if $X \le_{cx} Y$, then $E[X]=E[Y]$ and $Var[X]\le Var[Y]$
2. If $X \le_{cx} Y$ and Z is independent of X and Y then $X+Z \le_{cx} Y+Z$
3. Let X and Y be two random variables, then $X \le_{cx} Y \Leftrightarrow -X \le_{cx} -Y$
4. Let X and Y be two random variables such that $E[X]=E[Y]$. Then $X \le_{cx} Y$ if and only if $E|X-a| \le_{cx} E|Y-a|, \forall$ a $\in \Re$
5. The convex order is closed under mixtures: Let X, Y and Θ be random variables such that $[X|\Theta=\theta] \le_{cx} [Y|\Theta=\theta] \forall\theta$ in the support of Θ. Then $X \le_{cx} Y$.
6. The convex order is closed under convolution: Let $X_1,X_2,\ldots,X_m$ be a set of independent random variables and $Y_1,Y_2,\ldots,Y_n$ be another set of independent random variables. If $X_i \le_{cx} Y_i$, for $i=1,....,m$, then $\sum_{j=1}^{m} X_j \le_{cx} \sum_{j=1}^{m} Y_j$
7. Let X be a random variable with finite mean. Then $X+E[X] \le_{cx} 2X$
8. Let $X_1,X_2,\ldots,X_m$ and Y be (n+1) random variables. If $X_i \le_{cx} Y$, $i=1,....,n$, then $\sum_{i=1}^{n} a_i X_i \le_{cx} Y$, whenever $a_i \ge 0, i=1,\ldots,n$ and $\sum_{i=1}^{n} a_i=1$

9. Let X and Y be independent random variables. Then $X_i \leq_{cx} Y_i$ if and only if $E[\phi(X,Y)] \leq E[\phi(Y,X)] \forall \phi \in \wp_{cx}$, where

$$\wp_{cx} = \{\phi : \Re^2 \longrightarrow \Re : \phi(X,Y) - \phi(Y,X) \text{ is convex for all } x \in y\}.$$

10. Let X_1 and X_2 be a pair of independent random variables and let Y_1 and Y_2 be another pair of independent random variables. If $X_i \leq_{cx} Y_i$, $i = 1,2$ then $X_1 X_2 \leq_{cx} Y_1 Y_2$.

2. Main Result of Convex Ordering

Theorem 1. *For any random vector* $\mathbf{X} = (X_1, X_2, ...,X_m)$ *and any random variable* Λ*, which is assumed to be a function of X, we have,*

$$\sum_{i=1}^{n} E[X_i|\Lambda] \leq_{cx} \sum_{i=1}^{n} X_i$$

Proof. From the definition 1 we have, $X \leq_{cx} Y$ if and only if $E[\phi(X)] \leq E[\phi(Y)]$. In accordance with this definition we need to show that

$$E_\Lambda[\phi(\sum_{i=1}^{n} E[X_i|\Lambda])] \leq_{cx} E[\phi(\sum_{i=1}^{n} X_i)]$$

Now,

$$E[\phi(\sum_{i=1}^{n} X_i)] = E_\Lambda E[\phi(\sum_{i=1}^{n} X_i)|\Lambda] \geq E_\Lambda[\phi(E(\sum_{i=1}^{n} X_i|\Lambda))] = E_\Lambda[\phi(\sum_{i=1}^{n} E[X_i|\Lambda])]$$

The last inequality was obtained by Jensen's inequality, which states that for any convex function ϕ,

$$\phi(E(X)) \leq E(\phi(X))$$

Therefore,

$$E_\Lambda[\phi(\sum_{i=1}^{n} E[X_i|\Lambda])] \leq E[\phi(\sum_{i=1}^{n} X_i)]$$

Hence,

$$\sum_{i=1}^{n} E[X_i|\Lambda] \leq_{cx} \sum_{i=1}^{n} X_i$$

which completes the proof.

2.1. Lower Bound Approximations of the Distribution Sum of Random Variables with Convex Ordering

In this section we will describe two examples [follow from 1] that show how distribution function of the sum of random variables can be approximated by convex order of random variable.

Example 1 (Approximation of distribution sum of two independent standard normal random variables)**.** *Suppose X and Y be independent $N(0,1)$ random variables. We want to derive lower bounds for $S = X + Y$. In this case we know the exact distribution of S, i.e $S \sim N(0,2)$. Let us see how lower bound approximation works in this case.*

Let $Z = X + aY$ for some real a . Then $Z \sim N(0,1+a^2)$. The conditional distribution of $S|Z$ is

$$N[\mu_S + \frac{\rho_{S,Z}\sigma_S}{\sigma_Z}(Z-\mu_Z), \sigma_S^2(1-\rho_{S,Z}^2)]$$

Here $Cov(X+Y,X+aY) = 1.1cov(X,X)+1.acov(Y,Y) = 1+a$ and $\rho_{S,Z} = \frac{1+a}{\sqrt{2}\sqrt{1+a^2}}$. Therefore,

$$S|Z \sim N[Z\frac{1+a}{1+a^2}, \frac{(1-a)^2}{1+a^2}]$$

Hence

$$E[S|Z] = \frac{1+a}{1+a^2}Z \sim N[0, \frac{(1+a)^2}{1+a^2}]$$

Therefore, for some choices of a, we get the following distribution for the lower bound for S:

$$a = 0 \quad gives \quad N(0,1) \leq_{cx} S = X + Y \sim N(0,2)$$

$$a = 1 \quad gives \quad N(0,2) \leq_{cx} S = X + Y \sim N(0,2)$$

$$a = -1 \quad gives \quad N(0,0) \leq_{cx} S = X + Y \sim N(0,2)$$

Thus in this case best lower bound is obtained for $a = 1$ which is the exact distribution. The variance of the lower bound can be seen to have a maximum at $a = 1$ and a minimum at $a = -1$.

Example 2 (Approximation of distribution sum of two lognormal random variables)**.** *Suppose Y_1 and Y_2 be independent $N(0,1)$. Define $X_1 = e^{Y_1} \Rightarrow X_1 \sim lognormal(0,1)$ and $X_2 = e^{Y_1+Y_2} \Rightarrow X_2 \sim lognormal(0,2)$. We want to find the lower bound for the distribution of $S = X_1 + X_2$. Let $Z = Y_1 + Y_2$. As shown in example 1, the conditional distribution of Y_1 given Z is,*

$$Y_1|(Y_1+Y_2) = Z \sim N(\frac{1}{2}Z, \frac{1}{2})$$

Therefore, $E[X_1 = e^{Y_1}|(Y_1+Y_2) = z] = M_Y(1, \frac{1}{2}Z, \frac{1}{2})$, where $Y \sim N(\mu,\sigma^2)$

$$= exp(\frac{1}{2}Z + \frac{1}{4})$$

We also observe that $E[X_2|(Y_1+Y_2)=z]=e^Z$.Therefore the lower bound for approximating the distribution of $S=X_1+X_2$ is $S_l=E[(X_1+X_2)|Z]=exp(\frac{1}{2}Z+\frac{1}{4})$. It can be easily verified that $E(S_l)=E[exp(\frac{1}{2}Z+\frac{1}{4})]=e^{\frac{1}{2}}+e$ and $E((S_l)^2)=e^{\frac{3}{2}}+2e^{\frac{5}{2}}+e^4$. Thus the variance of the lower bound is 64.374 and is close to the variance of $S=67.281$. The idea is to obtain lower convex bound in such a way that the variance of the lower bound gets as close as possible to the variance of the sum. With this view in mind considering more general form of the conditioning variable as $Z=Y_1+aY_2$, it could be shown that optimal lower bound is reached for $a=1.27$ and the variance of S_l in this case is 66.082. Thus choosing the conditioning variable is crucial in determining the lower convex order bound.

3. Application of Convex Order of Random Variables in Finance and Economics

Let $\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_{n-1}$ be non-negative real numbers. Let $\mathbf{Y}=(Y_1, Y_2, \ldots, Y_n)^T$ be a multivariate skew normal random vector with specified mean vector and variance-covariance matrix and satisfying additive properties. Define, $Z_i=\sum_{k=i+1}^{n} Y_k, i=0,1,\ldots,n-1$, that is, Z_i's are linear combinations of the components $(Y_1, Y_2, \ldots, Y_n)$. With the components so defined, consider the sum

$$S=\sum_{i=0}^{n-1}\alpha_i e^{Z_i}=\sum_{i=0}^{n-1}\alpha_i e^{Y_{i+1}+,\ldots,+Y_n} \tag{1}$$

From economic or actuarial point of view, the sum S could be interpreted as the final wealth or terminal wealth or the accumulated value of a series of deterministic saving amounts or alternatively the accumulated value of a series of payments. In this situation, α_i $(i=0,\ldots,n-1)$ represents yearly saving in period i or amount invested in period i, Y_{i+1} refers to the random rate of return in period i for $i=0,\ldots,n-1$. The term $Y_k=log\frac{P_k}{P_{k-1}}=logP_k-logP_{k-1}$ i.e $e^{Y_k}=\frac{P_k}{P_{k-1}}$, where P_k is the price of the asset at period k , for $k=0,\ldots,n$; is called the random log-return in period k and Z_i denote the sum of stochastic or random returns in period i, $i=0,\ldots,n-1$. With suitable adjustment, S could also be referred as the present value of a series of payments. To be more precise, if $-Z_i$ denotes the stochastic log-return over the period $[0,i]$, then e^{Z_i},represents the stochastic discount factor over the period [0, i]. In this situation, the sum S is the present value of α_i [4].

The sum defined in (1) plays a central role in the actuarial and financial theory because it allows computation of risk measures such as value at risk or stop-loss premium. To calculate the risk measures we need to evaluate the distribution function of S. Unfortunately, the distribution of the sum S (of log-normally or log-skew normally distributed random variables) is not available in the closed-form. It is possible to use Monte Carlo simulation method to approximate the distribution function. However, Monte Carlo simulation of the distribution is often time-consuming. Thus one has to find alternate way to approximate the distribution of the sum. Among the proposed solutions, moment matching methods, lognormal and inverse gamma approximations are commonly used. Both methods approximate the unknown distribution function by a given one such that the first two moments coincide.

Kaas, et al. [2] and Dhaene, et al. [1] propose to approximate the distribution function of S by so called “convex lower bound”. The underlying idea of convex lower order bound is to replace an unknown or too complex distribution (for which no explicit form is found) by another one which is easier to determine. In this approach, the real distribution is known to be bounded in terms of convex ordering to the approximated distribution. To be more precise, by Theorem 1, the distribution function of $S = \sum_{i=0}^{n-1} \alpha_i e^{Z_i}$ is approximated by the distribution function of S_l, where S_l is defined by,

$$S_l = \sum_{i=0}^{n-1} \alpha_i E(e^{Z_i}|\Lambda) \tag{2}$$

An appropriate choice of the conditioning random variable Λ is required. This approach has two-fold advantages. Firstly, use of this approach transforms the multidimensionality problem caused by $(Z_0, Z_2, \dots, Z_{n-1})$ to a single dimension caused by Λ. Secondly, an appropriate choice of Λ (that makes the expectation in (2) non-decreasing or non-increasing function of the conditioning random variable Λ) will make a comonotonic sum i.e the elements of the sum in (2) posses the so called comonotonic dependence structure. Using additivity properties of sum of comonotonic random variables risk measures related to the distribution function of S is then approximated by the corresponding risk measures of S_l. According to [2], comonotonic *upper* bound for the sum in convex order sense can also be derived using the result

$$\sum_{i=0}^{n-1} X_i \leq_{cx} \sum_{i=0}^{n-1} F_{X_i}(U),$$

where U is the uniform random variable over $(0, 1)$. However, comonotonic upper bounds generally provide too conservative estimates of the cumulative distribution function [3]. Thus we only discussed convex *lower* bound here.

Remark 1. *In general, the random vector* $(E(X_0|\Lambda), E(X_1|\Lambda), \dots, E(X_{n-1}|\Lambda))$ *does not have the same marginal distribution as* $(X_0, X_1, \dots, X_{n-1})$*. However, if the conditioning random variable* Λ *is chosen in such a way that all random variables* $E(X_i|\Lambda)$*,* $(i = 0, 1, 2, \dots, n-1)$ *are non-decreasing functions of* Λ *(or non-increasing functions of* Λ*), then the sum* $\sum_{i=0}^{n-1} E[X_i|\Lambda]$ *is a sum of n comonotonous random variables and can be referred to as comonotonic lower bound. Hence risk measures for the sum could easily be obtained by summing the corresponding risk measures for the marginals involved.*

References

[1] J Dhaene, M Denuit, M Gooverts, R Kass, and D Vyncke. The concept of comonotonicity in actuarial science and finance: Applications. *Insurance: Mathematics and Economics*, 31:133–161, 2002.

[2] J Dhaene R Kaas, M Goovaerts and M Denuit. *Modern Actuarial Risk Theory*. Kluwer Academic Publishers, Dordrecht, 2001.

[3] O Roch and E Valdez. Lower convex order bound approximations for sums of log-skew normal random variables. Working Paper (2008).

[4] S Vanduffel, X Chen, J Dhaene, M Goovaerts, L Henrard, and R Kass. Optimal approximations for risk measures of sums of lognormals based on conditional expectations. *Journal of Computational and Applied Mathematics*, 221:202–218, 2008.