A NOTE ON NONPARAMETRIC ESTIMATION OF THE CTE

BY

Bangwon Ko, Ralph P. Russo and Nariankadu D. Shyamalkumar

Abstract

The α-level Conditional Tail Expectation (CTE) of a continuous random variable X is defined as its conditional expectation given the event $\{X > q_\alpha\}$ where q_α represents its α-level quantile. It is well known that the empirical CTE (the average of the $n(1-\alpha)$ largest order statistics in a sample of size n) is a negatively biased estimator of the CTE. This bias vanishes as the sample size increases, but in small samples can be significant. In this article it is shown that an unbiased nonparametric estimator of the CTE does not exist. In addition, the asymptotic behavior of the bias of the empirical CTE is studied, and a closed form expression for its first order term is derived. This expression facilitates the study of the behavior of the empirical CTE with respect to the underlying distribution, and suggests an alternative (to the bootstrap) approach to bias correction. The performance of the resulting estimator is assessed via simulation.

1. Introduction

Since the publication of Artzner et al. (1997, 1999), the Conditional Tail Expectation (CTE) has gained an increasing level of attention as a measure of risk, especially among academic actuaries (see Hardy (2003) and the references therein, and also Landsman and Valdez (2003), Bilodeau (2004), Cai and Li (2005) and Dhaene et al. (2003)). The main reason for preferring the CTE over the Value at Risk (the VaR being the most popular measure of risk in the financial world) is that the former is *coherent* while the latter is not (see Delbaen (2000)). That the CTE is now part of the insurance industry regulations in both Canada (for segregated fund contracts) and the US (C3 Phase 2 Report, American Academy of Actuaries) has increased its relevance as a risk measure among practicing actuaries.

The literature on the statistical estimation of the CTE includes Jones and Zitikis (2003), Manistre and Hancock (2005), and Kim and Hardy (2007). The main motivation for this article derives from the fact that the empirical CTE is biased as an estimator of the CTE. This bias was observed in Manistre and Hancock (2005), and was shown in Kim and Hardy (2007) to almost always exist. The study of the bias of the empirical CTE is of practical interest because

Astin Bulletin 39(2), 717-734. doi: 10.2143/AST.39.2.2044655

the relative bias (bias as a percentage of the CTE) has been observed to be at a significant level for distributions of actuarial interest and for sample sizes in the range of a few hundreds. As observed in Kim and Hardy (2007), practicing actuaries adopting a seriatim approach to simulations may often have to settle for sample sizes in this range due to heavy computational costs in terms of both time and money. Hence, it is of interest to understand the bias of the empirical CTE and to explore methods to correct for it.

We show that under general conditions a nonparametric unbiased estimator for the CTE does not exist (Theorem 1), and hence there is no perfect bias correction that can be applied to the empirical CTE. Nonetheless, as observed in Manistre and Hancock (2005) and Kim and Hardy (2007), the bias of the empirical CTE vanishes as the sample size increases. Corollary 2 implies that when sampling from a continuous distribution (under some non-restrictive conditions) the empirical CTE is an asymptotically unbiased estimator of the CTE. Moreover, Corollary 2 establishes that the bias converges to zero at the rate $1/n$, and provides a closed form expression for its first order term. This confirms the empirical observation in Manistre and Hancock (2005) and Kim and Hardy (2007) that the bias converges to zero faster than the standard error; as the standard error vanishes at the rate $1/\sqrt{n}$. It is worth mention that for discrete distributions the bias may converge at a rate slower than, or in some cases faster than, $1/n$ (see Russo and Shyamalkumar (2008)). In the course of proving Corollary 2, we derive an expression for the bias of the empirical CTE (Lemma 1) that yields a simple alternative proof of the known result that the CTE is negatively biased (Corollary 1).

The derivation of a closed form expression for the first order term of the bias allows us to understand its behavior with respect to the α level and the tail heaviness of the underlying distribution. In particular, we show that, rather unexpectedly, the relative bias is not monotone with respect to tail heaviness. Bias correction via the Bootstrap was proposed in Kim and Hardy (2007). Here we show that correcting for the first order bias term by using a plug-in estimator involving a *default* density estimator on the R computing platform (see R Development Core Team (2008)) works very well in that it reduces bias significantly while keeping the mean square error close to that of the empirical CTE. Another appeal of this approach is that it does not entail any additional computing as the asymptotic standard error of the empirical VaR (and in fact even the asymptotic covariance between the empirical VaR and empirical CTE, see Manistre and Hancock (2005)) is also a function of the density evaluated at the appropriate quantile.

1.1. Definitions and Notation

Let $X, X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables with a common loss distribution $F(\cdot)$. Let $S(\cdot) = 1 - F(\cdot)$ denote the survival function corresponding to $F(\cdot)$, and $f(\cdot)$ and $\mu(\cdot)$ the density and hazard rate functions if they exist. Also, let $Q(\cdot) = F^{-1}(\cdot)$ denote the quantile function associated with F, or more precisely:

$$Q(u) := \inf\{x \in \mathbb{R} : F(x) \geq u\}, \quad u \in (0,1). \tag{1}$$

For $\alpha \in (0,1)$, $q_\alpha := Q(\alpha)$ denotes the α-quantile. The order statistics of the sample are represented by $X_{1:n}, X_{2:n}, \ldots, X_{n:n}$, with $X_{1:n}$ and $X_{n:n}$ being the sample minimum and maximum, respectively. Also, let $U_{1:n}, \ldots, U_{n:n}$ represent the order statistics from a random sample of size n from a uniform distribution on (0,1). It will be convenient occasionally to assume that $U_{i:n} = Q(X_{i:n})$ for $i = 1, \ldots, n$.

The α-level CTE of X, denoted either by $\eta_\alpha(X)$ or $\eta_\alpha(F)$, is defined as

$$\eta_\alpha(X) := \mathbb{E}(X \mid X > q_\alpha).$$

Our focus is on the α-level empirical CTE defined as

$$\mathrm{CTE}_n^\alpha := \left(\frac{1}{k_n}\right) \sum_{i=n-k_n+1}^{n} X_{i:n} \tag{2}$$

where k_n is equal to $n(1-\alpha)$ if integer valued, else can be defined as either the integer immediately below or immediately above $n(1-\alpha)$. The above quantity is called the empirical CTE as for $n\alpha$ an integer we have,

$$\mathrm{CTE}_n^\alpha = \eta_\alpha(Y)$$

where Y is a random variable having the empirical CDF as its distribution function, i.e.

$$\Pr(Y \leq y) := \frac{1}{n} \sum_{i=1}^{n} \mathrm{I}_{(-\infty, y]}(X_i).$$

For real numbers x and y we denote by $x \wedge y$ (resp., $x \vee y$) the minimum (resp., maximum) of x and y. The sign function $\mathrm{sgn}(\cdot)$ is defined such that $\mathrm{sgn}(x)$ equals 1 (resp., –1) when x is non-negative (resp., negative). By $U(0,\beta)$ we denote the continuous uniform distribution on the interval $(0,\beta)$, for $\beta > 0$. A class $\mathcal{F}$ of distributions is said to be convex if for $F, G \in \mathcal{F}$ and $\beta \in [0,1]$ we have $\beta F + (1-\beta)G \in \mathcal{F}$.

2. Non-Existence of an Unbiased Non-Parametric Estimator

Recall that an estimator T is unbiased for some function $\theta(\cdot)$ defined on a given class of distributions $\mathcal{F}$, if

$$\mathbb{E}_F(T) = \theta(F), \quad \forall F \in \mathcal{F}.$$

Hence, results on the non-existence of an unbiased estimator need to specify the function $\theta(\cdot)$ and also the class $\mathcal{F}$. In this section, we show (see Theorem 1) that there does not exist an unbiased estimator of the α-level CTE, $\alpha \in (0,1)$, when the class of distributions $\mathcal{F}$ is taken to be all distributions for which that CTE is finite. More concisely, we establish the non-existence of an unbiased *non-parametric* estimator of the α-level CTE, for $\alpha \in (0,1)$. The qualifier *non-parametric* refers to the fact that the set of all distributions with a finite α-level CTE is not parametrizable with a finite dimensional parameter. Theorem 1 is akin to a result for density estimation in Rosenblatt (1956).

While it is well known that there does not exist a non-parametric unbiased estimator of the α-level VaR, for $\alpha \in (0,1)$, it is not clear that this can be used to easily conclude the non-existence of a non-parametric unbiased estimator of the α-level CTE. For example, in the case of a sample of size one from a Bernoulli(p) distribution, p^2 cannot be unbiasedly estimated but p (which is a one-to-one function of p^2) can be unbiasedly estimated.

The proof of the next theorem relies on Theorem 2.1 of Bickel and Lehmann (1969), a powerful tool for establishing the non-existence of unbiased estimators of a quantity of interest when the unknown distribution could be any member of a given convex class (closed under finite mixtures) of distributions. To establish the non-existence of an unbiased estimator of $\theta(\cdot)$ on $\mathcal{F}$ we identify a suitable convex sub-set $\mathcal{G}$ of $\mathcal{F}$ so that Bickel and Lehmann's result can be employed to establish the non-existence of an unbiased estimator of $\theta(\cdot)$ on $\mathcal{G}$. This suffices as the non-existence of an unbiased estimator for $\theta(\cdot)$ restricted to the subset $\mathcal{G}$ of $\mathcal{F}$ implies the non-existence of an unbiased estimator for $\theta(\cdot)$ on the whole of $\mathcal{F}$.

Before stating the theorem, it is instructive to rule out the empirical CTE as a possible candidate for an unbiased estimator of the CTE. We do this by showing that it is a biased estimator of the CTE for $U(0,1)$, and hence for any class of distributions $\mathcal{F}$ containing $U(0,1)$. The uniform distribution was also considered in Example 2.2 of Kim and Hardy (2007).

Example 1. [Uniform Distribution]
Suppose that X is uniformly distributed on $(0,1)$. It is easily checked that $\mathbb{E}(X_{k:n}) = k/(n+1)$ for $k = 1, \ldots, n$. Thus, when $n\alpha$ is an integer,

$$\mathbb{E}(\mathrm{CTE}_n^\alpha) = \frac{1}{2} + \frac{n\alpha}{2\,(n+1)}.$$

Since $\eta_\alpha(X) = (1+\alpha)/2$, it follows that the empirical CTE is downwardly biased by $\alpha/2(n+1)$. □

Example 1 while ruling out the empirical CTE, falls short of establishing the non-existence of a nonparametric unbiased estimator for the CTE.

Theorem 1. *For* $0 < \alpha < 1$ *there does not exist a finite sample unbiased estimator for the* α*-level CTE when the unknown distribution is allowed to be any member of the class of all distributions that possess a finite* α*-level CTE.*

PROOF OF THEOREM 1. Let $\alpha \in (0,1)$, and let $\mathcal{F}_0$ be the set of distributions with a finite α-level CTE. Let F_0 (resp., G_0) be the uniform distribution on (0,1) (resp., (1,2)), and let $\mathcal{G}$ be the convex hull of F_0 and G_0, *i.e.* $\mathcal{G}_0 := \{H \mid H = \beta F_0 + (1-\beta) G_0, \beta \in [0,1]\}$. As any $H \in \mathcal{G}_0$ is supported on (0,2), it has a finite α-level CTE, and this implies that $\mathcal{G}_0 \subseteqq \mathcal{F}_0$. Hence, it suffices to show that there does not exist a finite sample unbiased estimator for the α-level CTE when the unknown distribution is allowed to be any distribution in $\mathcal{G}_0$.

By definition $\mathcal{G}_0$ is convex, and hence by Theorem 2.1 of Bickel and Lehmann (1969) it suffices to show that the α-level CTE of the distribution $\beta F_0 + (1-\beta) G_0$ is not a polynomial in β, for $0 \le \beta \le 1$. It is easy to show that

$$\eta_\alpha(\beta F_0 + (1-\beta) G_0) = \begin{cases} \dfrac{3}{2} + \dfrac{\alpha - \beta}{2(1-\beta)} & \beta \le \alpha; \\[2ex] \dfrac{3}{2} - \dfrac{(\beta - \alpha)(2\beta - \alpha)}{2(1-\alpha)\beta} & \beta > \alpha; \end{cases}$$

Observing that the above is not a polynomial in β completes the proof. □

We note that the above theorem can be strengthened to establish the non-existence of an unbiased estimate for the CTE for classes of distributions much smaller than $\mathcal{F}_0$. In fact, from the proof it follows that any class of distributions containing $\mathcal{G}_0$ is large enough to rule out unbiased estimators for the CTE. One such class of interest is the class of all unimodal distributions with a finite α-level CTE, as the class of unimodal distribution on $(0,\infty)$ is the same as the class of mixtures of continuous uniform distributions. We note that while the class $\{U(0,\beta), \beta > 0\}$ contains both F_0 and G_0, it does not contain $\mathcal{G}_0$. In fact, if the unknown distribution is one of $\{U(0,\beta) : \beta > 0\}$ then it is easily seen from Example 1 that

$$\mathrm{CTE}_n^\alpha + \left(\frac{\alpha}{2n}\right) X_{n:n}$$

is an unbiased estimator for the α-level CTE.

The goal of the next section is to study the asymptotic behavior of the bias, $\mathbb{E}(\mathrm{CTE}_n^\alpha) - \eta_\alpha(X)$, for a continuous distribution $F(\cdot)$.

3. BOUNDS FOR THE BIAS OF THE EMPIRICAL CTE

In the remainder of this article we assume that F satisfies

1) F is differentiable with a continuous density f
2) $f(q_\alpha) > 0$, for $\alpha \in (0,1)$
3) $\mathbb{E}(|X|) < \infty$

We begin by finding an expression for $\mathbb{E}(\mathrm{CTE}_n^\alpha)$:

$$\begin{aligned}\mathbb{E}(\mathrm{CTE}_n^\alpha) &= \mathbb{E}(\mathbb{E}(\mathrm{CTE}_n^\alpha \,|\, X_{n-k_n:n})) \\ &= \mathbb{E}(\mathbb{E}(X \,|\, X > X_{n-k_n:n})) \\ &= \mathbb{E}(\psi(X_{n-k_n:n}))\end{aligned} \tag{3}$$

where the function $\psi(\cdot)$ is defined as

$$\psi(x) = \mathbb{E}(X \,|\, X > x) = \int_x^\infty \frac{S(t)}{S(x)}\,\mathrm{d}t + x. \tag{4}$$

Note that the above relation does not require X to be non-negative, and that $\psi(q_\alpha)$ is the α-level CTE under F, so that $\psi(q_\alpha) = \eta_\alpha(X)$.

Lemma 1. *Let α and n be such that $n\alpha$ is an integer. Then we have,*

$$\psi(q_\alpha) - \mathbb{E}(\mathrm{CTE}_n^\alpha) = \mathbb{E}\left(\frac{1}{1-U_{n\alpha:n}} \int_{q_\alpha \wedge Q(U_{n\alpha:n})}^{q_\alpha \vee Q(U_{n\alpha:n})} |\, U_{n\alpha:n} - F(t)| \,\mathrm{d}t\right). \tag{5}$$

Proof. See Appendix. □

From Lemma 1 we obtain:

Corollary 1. *Let α and n be such that $n\alpha$ is an integer. Then the α-level empirical CTE estimator based on a random sample of size n from F is negatively biased as an estimator of the α-level CTE.*

We note that for noninteger valued $n\alpha$, the negative biasedness of the empirical CTE estimator follows if k_n (as in (2)) is taken to be $\lfloor n\alpha \rfloor$. However, if k_n is taken to be $\lfloor n\alpha \rfloor + 1$, the estimator may be positively biased, as in the uniform case of Example 1 with $n = 4$ and $\alpha = 1/3$.

Theorem 2. *Let α and $n(\geq 4)$ be such that $n\alpha$ is an integer, and let A_n be defined as*

$$A_n := (\alpha + (2\alpha - 1)/(n-2) - \theta_n \sigma/\sqrt{n-3},\ \alpha + (2\alpha-1)/(n-2) + \theta_n\sigma/\sqrt{n-3})$$

where $\sigma^2 = (n\alpha - 1)(n(1-\alpha)-1)/(n-2)^2$. Then we have,

$$\begin{aligned}\left(1 - \frac{1}{n+1}\right)\frac{\alpha}{2\max_{u\in A_n} f(Q(u))} &\leq n\left(\psi(q_\alpha) - \mathbb{E}\left(\mathrm{CTE}_n^\alpha\right)\right) \\ &\leq \frac{\alpha}{2\min_{u\in A_n} f(Q(u))} + \varepsilon_n\end{aligned} \tag{6}$$

where ε_n is bounded above by

$$\frac{n(|q_\alpha| + \mathbb{E}(|X|))}{(1-\alpha)^2(\alpha - (1-\alpha)/(n-1))} \exp\left\{-\frac{\theta_n^2}{3\left(1 + \theta_n / \sigma\sqrt{n-3}\right)}\right\}.$$

Proof. See Appendix. □

Corollary 2. *For a sequence $\{m_n\}_{n\geq 1}$, such that $m_n \to \infty$ and $m_n\alpha$ is an integer for all $n \geq 1$, we have*

$$\lim_{n\to\infty} n\left(\psi(q_\alpha) - \mathbb{E}\left(\mathrm{CTE}^\alpha_{m_n}\right)\right) = \frac{\alpha}{2f(q_\alpha)}.$$

Proof. Defining $\theta_n = \sqrt{C\log n}$ in the statement of Theorem 2, for some $C > 3$, the result follows from Theorem 2. □

In Corollary 2 we restrict ourselves to sample sizes n for which $n\alpha$ is an integer. If we do not do so, the limit need not exist. For example, from the calculations in Example 1 we see that for rational α with $\alpha = p/q$ for p and q coprimes, $n = mq + r$ for some $r \in \{0, 1, \ldots, q-1\}$, and $k_n = \lfloor n\alpha \rfloor$ in the definition of CTE^α_n, we have,

$$n(\psi(q_\alpha) - \mathbb{E}(\mathrm{CTE}^\alpha_n)) = \left[\frac{\alpha(r+1) - \lfloor r\alpha \rfloor}{2}\right]\left(1 - \frac{1}{n+1}\right).$$

In particular, the above implies that the sequence $\{n(\psi(q_\alpha) - \mathbb{E}(\mathrm{CTE}^\alpha_n))\}_{n\geq 1}$ is oscillatory in nature. Nonetheless, under the conditions of Theorem 2, it can easily be shown that

$$\limsup_{n\to\infty} n(\psi(q_\alpha) - \mathbb{E}(\mathrm{CTE}^\alpha_n)) < \infty$$

for k_n in the definition of CTE^α_n such that $k_n - n\alpha$ is bounded.

For any F satisfying the conditions of Theorem 2, we define the *asymptotic bias* of the α-level empirical CTE estimator to be $-\alpha/(2f(q_\alpha))$, and the *asymptotic relative bias* to be

$$-\frac{\alpha}{2\psi(q_\alpha)f(q_\alpha)}.$$

The following example demonstrates that in the case of the exponential distribution, the asymptotic bias is a good approximation to the n-times magnified bias even for a moderate sample size n, and that the absolute value of the relative bias increases with α, for α close to 1.

Example 2. [Exponential Distribution]
Suppose that X is exponentially distributed with mean β. The exponential is particularly interesting as it is on the boundary of heavy and light tailed distributions. Since $X_{j:n}$ is distributed as the sum of j independent exponential distributions with means $\beta/n, \ldots, \beta/(n+1-j)$, we have

$$\begin{aligned}\mathbb{E}(\mathrm{CTE}_n^{\alpha}) &= \mathbb{E}(X \mid X > X_{n-k_n:n}) = \mathbb{E}(X - X_{n-k_n:n} \mid X > X_{n-k_n:n}) + \mathbb{E}(X_{n-k_n:n}) \\ &= \beta\left(1 + \frac{1}{n}\sum_{i=k_n+1}^{n}\frac{n}{i}\right).\end{aligned}$$

But as $\eta_\alpha(X) = \beta(1-\ln(1-\alpha))$, the bias is given by

$$\mathbb{E}(\mathrm{CTE}_n^{\alpha}) - \eta_\alpha(X) = \beta\left(\frac{1}{n}\sum_{i=k_n+1}^{n}\frac{n}{i} + \ln(1-\alpha)\right).$$

It is worth observing that

$$\frac{1}{n}\sum_{i=k_n+1}^{n}\frac{n}{i} \approx \int_{1-\alpha}^{1}\frac{1}{x}\,\mathrm{d}x = -\ln(1-\alpha)$$

with the approximation being an upper bound. By Theorem 2,

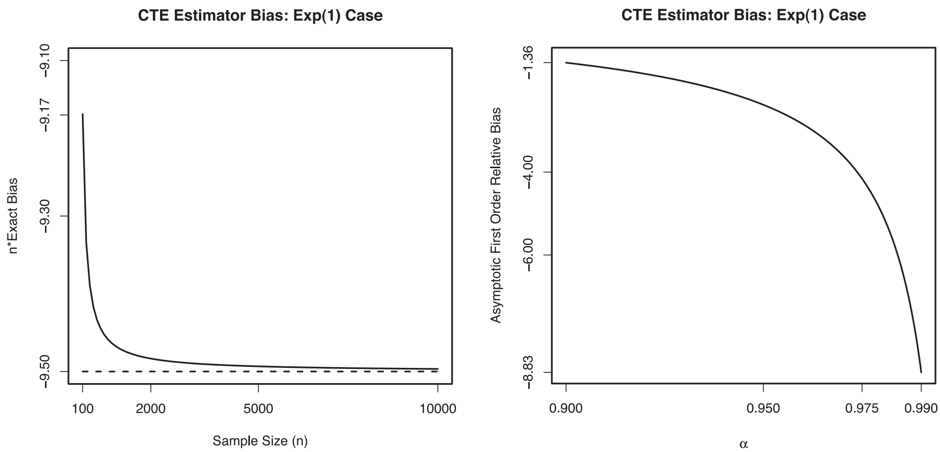
$$\lim_{n\to\infty} n(\mathbb{E}(\mathrm{CTE}_n^{\alpha}) - \eta_\alpha(X)) = \frac{-\alpha}{2f(q_\alpha)} = \frac{-\alpha\beta}{2(1-\alpha)}.$$

For the calculations in Figure 1a we used the base case of $\beta = 1$. The value for the asymptotic bias is -9.5 and is indicated by the dashed line. Figure 1a shows that the first order approximation for bias performs very well in this case, even for moderate values of n. For the asymptotic relative bias, we have the following expression:

$$\lim_{n\to\infty} n\left(\frac{\mathbb{E}\left(\mathrm{CTE}_n^{\alpha}\right) - \psi(q_\alpha)}{\psi(q_\alpha)}\right) = \frac{-\alpha}{2(1-\alpha)(1-\ln(1-\alpha))}.$$

Figure 1b shows that the relative bias worsens with increasing α. □

It is clear that the asymptotic relative bias depends on the tail of F. The next example studies this dependence within the class of Pareto distributions.



(a) Accuracy of the Asymptotic Bias: $\alpha = 0.95$.

(b) Asymptotic First Order Relative Bias

FIGURE 1: Exponential with mean 1.

We chose the Pareto because it is an important class of distributions in actuarial applications, and because it contains distributions with tails both heavier and lighter than the exponential.

Example 3. [Pareto Distribution]
For our purpose we will use the same parameterization as in Manistre and Hancock (2005). By Pareto(β, ξ) we mean the distribution with hazard rate function

$$\mu(x) = \begin{cases} 0 & x \leq 0; \\ \dfrac{1}{\beta + \xi x} & x > 0; \end{cases} .$$

It is clear that for β fixed, the lower the value of ξ, the lighter the tail of the Pareto, and that for $\xi = 0$ the Pareto reduces to the exponential distribution with mean β. Moreover,

$$\mathbb{E}(X \mid X > u) = \frac{\beta + u}{1 - \xi}, \quad \xi < 1.$$

Applying Theorem 2, we obtain

$$\lim_{n \to \infty} n \left(\frac{\mathbb{E}\left(\mathrm{CTE}_n^\alpha\right) - \psi(q_\alpha)}{\psi(q_\alpha)} \right) = \frac{-\alpha(1 - \xi)}{2(1 - \alpha)\left((1 - \alpha)^\xi + \xi^{-1}\left(1 - (1 - \alpha)^\xi\right)\right)}.$$

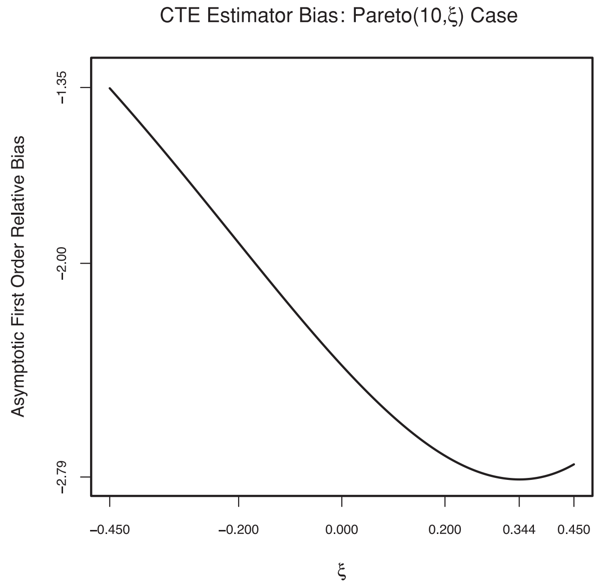


FIGURE 2: Asymptotic First Order Expression of Relative Bias: Pareto$(10, \xi)$ Case.

Figure 2 plots the asymptotic relative bias for varying values of ξ, for $\alpha = 0.95$. Note that the relative bias does not depend on the value of β as it is a scale parameter. The worst case value of the asymptotic relative bias is -2.799 which is attained at $\xi = 0.344$. It is interesting to note that with increasing ξ the tail gets heavier, but the relative bias does not necessarily get worse. □

4. First Order Bias Corrected Empirical CTE

As observed in Manistre and Hancock (2005), the bias of the empirical CTE could be significant for small samples. In Kim and Hardy (2007) the bootstrap was suggested as a way to estimate, and hence correct for, this bias. However, the closed form expression for the asymptotic bias (Corollary 2) suggests an alternative estimator obtained by substituting into this expression an estimate of $f(q_\alpha)$. In fact, Figure 1a suggests that this estimator could have good small sample performance. Moreover, the fact that the standard error of the empirical VaR and the covariance between the empirical CTE and the empirical VaR both involve $1/f(q_\alpha)$ implies that estimating the asymptotic bias does not add further to the computation, making this approach more appealing. In this section we present simulation results pertaining to three practical examples from Manistre and Hancock (2005) and Kim and Hardy (2007). In these simulations we compare the performance of the empirical CTE, Exact Bootstrap Bias Corrected empirical CTE, and the First Order Bias Corrected empirical CTE (FOBC).

4.1. The Examples

The first two examples consider the liability from a naked position in a written 10-year European put option with the initial price of the asset set at \$100, the strike price set at \$180, and the risk free rate set at 0.5% per month effective. Note that the strike price of 180 approximately equals $100 * 1.005^{120}$. In the first example, the monthly log-return on the asset is assumed to be normally distributed with mean 0.00947 and standard deviation 0.04167, these values being estimated from monthly S&P 500 data during 1956-2001 (see Hardy (2003)). The second example assumes instead that the monthly log-return follows a regime switching normal distribution with the parameters (estimated using the same S&P 500 data, see Hardy (2003)) being $(0.0127, -0.0162)$ for the mean vector, $(0.0351, 0.0691)$ for the two standard deviations, and the transition probabilities $p_{12} = 0.0468$ and $p_{21} = 0.3232$. We note that the mean and the standard deviation of the stationary distribution of the monthly log-return are 0.0091 and 0.0421, respectively, which expectedly is close to that of the first example. The liability in the Lognormal (LN) example has a .95-level VaR of 18.11 and a .95-level CTE of 31.26. The liability in the second, Regime Switching Lognormal distribution with two regimes (RSLN), is relatively heavier tailed and this is reflected in a higher .95-level VaR of 29.2 and higher CTE at the .95-level of 42.96.

The third example is that of a liability distributed as a polynomial tailed Pareto distribution (see Example 3) with parameters $\beta = 10$ and $\xi = 0.2$. This distribution, popular in actuarial applications, has mean 12.5 and standard deviation 16.14. Moreover, it has a .95-level VaR of 41.03 and a .95-level CTE of 63.79.

4.2. Simulation

In all three examples we simulated 100 million random samples for each of the sample sizes 200 and 1000. In Tables 1, 2, and 3 we report for each of the three estimators of the CTE, an estimate of their bias and an estimate of their root mean square error (RMSE) along with an estimate of the standard errors. The estimate of the bias is the difference between the average sample estimate (over the 100 million samples) and the true CTE.

The .95-level empirical CTE was calculated for each random sample using the average of the highest 5% of the values in the sample. To estimate the asymptotic bias, we estimate $f(q_{.95})$ by the estimate of the density at the 95-th percentile. We estimate the 95-th percentile by the corresponding empirical quantile. To estimate the density, we resort to a Gaussian kernel density estimator. For the bandwidth we use 0.9 times the minimum of the standard deviation and the interquartile range divided by 1.34 times the sample size to the negative one-fifth power (*i.e.* Silverman's 'rule of thumb', see Silverman (1986)) unless the quartiles coincide, in which case a positive value is used. We note that the bandwidth being $O(n^{-1/5})$ leads to the optimal rate of convergence of the mean square error to zero. The .95-level empirical CTE with first order bias correction (FOBC) was calculated as the difference between the empirical CTE and

TABLE 1

LOGNORMAL EXAMPLE: 95% CTE ESTIMATORS (CTE = 31.2552)

Sample Size	Estimator	Bias	S.E. of Bias	RMSE	S.E. of RMSE
200	Empirical	– 2.67652%	0.00166%	16.859%	0.00122%
	FOBC	– 0.37343%	0.00170%	17.046%	0.00125%
	Bootstrap	0.01192%	0.00170%	17.025%	0.00123%
1000	Empirical	– 0.54222%	0.00074%	7.4495%	0.00054%
	FOBC	– 0.01675%	0.00075%	7.4633%	0.00054%
	Bootstrap	0.00078%	0.00075%	7.4645%	0.00054%

TABLE 2

REGIME SWITCHING LOGNORMAL EXAMPLE: 95% CTE ESTIMATORS (CTE = 42.9634)

Sample Size	Estimator	Bias	S.E. of Bias	RMSE	S.E. of RMSE
200	Empirical	– 2.06071%	0.00126%	12.805%	0.00087%
	FOBC	– 0.19095%	0.00129%	12.891%	0.00086%
	Bootstrap	0.01909%	0.00129%	12.913%	0.00086%
1000	Empirical	– 0.41780%	0.00056%	5.6533%	0.00046%
	FOBC	– 0.01023%	0.00057%	5.6625%	0.00046%
	Bootstrap	0.00028%	0.00057%	5.6643%	0.00046%

TABLE 3

PARETO EXAMPLE: 95% CTE ESTIMATORS (CTE = 63.7853)

Sample Size	Estimator	Bias	S.E. of Bias	RMSE	S.E. of RMSE
200	Empirical	– 1.33789%	0.00180%	18.086%	0.00187%
	FOBC	– 0.17752%	0.00183%	18.276%	0.00192%
	Bootstrap	0.04017%	0.00184%	18.366%	0.00193%
1000	Empirical	– 0.27034%	0.00081%	8.1470%	0.00068%
	FOBC	– 0.01040%	0.00082%	8.1685%	0.00069%
	Bootstrap	0.00165%	0.00082%	8.1707%	0.00070%

$1/n$ times the estimator for the asymptotic bias derived by substituting for $f(q_\alpha)$ the Gaussian kernel density estimate at the empirical 95-th percentile. In Kim and Hardy (2007), using the exact expression of the bootstrap mean of the k-th order statistic of Hutson and Ernst (2000), an exact expression for the bootstrap mean of the empirical CTE is derived. Since, for large n, the bootstrap distribution is close to the exact distribution, one would expect that the difference between the empirical CTE and the bootstrap mean of the empirical CTE would be close to the negative bias of the empirical CTE. This leads to the exact bootstrap bias corrected empirical CTE of Kim and Hardy (2007) which is twice the empirical CTE minus the exact bootstrap mean of the empirical CTE.

All of the code was written on the R software environment for statistical computing and graphics (see R Development Core Team (2008)). It was run parallel on 10 nodes (40 processors) of a 22 node Beowulf cluster using the snow R package (Tierney et al. (2008, 2009)). For kernel density estimation we used the R function *density*, with default setting for both the bandwidth and the kernel.

4.3. Results

The results in Tables 1, 2, and 3 are presented as a percentage of the true .95-level CTE as this both facilitates comparison among examples and comparison of our results with those of Kim and Hardy (2007). We make the following observations:

1. As expected we see negative bias for the empirical CTE. That the bias converges to zero at the $1/n$ rate (Corollary 2) while the standard error converges to zero at the rate $1/\sqrt{n}$ means that bias is less of an issue at sample size 1000 than at sample size 200, a phenomenon that is seen in the tables.
2. Both the bootstrap and the first order bias correction significantly reduce bias, with the bootstrap correction doing significantly better. In Table 4, we have tabulated at the $\alpha = 0.95$ level the exact first order bias correction (as a percentage of the true CTE value) and the residual bias in the empirical CTE using this exact asymptotic bias. This suggests that the higher bias of the FOBC is due to the fact that the choice of the bandwidth is driven by the goal of reducing the MSE rather than the bias. Hence, using a bandwidth which goes to zero at a rate faster than $n^{-1/5}$ will reduce bias further at the cost of an increased RMSE.
3. Bias correction either by bootstrap or by first order bias correction leads to a higher RMSE than the empirical CTE. Among the two bias corrections, except in the lognormal example with sample size 200, FOBC leads to a slightly lower MSE than the bootstrap correction. This suggests that the implicit *bandwidth* in the bootstrap bias correction goes to zero faster than $n^{-1/5}$ leading to lower bias but higher MSE. This is something that should be explored further.

In Table 5, we report the results from similar simulations but at the 0.99-level, and for the sample size of 200. Here we observe that the FOBC consistently has a RMSE very close to that of the empirical CTE, whereas the bootstrap bias correction has significantly higher RMSE. Again, there is a tradeoff between lowering the bias and lowering the standard error. We observe that FOBC consistently keeps the RMSE close to that of the empirical CTE, while reducing the bias significantly. The Mixed estimator in Table 5 refers to the estimator proposed in Kim and Hardy (2007) which chooses between the empirical CTE estimator and its exact bootstrap mean based on the estimated bias

TABLE 4

First Order Bias – Sample Size of 200

Example	95%-Level			99%-Level		
	CTE	**First Order Bias**	**Residual Bias**	**CTE**	**First Order Bias**	**Residual Bias**
Lognormal	31.255	– 2.718%	0.042%	47.728	– 5.259%	0.3962%
RSLN	42.963	– 2.093%	0.032%	59.999	– 4.327%	0.3239%
Pareto	63.785	– 1.356%	0.018%	106.993	– 5.811%	0.4226%

TABLE 5

99% CTE Estimators – Sample Size of 200

Example	Estimator	Bias	S.E. of Bias	RMSE	S.E. of RMSE
Lognormal (CTE = 47.7281)	Empirical	– 4.86281%	0.00152%	15.967%	0.00111%
	FOBC	– 1.79214%	0.00159%	16.031%	0.00111%
	Bootstrap	0.21649%	0.00168%	16.817%	0.00116%
	Mixed	– 4.9277%	–	16.029%	–
RSLN (CTE = 59.9989)	Empirical	– 4.00327%	0.00123%	12.913%	0.00088%
	FOBC	– 1.15370%	0.00129%	12.972%	0.00087%
	Bootstrap	0.16416%	0.00135%	13.549%	0.00095%
	Mixed	– 4.0009%	–	12.731%	–
Pareto (CTE = 106.993)	Empirical	– 5.38801%	0.00319%	32.366%	0.00620%
	FOBC	– 3.15112%	0.00323%	32.455%	0.00626%
	Bootstrap	0.95143%	0.00362%	36.214%	0.00711%
	Mixed	– 6.8749%	–	30.879%	–

and standard error. In fact, the values reported for the Mixed estimator in Table 5 are from Kim and Hardy (2007). We point out that in Table 7 of Kim and Hardy (2007) the mixed estimator is compared with the exact bootstrap mean of the empirical CTE (with approximately twice the bias) whereas in Table 5 it is compared with the exact bootstrap bias corrected empirical CTE (with much reduced bias). Since the error in the RMSE for the Mixed estimator in Kim and Hardy (2007) is not provided, we can only say that the FOBC seems to compare well in terms of RMSE, and significantly reduces bias unlike the Mixed where the bias could be significantly higher than that of the empirical CTE (as in the Pareto example).

5. Conclusion

We first established theoretically the non-existence of a nonparametric unbiased estimator when the unknown distribution is allowed to be any distribution with a finite CTE. While this result is part of the folklore, there does not seem to be any published proof of it, and moreover our proof lends itself to further strengthening of the result. Second, we derived a closed form expression for the bias of the empirical CTE when the underlying distribution admits a density. Third, in the continuous case we derived finite sample bounds for the bias, showed that the bias is of order $1/n$, and derived a simple closed form expression for the first order bias.

The only unknown quantity in the first order term of the bias of the α-level empirical CTE is the density evaluated at the α-level quantile. This suggests a simple plug-in estimator as an alternate to the bootstrap method of bias correction. Fourth, we studied this proposed estimator and found that it has the advantage of being very easily computable on any statistical environment (in particular on R) that has a kernel density estimator, and tends to have reduced RMSE compared to the bootstrap method of bias correction. The flip side of having reduced RMSE is that the proposed method tends to have a larger bias than the bootstrap method of bias correction.

6. Acknowledgment

We thank the two referees for their thoughtful suggestions. Bangwon Ko's collaboration in this research was financially supported by Soongsil University, South Korea. The computations for this paper were performed on equipment funded by National Science Foundation grant DMS 06-18883.

References

Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1997) Thinking coherently. *RISK*, **10(11)**: 68-71.

Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1999) Coherent measures of risk. *Math. Finance*, **9(3)**: 203-228.

Bickel, P.J. and Lehmann, E.L. (1969) Unbiased estimation in convex families. *Ann. Math. Statist.*, **40**: 1523-1535.

Bilodeau, M. (2004) Discussion on: "Tail conditional expectations for elliptical distributions" [*N. Am. Actuar. J.* **7** (2003), no. 4, 55-71] by Z.M. Landsman and E.A. Valdez. *N. Am. Actuar. J*, **8(3)**: 118-123. With a reply by Landsman and Valdez.

Cai, J. and Li, H. (2005) Conditional tail expectations for multivariate phase-type distributions. *J. Appl. Probab.*, **42(3)**: 810-825.

Delbaen, F. (2000) *Coherent risk measures*. Lecture Notes, Pisa.

Dhaene, J., Goovaerts, M.J. and Kaas, R. (2003) Economic capital allocation derived from risk measures. *N. Am. Actuar. J.*, **7(2)**: 44-59.

Hardy, M. (2003) *Investment Guarantees: The New Science of Modeling and Risk Management for Equity-Linked Life Insurance*. Wiley, 1 edition.

Hutson, A.D. and Ernst, M.D. (2000) The exact bootstrap mean and variance of an L-estimator. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **62(1)**: 89-94.

Jones, B.L. and Zitikis, R. (2003) Empirical estimation of risk measures and related quantities. *N. Am. Actuar. J.*, **7(4)**: 44-54.

Kim, J.H.T. and Hardy, M. (2007) Quantifying and correcting the bias in estimated risk measures. *Astin Bulletin*, **37(2)**: 365-386.

Landsman, Z.M. and Valdez, E.A. (2003) Tail conditional expectations for elliptical distributions. *N. Am. Actuar. J.*, **7(4)**: 55-71.

MANISTRE, B.J. and HANCOCK, G.H. (2005) Variance of the CTE estimator. *N. Am. Actuar. J.*, **9(2)**: 129-156.
R DEVELOPMENT CORE TEAM (2008) *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
REISS, R.-D. (1989) *Approximate distributions of order statistics*. Springer Series in Statistics. Springer-Verlag, New York. With applications to nonparametric statistics.
ROSENBLATT, M. (1956) Remarks on some nonparametric estimates of a density function. *Ann. Math. Statist.*, **27**: 832-837.
RUSSO, R.P. and SHYAMALKUMAR, N.D. (2008) *Bounds for the bias of the empirical CTE*. Unpublished Report.
SILVERMAN, B.W. (1986) *Density estimation for statistics and data analysis*. Monographs on Statistics and Applied Probability. Chapman & Hall, London.
TIERNEY, L., ROSSINI, A.J., LI, N. and SEVCIKOVA, H. (2008) *snow: Simple Network of Workstations*. R package version 0.3-3.
TIERNEY, L., ROSSINI, A.J., LI, N. and SEVCIKOVA, H. (2009). Snow: A parallel computing framework for the R system. *Int. J. Parallel Prog.*, **37**: 78-90.

APPENDIX

Proof of Lemma 1. The negative bias of CTE_n^α can be expressed as

$$\begin{aligned}\left(\psi(q_\alpha) - \mathbb{E}\left(\mathrm{CTE}_n^\alpha\right)\right) &= \mathbb{E}\left(\psi(q_\alpha) - \psi(X_{n\alpha:n})\right)\\ &= \int_{q_\alpha}^{\infty}\frac{S(t)}{S(q_\alpha)}\,\mathrm{d}t - \mathbb{E}\left(\int_{X_{n\alpha:n}}^{\infty}\frac{S(t)}{S(X_{n\alpha:n})}\,\mathrm{d}t\right) + \mathbb{E}(q_\alpha - X_{n\alpha:n})\\ &= \left(\frac{1}{1-\alpha} - \mathbb{E}\left(\frac{1}{S(X_{n\alpha:n})}\right)\right)\int_{q_\alpha}^{\infty} S(t)\,\mathrm{d}t \\ &\quad + \mathbb{E}\left(\frac{\mathrm{sgn}(X_{n\alpha:n} - q_\alpha)}{S(X_{n\alpha:n})}\int_{q_\alpha\wedge X_{n\alpha:n}}^{q_\alpha\vee X_{n\alpha:n}} S(t)\,\mathrm{d}t\right) + \mathbb{E}(q_\alpha - X_{n\alpha:n}).\end{aligned} \tag{7}$$

Now, the first term in the final expression of (7) vanishes as we have

$$\mathbb{E}\left(\frac{1}{S(X_{n\alpha:n})}\right) = \mathbb{E}\left(\frac{1}{1-U_{n\alpha:n}}\right) = \frac{1}{1-\alpha}.$$

The other two terms in the final expression of (7) can be written as

$$\begin{aligned}&\mathbb{E}\left(\mathrm{sgn}(X_{n\alpha:n} - q_\alpha)\int_{q_\alpha\wedge X_{n\alpha:n}}^{q_\alpha\vee X_{n\alpha:n}}\left(\frac{S(t)}{S(X_{n\alpha:n})} - 1\right)\mathrm{d}t\right)\\ &\quad = \mathbb{E}\left(\frac{\mathrm{sgn}(X_{n\alpha:n} - q_\alpha)}{S(X_{n\alpha:n})}\int_{q_\alpha\wedge X_{n\alpha:n}}^{q_\alpha\vee X_{n\alpha:n}}\left(F(X_{n\alpha:n}) - F(t)\right)\mathrm{d}t\right)\\ &\quad = \mathbb{E}\left(\frac{1}{1-U_{n\alpha:n}}\int_{q_\alpha\wedge Q(U_{n\alpha:n})}^{q_\alpha\vee Q(U_{n\alpha:n})}\left|U_{n\alpha:n} - F(t)\right|\mathrm{d}t\right).\end{aligned} \tag{8}$$

Combining (7) with (8) completes the proof. □

Proof of Theorem 2. First, we split the expectation in (5) into two parts: the first where $U_{n\alpha:n}$ is restricted to A_n, where A_n is a neighborhood around α, and the second where $U_{n\alpha:n}$ is restricted to A_n^c. A lower bound for the first part can be derived as follows:

$$\begin{aligned}
&\mathbb{E}\left(\frac{I_{A_n}(U_{n\alpha:n})}{1-U_{n\alpha:n}}\int_{q_\alpha\wedge Q(U_{n\alpha:n})}^{q_\alpha\vee Q(U_{n\alpha:n})}|\,U_{n\alpha:n}-F(t)|\,\mathrm{d}t\right)\\
&\quad=\mathbb{E}\left(\frac{I_{A_n}(U_{n\alpha:n})}{1-U_{n\alpha:n}}\int_{\alpha\wedge U_{n\alpha:n}}^{\alpha\vee U_{n\alpha:n}}\frac{|U_{n\alpha:n}-s|}{f(Q(s))}\,\mathrm{d}s\right)\\
&\quad\geq\frac{1}{2\max_{s\in A_n}f(Q(s))}\,\mathbb{E}\left(\frac{(U_{n\alpha:n}-\alpha)^2}{1-U_{n\alpha:n}}\right)\\
&\quad=\frac{1}{2(1-\alpha)\max_{s\in A_n}f(Q(s))}\,\mathbb{E}\left((U_{n\alpha:n-1}-\alpha)^2\right)\\
&\quad=\frac{1}{2(1-\alpha)\max_{s\in A_n}f(Q(s))}\left[\frac{\alpha(1-\alpha)}{n+1}\right]\\
&\quad=\frac{\alpha}{2n\max_{s\in A_n}f(Q(s))}\left[1-\frac{1}{n+1}\right].
\end{aligned}\tag{9}$$

Similarly, it can be shown that the following is an upper bound for the above:

$$\frac{\alpha}{2n\min_{s\in A_n}f(Q(s))}.\tag{10}$$

Now, the remaining part of the final expectation in (5) can be bounded as follows:

$$\begin{aligned}
0&\leq\mathbb{E}\left(\frac{I_{A_n^c}(U_{n\alpha:n})}{1-U_{n\alpha:n}}\int_{q_\alpha\wedge Q(U_{n\alpha:n})}^{q_\alpha\vee Q(U_{n\alpha:n})}|\,U_{n\alpha:n}-F(t)|\,\mathrm{d}t\right)\\
&\leq\mathbb{E}\left(\frac{I_{A_n^c}(U_{n\alpha:n})(U_{n\alpha:n}-\alpha)(Q(U_{n\alpha:n})-q_\alpha)}{1-U_{n\alpha:n}}\right).
\end{aligned}\tag{11}$$

As $X\sim F$, we have

$$x(F(-x)+1-F(x))\leq\mathbb{E}(|X|I_{(x,\infty)}(|X|))\leq\mathbb{E}(|X|).$$

In particular, $u(1-u)Q(u)\leq\mathbb{E}(|X|)$. Using this in (11) and Lemma 2 (see below) we obtain

$$
\begin{aligned}
0 &\le \mathbb{E}\left(\frac{I_{A_n^c}(U_{n\alpha:n})}{1-U_{n\alpha:n}} \int_{q_\alpha \wedge Q(U_{n\alpha:n})}^{q_\alpha \vee Q(U_{n\alpha:n})} \left| U_{n\alpha:n} - F(t) \right| \mathrm{d}t \right) \\
&\le \left(|q_\alpha| + \mathbb{E}(|X|) \right) \mathbb{E}\left(\frac{I_{A_n^c}(U_{n\alpha:n}) \left| U_{n\alpha:n} - \alpha \right|}{U_{n\alpha:n}(1-U_{n\alpha:n})^2} \right) \\
&\le \frac{\left(|q_\alpha| + \mathbb{E}(|X|) \right) \Pr\left(U_{n\alpha-1:n-3} \notin A_n \right)}{(1-\alpha)(1-\alpha+(1-2\alpha)/(n-2))\,(\alpha-(1-\alpha)/(n-1))} \\
&\le \frac{\left(|q_\alpha| + \mathbb{E}(|X|) \right)}{(1-\alpha)^2\,(\alpha-(1-\alpha)/(n-1))} \exp\left\{ -\frac{\theta_n^2}{3\left(1+\theta_n/\sigma\sqrt{n-3}\right)} \right\}.
\end{aligned}
\tag{12}
$$

Combining (9), (10) and (12) completes the proof. □

The following is Lemma 3.1.1 from Reiss (1989).

Lemma 2. *For every $\varepsilon > 0$ and $r \in \{1, 2, \ldots, n\}$ we have*

$$
\Pr\left(\left| U_{r:n} - \frac{r}{n+1} \right| > \frac{\varepsilon\sigma}{\sqrt{n}} \right) \le \exp\left(-\frac{\varepsilon^2}{3(1+\varepsilon/(\sigma\sqrt{n}))} \right)
\tag{13}
$$

where $\sigma^2 = \dfrac{r}{n+1}\left(1-\dfrac{r}{n+1}\right)$.

Nariankadu D. Shyamalkumar
Dept. of Statistics and Actuarial Science
The University of Iowa
241 Schaeffer Hall, Iowa City, IA 52242, USA
Telephone: (319)-335-1980
E-Mail: shyamal-kumar@uiowa.edu

Bangwon Ko
Soongsil University
11 Sangdo-Dong, Dongjak-Gu
Seoul, 156-743, South Korea
Telephone: 82-2-820-0447
E-Mail: bko@ssu.ac.kr

Ralph P. Russo
Dept. of Statistics and Actuarial Science
The University of Iowa
241 Schaeffer Hall, Iowa City, IA 52242, USA
E-Mail: rrusso@stat.uiowa.edu