COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
2023, VOL. 52, NO. 11, 3652–3660
https://doi.org/10.1080/03610926.2021.1977960

Markov's inequality: Sharpness, renewal theory, finite samples, reliability theory

Mark Brown[a] and Joel E. Cohen[a,b,c,d]

[a]Department of Statistics, Columbia University, New York, NY, USA; [b]Laboratory of Populations, The Rockefeller University and Columbia University, New York, NY, USA; [c]Earth Institute, Columbia University, New York, NY, USA; [d]Department of Statistics, University of Chicago, Chicago, IL, USA

ABSTRACT
We examine Markov's inequality in the light of renewal theory and reliability theory. Suppose the non negative random variable (rv) X has cumulative distribution function (cdf) F with survival function $\overline{F} := \Pr(X > x) = 1 - F$ and left-continuous version of the survival function $\overline{F}(x^-) := \Pr(X \geq x)$, $x \geq 0$. We determine the points, if any, such that $x \geq \mu$ and $\overline{F}(x^-) = \mu/x$. We offer an alternative proof of Markov's inequality by observing that, if some collection of events $\{A_t : t \geq 0\}$ exists such that $\Pr(A_t) = t\overline{F}(t^-)/\mu$, then because $t\overline{F}(t^-)/\mu$ equals a probability, it must satisfy $t\overline{F}(t^-)/\mu \leq 1$, which is equivalent to Markov's inequality. We choose events connected to stationary renewal processes. When we know only the sample size n and the sample average $\overline{x}_n$ of n non negative observations $x_1, \ldots, x_n$, we establish an upper bound on the left-continuous version of the empirical survival function that improves Markov's inequality. We show that an upper bound of Markov type for the survival function is sharp when F is "new better than used in expectation" (NBUE) or has "decreasing mean residual life" (DMRL).

ARTICLE HISTORY
Received 4 November 2020
Accepted 3 September 2021

KEYWORDS
Bienaymé's inequality; Markov's inequality; Chebyshev–Markov inequality; mean residual life; reliability; renewal theory; survival curve

1. Introduction

Markov's inequality or the Bienaymé–Chebyshev–Markov inequality states that if X is a non negative random variable (rv) with expectation $0 < \mu := EX < \infty$ and if $a > 0$, then $\Pr(X \geq a) \leq \min(1, \mu/a)$. Markov's inequality has been refined and extended in many ways. For a very small sample of the relevant papers see, e.g., Ghosh (2002), Marshall and Olkin (2007), and Cohen (2015). Nevertheless, we offer several observations about Markov's inequality that we believe to be new.

In Section 2, we assume that the non negative rv X has cumulative distribution function (cdf) F with survival function $\overline{F} := \Pr(X > x) = 1 - F$ and left-continuous version of the survival function $\overline{F}(x^-) := \Pr(X \geq x)$, $x \geq 0$. Given F, we determine the points, if any, such that $x \geq \mu$ and $\overline{F}(x^-) = \mu/x$.

In Section 3, we offer an alternative proof of Markov's inequality by observing that, if some collection of events $\{A_t : t \geq 0\}$ exists such that $\Pr(A_t) = t\overline{F}(t^-)/\mu$, then because

CONTACT Joel E. Cohen cohen@rockefeller.edu Laboratory of Populations, The Rockefeller University, 1230 York Avenue, Box 20, New York, NY 10065, USA.

$t\overline{F}(t^-)/\mu$ equals a probability, it must satisfy $t\overline{F}(t^-)/\mu \leq 1$, which is equivalent to Markov's inequality. We choose $\{A_t : t \geq 0\}$ from events connected to stationary renewal processes.

In Section 4, we suppose we know only the sample size n and the sample average $\overline{x}_n$ of n non negative observations $x_1, \ldots, x_n$. We establish an upper bound on the left-continuous version of the empirical survival function,

$$\overline{F}_n{}^{\wedge}(x^-) := \#\{X_i \geq x : i = 1, \ldots, n\}/n, \quad x \geq 0.$$

As $\overline{F}_n{}^{\wedge}$ is a probability distribution over n unknown points with known mean $\overline{x}_n$, Markov's inequality applies. We use the additional information, beyond n and $\overline{x}_n$, that $\overline{F}_n{}^{\wedge}$ is atomic and the size of the atoms are multiples of $1/n$ to derive a sharp upper bound on $\overline{F}_n{}^{\wedge}(x^-)$ for all $x \geq 0$. This upper bound improves on Markov's inequality for this case. If it is known from context that $\Pr\{X \in [a, b], 0 \leq a < b < \infty\} = 1$, we derive a sharp upper bound on $\overline{F}_n{}^{\wedge}(x^-)$ for all $x \geq 0$.

In Section 5, when F is "new better than used in expectation" (NBUE), we show that an upper bound of Markov type for $\overline{F}(x^-)$ is sharp. Brown (2006, p. 206, Theorem 3.2, Eq. (3.5)) derived this bound but did not show it to be sharp. When Marshall and Olkin (2007, p. 198, Section 6B) presented this bound, citing Brown's earlier technical report, they asked whether it was sharp. We also show that the bound in sharp in the subclass within NBUE of survival functions with "decreasing mean residual life" (DMRL).

2. When is Markov's inequality sharp?

Let X be a non negative rv with cdf $F(x) = \Pr(X \leq x), x \geq 0$, such that $\Pr(X > 0) > 0$ and $0 < \mu := EX < \infty$. The survival function of X is $\overline{F}(x) := \Pr(X > x) = 1 - F(x)$, $x \geq 0$. The left-continuous version of the survival function is $\overline{F}(x^-) := \Pr(X \geq x)$, $x \geq 0$. In Markov-type inequalities for a distribution that may have atoms, using $\overline{F}(x^-)$ leads to a sharp bound (for which the upper bound is achievable), while using $\overline{F}(x)$ leads to a tight bound (the best bound but not achievable). Since $\overline{F}(x) \leq \overline{F}(x^-)$, it is esthetically preferable to upper bound the larger quantity, since the bound cannot be improved for the smaller quantity.

Markov's inequality says that

$$\overline{F}(x^-) \leq \min\left(1, \frac{\mu}{x}\right) \text{ for all } x \geq 0. \tag{1}$$

The bound is sharp for every $x \geq 0$. For $x \leq \mu$, the rv $X \equiv \mu$ has $\overline{F}(x^-) = 1$, so 1 is a sharp upper bound, and for $x > \mu$ the bound μ/x is sharp and non trivial in that $\mu/x < 1$. What are the points x, if any, such that $x \geq \mu$ and $\overline{F}(x^-) = \mu/x$?

We say X has the distribution of a Bernoulli rv $B(p)$ with parameter p if $\Pr(X = 1) = p = 1 - \Pr(\mathrm{X} = 0)$. To satisfy the assumption that $\Pr(X > 0) > 0$, we require that $0 < p \leq 1$. Let B denote the class $\{cB(p) : c > 0, 0 < p \leq 1\}$ of scaled Bernoulli rvs. We define a distribution on $[0, \infty)$ with $0 < \mu < \infty$ to be regular if it does not belong to B. The main result of this section is:

Theorem 1. *Define* $a(F) := \sup\{x\overline{F}(x^-) : x \in [0,\infty\}$.

(i) *If F is regular, then* $x\overline{F}(x^-) \le a(F) < \mu$ *for all* $x \in [0,\infty)$. *Thus the upper bound in Markov's inequality is not attained for any* $x \ge \mu$. *Moreover,*

$$\overline{F}(x^-) \le \frac{a(F)}{x} < \min\left(1, \frac{\mu}{x}\right) \text{ for all } x > a(F). \tag{2}$$

The improvement compared to Markov's inequality in (2) is $1 - a(F)/x$ *for* $a(F) < x \le \mu$ *and is* $(\mu - a(F))/x$ *for* $x \ge \mu$.

(ii) *If F is not regular, i.e., if* $F \overset{d}{=} cB(p)$ *for some* $c > 0,\ p \in (0,1]$ *(where* $\overset{d}{=}$ *means "has the same distribution as"), then* $x\overline{F}(x^-) < \mu$ *for all* $x \ge 0,\ x \ne c$. *For* $x = c$, $c\overline{F}(c^-) = \mu = cp$. *The upper bound is attained at the single point* $x = c$.

Proof. First we show that if F is regular, then $a(F) < \mu$. For fixed $x_0 > 0$, define $Y(x_0) := X - x_0 I(X \ge x_0)$. Then $Y(x_0) \ge 0$ and $EY(x_0) = \mu - x_0\overline{F}(x_0^-)$. A non negative rv equals 0 with probability 1 if and only if its mean is 0. Thus $\mu = x_0\overline{F}(x_0^-)$ if and only if (iff) $EY(x_0) = 0$ iff $Y(x_0) = 0$ with probability 1 iff $X = x_0 I(X \ge x_0)$ with probability 1 iff $X \overset{d}{=} x_0 B(\overline{F}(x_0^-))$. But because we assume F is regular, $X \notin B$. Thus for all $x \ge 0$, $EY(x) > 0$ and $\mu > x\overline{F}(x^-)$. If $X \overset{d}{=} cB(p)$, then $EY(x) = 0$ only at $x = c$, so $x\overline{F}(x^-) < \mu$ for $x \ne c$ and $c\overline{F}(c^-) = \mu$.

Because F is regular implies $x\overline{F}(x^-) < \mu$, we have from the definition $a(F) := \sup\{x\overline{F}(x^-) : x \in [0,\infty)\}$ that $a(F) \le \mu$. We now rule out the possibility that $a(F) = \mu$. Define $g(x) := x\overline{F}(x^-)$. Then $g(y^-) := \lim_{x\uparrow y}\ g(x) = g(y)$ and $g(y^+) := \lim_{x\downarrow y}\ g(x) = y\overline{F}(y) \le y\overline{F}(y^-) = g(y)$. Thus $\overline{\lim}_{x\to y} g(x) = g(y)$ so g is upper semi-continuous. (We recall that a function $g : [a,b] \to [-\infty,\infty)$, where $[a,b]$ is a closed interval, is upper semi-continuous at $y \in [a,b]$ iff $g(y) \ne +\infty$, $\overline{\lim}_{x\to y} g(x) \le g(y)$.)

As $\mu = \int_0^\infty \overline{F}(x^-)dx < \infty$, integration by parts implies that $\lim_{x\to\infty} g(x) = \lim_{x\to\infty} x\overline{F}(x^-) = 0$. Choose T such that $\sup_{x\ge T} g(x) < a(F)/2$. Then $\sup_{0\le x\le T} g(x) = \sup_{0\le x<\infty} g(x) = a(F)$. Thus $a(F)$ is the supremum of g, an upper semi-continuous function on the finite closed interval $[0,T]$. A generalization of the Bolzano–Weierstrass theorem to upper semi-continuous functions (Royden 1968, p. 161, Prop. 10) insures that for some x in the interval $[0,T]$, $g(x) = a(F)$. Since $g(x) = x\overline{F}(x^-) < \mu$, it follows that $a(F) < \mu$. □

Example 1. If X is exponentially distributed with mean 1, then $\overline{F}(x^-) = e^{-x},\ x \ge 0$. Then $\sup_{x\ge0} x\overline{F}(x^-) = \sup_{x\ge0} xe^{-x} = e^{-1}$ and the supremum is achieved at $x = 1$. Thus $a(F) = e^{-1}$ and $\overline{F}(x^-) \le (ex)^{-1} < \min(1, 1/x)$ for $x > e^{-1}$.

Example 2. If $n > 1$ and $\Pr\{X = n/j\} = n^{-1},\ j = 1,\dots,n$, then

$$x\overline{F}(x^-) = \begin{cases} x, & 0 \le x \le 1, \\ \dfrac{nx}{n-k}, & \dfrac{n}{n-k+1} < x \le \dfrac{n}{n-k}, k = 1,\dots,n-1, \\ 0, & x > n. \end{cases}$$

Here $a(F) = 1$ and $x\overline{F}(x^-) = a(F)$ at each of n points $x = n/j,\ j = 1, \ldots, n$. Since $EX = \mu_n = \sum_{j=1}^{n} j^{-1}$, for large n we have $\mu_n \approx \log n$, which is significantly greater than $a(F) = 1$. Then $\overline{F}(x^-) \le x^{-1} < \min(1, \mu_n/x)$ for $x > 1$. This example shows that, for any n, it is possible for $x\overline{F}(x^-)$ to equal $a(F)$ at n distinct points, by contrast with the case when $X \in B$, where $x\overline{F}(x^-) = \mu$ at at most one point, and the case when X is regular, where $x\overline{F}(x^-) = \mu$ cannot hold at any x.

3. Markov's inequality through renewal theory

Assume that the non negative rv X has cdf F with survival function $\overline{F}$ satisfying $\overline{F}(0) > 0$, $\overline{F}(0^-) = 1$, and $0 < \mu := EX < \infty$. Define $p(t) := t\overline{F}(t^-)/\mu$. Markov's inequality is equivalent to $p(t) \le 1$ for all $t \ge 0$. If, for every such F and for all $t \ge 0$, we can construct events $\{A_t : t \ge 0\}$ such that $\Pr(A_t) = p(t)$, then we have an alternative proof of Markov's inequality. The mathematical challenge is to construct such $\{A_t : t \ge 0\}$.

In a stationary renewal process with interarrival-time cdf F, let G be the cdf of forward and backward recurrence times. G is known as the equilibrium renewal distribution and is absolutely continuous with probability density function (pdf) $g(x) = \overline{F}(x)/\mu, x \ge 0$.

Suppose rv X has cdf F and rv Y has cdf G and X and Y are independent. Define rv L by its distribution conditional on $Y = x$:

$$L|(Y = x) \overset{d}{=} X|(X > x) \quad \text{for all } x \text{ with } \overline{F}(x) > 0.$$

L is well defined since $g(x) > 0$ iff $\overline{F}(x) > 0$. Then $L \overset{d}{=} X|(X > Y)$ and

$$\Pr\{L \ge t | Y = x\} = \begin{cases} 1, & t \le x, \\ \dfrac{\overline{F}(t^-)}{\overline{F}(x)}, & t > x. \end{cases}$$

Theorem 2. *Define $A_t := \{Y < t \le L\}$. Then $\Pr(A_t) = p(t) = t\overline{F}(t^-)/\mu$ for all $t \ge 0$ and any cdf F with survival function $\overline{F}$ satisfying $\overline{F}(0) > 0$, $\overline{F}(0^-) < 1$, and $0 < \mu := EX < \infty$.*

Proof.

$$\Pr(A_t) = \int_{x=0}^{t} g(x)\Pr\{L \ge t | Y = x\} dx = \int_{x=0}^{t} \frac{\overline{F}(x)}{\mu}\, \frac{\overline{F}(t^-)}{\overline{F}(x)} dx = \frac{t\overline{F}(t^-)}{\mu} = p(t).$$

□

If F is regular, Theorems 1 and 2 imply that $\sup_{t \ge 0} \Pr\{Y < t \le L\} \le a(F)/\mu < 1$.

When F is non arithmetic, i.e., when not all of the probability density falls on an arithmetic progression of points, then the backward and forward recurrence times converge in distribution as $t \to \infty$ (Feller 1971, p. 370). When F is arithmetic, the backward and forward recurrence times have a stationary distribution but do not converge in distribution, analogous to the situation in a finite-state irreducible periodic Markov chain (Thorisson 2000, Section 9.2).

The rv L is the distribution of the length of the renewal interval covering a fixed point for a stationary renewal distribution. If L has survival function $\overline{H}(t) := \Pr\{L \geq t\},\ t \geq 0,$ then

$$\overline{H}(t) = \overline{G}(t) + t\overline{F}(t)/\mu = \overline{G}(t) + t\overline{F}(t^-)/\mu$$

(Brown 2006, (3.3)), where the second equality holds by continuity. Therefore

$$p(t) := t\overline{F}(t^-)/\mu = \overline{H}(t) - \overline{G}(t).$$

In this case, $\Pr\{Y < t \leq L\} = \Pr\{L \geq t\} - \Pr\{Y \geq t\}$ because $\Pr\{Y \leq L\} = 1$. The probability $\Pr\{Y < t \leq L\}$ depends on the joint distribution of Y and L, not just on the marginal distributions, as this example will show.

Example 3. Let U, V be independent rvs with $\Pr\{U = 1\} = \frac{1}{2} = \Pr\{U = 2\}$ and $\Pr\{V = 1\} = \frac{3}{10} = 1 - \Pr\{V = 2\}$. Then U is stochastically smaller than V in the sense that, for all $x \geq 0$, $\Pr\{U \leq x\} \geq \Pr\{V \leq x\}$. However, $\Pr\{U \leq V\} < 1$ and $\Pr\{U < 2 \leq V\} = \frac{1}{2} \cdot \frac{7}{10} = 0.35 \neq \Pr\{V \geq 2\} - \Pr\{U \geq 2\} = \frac{7}{10} - \frac{1}{2} = 0.2$.

4. Markov's inequality given only the sample size and the sample average

Suppose $n > 1$ and $x_1, \ldots,\ x_n$ are non negative numbers with at least one x_i strictly positive. Suppose we know only n and the sum $s := x_1 + \ldots +\ x_n$ or the sample mean $\overline{x}_n := s/n$. Since at least one x_i is assumed to be strictly positive, we have $s > 0,\ \overline{x}_n > 0$. We establish an upper bound on the left-continuous version of the empirical survival function, $\overline{F}_n{}^\wedge(x^-) := \#\{X_i \geq x : i = 1, \ldots,\ n\}/n,\ x \geq 0$. Markov's inequality with $\overline{F} = \overline{F}_n{}^\wedge, \mu = \overline{x}_n$ gives $\overline{F}_n{}^\wedge(x^-) \leq \min\left(1, \frac{\overline{x}_n}{x}\right)$ for all $x \geq 0$. In this section, we show that the additional information, beyond n and $\overline{x}_n$, that $\overline{F}_n{}^\wedge$ is atomic and the size of the atoms are multiples of $1/n$ yields a sharp upper bound on $\overline{F}_n{}^\wedge(x^-)$ for all $x \geq 0$. This upper bound improves on Markov's inequality for this case.

For $s > 0$, define $A(s) := \{(x_1, \ldots,\ x_n) :\ x_i \geq 0,\ i = 1, \ldots, n$ and $x_1 + \ldots +\ x_n = s\}$.

Lemma. *Define $M_n(x, s)$ to be the maximum of the number of sample members x_i greater than or equal to x as $(x_1, \ldots,\ x_n)$ ranges over all samples of n non negative numbers with sample sum s. Then*

(i) $M_n(x, s) = n$ *if* $0 \leq x \leq \overline{x}_n$.
(ii) *For* $k = 1, \ldots, n - 1$, $M_n(x, s) = k$ *if* $\frac{s}{k+1} < x \leq \frac{s}{k}$.
(iii) $M_n(x, s) = 0$ *if* $s < x$

Proof. (i) For $x \leq \overline{x}_n$, in the particular case $x_i = \overline{x}_n,\ i = 1, \ldots, n,$ we have all sample members greater than or equal to x, which implies $M_n(x, s) = n$.

(ii) For $x \leq \frac{s}{k}$, in the particular case $x_i = s/k,\ i = 1, \ldots, k,$ and $x_j = 0,\ j = k + 1, \ldots, n,$ we have $\sum_{i=1}^n x_i = s$ and exactly k sample members greater than or equal to x, thus $M_n(x) \geq k$. For $\frac{s}{k+1} < x$, since $s < (k + 1)x$, we must have at most k elements greater than or equal to x for all non negative samples $(x_1, \ldots,\ x_n)$ with sum s, and thus $M_n(x, s) \leq k$. Consequently, for $\frac{s}{k+1} < x \leq \frac{s}{k}$, $M_n(x, s) = k$.

(iii) For $s < x$, it is impossible for any non negative number x_i to exceed the sum $x_1 + \ldots + x_n = s$ of a set of non negative numbers of which it is one member, so $M_n(x, s) = 0$. $\square$

$M_n(x, s)/n$ is the left-continuous version of the survival function of a discrete uniform distribution on the n points $\frac{s}{k}$, $k = 1, \ldots, n$. The Lemma implies that

$$\frac{M_n(x, s)}{n} = I\{0 \le x \le s\} + \sum_{k=1}^{n-1} \frac{k}{n} I\left\{\frac{s}{k+1} < x \le \frac{s}{k}\right\}, x \ge 0.$$

Since $M_n(x, s)$ is a sharp upper bound for the number of members that are greater than or equal to x of any non negative sample $(x_1, \ldots, x_n)$ with sum s, it follows that $M_n(x)/n$ is a sharp upper bound for $\overline{F}_n{}^{\wedge}(x^-)$. As $\min(1, \overline{x}_n/x)$ is an upper bound for $\overline{F}_n{}^{\wedge}(x^-)$, it follows that $M_n(x, s)/n \le \min(1, \overline{x}_n/x)$. Theorem 3 compares the two bounds $M_n(x, s)/n$ and $\min(1, \overline{x}_n/x)$ for $\overline{F}_n{}^{\wedge}(x^-)$.

Theorem 3. *Define $\overline{D}_n(x^-) := \min(1, \overline{x}_n/x) - M_n(x, s)/n$. Then $0 \le \overline{D}_n(x^-) < n^{-1}$ for all $x \ge 0$ and $\sup_{x \ge 0} \overline{D}_n(x^-) = n^{-1}$ but $\overline{D}_n(x^-)$ never equals n^{-1}.*

Proof. For $x \le \frac{s}{n}$, $\overline{D}_n(x^-) = 0$. For $x = s/k$, $k = 1, \ldots, n$, $\overline{D}_n(x^-) = 0$. For $k = 1, \ldots, n-1$ and $\frac{s}{k+1} < x < \frac{s}{k}$, $\overline{D}_n(x^-) > 0$ and $\overline{D}_n(x^-) \to 0$ as $x \uparrow \frac{s}{k}$, $\overline{D}_n(x^-) \to n^{-1}$ as $x \downarrow \frac{s}{k+1}$. For $x > s$, $\overline{D}_n(x^-) > 0$ and $\overline{D}_n(x^-) \to 0$ as $x \uparrow \infty$, $\overline{D}_n(x^-) \to n^{-1}$ as $x \downarrow s$. But $\overline{D}_n(x^-) \ne n^{-1}$ for all $x \ge 0$. $\square$

$F_n^{\wedge}$ is regular (in the sense of Section 2) iff for some $i \ne j$, $0 < x_i < x_j$, i.e., iff $\{x_1, \ldots, x_n\}$ contains at least two distinct positive values. If $F_n^{\wedge}$ is regular, then by Theorem 1, $\overline{F}_n{}^{\wedge}(x^-) < \min(1, \overline{x}_n/x)$ for all $x > \overline{x}_n$. But it can happen that $\overline{F}_n{}^{\wedge}(x^-) = M_n(x, s)/n$.

Example 4. Let $n = 4$, $s = 16$, $(x_1, x_2, x_3, x_4) = (0, 4, 6, 6)$. Since $M_n(x, s)/n = \frac{1}{2}$ for $x \in (\frac{16}{3}, 8]$, we have $M_n(6^-, 16)/n = \frac{1}{2} = \overline{F}_n{}^{\wedge}(6^-)$ while at $x = 6^-$, $\min(1, \overline{x}_n/x) = \min\left(1, \frac{16}{4 \cdot 6}\right) = \frac{2}{3}$, and $(0, 4, 6, 6)$ gives a regular $F_n^{\wedge}$.

Example 5. Suppose that $n = 50$ SAT scores graded on a scale of 200–800 had a mean of $s/n = 516$, so that $s = 25,800$. Consider a hypothetical score of $x = 1000$. Since the sum of all 50 scores is $s = 25,800$, at most 25 scores could equal 1000. From the lemma, as $s/26 < x = 1000 \le s/25$, we have $k = 25, M_n(1000^-)/n = 25/50 = 1/2$ while $\overline{F}_n{}^{\wedge}(1000^-) = 0$ since all scores are in $[a, b] = [200, 800]$.

This example suggests that there is room to improve the upper bound $M_n(x, s)/n$ when it is known that each $x_i \in [a, b], 0 \le a < b \le \infty, i = 1, \ldots, n$.

If $b = \infty$, so that we know only that each $x_i \ge a \ge 0$, let us assume $x_i > a$ for at least one i. Then we define

$$y_i := x_i - a, \ \ s(y) := s - na, \ \ \overline{y}_n := \overline{x}_n - a.$$

Using these y_i instead of x_i in $M_n(x, s)/n$ gives a sharp upper bound for the left-continuous empirical survival function of $y_1, \ldots, y_n$. We translate the intervals in the lemma by a to obtain a sharp upper bound for $\overline{F}_n{}^{\wedge}(x^-)$ corresponding to $x_1, \dots, x_n$. That bound is

$$\overline{A}_n(x^-) = \begin{cases} 1, & 0 \le x \le \overline{x}_n, \\ \dfrac{k}{n}, & a + \dfrac{s-na}{k+1} < x \le a + \dfrac{s-na}{k}, k = 1, \dots, n-1, \\ 0, & x > s-(n-1)a. \end{cases}$$

For a general a, b, the sharp upper bound is

$$\overline{B}_n(x^-) = \begin{cases} \overline{A}_n(x^-), & x \le b, \\ 0, & x > b. \end{cases}$$

Example 6. Continuing Example 5, suppose $a = 200$, $b = \infty$ in the SAT example above. Then $\overline{A}_n(1000^-) = 0.38$, a nice improvement over $M_n(x, s)/n = 0.5$. Since $1000 > b$, $\overline{B}_n(1000^-) = 0$. When we know that $b = 800$ so that $x = 600 \in [200, 800]$, then $M_n(x, s)/n = 0.86$ while $\overline{A}_n(600^-) = \overline{B}_n(600^-) = 0.78$.

5. Markov's inequality in reliability theory: NBUE and DMRL

Marshall and Olkin (2007, pp. 198–215, Section 6B) reviewed upper bounds for $\overline{F}(t^-)$ when given more information about the distribution F than its mean μ. Such additional information includes a description of the pattern of aging associated with the survival curve $\overline{F}$.

For example, a cdf F on $[0, \infty)$ with finite mean μ is defined to be "new better than used in expectation" (NBUE) iff

$$E(X - t \mid X > t) \le EX \quad \text{for all } t \ge 0 \quad \text{with } \overline{F}(t) > 0.$$

NBUE is equivalent to $F \ge_{st} G$ where G, the equilibrium cdf corresponding to F, has, by definition, pdf at $t \ge 0$ given by $\overline{F}(t)/\mu$. A rv that is NBUE deteriorates weakly with age. Subclasses of NBUE include new better than used, increasing failure rate, and decreasing mean residual life (DMRL) (Barlow and Proschan 1975; Marshall and Olkin 2007).

5.1. NBUE distributions

For NBUE distributions, Marshall and Olkin (2007) stated a bound of Brown (2006, p. 206),

$$\overline{F}(t) \le \min\{1, \exp\left(-\left[\frac{t}{\mu} - 1\right]\right)\}, \tag{3}$$

which is informative for $t > \mu$, and asked whether that bound is sharp. Here we show that it is sharp.

Theorem 4. *For every $t \geq 0$, a NBUE distribution F exists such that*

$$\overline{F}(t) = \min\left\{1, \exp\left(-\left[\frac{t}{\mu} - 1\right]\right)\right\}.$$

Proof. For $t \leq \mu$ let $X \equiv \mu$. This distribution is NBUE as $E(X - t \mid X > t) = \mu - t$ for $t < \mu$. Moreover, $\overline{F}(t^-) = 1$ for $0 \leq t \leq \mu$. Thus the bound 1 is sharp.

For every $t > \mu$, we produce a distribution that achieves the upper bound for that t. The upper bound is the value of the survival function of an exponential distribution with mean μ at $t - \mu$. The idea of the construction is that, if we want $\overline{F}(t^-)$ to equal $\min\{1, \exp\left(-\left[\frac{t}{\mu} - 1\right]\right)\}$, we gain nothing by allowing $\overline{F}(t) > 0$. So we assign F the pdf $f(x) = \mu^{-1}e^{-x/\mu}$ on $[0, t-\mu)$ and $f(x) = 0$ on $[t-\mu, t)$. Then, for a chosen t,

$$\overline{F}(x) = \begin{cases} \overline{G}(x) = e^{-x/\mu}, & \text{for } 0 \leq x \leq t - \mu, \\ \exp\left(-\left[\frac{t}{\mu} - 1\right]\right), & \text{for } t - \mu \leq x < t, \end{cases}$$

with $\overline{F}(t^-) = \exp\left(-\left[\frac{t}{\mu} - 1\right]\right)$, and $\overline{F}(x) = 0$ for $x > t$. Further,

$$\overline{G}(x) = \begin{cases} \frac{t-x}{\mu}\exp\left(-\left[\frac{t}{\mu} - 1\right]\right), & \text{for } t - \mu \leq x < t, \\ 0, & \text{for } x \geq t. \end{cases}$$

As $(t - x)/\mu$ goes from 1 to 0 as x goes from $t - \mu$ to t, we see that $F \geq_{st} G$ and F is NBUE. Also,

$$E_F X = \int_{x=0}^{t-\mu} \mu^{-1}e^{-x/\mu}dx + \int_{x=t-\mu}^{t} \exp\left(-\left[\frac{t}{\mu} - 1\right]\right)dx = \mu.$$

Thus F is NBUE, has mean μ and satisfies (3) with equality for the chosen t.

5.2. DMRL distributions

A distribution F on $[0, \infty)$ with $F(0) = 0$ is DMRL if $E(X - x \mid X > x)$ is non increasing in $x \geq 0$ for all x with $\overline{F}(x) > 0$. DMRL is a subclass of NBUE because $E(X - x \mid X > x) \leq E(X - 0 \mid X > 0) = EX$. The cdf F constructed in the proof of Theorem 4 is not only NBUE but is also DMRL because

$$E(X - x \mid X > x) = \begin{cases} \mu, & \text{for } 0 \leq x \leq t - \mu, \\ t - x, & \text{for } t - \mu \leq x < t. \end{cases}$$

Therefore this F also demonstrates that (3) is sharp for DMRL distributions.

In defining DMRL, we required $F(0) = 0$. To see why, consider rv X with $\overline{F}(t) = qe^{-t}$ for $t \geq 0$, $0 < q < 1$. For all $t \geq 0$, $E(X - t \mid X > t) = \mu := EX$, which is non increasing in t. But since $\overline{F}(0) = q < 1$, $F(0) = p := 1 - q > 0$, so this X does not satisfy the definition, and the bound (3) does not hold because (since $\mu = q$)

$$\frac{qe^{-t}}{\exp\left(-\left[\frac{t}{q}-1\right]\right)} = qe^{-1}\exp\left(\frac{pt}{q}\right) \to \infty \ \ as \ \ t \to \infty.$$

Thus for all sufficiently large t,

$$\overline{F}(t) > \exp\left(-\left[\frac{t}{q}-1\right]\right).$$

Likewise, if F is NBUE, then $F(0)=0$, since if $F(0)>0$, then $E(X-0\ |X>0) = \mu/\overline{F}(0) > \mu$ and F is not NBUE.

Acknowledgments

JEC thanks Roseanne Benjamin for help during this work. Both authors thank the anonymous referee for helpful comments.

ORCID

Joel E. Cohen http://orcid.org/0000-0002-9746-6725

References

Barlow, R. E., and F. Proschan. 1975. *Statistical theory of reliability and life testing*. New York: Holt.

Brown, M. 2006. Exploiting the waiting time paradox: Applications of the size-biasing transformation. *Probability in the Engineering and Informational Sciences* 20 (2):195–230. doi:10.1017/S026996480606013X.

Cohen, J. E. 2015. Markov's inequality and Chebyshev's inequality for tail probabilities: A sharper image. *The American Statistician* 69 (1):5–7. doi:10.1080/00031305.2014.975842.

Feller, W. 1971. *An introduction to probability theory and its applications*. Vol. 2, 2nd ed. New York: Wiley.

Ghosh, B. K. 2002. Probability inequalities related to Markov's theorem. *The American Statistician* 56 (3):186–90. doi:10.1198/000313002119.

Marshall, A. W., and I. Olkin. 2007. *Life distributions: Structure of nonparametric, semiparametric, and parametric families*. New York: Springer.

Royden, H. L. 1968. *Real analysis*. 2nd ed. New York: Macmillan.

Thorisson, H. 2000. *Coupling, stationarity, and regeneration*. New York: Springer.