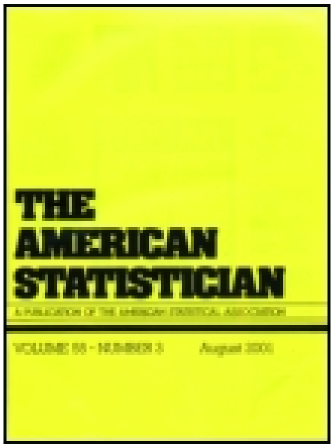
This article was downloaded by: [JOEL Cohen]
On: 27 March 2015, At: 11:28
Publisher: Taylor & Francis
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Click for updates

The American Statistician
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/utas20

Markov's Inequality and Chebyshev's Inequality for Tail Probabilities: A Sharper Image

Joel E. Cohen

Accepted author version posted online: 22 Oct 2014.Published online: 24 Mar 2015.

To cite this article: Joel E. Cohen (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

To link to this article: http://dx.doi.org/10.1080/00031305.2014.975842

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the “Content”) contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at http://www.tandfonline.com/page/terms-and-conditions

Markov's Inequality and Chebyshev's Inequality for Tail Probabilities: A Sharper Image

Joel E. COHEN

Markov's inequality gives an upper bound on the probability that a nonnegative random variable takes large values. For example, if the random variable is the lifetime of a person or a machine, Markov's inequality says that the probability that an individual survives more than three times the average lifetime in the population of such individuals cannot exceed one-third. Here we give a simple, intuitive geometric interpretation and derivation of Markov's inequality. These results lead to inequalities sharper than Markov's when information about conditional expectations is available, as in reliability theory, demography, and actuarial mathematics. We use these results to sharpen Chebyshev's tail inequality also.

KEY WORDS: Demography; Lifetime; Reliability; Remaining life expectancy; Survival curve.

1. INTRODUCTION

In reliability theory, mathematical demography (Keyfitz 1968), and probability theory generally (Feller 1971; Loève 1977; Janson et al. 2000; Haccou et al. 2005), Markov's inequality and Chebyshev's inequality (Ghosh 2002, Steele 2004, Haccou et al. 2005) give upper bounds on the probability that a nonnegative random variable takes large values. Here we derive Markov's inequality in simple ways, sharpen it by using information about conditional expectations, and interpret it geometrically. We use these results to sharpen Chebyshev's inequality also.

Let X be a random variable that takes nonnegative real values x in $\mathbb{R}_+ = [0, \infty)$. For x in $\mathbb{R}_+$, let $F(x) := \text{Prob}(X \leq x)$ be the cumulative distribution function (cdf) of X, and let $S(x) := 1 - F(x) = \text{Prob}(X > x)$ be the complement of the cumulative distribution function, or the survival function, of X. If X is lifetime, then $S(x)$ gives the probability of surviving to at least age x. Assuming no atom of probability at $x = 0$, we have $F(0) = 0$, $S(0) = 1$, $F(\infty) := \lim_{x\to\infty} F(x) = 1$, $S(\infty) := \lim_{x\to\infty} S(x) = 0$. The average, expectation, or expected value of X is defined as $\mathbb{E}X := \int_0^\infty x\,dF(x)$. Henceforth we assume that $\mathbb{E}X < \infty$ and that X is continuous, meaning that its probability density function $f(x) = F'(x)$ is absolutely continuous (Hong 2012). Then, $\mathbb{E}X = \int_0^\infty x f(x)dx$.

Markov's inequality states that $S(x) \leq \mathbb{E}X/x$ for all x in $\mathbb{R}_+$. Ghosh (2002) established the sharpest statements one can make about $\text{Prob}(X \geq x)$ for $-\infty < x < \infty$ for all distributions of X with $\mathbb{E}X = \mu > 0$, $\text{var}(X) = \sigma^2 > 0$, and $\text{Prob}(X \geq 0) = 1$. Here we assume nothing about the existence or positivity of the variance except in Section 5 on Chebyshev's inequality. Unlike Ghosh (2002), our purpose here is to sharpen Markov's inequality by using information about conditional expectations of the sort likely to be encountered in actuarial mathematics, demography, and reliability theory. Our results serve a different purpose from those of Ghosh (2002), who also gives useful references to much prior literature.

2. MAIN RESULT: ABC THEOREM

To prove and improve Markov's inequality, we decompose $\mathbb{E}X$ in the following ABC theorem. Let $\mathbb{E}(X|X \leq x)$ be the conditional expectation of X given $X \leq x$. When X is the lifetime, $\mathbb{E}(X|X \leq x)$ is the average lifetime of individuals who die at any age up to and including x. Let $\mathbb{E}(X - x|X > x)$ be the conditional expectation of X–x given $X > x$. When X is the lifetime, $\mathbb{E}(X - x|X > x)$ is the average remaining duration of life beyond age x of individuals who have just attained age x.

ABC Theorem. For all x in $\mathbb{R}_+$,

$$\mathbb{E}X = F(x)\mathbb{E}(X|X \leq x) + xS(x) + S(x)\mathbb{E}(X - x|X > x).$$

If $A = F(x)\mathbb{E}(X|X \leq x)$, $B = xS(x)$, $C = S(x)\mathbb{E}(X - x|X > x)$, then $\mathbb{E}X = A + B + C$.

A simple but intuitively opaque proof follows quickly from the definition of conditional expectation: $\mathbb{E}X = F(x)\mathbb{E}(X|X \leq x) + S(x)\mathbb{E}(X|X > x) = F(x)\mathbb{E}(X|X \leq x) + S(x)\mathbb{E}(X - x + x|X > x)$. Since the expectation is linear in its arguments, $\mathbb{E}(X - x + x|X > x) = \mathbb{E}(X - x\,|\,X > x) + \mathbb{E}(x|X > x) = \mathbb{E}(X - x|X > x) + x$. Substituting into the previous expression gives the theorem.

The ABC theorem has a nice geometric interpretation and derivation. A standard formula (proved in two ways by Hong (2012)) for the expectation of a nonnegative continuous random variable X is that

$$\mathbb{E}X = \int_0^\infty S(x)dx.$$

Joel E. Cohen is Professor of Populations, Laboratory of Populations, Rockefeller University and Columbia University, 1230 York Ave., Box 20, New York, NY 10065 (E-mail: *cohen@rockefeller.edu*). This work was partially supported by the U.S. National Science Foundation grants EF-1038337 and DMS-1225529. The author thanks the editor and referees for very helpful comments.

© 2015 American Statistical Association DOI: 10.1080/00031305.2014.975842

Downloaded by [JOEL Cohen] at 11:28 27 March 2015

We use this formula to derive integral expressions for $\mathbb{E}(X|X \leq x)$ and $\mathbb{E}(X - x|X > x)$. The event $\{X \leq x\}$ has probability $F(x)$. Hence, for $0 < a \leq x$, the cdf of X at a conditional on $\{X \leq x\}$ is $F(a)/F(x)$ and the survival function of X at a conditional on $\{X \leq x\}$ is $1 - F(a)/F(x)$. Applying the formula above to X conditional on $\{X \leq x\}$ then gives

$$\mathbb{E}(X|X \leq x) = \int_0^x \left(1 - \frac{F(a)}{F(x)}\right) da = \int_0^x \frac{[F(x) - F(a)]}{F(x)} da$$
$$= \frac{1}{F(x)} \int_0^x [S(a) - S(x)]\, da.$$

Likewise, the event $\{X > x\}$ has probability $S(x)$. Hence for $x < a < \infty$, the cdf of $X - x$ at $a - x$ conditional on $\{X > x\}$ is $(F(a) - F(x))/S(x)$ and the survival function of $X - x$ at $a - x$ conditional on $\{X > x\}$ is $1 - (F(a) - F(x))/S(x) = (S(x) - F(a) + F(x))/S(x) = (1 - F(a))/S(x) = S(a)/S(x)$. By the same formula for $\mathbb{E}X$ applied to $X - x$ conditional on $\{X > x\}$, we have

$$\mathbb{E}(X - x|X > x) = \frac{1}{S(x)} \int_x^\infty S(a)\, da.$$

This formula is well known in actuarial mathematics (e.g., Bowers et al. 1997, p. 68, their eq. (3.5.2)) and demography (e.g., Keyfitz 1968, p. 6).

We now decompose $\mathbb{E}X$. For all x in $\mathbb{R}_+$,

$$\begin{aligned}
\mathbb{E}X &= \int_0^\infty S(a)\, da = \int_0^x S(a)\, da + \int_x^\infty S(a)\, da \\
&= F(x) \cdot \frac{1}{F(x)} \int_0^x S(a)\, da + S(x) \cdot \frac{1}{S(x)} \int_x^\infty S(a)\, da \\
&= F(x) \cdot \frac{1}{F(x)} \int_0^x [S(a) - S(x) + S(x)]\, da \\
&\quad + S(x)\, \mathbb{E}(X - x|X > x) \\
&= F(x) \int_0^x \frac{[S(a) - S(x)]}{F(x)} da \\
&\quad + S(x) \int_0^x da + S(x)\, \mathbb{E}(X - x|X > x) \\
&= F(x)\, \mathbb{E}(X|X \leq x) + x \cdot S(x) + S(x)\, \mathbb{E}(X - x|X > x).
\end{aligned}$$

This expression shows that A, B, and C have geometric interpretations as areas under the survival function $S(x)$:

$$\begin{aligned}
A &= F(x)\mathbb{E}(X|X \leq x) = \int_0^x [S(a) - S(x)]\, da \\
&= \int_0^x S(a) da - x S(x), \\
B &= x S(x), \\
C &= S(x)\mathbb{E}(X - x|X > x) = \int_x^\infty S(a)\, da.
\end{aligned}$$

These three nonnegative quantities are represented by the areas of the regions labeled A, B, and C, respectively, in the numerical example given in Figure 1(a) and their sum is $\mathbb{E}X$. Thus Figure 1(a) illustrates the ABC theorem.

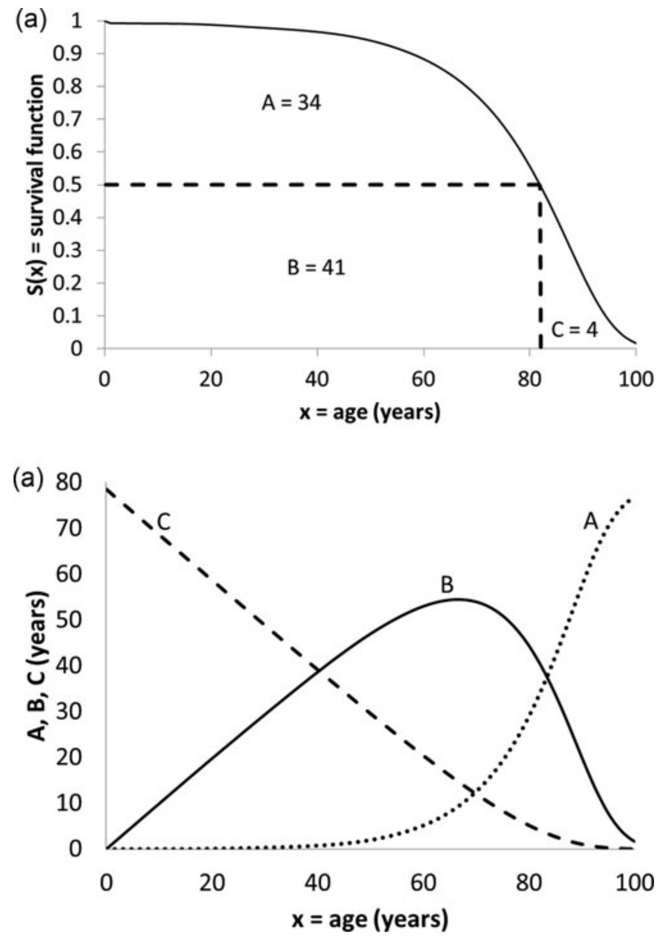


Figure 1. (a). Survival function *S(x)* (solid line) of the total United States human population in 2008 as a function of age *x* (years), and median age *m* at death, $S(m) = 1/2$, $m = 82$. $A = F(x)\mathbb{E}(X|X \leq x)$, $B = xS(x)$, $C = S(x)\mathbb{E}(X - x|X > x)$, and $\mathbb{E}X = A + B + C$. (b) A (dotted line), B (solid line), and C (dashed line) as functions of age x (years) for the total United States human population in 2008. Calculations by author based on data of Arias (2012).

3. DERIVATION OF MARKOV'S AND SHARPER INEQUALITIES

Corollary. $\mathbb{E}X$ is greater than or equal to the sum of any subset of $\{A, B, C\}$ for all x in $\mathbb{R}_+$ (assuming an empty subset has sum zero).

In particular, $\mathbb{E}X \geq B = xS(x)$ for all $x \geq 0$. Dividing this by x gives Markov's inequality $\mathbb{E}X/x \geq \text{Prob}(X > x)$. Although this proof of Markov's inequality is longer than some purely analytical alternatives (e.g., Ghosh 2002), Figure 1(a) gives an easily remembered, intuitively appealing visualization of why Markov's inequality is true: the area under the solid curve $\mathbb{E}X$ is greater than or equal to the area $B = xS(x)$ of the rectangle (with base x and height $S(x)$) bounded by the dashed lines and the axes.

Figure 1(a) also suggests intuitively why $\lim_{x\to\infty} xS(x) = 0$, which Hong (2012) proved rigorously without interpretation. The total area under $S(x)$ from $x = 0$ to $x = \infty$ is the life expectancy, which we assumed to be finite. From Figure 1(a), it seems clear intuitively that as $x \to \infty$, $A \to \mathbb{E}X$, $B = xS(x) \to 0$, and $C \to 0$.

Downloaded by [JOEL Cohen] at 11:28 27 March 2015

Downloaded by [JOEL Cohen] at 11:28 27 March 2015

The corollary implies two sharpenings of Markov's inequality. We propose to call these inequalities Hansel and Gretel (which are the names of characters in a German fairy tale published by the Brothers Grimm in 1812 and dramatized in an opera and a movie) because these siblings, offspring of a simple father, follow an interesting path. From $\mathbb{E}X \geq B + C$ (an inequality previously derived by Cohen 2011), we get Hansel:

$$\frac{\mathbb{E}X}{x + \mathbb{E}(X - x|X > x)} \geq \text{Prob}\,(X > x)\,.$$

From $\mathbb{E}X \geq A + B$, using $F(x) = 1 - S(x)$, we get Gretel:

$$\frac{\mathbb{E}X - \mathbb{E}(X|X \leq x)}{x - \mathbb{E}(X|X \leq x)} \geq \text{Prob}(X > x).$$

The three upper bounds in Markov's inequality, Hansel, and Gretel exceed 1 for some combinations of cdf and x. In such cases, the inequalities give no useful information about $\text{Prob}(X > x)$.

Other paths also lead to Hansel and Gretel. For example, to prove Hansel, observe that $\mathbb{E}X \geq \mathbb{E}(X \cdot 1_{\{X>x\}}) = \mathbb{E}[(X - x + x) \cdot 1_{\{X>x\}}] = \mathbb{E}\,(X - x|X > x)\,P(X > x) + x\,P(X > x)$, equivalent to Hansel.

The remaining four cases of the corollary are less interesting. $\mathbb{E}X \geq A + B + C$ is always an equality. $\mathbb{E}X \geq A$ gives a trivial bound, $S(x) \geq 0$. $\mathbb{E}X \geq C$ yields $\mathbb{E}X/\mathbb{E}(X - x|X > x) \geq \text{Prob}(X > x)$ and the ratio on the left of this inequality may be greater or less than 1. From $\mathbb{E}X \geq A + C$, if $\mathbb{E}(X - x|X > x) - \mathbb{E}(X|X \leq x) > 0$, then

$$\frac{\mathbb{E}X - \mathbb{E}(X|X \leq x)}{\mathbb{E}(X - x|X > x) - \mathbb{E}(X|X \leq x)} \geq \text{Prob}(X > x).$$

But if $\mathbb{E}(X - x|X > x) - \mathbb{E}(X|X \leq x) < 0$, then we get a lower bound on $S(x)$.

4. ILLUSTRATION OF INEQUALITIES BY U.S. LIFETIME

We illustrate Markov's inequality, Hansel and Gretel when X is interpreted as the length of human life in the United States, using data from 2008 (Arias 2012, table 1). The National Center for Health Statistics used death certificates and population estimates to estimate $S(x)$, the fraction of people who would survive from birth to age x according to age-specific death rates in the United States in 2008 (solid descending curve in Figure 1(a)). All the following numerical values are approximate because the number of deaths at ages 100 and over was small. Expectations of remaining life at very advanced ages were not directly measurable. All values here are rounded to the nearest whole year.

Life expectancy at birth in 2008 was the area under the curve $S(x)$. Crudely estimating this quantity by summing the tabulated values of $S(x)$ for the 101 values $x = 0, 1, \ldots, 100$ gives approximately $\mathbb{E}X = 79$ years. (The official life expectancy at birth was 78.1 years.) The median lifetime m is defined as the lifetime at which half of newborns would survive, that is, m is the solution for x of $S(x) = 1/2$. Approximately, $m = 82$ years, which is where the vertical dashed line through the point $S(x) = 1/2$ intersects the x-axis. To illustrate A, B, and C numerically, we set $x = m = 82$ years. Then $B = m{\cdot}S(m) = 82/2 = 41$ years and Markov's inequality becomes $79/82 \approx 0.96 \geq 1/2$. Hansel becomes $79/[82 + 8] \approx 0.88 \geq 1/2$, which is an upper bound on 1/2 that is closer than Markov's inequality provides. Since $A = F(m)\mathbb{E}(X|X \leq m) = 34$, we have $\mathbb{E}(X|X \leq m) = 2 \times 34 = 68$ and Gretel becomes $[79 - 68]/[82 - 68] \approx 0.79 \geq 1/2$, which is still better. At some ages x other than m, it appears numerically that Hansel is better than Gretel.

5. APPLICATIONS TO CHEBYSHEV'S INEQUALITY

Markov's inequality gives a very easy proof of Chebyshev's famous inequality (Ghosh 2002, Steele 2004, p. 86) for tail probabilities. Hansel and Gretel can follow the same path. Let Z be a random variable with finite mean $\mu = \mathbb{E}Z$ and finite variance $\sigma^2 = \mathbb{E}(|Z - \mu|^2)$. Then $X = (Z - \mu)^2$ is nonnegative and $\mathbb{E}X = \sigma^2$. Substituting this X into Markov's inequality gives $\sigma^2/x \geq \text{Prob}\left((Z - \mu)^2 > x\right)$ for all x in $\mathbb{R}_+$. Steele (2004, p. 247, Solution to Exercise 5.11) gives an equivalent proof. Using $X = (Z - \mu)^2$, Hansel and Gretel give, respectively,

$$\frac{\sigma^2}{x + \mathbb{E}((Z - \mu)^2 - x|\,(Z - \mu)^2 > x)} \geq \text{Prob}\left((Z - \mu)^2 > x\right),$$

$$\frac{\sigma^2 - \mathbb{E}(\,(Z - \mu)^2\,|\,(Z - \mu)^2 \leq x)}{x - \mathbb{E}(\,(Z - \mu)^2\,|\,(Z - \mu)^2 \leq x)} \geq \text{Prob}\left((Z - \mu)^2 > x\right).$$

[Received February 2014. Revised September 2014]

REFERENCES

Arias, E. (2012), "United States Life Tables, 2008," in *National Vital Statistics Reports* (Vol. 61, no. 3), Hyattsville, MD: National Center for Health Statistics. [7]

Bowers, N. L., Herber, H. U., Hickman, J. C., Jones, D. A., and Nesbot, C. J. (1997), *Actuarial Mathematics*, Schaumburg, IL: Society of Actuaries. [6]

Cohen, J. E. (2011), "Life Expectancy: Lower and Upper Bounds From Surviving Fractions and Remaining Life Expectancy," *Demographic Research*, 24, 251–256. [7]

Feller, W. (1971), *An Introduction to Probability Theory and Its Applications* (Vol. 2, 2nd ed.), New York: Wiley. [5]

Ghosh, B. K. (2002), "Probability Inequalities Related to Markov's Theorem," *The American Statistician*, 56, 186–190. [5,6,7]

Haccou, P., Jagers, P., and Vatutin, V. A. (2005), *Branching Processes: Variation, Growth, and Extinction of Populations*, Cambridge, UK, and New York: Cambridge University Press. [5]

Hong, L. (2012), "A Remark on the Alternative Expectation Formula," *The American Statistician*, 66, 232–233. [5,6]

Janson, S., Luczak, T., and Rucinski, A. (2000), *Random Graphs*, New York: Wiley. [5]

Keyfitz, N. (1968), *Introduction to the Mathematics of Population*, Reading, MA: Addison-Wesley. [5,6]

Loeve, M. (1977), *Probability Theory* (Vol. 1, 4th ed.), New York: Springer. [5]

Steele, J. M. (2004), *The Cauchy-Schwarz Master Class: An Introduction to the Art of Mathematical Inequalities*, Cambridge, UK, and New York: Cambridge University Press. [5,7]