Demystifying the Integrated Tail Probability Expectation Formula

Ambrose Lo*

Department of Statistics and Actuarial Science, The University of Iowa

241 Schaeffer Hall, Iowa City, IA 52242-1409, USA

Abstract

Calculating the expected values of different types of random variables is a central topic in mathematical statistics. Targeted towards students and instructors in both introductory probability and statistics courses and graduate-level measure-theoretic probability courses, this pedagogical note casts light on a general expectation formula stated in terms of distribution and survival functions of random variables and discusses its educational merits. Often consigned to an end-of-chapter exercise in mathematical statistics textbooks with minimal discussion and presented under superfluous technical assumptions, this unconventional expectation formula provides an invaluable opportunity for students to appreciate the geometric meaning of expectations, which is overlooked in most undergraduate and graduate curricula, and serves as an efficient tool for the calculation of expected values that could be much more laborious by traditional means. For students' benefit, this formula deserves a thorough in-class treatment in conjunction with the teaching of expectations. Besides clarifying some commonly held misconceptions and showing the pedagogical value of the expectation formula, this note offers guidance for instructors on teaching the formula taking the background of the target student group into account.

Keywords: Expected value; Fubini's theorem; Integration by parts; Covariance; Hoeffding's formula

*Ambrose Lo is an Assistant Professor at the Department of Statistics and Actuarial Science, The University of Iowa, Iowa City, IA 52242-1409 (e-mail: ambrose-lo@uiowa.edu). This work was supported by a research fund provided by the Department of Statistics and Actuarial Science, The University of Iowa, and a Centers of Actuarial Excellence (CAE) Research Grant (2018-2021) from the Society of Actuaries (SOA). Any opinions, finding, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the SOA. The author is grateful to the Editor, an Associate Editor, and two anonymous reviewers for their careful reading and constructive comments, and to Joyee Ghosh and Rui Huang at The University of Iowa for stimulating discussions.

1 Introduction

Inculcating students with the ability to calculate the expected values of a wide variety of random variables is one of the key objectives of an introductory mathematical statistics course. Along this line, this pedagogical note centers on the integral expectation formula which, in its simplest form, states that

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X > x)\,\mathrm{d}x \tag{1.1}$$

for any non-negative random variable X. This expectation formula does not seem to have received a standard name in the literature, to the best of the author's knowledge. Because (1.1) exhibits the expectation of a non-negative random variable as the integral of its survival function, a kind of tail probability, for expository convenience we refer to (1.1) as the *integrated tail probability expectation formula*, or the *tail expectation formula* in short. For many students in statistics and allied disciplines, their first encounter with this formula may be in an end-of-chapter exercise of many elementary calculus-based probability and statistics textbooks, often presented under the (in fact, unnecessary) assumption that the random variable X is (absolutely) continuous, see, e.g., Exercise 2.14 (a) on page 78 of Casella and Berger (2002), Exercise 1.9.20 on page 67 of Hogg et al. (2013), and Exercise 4.34 on page 173 of Wackerly et al. (2008). Under such a continuity assumption, (1.1) is typically proved by integration by parts (see, e.g., pages 215 and 216 of Rényi (2007)), which is susceptible to students' misuse (see Hong (2012) for the subtle technical issues when applying integration by parts in this context). Quite often, (1.1) is accompanied by an analogous tail sum formula for non-negative discrete random variables in the form of

$$\mathbb{E}[X] = \sum_{x=0}^{\infty} \mathbb{P}(X > x), \tag{1.2}$$

see, e.g., Exercise 2.14 (b) on page 78 of Casella and Berger (2002), Exercise 1.9.21 on page 67 of Hogg et al. (2013), and Exercise 3.29 on page 99 of Wackerly et al. (2008).

While (1.1) and (1.2) are true under their respective assumptions, the disparate way in which they are commonly presented conceals their interrelationship, encourages the misguided impression that (1.1) is true only for absolutely continuous random variables (see, e.g., Chakraborti et al. (2017)), and in turn undesirably undermines its applicability. In fact, (1.2) is a special case of (1.1) and the latter is a general result that holds true for all non-negative random variables, be their distributions discrete, continuous, or mixed in nature. For the benefit of students, instructors, and researchers alike, this message should be disseminated much more widely. Furthermore, by relegating the tail expectation formula to an end-of-chapter exercise, regrettably few texts discuss the probabilistic interpretation and practical usage of this unconventional expectation formula, which is radically different from the traditional way to calculate expectations by probability density or mass functions. This deprives students of the opportunity to appreciate the genuine meaning and realize the range of applications of the tail expectation formula, which, as this note shows, proves a versatile computational and theoretical tool in relation to expectations.

Motivated by the above two deficiencies in the existing treatment of expectations in mainstream textbooks, this note aims to demonstrate the pedagogical merits of the integrated tail probability expectation formula to undergraduate students in an introductory mathematical statistics course as well as first-year graduate students in a measure-theoretic probability class. The demonstration features the lesser-known geometric interpretation of the formula and its use for efficient calculations of expectations that would be much more arduous otherwise. To begin with, we formulate in Subsection 2.1 the tail expectation formula for general random variables without extraneous continuity assumptions and provide a simple proof that is accessible to even undergraduate students with a modest background in multi-variable calculus, because the main idea of the proof lies in interchanging the order of integration. On the basis of the correctly stated formula, we discuss in Subsection 2.2 how the tail expectation formula can be interpreted in graphical terms. It turns out that the tail expectation formula is amenable to a colorful probabilistic interpretation which furnishes the ideal context for students to visualize expectations. Whereas the traditional method of calculating expectations slices the region bounded by the distribution function of a random variable that represents its expected value horizontally, the tail expectation formula performs the same task vertically. Rarely found in textbooks, such an instructive geometric perspective translates seemingly abstruse algebraic terms into self-explanatory pictures and is instrumental in helping students put the genuine meaning of expectations into perspective, remember the tail expectation formula, and keep it in their statistics toolbox. Applications of the tail expectation formula are given in Subsection 2.3, where it is shown that the formula lends itself to calculating the expectations of a random variable with a mixed distribution and formalizes the relative size of two random variables given the relative order of their distribution or survival functions. As a natural extension, we present in Section 3 a generalization of the tail expectation formula to multiple random variables and an analogous integrated tail probability formula for the covariance between two random variables. An example in the context of the Marshall–Olkin common shock model provides a further illustration of the computational utility of the integrated tail probability (expectation or covariance) formula.

2 Integrated Tail Probability Expectation Formula: Univariate Version

2.1 Statement and Proof

We start by stating the integrated tail probability expectation formula for general random variables, followed by a simple and transparent proof and pertinent discussions.

Proposition 2.1. (Integrated tail probability expectation formula) *For any integrable (i.e., finite-mean) random variable X,*

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X > x)\,\mathrm{d}x - \int_{-\infty}^0 \mathbb{P}(X < x)\,\mathrm{d}x. \tag{2.1}$$

Proof. We first assume that X is a non-negative random variable. The key of the proof lies in writing X, in integral form, as (this is sometimes known as the layer cake representation of a non-negative measurable function; see page 26 of Lieb and Loss (2001))

$$X = \int_0^X \mathrm{d}x = \int_0^\infty 1_{\{X>x\}}\,\mathrm{d}x, \tag{2.2}$$

where 1_A is the indicator function of event A, i.e., $1_A = 1$ if A is true and $1_A = 0$ otherwise. By taking expectation and interchanging the order of expectation and integration (see the remarks at the bottom of page 6), we get

$$\mathbb{E}[X] = \int_0^\infty \mathbb{E}[1_{\{X>x\}}]\,\mathrm{d}x = \int_0^\infty \mathbb{P}(X > x)\,\mathrm{d}x,$$

which is (1.1).

If X is a general random variable, then we consider its positive part $X_+ := \max(X, 0)$ and negative part $X_- := \max(-X, 0)$ so that $X = X_+ - X_-$. Applying (1.1) to X_+ and X_-, which are non-negative, yields

$$\mathbb{E}[X_+] = \int_0^\infty \mathbb{P}(X_+ > x)\,\mathrm{d}x \qquad \text{and} \qquad \mathbb{E}[X_-] = \int_0^\infty \mathbb{P}(X_- > x)\,\mathrm{d}x.$$

For any $x > 0$, $X_+ > x$ is equivalent to $X > x$ and $X_- > x$ is equivalent to $X < -x$, so we further have

$$\mathbb{E}[X_+] = \int_0^\infty \mathbb{P}(X > x)\,\mathrm{d}x \quad \text{and} \quad \mathbb{E}[X_-] = \int_0^\infty \mathbb{P}(X < -x)\,\mathrm{d}x = \int_{-\infty}^0 \mathbb{P}(X < x)\,\mathrm{d}x, \tag{2.3}$$

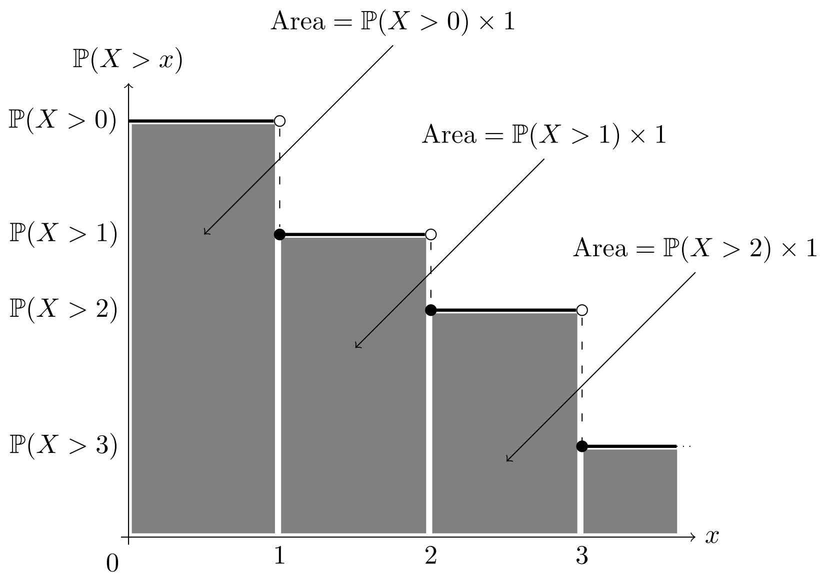


Figure 1: Graphical demonstration of how (1.1) retrieves (1.2) when X is a non-negative and integer-valued random variable.

where the last equality follows from a simple change of variables. By integrability, $\mathbb{E}[X] := \mathbb{E}[X_+] - \mathbb{E}[X_-]$ makes sense and equals

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X > x)\,\mathrm{d}x - \int_{-\infty}^0 \mathbb{P}(X < x)\,\mathrm{d}x,$$

as desired. $\square$

The proof of Proposition 2.1 above is simple, expeditious (at least for the first part, which has only a few lines), and, for that matter, easy for students to follow and digest. It also dispenses with all technical issues brought by integration by parts (see Hong (2012), where X was assumed to be absolutely continuous, for details). Observe that nowhere in the proof above did we use the fact that X is continuous with a probability density function. The only assumption for (1.1) and (2.1) to be true is that X is non-negative and X is integrable, respectively. Unlike the way in which many textbooks present the two formulas may suggest, (1.1) and (2.1) indeed provide a unifying treatment of all kinds of random variables, continuous, discrete, or mixed, all of which are of importance in different contexts in probability and statistics. Of course, if X is non-negative, then the second integral in (2.1) vanishes and (2.1) reduces to (1.1). In particular, if X is non-negative and integer-valued, then the survival function $\mathbb{P}(X > x)$ is a right-continuous piecewise constant function with a jump at each non-negative integer; see Figure 1. Because integrating a non-negative function is essentially about calculating the area under its graph, the integral

$\int_0^\infty \mathbb{P}(X > x)\,\mathrm{d}x$ amounts to the sum of the area of the rectangular strips in Figure 1, each having height $\mathbb{P}(X > x)$ for some non-negative integer x and unit width. In this case, (1.1) retrieves (1.2):

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X > x)\,\mathrm{d}x = \sum_{x=0}^\infty \mathbb{P}(X > x).$$

When teaching the integrated tail probability expectation formula to students in an undergraduate mathematical statistics class, it would be judicious of instructors to focus on (1.1), which is technically simpler than (2.1) but is sufficient for most applications, and present only the first half of the proof of Proposition 2.1 without the need for justifying rigorously why putting the expectation inside the integral is permissible. Students are likely to be receptive to the first part of the proof as interchanging the order of expectation and integration makes intuitive sense, keeping in mind that an integral (both Riemann and Lebesgue) is a limit of an approximating sum and that expectation is linear. They should also find swapping the order of two operations familiar as they should have encountered exchanging the order of integration in a multi-variable calculus class (mostly without seeing the underlying theory). Graduate students in a measure-theoretic probability course will have no difficulty in understanding the proof in full as they have learned that expectation is also a kind of integration (in a general sense) and that Fubini's theorem, which is typically covered in the chapter on product measures, is clearly applicable in this setting to justify interchanging the order of integration—the indicator function $(\omega, x) \mapsto 1_{\{X(\omega)>x\}}$ is non-negative and jointly measurable. In fact, (2.1) is often presented as a prominent probabilistic application of Fubini's theorem. They should also be able to master the technical underpinnings of (2.1) and realize that (2.1) involves Riemann integrals and that the events $\{X > x\}$ and $\{X < x\}$ can be replaced by $\{X \geq x\}$ and $\{X \leq x\}$, respectively, without affecting the validity of the formula. As a side note, it should be pointed out that (1.1) (resp. (2.3)) remains true even when X (resp. X_+ or X_-) is non-negative, but not integrable. This is because (2.2), along with the fact that the indicator function $1_{\{X>x\}}$ (resp. $1_{\{X_+>x\}}$ or $1_{\{X_->x\}}$) is non-negative, is enough to render the use of Fubini's theorem and, by extension, (1.1) (resp. (2.3)) valid. In this non-integrable case, the integral $\int_0^\infty \mathbb{P}(X > x)\,\mathrm{d}x$ (resp. $\int_0^\infty \mathbb{P}(X > x)\,\mathrm{d}x$ or $\int_{-\infty}^0 \mathbb{P}(X < x)\,\mathrm{d}x$) diverges. A consequence is that if at least one of the two integrals, $\int_0^\infty \mathbb{P}(X > x)\,\mathrm{d}x$ and $\int_{-\infty}^0 \mathbb{P}(X < x)\,\mathrm{d}x$, is finite, then $\mathbb{E}[X]$ exists and is given by (2.1). If, in addition, both integrals are finite, then so is $\mathbb{E}[X]$, i.e., X is integrable (see also the converse result in Theorem 5 on page 215 of Rényi (2007)).

Irrespective of the type of students that face (2.1), they will benefit from a much more thorough understanding of the formula if they are afforded the opportunity to glimpse its geometric meaning and practical usage, as the next two subsections will show.

2.2 Geometric Interpretation

While the integrated tail probability expectation formula has been given and proved algebraically in Proposition 2.1, a much deeper understanding of its genuine meaning can be acquired from a geometric perspective. Aesthetically appealing and theoretically sound, such a pictorial interpretation of the formula not only lends probabilistic insights to the two integrals constituting (2.1), but also makes the tail expectation formula much more memorable for students.

We begin by recalling the standard formula for computing the expectation of any random variable X in the Lebesgue–Stieltjes form of

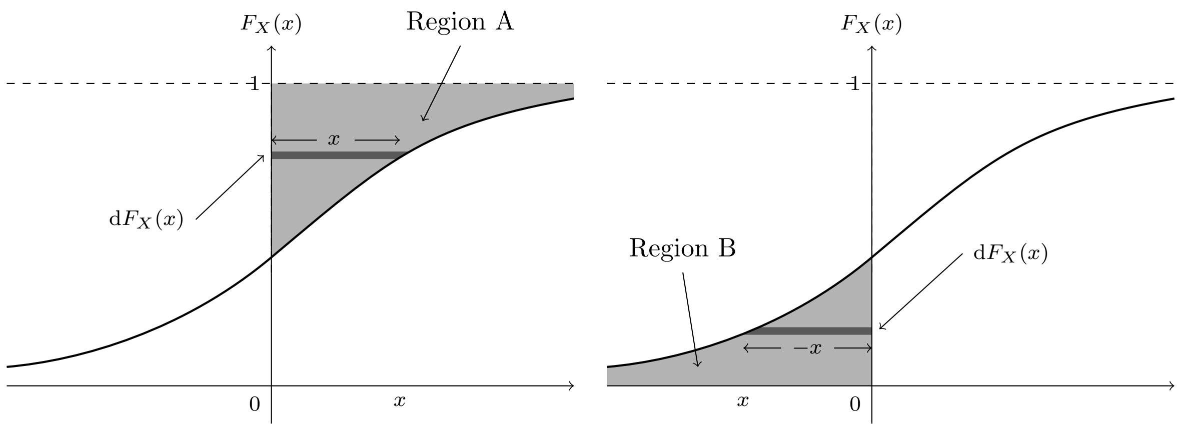
$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \, \mathrm{d}F_X(x), \tag{2.4}$$

where F_X is the distribution function of X. The integral $\int_{-\infty}^{\infty} x \, \mathrm{d}F_X(x)$ unifies continuous, discrete, and mixed distributions and, loosely speaking, can be interpreted as integrating $xf_X(x)$ over the set of all x at which X has a probability density function f_X (teachers in elementary undergraduate courses may focus on this case for simplicity) and summing $x\mathbb{P}(X = x)$ over all x at which the distribution function of X has a jump. To grasp the geometric meaning of (2.4), consider any given $x > 0$ in Figure 2 (a), where the rectangle has a width of x, a height of $\mathrm{d}F_X(x)$, and an area of $x \, \mathrm{d}F_X(x)$. The integral $\int_0^{\infty} x \, \mathrm{d}F_X(x)$ aggregates all such infinitesimally small rectangles and returns the area of region A in Figure 2 (a). Now, for any fixed $x < 0$ in Figure 2 (b), the height of the rectangle remains $\mathrm{d}F_X(x)$, but its width becomes $-x$, and its area $-x \, \mathrm{d}F_X(x)$. The area of region B is then captured by the integral $\int_{-\infty}^{0} (-x) \, \mathrm{d}F_X(x)$. Geometrically, (2.4) depicts the expectation of a random variable as the area of region A less the area of region B:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \, \mathrm{d}F_X(x) = \underbrace{\int_0^{\infty} x \, \mathrm{d}F_X(x)}_{\text{area of region A}} - \underbrace{\int_{-\infty}^{0} (-x) \, \mathrm{d}F_X(x)}_{\text{area of region B}}. \tag{2.5}$$

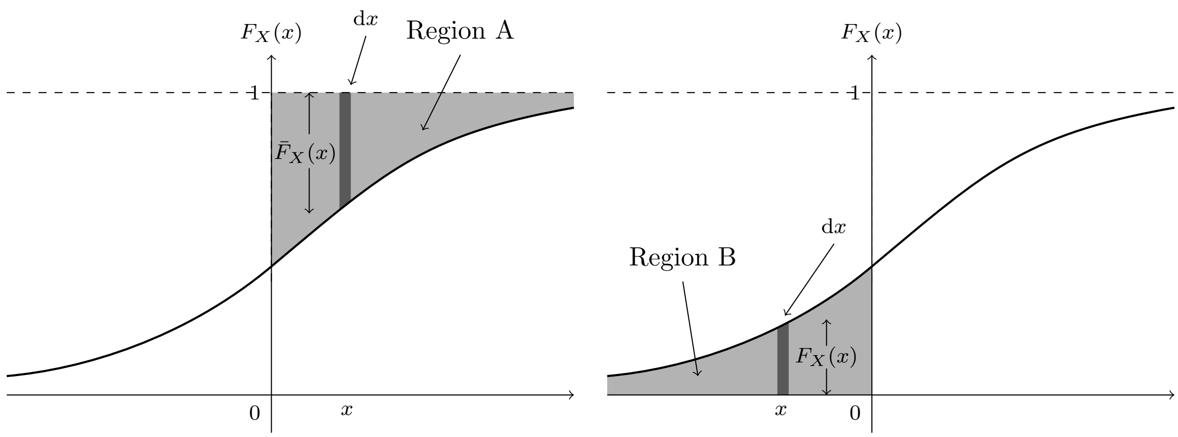
If X is non-negative, then region B vanishes, and the expected value of X simply equals the area of region A.

The integral tail probability expectation formula in the form of (2.1) offers an alternative but equally instructive way to sweep regions A and B. In Figure 3 (a), the shaded rectangle has an area of $[1 - F_X(x)] \, \mathrm{d}x = \bar{F}_X(x) \, \mathrm{d}x$, where $\bar{F}_X = 1 - F_X$ is the survival function of X, and that in Figure 3 (b) has an area of $F_X(x) \, \mathrm{d}x$. Aggregating all of these rectangles in the positive and negative real lines, we can write the areas of regions A and B as $\int_0^{\infty} \bar{F}_X(x) \, \mathrm{d}x$ and $\int_{-\infty}^{0} F_X(x) \, \mathrm{d}x$, respectively. Therefore, (2.1) confirms that the expectation of the random



(a) $x > 0$. (b) $x < 0$.

Figure 2: Geometric meaning of (2.4) (horizontal slicing). For simplicity, the distribution function is taken to be continuous.



(a) $x > 0$. (b) $x < 0$.

Figure 3: Geometric meaning of (2.1) (vertical slicing).

variable X is the excess of the area of region A over the area of region B, consistent with (2.5), and provides alternative expressions for these two areas (see Figure 24 on page 216 of Rényi (2007) for a similar graphical interpretation).

In essence, (2.1) and (2.4) share the same spirit as reversing the order of integration in a double integral (just as Fubini's theorem entails) and correspond to two different but equivalent methods of slicing the area above or under the graph of F_X which represents the expected value of the random variable X. In contrast to (2.4), which employs horizontal strips to capture the areas of regions A and B, (2.1) does so by virtue of vertical strips. Pedagogically, instructors who complement their teaching of expectations with (2.1) and its geometric interpretation will instill into students the geometry of the expectation of a

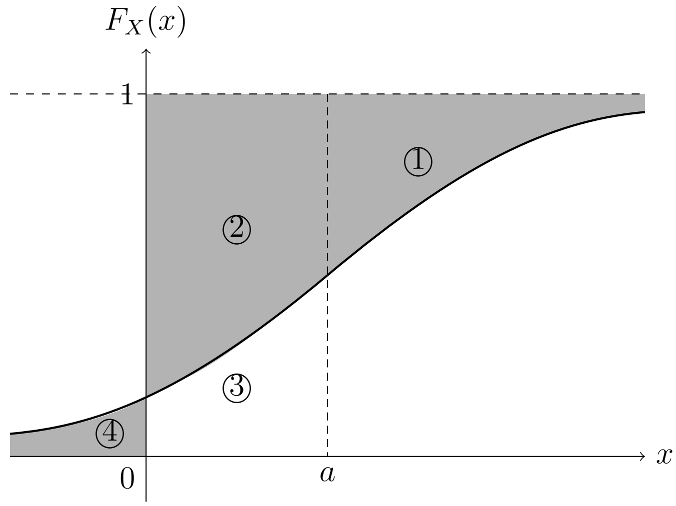


Figure 4: Graphical illustration that the expectation of a distribution symmetric about $a \geq 0$ is precisely a in Example 2.2.

random variable, which is fundamental in the theory of probability but often overlooked in both elementary and advanced probability courses. By turning the convoluted algebraic expressions in (2.1) and (2.4) into pictures with colorful probabilistic meaning, instructors will make it easy for students to remember and make sense of the two equivalent expectation formulas.

Example 2.2. *(Expectation of a symmetric distribution)* A simple but instructive application of the geometric interpretation of expectations presented in this subsection is a pictorial determination of the expectation of a general symmetric distribution, provided that its expectation exists. Given a random variable X, we say that its distribution F_X is *symmetric about* $a \in \mathbb{R}$ if $\mathbb{P}(X < a - x) = \mathbb{P}(X > a + x)$ for all $x \geq 0$. If the expectation of such a random variable exists, then it must be integrable with an expected value of a (see, e.g., Exercise 2.26 (c) on page 79 of Casella and Berger (2002), Exercise 1.9.5 on page 65 of Hogg et al. (2013), and Exercise 4.37 on page 174 of Wackerly et al. (2008), all of which are presented under the continuity assumption). To show this, we assume, without loss of generality, that $a \geq 0$[i]; the case of $a \leq 0$ requires only cosmetic adjustments. The distribution function of X is sketched in Figure 4 with four regions identified. Because of symmetry, we have

$$\int_a^\infty \mathbb{P}(X > x)\,\mathrm{d}x = \int_0^\infty \mathbb{P}(X > a + x)\,\mathrm{d}x = \int_0^\infty \mathbb{P}(X < a - x)\,\mathrm{d}x = \int_{-\infty}^a \mathbb{P}(X < x)\,\mathrm{d}x, \tag{2.6}$$

[i]Alternatively, it can be observed that if the distribution of X is symmetric about a, then X has the same distribution as $Y + a$ for some random variable Y whose distribution is symmetric about 0. Therefore, it suffices to focus on the case when $a = 0$.

which means that the area of region ① equals the sum of the area of region ③ and that of region ④. Furthermore, due to (2.6), the two integrals, $\int_0^\infty \mathbb{P}(X > x)\,\mathrm{d}x$ and $\int_{-\infty}^0 \mathbb{P}(X < x)\,\mathrm{d}x$, are either both finite, in which case X is integrable, or both infinite, in which case $\mathbb{E}[X]$ does not exist (see the discussions on page 6). In the former case, our graphical interpretation of expectations yields

$$\begin{aligned}\mathbb{E}[X] &= (\text{area of region ①}) + (\text{area of region ②}) - (\text{area of region ④})\\ &= (\text{area of region ①}) + (a - \text{area of region ③}) - (\text{area of region ④})\\ &= a,\end{aligned}$$

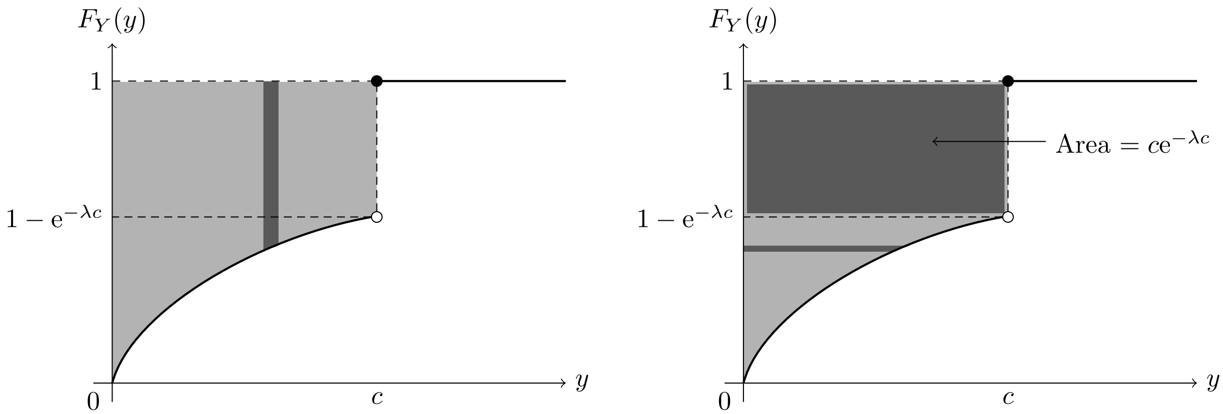
where the second equality follows from the fact that the area of region ② and that of region ③ sum to a. Note that because our analysis holds for general symmetric distributions, so does our conclusion (see page 216 of Rényi (2007) for related discussions). □

2.3 Pedagogical Applications

When faced with (2.1) or its simplified form (1.1), students' first reaction would inevitably be "why another formula for expectations?" This question is seldom tackled, if at all, squarely in textbooks, which tend to do little beyond stating the tail expectation formula as an end-of-chapter exercise. At the level of an elementary probability and statistics course, the greatest advantage of the tail expectation formula over the traditional formula (2.4) arguably lies in the convenience it brings when it comes to computing expectations by hand, a central topic in an undergraduate-level mathematical statistics class and a skill that almost all beginners need to master. Just as exchanging the order of integration can sometimes ease the computation of a multiple integral, the tail expectation formula has the potential of simplifying the analytic determination of expectations, especially when the survival function of the random variable in question takes a mathematically tractable form.

Example 2.3. *(A right-censored exponential random variable)* Let X be an exponential random variable with rate λ. For a given $c \geq 0$, we are interested in the expectation of the censored random variable $Y := \min(X, c)$. Such a kind of random variable bears particular relevance to survival analysis, where X represents the lifetime of the subject of interest in a clinical trial and c is the duration of the study. The distribution function of Y is

$$F_Y(y) = \mathbb{P}(Y \leq y) = \begin{cases} 0, & \text{if } y < 0,\\ \mathbb{P}(X \leq y) = 1 - \mathrm{e}^{-\lambda y}, & \text{if } 0 \leq y < c,\\ 1, & \text{if } c \leq y.\end{cases}$$



(a) Use of (1.1) (vertical slicing). (b) Use of (2.4) (horizontal slicing).

Figure 5: Illustration of the distribution function of the random variable Y in Example 2.3. The pale gray area represents the expected value of Y.

Exhibited in Figure 5, the distribution function of Y is a mixed one which is continuous on $[0, c)$ but has a discontinuity at $y = c$ with a size of $\mathrm{e}^{-\lambda c}$. To determine the expected value of Y, we appeal to (1.1) and get almost effortlessly

$$\mathbb{E}[Y] = \int_0^\infty \bar{F}_Y(y)\,\mathrm{d}y = \int_0^c \mathrm{e}^{-\lambda y}\,\mathrm{d}y = \frac{1-\mathrm{e}^{-\lambda c}}{\lambda}.$$

As $c \to \infty$, we retrieve X from Y and the preceding expectation formula simplifies as expected to $1/\lambda$, a result which many students have memorized (but may not have actually performed the integration once themselves).

Had we found the expected value of Y in the standard way by virtue of (2.4), we would have needed to distinguish the discrete part (at $y = c$) and absolutely continuous part (over $[0, c)$) of the distribution of Y, consider their contributions to $\mathbb{E}[Y]$ separately, and calculate

$$\mathbb{E}[Y] = \int_0^c y f_Y(y)\,\mathrm{d}y + c\,\mathbb{P}(Y = c) = \int_0^c y(\lambda \mathrm{e}^{-\lambda y})\,\mathrm{d}y + c\,\mathrm{e}^{-\lambda c}$$

more tediously via integration by parts.

As an exercise, interested readers may try to find the expected value of a Pareto random variable Z with probability density function $f_Z(z) = \alpha\theta^\alpha/(z+\theta)^{\alpha+1}$ for $z \geq 0$ by (1.1), then by (2.4), and see how much easier the integration is with the former compared with the latter. □

Example 2.4. *(Expected number of trials until the first success)* Another instance in which the integrated tail probability expectation formula is much more efficient than the conventional formula is to calculate the expected number of trials required to reach the first success in a sequence of independent and identically distributed Bernoulli trials. Let X be the number of trials until the first success, which is a geometric random variable, and $p \in (0,1)$ be the success probability of each trial. For any non-negative integer x, $X > x$ if and only if the first x trials all lead to a failure, which happens with a probability of $(1-p)^x$. By (1.2),

$$\mathbb{E}[X] = \sum_{x=0}^{\infty} \mathbb{P}(X > x) = \sum_{x=0}^{\infty} (1-p)^x = \frac{1}{1-(1-p)} = \frac{1}{p},$$

which is simply a geometric series. In contrast, the standard way of calculating the expected value of X would require

$$\mathbb{E}[X] = \sum_{x=1}^{\infty} x\mathbb{P}(X = x) = p\sum_{x=1}^{\infty} x(1-p)^{x-1},$$

which can be an intimidating series to undergraduate students. □

At a more theoretical level, the tail expectation formula serves as an indispensable tool that can be used to translate information about tail probabilities to moment-based conclusions and establish otherwise inaccessible technical results. We offer one illustrative example in this vein that may be of interest to graduate students in applied probability and survival analysis.

Example 2.5. *(Stochastic dominance or stochastic ordering)* Let X and Y be integrable random variables with respective distribution functions F_X and F_Y. We say that X precedes Y in *first-order stochastic dominance*, or X is *stochastically smaller* than Y, written as $X \preccurlyeq Y$, if $F_X(t) \geq F_Y(t)$ for all $t \in \mathbb{R}$, or equivalently, $\bar{F}_X(t) \leq \bar{F}_Y(t)$ for all $t \in \mathbb{R}$ (see, e.g., Section 1.2 of Müller and Stoyan (2002)). Intuitively, if $X \preccurlyeq Y$, then X is "smaller" than Y in the sense that X is more likely to assume small values and less likely to take large values than Y. This observation on the relative size of the two random variables is reflected by their expectations ordered as

$$\mathbb{E}[X] = \int_0^{\infty} \bar{F}_X(t)\,\mathrm{d}t - \int_{-\infty}^{0} F_X(t)\,\mathrm{d}t \leq \int_0^{\infty} \bar{F}_Y(t)\,\mathrm{d}t - \int_{-\infty}^{0} F_Y(t)\,\mathrm{d}t = \mathbb{E}[Y],$$

which is a direct consequence of (2.1). □

3 Multivariate Extensions

The integrated tail probability expectation formula admits generalizations in a number of directions. The ones most germane to this pedagogical note are its multivariate extension and, in the same spirit, an integrated tail probability formula for the covariance between two random variables. To avoid unnecessary notational complexity, we present and prove the multivariate counterpart of the tail expectation formula only for non-negative random variables, which are sufficient for most practical applications and for illustrating the essential ideas of the formula.

Proposition 3.1. (Integrated tail probability covariance formula) *For any non-negative random variables X and Y with finite variance,*

$$Cov(X,Y) = \int_0^\infty \int_0^\infty [\mathbb{P}(X > x, Y > y) - \mathbb{P}(X > x)\mathbb{P}(Y > y)]\,\mathrm{d}x\mathrm{d}y \tag{3.1}$$

Proof. Mimicking the proof of Proposition 2.1, we start by writing

$$X = \int_0^\infty 1_{\{X>x\}}\,\mathrm{d}x \quad \text{and} \quad Y = \int_0^\infty 1_{\{Y>y\}}\,\mathrm{d}y.$$

By Fubini's theorem, the product moment can be exhibited in integral form as

$$\begin{aligned}
\mathbb{E}[XY] &= \mathbb{E}\left[\left(\int_0^\infty 1_{\{X>x\}}\,\mathrm{d}x\right)\left(\int_0^\infty 1_{\{Y>y\}}\,\mathrm{d}y\right)\right] \\
&= \mathbb{E}\left[\int_0^\infty \int_0^\infty 1_{\{X>x,Y>y\}}\,\mathrm{d}x\mathrm{d}y\right] \\
&= \int_0^\infty \int_0^\infty \mathbb{P}(X > x, Y > y)\,\mathrm{d}x\mathrm{d}y. \qquad (3.2)
\end{aligned}$$

Together with Proposition 2.1, we have

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \\
&= \int_0^\infty \int_0^\infty \mathbb{P}(X > x, Y > y)\,\mathrm{d}x\mathrm{d}y - \left(\int_0^\infty \mathbb{P}(X > x)\,\mathrm{d}x\right)\left(\int_0^\infty \mathbb{P}(Y > y)\,\mathrm{d}y\right) \\
&= \int_0^\infty \int_0^\infty [\mathbb{P}(X > x, Y > y) - \mathbb{P}(X > x)\mathbb{P}(Y > y)]\,\mathrm{d}x\mathrm{d}y,
\end{aligned}$$

completing the proof. □

Note that (3.2) in the proof above is essentially the extension of (1.1) to two non-negative random variables, which was given in Hong (2015) under the assumption of absolute continuity. Compared to Hong (2015), the proof above is considerably less involved, more expeditious and is applicable to general non-negative random variables, not necessarily absolutely continuous ones. After presenting and discussing (2.1) in class, instructors may challenge students with (3.2) and Proposition 3.1 as closely related after-class exercises meant to test if students have mastered the techniques illustrated in proving the univariate expectation formula (i.e., displaying the random variable in integral form and interchanging the order of expectation and integration). In the bivariate setting, however, the simple geometric interpretation in Figures 2 and 3 may be lost.

Whereas (3.2) holds true only for non-negative random variables, the structure of (3.1) remains unchanged even for general random variables that take both positive and negative values. The only adjustment needed is that the integration is extended from the positive quadrant to the whole two-dimensional plane:

$$\mathrm{Cov}(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [\mathbb{P}(X > x, Y > y) - \mathbb{P}(X > x)\mathbb{P}(Y > y)]\,\mathrm{d}x\mathrm{d}y$$

This covariance formula can be traced back to Hoeffding (1940) and is justifiably named the *Hoeffding's formula* (see Shea (1983) for further discussions and Mardia (1967), Cuadras (2002), and Lo (2017) for extensions). Our proof above, while restricted to non-negative random variables, dispenses with the coupling technique which is typically used to prove the formula (see, e.g., Property 1.6.13 on page 28 of Denuit et al. (2005)) and which may be intimidating to undergraduate students.

Hoeffding's covariance formula has a host of applications in the study of dependence structures and stochastic orders (see, e.g., pages 266 and 267 of Denuit et al. (2005)) that may be of interest to graduate students and researchers, but are beyond the scope of this note. For our pedagogical purpose, we provide an illustrative example showing how (3.2) and Hoeffding's covariance formula may enter the conversations in an undergraduate probability and statistics class as an efficient tool for calculating covariances between random variables with an intricate joint distribution.

Example 3.2. *(Bivariate Marshall–Olkin exponential distribution)* Let (X, Y) be a pair of dependent exponential random variables with joint survival function

$$\bar{F}_{X,Y}(x, y) = \exp[-\mu x - \nu y - \lambda \max(x, y)], \qquad \text{for } x > 0 \text{ and } y > 0, \tag{3.3}$$

where μ, ν, and λ are positive parameters. The marginal distributions of X and Y are exponential with parameters $\mu + \lambda$ and $\nu + \lambda$, respectively. Such a bivariate exponential distribution arises most notably in shock models in reliability theory (see Section 2.1 of Marshall and Olkin (1967)) and insurance mathematics (see Subsection 9.6.1 of Bowers et al.

(1997)), where a "common shock" modeling the occurrence time of a widespread catastrophe with rate λ is at work. We are interested in the covariance between X and Y as an illustration of Hoeffding's formula. Distinguishing whether the integration variables satisfy $x > y$ or $x < y$, we obtain, by straightforward integration,

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \int_0^\infty \int_0^\infty \mathbb{P}(X > x, Y > y)\,\mathrm{d}x\mathrm{d}y - \left(\int_0^\infty \mathbb{P}(X > x)\,\mathrm{d}x\right)\left(\int_0^\infty \mathbb{P}(Y > y)\,\mathrm{d}y\right) \\
&= \int_0^\infty \int_y^\infty \exp[-(\mu+\lambda)x - \nu y]\,\mathrm{d}x\mathrm{d}y + \int_0^\infty \int_x^\infty \exp[-\mu x - (\nu+\lambda)y]\,\mathrm{d}y\mathrm{d}x \\
&\qquad - \left(\int_0^\infty \exp[-(\mu+\lambda)x]\,\mathrm{d}x\right)\left(\int_0^\infty \exp[-(\nu+\lambda)y]\,\mathrm{d}y\right) \\
&= \frac{1}{\mu+\lambda}\int_0^\infty \exp[-(\mu+\nu+\lambda)y]\,\mathrm{d}y + \frac{1}{\nu+\lambda}\int_0^\infty \exp[-(\mu+\nu+\lambda)x]\,\mathrm{d}x \\
&\qquad - \frac{1}{(\mu+\lambda)(\nu+\lambda)} \\
&= \frac{1}{(\mu+\lambda)(\mu+\nu+\lambda)} + \frac{1}{(\nu+\lambda)(\mu+\nu+\lambda)} - \frac{1}{(\mu+\lambda)(\nu+\lambda)} \\
&= \frac{\lambda}{(\mu+\lambda)(\nu+\lambda)(\mu+\nu+\lambda)}. \qquad (3.4)
\end{aligned}$$

When $\lambda = 0$, the common shock vanishes, in which case X and Y become independent, with a zero covariance. □

Note that a derivation of the covariance in Example 3.2 using the standard formula $\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ requires two double integrals, one corresponding to the absolutely continuous part of the joint distribution of (X,Y), together with an additional integral accounting for the simultaneous occurrence of X and Y (see Theorem 3.1 and Lemma 3.2 of Marshall and Olkin (1967) for the mathematical details), that is,

$$\mathbb{E}[XY] = \iint_{\{(x,y):x\geq 0, y\geq 0, x\neq y\}} xy\frac{\partial^2}{\partial x\partial y}\bar{F}_{X,Y}(x,y)\,\mathrm{d}x\mathrm{d}y + \int_0^\infty z^2(\lambda \mathrm{e}^{-(\mu+\nu+\lambda)z})\,\mathrm{d}z.$$

The development of the ingredients for these integrals and their evaluations are substantially more tedious and strenuous than those presented in Example 3.2. In addition to being applicable to general random vectors irrespective of their type of distributions, the use of Hoeffding's formula to evaluate $\mathrm{Cov}(X,Y)$ conveniently sidesteps the singularly continuous component of the distribution of (X,Y) and involves much more straightforward integration and algebraic simplification.

4 Concluding Remarks

The linchpin of this pedagogical note is an expectation formula coined as the integrated tail probability expectation formula. This formula is shown to be a vehicle of substantial pedagogical interest and a useful complement to the teaching of expectations in both introductory mathematical statistics classes and graduate-level measure-theoretic probability courses. At minimal technical cost, the tail expectation formula has the payoff of unifying the treatment of all integrable random variables, teaching students the geometric meaning of expectations, and simplifying the calculation of the expected values of random variables with complex distributions. Given the educational merits of the formula, it is somewhat unfortunate that it has received only scant attention from mainstream textbooks. It is the author's hope that this note can help promote the tail expectation formula to the wider community, offer useful guidelines to instructors on revitalizing the teaching of expectations with some geometric flavor, and in turn benefit students' lifelong learning in probability and statistics.

References

Bowers, N. L., H. U. Gerber, J. C. Hickman, D. A. Jones, and C. J. Nesbitt (1997). *Actuarial Mathematics* (Second ed.). The Society of Actuaries, Schaumburg, Illinois.

Casella, G. and R. L. Berger (2002). *Statistical Inference* (Second ed.). Duxbury.

Chakraborti, S., F. Jardim, and E. Epprecht (2017). Higher order moments using the survival function: The alternative expectation. *The American Statistician*, DOI: 10.1080/00031305.2017.1356374.

Cuadras, C. M. (2002). On the covariance between functions. *Journal of Multivariate Analysis 81*, 19–27.

Denuit, M., J. Dhaene, M. Goovaerts, and R. Kaas (2005). *Actuarial Theory for Dependent Risks: Measures, Orders and Models*. Chichester: John Wiley & Sons, Ltd.

Hoeffding, W. (1940). Masstabinvariante Korrelationstheorie. *Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin 5*, 181–233.

Hogg, R. V., J. W. McKean, and A. T. Craig (2013). *Introduction to Mathematical Statistics* (Seventh ed.). Pearson.

Hong, L. (2012). A remark on the alternative expectation formula. *The American Statistician 66*(4), 232–233.

Hong, L. (2015). Another remark on the alternative expectation formula. *The American Statistician 69*(3), 157–159.

Lieb, E. H. and M. Loss (2001). *Analysis* (Second ed.). American Mathematical Society.

Lo, A. (2017). Functional generalizations of Hoeffding's covariance lemma and a formula for Kendall's tau. *Statistics and Probability Letters 122*, 218–226.

Mardia, K. V. (1967). Some contributions to contingency-type bivariate distributions. *Biometrika 54*, 235–249.

Marshall, A. W. and I. Olkin (1967). A multivariate exponential distribution. *Journal of the American Statistical Association 62*(317), 30–44.

Müller, A. and D. Stoyan (2002). *Comparison Methods for Stochastic Models and Risks.* John Wiley & Sons Inc., England.

Rényi, A. (2007). *Probability Theory* (Dover Ed ed.). Dover Publications.

Shea, G. A. (1983). Hoeffding's lemma. In S. Kotz, N. Balakrishnan, C. Read, and B. Vidakovic. (Eds.), *Encyclopedia of Statistical Sciences*. John Wiley & Sons, Inc, New York.

Wackerly, D. D., W. Mendenhall, and R. L. Scheaffer (2008). *Mathematical Statistics with Applications* (Seventh ed.). Thomson.