VERSITA

TATRA MOUNTAINS
Mathematical Publications
DOI: 10.2478/v10127-012-0025-9
Tatra Mt. Math. Publ. **52** (2012), 53–63

THE DARTH VADER RULE

Pat Muldowney — Krzysztof Ostaszewski — Wojciech Wojdowski

Dedicated to Professor Władysław Wilczyński on his 65th birthday

ABSTRACT. Using Henstock's generalized Riemann integral, we show that, for any almost surely non-negative random variable X with probability density function f_X and survival function $s_X(x) := \int_x^\infty f_X(t)\,dt$, the expected value of X is given by $\mathbf{E}(X) = \int_0^\infty s_X(x)\,dx$.

In life insurance mathematics, the concept of a survival function is commonly used in life expectancy calculations. The survival function of a random variable X is defined at x as the probability that X is greater than a specific value x. *For a non-negative random variable whose expected value exists, the expected value equals the integral of the survival function.* We propose to designate this result as the *Darth Vader Rule*[1]. It holds for any type of random variable, although its most general form relies on the integration by parts formula for the Lebesgue-Stieltjes integral, fully developed by Hewitt [3]. This result, while known (and stated in Feller [1]), is not widely disseminated except in life insurance mathematics texts; but it is worth knowing and popularizing because it provides an efficient tool for calculation of expected value, and gives insight into a property common to all types of random variables.

We give a proof of the Darth Vader Rule which works for all random variables which are non-negative almost surely and whose expected value exists. The proof is based not on the Lebesgue integral formulation of [3], but on the generalized Riemann integration of Henstock and Kurzweil [2], [4]. Since every Lebesgue integrable function is also generalized Riemann integrable, the proof here includes all cases covered by [3].

While the result is simple to state and comprehend, its proof using Lebesgue integral theory is somewhat complex. We present the result in the traditional way,

2010 Mathematics Subject Classification: 28, 60.
Keywords: expected value, Henstock integral, integration by parts.
[1]This is not a reference to a discoverer. But the designation may capture the somewhat counterintuitive—if not slightly unsettling and surreal—impression which the result can evoke on first encounter.

and then demonstrate that its proof follows directly from a simple and elegant proof of the integration by parts formula for the generalized Riemann integral.

1. Traditional derivation

When an event has probability one, we say that it happens *almost surely.* Consider a random variable X that is non-negative almost surely, whose expected value exists. If X is absolutely continuous with probability density function f_X, then

$$\begin{aligned}\mathbf{E}(X) &= \int_{-\infty}^{\infty} x f_X(x)\,dx = \int_{0}^{\infty} x f_X(x)\,dx \\ &= x\big(-s_X(x)\big)\Big|_{x=0}^{x\to+\infty} + \int_0^{\infty} s_X(x)\,dx \\ &= -\lim_{x\to\infty} x s_X(x) + \int_0^{\infty} s_X(x)\,dx \\ &= \int_0^{\infty} s_X(x)\,dx.\end{aligned}$$

The second line is obtained by integrating by parts, taking

$$u = x, \quad du = dx, \quad v = -s_X(x), \quad dv = f_X(x)\,dx.$$

Note that $\mathbf{E}(X) = \int_0^{\infty} t f_X(t)\,dt$ exists, that $x < t < \infty$, and that

$$s_X(x) = \int_x^{\infty} f_X(t)\,dt;$$

so

$$0 \le \lim_{x\to\infty} x s_X(x) = \lim_{x\to\infty} x \int_x^{\infty} f_X(t)\,dt \le \lim_{x\to\infty} \int_x^{\infty} t f_X(t)\,dt = 0.$$

Therefore we can conclude that

$$\lim_{x\to\infty} x s_X(x) = 0.$$

This implies that $\mathbf{E}(X) = \int_0^{\infty} s_X(x)\,dx$ provided X is continuous and non--negative almost surely.

What if X is discrete and non-negative almost surely? Then, by definition of expected value of a discrete random variable,

$$\mathbf{E}(X) = \sum_{x \in \mathbf{R}} x f_X(x).$$

Assume first that we can form

$$x_0 = 0 < x_1 < x_2 < \cdots ,$$

the sequence of values where the probability function f_X is positive; but put $x_0 = 0$ at the beginning of this sequence regardless of whether X attains that value with probability zero or with positive probability. Then, because of the step-function structure of the survival function,

$$\begin{aligned}
\mathbf{E}(X) &= x_0 f_X(x_0) + x_1 f_X(x_1) + x_2 f_X(x_2) + x_3 f_X(x_3) + \cdots \\
&= x_0\big(f_X(x_0) + f_X(x_1) + f_X(x_2) + \cdots\big) \\
&\quad + (x_1 - x_0)\big(f_X(x_1) + f_X(x_2) + f_X(x_3) + \cdots\big) \\
&\quad + (x_2 - x_1)\big(f_X(x_2) + f_X(x_3) + f_X(x_4) + \cdots\big) \\
&\quad + (x_3 - x_2)\big(f_X(x_3) + f_X(x_4) + f_X(x_5) + \cdots\big) \\
&\quad \vdots \\
&= \sum_{j=1}^{\infty} (x_j - x_{j-1}) s_X(x_{j-1}) \\
&= \int_0^{\infty} s_X(x)\, dx.
\end{aligned}$$

In the case when X is discrete and assumes only positive integer values, we have the following special rule:

$$\begin{aligned}
\mathbf{E}(X) &= \sum_{n=1}^{\infty} \big(n - (n-1)\big) s_X(n-1) \\
&= \sum_{n=0}^{\infty} \mathrm{Prob}(X > n) \\
&= \sum_{n=1}^{\infty} \mathrm{Prob}(X \geq n).
\end{aligned}$$

The above proof assumes that the point masses can be put in an increasing sequence. But there may be discrete probability distributions that violate that assumption. The simplest example of such a distribution would be an assignment of positive probability to every positive rational number in such a way that their probabilities, as they should, add up to one.

However, the most general result follows from integration by parts for the Lebesgue-Stieltjes integral, proved by Edwin Hewitt [3]. The formula is given in the following form. Let μ and ν be measures defined for Borel subsets of $[a, b]$ and let

$$M(t) := \frac{1}{2}\Big(\mu\big([a,t]\big) + \mu\big([a,t[\big)\Big),$$
$$N(t) := \frac{1}{2}\Big(\nu\big([a,t]\big) + \nu\big([a,t[\big)\Big).$$

Then Hewitt's integration by parts formula is

$$\int_a^b M(t)\, d\nu(t) + \int_a^b N(t)\, d\mu(t) = \mu\big([a,b]\big)\nu\big([a,b]\big).$$

Consider a random variable X defined on the interval $[0, \infty[$. Assume that $\mathbf{E}(X)$ exists and f_X is well-defined. Define the measures μ and ν as follows:

$$\mu\big([0,x]\big) := 1 - s_X(x),$$
$$\nu\big([0,x]\big) := x.$$

Thus

$$M(x) := \frac{1}{2}\Big(\mu\big([0,x]\big) + \mu\big([0,x[\big)\Big) = 1 - s_X(x),$$
$$N(x) := \frac{1}{2}\Big(\nu\big([0,x]\big) + \nu\big([0,x[\big)\Big) = x.$$

Then Hewitt's formula implies, on an interval of the form $[0, b]$,

$$\int_0^b \big(1 - s_X(x)\big)\, dx + \int_0^b x f_X(x)\, dx = \left(\int_0^b f_X(x)\, dx\right)(b - 0),$$

or

$$\int_0^b x f_X(x)\, dx = \int_0^b s_X(x)\, dx + b\int_0^b f_X(x)\, dx - b$$
$$= \int_0^b s_X(x)\, dx - b\int_b^\infty f_X(x)\, dx.$$

Note that

$$0 \le b\int_b^\infty f_X(x)\, dx = \int_b^\infty b f_X(x)\, dx \le \int_b^\infty x f_X(x)\, dx.$$

Also

$$\lim_{b\to\infty}\int_b^\infty x f_X(x)\,dx = 0$$

since $\mathbf{E}(X)$ exists. We conclude that

$$\int_0^\infty x f_X(x)\,dx = \int_0^b s_X(x)\,dx.$$

F e l l e r [1] also notes that for a random variable whose nth moment exists,

$$\mathbf{E}(X^n) = n\int_0^\infty x^{n-1} s_X(x)\,dx.$$

Note also that if X is a mixed random variable which is nonnegative almost surely then the result applies, as the mean of X is a probability-weighted average of the means of the distributions that are mixed in its creation. Of course, Hewitt's formula applies to this case as well. This means that for any random variable X which is nonnegative almost surely, and whose expected value exists, the expected value equals the integral of the survival function.

We will call this important result the **Darth Vader Rule**.

2. The Henstock-Kurzweil generalized Riemann integral

For $\mathbf{R} =]-\infty,\infty[$, denote by $\mathcal{I}$ the family of intervals in $\mathbf{R}$ of the following generic type:

$$]-\infty,a] \quad \text{or} \quad]a,b] \quad \text{or} \quad]b,\infty[\quad \text{or} \quad]-\infty,\infty[$$

where a, b are any real numbers with $a < b$.

For $I \in \mathcal{I}$, a point-interval pair (x, I) is *associated* if, for any $a < b$,

$$\begin{array}{lllll} I &=&]-\infty,a] & \text{and} & x = -\infty; \quad \text{or} \\ I &=&]a,b] & \text{and} & x = a \ \text{or}\ b; \quad \text{or} \\ I &=&]b,\infty[& \text{and} & x = \infty. \end{array}$$

A *partition* of $\mathbf{R}$ is a finite collection of disjoint $I \in \mathcal{I}$ whose union is $\mathbf{R}$. A *division* of $\mathbf{R}$ is a finite collection $\mathcal{D}$ of associated pairs (x, I) such that the collection

$$\{I : (x, I) \in \mathcal{D}\}$$

is a partition of $\mathbf{R}$. If $J \in \mathcal{I}$ is given, partitions and divisions of J are similarly defined.

A *gauge* in $\mathbf{R}$ is a positive-valued function $\delta(x)$ defined for each $x \in \bar{\mathbf{R}}$. An associated pair (x, I) is *δ-fine* if

$$\begin{aligned} I &= \,]-\infty, a] \quad \text{and} \quad a < -\delta(x)^{-1}, \\ I &= \,]a, b] \quad \text{and} \quad b - a < \delta(x), \\ I &= \,]b, \infty[\quad \text{and} \quad b > \delta(x)^{-1}. \end{aligned}$$

The idea is that, for each x, the value $\delta(x)$ constraining any interval I associated with x can be taken progressively smaller. Thus, if I is a bounded interval $]a, b]$, then the length of I decreases in accordance with $\delta(x)$, x being one of end-points of I. If I is an infinite interval $]b, \infty[$ with $x = \infty$, then decreasing $\delta(\infty)$ forces b to be successively larger; while if $I =]-\infty, a]$, decreasing $\delta(x) = \delta(-\infty)$ forces a leftward, in the negative direction.

Given a gauge δ and $J \in \mathcal{I}$, a division $\mathcal{D}$ of J is δ-fine if each $(x, I) \in \mathcal{D}$ is δ-fine.

If a real- or complex-valued function f is defined on $\mathbf{R}$, its definition is extended to $\bar{\mathbf{R}} = \mathbf{R} \cup \{-\infty, \infty\}$ by taking

$$f(-\infty) = f(\infty) = 0.$$

A *distribution function* F is a non-negative function defined on $\mathcal{I}$, and finitely additive on intervals $I \in \mathcal{I}$. If $F(\mathbf{R}) = 1$, then F is a *probability distribution function.*

Given $J \in \mathcal{I}$, a real- or complex-valued function f defined on $\mathbf{R}$, and a distribution function F defined on $\mathcal{I}$, the function f is integrable in J with respect to F (in the generalized Riemann sense) with integral α, if, for any given $\varepsilon > 0$, there exists a gauge δ such that, for each δ-fine division $\mathcal{D}_\delta$ of J, the following inequality holds

$$\left| \alpha - \sum \{ f(x)F(I) : (x, I) \in \mathcal{D}_\delta \} \right| < \varepsilon.$$

We write the Riemann sum and the integral as follows

$$\sum \{ f(x)F(I) : (x, I) \in \mathcal{D}_\delta \} = (\mathcal{D}_\delta) \sum f(x)F(I),$$

$$\alpha = \int_J f(x)F(I).$$

In [2] it is shown that f is Lebesgue integrable with respect to F if and only if f is generalized Riemann integrable with respect to F, and the latter theory, like the former, has monotone and dominated convergence theorems, and other properties expected in an integral.

3. The Darth Vader Rule

To prove the Darth Vader Rule, suppose a probability space $(\Omega, \mathcal{A}, P)$ is given, and suppose X is a random variable, non-negative P-almost everywhere, with expected value $\mathbf{E}(X)$. For $I \in \mathcal{I}$, define the probability distribution of X,

$$F_X(I) := P(X \in I).$$

Then

$$\mathbf{E}(X) = \int_{\Omega} X(\omega) dP = \int_{-\infty}^{\infty} x dF_X = \int_{\mathbf{R}} x F_X(I).$$

The last integral is generalized Riemann, and the preceding two integrals are Lebesgue and Lebesgue-Stieltjes, respectively.

Define the survival function $s(v) := \int_{[v,\infty[} F_X(I)$. Since X is non-negative almost everywhere, $F_X\big(]-\infty, 0[\big) = 0$. For $I \in \mathcal{I}$, define

$$|I| = \begin{cases} v - u & \text{if} \quad I =]u, v], \\ 0 & \text{otherwise.} \end{cases}$$

We wish to prove the Darth Vader Rule,

$$\mathbf{E}(X) = \int_{\mathbf{R}} s(x)|I|.$$

That is, in traditional notation, we seek to prove

$$\mathbf{E}(X) = \int_{-\infty}^{\infty} x dF_X = \int_{-\infty}^{\infty} s(x)\, dx.$$

First we prove some lemmas.

Lemma 1. *If $0 < v < \infty$, and*

$$\mu(v) := \int_{[0,v]} x F_X(I), \qquad \bar{\mu}(v) := \int_{]v,\infty[} x F_X(I),$$

then these integrals exist and satisfy

$$\mu(v) + \bar{\mu}(v) = \mathbf{E}(X). \tag{3.1}$$

P r o o f. These results follow from Theorem 5.1 of [2]. □

Lemma 2. *$\mu(v)$ is monotone increasing to $\mathbf{E}(X)$ as $v \to \infty$, and $\bar{\mu}(v)$ is monotone decreasing to 0 as $v \to \infty$.*

P r o o f. We can write

$$\mu(v) = \int_{[0,\infty]} \mathbf{1}_{[0,v]}(x) x F_X(I) \quad \text{and} \quad \bar{\mu}(v) = \int_{[0,\infty]} \mathbf{1}_{[v,\infty]}(x) x F_X(I),$$

and since $\mathbf{1}_{[0,v]}(x) x F_X(I)$ converges monotonically to $x F_X(I)$ in $[0, \infty[$ as $v \to \infty$, the results then follow from (3.1) and the monotone convergence theorem [2, Theorem 8.1]. □

Lemma 3. *$\bar{\mu}(v) < \varepsilon$ implies $vs\,(v) \leq \varepsilon$.*

P r o o f. If we form Riemann sums over any division of $[v, \infty[$, we have

$$\sum x F_X(I) > v \sum F_X(I)$$

and the result follows from this. □

Theorem 1 (Darth Vader Rule)**.** *If X is an almost everywhere non-negative random variable whose expectation exists, then*

$$\mathbf{E}(X) = \int_{\mathbf{R}} s(x)|I|.$$

P r o o f. For intervals $J =]u, v]$, the survival function $s(v) = \int_{]v,\infty[} F_X(I)$ satisfies

$$s(u) - s(v) = \int_{]u,\infty[} F_X(I) - \int_{]v,\infty[} F_X(I) = \int_J F_X(I).$$

Every Riemann sum estimate of $\int_J F_X(I)$ has the form $\sum F_X(I)$ where the intervals I partition J. Therefore, by the additivity of F_X, we have

$$\sum F_X(I) = F_X(J) = \int_J F_X(I) \quad \text{and} \quad s(u) - s(v) = F_X(J).$$

Now suppose J has the form $]b, \infty[$. If we define $s(\infty)$ to be zero, then, using Riemann sum estimates as before, we also find that

$$s(b) - s(\infty) = F_X(J).$$

Define

$$s(I) := \begin{cases} s(v) - s(u) & \text{if} \quad I =]u, v], \\ s(\infty) - s(b) & \text{if} \quad I =]b, \infty[, \end{cases}$$

so $s(I) = -F_X(I)$ for all intervals. Now define

$$h(x, I) := -xs(I) - s(x)|I| \quad \text{with} \quad h(x, I) := 0 \qquad \text{if} \quad x = \infty.$$

Recall that $|I| = v - u$ whenever I has the form $]u, v]$ and $|I| = 0$ whenever I has the form $]b, \infty[$. Since $h(x, I)$ is the same as $x F_X(I) - s(x)|I|$, our object is to prove that $\int_{[0,\infty[} h(x, I) = 0$.

If $I=]u,v]$ and $x=u$, then

$$\begin{aligned} h(x,I) &= x\big(s(u)-s(v)\big)-s(x)(v-u) \\ &= (v-u)\big(s(v)-s(u)\big)+us(u)-vs(v). \end{aligned}$$

If $I=]u,v]$ and $x=v$, then

$$\begin{aligned} h(x,I) &= x\big(s(u)-s(v)\big)-s(x)(v-u) \\ &= (v-u)\big(s(u)-s(v)\big)+us(u)-vs(v). \end{aligned}$$

If $I=]b,\infty[$, with $x=\infty$, then $h(x,I)=0$ by definition. Thus, for any division $\mathcal{D}$ of $[0,\infty[$,

$$\begin{aligned} \left|(\mathcal{D})\sum h(x,I)\right| &\le (\mathcal{D})\sum \left|(v-u)\big(s(u)-s(v)\big)\right| \\ &\quad + \left|(\mathcal{D})\sum\big(us(u)-vs(v)\big)\right|. \end{aligned} \tag{3.2}$$

With $\varepsilon>0$ given, choose v_ε so that $\bar{\mu}(v_\varepsilon)<\varepsilon$. Now define a gauge δ so that

$$\delta(x) < \begin{cases} \varepsilon & \text{if } \quad x=0, \\ \min\{x,\varepsilon\} & \text{if } \quad 0<x<\infty, \\ v_\varepsilon^{-1} & \text{if } \quad x=\infty. \end{cases}$$

The definition of $\delta(0)$ ensures that any δ-fine division $\mathcal{D}$ of $[0,\infty[$ includes a term $\big(x,]0,v]\big)$ with $x=0$. Then

$$\mathcal{D}=\Big\{\big(0,]0,v_1]\big),\ldots,\big(x,]\bar{u},\bar{v}]\big),\big(\infty,]\bar{v},\infty[\big)\Big\}$$

where $\bar{v}$ is the largest of those v for which the partitioning intervals I are bounded (that is have the form $]u,v]$); and satisfies $\bar{v}>v_\varepsilon$. Therefore, by the monotonicity of $\bar{\mu}(v)$, we have

$$\varepsilon>\bar{\mu}(v_\varepsilon)>\bar{\mu}(\bar{v}),$$

and hence, by Lemma 3,

$$\bar{v}s(\bar{v})\le\varepsilon.$$

Let $\mathcal{E}$ denote the division

$$\Big\{\big(0,]0,v_1]\big),\ldots,\big(x,]\bar{u},\bar{v}]\big)\Big\}.$$

Then $\mathcal{E}$ is a δ-fine division of $[0,\bar{v}]$, and inequality (3.2) gives

$$\begin{aligned} \left|(\mathcal{D})\sum h(x,I)\right| &\le (\mathcal{E})\sum \left|(v-u)\big(s(u)-s(v)\big)\right| \\ &\quad + \left|(\mathcal{E})\sum\big(us(u)-vs(v)\big)\right|. \end{aligned}$$

The first Riemann sum on the right satisfies

$$(\mathcal{E})\sum\left|(v-u)\big(s(u)-s(v)\big)\right| = (\mathcal{E})\sum|I|F_X(I)<\varepsilon(\mathcal{E})\sum F_X(I)<\varepsilon,$$

and by cancelation the second Riemann sum on the right satisfies

$$\Big|(\mathcal{E})\sum\big(us(u)-vs(v)\big)\Big| = \bar{v}s(\bar{v}) \le \varepsilon.$$

Thus, for every δ-fine division $\mathcal{D}$ of $[0,\infty[$, we have

$$\Big|(\mathcal{D})\sum h(x,I)\Big| < 2\varepsilon,$$

and this completes the proof. □

4. "Use the Force, Łukasz!"

At first sight, the idea that the expected value of a variable can be established from the properties of the outlook function $s(x)$ suggests precognition. It has "something of the night" about it. The following example illustrates this.

Consider two random variables X and Y jointly uniformly distributed on the unit disk; that is, the region where $x^2+y^2 \le 1$. Pick a random point (x,y) in the unit disk, and let R be the random variable for distance from the origin, so

$$R=\sqrt{X^2+Y^2}, \qquad r=\sqrt{x^2+y^2}.$$

The expected value of R is given by

$$\mathbf{E}(R)=\int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{\sqrt{x^2+y^2}}{\pi}\,dy\,dx.$$

This, of course, can be calculated by transformation into polar coordinates. But it requires the student to expend effort, care, and time. On the other hand, for $0 \le r \le 1$, let $A(r)$ denote the area within the unit disk but outside the disk of radius r centered at (0,0), so the probability that R exceeds r is $A(r)$ divided by the area of the unit disk. That is,

$$s_R(r)=P(R>r)=F_R(R>r)=\frac{A(r)}{\pi}=\frac{\pi-\pi r^2}{\pi}=1-r^2$$

with $s_R(r)=0$ for $r>1$. Since R is non-negative with probability 1,

$$\mathbf{E}(R)=\int_0^1 s_R(r)\,dr=\int_0^1\big(1-r^2\big)\,dr=1-\frac{1}{3}=\frac{2}{3}.$$

Is this not conclusive evidence of the power of the Dark Side?

REFERENCES

[1] FELLER, W.: *An Introduction to Probability Theory and its Applications.* Vol. 2. John Wiley & Sons, Inc. New York, 1966.

[2] HENSTOCK, R.: *Lectures in the Theory of Integration.* World Scientific, Singapore, 1988.

[3] HEWITT, E.: *Integration by parts for Stieltjes integrals*, Amer. Math. Monthly **67** (1960), 419–423, `http://www.jstor.org/stable/2309287`

[4] KURZWEIL, J.: *Generalized ordinary differential equations and continuous dependence on a parameter*, Czechoslovak Math. J. **82** (1957), 418–446.

Received October 31, 2010

Pat Muldowney
University of Ulster
Lisselty
Dunmore East
County Waterford
IRELAND
E-mail: pat.muldowney@gmail.com

Krzysztof Ostaszewski
Department of Mathematics
Stevenson Hall 313G
Illinois State University
Normal, IL 61790–4520
U.S.A.
E-mail: krzysio@ilstu.edu

Wojciech Wojdowski
Institute of Mathematics
Technical University of Łódź
ul. Wólczańska 215
93-005 Łódź
POLAND
E-mail: wojwoj@gmail.com