Scientific
Research
Publishing

Probability Models with Discrete and Continuous Parts

James E. Marengo, David L. Farnsworth

School of Mathematical Sciences, Rochester Institute of Technology, Rochester, New York, USA
Email: jemsma@rit.edu, dlfsma@rit.edu

Abstract

In mathematical statistics courses, students learn that the quadratic function $E\left(\left(X-x\right)^2\right)$ is minimized when x is the mean of the random variable X, and that the graphs of this function for any two distributions of X are simply translates of each other. We focus on the problem of minimizing the function defined by $y(x)=E\left(\left|X-x\right|\right)$ in the context of mixtures of probability distributions of the discrete, absolutely continuous, and singular continuous types. This problem is important, for example, in Bayesian statistics, when one attempts to compute the decision function, which minimizes the expected risk with respect to an absolute error loss function. Although the literature considers this problem, it does so only under restrictive conditions on the distribution of the random variable X, by, for example, assuming that the corresponding cumulative distribution function is discrete or absolutely continuous. By using Riemann-Stieltjes integration, we prove a theorem, which solves this minimization problem under completely general conditions on the distribution of X. We also illustrate our result by presenting examples involving mixtures of distributions of the discrete and absolutely continuous types, and for the Cantor distribution, in which case the cumulative distribution function is singular continuous. Finally, we prove a theorem that evaluates the function $y(x)$ when X has the Cantor distribution.

Keywords

Mixed-Type Distribution Function, Riemann-Stieltjes Integration, Median of a Random Variable, Cantor Distribution

1. Introduction

There are commonly used, continuous probability distributions of one variable,

such as the normal distribution and the exponential distribution. Likewise, there are discrete distributions that are well established, such as the binomial distribution and the Poisson distribution. Here, the much less routine situation in which there is a discrete component and a continuous component in a single probability distribution is addressed. These are called mixed-type probability distributions. Often, the components can be entirely separated from each other, but sometimes it might reduce the effectiveness of the probability model to do so.

An example of a mixed-type distribution is the lifetimes of electronic components. Some components may have zero lifetimes, because they are defective from the onset, giving a discrete probability component at zero ([1], p. 72-73, 121), while a continuous component is used for the remaining lifetimes. Biological lifetimes can have this feature ([2], p. 34).

Another example is the elapsed time at a stop sign at an intersection on a street. Some drivers will spend no time at the sign after stopping, because there is no cross traffic, giving the discrete component at elapsed time zero. Other drivers will linger until traffic clears, supplying the continuous component ([3], p. 63) ([4], p. 98-99).

A third example of a mixed-type distribution is lifetimes in an experiment that is terminated at a predetermined time t_d. The complete lifetimes of those subjects still living or objects that have not yet failed cannot be known. That group produces a discrete component at time t_d ([2], p. 52-63) ([5], p. 97-98).

On occasion, there might be times during an experiment, or in the course of events, at which interventions can introduce discrete components. For instance, a planned medical procedure, mass vaccination, or military campaign could be such an intervention.

There are choices for the method for proceeding. One choice is to standardize the continuous portion, so that it has a probability density. This is usually expressed with conditional distributions ([2], p. 52-63) ([5], p. 97-98). The discrete portion might be standardized separately or ignored. Another choice, which is to proceed with a mixed-type probability distribution for the whole experiment, is focused on presently.

For a discrete distribution, sums are used to compute probabilities and expected values. For continuous distributions that have a probability density function, Riemann integration is used. All discrete, continuous, and mixed-type distributions, including continuous distributions without a probability density function, which are discussed in Section 3, are covered simultaneously with Riemann-Stieltjes integration ([1], p. 118-126), ([2], p. 11-14, 34), ([6], p. 281-284).

2. An Example of a Mixed-Type Distribution

Consider the following example.

Example 1 (mixed-type distribution). In a populous state in the USA, it has been determined by using surveys that there are three distinct types of voters for an upcoming ballot initiative. One type of voter definitely opposes the initiative and is believed to be 7/20 of the voters. These individuals are coded $x = 0$.

Another type is definitely in favor of the initiative and is 2/5 of the voters. Those individuals are coded $x = 1$. The third type is not polarized on the initiative, and their degree of support x is between zero and one. They are the remaining 1/4 of the voters. For these voters, the proportion of the population is modeled with $1/5 + (1/10)x$. The random variable X is an individual voter's degree of positive support for the ballot initiative.

The $1/5 + (1/10)x$ was obtained by first fitting, via a smoothed histogram from the sample's non-polarized voters, a curve that represents the number of votes as a function of degree of support. The fit was a linear function such that the intercept term is twice the slope coefficient. Then, the values 1/5 and 1/10 were determined from the requirement that the integral of the linear function from $x = 0$ to $x = 1$ must equal the remaining fraction, 1/4, of the voters.

This is a mixed-type distribution with a discrete part for $X = 0$ and $X = 1$ and a continuous part for $X \in (0,1)$. For every random variable X, the cumulative distribution function (cdf) is defined on $\mathbb{R}$ by $F(x) = \Pr(X \le x)$. A cdf $F(x)$ has a *jump discontinuity* at $x = a$ when $\Pr(X = a) > 0$. Any discrete cdf has at most a countable number of jump discontinuities ([1], p. 74) ([7], p. 71). The continuous part of Example 1 is absolutely continuous. A cdf $F(x)$ is defined to be *absolutely continuous* if there exists a nonnegative probability density function (pdf) $f(x)$ that has $\mathbb{R}$ as its domain and $F(x) = \Pr(X \le x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$, where the integral is a Riemann integral ([7], p. 127) ([8], p. 139-140). The derivative of an absolutely continuous cdf is the pdf.

The cumulative distribution function for Example 1 is

$$y = F(x) = \begin{cases} 0 & x < 0 \\ \dfrac{7}{20} + \dfrac{1}{5}x + \dfrac{1}{20}x^2 & 0 \le x < 1 \\ 1 & 1 \le x \end{cases}, \tag{1}$$

which is displayed in **Figure 1**.

Use the fact that any cdf can be decomposed uniquely into a convex sum of a discrete cdf and a continuous cdf giving the *Jordan decomposition*

$$F(x) = c_1 F^d(x) + c_2 F^c(x), \tag{2}$$

where $c_1 \ge 0$, $c_2 \ge 0$, $c_1 + c_2 = 1$, $F^d(x)$ is a discrete cdf, and $F^c(x)$ is a continuous cdf ([1], p. 121) ([5], p. 88-90) ([8], p. 138). Then,

$$F^d(x) = \begin{cases} 0 & x < 0 \\ \dfrac{7}{15} & 0 \le x < 1 \\ 1 & 1 \le x \end{cases} \quad \text{and} \quad F^c(x) = \begin{cases} 0 & x < 0 \\ \dfrac{4}{5}x + \dfrac{1}{5}x^2 & 0 \le x < 1 \\ 1 & 1 \le x \end{cases}, \tag{3}$$

which yield the probability mass function (pmf) and probability density function (pdf)

$$f^d(x) = \begin{cases} \dfrac{7}{15} & x = 0 \\ \dfrac{8}{15} & x = 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad f^c(x) = \begin{cases} \dfrac{4}{5} + \dfrac{2}{5}x & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}, \tag{4}$$
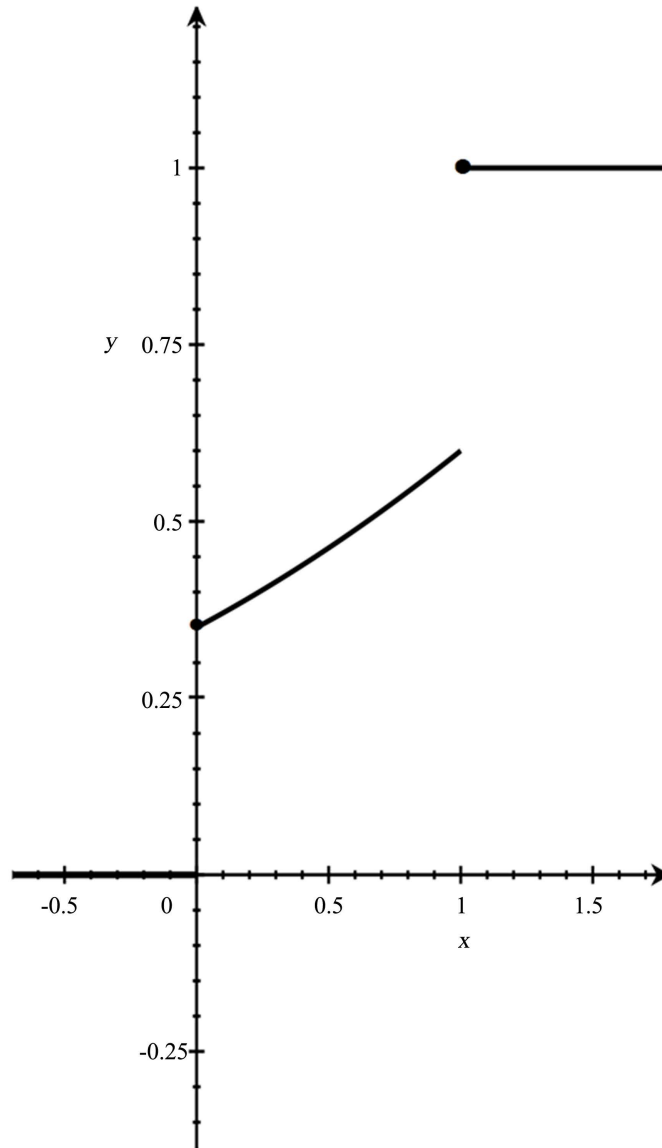
Figure 1. The cumulative distribution function $y = F(x)$ for X in Example 1.

respectively, and $c_1 = 3/4$ and $c_2 = 1/4$. The weight c_1 might be most easily computed from the jumps in the original cdf $F(x)$. Then, $\Pr(X = 0) + \Pr(X = 1) = 7/20 + 2/5 = 3/4$, which is the divisor of each of the jumps 7/20 and 2/5 from (1), in order to obtain the discrete parts of (3) and (4) and the multiplier c_1. For the continuous part, the normalizing divisor, and multiplier as well, is $c_2 = 1 - c_1 = 1/4$ or $c_2 = \int_0^1 \left(1/5 + (1/10)t \right) dt = 1/4$. Alternatively, obtain F^c from

$$F^c = \frac{F - c_1 F^d}{c_2}.$$

Use the property that the expectation of the function $g(X)$ is

$$E\big(g(X) \big) = c_1 E^d \big(g(X) \big) + c_2 E^c \big(g(X) \big).$$

where the expectations on the right-hand side are with respect to the similarly

superscripted pmf and pdf ([1], p. 121) ([3], p. 69). The left-hand side would be computed as a Riemann-Stieltjes integral, but, in Example 1, the right-hand side contains a summation and a Riemann integral. Thus, direct consideration of a Riemann-Stieltjes integration is sidestepped in this example. This formulation has the advantage of exhibiting the way that the expected value is a weighted average of the expectations with respect to the discrete and the continuous components. For $Y = g(X) = X$, the expected value is

$$\mu = E(X) = \frac{3}{4}\left((0)\frac{7}{15} + (1)\frac{8}{15}\right) + \frac{1}{4}\int_0^t t\left(\frac{4}{5} + \frac{2}{5}t\right)\mathrm{d}t = \frac{2}{5} + \frac{1}{4}\left(\frac{4}{5}\frac{1^2}{2} + \frac{2}{5}\frac{1^3}{3}\right) = \frac{8}{15}.$$

3. Singular Continuous CDFs

Any cdf $F(x)$ can be decomposed uniquely into a convex sum of a discrete cdf and a continuous cdf, as in (2). Further, the continuous component can be uniquely decomposed into an absolutely continuous component $F^{ac}(x)$ and a singular continuous component $F^{sc}(x)$, giving the *Lebesque decomposition*

$$F(x) = c_1 F^d(x) + c_2 F^{ac}(x) + c_3 F^{sc}(x),$$

where c_1, c_2, $c_3 \geq 0$, $c_1 + c_2 + c_3 = 1$ ([7], p. 131) ([8], p. 142-143) ([9], p. 10-12). A function is *singular continuous* if it is a continuous function that is not identically zero and whose first derivative exists and equals zero almost everywhere ([7], p. 131, 146-149) ([8], p. 141) ([9], p. 11). The main example is the Cantor distribution, but others, such as Minkowski's singular continuous distribution, are well-known [10] [11]. The phrase "continuous random variable" refers to a random variable that has a cdf that is everywhere continuous.

The Cantor set is created by an infinite process. Beginning with the closed interval [0,1], during the nth step of the process, remove the 2^{n-1} middle-third open intervals, each of which has length $1/3^n$. After doing that step, there remain 2^n disjoint, closed intervals. The infinite intersection of the closed sets is the Cantor set. The sum of the lengths of the deleted intervals is one, so the Cantor set has Lebesgue measure zero. The Cantor set is the support of the Cantor distribution, whose cdf is called the *devil's staircase*. This cdf fails to be differentiable at every point of the Cantor set, but its derivative is zero on the set's complement. Thus, probabilities cannot be recovered by integrating the derivative of the cdf. The devil's staircase has no jumps, and so it is continuous at every real number. It is singular continuous, because it assigns probability one to the Cantor set, which has Lebesgue measure, *i.e.* length, zero. The devil's staircase has no discrete component and no absolutely continuous component. The Cantor distribution and the devil's staircase appear in the probability and statistics literature ([7], p. 146-149) ([8], p. 35-36, 141, 146, 593), ([9], p. 13-15, 129, 174) [12] and the mathematical modeling and real analysis literature ([6], p. 80-84, 90) [10] [11] ([13], p. 249). It is the basis of Example 5 in Section 4.4.

The first seven omitted intervals, where the devil's staircase has slope zero, and the accompanying values of the cdf are

$$F(x) = \begin{cases} \dfrac{1}{8} & \dfrac{1}{27} < x < \dfrac{2}{27} \\[1mm] \dfrac{1}{4} & \dfrac{1}{9} < x < \dfrac{2}{9} \\[1mm] \dfrac{3}{8} & \dfrac{7}{27} < x < \dfrac{8}{27} \\[1mm] \dfrac{1}{2} & \dfrac{1}{3} < x < \dfrac{2}{3} \\[1mm] \dfrac{5}{8} & \dfrac{19}{27} < x < \dfrac{20}{27} \\[1mm] \dfrac{3}{4} & \dfrac{7}{9} < x < \dfrac{8}{9} \\[1mm] \dfrac{7}{8} & \dfrac{25}{27} < x < \dfrac{26}{27} \end{cases}$$

and are graphed in **Figure 2**.

4. Medians

A *median* of the random variable X, and therefore of F, is any real number m such that

$$\Pr(X \le m) \ge 1/2 \quad \text{and} \quad \Pr(X \ge m) \ge 1/2,$$

or, equivalently,

$$\Pr(X \le m) \ge 1/2 \quad \text{and} \quad \Pr(X < m) \le 1/2. \tag{5}$$

For Example 1, the median of X is

$$m = \sqrt{7} - 2 = 0.646, \tag{6}$$

which is obtained from (1) by solving

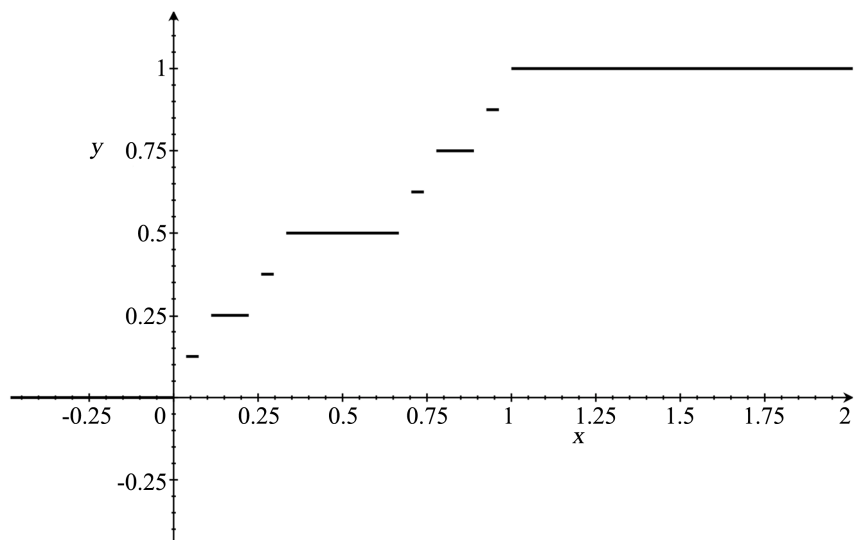$$\frac{7}{20} + \frac{1}{5}m + \frac{1}{20}m^2 = \frac{1}{2}.$$

Figure 2. Portions of the devil's staircase, which is the cdf of the Cantor distribution.

The main purpose of this section is to show that, if m is a median of the univariate random variable X and $a \in \mathbb{R}$, then $E\left(|X - a|\right) \geq E\left(|X - m|\right)$, and this inequality is strict if a is not a median of X. To avoid trivialities, assume that these expectations exist as finite real numbers, which is assured by presupposing that $E\left(|X|\right) \in \mathbb{R}$. Mood, Graybill, and Boes ([3], p. 83), Hogg, McKean, and Craig ([14], p. 58), and Parzen ([15], p. 213) consider this inequality for absolutely continuous cdfs. Rohatgi ([5], p. 170-171) and Dwass ([16], p. 341-342) consider it separately for discrete and for absolutely continuous cdfs. The advantage of using Riemann-Stieltjes integration is that it covers the inequality for any discrete, continuous, and mixed type cdfs without exception with a single argument, which is presented in Theorem 1.

4.1. Preliminaries

The expectation $E\left(|X - x|\right)$ is a convex function of x. Indeed, for $c, d \in \mathbb{R}$ and $t \in (0,1)$, using the triangle inequality and the linearity of expectation,

$$
\begin{aligned}
E\left(\left|X - \left(tc + (1-t)d\right)\right|\right) &= E\left(\left|t(X - c) + (1-t)(X - d)\right|\right) \\
&\leq E\left(\left|t(X - c)\right| + \left|(1-t)(X - d)\right|\right) \\
&= tE\left(\left|(X - c)\right|\right) + (1-t)E\left(\left|(X - d)\right|\right).
\end{aligned}
$$

All convex functions are continuous ([6], p. 199) ([17], p. 149-152).

For $c \in \mathbb{R}$, define

$$
F(c-) = \lim_{\substack{x \to c \\ x < c}} F(x).
$$

Because $F(c-) = \Pr(X < c)$,

$$
\Pr(X = c) = F(c) - F(c-)
$$

([14], p. 38). Since F is right continuous ([1], p. 71) ([7], p. 70-71),

$$
F(c+) = \lim_{\substack{x \to c \\ x > c}} F(x) = F(c).
$$

Thus, (5) can be expressed

$$
F(m-) \leq \frac{1}{2} \quad \text{and} \quad \frac{1}{2} \leq F(m). \tag{7}
$$

Assuming that g is a continuous positive function on the interval (c, d) and $F(c) < F(d-)$,

$$
\int_{(c,d)} g(x)\,\mathrm{d}F(x) > 0 \tag{8}
$$

([1], p. 118-119) ([6], p. 281-284). For $F(c) < F(d-)$,

$$
\int_{(c,d)} \mathrm{d}F(x) = F(d-) - F(c).
$$

4.2. Lemma 1

Lemma 1. *Let X be a random variable with the cumulative distribution func-*

tion of F. Suppose that $E(X) \in \mathbb{R}$. Then, for any a and $b \in \mathbb{R}$,

$$E(|X-a|) - E(|X-b|)$$

$$= \begin{cases} 2\int\limits_{(a,b)} (x-a)dF(x) + (b-a)(1-2F(b-)) & a < b \quad (9) \\ 2\int\limits_{(b,a)} (a-x)dF(x) + (a-b)(2F(b)-1) & b < a \quad (10) \end{cases}$$

Proof. Observe that the assumption that $E(X) \in \mathbb{R}$ is equivalent to $E(|X|) < \infty$. If $c \in \mathbb{R}$, then $E(|X-c|) \le E(|X|) + |c| < \infty$ and

$$E(|X-c|) = \int\limits_{(-\infty,\infty)} |x-c| dF(x).$$

Consider two cases.

Case 1 ($a < b$). Expanding gives

$$E(|X-a|) = \int\limits_{(-\infty,a)} (a-x)dF(x) + \int\limits_{(a,b)} (x-a)dF(x)$$
$$+ \int\limits_{\{b\}} (b-a)dF(x) + \int\limits_{(b,\infty)} (x-a)dF(x)$$

and

$$E(|X-b|) = \int\limits_{(-\infty,a)} (b-x)dF(x) + \int\limits_{\{a\}} (b-a)dF(x)$$
$$+ \int\limits_{(a,b)} (b-x)dF(x) + \int\limits_{(b,\infty)} (x-b)dF(x).$$

Substituting gives

$$E(|X-a|) - E(|X-b|)$$
$$= \int\limits_{(-\infty,a)} (a-b)dF(x) + \int\limits_{(a,b)} (2x-a-b)dF(x) + \int\limits_{(b,\infty)} (x-a)dF(x)$$
$$+ (b-a)(F(b)-F(b-)) - (b-a)(F(a)-F(a-))$$
$$= (a-b)F(a-) + \int\limits_{(a,b)} (2(x-a)+a-b)dF(x) + (b-a)(1-F(b))$$
$$+ (b-a)(F(b)-F(b-)) - (b-a)(F(a)-F(a-))$$
$$= (a-b)F(a-) + 2\int\limits_{(a,b)} (x-a)dF(x) + (a-b)(F(b-)-F(a))$$
$$+ (b-a)(1-F(b)) + (b-a)(F(b)-F(b-)),$$

which yields (9).

Case 2 ($b < a$). Interchanging the roles of *a* and *b* in (9) gives

$$E(|X-b|) - E(|X-a|)$$
$$= 2\int\limits_{(b,a)} (x-b)dF(x) + (a-b)(1-2F(a-))$$
$$= 2\int\limits_{(b,a)} ((x-a)-(b-a))dF(x) + (a-b)(1-2F(a-))$$
$$= 2\int\limits_{(b,a)} (x-a)dF(x) - 2(b-a)(F(a-)-F(b)) + (a-b)(1-2F(a-))$$
$$= 2\int\limits_{(b,a)} (x-a)dF(x) + (a-b)(1-2F(b)).$$

Multiplying by −1 yields (10).

4.3. Theorem 1

Theorem 1. *Let X be a random variable. Suppose that $E(X) \in \mathbb{R}$ and $m \in \mathbb{R}$ is a median of X. Then, for any $a \in \mathbb{R}$,*

$$E(|X-a|) \geq E(|X-m|). \tag{11}$$

The inequality is strict if a is not a median of X.

Proof. From the proof of Lemma 1, $E(|X-a|) < \infty$. Set $b = m$ in Lemma 1. The integrals in (9) and (10) are nonnegative. Inequality (11) follows, because, using (7), the second terms on the right-hand sides of (9) and (10) are also nonnegative. To show that the inequality (11) is strict when a is not a median of X, consider the two cases in Lemma 1.

Case 1 ($a < m$). From (7), either $F(m–) < 1/2$ or $F(m–) = 1/2$. If $F(m–) < 1/2$, then $(m – a)(1 – 2F(m–)) > 0$, so that the right-hand side of (9) is positive. If $F(m–) = 1/2$, then $(m – a)(1 – 2F(m–)) = 0$. Also, $F(a) \leq F(m–) = 1/2$. Because a is not a median, $F(a) \neq 1/2$ and, thus, $F(a) < 1/2$, $F(a) < F(m–)$, and the integral in (9) is positive from (8).

Case 2 ($m < a$). From (5), either $1/2 < F(m)$ or $1/2 = F(m)$. If $1/2 < F(m)$, then $(a – m)(2F(m) – 1) > 0$, so that the right-hand side of (10) is positive. If $F(m) = 1/2$, then $(a – m)(2F(m) – 1) = 0$. Also, $1/2 = F(m) \leq F(a–)$. Because a is not a median, $F(a–) \neq 1/2$ and, thus, $F(a–) > 1/2$, $F(m) < F(a–)$, and the integral in (10) is positive from (8).

4.4. Representative Examples

The following examples display the function $y = E(|X-x|)$ and the locations of the medians for various distributions. The graphs illustrate that $y = E(|X-x|)$ is a convex and continuous function and that medians occur as single points or as all of the values in an interval.

Example 1 revisited (mixed-type distribution). For the voter preference example,

$$y = E(|X-x|) = \begin{cases} \dfrac{8}{15} - x & x < 0 \\[2mm] \dfrac{1}{30}\left(x^3 + 6x^2 - 9x + 16\right) & 0 \leq x < 1, \\[2mm] x - \dfrac{8}{15} & 1 \leq x \end{cases}$$

which is displayed in **Figure 3**. The minimum occurs at $x = m = \sqrt{7} - 2 \approx 0.646$, as computed in (6).

Example 2 (absolutely continuous cdf). For the exponential random variable with pdf $f(x) = e^{-x/3}/3$ for $x > 0$ and zero otherwise, **Figure 4** displays

$$y = E(|X-x|) = \int_0^\infty |t-x| \frac{1}{3} e^{-t/3} \, dt = \begin{cases} 3 - x & x \leq 0 \\ x - 3 + 3e^{-x/3} & x > 0 \end{cases}.$$
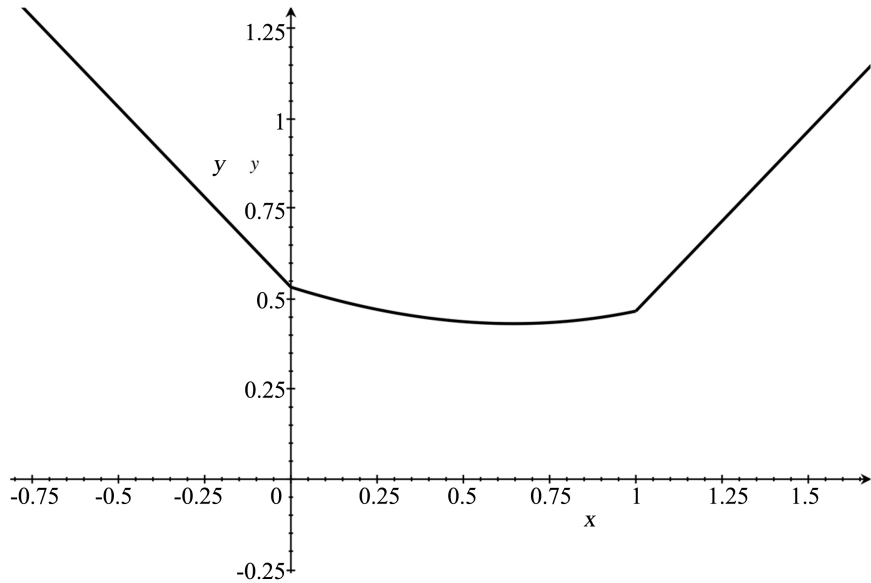
Figure 3. $y = E(|X - x|)$ for the mixed-type distribution in Example 1.
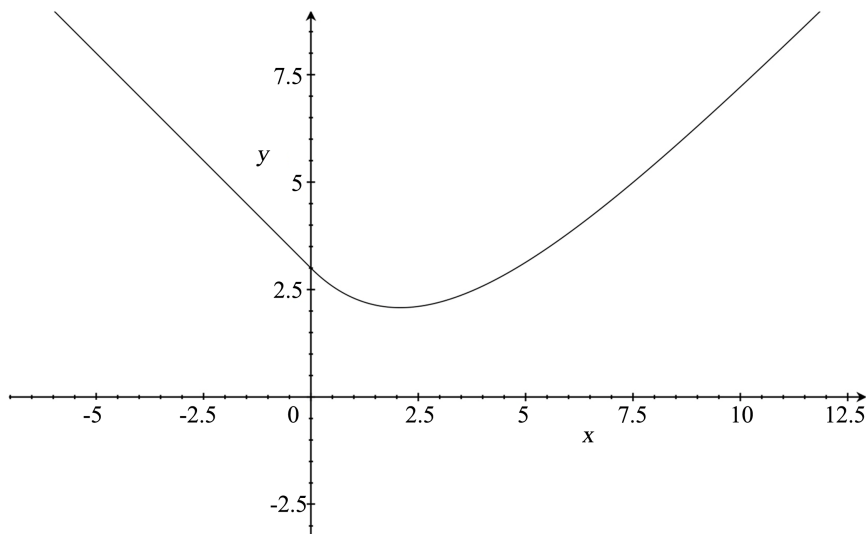

Figure 4. $y = E(|X - x|)$ for the exponential distribution with mean $\mu = 3$ in Example 2.

The expected value is $\mu = 3$. The minimum is at the unique median $m = 3\ln2 \approx 2.08$.

Example 3 (discrete cdf and a single median). For the binomial distribution with $n = 3$ and $p = 0.7$, the expected value is $\mu = np = 2.1$. **Figure 5** displays

$$y = E(|X - x|) = \begin{cases} -x + 2.1 & x < 0 \\ -0.946x + 2.1 & 0 \le x < 1 \\ -0.568x + 1.722 & 1 \le x < 2 \\ 0.314x - 0.042 & 2 \le x < 3 \\ x - 2.1 & 3 \le x \end{cases}.$$
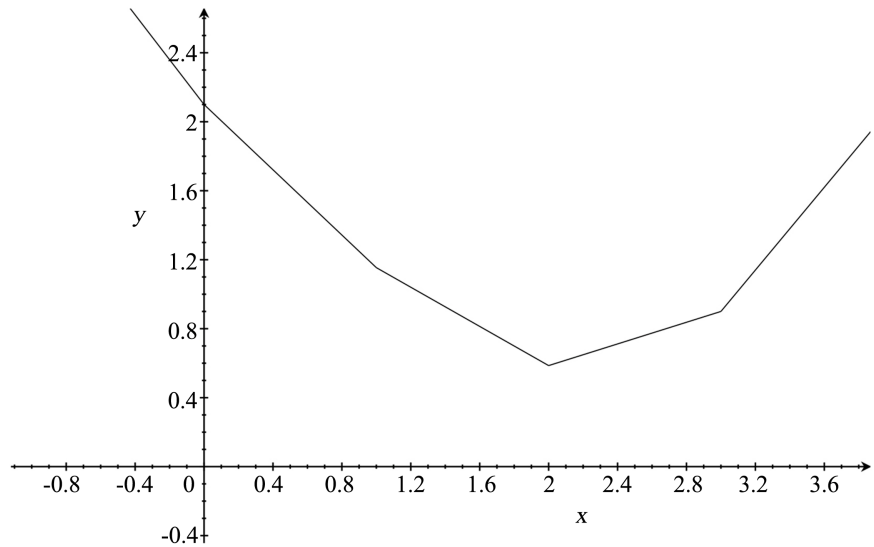
The minimum is at the unique median $m = 2$.

Figure 5. $y = E(|X - x|)$ for the binomial distribution with $n = 3$ and $p = 0.7$ in Example 3.

Example 4 (discrete cdf and an interval of medians). For the binomial distribution with $n = 5$ and $p = 0.5$, the expected value is $\mu = np = 2.5$. **Figure 6** displays

$$y = E(|X - x|) = \begin{cases} -x + 2.5 & x < 0 \\ -0.9375x + 2.5 & 0 \le x < 1 \\ -0.625x + 2.1875 & 1 \le x < 2 \\ 0.9375 & 2 \le x < 3 \\ 0.625x - 0.9375 & 3 \le x < 4 \\ 0.9375x - 2.1875 & 4 \le x < 5 \\ x - 2.5 & 5 \le x \end{cases}.$$

Note that every number in the interval [2, 3] is a median.

Example 5 (singular continuous distribution). The Cantor distribution has mean 0.5 and median any $m \in (1/3, 2/3)$. Because the derivation of an expression for $y = E(|X - x|)$ is more complicated than the previous examples, details are presented.

During the nth step of the process that leads to the Cantor set, remove the 2^{n-1} middle-third open intervals, each of which has length $1/3^n$. After doing such a step, there remain 2^n disjoint, closed intervals, which are denoted by

$$I_k^{(n)} = \left[a_k^{(n)}, b_k^{(n)} \right]$$

for $k = 1, 2, \cdots, 2^n$, where

$$a_1^{(n)} < a_2^{(n)} < \cdots < a_{2^n}^{(n)}.$$

The following lemma provides a recursive formula that is used for computing $\int_{I_k^{(n)}} x \mathrm{d}F(x)$.
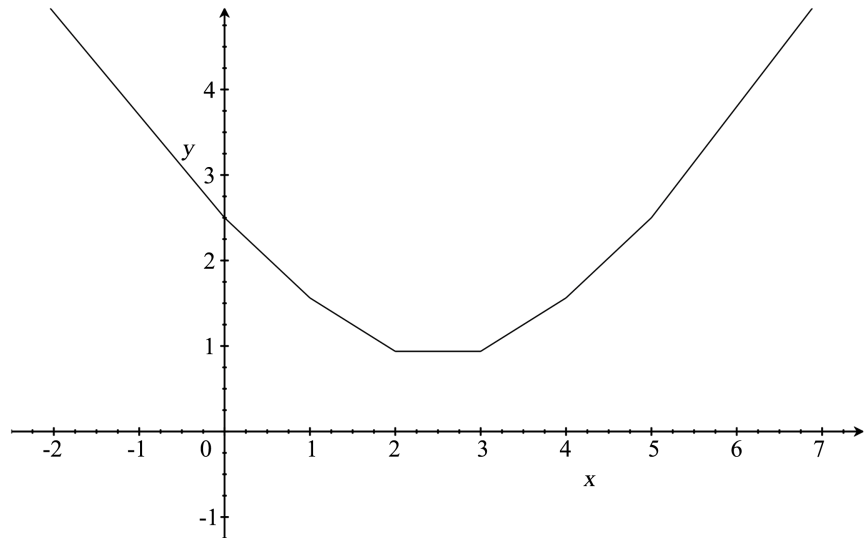
Figure 6. $y = E\left(\left|X - x\right|\right)$ for the binomial distribution with $n = 5$ and $p = 0.5$ in Example 4.

Lemma 2. *Define*

$$c_1^{(0)} = \int_0^1 x \mathrm{d}F(x) \quad and \quad c_k^{(n)} = \int_{a_k^{(n)}}^{b_k^{(n)}} x \mathrm{d}F(x)$$

for $n \geq 1$. Then

$$c_1^{(0)} = \frac{1}{2}$$

and

$$c_k^{(n)} = \begin{cases} \dfrac{1}{6} c_k^{(n-1)} & k = 1, 2, \cdots, 2^{n-1} \\ \dfrac{1}{2^n} - c_{2^n + 1 - k}^{(n)} & k = 2^{n-1} + 1, 2^{n-1} + 2, \cdots, 2^n \end{cases}.$$

Proof. Because

$$F\left(\frac{x}{3}\right) = \frac{F(x)}{2} \quad and \quad F(1-x) = 1 - F(x)$$

for $x \in [0,1]$ ([9], p. 15),

$$\mathrm{d}F\left(\frac{x}{3}\right) = \frac{\mathrm{d}F(x)}{2} \quad and \quad \mathrm{d}F(1-x) = -\mathrm{d}F(x).$$

Hence,

$$1 - c_1^{(0)} = 1 - \int_0^1 x \mathrm{d}F(x) = 1 - \int_0^1 (1-t) \mathrm{d}F(t) = \int_0^1 t \mathrm{d}F(t) = c_1^{(0)}.$$

Thus,

$$c_1^{(0)} = \frac{1}{2}.$$

Consider $n \geq 1$. For $1 \leq k \leq 2^{n-1}$,

$$a_k^{(n)} = \frac{1}{3} a_k^{(n-1)}, \, b_k^{(n)} = \frac{1}{3} b_k^{(n-1)},$$

and

$$c_k^{(n)} = \int_{a_k^{(n)}}^{b_k^{(n)}} x \mathrm{d}F(x) = \int_{a_k^{(n-1)}}^{b_k^{(n-1)}} \frac{t}{3} \mathrm{d}F\left(\frac{t}{3}\right) = \int_{a_k^{(n-1)}}^{b_k^{(n-1)}} \frac{t}{3} \frac{\mathrm{d}F(t)}{2} = \frac{1}{6} c_k^{(n-1)}.$$

Let $2^{n-1} + 1 \leq k \leq 2^n$. From

$$a_k^{(n)} = 1 - b_{2^n+1-k}^{(n)}, \, b_k^{(n)} = 1 - a_{2^n+1-k}^{(n)}, \quad \text{and} \quad F\left(b_{2^n+1-k}^{(n)}\right) - F\left(a_{2^n+1-k}^{(n)}\right) = \frac{1}{2^n},$$

it follows that

$$c_k^{(n)} = \int_{a_k^{(n)}}^{b_k^{(n)}} x \mathrm{d}F(x) = \int_{a_{2^n+1-k}^{(n)}}^{b_{2^n+1-k}^{(n)}} (1-t) \mathrm{d}F(t) = \frac{1}{2^n} - c_{2^n+1-k}^{(n)}.$$

The numerical values of the definite integrals $c_k^{(n)}$ for $n = 1, 2$, and 3 are in **Table 1**.

Theorem 2. *Let $J_1^{(n)}, J_2^{(n)}, \cdots, J_{2^{n-1}}^{(n)}$ be the disjoint open intervals that are removed during the nth step of the construction of the Cantor set, where the midpoints of these intervals are strictly increasing. For $x \in J_k^{(n)}$,*

$$y = y(x) = E\left(|X - x|\right) = \frac{2k - 1 - 2^{n-1}}{2^{n-1}} x - \sum_{j=1}^{2k-1} c_j^{(n)} + \sum_{j=2k}^{2^n} c_j^{(n)}. \tag{12}$$

Proof. For $x \in J_k^{(n)}$, it follows from Lemma 2 that

$$y = y(x) = E\left(|X - x|\right) = \int_{-\infty}^{\infty} |t - x| \mathrm{d}F(t) = \sum_{j=1}^{2^n} \int_{I_j^{(n)}} |t - x| \mathrm{d}F(t)$$

$$= \sum_{j=1}^{2k-1} \int_{I_j^{(n)}} (t - x) \mathrm{d}F(t) + \sum_{j=2k}^{2^n} \int_{I_j^{(n)}} (t - x) \mathrm{d}F(t)$$

$$= \sum_{j=1}^{2k-1} \left(\frac{x}{2^n} - c_j^{(n)}\right) + \sum_{j=2k}^{2^n} \left(c_j^{(n)} - \frac{x}{2^n}\right)$$

$$= \frac{2k - 1 - 2^{n-1}}{2^{n-1}} x - \sum_{j=1}^{2k-1} c_j^{(n)} + \sum_{j=2k}^{2^n} c_j^{(n)}.$$

Since the complement of the Cantor set in [0, 1] is dense in that interval, and since the value of a continuous function at any number in [0, 1] is determined by its values on a dense subset ([6], p. 121), Theorem 2 determines the value of $E\left(|X - x|\right)$ for $x \in [0, 1]$. Additionally, $y(x) = 1/2 - x$ for every $x < 0$ and $y(x) = x - 1/2$ for $x > 0$.

For $k = 1, 2, \cdots, 2^{n-1}$, the graph of y on $J_k^{(n)}$ is a line segment with slope

$$\frac{2k - 1 - 2^{n-1}}{2^{n-1}}.$$

Equation (13) and **Figure 7** display the values and give the graph of $y = y(x)$ for the Cantor distribution for $J_k^{(n)}$ with $n = 1, 2$, and 3.

Table 1. Numerical values of $c_k^{(n)} = \int_{a_k^{(n)}}^{b_k^{(n)}} x \, \mathrm{d}F(x)$ for $n = 1, 2, 3$ and $k = 1, 2, \cdots, 2^n$.

n	k	
1	1	$c_1^{(1)} = \dfrac{1}{6} c_1^{(0)} = \dfrac{1}{6} \dfrac{1}{2} = \dfrac{1}{12}$
	2	$c_2^{(1)} = \dfrac{1}{2} - c_1^{(1)} = \dfrac{1}{2} - \dfrac{1}{12} = \dfrac{5}{12}$
2	1	$c_1^{(2)} = \dfrac{1}{6} c_1^{(1)} = \dfrac{1}{6} \dfrac{1}{12} = \dfrac{1}{72}$
	2	$c_2^{(2)} = \dfrac{1}{6} c_2^{(1)} = \dfrac{1}{6} \dfrac{5}{12} = \dfrac{5}{72}$
	3	$c_3^{(2)} = \dfrac{1}{2^2} - c_2^{(2)} = \dfrac{1}{4} - \dfrac{5}{72} = \dfrac{13}{72}$
	4	$c_4^{(2)} = \dfrac{1}{2^2} - c_1^{(2)} = \dfrac{1}{4} - \dfrac{1}{72} = \dfrac{17}{72}$
3	1	$c_1^{(3)} = \dfrac{1}{6} c_1^{(2)} = \dfrac{1}{6} \dfrac{1}{72} = \dfrac{1}{432}$
	2	$c_2^{(3)} = \dfrac{1}{6} c_2^{(2)} = \dfrac{1}{6} \dfrac{5}{72} = \dfrac{5}{432}$
	3	$c_3^{(3)} = \dfrac{1}{6} c_3^{(2)} = \dfrac{1}{6} \dfrac{13}{72} = \dfrac{13}{432}$
	4	$c_4^{(3)} = \dfrac{1}{6} c_4^{(2)} = \dfrac{1}{6} \dfrac{17}{12} = \dfrac{17}{432}$
	5	$c_5^{(3)} = \dfrac{1}{2^3} - c_4^{(3)} = \dfrac{1}{8} - \dfrac{17}{432} = \dfrac{37}{432}$
	6	$c_6^{(3)} = \dfrac{1}{2^3} - c_3^{(3)} = \dfrac{1}{8} - \dfrac{13}{432} = \dfrac{41}{432}$
	7	$c_7^{(3)} = \dfrac{1}{2^3} - c_2^{(3)} = \dfrac{1}{8} - \dfrac{5}{432} = \dfrac{49}{432}$
	8	$c_8^{(3)} = \dfrac{1}{2^3} - c_1^{(3)} = \dfrac{1}{8} - \dfrac{1}{432} = \dfrac{53}{432}$

$$y = E\left(|X - x|\right) = \begin{cases} 1/2 - x & x < 0 \\ 107/216 - (3/4)x & 1/27 < x < 2/27 \\ 17/36 - (1/2)x & 1/9 < x < 2/9 \\ 89/216 - (1/4)x & 7/27 < x < 8/27 \\ 1/3 & 1/3 < x < 2/3 \\ (1/4)x + 35/216 & 19/27 < x < 20/27 \\ (1/2)x - 1/36 & 7/9 < x < 8/9 \\ (3/4)x - 55/216 & 25/27 < x < 26/27 \\ x - 1/2 & 1 < x \end{cases} \tag{13}$$
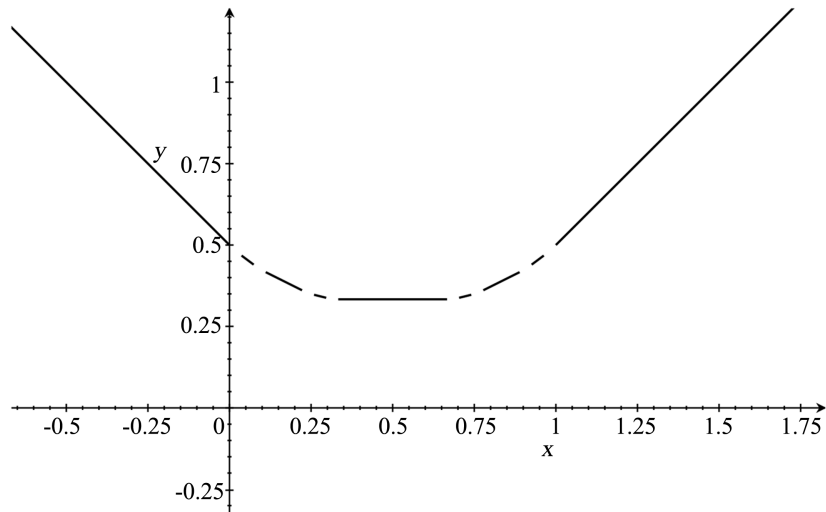
Figure 7. A portion of the graph for $y = E(|X - x|)$ for the Cantor distribution in Example 5.

For a sample calculation, using (12) and **Table 1**, take $n = 2$ and $k = 1$ in (13),

$$y = \frac{2(1) - 1 - 2^{2-1}}{2^{2-1}} x - \frac{1}{72} + \frac{5}{72} + \frac{13}{72} + \frac{17}{72} = \frac{17}{36} - \frac{1}{2} x.$$

5. Concluding Comments

The necessary and sufficient condition that x minimizes $y(x) = E(|X - x|)$ for medians of the distribution of X has been established under completely general conditions on the distribution of X and illustrated for various pure and mixed cumulative distribution functions, including the devil's staircase of the Cantor distribution. We have evaluated the function $y(x) = E(|X - x|)$ when X has the Cantor distribution, and we have also demonstrated the way in which the nature of this function depends heavily on the distribution of X in other cases. By way of contrast, the minimum value of the quadratic function $E((X - x)^2)$ is $Var(X)$, and the minimizing value of x is $E(X)$. Its graph depends only on this mean and variance, and any two distributions for X yield graphs that are translates of each other.

Acknowledgements

The authors want to thank the anonymous referees for many insightful comments.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1]   Lindgren, B.W. (1976) Statistical Theory. 3rd Edition, Macmillan, New York.

[2]   Lawless, J.F. (2003) Statistical Models and Methods for Lifetime Data. 2nd Edition, Wiley, Hoboken. https://doi.org/10.1002/9781118033005

[3]   Mood, A.M., Graybill, F.A. and Boes, D.C. (1974) Introduction to the Theory of Statistics. 3rd Edition, McGraw-Hill, New York.

[4]   Bain, L.J. and Engelhardt, M. (1987) Introduction to Probability and Mathematical Statistics. Duxbury Press, Boston.

[5]   Rohatgi, V.K. (2003) Statistical Inference. Dover, Mineola.

[6]   Stromberg, K.R. (1981) An Introduction to Classical Analysis. Wadsworth, Belmont.

[7]   Dudewicz, E.J. and Mishra, S.N. (1998) Modern Mathematical Statistics. Wiley, New York.

[8]   Feller, W. (1971) An Introduction to Probability Theory and Its Applications. Volume 2, 2nd Edition, Wiley, New York.

[9]   Chung, K.L. (2001) A Course in Probability Theory. 3rd Edition, Academic Press, San Diego.

[10]   Salem, R. (1943) On Some Singular Monotonic Functions Which Are Strictly Increasing. *Transactions of the American Mathematical Society*, **53**, 427-439. https://doi.org/10.1090/S0002-9947-1943-0007929-6

[11]   Bernstein, D. (2013) Algorithmic Definitions of Singular Functions. Department of Mathematics, Davidson College, Davidson.

[12]   Lad, F.R. and Taylor, W.F.C. (1992) The Moments of the Cantor Distribution. *Statistics and Probability Letters*, **13**, 307-310. https://doi.org/10.1016/0167-7152(92)90039-8

[13]   Apostol, T.M. (1957) Mathematical Analysis. Addison-Wesley, Reading.

[14]   Hogg, R.V., McKean, J.W. and Craig, A.T. (2005) Introduction to Mathematical Statistics. 6th Edition, Pearson Prentice Hall, Upper Saddle River.

[15]   Parzen, E. (1960) Modern Probability Theory and Its Applications. Wiley, New York. https://doi.org/10.1063/1.3056709

[16]   Dwass, M. (1970) Probability and Statistics. W.A. Benjamin, New York.

[17]   Bauschke, H.H. and Combettes, P.L. (2017) Convex Analysis and Monotone Operator Theory in Hilbert Spaces. 2nd Edition, Springer, Cham. https://doi.org/10.1007/978-3-319-48311-5