

Effect of Rounding on Expected Value and Standard Deviation

Anton Cedilnik¹ and Katarina Košmelj²

Abstract

We prove that the rounded expected value of the rounded random variable differs from the expected value of the original random variable by less than 2δ , where δ is the distance between two neighbouring rounded numbers. The same conclusion holds for standard deviation. Hence, in practice the influence of rounding on expected value and standard deviation is almost negligible.

1 Introduction

Usually, rounding (Everitt, 2002; Lisman, 1985) is considered as a procedure for reporting numerical information to fewer digits than used during data collection or analysis. The most frequent rounding procedure is to discard all decimal digits right of the d -th one which may be corrected in a well known way; the set of all rounded values form the arithmetic sequence of all rational numbers of the form $integer \cdot 10^{-d}$. In this paper, we understand rounding more generally: it is a procedure for reporting real-numerical information as a subset of an arbitrary arithmetic sequence.

Suppose that rounded values form an arithmetic sequence $(\epsilon + n\delta)_{n \in \mathbb{Z}}$. δ is the *rounding level* and is a positive real number, and the *shift* ϵ is the first non-negative rounded value ($0 \leq \epsilon < \delta$). The value $\epsilon + n\delta$ is the round-off of the interval

$$I_n = [\epsilon + (n - \eta)\delta, \epsilon + (n + 1 - \eta)\delta),$$

where η ($0 \leq \eta < 1$), if different of $\frac{1}{2}$, presents the *asymmetry* of rounding. This rounding is described by the function

$$x \mapsto \langle x \rangle := \epsilon + \text{int} \left(\frac{x - \epsilon}{\delta} + \eta \right) \cdot \delta,$$

where $\text{int}(\cdot)$ denotes the *integer function* (also known as *floor function*)

$$\text{int}(t) = [t] := \max\{m \mid m \in \mathbb{Z} \wedge m \leq t\}.$$

Example A. Rounding to d decimal places (or to $|d|$ integer places, if $d < 0$): $\delta = 10^{-d}$, $\epsilon = 0$, $\eta = \frac{1}{2}$. It is worth noting that the standard rounding procedure differs slightly from

¹Biotechnical Faculty, University of Ljubljana, Slovenia; anton.cedilnik@bf.uni-lj.si

²Biotechnical Faculty, University of Ljubljana, Slovenia; katarina.kosmelj@bf.uni-lj.si

our one if we cut off an exact 5. For example, for $d = -2$ the standard procedure rounds 7450 to 7400 (the last retained digit on the right should be even); but, according to our definition, $\langle 7450 \rangle = 7500$.

Example B. The value of a person's age is usually truncated to its integer part: $\delta = 1$, $\epsilon = 0$, $\eta = 0$.

Example C. The number of car accidents is given in classes, e.g., 0 to 9, 10 to 19, ... The values of each class are rounded to its mid-point. The class width determines the rounding level, $\delta = 10$, the shift (the mid-point of the first class) is $\epsilon = 4,5$, and $\eta = \frac{1}{2}$.

The problem of rounding random variable occurs most frequently in the following two cases.

1. A discrete random variable is given entirely, but with rounded values (as in Example B, the ages of some minor population).

2. By experimenting we obtain a huge number of rounded values of certain random variable, so huge that the Law of large numbers already works perfectly, and we would wish to know the relation between the expected value and the arithmetic mean of obtained values.

Suppose that X is a real random variable with the distribution function $F(x)$. Its rounded image $\langle X \rangle$ is then a discrete random variable

$$\langle X \rangle = \epsilon + \text{int} \left(\frac{X - \epsilon}{\delta} + \eta \right) \cdot \delta \sim \left[\begin{array}{ccc} \cdots & \epsilon + n\delta & \cdots \\ \cdots & v_n & \cdots \end{array} \right]_{n \in \mathbb{Z}} \quad (1.1)$$

$$v_n = F(\rightarrow \epsilon + (n + 1 - \eta)\delta) - F(\rightarrow \epsilon + (n - \eta)\delta)$$

(where \rightarrow indicates the left limit of possibly discontinuous distribution function).

The objective of our work is to assess the effect of rounding on the expected value and standard deviation. The following four differences will be under study: the two of theoretical significance

$$E(\langle X \rangle) - E(X), \quad (1.2)$$

$$\sigma(\langle X \rangle) - \sigma(X), \quad (1.3)$$

and the two of more practical significance

$$\langle E(\langle X \rangle) \rangle - E(X), \quad (1.4)$$

$$\langle \sigma(\langle X \rangle) \rangle - \sigma(X). \quad (1.5)$$

Let us write $\langle X \rangle = X + W$, where the perturbation W is a bounded random variable

$$W = \eta\delta - \left[\left(\frac{X - \epsilon}{\delta} + \eta \right) - \text{int} \left(\frac{X - \epsilon}{\delta} + \eta \right) \right] \delta$$

with the obvious range

$$W = \langle X \rangle - X \in (-(1 - \eta)\delta, \eta\delta]. \quad (1.6)$$

The existence of the expected value $E(\langle X \rangle)$ and standard deviation $\sigma(\langle X \rangle)$ has its justification in the following lemma with a simple proof which we omit.

LEMMA. *Let X be any random variable and W a bounded random variable. $X + W$ has an n -th moment about origin (and accordingly all other moments of the same or lower orders) if and only if X has it.*

2 Expected value

Taking into account (1.6), we immediately obtain the following result on the difference (1.2):

$$E(W) = E(\langle X \rangle) - E(X) \in (-(1 - \eta)\delta, \eta\delta], \quad (2.1)$$

which is an interval of the length δ .

Now, we consider the difference (1.4). Denote $a = (E(X) + E(W) - \epsilon)/\delta + \eta$. Then $\langle E(\langle X \rangle) \rangle - E(X) = \epsilon + \text{int}(a) \cdot \delta - E(X) = \epsilon + \text{int}(a) \cdot \delta - ((a - \eta)\delta + \epsilon - E(W)) = E(W) + \eta\delta - (a - \text{int}(a))\delta$. Since $\eta\delta - (a - \text{int}(a))\delta \in (-(1 - \eta)\delta, \eta\delta]$, then

$$\langle E(\langle X \rangle) \rangle - E(X) \in (-2(1 - \eta)\delta, 2\eta\delta], \quad (2.2)$$

which is an interval of the length 2δ .

The intervals in (2.1) and (2.2) are as narrow as possible for arbitrary values of the parameters δ , ϵ , and η of the rounding function (1.1). The example where the upper bound in (2.1) and (2.2) is attained is

$$X_1 \sim \begin{bmatrix} \epsilon - \eta\delta & \epsilon - \eta\delta + \delta \\ \eta & 1 - \eta \end{bmatrix},$$

and the example for both lower bounds is

$$X_2 \sim \begin{bmatrix} \epsilon - \eta\delta - \omega\delta & \epsilon - \eta\delta + \delta - \omega\delta \\ \eta + \omega & 1 - \eta - \omega \end{bmatrix},$$

where ω should be extremely small positive.

3 Standard deviation

Let S be a random variable with $P(\alpha \leq S \leq \beta) = 1$. Then $\sigma(S) \leq \frac{\beta - \alpha}{2}$, and $\sigma(S) = \frac{\beta - \alpha}{2}$ if and only if

$$S \sim \begin{bmatrix} \alpha & \beta \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Hence, if $P(\alpha < S \leq \beta) = 1$ then $\sigma(S) < \frac{\beta - \alpha}{2}$. We shall use this fact to estimate the standard deviation of W .

Suppose that X has a variance as well. Firstly, we estimate $\sigma(W) < \delta/2$, since the length of the interval $-(1 - \lambda)\delta, \lambda\delta]$ from (1.6) is precisely δ . Then, also using the Cauchy - Schwarz inequality (Jamnik, 1971),

$$\sigma(X) - \delta/2 < \sigma(X) - \sigma(W) \leq \sigma(X + W) \leq \sigma(X) + \sigma(W) < \sigma(X) + \delta/2.$$

Shortly,

$$\sigma(\langle X \rangle) - \sigma(X) \in (-\delta/2, \delta/2) \quad (3.1)$$

which is the result about the difference (1.3).

The following two random variables are examples which confirm that the interval in (2.2) is as narrow as possible, X_3 for the upper bound, and X_4 for the lower bound.

$$X_3 \sim \begin{bmatrix} \epsilon - \eta\delta - \omega & \epsilon - \eta\delta \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix},$$

$$X_4 \sim \begin{bmatrix} \epsilon - \eta\delta & \epsilon - \eta\delta + \delta - \omega \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix},$$

where ω is extremely small positive.

Now, let us deal with the last difference (1.5). Denote $a = (\sigma(\langle X \rangle) - \epsilon)/\delta + \eta$. Then

$$\begin{aligned} \langle \sigma(\langle X \rangle) \rangle - \sigma(X) &= \epsilon + \text{int} \left(\frac{\sigma(\langle X \rangle) - \epsilon}{\delta} + \eta \right) \cdot \delta - \sigma(X) = \\ &= \eta\delta + [\sigma(\langle X \rangle) - \sigma(X)] - [a - \text{int}(a)]\delta. \end{aligned}$$

Hence,

$$\langle \sigma(\langle X \rangle) \rangle - \sigma(X) \in (-(1.5 - \eta)\delta, (0.5 + \eta)\delta) \quad (3.2)$$

We do not know if the bounds of this interval are the best possible for any selection of δ , ϵ and η . However, for special selections of these numbers this is the case. For example: X_3 for $\epsilon = 0$ and $\eta = 1/2$, and X_4 for $\epsilon = 0$ and $\eta = 1 - \omega$.

The following theorem summarizes previous results.

THEOREM. Let X be a random variable. If $E(X)$ exists, then

$$\begin{aligned} -(1 - \eta)\delta &< E(\langle X \rangle) - E(X) \leq \eta\delta, \\ -2(1 - \eta)\delta &< \langle E(\langle X \rangle) \rangle - E(X) \leq 2\eta\delta, \end{aligned}$$

and if $\sigma(X)$ also exists, then

$$\begin{aligned} -\delta/2 &< \sigma(\langle X \rangle) - \sigma(X) < \delta/2, \\ -(1.5 - \eta)\delta &< \langle \sigma(\langle X \rangle) \rangle - \sigma(X) < (0.5 + \eta)\delta. \end{aligned}$$

The first three inequalities are the best possible for any selection of the parameters of the rounding function.

The conclusion is that, in case of rounding random variables, nothing very uncontrolled can happen with the expected values and standard deviations. The effect of rounding can be assessed in terms of the parameters of the rounding function. The most important is the rounding level δ .

In concrete cases, when the distribution of observed random variable is in hand, we expect that the intervals in Theorem are in fact much narrower. For symmetric rounding and in the case of normal distribution with standard deviation equal to the rounding level δ we calculated (numerically, since the exact calculation demands to evaluate an intricate series that cannot be found in standard handbooks like Gradshteyn and Ryzhik (1994)): $|E(\langle X \rangle) - E(X)| < 10^{-9}$ -times standard deviation. So, even in this case of standard deviation unusually small and/or the rounding extremely rough, the influence of rounding is negligible.

References

- [1] Everitt, B.S. (2002): *The Cambridge Dictionary of Statistics*. Cambridge: Cambridge University Press, 2nd edition.
- [2] Gradshteyn, I.S., Ryzhik, I.M. (1994): *Table of Integrals, Series, and Products*. Jeffrey, A. (Ed.), San Diego: Academic Press, Inc. (5th edition).
- [3] Jamnik, R. (1971): *Verjetnostni račun*. Ljubljana: Mladinska knjiga.
- [4] Lisman, J.H.C. (1985): Conditional rounding: Letter to the editor. *Statistica*, **XLV**, 571-573.