Subpixel Edge Location in Binary Images Using Dithering

Xiangdong Liu and Roger W. Ehrich

TR 93-11

Department of Computer Science
Virginia Polytechnic Institute and State University
Blacksburg, Virginia 24061

March 24, 1993

Subpixel Edge Location in Binary Images Using Dithering

by

Xiangdong Liu
Roger W. Ehrich

Technical Report TR 93-11

Department of Computer Science
Virginia Polytechnic Institute and State University
Blacksburg, Virginia 24061

March 24, 1993

Abstract

This paper concerns the problem of obtaining subpixel estimates of the locations of straight lines in digital images for purposes of machine vision. In particular, it presents a dithering method for improving the estimation accuracy on a rectangular sampling lattice. By adding uniformly distributed independent random noise it is shown that estimation bias may be removed and that the estimation variance is inversely proportional to the length of the line segment. The sensitivity to incorrect dither amplitude is calculated, and a novel approach is given for adding the dither by using grey-level image sensor and utilizing the imaging model.

Introduction

Subpixel edge location measurement is extremely important for machine inspection and mensuration applications. It permits highly accurate measurements to be made with inexpensive, low resolution vision sensors. In many cases physical constraints do not permit the use of higher resolution sensors, and one must strive to achieve the best possible accuracies with limited sensor capabilities. A large number of papers have appeared in the last decade concerning a broad range of subpixel edge location techniques for diverse applications. Generally these have been obtained with widely varying assumptions, and it is not always easy to understand what they are. Commonly shared among the various approaches is the idea that the longer the edge being measured, the better the estimates of edge position become. Central limit theorem-like arguments suggest that as n independent edge measurements are made, the variance of position estimators is reduced by the factor $1/n$, with the possible exceptions of rational slopes, most notably vertical, horizontal, or $45°$ diagonal lines where the measurements are completely correlated. Indeed, this behavior is supported by many theoretical and experimental papers.

This paper is motivated by applications in which objects are to be digitized in binary and measured with high accuracy. Binary digitization is desirable for reducing storage requirements and for speeding up computations in applications involving extremely large images. We first show that by adding random position noise to the line or by introducing dither to the sensor, the $1/n$ reduction in variance is preserved, regardless of line orientation. This can be implemented in various ways by actually making ideally straight scene edges fuzzy in the images themselves or equivalently by randomizing the sensor binarization threshold in just the right way. We calculate the worst-case performance and investigate the sensitivity of the estimate to slight deviations from the optimal dither noise statistics.

Review of Previous Work

Edge location algorithms appearing in the literature can be broadly classified into two categories — grey level and binary. In the case of grey level images, it is frequently the case that edge models are assumed, often without considering the characteristics of the imaging/digitizing equipment. In this category one finds two basic approaches — geometric and moment-based. In the geometric approach, the location of an edge is determined by using either an explicit or an implicit edge model. An edge surface may be interpolated explicitly using a curve or surface such as a cubic or integral of Gaussian and then differentiated analytically. Alternately the edge model may be applied implicitly by using a linear or nonlinear mask-type edge operator that is not derived by surface approximation. Then the edge is located by thresholding the derivative surface or by finding the zero crossings of the second derivative of the image [1–3]. In a moment-based approach, moments of the discrete pixel intensity distribution across an edge and those of the underlying continuous edge are assumed to be equal (moment preservation). Then by equating moments one obtains equations whose solution is by definition the desired edge location. Two definitions of moments have been proposed [4,5]. Algorithms in this category usually utilize information from pixels that are close to the 'true' edge, in addition to those pixels that are 'on' the edge. Most papers seem to fall into this category of grey level techniques.

In the second category, the underlying images are binary. Edges are determined by the boundary between adjacent regions with different intensities (black or white). It is the locations (coordinates) of those pixels on the boundary that are used to determine the edge, rather than the grey values of pixels in a neighborhood around the edge. The method presented in this paper belongs to this category.

Hyde and Davis [6] considered fitting lines to straight edges in a grey level image using information about both the (quantized) locations and the intensities of pixels through which the edge passes. They concluded from experiments that the incorporation of grey level information does not improve the accuracy of the fitted line and does not warrant the higher computational cost. Other authors did achieve higher accuracy by taking into account pixels that are adjacent to the edge in a grey level image [7].

Dorst and Smeulders [8] were interested specifically in the geometrical properties of binarized lines of finite length. For example, they characterized the set of continuous finite length line segments which, after digitization, give rise to the same digital line segment with a particular chain code. Later, Dorst and Duin [9] also pointed out that the maximum vertical distance between any two parallel line segments in this set is dependent upon the orientation of the lines, in addition to the number of pixels in the line segments. For certain orientations, notably 0, 45, and 90 degrees, the distance is especially large, up to one pixel spacing for the cases of 0 and 90 degrees. Based upon Dorst's work, Berenstein *et al* [10] proposed using a special line segment in the set as the estimator for all line segments in the set. They computed the corresponding average vertical offset error over all possible digital line segments of a given length under the assumption that all line segments are uniformly distributed in space and in orientation. Their assumption is that uniformly distributed in space implies uniform radial and angular distribution. For a specific line segment, the error can still be as large as half the pixel spacing.

Gordon and Seering [11] studied the estimation error in locating a binarized line by using a least squares approximation based upon the quantized coordinates of points along the line. They made a simplifying assumption that the fractional parts δy of the vertical coordinates of each point on the true continuous line segment are independent. As they point out, this is not always a valid assumption. They made different assumptions from Berenstein on the uniform randomness of line segments in the plane; their vertical shift and orientation estimates are based only upon the independence of δy and an assumption of a constant variance. In their experimental work they considered 18 equiprobable perturbations within a 2.5 degree range around particular orientations and show the comparison with their theoretical results. Their experiments show that the approximation errors are largest for the same line orientations at which Dorst found the largest errors, namely vertical, horizontal, and diagonal lines.

Cox, *et al*, approach a slightly different problem of finding the congruence or small perturbation that brings an image and a line model into agreement [12]. They make uniformness and independence assumptions similar to those of Gordon and Seering when considering the quantization error. They also noted that models with substantial orientations of 0, 45, and 90 degrees were especially subject to estimation error. They suggest that imaging equipment be positioned in such a way that orientations of edges are not at those special angles.

The above approaches attempt to obtain from the digitized edges an estimate of the original continuous edge with subpixel accuracy. Obviously, the digitization process itself destroys much of the subpixel information due both to position and to intensity quantization. Our approach is to perturb an edge randomly before quantization so that better information can be obtained about the fractional parts of the edge position measurements. This facilitates the statistical recovery of the edge position.

Locating Edges Using Unit Dither

Consider a thick ideal straight edge whose orientation is known. We represent the edge by $y = \alpha x + e$, where α is a known constant. The edge is digitized at horizontal coordinates $x = 0, 1, \ldots, n-1$. Our objective is to estimate the value of e so that the location of the boundary can be determined. First consider the case where $-45^\circ \leq \arctan(\alpha) \leq 45^\circ$. Let η_i, $i = 0, 1, \ldots, n-1$, be n i.i.d. random variables having uniform distribution in $[0, 1)$. We add η_i to the y value of the curve at $x = i$ before digitizing it. Then we have

$$y_i = \lfloor \alpha i + e + \eta_i \rfloor \qquad i = 0, 1, \ldots, n-1.$$

The η_i's represent a small dither added to the vertical coordinate of the edge before digitization. Note that y_i's are random variables and that the same continuous edge may give rise to different sets of y_i's. Note that the set of pixels $\{(i, y_i)\}$ in general no longer constitutes the noiseless digitization of a straight line. Next we compute a least squared error estimate of the value of e . That is, we choose an $\hat{e}$ that estimates e, such that

$$Err(\hat{e}) = \sum_{i=0}^{n-1} [\alpha i + \hat{e} - y_i]^2$$

is minimized. By taking the partial derivative of the above function with respect to $\hat{e}$ and equating the derivative to zero, we can solve for $\hat{e}$ (linear regression):

$$\hat{e} = \frac{1}{n} \sum_{i=0}^{n-1} y_i - \frac{(n-1)}{2} \alpha.$$

We can prove that the mean and variance of $\hat{e}$ satisfy

$$E(\hat{e}) = e \qquad and$$
$$Var(\hat{e}) < \frac{1}{4n}.$$

Note that the mean and variance are taken over all possible "digitizations" of the same continuous edge, and the result is independent of the orientation, α. It should be emphasized that this result provides us with an unbiased estimate of vertical shift of a straight edge for a known orientation and a carefully controlled unit dither.

Next this result can be generalized for the case where the orientation is unknown and must also be estimated. Specifically, let the boundary curve be $y = \alpha x + e$ where both α and e are unknown. We minimize the following squared error function that has two unknowns:

$$Err(\hat{\alpha}, \hat{e}) = \sum_{i=0}^{n-1} (\hat{\alpha} i + \hat{e} - y_i)^2.$$

Setting the two partial derivatives of the above function to 0, we can solve for $\hat{\alpha}$ and $\hat{e}$ (again, linear regression):

$$\hat{\alpha} = \frac{12}{n(n-1)(n+1)} \sum_{i=0}^{n-1} (i - \frac{n-1}{2}) y_i$$

$$\hat{e} = \frac{1}{n} \sum_{i=0}^{n-1} y_i - \frac{(n-1)}{2} \hat{\alpha}$$

Again, we can prove that

$$E(\hat{\alpha}) = \alpha \qquad and$$

$$E(\hat{e}) = e.$$

Also, we can compute the variances of $\hat{\alpha}$ and $\hat{e}$ in this case and find upper bounds for them:

$$Var(\hat{\alpha}) < \frac{3}{n(n-1)(n+1)},$$

$$Var(\hat{e}) < \frac{1}{n}.$$

We see that the variance of $\hat{\alpha}$ is fairly small and, therefore, $\hat{\alpha}$ is a fairly accurate estimate of α. The upper bound we derived for the variance of $\hat{e}$ is 4 times larger than that of the case in which α is known. This is quite understandable because it is $\hat{\alpha}$, the estimator of α, instead of the value of α itself, that is used in the calculation of $\hat{e}$.

For the case where $45° \leq \arctan(\alpha) \leq 135°$, more than one pixel on the digitized line may have the same x coordinate. However, the y coordinate of each pixel on the digitized line is unique. If we add dither to the x coordinates of those points on the continuous line that have integer y coordinates and quantize the line in x coordinate, we can get a result symmetric to the above. A simpler way to deal with such edges is to rotate the imaging system by 90°.

From the above results, we see that the accuracy of the estimated line, i.e., the accuracy of the values of $\hat{e}$ and $\hat{\alpha}$, depends only upon the number of points that are involved in the quantization of the original continuous line, that is, the length of the latter. The bound is independent of the orientation of the line, and is consistent for all lines. We emphasize here again that our mean and variance are taken over all possible digitizations of the same continuous line segment, not on all the unique digitizations of all possible continuous line segments.

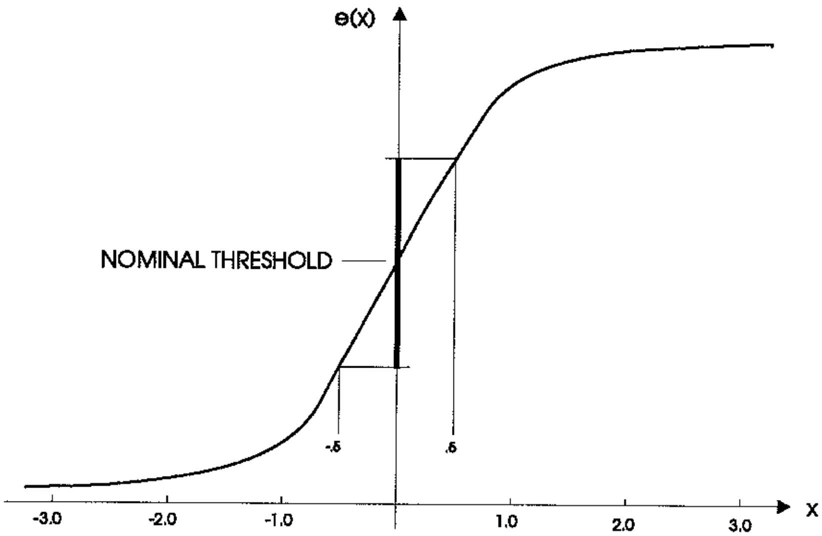

Figure 1. $e(x)$ vs. image plane distance (pixels).

Achieving the Dithering Effect

There are many factors that contribute to the response of a lens to an ideal edge. In the case where a simple lens is linear and spatially invariant, the image is the convolution of its point spread function $h_1(x, y)$ with the source image $i_s(x, y)$. If the lens is diffraction-limited and the light monochromatic, then the point spread function is angularly isotropic, i.e.,

$$h_2(r) = 4K[J_1(\alpha r)]^2/(\alpha r)^2,$$

where α depends upon the wavelength and the position of the image plane relative to the aperture and J_1 is the first order Bessel function. Using h_2 one may determine that the response of such a lens to a long thin line source oriented vertically is

$$h_3(x) = 4KS_1(2\alpha x)/(\alpha x)^2,$$

where S_1 is the Struve function of order 1 [13]. $h_3(x)$ looks much like a 1-dimensional Gaussian with slightly oscillatory tails. Then the familiar edge response shown in Figure 1 is given by

$$e(x) = \int_{-\infty}^{x} h_3(\tau)\ d\tau.$$

Suppose a nominal threshold is established at the midpoint of the function $e(x)$. Then by randomly disturbing the threshold so that $e^{-1}(x)$ is uniformly distributed in the interval $[-.5, .5)$ the desired dither is achieved without the necessity of actually dithering the image itself. Of course, actual optical systems will differ considerably from the simple case described above, and $e(x)$ will have to be determined empirically.

There are situations in which high resolution, gray level images are already available. We may want to convert such images to low resolution, binary formats to save storage and/or to speed up processing. We could incorporate a dithering procedure in the conversion software so that we can still obtain relatively accurate measurement of edges from the low resolution images. Writing such a dithering procedure in software is straightforward for edges oriented between $-45°$ to $45°$.

Practical Considerations

From a practical viewpoint, it may not be possible to have the variable threshold take on exactly all continuous values in the unit interval shown in Figure 1. Worse yet, the actual interval of the synthetic noise may not be exactly 1. For this reason it is useful to explore the sensitivity of $\hat{e}$ and $Var(\hat{e})$ to a deviation from the ideal distribution width. Our sensitivity analysis reveals the following mean and the variance of $\hat{e}$ when estimating the vertical shift of a straight edge with known orientation in which the η_i's now assume a uniform distribution in the interval $[-\epsilon, 1+\epsilon)$, with $|\epsilon| < \frac{1}{2}$:

$$E(\hat{e}) = e + r \qquad \textit{with} \qquad |r| < \frac{|\epsilon|}{1+2\epsilon},$$

$$Var(\hat{e}) < \frac{1}{4n} + \frac{3|\epsilon|}{2n}.$$

Notice that $Var(\hat{e})$ increases linearly with $\frac{\epsilon}{n}$, and that the estimate $\hat{e}$ becomes biased. We would also like to know how close to e an instance of $\hat{e}$ is expected to be. $E[(\hat{e}-e)^2]$ is such an indicator. We can show that

$$E[(\hat{e}-e)^2] < \frac{1}{4n} + \frac{3|\epsilon|}{2n} + \frac{\epsilon^2}{(1+2\epsilon)^2}.$$

We see that the estimator $\hat{e}$ is quite stable. The sensitivity analysis for the more complex case of estimating both α and e for a straight edge has not been done because the computation involved is rather complicated.

The dither makes y_i and y_j $(i \neq j)$ statistically independent, although we did not use this fact in our derivation (see Appendix). The effect of the dither that we have made use of is that when it has the exact unit interval distribution, it makes the y_i's have the following two properties:

$$E(y_i) = \alpha i + e$$

and

$$E(y_i y_j) = \begin{cases} (\alpha i + e)(\alpha j + e) & i \neq j \\ (\alpha i + e)^2 + \delta_i \sigma_i & i = j, \end{cases}$$

where

$$\delta_i = \lceil \alpha i + e \rceil - (\alpha i + e)$$

and

$$\sigma_i = (\alpha i + e) - \lfloor \alpha i + e \rfloor$$

are the subpixel fractional distances between the intersection of the line with $x = i$ and the next higher and lower sample points, respectively. Intuitively, the closer $\alpha i + e$ is to an integer (say, $\lfloor \alpha i + e \rfloor$), the more likely y_i is to take that integer as its value. Thus, the information about the fractional part of $\alpha i + e$ is retained to some degree. For example, consider the digitization of a continuous horizontal line $y = e$ at $x = 0, 1, \ldots, n-1$, where $0 < e < 1$. From the usual quantization process, we get $y_i = 0$ for $i = 0, 1, \ldots, n-1$. The best estimate we can propose is $y = 0.5$ and the vertical offset error is $|e - 0.5|$. The error is deterministic for a particular value of e. If dither is added prior to quantization, the least squares estimate $\hat{e}$ is non-deterministic for a fixed e. The usual situation is that some of the y_i's will have value 1 while others will have value 0. Then $\hat{e}$ has a value between 0 and 1 and is close to e as the mean and the variance of $\hat{e}$ indicate. An extreme situation is when all n dither values are so small that $y_i = \lfloor e + \eta_i \rfloor = 0$ for all i, in which case $\hat{e}$ will be 0. Fortunately the probability of this occurring is $(1-e)^n$ which is very unlikely when n is large.

From the statistical point of view, since there are n independent data points that are used to estimate the parameter of the edge, the variance of the result is on the order of v/n, where v is the variance of a single sample point about its true value. Our method provides one way to make the samples independent while keeping the estimate unbiased and the variance small. It turns out that as long as the dither is uniform over the interval $[-i, i+1)$, $i \geq 0$, the estimator $\hat{e}$ is unbiased. For example, we could use dither with uniform distribution in say, $[-1, 2)$. This would still preserve the independence among the y_i's and cause the mean of each y_i to be $\alpha i + e$, but it would make the variance of y_i unnecessarily large. Also notice that the central limit theorem tells us that the estimator $\hat{e}$ tends to have a normal distribution when α is known.

The method for estimating the vertical shift of a straight edge and the corresponding sensitivity analysis can be extended in two ways. First, it can be seen very easily that the abscissas of the digitized points on the edge do not have to be $0, 1, \ldots, n-1$; they can be $x_0, x_1, \ldots, x_{n-1}$, as long as the x_i's are distinct. Second, the edge does not have to be straight. It can have any known shape, in which case, the edge curve can be expressed in the form $y = F(x) + e$ where F is known. The derivation of this extension parallels that for the straight edge.

The main mathematical method used concerns the identification and use of the fractions δ_i and σ_i. We have used few approximations in our derivations despite the discontinuous floor and ceiling functions, and have only enlarged the error estimates to get bounds with simple forms at the ends of the derivations. Thus the results are good even when n, the number of pixels in the digitized edge, is not very large, which is said to be common [10].

In the sensitivity analysis, the exact values of the mean and the variance of $\hat{e}$ depend not only on n, but also on α and e. However, we have estimated upper bounds for the mean and the variance that are dependent only upon n, the length of the edge.

References

[1] V. S. Nalwa and T. O. Binford, "On Detecting Edges," *IEEE Trans. PAMI* (8), No. 6, pp. 699–714, 1986.

[2] E. De Micheli, B. Caprile, P. Ottonello, and V. Torre, "Localization and Noise in Edge Detection", *IEEE Trans. PAMI* (11), No. 10, pp. 1106–1117, 1989.

[3] A. Huertas and G. Medioni, "Edge Detection with Subpixel Precision," *Proceedings of the 3rd IEEE Workshop on Computer Vision: Representation and Control,* Bellaire, MI, 63–74, 1985.

[4] A. J. Tabatabai and O. R. Mitchell, "Edge Location to Subpixel Values in Digital Imagery," *IEEE Trans. PAMI* (6), No. 2, pp. 188–201, 1984.

[5] E. P. Lyvers, O. R. Mitchell, M. L. Akey, and A. P. Reeves, "Subpixel Measurements Using a Moment-Based Edge Operator," *IEEE Trans. PAMI* (11), No. 12, pp. 1293–1309, 1989.

[6] P. D. Hyde and L. S. Davis, "Subpixel Edge Estimation," *Pattern Recognition* (16), No. 4, pp. 413–420, 1983.

[7] R. A. Young, "Locating Industrial Parts with Subpixel Accuracies," *Proceedings of the SPIE* (728), pp. 2–9, 1986.

[8] L. Dorst and A. M. Smeulders, "Discrete Representation of Straight Lines," *IEEE Trans. PAMI* (6), No. 4, 450–463, 1984.

[9] L. Dorst and R. P. W. Duin, "Spirograph Theory: A Framework for Calculations on Digitized Straight Lines," *IEEE Trans. PAMI* (6), No. 5, 632–639, 1984.

[10] C. A. Berenstein, L. N. Kanal, D. Lavine, and E. C. Olson, "A Geometric Approach to Subpixel Registration Accuracy," *CVGIP* (40), 334–360, 1987.

[11] S. J. Gordon and W. P. Seering, "Accuracy Issues in Measuring Quantized Images of Straight-Line Features," *IEEE Int. Conf. on Robotics and Automation,* v. 2, 931–936, 1986.

[12] I. J. Cox, J. B. Kruskal, and D. A. Wallach, "Predicting and Estimating the Accuracy of a Subpixel Registration Algorithm," *IEEE Trans. PAMI* (12), No. 8, pp. 721–734, 1990.

[13] Charles S. Williams and Orville A. Becklund, *Introduction to the Optical Transfer Function,* New York: Wiley, 1989.

Appendix

In this Appendix we provide proofs of the three assertions made in the paper. Assertion 1) is actually a special case of assertion 3), but we prove it first because it is simple and some intermediate results can be used in the proof of other assertions. First we list the assertions again. Given

$$y_i = \lfloor \alpha i + e + \eta_i \rfloor, \qquad i = 0, 1, \ldots, n-1,$$

where α and e are constants, and η_i's are i.i.d. random variables with uniform distribution over interval I, then

1) if α is known and $I = [0, 1)$, then

$$\hat{e} = \frac{1}{n}\sum_{i=0}^{n-1} y_i - \frac{(n-1)}{2}\alpha, \tag{1}$$

the estimator of e, has the following properties:

$$E(\hat{e}) = e,$$

$$Var(\hat{e}) < \frac{1}{4n};$$

2) if both α and e are unknown and $I = [0, 1)$, then

$$\hat{\alpha} = \frac{12}{n(n-1)(n+1)}\sum_{i=0}^{n-1}(i - \frac{n-1}{2})y_i, \tag{2}$$

and

$$\hat{e} = \frac{1}{n}\sum_{i=0}^{n-1} y_i - \frac{(n-1)}{2}\hat{\alpha}, \tag{3}$$

the estimators of α and e respectively, have the following properties:

$$E(\hat{\alpha}) = \alpha,$$

$$E(\hat{e}) = e,$$

$$Var(\hat{\alpha}) < \frac{3}{n(n-1)(n+1)},$$

$$Var(\hat{e}) < \frac{1}{n};$$

3) if α is known and $I = [-\epsilon, 1+\epsilon)$ where $|\epsilon| < \frac{1}{2}$, then the estimator $\hat{e}$ of e in (1), satisfies

$$E(\hat{e}) = e + r$$

where $|r| < \frac{|\epsilon|}{1+2\epsilon}$, and

$$Var(\hat{e}) < \frac{1}{4n} + \frac{3|\epsilon|}{2n},$$

while

$$E[(\hat{e}-e)^2] < \frac{1}{4n} + \frac{3|\epsilon|}{2n} + \frac{\epsilon^2}{(1+2\epsilon)^2}.$$

Notation: in the following derivations, the index of a summation runs from 0 to $n-1$ if no limits are specified.

Proof of assertion 1)

It is easy to see from (1) that

$$E(\hat{e}) = \frac{1}{n}\sum E(y_i) - \frac{(n-1)}{2}\alpha. \tag{4}$$

To compute $E(y_i)$, notice that $\alpha i+e+\eta_i$ has equal probability of being any real number between $\alpha i+e$ and $\alpha i+e+1$. If $\alpha i+e$ is not an integer, then

$$y_i = \begin{cases} \lfloor \alpha i+e \rfloor & \text{if } \alpha i+e+\eta_i < \lceil \alpha i+e \rceil \\ \lceil \alpha i+e \rceil & \text{if } \alpha i+e+\eta_i \geq \lceil \alpha i+e \rceil. \end{cases}$$

More importantly, the probability distribution $Pr(y_i)$ of y_i is

$$Pr(y_i) = \begin{cases} \lceil \alpha i+e \rceil - (\alpha i+e) & y_i = \lfloor \alpha i+e \rfloor \\ (\alpha i+e) - \lfloor \alpha i+e \rfloor & y_i = \lceil \alpha i+e \rceil. \end{cases} \tag{5}$$

We see that

$$\begin{aligned} E(y_i) &= \lfloor \alpha i+e \rfloor[\lceil \alpha i+e \rceil - (\alpha i+e)] + \lceil \alpha i+e \rceil[(\alpha i+e) - \lfloor \alpha i+e \rfloor] \\ &= [\lceil \alpha i+e \rceil - \lfloor \alpha i+e \rfloor](\alpha i+e) \\ &= (\alpha i+e). \end{aligned}$$

If $\alpha i+e$ is an integer, then

$$y_i = \lfloor \alpha i+e+\eta_i \rfloor \equiv \lfloor \alpha i+e \rfloor = (\alpha i+e).$$

Again, we have

$$E(y_i) = (\alpha i+e). \tag{6}$$

Thus we have

$$\sum E(y_i) = \sum(\alpha i+e) = \frac{n(n-1)}{2}\alpha + ne \tag{7}$$

and the substitution of this formula into (4) gives $E(\hat{e}) = e$.

To compute the variance of $\hat{e}$, we notice that

$$Var(\hat{e}) = E(\hat{e}^2) - [E(\hat{e})]^2 = E(\hat{e}^2) - e^2$$

and

$$\begin{aligned} E(\hat{e}^2) &= E\{[\frac{1}{n}\sum y_i - \frac{(n-1)}{2}\alpha]^2\} \\ &= \frac{1}{n^2}E[(\sum y_i)^2] - \frac{(n-1)}{n}E(\sum y_i) + \frac{(n-1)^2\alpha^2}{4} \\ &= \frac{1}{n^2}E[(\sum y_i)^2] - \frac{(n-1)^2\alpha^2}{4} - (n-1)\alpha e. \end{aligned} \tag{8}$$

Since

$$E[(\sum_i y_i)^2] = E[(\sum_i y_i)(\sum_j y_j)] = \sum_i \sum_j E(y_i y_j), \tag{9}$$

we need to compute $E(y_i y_j)$, $i, j = 0, 1, \ldots, n-1$. We notice that y_i and y_j are, respectively, functions of η_i and η_j which are independent if $i \neq j$. From probability theory we have

$$E(y_i y_j) = E(y_i)E(y_j) = (\alpha i + e)(\alpha j + e) \qquad if \qquad i \neq j.$$

For the case of $i = j$, first assume that $\alpha i + e$ is not an integer. Now y_i^2 can take only one of two values, $\lfloor \alpha i + e \rfloor^2$ and $\lceil \alpha i + e \rceil^2$, with the same probabilities that y_i takes $\lfloor \alpha i + e \rfloor$ or $\lceil \alpha i + e \rceil$. Let

$$\begin{aligned} \delta_i &= \lceil \alpha i + e \rceil - (\alpha i + e) \\ \sigma_i &= (\alpha i + e) - \lfloor \alpha i + e \rfloor. \end{aligned} \tag{10}$$

We have

$$\begin{aligned} E(y_i^2) &= \lfloor \alpha i + e \rfloor^2 \delta_i + \lceil \alpha i + e \rceil^2 \sigma_i \\ &= [(\alpha i + e) - \sigma_i]^2 \delta_i + [(\alpha i + e) + \delta_i]^2 \sigma_i \\ &= (\alpha i + e)^2 + \delta_i \sigma_i. \end{aligned}$$

If $\alpha i + e$ is an integer, then $E(y_i^2) = (\alpha i + e)^2$ and the above formula still holds (with $\sigma_i = 0$). In summary, we have

$$E(y_i y_j) = \begin{cases} (\alpha i + e)(\alpha j + e) & i \neq j \\ (\alpha i + e)^2 + \delta_i \sigma_i & i = j. \end{cases} \tag{11}$$

Now the summation in (9) can be carried out:

$$\begin{aligned} E[(\sum_i y_i)^2] &= \sum_i \sum_j (\alpha i + e)(\alpha j + e) + \sum_i \delta_i \sigma_i \\ &= \sum_i (\alpha i + e) \sum_j (\alpha j + e) + \sum_i \delta_i \sigma_i \\ &= \frac{n^2(n-1)^2}{4} \alpha^2 + n^2(n-1)\alpha e + n^2 e^2 + \sum \delta_i \sigma_i. \end{aligned}$$

Substituting this last formula in (8) yields

$$E(\hat{e}^2) = e^2 + \frac{1}{n^2} \sum \delta_i \sigma_i,$$

and hence

$$Var(\hat{e}) = \frac{1}{n^2} \sum \delta_i \sigma_i. \tag{12}$$

Notice that $\delta_i, \sigma_i \geq 0$ and $\delta_i + \sigma_i = 0$ or 1. We have $\delta_i \sigma_i \leq \frac{1}{4}$. So

$$Var(\hat{e}) \leq \frac{1}{4n}.$$

Proof of assertion 2)

Again, the means are relatively easy to compute. From (2) one has

$$E(\hat{\alpha}) = \frac{12}{n(n-1)(n+1)} \sum [i - \frac{(n-1)}{2}] E(y_i).$$

Now

$$\begin{aligned}
\sum [i - \frac{(n-1)}{2}] E(y_i) &= \sum [i - \frac{(n-1)}{2}](\alpha i + e) \\
&= \sum \{\alpha i^2 + [e - \frac{(n-1)}{2}\alpha] i - \frac{(n-1)}{2} e\} \\
&= \frac{n(n-1)(2n-1)}{6}\alpha + [e - \frac{(n-1)}{2}\alpha]\frac{n(n-1)}{2} - \frac{n(n-1)}{2} e \\
&= \frac{n(n-1)(2n-1)}{6}\alpha + \frac{n(n-1)^2}{4}\alpha \\
&= \frac{n(n-1)(n+1)}{12}\alpha.
\end{aligned}$$

This leads to $E(\hat{\alpha}) = \alpha$. From (3) one has

$$E(\hat{e}) = \frac{1}{n}\sum E(y_i) - \frac{(n-1)}{2} E(\hat{\alpha}).$$

Combining this with (7) gives $E(\hat{e}) = e$.

We now compute $Var(\hat{\alpha})$ and $Var(\hat{e})$ in the same way we computed $Var(\hat{e})$ in the previous section. From (2) one has

$$\hat{\alpha} = \frac{1}{p(n)}[\sum i y_i - \frac{(n-1)}{2}\sum y_i],$$

where

$$p(n) = \frac{n(n-1)(n+1)}{12}. \tag{13}$$

Thus one has

$$E(\hat{\alpha}^2) = \frac{1}{p^2(n)}\{E[(\sum i y_i)^2] - (n-1)E[(\sum i y_i)(\sum y_i)] + \frac{(n-1)^2}{4} E[(\sum y_i)^2]\}. \tag{14}$$

Since we have computed the last term in the braces in the previous section, we only need to compute the first two terms. Using the same notation as in the previous section, one has

$$\begin{aligned}
E[(\sum i y_i)^2] &= E[(\sum_i i y_i)(\sum_j j y_j)] = \sum_i \sum_j ij E(y_i y_j) \\
&= \sum_i \sum_j ij(\alpha i + e)(\alpha j + e) + \sum_i i^2 \delta_i \sigma_i \\
&= [\sum_i i(\alpha i + e)][\sum_j j(\alpha j + e)] + \sum_i i^2 \delta_i \sigma_i \\
&= [\frac{n(n-1)(2n-1)}{6}\alpha + \frac{n(n-1)}{2} e]^2 + \sum i^2 \delta_i \sigma_i \\
&= \frac{n^2(n-1)^2(2n-1)^2}{36}\alpha^2 + \frac{n^2(n-1)^2(2n-1)^2}{6}\alpha e + \frac{n^2(n-1)^2}{4} e + \sum i^2 \delta_i \sigma_i,
\end{aligned}$$

and

$$
\begin{aligned}
E[(\sum iy_i)(\sum y_i)] &= E[(\sum_i iy_i)(\sum_j y_j)] = \sum_i \sum_j iE(y_i y_j) \\
&= \sum_i \sum_j i(\alpha i + e)(\alpha j + e) + \sum_i i\delta_i \sigma_i \\
&= [\sum_i i(\alpha i + e)][\sum_j (\alpha j + e)] + \sum_i i\delta_i \sigma_i \\
&= [\frac{n(n-1)}{2}\alpha + ne][\frac{n(n-1)(2n-1)}{6}\alpha + \frac{n(n-1)}{2}e] + \sum i\delta_i \sigma_i \\
&= \frac{n^2(n-1)^2(2n-1)}{12}\alpha^2 + [\frac{n^2(n-1)^2}{4} + \frac{n^2(n-1)(2n-1)}{6}]\alpha e + \frac{n^2(n-1)}{2}e + \sum i\delta_i \sigma_i.
\end{aligned}
$$

Putting everything together in (14) and collecting terms, we obtain

$$
E(\hat{\alpha}^2) = \alpha^2 + \frac{1}{p^2(n)}[\frac{(n-1)^2}{4}\sum \delta_i \sigma_i - (n-1)\sum i\delta_i \sigma_i + \sum i^2 \delta_i \sigma_i]. \qquad (15)
$$

Now we obtain $Var(\hat{\alpha})$ and can estimate an upper bound for it:

$$
\begin{aligned}
Var(\hat{\alpha}) &= E(\hat{\alpha}^2) - \alpha^2 \\
&= \frac{144}{n^2(n-1)^2(n+1)^2}[\frac{(n-1)^2}{4}\sum \delta_i \sigma_i - (n-1)\sum i\delta_i \sigma_i + \sum i^2 \delta_i \sigma_i] \\
&= \frac{144}{n^2(n-1)^2(n+1)^2}\sum [\frac{(n-1)^2}{4} - (n-1)i + i^2]\delta_i \sigma_i \\
&\le \frac{36}{n^2(n-1)^2(n+1)^2}\sum [\frac{(n-1)^2}{4} - (n-1)i + i^2] \\
&= \frac{36}{n^2(n-1)^2(n+1)^2}[\frac{(n-1)^2}{4}n - (n-1)\frac{(n-1)n}{2} + \frac{n(n-1)(2n-1)}{6}] \\
&= \frac{3}{n(n-1)(n+1)}.
\end{aligned}
$$

Last, we compute $Var(\hat{e})$. From (3) one has

$$
\begin{aligned}
E(\hat{e}^2) &= E\{[\frac{1}{n}\sum y_i - \frac{(n-1)}{2}\hat{\alpha}]^2\} \\
&= \frac{1}{n^2}E[(\sum y_i)^2] - \frac{(n-1)}{n}E(\hat{\alpha}\sum y_i) + \frac{(n-1)^2}{4}E(\hat{\alpha}^2).
\end{aligned}
$$

We have essentially computed the first term and the third term before. Let's list them here:

$$
\begin{aligned}
\frac{1}{n^2}E[(\sum y_i)^2] &= \frac{(n-1)^2}{4}\alpha^2 + (n-1)\alpha e + e^2 + \frac{1}{n^2}\sum \delta_i \sigma_i, \\
\frac{(n-1)^2}{4}E(\hat{\alpha}^2) &= \frac{(n-1)^2}{4}\alpha^2 + \frac{9(n-1)^2}{n^2(n+1)^2}\sum (1 - \frac{2i}{n-1})^2 \delta_i \sigma_i.
\end{aligned}
$$

The summation in the second term is

$$\begin{aligned}
E(\hat{\alpha}\sum y_i) &= E[\frac{1}{p(n)}\sum_j (j-\frac{n-1}{2})y_j \sum_i y_i] \\
&= \frac{1}{p(n)}\sum_j\sum_i (j-\frac{n-1}{2})E(y_j y_i) \\
&= \frac{1}{p(n)}[\sum_j\sum_i (j-\frac{n-1}{2})(\alpha i+e)(\alpha j+e) + \sum_i (i-\frac{n-1}{2})\delta_i\sigma_i] \\
&= \frac{1}{p(n)}[\sum_i (\alpha i+e)\sum_j (j-\frac{n-1}{2})(\alpha j+e) + \sum_i (i-\frac{n-1}{2})\delta_i\sigma_i] \\
&= \frac{1}{p(n)}\{[\frac{n(n-1)}{2}\alpha + ne]\frac{n(n-1)(n+1)}{12}\alpha + \sum_i (i-\frac{n-1}{2})\delta_i\sigma_i\} \\
&= \frac{n(n-1)}{2}\alpha^2 + n\alpha e + \frac{1}{p(n)}\sum_i (i-\frac{n-1}{2})\delta_i\sigma_i.
\end{aligned}$$

Thus the second term is

$$\frac{(n-1)}{n}E(\hat{\alpha}\sum y_i) = \frac{(n-1)^2}{2}\alpha^2 + (n-1)\alpha e + \frac{12}{n^2(n+1)}\sum_i (i-\frac{n-1}{2})\delta_i\sigma_i.$$

Combining the three terms, we obtain

$$E(\hat{e}^2) = e^2 + \frac{1}{n^2}\sum \delta_i\sigma_i - \frac{12}{n^2(n+1)}\sum (i-\frac{n-1}{2})\delta_i\sigma_i + \frac{9(n-1)^2}{n^2(n+1)^2}\sum (1-\frac{2i}{n-1})^2\delta_i\sigma_i.$$

Now we obtain $Var(\hat{e})$ and can estimate an upper bound for it:

$$\begin{aligned}
Var(\hat{e}) &= E(\hat{e}^2) - e^2 \\
&= \frac{1}{n^2}\sum [1 - \frac{12}{n+1}(i-\frac{n-1}{2}) + \frac{9(n-1)^2}{(n+1)^2}(1-\frac{2i}{n-1})^2]\delta_i\sigma_i \\
&= \frac{4}{n^2(n+1)^2}\sum (2n-1-3i)^2\delta_i\sigma_i \\
&\leq \frac{1}{n^2(n+1)^2}\sum (2n-1-3i)^2 \\
&= \frac{1}{n^2(n+1)^2}\sum [(2n-1)^2 - 6(2n-1)i + 9i^2] \\
&= \frac{2n-1}{2n(n+1)} \\
&< \frac{1}{(n+1)} < \frac{1}{n}.
\end{aligned}$$

Proof of assertion 3)

When the distribution interval of the dither is $[-\epsilon, 1+\epsilon)$, instead of exactly $[0,1)$, the y_i's may take one of three values, rather than taking at most one of two values as in the previous two sections. The number of values y_i takes depends on how close $\alpha i + e$, from which y_i is obtained, is to an integer. There are eight cases:

(I) $\epsilon > 0$, $\alpha i + e$ is not an integer

Case (a): $\epsilon < \delta_i$ and $\epsilon < \sigma_i$

This seems to be one of the most common cases. The probability distribution of y_i is:

$$Pr(y_i) = \begin{cases} \frac{\epsilon+\delta_i}{1+2\epsilon} & y_i = \lfloor \alpha i + e \rfloor \\ \frac{\epsilon+\sigma_i}{1+2\epsilon} & y_i = \lceil \alpha i + e \rceil. \end{cases}$$

Case (b): $\delta_i < \epsilon < \sigma_i$

Now $\alpha i + e + \eta_i$ may be greater than $\lceil \alpha i + e \rceil + 1$, and we have the following probability distribution of y_i:

$$Pr(y_i) = \begin{cases} \frac{\epsilon+\delta_i}{1+2\epsilon} & y_i = \lfloor \alpha i + e \rfloor \\ \frac{1}{1+2\epsilon} & y_i = \lceil \alpha i + e \rceil \\ \frac{\epsilon-\delta_i}{1+2\epsilon} & y_i = \lceil \alpha i + e \rceil + 1. \end{cases}$$

Case (c): $\sigma_i < \epsilon < \delta_i$

Now $\alpha i + e + \eta_i$ may be less than $\lfloor \alpha i + e \rfloor$, and we have the following probability distribution of y_i:

$$Pr(y_i) = \begin{cases} \frac{\epsilon-\sigma_i}{1+2\epsilon} & y_i = \lfloor \alpha i + e \rfloor - 1 \\ \frac{1}{1+2\epsilon} & y_i = \lfloor \alpha i + e \rfloor \\ \frac{\epsilon+\sigma_i}{1+2\epsilon} & y_i = \lceil \alpha i + e \rceil. \end{cases}$$

(II) $\epsilon > 0$, $\alpha i + e$ is an integer

Case (d): $0 = \delta_i = \sigma_i < \epsilon$

$$Pr(y_i) = \begin{cases} \frac{\epsilon}{1+2\epsilon} & y_i = (\alpha i + e) - 1 \\ \frac{1}{1+2\epsilon} & y_i = (\alpha i + e) \\ \frac{\epsilon}{1+2\epsilon} & y_i = (\alpha i + e) + 1. \end{cases}$$

(III) $\epsilon < 0$, $\alpha i + e$ is not an integer

Case (e): $|\epsilon| < \delta_i$ and $|\epsilon| < \sigma_i$

This seems to be the other most common case. The probability distribution of y_i is the same as in Case (a).

$$Pr(y_i) = \begin{cases} \frac{\epsilon+\delta_i}{1+2\epsilon} & y_i = \lfloor \alpha i + e \rfloor \\ \frac{\epsilon+\sigma_i}{1+2\epsilon} & y_i = \lceil \alpha i + e \rceil. \end{cases}$$

Case (f): $\delta_i < |\epsilon| < \sigma_i$

$$y_i \equiv \lceil \alpha i + e \rceil = (\alpha i + e) + \delta_i.$$

Case (g): $\sigma_i < |\epsilon| < \delta_i$

$$y_i \equiv \lfloor \alpha i + e \rfloor = (\alpha i + e) - \sigma_i.$$

(IV) $\epsilon < 0$, $\alpha i + e$ is an integer

Case (h): $0 = \delta_i = \sigma_i < |\epsilon|$

$$y_i \equiv \lfloor \alpha i + e \rfloor = \lceil \alpha i + e \rceil = (\alpha i + e).$$

Note: (1) Cases (a)–(h) occur depending upon the position where the line $y = \alpha x + e$ intercepts $x = i$, which is determined by the values of α and e. Thus, whether cases (a)–(h) occur is original-line-dependent. (2) We need $|\epsilon| < \frac{1}{2}$ so that cases (b) and (c) or cases (f) and (g) do not occur simultaneously.

Now we compute $E(y_i)$ as they are needed in the calculation of $E(\hat{e})$.

Cases (a) and (e):

$$\begin{aligned} E(y_i) &= \frac{\lfloor \alpha i + e \rfloor (\epsilon + \delta_i) + \lceil \alpha i + e \rceil (\epsilon + \sigma_i)}{(1 + 2\epsilon)} \\ &= \frac{\lfloor \alpha i + e \rfloor \delta_i + \lceil \alpha i + e \rceil \sigma_i + (\lfloor \alpha i + e \rfloor + \lceil \alpha i + e \rceil)\epsilon}{(1 + 2\epsilon)} \\ &= \frac{(\alpha i + e) + [(\alpha i + e) - \sigma_i + (\alpha i + e) + \delta_i]\epsilon}{(1 + 2\epsilon)} \\ &= (\alpha i + e) + \frac{(\delta_i - \sigma_i)\epsilon}{(1 + 2\epsilon)}. \end{aligned}$$

The magnitude of the numerator of the second term above satisfies

$$|(\delta_i - \sigma_i)\epsilon| < |\epsilon|.$$

Case (b):

$$\begin{aligned} E(y_i) &= \frac{\lfloor \alpha i + e \rfloor (\epsilon + \delta_i) + \lceil \alpha i + e \rceil + (\lceil \alpha i + e \rceil + 1)(\epsilon - \delta_i)}{(1 + 2\epsilon)} \\ &= \frac{\lfloor \alpha i + e \rfloor \epsilon + \lfloor \alpha i + e \rfloor \delta_i + \lceil \alpha i + e \rceil \sigma_i + \lceil \alpha i + e \rceil \delta_i + \lceil \alpha i + e \rceil \epsilon - \lceil \alpha i + e \rceil \delta_i + (\epsilon - \delta_i)}{(1 + 2\epsilon)} \\ &= \frac{(\alpha i + e) + [(\alpha i + e) - \sigma_i]\epsilon + [(\alpha i + e) + \delta_i]\epsilon + (\epsilon - \delta_i)}{(1 + 2\epsilon)} \\ &= (\alpha i + e) + \frac{(\delta_i - \sigma_i)\epsilon + (\epsilon - \delta_i)}{(1 + 2\epsilon)}. \end{aligned}$$

We can estimate the magnitude of the numerator of the second term:

$$|(\delta_i - \sigma_i)\epsilon + (\epsilon - \delta_i)| = |(2\delta_i - 1)\epsilon + (\epsilon - \delta_i)| = |(2\epsilon - 1)\delta_i| < \delta_i < \epsilon.$$

Case (c):

$$\begin{aligned} E(y_i) &= \frac{(\lfloor \alpha i + e \rfloor - 1)(\epsilon - \sigma_i) + \lceil \alpha i + e \rceil + \lceil \alpha i + e \rceil (\epsilon + \sigma_i)}{(1 + 2\epsilon)} \\ &= \frac{\lfloor \alpha i + e \rfloor \epsilon - \lfloor \alpha i + e \rfloor \sigma_i - (\epsilon - \sigma_i) + \lfloor \alpha i + e \rfloor \delta_i + \lfloor \alpha i + e \rfloor \sigma_i + \lceil \alpha i + e \rceil \epsilon + \lceil \alpha i + e \rceil \delta_i}{(1 + 2\epsilon)} \\ &= \frac{(\alpha i + e) + [(\alpha i + e) - \sigma_i]\epsilon + [(\alpha i + e) + \delta_i]\epsilon - (\epsilon - \sigma_i)}{(1 + 2\epsilon)} \\ &= (\alpha i + e) + \frac{(\delta_i - \sigma_i)\epsilon - (\epsilon - \sigma_i)}{(1 + 2\epsilon)}. \end{aligned}$$

Again, the magnitude of the numerator of the second term can be estimated as

$$|(\delta_i - \sigma_i)\epsilon - (\epsilon - \sigma_i)| = |(1 - 2\sigma_i)\epsilon - (\epsilon - \sigma_i)| = |(1 - 2\epsilon)\sigma_i| < \sigma_i < \epsilon.$$

Case (d)

$$E(y_i) = \frac{[(\alpha i + e) - 1]\epsilon + (\alpha i + e) + [(\alpha i + e) + 1]\epsilon}{(1 + 2\epsilon)} = (\alpha i + e).$$

Case (f)

$$E(y_i) = \lceil \alpha i + e \rceil = (\alpha i + e) + \delta_i, \qquad \delta_i < |\epsilon|.$$

Case (g)

$$E(y_i) = \lfloor \alpha i + e \rfloor = (\alpha i + e) - \sigma_i, \qquad \sigma_i < |\epsilon|.$$

Case (h)

$$E(y_i) = (\alpha i + e).$$

In summary for cases (a)–(h), we have

$$E(y_i) = (\alpha i + e) + \phi_i,$$

where

$$\phi_i = \begin{cases} \frac{(\delta_i - \sigma_i)\epsilon}{1+2\epsilon} & cases(a), (e) \\ \frac{(\delta_i - \sigma_i)\epsilon + (\epsilon - \delta_i)}{1+2\epsilon} & case(b) \\ \frac{(\delta_i - \sigma_i)\epsilon - (\epsilon - \sigma_i)}{1+2\epsilon} & case(c) \\ 0 & cases(d), (h) \\ \delta_i & case(f) \\ -\sigma_i & case(g), \end{cases} \tag{16}$$

and

$$|\phi_i| < \frac{|\epsilon|}{1 + 2\epsilon}.$$

Now it is easy to compute $E(\hat{e})$. We first have

$$E(\sum y_i) = \sum(\alpha i + e) + \sum \phi_i = \frac{n(n-1)}{2}\alpha + ne + \sum \phi_i.$$

Therefore

$$E(\hat{e}) = \frac{1}{n}E(\sum y_i) - \frac{(n-1)}{2}\alpha = e + \frac{1}{n}\sum \phi_i = e + r, \tag{17}$$

where r satisfies

$$|r| = |\frac{1}{n}\sum \phi_i| < \frac{1}{n}\sum \frac{|\epsilon|}{1 + 2\epsilon} = \frac{|\epsilon|}{1 + 2\epsilon}.$$

Now we will derive $Var(\hat{e})$ and then calculate $E[(\hat{e}-e)^2]$. As in the first two sections, we start with the same formula:

$$
\begin{aligned}
Var(\hat{e}) &= E(\hat{e}^2) - [E(\hat{e})]^2 \\
&= E\{[\frac{1}{n}\sum y_i - \frac{(n-1)}{2}\alpha]^2\} - (e + \frac{1}{n}\sum \phi_i)^2 \\
&= \frac{1}{n^2}E[(\sum y_i)^2] - \frac{(n-1)}{n}\alpha E(\sum y_i) + \frac{(n-1)^2}{4}\alpha^2 - e^2 - \frac{2e}{n}\sum \phi_i - \frac{1}{n^2}(\sum \phi_i)^2 \\
&= \frac{1}{n^2}E[(\sum y_i)^2] - \frac{(n-1)}{n}\alpha[\frac{n(n-1)}{2}\alpha + ne + \sum \phi_i] + \frac{(n-1)^2}{4}\alpha^2 - e^2 - \frac{2e}{n}\sum \phi_i - \frac{1}{n^2}(\sum \phi_i)^2 \\
&= \frac{1}{n^2}E[(\sum y_i)^2] - \frac{(n-1)^2}{4}\alpha^2 - (n-1)\alpha e - e^2 - \frac{2}{n}[\frac{(n-1)}{2}\alpha + e]\sum \phi_i - \frac{1}{n^2}(\sum \phi_i)^2.
\end{aligned}
$$

Once again, we need to compute (18)

$$E[(\sum y_i)^2] = \sum_i \sum_j E(y_i y_j). \tag{19}$$

Cases (a)–(h) lead to sixty-four possible combinations for $y_i y_j$ including square terms and cross terms. Fortunately, we don't need to enumerate all of them, thanks to the notation we introduced before. However, we still need to distinguish eight situations, among which seven are the square terms.

(1) Square terms, y_i^2

Cases (a), (e):

$$
\begin{aligned}
E(y_i^2) &= \frac{\lfloor \alpha i + e\rfloor^2(\epsilon + \delta_i) + \lceil \alpha i + e\rceil^2(\epsilon + \sigma_i)}{(1+2\epsilon)} \\
&= \frac{\lfloor \alpha i + e\rfloor^2\delta_i + \lceil \alpha i + e\rceil^2\sigma_i + [(\alpha i + e - \sigma_i)^2 + (\alpha i + e + \delta_i)^2]\epsilon}{(1+2\epsilon)} \\
&= \frac{(\alpha i + e)^2 + \delta_i\sigma_i + [(\alpha i + e)^2 - 2(\alpha i + e)\sigma_i + \sigma_i^2 + (\alpha i + e)^2 + 2(\alpha i + e)\delta_i + \delta_i^2]\epsilon}{(1+2\epsilon)} \\
&= (\alpha i + e)^2 + \frac{\delta_i\sigma_i + 2(\alpha i + e)(\delta_i - \sigma_i)\epsilon + (\delta_i^2 + \sigma_i^2)\epsilon}{(1+2\epsilon)}.
\end{aligned}
$$

Case (b):

$$
\begin{aligned}
E(y_i^2) &= \frac{\lfloor \alpha i + e\rfloor^2(\epsilon + \delta_i) + \lceil \alpha i + e\rceil^2(\delta_i + \sigma_i) + (\lceil \alpha i + e\rceil + 1)^2(\epsilon - \delta_i)}{(1+2\epsilon)} \\
&= \frac{(\alpha i + e)^2 + \delta_i\sigma_i + (\lfloor \alpha i + e\rfloor^2 + \lceil \alpha i + e\rceil^2)\epsilon + 2\lceil \alpha i + e\rceil(\epsilon - \delta_i) + (\epsilon - \delta_i)}{(1+2\epsilon)} \\
&= (\alpha i + e)^2 + \frac{\delta_i\sigma_i + 2(\alpha i + e)(\delta_i - \sigma_i)\epsilon + (\delta_i^2 + \sigma_i^2)\epsilon + 2(\alpha i + e)(\epsilon - \delta_i) + (2\delta_i + 1)(\epsilon - \delta_i)}{(1+2\epsilon)}.
\end{aligned}
$$

Case (c):

$$
\begin{aligned}
E(y_i^2) &= \frac{(\lfloor \alpha i + e\rfloor - 1)^2(\epsilon - \sigma_i) + \lfloor \alpha i + e\rfloor^2(\delta_i + \sigma_i) + (\lceil \alpha i + e\rceil)^2(\epsilon + \sigma_i)}{(1+2\epsilon)} \\
&= \frac{(\alpha i + e)^2 + \delta_i\sigma_i + (\lfloor \alpha i + e\rfloor^2 + \lceil \alpha i + e\rceil^2)\epsilon - 2\lfloor \alpha i + e\rfloor(\epsilon - \sigma_i) + (\epsilon - \sigma_i)}{(1+2\epsilon)} \\
&= (\alpha i + e)^2 + \frac{\delta_i\sigma_i + 2(\alpha i + e)(\delta_i - \sigma_i)\epsilon + (\delta_i^2 + \sigma_i^2)\epsilon - 2(\alpha i + e)(\epsilon - \delta_i) + (2\sigma_i + 1)(\epsilon - \sigma_i)}{(1+2\epsilon)}.
\end{aligned}
$$

Cases (d):

$$E(y_i^2) = \frac{[(\alpha i + e) - 1]^2\epsilon + (\alpha i + e)^2 + [(\alpha i + e) + 1]^2\epsilon}{(1 + 2\epsilon)} = (\alpha i + e)^2 + \frac{2\epsilon}{(1 + 2\epsilon)}.$$

Cases (f):

$$E(y_i^2) = \lceil \alpha i + e \rceil^2 = [(\alpha i + e) + \delta_i]^2 = (\alpha i + e)^2 + 2(\alpha i + e)\delta_i + \delta_i^2.$$

Cases (g):

$$E(y_i^2) = \lfloor \alpha i + e \rfloor^2 = [(\alpha i + e) - \sigma_i]^2 = (\alpha i + e)^2 - 2(\alpha i + e)\sigma_i + \sigma_i^2.$$

Cases (h):

$$E(y_i^2) = (\alpha i + e)^2.$$

In summary of cases (a)–(h), we have

$$E(y_i^2) = (\alpha i + e)^2 + 2(\alpha i + e)\phi_i + \theta_i,$$

where

$$\theta_i = \begin{cases} \frac{\delta_i\sigma_i + (\delta_i^2 + \sigma_i^2)\epsilon}{1+2\epsilon} + \begin{cases} 0 & cases(a),(e) \\ \frac{(2\delta_i+1)(\epsilon-\delta_i)}{(1+2\epsilon)} & case(b) \\ \frac{(2\sigma_i+1)(\epsilon-\sigma_i)}{(1+2\epsilon)} & case(c) \end{cases} \\ \frac{2\epsilon}{(1+2\epsilon)} & case(d) \\ \delta_i^2 & case(f) \\ \sigma_i^2 & case(g) \\ 0 & case(h). \end{cases}$$

(2) Cross terms, $y_i y_j$

Since y_i and y_j are, respectively, functions of η_i and η_j that are independent, we are able to write the cross terms in one form.

$$\begin{aligned} E(y_i y_j) &= [(\alpha i + e) + \phi_i][(\alpha j + e) + \phi_j] \\ &= (\alpha i + e)(\alpha j + e) + (\alpha i + e)\phi_j + (\alpha j + e)\phi_i + \phi_i\phi_j. \end{aligned}$$

Now we can calculate the summation in (19):

$$\begin{aligned} \sum_i\sum_j E(y_i y_j) &= \sum_i\sum_j(\alpha i + e)(\alpha j + e) + 2\sum_i\sum_j(\alpha i + e)\phi_j + \sum_i\sum_{j\neq i}\phi_i\phi_j + \sum_i\theta_i \\ &= \frac{n^2(n-1)^2}{4}\alpha^2 + n^2(n-1)\alpha e + n^2e^2 + 2[\frac{n(n-1)}{2}\alpha + ne]\sum_j\phi_j + \sum_i\sum_{j\neq i}\phi_i\phi_j + \sum_i\theta_i. \end{aligned}$$

Now substituting this last formula into (18) and collecting terms, we have

$$\begin{aligned} Var(\hat{e}) &= \frac{1}{n^2}[\sum_i \sum_{j \neq i} \phi_i \phi_j + \sum_i \theta_i - \sum_i \sum_j \phi_i \phi_j] \\ &= \frac{1}{n^2} \sum_i (\theta_i - \phi_i^2) \\ &\leq \frac{1}{n^2} \sum \theta_i. \end{aligned} \tag{20}$$

Let's examine θ_i carefully. If $\epsilon \geq 0$, the largest magnitude of θ_i occurs in case (b) or case (c). We have

$$\theta_i \leq \frac{\frac{1}{4} + \epsilon + \epsilon}{1 + 2\epsilon} = \frac{1 + 8\epsilon}{4(1 + 2\epsilon)} < \frac{1 + 6\epsilon}{4}.$$

If $\epsilon < 0$, then the largest magnitude of θ_i occurs in case (e). We have

$$\theta_i = \frac{\delta_i \sigma_i + (\delta_i^2 + \sigma_i^2)\epsilon}{1 + 2\epsilon} \leq \frac{\frac{1}{4} + \frac{1}{2}\epsilon}{1 + 2\epsilon} = \frac{1}{4}.$$

In the above two estimates, we used the following three inequalities:

$$\frac{1}{2} \leq \delta_i^2 + \sigma_i^2 \leq 1$$

$$(2\delta_i + 1)(\epsilon - \delta_i) \leq \epsilon \qquad for \qquad 0 \leq \delta_i \leq \epsilon$$

$$(2\sigma_i + 1)(\epsilon - \sigma_i) \leq \epsilon \qquad for \qquad 0 \leq \sigma_i \leq \epsilon$$

Thus we have

$$\begin{aligned} Var(\hat{e}) &\leq \begin{cases} \frac{1}{4n} & \epsilon < 0 \\ \frac{1}{4n} + \frac{3\epsilon}{2n} & \epsilon \geq 0 \end{cases} \\ &\leq \frac{1}{4n} + \frac{3|\epsilon|}{2n} \end{aligned}$$

Now we compute $E[(\hat{e} - e)^2]$. First, we give a useful formula. Let X be a r.v. with mean $E(X)$ and variance $Var(X)$. Let c be an arbitrary constant. Then

$$\begin{aligned} E[(X - c)^2] &= E[X - E(X) + E(X) - c]^2 = E[X - E(X)]^2 + 2[E(X) - c]E[X - E(X)] + [E(X) - c]^2 \\ &= Var(X) + [E(X) - c]^2. \end{aligned}$$

Using this formula in conjunction with (17) and (20), we obtain

$$\begin{aligned} E[(\hat{e} - e)^2] &= Var(\hat{e}) + [E(\hat{e}) - e]^2 \\ &= Var(\hat{e}) + (\frac{1}{n} \sum \phi_i)^2 \\ &< \frac{1}{4n} + \frac{3|\epsilon|}{2n} + \frac{\epsilon^2}{(1 + 2\epsilon)^2}. \end{aligned}$$

We can easily extend assertions 1) and 3) to arbitrary shaped edges. That is, given

$$y_i = \lfloor F(i) + e + \eta_i \rfloor, \qquad i = 0, 1, \ldots, n - 1,$$

where F satisfies $-1 \le \frac{dF(x)}{dx} \le 1$, e is an unknown constant, and η_i's are i.i.d. random variables with uniform distribution in the interval I, then

New assertion 1) if $I = [0, 1)$, then

$$\hat{e} = \frac{1}{n}[\sum_{i=0}^{n-1} y_i - \sum_{i=0}^{n-1} F(i)], \tag{21}$$

the estimator of e, has the following properties:

$$E(\hat{e}) = e,$$

$$Var(\hat{e}) < \frac{1}{4n};$$

New assertion 3) if $I = [-\epsilon, 1 + \epsilon)$ where $|\epsilon| < \frac{1}{2}$, then $\hat{e}$ in (21) satisfies

$$E(\hat{e}) = e + r,$$

where $|r| < \frac{|\epsilon|}{1+2\epsilon}$, and

$$Var(\hat{e}) < \frac{1}{4n} + \frac{3|\epsilon|}{2n},$$

while

$$E[(\hat{e} - e)^2] < \frac{1}{4n} + \frac{3|\epsilon|}{2n} + \frac{\epsilon^2}{(1 + 2\epsilon)^2}.$$

These two extended assertions can be proven by replacing terms of αi, $\alpha \sum i$, etc. with $F(i)$, $\sum F(i)$, etc., in the proof of assertions 1) and 3).

We can also show that y_i and y_j are independent when $i \neq j$. Thus we could first calculate the variance of y_i (which is $\delta_i \sigma_i$ in the case of assertion 1, for example). Then we could calculate the variance of $\hat{e}$ which is just one n-th of that of y_i in assertions 1) and 3), because $\hat{e}$ is the mean of the y_i's plus a constant. This does not work for assertion 2) because there $\hat{e}$ and $\hat{\alpha}$ are not simple means of y_i's. Our derivations are similar for all three assertions. The above approach does not lead to much simplification because most of the work in our approach involves calculating quantities that are also needed in calculating the variance of y_i. However, realizing that y_i and y_j are independent does facilitate our understanding and interpretation of the results.