South African Statistical Journal
Vol. 53, No. 2, 133–142

ON THE CONDITIONAL DISTRIBUTION OF THE MEAN OF THE TWO CLOSEST AMONG A SET OF THREE OBSERVATIONS

I. J. H. Visagie
Department of Statistics, University of Pretoria, Pretoria, South Africa

F. Lombard
Department of Statistics, University of Johannesburg, Johannesburg, South Africa
e-mail: *fred.lombard@gmail.com*

Chemical analyses of raw materials are often repeated in duplicate or triplicate. The assay values obtained are then combined using a predetermined formula to obtain an estimate of the true value of the material of interest. When duplicate observations are obtained, their average typically serves as an estimate of the true value. On the other hand, the "best of three" method involves taking three measurements and using the average of the two closest ones as estimate of the true value.

In this paper, we consider another method which potentially involves three measurements. Initially two measurements are obtained and if their difference is sufficiently small, their average is taken as estimate of the true value. However, if the difference is too large then a third independent measurement is obtained. The estimator is then defined as the average between the third observation and the one among the first two which is closest to it.

Our focus in the paper is the conditional distribution of the estimate in cases where the initial difference is too large. We find that the conditional distributions are markedly different under the assumption of a normal distribution and a Laplace distribution.

Key words: Conditional density, Normal distribution, Laplace distribution, Closest two out of three.

1. Introduction

Chemical analyses of raw materials are often repeated in duplicate or triplicate. The assay values obtained are then combined using a predetermined formula to obtain an estimate of the true value, μ, of the material of interest. When duplicate observations X_1 and X_2 are obtained, their average typically serves as an estimate of the true value. On the other hand, the "best of three" method involves taking three measurements X_1, X_2, and X_3 and using the average of the two closest of these values as estimate of the true value. The statistical properties of this estimator were worked out by Seth (1950) and Lieblein (1952).

In this paper, we consider another method which potentially involves three measurements. Initially two measurements, X_1 and X_2, are obtained. If the difference between X_1 and X_2 is sufficiently small, their average is taken as the estimate. If the difference is too large, then a third independent measurement, X_3, is obtained. Then the estimator, henceforth denoted by $\hat{\mu}$, is the average between

[1] Corresponding author.
MSC2010 subject classifications. 62P30, 62-07.

X_3 and the one among X_1 and X_2 which is closest to X_3. The rationale underlying the method is that whichever one of X_1 and X_2 is closest to X_3 is the least likely to contain a large measurement error.

The usual assumption made in laboratory analysis protocols is that the measurement error is normally distributed. However, Wilson (1923) draws attention to the fact that in some instances there are strong grounds for assuming that the errors follow a Laplace distribution. In the context of a series of observations that estimate the true value of a given parameter, Keynes (1911) asks the following question: "If the most probable value (maximum likelihood estimate in modern terminology) of the quantity is equal to the arithmetic mean of the measurements, what law of error does this imply?" Under the additional assumption that the resulting law of error is symmetric, Keynes shows that it is necessarily normal. Interestingly, he also shows that when the question is restated to enquire about the median instead of the mean, then the resulting law of error is the Laplace distribution which, in standardised form, has density function

$$f(x) = \frac{1}{\sqrt{2}} \exp\left(-\sqrt{2}\,|x-\mu|\right).$$

These facts provide motivation for studying the behaviour of the estimator $\hat{\mu}$ under both the normal and Laplace distribution assumptions.

Even if both X_1 and X_2 are unbiased estimators of μ, the measurement errors attached to each will result in a fixed proportion $\alpha \in (0,1)$ of unacceptably large differences. In this paper, we investigate the conditional distribution of $\hat{\mu}$ given that an unacceptably large difference was observed. On a purely intuitive level, one would expect this conditional distribution to be symmetric around μ. This is indeed the case. However, the form of the symmetry is quite surprising. For realistic values of α we have the following. It turns out that for the normal distribution $\hat{\mu}$ has a bimodal conditional distribution with modes to the left and the right of μ. For the Laplace distribution the surprise is that $\hat{\mu}$ has a unimodal distribution with mode μ.

The remainder of the paper is structured as follows. In Section 2, we define the estimator and derive its conditional density function in the general case where X_1, X_2 and X_3 are independent and identically distribution (i.i.d.) observations from a symmetric distribution. The conditional density function of the estimator is then computed specifically in the normal and Laplace cases and the surprising difference between the two is illustrated and its possible consequences discussed. In Section 3, we consider a data set and demonstrate that the Laplace rather than the normal distribution provides an acceptable fit to the observed data.

2. Conditional distribution of the estimator

In the application sketched in the Introduction, the difference between X_1 and X_2 is regarded as unacceptably large if

$$|X_1 - X_2| > r(\alpha),$$

where $r(\alpha)$ satisfies

$$P\left[|X_1 - X_2| > r(\alpha)\right] = \alpha$$

for an a priori given small positive α. In the following, the argument α in $r(\alpha)$ is suppressed in cases where this is unlikely to lead to confusion. Thus, in the absence of any change in the population

mean or standard deviation, the rate at which unacceptably large differences occur will be α. There are two possibilities, namely

(i) $|X_1 - X_2| \leq r$, in which case the estimate is $\hat{\mu} = (X_1 + X_2)/2$;

(ii) $|X_1 - X_2| > r$, in which case a third observation X_3 is obtained and

$$\hat{\mu} = \begin{cases} \dfrac{X_1 + X_3}{2} & \text{if } |X_1 - X_3| < |X_2 - X_3|, \\ \dfrac{X_2 + X_3}{2} & \text{if } |X_2 - X_3| < |X_1 - X_3|. \end{cases}$$

Since μ and the standard deviation of the error distribution, σ, are assumed to be fixed and known, we may assume without loss of generality that $\mu = 0$ and $\sigma = 1$.

Our interest centers on (ii), hence on the conditional distribution of $\hat{\mu}$ given that $|X_1 - X_2| > r$. Let

$$G(x, \alpha) = P\left[\frac{X_1 + X_3}{2} \leq x,\ X_1 - X_2 > r,\ X_3 > \frac{X_1 + X_2}{2}\right],$$

and

$$g(x, \alpha) = \frac{d}{dx} G(x, \alpha).$$

We show in Appendix 1 that the conditional density function of $\hat{\mu}$, given $|X_1 - X_2| > r$, is

$$h(x, \alpha) = \frac{2}{\alpha} \left[g(x, \alpha) + g(-x, \alpha)\right]. \tag{1}$$

The density h is symmetric around $x = 0$, which is what one would expect a priori. However, from a practitioner's point of view, it is the shape of this density that turns out to be the most interesting and important aspect of the conditional distribution. Given a density function f of the X_i, $g(x, \alpha)$ is given by the expression

$$g(x, \alpha) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} 2f(2x - x_1) \mathbb{J}(x, x_1, x_2) f(x_1) f(x_2) \, dx_1 dx_2, \tag{2}$$

where

$$\mathbb{J}(x, x_1, x_2) = \mathbb{I}\left(x > \frac{3x_1 + x_2}{4}, x_1 - x_2 > r\right), \tag{3}$$

with $\mathbb{I}(\cdot)$ the indicator function. Substitution of the normal or Laplace density functions into (2) does not lead to any substantial algebraic simplification of the expression for $h(x)$. Therefore, we obtain $g(x, \alpha)$ by numerical integration over a fine grid of x values using the MATLAB (Release 2018b) function `integral2.m` — see Appendix 2.

Figure 1 shows the conditional densities (1) of $\hat{\mu}$ in the normal and Laplace distributions. The density in the case of the normal distribution is bimodal, while in the case of the Laplace distribution it is unimodal. In both cases, the estimator is centered around the population average. Nevertheless, a process engineer is bound to be somewhat perplexed upon seeing the bimodal form in the normal distribution. This phenomenon can, to some extent, be explained as follows. First, the Laplace distribution differs from the normal distribution in some important respects. For instance, the

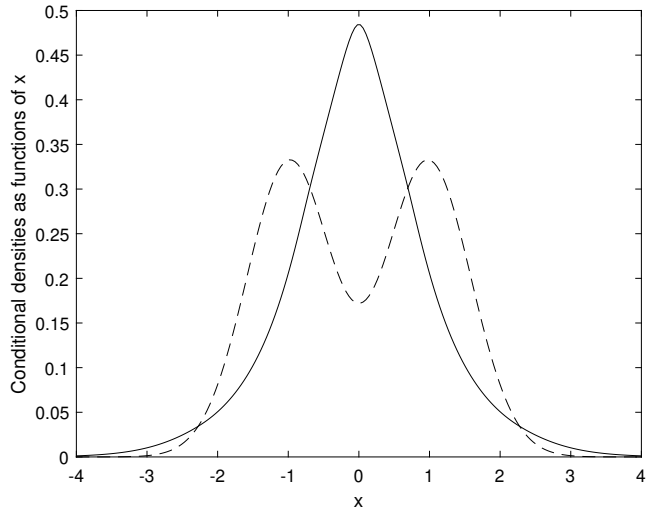

Figure 1. Density of $\hat{\mu}$ given that $|X_1 - X_2| > r$ under the normal (dashed line) and Laplace (solid line) assumptions.

Table 1. Comparison of tail thickness of the normal and Laplace densities.

α	**normal**	**Laplace**
0.100	1.645	1.628
0.050	1.960	2.118
0.025	2.241	2.608
0.010	2.576	3.256
0.005	2.807	3.746

Laplace density has a sharp peak at its point of symmetry, hence is not differentiable there. The tails of the Laplace density are also substantially thicker than those of the normal density. This is perhaps not obvious from visual inspection of Figure 2, which shows plots of the density functions of the two standardised densities.

In order to better appreciate the differences between the tails of the distributions, consider Table 1, which shows the numbers $r(\alpha)$ which satisfy $P(|X_1 - X_2| > r(\alpha)) = \alpha$ for a range of values of α. The indications are that the Laplace distribution has substantially heavier tails than the normal distribution. In fact, the kurtosis of the Laplace distribution is 6, twice that of the normal distribution.

Second, we now argue that, as a consequence of the preceding remark, the resulting density is bimodal in the case where the separation between $g(x, \alpha)$ and $g(-x, \alpha)$ per unit standard deviation is large and unimodal when this separation is small.

Figure 3 shows plots of $g(x, \alpha)$ and $g(-x, \alpha)$ for the normal distribution while Figure 4 shows the corresponding plots for the Laplace distribution. The figures clearly indicate that the separation

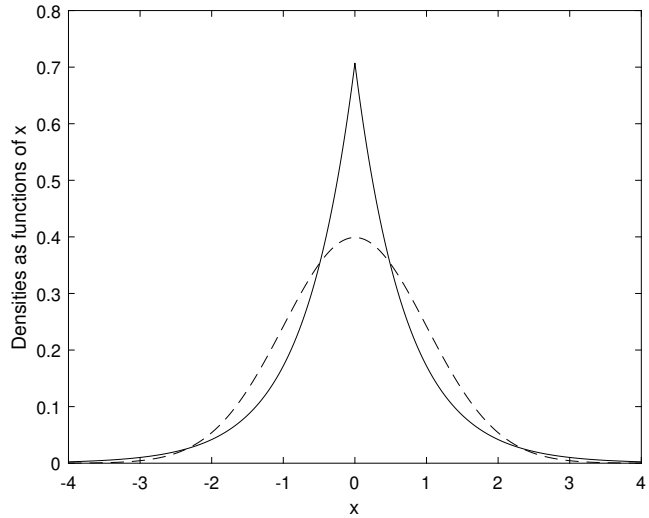

Figure 2. Standardised normal (dashed line) and Laplace (solid line) density functions.

between $g(x, \alpha)$ and $g(-x, \alpha)$ is substantially larger under the normal distribution than under the Laplace distribution.

We now discuss some possible consequences of this difference between the two conditional distributions. The quality of coal is determined, in part, by its ash content. The lower the ash content, the greater is the release of energy when the coal is burnt. As a result, the price of coal is often linked to its ash content. Typically, two determinations, X_1 and X_2, of the ash content of a batch of coal are made and the estimate $\hat{\mu}$ is computed as shown above. As pointed out above, even if both determinations are unbiased estimators of μ, unacceptably large deviations would occur in a proportion α of batches. If μ denotes the contractual ash content, then ash contents in excess of μ could attract penalties, i.e., a lower price than that originally agreed upon.

Figure 5 shows conditional exceedance probabilities

$$P(\hat{\mu} > x \,||X_1 - X_2| > r)$$

over a range of x values for the normal and Laplace distributions.From the figure it is clear that deviations up to 1.5 standard deviations in a normal distribution will tend to attract larger penalties than in a Laplace distribution. This is also rather clear from Figure 1. The economic implications of this are greater than would seem to be apparent at first glance. A batch of coal could consist of several hundreds of tons, which means that the penalty of, for example, 1% of the contractual price could involve hundreds of thousands of dollars.

3. Application to some data

If an enormous amount of data were available, it would be possible to assess empirically which of the conditional densities seen in Figure 1 is the valid one. In the absence of a large amount of data we will have to be satisfied with something less, namely a test of sorts to decide which of the

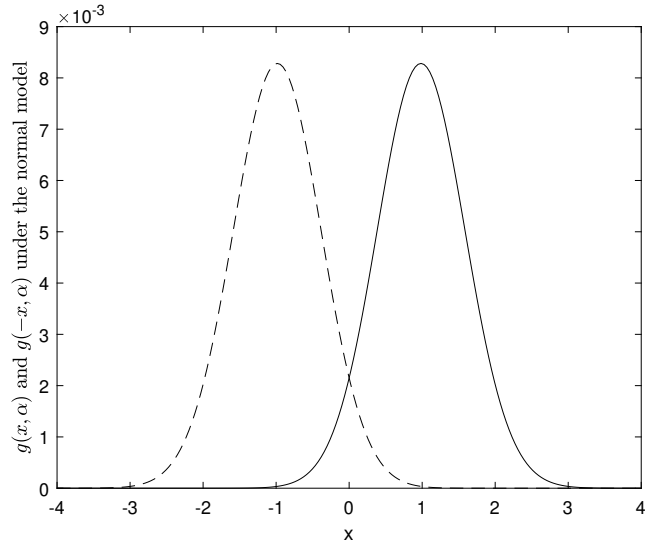

Figure 3. $g\,(x,\alpha)$ (solid line) and $g\,(-x,\alpha)$ (dashed line) under the normal distribution.

normal or Laplace error distributions is applicable. Towards this, Figure 6 shows the differences $X_{1,j} - X_{2,j}$, $j = 1, ..., 199$, for 199 batches of coal. Typically, a prescribed value of σ, the common standard deviation of X_1 and X_2, is attained by following a standard operating procedure. In the present instance, the prescribed value was $\sigma = 0.4$. Thus, we standardise the observed differences as follows:

$$Z_j = \frac{X_{1,j} - X_{2,j}}{0.4}.$$

The resulting sample mean and standard deviation are -0.06 and $1.40\ (\approx \sqrt{2})$ respectively.

In order to determine which of the two distributions is most appropriate we use the standardised Kolmogorov-Smirnov statistic:

$$T_n = \max_{1 \le j \le n} \left| \frac{F\left(Z_j\right) - F_n\left(Z_j\right)}{\sqrt{F\left(Z_j\right)\left(1 - F\left(Z_j\right)\right)}} \right|,$$

where F denotes the hypothesised cumulative distribution function and F_n denotes the usual empirical distribution function

$$F_n\left(x\right) = \frac{1}{n} \sum_{j=1}^{n} \mathbb{I}\left(Z_j \le x\right).$$

The observed values of T_n in the dataset are $T_n = 0.27$ and $T_n = 0.21$ when F is based on the normal and Laplace error distributions respectively. The corresponding p-values obtained from 100 000 Monte Carlo simulations are 0.09 and 0.21 respectively. The first of these is borderline significant at 10% which seems to suggest more support for the Laplace assumption than for the normal in this particular instance. However, due to the heterogeneous nature of batches of coal, it is not possible to arrive at definitive conclusions regarding the validity of the two models from just one data set.

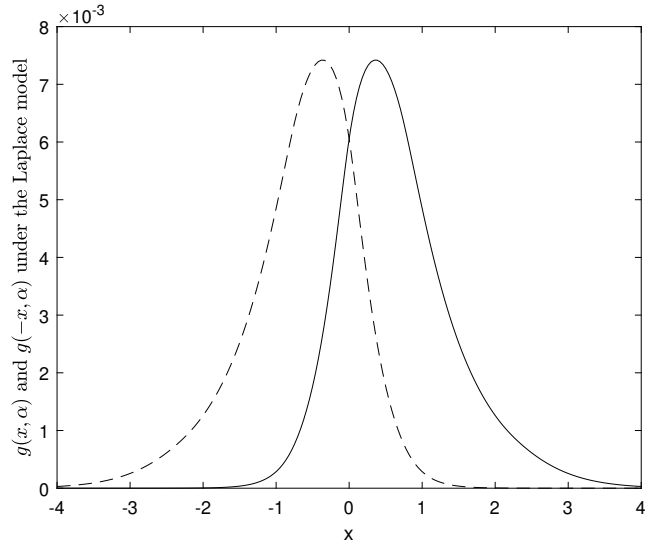

Figure 4. $g\,(x,\alpha)$ (solid line) and $g\,(-x,\alpha)$ (dashed line) under the Laplace distribution.

4. Summary

We consider the distribution of the mean of the two closest among a set of three observations, conditional on the distance between the first two observations being relatively large. The modality of this conditional distribution depends on the distributional assumptions of the measurement error, which is the difference between the value of an observation and the true value. We show that if the errors are from the normal distribution, then the conditional distribution is bimodal, while when the errors are from the Laplace distribution the conditional distribution is unimodal.

In this paper, the phrase "relatively large" refers the absolute difference between two observations. This definition is appropriate when the standard deviation of the observations is fixed, regardless of the mean. However, there are certain quality characteristics of coal for which the standard deviation seems to be an increasing function of the mean. Then it would be more appropriate to work with relative differences, that is, the absolute difference between the two observations divided by their mean. The authors are currently investigating what effect, if any, this change produces in the results of the analysis.

Acknowledgement. We thank the referee for some useful suggestions which led to an improved presentation of the work.

Appendix 1: Derivation of (1)

Let X_1 and X_2 denote the first two observations and let X_3 denote the third sample observation. Given x and a small $\delta > 0$, let dx denote the interval $(x-\delta, x+\delta)$. Then

$$P\left[X \in dx \,||X_1 - X_2| > r\right] = \frac{P\left[X \in dx, |X_1 - X_2| > r\right]}{P\left[|X_1 - X_2| > r\right]} = \frac{1}{\alpha} P\left[X \in dx, |X_1 - X_2| > r\right]. \quad (4)$$

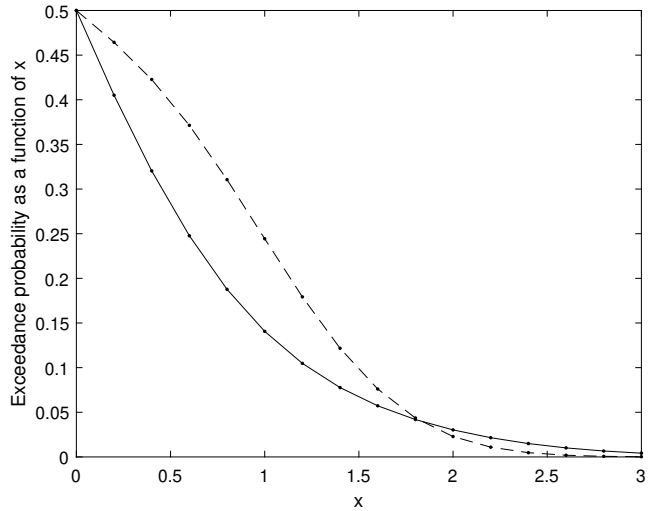

Figure 5. Exceedance probabilities for the normal (dashed line) and Laplace (solid line) distributions.

Furthermore, since (X_1, X_2, X_3) has the same distribution as (X_2, X_1, X_3),

$$P\left[X \in dx, |X_1 - X_2| > r\right] = 2P\left[X \in dx, X_1 - X_2 > r\right]. \tag{5}$$

Now,

$$\begin{aligned}
&P\left[X \in dx, X_1 - X_2 > r\right] \\
&= P\left[\frac{X_1 + X_3}{2} \in dx,\ X_1 - X_2 > r, |X_1 - X_3| < |X_2 - X_3|\right] \\
&\quad + P\left[\frac{X_2 + X_3}{2} \in dx,\ X_1 - X_2 > r,\ |X_1 - X_3| > |X_2 - X_3|\right] \\
&= P\left[\frac{X_1 + X_3}{2} \in dx,\ X_1 - X_2 > r,\ X_3 > \frac{X_1 + X_2}{2}\right] \\
&\quad + P\left[\frac{X_2 + X_3}{2} \in dx,\ X_1 - X_2 > r,\ X_3 < \frac{X_1 + X_2}{2}\right] \\
&= G(x + \delta, \alpha) - G(x - \delta, \alpha) + P\left[\frac{X_2 + X_3}{2} \in dx,\ X_1 - X_2 > r,\ X_3 < \frac{X_1 + X_2}{2}\right],
\end{aligned} \tag{6}$$

with the next to last equality following because

$$\begin{aligned}
&X_1 - X_2 > r \text{ and } |X_1 - X_3| < |X_2 - X_3| \\
&\iff X_1 > X_2 + r \text{ and } X_3 \text{ closer to } X_1 \text{ than to } X_2 \\
&\iff X_1 > X_2 + r \text{ and } X_3 > \frac{X_1 + X_2}{2},
\end{aligned}$$

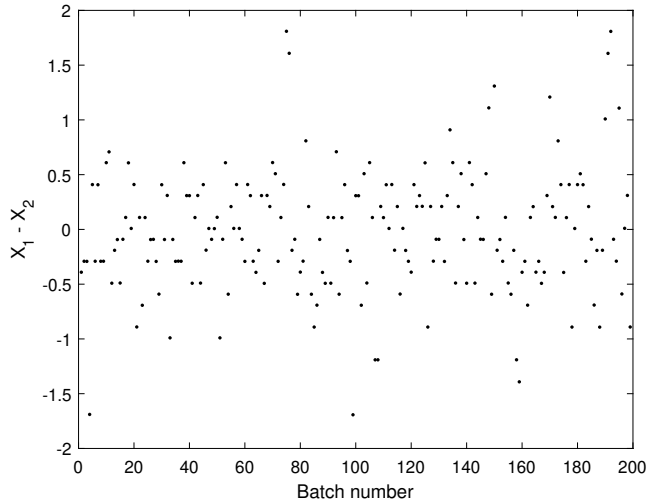

Figure 6. Difference in observed ash contents for 199 batches of coal.

and

$$\begin{aligned} & X_1 - X_2 > r \text{ and } |X_1 - X_3| > |X_2 - X_3| \\ & \iff X_1 > X_2 + r \text{ and } X_3 \text{ closer to } X_2 \text{ than to } X_1 \\ & \iff X_1 > X_2 + r \text{ and } X_3 < \frac{X_1 + X_2}{2}. \end{aligned}$$

Next, the second term in (6) is

$$\begin{aligned} & P\left[\frac{X_2 + X_3}{2} \in dx,\ X_1 - X_2 > r,\ X_3 < \frac{X_1 + X_2}{2}\right] \\ & = P\left[\frac{-X_2 - X_3}{2} \in d(-x),\ (-X_2) - (-X_1) > r,\ -X_3 > \frac{-X_1 - X_2}{2}\right] \\ & = P\left[\frac{X_1 + X_3}{2} \in d(-x),\ X_1 - X_2 > r,\ X_3 > \frac{X_1 + X_2}{2}\right] \\ & = G(-x + \delta, \alpha) - G(-x - \delta, \alpha), \end{aligned} \tag{7}$$

with the next to last equality following because $(-X_2, -X_1, -X_3)$ has the same distribution as (X_1, X_2, X_3). Putting (4), (5), (6) and (7) together, we see that

$$\frac{P\left[X \in dx \,||X_1 - X_2| > r\right]}{2\delta} = \frac{2}{\alpha}\left(\frac{G(x + \delta, \alpha) - G(x - \delta, \alpha)}{2\delta} + \frac{G(-x + \delta, \alpha) - G(-x - \delta, \alpha)}{2\delta}\right).$$

Letting $\delta \downarrow 0$ gets us to (1):

$$h\left(x, \alpha\right) = \lim_{\delta \downarrow 0} \frac{1}{2\delta} P\left[X \in dx \,||X_1 - X_2| > r\right] = \frac{2}{\alpha}\left(g(x, \alpha) + g(-x, \alpha)\right).$$

Appendix 2: Derivation of (2)

Let X_1, X_2 and X_3 be independent random variables with common distribution function F and density function f. Then, for fixed x_1 and x_2,

$$\begin{aligned}
&P\left(\frac{X_1+X_3}{2} \leq x, X_1 - X_2 > r, X_3 > \frac{X_1+X_2}{2} \middle| X_1 = x_1,\ X_2 = x_2\right) \\
&= P\left(\frac{x_1+X_3}{2} \leq x, X_3 > \frac{x_1+x_2}{2}\right)\mathbb{I}(x_1 - x_2 > r) \\
&= P\left(\frac{x_1+x_2}{2} < X_3 < 2x - x_1\right)\mathbb{I}\left(2x - x_1 > \frac{x_1+x_2}{2},\ x_1 - x_2 > r\right) \\
&= \left(F(2x - x_1) - F\left(\frac{x_1+x_2}{2}\right)\right)\mathbb{J}(x, x_1, x_2),
\end{aligned}$$

where $\mathbb{J}$ is defined in (3). Consequently,

$$\begin{aligned}
G(x,\alpha) &= E\left[P\left(\frac{X_1+X_3}{2} \leq x,\ X_1 - X_2 > r,\ X_3 > \frac{X_1+X_2}{2} \middle| X_1, X_2\right)\right] \\
&= E\left[\left(F(2x - X_1) - F\left(\frac{X_1+X_2}{2}\right)\right)\mathbb{J}(x, X_1, X_2)\right].
\end{aligned}$$

Taking the derivative with respect to x, we obtain

$$g(x,\alpha) = E\left[2f(2x - X_1)\,\mathbb{J}(x, X_1, X_2)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} 2f(2x - x_1)\mathbb{J}(x, x_1, x_2)\, f(x_1)\, f(x_2)\, dx_1 dx_2.$$

References

Keynes, J. M. (1911). The principal averages and the laws of error which lead to them. *Journal of the Royal Statistical Society*, **74**, 322–331.

Lieblein, J. (1952). Properties of certain statistics involving the closest pair in a sample of three observations. *Journal of Research of the National Bureau of Standards*, **48**, 255–268.

MATLAB (Release 2018b). The Mathworks, Inc., Natick, Massachusetts, United States.

Seth, G. R. (1950). On the distribution of the two closest among a set of three observations. *The Annals of Mathematical Statistics*, **21**, 298–301.

Wilson, E. B. (1923). First and second laws of error. *Journal of the American Statistical Association*, **18**, 841–851.

Manuscript received 2019-06-28, revised 2019-08-06, accepted 2019-08-12.