CSGB CENTRE FOR STOCHASTIC GEOMETRY AND ADVANCED BIOIMAGING

Linda Vadgård, Jens R. Nyengaard, Johnnie Bremholm and Eva B. Vedel Jensen

The semi-automatic nucleator

No. 03, June 2010

The semi-automatic nucleator

Linda Vadgård Hansen[*], Jens R. Nyengaard[**,***],
Johnnie Bremholm Andersen[***] and Eva B. Vedel Jensen[*,**]

[*] *Department of Mathematical Sciences*
Aarhus University, Denmark
[**] *Centre for Stochastic Geometry and Advanced Bioimaging*
Aarhus University, Denmark
[***] *Stereology and EM Laboratory*
Aarhus University, Denmark

Summary

The nucleator is a well-established manual stereological method of estimating mean cell volume from observations on random cell sections through reference points of the cells. In this paper, we present an automated version of the nucleator that uses automatic segmentation of the boundaries of the cell sections. An expert supervises the process. If the segmentation is judged to be satisfactory, an estimate of the cell volume is calculated automatically on the basis of the whole cell section. In the remaining cases, the expert intervenes and uses the classical nucleator. The resulting estimator is called the semi-automatic nucleator. In the present paper, we study the statistical properties of the semi-automatic nucleator. Formulae for the bias and mean square error are derived. The semi-automatic nucleator may have a small bias but will still in most cases be more efficient than the classical nucleator. Procedures for estimating bias and mean square error (MSE) from a pilot study are provided. The application of the semi-automatic nucleator is illustrated in a study of somatostatin positive inhibitory interneurons which were genetically labeled with green fluorescent protein (GFP). It is found in this study that the number of cells needed for obtaining, for instance, a 5 % precision of the estimate of mean cell volume is 150 and 189 for the semi-automatic and the classical nucleator, respectively. Taking into account that the time spent analyzing one cell is shorter for the semi-automatic nucleator than for the classical nucleator, the semi-automatic nucleator is superior to the classical nucleator.

Key words: computerized image analysis, local stereology, nucleator, volume.

1 Introduction

By means of local stereology, it is possible to determine the size of an object from random sections through a reference point ([11]). Local stereological techniques can be applied without specific assumptions about the shape of the object, which is an important advantage compared to earlier methods, depending on shape assumptions such as spherical, ellipsoidal or some other simplistic shape ([3, 4, 15, 16]). The local stereological methods do not have these severe shape restrictions, but local stereological estimators may have a high variability if the object is far from being

spherical and/or the reference point is not centrally positioned within the object. In case the reference point is very far from being centrally positioned, the recent pivotal estimators based on the invariator principle are to be preferred, see [5] and the accompagnying papers [6] and [7].

The nucleator in its original form (the 'classical' nucleator, see [10]) is used for estimating cell volumes from observations in thick transparent sections. Cells are sampled when their reference point comes into focus. In the case of isotropic sections, two perpendicular lines are normally used in the section through a sampled cell and the expert indicates by the computer mouse the four (or perhaps more) intersection points between the lines and the cell boundary, see the lower right illustration of Figure 1. It is the expert that decides the position of the intersection points.

If it is possible by automatic image analysis to identify the boundaries of the cell sections through the reference points, it is more powerful to use the information available in the whole cell section than the information from two lines. There exists such an estimator of cell volume that uses the whole cell section. The theory of this estimator has been known for quite long, see [12] and references therein, but the estimator has not been in common use because automatic identication of the cell boundaries has not earlier appeared to be a realistic possibility. It can be shown that if the classical nucleator estimate is calculated on the basis of an increasing number of lines, then the estimate will come closer and closer to the value obtained when the whole cell section is used directly. A natural name for the estimator that uses the whole cell section is, therefore, *the integrated nucleator*.

It is important to know how much more precise the integrated nucleator is compared to the classical nucleator based on measurements along two lines. In fact, the classical nucleator is already quite precise if the reference point is centrally positioned because of the antithetic effect of the two perpendicular lines. The gain in precision by using the integrated nucleator depends on the shape of the cells. If the cells are of perfect spherical shape with the centres as reference points, there is no gain. A simulation study in [12] showed that if the cells are prolate ellipsoids with centres as reference points and ratio between major and minor axis equal to 2, then the CE of the integrated nucleator will be 83 % of that of the classical nucleator. If the cells are prolate ellipsoids with ratio equal to 4, then the percent will be 64. Furthermore, using the integrated nucleator, it is not needed to spend time indicating the position of the intersection points between the lines and the cell boundary. All this is under the assumption that the automatic identification of the cell boundaries is correct.

In the present paper, we study the performance of an intermediate option between the classical and the integrated nucleator: *the semi-automatic nucleator* where the expert supervises the process, see Figure 1. The first step of the semi-automatic nucleator is an automatic identification of the cell boundary but it is not a requirement that the identification is correct. If the expert judges that the identification of the cell boundary is satisfactory, then the integrated nucleator based on the automatically segmented cell section is used. If instead the identification of the cell boundary is judged unsatisfactory, the expert intervenes and indicates by the mouse the four (or more) intersection points between the lines and the real cell boundary. The semi-automatic nucleator may have a small bias due to non perfect identifica-

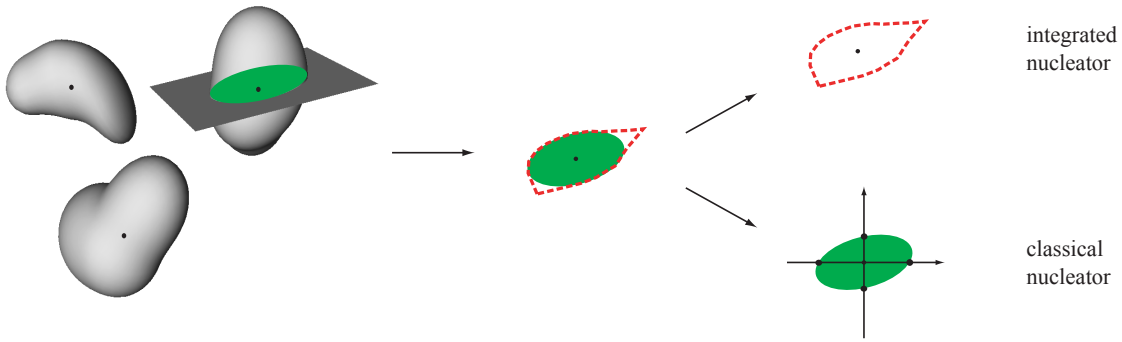

Figure 1: Illustration of the semi-automatic nucleator. The true cell section is shown as the green set while the segmented cell section is delineated by the red broken curve. The integrated nucleator is used if the segmentation is judged to be satisfactory, otherwise the classical nucleator is used.

tion of cell boundaries. As we shall see in a concrete example, this small bias may not be important due to an important gain in precision.

In Section 2 we introduce the various estimators of volume mentioned above. As not all estimators are guaranteed unbiased we discuss mean square error as well as variance relations. In Section 3 we compare the various estimators in a study of somatostatin positive inhibitory interneurons from mice hippocampi, observed by optical fluorescent microscopy. Further aspects are discussed in Section 4. Some mathematical derivations are deferred to an Appendix.

2 Theoretical background

In the following we describe various procedures for estimating mean particle volume. They all consist of a sampling step followed by a measurement step and use a reference point associated with each particle in the sample.

Throughout this paper, Y will denote a random particle with the origin O as reference point, i.e. Y is a compact subset of $\mathbb{R}^3$ containing O. The aim is to estimate the mean particle volume $\mu = EV(Y)$. The estimators to be considered use an isotropic plane L_2 through O. The normal vector of such a random plane is uniformly distributed on a unit half-sphere. Furthermore, we will let L_1 be a line in L_2 through O. Every such line can be uniquely determined by the angle $\theta \in [0, \pi)$ it generates with a fixed axis within L_2. When we need to be specific this line is denoted $L_1(\theta)$.

2.1 The estimators

2.1.1 The classical nucleator

The classical nucleator m_{cl_1} uses measurements along one isotropic line L_1 in L_2 and the unbiased estimator is given by, see e.g. [11],

$$m_{\mathrm{cl}_1} = m_{\mathrm{cl}_1}(Y \cap L_1) = 2\pi \int_{Y_\cap L_1} d(y, O)^2 \, \mathrm{d}y.$$

Here, $d(\cdot, \cdot)$ is the Euclidean distance. If $Y \cap L_1$ consists of a single line segment $[y_-, y_+]$, then m_{cl_1} takes the simple form

$$m_{\mathrm{cl}_1} = \frac{2\pi}{3}\Big(d(y_-, O)^3 + d(y_+, O)^3\Big). \tag{1}$$

If $Y \cap L_1$ is a union of such line segments, m_{cl_1} involves the measurements of distances from O to the endpoints of all line segments. In practice, usually an expert decides where the intersection points between the line and the boundary of Y are positioned.

In [10] it was suggested to use measurements along more than one line to increase the efficiency of the estimator. The resulting estimator m_{cl_2} is based on measurements along two perpendicular lines. This unbiased estimator is given by

$$m_{\mathrm{cl}_2} = \tfrac{1}{2}\big(m_{\mathrm{cl}_1}(Y \cap L_1(\Theta)) + m_{\mathrm{cl}_1}(Y \cap L_1(\Theta + \pi/2))\big),$$

where Θ is uniform on $[0, \pi/2)$. Introducing a perpendicular line usually implies a significant reduction in estimator variance because of the antithetic effect and m_{cl_2} is much used in practice today. Recent references are [1] and [13].

2.1.2 The integrated nucleator

The integrated nucleator is an unbiased estimator of particle volume based on the whole section $Y \cap L_2$, see e.g. [11] and references therein. The estimator takes the following form

$$m_{\mathrm{int}} = m_{\mathrm{int}}(Y \cap L_2) = 2 \int_{Y \cap L_2} d(y, O) \, \mathrm{d}y^2.$$

Here, $\mathrm{d}y^2$ denotes the element of the area measure on L_2. This estimator has not been in common use in the study of biological cell populations because accurate recognition of $Y \cap L_2$ has not appeared to be a realistic possibility. Today, due to improved labelling techniques, automatic identification of $Y \cap L_2$ is no longer unrealistic.

Using polar decomposition in L_2, we find that

$$\begin{aligned} m_{\mathrm{int}} &= 2 \int_0^\pi \int_{Y \cap L_1(\theta)} d(y, O)^2 \, \mathrm{d}y^1 \, \mathrm{d}\theta \\ &= \int_0^\pi m_{\mathrm{cl}_1}(Y \cap L_1(\theta)) \frac{\mathrm{d}\theta}{\pi}. \end{aligned} \tag{2}$$

It follows that

$$E\,(m_{\mathrm{cl}_1} | Y, L_2) = E\,(m_{\mathrm{cl}_2} | Y, L_2) = m_{\mathrm{int}}(Y \cap L_2). \tag{3}$$

Equation (2) also shows that the integrated nucleator can be regarded as a classical nucleator based on an infinite number of lines.

2.1.3 The automatic nucleator

If it is possible by automatic image analysis to recognize the particle section $Y \cap L_2$, it appears natural to use this information in connection with the integrated nucleator. Let $\tilde{Y}_2$ be the readily available estimate of the cell section $Y \cap L_2$ obtained using computerized image analysis and assume that $\tilde{Y}_2$ contain O. Furthermore, let

$$m_{\text{aut}} = m_{\text{int}}(\tilde{Y}_2).$$

If the segmentation is perfect, i.e. $\tilde{Y}_2 = Y \cap L_2$, then the automatic nucleator m_{aut} provides an unbiased estimator of the volume of Y. If, on the other hand, the segmentation is of poor quality, then m_{aut} may be heavily biased.

2.1.4 The semi-automatic nucleator

As mentioned above, the advantages of the automatic nucleator depend on a satisfactory automatic identification of the cell boundaries. Below we present an intermediate possibility between the classical nucleator and the automatic nucleator where the expert supervises the measurement process performed using the automatic nucleator and only interferes if the recognition of the cell boundaries is not satisfactory. In such cases, the expert will perform the measurements needed for the classical nucleator. We will refer to this estimator as the semi-automatic nucleator m_{semi}.

If the segmentation is judged satisfactory, then $m_{\text{semi}} = m_{\text{aut}}$, otherwise $m_{\text{semi}} = m_{\text{cl}_2}$. In contrast to the classical and integrated nucleator, the semi-automatic nucleator may have a small bias. Let A be the event that the segmentation is accepted, i.e. judged satisfactory. It can be shown, see the Appendix, that

$$\text{bias}\,(m_{\text{semi}}) := E\,(m_{\text{semi}}) - \mu = p\,E\,(m_{\text{aut}} - m_{\text{int}}|A),$$

where $p = P(A)$ is the probability that the segmentation is accepted and $E\,(m_{\text{aut}} - m_{\text{int}}|A)$ is the mean difference between the automatic and integrated nucleator among cells for which the segmentation is accepted. The bias will be small because $m_{\text{aut}} - m_{\text{int}}$ is small for a cell section with accepted segmentation.

2.2 Variance and mean square error relations

If m denotes either m_{cl_1} or m_{cl_2}, then we have the following equation and lower bound for the variance of the estimator

$$\begin{aligned} \text{Var}(m) &= \text{Var}(E\,(m|Y, L_2)) + E\ \text{Var}(m|Y, L_2) \\ &= \text{Var}(m_{\text{int}}) + E\ \text{Var}(m|Y, L_2) \\ &\geq \text{Var}(m_{\text{int}}). \end{aligned} \tag{4}$$

At the second equality sign, we have used (3).

To compare estimators that are not necessarily unbiased the mean square error (MSE) is a more appropriate measure of variability. The MSE of an estimator m of μ is given by, see e.g. [2],

$$\begin{aligned} \text{MSE}\,(m) &= E\,(m - \mu)^2 \\ &= \text{Var}(m) + \text{bias}\,(m)^2. \end{aligned}$$

As m_{int} is unbiased we have that

$$\text{bias}\,(m) = E\,m - E\,m_{\text{int}}.$$

If m is unbiased then $\text{MSE}\,(m) = \text{Var}(m)$. Therefore the variance relations in (4) also holds for the corresponding MSEs, when m is unbiased.

For m_{semi}, we find, cf. the Appendix,

$$\text{MSE}\,(m_{\text{semi}}) = p\,\text{MSE}\,(m_{\text{aut}}|A) + (1-p)\,\text{MSE}\,(m_{\text{cl}_2}|A^c), \tag{5}$$

where $\text{MSE}\,(m_{\text{aut}}|A)$ is the mean square error of the automatic nucleator among cells for which a satisfactory segmentation is obtained. Likewise, $\text{MSE}\,(m_{\text{cl}_2}|A^c)$ is the mean square error of m_{cl_2} among cells for which the segmentation is not satisfactory. The semi-automatic nucleator m_{semi} will be more precise than the classical nucleator m_{cl_2} if

$$\text{MSE}\,(m_{\text{semi}}) \leq \text{Var}(m_{\text{cl}_2}) = \text{MSE}\,(m_{\text{cl}_2}). \tag{6}$$

Using (5), it follows that (6) is equivalent to

$$\text{MSE}\,(m_{\text{aut}}|A) \leq \text{MSE}\,(m_{\text{cl}_2}|A). \tag{7}$$

The inequality (7) is satisfied if m_{aut} is replaced by m_{int}, cf. the Appendix, and is therefore likely to hold since $\text{MSE}\,(m_{\text{aut}}|A)$ is calculated for cells with satisfactory segmentation. The magnitude of the gain in efficiency by using m_{semi} instead of m_{cl_2} will depend on the shapes of the cells and, for a given cell population, on how large the fraction p of cells with satisfactory segmentation is. Note also that the workload associated with determining m_{semi} will not be larger than that associated with m_{cl_2}.

2.3 Estimation of bias and MSE

In order to determine which one of the estimators will be the most efficient one in an actual study, it is advisable to perform a pilot study.

Let $Y_1, \ldots, Y_M$ denote the sample of cells in such a study. For all estimators m except m_{semi}, we imagine that we perform N replicated determinations $m_{i1}, \ldots, m_{iN}$ of the estimator for each cell Y_i, $i = 1, \ldots, M$. The MSE of m can then be estimated by

$$\widehat{\text{MSE}}\,(m) = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(m_{ij} - \hat{\mu})^2. \tag{8}$$

In (8) and the following, we use $\hat{\mu} = \bar{m}_{\text{int}\cdot}$, but an estimate of μ based on one of the other unbiased estimators might be used as well.

Since m_{int} and m_{aut} depend on the planar section as a whole, we will typically have $N = 1$ for these estimators. For m_{cl_1} and m_{cl_2}, we may be interested in estimating the within section variance due to the random positioning of the lines inside the sections and in this case $N > 1$. The within section variance can be estimated by

$$\frac{1}{M(N-1)}\sum_{i=1}^{M}\sum_{j=1}^{N}(m_{ij} - \bar{m}_{i\cdot})^2$$

For m_{semi}, we let I ($\#I$) denote the (number of cells in the) subpopulation of the sampled cells that are judged to have a satisfactory segmentation. The probability of a satisfactory segmentation is estimated by $\hat{p} = \#I/M$. Then, an estimate of $\text{MSE}\,(m_{\text{semi}})$ is, cf. (5),

$$\begin{aligned}
&\widehat{\text{MSE}}\,(m_{\text{semi}}) \\
&\quad = \hat{p}\,\widehat{\text{MSE}}\,(m_{\text{aut}}|A) + (1-\hat{p})\,\widehat{\text{MSE}}\,(m_{\text{cl}_2}|A^c) \\
&\quad = \frac{\#I}{M}\frac{1}{\#I}\sum_{i\in I}(m_{\text{aut},i}-\hat{\mu})^2 + \left(1-\frac{\#I}{M}\right)\frac{1}{(M-\#I)N}\sum_{i\notin I}\sum_{j=1}^{N}(m_{\text{cl}_2,ij}-\hat{\mu})^2 \\
&\quad = \frac{1}{M}\left(\sum_{i\in I}(m_{\text{aut},i}-\hat{\mu})^2 + \frac{1}{N}\sum_{i\notin I}\sum_{j=1}^{N}(m_{\text{cl}_2,ij}-\hat{\mu})^2\right).
\end{aligned}$$

Among the estimators considered, only m_{aut} and m_{semi} may be biased. For $m = m_{\text{aut}}$, the bias is estimated by

$$\widehat{\text{bias}}\,(m_{\text{aut}}) = \bar{m}_{\text{aut}\cdot} - \hat{\mu},$$

while the bias of m_{semi} can be estimated by

$$\widehat{\text{bias}}\,(m_{\text{semi}}) = \frac{1}{M}\sum_{i\in I}(m_{\text{aut},i} - m_{\text{int},i}).$$

2.4 Discrimination between estimators

A yardstick for the precision of an estimator of the mean cell volume μ is the number n of cells needed for obtaining a given precision. The estimator

$$\hat{m} = \frac{1}{n}\sum_{i=1}^{n} m(Y_i)$$

of mean cell volume has MSE of the form

$$\text{MSE}\,(\hat{m}) = \tfrac{1}{n}\,\text{Var}(m) + \text{bias}\,(m)^2. \tag{9}$$

It is seen that in order to obtain a relative error $\rho = \sqrt{\text{MSE}\,(\hat{m})}/\mu$ of the estimate $\hat{m}$ of mean cell volume, we need to sample

$$n = \frac{\text{MSE}\,(m) - \text{bias}\,(m)^2}{\rho^2\mu^2 - \text{bias}\,(m)^2} \tag{10}$$

cells. If we let

$$\text{Relative bias}\,(m) = \text{bias}\,(m)/\mu$$

and

$$\text{Relative error}\,(m) = \sqrt{\text{MSE}\,(m)}/\mu,$$

then (10) reduces to

$$n = \frac{\text{Relative error}^2(m) - \text{Relative bias}^2(m)}{\rho^2 - \text{Relative bias}^2(m)}. \tag{11}$$

Estimates of μ, $\text{bias}\,(m)$ and $\text{MSE}\,(m)$ are available from the pilot study, cf. Section 2.3.

3 A comparative study

In this section we will compare the described estimators in a study of somatostatin positive inhibitory interneurons from transgenic GFP-GAD (FVB-TgN (GadGFP) 45704Swn) mice hippocampi, observed by optical fluorescent microscopy.

3.1 Materials and preparation methods

The animal study was approved by the Danish Animal Experiments Inspectorate. Two GFP-GAD (FVB-TgN (GadGFP) 45704Swn) mice were anesthetized by sodium pentobarbital (50 mg/kg i.p.) and transcardially perfused with phosphate-buffered 4 % paraformaldehyde. Following post-fixation in 4 % paraformaldehyde overnight at 4 °C, the brains were removed, and the hippocampi were cut out and embedded in 5 % agar in the isector ([14]) to generate isotropic sections. Using a Vibratome (Vibratome, St. Louis, USA), the brains were sectioned exhaustively at 65 µm. Every sixth section was selected for sampling. For counterstaining, these sections were transferred into DAPI (Sigma, St. Louis, USA) solution. Each section was wet mounted on a super frost slide and was dried at room temperature for only 10 min. An aqueous mounting media was used to adhere the coverglass and care was taken to remove the excess mounting media. Z-stacks were recorded at a confocal microscope (Zeiss LSM 510 META system) using a 40× NA 1.2 C-Apochromat objective and systematic sampling. The laser line used was 488 nm, image size 225 µm × 225 µm and voxel size 0.44 µm × 0.44 µm × 0.44 µm (undersampling in the XY plane wrt. the optimum resolution which is 0.1 µm × 0.1 µm at this wavelength and NA).

3.2 Sampling and segmentation

Figure 2 shows examples of somatostatin positive inhibitory interneurons (green) in an optical section of tissue as seen under a confocal microscope. A characteristic of these cells is the dendrites/axons which are also visible in Figure 2. The dendrites will not be regarded as part of the actual cell body and, accordingly, they do not contribute to the cell volume.

A total of 91 cells were sampled using an optical disector within an isotropic thick section and an unbiased counting frame. For each sampled cell a segmentation of the boundary of the central cell section was performed using the max-vol function in Visiomorph (Visiopharm, Hørsholm, Denmark). The segmentation results in a set of xy-coordinates on the segmentation boundary. These 91 sets of xy-coordinates constitute the data to be used in the following analysis.

3.3 Constructing the true cell section

In order to study the performance of the different types of estimators, we fitted for each sampled cell section a spline to those xy-coordinates not originating from the dendrites or from other cells visible from other layers of the thick section. Subsequently, the spline curve representing $Y \cap L_2$ was approved by an expert and used in the following as the true cell section.

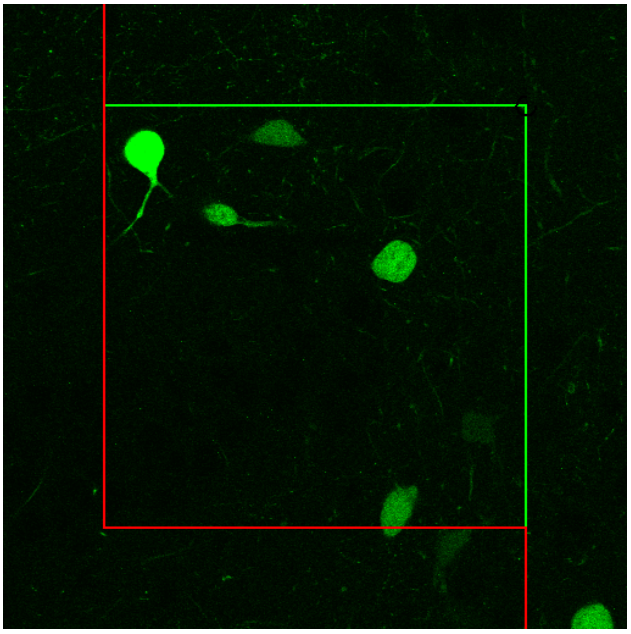

Figure 2: Somatostatin positive inhibitory interneurons in mice hippocampi have been genetically labelled with green fluorescent protein (GFP) and observed under a confocal microscope. The optical disector with the unbiased counting frame is used to sample the interneurons.

Figure 3 shows an example of a sampled cell (green) as seen under the confocal laser microscope. The observed xy-coordinates (red crosses) are superimposed along with the calculated spline (blue line) and the centre of mass of the spline (white dot). As these nerve cells do not have a natural reference point we will use the centre of mass for this purpose. This centre of mass in the central section of the cell may be regarded as an approximation to the centre of mass in 3D.

Figure 4 shows for each cell the true cell section (green) and the boundary of the segmented cell (red broken line). The dot indicates the centre of mass according to the true cell section. The cell sections above the horizontal line of Figure 4 have satisfactory segmentation while for those below the segmentated cell sections are too large, either because parts of the dendrites are included or neighbouring cells interfer.

3.4 Comparison of estimators

These visual observations are reflected in the histograms in Figure 5 displaying m_{int} (left) and m_{aut} (right), respectively. Furthermore, the histograms indicate a right skewed distribution, especially for m_{aut}, with means $E(m_{\text{int}})$ and $E(m_{\text{aut}})$ estimated by $1446\,\mu\text{m}^3$ and $1904\,\mu\text{m}^3$, respectively. Furthermore, the estimates of $\text{Var}(m_{\text{int}})$ and $\text{Var}(m_{\text{aut}})$ are $706\,620\,\mu\text{m}^6$ and $1\,944\,011\,\mu\text{m}^6$, respectively. The estimate of the bias of m_{aut} is $458\,\mu\text{m}^3$ resulting in an estimated MSE of m_{aut} of $2\,132\,314\,\mu\text{m}^6$. The estimated relative bias of m_{aut} is quite large, i.e. $(\bar{m}_{\text{aut}\cdot} - \hat{\mu})/\hat{\mu} = 0.32$. It is therefore not advisable to use the automatic nucleator in this study.

For the implementation of the semi-automatic nucleator, we used a distance in

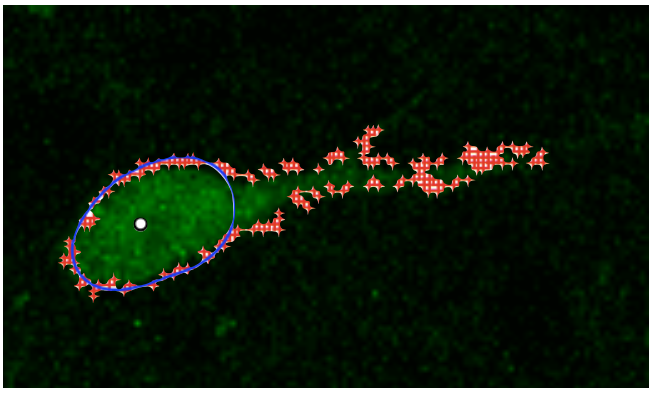

Figure 3: A somatostatin positive inhibitory interneuron as observed under the confocal laser microscope. The observed xy-coordinates on the segmentation boundary are shown (red crosses) along with the calculated spline (blue line) and the centre of mass relative to the spline (white dot).

the judgement of the discrepancy between the true cell section $Y \cap L_2$ and the automatically segmented cell section called $\tilde{Y}_2$. The segmentation was judged satisfactory if

$$d(Y \cap L_2, \tilde{Y}_2) < \varepsilon$$

where $\varepsilon \geq 0$ and d denotes a distance on the set of subsets of L_2. Two examples of distances were considered

$$d_1(B_1, B_2) = \frac{|A(B_1) - A(B_2)|}{A(B_1)} \quad \text{and} \quad d_2(B_1, B_2) = \frac{|m_{\text{int}}(B_1) - m_{\text{int}}(B_2)|}{m_{\text{int}}(B_1)}.$$

Here, A denotes area and $B_1, B_2 \subset L_2$. In practice, the distance d_1 seems straightforward for an expert to evaluate since the distance only involves differences in area between the true cell section and the segmented cell. But as it is the difference in estimated volume that is important d_2 seems more appropriate. Figure 6 shows $d_2(Y \cap L_2, \tilde{Y}_2)$ plotted against $d_1(Y \cap L_2, \tilde{Y}_2)$ in a double logarithmic scale along with an estimated regression line $y = 0.91x + 0.38$. It is seen from this plot, that the two distances provide quite similar results.

In the analysis below we used d_2 and $\varepsilon = 0.15$. A total of 66 cells had a satisfactory segmentation according to this criterion. They are shown above the horizontal line of Figure 4. The probability p is estimated by $66/91 = 0.73$. Using the empirical relationship between d_1 and d_2, $d_2(Y \cap L_2, \tilde{Y}_2) = 0.15$ corresponds to $d_1(Y \cap L_2, \tilde{Y}_2) = 0.08$.

For each sampled cell, Figure 7 shows the volume estimates m_{aut}, m_{cl_1}, m_{cl_2} and m_{semi} plotted against m_{int}, respectively. The scale is $\ln(\mu\text{m}^3)$. The plots confirm that m_{aut} is quite heavily biased while m_{int}, m_{cl_1} and m_{cl_2} are unbiased. Also, m_{semi} appears virtually unbiased and somewhat more precise than m_{cl_2} which again is more precise than m_{cl_1}.

Table 1 shows, for each estimator considered, the relative bias and the relative error of the estimated cell volume, when analyzing one cell. For m_{cl_1} and m_{cl_2}, we also estimated the within section variance which was found to constitute 89 % and

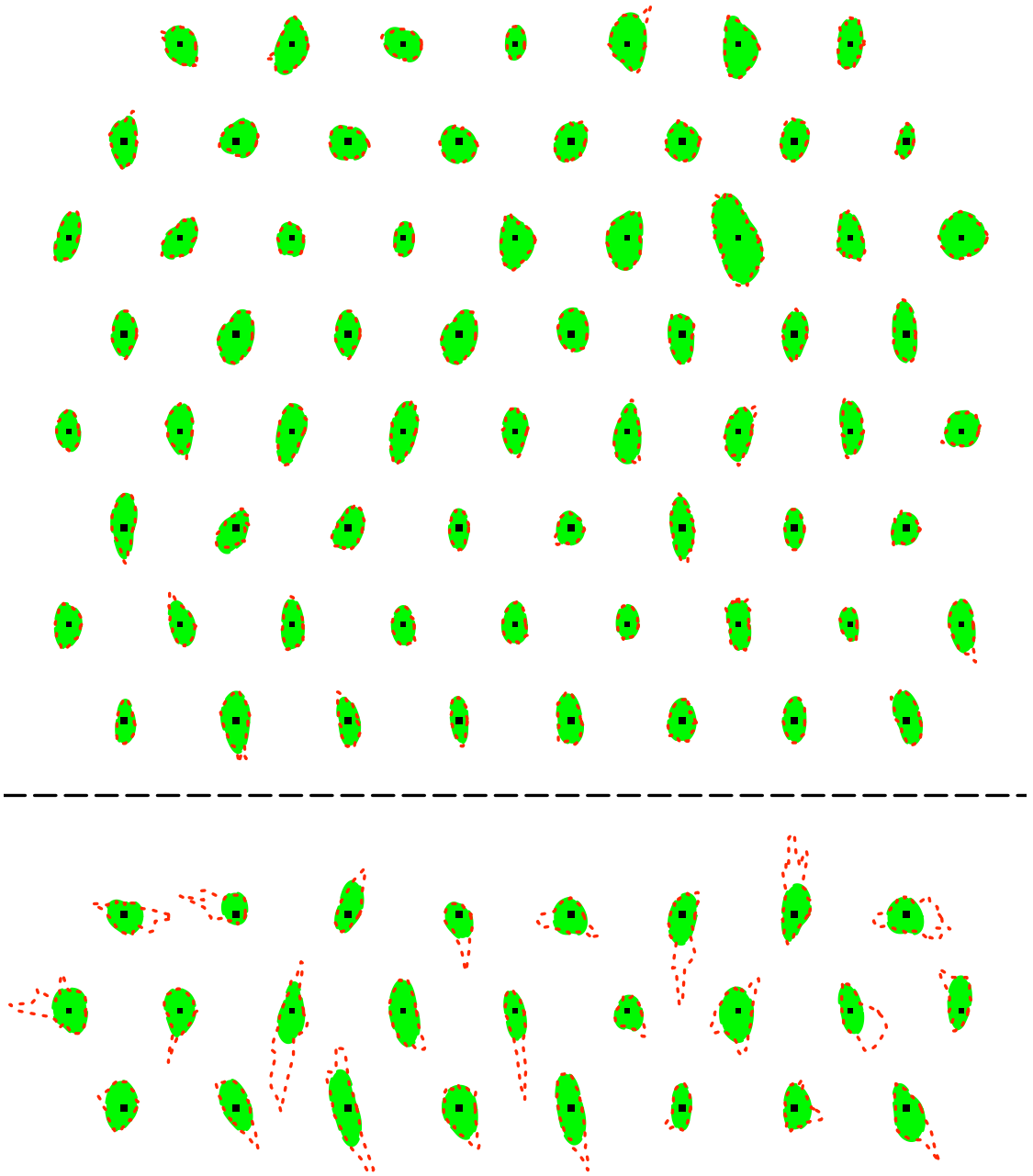

Figure 4: The 'true' cell section (green) and the boundary of the segmented cell (red broken line) for each of the 91 sampled somatostatin positive inhibitory interneurons. The dot indicates the reference point of the cell. The cell sections above the horizontal line have satisfactory segmentation.

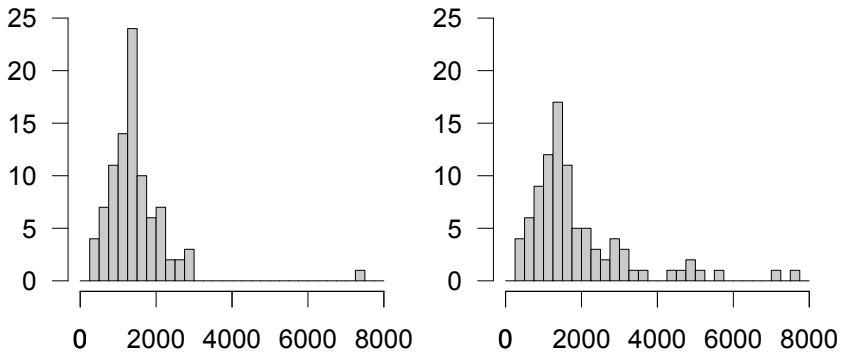

Figure 5: Histograms of the estimated volumes (μm^3) based on the integrated nucleator (left) and the automatic nucleator (right).

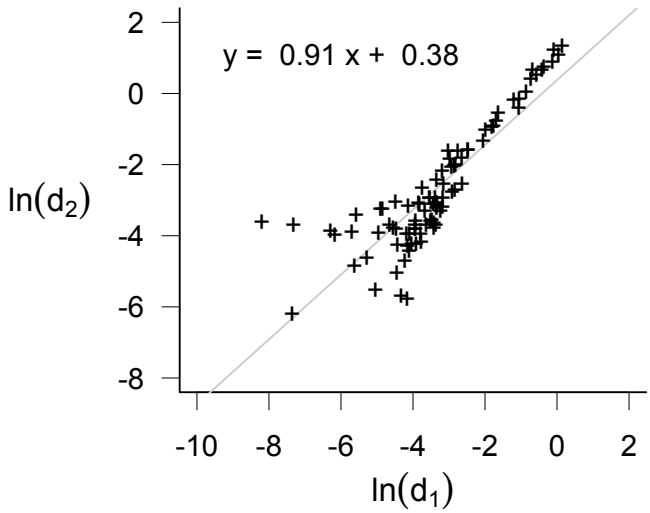

Figure 6: Plot of $d_2(Y \cap L_2, \tilde{Y}_2)$ against $d_1(Y \cap L_2, \tilde{Y}_2)$ in a double logarithmic scale.

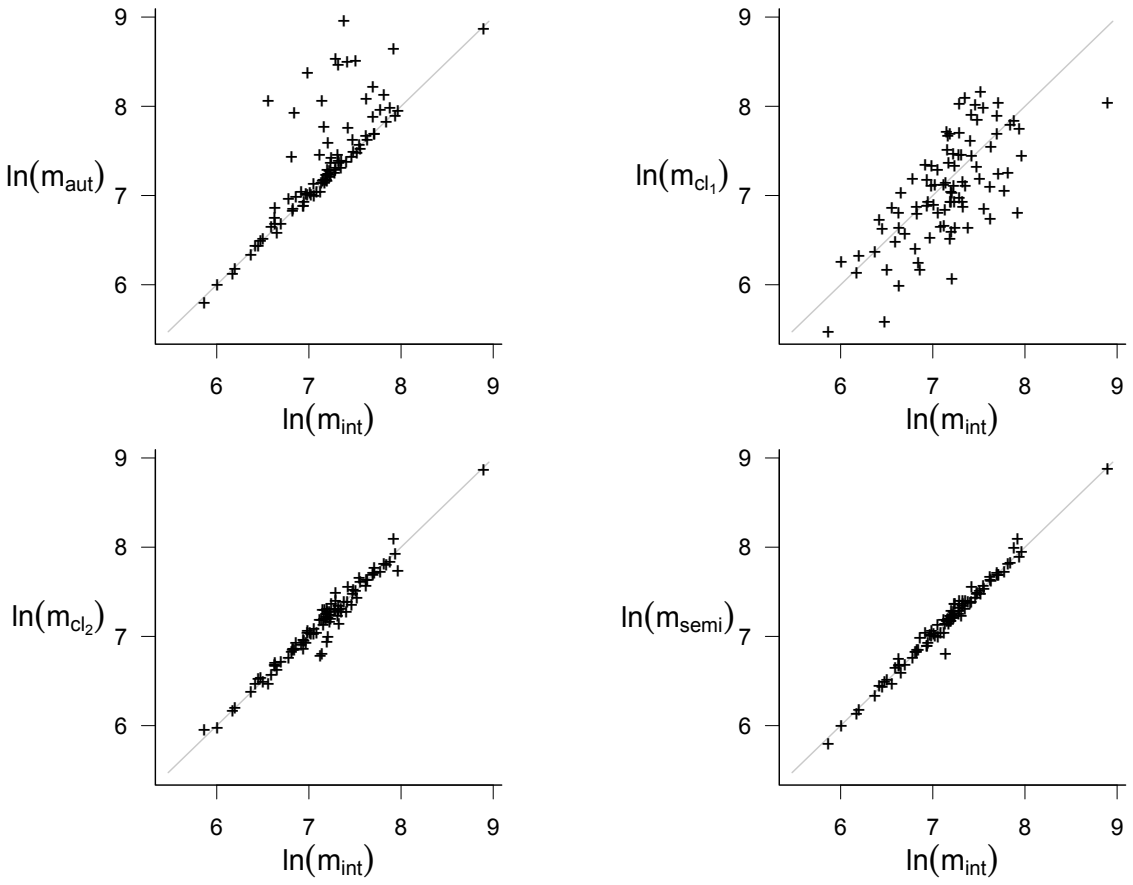

Figure 7: Volume estimates $\hat{m}_{\text{aut}}$, $\hat{m}_{\text{cl}_1}$, $\hat{m}_{\text{cl}_2}$ and $\hat{m}_{\text{semi}}$ plotted against $\hat{m}_{\text{int}}$ in a double logarithmic scale. The scale is $\ln(\mu\text{m}^3)$.

29 % of the total variance, respectively. Still m_{semi} is more precise than m_{cl_2}, see Table 1.

	m_{int}	m_{aut}	m_{cl_1}	m_{cl_2}	m_{semi}
Relative bias	—	0.32	—	—	0.004
Relative error	0.58	1.01	1.79	0.69	0.61

Table 1: For each of the estimators considered, the relative bias and the relative error of the estimated cell volume, when analyzing one cell, are shown. Note that m_{semi} is virtually unbiased.

Using (11) and the results shown in Table 1, we can estimate the number of cells needed to analyze in order to obtain a given relative error ρ of the estimated mean cell volume. Figure 8 shows the number of cells needed to be analyzed for m_{int} (full), m_{semi} (dashed) and m_{cl_2} (dotted), respectively, to obtain a given relative error of the estimated mean cell volume between 0.02 and 0.10. For a given relative error in this interval it is seen that we always need to include less cells for m_{semi} than for m_{cl_2}. As an example, in order to obtain a relative error of the estimate of mean cell volume of 0.05, it is needed to sample and analyze 134, 150 and 189 cells using the integrated nucleator, semi-automatic nucleator and classical nucleator, respectively. The use of m_{int} is not feasible in this study because it requires the correct segmentation of *all* cells. The semi-automatic nucleator is more efficient than the classical nucleator, since the same precision can be obtained by sampling fewer cells in the case of the semi-automatic nucleator and in addition the time spent analyzing a cell is shorter.

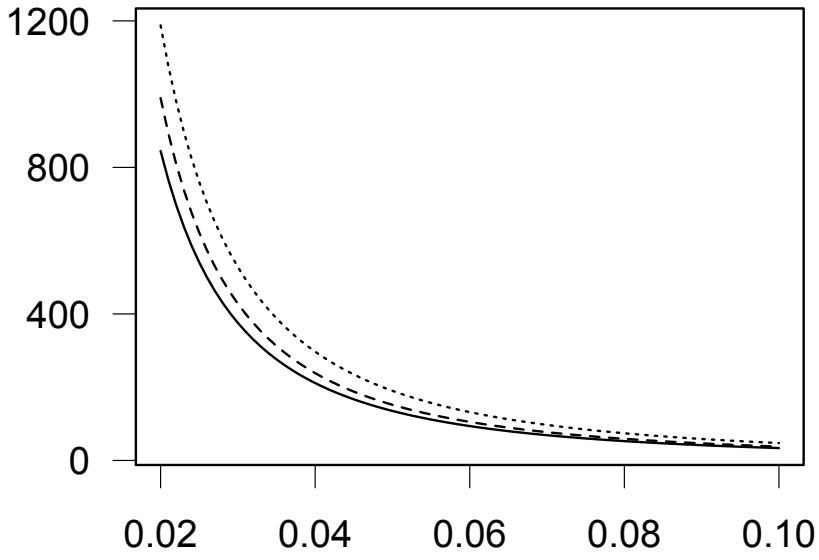

Figure 8: The number of cells needed to be analyzed for m_{int} (full), m_{semi} (dashed) and m_{cl_2} (dotted), respectively, as a function of the relative error of the estimate of the mean cell volume.

4 Discussion

In the present paper, we have proposed a new method of estimating mean cell volume, the semi-automatic nucleator. The method uses sections through reference points of the cells. Automatic segmentation of the cell sections is an integral part of the method. A gain in efficiency compared to the classical nucleator can generally be expected. The magnitude of the gain in efficiency increases with the fraction of analyzed cell sections with satisfactory segmentation.

An expert supervises the process and interferes if the segmentation of a particular cell section is judged unsatisfactory. It is important that the expert is trained in using an appropriate threshold for deciding whether a segmentation is satisfactory.

The efficiency of nucleator type estimators depends on the positioning of the reference point inside the cell. If the reference point is very far from being centrally positioned, we advice to use the pivotal estimators suggested in [5, 6, 7] instead.

Recently, weighted sampling has been introduced in stereology under the name of the *proportionator*, as a successful method of reducing variance in the stereological estimation of number ([8, 9]). During the research work reported in the present paper, we also considered to reduce the within section variance of the classical nucleator by changing the distribution of the lines in the section from uniform to a weighted orientation distribution, using the information available in the segmentated cell section $\tilde{Y}_2$.

More specifically, we considered lines $L_1 = L_1(\theta)$ with the following orientation density distribution p, depending on the shape of $\tilde{Y}_2$,

$$p(\theta) \propto \int_{\tilde{Y}_2 \cap L_1(\theta)} d(y, O)^2 \, \mathrm{d}y,$$

for $\theta \in [0, \pi)$. This density favors lines with a large intersection with the cell. The following estimator

$$\frac{m_{\mathrm{cl}_1}(Y \cap L_1(\Theta))}{\pi \, p(\Theta)}, \tag{12}$$

where Θ has density p, is an unbiased estimator of $V(Y)$. In the case of a perfect segmentation, i.e. $\tilde{Y}_2 = Y \cap L_2$, the values of this estimator will not depend on Θ and therefore its within section variability will be zero.

When comparing the estimator in (12) to the classical nucleator estimators, using the data analyzed in Section 3, we found that the estimator in (12) had smaller variance than m_{cl_1} but larger variance than m_{cl_2}. Two reasons that the estimator (12) failed to outperform m_{cl_2} might be: *i*) the estimator m_{cl_2} has already a small within section variability due to the antithetic effect of the two perpendicuclar lines and *ii*) the observed cell sections are not very irregular in shape and the variance reducing effect of weighted sampling cannot compensate for the effect that some extra within section variability is introduced when (12) is used in case of non-perfect segmentation.

If a rough voxel image of a sampled cell is available, it might instead be worthwhile to use a distribution of the orientation of the section plane L_2 that depends

on this rough voxel image. Implementation of weighted sampling in this connection might lead to a further reduction in estimator variance that will affect all the estimators presented in this paper.

5 Acknowledgements

This project has been supported by the Danish Council for Strategic Research and by Centre for Stochastic Geometry and Advanced Bioimaging, funded by a grant from THE VILLUM FOUNDATION.

References

[1] Abrahão, L.M., Nyengaard, J.R., Sasahara, T.H., Gomes, S.P., Oliveira Fda, R., Ladd, F.V., Ladd, A.A., Melo, M.P., Machado, M.R., Melo, S.R. and Ribeiro, A.A. (2009): Asymmetric post-natal development of superior cervical ganglion of paca (Agouti paca). *International Journal of Developmental Neuroscience* **27**, 37–45.

[2] Cochran, W.G. (1977): *Sampling Techniques*. John Wiley & Sons, New York.

[3] Cruz-Orive, L.M. (1976): Particle size-shape distributions: the generel spheroid problem. I. Matematical model. *Journal of Microscopy* **107**, 235–253.

[4] Cruz-Orive, L.M. (1978): Particle size-shape distributions: the generel spheroid problem. II. Stochastic model and particle guide. *Journal of Microscopy* **112**, 153–167.

[5] Cruz-Orive, L.M. (2008): Comparative precision of the pivotal estimators of particle size. *Image Analysis and Stereology* **27**, 17–22.

[6] Cruz-Orive, L.M. (2009): The pivotal tesselation. *Image Analysis and Stereology* **28**, 63–67.

[7] Cruz-Orive, L.M., Ramos-Herrera, M.L. and Artacho-Pérula, E. (2010): Stereology of isolated objects with the invariator. To appear in *Journal of Microscopy*.

[8] Gardi, J.E., Wulfsohn, D. and Nyengaard, J.R. (2007): A handheld support to facilitate stereological measurements and mapping of branching structures. *J. Microsc.* **227**, 127–139.

[9] Gardi, J.E., Nyengaard, J.R. and Gundersen, H.J.G. (2008a): The proportionator: unbiased stereological estimation using biased automatic image analysis and non-uniform probability proportional to size sampling. *Comput. Biol. Med.* **38**, 313–328.

[10] Gundersen, H.J.G. (1988): The nucleator. *Journal of Microscopy* **151**, 3–21.

[11] Jensen, E.B.V. (1998): *Local Stereology*. Singapore: World Scientific Publishing.

[12] Jensen, E.B.V. (2000): On the variance of local stereological volume estimators. *Image Analysis and Stereology* **19**, 15–18.

[13] Melo, S.R., Nyengaard, J.R., da Roza Oliveira, F., Ladd, F.V., Abrahão, L.M., Machado, M.R., Sasahara, T.H., de Melo, M.P. and Ribeiro, A.A. (2009): The developing left superior cervical ganglion of Pacas (Agouti paca). *The Anatomical Record (Hoboken)* **292**, 966–975.

[14] Nyengaard, J.R. and Gundersen, H.J.G. (1992): The isector: a simple and direct method for generating isotropic, uniform random sections from small specimens. *Journal of Microscopy* **165**, 427–431.

[15] Wicksell, S.D (1925): The corpuscle problem. A mathematical study of a biometric problem. *Biometrika* **17**, 84–89.

[16] Wicksell, S.D (1926): The corpuscle problem. Second memoir. Case of ellipsoidal corpuscles. *Biometrika* **18**, 152–172.

Appendix

In this Appendix, we will derive the results on bias and MSE of m_{semi} presented in Section 2.1.4 and Section 2.2 of the main text.

Let $A_\varepsilon = \{d(Y \cap L_2, \tilde{Y}_2) < \varepsilon\}$. Then, the semi-automatic nucleator is given by

$$m_{\text{semi}} = \mathbf{1}_{A_\varepsilon} m_{\text{aut}} + \mathbf{1}_{A_\varepsilon^c} m_{\text{cl}_2},$$

where $\mathbf{1}_A$ is the indicator function of A. Given Y and L_2, the distribution of m_{cl_2} depends only on the lines $L_1(\Theta)$ and $L_1(\Theta + \pi/2)$ where Θ is uniform on $[0, \pi/2)$. It follows that m_{cl_2} and A_ε are conditional independent given Y and L_2 and we find, using (3),

$$\begin{aligned} E\,(m_{\text{semi}}|Y, L_2) &= E\,(\mathbf{1}_{A_\varepsilon} m_{\text{aut}}|Y, L_2) + E\,(\mathbf{1}_{A_\varepsilon^c}|Y, L_2) E\,(m_{\text{cl}_2}|Y, L_2) \\ &= E\,(\mathbf{1}_{A_\varepsilon} m_{\text{aut}}|Y, L_2) + E\,(\mathbf{1}_{A_\varepsilon^c}|Y, L_2) m_{\text{int}} \\ &= E\,(\mathbf{1}_{A_\varepsilon} m_{\text{aut}} + \mathbf{1}_{A_\varepsilon^c} m_{\text{int}}|Y, L_2). \end{aligned}$$

Thereby,

$$E\,(m_{\text{semi}}) = E\,(\mathbf{1}_{A_\varepsilon} m_{\text{aut}} + \mathbf{1}_{A_\varepsilon^c} m_{\text{int}}).$$

It follows that

$$\begin{aligned} \text{bias}\,(m_{\text{semi}}) &= E\,(m_{\text{semi}}) - E\,(m_{\text{int}}) \\ &= E\,(\mathbf{1}_{A_\varepsilon} m_{\text{aut}} + \mathbf{1}_{A_\varepsilon^c} m_{\text{int}}) - E\,(\mathbf{1}_{A_\varepsilon} m_{\text{int}} + \mathbf{1}_{A_\varepsilon^c} m_{\text{int}}) \\ &= E\,(\mathbf{1}_{A_\varepsilon} (m_{\text{aut}} - m_{\text{int}})) \\ &= p\,E\,(m_{\text{aut}} - m_{\text{int}}|A_\varepsilon). \end{aligned}$$

Using the distance d_2, we have the following inequality between estimators

$$\mathbf{1}_{A_\varepsilon} (1 - \varepsilon)\, m_{\text{int}} < \mathbf{1}_{A_\varepsilon} m_{\text{aut}} < \mathbf{1}_{A_\varepsilon} (1 + \varepsilon)\, m_{\text{int}}. \tag{13}$$

Therefore,

$$\begin{aligned} E\left(m_{\text{semi}}\right) &< (1+\varepsilon)\, E\left(\mathbf{1}_{A_\varepsilon} m_{\text{int}}\right) + E\left(\mathbf{1}_{A_\varepsilon^c} m_{\text{int}}\right) \\ &= \varepsilon\, E\left(\mathbf{1}_{A_\varepsilon} m_{\text{int}}\right) + E\left(m_{\text{int}}\right) \\ &= \varepsilon\, p\, E\left(m_{\text{int}} | A_\varepsilon\right) + \mu. \end{aligned}$$

Likewise, $E\left(m_{\text{semi}}\right) > -\varepsilon\, p\, E\left(m_{\text{int}} | A_\varepsilon\right) + \mu$ and we get a bound for the relative bias

$$\frac{\left|E\left(m_{\text{semi}}\right) - \mu\right|}{\mu} \leq \varepsilon\, p\, \frac{E\left(m_{\text{int}} | A_\varepsilon\right)}{\mu}.$$

The results on the MSE of m_{semi} is derived as follows

$$\begin{aligned} \text{MSE}\left(m_{\text{semi}}\right) &= E\left(\mathbf{1}_{A_\varepsilon}[m_{\text{aut}} - \mu]^2 + \mathbf{1}_{A_\varepsilon^c}[m_{\text{cl}_2} - \mu]^2\right) \\ &= p\, \text{MSE}\left(m_{\text{aut}} | A_\varepsilon\right) + (1-p)\, \text{MSE}\left(m_{\text{cl}_2} | A_\varepsilon^c\right). \end{aligned}$$

Finally, we show that (7) is satisfied when m_{aut} is replaced by m_{int}, i.e.

$$\text{MSE}\left(m_{\text{int}} | A_\varepsilon\right) \leq \text{MSE}\left(m_{\text{cl}_2} | A_\varepsilon\right).$$

Utilizing that m_{cl_2} and A_ε are conditionally independent given Y and L_2, we find

$$E\left(m_{\text{cl}_2} | A_\varepsilon\right) = E\left(E\left(m_{\text{cl}_2} | A_\varepsilon, Y, L_2\right) | A_\varepsilon\right) = E\left(m_{\text{int}} | A_\varepsilon\right)$$

and

$$\begin{aligned} E\left(m_{\text{cl}_2}^2 | A_\varepsilon\right) &= \text{Var}(m_{\text{cl}_2} | A_\varepsilon) + \left(E\left(m_{\text{cl}_2} | A_\varepsilon\right)\right)^2 \\ &\geq \text{Var}(E\left(m_{\text{cl}_2} | A_\varepsilon, Y, L_2\right) | A_\varepsilon) + \left(E\left(m_{\text{int}} | A_\varepsilon\right)\right)^2 \\ &= \text{Var}(m_{\text{int}} | A_\varepsilon) + \left(E\left(m_{\text{int}} | A_\varepsilon\right)\right)^2 \\ &= E\left(m_{\text{int}}^2 | A_\varepsilon\right). \end{aligned}$$

Therefore,

$$\begin{aligned} \text{MSE}\left(m_{\text{cl}_2} | A_\varepsilon\right) &= E\left((m_{\text{cl}_2} - \mu)^2 | A_\varepsilon\right) \\ &= E\left(m_{\text{cl}_2}^2 | A_\varepsilon\right) - 2\,\mu\, E\left(m_{\text{cl}_2} | A_\varepsilon\right) + \mu^2 \\ &\geq E\left(m_{\text{int}}^2 | A_\varepsilon\right) - 2\,\mu\, E\left(m_{\text{int}} | A_\varepsilon\right) + \mu^2 \\ &= \text{MSE}\left(m_{\text{int}} | A_\varepsilon\right). \end{aligned}$$