THIELE CENTRE
FOR APPLIED MATHEMATICS IN NATURAL SCIENCE

Image-based empirical importance sampling: an efficient way of estimating intensities

Linda V. Hansen, Markus Kiderlen
and Eva B. Vedel Jensen

Research Report

No. 08 I July 2009

Image-based empirical importance sampling: an efficient way of estimating intensities

Linda V. Hansen, Markus Kiderlen and Eva B. Vedel Jensen

T.N. Thiele Centre for Applied Mathematics
Department of Mathematical Sciences
University of Aarhus, Denmark

Summary

Very recently, it has been suggested in the biomedical literature to combine computerized image analysis with non-uniform sampling in order to increase the efficiency of estimators of intensities of biological cell populations. We give this ingenious idea of empirical importance sampling a stochastic formulation, using point process theory and modern sampling theory. We develop statistical tools for assessing its efficiency and construct optimal model-based estimators of intensities. Examples of applications of empirical importance sampling in microscopy are provided.

Keywords: image analysis, importance sampling, probability proportional to size, proportionator, stereology.

1 Introduction

Importance sampling is a general statistical technique for estimating properties of a particular distribution, based on samples from a different distribution than the one of interest, cf. [1] and references therein. Depending on the application, the term may refer to the process of sampling from this alternative distribution, the process of inference, or both. By choosing the alternative distribution appropriately, importance sampling may result in a marked increase in estimator efficiency. The basic idea is to choose an alternative sampling distribution in such a way that most of the sampling is done in the part of the state space that contributes the most to the parameter of interest.

This type of sampling technique has recently been introduced in computerized analysis of microscopy images under the name of *the proportionator*, [7, 8], see also the early paper [4]. This technique addresses the essential problem in the biomedical sciences that cell populations often show pronounced inhomogeneity, being present only in structured layers or showing marked gradients. Observing such a cell population, using a systematic set of fields of view, will be highly inefficient because most fields will contain no or very few cells.

An example of application of the technique suggested in [7, 8] is shown in Figure 1. The aim is here to estimate the total number of green GFP-expressing neurons, see panel B. Under low magnification, the complete region (panel A) of interest is delineated, and by automatic image analysis every field of view inside this region is given a weight proportional to the amount of green colour observed under fluorescence illumination. A field of view is automatically sampled with a

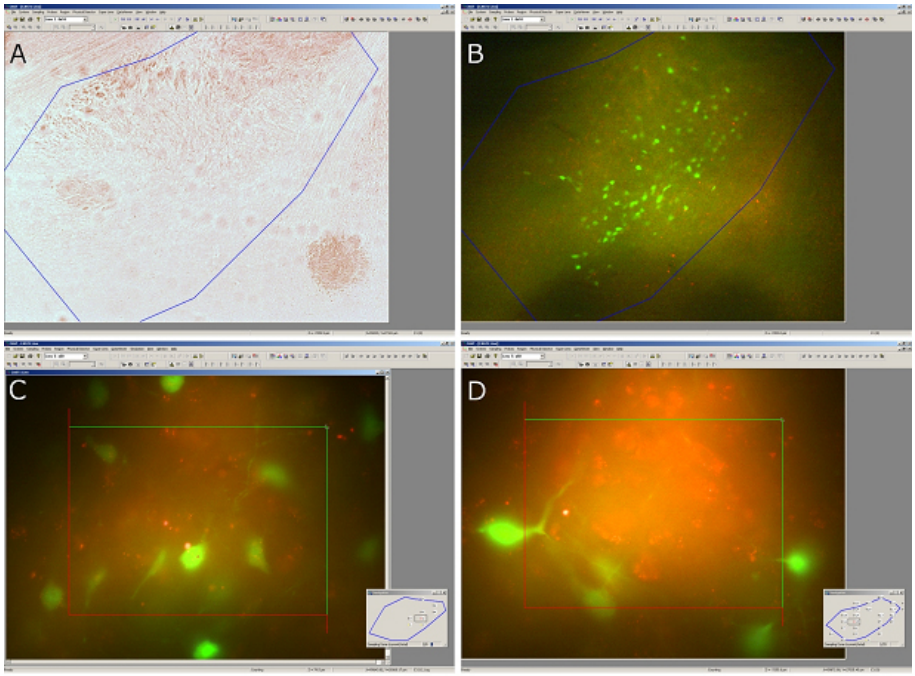

Figure 1: Estimating total number of GFP-expressing neurons in transgenic mice brain. Panels A and B show the same region of interest at 10× magnification in bright field (panel A) and during colour identification, using fluorescence light (panel B). Counting is performed by an expert using a 60× magnification, as shown in panels C and D. The small inserts indicate the positions of the sampled windows. For more details, see [8].

probability proportional to its weight and the number of neurons seen in the sampled field is determined under high magnification by an expert, see panels C and D. Further examples of computer-assisted spatial sampling may be found in [5, 6, 9].

If the weight assigned to a field of view is positively correlated with the number of cells seen in the field, the sampling is directed towards fields of view with high number of cells and an increase in efficiency is expected. In [8], increase in efficiency ranging from 8× to 25× was indeed observed in three biological examples without increasing the workload. This finding was supported by extensive simulation studies [7], demonstrating the beneficial effect of this type of empirical importance sampling.

In this paper, we give this ingeneous idea of empirical importance sampling a stochastic formulation. We derive statistical tools for assessing its efficiency and construct optimal model-based estimators of intensities. It will be demonstrated that even for homogeneous point processes there will be a gain in efficiency of intensity estimators by utilizing the inhomogeneity of the realized point pattern.

The paper is organized as follows. In Section 2, we summarize the concepts needed from sampling theory while Section 3 gives mild conditions under which empirical importance sampling will result in a gain in efficiency. In Section 4, the gain in efficiency is assessed for homogeneous point processes, including processes from the flexible class of Lévy driven Cox processes. Systematic weighted sampling is discussed in Section 5. Model-based inference for such data is developed and exemplified by the estimation of granule cell number in Section 6. Section 7 contains

concluding remarks. Technical details are deferred to three appendices.

2 The set-up

Throughout the paper, we will assume that a realization of a point process Φ is available for observation in a bounded subset X of $\mathbb{R}^2$ of area $A(X)$. In addition, a non-negative random field

$$Z = \{Z(u) : u \in X\},$$

associated with Φ, is observed in X. In the applications we have in mind, the points of Φ represent the positions of the objects in a digital or analog image. In the example of Figure 1, X corresponds to the region delineated at low magnification. Determination of the total number $N(\Phi \cap X)$ of points in X can only be performed at high magnification of the image and is impracticable. In contrast, the random field Z is readily available, e.g. from observation of colour proportions at low magnification by automatic image analysis.

Our aim is to predict $N(\Phi \cap X)$ or equivalently $N_A = N(\Phi \cap X)/A(X)$ from observation in a randomly placed window $Q_U = U + Q$, hitting X. Here, $Q \subset \mathbb{R}^2$ is bounded and is assumed to contain the origin O, cf. Figure 2. The position of the window is determined by the random vector $U \in \mathbb{R}^2$.

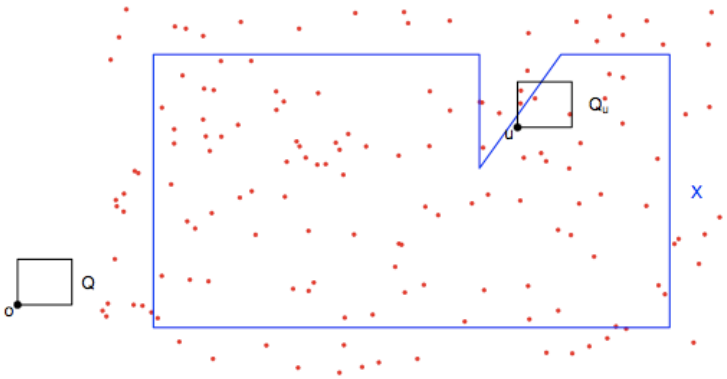

Figure 2: Illustration of the stochastic set-up.

In [7, 8], design-unbiased predictors $\widehat{N}_A$ of N_A are proposed, based on information in windows Q_U with different types of distribution of U. A predictor $\widehat{N}_A$ of N_A is said to be design-unbiased if

$$\mathbb{E}(\widehat{N}_A|\Phi, Z) = N_A. \tag{1}$$

The conditional mean value in (1) is calculated with respect to the conditional distribution of U given Φ, Z. In particular, a design-unbiased predictor of N_A is an (unconditionally) unbiased estimator of $\mathbb{E}(N_A)$. Likewise, a predictor $\hat{\sigma}^2_{\Phi,Z}$ of the conditional variance

$$\sigma^2_{\Phi,Z} = \mathbb{V}(\widehat{N}_A|\Phi, Z)$$

is said to be design-unbiased if

$$\mathbb{E}(\hat{\sigma}^2_{\Phi,Z}|\Phi, Z) = \sigma^2_{\Phi,Z}.$$

In most cases, it is of interest to make statements about the quality of a predictor across different realizations of Φ and Z. For this purpose, we will consider the prediction error $\mathbb{E}(\widehat{N}_A - N_A)^2$. A predictor $\widehat{N}_A^{(1)}$ is said to be more efficient than $\widehat{N}_A^{(2)}$ if $\widehat{N}_A^{(1)}$ has smaller prediction error than $\widehat{N}_A^{(2)}$. For a design-unbiased predictor, the prediction error can be expressed as

$$\mathbb{E}(\widehat{N}_A - N_A)^2 = \mathbb{V}(\widehat{N}_A) - \mathbb{V}(N_A). \tag{2}$$

If $\widehat{N}_A$ and $\hat{\sigma}^2_{\Phi,Z}$ are design-unbiased, then

$$\mathbb{E}(\widehat{N}_A - N_A)^2 = \mathbb{E}(\hat{\sigma}^2_{\Phi,Z})$$

and $\hat{\sigma}^2_{\Phi,Z}$ or an average of such estimators for a set of realizations of Φ and Z can be regarded as an unbiased estimator of the prediction error.

3 A simple condition for gain in efficiency

The weighted sampling suggested in [7, 8] will not always result in a gain in efficiency. In this section, we will give a simple condition on Φ and Z under which the predictor based on weighted sampling is more efficient than the one based on standard uniform random sampling. We first present the two different types of sampling to be considered.

In traditional sampling, the position u of the window is selected uniform randomly without any reference to the random field Z. In empirical importance sampling, the position u of the window is selected with a probability proportional to $Z(u)$. The two types of random windows are denoted uniform random (UR) and proportional random (PR) windows, respectively.

Below, we need the following versions of the principal kinematic formula, cf. e.g. [2, (4.22)],

$$\int_{\mathbb{R}^2} A(X \cap Q_u)\,\mathrm{d}u = A(X)A(Q), \tag{3}$$

$$\int_{\mathbb{R}^2} N(\Phi \cap X \cap Q_u)\,\mathrm{d}u = N(\Phi \cap X)A(Q). \tag{4}$$

A UR window is distributed as $Q_U = U + Q$ where U is independent of Φ, Z and uniform in

$$\{u \in \mathbb{R}^2 | X \cap Q_u \neq \emptyset\} = X \oplus \check{Q}.$$

Here, $\check{Q} = \{v \in \mathbb{R}^2 | -v \in Q\}$ and $\oplus$ denotes vector (Minkowski) addition. Let

$$C = \frac{A(X \oplus \check{Q})}{A(X)\,A(Q)}.$$

It then follows from the principal kinematic formula that

$$\widehat{N}_A^{UR} = C\,N(\Phi \cap X \cap Q_U) \tag{5}$$

is a design-unbiased predictor of N_A. In fact, using (4),

$$\mathbb{E}[\widehat{N}_A^{UR}|\Phi, Z] = \mathbb{E}[\widehat{N}_A^{UR}|\Phi] = \frac{N(\Phi \cap X)}{A(X)} = N_A.$$

We now consider PR windows Q_U. In order to avoid problems with edge-effects, we assume that $Z(u)$ is defined for all u with $X \cap Q_u \neq \emptyset$. A PR window is distributed as Q_U where U is a stochastic vector with density p given by

$$p(u|Z) = \begin{cases} \dfrac{Z(u)}{\displaystyle\int_{X \oplus \check{Q}} Z(u)\, \mathrm{d}u} & \text{if } X \cap Q_u \neq \emptyset, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Here and in the following, we assume that $\int_{X \oplus \check{Q}} Z(u)\, \mathrm{d}u > 0$ and that

$$Z(u) = 0 \Rightarrow N(\Phi \cap X \cap Q_u) = 0.$$

The predictor suggested in [7, 8] takes the form

$$\widehat{N}_A^{PR} = \frac{\int_{X \oplus \check{Q}} Z(u)\, \mathrm{d}u}{A(X)A(Q)} \frac{N(\Phi \cap X \cap Q_U)}{Z(U)} \tag{7}$$

and is design-unbiased for N_A, as can easily be seen, using the principal kinematic formula (4).

The predictor $\widehat{N}_A^{PR}$ is not in general more efficient than $\widehat{N}_A^{UR}$. Using that $\widehat{N}_A^{UR}$ and $\widehat{N}_A^{PR}$ are both design-unbiased and that (2) holds for any design-unbiased predictor, it is seen that $\widehat{N}_A^{PR}$ is more efficient than $\widehat{N}_A^{UR}$ if and only if

$$\mathbb{V}(\widehat{N}_A^{PR}) \leq \mathbb{V}(\widehat{N}_A^{UR})$$

or, equivalently,

$$\mathrm{Cov}\left(\frac{N(\Phi \cap X \cap Q_U)^2}{Z(U)}, Z(U)\right) \geq 0 \tag{8}$$

for U uniform in $X \oplus \check{Q}$.

A simple pilot study can reveal whether (8) is likely to be satisfied.

4 Efficiency for homogeneous point processes

Since PR windows use the inhomogeneity of the *realized* point pattern, a gain in efficiency may even be obtained for homogeneous point processes.

Throughout this section, Φ will be a homogeneous point process with intensity λ. Then, $\mathbb{E}N(\Phi \cap X) = \lambda A(X)$ and its pair correlation function is translation invariant. For homogeneous Φ, we have, cf. [12, (4.5.3)],

$$\mathbb{V}(N(\Phi \cap X)) = \lambda^2 \int_{\mathbb{R}^2} \gamma_X(y) g(y)\, \mathrm{d}y + \lambda A(X) - \lambda^2 A^2(X), \tag{9}$$

where g is the pair correlation function of Φ and

$$\gamma_X(y) = A(X \cap (X + y))$$

is the set covariogram of X. In the following, we will, in particular, consider Cox processes. Recall that Φ is a Cox process with driving field

$$\Lambda = \{\Lambda(u) : u \in \mathbb{R}^2\}$$

if, conditionally on Λ, Φ is a Poisson point process with intensity function Λ.

It can be shown, cf. Appendix A, that for a homogeneous process Φ, $\widehat{N}_A^{PR}$ will be more efficient than $\widehat{N}_A^{UR}$ if the counts satisfy the following proportional regression model

$$\mathbb{E}(N(\Phi \cap X \cap Q_U)|U, Z) = aZ(U), \tag{10}$$

$$\mathbb{V}(N(\Phi \cap X \cap Q_U)|U, Z) = bZ(U)^p, \quad p \geq 1. \tag{11}$$

Note that in the particular case where Φ is a Cox process with driving field $\Lambda = \{\Lambda(u) : u \in \mathbb{R}^2\}$ and with $Z(u)$ equal to the cumulated intensity in Q_u,

$$Z(u) = \int_{X \cap Q_u} \Lambda(v) \mathrm{d}v,$$

the relations (10) and (11) are satisfied with $a = b = p = 1$.

In order to assess the relative efficiency of $\widehat{N}_A^{PR}$ to $\widehat{N}_A^{UR}$ for specific point process models, we need to express the variances in terms of second-order properties of the point process. Using (3) and (9), we get

$$\begin{aligned}
\mathbb{V}(\widehat{N}_A^{UR}) &= C^2\Big(\mathbb{V}\big(\mathbb{E}[N(\Phi \cap X \cap Q_U)|U]\big) + \mathbb{E}\big(\mathbb{V}[N(\Phi \cap X \cap Q_U)|U]\big)\Big) \\
&= C^2\lambda^2 \int_{\mathbb{R}^2} \mathbb{E}\gamma_{X \cap Q_U}(y) g(y) \mathrm{d}y + \lambda C - \lambda^2 \\
&= \frac{A(X \oplus \check{Q})}{A(Q)^2} \frac{\lambda^2}{A(X)^2} \int_{\mathbb{R}^2} \gamma_X(y)\gamma_Q(y) g(y) \mathrm{d}y + \lambda C - \lambda^2.
\end{aligned} \tag{12}$$

The last equality sign follows from (3) with X and Q replaced by $X \cap (X + y)$ and $Q \cap (Q + y)$, respectively. Since, cf. (9),

$$\mathbb{V}(N_A) = \frac{\lambda^2}{A(X)^2} \int_{\mathbb{R}^2} \gamma_X(y) g(y) \mathrm{d}y + \frac{\lambda}{A(X)} - \lambda^2, \tag{13}$$

the prediction error of $\widehat{N}_A^{UR}$ is easily obtained

$$\begin{aligned}
\mathbb{E}(\widehat{N}_A^{UR} - N_A)^2 = \frac{\lambda^2}{A(X)^2} \int_{\mathbb{R}^2} \gamma_X(y) g(y) \left[\frac{A(X \oplus \check{Q})}{A(Q)^2} \gamma_Q(y) - 1\right] \mathrm{d}y \\
+ \frac{\lambda}{A(X)} \left[\frac{A(X \oplus \check{Q})}{A(Q)} - 1\right].
\end{aligned} \tag{14}$$

The variance of $\widehat{N}_A^{PR}$ depends on the random field Z and its interplay with Φ. If Φ is a Cox process with driving field Λ such that Z is the cumulated intensity in Q_u,

$$Z(u) = \int_{X \cap Q_u} \Lambda(v) \mathrm{d}v,$$

then, cf. Appendix B,

$$\mathbb{V}(\widehat{N}_A^{PR}) = \frac{\lambda^2}{A(X)^2} \int_{\mathbb{R}^2} \gamma_X(y) g(y) \mathrm{d}y + \lambda C - \lambda^2. \tag{15}$$

Combining with (13), we find

$$\mathbb{E}(\widehat{N}_A^{PR} - N_A)^2 = \frac{\lambda}{A(X)} \left[\frac{A(X \oplus \check{Q})}{A(Q)} - 1 \right]. \tag{16}$$

Note that in this case the prediction error of $\widehat{N}_A^{PR}$ only depends on λ.

We now compare the efficiency of $\widehat{N}_A^{UR}$ and $\widehat{N}_A^{PR}$ within the flexible class of homogeneous Lévy driven Cox processes ([10]). Such Cox processes have driving fields of the form

$$\Lambda(y) = \int_{\mathbb{R}^2} k(y - n) L(\mathrm{d}n), \quad y \in \mathbb{R}^2,$$

where L is a so-called non-negative homogeneous Lévy basis on $\mathbb{R}^2$ (an independently scattered, infinitely divisible random measure) and k is a density.

It can be shown that Lévy driven Cox processes are, under mild regularity conditions, shot noise Cox processes with additional random noise ([10]). One of the advantages of this class of Cox processes is that the intensity and pair correlation function can be calculated explicitly. By [10, Corollary 2], one can associate a random variable L', the so-called spot-variable, to the Cox process such that

$$\lambda = c\, \mathbb{E}L',$$
$$g(y) = 1 + \frac{\mathbb{V}(L')}{(\mathbb{E}L')^2} \frac{I_k(y)}{c},$$

where $I_k(y) = \int_{\mathbb{R}^2} k(y + n) k(n)\, \mathrm{d}n$ and $c > 0$ is an explicitly known constant. For a Gaussian kernel

$$k(y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\|y\|^2}{2\sigma^2} \right), \quad y \in \mathbb{R}^2, \tag{17}$$

we get

$$g(y) = 1 + \frac{\mathbb{V}(L')}{\mathbb{E}L'} \frac{\exp\left(-\frac{\|y\|^2}{4\sigma^2} \right)}{4\pi\sigma^2\lambda}, \quad y \in \mathbb{R}^2. \tag{18}$$

In Figure 3, simulations of Lévy driven Cox processes are shown for a Poisson Lévy basis ($L' \sim \mathbb{E}L' Po(1)$), a compound gamma Lévy basis and a compound inverse Gaussian Lévy basis, respectively, in a $[0, 10] \times [0, 10]$ window with $c = 3$, $\mathbb{E}L' = 2$ and $\sigma = 0.4$. The variance of L' is equal to 4, 8 and 12, respectively. Notice the increasing heterogeneity in the realized point patterns.

We have calculated the prediction error of $\widehat{N}_A^{UR}$ and $\widehat{N}_A^{PR}$, using (14) and (16), for the case where Q is a square with unit side length while X is a square with side length

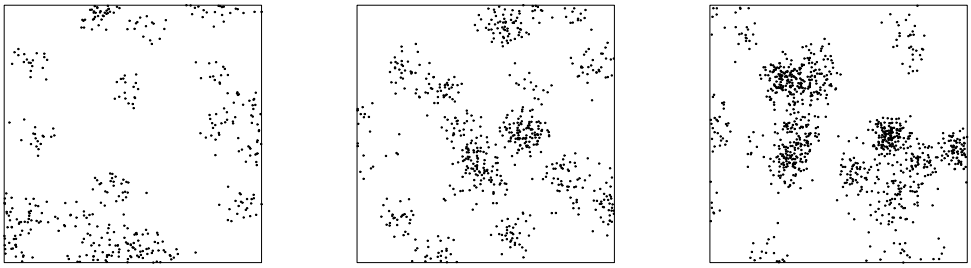

Figure 3: Examples of realizations of homogeneous Lévy driven Cox processes with Poisson (left), compound gamma (middle) and compound inverse Gaussian (right) Lévy bases. For details, see the text.

$D \geq 1$. Furthermore, we use the Gaussian kernel (17). For some technical details, see Appendix C. In Figure 4, we have plotted the ratio $\mathbb{E}(\widehat{N}_A^{UR} - N_A)^2 / \mathbb{E}(\widehat{N}_A^{PR} - N_A)^2$ as a function of $D \in [1, 10]$ for Lévy driven Cox processes with Poisson, compound gamma and compound inverse Gaussian Lévy bases and $\lambda = 1, 10, 25$. The value of σ is 0.4 and $\mathbb{V}(L')/\mathbb{E}(L')$ is 2, 4 and 6, respectively, for the three types of Lévy bases. The gain in efficiency is largest for the most pronounced clustered point patterns. The efficiency also increases with increasing intensity of the point process.

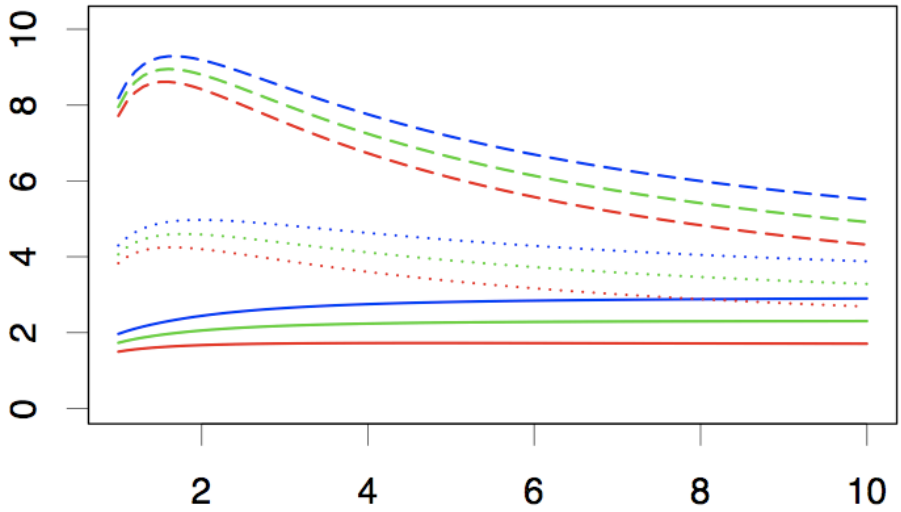

Figure 4: The ratio $\mathbb{E}(\widehat{N}_A^{UR} - N_A)^2 / \mathbb{E}(\widehat{N}_A^{PR} - N_A)^2$ as a function of the side length D of X for the Poisson (red), compound gamma (green) and compound inverse Gaussian (blue) cases, respectively. The intensities are $\lambda = 1$ (full-drawn), $\lambda = 10$ ($\cdot\,\cdot\,\cdot$) and $\lambda = 25$ (– – –), respectively.

5 Replication

Replicated generation of PR windows in a systematic set-up has in [7, 8] been implemented for the analysis of microscopy sections as the one shown in Figure 1. The sampling is denoted systematic proportional random sampling (SPRS). Below,

we describe SPRS sampling and discuss the statistical analysis of data sampled in this fashion.

The window $Q = [0, l_1) \times [0, l_2)$ is a rectangle. The region X is covered by rectangles

$$G = \bigcup_{(s_1,s_2)\in\mathcal{S}} \{Q + (s_1 l_1, s_2 l_2)\},$$

where $\mathcal{S} \subset \mathbb{Z}^2$. The windows in G are ordered lexicographically $Q_1, \ldots, Q_N$, where N is the number of windows in G. The windows $Q_1, \ldots, Q_N$ are translations of Q by $u_1, \ldots u_N$, say, see Figure 5. The weight $Z(u_i)$ is assigned to Q_i, $i = 1, \ldots, N$.

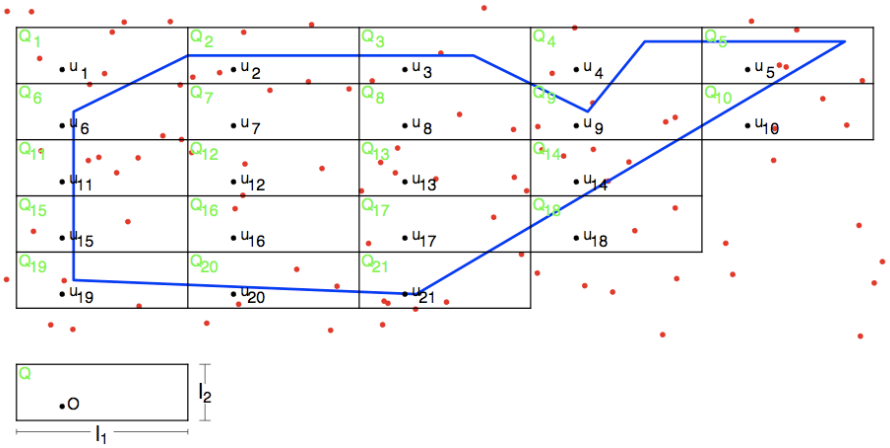

Figure 5: The set-up for systematic proportional random sampling (SPRS). The point pattern Φ (red dots) is available for observation in X (delineated by blue lines). The set X is covered by a family of lexicographically ordered non-overlapping rectangles. These rectangles are translations of the given rectangle Q.

Sampling in a systematic $Z-$weighted fashion is performed as follows. Let $W_j = \sum_{i=1}^{j} Z(u_i)$ denote the cumulated weight, with the convention that $W_0 = 0$, and let $S = \sum_{i=1}^{N} Z(u_i)$. A sample of $n \in \{1, \ldots, N\}$ windows is selected by choosing V_1 uniformly in $[0, \frac{S}{n}]$, independently of Φ and Z, and let $V_j = V_1 + (j-1)\frac{S}{n}$ for $j = 2, 3, \ldots, n$. The sampled windows are those with index in

$$J = \bigcup_{i:V_i\in[W_{j_i-1},W_{j_i}]} \{j_i\},$$

see also Figure 6. A window may be sampled more than once. Notice that ordinary

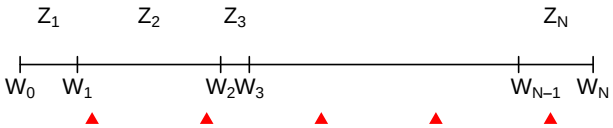

Figure 6: Illustration of SPRS sampling. The 1st and 3rd window are not sampled, the Nth window is sampled exactly once while the 2nd window is sampled twice. We use the abbreviation $Z_j = Z(u_j)$.

systematic uniform random sampling (SURS) is a special case of SPRS where Z is a constant field.

In [8], it is suggested to use the following predictor of N_A

$$\widehat{N}_A^{SPRS} = \frac{S}{n\,A(X)} \sum_{j=1}^{N} \#\{i \,|\, V_i \in [W_{j-1}, W_j]\} \frac{N(\Phi \cap X \cap Q_j)}{Z(u_j)}.$$

This predictor is design-unbiased. In fact, since $V_i \sim \mathrm{U}\big([(i-1)\frac{S}{n}, i\frac{S}{n}]\big)$,

$$\begin{aligned}
\mathbb{E}\big[\#\{i \,|\, V_i \in [W_{j-1}, W_j]\} | \Phi, Z\big] &= \sum_{i=1}^{n} \mathbb{P}\big(V_i \in [W_{j-1}, W_j] | \Phi, Z\big) \\
&= \frac{n}{S} \sum_{i=1}^{n} \int_{(i-1)\frac{S}{n}}^{i\frac{S}{n}} \mathbf{1}_{[W_{j-1}, W_j]}(v) \mathrm{d}v \\
&= \frac{n}{S}(W_j - W_{j-1}) \\
&= \frac{n}{S} Z(u_j),
\end{aligned}$$

which gives us

$$\mathbb{E}\left[\widehat{N}_A^{SPRS} | \Phi, Z\right] = \frac{1}{A(X)} \sum_{j=1}^{N} N(\Phi \cap X \cap Q_j) = \frac{N(\Phi \cap X)}{A(X)}.$$

6 Model-based inference

In this section, we will discuss how to use the information available in the data

$$(Z(u_j), N(\Phi \cap X \cap Q_j)), \quad j \in J,$$

to construct optimal predictors of N_A. For this purpose, we will take a model-based approach, cf. e.g. [13, Section 2.7 and Chapter 8].

Under a proportional regression model

$$\mathbb{E}(N(\Phi \cap X \cap Q_j) | J, Z) = a Z(u_j), \tag{19}$$
$$\mathbb{V}(N(\Phi \cap X \cap Q_j) | J, Z) = b Z(u_j)^p, \tag{20}$$

$j \in J$, we can construct a model-unbiased predictor with minimal model-variance among predictors of the form

$$\widehat{N}_A = \frac{S}{A(X)} \sum_{j \in J} \alpha_j(J, Z) \frac{N(\Phi \cap X \cap Q_j)}{Z(u_j)}, \tag{21}$$

where $\alpha_j(J, Z) \geq 0$ and $\sum_{j \in J} \alpha_j(J, Z) = 1$, provided that the counts

$$N(\Phi \cap X \cap Q_j), \ j \in J,$$

can be regarded as uncorrelated, given J and Z.

First, notice that for a predictor of the form (21)

$$\mathbb{E}(\widehat{N}_A | J) = \frac{a}{A(X)} \mathbb{E}(S).$$

Furthermore, because of (19),

$$\begin{aligned}\mathbb{E}(N_A) &= \frac{1}{A(X)}\mathbb{E}N(\Phi \cap X) \\ &= \frac{1}{A(X)}\sum_{j=1}^{N}\mathbb{E}N(\Phi \cap X \cap Q_j) \\ &= \frac{a}{A(X)}\mathbb{E}(S).\end{aligned}$$

It follows that $\mathbb{E}(\widehat{N}_A|J) = \mathbb{E}(N_A)$, so any predictor of the form (21) is indeed model-unbiased. Using the assumption of uncorrelatedness, the model-variance becomes

$$\begin{aligned}\mathbb{V}(\widehat{N}_A|J) &= \mathbb{E}(\mathbb{V}(\widehat{N}_A|J,Z)) + \mathbb{V}(\mathbb{E}(\widehat{N}_A|J,Z)) \\ &= \frac{b}{A(X)^2}\mathbb{E}(S^2\sum_{j\in J}\alpha_j(J,Z)^2 Z(u_j)^{p-2}) + \frac{a^2}{A(X)^2}\mathbb{V}(S).\end{aligned}$$

Using Cauchy-Schwartz' inequality it follows that $\mathbb{V}(\widehat{N}_A|J)$ is minimized for

$$\alpha_j(J,Z) = \frac{Z(u_j)^{2-p}}{\sum_{j\in J} Z(u_j)^{2-p}}.$$

The minimal variance becomes

$$\mathbb{V}(\widehat{N}_A|J) = \frac{b}{A(X)^2}\mathbb{E}\left(\frac{S^2}{\sum_{j\in J} Z(u_j)^{2-p}}\right) + \frac{a^2}{A(X)^2}\mathbb{V}(S).$$

If $p = 2$ and the sampling is performed such that a window is only sampled once, then the minimal model-variance predictor of the form (21) coincides with $\widehat{N}_A^{SPRS}$.

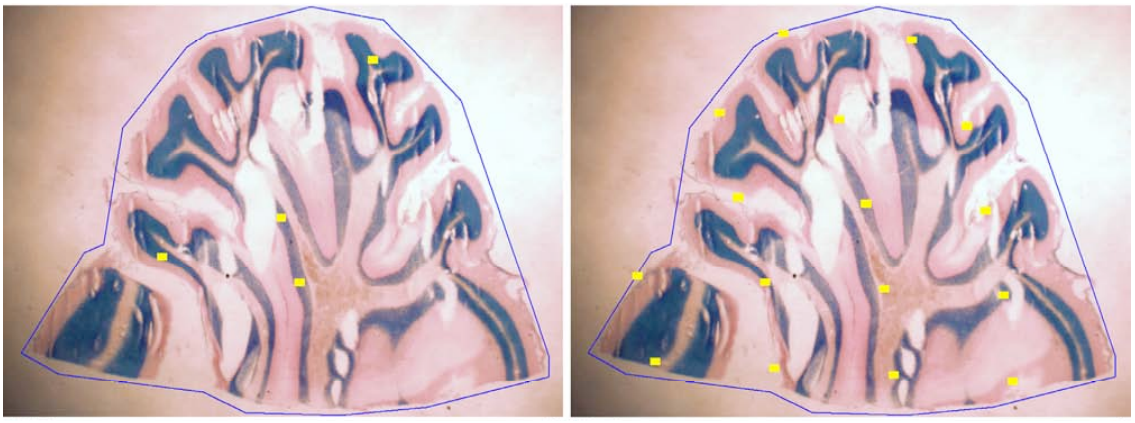

Figure 7: Left: A systematic proportional random sample (SPRS) of the granule cell layer (blue). The selected sampling windows are shown as small yellow rectangles. Right: A systematic uniform random sample (SURS).

We have applied this type of model-based inference on data collected by SPRS sampling for the estimation of the intensity of granule cells in rat cerebellum. SPRS

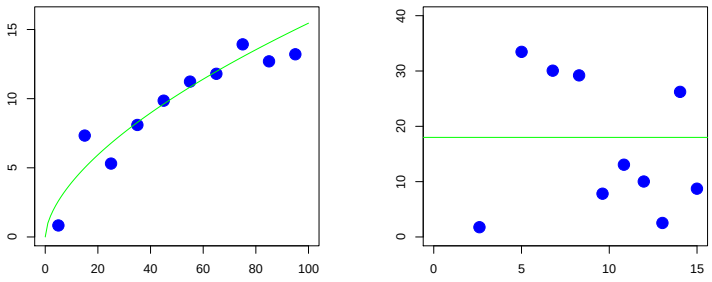

Figure 8: Left: Mean counts as a function of weights together with the fitted relationship. Right: Empirical variances of counts as a function of transformed weights $\tilde{Z}$. For details, see text.

sampling of the granule cell layer (blue) is shown in Figure 7 (left) where almost all selected sampling windows (small yellow rectangles) lie in the granule layer. For comparison, Figure 7 (right) shows SURS sampling where many of the sampling windows fall outside the blue region of interest. For more details, see [8].

A plot of the observed counts versus weights shows that a proportional regression model is not appropriate. Mean counts are shown in Figure 8 as a function of grouped weights. A relationship of the type

$$\mathbb{E}(N(\Phi \cap X \cap Q_j)|J, Z) = a_1 \, Z(u_j)^{a_2}$$

with $a_2 \neq 1$ is more appropriate. The fitted curve in Figure 8 (left) has parameters $a_1 = 1$ and $a_2 = 22/37$. We therefore transform the weights

$$\tilde{Z}(u_j) = Z(u_j)^{a_2},$$

such that the proportional regression is fulfilled for the transformed weights. Figure 8 (right) shows the empirical variances of the counts as a function of the transformed weigths. A relationship of the type (20) with $p = 0$ seems to be appropriate. Under this type of model, the optimal predictor takes the form

$$\widehat{N}_A^{\text{opt}} = \frac{\sum_{i=1}^N \tilde{Z}(u_i)}{A(X)} \frac{\sum_{j\in J} \tilde{Z}(u_j) N(\Phi \cap X \cap Q_j)}{\sum_{j\in J} \tilde{Z}(u_j)^2},$$

while the predictor suggested in [8] based on the transformed weights becomes (assuming no window is sampled more than once)

$$\widehat{N}_A^{SPRS} = \frac{\sum_{i=1}^N \tilde{Z}(u_i)}{nA(X)} \sum_{j\in J} N(\Phi \cap X \cap Q_j)/\tilde{Z}(u_j).$$

The ratio between the conditional variances is for $p = 0$

$$\frac{\mathbb{V}(\widehat{N}_A^{SPRS}|J, Z)}{\mathbb{V}(\widehat{N}_A^{\text{opt}}|J, Z)} = \frac{1}{n} \sum_{j\in J} \tilde{Z}(u_j)^2 \cdot \frac{1}{n} \sum_{j\in J} \tilde{Z}(u_j)^{-2}.$$

In the concrete example, this ratio takes the value 2.90 and the model-based approach represents an increase in efficiency of a factor 3 in terms of the conditional variance.

7 Concluding remarks

In this paper, we have studied how computerized image analysis can be combined with non-uniform sampling in order to increase the efficiency of estimators of intensities of biological cell populations. We have provided conditions under which the proposed non-uniform sampling results in a gain in efficiency and constructed optimal model-based estimators of intensities.

We believe that the principle of empirical importance sampling has a much wider range of applications than in microscopy. It is likely to be useful in other areas of spatial sampling where point patterns show realized inhomogeneity, e.g. in precision farming and satellite image analysis. It might also be the solution to the problem of low resolution in modern MR and PET scanners. An initial complete scan at low resolution may be used to direct the sampling towards the region of interest which is subsequently scanned at a high resolution.

8 Acknowledgements

This work has been supported by the Danish Council for Strategic Research.

References

[1] Asmussen, S. and Glynn P.W. (2007): *Stochastic Simulation*. Springer Science+Business Media, LLC, New York.

[2] Baddeley, A. and Jensen, E.B.V. (2005): *Stereology for Statisticians*. CRC Press, Boca Raton.

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

[4] Dorph-Petersen, K.A., Gundersen, H.J.G. and Jensen, E.B.V. (2000): Non-uniform systematic sampling in stereology. *J. Microsc.* **200**, 148-157.

[5] Gardi, J.E., Nyengaard, J.R. and Gundersen, H.J.G. (2006): Using biased image analysis for improving unbiased stereological number estimation - a pilot study of the smooth fractionator. *J. Microsc.* **222**, 242-250.

[6] 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.

[7] 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.

[8] Gardi, J.E., Nyengaard, J.R. and Gundersen, H.J.G. (2008b): Automatic sampling for unbiased and efficient stereological estimation using the proportionator in biological studies. *J. Microsc.* **230**, 108-120.

[9] Gundersen, H.J.G. (2002): The smooth fractionator. *J. Microsc.* **207**, 191-210.

[10] Hellmund, G., Prokešová, M. and Jensen, E.B.V. (2008): Lévy based Cox point processes. *Adv. Appl. Prob.* **40**, 603-629.

[11] Jónsdóttir, K.Y., Schmiegel, J. and Jensen, E.B.V. (2008): Lévy based growth models. *Bernoulli* **14**, 62-90.

[12] Stoyan, D., Kendall, W.S. and Mecke, J. (1995): *Stochastic Geometry and Its Applications*. Akademie-Verlag, Berlin. Second edition.

[13] Thompson, S.K. (1992): *Sampling*. John Wiley & Sons, New York.

Appendix A

We suppose that the proportional regression model given by (10) and (11) is satisfied. Since $\widehat{N}_A^{UR}$ and $\widehat{N}_A^{PR}$ are both design-unbiased, it suffices to show that

$$\mathbb{V}(\widehat{N}_A^{PR}) \leq \mathbb{V}(\widehat{N}_A^{UR}). \tag{22}$$

Throughout this appendix, U will denote a uniform random vector on $X \oplus \check{Q}$ and W a random vector with density (6). We show (22) by showing the following two relations

$$\mathbb{V}(\mathbb{E}(\widehat{N}_A^{PR}|W,Z)) \leq \mathbb{V}(\mathbb{E}(\widehat{N}_A^{UR}|U,Z)) \tag{23}$$

$$\mathbb{E}(\mathbb{V}(\widehat{N}_A^{PR}|W,Z)) \leq \mathbb{E}(\mathbb{V}(\widehat{N}_A^{UR}|U,Z)). \tag{24}$$

First we show (23). Since $\widehat{N}_A^{UR}$ and $\widehat{N}_A^{PR}$ are both design-unbiased, it suffices to show that

$$\mathbb{E}(\mathbb{E}^2(\widehat{N}_A^{PR}|W,Z)) \leq \mathbb{E}(\mathbb{E}^2(\widehat{N}_A^{UR}|U,Z)).$$

We find

$$\mathbb{E}(\widehat{N}_A^{PR}|W,Z) = Ca \int_{X\oplus\check{Q}} Z(u) \frac{\mathrm{d}u}{A(X\oplus\check{Q})} = Ca\,\mathbb{E}(Z(U)|Z),$$

and

$$\mathbb{E}(\widehat{N}_A^{UR}|U,Z) = Ca\,Z(U).$$

Since

$$\mathbb{E}(\mathbb{E}^2(Z(U)|Z)) \leq \mathbb{E}(\mathbb{E}(Z(U)^2|Z))$$

(23) follows.

In order to show (24), we use (11) and find

$$\mathbb{V}(\widehat{N}_A^{UR}|U,Z) = C^2 b Z(U)^p,$$

$$\mathbb{V}(\widehat{N}_A^{PR}|W,Z) = \left(\frac{\int_{X\oplus\check{Q}} Z(u)\mathrm{d}u}{A(X)A(Q)}\right)^2 b Z(W)^{p-2}.$$

Using that

$$\mathbb{E}(Y^p) \geq \mathbb{E}(Y)\mathbb{E}(Y^{p-1}), \quad p \geq 1, \tag{25}$$

we finally get

$$\begin{aligned}\mathbb{E}(\mathbb{V}(\widehat{N}_A^{PR}|W,Z)) &= \mathbb{E}\left(\left(\frac{\int_{X\oplus\check{Q}} Z(u)\mathrm{d}u}{A(X)A(Q)}\right)^2 b\mathbb{E}(Z(W)^{p-2}|Z)\right)\\ &= C^2 b\mathbb{E}\left(\left(\int_{X\oplus\check{Q}} Z(u)\frac{\mathrm{d}u}{A(X\oplus\check{Q})}\right)^2 \int_{X\oplus\check{Q}} \frac{Z(u)^{p-1}}{\int_{X\oplus\check{Q}} Z(u)\mathrm{d}u}\mathrm{d}u\right)\\ &= C^2 b\mathbb{E}\left(\mathbb{E}(Z(U)|Z)\mathbb{E}(Z(U)^{p-1}|Z)\right)\\ &\leq C^2 b\mathbb{E}\left(\mathbb{E}(Z(U)^p|Z)\right) = \mathbb{E}(\mathbb{V}(\widehat{N}_A^{UR}|U,Z)).\end{aligned}$$

The relation (25) can be shown as follows. For $p, q \geq 0$, we have

$$\begin{aligned}\mathbb{E}(Y^q)\mathbb{E}(Y^p) &= \left((\mathbb{E}(Y^q))^{\frac{1}{q}}\right)^q \left((\mathbb{E}(Y^p))^{\frac{1}{p}}\right)^p\\ &\leq (\mathbb{E}(Y^{p+q}))^{\frac{q}{p+q}}(\mathbb{E}(Y^{p+q}))^{\frac{p}{p+q}} = \mathbb{E}(Y^{p+q}),\end{aligned}$$

using that $p \mapsto (\mathbb{E}(Y)^p)^{1/p}$ is increasing for $p \geq 0$, due to Jensen's inequality. Replacing q and p by 1 and $p-1$, respectively, we have shown (25).

Appendix B

In this appendix, we show (15). Note that, given U and Λ, $N(\Phi \cap X \cap Q_U)$ is Poisson distributed with mean

$$Z(U) = \int_{X\cap Q_U} \Lambda(v)\mathrm{d}v.$$

It follows that

$$\begin{aligned}\mathbb{E}(\widehat{N}_A^{PR}|U,\Lambda) &= \frac{\int_{X\oplus\check{Q}} Z(u)\,\mathrm{d}u}{A(X)A(Q)}\\ &= \int_X \Lambda(v)\,\frac{\mathrm{d}v}{A(X)}\end{aligned}$$

and

$$\mathbb{V}(\widehat{N}_A^{PR}|U,\Lambda) = \left(\int_X \Lambda(v)\frac{\mathrm{d}v}{A(X)}\right)^2 \frac{1}{Z(U)}.$$

Therefore,

$$\mathbb{E}(\mathbb{V}(\widehat{N}_A^{PR}|U,\Lambda)) = \frac{A(X\oplus\check{Q})}{A(X)^2A(Q)}\mathbb{E}\left(\int_X \Lambda(v)\,\mathrm{d}v\right) = C\lambda. \tag{26}$$

Furthermore,

$$\begin{aligned}\mathbb{V}(\mathbb{E}(\widehat{N}_A^{PR}|U,\Lambda)) &= \mathbb{E}\left(\frac{1}{A(X)^2}\int_X\int_X \Lambda(v_1)\Lambda(v_2)\,\mathrm{d}v_1\,\mathrm{d}v_2\right) - \lambda^2\\ &= \frac{\lambda^2}{A(X)^2}\int_X\int_{X-v} g(u)\,\mathrm{d}u\,\mathrm{d}v - \lambda^2\\ &= \frac{\lambda^2}{A(X)^2}\int_{\mathbb{R}^2} \gamma_X(u)g(u)\,\mathrm{d}u - \lambda^2.\end{aligned} \tag{27}$$

Combining (26) and (27), we finally get (15).

Appendix C

Throughout this appendix, $D \geq 1$. For $Q = [0,1]^2$ and $X = [0,D]^2$, we have

$$A(Q) = 1,\ A(X) = D^2,\ A(X \oplus \check{Q}) = (D+1)^2.$$

Furthermore,

$$\gamma_X(y_1,y_2) = \begin{cases} (D-|y_1|)(D-|y_2|) & \text{if } (y_1,y_2) \in [-D,D]^2 \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\gamma_Q(y_1,y_2) = \begin{cases} (1-|y_1|)(1-|y_2|) & \text{if } (y_1,y_2) \in [-1,1]^2 \\ 0 & \text{otherwise.} \end{cases}$$

For the calculation of the prediction error of $\widehat{N}_A^{UR}$ we use that

$$\int_{\mathbb{R}^2} \gamma_X(y)\gamma_Q(y)\,\mathrm{d}y = \left(\int_{-1}^{1}(D-|y_1|)(1-|y_1|)\,\mathrm{d}y_1\right)^2 = \left(D-\frac{1}{3}\right)^2,$$

$$\begin{aligned}\int_{\mathbb{R}^2} \gamma_X(y)\gamma_Q(y)e^{-\frac{\|y\|^2}{4\sigma^2}}\,\mathrm{d}y &= \left(\int_{-1}^{1}(D-|y_1|)(1-|y_1|)e^{-\frac{y_1^2}{4\sigma^2}}\,\mathrm{d}y_1\right)^2 \\ &= 4\Bigg(D\sqrt{\pi\sigma^2}\left[2\Phi\left(\frac{1}{\sqrt{2\sigma^2}}\right)-1\right] \\ &\qquad + 2\sigma^2(D+1)\left[e^{\frac{-1}{4\sigma^2}}-1\right] + \int_0^1 y^2 e^{\frac{-y^2}{4\sigma^2}}\,\mathrm{d}y\Bigg)^2.\end{aligned}$$

Furthermore,

$$\int_{\mathbb{R}^2} \gamma_X(y)\,\mathrm{d}y = \left(\int_{-D}^{D}(D-|y_1|)\,\mathrm{d}y_1\right)^2 = D^4,$$

$$\begin{aligned}\int_{\mathbb{R}^2} \gamma_X(y)e^{-\frac{\|y\|^2}{4\sigma^2}}\,\mathrm{d}y &= \left(\int_{-D}^{D}(D-|y_1|)e^{-\frac{y_1^2}{4\sigma^2}}\,\mathrm{d}y_1\right)^2 \\ &= 4\left(D\sqrt{\pi\sigma^2}\left[2\Phi\left(\frac{D}{\sqrt{2\sigma^2}}\right)-1\right] + 2\sigma^2\left[e^{\frac{-D^2}{4\sigma^2}}-1\right]\right)^2.\end{aligned}$$

Using (14) and (18), we finally get

$$\begin{aligned}\mathbb{E}(\widehat{N}_A^{UR}-\widehat{N}_A)^2 &= \lambda(1+\frac{2}{D}) - \lambda^2 + \lambda^2\frac{(D+1)^2\left(D-\frac{1}{3}\right)^2}{D^4} \\ &\quad + \lambda\frac{(D+1)^2\mathbb{V}(L')}{D^4\,\pi\,\sigma^2\,\mathbb{E}L'}\Bigg(D\sqrt{\pi\sigma^2}\left[2\Phi\left(\frac{1}{\sqrt{2\sigma^2}}\right)-1\right] \\ &\quad + 2\sigma^2(D+1)\left[e^{-\frac{1}{4\sigma^2}}-1\right] + \int_0^1 y^2 e^{-\frac{y^2}{4\sigma^2}}\,\mathrm{d}y\Bigg)^2 \\ &\quad - \lambda\frac{\mathbb{V}(L')}{D^4\,\pi\,\sigma^2\,\mathbb{E}L'}\Bigg(D\sqrt{\pi\sigma^2}\left[2\Phi\left(\frac{D}{\sqrt{2\sigma^2}}\right)-1\right] \\ &\quad + 2\sigma^2\left[e^{-\frac{D^2}{4\sigma^2}}-1\right]\Bigg)^2.\end{aligned}$$