Monte Carlo integration of Sobolev functions under systematic sampling

L. Barabesi and M. Marcheselli

Dipartimento di Metodi Quantitativi, *Università di Siena*

Working Paper n. 60, 2006

Monte Carlo integration of Sobolev functions under systematic sampling

L. Barabesi and M. Marcheselli

Dipartimento di Metodi Quantitativi, Università di Siena,

P.zza S.Francesco 17, 53100 Siena, Italy, barabesi@unisi.it

Summary. The Monte Carlo integration of a suitable function by means of systematic sampling is required in many settings, such as in stereological and environmental designs. The properties of the corresponding standard Monte Carlo estimator are well-established when the integrand function satisfies some regularity conditions. However, these assumptions may not hold even for quite simple cases. Hence, some general results are given when the integrand function is defined in the Sobolev space.

Key words: Monte Carlo integration; Systematic sampling; Stereology; Replicated environmental design; Cavalieri estimator; Weak derivative.

1. Introduction. The use of systematic sampling is popular among field scientists in order to analyze geometrical structures embedded in $\mathbb{R}^2$ or $\mathbb{R}^3$. For example, in stereology the target object is often sampled by means of a set of equally spaced probes (such as test points, lines or planes) with a random start (for an introduction to stereology see *e.g.* Baddeley and Jensen, 2004). Indeed, systematic sampling is a convenient way of introducing replications owing to physical reasons, since the probes must be often placed at a minimum distance apart. Similarly, in environmental studies the region of interest is commonly surveyed by using randomly placed systematic grids of points or transects (see *e.g.* U.S. EPA QA/G-5S Guidance, 2002, and Thompson, 2002). In turn, ecological scientists judge systematic sampling as a practical replicated strategy to collect data on the field.

In this setting, the key parameter is often the integral of a suitable function defined on a compact support contained in $\mathbb{R}$. For example, in stereology the parameter is frequently given by the volume of the target object, which may be obviously represented as the integral of the so-called measurement function on the basis of the Cavalieri principle. Since in this case the object is sectioned by using of a set of parallel planes with a random start, the volume is estimated on the basis of the corresponding plane-section areas (Cruz-Orive, 1999). As a further example, in environmental designs such as the line-intercept sampling (see *e.g.* Thompson, 2002) the parameter is usually represented by the total of an attribute scattered on the population units. Moreover, it can be shown that the parameter may expressed as the integral of the Horvitz-Thompson estimating function or, more generally, of the linear homogeneous estimating function (Barabesi, 2003). Since the total is estimated by replicating the design, field researchers often adopt systematic sampling of transects. However, it should be remarked that non-aligned systematic sampling may be theoretically preferable to aligned systematic sampling in this context as remarked by Barabesi (2003), Barabesi and Marcheselli (2003, 2005).

The properties of the Monte Carlo estimator of the integral under systematic sampling obviously rely on the characteristic of the integrand function. It is at once apparent that the estimator displays an increasing accuracy as the integrand becomes more regular, as clearly shown by Kiêu, Souchet and Istas (1999). However, since in stereology a less regular integrand is likely to appear in practice, García-Fiñana and Cruz-Orive (2000, 2004) have considered less stringent assumptions on the integrand class. Nevertheless, even in this case the class is not large enough to encompass even quite simple functions, as pointed out by García-Fiñana and Cruz-Orive (2004). Accordingly, the aim of this paper is to assess the property of the Monte Carlo estimator when the integrand function is defined in a Sobolev space, which may contain functions displaying very irregular patterns.

2. Preliminaries. Let $f : \mathbb{R} \mapsto \mathbb{R}^+$ be a bounded integrable function (usually referred to as the measurement function in stereology). Moreover, let $[a, b]$ be the finite support of f. Hence, the target parameter is given by

$$Q = \int_{-\infty}^{\infty} f(x)\, dx \; . \tag{1}$$

The estimation of (1) is performed by means of systematic sampling with a random start based on a sampling period T $(T > 0)$. In order to avoid triviality, it is supposed that $T \leq b - a$. Thus, if U is a uniform random variable on $[0,1]$, the Monte Carlo estimator of Q is given by

$$\widehat{Q}_T = T \sum_{k \in \mathcal{Z}} f(T(U+k)) \; . \tag{2}$$

It is straightforward to prove that (2) is an unbiased estimator of Q. However, in order to assess the behavior of $\mathrm{Var}[\widehat{Q}_T] = \sigma_T^2$ some regularity conditions on f must be assumed.

First, Kiêu, Souchet and Istas (1999) have introduced a general class of measurement functions. Subsequently, García-Fiñana and Cruz-Orive (2000, 2004) have considered an extension of the same class, *i.e.* the so-called q-smooth functions. More specifically, if $Df^{(k)}$ represents the set for which $f^{(k)}$ is not continuous, then f is defined to be a q-smooth function if:

i) $Df^{(k)} = \emptyset$ $(k = 0, 1, \ldots, \lceil q \rceil - 1)$;

ii) $Df^{(\lceil q \rceil)} = \{c_i \, , i = 1, \ldots, N\}$ is a finite set and there exist constants $d_i^-, d_i^+ \in \mathbb{R}$, $\alpha_i^-, \alpha_i^+ \in [0,1)$ such that

$$f^{(\lceil q \rceil)}(x) = \begin{cases} d_i^- |x - c_i|^{-\alpha_i^-} + o(1) & x \in [c_i - \delta, c_i) \\ d_i^+ |x - c_i|^{-\alpha_i^+} + o(1) & x \in (c_i, c_i + \delta] \end{cases}$$

where $\delta > 0$. In the previous expressions, $\lceil q \rceil$ denotes the smallest integer greater than (or equal to) q. Moreover, the quantity $q = \lceil q \rceil - \alpha$ with $\alpha = \max(\alpha_1^-, \alpha_1^+, \ldots, \alpha_N^-, \alpha_N^+)$ is called the fractional smoothness constant of f. In this case, García-Fiñana and Cruz-Orive (2000, 2004) prove that

$$\sigma_T^2 = \mathrm{Var}_E[\widehat{Q}_T] + Z(T) + o(T^{2q+2}) \; .$$

The so-called extension term $\mathrm{Var}_E[\widehat{Q}_T] = O(T^{2q+2})$ explains the overall trend of the variance, while $Z(T)$ is the *Zitterbewegung* term, *i.e.* an oscillating function about zero. However, as remarked by García-Fiñana and Cruz-Orive (2004), even simple measurement functions such as $f_1(x) = -x\log(x)\mathrm{I}_{[0,1]}(x)$ are not q-smooth. Hence, a suitable functional space must be considered to exploit the properties of σ_T^2 in a more general setting.

Let us suppose that $f \in W^{1,p}(\mathbb{R})$ with $p \in [1,\infty)$, *i.e.* the measurement function is required to be a member of a Sobolev space. This space contains the functions $f \in L^p(\mathbb{R})$ such that it exists $g \in L^p(\mathbb{R})$ (the so-called weak derivative of order p) for which

$$\int_{\mathbb{R}} f(x)\varphi'(x)\,dx = -\int_{\mathbb{R}} g(x)\varphi(x)\,dx$$

for each test function $\varphi \in C^1(\mathbb{R})$ (see *e.g.* Brézis, 1983). Equivalently, it may be proven that $f \in W^{1,p}(\mathbb{R})$ if and only if f is almost everywhere differentiable and $f' \in L^p(\mathbb{R})$ in such a way that for each $x, y \in \mathbb{R}$ it holds

$$f(x) - f(y) = \int_y^x f'(t)\,dt\,.$$

Owing to the boundedness of the support of f, it can be proven that $W^{1,p_1}(\mathbb{R}) \subset W^{1,p_2}(\mathbb{R})$ if $p_1 \geq p_2$ and therefore it is at once apparent that the Sobolev space $W^{1,1}(\mathbb{R})$ may contain function with rather irregular behavior. Indeed, $W^{1,1}(\mathbb{R})$ actually coincides with the space of the absolutely continuous functions. An extension of $W^{1,p}(\mathbb{R})$ is given by the Sobolev space $W^{s,p}(\mathbb{R})$ where $s \geq 2$ is an integer. The space $W^{s,p}(\mathbb{R})$ may be defined by recurrence, *i.e.* the members of $W^{2,p}(\mathbb{R})$ are such that $f, f' \in W^{1,p}(\mathbb{R})$, the members of $W^{3,p}(\mathbb{R})$ are such that $f, f', f'' \in W^{1,p}(\mathbb{R})$, etc. (see *e.g.* Brézis, 1983). In turn, it is at once apparent that $W^{s,p_1}(\mathbb{R}) \subset W^{s,p_2}(\mathbb{R})$ if $p_1 \geq p_2$.

It is worth noting that a q-smooth measurement function belongs to $W^{s,p}(\mathbb{R})$ with $s = \lceil q \rceil$ and for each $p < 1/\alpha$. Obviously, the converse is not true since Sobolev spaces contain continuous functions which may be not differentiable on dense sets. As a further example, it should be emphasized that the function f_1 is “quite” regular since it is a member of $W^{1,p}(\mathbb{R})$ for each p, *i.e.* it may be considered a “nearly” Lipschitz function. Indeed, a Lipschitz function is equivalent to a $W^{1,\infty}(\mathbb{R})$ function.

3. Estimator variance properties. First, it is convenient to introduce some notation, *i.e.* let us assume that

$$\widehat{Q}_T^{(s)}(x) = T \sum_{k \in \mathcal{Z}} f^{(s)}(T(x+k)) \, .$$

In order to analyze the behaviour of the variance of (2) when $f \in W^{s,p}(\mathbb{R})$, the following preliminary result produces a suitable alternative expression of σ_T^2, as well as a tigth inequality of σ_T.

Lemma 1. *For each $f \in W^{1,1}(\mathbb{R})$, it turns out that*

$$\sigma_T^2 = T^2 \int_{[0,1]^2} \phi(x,y) \widehat{Q}_T^{(1)}(x) \widehat{Q}_T^{(1)}(y) \, dxdy \, ,$$

where $\phi(x,y) = x \wedge y - xy$. In particular, it follows that

$$\sigma_T \leq \frac{T}{2} \, \mathrm{E}[|\widehat{Q}_T^{(1)}(U)|] \, . \tag{3}$$

Moreover, if $f \in W^{s,1}(\mathbb{R})$ with $s \geq 2$, then

$$\sigma_T \leq \frac{T^s}{2^s} \, \mathrm{E}[|\widehat{Q}_T^{(s)}(U)|] \, . \tag{4}$$

Proof. Since for $f \in W^{1,1}(\mathbb{R})$

$$f(T(U+k)) - f(Tk) = \, T \int_0^U f'(T(x+k)) \, dx$$

and

$$\mathrm{E}[f(T(U+k))] - f(Tk) = \, T \int_0^1 P(U > x) f'(T(x+k)) \, dx \, ,$$

it follows that

$$\widehat{Q}_T - Q = T^2 \sum_{k \in \mathcal{Z}} Z_k \, ,$$

where

$$Z_k = \int_0^1 [\mathrm{I}_{\{x<U\}} - P(x < U)] f'(T(x+k)) \, dx \, .$$

Moreover, since

$$Z_k Z_h = \int_{[0,1]^2} [\mathrm{I}_{\{x<U\}} - P(x<U)][\mathrm{I}_{\{y<U\}} - P(y<U)] f'(T(x+k)) f'(T(y+k))\,dxdy$$

and

$$\begin{aligned}\mathrm{E}[Z_k Z_h] &= \int_{[0,1]^2} (1 - x \vee y - (1-x)(1-y)) f'(T(x+k)) f'(T(y+k))\,dxdy \\ &= \int_{[0,1]^2} (x \wedge y - xy) f'(T(x+k)) f'(T(y+k))\,dxdy\,,\end{aligned}$$

it turns out that

$$\sigma_T^2 = T^4 \sum_{k,h\in\mathcal{Z}} \int_{[0,1]^2} \phi(x,y) f'(T(x+k)) f'(T(y+k))\,dxdy\,.$$

Hence, the first part follows. Expression (3) holds since $\phi(x,y) \le 1/4$ for $x, y \in [0,1]$ and

$$\int_{[0,1]^2} \widehat{Q}_T^{(1)}(x)\widehat{Q}_T^{(1)}(y)\,dxdy \le \int_{[0,1]^2} |\widehat{Q}_T^{(1)}(x)||\widehat{Q}_T^{(1)}(y)|\,dxdy = \mathrm{E}[|\widehat{Q}_T^{(1)}(U)|]^2\,.$$

Expression (4) may be proven recursively. Indeed, since

$$\mathrm{E}[|\widehat{Q}_T^{(1)}(U)|]^2 \le \mathrm{E}[\widehat{Q}_T^{(1)}(U)^2] = \mathrm{Var}[\widehat{Q}_T^{(1)}(U)]\,,$$

from (3) it turns out that

$$\mathrm{Var}[\widehat{Q}_T^{(1)}(U)] \le \frac{T}{2}\,\mathrm{E}[|\widehat{Q}_T^{(2)}(U)|]\,,$$

and hence by iterating expression (4) follows. □

Obviously, since the previous results hold for each $f \in W^{s,1}(\mathbb{R})$, in turn they hold for a general $f \in W^{s,p}(\mathbb{R})$. The following result gives rise to a more accurate inequality for σ_T when it is assumed a further mild requirement on f, *i.e.* $f' \in W^{1,1}([c,d])$ for each $[c,d] \subset (a,b)$.

Lemma 2. *Let us assume that* $f \in W^{1,1}(\mathbb{R})$ *in such a way that* $f' \in W^{1,1}([c,d])$ *for each* $[c,d] \subset (a,b)$ *and that* $\alpha_T(x) = a + Tx$ *and* $\beta_T(x) = b + T(x-1)$ *with* $x \in (0,1)$. *If* $(b-a)/T = m+1$ *with* $m \in \mathcal{Z}^+$ *, it follows that*

$$\widehat{Q}_T^{(1)}(x) = [f(\beta_T(x)) - f(\alpha_T(x))] + \frac{T}{2}\,[f'(\beta_T(x)) + f'(\alpha_T(x))] + g_T(x)\,,$$

where

$$g_T(x) = T^2 \int_0^m p(u) f''(T(u+x))\,du = T \int_{\alpha_T(x)}^{\beta_T(x)} p((u-a)/T - x) f''(u)\,du$$

and $p : u \mapsto u - \lfloor u \rfloor - 1/2$*. In particular, if the sign of* f'' *is constant in* (a,c) *and* (d,b) *for a fixed* (c,d)*, it exists a constant* $L \geq 0$*, such that for each* $x \in (0,1)$*, it turns out that*

$$|\widehat{Q}_T^{(1)}(x) - [f(\beta_T(x)) - f(\alpha_T(x))]| \leq T\,[|f'(\beta_T(x))| + |f'(\alpha_T(x))| + L]$$

(*if* f *is a concave function,* L *may be taken equal* 0 *for a suitable interval* $[c,d]$). *Hence, if* $T \leq (c-a) \wedge (b-d)$*, it finally holds*

$$\sigma_T \leq T\,[|f(b-T)| + |f(a+T)|] + \frac{LT^2}{2}\,. \tag{5}$$

Proof. For sake of simplicity, it is assumed $a = 0$ without loss of generality. Since the family $C^2(\mathbb{R})$ is dense in $W^{2,1}(\mathbb{R})$ it suffices to consider functions in this class. Moreover, it should be remarked that the Euler-McLaurin expansion implies that for each $f \in C^1$

$$\sum_{k=0}^{m} f(k) = \int_0^m f(x)\,dx + \frac{1}{2}\,(f(0) + f(m)) + \int_0^m p(x) f'(x)\,dx\,.$$

By applying the previous expansion to $F : u \mapsto Tf'(T(u+x))$ on $[0,m]$ the first result follows. Indeed, it turns out that $F(0) = Tf'(\alpha_T(x))$, $F(m) = Tf'(\beta_T(x))$ and $\int_0^m F(u)du = f(\beta_T(x)) - f(\alpha_T(x))$.

In particular, if f'' has a constant sign on $(0,c)$ and (d,b) for a fixed (c,d), in order to obtain (5) it should be remarked that p is bounded by $1/2$ and let

$$L \geq \frac{1}{2}\,(|f'(c)| + |f'(d)| + \int_c^d |f''(u)|\,du)\,.$$

Moreover, if $T \leq c \wedge (b-d)$ the function f' has in turn a constant sign on $(0,T)$ and $(b-T,T)$, from which

$$\int_0^1 (|f'(\beta_T(x))| + |f'(\alpha_T(x))|)dx = \frac{1}{T}\,(|f(b-T)| + |f(T)|)\,.$$

Finally, on the basis of (4) and since $x \mapsto |f'(\beta_T(x))|$ is increasing and $x \mapsto |f'(\alpha_T(x))|$ is decreasing on $(0,1)$, inequality (5) follows. $\square$

In order to give more insigth into the previous results, it is worthwhile to consider a simple example by considering again the function f_1. It is at once apparent that $f_1' \in W^{1,1}([c,d])$ for each $[c,d] \subset (0,1)$ and that f_1 is concave. Hence from Lemma 2 it follows that for $T \leq 1/e$

$$\sigma_T \leq T\,[-(1-T)\log(1-T) - T\log T] \leq T^2 - T^2\log T\ ,$$

from which $\sigma_T^2 \leq T^4(1-\log T)^2$.

Lemma 2 gives rises to the following Corollary.

Corollary 3. *Let $f \in W^{s,1}(\mathbb{R})$ in such a way that $f^{(s)} \in W^{1,1}([c,d])$ for each $[c,d] \subset (a,b)$. If the sign of $f^{(s+1)}$ is constant in (a,c_0) and (d_0,b) for a fixed (c_0,d_0), it exists a constant $L \geq 0$, such that for each $T \leq (c_0-a) \wedge (b-d_0)$, it holds*

$$\sigma_T \leq \frac{T^s}{2^{s-1}}\,[|f^{(s-1)}(b-T)| + |f^{(s-1)}(a+T)|] + \frac{LT^{s+1}}{2^s}\ .$$

Remark 1. If $f^{(s-1)}$ is supposed to be a Hölderian function with parameter β in Corollary 3, it follows that

$$\sigma_T \leq C\,\frac{T^{s+\beta}}{2^{s-2}} + \frac{LT^{s+1}}{2^s}\ ,$$

for a suitable $C \geq 0$. It should be remarked that $f^{(s-1)} \in W^{1,p}(\mathbb{R})$, then $f^{(s-1)}$ is a Hölderian function with parameter $\beta < 1 - 1/p$. For a q-smooth measurement function $s = \lceil q \rceil$ and $\beta = 1 - \alpha$.

Remark 2. In an anagolous way, the previous results hold even if f is a finite sum of functions with singularities for f' in the points a' and b' in $[a,b]$, where a' and b' are not necessarily equal to the support endpoints.

As to the estimation of σ_T^2, from the representation in Lemma 1 an estimator may be given by

$$\widehat{\sigma}_T^2 = T^2\phi(U,V)\sum_{k,h\in\mathcal{Z}}[f(T(k+1+U))-f(T(k+U)][f(T(k+1+V))-f(T(k+V)]$$

where U and V are independent uniform random variables on $(0,1)$. However, this estimator obviously requires two independent sets of systematic observations and this may be not practically feasible in stereological applications. In contrast, inequality (5) may give rise to a conservative estimator of σ_T^2, which may be obtained as

$$\widehat{\sigma}_T^2 = \frac{T^2}{4}\,\phi(U,V)(1-U)(1-V)\,[\frac{f(\alpha_T(U))}{U}+\frac{f(\beta_T(U))}{1-U}][\frac{f(\alpha_T(V))}{V}+\frac{f(\beta_T(V))}{1-V}]\,,$$

where in turn U and V are independent uniform random variables on $(0,1)$. Even if this estimator is obviously based on a second replication of the systematic sampling, it is at once apparent that the measurements corresponding to the second replication are solely carried out in the interval $[a,a+T]$ and $[b-T,b]$ and hence the extra sampling effort could be trascurable.

4. Some applications. In this Section, we emphasize the application of the systematic sampling in stereology and in some environmental designs. As to the stereology setting, as previously remarked in the introduction, f usually represents the area section of an object along a given axis and hence Q turns out to be a volume. Thus, $\widehat{Q}_T$ reduces to the well-known Cavalieri estimator (Cruz-Orive, 1993). The sections may be physical or obtained by means of non-invasive methods such as the magnetic risonance imaging (Garcia-Fiñana *et al.*, 2003).

When a quantitative ecological study is carried out, the target parameter is usually the total of the variable under study in a delineated region, which may be assumed to be the unit square for sake of simplicity. Accordingly, let us consider a population of N units scattered at fixed locations in the region. In addition, let $q_1,\ldots,q_N$ be the values of the variable corresponding to each unit. Hence, the target parameter is given by

$$Q = \sum_{l=1}^{N} q_l \ .$$

In thi setting, a large family of practical environmental designs are carried out by throwing a point on the baseline, *i.e.* the projection of the region on a line of fixed or random direction (Barabesi, 2003, Barabesi and Pisani, 2002). Thus, let x be the coordinate of the selected point on the baseline which may be identified with the interval $[a, b]$ without loss of generality. In this setting, the inclusion set $\mathcal{P}_l$ of the l-th unit is a suitable interval contained within the baseline. Hence, the l-th unit is selected and q_l is measured if $x \in \mathcal{P}_l$. As an example, let us consider a population of plants in a delineated forest and let q_l be the canopy coverage of the l-th plant. Therefore, the target parameter is the total canopy coverage biomass in the forest. If the line intercept design is adopted, a transect perpendicular to the baseline at x is thrown across the study region (see *e.g.* the recent paper of Affleck, Gregoire, and Valentine). Subsequently, the plants intersected by the transect are included in the sample. Obviously, the inclusion sets are the canopy projections on the baseline. Hence, a plant is selected if x is located in the corresponding inclusion set.

In this framework, it is worthwhile to consider a general class of estimates of the total. The linear homogeneous estimate at the point x is given by (see *e.g.* Kaiser, 1983)

$$f(x) = \sum_{l=1}^{N} \frac{q_l}{w_l} \psi_l(x) \ ,$$

where $w_l = \int_a^b \psi_l(x) dx$ and ψ_l is a suitable function vanishing if $x \notin \mathcal{P}_l$. However, it is at once apparent that

$$\int_a^b f(x)\, dx = \int_a^b \sum_{l=1}^{N} \frac{q_l}{w_l} \psi_l(x)\, dx = \sum_{l=1}^{N} q_l = Q \ ,$$

i.e. an integral representation is achieved for Q and hence its estimation reduces to an integration problem. Obviously, estimator (2) must be considered when a replicated systematic sampling is adopted. In this case, the properties of (2) obviously depend on the structure of the components ψ_ls of f. As an example, if Q represents the coverage, the usual adopted linear homogeneous estimate assumes that $\psi_l(x)$ is the length of the intersection between the l-th plant and the transect. If the plants are assumed to be connected compact sets with a "smooth" boundary, it is at once apparent that the ψ_ls are generally $W^{1,p}(\mathbb{R})$ functions and hence f is in turn a $W^{1,p}(\mathbb{R})$ function (see Barabesi and Marcheselli, 2006). Other examples of estimator that are likely to produce ψ_l functions in $W^{1,p}(\mathbb{R})$ are contained in Kaiser (1983).

References

Affleck, D.R.L., Gregoire, T.G. and Valentine, H.T. (2005) Design unbiased estimation in line intersect sampling using segmented transects, *Environmental and Ecological Statistics*, in the press.

Baddeley, A.J. and Jensen, E.B.V. (2004) *Stereology for statisticians*, Chapman and Hall/CRC, London.

Barabesi, L. (2003) A Monte Carlo integration approach to Horvitz-Thompson estimation in replicated environmental designs, *Metron* **LXI**, 355-374.

Barabesi, L. and Marcheselli, M. (2003) A modified Monte Carlo integration, *International Mathematical Journal* **3**, 555-565.

Barabesi, L. and Marcheselli, M. (2005) Some large-sample results on a modified Monte Carlo integration method, *Journal of Statistical Planning and Inference* **135**, 420-432.

Barabesi, L. e Marcheselli, M. (2006) Improved strategies for coverage estimation by using replicated line-intercept sampling, *Environmental and Ecological Statistics*, under revision.

Barabesi, L. and Pisani, C. (2002) Ranked set sampling for replicated sampling designs, *Biometrics* **58**, 586-592.

Brézis, H. (1983) *Analyse fonctionelle*, Masson, Paris.

Cruz-Orive, L.M. (1993) Systematic sampling in stereology, *Bulletin of the International Statistical Institute*, Proceedings of 49th Session, Florence 1993, **53,** 451-468.

García-Fiñana, M. and Cruz-Orive, L.M. (2000) Fractional trend of the variance in Cavalieri sampling, *Image Analysis and Stereology* **19**, 71-79.

García-Fiñana, M. and Cruz-Orive, L.M. (2004) Improved variance prediction for systematic sampling on $\mathbb{R}$, *Statistics* **38**, 243-272.

Garcia-Fiñana, M., Cruz-Orive, L.M., Mackay, C.E., Pakkenberg, E.B. and Roberts, N. (2003) Comparison of MR imaging against physical sectioning to estimate the volume of human cerebral compartments, *NeuroImage* **18**, 505-516.

Kaiser, L. (1983) Unbiased estimation in line-intercept sampling, *Biometrics* **39**, 965-976.

Kiêu, K., Souchet, S. and Istas, J. (1999) Precision of systematic sampling and transitive methods, *Journal of Statistical Planning and Inference* **77,** 263-279.

Thompson, S.K. (2002) *Sampling*, 2nd edition, Wiley, New York.

U.S. Environmental Protection Agency (2002) Guidance on choosing a sampling design for environmental data collection, EPA QA/G-5S, Washington DC, 1-166.