CENTRE FOR **STOCHASTIC GEOMETRY**
AND ADVANCED **BIOIMAGING**

Rasmus Waagepetersen and Abdollah Jalilian

Fast bandwidth selection for estimation of the pair correlation function

No. 07, June 2017

Fast bandwidth selection for estimation of the pair correlation function

Rasmus Waagepetersen[1] and Abdollah Jalilian[2]

[1]Department of Mathematical Sciences, Aalborg University, Denmark, email: rw@math.aau.dk
[2]Department of Statistics, Razi University, Iran, email: jalilian@razi.ac.ir

Abstract

Kernel estimation is a popular approach to estimation of the pair correlation function function which is a fundamental spatial point process characteristic. Least squares cross validation was suggested by Guan (2007a) as a data-driven approach to select the kernel bandwidth. The method can, however, be computationally demanding for large point pattern data sets. We suggest a modified least squares cross validation approach that is asymptotically equivalent to the one proposed by Guan (2007a) but is computationally much faster.

Keywords: bandwidth, kernel estimator, pair correlation function, spatial point process

1 Introduction

The pair correlation function is an important spatial point process characteristic which is often estimated using kernel smoothing methods (Stoyan and Stoyan, 1994; Møller and Waagepetersen, 2003; Illian et al., 2008) related to kernel estimation of probability densities. The bias and variance of a kernel estimate depend critically on the choice of kernel bandwidth. In the stationary case, based on practical experience, Stoyan and Stoyan (1994) and Illian et al. (2008) suggest to use a bandwidth inversely proportional to the estimated intensity. Guan (2007a) and Guan (2007b) provide more principled data driven approaches where the bandwidth is chosen either by least squares cross validation or composite likelihood cross validation. It is, however, our impression that these suggestions have not found widespread use. We guess that this is due to lack of userfriendly software compatible with the predominant spatial statistics software package `spatstat` (Baddeley et al., 2015). Another reason could be that the cross validation approaches can be quite time consuming for large point pattern data sets. In this paper we suggest a modification of the least squares cross validation method. The modified approach is asymptotically equivalent with the method in Guan (2007a) but computationally much faster. We have implemented the new method in a `spatstat` procedure `bw.pcf` that covers both the original cross validation method in Guan (2007a) as well as the fast alternative.

2 Kernel estimation of the pair correlation function

Let X denote a point process on $\mathbb{R}^d$, $d \geq 1$, that is, X is a locally finite random subset of $\mathbb{R}^d$. We denote by $\rho(\cdot)$ the intensity function of X where $\rho(u)\mathrm{d}u$, $u \in \mathbb{R}^d$, is the probability of observing a point from X in a neighbourhood of u of infinitesimal volume $\mathrm{d}u$. Consider two locations $u, v \in \mathbb{R}^d$. Then the pair correlation function $g(u, v)$ evaluated at u, v can be interpreted as the ratio $\rho(u|v)/\rho(u)$ where $\rho(\cdot|v)$ denotes the intensity function of X conditional on that $v \in X$, see e.g. Coeurjolly et al. (2017). The pair correlation function thus quantifies how much the presence of a point at v changes the intensity at another location u. In this paper we will assume that g depends on u, v only through their distance and write with an abuse of notation, $g(u, v) = g(\|v - u\|)$.

Suppose X is observed within a bounded observation window $W \subset \mathbb{R}^d$ and let $X_W = X \cap W$. Further, $k_b(\cdot)$ denotes a kernel of the form $k_b(r) = k(r/b)/b$, where k is a probability density with bounded support $[-1, 1]$ and $b > 0$ is the bandwidth. Typical choices of k are the density of a uniform distribution or the Epanecnikov kernel $k(r) = (3/4)(1-r^2)1(|r| \leq 1)$. A kernel density estimator (Stoyan and Stoyan, 1994; Baddeley et al., 2000) of $g(r)$ is

$$\frac{1}{\varsigma_d r^{d-1}} \sum_{u,v \in X_W}^{\neq} \frac{k_b(r - \|v - u\|)}{\rho(u)\rho(v)|W \cap W_{v-u}|}, \qquad r \geq 0,$$

where ς_d is the surface area of the unit sphere in $\mathbb{R}^d$, $\sum^{\neq}$ denotes sum over pairs of distinct points, $1/|W \cap W_h|$, $h \in \mathbb{R}^d$, is the translation edge correction factor with $W_h = \{u + h : u \in W\}$, and $|A|$ is the volume of $A \subset \mathbb{R}^d$. A closely related estimator is (Guan, 2007a)

$$\hat{g}(r; b) = \frac{1}{\varsigma_d} \sum_{u,v \in X_W}^{\neq} \frac{k_b(r - \|v - u\|)}{\|v - u\|^{d-1} \rho(u)\rho(v)|W \cap W_{v-u}|}, \qquad r \geq 0.$$

The two estimators mainly differ for r close to zero where the former estimator tends to have strong positive bias while the latter may exhibit negative bias. We consider in the following the second alternative.

2.1 Least squares cross validation selection of the bandwidth

Guan (2007a) suggests to choose b by minimizing an estimate of the mean integrated squared error. For some upper lag R this is defined as

$$\mathrm{MISE}(b) = \varsigma_d \int_0^R \mathbb{E}\left\{\hat{g}(r; b) - g(r)\right\}^2 r^{d-1}\mathrm{d}r = M(b) + \varsigma_d \int_0^R g(r)^2 r^{d-1}\mathrm{d}r \qquad (2.1)$$

where

$$M(b) = \varsigma_d \int_0^R \mathbb{E}\,\hat{g}(r; b)^2 r^{d-1}\mathrm{d}r - 2\varsigma_d \int_0^R g(r)\,\mathbb{E}\,\hat{g}(r; b) r^{d-1}\mathrm{d}r.$$

Guan (2007a) disregards the term in (2.1) not depending on b and estimates $M(b)$ by

$$\hat{M}(b) = \varsigma_d \int_0^R \hat{g}(r;b)^2 r^{d-1} \mathrm{d}r - 2 \sum_{\substack{u,v \in X_W \\ \|v-u\| \leq R}}^{\neq} \frac{\hat{g}^{-\{u,v\}}(\|v-u\|;b)}{\rho(u)\rho(v)|W \cap W_{v-u}|}, \tag{2.2}$$

where $\hat{g}^{-\{u,v\}}$, is defined as $\hat{g}$ but based on the reduced data $(X \setminus \{u,v\}) \cap W$. That is,

$$\hat{g}^{-\{u,v\}}(r) = \frac{1}{\varsigma_d} \sum_{\substack{u',v' \in X_W: \\ \{u',v'\} \cap \{u,v\} = \emptyset}}^{\neq} \frac{k_b(r - \|v'-u'\|)}{\|v'-u'\|^{d-1} \rho(u')\rho(v')|W \cap W_{v'-u'}|}. \tag{2.3}$$

Imposing conditions on the observation window and boundedness and translation invariance for the second to fourth order cumulant functions, Guan (2007a) showed consistency in the sense that $\mathbb{E}\,\hat{M}(b) - M(b)$ tends to zero as the observation window increases. The use of conditions on the cumulant functions leads to a fairly long proof.

Define the k'th order normalized joint intensity $g^{(k)}$ of X by the identity

$$\mathbb{E} \sum_{u_1,\ldots,u_k \in X}^{\neq} h(u_1,\ldots,u_k) = \int_{(\mathbb{R}^d)^k} h(v_1,\ldots,v_k) \Big[\prod_{i=1}^k \rho(v_i)\Big] g^{(k)}(v_1,\ldots,v_k) \mathrm{d}v_1 \cdots \mathrm{d}v_k \tag{2.4}$$

for non-negative functions h on $(\mathbb{R}^d)^k$, where the sum is over distinct $u_1,\ldots,u_k$. As a simpler alternative to the conditions in Guan (2007a) on the cumulant functions we just assume that the fourth order normalized joint intensity $g^{(4)}$ is translation invariant and satisfies

$$\sup_{\|h_1\|,\|h_2\| \leq R+b} \int_{\mathbb{R}^d} \big| g^{(4)}(0,h_1,h_3,h_2+h_3) - g(\|h_1\|)g(\|h_2\|) \big| \mathrm{d}h_3 \leq C_1 \tag{2.5}$$

for some constant C_1. Intuitively this means that spatial dependence between two pairs of points decreases as a function of distance between the pairs. We further assume that the volume $|W|$ of the observation window tends to infinity in such a way that for some constant C_2,

$$|W|/|W \cap W_l| \leq C_2 \tag{2.6}$$

for any $l \in \mathbb{R}^d$ with $\|l\| \leq R+b$. This is easily seen to be satisfied e.g. for a sequence of observation windows nW_0, $n = 1,2,\ldots$, where W_0 is a rectangular subset of $\mathbb{R}^d$. For completeness we provide under these conditions a short proof of consistency of $\hat{M}(b)$ in the appendix.

For large point pattern data sets the evaluation of the second term in $\hat{M}(b)$ can be quite time consuming, see Section 4 for further details.

3 A fast estimate of the mean integrated squared error

The kernel estimate can be viewed as a density estimate based on distance observations $d = \|v-u\|$, $u,v \in X$. Given this perspective, a natural approach to cross validation is to leave out in turn the contribution from each observed distance. Leaving out a distance observation d corresponding to a pair $\{u,v\}$ leads to the modified leave-one-out estimate

$$\tilde{g}^{-\{u,v\}}(r) = \frac{1}{\varsigma_d} \sum_{\substack{u',v' \in X_W: \\ \{u',v'\} \neq \{u,v\}}}^{\neq} \frac{k_b(r - \|v'-u'\|)}{\|v'-u'\|^{d-1} \rho(u')\rho(v')|W \cap W_{v'-u'}|}.$$

As explained in Section 4 this is very convenient computationally. We thus suggest to use a modified least squares cross validation criterion

$$\tilde{M}(b) = \varsigma_d \int_0^R \hat{g}(r;b)^2 r^{d-1} \mathrm{d}r - 2 \sum_{\substack{u,v \in X_W \\ \|v-u\| \leq R}}^{\neq} \frac{\tilde{g}^{-\{u,v\}}(\|v-u\|;b)}{\rho(u)\rho(v)|W \cap W_{v-u}|} \tag{3.1}$$

that just differs from (2.2) by the replacement of $\hat{g}^{-\{u,v\}}$ by $\tilde{g}^{-\{u,v\}}$.

In addition to (2.5) and (2.6) assume further that

$$g^{(3)}(\cdot,\cdot,\cdot) \leq C_3 \text{ and } 1/\rho(\cdot) \leq C_4 \tag{3.2}$$

for constants C_3 and C_4, i.e. $g^{(3)}$ is bounded from above and the intensity is bounded away from zero. We then show that $\hat{M}(b)$ and $\tilde{M}(b)$ are asymptotically equivalent meaning that their expected difference converges to zero.

3.1 Asymptotic equivalence of cross validation criteria

The first terms in $\hat{M}(b)$ and $\tilde{M}(b)$ coincide so we focus on the last terms. By inspection of (2.2) and (2.3), the last term in $\hat{M}(b)$ is a sum over quadruples of distinct points $u,v,u',v' \in X_W$. The last term in $\tilde{M}(b)$ contains the same sum as well as a sum over $u,v,u',v' \in X_W$ where either $u=u'$, $u=v'$, $v=u'$ or $v=v'$. We need to show that the expected value of the latter sum converges to zero. Without loss of generality consider the sum of terms where $u=u'$. The other terms are handled similarly. By (2.4) with $k=3$, the expectation of

$$\sum_{\substack{u,v \in X_W \\ \|v-u\| \leq R}}^{\neq} \sum_{v' \in X_W} \frac{k_b(\|v-u\| - \|v'-u\|)}{\rho(u)\rho(v)\rho(u)\rho(v')\|v'-u\|^{d-1}|W \cap W_{v-u}||W \cap W_{v'-u}|}$$

is

$$\int_{W^3} 1[\|v-u\| \leq R] \frac{k_b(\|v-u\| - \|v'-u\|) g^{(3)}(u,v,v')}{\rho(u)\|v'-u\|^{d-1}|W \cap W_{v-u}||W \cap W_{v'-u}|} \mathrm{d}u\mathrm{d}v\mathrm{d}v'$$

Letting $K = \sup_{-b \leq r \leq b} k_b(r)$ and $V_d(R)$ the volume of the d-dimensional sphere, and using (2.5), (2.6) and (3.2), this is bounded by

$$\begin{aligned} &\frac{KC_2^2C_3C_4}{|W|^2} \int_{W^3} 1[\|v-u\| \leq R] \frac{1[\|v'-u\| \leq R+b]}{\|v'-u\|^{d-1}} \mathrm{d}u\mathrm{d}v\mathrm{d}v' \\ \leq &\frac{KC_2^2C_3C_4}{|W|^2} |W| V_d(R) \varsigma_d (b+R) \end{aligned}$$

which tends to zero as $|W|$ tends to infinity.

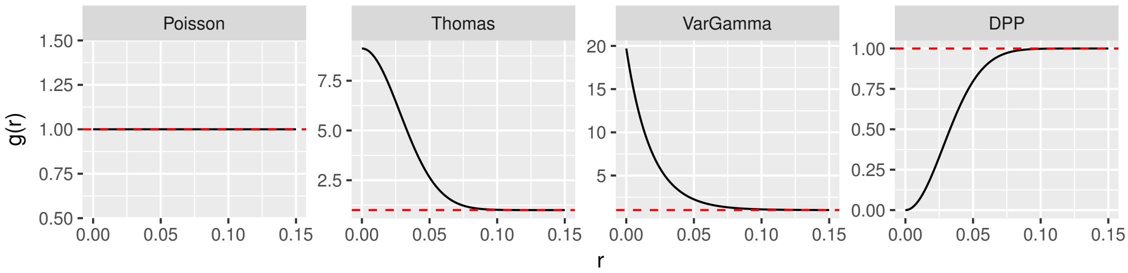

Figure 1: Pair correlation functions of the considered Poisson, Thomas, Variance-Gamma and Gaussian determinantal point processes.

4 Implementation and simulation study

To evaluate the cross validation criteria (2.2) and (3.1) we compute $\hat{g}(r;b)$ for $r \in G$ where G is a fine grid of spatial lags between 0 and R containing the set O of observed pairwise distances between 0 and R. This enables us to evaluate the first term in the cross validation criteria by numerical quadrature. To evaluate the second term we look up $\hat{g}(r;b)$ for each observed distance $r \in O$. Each such distance corresponds to a pair $\{u,v\}$. To evaluate $g^{-\{u,v\}}(r;b)$ appearing in (2.2) we need to subtract from $\hat{g}(r;b)$ all terms

$$T(\{u',v'\}) = \frac{1}{\varsigma_d} \frac{k_b(r - \|v' - u'\|)}{\|v' - u'\|^{d-1} \rho(u')\rho(v')|W \cap W_{v'-u'}|},$$

$u', v' \in X_W$, where $\{u,v\} \cap \{u',v'\} \neq \emptyset$. To evaluate $\tilde{g}^{-\{u,v\}}(r;b)$ in (3.1) we just need to subtract $2T(\{u,v\})$. This does not require any inspection of other pairs of points. The computation of $\tilde{g}^{-\{u,v\}}(r;b)$ for each observed pair $\{u,v\}$ with $r = \|u - v\| \leq R$ can therefore easily be done in a vectorized manner which is very fast in `R`.

4.1 Simulation study

To study the performance of the fast bandwidth selection method we perform a simulation study involving the following point process models:

P: a Poisson process with $\rho = 100$ and $g(r) = 1$.

T: a Thomas process with $\rho = 100$ and $g(r) = 1 + \exp\{-r^2/(4\omega^2)\}/(4\pi\omega^2\kappa)$, where $\kappa = 25$ and $\omega = 0.0198$.

V: a Variance-Gamma process with $\rho = 100$ and $g(r) = 1 + \exp(-r/\omega)/(2\pi\omega^2\kappa)$, where $\kappa = 25$ and $\omega = 0.01845$.

D: a Gaussian determinantal point process with $\rho = 100$ and $g(r) = 1 - \exp\{-2(r/\alpha)^2\}$, where $\alpha = 0.056$.

The pair correlation functions of these processes are shown in Figure 1. For the Poisson process there is no interaction between points, the Thomas and Var-Gamma processes produce clustered point pattern realizations, and the determinantal point process is regular.

For each of the observation windows $W_1 = [0,1]^2$ and $W_2 = [0,2]^2$ we generate 1000 realizations from each of the point process models. The average number of points is 100 on W_1 and 400 on W_2. For each simulated realization, the bandwidth b of the kernel density estimator $\hat{g}(r;b)$ is estimated using the `spatstat` default choice $\hat{b}_{\mathtt{st}} = 0.15/\sqrt{\hat{\rho}}$ due to Stoyan and Stoyan (1994), the fast least squares cross validation band width $\hat{b}_{\mathtt{fa}}$ obtained by minimizing (3.1), and the original least squares cross validation band width $\hat{b}_{\mathtt{ls}}$ obtained by minimizing (2.2). In $\hat{b}_{\mathtt{st}}$, $\hat{\rho}$ is the empirical intensity estimate given by the number of simulated points divided by the area of the observation window. Then the mean integrated squared errors (MISE) of $\hat{g}(r;\hat{b}_{\mathtt{st}})$, $\hat{g}(r;\hat{b}_{\mathtt{fa}})$ and $\hat{g}(r;\hat{b}_{\mathtt{ls}})$ are estimated by approximating the integral in (2.1) using numerical quadrature over $[0,R]$ and approximating the expectation in (2.1) by average over simulations. For W_1 we use $R = 0.25$ and for W_2 we use $R = 0.5$. Table 1 summarizes $\mathrm{MISE}(\hat{b}_{\mathtt{ls}})$, the relative differences in MISE in %, $d_M(\hat{b}_{\mathtt{st}}|\hat{b}_{\mathtt{ls}}) = 100\big(\mathrm{MISE}(\hat{b}_{\mathtt{st}}) - \mathrm{MISE}(\hat{b}_{\mathtt{ls}})\big)/\mathrm{MISE}(\hat{b}_{\mathtt{ls}})$ and $d_M(\hat{b}_{\mathtt{fa}}|\hat{b}_{\mathtt{ls}}) = 100\big(\mathrm{MISE}(\hat{b}_{\mathtt{fa}}) - \mathrm{MISE}(\hat{b}_{\mathtt{ls}})\big)/\mathrm{MISE}(\hat{b}_{\mathtt{ls}})$, and average computational times $t(\hat{b}_{\mathtt{fa}})$ and $t(\hat{b}_{\mathtt{ls}})$ for obtaining $\hat{b}_{\mathtt{fa}}$ and $\hat{b}_{\mathtt{ls}}$.

Table 1: Estimated values of $\mathrm{MISE}(\hat{b}_{\mathtt{ls}})$, the relative differences in MISE $d_M(\hat{b}_{\mathtt{st}}|\hat{b}_{\mathtt{ls}}) = 100\big(\mathrm{MISE}(\hat{b}_{\mathtt{st}}) - \mathrm{MISE}(\hat{b}_{\mathtt{ls}})\big)/\mathrm{MISE}(\hat{b}_{\mathtt{ls}})$ and $d_M(\hat{b}_{\mathtt{fa}}|\hat{b}_{\mathtt{ls}}) = 100\big(\mathrm{MISE}(\hat{b}_{\mathtt{fa}}) - \mathrm{MISE}(\hat{b}_{\mathtt{ls}})\big)/\mathrm{MISE}(\hat{b}_{\mathtt{ls}})$, and the average computational times $t(\hat{b}_{\mathtt{fa}})$ and $t(\hat{b}_{\mathtt{ls}})$ for obtaining $\hat{b}_{\mathtt{fa}}$ and $\hat{b}_{\mathtt{ls}}$. Windows $W_1 = [0,1]^2$ and $W_2 = [0,2]^2$ are considered.

W		$\mathrm{MISE}(\hat{b}_{\mathtt{ls}})$	$d_M(\hat{b}_{\mathtt{st}}\|\hat{b}_{\mathtt{ls}})$	$d_M(\hat{b}_{\mathtt{fa}}\|\hat{b}_{\mathtt{ls}})$	$\mathbb{E}\,t(\hat{b}_{\mathtt{fa}})$	$\mathbb{E}\,t(\hat{b}_{\mathtt{ls}})$
W_1	P	1.46×10^{-2}	1644.10	116.20	0.59	2.270
	T	143.31×10^{-2}	100.84	8.73	0.50	2.487
	V	344.66×10^{-2}	118.66	11.92	0.50	2.590
	D	5.31×10^{-2}	−48.09	−47.97	0.52	2.080
W_2	P	0.40×10^{-2}	3110.18	84.79	2.97	170.990
	T	34.25×10^{-2}	328.78	2.70	1.96	158.250
	V	158.35×10^{-2}	397.28	1.59	2.01	168.640
	D	7.36×10^{-2}	−81.81	−72.26	2.15	154.750

The simulation results are obtained on a Linux server with 2.3 GHz 12-Core Intel® Xeon® processor and 10GB of RAM.

Considering the results in columns 2 and 3 of Table 1, the `spatstat` default choice of band width is much inferior to both of the cross validation band widths in case of the Poisson and cluster point processes (increase in MISE at least 100% relative to MISE with $\hat{b}_{\mathtt{ls}}$). For the determinantal point process, the performance of the `spatstat` default and the fast cross validation method are similar and clearly better than for the original cross validation method. In case of the Thomas and Variance-Gamma processes, the relative increase in MISE by using $\hat{b}_{\mathtt{fa}}$ instead of $\hat{b}_{\mathtt{ls}}$ is moderate and at most 12% for W_1. For the larger observation window W_2, in accordance with the theoretical result in Section 3.1, the relative increase in MISE is smaller and at most 3%. For the Poisson process, the original cross validation method performs better than the fast one while the opposite is the case for the determinantal point process. Both for the Poisson and the determinantal process the absolute value of the MISE is small relative to the MISE for the clustered point processes regardless of whether the fast or the original cross validation method is used. Overall, the fast and the original cross validation methods have a similar performance in terms of MISE within the settings of the simulation study. The fast method on the other hand leads to substantial savings in computing time since the fast cross validation method is at least 4 or 57 times faster than the original cross validation method for W_1 respectively W_2.

5 Conclusion

In this paper we introduce a fast approach to cross validation selection of the band width for kernel estimation of the pair correlation function of a spatial point process. The method is asymptotically equivalent to the method introduced by Guan (2007a). In a simulation study involving a variety of point process models, the performance in terms of mean integrated squared error of the fast method is similar to the method proposed in Guan (2007a) and better than the `spatstat` default method. Already for moderately sized point patterns with on average 400 points, the computing time for the fast cross validation method is much smaller than for the original cross validation method.

Acknowledgment

This work was supported by The Danish Council for Independent Research – Natural Sciences, grant 7014-00074B, "Statistics for point processes in space and beyond", and by the "Centre for Stochastic Geometry and Advanced Bioimaging", funded by grant 8721 from the Villum Foundation.

References

Baddeley, A., Rubak, E., and Turner, R. (2015). *Spatial Point Patterns: Methodology and Applications with R.* Chapman & Hall/CRC Interdisciplinary Statistics. Chapman & Hall/CRC, Boca Raton.

Baddeley, A. J., Møller, J., and Waagepetersen, R. (2000). Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54:329–350.

Coeurjolly, J.-F., Møller, J., and Waagepetersen, R. (2017). A tutorial on Palm distributions for spatial point processes. *International Statistical Review.* to appear.

Guan, Y. (2007a). A least-squares cross-validation bandwidth selection approach in pair correlation function estimations. *Statistics and Probability Letters*, 77(18):1722–1729.

Guan, Y. (2007b). A composite likelihood cross-validation approach in selecting bandwidth for the estimation of the pair correlation function. *Scandinavian Journal of Statistics*, 34(2):336–346.

Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*, volume 76. Wiley, London.

Møller, J. and Waagepetersen, R. P. (2003). *Statistical inference and simulation for spatial point processes.* Chapman and Hall/CRC, Boca Raton.

Stoyan, D. and Stoyan, H. (1994). *Fractals, Random Shapes and Point Fields: Methods of Geometrical Statistics.* Wiley.

A Consistency of Guan's estimate

We need to show that the expectation of

$$\varsigma_d \int_0^R g(r)\,\mathbb{E}\,\hat{g}(r;b)r^{d-1}\mathrm{d}r - \sum_{\substack{u,v\in X_W \\ \|v-u\|\le R}}^{\neq} \frac{\hat{g}^{-\{u,v\}}(\|v-u\|;b)}{\rho(u)\rho(v)|W\cap W_{v-u}|}$$

converges to zero. By (2.4) with $k=2$, the integral term is

$$\begin{aligned}
&\varsigma_d \int_0^R g(r)\int_{W^2} \frac{k_b(r-\|v'-u'\|)}{\|v'-u'\|^{d-1}|W\cap W_{v'-u'}|} g(\|v'-u'\|)\mathrm{d}u'\mathrm{d}v' r^{d-1}\mathrm{d}r \\
&\quad = \varsigma_d \int_0^R g(r)\int_{\mathbb{R}^d\times\mathbb{R}^d} \frac{k_b(r-\|l\|)}{\|l\|^{d-1}|W\cap W_l|} g(\|l\|)1[u'\in W\cap W_{-l}]\mathrm{d}u'\mathrm{d}l r^{d-1}\mathrm{d}r \\
&\quad = \varsigma_d \int_0^R g(r)\int_{\mathbb{R}^d} \frac{k_b(r-\|l\|)}{\|l\|^{d-1}} g(\|l\|)\mathrm{d}l\mathrm{d}r^{d-1}\mathrm{d}r. \qquad (A.1)
\end{aligned}$$

The expectation of the sum is evaluated using (2.4) with $k=4$ and can be partitioned into the sum of

$$\int_{W^4} \frac{1[\|v-u\|\le R]}{\|v'-u'\|^{d-1}} \frac{k_b(\|v-u\|-\|v'-u'\|)}{|W\cap W_{v-u}||W\cap W_{v'-u'}|} g(\|v-u\|)g(\|v'-u'\|)\mathrm{d}u\mathrm{d}v\mathrm{d}u'\mathrm{d}v'$$

and

$$\int_{W^4} \frac{1[\|v-u\|\le R]}{\|v'-u'\|^{d-1}} \frac{k_b(\|v-u\|-\|v'-u'\|)}{|W\cap W_{v-u}||W\cap W_{v'-u'}|} \\ [g^{(4)}(0,v-u,u'-u,v'-u) - g(\|v-u\|)g(\|v'-u'\|)]\mathrm{d}u\mathrm{d}v\mathrm{d}u'\mathrm{d}v'$$

The first term is upon a change of variable equal to (A.1). Letting

$$K = \sup_{-b\le r\le b} k_b(r),$$

the second term is less than

$$
\begin{aligned}
&K\int_{(\mathbb{R}^d)^4} 1[\|h\|\leq R, \|l\|\leq R+b]\frac{1[u\in W\cap W_{-h}, u'\in W\cap W_{-l}]}{\|l\|^{d-1}|W\cap W_h||W\cap W_l|}\\
&\qquad\times |g^{(4)}(0,h,u'-u,u'-u+l)-g(\|h\|)g(\|l\|)|\mathrm{d}u\mathrm{d}u'\mathrm{d}h\mathrm{d}l\\
&\qquad\leq K\int_{(\mathbb{R}^d)^3} 1[\|h\|\leq R, \|l\|\leq R+b]\frac{1[u\in W\cap W_{-h}]}{\|l\|^{d-1}|W\cap W_h||W\cap W_l|}\\
&\qquad\times\left[\int_{\mathbb{R}^d}|g^{(4)}(0,h,k,k+l)-g(\|h\|)g(\|l\|)|\mathrm{d}k\right]\mathrm{d}u\mathrm{d}h\mathrm{d}l\\
&\qquad\leq KC_1\int_{(\mathbb{R}^d)^2}\frac{1[\|h\|\leq R, \|l\|\leq R+b]}{\|l\|^{d-1}|W\cap W_l|}\mathrm{d}l\mathrm{d}h\leq KC_1C_2\frac{1}{|W|}V_d(R)\varsigma_d(R+b)
\end{aligned}
$$

where $V_d(R)$ is the volume of the d-dimensional sphere of radius R. Thus the second term tends to zero as $|W|$ tends to infinity.