

Estimating the power-law of the two-point correlation function

An analytical investigation of several estimation procedures

M. Snethlage*

Mathematical Department, Silesian University at Opava, 74601 Opava 1, Czech Republic

Received 26 January 2001 / Accepted 1 May 2001

Abstract. There is empirical evidence that the two-point correlation function of the galaxy point pattern follows for small r reasonably well a power-law expression $\xi(r) \propto r^{-\gamma}$. This paper analyses the estimation procedure for γ and suggests an estimator for $\xi(r)$ which is well suited for the estimation of γ .

Key words. large-scale structure of Universe – cosmology: theory

1. Introduction

The two-point correlation function $\xi(r)$ has been the primary tool for quantifying large-scale cosmic structure (see Peebles 1980). Several estimates of this statistical quantity from red-shift surveys suggest a power-law shape

$$\xi(r) = \left(\frac{r}{r_0}\right)^{-\gamma},$$

with $0.5 h^{-1}\text{Mpc} < r < 10 h^{-1}\text{Mpc}$ (h being the Hubble constant in units of $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$). The true exponent γ for the galaxy point field has often been investigated. The detected values fluctuate between 1.5 and 1.9. Apart from different samples, the fluctuation may be also a consequence of different estimation procedures. The detection of γ is essentially based on an estimate of $\xi(r)$. Therefore, the attention has often been directed only to the accuracy of several $\xi(r)$ -estimators. Especially, biased estimators have been excluded.

This paper analyses the whole estimation procedure for γ based on an estimate of $\xi(r)$ and a doubly logarithmic linear regression procedure. It presents an adaptive kernel function which leads to a larger bias. Nevertheless, the adaptive kernel function is better suited for the estimation of γ than a constant bandwidth.

2. Expectation of regression

Let Φ be a point field whose two-point correlation function follows for small r the power-law

$$\xi(r) \propto r^{-\gamma}. \quad (1)$$

As $\xi(r)$ is very large for small r , this is approximately equivalent to

$$\varrho^{(2)}(r) \propto r^{-\gamma}, \quad (2)$$

where $\varrho^{(2)}(r)$ is the second order product density function. It is connected with $\xi(r)$ over

$$\varrho^{(2)}(r) = \lambda^2 (\xi(r) + 1).$$

The following calculations are based on (2) instead of (1) as this leads to less difficulties. Clearly, they could also be carried out based directly on the power-law for $\xi(r)$. The result is not considerably different.

Starting from an estimate of $\varrho^{(2)}(r)$, the detection of the exponent γ can be carried out by a doubly logarithmic linear regression. Introduce the function p defined by

$$p : s \mapsto \ln(\varrho^{(2)}(e^s)), \quad -\infty < s < \infty,$$

or equivalently

$$p : \ln r \mapsto \ln(\varrho^{(2)}(r)), \quad r > 0.$$

As the latter representation is more intuitive, it is used in the following. In the given case, $p(\ln r)$ is a straight line with slope $-\gamma$

$$\ln \varrho^{(2)}(r) = -\gamma \ln r + c,$$

where c is a suitable number. Thus, the exponent γ may be estimated by a simple linear regression procedure applied to an estimate of p . In the following, this estimate for γ is denoted by $\hat{\gamma}$.

As the power-law breaks down at separations r larger than some r_0 , one has to choose the regression window carefully and not too large. But in many cases, especially for the galaxy point pattern, a visual inspection seems to be sufficient. Thus, from now on $\xi(r)$ is assumed to follow the power-law on the whole regression window.

* e-mail: Martin.Snethlage@math.slu.cz

2.1. Approximating the regression procedure

The linear regression is usually based on a lot of sample points situated within an interval $[a, b]$. As this procedure is very complex, it is approximated here by the corresponding procedure including only the boundary points of the interval. When computing the mean of the slope of the regression line, this is only a slight simplification. While the corresponding regression lines in a single sample may considerably diverge, their average slopes will approximately coincide.

The boundary points of the regression window can be written as $r - \varepsilon$ and $r + \varepsilon$ with r being the centre of $[a, b]$. The regression is then equivalent to the determination of the secant between the two boundary points

$$\mathbf{E}\hat{\gamma}(r) \approx -\mathbf{E} \frac{\Delta\hat{p}(\ln r)}{\Delta \ln r} \text{ where } \frac{\Delta f}{\Delta r} = \frac{f(r + \varepsilon) - f(r - \varepsilon)}{2\varepsilon}.$$

Here, $\hat{p}(\ln r)$ is an estimate of $p(\ln r)$.

As the secant is a linear functional, it is

$$\mathbf{E} \frac{\Delta\hat{p}(\ln r)}{\Delta \ln r} = \frac{\Delta \mathbf{E}\hat{p}(\ln r)}{\Delta \ln r},$$

what is approximately equal to

$$\mathbf{E} \frac{\Delta\hat{p}(\ln r)}{\Delta \ln r} \approx \frac{\partial \mathbf{E}\hat{p}(\ln r)}{\partial \ln r} = r \frac{\partial \mathbf{E} \ln(\hat{\varrho}(r))}{\partial r}.$$

It is

$$\mathbf{E} \ln(\hat{\varrho}(r)) = \ln(\mathbf{E}\hat{\varrho}(r)) + \mathbf{E} \ln \frac{\hat{\varrho}(r)}{\mathbf{E}\hat{\varrho}(r)}.$$

A Taylor expansion (only the first two terms) for the second summand leads to

$$\begin{aligned} \mathbf{E} \ln \frac{\hat{\varrho}(r)}{\mathbf{E}\hat{\varrho}(r)} &\approx \mathbf{E} \left(\frac{\hat{\varrho}(r)}{\mathbf{E}\hat{\varrho}(r)} - 1 \right) - \frac{1}{2} \mathbf{E} \left(\frac{\hat{\varrho}(r)}{\mathbf{E}\hat{\varrho}(r)} - 1 \right)^2 \\ &= -\frac{1}{2} \text{Var} \frac{\hat{\varrho}(r)}{\mathbf{E}\hat{\varrho}(r)}. \end{aligned}$$

Landy & Szalay (1993) have shown that the variance of $\hat{\varrho}(r) / \mathbf{E}\hat{\varrho}(r)$ is an O -term of $1/N^2$, where N is the number of observed points. Thus, with growing observation window, this term can be neglected what results in

$$\mathbf{E}\hat{\gamma}(r) = -\frac{r}{\mathbf{E}\hat{\varrho}(r)} \frac{\partial \mathbf{E}\hat{\varrho}(r)}{\partial r}. \quad (3)$$

2.2. Expectation of $\hat{\varrho}(r)$

A popular estimator of $\varrho^{(2)}(r)$ is

$$\hat{\varrho}(r) = \frac{1}{C(r)} \sum_{x,y \in \Phi \cap W}^{\neq} \mathbf{1}_{[r-h,r+h]}(|x-y|),$$

see Ohser (1991). Here, W is the observation window, $\mathbf{1}_A(x)$ the indicator function of a set A , and h a suitably small number, called bandwidth, which is usually much smaller than r . The \neq sign means that the summation goes over all point pairs with different members. The number $C(r)$ depends on the given dimensionality.

In the most interesting case of three dimensions it is $C(r) = 4\pi r^2 2h \bar{\gamma}_W(r)$, where $\bar{\gamma}_W(r)$ is the isotropized set covariance function.

Among astronomers, this estimator is well-known under the designation

$$\hat{\varrho}(r) = \frac{DD(r)}{RR(r)} \cdot \frac{N_{rd}^2}{N^2} (\bar{\varrho})^2.$$

For more detailed information especially about the notation, see e.g. Davis & Peebles (1983).

There are a lot of other estimators existing. As most of them differ substantially in the edge correction, they behave similarly for small r (see Kerscher 1999). Thus, the results of the following calculations hold approximately also for many other estimators.

The expectation of $\hat{\varrho}(r)$ is

$$\begin{aligned} \mathbf{E}\hat{\varrho}(r) &= \frac{1}{C(r)} \int_{W^2} \varrho^{(2)}(|x-y|) \mathbf{1}_{[r-h,r+h]}(|x-y|) d(x,y) \\ &= \frac{1}{C(r)} \int \varrho^{(2)}(|x|) \mathbf{1}_{[r-h,r+h]}(|x|) \\ &\quad \cdot \int \mathbf{1}_W(y) \mathbf{1}_W(x+y) dy dx \\ &= \frac{1}{C(r)} \int_{(r-h)_+}^{r+h} \varrho^{(2)}(v) 4\pi v^2 \bar{\gamma}_W(v) dv \\ &\approx \frac{1}{2hr^2} \int_{(r-h)_+}^{r+h} \varrho^{(2)}(v) v^2 dv, \end{aligned} \quad (4)$$

where $V(A)$ is the volume of the set A and x_+ is the maximum of x and 0. In the given case ($\varrho^{(2)}(r) = cr^{-\gamma}$ with c being a suitable number), (4) yields

$$\begin{aligned} \mathbf{E}\hat{\varrho}(r) &= \frac{c}{2hr^2} \int_{(r-h)_+}^{r+h} v^{2-\gamma} dv \\ &= c \frac{(r+h)^{3-\gamma} - ((r-h)_+)^{3-\gamma}}{2hr^2(3-\gamma)}. \end{aligned}$$

Its derivative with respect to r is

$$\frac{\partial \mathbf{E}\hat{\varrho}(r)}{\partial r} = -\frac{2}{r} \mathbf{E}\hat{\varrho}(r) + c \frac{(r+h)^{2-\gamma} - ((r-h)_+)^{2-\gamma}}{2hr^2},$$

thus (3) yields

$$\mathbf{E}\hat{\gamma}(r) = 2 - (3-\gamma)r \frac{(r+h)^{2-\gamma} - ((r-h)_+)^{2-\gamma}}{(r+h)^{3-\gamma} - ((r-h)_+)^{3-\gamma}}. \quad (5)$$

To simplify this expression introduce $q = h/r$. Then (5) yields

$$\mathbf{E}\hat{\gamma}(r) = 2 - (3-\gamma) \frac{(1+q)^{2-\gamma} - ((1-q)_+)^{2-\gamma}}{(1+q)^{3-\gamma} - ((1-q)_+)^{3-\gamma}}.$$

Obviously, $\mathbf{E}\hat{\gamma}(r)$ depends only on the quotient of h and r and not on their absolute values. If q is much smaller than

1, the last quotient can be approximated by

$$\begin{aligned} & \lim_{q \rightarrow 0} \frac{(1+q)^{2-\gamma} - (1-q)^{2-\gamma}}{(1+q)^{3-\gamma} - (1-q)^{3-\gamma}} \\ &= \lim_{q \rightarrow 0} \left(\frac{(1+q)^{2-\gamma} - (1-q)^{2-\gamma}}{(1+q)^{3-\gamma} - (1-q)^{3-\gamma}} \right) / q \\ &\quad \cdot \lim_{q \rightarrow 0} q / ((1+q)^{3-\gamma} - (1-q)^{3-\gamma}) \\ &= \frac{(x^{2-\gamma})'(1)}{(x^{3-\gamma})'(1)}. \end{aligned}$$

This is the quotient of the derivatives of the two functions $x^{2-\gamma}$ and $x^{3-\gamma}$ at $x = 1$. It is

$$\frac{(x^{2-\gamma})'(1)}{(x^{3-\gamma})'(1)} = \frac{2-\gamma}{3-\gamma},$$

hence

$$\mathbf{E}\hat{\gamma}(r) = 2 - (3-\gamma)\frac{2-\gamma}{3-\gamma} = \gamma.$$

Thus, if h is small in comparison with r (and hence q small), $\hat{\gamma}$ is approximately unbiased.

Note that this is not true in the reversal case when r is much smaller than h . This can happen when uncarefully estimating $\varrho^{(2)}(r)$ for $r \rightarrow 0$. Then, (5) yields

$$\begin{aligned} \mathbf{E}\hat{\gamma}(r) &= 2 - (3-\gamma)r \frac{(r+h)^{2-\gamma}}{(r+h)^{3-\gamma}} \\ &\approx 2 - \frac{r}{h}(3-\gamma) \\ &\approx 2, \end{aligned} \tag{6}$$

independent of the true value of γ . Thus, the naive estimation of $\varrho^{(2)}(r)$ for $r \rightarrow 0$ may cause serious errors.

Apart from a constant bandwidth also an adaptive bandwidth is imaginable. If $h = h(r) = r$, (4) yields

$$\mathbf{E}\hat{\varrho}(r) = c \frac{(2r)^{3-\gamma}}{2r^3(3-\gamma)} = \frac{2^{2-\gamma}c}{3-\gamma} r^{-\gamma}.$$

Its derivative with respect to r is

$$\frac{\partial}{\partial r} \mathbf{E}\hat{\varrho}(r) = -\gamma \frac{2^{2-\gamma}c}{3-\gamma} r^{-\gamma-1},$$

thus

$$\mathbf{E}\hat{\gamma}(r) = \gamma.$$

Although the use of $h(r)$ increases the bias of $\hat{\varrho}(r)$, it leads to a γ -estimator which is in the mean as good as in case of a small constant bandwidth.

3. Variance of $\hat{\gamma}$

The preceding calculations have shown that both the adaptive and constant bandwidth lead to an approximately unbiased γ -estimator. Just as important as the unbiasedness is the variance of the γ -estimator. This section shows that the estimation variance is considerably reduced when using the adaptive bandwidth.

Recall the estimator for $\varrho^{(2)}(r)$

$$\hat{\varrho}(r) = \frac{1}{4\pi r^2 2h \bar{\gamma}_W(r)} \sum_{x,y \in \Phi \cap W} \mathbf{1}_{[r-h,r+h]}(|x-y|).$$

Essentially, the estimation is carried out by counting all those point pairs whose interpoint distance is situated within the interval $[r-h, r+h]$. In case of the adaptive bandwidth $h(r) = r$ this interval is much larger than for a constant bandwidth, as a constant bandwidth has to be much smaller than r . (This ensures the approximate unbiasedness of $\hat{\gamma}$, cf. Sect. 2.2. Otherwise, according to (6), considerable errors may appear.) Therefore, the estimation includes more point pairs when using the adaptive bandwidth and it is plausible to assume that the $\varrho^{(2)}(r)$ -estimator as well as the γ -estimator becomes more robust. Consequently, the use of the adaptive bandwidth should lead to a smaller estimation variance than a constant bandwidth.

This hypothesis was verified by several simulation studies. The description of all details and results of the studies would be beyond the scope of this paper. Therefore, only a summary of them is given here.

In more than 3000 different situations the studies have investigated whether or not the adaptive bandwidth leads to a smaller estimation variance than a constant bandwidth. The studies are based on simulations of the Gauss-Poisson process. This is a two-point cluster process with iid. clusters applied to a homogeneous Poisson process (for more details, see Stoyan 1994 Example 2). Certainly, the corresponding two-point correlation function fulfills the power-law within a certain interval.

The length and the position of the regression interval has been varied as well as the length of the constant bandwidth. To ensure the (approximate) unbiasedness of the γ -estimator, the constant bandwidth was always smaller than or equal to the left boundary of the regression interval.

Table 1 refers only to a small part of all studies. Nevertheless, this part is exemplarily for the whole studies. In this case, the intensity of the Poisson process is $\lambda = 100$ and 1000 of its realisations are observed in the unit cube. The distance between the two cluster points is distributed according to the probability density function

$$f(r) \propto 1_{[0,R]}(r) \cdot \sqrt{r},$$

where $1_{[0,R]}(r)$ is the indicator function of the interval $[0, R]$ and R , the cluster diameter, varies from 0.01 to 0.1. This leads to the two-point correlation function

$$\xi(r) \propto 1_{[0,R]}(r) \cdot r^{-1.5},$$

thus $\gamma = 1.5$. Each regression is based on estimates of $\xi(r)$ at 100 equidistant positions $\ln r = \ln R/10, \dots, \ln R/2$. The used bandwidths are the adaptive bandwidth $h(r) = r$ and the constant bandwidths $h = R/50, R/20, R/10$.

The result of the studies is unequivocal. Table 1 lists the mean value, the standard deviation and the mean error (the square root of the mean squared error $\sqrt{\mathbf{E}(\hat{\gamma} - \gamma)^2}$)

Table 1. Comparision of two different kinds of bandwidths when estimating γ : the adaptive bandwidth $h(r) = r$ (first column) and constant bandwidth $h = R/50, R/20, R/10$ (other columns). In all cases, the variance as well as the mean error (square root of the MSE) is minimal for the adaptive bandwidth.

$R = 0.1$	adap.	$R/50$	$R/20$	$R/10$
$E\hat{\gamma}$	1.40	1.43	1.48	1.47
Std. dev.	0.24	0.68	0.51	0.39
$\sqrt{\text{MSE}}$	0.26	0.68	0.51	0.39

$R = 0.03$	adap.	$R/50$	$R/20$	$R/10$
$E\hat{\gamma}$	1.45	0.97	1.42	1.45
Std. dev.	0.22	1.25	0.53	0.33
$\sqrt{\text{MSE}}$	0.22	1.36	0.54	0.33

$R = 0.01$	adap.	$R/50$	$R/20$	$R/10$
$E\hat{\gamma}$	1.46	0.76	1.43	1.45
Std. dev.	0.22	1.74	0.62	0.33
$\sqrt{\text{MSE}}$	0.22	1.89	0.62	0.33

of the estimator $\hat{\gamma}$. As predicted in the preceding discussion, the adaptive bandwidth leads always to a considerably smaller estimation variance than the constant bandwidth. Correspondingly, also the mean error is always minimal for the adaptive bandwidth.

4. Conclusions

1. This paper is based on the empirical evidence that the two-point correlation function $\xi(r)$ of the galaxy point field follows for small r a power-law $\xi(r) \propto r^{-\gamma}$. The complex estimation procedure for γ has been analysed.

It has been shown that a constant bandwidth leads to an approximately unbiased estimator for γ . The same is true for the adaptive bandwidth $h(r) = r$, even though this bandwidth leads to a biased estimator for $\xi(r)$.

2. In case of the constant bandwidth the unbiasedness is ensured only if h is considerably smaller than r . An unwished consequence of this condition is that only a relatively small number of point pairs are taken into account when estimating γ . In contrast, using the adaptive bandwidth, the estimation of γ is based on considerably more point pairs. Thus, one can assume that the estimation is more robust when using the adaptive bandwidth.
3. Several simulation studies support this assumption. They have shown that a constant bandwidth leads to a higher estimation variance and consequently to a higher mean error than the adaptive bandwidth. Thus, the adaptive bandwidth is absolutely recommended for the estimation of γ .

Acknowledgements. I am grateful to Dietrich Stoyan, Martin Kerscher and Vicent J. Martínez for interesting and helpful discussions. I would like to thank the referee Jon Loveday for his helpful comments.

References

- Davis, M., & Peebles, P. 1983, ApJ, 267, 465
 Kerscher, M. 1999, A&A, 343, 333
 König, D., & Schmidt, V. 1992, Zufällige Punktprozesse (Stuttgart: B G Teubner)
 Landy, S., & Szalay, A. 1993, ApJ, 412, 64
 Ohser, J. 1991, Res. Inform., 4, 147
 Peebles, P. 1980, The large scale structure of the universe (Princeton: Princeton Univ. Press)
 Peebles, P., & Hauser, M. 1974, ApJS, 28, 19
 Pons-Bordería, M.-J., Martínez, V., Stoyan, D., Stoyan, H., & Saar, E. 1999, ApJ, 523, 480
 Stoyan, D. 1994, Statistics, 25, 267
 Stoyan, D., Kendall, W., & Mecke, J. 1995, Stochastic geometry and its applications (Chichester: J. Wiley & Sons)