

Chapter 14

Geographic Dependency of Population Distribution

Shouji Fujimoto, Takayuki Mizuno, Takaaki Ohnishi, Chihiro Shimizu,
and Tsutomu Watanabe

Abstract The agglomeration effect of population, which explains why many people live near places where many other people also live, is one important interaction that influences human population. We examine the agglomeration effect by measuring the distribution of the logarithmic differences between populations living in two places separated by some distance. The shapes of the distributions of the logarithmic differences closely resemble each other without depending on the regions or the country in cases of small scale of separation distance. This result suggests a unified explanation to understand the population distributions of various regions.

S. Fujimoto (✉)

Faculty of Business Administration and Information Science, Kanazawa Gakuin University,
Kanazawa, Ishikawa, Japan
e-mail: fujimoto@kanazawa-gu.ac.jp

T. Mizuno

Department of Informatics, National Institute of Informatics, Chiyoda-ku, Tokyo, Japan
Graduate University for Advanced Studies, Tokyo, Japan

The Canon Institute for Global Studies, Marunouchi, Chiyoda-ku, Tokyo, Japan
e-mail: mizuno@nii.ac.jp

T. Ohnishi

Graduate School of Information Science and Technology, The University of Tokyo, Bunkyo-ku,
Tokyo, Japan

The Canon Institute for Global Studies, Chiyoda-ku, Tokyo, Japan
e-mail: ohnishi.takaaki@i.u-tokyo.ac.jp

C. Shimizu

Institute of Real Estate Studies, National University of Singapore, Singapore
e-mail: cshimizu@nus.edu.sg

T. Watanabe

Graduate School of Economics, The University of Tokyo, Bunkyo-ku, Tokyo, Japan

The Canon Institute for Global Studies, Chiyoda-ku, Tokyo, Japan
e-mail: watanabe@e.u-tokyo.ac.jp

© The Author(s) 2015

H. Takayasu et al. (eds.), *Proceedings of the International Conference on Social Modeling and Simulation, plus Econophysics Colloquium 2014*, Springer
Proceedings in Complexity, DOI 10.1007/978-3-319-20591-5_14

14.1 Introduction

Population distribution has been studied for many decades. Zipf's law [1], which argues that the size distribution of a city's population is a power-law, is known well [2–6]. However, a problem exists: how to define the area of cities when we observe population distributions. The tail of a power-law distribution is composed of megacities. By dividing megacities into several smaller cities, the distribution's tail becomes thin. Because of the different definitions of a city, population distribution is not a power-law distribution but a log-normal one [7–9]. City areas have been decided by geographical, historical, and administrative factors. Rozenfeld et al. proposed a method that decided a city's area by a city clustering algorithm [10]. In this research, we divide spatial regions by a method that ignores the shape of cities to find the properties of population distribution that do not depend on countries or local regions.

We investigated population distribution using a spatial division method by identically sized squares. This approach resembles a previous method [9]. In our case, we control the scale of the spatial division by changing the size of the squares and clarify the universal properties concerned with population agglomeration. Population's universal properties can be observed by changing the scale of the spatial division.

We introduce logarithmic differences between the nearest neighbor two square blocks in terms of population. The regional dependence of these values in terms of the shape of the distributions vanishes for small size scales. The property of the distribution of logarithmic differences is concerned with the correlation coefficient of the population in two squares. This correlation is one index to measure population agglomeration.

In this research, we investigate Japanese population data. In Sect. 14.2, we introduce eight regions to investigate local properties inside Japan. In Sect. 14.3, we compare several distributions concerned with population among these eight regions. Next we compare Japan and Europe in Sect. 14.5 and show the universal properties concerned with population in both cases.

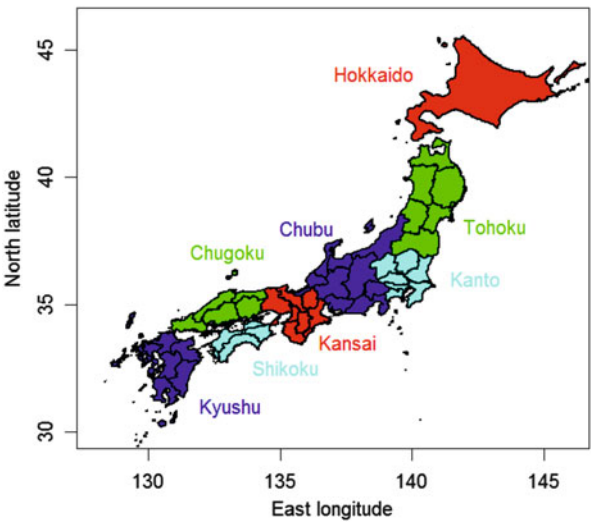
14.2 Basic Information of Japanese Population

The Statistics Bureau of the Japanese Ministry of Internal Affairs and Communications conducts a census every five years. Much census data can be obtained in a mesh data format from its websites [11], including population data from 2000, 2005, and 2010. The mesh data are raster data that are obtained by equally dividing latitudes and longitudes. A mesh size of about $500 \times 500 \text{ m}^2$ provides the highest accuracy for population. Mesh codes are assigned to each bit of data, and we can specify the data's position on the map from this code.

Table 14.1 Land use code assignment

Code : land use description	Code : land use description
1 : Paddy fields	A : Other land
2 : Other agricultural land	B : Rivers and lakes
5 : Forest	E : Beach
6 : Wasteland	F : Body of seawater
7 : Land for building	G : Golf course
9 : Trunk transportation land	0 : outside of the analysis range

Fig. 14.1 Map of the eight regions of Japan. From north to south: Hokkaido (red), Tohoku (green), Kanto (cyan), Chubu (blue), Kansai (red), Chugoku (green), Shikoku (cyan) and Kyushu (blue). Black lines indicate prefecture borders



The Japanese Ministry of Land, Infrastructure, Transport and Tourism provides land use data on its website [12]. Such data are also provided in a mesh data format. A mesh of about 100 [m]×100 [m] provides the highest accuracy for land use. In these data, a land use code (see Table 14.1) is assigned to each mesh. An inhabitable place is defined as any place that is fit for humans to live in. Inhabitable areas can be estimated by subtracting such uninhabitable areas as forests and lakes from the land area. We estimated the inhabitable areas by totaling the areas whose land use codes are 1, 2, 7, 9, A, and G. Only about 33 % of Japan’s land area is inhabitable because it has many mountainous areas. This percentage is smaller than European countries. For example, the inhabitable area percentages of Germany, France, and the United Kingdom are 68 %, 71 %, and 88 %, respectively. We have to use inhabitable areas instead of land areas to more precisely evaluate population density.

To investigate locality and universality, we divided Japan into the following eight regions based on traditional ways of combining several prefectures (see Fig. 14.1): Hokkaido, Tohoku, Kanto, Chubu, Kansai, Chugoku, Shikoku, and Kyushu. Table 14.2 shows the basic information of the eight regions. Population densities depend on the regions for various reasons.

Table 14.2 Basic information of eight Japanese regions

Region	Population in 2010	Land area km ²	Inhabitable area km ²	Density /km ²
All JP data	128,045,367	372, 907	121,941	1050.1
Hokkaido	5,506,197	83, 456	27,046	203.6
Tohoku	9,337,024	66, 890	20,306	459.8
Kanto	42,608,322	32, 424	18,256	2333.9
Chubu	21,721,795	66, 799	22,970	945.7
Kansai	22,758,142	33, 118	8,495	2679.0
Chugoku	7,563,164	31, 920	8,427	897.5
Shikoku	3,977,562	18, 805	4,860	818.4
Kyushu	4,589,693	44, 455	16,542	882.0

Each region’s population density is estimated by population per inhabitable area

14.3 Population Distribution in Japan

How to divide space is critical when examining population’s size distribution. Dividing space by municipal level is standard for investigating the size distributions of cities. In this study we do not use such spatial division method. We adopted square blocks of the same size as a spatial division method and divided a particular region into identical sized square lattices. Then we aggregated the population inside the square blocks and observed its population distribution. We can control the spatial division’s scale using this method. We use parameter BS [km], which denotes the size of one side of the square blocks.

Figure 14.2 shows a complementary cumulative distribution function (CCDF)

$$\Pr\{X \geq x\} \tag{14.1}$$

of Japan’s population in 2010. The distributions of the regions with high-density populations such as Kanto and Kansai are plotted on the right side compared to other regions. The distributions of the regions with low-density populations such as Hokkaido are plotted on the left side compared to other regions. These properties denote the distribution locality. The population distributions vary by region.

To find the distribution quantities that do not depend on the region, we focused on the population distribution’s shape. For a small scale (BS = 0.5 [km]), the right tail of the distributions rapidly falls. As BS becomes larger, the right tail of the distributions becomes gentler. The slopes of the right tail seem close to each other for a small BS. The value of the logarithmic differences between populations whose values are close to each other seems to share similar quantities of population distribution slopes.¹

¹It is possible to confirm of this expectation by left figure of Fig. 14.6.

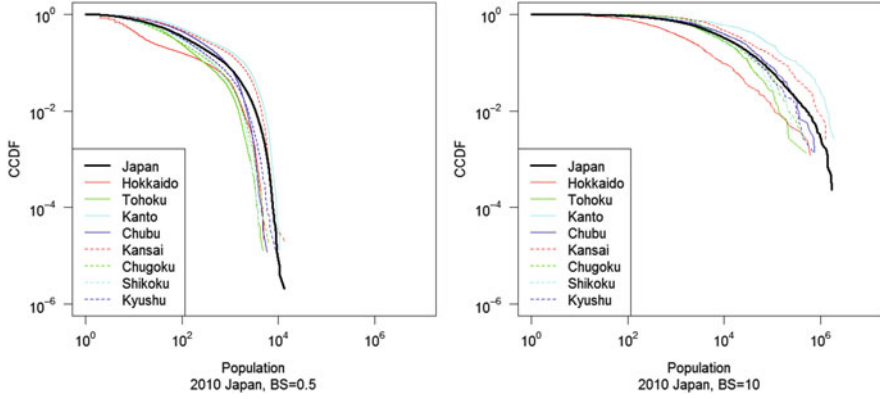


Fig. 14.2 Log-log plot of population distributions. *Left figure* is for $BS = 0.5$ [km]. *Right figure* is for $BS = 10$ [km]. Thick black curves show all Japanese distributions in 2010. Thin colored curves show all distributions of eight regions in 2010

We use $S(x, y)$ to denote the population inside a square whose vertex coordinates are (x, y) , $(x + BS, y)$, $(x + BS, y + BS)$, and $(x, y + BS)$. The logarithmic difference between the populations of nearest neighbors in x -direction is represented by

$$\ln S(x + BS, y) - \ln S(x, y), \quad (14.2)$$

and the logarithmic difference in y -direction is represented by

$$\ln S(x, y + BS) - \ln S(x, y). \quad (14.3)$$

The logarithmic difference is a value that is frequently used in such time-series analyses as stock prices [13]. In this paper we apply it to spatial directions. The effects of the differences are the same regardless whether the difference direction is positive or negative in terms of the spatial direction. Next we investigate the distributions of the absolute value of the logarithmic differences.

Figure 14.3 shows the CCDF of the absolute value of the logarithmic differences between the nearest neighbor populations in Japan in 2010. For small scale ($BS = 0.5$ [km]), the distributions almost overlap. As BS becomes larger, the right tail of the distributions becomes gentler, and they no longer overlap.

Figure 14.4 shows the BS dependence of the moments values of the distributions of absolute value of logarithmic differences. Where n -th order moments is defined by mean of n -th powered of the stochastic variable. These values are one of the quantitative index of the overlapping of the distributions.

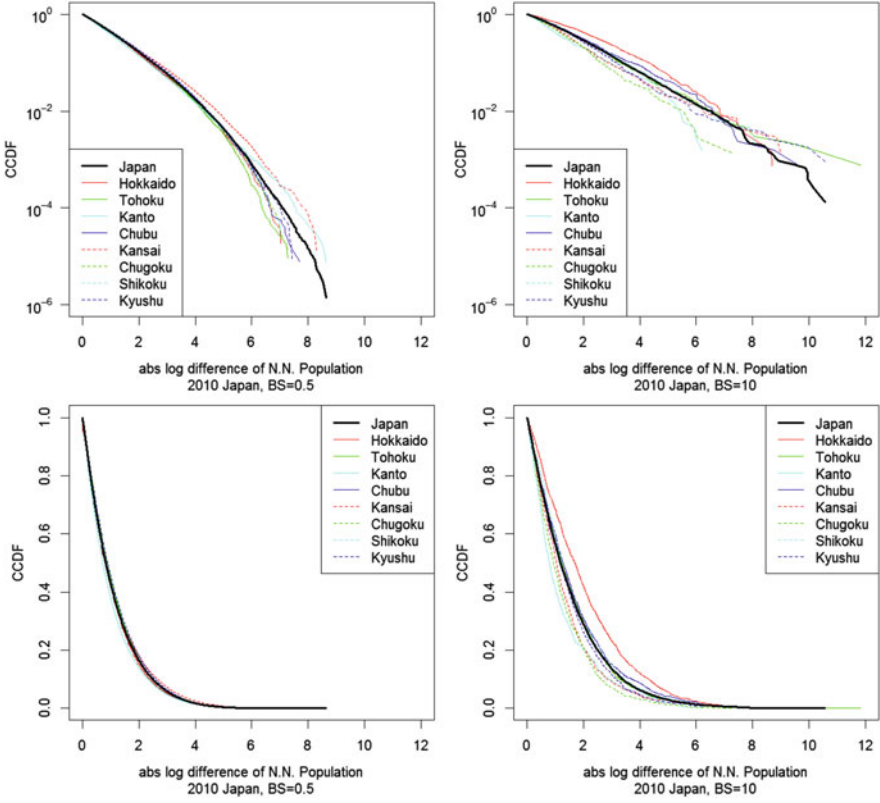


Fig. 14.3 Distributions of absolute value of logarithmic differences between nearest neighbor populations. *Top figures* are semi-log plot. *Bottom figures* are non-log plot. *Left figures* are for $BS = 0.5$ [km]. *Right figures* are for $BS = 10$ [km]. *Thick black curves* show all Japanese distribution in 2010. *Thin colored curves* show distributions of all eight regions in 2010

Figure 14.5 compares the observed distribution and the distributions represented by analytic functions. The red lines show an exponential distribution whose CCDF is defined by

$$\Pr\{X \geq x\} = \int_x^\infty \frac{1}{\mu} \exp\left(-\frac{t}{\mu}\right) dt. \quad (14.4)$$

Here parameter μ is the distribution's mean. The estimated values from the data are $\mu = 1.1022$ for $BS = 0.5$ and $\mu = 1.5629$ for $BS = 10$. The blue curves show truncated normal distribution, whose CCDF is defined by

$$\Pr\{X \geq x\} = \int_x^\infty \sqrt{\frac{2}{\pi\sigma^2}} \exp\left(-\frac{t^2}{2\sigma^2}\right) dt. \quad (14.5)$$

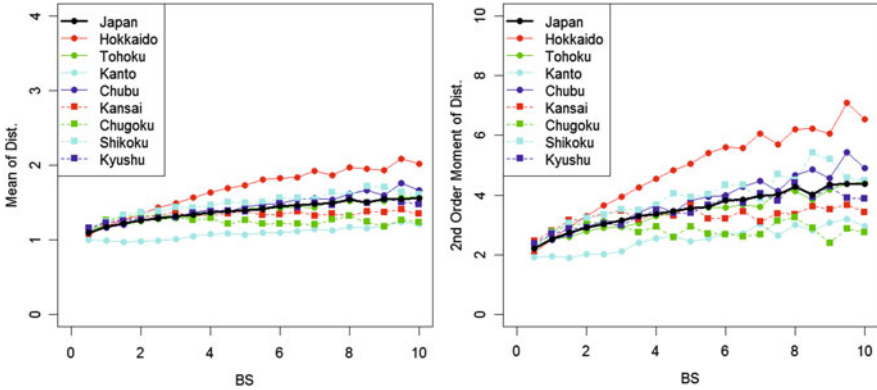


Fig. 14.4 BS dependence of the moments values of the distributions of absolute value of logarithmic differences. *Left figure* shows the 1st order moments. *Right figure* shows the 2nd order moments. *Black symbols* show the moments values of the Japanese distributions. *Colored symbols* show the moments values of the distributions of eight regions

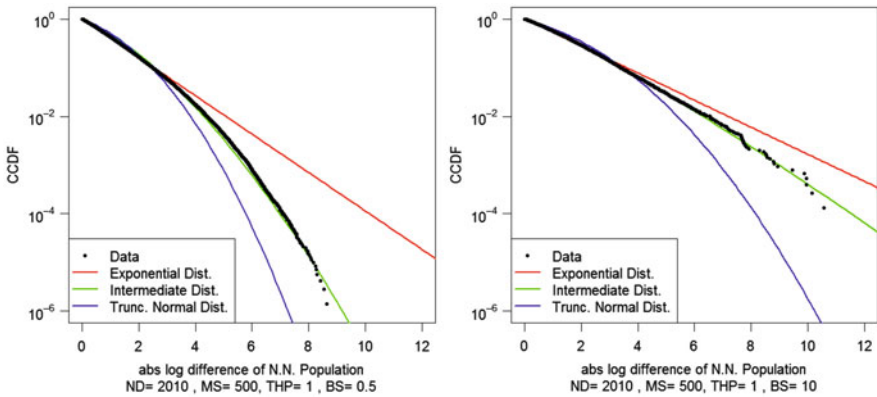


Fig. 14.5 Distributions of absolute value of logarithmic differences between nearest neighbor populations. *Left figure* is for $BS = 0.5$ [km]. *Right figure* is for $BS = 10$ [km]. *Black circles* show distributions observed from all Japanese data in 2010. *Red lines* show exponential distributions whose means match observed data. *Blue curves* show truncated normal distributions whose standard deviations match observed data. *Green curves* show intermediate distributions between *red curves* and *blue curves*

Here parameter σ is the standard deviation from the $x = 0$ of the distribution. The estimated values from the data are $\sigma = 1.4853$ for $BS = 0.5$ and $\sigma = 2.0944$ for $BS = 10$. The shape of the distributions seems to be intermediate between the exponential and the truncated normal distributions. The distributions resemble a truncated normal distribution in a small BS scale. As BS becomes larger, the

distribution becomes an exponential distribution. Intermediate distribution between Eq. (14.4) and Eq. (14.5) is represented by

$$\Pr\{X \geq x\} = \int_x^\infty \frac{\alpha}{\lambda \Gamma(\frac{1}{\alpha})} \exp\left(-\frac{t^\alpha}{\lambda^\alpha}\right) dt. \quad (14.6)$$

Where α is a shape parameter and λ is a scale parameter. If $\alpha = 1$, Eq. (14.6) corresponds to Eq. (14.4). If $\alpha = 2$, Eq. (14.6) corresponds to Eq. (14.5). The green curves in Fig. 14.5 show distributions of Eq. (14.6). We selected the parameters $\alpha = 1.6, \lambda = 0.9$ for $BS = 0.5$ and $\alpha = 1.2, \lambda = 0.9$ for $BS = 10$.

The shape of the distributions of the logarithmic differences of two values is concerned with the correlation between those two values. The left side of Fig. 14.6 shows a scatter plot of $\ln S(x, y)$ versus $\ln S(x + BS, y)$ or $\ln S(x, y + BS)$. From this figure, we observe agglomeration effect that many people live near places where many other people also live. The correlation coefficient is able to interpret as an index of agglomeration effect. The right side figure's data are transformed from the left side figure's data by dilating both axis data $\sqrt{2}$ and rotating clockwise 45° . The horizontal axis of the right side figure is the logarithmic summation between the nearest neighbor populations. The vertical axis of the right side figure is the logarithmic difference between the nearest neighbor populations. The red bars are the standard deviation inside each segment, which is equally divided by the horizontal axis. The correlation of the left side figure represents the correlation between the population and the nearest neighbor population. If this correlation is strong, the population near the large population is large. It is considered that the strengthen of this correlation is one of the indices which represents degree of the agglomeration of population. The deviation of the distribution of the vertical axis of

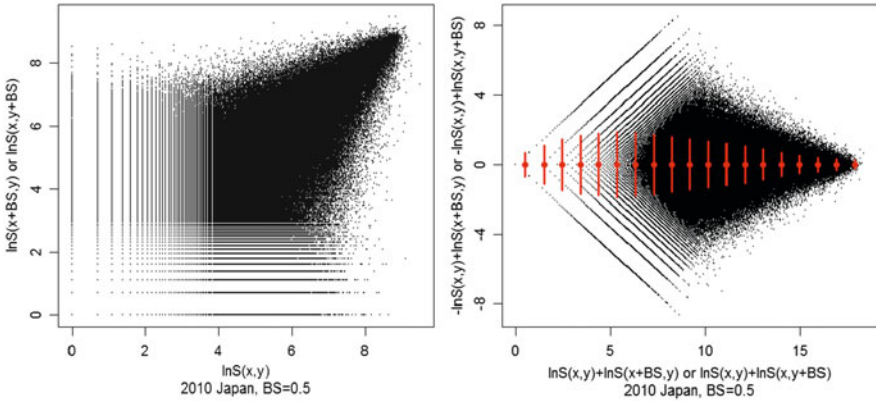


Fig. 14.6 Left side figure shows scatter plot of $\ln S(x, y)$ versus $\ln S(x + BS, y)$ or $\ln S(x, y + BS)$. Correlation coefficient of these data is 0.69. Right side figure's data are transformed from left side figure's data by expansion and rotation. Red circles are means inside each segment that is equally divided by horizontal axis. Red bars are standard deviation inside each segment

Table 14.3 Basic information of top seven European countries by population

Region	Population in 2011	Land area km ²	Inhabitable area km ²	Density /km ²
All EU data	514,988,832			
Germany	80,122,036	348,560	237,800	336.9
France	62,623,425	547,557	387,537	161.6
U.K.	62,583,331	241,930	213,048	293.8
Italy	59,315,222	294,140	201,870	293.8
Spain	46,802,562	498,800	315,307	148.4
Poland	38,449,414	306,230	212,586	180.9
Romania	16,609,793	230,170	164,076	101.2

Population density in each region is estimated by population per inhabitable area

the right side figure concerns the correlation of the left side figure. The deviation of the distribution of the vertical axis of the right side figure shrinks when the correlation of the left side figure becomes strong. It is possible to estimate the degree of agglomeration of the population by observing the deviation of the distribution of the logarithmic difference.

14.4 Basic Information of European Populations

The European Union provides several kinds of statistical data from eurostat. The GEOSTAT project provides European countries' population dataset representing in a 1 km² grid dataset. Population data for 2006 and 2011 are available on their website [14].

The food and agriculture organization of the United Nations statistics division (FOSTAT) [15] provides land and forest area data from most countries. We can roughly estimate the inhabitable areas by subtracting forest areas from land areas.

Table 14.3 shows the basic information of the top seven European countries by population. Their population density is lower than Japan. The variation of the population density of each country is smaller than the variation of all eight Japanese regions.

14.5 Comparison between Japan and European Countries

In this section we compare Japan and European countries in terms of the distribution of log differences of population. Figure 14.7 shows the CCDF of the absolute value of the logarithmic differences between the nearest neighbor population of Japan and EU countries. The results are almost the same as those among Japan's eight

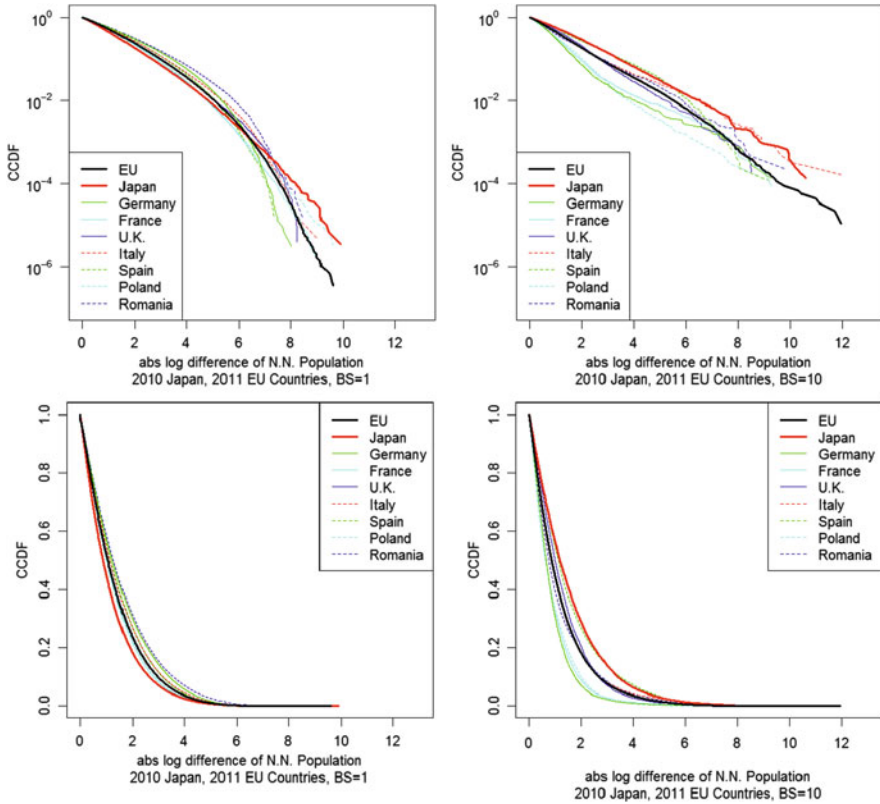


Fig. 14.7 Distributions of absolute value of logarithmic differences between nearest neighbor populations. *Top figures* are semi-log plot. *Bottom figures* are non-log plot. *Left figures* are for $BS = 0.5$. *Right figures* are for $BS = 10$. *Thick black curves* show all EU distributions in 2011. *Thick red curves* show Japanese distribution in 2010. *Thin colored curves* show all seven European countries' distributions in 2011

regions. As BS becomes larger, the right tail of the distributions becomes gentler. The overlapping of the distributions for $BS = 1$ is better than for $BS = 10$. If we observed data whose scale $BS = 0.5$, the overlapping would be better than for $BS = 1$.

The transitions of the distributions due to changes by BS are shown in Fig. 14.8. Japan's distribution shape is almost the same as that of EU at a small BS . The difference of Japan and EU becomes larger as BS increases.

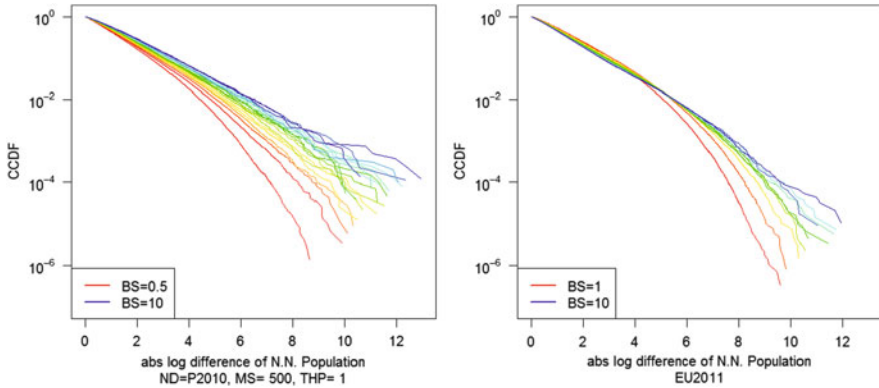


Fig. 14.8 Distributions of absolute value of logarithmic differences between nearest neighbor populations. Color gradation of curves represents size of BS. Red, green, and blue curves show small, intermediate, and large sizes, respectively. Left figure shows distributions of Japan with BS from 0.5 to 10 by 0.5 increments. Right figure shows distribution of all EU with BS from 1 to 10 by 1

14.6 Conclusion

We investigated population distributions using Japan and EU data. Using a spatial division method with same size squares, we can easily control the division scale. The shape of the population distribution differs by country or region. We introduce logarithmic differences between nearest neighbor populations to identify distributions that do not depend on country or region. When the division scale is large, the distribution of logarithmic differences depends on the country or region. The local dependence of the distribution disappears as the division scale becomes smaller. The distribution's shape closely resembles a normal distribution when the division scale is small; it is close to exponential distribution when the division scale is large.

This study investigated population distributions from a universal standpoint that does not depend on country or region. In general, various interactions determine population distribution. These interactions can be divided into two types. One is internal interactions, and the other is external interactions. External interactions are such environmental elements as topography and habitability. Internal interactions are interactions between people. Our results suggest that a universal feature exists for interaction with a small-scale neighboring population.

The next stage of our study will reproduce the results of Fig. 14.8 using a simple model. If we generate population data randomly, BS dependence of the shape of the distributions of logarithmic differences are quite different from Fig. 14.8. To reproduce the BS dependence of Fig. 14.8, we have to generate population configuration which satisfy the left figures of Fig. 14.6. We will have to introduce interactions between people to generate the agglomeration effect.

It would be interesting if the local features of population distribution could be explained by the interaction between people and environmental factors. We consider that the inhabitable area is most important in the environmental factor. We expect that the interaction between people and geometrical environmental factor is to be detected from relations between fluctuation of the population and the population density per inhabitable area.

Acknowledgements The authors are grateful to SMSEC 2014, where this work was completed.

Open Access This book is distributed under the terms of the Creative Commons Attribution Non-commercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Zipf G (1949) Human behavior and the principle of least-effort. Addison-Wesley, Cambridge
2. Hill BM (1970) J Am Stat Assoc 65(331):1220
3. Kruguman P (1996) J Japanese Int Economies 10:399
4. Zanette DH, Manrubia SC (1997) Phys Rev Lett 79:523
5. Marsili M, Zhang YC (1998) Phys Rev Lett 80(12):2741
6. Gabaix X (1999) Q J Econ 114:739
7. Eeckhout BJ (2004) Am Econ Rev 94(5):1429
8. Kuninaka H, Matsushita M (2008) J Phys Soc Japan 77(11):114801
9. Holmes TJ, Lee S (2009) In: Glaeser E (ed) In the economics of agglomerations, University of Chicago Press, pp 105–131
10. Rozenfeld HD, Rybski D, Gabaix X, Makse HA (2011) Am Econ Rev 101(5):2205
11. Portal site of official statistics of japan. URL <http://www.e-stat.go.jp/SG1/estat/eStatTopPortalIE.do>
12. National land numerical information download service. URL <http://nlftp.mlit.go.jp/ksj-e/gml/datalist/KsjTmplt-L03-b.html>
13. Mantegna RN, Stanley HE (2002) Nature 376:46
14. Eurostat geostat project. <http://ec.europa.eu/eurostat/web/gisco/geostat-project/>
15. FOSTAT. <http://faostat3.fao.org/download/R/RL/E>