Upper limit of the human lifetime distributions: Analysis of a list of the oldest olds.

Masaaki SIBUYA, Keio University
sibuyam@1986.jukuin.keio.ac.jp

Cherry Bud Workshop, 2007-03-13

Abstract

The generalized Pareto models are well fitted to a sequence of lists of persons whose age is 100 and over, recorded every year. The 'finite' upper limits of lifetime distributions of Japanese males and females are estimated.

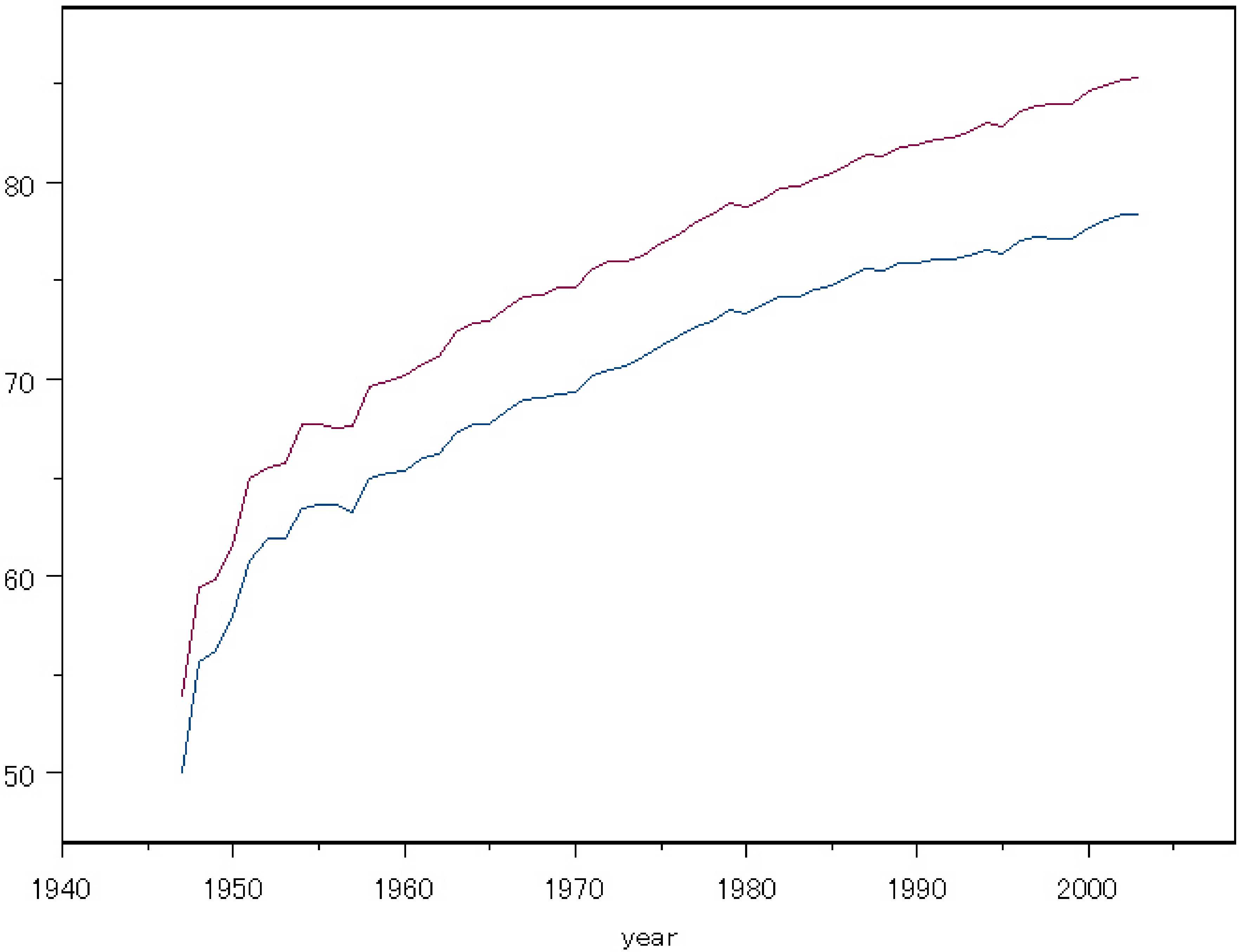
Keywords Cohort analysis, mean residual lifetime, pseudo random sample, survival statistics, tabulated data, vital statistics.

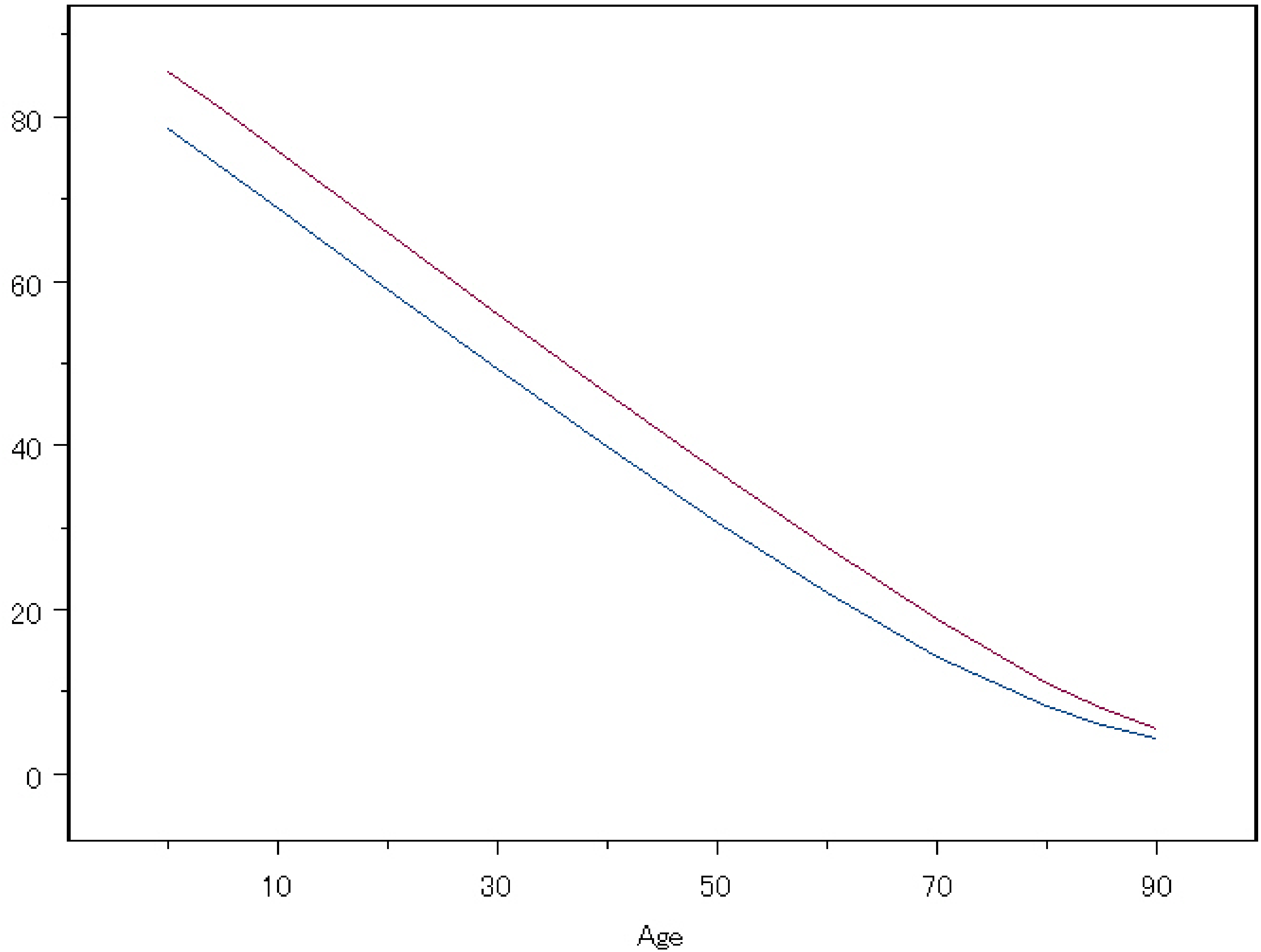
Contents

1 A List of The Oldest Olds (Background)

- Inceasing lifetime of Japanese people, causing social problems.
- What is really happening at the upper tail of lifetime distribution. (Disregarding Quality Of Life.)

- **Senior Citizen's Day**, The 3rd Monday of September, a National Holiday.
- Prime Minister sends a congratulation letter and a memorial gift to those who celebrated centennial birthday within a year.
- For that purpose Ministry of Health, Labor and Welfare (MHLW) prepare the list of the qualified people based on the reports from local goverments.
- Incidentally MHLW collects the number of peole over one hundred, and publish a list.


80
70
60
50
1940
1950
1960
1970
1980
1990
2000
year



80
60
40
20
0
10
30
50
70
90
Age

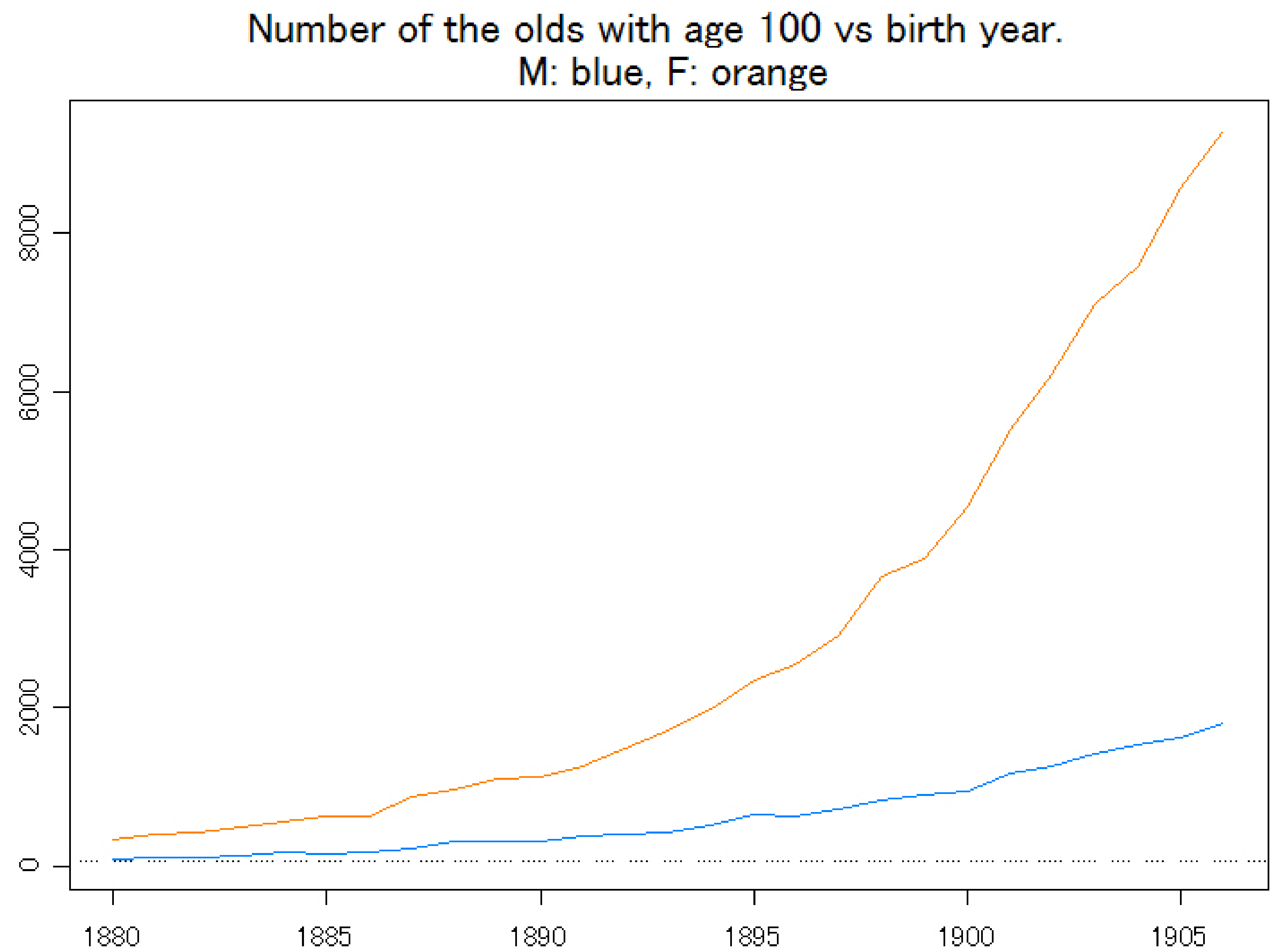

Number of the olds with age 100 vs birth year.
M: blue, F: orange
0
2000
4000
6000
8000
1880
1885
1890
1895
1900
1905

Table 1: The number of males with age 100 and over, 2002–2006.

	age																
year	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116
2002	1265	747	401	229	106	65	34	16	5	1	3	2	0	1	0	0	
2003	1420	799	457	230	119	57	46	20	10	0	0	0	0	0	1	0	
2004	1537	895	505	290	149	69	36	30	8	4	0	0	0	0	0	0	
2005	1628	991	548	295	155	87	33	19	15	5	3	0	0	0	0	0	
2006	1804	1046	619	355	161	88	44	13	11	7	1	1	0	0	0	0	

2 Generalized Pareto distributions

Extreme Value Theory

1. block maxima model: extreme value distributions.
2. threshold exceedances model (Peak Over Threshold): generalized Pareto distributions.

Embrechts, Klüppelburg and Mikosch (1998), Coles (2001), Beirlant, Goegebeur, Segers and Teugels (2004), McNeil, Frey and Embrechts (2005).

Threshold Exceedances

Generalized Pareto distributions: $\mathsf{GnPrt}\ (\gamma, a),\ \gamma \in \mathcal{R},\ a \in \mathcal{R}_+$, defined by the survival function,

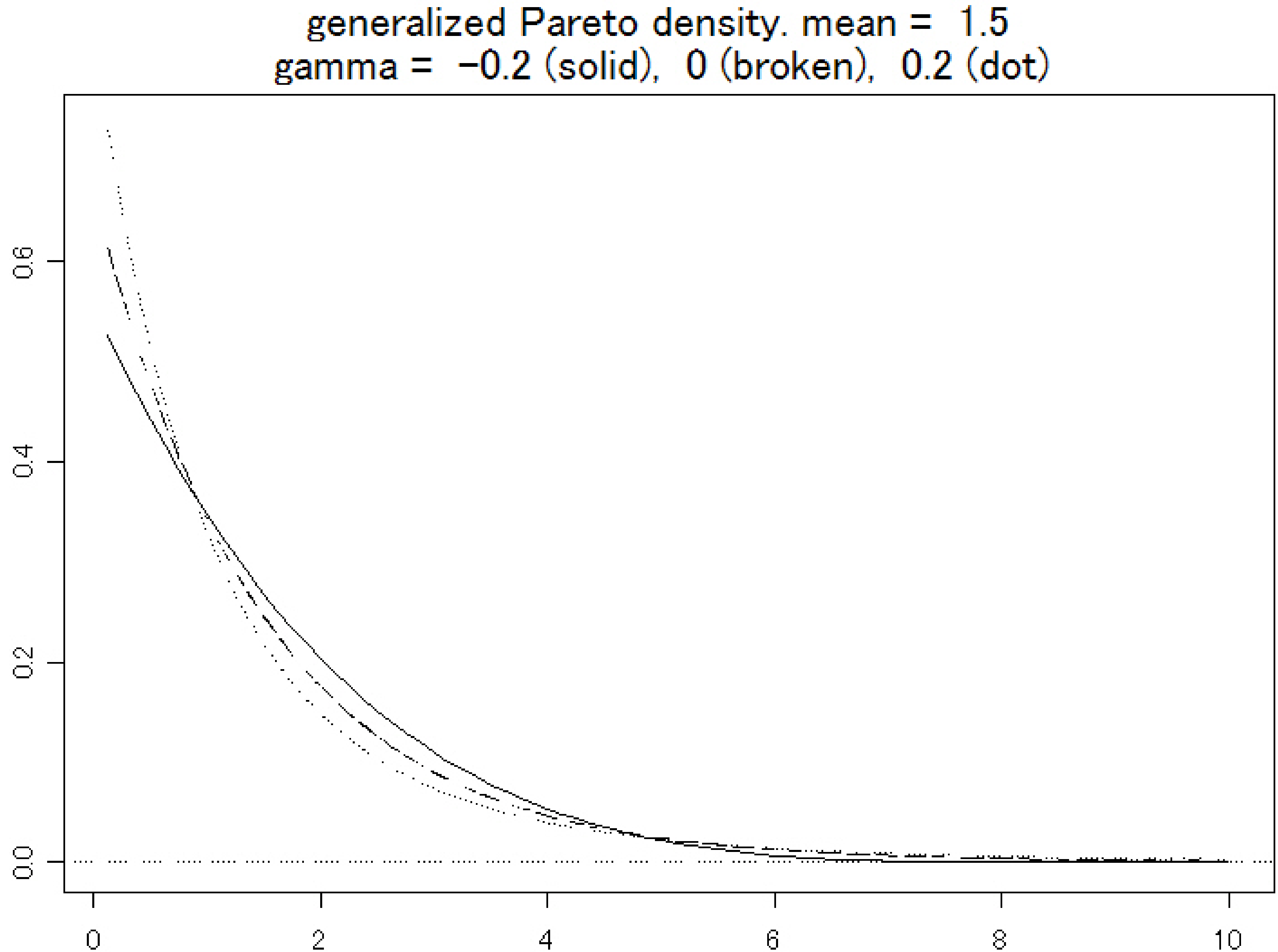
$$\overline{F}(x) = \overline{F}(x; \gamma, a) = \begin{cases} (1 + \gamma x/a)^{-1/\gamma}, & \gamma \neq 0, \\ \exp(-x/a), & \gamma = 0, \end{cases} \tag{1}$$

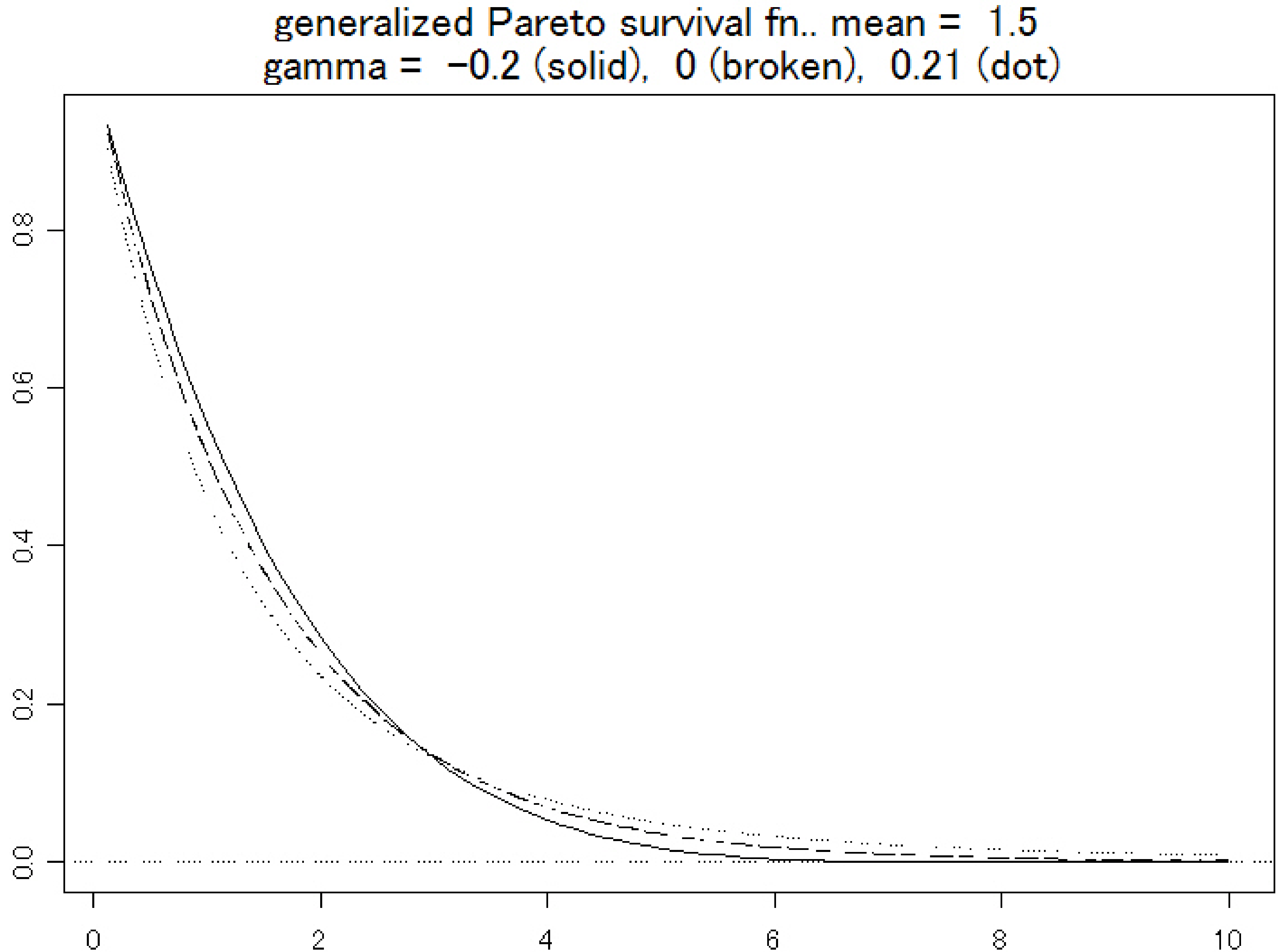
with the range

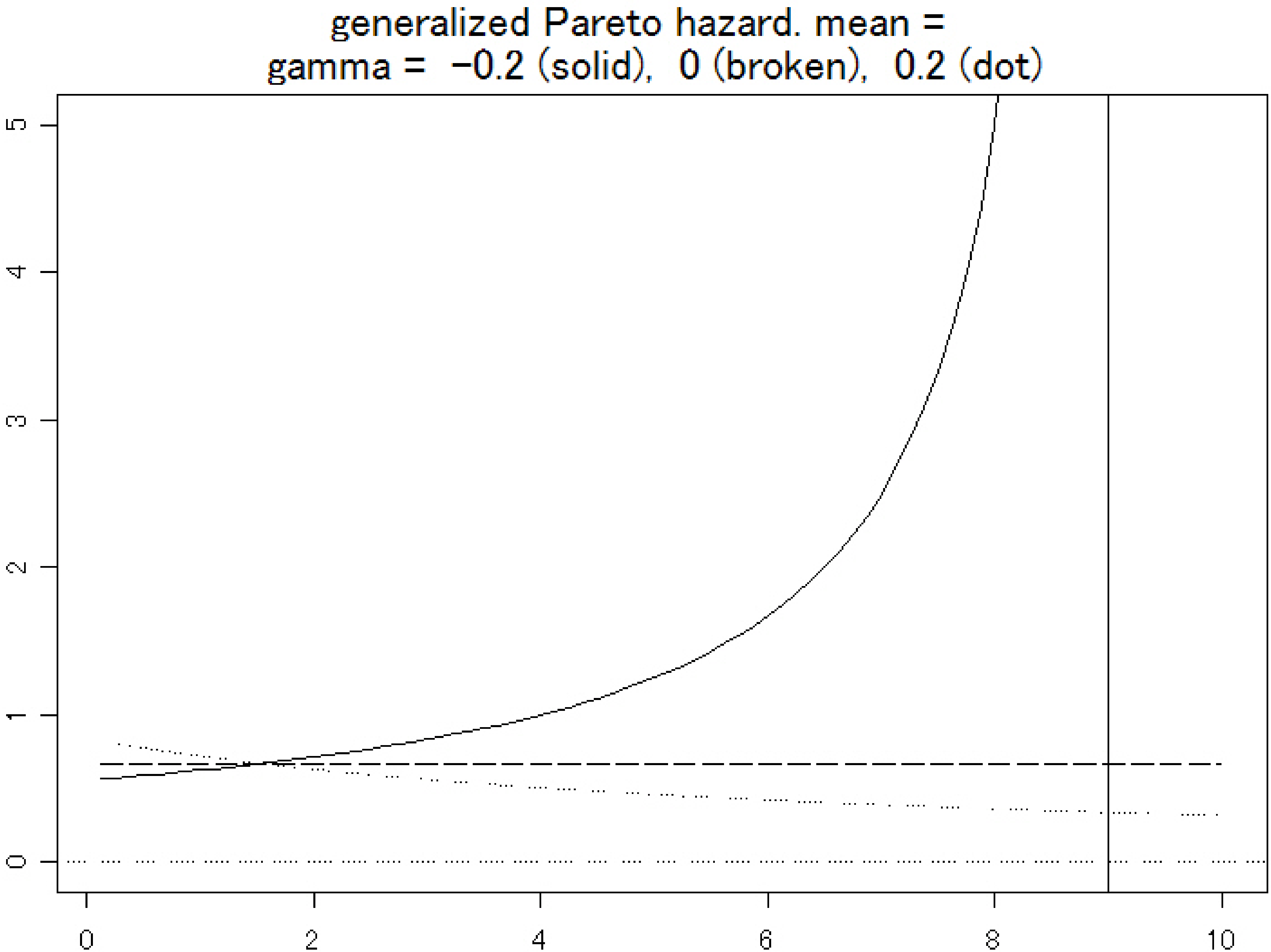
$$\begin{cases} 0 < x < \infty, & \gamma \geq 0, \\ 0 < x < a/(-\gamma) =: \omega, & \gamma < 0. \end{cases}$$

Its hazard function is

$$h(x; \gamma, a) := f(x)/\overline{F}(x) = 1/(a + \gamma x).$$


generalized Pareto density. mean = 1.5
gamma = −0.2 (solid), 0 (broken), 0.2 (dot)
0.0
0.2
0.4
0.6
0
2
4
6
8
10


generalized Pareto survival fn.. mean = 1.5
gamma = -0.2 (solid), 0 (broken), 0.21 (dot)
0.0
0.2
0.4
0.6
0.8
0
2
4
6
8
10


generalized Pareto hazard. mean =
gamma = −0.2 (solid), 0 (broken), 0.2 (dot)
0
1
2
3
4
5
0
2
4
6
8
10

Linear mean residual lifetime

$X \sim \mathsf{GnPrt}(\gamma, a) \quad \Rightarrow \quad (X-u)|(X \geq u,\ u>0) \sim \mathsf{GnPrt}(\gamma, a+\gamma u)$.

That is, only the scale changes: $a \Rightarrow a+\gamma u$. A generalization of the memoryless property of the exponential distribution.

Since $E(X) = a/(1-\gamma),\ \gamma < 1$, $E(X-u|X>u)$, the "mean residual lifetime" of X at $X = u$, is

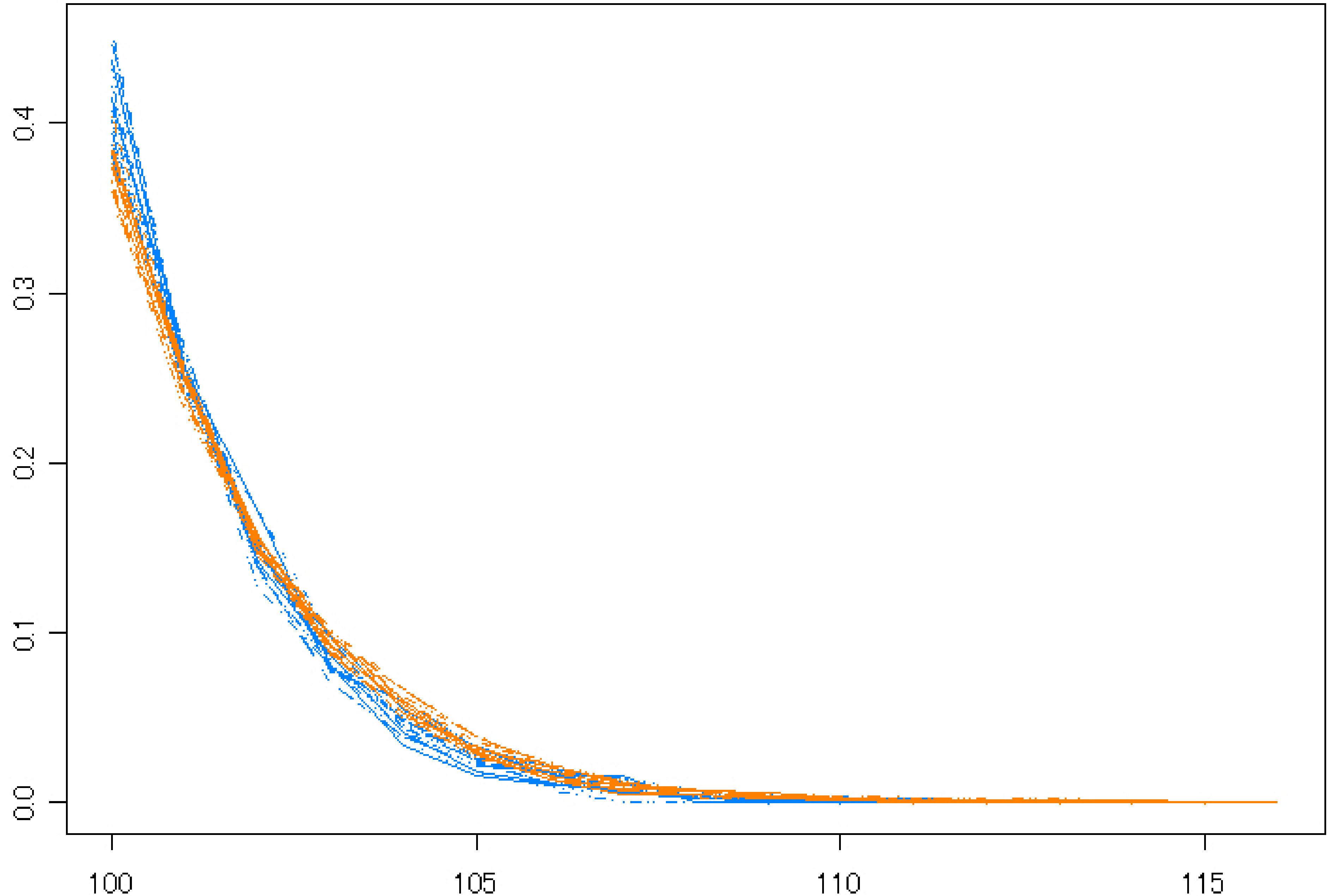
$$m(u) := E(X-u|X>u) = (a+\gamma u)/(1-\gamma), \quad \gamma < 1, \quad 0 < u,$$

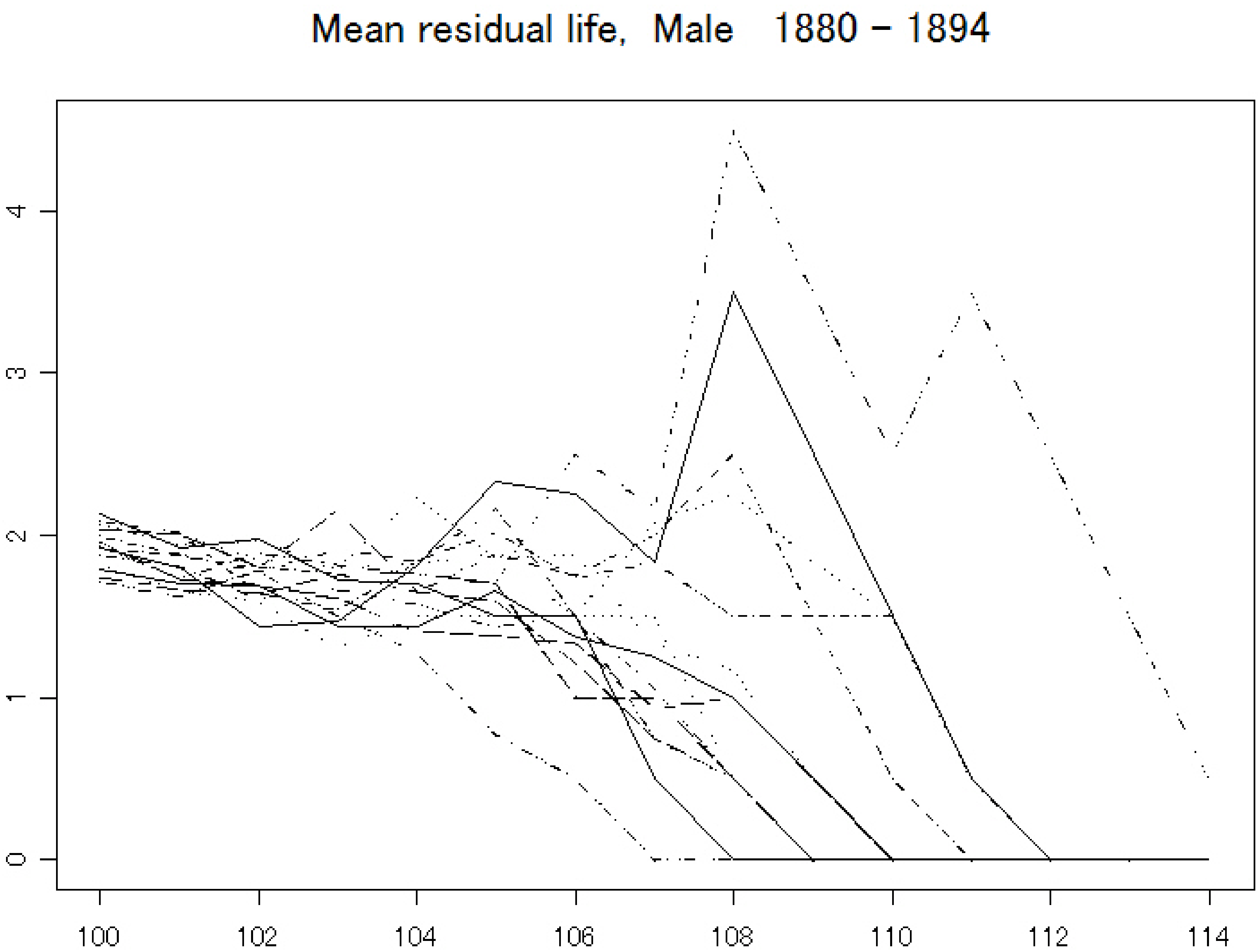
The linearity of the mean residual lifetime characterizes $\mathsf{GnPrt}(\gamma, a),\ \gamma < 1$.

$\gamma < 0$ or not, that is whether the human lifetime distribution has a finite upper limit or not.

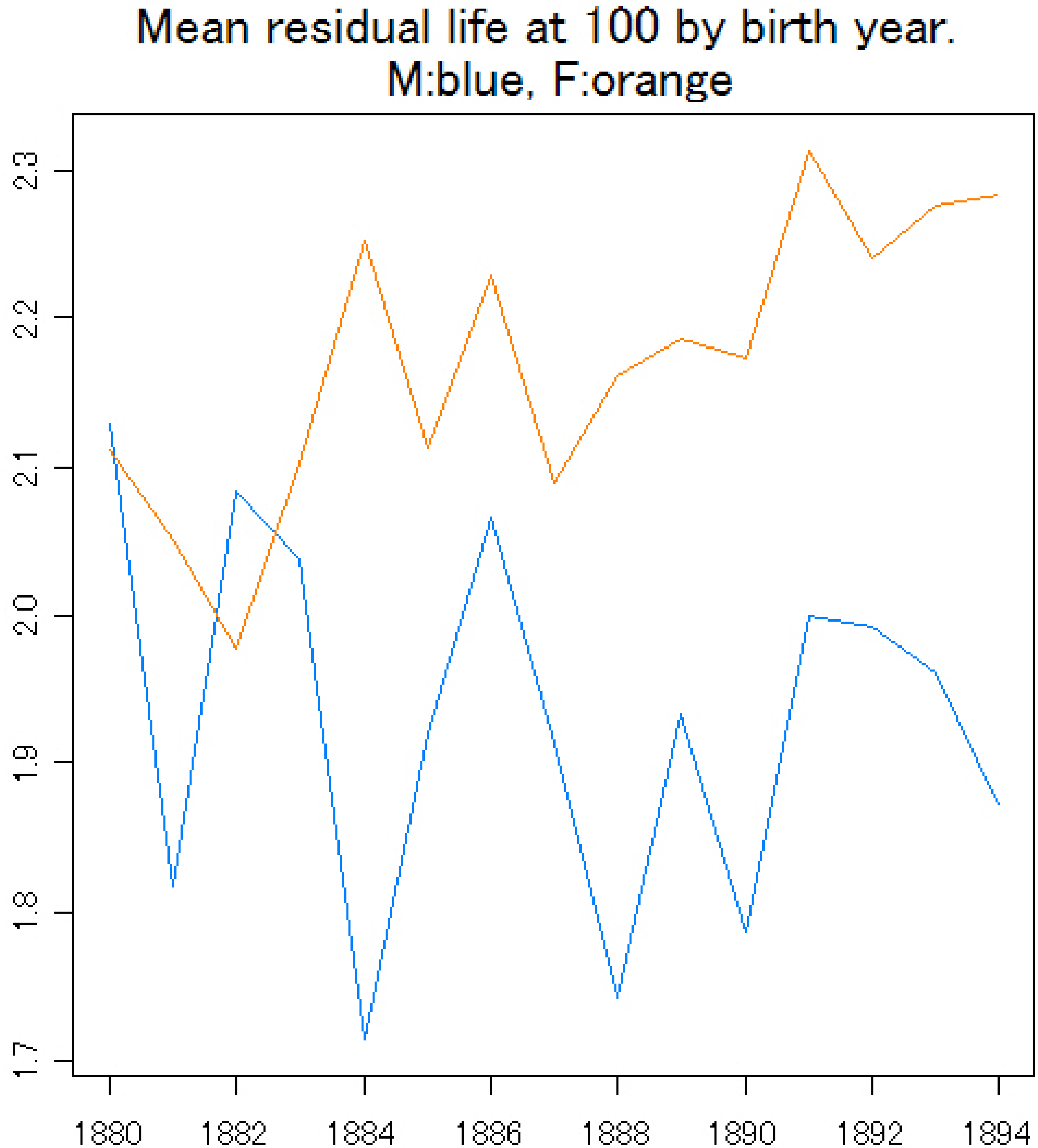
In the demography, human lifetime distributions are conventionally assumed to be Gompertz curves, with the semi-infinite range, being fitted to the whole human lifetime, or to the adult lifetime. See, e.g. Kanisto (1999).

Lifetime frequency of the group 1880 - 1894 .
M: blue, F: orange

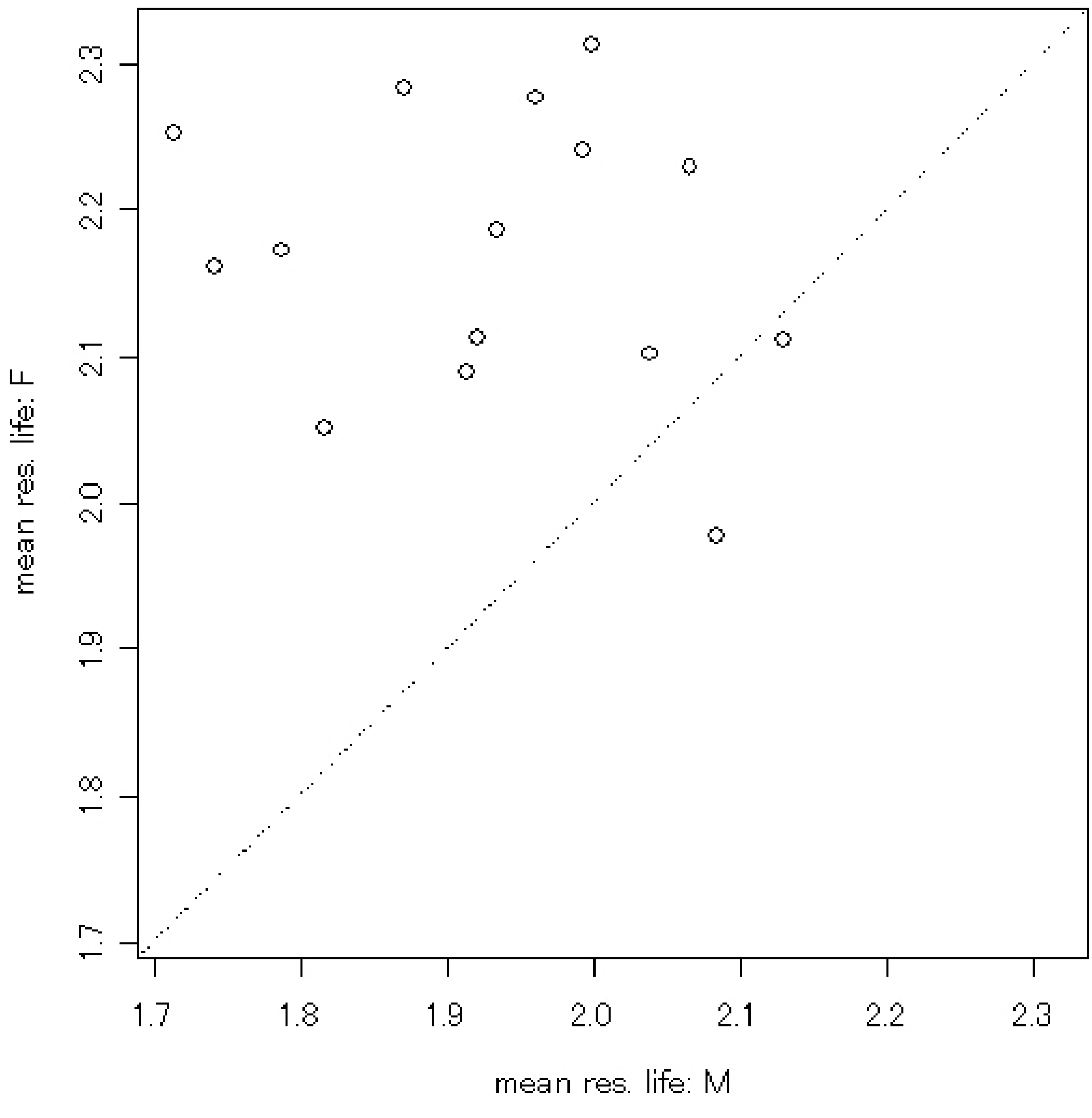



Mean residual life, Male 1880 – 1894
0
1
2
3
4
100
102
104
106
108
110
112
114

Mean residual life, Female 1880 - 1894
0
1
2
3
4
100
105
110
115


Mean residual life at 100 by birth year.
M:blue, F:orange
2.3
2.2
2.1
2.0
1.9
1.8
1.7
1880
1882
1884
1886
1888
1890
1892
1894

Mean residual life at 100 by birth year.



3 Two Types of Population Statistics

Cohort by birth and by registry

1. official statistics of deaths: vital statistics.
2. official statistics of survivors: census and our datasets. (In census, population above 85 is put in a class, “top coding” for SDC)

Our data: the number M_{ij} of the people (males or females) of the age i (in year unit), on a specific day of the year j.

V_{ij} denote the event that persons in this group will die within one year.

the number of occurrence of V_{ij}:

$$D_{ij} := M_{ij} - M_{i+1,j+1}.$$

“survival statistics” for a group,“cohort by registry”,

a group of people who were recorded for the first time in the same year, $j = k + 100$.

Observing M_{ij}, $k = j - i$, $i = 100, 101, \ldots,$ until M_{ij} vanishes, a complete set of the survival statistics of the k-th cohort is obtained.

“cohort by birth”, that is a generation.

the difference of the lifetime distributions between cohort by age and cohort by birth.

In calculating probabilities in cohort analysis, one easily makes mistakes. To avoid such mistakes the use of the classical Lexis diagram is recommended. See Keiding(1999). See Figures 1 and 2.

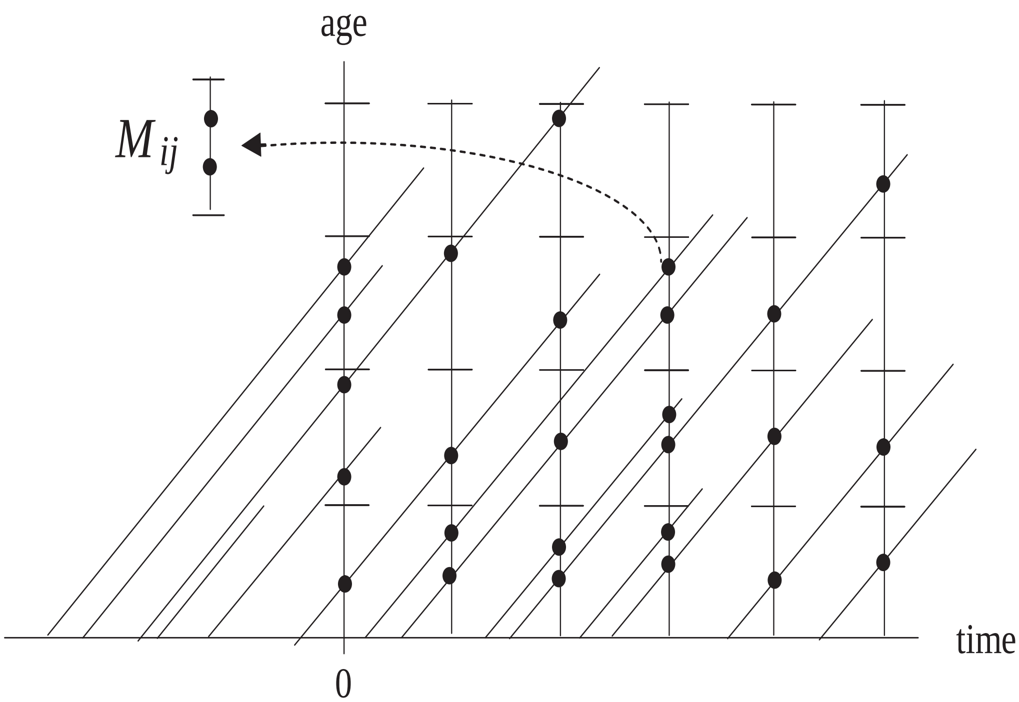


Figure 1: M_{ij} on Lexis diagram.

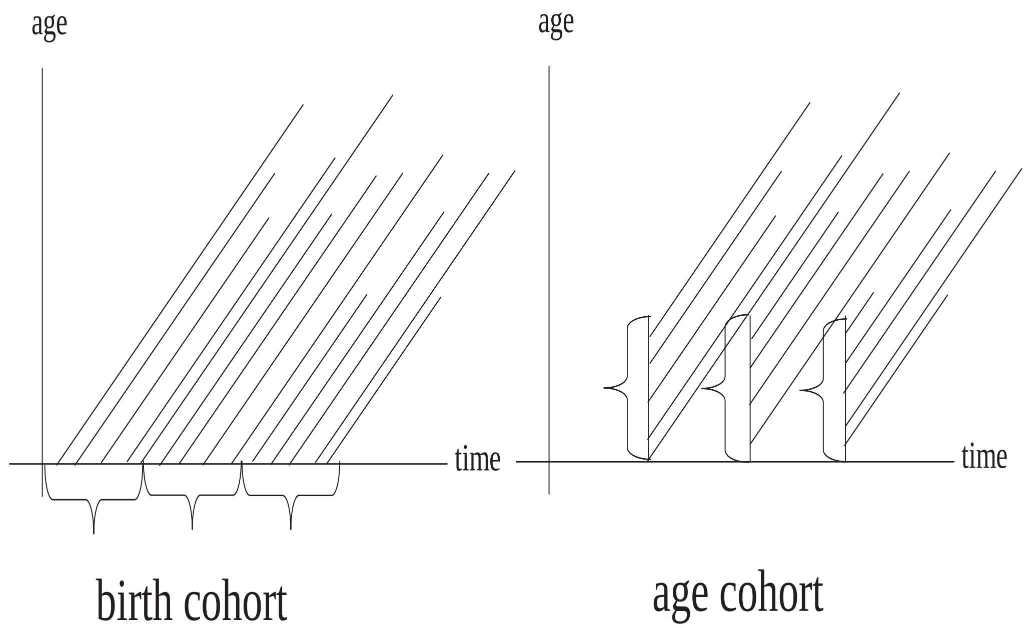


Figure 2: Cohort by birth and by age on Lexis diagram.

Cohort by registry

$$\overline{G}_0(x) := \int_0^1 \overline{F}(x-u)du, \quad \overline{F}(x) = \begin{cases} 1, \quad x < 0, \\ (1-x/\omega)^\eta, \quad 0 \le x < \omega, \ \gamma < 0, \\ 0, \quad \omega \le x, \end{cases}$$

where $\eta = -1/\gamma > 0$ and $\omega = -a/\gamma > 0$ and,

$$\begin{aligned} \overline{G}_0(x) &= \int_{\max(0,x-\omega)}^{\min(1,x)} \left(1 - \frac{x-u}{\omega}\right)^\eta du + \int_{\min(1,x)}^1 du \\ &= \begin{cases} 1 - x + \frac{\omega}{\eta+1}\left(1 - \left(1-\frac{x}{\omega}\right)^{\eta+1}\right), & 0 \le x \le 1, \\ \frac{\omega}{\eta+1}\left(\left(1-\frac{x-1}{\omega}\right)^{\eta+1} - \left(1-\frac{x}{\omega}\right)^{\eta+1}\right), & 1 < x \le \omega, \\ \frac{\omega}{\eta+1}\left(1-\frac{x-1}{\omega}\right)^{\eta+1}, & \omega < x \le \omega+1. \end{cases} \end{aligned} \tag{2}$$

Hence the survival function of cohort by age is

$$\overline{G}(x) = \overline{G}_0(x+1)/\overline{G}_0(1), \quad x > 0. \tag{3}$$

The numbers $D_{i,k-j}$ of the events $V_{i,k-j}$; $k = j - 100$; $i = 100, 101, \ldots$ are multinomially distributed with the probabilities $P(V_{ij}) = \overline{G}(i-1) - \overline{G}(i)$, $\quad i = 1, 2, \ldots,$ and the trial number $M_{100,\,j}$.

4 Estimation Methods and Conclusion

Estimation Methods

the data of people born before 1880 are not reliable: the birthdays are doubtful, (Keio University, since 1858 during the civil war for the Meiji Restoration, 1876) the data of people born from 1880 to 1894 (not the calendar years),

Method MN1 The typical estimation method is the maximum likelihood estimation of the multinomial model based on the survival function $\overline{G}$, (3), in the previous section. The estimation function is rather complicated in this model.

Method MN2 Another estimation is to approximate the probabilities of the event V_{ij} by the GnPrt distributions (1)

$$P(V_{ij}) \doteq \overline{F}(i-1) - \overline{F}(i).$$

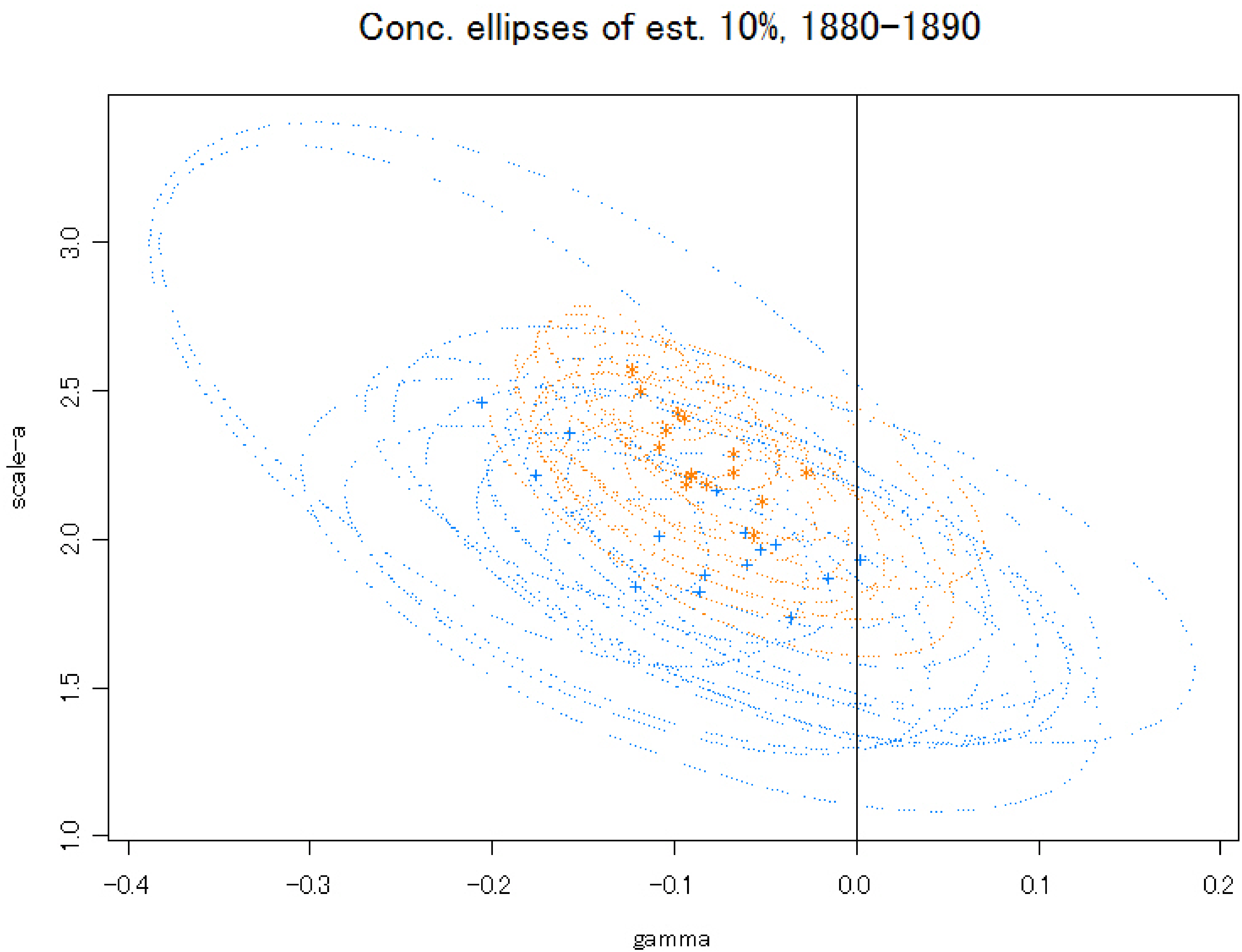
Further, a different approach is used. Similar to the birthdays, the deathdays can be assumed to be distributed uniformly within one year. The seasonal effect on deathdays is diminishing because the living environment is becoming comfortable, and thus can be neglected.

Estimation Methods (continued)

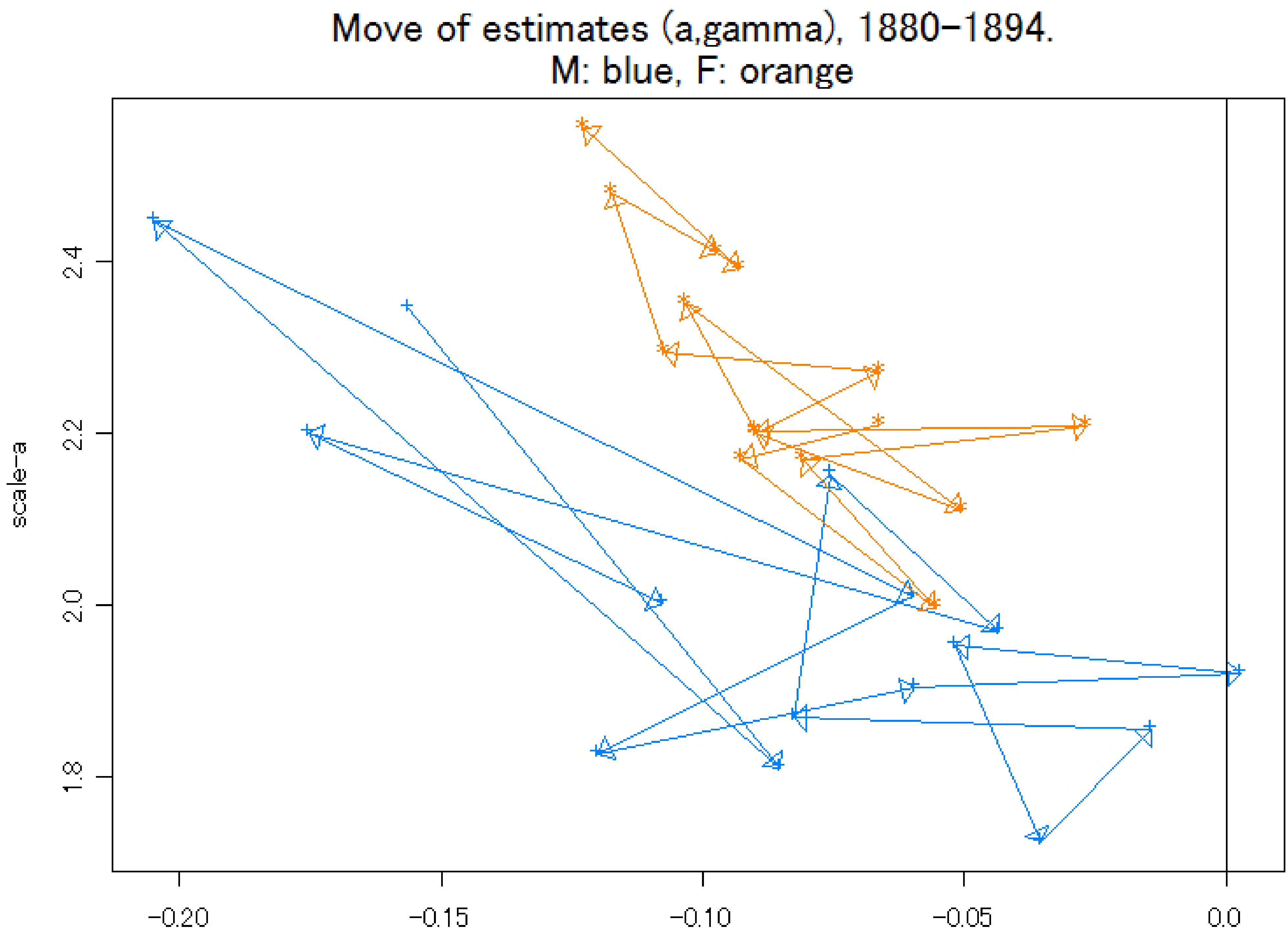
Method PS3 Another pseudo continuous random sample technique is based on the assumption that the lifetime of the cohort by birth follows GnPrt. Let u_ν and v_ν be random numbers uniformly distributed on $(0,1)$, corresponding to birthdays and deathdays within one year, respectively. Generate M_{ij} random pairs (u_ν, v_ν) to obtain the lifetime data

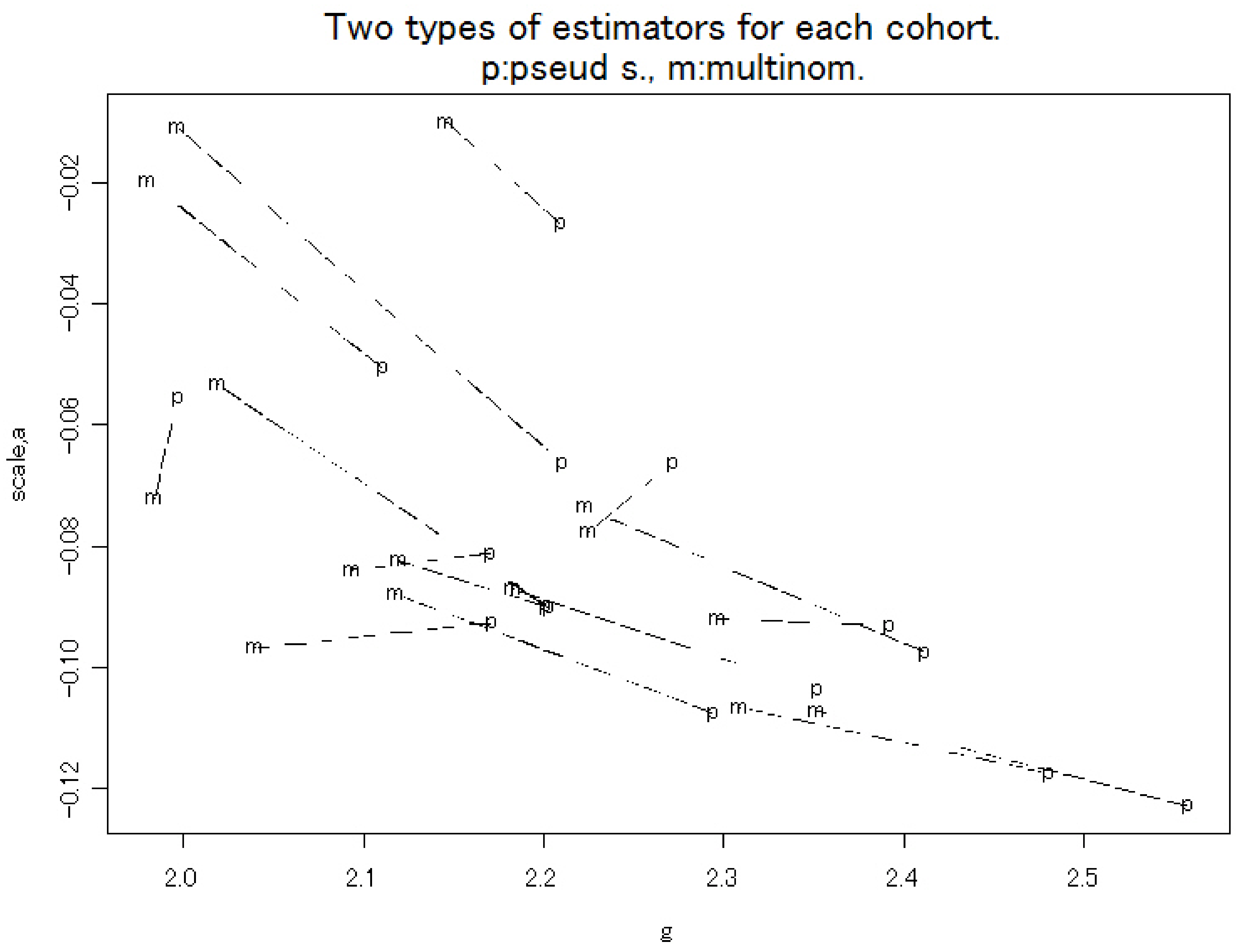
$$y_{i\nu} = i + u_\nu + v_\nu; \quad \nu = 1, \ldots, M_{ij}; \quad i = 100, 101, \ldots$$

of the i_0-th, $i_0 = i - 100$, cohort by birth. Since the lifetime in $(100, 101)$ is not completely observed, the data $y_{i\nu} < 101$ is censored, and GnPrt is fitted to the reserved dataset. The ML estimation of GnPrt models is thoroughly studied by Smith (1985, 1987).



Conc. ellipses of est. 10%, 1880-1890
scale-a
gamma
1.0
1.5
2.0
2.5
3.0
-0.4
-0.3
-0.2
-0.1
0.0
0.1
0.2


Move of estimates (a,gamma), 1880-1894.
M: blue, F: orange
scale-a
1.8
2.0
2.2
2.4
gamma
-0.20
-0.15
-0.10
-0.05
0.0


Two types of estimators for each cohort.
p:pseud s., m:multinom.
scale,a
g
-0.02
-0.04
-0.06
-0.08
-0.10
-0.12
2.0
2.1
2.2
2.3
2.4
2.5

Conclusion

sex	methods	γ	a	η	ω	mean	std	$\widehat{\text{std}(\hat{\gamma})}$
males	MN1	-0.07246	2.011	13.80	27.76	1.875	1.753	0.0800
size=	MN2	-0.07230	1.978	13.83	27.35	1.844	1.724	0.0778
2050	PS2	-0.08089	2.035	12.36	25.16	1.883	1.747	
or 1653	PS3	-0.06065	1.906	16.49	31.42	1.797	1.697	
females	MN1	-0.08771	2.332	11.40	26.59	2.144	1.977	0.0522
size=	MN2	-0.08750	2.291	11.43	26.18	2.106	1.943	0.0459
7486	PS2	-0.09963	2.356	10.04	23.65	2.143	1.957	
or 6229	PS3	-0.07413	2.201	13.49	29.69	2.049	1.912	

mean= $a/(1-\gamma)$, std = mean$/\sqrt{1-2\gamma}$

Table 2: Estimates of parameters.

Arssen and de Haan (1994) studied the same problem using datasets of both birthdays and deathdays. Their data consists of the total life span (in days), sex, and year of birth of all people born in the years 1877–1881, still alive on January 1, 1971 and who died as residents of the Netherlands. The size of table is 4131 men and 6260 women. Their results are consistent with that of the present paper.

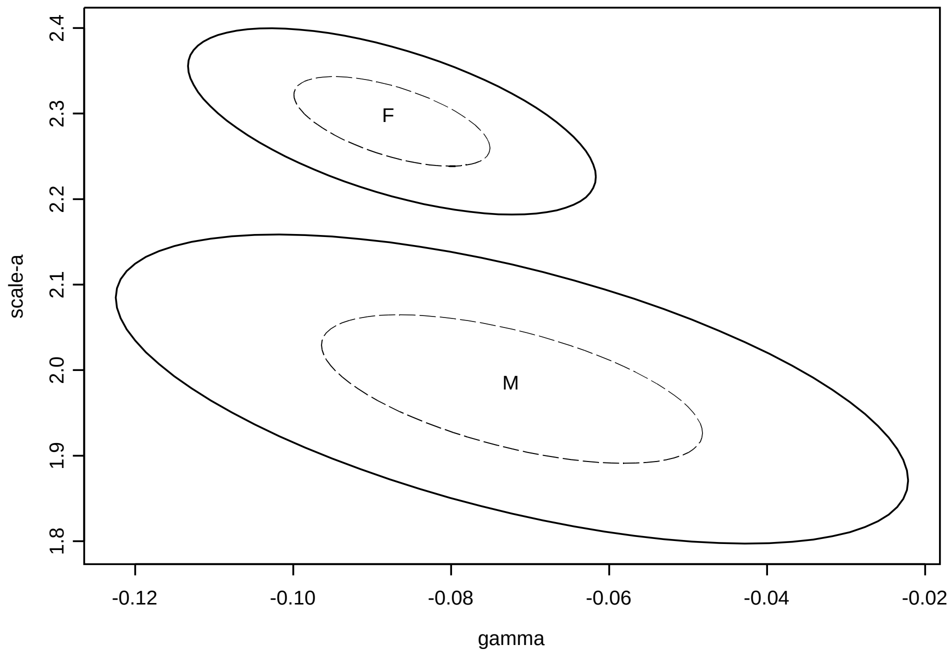


Figure 3: Estimates and concentration ellipses of $(\hat{\gamma}, \hat{a})$ with 50% and 95% contours based on Tables 2 and **??** (MN2).

5 Inference Problems

- nonregular case: the distribution range depends on parameters.
- max value is not sufficient.

$$c(x/\omega)^\eta, \quad 0 < x < \omega$$

- any better estimators.
- the exponentiality test.
- tabulation loss of information.

References

[1] Arssen, K. and de Haan, L. (1994) On the maximal life span of humans. *Mathematical Population Studies*, **4**, 259–281.

[2] Coles, S. (2001) *An Introduction to Statistical Modeling of Extreme Values*, Springer-Verlag, London.

[3] Embrechts, P., Klüppelburg, C., and Mikosch, T. (1998) *Modelling Extremal Events for Insurance and Finance*, Springer, New York.

[4] Kannisto, V. (1999) Trends in the mortality of the oldest-old, in Alho, J. (ed.) *Statistics, Registries, and Science: Experiences from Finland*, 177–194, Statistics Finland.

[5] Keiding, N. (1999). Lexis diagram. S. Kotz, C.B. Read, and D.L. Banks (eds.) *Encyclopedia of Statistical Sciences*, Update Volume 3, Wiley, New York, NY, 411–413.

[6] Smith, R.L. (1985). Maximum likelihood estimation in a class of nonregular cases, *Biometrika*, **72**, 67–90.

[7] Smith, R.L. (1987) Estimating tails of probability distributions, *Annals of Statistics*, **15**, 1174–1207.