NOTES ET DOCUMENTS

PROBABILITIES, FREQUENCIES AND EVENTS

> "If there were no difference in mortality and mobility between [the studied sub-group] and the whole of the population and if there were no errors in the numbers, [events] and population size, [...] it would make no difference which of the three series [making up the decrement table] we calculated, the *probabilities*, the *frequencies* [in the state prior to occurrence of the event] or the *events*."

This was how Louis Henry concluded the chapter on nuptiality in his manual of demography[1]. But he was referring to the analysis of demographic processes within cohorts. What happens when we synthesize cohort data transversally?

To understand more clearly the relations existing between each of these three series in a transversal perspective, let us consider the ideal situation in which the events observed are those of tables. To simplify even further, let us use life tables, since, death being the lot of every man and beast, the intensity of mortality is constant and equal to unity. In each elementary parallelogram of the Lexis diagram, we then have the number of deaths occurring at the age considered in the cohort concerned (Figure 1). From these data, we can calculate each of the three series which constitute a *period* decrement table. Thus, we obtain three entries by which we can construct three period life tables.

Entry by the probabilities[2] By relating the number of deaths in each elementary parallelogram to the number of survivors at the beginning of the interval studied, we obtain the probabilities of dying relative to each parallelogram. From the series of age-specific probabilities relative to a given calendar year, we can construct the period life table by deriving, step by step, the other series in the table: survivors on each birthday (or each January 1st for cohort probabilities of dying within a calendar year), then deaths between successive birthdays (or January 1sts).

This is the most conventional way of calculating a table, but is not, as we have seen, the only one. We could alternatively calculate the life-table events directly.

[1] L. Henry, *Démographie. Analyse et modèles*, Paris, Larousse, 1972, p. 90.
[2] Hereafter referred to as "Probabilities".

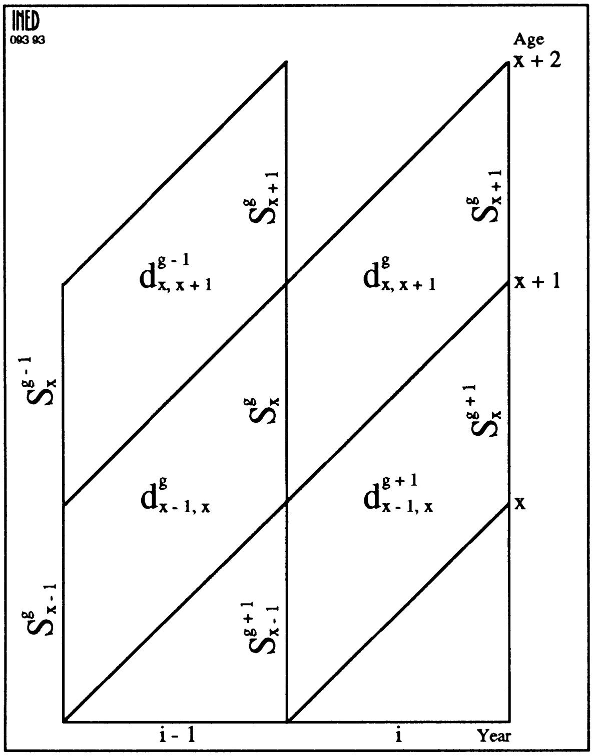

Figure 1. – Diagram of the data

Entry by the events[3] The cohort life-table deaths can be used directly to construct period life tables, since the deaths occurring in a given year constitute the series of events of the life table relative to the year considered.

By subtracting, step by step, the deaths at a given age from the number of survivors at the beginning of the interval, we obtain the number of survivors at beginning of next interval. By relating the deaths to the survivors at beginning of interval, we calculate the series of probabilities of dying.

But the life table constructed in this way will differ from that constructed by taking the probabilities as entry, unless the timing of mortality is invariant from one cohort to the next.

(3) Referred to as "Events".

Entry by the frequencies[4] The series of survivors in the cohort life tables enables us to perform the third type of calculation. The survivors in the different cohorts, in the middle of the year considered, can, when mortality is stable, be considered as the series of survivors in the period life table. By subtraction between survivors at successive ages, we derive the series of life-table deaths. The series of probabilities of dying can be obtained simply by relating the deaths at a given age to the survivors at the beginning of this age interval, or from the complement of the ratio of survivors at two successive ages.

Relationships between the three entries These three methods, which yield strictly equivalent results in the case of cohort measures[5], do so in the case of period constructs only if the timing (tempo) and intensity (quantum) of the process investigated remain invariable over the cohorts concerned.

For mortality, where the quantum is always 1, but the tempo becomes later cohortwise, we observe the following differences between the three sets of results: the sum of events is equal to the radix of the table when the life table is entered by the probabilities or frequencies, but is lower when it is entered by the reduced events. The mean age from the distribution of events in the three tables shows that life expectancy increases from entry by the frequencies of survival to entry by reduced deaths to entry by probabilities.

But if the three sets of results are no longer equivalent when tempo and/or quantum changes are observed cohortwise, that does not mean they have no relationship.

In fact, there is a similar relationship to that existing between the mode, the median and the mean of a distribution: when the latter follows the normal distribution[6], the three central moments merge, like our three sets of results when the cohort rates do not change.

But when the distribution becomes increasingly left-tailed, the following relationship is observed:

$$\text{Mean} < \text{Median} < \text{Mode}$$

and when it is increasingly right-tailed:

$$\text{Mode} < \text{Median} < \text{Mean}.$$

Similar changes are observed when the age-specific rates vary over cohorts.

---

[4] Referred to as "Frequency"; also known as 'Hajnal's method' (see J. Hajnal, «Age at marriage and proportions marrying», *Population Studies*, 2, 1953, 115-136).

[5] Since in the framework of tables, there are no observation problems or disturbances to upset the equivalence.

[6] Or, more generally, if it is symmetrical.

When intensity remains constant and timing grows later, the mean time elapsed before occurrence of the studied event, calculated using each of the three entries, is in the following relationship[7]:

Frequency < Events < Probabilities

When intensity is constant but timing becomes earlier, we can suppose that the relationship will be inversed:

Probabilities < Events < Frequency

To clarify these relationships, we can leave the field of mortality and simulate processes in which tempo and quantum vary uniformly and independently across cohorts. With the three different entries, the simulations, in terms of sum of events and mean time elapsed before occurrence of event, give the following results[8]:

TABLE 1. – RELATIONSHIPS BETWEEN THE THREE ENTRIES IN DIFFERENT TEMPO AND QUANTUM SITUATIONS

Tempo	Quantum lower		Quantum stable		Quantum higher	
	Sum	Mean	Sum	Mean	Sum	Mean
Earlier	R < P < F	R < P < F	F < P < R	P < R < F	F < P < R	F < P < R
Stable	R < P < F	R < P < F	F = R = P	F = R = P	F < P < R	F < P < R
Later	R < P < F	F < R < P	R < P < F	F < R < P	F < P < R	F < P < R

F = Frequency; R = Reduced events; P = Probabilities.

We note that when quantum increases across cohorts, the period indicator (*sum* of events) calculated from the frequencies is always lower than from the probabilities, which is lower again than from the reduced events. This order is the same whatever the cohort timing changes. It is also observed in the case of earlier timing, with no change in quantum. In all other cases, except when the rates are constant, as we have seen above, the order is inversed.

Thus, the order is apparently determined by changes in cohort quantum, rather than tempo.

Regarding the *mean* duration at occurrence of event, the relationships are identical to those for the sum of events, except when quantum does not vary across cohorts or when quantum is decreasing and timing becoming later: in these cases, the position of the 'Frequencies' measure is inversed.

(7) In the case of inescapable phenomena, attrition is ultimately complete, except when the table is constructed from the reduced events.

(8) We could go further and measure the differences between the three sets depending on the pace of change, since in the framework of linear trends the indices measured by the three methods are strictly related. We could also compare these indices to those observed in cohorts to determine whether, in a given situation, one is not preferable to the others.

These relationships could therefore be used to predict the direction of cohort developments, at least in cases where we can suppose there has been no about-turn.

In all the different situations, the measures calculated from frequencies are relatively distant from the others, except when cohort quantum is stable.

The example of the French life tables for 1980

To look into the origins of these relationships, let us come back to mortality. We shall consider only the mean ages at death derived from the three life tables, since in the tables constructed from probabilities of dying or frequencies of survival, attrition is ultimately complete.

Table 2 displays, for the year 1980, the numbers of survivors and deaths in the three period life tables that can be constructed using the different entries.

It confirms the results of Table 1, which indicated that, when the timing of a process grows later, while its intensity is stable – this is the case of mortality in France at least since the beginning of the 20th century – the mean age calculated from the probabilities should be the highest of the three, and that calculated from the frequencies the lowest. If this is

TABLE 2. – LIFE TABLES FOR THE YEAR 1980, BY SELECTED ENTRY

	Reduced deaths		Probabilities of dying		Frequency of survival	
	Survivors	Deaths	Survivors	Deaths	Survivors	Deaths
0	100 000	1 055	100 000	1 055	100 000	1 370
5	98 945	136	98 945	128	98 630	597
10	98 809	119	98 817	118	98 033	495
15	98 690	254	98 699	259	97 538	849
20	98 436	269	98 440	289	96 689	1 203
25	98 167	286	98 151	294	95 486	1 821
30	97 881	364	97 857	394	93 665	5 586
35	97 517	504	97 463	559	88 079	– 70
40	97 013	774	96 904	859	88 149	410
45	96 239	1 132	96 045	1 282	87 739	3 819
50	95107	1 587	94 763	1 875	83 920	4 720
55	93 520	2 150	92 888	2 596	79 200	6 485
60	91 370	2 786	90 292	3 629	72 715	6 531
65	88 584	4 053	86 663	5 528	66 184	7 396
70	84 531	6 378	81 135	9 083	58 788	10 322
75	78 153	9 357	72 052	14 380	48 466	13 404
80	68 796	11 664	57 672	19 902	35 062	15 160
85	57 132	10 273	37 770	19 953	19 902	11 975
90	46 859	5 847	17 817	12 579	7 927	6 378
95	41 012	1 425	5 238	4 129	1 549	1 466
100	39 587	83	1 109	1 109	83	83
Total	–	60 496	–	100 000	–	100 000
Mean length of life	–	74.40	–	77.65	–	67.81

the case, it is because, *as we move from probabilities to reduced events to observed proportions, the centre of gravity of the calculation gradually shifts further back into the past.*

Indeed, in the table constructed from *probabilities,* the mean is calculated from observation during a single year. The table built on *reduced events* "links the current and past observations"[9] and uses both the events in the given year and the current state of the cohort regarding the process studied. In the table derived from observed *frequencies*, the mean hardly reflects the current events, since they are incorporated into the sum of the past occurrences in each cohort. The events of the year are thus the differences in progression in the studied process[10] between cohorts. This amounts to taking into account, for this calculation, only the cohort's past history.

We can conclude that the construction of tables from observed frequencies should be avoided whenever there is cohort change in intensity of the studied process, if the aim is to obtain a period measure giving an idea of the *levels* of cohort quantum, since this calculation reveals essentially the *differences* between cohorts. Its principal contribution would be to refine estimation of the direction of change in cohorts, provided the other two types of table have been calculated.

Jean-Paul SARDON
INED

APPENDIX
Calculation of period life-table deaths using three types of entry[11]

— Entry by the probabilities:

$$ {}_qD^g_{x,x+1} = [\dots]\left(1-\frac{d^{g-2}_{x+2,x+3}}{S^{g-2}_{x+2}}\right)\left(1-\frac{d^{g-1}_{x+1,x+2}}{S^{g-1}_{x+1}}\right) - [\dots]\left(1-\frac{d^{g-2}_{x+2,x+3}}{S^{g-2}_{x+2}}\right)\left(1-\frac{d^{g-1}_{x+1,x+2}}{S^{g-1}_{x+1}}\right)\left(1-\frac{d^{g}_{x,x+1}}{S^{g}_{x}}\right) $$

— Entry by the events:

$$ {}_ED^g_{x,\,x+1} = d^g_{x,\,x+1} = S^g_x - S^g_{x+1} $$

— Entry by the frequencies:

$$ {}_FD^g_{x,\,x+1} = S^{g+1}_x - S^g_{x+1} $$

(9) Y. Péron, «Les indices du moment de la nuptialité des célibataires», *Population*, 6, 1991, 1429-1440.

(10) With the possibility of negative events.

(11) The data being derived from cohort life tables.

Generalizing the formulation to any type of observed events, we have:

- the probabilities

$$q_x^g = \frac{E_{x,\,x+1}^g}{{}_EP_x^g - \dfrac{S_{x,\,x+1}^g}{2}}$$

where ${}_EP_x^g$ = the population of age x in cohort g having 'survived' the event and $S_{x,\,x+1}^g$ the exits from observation at age x

- the reduced events:

$$e_x^g = \frac{E_{x,\,x+1}^g}{\left(P_x^g + P_{x+1}^g\right) \times \dfrac{1}{2}}$$

where P_x^g = the population of age x in cohort g, having survived or not survived the studied event[12].

- the frequencies:

$$f_x^g = \frac{{}_EP_x^g}{P_x^g}$$

(12) In the case of mortality, $P_x^g \approx P_0^g - \sum_0^x S_x^g$ where P_0^g = initial population of cohort g.