Hazard Rate Functions

General Discussion

Definition. If T is an *absolutely continuous* non-negative random variable, its hazard rate function $h(t),\ t \geq 0$, is defined by

$$h(t) \ = \ \frac{f(t)}{S(t)}, \quad t \geq 0,$$

where $f(t)$ is the density of T and $S(t)$ is the survival function: $S(t) = \int_t^\infty f(u)du$.
Note that $\mathrm{P}\{T \leq t + \Delta \,|\, T > t\} \approx h(t) \cdot \Delta$.
If T is a *discrete* non-negative random variable that takes values $t_1 < t_2 < \ldots$ with corresponding probabilities $\{p_i,\ i \geq 1\}$, then its hazard-sequence $\{h(t_i)\}$ is defined by

$$h(t_i) \ = \ \frac{p_i}{\sum_{j \geq i} p_j} \ = \ \frac{p_i}{S(t_i -)}, \quad i \geq 1.$$

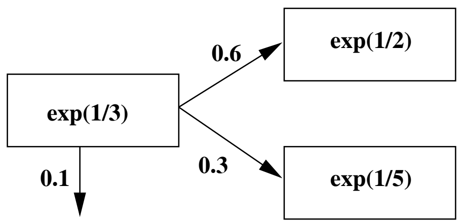
Note that $\mathrm{P}\{T = t_i \,|\, T > t_{i-1}\} = h(t_i)$.

Why estimate the hazard rates of service times or patience?

- The hazard rate is a *dynamic* characteristic of a distribution.
- The hazard rate is a more precise "fingerprint" of a distribution than the cumulative distribution function, the survival function, or density (for example, unlike the density, its tail need not converge to zero; the tail can increase, decrease, converge to some constant etc.)
- The hazard rate provides a tool for comparing the tail of the distribution in question against some "benchmark": the exponential distribution, in our case.
- The hazard rate arises naturally when we discuss "strategies of abandonment", either rational (as in Mandelbaum & Shimkin) or ad-hoc (Palm).

Theoretical Calculation and Statistical Estimation of Hazard Rate

Example: consider the following service time distribution:



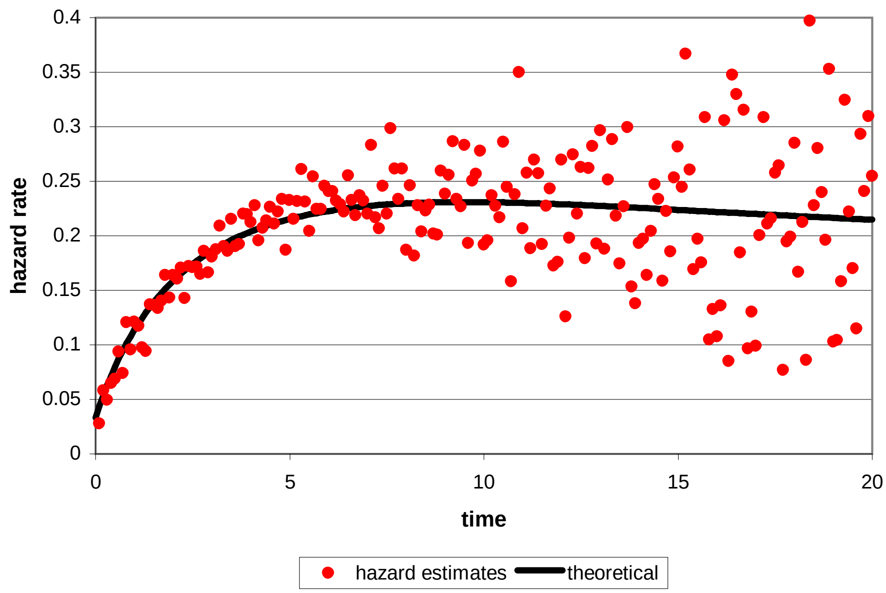
Its hazard rate can be calculated theoretically:

$$h(x) \;=\; \frac{0.15 \cdot e^{-x/5} + (29/60) \cdot e^{-x/3} - 0.6 \cdot e^{-x/2}}{0.75 \cdot e^{-x/5} + 1.45 \cdot e^{-x/3} - 1.2 \cdot e^{-x/2}}\,.$$

How do we estimate hazard rate from data?

Description of simulation experiment.

10,000 independent realizations of service time above were simulated in Excel. The theoretical hazard rates were plotted and compared against estimates of the hazard rate, based on the simulation data. (The method used for hazard rate estimation is described on the next page.)



Comments:

- The hazard-rate is neither increasing nor decreasing: **"hump"** pattern.
- Value at zero: 1/3*0.1 – product of rate of the initial phase and exit probability.
- Limit at infinity: 1/5 – rate of the longest final phase.

Estimation of the Hazard Rate: Technicalities

The hazard rate is assumed to be constant on successive time intervals of length 0.1 between 0 and 20 (200 intervals overall). Formally, interval j is $\left(\frac{j-1}{10}, \frac{j}{10}\right]$, $j = 1, 2, \ldots, 200$.

The hazard estimate $\hat{h}_j$ for interval number j is calculated using the following formula:

$$\hat{h}_j = \frac{d_j}{b_j\left(r_{j-1} - \frac{1}{2}d_j\right)},$$

where

d_j = *number of events* (service terminations) in interval number j;

r_{j-1} = *number at risk* at the beginning of interval number j (number of services that have not terminated yet at time $\frac{j-1}{10}$);

b_j = length of interval number j (0.1 for all intervals, in our case).

The following provides some intuition for the above formula:
Let n denote the sample size. Then $\dfrac{d_j}{b_j \cdot n}$ is a reasonable estimate of the average density in interval number j and $\dfrac{r_{j-1} - 0.5 \cdot d_j}{n}$ is an approximation for the survival function in the center of this interval.

Remark. This estimation procedure is also valid for the **censored data**.

Remark. Handout that we install in "Related Materials" contains additional examples of phase-type distributions and their hazard rates.

Part of Excel Table

Time	events	at risk	Hazard Estimate	Theoretical
0		10000		0.033
0.1	28	9972	0.028	0.044
0.2	58	9914	0.058	0.054
0.3	49	9865	0.050	0.063
0.4	64	9801	0.065	0.072
0.5	67	9734	0.069	0.080
0.6	91	9643	0.094	0.087
0.7	71	9572	0.074	0.095
0.8	115	9457	0.121	0.101
0.9	90	9367	0.096	0.108
1	113	9254	0.121	0.114
1.1	108	9146	0.117	0.119
1.2	89	9057	0.098	0.125
1.3	85	8972	0.094	0.130
1.4	122	8850	0.137	0.135
1.5	120	8730	0.137	0.139
1.6	116	8614	0.134	0.144
1.7	120	8494	0.140	0.148
1.8	138	8356	0.164	0.152
1.9	119	8237	0.143	0.155
2	134	8103	0.164	0.159