Analysis of Survival Data

Notes and Exercises

2020

F. P. Treasure

January 23, 2020

Time-to-Event Distributions

1 Survivor and Hazard

1.1 Definitions

Given a continuous random time-to-event variable T, the *survivor function* $\mathrm{F}(t)$ is defined by $\mathrm{F}(t) = \mathbb{P}[t < T]$.

The *density* $\mathrm{f}(t)$ has the standard definition:

$$\mathrm{f}(t) = \lim_{\Delta \downarrow 0} \left\{ \frac{1}{\Delta} \mathbb{P}\left[t \leq T < t + \Delta\right] \right\}.$$

The *hazard* $\mathrm{h}(t)$ is the density at t given that there has been no event before t:

$$\mathrm{h}(t) = \lim_{\Delta \downarrow 0} \left\{ \frac{1}{\Delta} \mathbb{P}\left[t \leq T < t + \Delta | t < T\right] \right\}$$

and the *integrated hazard* $\mathrm{H}(t)$ is simply the integral of the hazard:

$$\mathrm{H}(t) = \int_0^t \mathrm{h}(t')\mathrm{d}t'.$$

Note: The definition of $\mathrm{F}(t)$ does not rely on T having a continuous distribution.

1.2 Exercises

(1) Verify the following key relationships:

1. $\mathrm{f}(t) = -\mathrm{F}'(t)$;
2. $\mathrm{F}(t) = \int_t^\infty \mathrm{f}(t')\mathrm{d}t'$;
3. $\mathrm{f}(t) = \mathrm{h}(t)\mathrm{F}(t)$;
4. $\mathrm{F}(t) = \exp\{-\mathrm{H}(t)\}$;
5. $\mathrm{H}(t) = -\log \mathrm{F}(t)$.

Note: The relationship $\mathrm{H}(t) = -\log \mathrm{F}(t)$ can be used to provide a definition of $H(t)$ for discrete data.

2 Likelihood

2.1 Setting up the Likelihood Function

A typical time-to-event dataset comprises n individuals: x_i being either the time of the observed event ($v_i = 1$) or the time of censoring ($v_i = 0$) for the ith individual. The common density and survivor function are $\mathrm{f}(t;\theta)$ and $\mathrm{F}(t;\theta)$ respectively – where θ is a parameter vector.

If the ith individual has an event at x_i then that individual contributes $\mathrm{f}(x_i;\theta)$ to the likelihood: we know that $T_i = x_i$.

If the ith individual has is right-censored at x_i then that individual contributes $\mathrm{F}(x_i;\theta)$ to the likelihood: all we know is that $T_i > x_i$.

The likelihood function $\mathcal{L}(\theta)$ is therefore given by:

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} \big(\mathrm{f}(x_i;\theta)\big)^{v_i} \big(\mathrm{F}(x_i;\theta)\big)^{1-v_i}.$$

2.2 Exercises

(2) Derive the following equivalent expressions for the log-likelihood or *support* function $\mathcal{S}(\theta)$:

1. $\mathcal{S}(\theta) = \sum_{i=1}^{n} v_i \log \mathrm{f}(x_i;\theta) + (1 - v_i) \log \mathrm{F}(x_i;\theta)$;
2. $\mathcal{S}(\theta) = \sum_{i=1}^{n} v_i \log \mathrm{h}(x_i;\theta) - \mathrm{H}(x_i;\theta)$.

3 Exponential Distribution

A time-to-event variable has an *exponential* distribution with rate parameter λ ($\lambda > 0$) if the density $\mathrm{f}(t;\lambda) = \lambda \exp(-\lambda t)$.

3.1 Exercises

(3) Obtain the following expressions for the survivor, hazard and integrated hazard functions of a time-to-event variable with an `exponential`(λ) distribution:

1. $\mathrm{F}(t;\lambda) = \exp(-\lambda t)$;
2. $\mathrm{h}(t;\lambda) = \lambda$;
3. $\mathrm{H}(t;\lambda) = \lambda t$.

(4) Show that the support function for the `exponential`(λ) distribution is given by:

$$\mathcal{S}(\lambda) = v_+ \log \lambda - \lambda x_+$$

where $v_+ = \sum_{i=1}^n v_i$ is the number of observed events and $x_+ = \sum_{i=1}^n x_i$ is the 'total time at risk'. Show that the maximum likelihood estimate of λ is given by

$$\hat{\lambda} = \frac{v_+}{x_+}$$

and the information function calculated at the maximum likelihood estimate is:

$$\mathcal{I}(\hat{\lambda}) = \frac{v_+}{\hat{\lambda}^2}.$$

Note: The observed information – and therefore the precision of estimates of λ – is a function of v_+ (the number of events), not n (the number of individuals).

3.2 An Interval Estimate for the Rate Parameter

By Wilks's lemma we know that when the true value of λ is $\tilde{\lambda}$ and its maximium likelihood estimate is $\hat{\lambda}$:

$$2[\mathcal{S}(\hat{\lambda}) - \mathcal{S}(\tilde{\lambda})] \sim \texttt{chi-square}(1)$$

approximately. An approximate $1-\alpha$ confidence interval for λ is therefore given by:

$$\{\lambda : 2[\mathcal{S}(\hat{\lambda}) - \mathcal{S}(\lambda)] \leq c_{1,1-\alpha}\}$$

where $c_{k,q}$ is the inverse cumulative distribution function of a `chi-square`(k) distribution. (That is: if $Z \sim$ `chi-square`(k) then $\mathbb{P}[Z \leq c_{k,q}] = q$.)

4 Weibull Distribution

A time-to-event variable has an *Weibull* distribution with rate parameter λ and shape parameter k ($\lambda > 0$, $k > 0$) if the integrated hazard is $\mathrm{H}(t; \lambda, k) = (\lambda t)^k$.

Note: A $\texttt{Weibull}(\lambda, 1)$ distribution is the same as an $\texttt{exponential}(\lambda)$ distribution.

4.1 Exercises

(5) Derive the hazard, survivor and density functions for a $\texttt{Weibull}(\lambda, k)$ distribution. In what circumstances does the hazard decrease with time?

(6) Show that two Weibull distributions with the same shape parameter k belong (i) to the same proportional hazards family and (ii) to the same accelerated life family.

Reminder: Two time-to-event distributions belong to the same *proportional hazards* family if their hazard functions $\mathrm{h}_1(t)$ and $\mathrm{h}_2(t)$ are related by $\mathrm{h}_2(t) = \beta \mathrm{h}_1(t)$ for some $\beta > 0$. Two time-to-event distributions belong to the same *accelerated life* family if their survivor functions $\mathrm{F}_1(t)$ and $\mathrm{F}_2(t)$ are related by $\mathrm{F}_2(t) = \mathrm{F}_1(\gamma t)$ for some $\gamma > 0$.

(7) Show that if Z has a $\texttt{uniform}(0, 1)$ density then $\frac{1}{\lambda} \sqrt[k]{-\log Z}$ has a $\texttt{Weibull}(\lambda, k)$ distribution.

(8) Show that if T has a Weibull distribution then U, defined by $\log U = \alpha \log T + \beta$ ($\alpha > 0$), also has a Weibull distribution.

(9) The *characteristic graph* of a time-to-event distribution is the graph of $\log\big(-\log \mathrm{F}(t)\big)$ against $\log t$.

1. What is the characteristc graph of a Weibull distribution?
2. If two time-to-event distributions belong to the same proportional hazards family, how are their characteristic graphs related?
3. If two time-to-event distributions belong to the same accelerated life family, how are their characteristic graphs related?
4. Show how the shape of the characteristic graph of a Weibull distribution is consistent with the family of Weibull distributions with the same k being both a proportional hazards family and an accelerated life family.

(10) *Harder:* If the proportional hazards family:

$$\mathrm{h}(t;\beta) = \exp(\beta)h_0(t) \text{ with } \beta \in \mathbb{R}$$

is also an accelerated life family, show that $h_0(t)$ must be the hazard function of a Weibull distribution.

5 Kaplan-Meier

5.1 Exercises

(11) The first five observations of a time-to-event dataset comprising 21 individuals are 6, 6, 6, 6+ and 7 weeks respectively where the '+' represents a censored observation. All other individuals have event or right censoring times strictly greater than 8 weeks.

1. Explain why we have enough data to calculate the Kaplan-Meier estimate for F(8).
2. Calculate the Kaplan-Meier estimate of F(8).
3. Suppose further information was obtained about the censored individual. Calculate the Kaplan-Meier estimate of F(8) in the cases (a) the censored individual in fact had an event at 7 weeks and (b) the censored individual in fact had an event at 9 weeks.
4. Comment on the relative magnitudes of the three estimates.

6 Empirical Likelihood

6.1 Exercises

(12) Using the same dataset as in question **??**, obtain the empirical log-likelihood function for F(6) and F(8) as

$$\mathcal{S}(\alpha, \beta) = 3\log(1-\alpha) + \log(\alpha) + \log(\alpha - \beta) + 16\log(\beta)$$

justifying each term, with F(6) represented by α and F(8) represented by β.

1. Maximise $\mathcal{S}(\alpha, \beta)$ to obtain maximum empirical likelihood estimates of F(6) and F(8). Compare your answers with the Kaplan-Meier estimates obtained previously.
2. Suppose it was discovered that the individual with an event recorded at 7 weeks in fact is only known to have had an event before or at 7 weeks. How should the log-likelihood be modified to take account of this new information? Maximise the new log-likelihood with respect to α and β. Why is your answer inappropriate? Obtain a better answer.

(13) A student intended to record the time of arrival of a lecturer at five consecutive 10 a.m. lectures.

(a) on Monday, Tuesday and Friday the lecturer arrived at 10:02, 10:18 and 10:07 respectively.

(b) on Wednesday the student had started reading a newspaper at 10:03 (at which time the lecturer had not arrived) and finished at 10:15 (at which time the lecturer was in the lecture room). The student did not notice the actual arrival of the lecturer.

(c) on Thursday the student was hungry and left the lecture room at 10:06, not to return. The lecturer had not arrived at that time.

Using the information given, obtain the non-parametric maximum likelihood estimator of F(t), the probability that the lecturer's time of arrival is after t. [Diploma 1999Q2]

7 Proportional Hazards

7.1 Exercises

(14) If two time-to-event distributions have hazard functions $h(t)$ and $kh(t)$ (with $k > 0$) respectively what is the relationship between their survivor functions?

Reminder: The *score* test of the hypothesis $\beta = \beta_0$ is an approximate likelihood-ratio test. It is obtained by applying Wilks's lemma to a quadratic approximation to the log-likelihood at β_0. The ratio $[\mathcal{S}'(\beta_0)]^2 \,/\, |\mathcal{S}''(\beta_0)|$ is compared to a `chi-square(1)` distribution.

(15) A time-to-event dataset comprises n individuals with the observed time (event or censoring), visibility indicator (indicator that the observed time is an event time) and explanatory variable ($\in \{0, 1\}$) for the ith individual being denoted by x_i, v_i and z_i respectively. There are no ties in the dataset.

1. Construct a proportional hazards model in terms of a parameter β for the effect of z on the time-to-event distribution. Obtain an expression for the log-likelihood function $\mathcal{S}(\beta)$.
2. Show that the score test of the hypothesis $\beta = 0$ has the same form as the log-rank test of difference between the two groups defined by z_i.

8 Competing Risks

8.1 Exercises

(16) The time to failure of the Deuteron Powered Moon and Mars Shuttle is a continuous random variable. The shuttle can fail in either of two ways: the Kinetic Entropy Terawatt Transit Laser Engine may burn out or the Tetrahedral Evolutionary Asynchronous Potential Oxygen Transducer may fracture. It is not possible for both types of failure to occur simultaneously.

The cause-specific hazards for engine burn out and transducer fracture are $h_A(t)$ and $h_B(t)$ respectively. Obtain in terms of h_A and h_B an expression for the probability that the engine burns out before the transducer fractures.

Suppose that h_A is constant and h_B proportional to time. Choose the unit of time such that:

$$h_A(t) = c \text{ and } h_B(t) = t$$

and show that if $c = 0.6120$ the probability that the engine will burn out before the transducer fails is approximately one-half.

[Diploma 1991 P2/Q12: candidates were permitted to use calculators but only for addition, subtraction, multiplication and division. They were given the

free information that if Φ is the standard Normal distribution function then $\Phi(0.6120) \simeq 0.7298$ and $\Phi'(0.6120) \simeq 0.3308$.]