Event history analysis

An introduction

Alessandro Di Nallo

Università Bocconi

Université de Lausanne, 2 November 2022

This workshop

A brief introduction to event history analysis

1. Theoretical concepts
   - Why event history analysis?
   - Data for survival analysis
   - Survival models vs. linear regression
   - Functions in continuous and discrete time
2. Survival analysis in STATA
   - Continuous time models
   - Discrete time models
3. Some extensions
   - Unobserved heterogeneity
   - Discrete time models with time-varying covariates
   - Competing risk models

Material

Go to https://github.com/alessandrodinallo/EHA

- These slides
- Lecture notes (a more detailed version of the slides)
- STATA do files and data

PART I

What is event history analysis

What is event history analysis about

- Linear regression model
  - How much variance of y is explained by x ?

What is event history analysis about

- Linear regression model
  - How much variance of y is explained by x ?
- Event history analysis
  - When does a subject transition from state T_0 to T_1?
  - Does x influence this transition ?

What is event history analysis about

- It is used to model $\{$time to event $\|$ transition $\|$ survival time $\|$ duration$\}$ data
- We can use ‘event history analysis’ and ‘survival analysis’ interchangeably
- Consider a particular life-course domain, partitioned into a number of mutually-exclusive states at each point in time.
- With the passage of time, individuals move (or do not move) between states.

Examples of event history analysis

Lifecourse domains

- Partnership
  - Married
  - Cohabiting
  - Separated
  - Widowed
  - Single & never-married

Examples of event history analysis

Lifecourse domains

- Partnership
  - Married
  - Cohabiting
  - Separated
  - Widowed
  - Single & never-married

Single
In a union

Examples of event history analysis

Lifecourse domains

- Paid work
    - Employed
    - Self-employed
    - Unemployed
    - Inactive
    - Retired

Examples of event history analysis

Lifecourse domains

- Paid work
  - Employed
  - Self-employed
  - Unemployed
  - Inactive
  - Retired

Examples

1. Unemployed ⇒ Employed

Examples of event history analysis

Lifecourse domains

- Paid work
  - Employed
  - Self-employed
  - Unemployed
  - Inactive
  - Retired

Examples

1. Unemployed ⇒ Employed
2. $\text{Unemployed} \Rightarrow \begin{cases} \textit{Employed} \\ \textit{Inactive} \end{cases}$

Examples of event history analysis

Lifecourse domains

- Paid work
  - Employed
  - Self-employed
  - Unemployed
  - Inactive
  - Retired

Examples of event history analysis

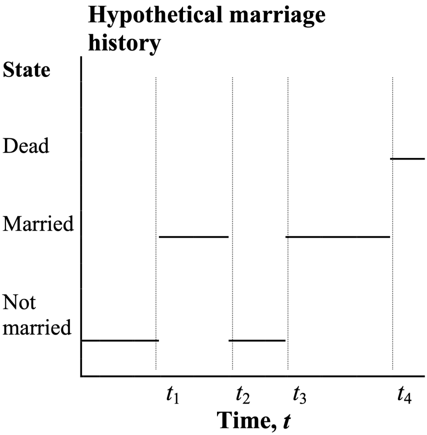
Lifecourse domains

- Paid work
  - Employed
  - Self-employed
  - Unemployed
  - Inactive
  - Retired

- Each of these states at each t (point / interval in time) is unique = a persons *cannot* be in two states at the same time.

- The transition patterns must be characterised by :
  - the time spent within each state;
  - or, the dates of any transitions made between states (if any)

How to define states



- Length of time spent within each state ≈ length of horizontal line
- Spells within a given state marked out by dates (start, and end)

Event history analysis in a nutshell

- Survival data are applicable to a large number of subjects (e.g. individuals, firms, . . .)

- Goals
  - We can describe data and to predict spell lengths
  - We can combine spells with other information about the subjects (their characteristics) → explanatory variables for multivariate modelling. ⇒ Explain the causes of the transitions.

Event history analysis in a nutshell

- Survival time within a single state
  - single → union ✓
  - single → cohabiting → married

Event history analysis in a nutshell

- Survival time within a single state
  - single → union ✓
  - single → cohabiting → married
- Single spell observed for each subject
  - single → union ✓
  - single → union → single → union → . . .

Event history analysis in a nutshell

- Survival time within a single state
  - single → union ✓
  - single → cohabiting → married
- Single spell observed for each subject
  - single → union ✓
  - single → union → single → union → . . .
- No state dependence
  - The chances of making a transition from current state do not depend on transition history prior to entry to current state

Event history analysis in a nutshell

- Survival time within a single state
  - single → union ✓
  - single → cohabiting → married
- Single spell observed for each subject
  - single → union ✓
  - single → union → single → union → . . .
- No state dependence
  - The chances of making a transition from current state do not depend on transition history prior to entry to current state
- No initial conditions issues
  - Entry to state being modelled is treated as exogenous (otherwise we would have to model the chances of having arrived in the state in the first place)

Event history analysis in a nutshell

- Survival time within a single state
  - single → union ✓
  - single → cohabiting → married
- Single spell observed for each subject
  - single → union ✓
  - single → union → single → union → . . .
- No state dependence
  - The chances of making a transition from current state do not depend on transition history prior to entry to current state
- No initial conditions issues
  - Entry to state being modelled is treated as exogenous (otherwise we would have to model the chances of having arrived in the state in the first place)
- Stationary process
  - Model parameters are fixed constant, or can be characterised using explanatory variables, or parametrically

Data for survival analysis

Data collection methods

1. Stock sample
   - Data collection is based upon a random sample of the individuals that are currently in the state of interest, who are typically (but not always) interviewed at some time later, and one also determines when they entered the state (the spell start date).
   - **Example**: When modelling the length of spells of unemployment insurance (UI) receipt, one might sample all the individuals who were in receipt of UI at a given date, and also find out when they first received UI (and other characteristics).

Data collection methods

1. Stock sample
   - Data collection is based upon a random sample of the individuals that are currently in the state of interest, who are typically (but not always) interviewed at some time later, and one also determines when they entered the state (the spell start date).
   - **Example**: When modelling the length of spells of unemployment insurance (UI) receipt, one might sample all the individuals who were in receipt of UI at a given date, and also find out when they first received UI (and other characteristics).

2. Inflow sample
   - Data collection is based on a random sample of all persons entering the state of interest, and individuals are followed until some pre-specifed date (which might be common to all individuals), or until the spell ends.
   - **Example**: When modelling the length of spells of receipt of unemployment insurance (UI), one might sample all the individuals who began a UI spell.

Data collection methods

3. Outflow sample
    - Data collection is based on a random sample of those leaving the state of interest, and one also determines when the spell began.
    - **Example**: The sample would consist of individuals leaving UI recept.

Data collection methods

3. Outflow sample
    - Data collection is based on a random sample of those leaving the state of interest, and one also determines when the spell began.
    - **Example**: The sample would consist of individuals leaving UI recept.

4. Population sample
    - Data collection is based on a general survey of the population (i.e. where sampling is not related to the process of interest), and respondents are asked about their current and/or previous spells of the type of interest (starting and ending dates).
    - **Example**: Census asking people if/when they entered/left UI recept.

Data collection methods

- Administrative records (e.g. current recipients, or ever a recipient within some observation window) + interview
- Sample survey (often one-off) of population, with retrospective questions, e.g. Understanding Society (waves 1) reconstruct past union histories; NSFG (US) etc. . .
- Panel and cohort surveys follow a population – spell info built up from repeated observations on persons, e.g. BHPS, Understanding Society, GSOEP

Censoring

A survival time is **censored** if all that is known is that it began or ended within some particular interval of time, and thus *the total spell length (from entry time until transition) is not known exactly*. We may distinguish the following types of censoring:

Censoring

A survival time is **censored** if all that is known is that it began or ended within some particular interval of time, and thus *the total spell length (from entry time until transition) is not known exactly*. We may distinguish the following types of censoring:

- **Right censoring**: at the time of observation, the transition out of the current state had not yet occurred (the spell end date is unknown), and so the total length of time between entry to and exit from the state is unknown. Given entry at time 0 and observation at time t, we only know that the completed spell is of length $T > t$.

Censoring

A survival time is **censored** if all that is known is that it began or ended within some particular interval of time, and thus *the total spell length (from entry time until transition) is not known exactly*. We may distinguish the following types of censoring:

- **Right censoring**: at the time of observation, the transition out of the current state had not yet occurred (the spell end date is unknown), and so the total length of time between entry to and exit from the state is unknown. Given entry at time 0 and observation at time t, we only know that the completed spell is of length $T > t$.

- **Left censoring**: the start date of the spell was not observed, so again the exact length of the spell (whether completed or incomplete) is not known.

Truncation

Truncation refers to *whether or not we observe a spell or not* in our data (sample selection on dependent variable), whereas censoring means that we don't know the exact length of a completed spell in total

- **Left truncation** (or 'delayed entry'): Only those who have survived more than some minimum amount of time are included in the observation sample ('small' survival times – those below the threshold – are not observed).

Truncation

Truncation refers to *whether or not we observe a spell or not* in our data (sample selection on dependent variable), whereas censoring means that we don't know the exact length of a completed spell in total

- **Left truncation** (or 'delayed entry'): Only those who have survived more than some minimum amount of time are included in the observation sample ('small' survival times – those below the threshold – are not observed).
  - If one samples from the stock of persons in the relevant state at some time s, and interviews them some time later, then persons with short spells are systematically excluded.

Truncation

Truncation refers to *whether or not we observe a spell or not* in our data (sample selection on dependent variable), whereas censoring means that we don't know the exact length of a completed spell in total

- **Left truncation** (or 'delayed entry'): Only those who have survived more than some minimum amount of time are included in the observation sample ('small' survival times – those below the threshold – are not observed).
  - If one samples from the stock of persons in the relevant state at some time s, and interviews them some time later, then persons with short spells are systematically excluded.
  - Of all those who began a spell at time $r < s$, only those with relatively long spells survived long enough to be found in the stock at time s and thence available to be sampled.

Truncation

Truncation refers to *whether or not we observe a spell or not* in our data (sample selection on dependent variable), whereas censoring means that we don't know the exact length of a completed spell in total

- **Left truncation** (or 'delayed entry'): Only those who have survived more than some minimum amount of time are included in the observation sample ('small' survival times – those below the threshold – are not observed).
  - If one samples from the stock of persons in the relevant state at some time s, and interviews them some time later, then persons with short spells are systematically excluded.
  - Of all those who began a spell at time $r < s$, only those with relatively long spells survived long enough to be found in the stock at time s and thence available to be sampled.
  - Note that the spell start is assumed known in this case (cf. left censoring), but the subject's survival is only observed from some later date – hence 'delayed entry'.

Truncation

Truncation refers to *whether or not we observe a spell or not* in our data (sample selection on dependent variable), whereas censoring means that we don't know the exact length of a completed spell in total

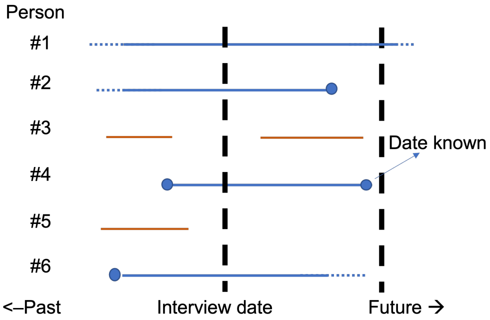
- **Right truncation** : It is the case when only those persons who have experienced the exit event by some particular date are included in the sample, and so relatively 'long' survival times are systematically excluded.

Truncation

Truncation refers to *whether or not we observe a spell or not* in our data (sample selection on dependent variable), whereas censoring means that we don't know the exact length of a completed spell in total

- **Right truncation** : It is the case when only those persons who have experienced the exit event by some particular date are included in the sample, and so relatively 'long' survival times are systematically excluded.
  - Right truncation occurs, for example, when a sample is drawn from the persons who exit from the state at a particular date (e.g. an *outflow* sample from the unemployment register).
  - Only those with a transition by a particular time are included in the sample (e.g. sample from the outflow from a state).

Censoring and truncation, compared



1. Left censored and right censored
2. Left censored
3. Left and right truncated
4. Not censored
5. Left truncated
6. Right censored

Survival models vs. OLS

Continuous & discrete data

- Characteristics of subjects (persons, firms, etc...) & characteristics of socio-economic environment
  - not an issue analytically; may be empirically
- Fixed versus time-varying covariates (TVCs), where TVCs may vary with:
  - survival time in state and/or
  - calendar time
- Analytics and interpretation easier if there are no TVCs
  - estimating models with TVCs means data re-organisation ('episode-splitting', we'll get there)

Econometric methods for survival time data: some motivation

- At this stage, you may still wonder “Why not use OLS?”
  - Regress each survival time (T or logT) on covariates
- Problems with OLS:
  - (right-)censoring of spell data
  - time-varying covariates
  - ‘structural’ modelling (e.g., multi-state models, recurrent events, frailty models)

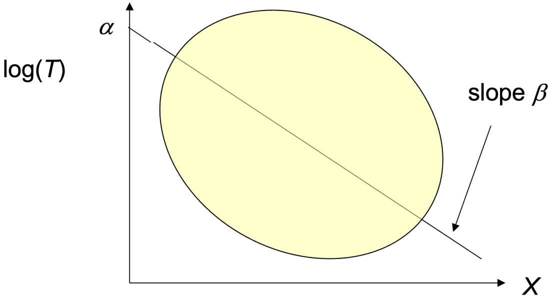
OLS and right-censored data

Suppose that log survival times T_i are a linear function of a single characteristic ${}_i, i = 1, \ldots, N$:

$$log(T_i) = \alpha + \beta X_i + e_i$$

and

$$\alpha > 0, \beta < 0.$$

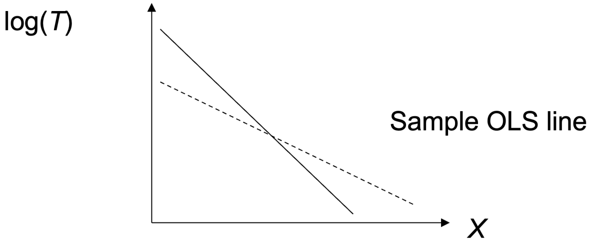


OLS: choose estimators a and b that minimise the sum of the squared

OLS and right-censored data

Suppose prevalence of right-censoring greater at *longer* durations than shorter durations:

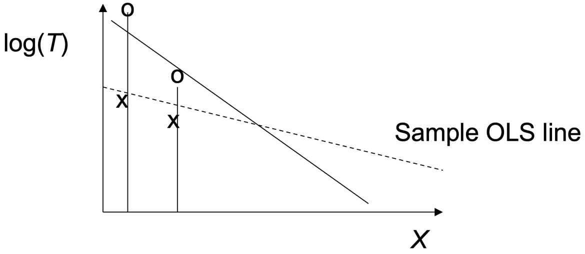
1. Exclude censored spells altogether from OLS estimation → sample data cloud less dense everywhere, *but especially so at higher values of* $log(T)$ → estimated slope not as negative as true slope, i.e. over-estimate.



OLS and right-censored data

Suppose prevalence of right-censoring greater at *longer* durations than shorter durations:

2. Treat censored spells as if they were complete → under-recording, especially so at higher values of $log(T)$ = ‘like non-random mis-measurement of depvar’ → estimated slope not as negative as true slope, i.e. over-estimate.



OLS and time-varying covariates

How can OLS handle them, given that each observation (one spell length) contributes one observation to the regression?
If one were to choose one value of the TVC for each person, which one would one choose?

- That just before the transition (but this varies by person, and what about censored observations?)
- Might use value of TVC at start of spell? (Consistent definition for all spell, but now fixed covariate and lose information)

Why not use a binary dependent variable model?

Use logit (or probit) regression of whether or not experience a transition or not against characteristics? (This would deal with right-censored obs.)

But ...

- But would take no account of the differences in length of time each person was at risk of experiencing the transition, and so loses information (when left, if did so).
- How to handle TVCs?

Bottomline

- **Problem**. We need methods recognising:
  - the longitudinal structure (passage of time);
  - the nature of the spell data;
  - the ability to handle censored (and truncated) spells, and also time-varying covariates
- **Solution**: Use estimation methods other than OLS (typically ML), and also *re-organise* the data set (so that get likelihood right, and can handle TVCs)
- Definition of time
  - Continuous
  - Discrete

Bottomline

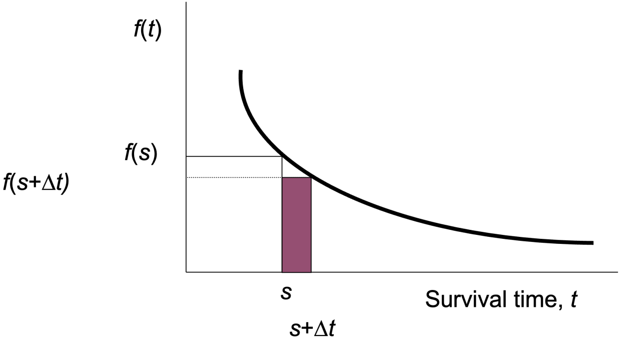
Definition of **time**

- Continuous
  - A process consists of ‘a lot’ of episodes (or ‘*spells*’)
  - There are ‘a lot’ of transition in each episode
  - There a no time-varying covariates
- Discrete
  - A process does not consist of ‘a lot’ of episodes OR there are not ‘lots’ of transitions in each episode
  - There a time-varying covariates

Functions in continuous time

Probability density function

The length of a spell is a realisation of a continuous random variable T with probability density function (PDF): $f(t)$



Quite different from a bell-shape of Normal distribution

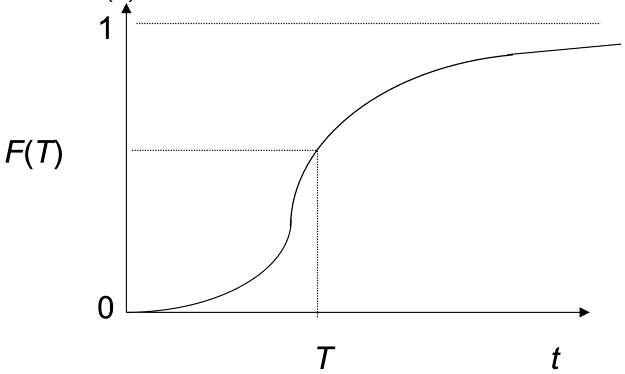
Areas under PDFs are probabilities

$area(rectangle) = height \times base \rightarrow$

$f(s) = height = area/base = Prob/\Delta t$

Failure function

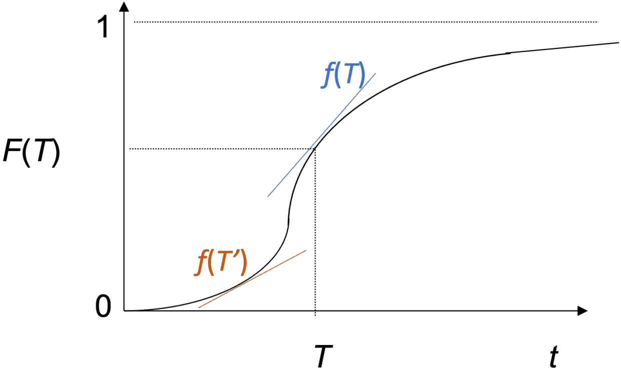
Cumulative density function (CDF), probability distribution function, or "failure function": $F(t)$:



$$F(t) = Pr(T \leq t) = \text{area under } f(t)$$
$$\text{(PDF) up to } t = T$$

Failure function, F(t)

Cumulative density function (CDF), probability distribution function, or "failure function": $F(t)$:

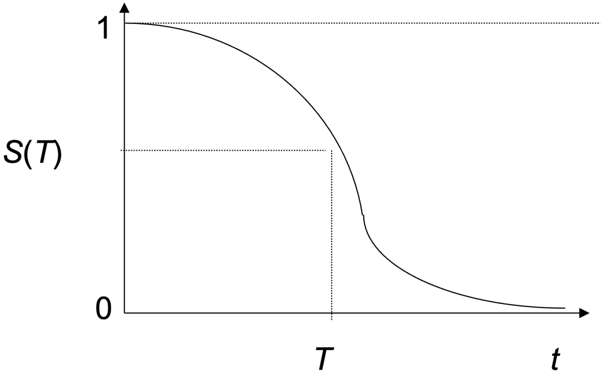


$f(t)$ = slope of CDF at t:

$$f(t) = \lim_{\Delta t \to 0} \frac{Pr(t \leq T \leq t + \Delta t)}{\Delta t} = \frac{\partial F(t)}{\partial t}$$

Survivor function, S(t)

$S(t)$ is probability of survival (i.e. remaining in state) at least t units of time since entry at $t = 0$



Survivor function, S(t)

$$Pr(T > t) = 1 - F(t) \equiv S(t)$$

- The survivor function equals 1 – failure function
- S(t) is probability of survival (i.e. remaining in state) at least t units of time since entry at t = 0
- S(t) and F(t) are probabilities, so $0 \leq$ S(t) ≤ 1, and $0 \leq$ F(t) ≤ 1.
- f(t) is not a probability (it's a density); f(t) ≥ 0.

What do you think of when I say the word **hazard**?

What do you think of when I say the word **hazard**?

peril

What do you think of when I say the word **hazard**?

peril

threat

What do you think of when I say the word **hazard**?

peril

threat

risk

Hazard rate function, θ(t)

$$\theta(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)}$$

- Hazard rate at t equals the ratio of the pdf at t to the survivor function at t.
- Properties of hazard rate: $\theta(t) \geq 0$, but may be > 1!

Interpretation of θ(t)

$$\theta(t)\Delta(t) = \frac{f(t)\Delta(t)}{S(t)}$$

for some tiny interval of time Δt.

- Numerator is like a probability (recall areas under PDFs = probabilities):
  - $f(t)\Delta(t) \approx$ Pr(leaving the state in the interval $[t, t + \Delta t]$)
- Denominator is a probability and implies 'surviving at to time t'
- So, the expression for the hazard rate looks a bit like a conditional probability.

Pr(leaving the state in the interval $[t, t + \Delta t]$, **conditional on** survival time until t)

Interpretation of θ(t)

Recall the rules of conditional probability:

$$Pr(A|B) = Pr(A \cap B)/Pr(B)$$

$$Pr(A|B) = Pr(A \cap B)/Pr(B) = Pr(B|A)Pr(A)/Pr(B)$$

Now, let:

- A: "leaving the state in the interval $[t, t + \Delta(t)]$"
- B: "survival to time t"

It follows that:

Pr(leaving the state in the interval [t, t+ Δt], conditional on survival time until t)

is equivalent to:

Pr(A|B) = Pr(A)/Pr(B) since *Pr(B|A)* = 1

So, the continuous-time hazard rate has similarities to a conditional probability, but isn't a 'genuine' probability!

Conditional vs. unconditional probabilities

Contrast:

1. $\theta(\mathrm{t})\Delta(t)$
    - Probability for a person who has been unemployed for 120 days of leaving unemployment on the 121st day
    - Probability of dying at age 12 for someone who is aged 12

2. $\mathrm{f(t)}\Delta(t)$
    - Probability for persons entering unemployment of having a spell length of 121 days
    - Probability for a new-born baby of dying at age 12

Functions in discrete time

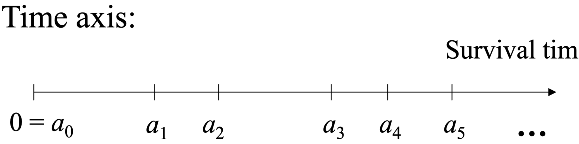
Discrete time concepts

Time scales might be:

1. Continuous time process but survival times measured in bands (grouped data; interval censoring)
2. Intrinsically discrete (e.g. time to conception measured as # menstrual cycles; time to machine breakdown measured as # machine cycles)

(a) Grouped data

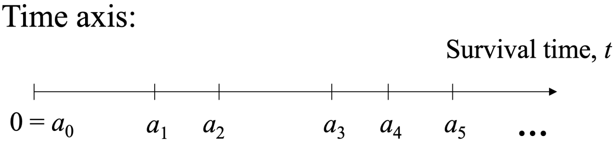
Consider the first case (grouped data; interval censoring) for now, and suppose underlying continuous time survival time T recorded in disjoint intervals (need not be of same length):



Intervals of time, indexed by $0 = a_0, a_1, a_2, a_3, \ldots, \ldots,$ where the intervals are $[0{=}a_0, a_1], (a_1, a_2], (a_2, a_3], \ldots, (a_{k-1}, a_k = \infty)$.

(a) Grouped data

Consider the first case (grouped data; interval censoring) for now, and suppose underlying continuous time survival time T recorded in disjoint intervals (need not be of same length):



Survivor function at start of j^{th} interval (just after end interval *j–1*):

$$S(t) = 1 - F(t) = Pr(T \geq a_{j-1}).$$

Probability of exit from state in j^{th} interval (interval density):

$$Pr(T \in (a_{j-1}, a_j]) = F(a_j) - F(a_{j-1}) = S(a_{j-1}) - S(a_j)$$

Discrete hazard rate of exit during jth interval: h(a$_j$)

$$h(a_j) = Pr(a_{j-1} < T \leq a_j | T > a_{j–1}) = \frac{Pr(a_{j-1} < T \leq a_j)}{Pr(T > a_{j–1})} = \frac{S(a_{j-1}) - S(a_j)}{S(a_{j–1})}$$

$h(a_j)$ is a probability (*unlike* continuous time hazard), and so $0 \leq h(a_j) \leq 1$

Easiest to consider the case in which every interval is of **unit length** so recorded duration intervals become $(t–1, t]$ with $t = 1, 2, 3, ...$ (positive integer); $T \in (t–1, t]$

Discrete time hazard rate and survivor functions

- Instead of indexing intervals using the date at end of each interval, let us index each interval directly.
- Thus refer to a spell of length j (i.e. one lasting to end of the j^{th} interval)
- Survivor function: Probability of survival to the end of interval j is the product of the probabilities of **not** experiencing the event in each of the intervals up to and including the current one, i.e. product of discrete hazards. For instance:

  $S_3 =$ (probability of survival through interval 1) ×
  (prob. of survival through int. 2, given survival through int. 1) ×
  (prob. of survival through int. 3, given survival through int. 2)

Discrete time hazard rate and survivor functions

- Instead of indexing intervals using the date at end of each interval, let us index each interval directly.
- Thus refer to a spell of length j (i.e. one lasting to end of the j^{th} interval)
- Survivor function: Probability of survival to the end of interval j is the product of the probabilities of **not** experiencing the event in each of the intervals up to and including the current one, i.e. product of discrete hazards:

$$S_j = (1 - h_1)(1 - h_2)...(1 - h_j) = \prod_{k=1}^{j}(1 - h_k)$$

Functional forms for the hazard rate

Why do we describe event history models with hazard functions?

- Given 1:1 relationships between hazard and density, failure, and survivor functions, we could specify our models in terms of any one of these.
- But typically done in terms of the hazard rate function (more closely related to the underlying behavioural processes)
- Shape that is empirically relevant, or suggested by theoretical models
  - likely to differ between applications (cf. human mortality, unemployment spell lengths, failure times of machine tools)
- Specification with convenient mathematical properties

Taxonomy of specifications

- Continuous time *vs.* discrete time models
  - differences in the assumptions about the survival time metric (whether underlying process, or way the data are recorded)

- Proportional Hazard (PH) *vs.* Accelerated Failure Time (AFT) *vs.* Proportional Odds models
  - differences in *interpretation* of a model and its parameters
  - some models have the PH property; others the AFT one

Continuous *vs.* discrete

- Continuous time parametric
  - Weibull (including Exponential)
  - Log-logistic
  - Log-normal
  - Gompertz
  - Generalised Gamma
- Continuous time semi-parametric
  - Piecewise Constant Exponential (PCE)
  - Cox model
- Discrete time (parametric & semiparametric)
  - Logistic
  - Complementary log-log ('cloglog')

A simple linear regression model vs. an event history model

A linear regression model (e.g., OLS):

$$y = \beta\mathbf{X} \equiv \beta_0 + \beta_1 X_1 + \beta_1 X_2 + ... + \beta_k X_k$$

A simple linear regression model vs. an event history model

A linear regression model (e.g., OLS):

$$y = \beta\mathbf{X} \equiv \beta_0 + \beta_1 X_1 + \beta_1 X_2 + ... + \beta_k X_k$$

A continuous time event-history model (e.g., Weibull):

$$\theta(t, \mathbf{X}) = f(t)\lambda(\mathbf{X}) = \underbrace{\alpha t^{\alpha-1}}_{f(t)} \underbrace{exp(\beta\mathbf{X})}_{\lambda(\mathbf{X})}$$

where

$$exp(\beta\mathbf{X}) \equiv exp(\beta_0 + \beta_1 X_1 + \beta_1 X_2 + ... + \beta_k X_k)$$

A simple linear regression model vs. an event history model

A linear regression model (e.g., OLS):

$$y = \beta \mathbf{X} \equiv \beta_0 + \beta_1 X_1 + \beta_1 X_2 + ... + \beta_k X_k$$

A continuous time event-history model (e.g., Weibull):

$$\theta(t, \mathbf{X}) = f(t)\lambda(\mathbf{X}) = \underbrace{\alpha t^{\alpha-1}}_{f(t)} \underbrace{exp(\beta \mathbf{X})}_{\lambda(\mathbf{X})}$$

where $\lambda(\mathbf{X})$ or, for simplicity, $\lambda \equiv \exp(\beta \mathbf{X}) > 0$

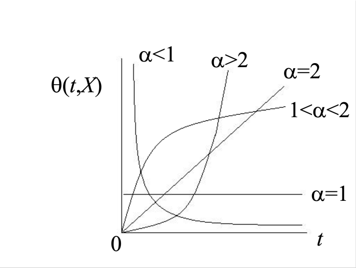
λ is a *scaling factor*: larger $\lambda \rightarrow$ larger hazard, at each t

$\alpha > 0$ is the *shape parameter*

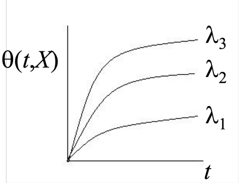
- $\alpha = 1$: *Exponential* model: hazard rate constant over time
- $\alpha > 1$: hazard monotonically increases with survival time
- $\alpha < 1$: hazard monotonically decreases with survival time

Weibull hazard function

- Variation in α keeping λ constant



- Variation in λ ($\lambda_3 > \lambda_2 > \lambda_1$) keeping α constant

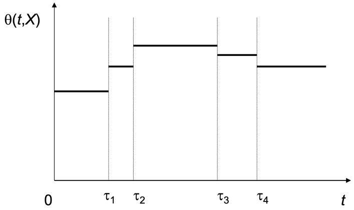


Intuitively, 'proportional hazard' means that hazard functions are multiplicative.

Piecewise constant Exponential (PCE) hazard

Allows flexibility in shape (of a sort).

Hazard constant within (user-specified) intervals, but may differ between them:



$$\theta(t,X) = \begin{cases} \bar{\theta}_1 exp(\beta \mathbf{X}) & t \in (0, \tau_1] \\ \bar{\theta}_2 exp(\beta \mathbf{X}) & t \in (\tau_1, \tau_2] \\ \vdots & \vdots \\ \bar{\theta}_k exp(\beta \mathbf{X}) & t \in (\tau_{k-1}, \tau_k] \end{cases}$$

Thus, PCE model is equivalent to having interval-specific intercept terms in the overall hazard rate $\theta(t,X)$

Interpreting models: proportional hazards (PH)

- PH models = 'multiplicative hazard' models = 'log relative hazard' models (for reasons, see below)
- All PH models satisfy a separability condition:

$$\theta(t, X) = \theta_0(t) exp(\beta \mathbf{X}) \rightarrow log[\theta(t, \mathbf{X})] = log[\theta_0(t)] + \beta \mathbf{X}$$

where

- θ_0: baseline hazard function depending on t, but not on $\mathbf{X}$. It summarises the pattern of *duration dependence*.
- $\exp(\beta \mathbf{X})$: non-negative function of $\mathbf{X}$, but not t.

PH model interpretations

Absolute differences in **X** imply proportionate differences in hazard (at each t)
For some $t = \bar{t}$, and for two persons i and j with characteristics vectors X_i and X_j,

$$\frac{\theta(\bar{t}, X_i)}{\theta(\bar{t}, X_j)} = \frac{exp(\beta X_i)}{exp(\beta X_j)} \equiv exp(\beta X_i - \beta X_j)$$

Likewise, in *log relative hazard* form

$$log\left[\frac{\theta(\bar{t}, X_i)}{\theta(\bar{t}, X_j)}\right] = \beta(X_i - X_j)$$

PH model interpretations

Each *regression coefficient* β_k (or log odds) summarises the proportional effect on the hazard of a **unit change** in the corresponding covariate X_k:

$$\beta_k = \partial log\theta(t, X)/\partial X_k$$

This proportional effect *does not vary* with survival time: β_k **does not vary with** t.

Proportionate change in the hazard given a one unit change in X_k, with all other covariates held fixed, i.e. *Hazard ratio* for X_k is $exp(\beta_k)$

PH model interpretations

Each *regression coefficient* β_k (or log odds) summarises the proportional effect on the hazard of a **unit change** in the corresponding covariate X_k:

$$\beta_k = \partial log\theta(t, X)/\partial X_{_k}$$

This proportional effect *does not vary* with survival time: β_k **does not vary with** t.

Proportionate change in the hazard given a one unit change in X_k, with all other covariates held fixed, i.e. *Hazard ratio* for X_k is $exp(\beta_k)$

Elasticity of the hazard with respect to X_k

$$\partial log[\theta(t, X)]/\partial[log(X_{_k})]$$

If covariate measured in logs: $X_k \equiv$log(Z$_k$), then it follows that β_k is the elasticity (% **change**) of the hazard with respect to Z_k .

Discrete time models

1. Complementary log-log (cloglog) model
   - We can interpret it as specification of a proportional hazards model: underlying survival process is continuous, but survival time data are recorded in bands (‘grouped’).
2. Discrete time logistic model
   - We can interpret it as specification of a proportional odds model for an intrinsically discrete survival process

We can apply both models to discrete time data.

We will focus on discrete time logistic, or logit, models, which are more frequently used.

Are you familiar with the odds?

p is a probability of an event P

$1 - p$ is the probability that an event P does not happen

The odds of an event P are :

$$\frac{p}{1-p}$$

We can also express the odds of hazards like this....

$$\frac{h(j, \mathbf{X})}{1 - h(j, \mathbf{X})}$$

Logistic hazard model as a discrete time proportional odds model

Suppose that the relative odds of failure in interval j, conditional on survival to end of interval j–1, take the proportional odds form:

$$\frac{h(j,\mathbf{X})}{1-h(j,\mathbf{X})} = \left[\frac{h_0(j)}{1-h_0(j)}\right] exp(\beta\mathbf{X})$$

for discrete hazard $h(j,\mathbf{X})$ for interval j. Hence,

$$logit[h(j,\mathbf{X})] = \log\left[\frac{h(j,\mathbf{X})}{1-h(j,\mathbf{X})}\right] = \log\left[\frac{h_0(j)}{1-h_0(j)}\right] + (\beta\mathbf{X}) = \alpha_j + \beta\mathbf{X}$$

Interpretation: logit(hazard for interval j) = linear function of characteristics ($\beta\mathbf{X}$), plus a duration-interval-specific parameter (α_j)

For $h \rightarrow 0$, then $log(1-h) \rightarrow 0$ too. In this case, like in a PH model: $log[\theta(t)] = log[\theta_0(t)] + \beta X$