Survival Models

Patrick Lam

February 1, 2008

Outline

Outline

What are survival models used for?

- Survival models aka duration models aka event history models

What are survival models used for?

- Survival models aka duration models aka event history models
- Dependent variable Y is the duration of time units spend in some state before experiencing an event (aka failure, death)

What are survival models used for?

- Survival models aka duration models aka event history models
- Dependent variable Y is the duration of time units spend in some state before experiencing an event (aka failure, death)
- Used in biostatistics and engineering: i.e. how long until a patient dies

What are survival models used for?

- Survival models aka duration models aka event history models
- Dependent variable Y is the duration of time units spend in some state before experiencing an event (aka failure, death)
- Used in biostatistics and engineering: i.e. how long until a patient dies
- Models the relationship between duration and covariates (how does an increase in X affect the duration Y)

What are survival models used for?

- Survival models aka duration models aka event history models
- Dependent variable Y is the duration of time units spend in some state before experiencing an event (aka failure, death)
- Used in biostatistics and engineering: i.e. how long until a patient dies
- Models the relationship between duration and covariates (how does an increase in X affect the duration Y)
- In political science, used in questions such as how long a coalition government lasts, how long a war lasts, how long a regime stays in power, or how long until a legislator leaves office.

What are survival models used for?

- Survival models aka duration models aka event history models
- Dependent variable Y is the duration of time units spend in some state before experiencing an event (aka failure, death)
- Used in biostatistics and engineering: i.e. how long until a patient dies
- Models the relationship between duration and covariates (how does an increase in X affect the duration Y)
- In political science, used in questions such as how long a coalition government lasts, how long a war lasts, how long a regime stays in power, or how long until a legislator leaves office.
- Observations should be on the same clock time, but not necessarily calendar time.

Why not use OLS?

Why not use OLS?

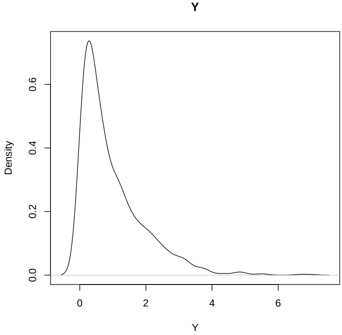
1. OLS assumes Y is Normal but duration dependent variables are always positive (number of years, number of days. etc.)

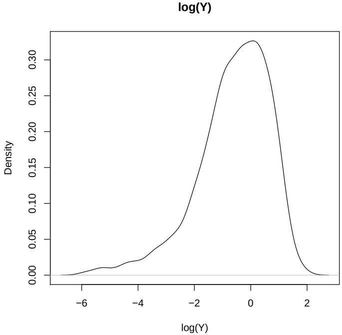
Why not use OLS?

1. OLS assumes Y is Normal but duration dependent variables are always positive (number of years, number of days. etc.)
   - Can possibly transform Y (log) to make it look Normal

Why not use OLS?

1. OLS assumes Y is Normal but duration dependent variables are always positive (number of years, number of days. etc.)
   - Can possibly transform Y (log) to make it look Normal



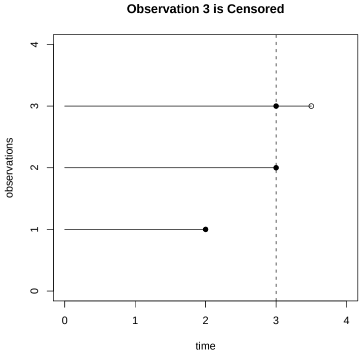


Why not use OLS?

2. Survival models can handle censoring.

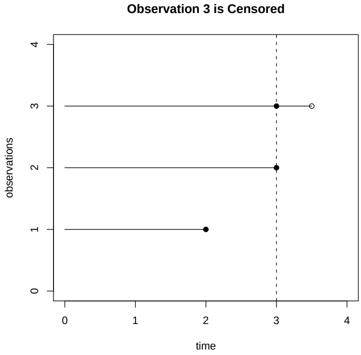
Why not use OLS?

2. Survival models can handle censoring.



Why not use OLS?

2. Survival models can handle censoring.



Observation 3 is censored in that it has not experienced the event at the time we stop collecting data, so we don't know its true duration.

Why not use OLS?

3. Survival models can handle time-varying covariates (TVCs).

Why not use OLS?

3. Survival models can handle time-varying covariates (TVCs).
   - If Y is duration of a regime, GDP may change during the duration of the regime.

Why not use OLS?

3. Survival models can handle time-varying covariates (TVCs).
   - If Y is duration of a regime, GDP may change during the duration of the regime.
   - OLS cannot handle multiple values of GDP per observation.

Why not use OLS?

3. Survival models can handle time-varying covariates (TVCs).
   - If Y is duration of a regime, GDP may change during the duration of the regime.
   - OLS cannot handle multiple values of GDP per observation.
   - You can set up data in a special way with survival models such that you can accomodate TVCs (not going to talk about this today).

Outline

Probability Density Function

Probability Density Function

Let T be a continuous positive random variable denoting the duration/survival times ($T = Y$).

Probability Density Function

Let T be a continuous positive random variable denoting the duration/survival times ($T = Y$).

T has a **probability density function** $f(t)$: roughly *probability that an event occurs at time* t.

Probability Density Function

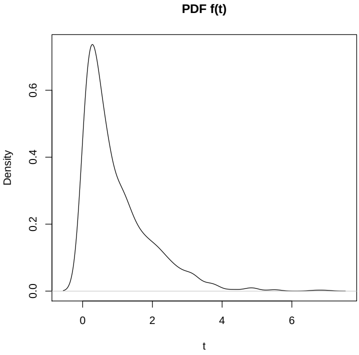
Let T be a continuous positive random variable denoting the duration/survival times ($T = Y$).

T has a **probability density function** $f(t)$: roughly *probability that an event occurs at time* t. Because $f(t)$ is continuous, to get the probability, we would have to find the area under the density for an infinitely small interval around t.

Probability Density Function

Let T be a continuous positive random variable denoting the duration/survival times ($T = Y$).

T has a **probability density function** $f(t)$: roughly *probability that an event occurs at time* t. Because $f(t)$ is continuous, to get the probability, we would have to find the area under the density for an infinitely small interval around t.



Survivor Function

Survivor Function

$f(t)$ has a corresponding CDF $F(t) = \int_0^t f(u)du = P(T \leq t)$, which is the probability of an event occurring by time t.

Survivor Function

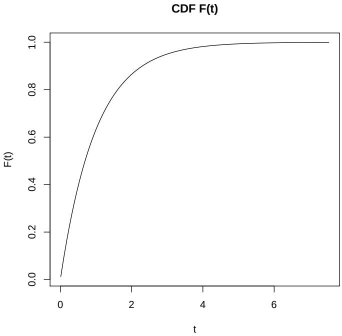
$f(t)$ has a corresponding CDF $F(t) = \int_0^t f(u)du = P(T \leq t)$, which is the probability of an event occurring by time t.

Then the **survivor function** is $S(t) = 1 - F(t)$: *probability of surviving (no event) until at least time t*

Survivor Function

$f(t)$ has a corresponding CDF $F(t) = \int_0^t f(u)du = P(T \leq t)$, which is the probability of an event occurring by time t.

Then the **survivor function** is $S(t) = 1 - F(t)$: *probability of surviving (no event) until at least time t*



Hazard Rate

Hazard Rate

The **hazard rate** (or hazard function) $h(t)$ is roughly *the probability of an event at time t given survival up to time t*.

Hazard Rate

The **hazard rate** (or hazard function) $h(t)$ is roughly *the probability of an event at time t given survival up to time t*.

$$h(t) = P(\text{event at } t | \text{survival up to } t)$$

Hazard Rate

The **hazard rate** (or hazard function) $h(t)$ is roughly *the probability of an event at time t given survival up to time t*.

$$
\begin{aligned}
h(t) &= P(\text{event at } t | \text{survival up to } t) \\
&= \frac{P(\text{survival up to } t | \text{event at } t) P(\text{event at } t)}{P(\text{survival up to } t)}
\end{aligned}
$$

Hazard Rate

The **hazard rate** (or hazard function) $h(t)$ is roughly *the probability of an event at time t given survival up to time t.*

$$
\begin{aligned}
h(t) &= P(\text{event at } t | \text{survival up to } t) \\
&= \frac{P(\text{survival up to } t | \text{event at } t) P(\text{event at } t)}{P(\text{survival up to } t)} \\
&= \frac{P(\text{event at } t)}{P(\text{survival up to } t)}
\end{aligned}
$$

Hazard Rate

The **hazard rate** (or hazard function) $h(t)$ is roughly *the probability of an event at time t given survival up to time t*.

$$\begin{aligned} h(t) &= P(\text{event at } t | \text{survival up to } t) \\ &= \frac{P(\text{survival up to } t | \text{event at } t) P(\text{event at } t)}{P(\text{survival up to } t)} \\ &= \frac{P(\text{event at } t)}{P(\text{survival up to } t)} \\ &= \frac{f(t)}{S(t)} \end{aligned}$$

Hazard Rate

The **hazard rate** (or hazard function) $h(t)$ is roughly *the probability of an event at time t given survival up to time t*.

$$
\begin{aligned}
h(t) &= P(\text{event at } t | \text{survival up to } t) \\
&= \frac{P(\text{survival up to } t | \text{event at } t) P(\text{event at } t)}{P(\text{survival up to } t)} \\
&= \frac{P(\text{event at } t)}{P(\text{survival up to } t)} \\
&= \frac{f(t)}{S(t)} \\
f(t) &= h(t) S(t)
\end{aligned}
$$

- The hazard rate has a substantive interpretation:

- The hazard rate has a substantive interpretation:
  - Given a government has survived 2 years, what is the probability it will collapse?

- The hazard rate has a substantive interpretation:
  - Given a government has survived 2 years, what is the probability it will collapse?
- Note that the hazard rate is a function of time:

- The hazard rate has a substantive interpretation:
  - Given a government has survived 2 years, what is the probability it will collapse?
- Note that the hazard rate is a function of time:
  - Increasing hazard: A government that has survived 2 years is more likely to collapse than one that has survived 1 year.

- The hazard rate has a substantive interpretation:
  - Given a government has survived 2 years, what is the probability it will collapse?
- Note that the hazard rate is a function of time:
  - Increasing hazard: A government that has survived 2 years is more likely to collapse than one that has survived 1 year.
  - Decreasing hazard: A government that has survived 2 years is less likely to collapse than one that has survived 1 year.

- The hazard rate has a substantive interpretation:
  - Given a government has survived 2 years, what is the probability it will collapse?
- Note that the hazard rate is a function of time:
  - Increasing hazard: A government that has survived 2 years is more likely to collapse than one that has survived 1 year.
  - Decreasing hazard: A government that has survived 2 years is less likely to collapse than one that has survived 1 year.
  - Flat hazard: A government that has survived 2 years is no more or less likely to collapse than one that has survived 1 year.

- The hazard rate has a substantive interpretation:
  - Given a government has survived 2 years, what is the probability it will collapse?
- Note that the hazard rate is a function of time:
  - Increasing hazard: A government that has survived 2 years is more likely to collapse than one that has survived 1 year.
  - Decreasing hazard: A government that has survived 2 years is less likely to collapse than one that has survived 1 year.
  - Flat hazard: A government that has survived 2 years is no more or less likely to collapse than one that has survived 1 year.
- Parametric models usually assume some shape for the hazard rate (i.e. flat, monotonic, etc.).

Modeling with Covariates

Modeling with Covariates

We usually model the hazard rate as a function of covariates.

Modeling with Covariates

We usually model the hazard rate as a function of covariates.

$$h_i(t) = g(X_i\beta)$$

Modeling with Covariates

We usually model the hazard rate as a function of covariates.

$$h_i(t) = g(X_i\beta)$$

We can then interpret the change in the hazard rate and also the change in average duration given a change in X (via the math from before).

Modeling with Covariates

We usually model the hazard rate as a function of covariates.

$$h_i(t) = g(X_i\beta)$$

We can then interpret the change in the hazard rate and also the change in average duration given a change in X (via the math from before).

When all the covariates are 0, $h_i(t) = g(\beta_0)$. We call this the **baseline hazard**.

Estimation of the Parameters

Question: How do we estimate the parameters of interest?

Estimation of the Parameters

Question: How do we estimate the parameters of interest?

Answer: Use maximum likelihood estimation.

A Brief Review (or Preview) of Maximum Likelihood

A Brief Review (or Preview) of Maximum Likelihood

We want to find a set of parameters θ (which include β and possibly other ancillary parameters) given that we have data y.

A Brief Review (or Preview) of Maximum Likelihood

We want to find a set of parameters θ (which include β and possibly other ancillary parameters) given that we have data y.

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)}$$

A Brief Review (or Preview) of Maximum Likelihood

We want to find a set of parameters θ (which include β and possibly other ancillary parameters) given that we have data y.

$$\begin{aligned} P(\theta|y) &= \frac{P(y|\theta)P(\theta)}{P(y)} \\ &= k(y)P(y|\theta) \end{aligned}$$

A Brief Review (or Preview) of Maximum Likelihood

We want to find a set of parameters θ (which include β and possibly other ancillary parameters) given that we have data y.

$$\begin{aligned} P(\theta|y) &= \frac{P(y|\theta)P(\theta)}{P(y)} \\ &= k(y)P(y|\theta) \\ &\propto P(y|\theta) \end{aligned}$$

A Brief Review (or Preview) of Maximum Likelihood

We want to find a set of parameters θ (which include β and possibly other ancillary parameters) given that we have data y.

$$\begin{aligned} P(\theta|y) &= \frac{P(y|\theta)P(\theta)}{P(y)} \\ &= k(y)P(y|\theta) \\ &\propto P(y|\theta) \\ \mathcal{L}(\theta|y) &= \prod_{i=1}^{n} P(y_i|\theta) \text{ by i.i.d} \end{aligned}$$

A Brief Review (or Preview) of Maximum Likelihood

We want to find a set of parameters θ (which include β and possibly other ancillary parameters) given that we have data y.

$$
\begin{aligned}
P(\theta|y) &= \frac{P(y|\theta)P(\theta)}{P(y)} \\
&= k(y)P(y|\theta) \\
&\propto P(y|\theta) \\
\mathcal{L}(\theta|y) &= \prod_{i=1}^{n} P(y_i|\theta) \text{ by i.i.d}
\end{aligned}
$$

The maximum point of $\mathcal{L}(\theta|y)$ in multi-dimensional space is the MLE of our parameters.

A Brief Review (or Preview) of Maximum Likelihood

We want to find a set of parameters θ (which include β and possibly other ancillary parameters) given that we have data y.

$$\begin{aligned} P(\theta|y) &= \frac{P(y|\theta)P(\theta)}{P(y)} \\ &= k(y)P(y|\theta) \\ &\propto P(y|\theta) \\ \mathcal{L}(\theta|y) &= \prod_{i=1}^{n} P(y_i|\theta) \text{ by i.i.d} \end{aligned}$$

The maximum point of $\mathcal{L}(\theta|y)$ in multi-dimensional space is the MLE of our parameters. In our case,

$$P(y|\theta) = f(t)$$

What about Censoring?

What about Censoring?

Observations that are censored give us no information about when the event occurs, but they do give us information about how long they survive.

What about Censoring?

Observations that are censored give us no information about when the event occurs, but they do give us information about how long they survive.

For censored observations, we know that they survived at least up to some observed time t^c and that their true duration is some $t^* \geq t^c$.

What about Censoring?

Observations that are censored give us no information about when the event occurs, but they do give us information about how long they survive.

For censored observations, we know that they survived at least up to some observed time t^c and that their true duration is some $t^* \geq t^c$.

For each observation, create a censoring indicator c_i such that

$$c_i = \begin{cases} 1 & \text{if not censored} \\ 0 & \text{if censored} \end{cases}$$

We can incorporate the information from the censored observations into the likelihood function.

We can incorporate the information from the censored observations into the likelihood function.

$$\mathcal{L} = \prod_{i=1}^{n} [f(t_i)]^{c_i} \; [P(T_i^* \geq t_i^c)]^{1-c_i}$$

We can incorporate the information from the censored observations into the likelihood function.

$$
\begin{aligned}
\mathcal{L} &= \prod_{i=1}^{n} [f(t_i)]^{c_i} \; [P(T_i^* \geq t_i^c)]^{1-c_i} \\
&= \prod_{i=1}^{n} [f(t_i)]^{c_i} \; [1 - F(t_i)]^{1-c_i}
\end{aligned}
$$

We can incorporate the information from the censored observations into the likelihood function.

$$
\begin{aligned}
\mathcal{L} &= \prod_{i=1}^{n} [f(t_i)]^{c_i} \; [P(T_i^* \geq t_i^c)]^{1-c_i} \\
&= \prod_{i=1}^{n} [f(t_i)]^{c_i} \; [1 - F(t_i)]^{1-c_i} \\
&= \prod_{i=1}^{n} [f(t_i)]^{c_i} \; [S(t_i)]^{1-c_i}
\end{aligned}
$$

We can incorporate the information from the censored observations into the likelihood function.

$$
\begin{aligned}
\mathcal{L} &= \prod_{i=1}^{n} [f(t_i)]^{c_i} \; [P(T_i^* \geq t_i^c)]^{1-c_i} \\
&= \prod_{i=1}^{n} [f(t_i)]^{c_i} \; [1 - F(t_i)]^{1-c_i} \\
&= \prod_{i=1}^{n} [f(t_i)]^{c_i} \; [S(t_i)]^{1-c_i}
\end{aligned}
$$

So uncensored observations contribute to the density function and censored observations contribute to the survivor function in the likelihood.

Outline

Estimating Parametric Survival Models

Estimating Parametric Survival Models

1. Make an assumption that T_i follows a specific distribution $f(t)$ (stochastic component).

Estimating Parametric Survival Models

1. Make an assumption that T_i follows a specific distribution $f(t)$ (stochastic component).
    - By making an assumption about $f(t)$, you are also making an assumption about the shape of $h(t)$.

Estimating Parametric Survival Models

1. Make an assumption that T_i follows a specific distribution $f(t)$ (stochastic component).
   - By making an assumption about $f(t)$, you are also making an assumption about the shape of $h(t)$.
2. Model the hazard rate with covariates (systematic component).

Estimating Parametric Survival Models

1. Make an assumption that T_i follows a specific distribution $f(t)$ (stochastic component).
   - By making an assumption about $f(t)$, you are also making an assumption about the shape of $h(t)$.
2. Model the hazard rate with covariates (systematic component).
3. Estimate via ML.

Estimating Parametric Survival Models

1. Make an assumption that T_i follows a specific distribution $f(t)$ (stochastic component).
   - By making an assumption about $f(t)$, you are also making an assumption about the shape of $h(t)$.
2. Model the hazard rate with covariates (systematic component).
3. Estimate via ML.
4. Interpret quantities of interest (hazard ratios, expected survival times).

Running Example: Duration of Parliamentary Cabinets

Running Example: Duration of Parliamentary Cabinets

- Example taken from King et al. (1990).

Running Example: Duration of Parliamentary Cabinets

- Example taken from King et al. (1990).
- Dependent variable: number of months a coalition government stays in power

Running Example: Duration of Parliamentary Cabinets

- Example taken from King et al. (1990).
- Dependent variable: number of months a coalition government stays in power
- Event: fall of a coalition government

Running Example: Duration of Parliamentary Cabinets

- Example taken from King et al. (1990).
- Dependent variable: number of months a coalition government stays in power
- Event: fall of a coalition government
- Independent variables:

Running Example: Duration of Parliamentary Cabinets

- Example taken from King et al. (1990).
- Dependent variable: number of months a coalition government stays in power
- Event: fall of a coalition government
- Independent variables:
  - investiture (`invest`): legal requirement for legislature to approve cabinet

Running Example: Duration of Parliamentary Cabinets

- Example taken from King et al. (1990).
- Dependent variable: number of months a coalition government stays in power
- Event: fall of a coalition government
- Independent variables:
  - investiture (`invest`): legal requirement for legislature to approve cabinet
  - fractionalization (`fract`): index characterizing the number and size of parties in parliament, where higher numbers indicate larger number of small blocs

Running Example: Duration of Parliamentary Cabinets

- Example taken from King et al. (1990).
- Dependent variable: number of months a coalition government stays in power
- Event: fall of a coalition government
- Independent variables:
    - investiture (`invest`): legal requirement for legislature to approve cabinet
    - fractionalization (`fract`): index characterizing the number and size of parties in parliament, where higher numbers indicate larger number of small blocs
    - polarization (`polar`): measure of support for extremist parties

Running Example: Duration of Parliamentary Cabinets

- Example taken from King et al. (1990).
- Dependent variable: number of months a coalition government stays in power
- Event: fall of a coalition government
- Independent variables:
  - investiture (`invest`): legal requirement for legislature to approve cabinet
  - fractionalization (`fract`): index characterizing the number and size of parties in parliament, where higher numbers indicate larger number of small blocs
  - polarization (`polar`): measure of support for extremist parties
  - numerical status (`numst2`): dummy variable coding majority (1) or minority (0) government

Running Example: Duration of Parliamentary Cabinets

- Example taken from King et al. (1990).
- Dependent variable: number of months a coalition government stays in power
- Event: fall of a coalition government
- Independent variables:
  - investiture (`invest`): legal requirement for legislature to approve cabinet
  - fractionalization (`fract`): index characterizing the number and size of parties in parliament, where higher numbers indicate larger number of small blocs
  - polarization (`polar`): measure of support for extremist parties
  - numerical status (`numst2`): dummy variable coding majority (1) or minority (0) government
  - crisis duration (`crisis`): number of days of "crisis" before government is formed

Running Example: Duration of Parliamentary Cabinets

- Example taken from King et al. (1990).
- Dependent variable: number of months a coalition government stays in power
- Event: fall of a coalition government
- Independent variables:
  - investiture (`invest`): legal requirement for legislature to approve cabinet
  - fractionalization (`fract`): index characterizing the number and size of parties in parliament, where higher numbers indicate larger number of small blocs
  - polarization (`polar`): measure of support for extremist parties
  - numerical status (`numst2`): dummy variable coding majority (1) or minority (0) government
  - crisis duration (`crisis`): number of days of "crisis" before government is formed
- Censoring occurs because of constitutionally mandated elections: governments fall apart in anticipation of such elections

The Exponential Model

The Exponential Model

Assume:

$$T_i \quad \sim \quad \text{Exponential}(\lambda_i)$$

The Exponential Model

Assume:

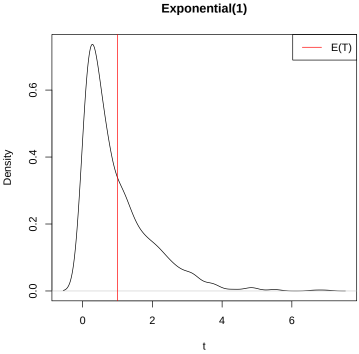
$$
\begin{aligned}
T_i &\sim \text{Exponential}(\lambda_i) \\
E(T_i) &= \frac{1}{\lambda_i}
\end{aligned}
$$

The Exponential Model

Assume:

$$\begin{aligned} T_i &\sim \text{Exponential}(\lambda_i) \\ E(T_i) &= \frac{1}{\lambda_i} \end{aligned}$$

$\lambda_i > 0$ and is known as the rate parameter.



$$f(t) = \lambda e^{-\lambda t}$$

$$
\begin{aligned}
f(t) &= \lambda e^{-\lambda t} \\
S(t) &= 1 - F(t)
\end{aligned}
$$

$$
\begin{aligned}
f(t) &= \lambda e^{-\lambda t} \\
\\
S(t) &= 1 - F(t) \\
&= 1 - (1 - e^{-\lambda t})
\end{aligned}
$$

$$
\begin{aligned}
f(t) &= \lambda e^{-\lambda t} \\
\\
S(t) &= 1 - F(t) \\
&= 1 - (1 - e^{-\lambda t}) \\
&= e^{-\lambda t}
\end{aligned}
$$

$$f(t) = \lambda e^{-\lambda t}$$

$$\begin{aligned} S(t) &= 1 - F(t) \\ &= 1 - (1 - e^{-\lambda t}) \\ &= e^{-\lambda t} \end{aligned}$$

$$h(t) = \frac{f(t)}{S(t)}$$

$$
\begin{aligned}
f(t) &= \lambda e^{-\lambda t} \\[2em]
S(t) &= 1 - F(t) \\
&= 1 - (1 - e^{-\lambda t}) \\
&= e^{-\lambda t} \\[2em]
h(t) &= \frac{f(t)}{S(t)} \\
&= \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}}
\end{aligned}
$$

$$
\begin{aligned}
f(t) &= \lambda e^{-\lambda t}
\end{aligned}
$$

$$
\begin{aligned}
S(t) &= 1 - F(t) \\
&= 1 - (1 - e^{-\lambda t}) \\
&= e^{-\lambda t}
\end{aligned}
$$

$$
\begin{aligned}
h(t) &= \frac{f(t)}{S(t)} \\
&= \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} \\
&= \lambda
\end{aligned}
$$

Modeling $h(t)$ with covariates:

Modeling $h(t)$ with covariates:

$$h(t_i) = \lambda_i$$

Modeling $h(t)$ with covariates:

$$h(t_i) \quad = \quad \lambda_i$$

We want to constrain the $h(t)$ to be positive.

Modeling $h(t)$ with covariates:

$$h(t_i) \quad = \quad \lambda_i$$

We want to constrain the $h(t)$ to be positive.

$$\lambda_i \quad = \quad e^{-\mathbf{x}_i\beta}$$

Modeling $h(t)$ with covariates:

$$h(t_i) \quad = \quad \lambda_i$$

We want to constrain the $h(t)$ to be positive.

$$\lambda_i \quad = \quad e^{-\mathbf{x}_i \beta}$$

Positive β implies that hazard decreases and average survival time increases as x increases.

Modeling $h(t)$ with covariates:

$$h(t_i) \quad = \quad \lambda_i$$

We want to constrain the $h(t)$ to be positive.

$$\lambda_i \quad = \quad e^{-\mathbf{x}_i \beta}$$

Positive β implies that hazard decreases and average survival time increases as x increases.

Because $h(t_i)$ is modeled with only one parameter λ_i and is not a function of t_i, the exponential model assumes a **flat hazard** (no time dependence).

Modeling $h(t)$ with covariates:

$$h(t_i) = \lambda_i$$

We want to constrain the $h(t)$ to be positive.

$$\lambda_i = e^{-\mathbf{x}_i \beta}$$

Positive β implies that hazard decreases and average survival time increases as x increases.

Because $h(t_i)$ is modeled with only one parameter λ_i and is not a function of t_i, the exponential model assumes a **flat hazard** (no time dependence). This corresponds to the memoryless property of the exponential distribution.

Estimation via ML:

Estimation via ML:

$$\mathcal{L} = \prod_{i=1}^{n} [f(t_i)]^{c_i} \; [S(t_i)]^{1-c_i}$$

Estimation via ML:

$$
\begin{aligned}
\mathcal{L} &= \prod_{i=1}^{n} [f(t_i)]^{c_i} \; [S(t_i)]^{1-c_i} \\
&= \prod_{i=1}^{n} \left[\lambda_i e^{-\lambda_i t_i}\right]^{c_i} \; \left[e^{-\lambda_i t_i}\right]^{1-c_i}
\end{aligned}
$$

Estimation via ML:

$$\begin{aligned}
\mathcal{L} &= \prod_{i=1}^{n} [f(t_i)]^{c_i} \; [S(t_i)]^{1-c_i} \\
&= \prod_{i=1}^{n} \left[\lambda_i e^{-\lambda_i t_i}\right]^{c_i} \; \left[e^{-\lambda_i t_i}\right]^{1-c_i} \\
\ln \mathcal{L} &= \sum_{i=1}^{n} c_i \left(\ln \lambda_i - \lambda_i t_i\right) + (1 - c_i)(-\lambda_i t_i)
\end{aligned}$$

Estimation via ML:

$$
\begin{aligned}
\mathcal{L} &= \prod_{i=1}^{n} [f(t_i)]^{c_i} \; [S(t_i)]^{1-c_i} \\
&= \prod_{i=1}^{n} \left[\lambda_i e^{-\lambda_i t_i}\right]^{c_i} \; \left[e^{-\lambda_i t_i}\right]^{1-c_i} \\
\ln \mathcal{L} &= \sum_{i=1}^{n} c_i \left(\ln \lambda_i - \lambda_i t_i\right) + (1-c_i)(-\lambda_i t_i) \\
&= \sum_{i=1}^{n} c_i \left(\ln e^{-\mathbf{x}_i\beta} - e^{-\mathbf{x}_i\beta} t_i\right) + (1-c_i)(-e^{-\mathbf{x}_i\beta} t_i)
\end{aligned}
$$

Estimation via ML:

$$
\begin{aligned}
\mathcal{L} &= \prod_{i=1}^{n} [f(t_i)]^{c_i} \; [S(t_i)]^{1-c_i} \\
&= \prod_{i=1}^{n} \left[\lambda_i e^{-\lambda_i t_i}\right]^{c_i} \; \left[e^{-\lambda_i t_i}\right]^{1-c_i} \\
\ln \mathcal{L} &= \sum_{i=1}^{n} c_i \left(\ln \lambda_i - \lambda_i t_i\right) + (1-c_i)(-\lambda_i t_i) \\
&= \sum_{i=1}^{n} c_i \left(\ln e^{-\mathbf{x}_i \beta} - e^{-\mathbf{x}_i \beta} t_i\right) + (1-c_i)(-e^{-\mathbf{x}_i \beta} t_i) \\
&= \sum_{i=1}^{n} c_i \left(\ln e^{-\mathbf{x}_i \beta} - e^{-\mathbf{x}_i \beta} t_i + e^{-\mathbf{x}_i \beta} t_i\right) - e^{-\mathbf{x}_i \beta} t_i
\end{aligned}
$$

Estimation via ML:

$$
\begin{aligned}
\mathcal{L} &= \prod_{i=1}^{n} [f(t_i)]^{c_i} \; [S(t_i)]^{1-c_i} \\
&= \prod_{i=1}^{n} \left[\lambda_i e^{-\lambda_i t_i}\right]^{c_i} \; \left[e^{-\lambda_i t_i}\right]^{1-c_i} \\
\ln \mathcal{L} &= \sum_{i=1}^{n} c_i \left(\ln \lambda_i - \lambda_i t_i\right) + (1 - c_i)(-\lambda_i t_i) \\
&= \sum_{i=1}^{n} c_i \left(\ln e^{-\mathbf{x}_i \beta} - e^{-\mathbf{x}_i \beta} t_i\right) + (1 - c_i)(-e^{-\mathbf{x}_i \beta} t_i) \\
&= \sum_{i=1}^{n} c_i \left(\ln e^{-\mathbf{x}_i \beta} - e^{-\mathbf{x}_i \beta} t_i + e^{-\mathbf{x}_i \beta} t_i\right) - e^{-\mathbf{x}_i \beta} t_i \\
&= \sum_{i=1}^{n} c_i \left(-\mathbf{x}_i \beta\right) - e^{-\mathbf{x}_i \beta} t_i
\end{aligned}
$$

$$\ln \mathcal{L} = \sum_{i=1}^{n} c_i \left(-\mathbf{x}_i\beta\right) - e^{-\mathbf{x}_i\beta} t_i$$

$$\ln \mathcal{L} = \sum_{i=1}^{n} c_i \left(-\mathbf{x}_i\beta\right) - e^{-\mathbf{x}_i\beta} t_i$$

```
> data(coalition)

> X <- as.matrix(cbind(1, coalition[, c("invest", "fract", "polar",
+     "numst2", "crisis")]))
> T <- coalition$duration
> C <- coalition$ciep12
```

$$\ln \mathcal{L} = \sum_{i=1}^{n} c_i \left(-\mathbf{x}_i\beta\right) - e^{-\mathbf{x}_i\beta} t_i$$

```
> data(coalition)

> X <- as.matrix(cbind(1, coalition[, c("invest", "fract", "polar",
+     "numst2", "crisis")]))
> T <- coalition$duration
> C <- coalition$ciep12

> expo.lik <- function(par, T, X, C) {
+     beta <- par
+     lambda <- exp(-(X %*% beta))
+     log.lik <- sum(C * (-(X %*% beta)) - (lambda * T))
+     return(log.lik)
+ }
```

$$\ln \mathcal{L} = \sum_{i=1}^{n} c_i \left(-\mathbf{x}_i\beta\right) - e^{-\mathbf{x}_i\beta} t_i$$

```
> data(coalition)

> X <- as.matrix(cbind(1, coalition[, c("invest", "fract", "polar",
+     "numst2", "crisis")]))
> T <- coalition$duration
> C <- coalition$ciep12

> expo.lik <- function(par, T, X, C) {
+     beta <- par
+     lambda <- exp(-(X %*% beta))
+     log.lik <- sum(C * (-(X %*% beta)) - (lambda * T))
+     return(log.lik)
+ }

> my.coef <- optim(par = c(0, 0, 0, 0, 0, 0), fn = expo.lik, T = T,
+     X = X, C = C, method = "BFGS", control = list(fnscale = -1))$par
```

Or we can use the `survival` (or `Zelig`) package in R:

Or we can use the `survival` (or `Zelig`) package in R:

```
> library(survival)
```

Or we can use the `survival` (or `Zelig`) package in R:

```
> library(survival)

> expo.surv <- survreg(Surv(duration, ciep12) ~ invest + fract +
+     polar + numst2 + crisis, data = coalition, dist = "exponential")
```

Or we can use the `survival` (or `Zelig`) package in R:

```
> library(survival)

> expo.surv <- survreg(Surv(duration, ciep12) ~ invest + fract +
+     polar + numst2 + crisis, data = coalition, dist = "exponential")

> expo.surv$coef

(Intercept)      invest       fract       polar      numst2      crisis
   4.826723   -0.504758   -2.250355   -0.028796    0.461321    0.005587

> my.coef

[1]  4.828623 -0.504985 -2.253515 -0.028797  0.461015  0.005603
```

Or we can use the survival (or Zelig) package in R:

```
> library(survival)

> expo.surv <- survreg(Surv(duration, ciep12) ~ invest + fract +
+     polar + numst2 + crisis, data = coalition, dist = "exponential")

> expo.surv$coef

(Intercept)      invest       fract       polar      numst2      crisis
   4.826723   -0.504758   -2.250355   -0.028796    0.461321    0.005587

> my.coef

[1]  4.828623 -0.504985 -2.253515 -0.028797  0.461015  0.005603

> expo.surv$loglik[2]

[1] -1046

> expo.lik(par = my.coef, X = X, T = T, C = C)

[1] -1046
```

How do we get quantities of interest?

How do we get quantities of interest?

Variable of interest: majority versus minority governments (`numst2`), with all other variables set at mean or mode.

How do we get quantities of interest?

Variable of interest: majority versus minority governments (`numst2`), with all other variables set at mean or mode.

```
> x.min <- colMeans(model.matrix(expo.surv))
> x.min[c("invest", "numst2")] <- 0
> x.maj <- x.min
> x.maj["numst2"] <- 1
```

How do we get quantities of interest?

Variable of interest: majority versus minority governments (`numst2`), with all other variables set at mean or mode.

```
> x.min <- colMeans(model.matrix(expo.surv))
> x.min[c("invest", "numst2")] <- 0
> x.maj <- x.min
> x.maj["numst2"] <- 1

> x.min

(Intercept)      invest       fract       polar      numst2      crisis
     1.0000      0.0000      0.7188     15.2898      0.0000     22.3822

> x.maj

(Intercept)      invest       fract       polar      numst2      crisis
     1.0000      0.0000      0.7188     15.2898      1.0000     22.3822
```

Simulating Estimation Uncertainty:

How do we get quantities of interest?

Variable of interest: majority versus minority governments (`numst2`), with all other variables set at mean or mode.

```
> x.min <- colMeans(model.matrix(expo.surv))
> x.min[c("invest", "numst2")] <- 0
> x.maj <- x.min
> x.maj["numst2"] <- 1

> x.min

(Intercept)      invest       fract       polar      numst2      crisis
     1.0000      0.0000      0.7188     15.2898      0.0000     22.3822

> x.maj

(Intercept)      invest       fract       polar      numst2      crisis
     1.0000      0.0000      0.7188     15.2898      1.0000     22.3822
```

Simulating Estimation Uncertainty:

```
> betas <- mvrnorm(1000, mu = expo.surv$coef, Sigma = vcov(expo.surv))
```

Hazard Ratios:

$$\mathrm{HR} = \frac{h(t|\mathbf{x}_{\mathrm{maj}})}{h(t|\mathbf{x}_{\mathrm{min}})}$$

Hazard Ratios:

$$
\begin{aligned}
\mathrm{HR} &= \frac{h(t|\mathbf{x}_{\mathrm{maj}})}{h(t|\mathbf{x}_{\mathrm{min}})} \\
&= \frac{e^{-\mathbf{x}_{\mathrm{maj}}\beta}}{e^{-\mathbf{x}_{\mathrm{min}}\beta}}
\end{aligned}
$$

Hazard Ratios:

$$\begin{aligned}
\mathrm{HR} &= \frac{h(t|\mathbf{x}_{\mathrm{maj}})}{h(t|\mathbf{x}_{\mathrm{min}})} \\
&= \frac{e^{-\mathbf{x}_{\mathrm{maj}}\beta}}{e^{-\mathbf{x}_{\mathrm{min}}\beta}} \\
&= \frac{e^{-\beta_0}\, e^{-x_1\beta_1}\, e^{-x_2\beta_2}\, e^{-x_3\beta_3}\, e^{-x_{\mathrm{maj}}\beta_4}\, e^{-x_5\beta_5}}{e^{-\beta_0}\, e^{-x_1\beta_1}\, e^{-x_2\beta_2}\, e^{-x_3\beta_3}\, e^{-x_{\mathrm{min}}\beta_4}\, e^{-x_5\beta_5}}
\end{aligned}$$

Hazard Ratios:

$$\begin{aligned}
\mathrm{HR} &= \frac{h(t|\mathbf{x}_{\mathrm{maj}})}{h(t|\mathbf{x}_{\mathrm{min}})} \\
&= \frac{e^{-\mathbf{x}_{\mathrm{maj}}\beta}}{e^{-\mathbf{x}_{\mathrm{min}}\beta}} \\
&= \frac{e^{-\beta_0}\, e^{-x_1\beta_1}\, e^{-x_2\beta_2}\, e^{-x_3\beta_3}\, e^{-x_{\mathrm{maj}}\beta_4}\, e^{-x_5\beta_5}}{e^{-\beta_0}\, e^{-x_1\beta_1}\, e^{-x_2\beta_2}\, e^{-x_3\beta_3}\, e^{-x_{\mathrm{min}}\beta_4}\, e^{-x_5\beta_5}} \\
&= \frac{e^{-x_{\mathrm{maj}}\beta_4}}{e^{-x_{\mathrm{min}}\beta_4}}
\end{aligned}$$

Hazard Ratios:

$$
\begin{aligned}
\mathrm{HR} &= \frac{h(t|\mathbf{x}_{\mathrm{maj}})}{h(t|\mathbf{x}_{\mathrm{min}})} \\
&= \frac{e^{-\mathbf{x}_{\mathrm{maj}}\beta}}{e^{-\mathbf{x}_{\mathrm{min}}\beta}} \\
&= \frac{e^{-\beta_0}\, e^{-x_1\beta_1}\, e^{-x_2\beta_2}\, e^{-x_3\beta_3}\, e^{-x_{\mathrm{maj}}\beta_4}\, e^{-x_5\beta_5}}{e^{-\beta_0}\, e^{-x_1\beta_1}\, e^{-x_2\beta_2}\, e^{-x_3\beta_3}\, e^{-x_{\mathrm{min}}\beta_4}\, e^{-x_5\beta_5}} \\
&= \frac{e^{-x_{\mathrm{maj}}\beta_4}}{e^{-x_{\mathrm{min}}\beta_4}} \\
&= e^{-\beta_4}
\end{aligned}
$$

Hazard Ratios:

$$
\begin{aligned}
\mathrm{HR} &= \frac{h(t|\mathbf{x}_{\mathrm{maj}})}{h(t|\mathbf{x}_{\mathrm{min}})} \\
&= \frac{e^{-\mathbf{x}_{\mathrm{maj}}\beta}}{e^{-\mathbf{x}_{\mathrm{min}}\beta}} \\
&= \frac{e^{-\beta_0}e^{-x_1\beta_1}e^{-x_2\beta_2}e^{-x_3\beta_3}e^{-x_{\mathrm{maj}}\beta_4}e^{-x_5\beta_5}}{e^{-\beta_0}e^{-x_1\beta_1}e^{-x_2\beta_2}e^{-x_3\beta_3}e^{-x_{\mathrm{min}}\beta_4}e^{-x_5\beta_5}} \\
&= \frac{e^{-x_{\mathrm{maj}}\beta_4}}{e^{-x_{\mathrm{min}}\beta_4}} \\
&= e^{-\beta_4}
\end{aligned}
$$

Hazard ratio greater than 1 implies that majority governments fall faster (shorter survival time) than minority governments.

Hazard Ratios:

$$
\begin{aligned}
\mathrm{HR} &= \frac{h(t|\mathbf{x}_{\mathrm{maj}})}{h(t|\mathbf{x}_{\mathrm{min}})} \\
&= \frac{e^{-\mathbf{x}_{\mathrm{maj}}\beta}}{e^{-\mathbf{x}_{\mathrm{min}}\beta}} \\
&= \frac{e^{-\beta_0} e^{-x_1\beta_1} e^{-x_2\beta_2} e^{-x_3\beta_3} e^{-x_{\mathrm{maj}}\beta_4} e^{-x_5\beta_5}}{e^{-\beta_0} e^{-x_1\beta_1} e^{-x_2\beta_2} e^{-x_3\beta_3} e^{-x_{\mathrm{min}}\beta_4} e^{-x_5\beta_5}} \\
&= \frac{e^{-x_{\mathrm{maj}}\beta_4}}{e^{-x_{\mathrm{min}}\beta_4}} \\
&= e^{-\beta_4}
\end{aligned}
$$

Hazard ratio greater than 1 implies that majority governments fall faster (shorter survival time) than minority governments.

Constant hazard ratio across time is the *proportional hazards assumption*.

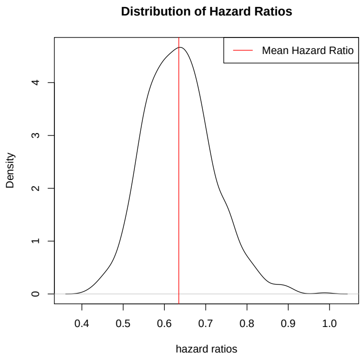
```
> hr.1 <- exp(-betas[, "numst2"])
```

```
> hr.1 <- exp(-betas[, "numst2"])

> hr.2 <- exp(-x.maj %*% t(betas))/exp(-x.min %*% t(betas))
```

```
> hr.1 <- exp(-betas[, "numst2"])
> hr.2 <- exp(-x.maj %*% t(betas))/exp(-x.min %*% t(betas))
> all.equal(hr.1, as.numeric(hr.2))
[1] TRUE
```

```
> hr.1 <- exp(-betas[, "numst2"])
> hr.2 <- exp(-x.maj %*% t(betas))/exp(-x.min %*% t(betas))
> all.equal(hr.1, as.numeric(hr.2))
[1] TRUE
```

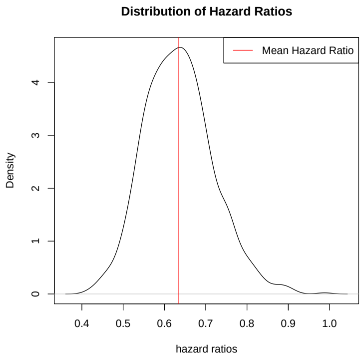


```
> hr.1 <- exp(-betas[, "numst2"])

> hr.2 <- exp(-x.maj %*% t(betas))/exp(-x.min %*% t(betas))

> all.equal(hr.1, as.numeric(hr.2))

[1] TRUE
```



Majority governments survive longer than minority governments.

Expected (average) Survival Time:

Expected (average) Survival Time:

$$E(T|\mathbf{x}_i) = \frac{1}{\lambda_i}$$

Expected (average) Survival Time:

$$
\begin{aligned}
E(T|\mathbf{x}_i) &= \frac{1}{\lambda_i} \\
&= \frac{1}{e^{-\mathbf{x}_i\beta}}
\end{aligned}
$$

Expected (average) Survival Time:

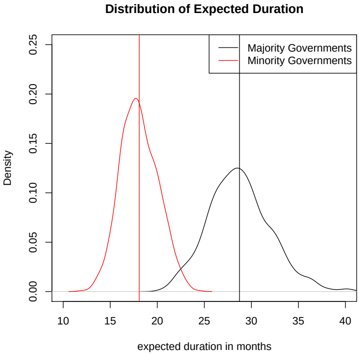
$$
\begin{aligned}
E(T|\mathbf{x}_i) &= \frac{1}{\lambda_i} \\
&= \frac{1}{e^{-\mathbf{x}_i\beta}}
\end{aligned}
$$

```
> expect.maj <- 1/exp(-x.maj %*% t(betas))
> expect.min <- 1/exp(-x.min %*% t(betas))
```

Expected (average) Survival Time:

$$E(T|\mathbf{x}_i) = \frac{1}{\lambda_i} = \frac{1}{e^{-\mathbf{x}_i\beta}}$$

```
> expect.maj <- 1/exp(-x.maj %*% t(betas))
> expect.min <- 1/exp(-x.min %*% t(betas))
```



Predicted Survival Time:

Predicted Survival Time:

Draw predicted values from the exponential distribution.

Predicted Survival Time:

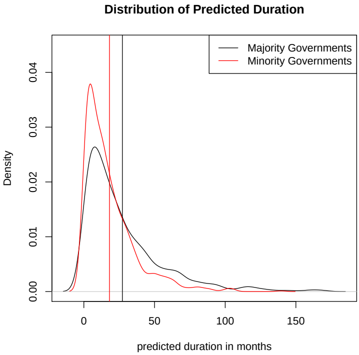
Draw predicted values from the exponential distribution.

```
> predict.maj <- apply(X = 1/expect.maj, MARGIN = 2, FUN = rexp,
+     n = 1)
> predict.min <- apply(X = 1/expect.min, MARGIN = 2, FUN = rexp,
+     n = 1)
```

Predicted Survival Time:

Draw predicted values from the exponential distribution.

```
> predict.maj <- apply(X = 1/expect.maj, MARGIN = 2, FUN = rexp,
+     n = 1)
> predict.min <- apply(X = 1/expect.min, MARGIN = 2, FUN = rexp,
+     n = 1)
```



First Differences:

First Differences:

$$E(T|\mathbf{x}_{\mathrm{maj}}) - E(T|\mathbf{x}_{\mathrm{min}})$$

First Differences:

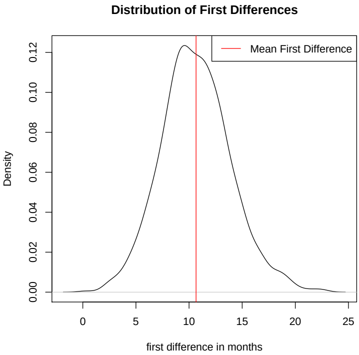
$$E(T|\mathbf{x}_{\mathrm{maj}}) - E(T|\mathbf{x}_{\mathrm{min}})$$

```
> first.diff <- expect.maj - expect.min
```

First Differences:

$$E(T|\mathbf{x}_{\mathrm{maj}}) - E(T|\mathbf{x}_{\mathrm{min}})$$

```
> first.diff <- expect.maj - expect.min
```



The exponential model is nice and simple, but the assumption of a flat hazard may be too restrictive.

The exponential model is nice and simple, but the assumption of a flat hazard may be too restrictive.

What if we want to loosen that restriction by assuming a monotonic hazard?

The exponential model is nice and simple, but the assumption of a flat hazard may be too restrictive.

What if we want to loosen that restriction by assuming a monotonic hazard?

We can use the Weibull model.

The Weibull Model

Assume:

$$T_i \quad \sim \quad \mathrm{Weibull}(\lambda_i, p)$$

The Weibull Model

Assume:

$$
\begin{aligned}
T_i &\sim \mathrm{Weibull}(\lambda_i, p) \\
E(T_i) &= \lambda_i \Gamma\left(1 + \frac{1}{p}\right)
\end{aligned}
$$

The Weibull Model

Assume:

$$
\begin{aligned}
T_i &\sim \mathrm{Weibull}(\lambda_i, p) \\
E(T_i) &= \lambda_i \Gamma\left(1 + \frac{1}{p}\right)
\end{aligned}
$$

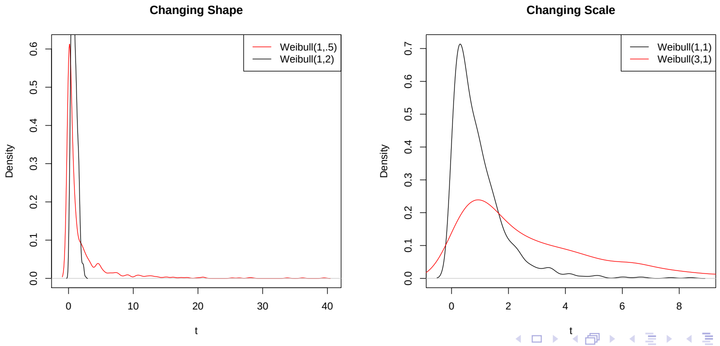
$\lambda_i > 0$ is known as the scale parameter and $p > 0$ is the shape parameter.

The Weibull Model

Assume:

$$
\begin{aligned}
T_i &\sim \mathrm{Weibull}(\lambda_i, p) \\
E(T_i) &= \lambda_i \Gamma\left(1 + \frac{1}{p}\right)
\end{aligned}
$$

$\lambda_i > 0$ is known as the scale parameter and $p > 0$ is the shape parameter.



$$f(t) = \left(\frac{p}{\lambda}\right)\left(\frac{t}{\lambda}\right)^{p-1} e^{-(t/\lambda)^p}$$

$$
\begin{aligned}
f(t) &= \left(\frac{p}{\lambda}\right)\left(\frac{t}{\lambda}\right)^{p-1} e^{-(t/\lambda)^p} \\
S(t) &= 1 - F(t)
\end{aligned}
$$

$$
\begin{aligned}
f(t) &= \left(\frac{p}{\lambda}\right)\left(\frac{t}{\lambda}\right)^{p-1} e^{-(t/\lambda)^p} \\
\\
S(t) &= 1 - F(t) \\
&= 1 - (1 - e^{-(t/\lambda)^p})
\end{aligned}
$$

$$
\begin{aligned}
f(t) &= \left(\frac{p}{\lambda}\right)\left(\frac{t}{\lambda}\right)^{p-1} e^{-(t/\lambda)^p} \\
S(t) &= 1 - F(t) \\
&= 1 - \left(1 - e^{-(t/\lambda)^p}\right) \\
&= e^{-(t/\lambda)^p}
\end{aligned}
$$

$$
\begin{aligned}
f(t) &= \left(\frac{p}{\lambda}\right)\left(\frac{t}{\lambda}\right)^{p-1} e^{-(t/\lambda)^p}
\end{aligned}
$$

$$
\begin{aligned}
S(t) &= 1 - F(t) \\
&= 1 - (1 - e^{-(t/\lambda)^p}) \\
&= e^{-(t/\lambda)^p}
\end{aligned}
$$

$$
h(t) = \frac{f(t)}{S(t)}
$$

$$
\begin{aligned}
f(t) &= \left(\frac{p}{\lambda}\right)\left(\frac{t}{\lambda}\right)^{p-1} e^{-(t/\lambda)^p} \\
\\
S(t) &= 1 - F(t) \\
&= 1 - \left(1 - e^{-(t/\lambda)^p}\right) \\
&= e^{-(t/\lambda)^p} \\
\\
h(t) &= \frac{f(t)}{S(t)} \\
&= \frac{\left(\frac{p}{\lambda}\right)\left(\frac{t}{\lambda}\right)^{p-1} e^{-(t/\lambda)^p}}{e^{-(t/\lambda)^p}}
\end{aligned}
$$

$$
\begin{aligned}
f(t) &= \left(\frac{p}{\lambda}\right)\left(\frac{t}{\lambda}\right)^{p-1} e^{-(t/\lambda)^p} \\
\\
S(t) &= 1 - F(t) \\
&= 1 - (1 - e^{-(t/\lambda)^p}) \\
&= e^{-(t/\lambda)^p} \\
\\
h(t) &= \frac{f(t)}{S(t)} \\
&= \frac{\left(\frac{p}{\lambda}\right)\left(\frac{t}{\lambda}\right)^{p-1} e^{-(t/\lambda)^p}}{e^{-(t/\lambda)^p}} \\
&= \left(\frac{p}{\lambda}\right)\left(\frac{t}{\lambda}\right)^{p-1}
\end{aligned}
$$

$$
\begin{aligned}
f(t) &= \left(\frac{p}{\lambda}\right)\left(\frac{t}{\lambda}\right)^{p-1} e^{-(t/\lambda)^p}
\end{aligned}
$$

$$
\begin{aligned}
S(t) &= 1 - F(t) \\
&= 1 - (1 - e^{-(t/\lambda)^p}) \\
&= e^{-(t/\lambda)^p}
\end{aligned}
$$

$$
\begin{aligned}
h(t) &= \frac{f(t)}{S(t)} \\
&= \frac{\left(\frac{p}{\lambda}\right)\left(\frac{t}{\lambda}\right)^{p-1} e^{-(t/\lambda)^p}}{e^{-(t/\lambda)^p}} \\
&= \left(\frac{p}{\lambda}\right)\left(\frac{t}{\lambda}\right)^{p-1} \\
&= \left(\frac{p}{\lambda^p}\right) t^{p-1}
\end{aligned}
$$

Modeling $h(t)$ with covariates:

Modeling $h(t)$ with covariates:

$$h(t_i) \quad = \quad \left(\frac{p}{\lambda_i^p}\right) t_i^{p-1}$$

Modeling $h(t)$ with covariates:

$$h(t_i) \quad = \quad \left(\frac{p}{\lambda_i^p}\right) t_i^{p-1}$$

We want to constrain the $h(t)$ to be positive.

Modeling $h(t)$ with covariates:

$$h(t_i) = \left(\frac{p}{\lambda_i^p}\right) t_i^{p-1}$$

We want to constrain the $h(t)$ to be positive.

$$\lambda_i = e^{\mathbf{x}_i \beta}$$

Modeling $h(t)$ with covariates:

$$h(t_i) \quad = \quad \left(\frac{p}{\lambda_i^p}\right) t_i^{p-1}$$

We want to constrain the $h(t)$ to be positive.

$$\lambda_i \quad = \quad e^{\mathbf{x}_i \beta}$$

Note that the link is different from the exponential model.

Modeling $h(t)$ with covariates:

$$h(t_i) \quad = \quad \left(\frac{p}{\lambda_i^p}\right) t_i^{p-1}$$

We want to constrain the $h(t)$ to be positive.

$$\lambda_i \quad = \quad e^{\mathbf{x}_i \beta}$$

Note that the link is different from the exponential model. Positive β implies that hazard decreases and average survival time increases as x increases.

$h(t_i)$ is modeled with both λ_i and p and is a function of t_i. Thus, the Weibull model assumes a **monotonic hazard**.

$h(t_i)$ is modeled with both λ_i and p and is a function of t_i. Thus, the Weibull model assumes a **monotonic hazard**.

- If $p = 1$, $h(t)$ is flat and the model is the exponential model.

$h(t_i)$ is modeled with both λ_i and p and is a function of t_i. Thus, the Weibull model assumes a **monotonic hazard**.

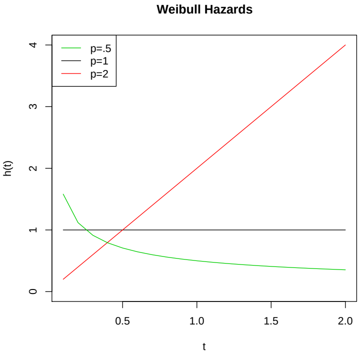
- If $p = 1$, $h(t)$ is flat and the model is the exponential model.
- If $p > 1$, $h(t)$ is monotonically increasing.

$h(t_i)$ is modeled with both λ_i and p and is a function of t_i. Thus, the Weibull model assumes a **monotonic hazard**.

- If $p = 1$, $h(t)$ is flat and the model is the exponential model.
- If $p > 1$, $h(t)$ is monotonically increasing.
- If $p < 1$, $h(t)$ is monotonically decreasing.

$h(t_i)$ is modeled with both λ_i and p and is a function of t_i. Thus, the Weibull model assumes a **monotonic hazard**.

- If $p = 1$, $h(t)$ is flat and the model is the exponential model.
- If $p > 1$, $h(t)$ is monotonically increasing.
- If $p < 1$, $h(t)$ is monotonically decreasing.



$$\mathcal{L} \;=\; \prod_{i=1}^{n} \left[f(t_i)\right]^{c_i} \; \left[S(t_i)\right]^{1-c_i}$$

$$
\begin{aligned}
\mathcal{L} &= \prod_{i=1}^{n} [f(t_i)]^{c_i} \; [S(t_i)]^{1-c_i} \\
&= \prod_{i=1}^{n} \left[\left(\frac{p}{\lambda}\right) \left(\frac{t}{\lambda}\right)^{p-1} e^{-(t/\lambda)^p} \right]^{c_i} \; \left[e^{-(t/\lambda)^p} \right]^{1-c_i}
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{L} &= \prod_{i=1}^{n} [f(t_i)]^{c_i} \; [S(t_i)]^{1-c_i} \\
&= \prod_{i=1}^{n} \left[\left(\frac{p}{\lambda}\right) \left(\frac{t}{\lambda}\right)^{p-1} e^{-(t/\lambda)^p} \right]^{c_i} \; \left[e^{-(t/\lambda)^p} \right]^{1-c_i} \\
&= \prod_{i=1}^{n} \left[\left(\frac{p}{\lambda^p}\right) t^{p-1} e^{-(t/\lambda)^p} \right]^{c_i} \; \left[e^{-(t/\lambda)^p} \right]^{1-c_i}
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{L} &= \prod_{i=1}^{n} [f(t_i)]^{c_i} \; [S(t_i)]^{1-c_i} \\
&= \prod_{i=1}^{n} \left[\left(\frac{p}{\lambda}\right) \left(\frac{t}{\lambda}\right)^{p-1} e^{-(t/\lambda)^p} \right]^{c_i} \; \left[e^{-(t/\lambda)^p} \right]^{1-c_i} \\
&= \prod_{i=1}^{n} \left[\left(\frac{p}{\lambda^p}\right) t^{p-1} e^{-(t/\lambda)^p} \right]^{c_i} \; \left[e^{-(t/\lambda)^p} \right]^{1-c_i} \\
\ln \mathcal{L} &= \sum_{i=1}^{n} c_i \left[\ln p - p \ln \lambda + (p-1) \ln t_i\right] - \left(\frac{t_i}{\lambda}\right)^p
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{L} &= \prod_{i=1}^{n} \, [f(t_i)]^{c_i} \; [S(t_i)]^{1-c_i} \\
&= \prod_{i=1}^{n} \left[\left(\frac{p}{\lambda}\right) \left(\frac{t}{\lambda}\right)^{p-1} e^{-(t/\lambda)^p} \right]^{c_i} \; \left[e^{-(t/\lambda)^p} \right]^{1-c_i} \\
&= \prod_{i=1}^{n} \left[\left(\frac{p}{\lambda^p}\right) t^{p-1} e^{-(t/\lambda)^p} \right]^{c_i} \; \left[e^{-(t/\lambda)^p} \right]^{1-c_i} \\
\ln \mathcal{L} &= \sum_{i=1}^{n} c_i \left[\ln p - p \ln \lambda + (p-1) \ln t_i \right] - \left(\frac{t_i}{\lambda}\right)^p \\
&= \sum_{i=1}^{n} c_i \left[\ln p - p \, \mathbf{x}_i \beta + (p-1) \ln t_i \right] - \left(\frac{t_i}{e^{\mathbf{x}_i \beta}}\right)^p
\end{aligned}
$$

Maximizing your own likelihood:

Maximizing your own likelihood:

```
> weib.lik <- function(par, T, X, C) {
+     beta <- par[1:ncol(X)]
+     p <- exp(par[(ncol(X) + 1)])
+     lambda <- exp((X %*% beta))
+     log.lik <- sum(C * (log(p) - p * log(lambda) + (p - 1) *
+         log(T)) - (T/lambda)^p)
+     return(log.lik)
+ }
```

Maximizing your own likelihood:

```
> weib.lik <- function(par, T, X, C) {
+     beta <- par[1:ncol(X)]
+     p <- exp(par[(ncol(X) + 1)])
+     lambda <- exp((X %*% beta))
+     log.lik <- sum(C * (log(p) - p * log(lambda) + (p - 1) *
+         log(T)) - (T/lambda)^p)
+     return(log.lik)
+ }

> my.max <- optim(par = c(0, 0, 0, 0, 0, 0, 0), fn = weib.lik,
+     T = T, X = X, C = C, method = "BFGS", control = list(fnscale = -1))$par
> my.coef <- my.max[1:ncol(X)]
> my.p <- exp(my.max[(ncol(X) + 1)])
```

Maximizing your own likelihood:

```
> weib.lik <- function(par, T, X, C) {
+     beta <- par[1:ncol(X)]
+     p <- exp(par[(ncol(X) + 1)])
+     lambda <- exp((X %*% beta))
+     log.lik <- sum(C * (log(p) - p * log(lambda) + (p - 1) *
+         log(T)) - (T/lambda)^p)
+     return(log.lik)
+ }

> my.max <- optim(par = c(0, 0, 0, 0, 0, 0, 0), fn = weib.lik,
+     T = T, X = X, C = C, method = "BFGS", control = list(fnscale = -1))$par
> my.coef <- my.max[1:ncol(X)]
> my.p <- exp(my.max[(ncol(X) + 1)])
```

Using the survival package:

```
> weib.surv <- survreg(Surv(duration, ciep12) ~ invest + fract +
+     polar + numst2 + crisis, data = coalition, dist = "weibull")
```

```
> summary(weib.surv)

Call:
survreg(formula = Surv(duration, ciep12) ~ invest + fract + polar +
    numst2 + crisis, data = coalition, dist = "weibull")
               Value Std. Error     z        p
(Intercept)  4.75007    0.53072  8.95 3.55e-19
invest      -0.47160    0.11643 -4.05 5.11e-05
fract       -2.11762    0.75876 -2.79 5.26e-03
polar       -0.02792    0.00506 -5.52 3.33e-08
numst2       0.42746    0.11025  3.88 1.06e-04
crisis       0.00538    0.00183  2.94 3.28e-03
Log(scale)  -0.15644    0.04971 -3.15 1.65e-03

Scale= 0.855

Weibull distribution
Loglik(model)= -1042   Loglik(intercept only)= -1101
Chisq= 117.8 on 5 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 5
n= 314
```

The shape parameter p for the Weibull distribution is the inverse of the scale parameter given by `survreg()`.

The shape parameter p for the Weibull distribution is the inverse of the scale parameter given by `survreg()`.

```
> surv.p <- 1/weib.surv$scale
```

The shape parameter p for the Weibull distribution is the inverse of the scale parameter given by `survreg()`.

```
> surv.p <- 1/weib.surv$scale

> surv.p

[1] 1.169

> my.p

[1] 1.169
```

The shape parameter p for the Weibull distribution is the inverse of the scale parameter given by `survreg()`.

```
> surv.p <- 1/weib.surv$scale

> surv.p

[1] 1.169

> my.p

[1] 1.169
```

The scale parameter given by `survreg()` is NOT the same as the scale parameter in the Weibull distribution, which should be $\lambda_i = e^{\mathbf{x}_i \beta}$.

The shape parameter p for the Weibull distribution is the inverse of the scale parameter given by `survreg()`.

```
> surv.p <- 1/weib.surv$scale

> surv.p

[1] 1.169

> my.p

[1] 1.169
```

The scale parameter given by `survreg()` is NOT the same as the scale parameter in the Weibull distribution, which should be $\lambda_i = e^{\mathbf{x}_i \beta}$.

```
> rbind(weib.surv.coef = weib.surv$coef, my.coef)

               (Intercept)  invest  fract    polar numst2   crisis
weib.surv.coef       4.750 -0.4716 -2.118 -0.02792 0.4275 0.005377
my.coef              4.753 -0.4719 -2.122 -0.02792 0.4271 0.005405
```

Other Parametric Models

Other Parametric Models

- Gompertz or gamma model: monotonic hazard

Other Parametric Models

- Gompertz or gamma model: monotonic hazard
- Log-logistic or log-normal model: nonmonotonic hazard

Other Parametric Models

- Gompertz or gamma model: monotonic hazard
- Log-logistic or log-normal model: nonmonotonic hazard
- Generalized gamma model: nests the exponential, Weibull, log-normal, and gamma models with an extra parameter

Other Parametric Models

- Gompertz or gamma model: monotonic hazard
- Log-logistic or log-normal model: nonmonotonic hazard
- Generalized gamma model: nests the exponential, Weibull, log-normal, and gamma models with an extra parameter

But what if we don't want to make an assumption about the shape of the hazard?

Outline

The Cox Proportional Hazards Model

The Cox Proportional Hazards Model

- Often described as a semi-parametric model.

The Cox Proportional Hazards Model

- Often described as a semi-parametric model.
- Makes no assumptions about the shape of the hazard or the distribution of T_i.

The Cox Proportional Hazards Model

- Often described as a semi-parametric model.
- Makes no assumptions about the shape of the hazard or the distribution of T_i.
- Takes advantage of the proportional hazards assumption.

1. Reconceptualize each t_i as a discrete event time rather than a duration or survival time (non-censored observations only).

1. Reconceptualize each t_i as a discrete event time rather than a duration or survival time (non-censored observations only).
    - $t_i = 5$: An event occurred at month 5, rather than observation i surviving for 5 months.

1. Reconceptualize each t_i as a discrete event time rather than a duration or survival time (non-censored observations only).
   - $t_i = 5$: An event occurred at month 5, rather than observation i surviving for 5 months.
2. Assume there are no tied event times in the data.

1. Reconceptualize each t_i as a discrete event time rather than a duration or survival time (non-censored observations only).
   - $t_i = 5$: An event occurred at month 5, rather than observation i surviving for 5 months.
2. Assume there are no tied event times in the data.

1. Reconceptualize each t_i as a discrete event time rather than a duration or survival time (non-censored observations only).
   - $t_i = 5$: An event occurred at month 5, rather than observation i surviving for 5 months.
2. Assume there are no tied event times in the data.
   - No two events can occur at the same instant. It only seems that way because our unit of measurement is not precise enough.

1. Reconceptualize each t_i as a discrete event time rather than a duration or survival time (non-censored observations only).
   - $t_i = 5$: An event occurred at month 5, rather than observation i surviving for 5 months.
2. Assume there are no tied event times in the data.
   - No two events can occur at the same instant. It only seems that way because our unit of measurement is not precise enough.
   - There are ways to adjust the likelihood to take into account observed ties.

1. Reconceptualize each t_i as a discrete event time rather than a duration or survival time (non-censored observations only).
    - $t_i = 5$: An event occurred at month 5, rather than observation i surviving for 5 months.
2. Assume there are no tied event times in the data.
    - No two events can occur at the same instant. It only seems that way because our unit of measurement is not precise enough.
    - There are ways to adjust the likelihood to take into account observed ties.
3. Assume no events can happen between event times.

We know that exactly one event occurred at each t_i for all non-censored i.

We know that exactly one event occurred at each t_i for all non-censored i.

Define a risk set R_i as the set of all possible observations at risk of an event at time t_i.

We know that exactly one event occurred at each t_i for all non-censored i.

Define a risk set R_i as the set of all possible observations at risk of an event at time t_i.

What observations belong in R_i?

We know that exactly one event occurred at each t_i for all non-censored i.

Define a risk set R_i as the set of all possible observations at risk of an event at time t_i.

What observations belong in R_i?

All observations (censored and non-censored) j such that $t_j \geq t_i$

We know that exactly one event occurred at each t_i for all non-censored i.

Define a risk set R_i as the set of all possible observations at risk of an event at time t_i.

What observations belong in R_i?

All observations (censored and non-censored) j such that $t_j \geq t_i$

For example, if $t_i = 5$ months, then all observations that do not experience the event or are not censored before 5 months are at risk.

We can then create a *partial likelihood* function:

We can then create a *partial likelihood* function:

$$\mathcal{L} = \prod_{i=1}^{n} [P(\text{event occurred in } i | \text{event occurred in } R_i)]^{c_i}$$

We can then create a *partial likelihood* function:

$$\begin{aligned}
\mathcal{L} &= \prod_{i=1}^{n} \left[P(\text{event occurred in } i | \text{event occurred in } R_i)\right]^{c_i} \\
&= \prod_{i=1}^{n} \left[\frac{P(\text{event occurred in } i)}{P(\text{event occurred in } R_i)}\right]^{c_i}
\end{aligned}$$

We can then create a *partial likelihood* function:

$$
\begin{aligned}
\mathcal{L} &= \prod_{i=1}^{n} [P(\text{event occurred in } i | \text{event occurred in } R_i)]^{c_i} \\
&= \prod_{i=1}^{n} \left[\frac{P(\text{event occurred in } i)}{P(\text{event occurred in } R_i)} \right]^{c_i} \\
&= \prod_{i=1}^{n} \left[\frac{h(t_i)}{\sum_{j \in R_i} h(t_j)} \right]^{c_i}
\end{aligned}
$$

We can then create a *partial likelihood* function:

$$
\begin{aligned}
\mathcal{L} &= \prod_{i=1}^{n} [P(\text{event occurred in } i|\text{event occurred in } R_i)]^{c_i} \\
&= \prod_{i=1}^{n} \left[\frac{P(\text{event occurred in } i)}{P(\text{event occurred in } R_i)}\right]^{c_i} \\
&= \prod_{i=1}^{n} \left[\frac{h(t_i)}{\sum_{j \in R_i} h(t_j)}\right]^{c_i} \\
&= \prod_{i=1}^{n} \left[\frac{h_0(t)h_i(t_i)}{\sum_{j \in R_i} h_0(t)h_j(t_j)}\right]^{c_i}
\end{aligned}
$$

We can then create a *partial likelihood* function:

$$
\begin{aligned}
\mathcal{L} &= \prod_{i=1}^{n} [P(\text{event occurred in } i | \text{event occurred in } R_i)]^{c_i} \\
&= \prod_{i=1}^{n} \left[\frac{P(\text{event occurred in } i)}{P(\text{event occurred in } R_i)} \right]^{c_i} \\
&= \prod_{i=1}^{n} \left[\frac{h(t_i)}{\sum_{j \in R_i} h(t_j)} \right]^{c_i} \\
&= \prod_{i=1}^{n} \left[\frac{h_0(t) h_i(t_i)}{\sum_{j \in R_i} h_0(t) h_j(t_j)} \right]^{c_i} \\
&= \prod_{i=1}^{n} \left[\frac{h_i(t_i)}{\sum_{j \in R_i} h_j(t_j)} \right]^{c_i}
\end{aligned}
$$

We can then create a *partial likelihood* function:

$$
\begin{aligned}
\mathcal{L} &= \prod_{i=1}^{n} [P(\text{event occurred in } i | \text{event occurred in } R_i)]^{c_i} \\
&= \prod_{i=1}^{n} \left[\frac{P(\text{event occurred in } i)}{P(\text{event occurred in } R_i)} \right]^{c_i} \\
&= \prod_{i=1}^{n} \left[\frac{h(t_i)}{\sum_{j \in R_i} h(t_j)} \right]^{c_i} \\
&= \prod_{i=1}^{n} \left[\frac{h_0(t) h_i(t_i)}{\sum_{j \in R_i} h_0(t) h_j(t_j)} \right]^{c_i} \\
&= \prod_{i=1}^{n} \left[\frac{h_i(t_i)}{\sum_{j \in R_i} h_j(t_j)} \right]^{c_i}
\end{aligned}
$$

$h_0(t)$ is the baseline hazard, which is the same for all observations, so it cancels out.

Like in parametric models, $h(t)$ is modeled with covariates:

Like in parametric models, $h(t)$ is modeled with covariates:

$$h_i(t_i) = e^{\mathbf{x}_i \beta}$$

Like in parametric models, $h(t)$ is modeled with covariates:

$$h_i(t_i) = e^{\mathbf{x}_i \beta}$$

Note that a positive β now suggests that an increase in x increases the hazard and decreases survival time.

Like in parametric models, $h(t)$ is modeled with covariates:

$$h_i(t_i) = e^{\mathbf{x}_i \beta}$$

Note that a positive β now suggests that an increase in x increases the hazard and decreases survival time.

$$\mathcal{L} = \prod_{i=1}^{n} \left[\frac{e^{\mathbf{x}_i \beta}}{\sum_{j \in R_i} e^{\mathbf{x}_j \beta}} \right]^{c_i}$$

Like in parametric models, $h(t)$ is modeled with covariates:

$$h_i(t_i) = e^{\mathbf{x}_i \beta}$$

Note that a positive β now suggests that an increase in x increases the hazard and decreases survival time.

$$\mathcal{L} = \prod_{i=1}^{n} \left[\frac{e^{\mathbf{x}_i \beta}}{\sum_{j \in R_i} e^{\mathbf{x}_j \beta}} \right]^{c_i}$$

There is no β_0 term estimated.

Like in parametric models, $h(t)$ is modeled with covariates:

$$h_i(t_i) = e^{\mathbf{x}_i \beta}$$

Note that a positive β now suggests that an increase in x increases the hazard and decreases survival time.

$$\mathcal{L} = \prod_{i=1}^{n} \left[\frac{e^{\mathbf{x}_i \beta}}{\sum_{j \in R_i} e^{\mathbf{x}_j \beta}} \right]^{c_i}$$

There is no β_0 term estimated. This implies that the shape of the baseline hazard is left unmodeled.

Pros:

Pros:

- Makes no restrictive assumption about the shape of the hazard.

Pros:

- Makes no restrictive assumption about the shape of the hazard.
- A better choice if you want the effects of the covariates and the nature of the time dependence is unimportant.

Pros:

- Makes no restrictive assumption about the shape of the hazard.
- A better choice if you want the effects of the covariates and the nature of the time dependence is unimportant.

Cons:

Pros:

- Makes no restrictive assumption about the shape of the hazard.
- A better choice if you want the effects of the covariates and the nature of the time dependence is unimportant.

Cons:

- Only quantities of interest are hazard ratios.

Pros:

- Makes no restrictive assumption about the shape of the hazard.
- A better choice if you want the effects of the covariates and the nature of the time dependence is unimportant.

Cons:

- Only quantities of interest are hazard ratios.
- Can be subject to overfitting

Pros:

- Makes no restrictive assumption about the shape of the hazard.
- A better choice if you want the effects of the covariates and the nature of the time dependence is unimportant.

Cons:

- Only quantities of interest are hazard ratios.
- Can be subject to overfitting
- Shape of hazard is unknown (although there are semi-parametric ways to derive the hazard and survivor functions)

How do I run a Cox proportional hazards model in R?

How do I run a Cox proportional hazards model in R?

Use the `coxph()` function in the `survival` package (also in the `Design` and `Zelig` packages).

Outline

How Do Survival Models Relate to Duration Dependence in a Logit Model?

How Do Survival Models Relate to Duration Dependence in a Logit Model?

- Based on Beck, Katz, and Tucker (1998)

How Do Survival Models Relate to Duration Dependence in a Logit Model?

- Based on Beck, Katz, and Tucker (1998)
- Suppose we have Time-Series Cross-Sectional Data with a binary dependent variable.

How Do Survival Models Relate to Duration Dependence in a Logit Model?

- Based on Beck, Katz, and Tucker (1998)
- Suppose we have Time-Series Cross-Sectional Data with a binary dependent variable.
  - For example, if we had data on country dyads over 50 years, with the dependent variable being whether there was a war between the two countries in each year.

How Do Survival Models Relate to Duration Dependence in a Logit Model?

- Based on Beck, Katz, and Tucker (1998)
- Suppose we have Time-Series Cross-Sectional Data with a binary dependent variable.
  - For example, if we had data on country dyads over 50 years, with the dependent variable being whether there was a war between the two countries in each year.
- Not all observations are independent. We may see some duration dependence.

How Do Survival Models Relate to Duration Dependence in a Logit Model?

- Based on Beck, Katz, and Tucker (1998)
- Suppose we have Time-Series Cross-Sectional Data with a binary dependent variable.
  - For example, if we had data on country dyads over 50 years, with the dependent variable being whether there was a war between the two countries in each year.
- Not all observations are independent. We may see some duration dependence.
  - Perhaps countries that have been at peace for 100 years may be less likely to go to war than countries that have been at peace for only 2 years.

How Do Survival Models Relate to Duration Dependence in a Logit Model?

- Based on Beck, Katz, and Tucker (1998)
- Suppose we have Time-Series Cross-Sectional Data with a binary dependent variable.
  - For example, if we had data on country dyads over 50 years, with the dependent variable being whether there was a war between the two countries in each year.
- Not all observations are independent. We may see some duration dependence.
  - Perhaps countries that have been at peace for 100 years may be less likely to go to war than countries that have been at peace for only 2 years.

How can we account for this duration dependence in a logit model?

Think of the observations as grouped duration data:

Think of the observations as grouped duration data:

Year	t_k	Dyad	Y_i	T_i
1992	1	US-Iraq	0	
1993	2	US-Iraq	0	
1994	3	US-Iraq	0	
1995	4	US-Iraq	0	
1996	5	US-Iraq	0	
1997	6	US-Iraq	0	
1998	7	US-Iraq	0	12
1999	8	US-Iraq	0	
2000	9	US-Iraq	0	
2001	10	US-Iraq	0	
2002	11	US-Iraq	0	
2003	12	US-Iraq	1	

Then we end up with:

$$P(y_{i,t_k} = 1|\mathbf{x}_{i,t_k}) \quad = \quad h(t_k|\mathbf{x}_{i,t_k})$$

Then we end up with:

$$\begin{aligned} P(y_{i,t_k} = 1 | \mathbf{x}_{i,t_k}) &= h(t_k | \mathbf{x}_{i,t_k}) \\ &= 1 - P(\text{surviving beyond } t_k | \text{survival up to } t_{k-1}) \end{aligned}$$

Then we end up with:

$$\begin{aligned} P(y_{i,t_k} = 1|\mathbf{x}_{i,t_k}) &= h(t_k|\mathbf{x}_{i,t_k}) \\ &= 1 - P(\text{surviving beyond } t_k|\text{survival up to } t_{k-1}) \end{aligned}$$

It can be shown in general that

$$S(t) = e^{-\int_0^t h(u)du}$$

Then we end up with:

$$\begin{aligned} P(y_{i,t_k} = 1|\mathbf{x}_{i,t_k}) &= h(t_k|\mathbf{x}_{i,t_k}) \\ &= 1 - P(\text{surviving beyond } t_k|\text{survival up to } t_{k-1}) \end{aligned}$$

It can be shown in general that

$$S(t) = e^{-\int_0^t h(u)du}$$

So then we get

$$P(y_{i,t_k} = 1|\mathbf{x}_{i,t_k}) = 1 - e^{-\int_{t_{k-1}}^{t_k} h(u)du}$$

where we take the integral from t_{k-1} to t_k in order to get the conditional survival.

$$P(y_{i,t_k} = 1 | \mathbf{x}_{i,t_k}) \quad = \quad 1 - \exp\left(- \int_{t_{k-1}}^{t_k} h(u) du \right)$$

$$
\begin{aligned}
P(y_{i,t_k} = 1 | \mathbf{x}_{i,t_k}) &= 1 - \exp\left(-\int_{t_{k-1}}^{t_k} h(u)du\right) \\
&= 1 - \exp\left(-\int_{t_{k-1}}^{t_k} e^{\mathbf{x}_{i,t_k}\beta} h_0(u)du\right)
\end{aligned}
$$

$$
\begin{aligned}
P(y_{i,t_k} = 1 | \mathbf{x}_{i,t_k}) &= 1 - \exp\left(- \int_{t_{k-1}}^{t_k} h(u) du \right) \\
&= 1 - \exp\left(- \int_{t_{k-1}}^{t_k} e^{\mathbf{x}_{i,t_k}\beta} h_0(u) du \right) \\
&= 1 - \exp\left(- e^{\mathbf{x}_{i,t_k}\beta} \int_{t_{k-1}}^{t_k} h_0(u) du \right)
\end{aligned}
$$

$$
\begin{aligned}
P(y_{i,t_k} = 1 | \mathbf{x}_{i,t_k}) &= 1 - \exp\left(- \int_{t_{k-1}}^{t_k} h(u) du \right) \\
&= 1 - \exp\left(- \int_{t_{k-1}}^{t_k} e^{\mathbf{x}_{i,t_k}\beta} h_0(u) du \right) \\
&= 1 - \exp\left(- e^{\mathbf{x}_{i,t_k}\beta} \int_{t_{k-1}}^{t_k} h_0(u) du \right) \\
&= 1 - \exp\left(- e^{\mathbf{x}_{i,t_k}\beta} \alpha_{t_k} \right)
\end{aligned}
$$

$$
\begin{aligned}
P(y_{i,t_k} = 1 | \mathbf{x}_{i,t_k}) &= 1 - \exp\left(- \int_{t_{k-1}}^{t_k} h(u) du \right) \\
&= 1 - \exp\left(- \int_{t_{k-1}}^{t_k} e^{\mathbf{x}_{i,t_k} \beta} h_0(u) du \right) \\
&= 1 - \exp\left(- e^{\mathbf{x}_{i,t_k} \beta} \int_{t_{k-1}}^{t_k} h_0(u) du \right) \\
&= 1 - \exp\left(- e^{\mathbf{x}_{i,t_k} \beta} \alpha_{t_k} \right) \\
&= 1 - \exp\left(- e^{\mathbf{x}_{i,t_k} \beta + \kappa_{t_k}} \right)
\end{aligned}
$$

$$
\begin{aligned}
P(y_{i,t_k} = 1|\mathbf{x}_{i,t_k}) &= 1 - \exp\left(-\int_{t_{k-1}}^{t_k} h(u)du\right) \\
&= 1 - \exp\left(-\int_{t_{k-1}}^{t_k} e^{\mathbf{x}_{i,t_k}\beta} h_0(u)du\right) \\
&= 1 - \exp\left(-e^{\mathbf{x}_{i,t_k}\beta} \int_{t_{k-1}}^{t_k} h_0(u)du\right) \\
&= 1 - \exp\left(-e^{\mathbf{x}_{i,t_k}\beta} \alpha_{t_k}\right) \\
&= 1 - \exp\left(-e^{\mathbf{x}_{i,t_k}\beta + \kappa_{t_k}}\right)
\end{aligned}
$$

This is equivalent to a model with a complementary log-log (cloglog) link and time dummies κ_{t_k}.

- BKT suggest using a logit link instead of a cloglog link because logit is more widely used (and nobody knows what a cloglog link is).

- BKT suggest using a logit link instead of a cloglog link because logit is more widely used (and nobody knows what a cloglog link is).
- As long as probability of an event does not exceed 50 percent, logit and cloglog links are very similar.

- BKT suggest using a logit link instead of a cloglog link because logit is more widely used (and nobody knows what a cloglog link is).
- As long as probability of an event does not exceed 50 percent, logit and cloglog links are very similar.
- The use of time dummies means that we are imposing no structure on the nature of duration dependence (structure of the hazard).

- BKT suggest using a logit link instead of a cloglog link because logit is more widely used (and nobody knows what a cloglog link is).
- As long as probability of an event does not exceed 50 percent, logit and cloglog links are very similar.
- The use of time dummies means that we are imposing no structure on the nature of duration dependence (structure of the hazard).
- If we don't use time dummies, we are assuming no duration dependence (flat hazard)

- BKT suggest using a logit link instead of a cloglog link because logit is more widely used (and nobody knows what a cloglog link is).
- As long as probability of an event does not exceed 50 percent, logit and cloglog links are very similar.
- The use of time dummies means that we are imposing no structure on the nature of duration dependence (structure of the hazard).
- If we don't use time dummies, we are assuming no duration dependence (flat hazard)
- Using a variable such as "number of years at peace" instead of time dummies imposes a monotonic hazard.

- BKT suggest using a logit link instead of a cloglog link because logit is more widely used (and nobody knows what a cloglog link is).
- As long as probability of an event does not exceed 50 percent, logit and cloglog links are very similar.
- The use of time dummies means that we are imposing no structure on the nature of duration dependence (structure of the hazard).
- If we don't use time dummies, we are assuming no duration dependence (flat hazard)
- Using a variable such as "number of years at peace" instead of time dummies imposes a monotonic hazard.
- The use of time dummies may use up a lot of degrees of freedom, so BKT suggest using restricted cubic splines.

Possible complications:

Possible complications:

- Multiple events

Possible complications:

- Multiple events
    - Assumes that multiple events are independent (independence of observations assumption in a survival model).

Possible complications:

- Multiple events
  - Assumes that multiple events are independent (independence of observations assumption in a survival model).
- Left censoring

Possible complications:

- Multiple events
  - Assumes that multiple events are independent (independence of observations assumption in a survival model).
- Left censoring
  - Countries may have been at peace long before we start observing data, and we don't know when that "peace duration" began.

Possible complications:

- Multiple events
    - Assumes that multiple events are independent (independence of observations assumption in a survival model).
- Left censoring
    - Countries may have been at peace long before we start observing data, and we don't know when that "peace duration" began.
- Variables that do not vary across units

Possible complications:

- Multiple events
  - Assumes that multiple events are independent (independence of observations assumption in a survival model).
- Left censoring
  - Countries may have been at peace long before we start observing data, and we don't know when that "peace duration" began.
- Variables that do not vary across units
  - May be collinear with time dummies.

Outline

- Survival models are cool . . .

- Survival models are cool . . . but hard.

- Survival models are cool . . . but hard.
- There are other things you can model:

- Survival models are cool ... but hard.
- There are other things you can model:
  - Perhaps some observations are more likely to fail than others: frailty models

- Survival models are cool ... but hard.
- There are other things you can model:
  - Perhaps some observations are more likely to fail than others: frailty models
  - Perhaps some observations you don't expect to fail at all: split population models

- Survival models are cool . . . but hard.
- There are other things you can model:
  - Perhaps some observations are more likely to fail than others: frailty models
  - Perhaps some observations you don't expect to fail at all: split population models
  - Perhaps there can be more than one type of event: competing risks model

- Survival models are cool ... but hard.
- There are other things you can model:
  - Perhaps some observations are more likely to fail than others: frailty models
  - Perhaps some observations you don't expect to fail at all: split population models
  - Perhaps there can be more than one type of event: competing risks model

Go forth and learn.

References:

Box-Steffensmeier, Janet M. and Bradford S. Jones. 2004. *Event History Modeling.* Cambridge University Press.

King, Gary, James E. Alt, Nancy E. Burns, and Michael Laver. 1990. "A Unified Model of Cabinet Dissolution in Parliamentary Democracies." *American Journal of Political Science* 34(3): 846-971