GOV 2001/ 1002/ E-2001 Section 10[1]
Duration II and Matching

Mayya Komisarchik

Harvard University

April 13, 2016

[1] Heartfelt thanks to all of the Gov 2001 TFs of yesteryear; this section draws heavily upon materials created by Stephen Pettigrew, Solé Prillaman, and Patrick Lam.

Outline

Gov 2001 End-of-Semester Party

- **Saturday, May 7, 2016 at 12:00pm**
- Gary's House (see Canvas for address and directions)
- Please RSVP for yourself and guest(s) using the poll we've posted on Canvas.
- Bring your match!

Logistics

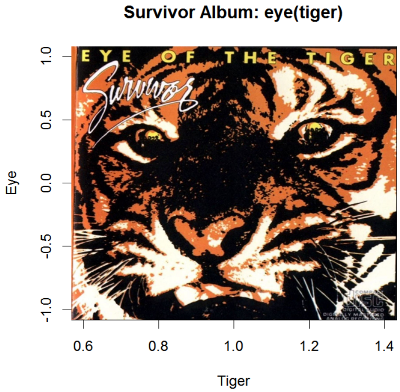
Reading for Next Week:

- "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs"
- "Misunderstandings Between Experimentalists and Observationalists about Causal Inference"
- "A Theory of Statistical Inference for Matching Methods in Applied Causal Research" (recommended)

Replication Paper:

- Final papers are due **Wednesday, April 27** (automatic extension to **Thursday, May 5** if you need it but that's actually the last day we will accept papers)
- Submit **two versions** of your final paper:
  - Complete final version with your names on it
  - Final version with your names and identifying information about you removed for anonymous student ranking

FROM LAST WEEK



Review of Duration Model Structure

Duration density functions consist of the following:

$$\underbrace{f(t)}_{\text{density function}} = \underbrace{h(t)}_{\text{hazard function}} \cdot \underbrace{S(t)}_{\text{survival function}}$$

$$\underbrace{f(t)}_{\text{density function}} = \underbrace{\frac{f(t)}{S(t)}}_{\text{hazard function}} \cdot \underbrace{S(t)}_{\text{survival function}}$$

Review of Duration Model Structure

Where:

- T is a continuous, positive random variable representing survival time
- $f(t)$ is the PDF (stochastic component) of T
  - Approximately the probability that an event occurs at time $T = t$
- $F(t)$ is the CDF of T, such that $\int_0^t f(u)du = P(T \leq t)$
  - The probability of an event occurring at or before time $T = t$
- $S(t)$ is the survival function for T, where $S(t) = 1 - F(t) = P(T > t)$
  - The probability that no event occurs until at least time $T = t$
- $h(t)$ is the hazard function for T, where $h(t) = P(t \leq T < t + \tau | T \geq t)$
  - The probability of an event occurring at $T = t$ given survival up to time $T = t$

THE EXPONENTIAL DISTRIBUTION AS A DURATION MODEL

The Poisson Process

- Particular example of a stochastic process
- Principles:
  - The numbers of events occurring in two disjoint intervals of time or space are independent (independent increments)
  - The probability distribution of the number of events which occur in a particular interval of time or unit of space depends only on the length of the interval or size of the space because the arrival rate is constant
  - Events occur at a rate of λ ("arrival rate"; expected occurences per unit of time or space)
- Let $N_\tau =$ number of arrivals in a time period of length τ
  - $N_\tau \sim \text{Poisson}(\lambda\tau)$

The Exponential Distribution as a Duration Model

Connecting the Poisson to the Exponential

The exponential distribution measures the intervals between arrivals in a Poisson process.

Principles:

- T = the amount of time until the next event occurs in a Poisson process with rate λ
- $T \sim \text{Expo}(\lambda)$
- Memorylessness
  - Your expected wait time for the next event, $E[T]$, is not affected by the time you have already waited, or:
  - $P(T > t + k | T > t) = P(T > k)$

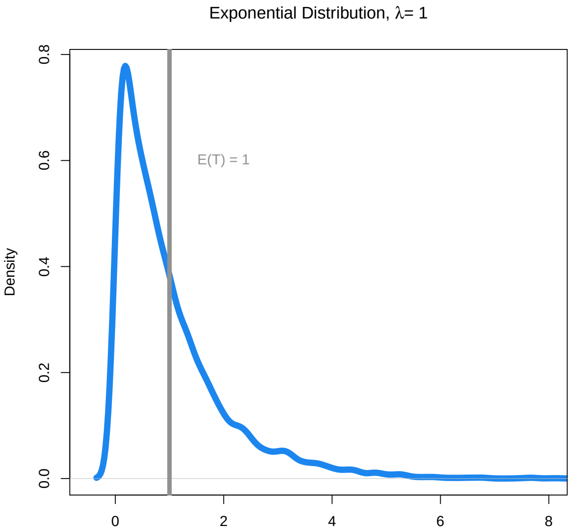
The Exponential Distribution as a Duration Model

Parameterizing the Exponential Distribution

We can parameterize the exponential distribution in two different ways:

- Rate Parameter
  - $T \sim \text{Expo}(\lambda)$
  - $f(t) = \lambda e^{-\lambda t}$
  - $\lambda > 0$ is the rate parameter
  - $E(T) = \frac{1}{\lambda} = \frac{1}{\exp(X\beta)}$
- Scale Parameter
  - $T \sim \text{Expo}(\theta)$
  - $f(t) = \frac{1}{\theta} e^{-\frac{t}{\theta}}$
  - $\theta > 0$ is the scale parameter, where $\theta = \frac{1}{\lambda}$
  - $E(T) = \theta = \exp(X\beta)$

The Exponential Distribution as a Duration Model



The Exponential Distribution as a Duration Model

Survival Model Structure with Rate Parameterization

Assuming $T \sim \text{Expo}(\lambda)$:

- $f(t) = \lambda e^{-\lambda t}$
- $F(t) = 1 - e^{-\lambda t}$
- $S(t) = 1 - F(t)$
  - $1 - (1 - e^{\lambda t})$
  - $e^{-\lambda t}$
- $h(t) = \frac{f(t)}{S(t)}$
  - $\frac{\lambda e^{-\lambda t}}{e^{-\lambda t}}$
  - λ

The Exponential Distribution as a Duration Model

Survival Model Structure with Scale Parameterization

Assuming $T \sim \text{Expo}(\theta)$:

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

The Exponential Distribution as a Duration Model

Interpreting the Hazard Function

Exponential models assume a flat, or constant, hazard.

- Every unit has its own hazard rate, but that rate doesn't change over time
- $h(t) = \lambda$ or $h(t) = \frac{1}{\theta}$, but neither depends on T!

In the rate parameterization:

- $h(t) = \lambda = \exp(X\beta)$
  - Positive coefficients (β) imply that the hazard rate increases, so average survival time is *decreasing* with X.

In the scale parameterization:

- $h(t) = \frac{1}{\theta} = \exp(-X\beta)$
  - Positive coefficients (β) imply that the hazard rate decreases, so average survival time is *increasing* with X.

The Exponential Distribution as a Duration Model

Estimating the Model, Scale Parameterization Example

Recall from last week that we can use maximum likelihood estimation to calculate estimates for our βs. The general structure of a likelihood function for a duration model is:

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

Quiz: What are the c_i and why are we raising the PDF and 1 - the CDF of T to the power of $1 - c_i$ and c_i, respectively?

The Exponential Distribution as a Duration Model

Estimating the Model, Scale Parameterization Example

Assuming $T \sim \text{Expo}(\theta)$, let's plug in $f(t)$ and $1 - F(t)$:

$$\mathcal{L} = \prod_{i=1}^{n} \left[\frac{1}{\theta_i} e^{-\frac{t_i}{\theta_i}} \right]^{1-c_i} \left[e^{-\frac{t_i}{\theta_i}} \right]^{c_i}$$

And then take the log of that expression:

$$\ell = \sum_{i=1}^{n} (1 - c_i) \left(\ln \left[\frac{1}{\theta_i} \right] - \frac{t_i}{\theta_i} \right) + c_i \left(-\frac{t_i}{\theta_i} \right)$$

The Exponential Distribution as a Duration Model

Estimating the Model, Scale Parameterization Example

Now we can simplify:

$$
\begin{aligned}
\ell &= \sum_{i=1}^{n}(1-c_i)(\ln\left[\frac{1}{\theta_i}\right]-\frac{t_i}{\theta_i})+c_i(-\frac{t_i}{\theta_i}) \\
&= \sum_{i=1}^{n}(1-c_i)(\ln\left[e^{-\mathbf{x}_i\beta}\right]-e^{-\mathbf{x}_i\beta}t_i)+c_i(-e^{-\mathbf{x}_i\beta}t_i) \\
&= \sum_{i=1}^{n}(1-c_i)(-\mathbf{x}_i\beta-e^{-\mathbf{x}_i\beta}t_i)-c_i(e^{-\mathbf{x}_i\beta}t_i) \\
&= \sum_{i=1}^{n}(1-c_i)(-\mathbf{x}_i\beta)-e^{-\mathbf{x}_i\beta}t_i
\end{aligned}
$$

The Exponential Distribution as a Duration Model

Example: Duration of Parliamentary Cabinets

- Example taken from King et al. (1990) (located in the `Zelig` package)
- 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 numbers of small blocs
  - polarization (`polar`): measure of support for extremist parties
  - numerical status (`numst2`): dummy variable coding majority (1) or minority (0) government

The Exponential Distribution as a Duration Model

Quantities of Interest

If our outcome variable is how long a parliamentary government lasts, and we're interested in the effect of majority versus minority governments. We could calculate:

- Find the hazard ratio of majority to minority governments
- Expected survival time for majority and minority governments
- Predicted survival times for majority and minority governments
- First differences in expected survival times between majority and minority governments

The Exponential Distribution as a Duration Model

Calculating Hazard Ratios

In our example, the hazard ratio for majority vs. minority governments would be given as:

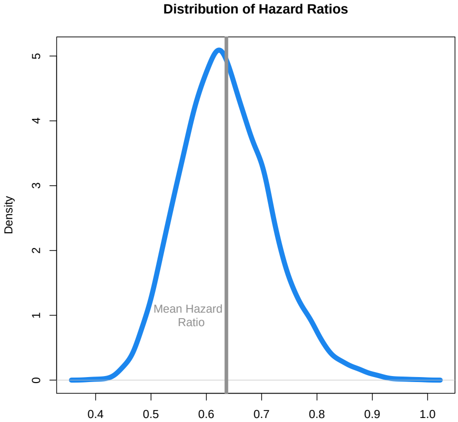
- $\text{HR} = \frac{h(t|\mathbf{x}_{\text{maj}})}{h(t|\mathbf{x}_{\text{min}})}$
- $\text{HR} = \frac{e^{-\mathbf{x}_{\text{maj}}\beta}}{e^{-\mathbf{x}_{\text{min}}\beta}}$
- $\text{HR} = \frac{e^{-\beta_0}e^{-x_1\beta_1}e^{-x_2\beta_2}e^{-x_3\beta_3}e^{-x_{\text{maj}}\beta_4}e^{-x_5\beta_5}}{e^{-\beta_0}e^{-x_1\beta_1}e^{-x_2\beta_2}e^{-x_3\beta_3}e^{-x_{\text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

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

The Exponential Distribution as a Duration Model

Calculating Hazard Ratios

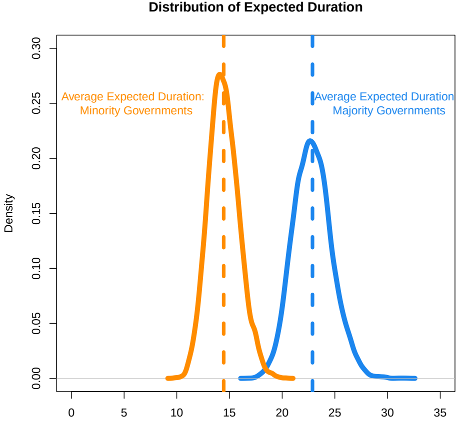
Majority governments survive longer than minority governments.



The Exponential Distribution as a Duration Model

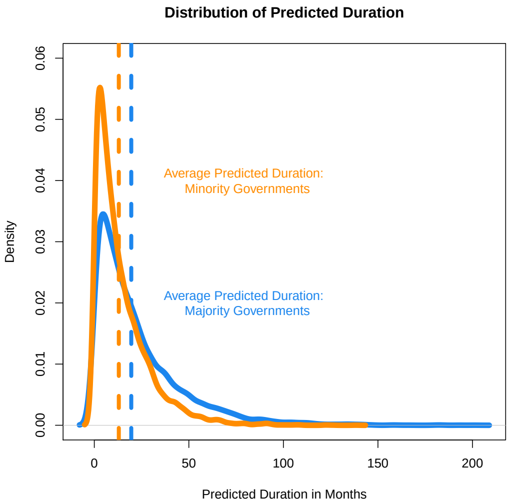
Calculating Expected Average Survival Time

$$E(T|X) = \theta = \exp(X\beta)$$



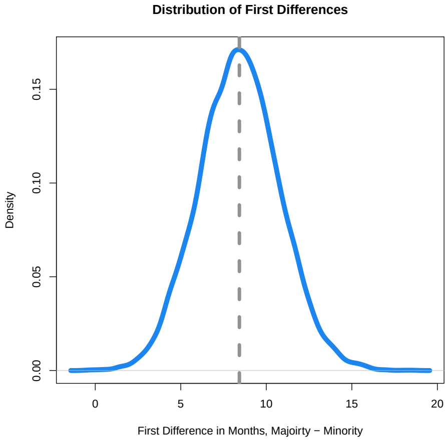
The Exponential Distribution as a Duration Model

Calculating Predicted Survival Time



The Exponential Distribution as a Duration Model

Calculating First Differences



The Weibull Distribution

When might we want to use the Weibull?

- Remember that the exponential distribution assumes a constant hazard (λ doesn't vary with time)
- But that might not be a realistic assumption to make given our DGP.
- The Weibull model lets us relax the constant hazard assumption and replace it with a monotonic hazard rate.
- This is somewhat analagous to generalizing the Poisson to a negative binomial model - we're just adding a parameter to account for additional variation in our data

The Weibull Distribution

Weibull Model Structure

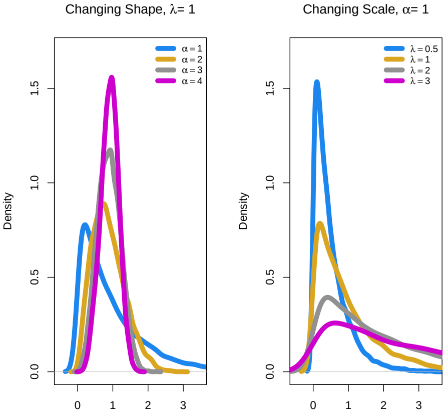
Assuming $T \sim \text{Weibull}(\lambda, \alpha)$:

- $f(t) = \frac{\alpha}{\lambda^\alpha} t^{\alpha-1} \exp\left[-\left(\frac{t}{\lambda}\right)^\alpha\right]$
- $F(t) = 1 - e^{-\left(\frac{t}{\lambda}\right)^\alpha}$
- $S(t) = 1 - F(t)$
  - $1 - (1 - e^{-\left(\frac{t}{\lambda}\right)^\alpha})$
  - $e^{-\left(\frac{t}{\lambda}\right)^\alpha}$
- $h(t) = \frac{f(t)}{S(t)}$
  - $\frac{\frac{\alpha}{\lambda^\alpha} t^{\alpha-1} \exp\left[-\left(\frac{t}{\lambda}\right)^\alpha\right]}{e^{-\left(\frac{t}{\lambda}\right)^\alpha}}$
  - $\left(\frac{\alpha}{\lambda_i}\right)\left(\frac{t_i}{\lambda_i}\right)^{\alpha-1}$
  - $\left(\frac{\alpha}{\lambda_i^\alpha}\right) t_i^{\alpha-1}$

The Weibull Distribution

What is this mess?

- $\lambda > 0$ is the scale parameter
- $\alpha > 0$ is the shape parameter



The Weibull Distribution

Properties

- $E(T) = \lambda\Gamma\left(1 + \frac{1}{\alpha}\right)$
- $\lambda = \exp(X\beta)$
- Positive βs imply that expected duration time increases with X

THE WEIBULL DISTRIBUTION

Estimating the Model

- $\mathcal{L} = \prod_{i=1}^{n} [f(t)]^{c} [S(t)]^{1-c}$
- $\mathcal{L} = \prod_{i=1}^{n} \left[\frac{\alpha}{\lambda^{\alpha}} t^{\alpha-1} \exp\left[-\left(\frac{t}{\lambda}\right)^{\alpha}\right]\right]^{c} \left[e^{-\left(\frac{t}{\lambda}\right)^{\alpha}}\right]^{1-c}$
- $\ell = \sum_{i=1}^{n} c\left[\ln(\alpha) - \alpha\ln(\lambda) + (\alpha-1)\ln(t)\right] - \left(\frac{t}{\lambda}\right)^{\alpha}$
- $\ell = \sum_{i=1}^{n} c\left[\ln(\alpha) - \alpha X\beta + (\alpha-1)\ln(t)\right] - \left(\frac{t}{\exp(X\beta)}\right)^{\alpha}$

The Weibull Distribution

Interpreting The Hazard Function

Recall from the earlier slides that $h(t_i) = \left(\frac{\alpha}{\lambda_i^{\alpha}}\right) t_i^{\alpha-1}$

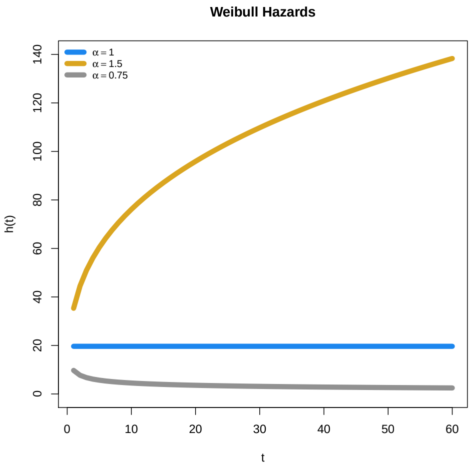
So $h(t)$ is modeled with both λ_i and α and is a function of t_i. Thus, the Weibull model assumes a **monotonic hazard**.

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

The Weibull Distribution

Interpreting The Hazard Function

Using the values from our coalition data example:



THE WEIBULL DISTRIBUTION

Quantities of Interest

- Hazard Ratios
  - Note that the Weibull distribution includes a proportional hazards assumption (that is, your hazard ratio does not depend on t):
    - HR = $\frac{h(t|x=1)}{h(t|x=0)}$

The Weibull Distribution

Quick Notes on Weibull and The Survival Package

If you use the `Survival` package in R to estimate your Weibull models, there are two things you should be aware of:

- The shape parameter α for the Weibull distribution is the reciprocal of the scale parameter given by `survreg()` (see the code for this section for how to use `survreg()` objects to estimate and plot).
- 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 Cox Proportional Hazard Model

- Often described as a semi-parametric model
- This model makes no assumptions about the shape of the hazard function or the distribution of T
- Like the Weibull, this model makes the proportional hazards assumption.

The Cox Proportional Hazard Model

Implementing the Proportional Hazard Model

1. Reconceptualize each t as a discrete event time rather than a duration or a survival time (for non-censored observations only)
   - For instance, instead of indicating that observation i survived for 5 months, $t_i = 5$ means that an event occurs at month 5
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 observed ties into account
3. Assume no events can occur between event times
4. Define a risk set, R, as the set of all possible observations in your data at risk of an event at time t.

The Cox Proportional Hazard Model

Implementing the Proportional Hazard Model

5. What observations belong in R?
   - All observations (censored or non-censored) j such that $t_j \geq t_i$
     - For instance, if $t_i = 5$, then all observations that do not experience the event OR are not censored before 5 months are in the risk set, R because the event can still happen to those observations.

The Cox Proportional Hazard Model

Cox Proportional Hazard Model Structure

- $f(t) = h(t)e^{X\beta}$
- $S(t) = e^{X\beta}$
- $h(t) = h(t)$
  - We don't specify or estimate the form of $h(t)$ in the Cox model.
  - Note also that $h(t)$ doesn't actually depend on the covariates in this model. It only depends on t.
  - Accordingly, we can estimate a *partial likelihood* for the Cox model

The Cox Proportional Hazard Model

Cox Proportional Hazard Model Partial Likelihood

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

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

The Cox Proportional Hazard Model

Modeling the Hazard Ratio

- Just like in parametric models, we model $h(t)$ using covariates:
  - $h_i(t_i) = e^{X\beta}$
  - $\mathcal{L} = \prod_{i=1}^{n} \left[\frac{e^{X\beta}}{\sum_{j \in R_i} e^{X\beta}} \right]^{c_i}$
- Note that unlike in previous parametric models, a positive β now indicates that increases in X are associated with *decreases* in the survival time (and an increase in the hazard rate).
- We don't estimate a β_0 term, which implies that we aren't modeling the shape of the baseline hazard rate.

The Cox Proportional Hazard Model

Considering the Cox Proportional Hazard Model

Pros:

- Makes no restrictive assumptions about the shape of the hazard function
- A good choice of model if you just want to see the effects of the covariates on survival and the nature of the time dependence is unimportant to you

Cons:

- The only quantities of interest you can calculate here are hazard ratios
- This can be subject to overfitting
- Sape of the hazard is unknown (though there are semi-parametric ways to derive the hazard and survivor functions, I'll show you that in the next slide)

The Cox Proportional Hazard Model

Cox Proportional Hazard Model in R

In R, the easiest way to estimate a Cox model is to use the `coxph()` function in the `survival` package.

Using our running example of coalition governments, this would be:

```
coxmod<-coxph(Surv(duration, ciep12)~invest+fract+polar+numst2+crisis, data=coalition)
```

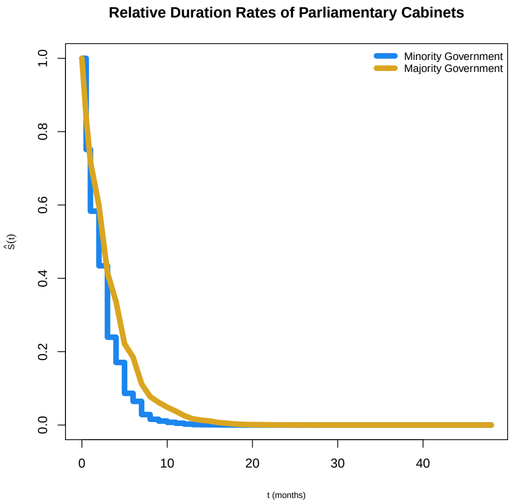
Cox Proportional Hazard Model in R

It's actually difficult to plot quantities of interest from the model because what you're interested in is largely unobserved and you have to rely on post-estimation techniques to get things like the baseline hazard and survival functions.

Still, it's possible to plot things like comparative survival functions:

The Cox Proportional Hazard Model

Cox Proportional Hazard Model in R



Other Parametric Duration Models

These are not the only duration models you can use to model survival times!

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

More resources about survival modeling

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

Long, S. J. (1997) *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: SAGE Publications, Inc.

McCullagh, Peter; Nelder, John (1989). *Generalized Linear Models*, Second Edition. Boca Raton: Chapman and Hall/CRC

Lam, Patrick. Survival Model Notes. http://www.patricklam.org/teaching.html