Advanced Quantitative Research Methodology, Lecture Notes: Models for Time Varying Events Data[1]

Gary King
http://GKing.Harvard.Edu

January 28, 2012

Case-Control Data Collection Designs

Readings:

1. King, Gary and Langche Zeng. "Logistic Regression in Rare Events Data," *Political Analysis*, Vol. 9, No. 2 (Spring, 2001): Pp. 137–163.
2. King, Gary and Langche Zeng. "Explaining Rare Events in International Relations," *International Organization*, Vol. 55, No. 3 (Summer, 2001): Pp. 693–715. [a less technical version of the PA article.]
3. King, Gary and Langche Zeng. 2001. "Estimating Risk and Rate Levels, Ratios, and Differences in Case-Control Studies," *Statistics in Medicine*, in press.
4. Tomz, Michael; Gary King; and Langche Zeng. *ReLogit: Rare Events Logistic Regression software*. for Gauss and Stata.
5. Copies of all are at `http://GKing.Harvard.edu`.

Notation for Time-Varying Rare Events

Notation for Time-Varying Rare Events

1. Let T be a random variable representing the duration until the next event (spells of peace, employment, longivity, etc.)

Notation for Time-Varying Rare Events

1. Let T be a random variable representing the duration until the next event (spells of peace, employment, longivity, etc.)
2. t is the realization

Notation for Time-Varying Rare Events

1. Let T be a random variable representing the duration until the next event (spells of peace, employment, longivity, etc.)
2. t is the realization
3. Probability density:

Notation for Time-Varying Rare Events

1. Let T be a random variable representing the duration until the next event (spells of peace, employment, longivity, etc.)
2. t is the realization
3. Probability density:

$$T \sim \mathrm{P}(t)$$

Notation for Time-Varying Rare Events

1. Let T be a random variable representing the duration until the next event (spells of peace, employment, longivity, etc.)
2. t is the realization
3. Probability density:

$$T \sim \mathrm{P}(t)$$

4. Cumulative density: probability of dying by time t

Notation for Time-Varying Rare Events

1. Let T be a random variable representing the duration until the next event (spells of peace, employment, longivity, etc.)
2. t is the realization
3. Probability density:
$$T \sim \mathrm{P}(t)$$
4. Cumulative density: probability of dying by time t
$$F(t) = \int_0^t \mathrm{P}(s)ds = \Pr(T \leq t)$$

Notation for Time-Varying Rare Events

1. Let T be a random variable representing the duration until the next event (spells of peace, employment, longivity, etc.)
2. t is the realization
3. Probability density:

$$T \sim \mathrm{P}(t)$$

4. Cumulative density: probability of dying by time t

$$F(t) = \int_0^t \mathrm{P}(s) ds = \Pr(T \leq t)$$

5. Survival function: probability of surviving (without an event) until at least t

Notation for Time-Varying Rare Events

1. Let T be a random variable representing the duration until the next event (spells of peace, employment, longivity, etc.)
2. t is the realization
3. Probability density:
$$T \sim \mathrm{P}(t)$$
4. Cumulative density: probability of dying by time t
$$F(t) = \int_0^t \mathrm{P}(s)ds = \Pr(T \leq t)$$
5. Survival function: probability of surviving (without an event) until at least t
$$S(t) = 1 - F(t) = \Pr(T \geq t)$$

Notation for Time-Varying Rare Events

6. Conditional probability of making it to $t + \Delta$ after having made it to t:

Notation for Time-Varying Rare Events

6. Conditional probability of making it to $t + \Delta$ after having made it to t:

$$\Pr(t \leq T \leq t + \Delta | T \geq t) = \frac{F(t + \Delta) - F(t)}{S(t)}$$

Notation for Time-Varying Rare Events

6. Conditional probability of making it to $t + \Delta$ after having made it to t:

$$\Pr(t \leq T \leq t + \Delta | T \geq t) = \frac{F(t+\Delta) - F(t)}{S(t)}$$

which follows the rule: $\Pr(A|B) = \Pr(AB)/\Pr(B)$.

Notation for Time-Varying Rare Events

6. Conditional probability of making it to $t + \Delta$ after having made it to t:

$$\Pr(t \leq T \leq t + \Delta | T \geq t) = \frac{F(t + \Delta) - F(t)}{S(t)}$$

   which follows the rule: $\Pr(A|B) = \Pr(AB)/\Pr(B)$.

7. Hazard rate (hard to understand *at first*, but important):

Notation for Time-Varying Rare Events

6. Conditional probability of making it to $t + \Delta$ after having made it to t:

$$\Pr(t \leq T \leq t + \Delta | T \geq t) = \frac{F(t + \Delta) - F(t)}{S(t)}$$

   which follows the rule: $\Pr(A|B) = \Pr(AB)/\Pr(B)$.

7. Hazard rate (hard to understand *at first*, but important):

$$\lambda(t) = \lim_{\Delta \to 0} \frac{\Pr(t \leq T \leq t + \Delta | T \geq t)}{\Delta}$$

Notation for Time-Varying Rare Events

6. Conditional probability of making it to $t + \Delta$ after having made it to t:

$$\Pr(t \leq T \leq t + \Delta | T \geq t) = \frac{F(t+\Delta) - F(t)}{S(t)}$$

   which follows the rule: $\Pr(A|B) = \Pr(AB)/\Pr(B)$.

7. Hazard rate (hard to understand *at first*, but important):

$$\begin{aligned} \lambda(t) &= \lim_{\Delta \to 0} \frac{\Pr(t \leq T \leq t + \Delta | T \geq t)}{\Delta} \\ &= \lim_{\Delta \to 0} \frac{F(t+\Delta) - F(t)}{S(t)\Delta} \end{aligned}$$

Notation for Time-Varying Rare Events

6. Conditional probability of making it to $t + \Delta$ after having made it to t:

$$\Pr(t \leq T \leq t + \Delta | T \geq t) = \frac{F(t + \Delta) - F(t)}{S(t)}$$

   which follows the rule: $\Pr(A|B) = \Pr(AB)/\Pr(B)$.

7. Hazard rate (hard to understand *at first*, but important):

$$\begin{aligned} \lambda(t) &= \lim_{\Delta \to 0} \frac{\Pr(t \leq T \leq t + \Delta | T \geq t)}{\Delta} \\ &= \lim_{\Delta \to 0} \frac{F(t + \Delta) - F(t)}{S(t)\Delta} \\ &= \frac{\mathrm{P}(t)}{F(t)} \end{aligned}$$

Notation for Time-Varying Rare Events

8. Understanding the hazard rate:

Notation for Time-Varying Rare Events

8. Understanding the hazard rate:
   (a) Hazard rates are a *probability per unit of time*.

Notation for Time-Varying Rare Events

8. Understanding the hazard rate:
   (a) Hazard rates are a *probability per unit of time*.
   (b) Speed in a car can be measured by number of miles driven in one hour (analogous to the *average* hazard rate). But how do we measure speed at any one *instant*?

Notation for Time-Varying Rare Events

8. Understanding the hazard rate:
   (a) Hazard rates are a *probability per unit of time*.
   (b) Speed in a car can be measured by number of miles driven in one hour (analogous to the *average* hazard rate). But how do we measure speed at any one *instant*?
   (c) The hazard rate is like the number on a car speedometer at one instant. The MPH on the speedometer & the hazard rate both change *continuously*.

Notation for Time-Varying Rare Events

8. Understanding the hazard rate:
   (a) Hazard rates are a *probability per unit of time*.
   (b) Speed in a car can be measured by number of miles driven in one hour (analogous to the *average* hazard rate). But how do we measure speed at any one *instant*?
   (c) The hazard rate is like the number on a car speedometer at one instant. The MPH on the speedometer & the hazard rate both change *continuously*.
   (d) A raw probability is unhelpful with continuous T because at the limit,

Notation for Time-Varying Rare Events

8. Understanding the hazard rate:
   (a) Hazard rates are a *probability per unit of time*.
   (b) Speed in a car can be measured by number of miles driven in one hour (analogous to the *average* hazard rate). But how do we measure speed at any one *instant*?
   (c) The hazard rate is like the number on a car speedometer at one instant. The MPH on the speedometer & the hazard rate both change *continuously*.
   (d) A raw probability is unhelpful with continuous T because at the limit,

$$\lim_{\Delta \to 0} \Pr(t \leq T \leq t + \Delta | T \geq t) = 0$$

Notation for Time-Varying Rare Events

8. Understanding the hazard rate:
   (a) Hazard rates are a *probability per unit of time*.
   (b) Speed in a car can be measured by number of miles driven in one hour (analogous to the *average* hazard rate). But how do we measure speed at any one *instant*?
   (c) The hazard rate is like the number on a car speedometer at one instant. The MPH on the speedometer & the hazard rate both change *continuously*.
   (d) A raw probability is unhelpful with continuous T because at the limit,

$$\lim_{\Delta \to 0} \Pr(t \leq T \leq t + \Delta | T \geq t) = 0$$

   (e) For the same reason, we define densities (not probabilities) for continuous variables and then compute probabilities from them.

Notation for Time-Varying Rare Events

8. Understanding the hazard rate:
   (a) Hazard rates are a *probability per unit of time*.
   (b) Speed in a car can be measured by number of miles driven in one hour (analogous to the *average* hazard rate). But how do we measure speed at any one *instant*?
   (c) The hazard rate is like the number on a car speedometer at one instant. The MPH on the speedometer & the hazard rate both change *continuously*.
   (d) A raw probability is unhelpful with continuous T because at the limit,

$$\lim_{\Delta \to 0} \Pr(t \leq T \leq t + \Delta | T \geq t) = 0$$

   (e) For the same reason, we define densities (not probabilities) for continuous variables and then compute probabilities from them.
   (f) Think of it as *the rate of event occurrence*.

Notation for Time-Varying Rare Events

9. Computing other quantities from the hazard rate: Let the *integrated hazard* (usually an intermediate quantity, not of ultimate interest) be

Notation for Time-Varying Rare Events

9. Computing other quantities from the hazard rate: Let the *integrated hazard* (usually an intermediate quantity, not of ultimate interest) be

$$\Lambda(t) = \int_0^t \lambda(t)dt$$

Notation for Time-Varying Rare Events

9. Computing other quantities from the hazard rate: Let the *integrated hazard* (usually an intermediate quantity, not of ultimate interest) be

$$\Lambda(t) = \int_0^t \lambda(t)dt$$

Then the survival function is

Notation for Time-Varying Rare Events

9. Computing other quantities from the hazard rate: Let the *integrated hazard* (usually an intermediate quantity, not of ultimate interest) be

$$\Lambda(t) = \int_0^t \lambda(t)dt$$

Then the survival function is

$$S(t) = e^{-\Lambda(t)}$$

Notation for Time-Varying Rare Events

9. Computing other quantities from the hazard rate: Let the *integrated hazard* (usually an intermediate quantity, not of ultimate interest) be

   $$\Lambda(t) = \int_0^t \lambda(t)dt$$

   Then the survival function is

   $$S(t) = e^{-\Lambda(t)}$$

   and the pdf is

Notation for Time-Varying Rare Events

9. Computing other quantities from the hazard rate: Let the *integrated hazard* (usually an intermediate quantity, not of ultimate interest) be

$$\Lambda(t) = \int_0^t \lambda(t)dt$$

Then the survival function is

$$S(t) = e^{-\Lambda(t)}$$

and the pdf is

$$\mathrm{P}(t) = S(t)/\lambda(t)$$

Notation for Time-Varying Rare Events

9. Computing other quantities from the hazard rate: Let the *integrated hazard* (usually an intermediate quantity, not of ultimate interest) be

$$\Lambda(t) = \int_0^t \lambda(t)dt$$

Then the survival function is

$$S(t) = e^{-\Lambda(t)}$$

and the pdf is

$$\mathrm{P}(t) = S(t)/\lambda(t)$$

Common practice: model the hazard rate directly and compute the density and then log-likelihood function from it.

Time Varying Quantities of Interest in Analyzing Rare Events

Time Varying Quantities of Interest in Analyzing Rare Events

1. Hazard rates:

Time Varying Quantities of Interest in Analyzing Rare Events

1. Hazard rates:

$$\lambda_0(t) \equiv \lim_{\Delta \to 0} \frac{\Pr(Y_{i,(t,t+\Delta)} = 1 | Y_{0s} = 0, \forall s < t, X_0)}{\Delta}$$

Time Varying Quantities of Interest in Analyzing Rare Events

1. Hazard rates:

$$\lambda_0(t) \equiv \lim_{\Delta\to 0} \frac{\Pr(Y_{i,(t,t+\Delta)} = 1 | Y_{0s} = 0, \forall s < t, X_0)}{\Delta}$$
$$\lambda_\ell(t) \equiv \lim_{\Delta\to 0} \frac{\Pr(Y_{i,(t,t+\Delta)} = 1 | Y_{\ell s} = 0, \forall s < t, X_\ell)}{\Delta}$$

Time Varying Quantities of Interest in Analyzing Rare Events

1. Hazard rates:

$$\lambda_0(t) \equiv \lim_{\Delta \to 0} \frac{\Pr(Y_{i,(t,t+\Delta)} = 1 | Y_{0s} = 0, \forall s < t, X_0)}{\Delta}$$

$$\lambda_\ell(t) \equiv \lim_{\Delta \to 0} \frac{\Pr(Y_{i,(t,t+\Delta)} = 1 | Y_{\ell s} = 0, \forall s < t, X_\ell)}{\Delta}$$

2. Rate ratio,

$$\text{rr}_t \equiv \lambda_\ell(t)/\lambda_0(t)$$

Time Varying Quantities of Interest in Analyzing Rare Events

1. Hazard rates:

$$\lambda_0(t) \equiv \lim_{\Delta \to 0} \frac{\Pr(Y_{i,(t,t+\Delta)} = 1 | Y_{0s} = 0, \forall s < t, X_0)}{\Delta}$$

$$\lambda_\ell(t) \equiv \lim_{\Delta \to 0} \frac{\Pr(Y_{i,(t,t+\Delta)} = 1 | Y_{\ell s} = 0, \forall s < t, X_\ell)}{\Delta}$$

2. Rate ratio,

$$\text{rr}_t \equiv \lambda_\ell(t)/\lambda_0(t)$$

3. Rate difference,

$$\text{rd}_t \equiv \lambda_\ell(t) - \lambda_0(t)$$

Time Varying Quantities of Interest in Analyzing Rare Events

4. Risk (or the Probability),

Time Varying Quantities of Interest in Analyzing Rare Events

4. Risk (or the Probability),

$$\pi_0 \equiv \Pr(Y = 1|X_0) = 1 - \exp\left(-\int_t^{t+\Delta} \lambda_0(s)ds\right)$$

Time Varying Quantities of Interest in Analyzing Rare Events

4. Risk (or the Probability),

$$\pi_0 \equiv \Pr(Y = 1|X_0) = 1 - \exp\left(-\int_t^{t+\Delta} \lambda_0(s)ds\right)$$

$$\pi_\ell \equiv \Pr(Y = 1|X_\ell) = 1 - \exp\left(-\int_t^{t+\Delta} \lambda_\ell(s)ds\right)$$

Time Varying Quantities of Interest in Analyzing Rare Events

4. Risk (or the Probability),

$$\pi_0 \equiv \Pr(Y = 1|X_0) = 1 - \exp\left(-\int_t^{t+\Delta} \lambda_0(s)ds\right)$$

$$\pi_\ell \equiv \Pr(Y = 1|X_\ell) = 1 - \exp\left(-\int_t^{t+\Delta} \lambda_\ell(s)ds\right)$$

5. Risk ratio,

Time Varying Quantities of Interest in Analyzing Rare Events

4. Risk (or the Probability),

$$\pi_0 \equiv \Pr(Y = 1|X_0) = 1 - \exp\left(-\int_t^{t+\Delta} \lambda_0(s)ds\right)$$

$$\pi_\ell \equiv \Pr(Y = 1|X_\ell) = 1 - \exp\left(-\int_t^{t+\Delta} \lambda_\ell(s)ds\right)$$

5. Risk ratio,

$$\mathrm{RR} \equiv \Pr(Y = 1|X_\ell)/\Pr(Y = 1|X_0)$$

Time Varying Quantities of Interest in Analyzing Rare Events

4. Risk (or the Probability),

$$\pi_0 \equiv \Pr(Y = 1|X_0) = 1 - \exp\left(-\int_t^{t+\Delta} \lambda_0(s)ds\right)$$

$$\pi_\ell \equiv \Pr(Y = 1|X_\ell) = 1 - \exp\left(-\int_t^{t+\Delta} \lambda_\ell(s)ds\right)$$

5. Risk ratio,

$$\text{RR} \equiv \Pr(Y = 1|X_\ell)/\Pr(Y = 1|X_0)$$

6. Risk difference (or First difference or Attributable risk),

Time Varying Quantities of Interest in Analyzing Rare Events

4. Risk (or the Probability),

$$\pi_0 \equiv \Pr(Y = 1|X_0) = 1 - \exp\left(-\int_t^{t+\Delta} \lambda_0(s)ds\right)$$

$$\pi_\ell \equiv \Pr(Y = 1|X_\ell) = 1 - \exp\left(-\int_t^{t+\Delta} \lambda_\ell(s)ds\right)$$

5. Risk ratio,

$$\mathrm{RR} \equiv \Pr(Y = 1|X_\ell)/\Pr(Y = 1|X_0)$$

6. Risk difference (or First difference or Attributable risk),

$$\mathrm{RD} \equiv \Pr(Y = 1|X_\ell) - \Pr(Y = 1|X_0)$$

Deriving the Exponential Duration Model from a constant Hazard Rate Model

Deriving the Exponential Duration Model from a constant Hazard Rate Model

1. *The model* by specifying the hazard as *not* a function of time: Let y_i be the duration until an event. Then

Deriving the Exponential Duration Model from a constant Hazard Rate Model

1. *The model* by specifying the hazard as *not* a function of time: Let y_i be the duration until an event. Then

$$\lambda_i(t) = \lambda_i = e^{x_i\beta}$$

Deriving the Exponential Duration Model from a constant Hazard Rate Model

1. *The model* by specifying the hazard as *not* a function of time: Let y_i be the duration until an event. Then

$$\lambda_i(t) = \lambda_i = e^{x_i \beta}$$

and independence over i (after conditioning on x_i).

Deriving the Exponential Duration Model from a constant Hazard Rate Model

1. *The model* by specifying the hazard as *not* a function of time: Let y_i be the duration until an event. Then

$$\lambda_i(t) = \lambda_i = e^{x_i\beta}$$

   and independence over i (after conditioning on x_i).
2. The model implies a stochastic component:

Deriving the Exponential Duration Model from a constant Hazard Rate Model

1. *The model* by specifying the hazard as *not* a function of time: Let y_i be the duration until an event. Then

$$\lambda_i(t) = \lambda_i = e^{x_i\beta}$$

   and independence over i (after conditioning on x_i).

2. The model implies a stochastic component:

$$\mathrm{P}(y_i) = S(y_i)\lambda(y_i)$$

Deriving the Exponential Duration Model from a constant Hazard Rate Model

1. *The model* by specifying the hazard as *not* a function of time: Let y_i be the duration until an event. Then

$$\lambda_i(t) = \lambda_i = e^{x_i \beta}$$

   and independence over i (after conditioning on x_i).

2. The model implies a stochastic component:

$$\begin{aligned} \mathrm{P}(y_i) &= S(y_i)\lambda(y_i) \\ &= e^{-\Lambda(y_i)}\lambda(y_i) \end{aligned}$$

Deriving the Exponential Duration Model from a constant Hazard Rate Model

1. *The model* by specifying the hazard as *not* a function of time: Let y_i be the duration until an event. Then

$$\lambda_i(t) = \lambda_i = e^{x_i\beta}$$

and independence over i (after conditioning on x_i).

2. The model implies a stochastic component:

$$\begin{aligned} \mathrm{P}(y_i) &= S(y_i)\lambda(y_i) \\ &= e^{-\Lambda(y_i)}\lambda(y_i) \\ &= \lambda(y_i)e^{-\int_0^t \lambda(y_i)dt} \quad \text{i.e., using only } \lambda_i(t) \end{aligned}$$

Deriving the Exponential Duration Model from a constant Hazard Rate Model

1. *The model* by specifying the hazard as *not* a function of time: Let y_i be the duration until an event. Then

$$\lambda_i(t) = \lambda_i = e^{x_i\beta}$$

and independence over i (after conditioning on x_i).

2. The model implies a stochastic component:

$$\begin{aligned}
\mathrm{P}(y_i) &= S(y_i)\lambda(y_i) \\
&= e^{-\Lambda(y_i)}\lambda(y_i) \\
&= \lambda(y_i)e^{-\int_0^t \lambda(y_i)dt} \quad \text{i.e., using only } \lambda_i(t) \\
&= \lambda_i e^{-\lambda_i y_i}
\end{aligned}$$

Deriving the Exponential Duration Model from a constant Hazard Rate Model

1. *The model* by specifying the hazard as *not* a function of time: Let y_i be the duration until an event. Then

$$\lambda_i(t) = \lambda_i = e^{x_i \beta}$$

and independence over i (after conditioning on x_i).

2. The model implies a stochastic component:

$$\begin{aligned}
\text{P}(y_i) &= S(y_i)\lambda(y_i) \\
&= e^{-\Lambda(y_i)}\lambda(y_i) \\
&= \lambda(y_i)e^{-\int_0^t \lambda(y_i)dt} \quad \text{i.e., using only } \lambda_i(t) \\
&= \lambda_i e^{-\lambda_i y_i} \\
&= \text{expon}(y_i|\lambda_i)
\end{aligned}$$

Deriving the Exponential Duration Model from a constant Hazard Rate Model

1. *The model* by specifying the hazard as *not* a function of time: Let y_i be the duration until an event. Then

$$\lambda_i(t) = \lambda_i = e^{x_i\beta}$$

and independence over i (after conditioning on x_i).

2. The model implies a stochastic component:

$$\begin{aligned}
\mathrm{P}(y_i) &= S(y_i)\lambda(y_i) \\
&= e^{-\Lambda(y_i)}\lambda(y_i) \\
&= \lambda(y_i)e^{-\int_0^t \lambda(y_i)dt} \quad \text{i.e., using only } \lambda_i(t) \\
&= \lambda_i e^{-\lambda_i y_i} \\
&= \text{expon}(y_i|\lambda_i)
\end{aligned}$$

with systematic component $\lambda_i = e^{x_i\beta}$.

Deriving the Exponential Duration Model from a constant Hazard Rate Model

3. The likelihood is

Deriving the Exponential Duration Model from a constant Hazard Rate Model

3. The likelihood is

$$L(\beta|y) = \prod_{i=1}^{n} \lambda_i e^{-\lambda_i y_i}$$

Deriving the Exponential Duration Model from a constant Hazard Rate Model

3. The likelihood is

$$L(\beta|y) = \prod_{i=1}^{n} \lambda_i e^{-\lambda_i y_i}$$

4. The Log-likelihood is

Deriving the Exponential Duration Model from a constant Hazard Rate Model

3. The likelihood is

$$L(\beta|y) = \prod_{i=1}^{n} \lambda_i e^{-\lambda_i y_i}$$

4. The Log-likelihood is

$$\ln L(\beta|y) = \sum_{i=1}^{n} \{\ln \lambda_i - \lambda_i y_i\}$$

Deriving the Exponential Duration Model from a constant Hazard Rate Model

3. The likelihood is

$$L(\beta|y) = \prod_{i=1}^{n} \lambda_i e^{-\lambda_i y_i}$$

4. The Log-likelihood is

$$\ln L(\beta|y) = \sum_{i=1}^{n} \{\ln \lambda_i - \lambda_i y_i\}$$

and using the systematic component gives

Deriving the Exponential Duration Model from a constant Hazard Rate Model

3. The likelihood is

$$L(\beta|y) = \prod_{i=1}^{n} \lambda_i e^{-\lambda_i y_i}$$

4. The Log-likelihood is

$$\ln L(\beta|y) = \sum_{i=1}^{n} \{\ln \lambda_i - \lambda_i y_i\}$$

and using the systematic component gives

$$= \sum_{i=1}^{n} \left\{ X_i\beta - e^{X_i\beta} y_i \right\}$$

Deriving the Exponential Duration Model from a constant Hazard Rate Model

5. Hence, a constant hazard rate is equivalent to the exponential duration model:

Deriving the Exponential Duration Model from a constant Hazard Rate Model

5. Hence, a constant hazard rate is equivalent to the exponential duration model:

$$Y_i \sim \text{expon}(y_i|\lambda_i)$$

Deriving the Exponential Duration Model from a constant Hazard Rate Model

5. Hence, a constant hazard rate is equivalent to the exponential duration model:

$$Y_i \sim \text{expon}(y_i|\lambda_i)$$
$$E(Y_i) \equiv \frac{1}{\lambda_i} = \frac{1}{e^{-x_i\beta}} = e^{x_i\beta}$$

Deriving the Exponential Duration Model from a constant Hazard Rate Model

5. Hence, a constant hazard rate is equivalent to the exponential duration model:

$$Y_i \sim \text{expon}(y_i|\lambda_i)$$
$$E(Y_i) \equiv \frac{1}{\lambda_i} = \frac{1}{e^{-x_i\beta}} = e^{x_i\beta}$$

6. Censoring could be added as described previously.

Interpreting a constant hazard rate model (and the exponential duration model assumption)

Interpreting a constant hazard rate model (and the exponential duration model assumption)

1. The hazard rate varies over observations i.

Interpreting a constant hazard rate model (and the exponential duration model assumption)

1. The hazard rate varies over observations i.
2. The hazard rate is constant over time for any observation.

Interpreting a constant hazard rate model (and the exponential duration model assumption)

1. The hazard rate varies over observations i.
2. The hazard rate is constant over time for any observation.
3. In other words, no matter how long we watch an observation the rate of event occurance, or hazard of an event occuring, is constant.

Interpreting a constant hazard rate model (and the exponential duration model assumption)

1. The hazard rate varies over observations i.
2. The hazard rate is constant over time for any observation.
3. In other words, no matter how long we watch an observation the rate of event occurance, or hazard of an event occuring, is constant.
4. How to violate the constant hazard rate assumption:

Interpreting a constant hazard rate model (and the exponential duration model assumption)

1. The hazard rate varies over observations i.
2. The hazard rate is constant over time for any observation.
3. In other words, no matter how long we watch an observation the rate of event occurance, or hazard of an event occuring, is constant.
4. How to violate the constant hazard rate assumption:
   (a) *Positive duration dependence* or a *rising hazard rate*

Interpreting a constant hazard rate model (and the exponential duration model assumption)

1. The hazard rate varies over observations i.
2. The hazard rate is constant over time for any observation.
3. In other words, no matter how long we watch an observation the rate of event occurance, or hazard of an event occuring, is constant.
4. How to violate the constant hazard rate assumption:
    (a) *Positive duration dependence* or a *rising hazard rate*
        i. Things that wear out or "rust"

Interpreting a constant hazard rate model (and the exponential duration model assumption)

1. The hazard rate varies over observations i.
2. The hazard rate is constant over time for any observation.
3. In other words, no matter how long we watch an observation the rate of event occurance, or hazard of an event occuring, is constant.
4. How to violate the constant hazard rate assumption:
   (a) *Positive duration dependence* or a *rising hazard rate*
      i. Things that wear out or "rust"
      ii. e.g., the longer people live (after $\approx$ 5 years of age), the higher their risk of death.

Interpreting a constant hazard rate model (and the exponential duration model assumption)

1. The hazard rate varies over observations i.
2. The hazard rate is constant over time for any observation.
3. In other words, no matter how long we watch an observation the rate of event occurance, or hazard of an event occuring, is constant.
4. How to violate the constant hazard rate assumption:
   (a) *Positive duration dependence* or a *rising hazard rate*
      i. Things that wear out or “rust”
      ii. e.g., the longer people live (after $\approx$ 5 years of age), the higher their risk of death.
   (b) *Negative duration dependence* or a *decreasing hazard rate*

Interpreting a constant hazard rate model (and the exponential duration model assumption)

1. The hazard rate varies over observations i.
2. The hazard rate is constant over time for any observation.
3. In other words, no matter how long we watch an observation the rate of event occurance, or hazard of an event occuring, is constant.
4. How to violate the constant hazard rate assumption:
   (a) *Positive duration dependence* or a *rising hazard rate*
      i. Things that wear out or "rust"
      ii. e.g., the longer people live (after $\approx$ 5 years of age), the higher their risk of death.
   (b) *Negative duration dependence* or a *decreasing hazard rate*
      i. Things that get better with age

Interpreting a constant hazard rate model (and the exponential duration model assumption)

1. The hazard rate varies over observations i.
2. The hazard rate is constant over time for any observation.
3. In other words, no matter how long we watch an observation the rate of event occurance, or hazard of an event occuring, is constant.
4. How to violate the constant hazard rate assumption:
   (a) *Positive duration dependence* or a *rising hazard rate*
      i. Things that wear out or "rust"
      ii. e.g., the longer people live (after $\approx$ 5 years of age), the higher their risk of death.

   (b) *Negative duration dependence* or a *decreasing hazard rate*
      i. Things that get better with age
      ii. e.g., the risk of being fired from a job drops the longer you have it.

Interpreting a constant hazard rate model (and the exponential duration model assumption)

5. Generalizations of the exponential model that allow duration dependence:

Interpreting a constant hazard rate model (and the exponential duration model assumption)

5. Generalizations of the exponential model that allow duration dependence:
   (a) Weibull model: monotonically increasing or decreasing hazard (depending on the value of an extra parameter it has)

Interpreting a constant hazard rate model (and the exponential duration model assumption)

5. Generalizations of the exponential model that allow duration dependence:
   (a) Weibull model: monotonically increasing or decreasing hazard (depending on the value of an extra parameter it has)
   (b) Log-normal model: hazard increases and then decreases

Interpreting a constant hazard rate model (and the exponential duration model assumption)

5. Generalizations of the exponential model that allow duration dependence:
   (a) Weibull model: monotonically increasing or decreasing hazard (depending on the value of an extra parameter it has)
   (b) Log-normal model: hazard increases and then decreases
   (c) Many others with different patterns parameterized in different ways.

Interpreting a constant hazard rate model (and the exponential duration model assumption)

5. Generalizations of the exponential model that allow duration dependence:
   (a) Weibull model: monotonically increasing or decreasing hazard (depending on the value of an extra parameter it has)
   (b) Log-normal model: hazard increases and then decreases
   (c) Many others with different patterns parameterized in different ways.
   (d) Some nonparametric methods exist. Leading example: Cox's proportional hazards model

Interpreting a constant hazard rate model (and the exponential duration model assumption)

5. Generalizations of the exponential model that allow duration dependence:
   (a) Weibull model: monotonically increasing or decreasing hazard (depending on the value of an extra parameter it has)
   (b) Log-normal model: hazard increases and then decreases
   (c) Many others with different patterns parameterized in different ways.
   (d) Some nonparametric methods exist. Leading example: Cox's proportional hazards model
   (e) Frailty models: duration models with random effects

The Weibull Duration Model

The Weibull Duration Model

1. *The Weibull Model*, specified from the hazard rate, is:

The Weibull Duration Model

1. *The Weibull Model*, specified from the hazard rate, is:

$$\lambda_i(t) = \lambda_i p(\lambda_i t)^{p-1}$$

The Weibull Duration Model

1. *The Weibull Model*, specified from the hazard rate, is:

$$\lambda_i(t) = \lambda_i p(\lambda_i t)^{p-1}$$

with parameters $\lambda_i \equiv e^{-X_i\beta}$ and p.

The Weibull Duration Model

1. *The Weibull Model*, specified from the hazard rate, is:

$$\lambda_i(t) = \lambda_i p(\lambda_i t)^{p-1}$$

   with parameters $\lambda_i \equiv e^{-X_i\beta}$ and p.
2. For any i, the hazard is not constant over t (like in the exponential model). Its either monotonically increasing or decreasing, depending on parameter p (which is estimated).

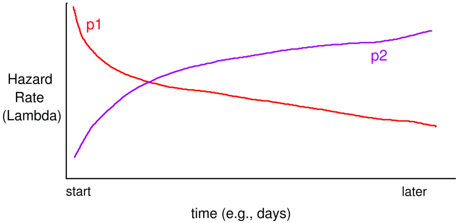
The Weibull Duration Model

1. *The Weibull Model*, specified from the hazard rate, is:

$$\lambda_i(t) = \lambda_i p(\lambda_i t)^{p-1}$$

   with parameters $\lambda_i \equiv e^{-X_i\beta}$ and p.

2. For any i, the hazard is not constant over t (like in the exponential model). Its either monotonically increasing or decreasing, depending on parameter p (which is estimated).



3. Weibull is more general than the exponential, but it does not encompass all functions. E.g., if a hazard for some application goes up and then down, the Weibull would not be appropriate. Can you provide a substantive example?

3. Weibull is more general than the exponential, but it does not encompass all functions. E.g., if a hazard for some application goes up and then down, the Weibull would not be appropriate. Can you provide a substantive example?
4. Assume X_i is a constant term and covariates that do not vary over time t. Let y_i be the duration until an event (number of days, etc.)

3. Weibull is more general than the exponential, but it does not encompass all functions. E.g., if a hazard for some application goes up and then down, the Weibull would not be appropriate. Can you provide a substantive example?
4. Assume X_i is a constant term and covariates that do not vary over time t. Let y_i be the duration until an event (number of days, etc.)
5. The model for the hazard implies a stochastic component:

3. Weibull is more general than the exponential, but it does not encompass all functions. E.g., if a hazard for some application goes up and then down, the Weibull would not be appropriate. Can you provide a substantive example?
4. Assume X_i is a constant term and covariates that do not vary over time t. Let y_i be the duration until an event (number of days, etc.)
5. The model for the hazard implies a stochastic component:

$$\mathrm{P}(y_i) = S(y_i)\lambda(y_i)$$

3. Weibull is more general than the exponential, but it does not encompass all functions. E.g., if a hazard for some application goes up and then down, the Weibull would not be appropriate. Can you provide a substantive example?
4. Assume X_i is a constant term and covariates that do not vary over time t. Let y_i be the duration until an event (number of days, etc.)
5. The model for the hazard implies a stochastic component:

$$\begin{aligned} \mathrm{P}(y_i) &= S(y_i)\lambda(y_i) \\ &= p\lambda_i(\lambda_i y_i)^{p-1} e^{-(\lambda_i y_i)^p} \end{aligned}$$

3. Weibull is more general than the exponential, but it does not encompass all functions. E.g., if a hazard for some application goes up and then down, the Weibull would not be appropriate. Can you provide a substantive example?
4. Assume X_i is a constant term and covariates that do not vary over time t. Let y_i be the duration until an event (number of days, etc.)
5. The model for the hazard implies a stochastic component:

$$\begin{aligned} \mathrm{P}(y_i) &= S(y_i)\lambda(y_i) \\ &= p\lambda_i(\lambda_i y_i)^{p-1} e^{-(\lambda_i y_i)^p} \\ &= \mathrm{Weibull}(y_i|\lambda_i, p) \end{aligned}$$

3. Weibull is more general than the exponential, but it does not encompass all functions. E.g., if a hazard for some application goes up and then down, the Weibull would not be appropriate. Can you provide a substantive example?
4. Assume X_i is a constant term and covariates that do not vary over time t. Let y_i be the duration until an event (number of days, etc.)
5. The model for the hazard implies a stochastic component:

$$\begin{aligned} \mathrm{P}(y_i) &= S(y_i)\lambda(y_i) \\ &= p\lambda_i(\lambda_i y_i)^{p-1} e^{-(\lambda_i y_i)^p} \\ &= \mathrm{Weibull}(y_i|\lambda_i, p) \end{aligned}$$

6. The systematic component is $\lambda_i = e^{-X_i\beta}$, but note that the expected duration is

3. Weibull is more general than the exponential, but it does not encompass all functions. E.g., if a hazard for some application goes up and then down, the Weibull would not be appropriate. Can you provide a substantive example?
4. Assume X_i is a constant term and covariates that do not vary over time t. Let y_i be the duration until an event (number of days, etc.)
5. The model for the hazard implies a stochastic component:

$$\begin{aligned} \mathrm{P}(y_i) &= S(y_i)\lambda(y_i) \\ &= p\lambda_i(\lambda_i y_i)^{p-1}e^{-(\lambda_i y_i)^p} \\ &= \mathrm{Weibull}(y_i|\lambda_i, p) \end{aligned}$$

6. The systematic component is $\lambda_i = e^{-X_i\beta}$, but note that the expected duration is

$$E(Y_i|X_i) = \int_0^\infty y \times \mathrm{Weibull}(y|\lambda_i, p)dy$$

3. Weibull is more general than the exponential, but it does not encompass all functions. E.g., if a hazard for some application goes up and then down, the Weibull would not be appropriate. Can you provide a substantive example?
4. Assume X_i is a constant term and covariates that do not vary over time t. Let y_i be the duration until an event (number of days, etc.)
5. The model for the hazard implies a stochastic component:

$$\begin{aligned} \mathrm{P}(y_i) &= S(y_i)\lambda(y_i) \\ &= p\lambda_i(\lambda_i y_i)^{p-1} e^{-(\lambda_i y_i)^p} \\ &= \mathrm{Weibull}(y_i|\lambda_i, p) \end{aligned}$$

6. The systematic component is $\lambda_i = e^{-X_i\beta}$, but note that the expected duration is

$$\begin{aligned} E(Y_i|X_i) &= \int_0^\infty y \times \mathrm{Weibull}(y|\lambda_i, p)dy \\ &= e^{X_i\beta}\Gamma[1 + 1/p]. \end{aligned}$$

3. Weibull is more general than the exponential, but it does not encompass all functions. E.g., if a hazard for some application goes up and then down, the Weibull would not be appropriate. Can you provide a substantive example?
4. Assume X_i is a constant term and covariates that do not vary over time t. Let y_i be the duration until an event (number of days, etc.)
5. The model for the hazard implies a stochastic component:

$$\begin{aligned} \mathrm{P}(y_i) &= S(y_i)\lambda(y_i) \\ &= p\lambda_i(\lambda_i y_i)^{p-1}e^{-(\lambda_i y_i)^p} \\ &= \mathrm{Weibull}(y_i|\lambda_i, p) \end{aligned}$$

6. The systematic component is $\lambda_i = e^{-X_i\beta}$, but note that the expected duration is

$$\begin{aligned} E(Y_i|X_i) &= \int_0^\infty y \times \mathrm{Weibull}(y|\lambda_i, p)dy \\ &= e^{X_i\beta}\Gamma[1 + 1/p]. \end{aligned}$$

How would you simulate Quantities of Interest?

7. The likelihood function is

7. The likelihood function is

$$L(\beta, p|y) = \prod_{i=1}^{n} \text{Weibull}(y_i|\lambda_i, p)$$

7. The likelihood function is

$$\begin{aligned} L(\beta, p|y) &= \prod_{i=1}^{n} \text{Weibull}(y_i|\lambda_i, p) \\ &= \prod_{i=1}^{n} p\lambda_i(\lambda_i y_i)^{p-1} e^{-(\lambda_i y_i)^p} \end{aligned}$$

7. The likelihood function is

$$\begin{aligned} L(\beta, p|y) &= \prod_{i=1}^{n} \text{Weibull}(y_i|\lambda_i, p) \\ &= \prod_{i=1}^{n} p\lambda_i(\lambda_i y_i)^{p-1} e^{-(\lambda_i y_i)^p} \end{aligned}$$

Then substitute in for the systematic component $\lambda_i = e^{-X_i\beta}$ and take logs to get the log-likelihood.

The Cox Proportional Hazards Model

The Cox Proportional Hazards Model

1. Suppose you don't know the hazard rate and are not willing to make an assumption about it.

The Cox Proportional Hazards Model

1. Suppose you don't know the hazard rate and are not willing to make an assumption about it.
2. The basic assumption: the hazard factors into a piece that varies over t but not i and a piece that varies with the covariates X_i over i but not t:

The Cox Proportional Hazards Model

1. Suppose you don't know the hazard rate and are not willing to make an assumption about it.
2. The basic assumption: the hazard factors into a piece that varies over t but not i and a piece that varies with the covariates X_i over i but not t:

$$\lambda_i(t) = \lambda(t) \times \lambda_i$$

The Cox Proportional Hazards Model

1. Suppose you don't know the hazard rate and are not willing to make an assumption about it.
2. The basic assumption: the hazard factors into a piece that varies over t but not i and a piece that varies with the covariates X_i over i but not t:

$$\lambda_i(t) = \lambda(t) \times \lambda_i$$

and by assumption $\lambda_i = e^{X_i\beta}$,

$$= \lambda(t)e^{X_i\beta}$$

The Cox Proportional Hazards Model

1. Suppose you don't know the hazard rate and are not willing to make an assumption about it.
2. The basic assumption: the hazard factors into a piece that varies over t but not i and a piece that varies with the covariates X_i over i but not t:

$$\lambda_i(t) = \lambda(t) \times \lambda_i$$

and by assumption $\lambda_i = e^{X_i\beta}$,

$$= \lambda(t)e^{X_i\beta}$$

where $\lambda(t)$ is known as the *baseline hazard*.

3. This is known as the *proportional hazards* assumption because if we are mainly interested in variation over i:

3. This is known as the *proportional hazards* assumption because if we are mainly interested in variation over i:

$$\lambda_i(t) = \lambda(t) \times \lambda_i$$

3. This is known as the *proportional hazards* assumption because if we are mainly interested in variation over i:

$$\lambda_i(t) = \lambda(t) \times \lambda_i$$
$$\propto \lambda_i$$

3. This is known as the *proportional hazards* assumption because if we are mainly interested in variation over i:

$$\lambda_i(t) = \lambda(t) \times \lambda_i$$
$$\propto \lambda_i$$

or in other words for two values 0 and ℓ, the *rate ratio* is:

3. This is known as the *proportional hazards* assumption because if we are mainly interested in variation over i:

$$\lambda_i(t) = \lambda(t) \times \lambda_i$$
$$\propto \lambda_i$$

or in other words for two values 0 and ℓ, the *rate ratio* is:

$$\text{rr} = \frac{\lambda_0(t)}{\lambda_\ell(t)}$$

3. This is known as the *proportional hazards* assumption because if we are mainly interested in variation over i:

$$\begin{aligned} \lambda_i(t) &= \lambda(t) \times \lambda_i \\ &\propto \lambda_i \end{aligned}$$

or in other words for two values 0 and ℓ, the *rate ratio* is:

$$\begin{aligned} \text{rr} &= \frac{\lambda_0(t)}{\lambda_\ell(t)} \\ &= \frac{\lambda(t) \times \lambda_0}{\lambda(t) \times \lambda_\ell} \end{aligned}$$

3. This is known as the *proportional hazards* assumption because if we are mainly interested in variation over i:

$$\lambda_i(t) = \lambda(t) \times \lambda_i$$
$$\propto \lambda_i$$

or in other words for two values 0 and ℓ, the *rate ratio* is:

$$\text{rr} = \frac{\lambda_0(t)}{\lambda_\ell(t)} = \frac{\lambda(t) \times \lambda_0}{\lambda(t) \times \lambda_\ell} = \frac{\lambda_0}{\lambda_\ell}$$

3. This is known as the *proportional hazards* assumption because if we are mainly interested in variation over i:

$$\begin{aligned}\lambda_i(t) &= \lambda(t) \times \lambda_i \\ &\propto \lambda_i\end{aligned}$$

or in other words for two values 0 and ℓ, the *rate ratio* is:

$$\begin{aligned}\text{rr} &= \frac{\lambda_0(t)}{\lambda_\ell(t)} \\ &= \frac{\lambda(t) \times \lambda_0}{\lambda(t) \times \lambda_\ell} \\ &= \frac{\lambda_0}{\lambda_\ell}\end{aligned}$$

and so the time component drops out.

3. This is known as the *proportional hazards* assumption because if we are mainly interested in variation over i:

$$\lambda_i(t) = \lambda(t) \times \lambda_i$$
$$\propto \lambda_i$$

or in other words for two values 0 and ℓ, the *rate ratio* is:

$$\text{rr} = \frac{\lambda_0(t)}{\lambda_\ell(t)} = \frac{\lambda(t) \times \lambda_0}{\lambda(t) \times \lambda_\ell} = \frac{\lambda_0}{\lambda_\ell}$$

and so the time component drops out.

4. Thus, if we are interested in rr, we do not need to know anything about $\lambda(t)$ other than that it does not vary by i — the proportional hazards assumption.

3. This is known as the *proportional hazards* assumption because if we are mainly interested in variation over i:

$$\begin{aligned} \lambda_i(t) &= \lambda(t) \times \lambda_i \\ &\propto \lambda_i \end{aligned}$$

or in other words for two values 0 and ℓ, the *rate ratio* is:

$$\begin{aligned} \text{rr} &= \frac{\lambda_0(t)}{\lambda_\ell(t)} \\ &= \frac{\lambda(t) \times \lambda_0}{\lambda(t) \times \lambda_\ell} \\ &= \frac{\lambda_0}{\lambda_\ell} \end{aligned}$$

and so the time component drops out.

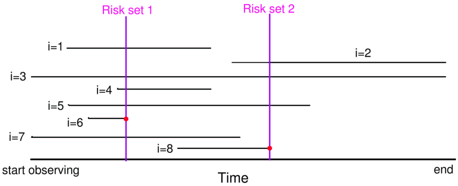
4. Thus, if we are interested in rr, we do not need to know anything about $\lambda(t)$ other than that it does not vary by i — the proportional hazards assumption.
5. Cox also devised an estimation strategy that made estimating the baseline hazard (the constant term) unnecessary. We will discuss this shortly.

6. If the proportional hazards assumption does not hold, using this model will lead to biased inferences. An example violation for government durations: democratic governments build a coalitions of minorities and have increasing hazard rates, whereas autocratic governments eliminate opposition, consolidate support, and so have declining hazard rates.

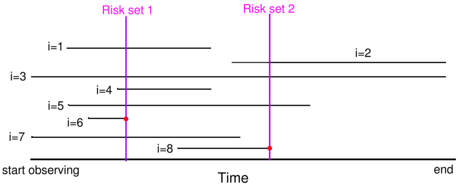
6. If the proportional hazards assumption does not hold, using this model will lead to biased inferences. An example violation for government durations: democratic governments build a coalitions of minorities and have increasing hazard rates, whereas autocratic governments eliminate opposition, consolidate support, and so have declining hazard rates.
7. The relationship between proportional hazards on the duration data and logistic regression on binary data:

6. If the proportional hazards assumption does not hold, using this model will lead to biased inferences. An example violation for government durations: democratic governments build a coalitions of minorities and have increasing hazard rates, whereas autocratic governments eliminate opposition, consolidate support, and so have declining hazard rates.
7. The relationship between proportional hazards on the duration data and logistic regression on binary data:
   (a) Start by discretizing the horizontal axis in this figure. The size of each bin must be small enough so that the hazard is essentially constant in the bin, but not small enough to run into zero probabilities (set of measure zero) problems.

6. If the proportional hazards assumption does not hold, using this model will lead to biased inferences. An example violation for government durations: democratic governments build a coalitions of minorities and have increasing hazard rates, whereas autocratic governments eliminate opposition, consolidate support, and so have declining hazard rates.
7. The relationship between proportional hazards on the duration data and logistic regression on binary data:
   (a) Start by discretizing the horizontal axis in this figure. The size of each bin must be small enough so that the hazard is essentially constant in the bin, but not small enough to run into zero probabilities (set of measure zero) problems.



6. If the proportional hazards assumption does not hold, using this model will lead to biased inferences. An example violation for government durations: democratic governments build a coalitions of minorities and have increasing hazard rates, whereas autocratic governments eliminate opposition, consolidate support, and so have declining hazard rates.
7. The relationship between proportional hazards on the duration data and logistic regression on binary data:
   (a) Start by discretizing the horizontal axis in this figure. The size of each bin must be small enough so that the hazard is essentially constant in the bin, but not small enough to run into zero probabilities (set of measure zero) problems.



   (b) Turn all intersections between a bin and a line (representing a unit) as a 1 if we're at the end point (representing an "event") or 0 otherwise. These 0/1 observations then are the data for the logit model.

(c) Using the rule for creating risks from rates, for bin of width Δ,

(c) Using the rule for creating risks from rates, for bin of width Δ,

$$\pi_0 \equiv \Pr(Y = 1|X_0)$$

(c) Using the rule for creating risks from rates, for bin of width Δ,

$$\begin{aligned}\pi_0 &\equiv \Pr(Y = 1|X_0) \\ &= 1 - \exp\left(-\int_t^{t+\Delta} \lambda_i(s)ds\right)\end{aligned}$$

(c) Using the rule for creating risks from rates, for bin of width Δ,

$$\begin{aligned}\pi_0 &\equiv \Pr(Y = 1|X_0) \\ &= 1 - \exp\left(-\int_t^{t+\Delta} \lambda_i(s) ds\right) \\ &= 1 - \exp\left(-\int_t^{t+\Delta} \lambda(s) e^{X_i\beta} ds\right)\end{aligned}$$

(c) Using the rule for creating risks from rates, for bin of width Δ,

$$\begin{aligned}
\pi_0 &\equiv \Pr(Y=1|X_0)\\
&= 1-\exp\left(-\int_t^{t+\Delta}\lambda_i(s)ds\right)\\
&= 1-\exp\left(-\int_t^{t+\Delta}\lambda(s)e^{X_i\beta}ds\right)\\
&= 1-\exp\left(-e^{X_i\beta}\int_t^{t+\Delta}\lambda(s)ds\right)
\end{aligned}$$

(c) Using the rule for creating risks from rates, for bin of width Δ,

$$\begin{aligned}
\pi_0 &\equiv \Pr(Y = 1|X_0) \\
&= 1 - \exp\left(-\int_t^{t+\Delta} \lambda_i(s)ds\right) \\
&= 1 - \exp\left(-\int_t^{t+\Delta} \lambda(s)e^{X_i\beta}ds\right) \\
&= 1 - \exp\left(-e^{X_i\beta}\int_t^{t+\Delta} \lambda(s)ds\right)
\end{aligned}$$

Letting the constant integral equal e^{β_0},

$$= 1 - \exp\left(-e^{X_i\beta}e^{\beta_0}\right)$$

(c) Using the rule for creating risks from rates, for bin of width Δ,

$$\begin{aligned}\pi_0 &\equiv \Pr(Y=1|X_0)\\ &= 1-\exp\left(-\int_t^{t+\Delta}\lambda_i(s)ds\right)\\ &= 1-\exp\left(-\int_t^{t+\Delta}\lambda(s)e^{X_i\beta}ds\right)\\ &= 1-\exp\left(-e^{X_i\beta}\int_t^{t+\Delta}\lambda(s)ds\right)\end{aligned}$$

Letting the constant integral equal e^{β_0},

$$\begin{aligned}&= 1-\exp\left(-e^{X_i\beta}e^{\beta_0}\right)\\ &= 1-\exp\left(-e^{X_i\beta+\beta_0}\right)\end{aligned}$$

and we can make a switch because

and we can make a switch because

$$e^{X_i\beta}/\ln[1+e^{X_i\beta}] \rightarrow 1$$

given rare events in small time interval $t, t+\Delta$,

and we can make a switch because

$$e^{X_i\beta}/\ln[1+e^{X_i\beta}] \rightarrow 1$$

given rare events in small time interval $t, t+\Delta$,

$$= 1 - \exp\left(-\ln[1+e^{X_i\beta+\beta_0}]\right)$$

and we can make a switch because

$$e^{X_i\beta}/\ln[1+e^{X_i\beta}] \rightarrow 1$$

given rare events in small time interval $t, t+\Delta$,

$$\begin{aligned} &= 1 - \exp\left(-\ln[1+e^{X_i\beta+\beta_0}]\right) \\ &= \frac{1}{1+e^{-\beta_0 - X_i\beta}}, \qquad \text{the logit model.} \end{aligned}$$

and we can make a switch because

$$e^{X_i\beta} / \ln[1 + e^{X_i\beta}] \to 1$$

given rare events in small time interval $t, t + \Delta$,

$$= 1 - \exp\left(-\ln[1 + e^{X_i\beta+\beta_0}]\right)$$

$$= \frac{1}{1 + e^{-\beta_0 - X_i\beta}}, \qquad \text{the logit model.}$$

(d) *Thus, the Cox proportional hazards model on durations is a logit model on the discretized binary data.*

and we can make a switch because

$$e^{X_i\beta}/\ln[1+e^{X_i\beta}] \to 1$$

given rare events in small time interval $t, t+\Delta$,

$$= 1 - \exp\left(-\ln[1+e^{X_i\beta+\beta_0}]\right)$$
$$= \frac{1}{1+e^{-\beta_0 - X_i\beta}}, \qquad \text{the logit model.}$$

(d) *Thus, the Cox proportional hazards model on durations is a logit model on the discretized binary data.*

8. Estimating β without the baseline (constant) term:

and we can make a switch because

$$e^{X_i\beta}/\ln[1+e^{X_i\beta}] \to 1$$

given rare events in small time interval $t, t+\Delta$,

$$\begin{aligned} &= 1 - \exp\left(-\ln[1+e^{X_i\beta+\beta_0}]\right) \\ &= \frac{1}{1+e^{-\beta_0-X_i\beta}}, \qquad \text{the logit model.} \end{aligned}$$

(d) *Thus, the Cox proportional hazards model on durations is a logit model on the discretized binary data.*

8. Estimating β without the baseline (constant) term:
   (a) The model assigns a hazard rate $\lambda_i(t)$ to each point on each line (i.e., for each i and t).

and we can make a switch because

$$e^{X_i\beta} / \ln[1 + e^{X_i\beta}] \to 1$$

given rare events in small time interval $t, t+\Delta$,

$$= 1 - \exp\left(-\ln[1 + e^{X_i\beta+\beta_0}]\right)$$
$$= \frac{1}{1 + e^{-\beta_0 - X_i\beta}}, \qquad \text{the logit model.}$$

(d) *Thus, the Cox proportional hazards model on durations is a logit model on the discretized binary data.*

8. Estimating β without the baseline (constant) term:
   (a) The model assigns a hazard rate $\lambda_i(t)$ to each point on each line (i.e., for each i and t).
   (b) Reconceptualize the data from durations into risk sets, one defined at each event occurrence. A subject may appear in more than one.

and we can make a switch because

$$e^{X_i\beta} / \ln[1 + e^{X_i\beta}] \to 1$$

given rare events in small time interval $t, t + \Delta$,

$$= 1 - \exp\left(-\ln[1 + e^{X_i\beta+\beta_0}]\right)$$
$$= \frac{1}{1 + e^{-\beta_0 - X_i\beta}}, \qquad \text{the logit model.}$$

(d) *Thus, the Cox proportional hazards model on durations is a logit model on the discretized binary data.*

8. Estimating β without the baseline (constant) term:
   (a) The model assigns a hazard rate $\lambda_i(t)$ to each point on each line (i.e., for each i and t).
   (b) Reconceptualize the data from durations into risk sets, one defined at each event occurrence. A subject may appear in more than one.
   (c) For each unit within a risk set, compute the probability of being an event, *conditional* on the total number of events in that risk set.

(d) To do this, we use almost the same logic as Chamberlain's clogit or multinomial logit (see notes Part 3). Suppose we have exactly 1 event in each risk set: $\sum_i Y_{it} = 1$, where $i = 1, \ldots, n_t$ indexes events in the risk set, and t indexes risk sets. This is typical, since events happen at discrete points in a continuum of time.

(d) To do this, we use almost the same logic as Chamberlain's clogit or multinomial logit (see notes Part 3). Suppose we have exactly 1 event in each risk set: $\sum_i Y_{it} = 1$, where $i = 1, \ldots, n_t$ indexes events in the risk set, and t indexes risk sets. This is typical, since events happen at discrete points in a continuum of time.

(e) Then calculate the conditional probability:

(d) To do this, we use almost the same logic as Chamberlain's clogit or multinomial logit (see notes Part 3). Suppose we have exactly 1 event in each risk set: $\sum_i Y_{it} = 1$, where $i = 1, \ldots, n_t$ indexes events in the risk set, and t indexes risk sets. This is typical, since events happen at discrete points in a continuum of time.

(e) Then calculate the conditional probability:

$$\Pr(Y_{it} = 1 | R_t) = \frac{\Pr(Y_{it} = 1, R_t)}{\Pr(R_t)}$$

(d) To do this, we use almost the same logic as Chamberlain's clogit or multinomial logit (see notes Part 3). Suppose we have exactly 1 event in each risk set: $\sum_i Y_{it} = 1$, where $i = 1, \ldots, n_t$ indexes events in the risk set, and t indexes risk sets. This is typical, since events happen at discrete points in a continuum of time.

(e) Then calculate the conditional probability:

$$\Pr(Y_{it} = 1|R_t) = \frac{\Pr(Y_{it} = 1, R_t)}{\Pr(R_t)}$$
$$= \frac{\pi_i \prod_{j \in R_t, j \neq i}(1 - \pi_j)}{\sum_{k \in R_t} \pi_k \prod_{j \neq k}(1 - \pi_j)}$$

(d) To do this, we use almost the same logic as Chamberlain's clogit or multinomial logit (see notes Part 3). Suppose we have exactly 1 event in each risk set: $\sum_i Y_{it} = 1$, where $i = 1, \ldots, n_t$ indexes events in the risk set, and t indexes risk sets. This is typical, since events happen at discrete points in a continuum of time.

(e) Then calculate the conditional probability:

$$\Pr(Y_{it} = 1|R_t) = \frac{\Pr(Y_{it} = 1, R_t)}{\Pr(R_t)}$$

$$= \frac{\pi_i \prod_{j \in R_t, j \neq i}(1 - \pi_j)}{\sum_{k \in R_t} \pi_k \prod_{j \neq k}(1 - \pi_j)}$$

Where $\pi_i = \Pr(y_i = 1)$. so the numerator is the probability that $Y_i = 1$, and all other $Y_j = 0$. The denominator is the probability that the total number of 1s in R_t is 1, which happens if Y_{kt} is 1 and all other Y_{jt} $(j \neq k)$ are 0, for any k in the risk set (hence the sum).

(d) To do this, we use almost the same logic as Chamberlain's clogit or multinomial logit (see notes Part 3). Suppose we have exactly 1 event in each risk set: $\sum_i Y_{it} = 1$, where $i = 1, \ldots, n_t$ indexes events in the risk set, and t indexes risk sets. This is typical, since events happen at discrete points in a continuum of time.

(e) Then calculate the conditional probability:

$$\begin{aligned}\Pr(Y_{it} = 1 | R_t) &= \frac{\Pr(Y_{it} = 1, R_t)}{\Pr(R_t)} \\ &= \frac{\pi_i \prod_{j \in R_t, j \neq i}(1 - \pi_j)}{\sum_{k \in R_t} \pi_k \prod_{j \neq k}(1 - \pi_j)}\end{aligned}$$

Where $\pi_i = \Pr(y_i = 1)$. so the numerator is the probability that $Y_i = 1$, and all other $Y_j = 0$. The denominator is the probability that the total number of 1s in R_t is 1, which happens if Y_{kt} is 1 and all other Y_{jt} $(j \neq k)$ are 0, for any k in the risk set (hence the sum).

If $\pi_i = [1 + e^{X_i \beta}]^{-1}$, then the above reduces to:

(d) To do this, we use almost the same logic as Chamberlain's clogit or multinomial logit (see notes Part 3). Suppose we have exactly 1 event in each risk set: $\sum_i Y_{it} = 1$, where $i = 1, \ldots, n_t$ indexes events in the risk set, and t indexes risk sets. This is typical, since events happen at discrete points in a continuum of time.

(e) Then calculate the conditional probability:

$$\Pr(Y_{it} = 1|R_t) = \frac{\Pr(Y_{it} = 1, R_t)}{\Pr(R_t)}$$

$$= \frac{\pi_i \prod_{j\in R_t, j\neq i}(1-\pi_j)}{\sum_{k\in R_t} \pi_k \prod_{j\neq k}(1-\pi_j)}$$

Where $\pi_i = \Pr(y_i = 1)$. so the numerator is the probability that $Y_i = 1$, and all other $Y_j = 0$. The denominator is the probability that the total number of 1s in R_t is 1, which happens if Y_{kt} is 1 and all other Y_{jt} ($j \neq k$) are 0, for any k in the risk set (hence the sum).

If $\pi_i = [1 + e^{X_i\beta}]^{-1}$, then the above reduces to:

$$\Pr(Y_{it} = 1|R_t) = \frac{e^{X_i\beta}}{\sum_{j\in R_t} e^{X_j\beta}}$$

(d) To do this, we use almost the same logic as Chamberlain's clogit or multinomial logit (see notes Part 3). Suppose we have exactly 1 event in each risk set: $\sum_i Y_{it} = 1$, where $i = 1, \ldots, n_t$ indexes events in the risk set, and t indexes risk sets. This is typical, since events happen at discrete points in a continuum of time.

(e) Then calculate the conditional probability:

$$\Pr(Y_{it} = 1|R_t) = \frac{\Pr(Y_{it} = 1, R_t)}{\Pr(R_t)}$$

$$= \frac{\pi_i \prod_{j \in R_t, j \neq i}(1 - \pi_j)}{\sum_{k \in R_t} \pi_k \prod_{j \neq k}(1 - \pi_j)}$$

Where $\pi_i = \Pr(y_i = 1)$. so the numerator is the probability that $Y_i = 1$, and all other $Y_j = 0$. The denominator is the probability that the total number of 1s in R_t is 1, which happens if Y_{kt} is 1 and all other Y_{jt} $(j \neq k)$ are 0, for any k in the risk set (hence the sum).

If $\pi_i = [1 + e^{X_i \beta}]^{-1}$, then the above reduces to:

$$\Pr(Y_{it} = 1|R_t) = \frac{e^{X_i \beta}}{\sum_{j \in R_t} e^{X_j \beta}}$$

$$= \frac{\lambda_i}{\sum_{j \in R_t} \lambda_j}$$

(f) The likelihood is then the product over risk sets $k = 1, \ldots, K$ (each k taking place at one time t):

(f) The likelihood is then the product over risk sets $k = 1, \ldots, K$ (each k taking place at one time t):

$$L(\beta|y) = \prod_{k=1}^{K} \Pr(Y_{it} = 1|R_t)$$

(f) The likelihood is then the product over risk sets $k = 1, \ldots, K$ (each k taking place at one time t):

$$L(\beta|y) = \prod_{k=1}^{K} \Pr(Y_{it} = 1|R_t)$$
$$= \prod_{k=1}^{K} \frac{e^{X_i\beta}}{\sum_{j\in R_k} e^{X_i\beta}}$$

(f) The likelihood is then the product over risk sets $k = 1, \ldots, K$ (each k taking place at one time t):

$$L(\beta|y) = \prod_{k=1}^{K} \Pr(Y_{it} = 1|R_t)$$
$$= \prod_{k=1}^{K} \frac{e^{X_i\beta}}{\sum_{j\in R_k} e^{X_i\beta}}$$

9. Unfortunately, no quantity of interest other than rr can be calculated without some estimate of the baseline hazard, $\lambda(t)$.

(f) The likelihood is then the product over risk sets $k = 1, \ldots, K$ (each k taking place at one time t):

$$L(\beta|y) = \prod_{k=1}^{K} \Pr(Y_{it} = 1|R_t)$$
$$= \prod_{k=1}^{K} \frac{e^{X_i\beta}}{\sum_{j\in R_k} e^{X_i\beta}}$$

9. Unfortunately, no quantity of interest other than rr can be calculated without some estimate of the baseline hazard, $\lambda(t)$.
10. Estimates of $\lambda(t)$ can come from models like the Weibull or exponential, or under Cox's model via a post-hoc analysis. To do the latter:

(f) The likelihood is then the product over risk sets $k = 1, \ldots, K$ (each k taking place at one time t):

$$L(\beta|y) = \prod_{k=1}^{K} \Pr(Y_{it} = 1|R_t)$$
$$= \prod_{k=1}^{K} \frac{e^{X_i\beta}}{\sum_{j \in R_k} e^{X_i\beta}}$$

9. Unfortunately, no quantity of interest other than rr can be calculated without some estimate of the baseline hazard, $\lambda(t)$.
10. Estimates of $\lambda(t)$ can come from models like the Weibull or exponential, or under Cox's model via a post-hoc analysis. To do the latter:
    (a) Information comes from the proportion of units within each risk set that are events: (1 event)/(number of non-events $+ 1 event$).

(f) The likelihood is then the product over risk sets $k = 1, \ldots, K$ (each k taking place at one time t):

$$L(\beta|y) = \prod_{k=1}^{K} \Pr(Y_{it} = 1|R_t) = \prod_{k=1}^{K} \frac{e^{X_i\beta}}{\sum_{j \in R_k} e^{X_i\beta}}$$

9. Unfortunately, no quantity of interest other than rr can be calculated without some estimate of the baseline hazard, $\lambda(t)$.
10. Estimates of $\lambda(t)$ can come from models like the Weibull or exponential, or under Cox's model via a post-hoc analysis. To do the latter:
    (a) Information comes from the proportion of units within each risk set that are events: (1 event)/(number of non-events $+ 1 event$).
    (b) For the baseline hazard estimate, let b be the MLE of β, and

(f) The likelihood is then the product over risk sets $k = 1, \ldots, K$ (each k taking place at one time t):

$$L(\beta|y) = \prod_{k=1}^{K} \Pr(Y_{it} = 1|R_t) = \prod_{k=1}^{K} \frac{e^{X_i\beta}}{\sum_{j\in R_k} e^{X_i\beta}}$$

9. Unfortunately, no quantity of interest other than rr can be calculated without some estimate of the baseline hazard, $\lambda(t)$.
10. Estimates of $\lambda(t)$ can come from models like the Weibull or exponential, or under Cox's model via a post-hoc analysis. To do the latter:
    (a) Information comes from the proportion of units within each risk set that are events: (1 event)/(number of non-events $+ 1 event$).
    (b) For the baseline hazard estimate, let b be the MLE of β, and

$$\lambda(t_j) = \frac{1}{\sum_{k\in R_j}(e^{X_k b})} = \frac{1}{\sum_{k\in R_j} e^{X_k b}},$$

(c) And with this we can estimate the rate,

(c) And with this we can estimate the rate,

$$\lambda_i(t_j) = \lambda(t_j)\lambda_i = \frac{e^{X_i b}}{\sum_{k \in R_j} e^{X_k b}},$$

(c) And with this we can estimate the rate,

$$\lambda_i(t_j) = \lambda(t_j)\lambda_i = \frac{e^{X_i b}}{\sum_{k \in R_j} e^{X_k b}},$$

(d) And from this we can get the risk by computing the cumulative rate,

(c) And with this we can estimate the rate,

$$\lambda_i(t_j) = \lambda(t_j)\lambda_i = \frac{e^{X_i b}}{\sum_{k \in R_j} e^{X_k b}},$$

(d) And from this we can get the risk by computing the cumulative rate,

$$H(T_i, X_i) = \sum_{t_j \in T_i} \lambda_i(t_j) = \sum_{t_j \in T_i} \frac{e^{X_i b}}{\sum_{k \in R_j} e^{X_k b}}$$

(c) And with this we can estimate the rate,

$$\lambda_i(t_j) = \lambda(t_j)\lambda_i = \frac{e^{X_i b}}{\sum_{k \in R_j} e^{X_k b}},$$

(d) And from this we can get the risk by computing the cumulative rate,

$$H(T_i, X_i) = \sum_{t_j \in T_i} \lambda_i(t_j) = \sum_{t_j \in T_i} \frac{e^{X_i b}}{\sum_{k \in R_j} e^{X_k b}}$$

and from that the risk:

(c) And with this we can estimate the rate,

$$\lambda_i(t_j) = \lambda(t_j)\lambda_i = \frac{e^{X_i b}}{\sum_{k \in R_j} e^{X_k b}},$$

(d) And from this we can get the risk by computing the cumulative rate,

$$H(T_i, X_i) = \sum_{t_j \in T_i} \lambda_i(t_j) = \sum_{t_j \in T_i} \frac{e^{X_i b}}{\sum_{k \in R_j} e^{X_k b}}$$

and from that the risk:

$$\pi_i = \Pr(Y = 1|X_i) = 1 - e^{-H(T_i, X_i)} = 1 - \exp\left(-\sum_{t_j \in T_i} \frac{e^{X_i b}}{\sum_{k \in R_j} e^{X_k b}}\right)$$

(c) And with this we can estimate the rate,

$$\lambda_i(t_j) = \lambda(t_j)\lambda_i = \frac{e^{X_i b}}{\sum_{k \in R_j} e^{X_k b}},$$

(d) And from this we can get the risk by computing the cumulative rate,

$$H(T_i, X_i) = \sum_{t_j \in T_i} \lambda_i(t_j) = \sum_{t_j \in T_i} \frac{e^{X_i b}}{\sum_{k \in R_j} e^{X_k b}}$$

and from that the risk:

$$\pi_i = \Pr(Y = 1|X_i) = 1 - e^{-H(T_i, X_i)} = 1 - \exp\left(-\sum_{t_j \in T_i} \frac{e^{X_i b}}{\sum_{k \in R_j} e^{X_k b}}\right)$$

(e) From the rate or the risk, we can then compute any other quantity of interest.

Density Case-Control Designs

Density Case-Control Designs

1. Combines the idea of case-control with Cox's proportional hazards model.

Density Case-Control Designs

1. Combines the idea of case-control with Cox's proportional hazards model.
2. Makes data collection easier when cases are occurring in real time and you need to find appropriate controls

Density Case-Control Designs

1. Combines the idea of case-control with Cox's proportional hazards model.
2. Makes data collection easier when cases are occurring in real time and you need to find appropriate controls
3. We can control for some confounders nonparametrically.

Density Case-Control Designs

1. Combines the idea of case-control with Cox's proportional hazards model.
2. Makes data collection easier when cases are occurring in real time and you need to find appropriate controls
3. We can control for some confounders nonparametrically.
4. Collect data as *sampled risk sets*: R_j $(j = 1, \ldots, M)$

Density Case-Control Designs

1. Combines the idea of case-control with Cox's proportional hazards model.
2. Makes data collection easier when cases are occurring in real time and you need to find appropriate controls
3. We can control for some confounders nonparametrically.
4. Collect data as *sampled risk sets*: R_j $(j = 1, \ldots, M)$
   (a) M is the total number of cases $(Y_{it} = 1)$ in the data

Density Case-Control Designs

1. Combines the idea of case-control with Cox's proportional hazards model.
2. Makes data collection easier when cases are occurring in real time and you need to find appropriate controls
3. We can control for some confounders nonparametrically.
4. Collect data as *sampled risk sets*: R_j $(j = 1, \ldots, M)$
   (a) M is the total number of cases $(Y_{it} = 1)$ in the data
   (b) A sampled risk set in this case includes one case matched with a small $(\approx 6 - 7)$ set of n_j controls $(Y_{1t}, \ldots, Y_{n_j t})$ randomly sampled from all those at risk

Density Case-Control Designs

1. Combines the idea of case-control with Cox's proportional hazards model.
2. Makes data collection easier when cases are occurring in real time and you need to find appropriate controls
3. We can control for some confounders nonparametrically.
4. Collect data as *sampled risk sets*: R_j $(j = 1, \ldots, M)$
   (a) M is the total number of cases ($Y_{it} = 1$) in the data
   (b) A sampled risk set in this case includes one case matched with a small ($\approx 6 - 7$) set of n_j controls ($Y_{1t}, \ldots, Y_{n_j t}$) randomly sampled from all those at risk
   (c) A subject may appear in multiple risk sets

Density Case-Control Designs

1. Combines the idea of case-control with Cox's proportional hazards model.
2. Makes data collection easier when cases are occurring in real time and you need to find appropriate controls
3. We can control for some confounders nonparametrically.
4. Collect data as *sampled risk sets*: R_j $(j = 1, \ldots, M)$
   (a) M is the total number of cases ($Y_{it} = 1$) in the data
   (b) A sampled risk set in this case includes one case matched with a small ($\approx 6 - 7$) set of n_j controls ($Y_{1t}, \ldots, Y_{n_j t}$) randomly sampled from all those at risk
   (c) A subject may appear in multiple risk sets
   (d) t is usually time, but can be any continuous variable or variables.

Density Case-Control Designs

1. Combines the idea of case-control with Cox's proportional hazards model.
2. Makes data collection easier when cases are occurring in real time and you need to find appropriate controls
3. We can control for some confounders nonparametrically.
4. Collect data as *sampled risk sets*: R_j $(j = 1, \ldots, M)$
   (a) M is the total number of cases $(Y_{it} = 1)$ in the data
   (b) A sampled risk set in this case includes one case matched with a small $(\approx 6 - 7)$ set of n_j controls $(Y_{1t}, \ldots, Y_{n_j t})$ randomly sampled from all those at risk
   (c) A subject may appear in multiple risk sets
   (d) t is usually time, but can be any continuous variable or variables.
5. Through matching, the procedure controls, without functional form assumptions, for any confounder related to t.

6. In classic case-control data, the missing information is the population fraction of ones. In density case-control, the information is the *risk set sampling fraction*: the fraction of each risk set included in each sampled risk set.

6. In classic case-control data, the missing information is the population fraction of ones. In density case-control, the information is the *risk set sampling fraction*: the fraction of each risk set included in each sampled risk set.
7. The statistical model is built in two stages.

6. In classic case-control data, the missing information is the population fraction of ones. In density case-control, the information is the *risk set sampling fraction*: the fraction of each risk set included in each sampled risk set.
7. The statistical model is built in two stages.
   (a) Estimate β by predicting which of the n_j sampled observations in the risk set is the case, using Cox's proportional hazard model.

6. In classic case-control data, the missing information is the population fraction of ones. In density case-control, the information is the *risk set sampling fraction*: the fraction of each risk set included in each sampled risk set.
7. The statistical model is built in two stages.
   (a) Estimate β by predicting which of the n_j sampled observations in the risk set is the case, using Cox's proportional hazard model.
   (b) Use the post-hoc method of estimating the constant term, but correct it with knowledge of the risk set sampling fraction at each risk set τ_j. To do this, estimate the baseline hazard as

6. In classic case-control data, the missing information is the population fraction of ones. In density case-control, the information is the *risk set sampling fraction*: the fraction of each risk set included in each sampled risk set.
7. The statistical model is built in two stages.
   (a) Estimate β by predicting which of the n_j sampled observations in the risk set is the case, using Cox's proportional hazard model.
   (b) Use the post-hoc method of estimating the constant term, but correct it with knowledge of the risk set sampling fraction at each risk set τ_j. To do this, estimate the baseline hazard as

$$\lambda(t_j) = \frac{1}{\sum_{k \in R_j} (e^{X_k b})(1/\tau_j)} = \frac{1}{\sum_{k \in R_j} e^{X_k b - \ln(\tau_j)}},$$

6. In classic case-control data, the missing information is the population fraction of ones. In density case-control, the information is the *risk set sampling fraction*: the fraction of each risk set included in each sampled risk set.
7. The statistical model is built in two stages.
   (a) Estimate β by predicting which of the n_j sampled observations in the risk set is the case, using Cox's proportional hazard model.
   (b) Use the post-hoc method of estimating the constant term, but correct it with knowledge of the risk set sampling fraction at each risk set τ_j. To do this, estimate the baseline hazard as

$$\lambda(t_j) = \frac{1}{\sum_{k \in R_j} (e^{X_k b})(1/\tau_j)} = \frac{1}{\sum_{k \in R_j} e^{X_k b - \ln(\tau_j)}},$$

   (c) If $\tau_j = 1$, this estimate is the same as under Cox's model.

8. If τ_j is not known, we can follow the same procedures as under classic case-control. The options:

8. If τ_j is not known, we can follow the same procedures as under classic case-control. The options:
   (a) assume knowledge of τ_j.

8. If τ_j is not known, we can follow the same procedures as under classic case-control. The options:
   (a) assume knowledge of τ_j.
   (b) assume τ_j could be anything and from that compute bounds on the quantities of interest (like Manski).

8. If τ_j is not known, we can follow the same procedures as under classic case-control. The options:
   (a) assume knowledge of τ_j.
   (b) assume τ_j could be anything and from that compute bounds on the quantities of interest (like Manski).
   (c) assume τ_j falls within a given range and apply Robust Bayesian methods.

8. If τ_j is not known, we can follow the same procedures as under classic case-control. The options:
   (a) assume knowledge of τ_j.
   (b) assume τ_j could be anything and from that compute bounds on the quantities of interest (like Manski).
   (c) assume τ_j falls within a given range and apply Robust Bayesian methods.
9. All methods are part of ReLogit Software, available at `http://GKing.Harvard.edu`.

8. If τ_j is not known, we can follow the same procedures as under classic case-control. The options:
   (a) assume knowledge of τ_j.
   (b) assume τ_j could be anything and from that compute bounds on the quantities of interest (like Manski).
   (c) assume τ_j falls within a given range and apply Robust Bayesian methods.
9. All methods are part of ReLogit Software, available at `http://GKing.Harvard.edu`.
10. When a risk set includes multiple cases, because of timing ties, the conditional probability expression is more complicated, but the approach remains the same.