A "Survivable" Introduction to Survival Analysis

Stephen D. Kachman
Department of Biometry
University of Nebraska–Lincoln

http://ianrwww.unl.edu/ianr/biometry/faculty/steve/skachman.html

Introduction

- Length of time between two events
  - Define:
    - Start point (birth, enter production, ...)
      - Time $t = 0$
    - End point (death, sale, illness, ...)

- Incomplete record
  - End point hasn't occurred yet
  - Animal is removed from the herd before the end

- Distribution is heavily skewed

Survival analysis

- Length of time an individual "survives"
- Packages
  - `SAS`: `Proc LIFEREG`
    - Fixed effects models
  - `Survival Kit`
    - Mixed models
- Require a basic understanding
- Breeder's test the limits of general packages

Objective

- Quick introduction to the analysis of survival data
  - Survival function
  - Hazard function as a function for building survival functions
  - Interpretation of risk factors under a Weibull model
  - Estimating equations compared to the mixed model equations
  - Censoring

Model

- T_i: Failure time of animal i
- Influenced by
- Risk Factor

$$\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u}$$

  - Risk factor for each animal
  - Increased risk leads to shorter survival times

- Fixed Effects $\boldsymbol{\beta}$
  - Breed, season of calving, heterosis
- Random Effects $\boldsymbol{u} \sim \mathrm{N}(0, \boldsymbol{G})$
  - Genetic merit $\boldsymbol{G} = \boldsymbol{A}\sigma_a^2$

Survival Function

$$S(t; \boldsymbol{\eta}_i) = \Pr(T_i \geq t)$$

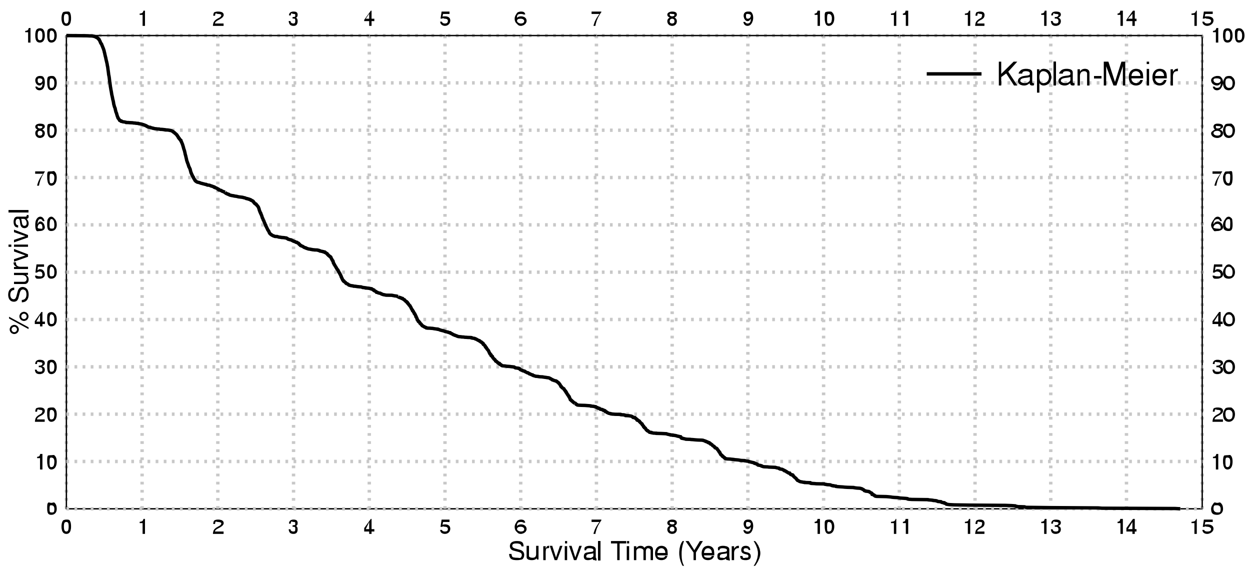
- Probability that an individual with a given risk factor η_i will survive till time t
- Length of time, implies that Survival is 100% at time 0
- Decreasing function
  – Nobody "unfails"
- Features
  – Shape of the survival function
  – Location "Stretching of the time scale"

Simmental Data Set

- 7,429 cows
  - Length of productive life
  - Censor $1 =$uncensored, $0 =$censored
  - $\sim$25% censored
  - Sire and maternal grandsire
  - Herd
  - Season
  - Percent Simmental

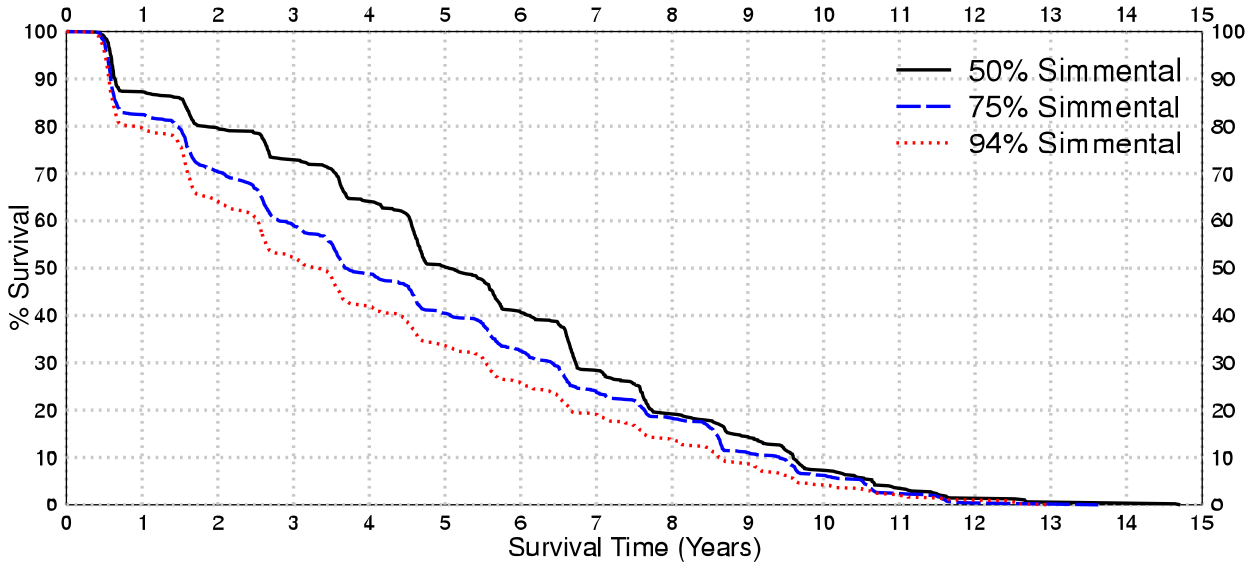
- Simple Model
  - Contemporary Group: Herd*Year*Season
  - Percent Simmental
  - Sire
  - 1,019 equations

Simmental survival function



Challenge: Find a reasonable model for the survival function

Survival Function: Percent Simmental



- Similar shape
  - Median survival time

50% 5.0 years or ∼150% of the 94% Simmental
75% 3.7 years or ∼110% of the 94% Simmental
94% 3.3 years

  - Between 10-20% of cows are culled each year
  - About 28% of the "50% Simmental" cows which enter their 4th year are culled in the next year of production

Hazard Function

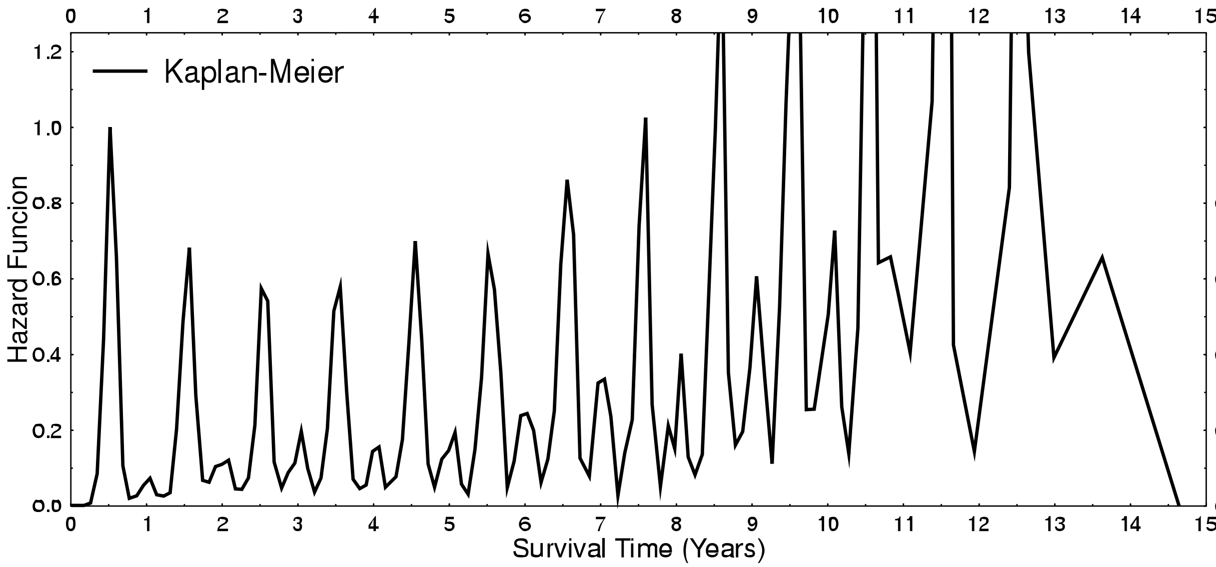
Short term risk of failure for animal alive at time t

$$\lambda(t; \boldsymbol{\eta}_i)$$

Over short periods of time, the probability that an animal fails is approximately equal to the hazard rate times the period of time.

- The higher the hazard rate the shorter the time period that this approximation is reasonable.
- Dramatic shifts will also make the approximate relationship poorer.

Simmental hazard function (monthly)



- Sharp peaks at "Weaning" with smaller peaks at "Calving"
- General rise as time increases
- Focus on short term effects
- Hazard rates are nonnegative
- Can be greater than one

Simmental hazard function (yearly)

- Smoothes over short term effects
- Long term effects are more evident

Hazard and Survival functions

- Given either the survival function, the density function, or the hazard function the other two can be found

$$
\begin{aligned}
S(t) &= e^{-\Lambda(t)} \\
f(t) &= \lambda(t)\ S(t) \\
\lambda(t) &= f(t)/S(t) \\
\Lambda(t) &= \int_0^t \lambda(w) dw
\end{aligned}
$$

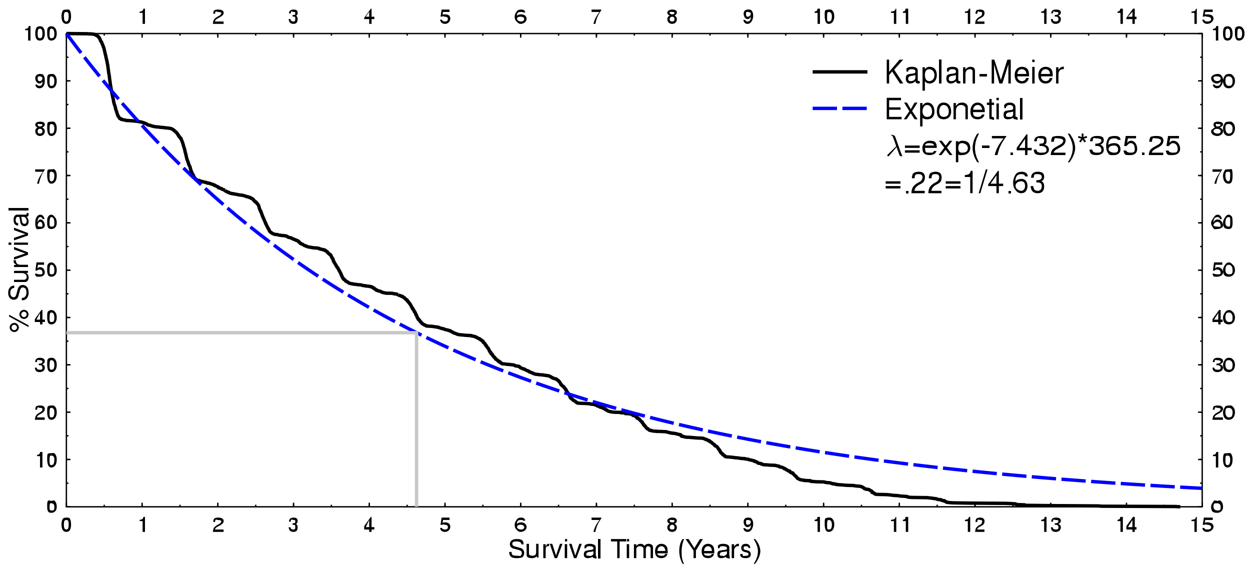
Exponetial model

Constant hazard $\lambda(t; \boldsymbol{\eta}_i) = \lambda$

$$S(t; \boldsymbol{\eta}_i) = e^{-\lambda\, t}$$

- Animal's chance of surviving an additional 5 years is the same when the animal
  - enters production
  - 5 years after entering production
  - 10 years after entering production

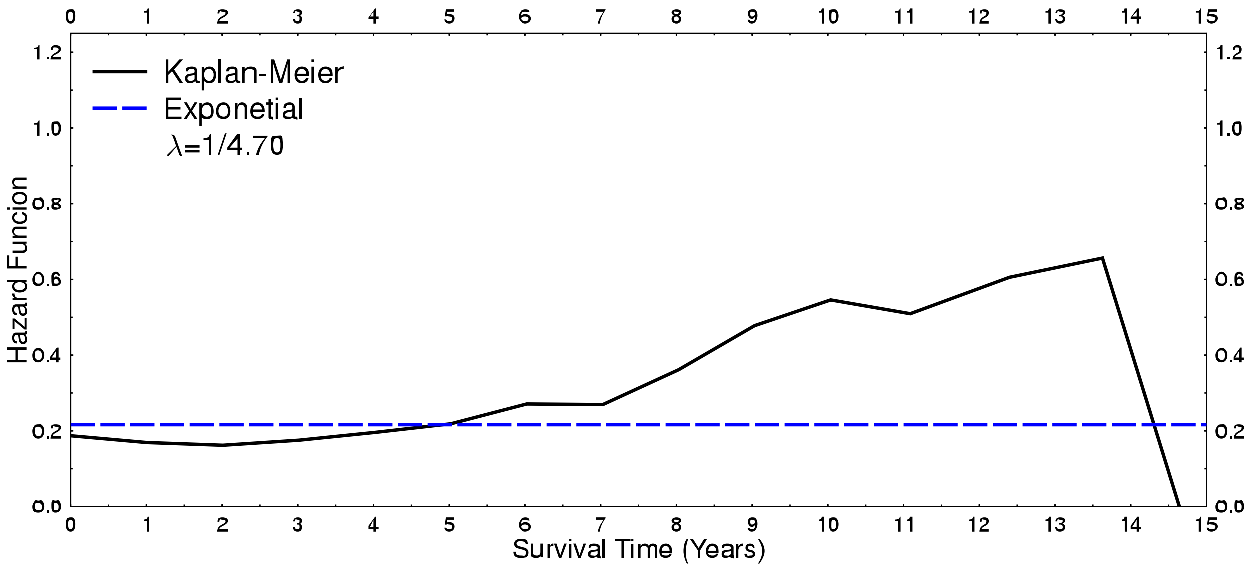
Simmental exponential survival function



- Underestimate the tail

- Approximately 2/3 of the cows are culled within $1/\lambda = 4.63$ years of entering production

Simmental exponential hazard function



- It is clear that the exponential hazard function underestimates the culling rate for old cows.

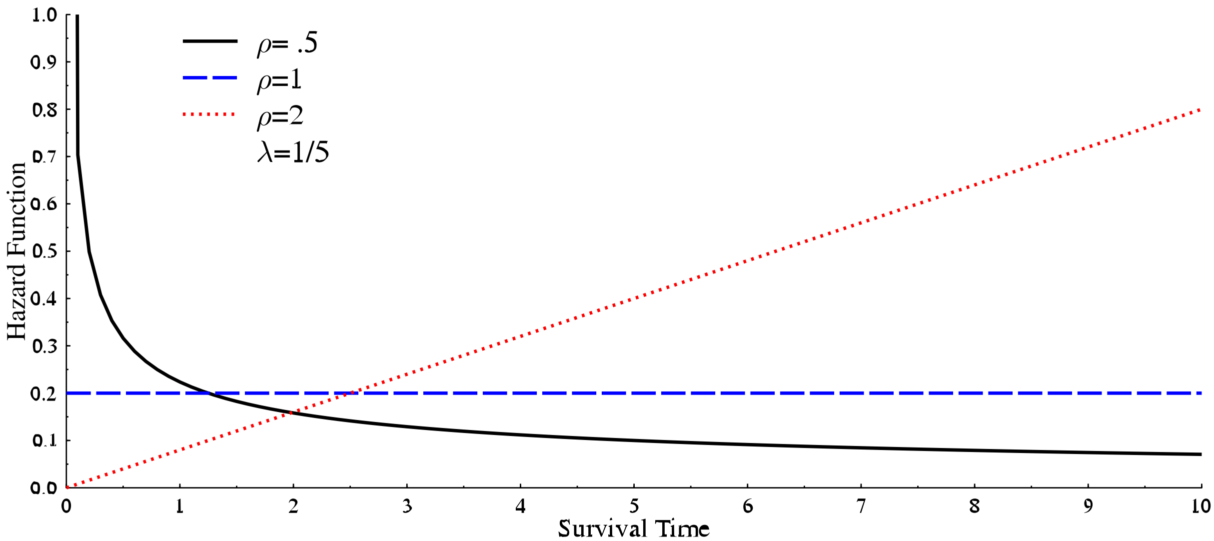
Weibull

- Popular generalization of the exponential model
- Hazard function is nonnegative
- Look at the log of the hazard function $\ln(\lambda(t, \boldsymbol{\eta}_i))$
- Linear in log time
- Parameterize so the Survival function looks nice

$$
\begin{aligned}
\ln(\lambda(t; \boldsymbol{\eta}_i)) &= [\ln(\rho) + (\rho)\ln(\lambda)] + (\rho - 1)\ln(t) \\
S(t; \boldsymbol{\eta}_i) &= e^{-(\lambda t)^{\rho}}
\end{aligned}
$$

- Watch out for time zero!

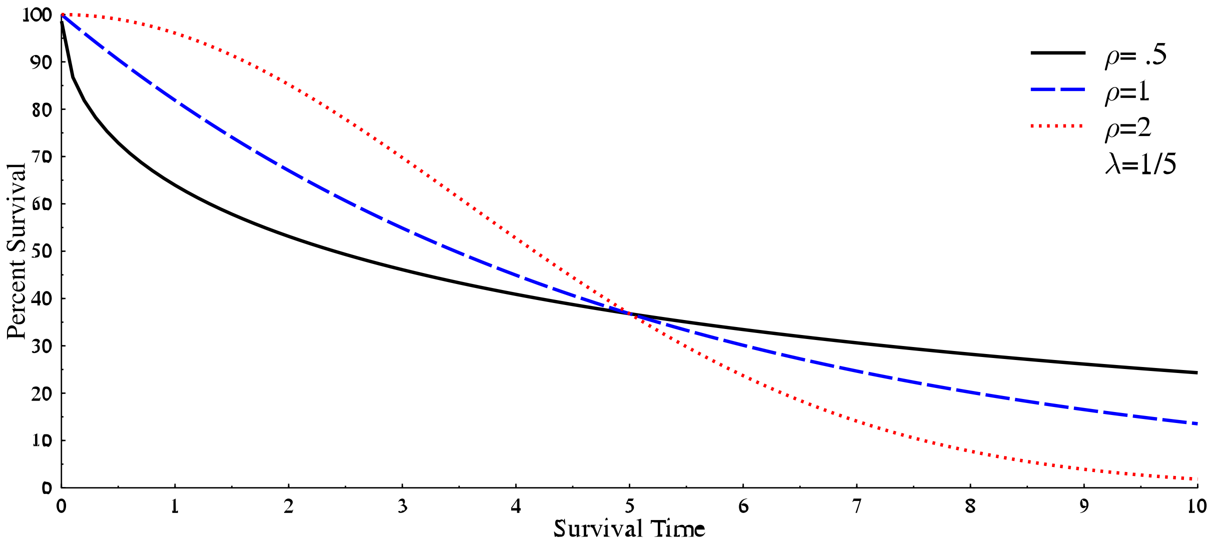
Weibull hazard function



- $\rho > 1$
  - Increasing Hazard function
  - $\lambda(0) \to 0$

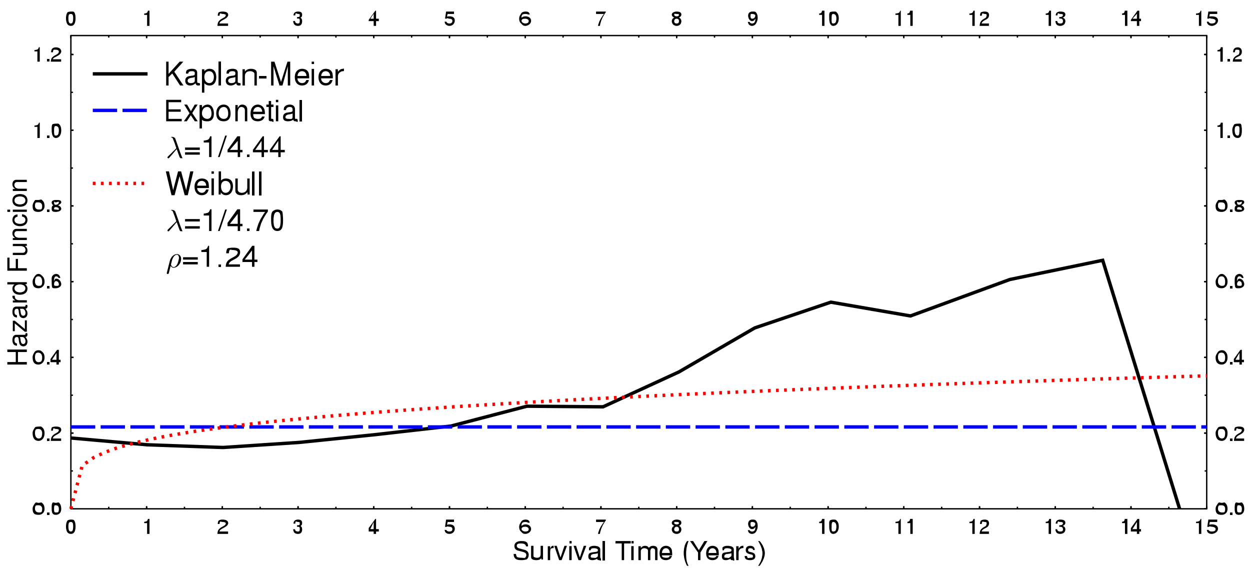
- $\rho < 1$
  - Decreasing Hazard function
  - $\lambda(0) \to \infty$

Weibull survival function



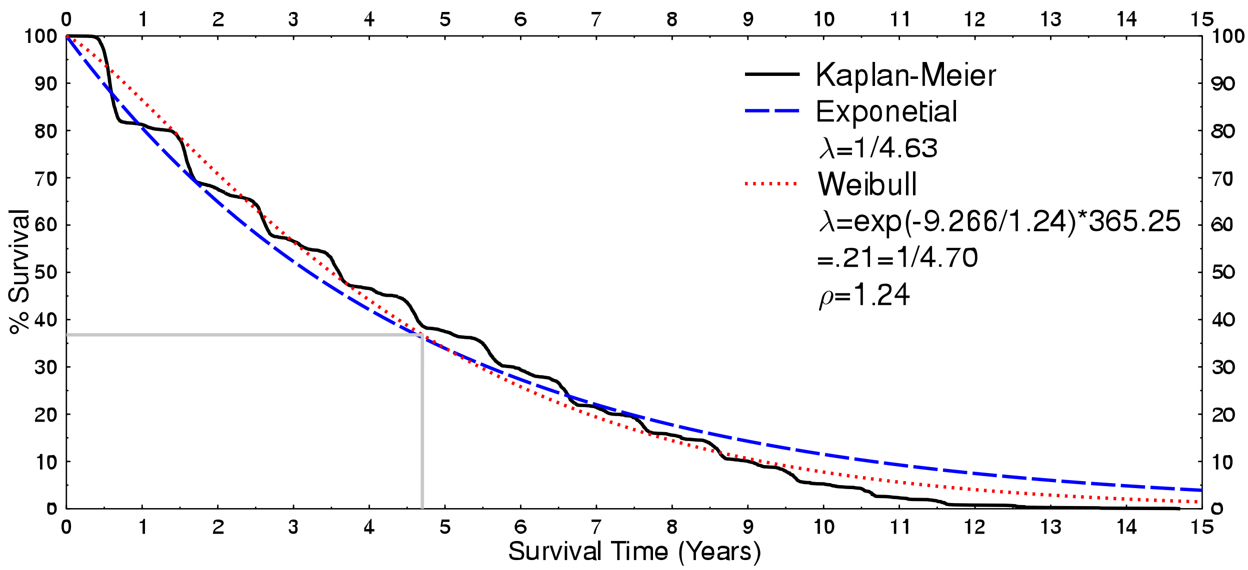
- – $\rho > 1$ starts out flat and speeds up
  – $\rho < 1$ starts out steep and flattens out
  – $1/\lambda$ is approximately equal to the time when 2/3 of the animals have been culled

Simmental Weibull hazard function



- Starts at 0
  - Failures at "Zero" will restrict how large the rate parameter can get.
- As a result it also has trouble picking up the tail.

Simmental Weibull survival function



- In the Weibull λ plays the role of an intercept and ρ is a rate parameter

- To emphasize this we can reparameterize as

$$S(t; \boldsymbol{\eta}_i) = e^{-\exp[\rho \ln(t) + \rho \ln(\lambda)]} = e^{-t^{\rho}\, e^{\eta}}$$

- where $\eta = \rho \ln(\lambda)$

Proportional Hazard Models

$$\lambda(t; \boldsymbol{\eta}_i) = \rho\, t^{\rho-1} e^{\eta}$$

- Baseline hazard $\lambda_0(t)$
- Scaling factor e^{η}

$$\lambda(t; \boldsymbol{\eta}_i) = \lambda_0(t) e^{\eta}$$
$$\Lambda(t; \boldsymbol{\eta}_i) = \Lambda_0(t) e^{\eta}$$

Risk Factor η

$$\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u}$$

- No residual
- $\mathrm{E}(\boldsymbol{y}|\boldsymbol{u}) \neq \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u}$
- Larger $\eta_i \Rightarrow$ shorter median survival times

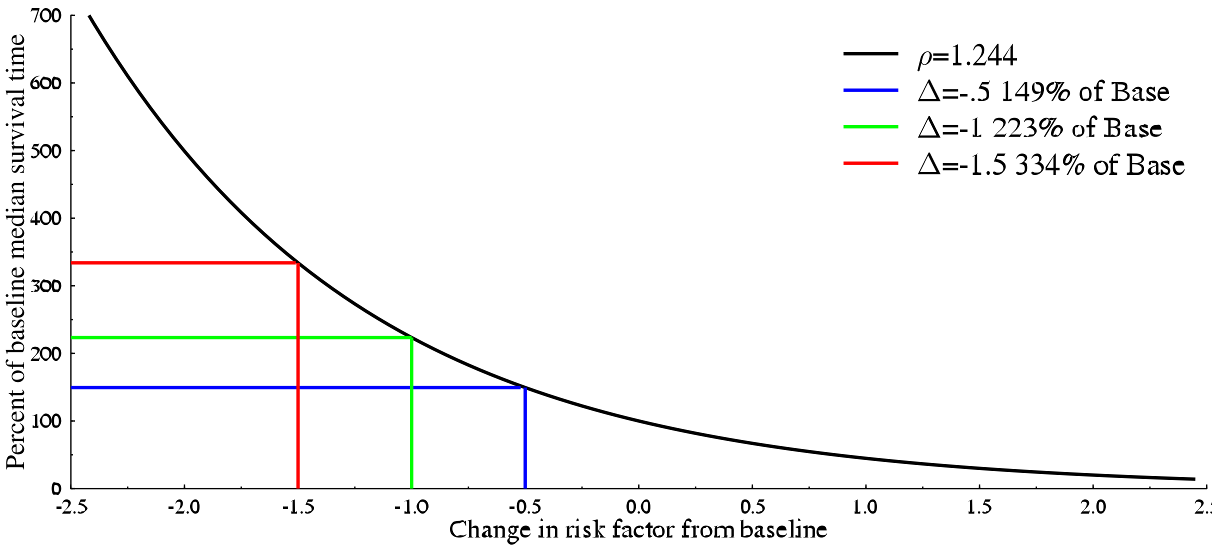
Effect of η

- Solve for median survival time, m_η, given risk factor, η

- $m_\eta = [-\ln(.5)]^{1/\rho} e^{-\eta/\rho}$

- $m_{\eta+\Delta} = m_\eta e^{-\Delta/\rho}$

- So if the risk factor for one group is Δ more than the risk factor for another group then their median survival time will be $e^{-\Delta/\rho}$ of the other groups median survival time.



- Suppose the rate parameter is 1.244 and the we wish to examine the effect of decreasing the risk factor by one.

- The median survival time would then be $e^{1/1.244} = 223\%$ of the original median survival time.

- Said another way, a decrease of one in the risk factor would result in the median survival time being increased by 123%

Percent Simmental

- Three levels: 50%, 75%, 94%

Contrast	Estimate	Median Survival Time % of 94%
50% vs 94%	$-.35$	$e^{.35/1.36} = 129\%$
75% vs 94%	$-.13$	$e^{.13/1.36} = 110\%$

Estimation

- Non-parametric
- Semi-parametric
- Parametric
  - Likelihood based procedures
- Based on the conditional log likelihood

$$\ell(\boldsymbol{\beta}, \boldsymbol{u}, \rho) = \sum_i \left[\ln(\lambda_0(y_i) + \eta_i - \Lambda_0(t_i)\exp(\eta_i)\right] - 1/2 \ln|\boldsymbol{G}| - 1/2 \boldsymbol{u}'\boldsymbol{G}^{-1}\boldsymbol{u}$$

- Depending on the complexity
  - Exact likelihood
  - Approximate likelihood

- Maximize

$$\begin{pmatrix} \boldsymbol{X}'\boldsymbol{R}\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{R}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{R}\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{R}\boldsymbol{Z} + \boldsymbol{G}^{-1} \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}'\boldsymbol{y}^* \\ \boldsymbol{Z}'\boldsymbol{y}^* \end{pmatrix}$$

where

$$\begin{aligned} \boldsymbol{R} &= -\frac{\partial^2 \ell}{\partial \boldsymbol{\eta} \partial \boldsymbol{\eta}'} \\ R_{ii} &= \Lambda_0(T_i) e^{\eta_i} \\ \boldsymbol{y}^* &= \frac{\partial \ell}{\partial \eta} + \boldsymbol{R}\boldsymbol{\eta} \\ y_i^* &= 1 - \Lambda_0(T_i) e_i^{\eta} + R_{ii}\eta_i. \end{aligned}$$

Joint Estimation

$$R_{ii} = \begin{pmatrix} \Lambda(T_i;\eta_i,\rho_i) & \Lambda(T_i;\eta_i,\rho_i)\ln(T_i) \\ \Lambda(T_i;\eta_i,\rho_i)\ln(T_i) & \Lambda(T_i;\eta_i,\rho_i)\ln(T_i)^2 + \frac{1}{\rho_i^2} \end{pmatrix}$$

$$y_i^* = \begin{pmatrix} 1 - \Lambda(T_i;\eta_i,\rho_i) \\ (1 - \Lambda(T_i;\eta_i,\rho_i))\ln(T_i) + \frac{1}{\rho_i} \end{pmatrix} + R_{ii}\begin{pmatrix} \eta_i \\ \rho_i \end{pmatrix}$$

- Linear predictors for both the risk factor and the rate parameter
- Typically, the linear predictor for the rate parameter only contains an intercept
- Multiple trait mixed model equations
- Replace both the residual covariance matrix and the dependent variable

Censoring

- Records on survival traits are often incomplete
- Animal survives past the point when data collection stopped
  - Right censored $T_i > C_i$
- Animal fails before data collection starts
  - Left censored $T_i < C_i$
- Animal fail at unknown point within an interval
  - Double censored $L_i < T_i < U_i$
- Assuming time of censoring and time of failure are independent
- Basic approach is to obtain the marginal likelihood by integration

- Focus on right censoring
  - Records with T_i
  - Records with $T_i > C_i$
  - $S(C_i; \eta_i) = \int_{C_i}^{\infty} f(w; \eta_i) dw$

The resulting contribution to the log likelihood for animal i

$$\ell_i = W_i \ln(\lambda(T_i; \eta_i)) - \Lambda(T_i; \eta_i)$$

where W_i equal zero if a record right censored and one if it is not censored.

- Weight matrix and recoded dependent variables will depend on the value of W_i.

- With right censoring some of the "one terms" drop out.

Summary

- Time till event traits
- Weibull is flexible enough to handle a wide variety of survival traits
- However, there are alternatives to the Weibull which may be more appropriate
- Modifications to existing programs are relatively *minor*
- Programs such as `SURVIVAL KIT` are available
- Other approaches are available