Martingale

A Martingale is a sequence of random variables $X_1, X_2, \ldots,$ such that

$$E[X_{n+1} \mid X_1, X_2, \ldots, X_n] = X_n \quad \forall n$$

The expected value of X_{n+1} is X_n given X_1 to X_n

A martingale can also exist in continuous time,

$$E[X_{t+s} \mid X_t, X_{[0,t]}] = X_t$$

A submartingale has the property that

$$E[X_{n+1} \mid X_1, \ldots, X_n] > X_n \quad \forall n$$

A supermartingale has the property that

$$E[X_{n+1} \mid X_1, \ldots, X_n] < X_n \quad \forall n$$

Time-to-Event Data

Time-to-event data is a random vector of variables $T_1, \ldots, T_n$ such that each r.v. i corresponds to the time it takes for person i to have the event of interest

Often, for some people, say person i, the event of interest is not observed during the observation period (the time for which person i is "watched" to see if they have an event). When this occurs, we say person i's outcome is censored.

Hazard Function

The Hazard Function represents the instantaneous likelihood (density) that a person has an event at time t given that they have survived up to time T

Let T_j be a random variable representing the survival time of individual j, T_j having pdf $f(t)$ and CDF $F(t)$

The Hazard Function $h(t)$ is defined as

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

$1 - F(t) = P(T \geq t)$, which is often called the Survival function $S(t) = 1 - F(t)$

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

By chain rule, $-\frac{d}{dt} \log(S(t)) = \frac{F(t)}{S(t)} = h(t)$

We call $H(t) = \int_0^t h(u)\,du$ the Cumulative Hazard Fn.

It can be shown that $S(t) = \exp(-H(t))$

Kaplan-Meier Curve

The Kaplan-Meier "curve" is a non-parametric estimator of the survival function that accomodates censored data. It is NPMLE. Whenever an event occurs among one of the iid time to events $T_1, \dots, T_n$ in the sample, say an event occurs at time a, the survival curve shrinks by a factor of the proportion of people removed by the event. When someone is censored, that person is removed without changing the KM curve.

Example: 6, 6, 6*, 7, 8*, 9, 10, 12*

time	$\frac{d}{t}$	$\frac{m}{t}$	$(1-\frac{d}{t})$	$(1-\frac{m}{t})$	$\prod(1-\frac{d}{t})(1-\frac{m}{t})$	KMC $\prod(1-\frac{d}{t})$
6	$1/8$	0	$7/8$	1	$7/8$	$7/8$
6	$1/7$	0	$6/7$	1	$6/8$	$6/8$
6*	0	$1/6$	1	$5/6$	$5/8$	$6/8$
7	$1/5$	0	$4/5$	1	$4/8$	0.6
8*	0	$1/4$	1	$3/4$	$3/8$	0.6
9	$1/3$	0	$2/3$	1	$2/8$	0.4
10	$1/2$	0	$1/2$	1	$1/8$	0.2

Sand Wich Variance

The "Sand Wich" variance of a parameter estimate $\hat{\beta}$ obtained with maximum likelihood estimation is an estimate of the variance-covariance matrix of $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_j)$ of the following form:

$$\widehat{Var}(\hat{\beta}) = \left(H(\hat{\beta})\right)^{-1} \left(\sum_{i=1}^{n} \frac{\partial \ell_i(\beta)}{\partial \beta} \left(\frac{\partial \ell_i(\beta)}{\partial \beta} \right)^T \Big|_{\beta=\hat{\beta}} \right) \left(H(\hat{\beta})\right)^{-1}$$

Bread, Hessian

Meat ($p \times p$, each $j \times k$ is $\sum_{i=1}^{n} \frac{\partial \ell_i(\beta_j)}{\partial \beta_j} \frac{\partial \ell_i(\beta_k)}{\partial \beta_k}$)

Bread, Hessian

Where $H(\hat{\beta}) = \frac{-\partial^2 \ell(\beta)}{\partial \beta \, \partial \beta^T} \Big|_{\beta=\hat{\beta}}$, also $p \times p$,

each $j \times k$ is $\frac{-\partial^2 \ell(\beta)}{\partial \beta_j \, \partial \beta_k} \Big|_{\beta=\hat{\beta}}$

Discrete Hazard Function

If the time to event is distributed discretely, with mass at $t = a_j$ is $f_j = P(T = a_j)$ $j = 1, \ldots$, the <u>discrete hazard func.</u> is

$$P(T = a_j \mid T \geq a_j) = \frac{f_j}{f_j + f_{j+1} + f_{j+2} + \ldots}$$

So $f_i = h_i \prod_{j=1}^{i-1} (1 - h_j)$ and $S(t) = \prod_{j;\, a_j \leq t} (1 - h_j)$

(much like Kaplan–Meier curve)

The <u>cumulative hazard</u> is $H(a_j) = \sum_{i=1}^{j} h_j$

Stieljes notation gives general F and H:

$$F(t) = \int_0^u dF(u) \qquad H(t) = \int_0^t \frac{dF(u)}{1 - F(u^-)}$$

$F(u^-)$ is $\lim_{\varepsilon \to 0+} F(u - \varepsilon)$, left hand limit of F at u

Exponential & Weibull

If $h(t)$ is constant, $h(t)=\lambda$, then the event times must have <u>exponential distribution</u> with parameter λ. Some info about exponential:

$f(t)=\lambda e^{-\lambda t}$ "memoryless" $\leftrightarrow h(t)=\lambda$

$S(t)=e^{-\lambda t}$ $H(t)=-\log(S(t))=\lambda t$

A more general distribution that is exponential when $K=1$ is the <u>weibull distribution</u>, K, λ

$F(t)=K\lambda(\lambda t)^{K-1}e^{-(\lambda t)^K}$ $h(t)=K\lambda(\lambda t)^{K-1}$

$S(t)=e^{-(\lambda t)^K}$ $H(t)=(\lambda t)^K$ $(\lambda>0, K>0)$

The plot of $\log(-\log(S(t)))$ vs $\log(t)$ is linear with slope K and intercept $K\log(\lambda)$ $(\lambda T)^K$ is Unit exponential. The hazard function is monotone increasing if $K>1$, constant if $K=1$, and monotone decreasing if $K<1$

Extreme Value Distribution

If Y is Weibull, $Y = \log(T)$ has an "extreme value distribution"

Write $y = \log(t)$, $w = -\log(\lambda)$, and $\sigma = \frac{1}{k}$, then

$$(\lambda t)^k = \left(e^{y-w}\right)^k \approx \exp\left(\frac{y-w}{\sigma}\right)$$ thus

$$Y = \frac{\log T - w}{\sigma}$$ has the standard extreme value distribution, with survival function

$$P(Y > x) = e^{-e^x}$$ and density $e^{(x - e^x)}$

So $\log T$ is in a <u>location-scale family</u>

Integrated Hazard Transform

Let T be a survival time with arbitrary continuous survivor function $S_T(t)$ and cum hazard function $H(t)$. $H(T)$ is the <u>integrated hazard transform</u> of T. The survivor function of $U = H(T)$ is $S_U(u) = P(U > u) = P(H(T) > u)$ $= P(T > H^{-1}(u))$, assuming H has an inverse. This is the probability that T is larger than T's own hazard function at u. Now, since T has survivor function S, we have $= S_T(H^{-1}(u))$. Now, recall that $S(t) = \exp(-H(t))$, so $S_U(u) = S_T(H^{-1}(u)) = \exp(-H(H^{-1}(u))) = e^{-u}$

Thus $H_T(T)$ always has exponential distribution with $\lambda = 1$. This transform is similar to the <u>probability integral transform</u> that says that $F_X(X)$ must be uniform $(0,1)$ (provided X is continuous)

Censoring

Censoring occurs when for some units the outcome of interest is not observed, but the unit is known not to have had an event in the time interval $(0, c)$ where c is called the censoring time.

If for unit i, T_i is the event time and C_i is the censoring time, let X_i be

$$X_i \equiv \min(T_i, C_i)$$

$$\delta = \begin{cases} 1 & \text{if } T_i \leq C_i \\ 0 & \text{ow} \end{cases}$$

C_i is viewed as constant, not random, and thus independent of T_i, "non-informative" censoring

Type I censoring: All C_i are equal

Type II censoring: Censor all remaining subj after d events

Restricted Mean Survival & Residual Life

The restricted mean survival, which gives the mean survival of all individuals up to time ℓ, is defined as

$$RMS(\ell) = \int_0^{\ell} S(t)\,dt$$

The conditional distribution of the remaining life U of an individual known to have survived to t has density

$$f_{U|t}(u|t) = \frac{f(u+t)}{S(t)}, \quad u \geq 0$$

The expected value of U as a function of t is called the <u>mean residual life function</u>, and is given by

$$r(t) = E[T - t \mid T \geq t] = \frac{\int_{u=t}^{\infty} (u-t) f(u)\,du}{S(t)}$$

Note that $r(0) = E[T]$, and expectation of total life $E[T \mid T \geq t]$ of T given the person survived to t is just $r(t) + t$

Pyke's formula: $Var(T) = E[r_T(T)]$

Pareto & Log-Logistic

Pareto:

$$S(t) = \left(\frac{K}{K+Pt}\right)^K \qquad h(t) = \frac{KP}{K+Pt}$$

if $K > 1$:

$$E[T] = \frac{K}{P(K-1)}$$

Provides a model for heterogeneous pop. where each lifetime is exp. dist. w/ diff λ

Log-logistic:

The standard logistic density over $-\infty < Y < \infty$ is

$$f(Y) = \frac{e^Y}{(1+e^Y)^2}$$ and survivor function $\frac{1}{1+e^Y}$

As a location scale family: $f(Y|\sigma,\mu) = \frac{\frac{1}{\sigma} e^{(Y-\mu)/\sigma}}{(1+e^{(Y-\mu)/\sigma})^2}$

When $P = e^{-\mu}$ and $K = \frac{1}{\sigma}$, exp(logistic) has

$$S(t) = \frac{1}{1+(Pt)^K} \qquad f(t) = \frac{KP(Pt)^{K-1}}{(1+(Pt)^K)^2} \qquad h(t) = \frac{KP(Pt)^{K-1}}{1+(Pt)^K}$$

Standard log-logistic is: $f(t) = \frac{1}{(1+t)^2}$, $K=1$, $P=1$

Scale & Proportional Hazard Families

A <u>Scale</u> family is a distribution that, when multiplies by a constant, the random variable is still in that distributional family. A scale family with scale parameter β satisfies

$S(t|\beta) = S_0(e^{\beta}t)$ where S_0 is the survivor function of the "standard" version where the scale is 1

Likewise, the hazard satisfies $h(t,\beta) = e^{\beta} h_0(e^{\beta}t)$

A <u>Proportional hazards</u> family with "standard" hazard h_0 has parameter β such that all other hazards in family have

$h(t) = e^{\beta} h_0(t)$ implying $S(t) = S_0(t)^{e^{\beta}}$

Scale and Prop-hazards families based on S_0 are identical iff S_0 is the survivor function of a Weibull

Nelson Aalen Plot

Simple graphical procedure for estimating the integrated hazard function $H(t)$ from a sample of iid variables subject to non-informative censoring

Analogous to Kaplan-Meier estimator of $S(t)$

Data are in form (x_i, δ_i), $x_i = \min(t_i, c_i)$, $\delta_i = 1_{[t_i \leq c_i]}$

Put data in order $x_{(1)} < \dots < x_{(m)}$ are observed, $d_{(i)} = \#\{i : t_i = t_{(i)}, \delta_i = 1\}$ is the number of observed failures at $x_{(i)}$, let $r(t) = \#\{i : t_i \geq t\}$ be the size of the risk set at time t and $r_{(i)} = r(t_{(i)})$

The Nelson-Aalen estimator $\hat{H}(t)$ of $H(t)$ is

$$\hat{H}(t) = \sum_{j : t_{(j)} \leq t} \frac{d_{(j)}}{r_{(j)}}$$

Step function with jumps at the ordered failure times $t_{(i)}$. It can be shown that $\hat{H}(t)$ is consistent for $H(t)$.