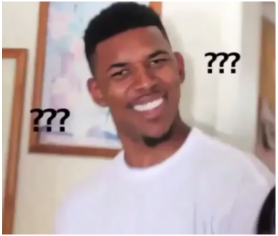
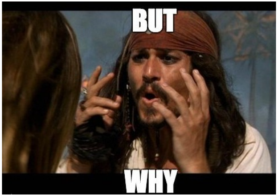
Modelagem Estatística da incidência de cânceres via um GLMM Multinomial

Henrique Laureano
http://leg.ufpr.br/~henrique

CiDWeek I, 03-07/02/2020



O quê?



Por quê?

Como?

Some continuous parametric failure time models

Random variable: failure time, $T > 0$.
Type: continuous.

Common failure time distributions for homogeneous populations:

Now, let's better understand how this works.

Generalized F

Advantage: it can adapt to a wide variety of distributional shapes.

Context

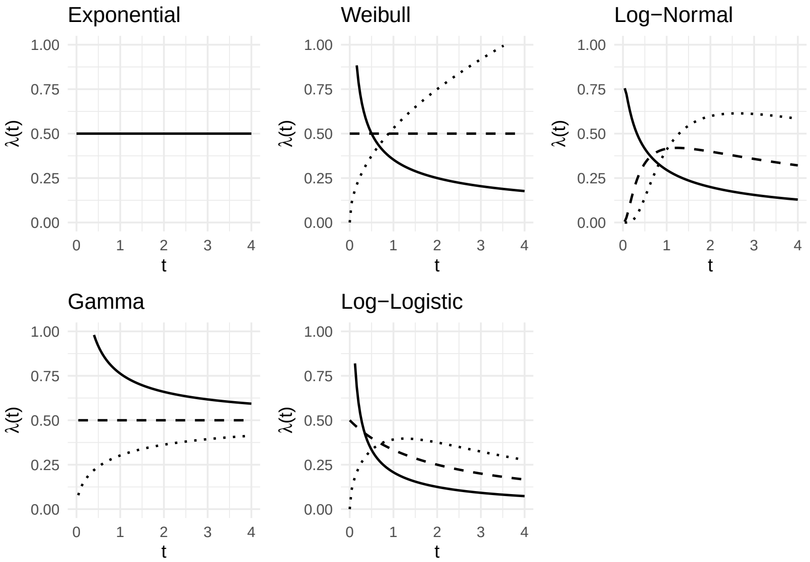
A location and scale model for $Y = \log T$ in which the error distribution is assumed to be that of the logarithm of an F variate on $2m_1$ and $2m_2$ degrees of freedom.

That is, $Y = \mu + \sigma W$, where

$$f_W(w) = \frac{(m_1/m_2)^{m_1} e^{wm_1} (1 + m_1 e^w/m_2)^{-(m_1+m_2)}}{B(m_1, m_2)}.$$

The resulting model for T is the generalized F distribution.

Shapes of the hazard functions



A door to another world

To be able to see these Generalized F special cases,
the transformation $Y = \mu + \sigma W$ is necessary.

However, this open a door for another world: Extreme Value Theory.

In the Generalized Gamma case and special cases,
W is an extreme value (minimum) distribution.

Extreme Value Theory
- ↳ Generalized extreme value (GEV) distribution
  - ↳ Type I extreme value distribution: Gumbel family
  - ↳ Type II extreme value distribution: Fréchet family
  - ↳ Type III extreme value distribution: Weibull family

Regression models
↳ Exponential and Weibull

Goal: obtain a regression model by allowing the failure rate to be a function of the derived covariates Z.

The hazard at time t for an individual can be written as

$$\lambda(t; x) = \text{hazard} \times c(Z^\top \beta),$$

three forms have been used for c:

- $c(s) = 1 + s$, corresponding to the failure rate;
- $c(s) = (1 + s)^{-1}$, corresponding to the mean survival time;
- $c(s) = \exp(s)$.

Exponential regression model

$$\lambda(t;x) = \lambda \exp(Z^\top \beta)$$
$$Y = -\log \lambda - Z^\top \beta + W$$
$$W \sim \text{Extreme Value dist.}$$

Weibull regression model

$$\lambda(t;x) = \gamma(\lambda t)^{\gamma-1} \exp(Z^\top \beta)$$
$$Y = -\log \lambda - Z^\top \sigma \beta + \gamma^{-1} W$$
$$W \sim \text{Extreme Value dist.}$$

Accelerated failure time models
- ↳ general class of log-linear models
  - ↳ covariates act additively on Y, or multiplication on T
    - ↳ log survival time, $Y = \log T$

More general model: Relative Risk or Cox Model.

Relative risk model

Cox, 1972

$$\lambda(t;x) = \lambda_0(t)\exp(Z^\top\beta),$$

where $\lambda_0(\cdot)$ is an arbitrary unspecified baseline hazard function for continuous T.

The conditional survivor function for T given Z is

$$F(t;x) = F_0^{\exp(Z^\top\beta)}(t), \quad \text{where} \quad F_0(t) = \exp\left[-\int_0^t \lambda_0(u)\mathrm{d}u\right].$$

Thus the survivor function of t for a covariate value, x, is obtained by raising the baseline survivor function $F_0(t)$ to a power.

Nice generalizations,

- stratified Cox model;
- time-dependent Cox model: *relative* risk model.

Accelerated failure time model

Suppose $Y = \log T$ and consider the linear model

$$Y = Z^\top \beta + W.$$

Exponentiation gives $T = \exp(Z^\top \beta)\ S$, where $S = \exp(W) > 0$ has hazard function $\lambda_0(s)$, say, that is independent of β.

The hazard function for T can be written as

$$\lambda(t; x) = \exp(-Z^\top \beta)\lambda_0[t \exp(-Z^\top \beta)].$$

The effect of the covariate is multiplicative on t rather than on the hazard function.

i.e.,

The role of Z is to accelerate (or decelerate) the time to failure.

Comparison of regression models

note

Exponential and Weibull regression models can be considered as special cases of both models.

Discrete failure time models

Discrete failure time?

- Grouping of continuous data due to imprecise measurement;
- Time itself may be discrete
  - e.g., when the response time represents the number of episodes that occur prior to a terminal event.

Discrete regression models?

- Grouped relative risk model;
- Discrete and continuous relative risk model;
- Discrete logistic model.

Discrete regression models

» Grouped relative risk model:

$$\text{Discrete baseline cumulative hazard function : } \Lambda_0(t) = \sum_{a_i \leq t} \lambda_i,$$

this model is the uniquely appropriate one for grouped data from the continuous relative risk model.

» Discrete and continuous relative risk model:

$$\mathrm{d}\Lambda(t; x) = \exp(Z^\top \beta)\, \mathrm{d}\Lambda_0(t),$$

which retains the multiplicative hazard relationship.

» Discrete logistic model:

$$\frac{\mathrm{d}\Lambda(t; x)}{1 - \mathrm{d}\Lambda(t; x)} = \frac{\mathrm{d}\Lambda_0(t)}{1 - \mathrm{d}\Lambda_0(t)} \exp(Z^\top \beta),$$

specifies a linear log odds model for the hazard probability at each potential failure time.


THANK YOU
memegenerator.net