Survival models in Stan

D.W. Bester

1 Model descriptions

Model 1

This is a survival model, where events happen according to the the hazard rate:

$$\begin{aligned} h(t) &= Be^{\theta t} \\ H(t) &= \int_0^t h(x)dx \\ &= \frac{B}{\theta}\left(e^{\theta t} - 1\right) \end{aligned}$$

The events are simulated using $H^{-1}(t)$, and then subjected to random (uninformative) censoring.

Model 2

This is the same as model 1, but we have multiple gompertz parameters. We try to use a `group` variable to assign the right gompertz parameters within Stan. Thus, this model fits the hazard rate

$$h(t) = B_j e^{\theta_j t} \tag{1}$$

where j is the group number of the subject in question. It is a covariate in our model.

Model 3

This is the same as model 2, but we add a "linear predictor":

$$\eta = \beta x \tag{2}$$

where x is a covariate. The hazard rate then becomes

$$\begin{aligned} h(t) &= B_j e^{\theta_j t} e^{\eta} \\ &= B_j e^{\theta_j t} e^{\beta x} \end{aligned}$$

where j is the group number of the subject in question. This is similar to the cox proportional hazards model. Notice that the linear predictor does not contain an intercept term. This is since the gompertz parameter B_j already acts as an intercept for that group. The β parameter measures the influence of of covariate x measured against the baseline. That is, we can rewrite the hazard rate as:

$$h(t) = e^{\theta_j t + \beta x + log(B_j)}$$

Thus, $log\,(B_j)$ is the intercept for group j, θ_j measures the influence of time on the hazard rate for group j, and β measures the influence of covariate x, assuming the influence is global over all groups.

2 Likelihood, estimation, and choice of prior parameters

All of our models assume that survival times come from a process with hazard rate $h(t)$, and we can use a Poisson process to show the likelihood contribution of each observation. Let T be the event-time of a subject in our model, then

$$T > t \quad \longleftrightarrow \quad N(t) < 1$$

where $N(t)$ is the number of events at time t, which will always be 0 or 1 in our model, then we have

$$N(t) \sim \text{Poisson}\left(\int_0^t h(x)dx \right).$$

We don't observe T_i directly, since our observations are subject to censoring. We observe $V_i = \min(T_i, C_i)$ where C_i is a censoring time, assumed to be random and independent of T (non-informative censoring). Our observations are the pairs (v_i, δ_i) where δ_i is an indicator taking 1 if $T_i \leq C_i$ and 0 otherwise. Thus, δ_i is an event indicator, and it will be equal to the number of events at time t.

The Poisson process (and some algebra) provides the the likelihood contribution of the observation pair (v_i, δ_i):

$$\begin{aligned}
\mathcal{L}\left(\theta|(v_i, \delta_i)\right) &= S(v_i)^{\delta_i} f(v_i)^{(1-\delta_i)} \\
&= h(v_i)^{\delta_i} \exp\left(-\int_0^{v_i} h(x)dx\right) \\
&= h(v_i)^{\delta_i} \exp\left(-H(v_i)\right)
\end{aligned}$$

where θ is the set of all parameters. This is the likelihood we need to add to MCMC software if we want to fit survival models.[1] The point of this repository is to show how this can done using Stan's `functions` block.

Stan was able to recover the chosen parameters for all of the above models. For model 3, the more complicated of the three models in terms of number of parameters, convergence was very sensitive to the choice of priors assumed on the Gompertz parameters, $\boldsymbol{B}$ and $\boldsymbol{\theta}$.

Add discussion here about the nuance of the prior parameters, how it effected the convergence of model 2 vs model 3, and how to avoid this problem in general by thinking carefully about prior parameters.

[1]MCMC software like Stan and JAGS do provide Poisson densities, but we can only use them if the hazard rate is constant. For more complicated hazard rate functions, we have to resort to more tedious measures. Since the likelihood contains both $h(t)$ and $H(t)$, we need to add these time-varying functions to the sampler.