Gompertz Cure Rate Survival Models with Stan and brms

Todd Yoder, Andrew Tredennick, Timothy Waterhouse
Metrum Research Group, Boston, MA, USA

METRUM RESEARCH GROUP

20 YEARS Anniversary

Abstract

Objectives: In drug safety data it is common for a proportion of the population to never experience a specified adverse event. Standard parametric survival models assume the survival curve approaches zero and fail to characterize this aspect of safety data. A Gompertz distribution with a negative scale parameter allows the survival curve to have a non-zero asymptote and may better characterize survival data where a proportion of the population never experiences the event of interest [1]. The brms package in R enables Bayesian modeling with Stan [2, 3]. The Gompertz distribution is not natively supported in brms or Stan, and a custom probability distribution was implemented in the software to enable cure rate model fitting and inference.
Methods: The following Gompertz distribution functions were defined as custom Stan functions: log probability density, log cumulative distribution, log complementary cumulative distribution, and random generation function. The Stan functions and the log likelihood of the survival model were used to define the Gompertz model as a custom family in brms. Data were simulated from a mixture cure rate model with a constant hazard to demonstrate implementation of the custom brms family. With the custom brms family in place, standard brms tooling was used for model fitting and inference.
Results: The Gompertz distribution better characterized the simulated survival curve than an Exponential, Lognormal, or Weibull distribution. The leave-one-out expected log pointwise predictive density (ELPD) model criterion identified the Gompertz model as the most favorable model, and visual predictive checks showed that the asymptotic assumption of the standard survival models led to an overprediction of adverse events in the latter part of the simulated study. The Gompertz model did not suffer from such a misspecification and characterized the data well.
Conclusions: The Gompertz distribution characterized the simulated survival data. The custom brms family can be used to model exposure-response data where a proportion of the population never experience the event of interest.

Methods

As a case study, time-to-event data were generated for 1000 virtual subjects, with 60% of the population susceptible to the event of interest. Events were simulated among the susceptible population by an Exponential distribution with a hazard of 0.08 (1/day). Censoring was simulated by an Exponential distribution with a hazard of 0.06 (1/day).

Four parametric survival models were fit to the data via the brms R package: Gompertz, Exponential, Lognormal, and Weibull. The Gompertz distribution is not natively supported by Stan or brms and was implemented as a custom distribution in brms, called a custom "family". The Gompertz survival model can be defined by the hazard function:

$$h(t) = \mu e^{\gamma t},$$

where $\mu > 0$ is the shape parameter, and γ is the scale parameter. When $\gamma < 0$ the Gompertz distribution describes a cure model where the survival function is bounded below by the cure fraction $\exp(\mu/\gamma)$.

Three Stan functions are needed for a custom continuous family in brms: the probability density (f), cumulative distribution (F), and complementary cumulative distribution ($\bar{F}$), each defined on a log scale. For the Gompertz distribution these are defined by Equations (1–3).

$$\log(f(t)) = \log(\mu) + \gamma t + \frac{\mu}{\gamma} - \frac{\mu}{\gamma} e^{\gamma t}, \tag{1}$$

$$\log(F(t)) = \log\left(1 - \exp\left(\frac{\mu}{\gamma}(1 - e^{\gamma t})\right)\right), \tag{2}$$

$$\log(\bar{F}(t)) = \frac{\mu}{\gamma}(1 - e^{\gamma t}). \tag{3}$$

A fourth function, a random generation function, was needed for simulation from the Gompertz model. Solving Equation (3) for t and sampling from a Uniform$(0, 1)$ distribution in place of $F(t)$ yielded

$$t = \frac{1}{\gamma} \log\left(1 - \frac{\gamma}{\mu} \log(1 - \text{Uniform}(0, 1))\right). \tag{4}$$

If $\gamma < 0$ then $F(t)$ is bounded above by a value less than 1, and the argument of the outer log in Equation (4) could be negative. This occurs when the uniform random variable is greater than $1 - \exp(\mu/\gamma)$, i.e., when the sampled proportion is greater than the proportion of the population who would ever have an event. In this case, an event time of infinity was assigned. Equations (1–4) were defined with Stan syntax and stored as a string in R (Code 1).

```
stan_funs <- "
  real gompertz_lpdf(real t, real mu, real gamma) {
    return log(mu) + gamma*t + mu/gamma - mu/gamma*exp(gamma*t);
  }
  real gompertz_lcdf(real t, real mu, real gamma) {
    return log(1 - exp(mu/gamma*(1 - exp(gamma*t))));
  }
  real gompertz_lccdf(real t, real mu, real gamma) {
    return mu/gamma*(1 - exp(gamma*t));
  }
  real gompertz_rng(real mu, real gamma) {
    real sim_time;
    sim_time = 1/gamma*log(1 - gamma/mu*log(1-uniform_rng(0, 1)));
    if (is_nan(sim_time))
      sim_time = positive_infinity();
    return sim_time;
  }
"
```

Code 1. Custom Stan functions defined in R.

```
gompertz <- brms::custom_family(
  name = "gompertz",
  dpars = c("mu", "gamma"),
  links = c("log", "identity"),
  lb = c(0, NA),
  type = "real",
  log_lik = function(i, prep) {
    mu <- brms::get_dpar(prep, "mu", i = i)
    gamma <- brms::get_dpar(prep, "gamma", i = i)
    t <- prep$data$Y[i]
    cens <- prep$data$cens[i]
    if (cens == 0) x <- gompertz_lpdf(t, mu, gamma)
    if (cens == 1) x <- gompertz_lccdf(t, mu, gamma)
    return(x)
  },
  posterior_predict = function(i, prep, ...) {
    mu <- brms::get_dpar(prep, "mu", i = i)
    gamma <- brms::get_dpar(prep, "gamma", i = i)
    return(gompertz_rng(mu, gamma))
  }
)
```

Code 2. Custom brms family defined in R.

The Gompertz family was defined in brms with the `custom_family()` function (Code 2). A name argument is required and must match the distribution name used in the custom Stan functions, i.e., "gompertz". The distribution parameter names, their link functions, and bounds on the parameters were also specified. Each brms family must have a μ (intercept) parameter. The distribution was specified as continuous by setting the type to "real".

Two additional post-processing functions were defined as part of the custom brms family: the log-likelihood (`log_lik()`) and a posterior response prediction function (`posterior_predict()`). The response prediction function used the previously defined Stan function, `gompertz_rng()` (Code 1). Posterior draws of the expected value (`posterior_epred()`) can also be defined with the custom brms family but was omitted because the Gompertz distribution with a negative scale parameter has a non-finite mean survival time.

A prior with mass around $\log(\mu) \in (-6, -3)$ and $\gamma \in (-0.1, -1e-8)$ allowed for a flexible survival function that captured a large range of cure fractions and event rates (Figure 1). Priors were centered at the middle of the intervals stated above with standard deviations approximately one-third the interval width (Code 3):

$$\log(\mu) \sim \text{Normal}(-4.5, 1^2), \qquad \gamma \sim \text{Normal}(-0.05, 0.03^2).$$

With custom Stan functions and the custom brms family in place, model-fitting was done in the usual way with brms, making sure to pass the custom functions to Stan (Code 4). After `log_lik()` and `posterior_predict()` were made available to the R environment with `expose_functions()` (Code 4), post-processing was done with typical brms tools.

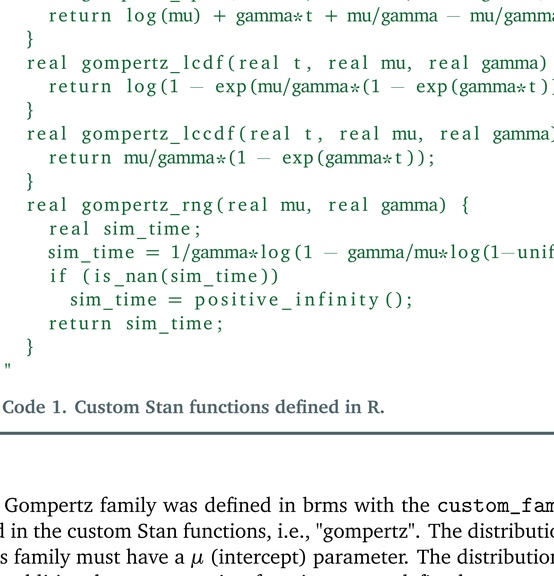

Figure 1. Survival function for a selection of Gompertz distribution parameter values.

```
priors <- brms::prior(normal(-4.5, 1), class = "Intercept") +
  brms::prior(normal(-0.05, 0.03), class = "gamma")
```

Code 3. Cure rate priors for Gompertz family.

```
fit <- brms::brm(
  formula = time | cens(1 - status) ~ 1,
  family = gompertz,
  prior = priors,
  data = dat,
  stanvars = brms::stanvar(scode = stan_funs, block = "functions")
)

brms::expose_functions(fit, vectorize = TRUE)
```

Code 4. Gompertz family brms fit.

Results

The Kaplan-Meier curve plateaued at 41%, consistent with 40% of the simulated population being insusceptible to the event of interest (Figure 2).

Of the four tested models, the Gompertz model was selected by the ELPD model criterion as the most favorable and had an ELPD value more than three standard errors higher than the second most favorable model (Table 1).

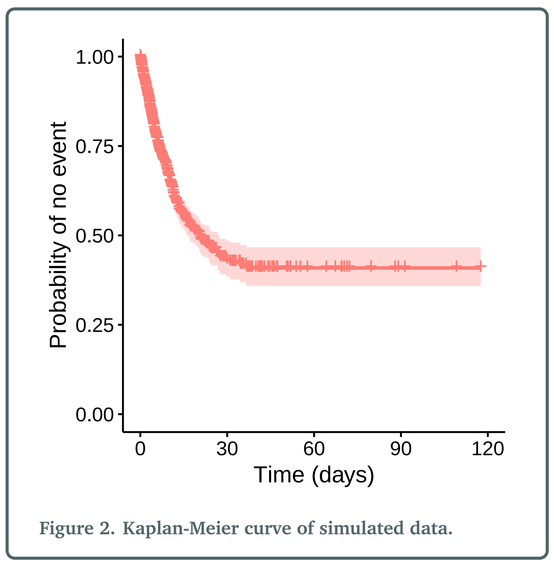

Figure 2. Kaplan-Meier curve of simulated data.

Model	ELDP	SE_{ELPD}	ΔELPD	$SE_{\Delta ELPD}$
Gompertz	-1487.1	49.5	0.00	0.00
Lognormal	-1497.8	49.8	-10.7	3.06
Weibull	-1511.3	50.1	-24.3	5.23
Exponential	-1530.7	50.5	-43.6	9.57

ELPD: leave-one-out expected log pointwise predictive density
SE_{ELPD}: standard error of ELPD
ΔELPD: difference between ELPD and most favorable ELPD
$SE_{\Delta ELPD}$: standard error of ΔELPD

Table 1. Model criterion.

Posterior predictive checks demonstrated the inability of standard survival models to adequately characterize the insusceptible proportion of subjects, with events overpredicted in the latter part of the simulated study (Figure 3). The Gompertz model did not suffer from such overprediction and characterized the data well (Figure 4).

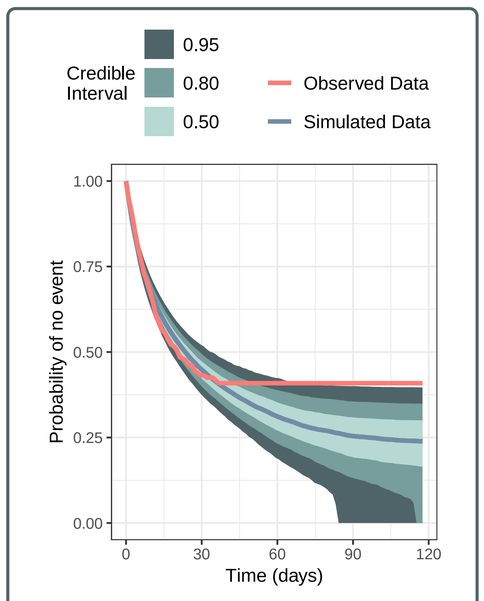

Figure 3. Lognormal model posterior predictive check.

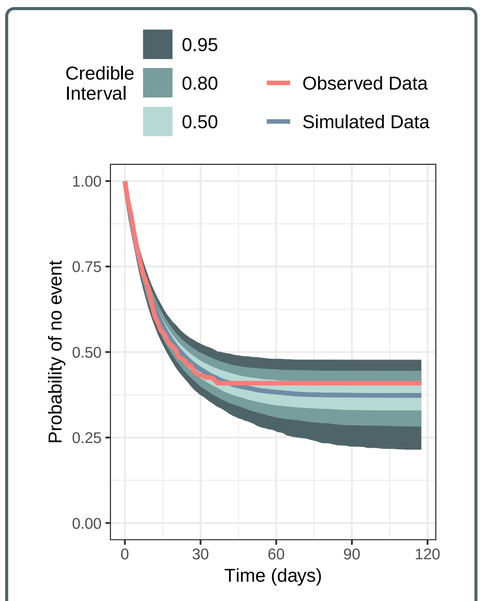

Figure 4. Gompertz model posterior predictive check.

The Gompertz model estimated that 36.8% of the population was insusceptible to the event of interest, consistent with the 40% used in the simulated data (Table 2).

Parameter	Estimand	Estimate	95% Credible Interval
shape	$\log(\mu)$	-2.93	(-3.08, -2.80)
scale	γ	-0.0531	(-0.0670, -0.0397)
cure rate	$\exp(\mu/\gamma)$	0.368	(0.215, 0.502)

Table 2. Gompertz model parameter estimates.

Conclusion

The Gompertz brms family enables time-to-event exposure-response modeling in a Bayesian framework, allowing for a proportion of the population to never have an event. The custom family allows model-fitting, covariate effects, and post-processing inference to proceed in the usual brms workflow.

References

[1] Gieser, P.W., Chang, M.N., Rao, P.V., Shuster, J.J. and Pullen, J. Modelling cure rates using the Gompertz model with covariate information. *Stat. Med.* **17** (1998):831–839.

[2] Bürkner, P.C. brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software* **80** (2017):1–28.

[3] Stan Development Team. Stan Modeling Language Users Guide and Reference Manual, 2.35. `https://mc-stan.org` (2024).

Metrum Research Group Publications and Posters