

Supplementary Material to

Nonparametric Bayesian inference for mean residual life functions in survival analysis

VALERIE POYNOR¹ and ATHANASIOS KOTTAS²

¹*California State University, Fullerton*

²*University of California, Santa Cruz*

Section A. Prior Specification Details

As discussed in Section 3.2 of the paper, we specify the hyperparameters using an approximate range of the data a priori. In order to reduce the number of hyperparameters to specify, we take B_μ and B_Σ to be diagonal with the same diagonal element b_μ and b_Σ , respectively. We also set $a_\Sigma = 4$, which is the smallest integer value that ensures finite prior expectation for the inverse-Wishart prior on Σ . Using the expression for the bivariate normal moment generating function at points $t_1 = (1, -2)'$, $t_2 = (2, -2)'$, $t_3 = (1, -1)'$, and the prior independence for μ and Σ , we approximate the marginal variance as follows:

$$\begin{aligned}
 \text{Var}(T) &= \text{E}\{\text{Var}(T \mid \mu, \Sigma)\} + \text{Var}\{\text{E}(T \mid \mu, \Sigma)\} \\
 &= \text{E}\{\text{Var}(\text{E}(T \mid \theta, \phi)) + \text{E}(\text{Var}(T \mid \theta, \phi))\} + \text{Var}\{\text{E}(\text{E}(T \mid \theta, \phi))\} \\
 &= \text{E}\{\text{Var}(e^{\theta-\phi} \mid \mu, \Sigma)\} + \text{E}\{\text{E}(e^{\theta-2\phi} \mid \mu, \Sigma)\} + \text{Var}\{\text{E}(e^{\theta-\phi} \mid \mu, \Sigma)\} \\
 &= \text{E}(e^{t_2'\mu})\text{E}(e^{0.5t_2'\Sigma t_2}) + \text{E}(e^{t_1'\mu})\text{E}(e^{0.5t_1'\Sigma t_1}) - \{\text{E}(e^{t_3'\mu})\}^2\{\text{E}(e^{0.5t_3'\Sigma t_3})\}^2 \\
 &\approx \exp(t_2'a_\mu + 0.5t_2'B_\mu t_2 + 0.5t_2'B_\Sigma t_2) + \exp(t_1'a_\mu + 0.5t_1'B_\mu t_1 + 0.5t_1'B_\Sigma t_1) \\
 &\quad - \exp(2t_3'a_\mu + t_3'B_\mu t_3 + t_3'B_\Sigma t_3)
 \end{aligned} \tag{1}$$

where the last expression is obtained by using again the bivariate normal moment generating func-

tion, and by approximating $E\{\exp(0.5t'\Sigma t)\}$ with $\exp\{0.5t'E(\Sigma)t\}$. Analogously, for the marginal mean: $E(T) \approx \exp(t'_3 a_\mu + 0.5t'_3 B_\mu t_3 + 0.5t'_3 B_\Sigma t_3) = \exp(a_{\mu_1} - a_{\mu_2} + b_\mu + b_\Sigma)$, where $a_\mu = (a_{\mu_1}, a_{\mu_2})'$. Considering $E(T) = \text{midrange}$ and $\text{Var}(T) = (R/4)^2$, we now have two equations with four unknowns, a_{μ_1} , a_{μ_2} , b_μ and b_Σ . We can finally specify the prior hyperparameters by, for instance, allocating portions of $(R/4)^2$ to the different terms in the variance decomposition in (1), or by setting each term to $(R/4)^2$ to induce a more dispersed prior predictive distribution.

An important note about (1) is that the approximation, $E(e^{0.5t'\Sigma t}) \approx e^{0.5t'E(\Sigma)t}$ under $a_\Sigma = 4$ is only accurate when $|B_\Sigma| \lesssim 1 \times 10^{-6}$. If this is not the case the approximation will underestimate the variance of the model with increasing error as $|B_\Sigma|$ becomes larger. In other words, the prior variance tends to be much more dispersed than the approximation in (1) suggests. In our prior sensitivity analysis, this has only been problematic when data exhibits a long sparse tail. When data is moderately sparse in the tail, the over dispersed prior only effects the upper quantile estimates of the tail of the MRL function. When data is extremely sparse, the median may also be effected. In such cases where predictions over sparse data is necessary, we suggest increasing the degrees of freedom parameter a_Σ . Samples from the prior predictive distribution can be obtained to monitor the true variance of the model under different values of a_Σ to determine the most appropriate prior specification.

We note that sensitivity analysis for a_Σ was performed for the two real data examples of the paper, and the effect was negligible. In order to demonstrate a scenario in which a_Σ plays a significant role, we simulate from an extremely sparse long tailed population and fit the gamma DPMM under two values of a_Σ . Results are discussed in more detail in Section B.1.

Section B. Simulations

In this section we consider three simulated data sets to assess properties of the gamma DPMM such as posterior convergence, computation time, effects of censoring, and posterior accuracy and precision under three different populations. The first simulation is presented in Section B.1 and demonstrates the ability of the gamma DPMM to fit more standard shapes for the MRL function without overfitting. The data are generated from an exponentiated Weibull population that exhibits a long sparse tail. Sensitivity analysis around a_Σ is discussed along with convergence and computation time. In Section B.2, we consider two simulated data examples arising from a mixture of gamma distributions to illustrate the ability of the gamma DPMM to capture non-standard

MRL function shapes as well as the correlation between kernel parameters θ and ϕ . These data are synthesized from a population having the same structure as the model to purposely create a positive and negative correlation of the parameters in the gamma components to test how well the model identifies the true correlation. We also illustrate the effects of (independent random) right censoring under both simulation settings.

B.1 MCMC Diagnostics and Prior Sensitivity

In this simulation example, we generate data from the exponentiated Weibull distribution (discussed in Section 2 of the paper) in which the mean residual life function takes on an upside-down bathtub shape: $S(t \mid \alpha, \theta, \sigma) = 1 - [1 - \exp\{-(t/\sigma)^\alpha\}]^\theta$ with $\alpha = 3$, $\theta = 0.08$, and $\sigma = 10$. This population distribution also exhibits an extremely sparse tail, which is challenging to estimate particularly for the MRL function. We consider two sample sizes, one smaller ($n = 60$) and one larger ($n = 600$), to illustrate the effect of the sample size on the model fit as well as the computation time. We also discuss the role of the degrees of freedom hyperparameter, a_Σ , in the prior predictive distribution and on posterior inference.

For this data set we choose the mean and variance by considering an exponential distribution that covers the range of the data, $(0, 16)$. In particular, an exponential distribution with rate 0.5 encompasses the range $(0, 16)$, leading to a prior mean of 2 and prior variance of 4. Under (1), we have the following prior set: $a_\mu = (2.33, 1.86)$, $B_\mu = B_\Sigma = \text{diag}(0.15, 0.15)$, $a_\alpha = 3$, $b_\alpha = 1$, and $L = 60$. Under this prior set, we simulate 2000 samples from the prior predictive distribution and find the range to be $(\approx 0, 15720)$. Clearly this range is much larger than our prior beliefs call for. As per our discussion in Section A, we increase the degrees of freedom parameter, a_Σ , until the range is approximately the range of the data. In particular when $a_\Sigma = 10$, the range of the prior predictive samples becomes $(\approx 0, 20)$.

We fit the gamma DPMM under both prior sets to each synthetic data set. The MCMC converged within 2000 iterations under three independent chains having diverse initial values. Convergence was informally monitored by visually examining trace plots of the 25th, 50th, and 90th quantiles of the cumulative distribution function. Computation time was measured using the *proc.time()* function in the base package of the R statistical computing language. Table 1 provides the CPU time in minutes for both simulated data sets.

Sample Size	MCMC Samples	
	100000	1200000
$n = 60$	0.35	3.64
$n = 600$	3.13	37.10

Table 1: Exponentiated Weibull Simulation. The central processing unit (CPU) time (in minutes) required to obtain the indicated number of posterior samples under two different sample sizes.

Posterior results for the MRL functions and the survival functions under the $a_\Sigma = 10$ prior are shown in Figure 1. Although not pictured, the effect of the prior value for a_Σ was very obvious in the tail of the MRL functions particularly for the larger sample size. In general, the 90th quantile for the MRL functions are much higher under $a_\Sigma = 4$ than under $a_\Sigma = 10$, illustrating the tendency of the MRL to be overestimated with a lower degrees of freedom parameter when data is sparse. The effect in the point estimate was more apparent in the larger data set than the smaller. Moreover, the truth in the larger data set was not captured in end tail region of the MRL function. The results under $a_\Sigma = 10$, showed a decrease in the size of the interval bands, or precision, as well as an increase in accuracy. The difference across the four scenarios in the inferential results for the survival function was negligible.

B.2 Correlation in Kernel Parameters and Effects of Censoring

We consider a mixture of four gamma mixture distributions in which the shape and scale parameters are positively associated: $0.35\Gamma(10, 0.5) + 0.4\Gamma(20, 1) + 0.15\Gamma(30, 5) + 0.1\Gamma(40, 8)$. We simulate 200 random draws to be our data set, and fit the gamma DPMM with priors $a_\mu = (1.6, 0.4)$, $B_\mu = B_\Sigma = \text{diag}(0.39, 0.39)$, and $a_\alpha = 2$, $b_\alpha = 1$.

Results are shown in Figure 2. Each of the density, survival, hazard, and MRL function of the true underlying distribution are contained within the gamma DPMM interval estimates, and the point estimate recovers well the shape of the true function within the range of the data. Moreover, the model captures the correlation between the kernel parameters, even though the prior does not favor either negative or positive correlation.

We demonstrate the effect of random right censoring for the MRL function under 25% and 50% censoring. In particular, we sample independently censored times from a gamma distribution. If the paired survival time was greater than the censored value, the survival time was replaced by the censored time and indicated as right censored. The parameters of the censoring distribution

were chosen such that probabilistically, 25% and 50% of the survival times were greater than their paired censored times. Results are shown in Figure 3. As expected the upper interval estimate is much larger when censoring is present. The interval estimate in the tail of the MRL is particularly affected by censoring. The point estimate maintains approximately the same accuracy under each censoring scenario.

We now consider a data set consisting of 200 draws from a gamma mixture distribution with negative correlation between the shape and rate parameters: $0.3\Gamma(15, 0.2) + 0.25\Gamma(12, 0.5) + 0.35\Gamma(8, 2) + 0.1\Gamma(3, 6)$. This population was chosen to test the model’s ability to estimate modes that are close together, as well as to capture the tail of a heavy-tailed distribution. A gamma DPMM was fit to the data with priors $a_\mu = (2.4, -1)$, $B_\mu = B_\Sigma = \text{diag}(0.18, 0.18)$, and $a_\alpha = 2$, $b_\alpha = 1$. Results are shown in Figure 4. Again, the model interval estimates contain the true density, survival, hazard, and MRL functional. The point estimates are close to the truth, even in the tail where less data is available. The uncertainty bands for the MRL estimate show a wider upper bound, which is likely an effect of the smaller sample size and the sparsity of the data from the tail of the underlying distribution. The model estimates a strong negative correlation between the kernel parameters.

We also consider the effect of censoring for this synthetic data set. We simulate censored observations as described in the previous example, again under two censoring levels: 25% and 50%. Results for the MRL are displayed in Figure 5. We can see the censoring has a much larger effect in this simulation example in comparison to the previous. In this example, the uncertainty is much larger specifically in the upper interval bands. The model seems to struggle to maintain precision. In addition, the point estimate in the tail is pulled up and away from the true MRL function. These results agree with our discussions regarding long tailed populations with sparse data in Section A and Section B.1. Here, our population is not as sparse in the tail as in Section B.1, however, when censoring is present we see similar behavior. We would suggest increasing the a_Σ in the presence of censoring particularly in long tailed populations to avoid placing an over dispersed prior.

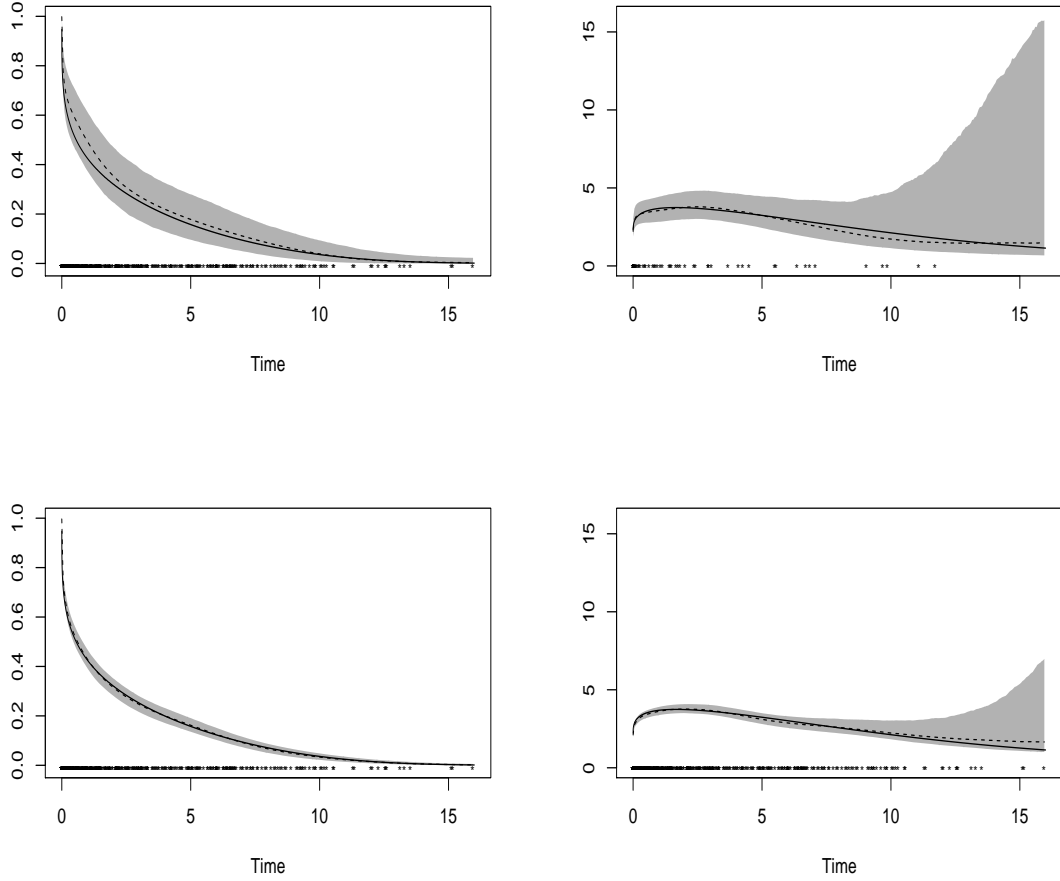


Figure 1: Exponentiated Weibull Simulation. Posterior point (dashed line) and 95% interval estimates (gray bands) for the Survival function (left column), and 80% interval estimates (gray bands) for the MRL function (right column) for simulated data having sample size of 60 (top row) and 600 (bottom row). The truth is depicted as the solid line. The simulated data are plotted along the bottom of each plot as an asterisk.

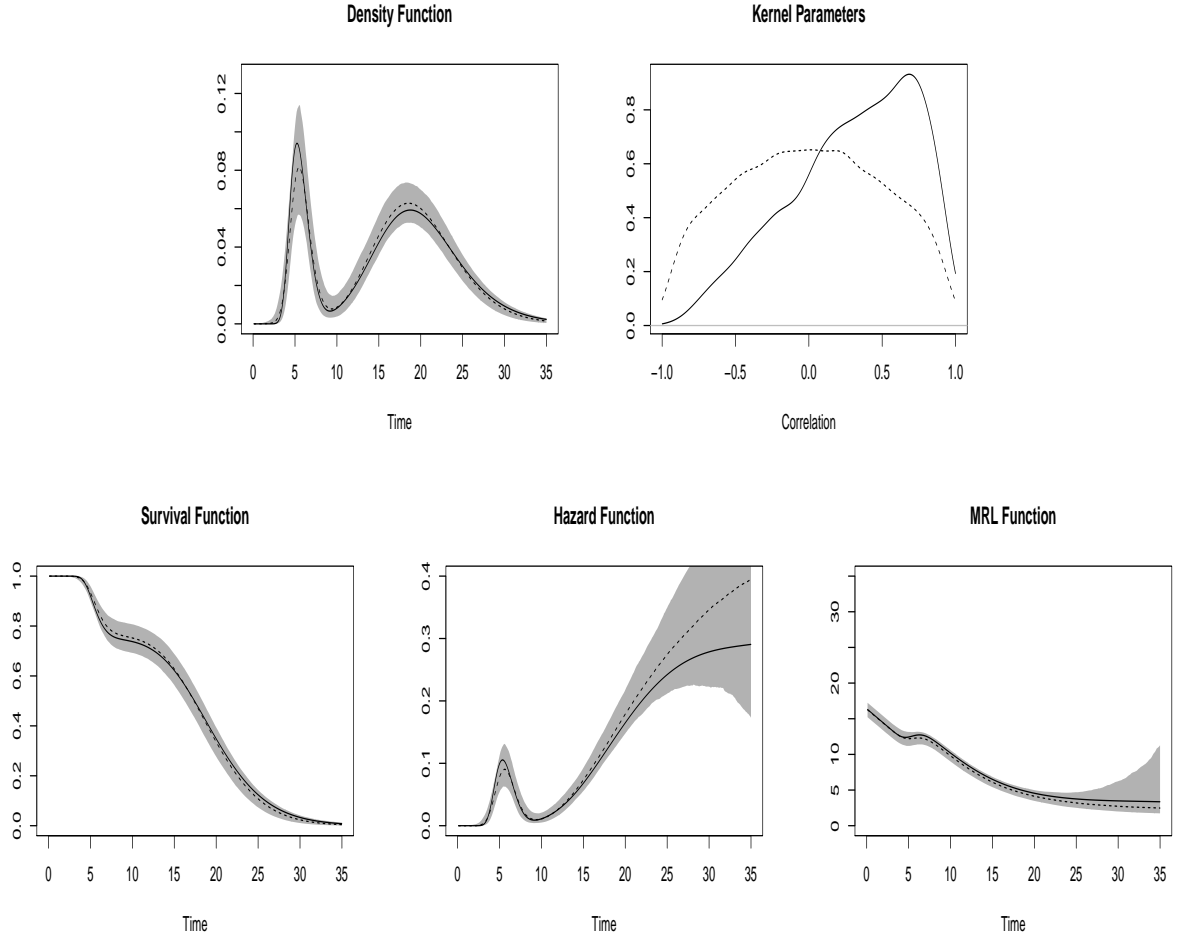


Figure 2: Positive Gamma Mixture Simulation. Posterior point (dashed line) and 95% interval estimates (gray bands) for the density function (top left panel, overlaying the sample histogram), survival function (bottom left panel), hazard function (bottom middle panel), and MRL function (bottom right panel). In all cases, the solid line indicates the corresponding true function. The top right panel plots the posterior density (solid line) and prior density (dashed line) of the correlation between θ and ϕ in the DP centering distribution.

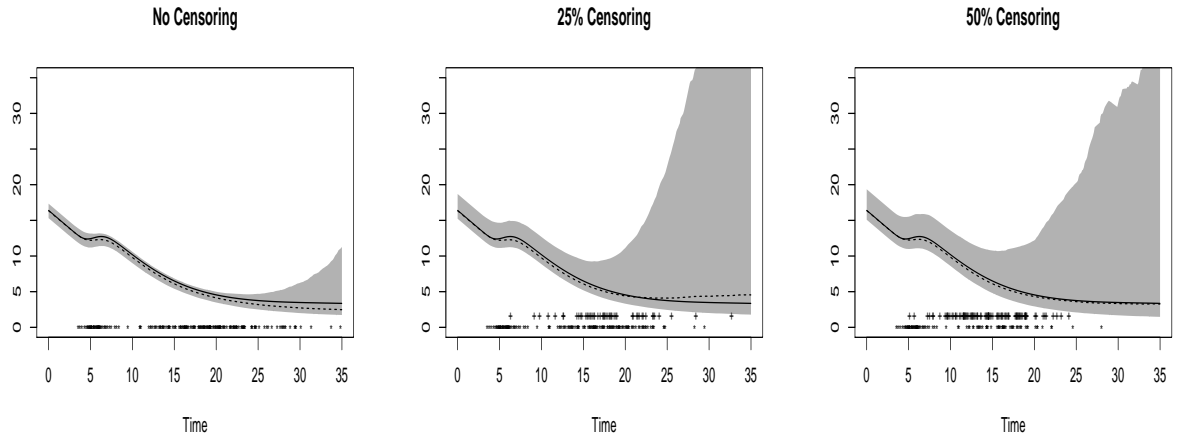


Figure 3: Positive Gamma Mixture Simulation. Point (dashed line) and 95% interval estimates (gray bands) for the MRL function for simulated data under 0% (left), 25% (middle), and 50% (right) censoring. The truth is depicted as the solid line. The simulated data are plotted along the bottom of each plot. The asterisk “*” and “+” are the fully observed and the right censored observations, respectively.

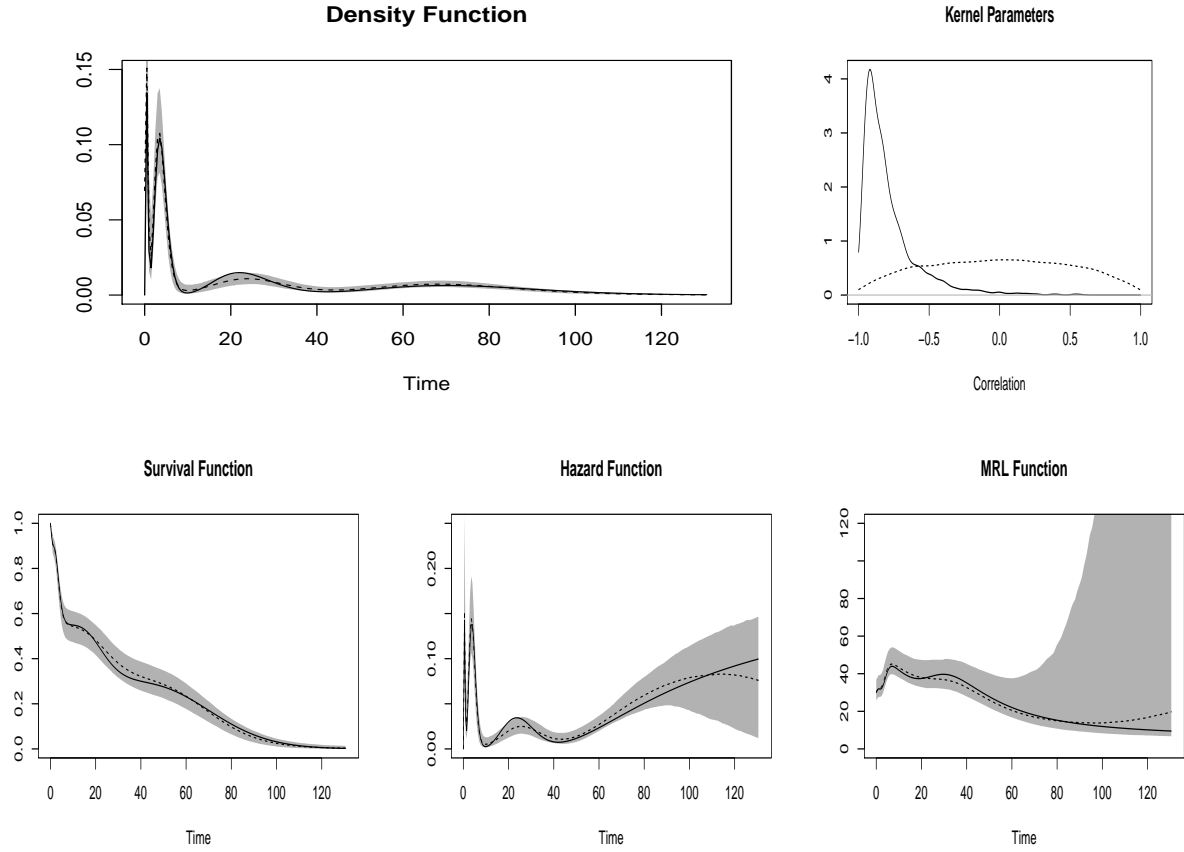


Figure 4: Negative Gamma Mixture Simulation. Posterior point (dashed line) and 95% interval estimates (gray bands) for the density function (top left panel, overlaying the sample histogram), survival function (bottom left panel), hazard function (bottom middle panel), and MRL function (bottom right panel). In all cases, the solid line indicates the corresponding true function. The top right panel plots the posterior density (solid line) and prior density (dashed line) of the correlation between θ and ϕ in the DP centering distribution.

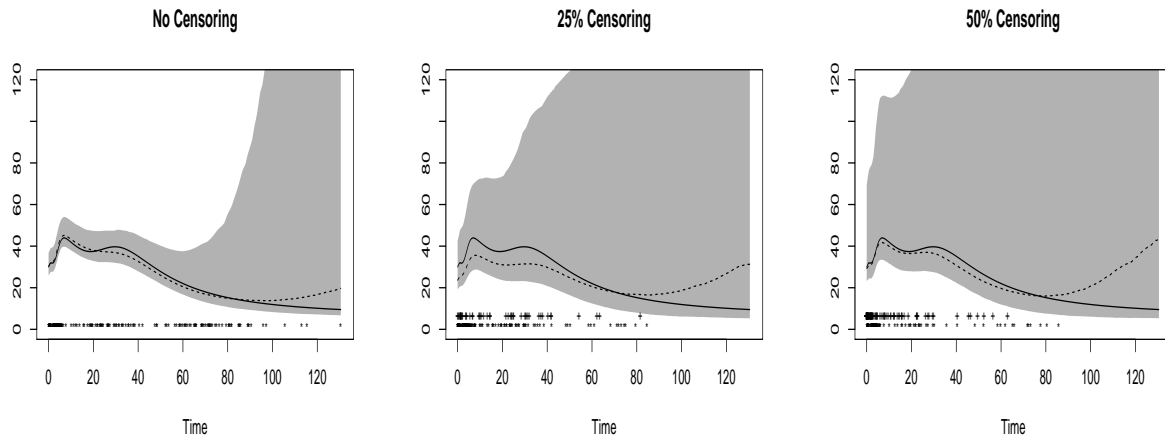


Figure 5: Negative Gamma Mixture Simulation. Point (dashed line) and 95% interval estimates (gray bands) for the MRL function for simulated data under 0% (left), 25% (middle), and 50% (right) censoring. The truth is depicted as the solid line. The simulated data are plotted along the bottom of each plot. The asterisk “*” and “+” are the fully observed and the right censored observations, respectively.