Aalto University
School of Engineering

Lung Cancer Survival Prediction with Bayesian Generalised Linear Models

Yann McLatchie `yann.mclatchie@aalto.fi`

Arina Odnoblyudova `arina.odnoblyudova@aalto.fi`

December 13, 2021

1 Introduction

1 Introduction

Introduction

Problem statement and data description

Related studies

1 Problem statement and data description

We look to predict the survival time of patients with advanced lung cancer.

- We start with 9 features and 149 observations, which we divide into 7 institutional groups;
- Two important features: `institution` (hierarchical covariate), `status` (censoring covariate);
- 44 patients with missing data, which we remove leaving 105 observations.

1 Related studies

- The original data were presented in Loprinzi et al. [1994];
- In Kumar and Sonker [2020] a comparison of semi-parametric and non-parametric models for survival analysis was presented;
- Abujarad and Khan [2018] provided a description of exponential models applied to a lung cancer survival time analysis with Stan code.

2

Description of models

2 Description of models

Description of models

- Weibull without censored data
- Weibull with censored data
- Hierarchical Weibull

2 Weibull without censored data

$$y_i \sim \text{Weibull}(\alpha, \sigma)\,, \quad i = 1, \dots, N,$$
$$\alpha \sim \text{Half-Cauchy}(5),$$
$$\sigma = \exp\left(-\frac{\boldsymbol{X}\beta}{\alpha}\right),$$
$$\beta_m \sim N(0, 1)\,, \quad m = 1, \dots, M.$$

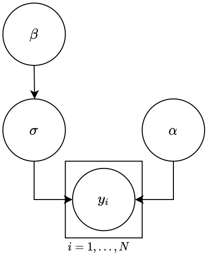

Figure: Pooled model DAG.

2 Weibull with censored data

Density function for Weibull distributed survival times:

$$p(t_i|\alpha,\gamma_i) = \exp\left(-\left(\frac{t_i}{\gamma_i}\right)^{\alpha}\right)\frac{\alpha}{\gamma_i}\left(\frac{t_i}{\gamma_i}\right)^{\alpha-1},$$

where α – shape, γ – scale, $\lambda : \lambda = -\alpha\log\gamma$. Survival function:

$$S(t_i|\alpha,\lambda_i) = \exp(-\exp(\lambda_i)t_i^{\alpha})$$

Log-likelihood function:

$$l(\alpha,\beta|t,x) = \sum_{i=1}^{n} v_i(\log(\alpha) + (\alpha-1)log(t_i) + X_i\beta) - \exp(X_i\beta)t_i^{\alpha},$$

where v_i an indicator showing 0 if the data are censored and 1 if not.

2 Weibull with censored data

$$
\begin{aligned}
y_i &= \begin{cases} \mathit{Weibull}(\alpha, \sigma), & \text{if } y_i \text{ observed} \\ \mathit{LogWeibullCompCum}(\alpha, \sigma), & \text{if } y_i \text{ unobserved} \end{cases} \\
\alpha &\sim \text{Half-Cauchy}(5), \\
\sigma &= \exp\left(-\frac{\boldsymbol{X}\beta}{\alpha}\right), \\
\beta_j &\sim N(0, 1)\,, \quad j = 1, \dots, M.
\end{aligned}
$$

2 Hierarchical Weibull model

$$
\begin{aligned}
y_{ij} &\sim \text{Weibull}(\alpha, \sigma_j)\,, \quad j = 1, \dots, J \\
\alpha &\sim \text{Half-Cauchy}(5), \\
\sigma_j &\propto \boldsymbol{X}\beta_j, \\
\beta_{mj} &\sim N(\mu_\beta, \sigma_\beta)\,, \quad m = 1, \dots, M, \\
\mu_\beta &\sim N(0, 1), \\
\sigma_\beta &\sim \text{Gamma}(1, 1).
\end{aligned}
$$

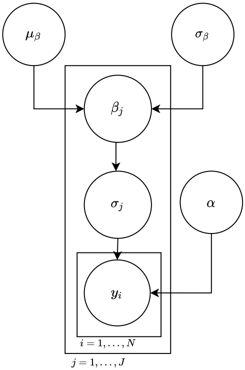

Figure: Hierarchical model DAG.

3

Diagnostics and performance

3 Diagnostics and performance

Diagnostics and performance

- Convergence and effective sample size
- Posterior predictive checks
- Divergence diagnostics
- Model comparison and predictive performance
- Prior sensitivity analysis

3 $\hat{R}$ and effective sample size

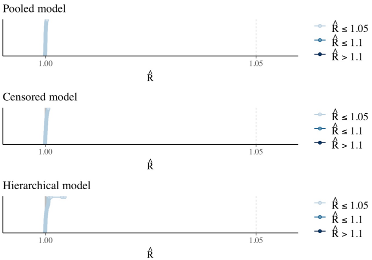

Figure: Rank-normalised $\hat{R}$ for the three models.

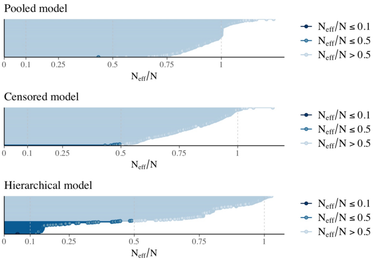

Figure: n_{eff}/N for the three models.

3 Posterior predictive checks

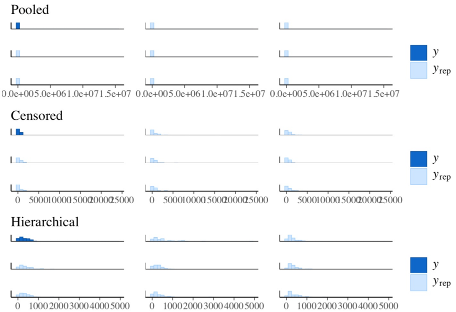

Figure: Posterior predictive checks for all three models.

3 Divergence diagnostics

	Mean tree depth	Mean divergent transitions
Pooled model	4.2	0.24
Censored model	5.1	0.021
Hierarchical model	6.5	0.038

Table: Mean tree depths and divergent transitions in all three models.

3 Model comparison and predictive performance

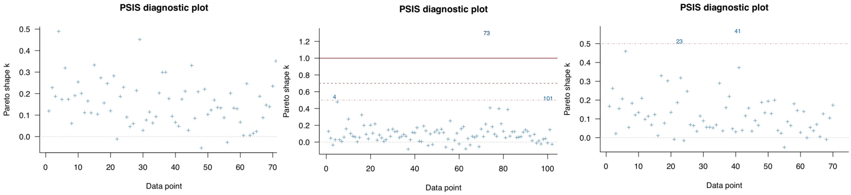

Figure: Pareto *k* values for the pooled model, censored data model, and hierarchical model (left to right).

3 Model comparison and predictive performance

```
##
## Computed from 20000 by 72 log-likelihood matrix
##
##          Estimate   SE
## elpd_loo   -631.9 10.0
## p_loo        5.4  0.5
## looic     1263.8 20.0
## ------
## Monte Carlo SE of elpd_loo is 0.0.
##
```

```
##
## Computed from 20000 by 105 log-likelihood matrix
##
##          Estimate   SE
## elpd_loo   -527.0 31.0
## p_loo       12.9  5.8
## looic     1053.9 62.0
## ------
## Monte Carlo SE of elpd_loo is NA.
##
```

```
##
## Computed from 20000 by 72 log-likelihood matrix
##
##          Estimate   SE
## elpd_loo   -485.6  6.1
## p_loo        6.5  1.1
## looic      971.3 12.1
## ------
## Monte Carlo SE of elpd_loo is 0.0.
##
```

Figure: ELPD LOO values for the pooled model, censored data model, and hierarchical model (left to right).

3 Prior sensitivity analysis

- Not particularly sensitive to the changes in prior distributions' parameters;
- Global scale parameter α prior change from Half-Cauchy to Gamma damages convergence and other diagnostics;
- Variance hyperparameter σ_β prior change from Gamma to Half-Cauchy damages convergence and other diagnostics.

4

Conclusion

4 Conclusion

4 Issues and improvements

- Experiment with further feature pre-processing;
- Try another distribution (gamma or such extensions of exponential, such as exponentiated exponential);
- Try other categorical features for the hierarchical approach;
- Low n_{eff} value for hierarchical model parameter.

4 Conclusion

- The survival time of patients with advanced lung cancer can be modelled with Weibull distribution;
- A hierarchical Weibull model is adept to predicting patient survival times;
- The institution to which a patient is admitted changes that patient's survival dynamics.

4 References

Abujarad, M. H. and Khan, A. (2018). Exponential model: A bayesian study with stan. *International Journal of Scientific Research*.

Kumar, M. and Sonker, P. (2020). Parametric survival analysis using r: Illustration with lung cancer data. *CANCER REPORTS*, 3.

Loprinzi, C., Laurie, J., Wieand, H., Krook, J., Novotny, P., Kugler, J., Bartel, J., Law, M., Bateman, M., and Klatt, N. (1994). Prospective evaluation of prognostic variables from patient-completed questionnaires. north central cancer treatment group. *Journal of Clinical Oncology*, 12:601–607.