Online Supplementary Material to the Paper: Variable Selection in a Log-linear Birnbaum-Saunders Regression Model for High-dimensional Survival Data via the Elastic-Net and Stochastic EM

Yukun Zhang, Xuewen Lu ^{*}
Department of Mathematics and Statistics
University of Calgary

Anthony F. Desmond
Department of Mathematics and Statistics
University of Guelph

December 10, 2015

1 Selection of tuning parameters

We want to select the optimal tuning parameter λ to minimize risk in estimation and prediction such as the population mean squared error that measures the average of the squared difference between the estimator and what is estimated. In practice, we can select tuning parameters by minimizing the cross-validation error.

Since there are actually two tuning parameters (τ and λ) in the elastic-net procedure, we cross-validate on a two dimensional surface. For example, in a K -fold cross-validation for

^{*}Corresponding author and presenter. Email: lux@math.ucalgary.ca. Phone: (1-403)220-6620. Fax: (1-403)282-5150.

selecting λ , we fix τ first, for each $k = 1, 2, \dots, K$, we fit the model with tuning parameter λ to the other $K - 1$ parts to obtain $(\hat{\boldsymbol{\theta}}, \hat{\alpha})^{-k}(\lambda)$ and compute its error in predicting the k th part:

$$E_k(\lambda) = \sum_{i \in kth \text{ part}} (Y_i^* - \mathbf{x}_i' \hat{\boldsymbol{\theta}}^{-k}(\lambda))^2.$$

This gives a cross-validation error:

$$CV(\lambda) = \frac{1}{K} \sum_{k=1}^K E_k(\lambda).$$

In practice, we use ten-fold cross-validation. As introduced in Section 2.2, τ is chosen between 0 and 1. We partition 0 to 1 to 51 equally spaced grid points, that is, $(0, 0.02, 0.04, \dots, 0.96, 0.98, 1)$. For each τ value in this set of grid points, we calculate λ using the function provided by the R package **glmnet** (Simon et al., 2011), and finally select τ and λ with the smallest cross-validation error.

2 Coordinate descent algorithm

We introduce a pathwise algorithm (Simon et al., 2011) via cyclical coordinate descent based on the elastic-net penalty for solving (15) in the paper. This algorithm gives us a solution path and is stable by exploiting warm starts. Suppose we have estimates $\tilde{\theta}_0$ and $\tilde{\theta}_l$ for $l \neq j$ and want to minimize the objective function with respect to θ_j . We can compute the gradient at $\theta_j = \tilde{\theta}_j$.

$$\frac{\partial M_\lambda}{\partial \theta_j} = -\frac{1}{n} \sum_{i=1}^n x_{ij}(Y_i^* - \tilde{\theta}_0 - \ddot{\mathbf{x}}_i' \tilde{\boldsymbol{\theta}}) + \lambda \tau \cdot \text{sgn}(\theta_j) + \lambda(1 - \tau)\theta_j, \quad (1)$$

where $\ddot{\mathbf{x}}_i$ is denoted in $\mathbf{x}_i = (1, \ddot{\mathbf{x}}_i)'$ and $\text{sgn}(\theta_j)$ is a sign function of θ_j .

Simple calculus (Donoho and Johnstone, 1994) shows that the coordinate-wise update has the form

$$\tilde{\theta}_j = \frac{S\left(\frac{1}{n} \sum_{i=1}^n x_{ij}(Y_i^* - \tilde{Y}_i^{(j)}), \lambda \tau\right)}{1 + \lambda(1 - \tau)}, \quad (2)$$

with $\tilde{Y}_i^{(j)} = \tilde{\theta}_0 + \sum_{l \neq j} x_{il} \tilde{\theta}_l$ being the fitted value by excluding the contribution from x_{ij} , and $S(x, \lambda) = \text{sgn}(x)(|x| - \lambda)_+$. Applying Equation (2) to the coordinates of $\boldsymbol{\theta}$ in a cyclic fashion until convergence minimizes the objective function $M_\lambda(\boldsymbol{\theta})$ given in (15) in the paper. This procedure can also be done using the R package **glmnet**.

3 Example 3: Variable selection for censored correlated data

Example 3: Based on Example 1, generate $\mathbf{X}$ from a multivariate normal distribution with mean zero and covariance matrix $(\rho^{|i-j|})_{i,j=1,\dots,n}$, ρ is set to 0.1, 0.5 and 0.8 respectively to control the amount of correlation between covariates.

In this example, we fix α at 0.5. From Table 6, we see that under moderate correlation ($\rho = 0.5$), LASSO and elastic-net have similar performance. When under small correlation ($\rho = 0.1$), elastic-net selects less variables than LASSO; while under high correlation ($\rho = 0.8$), elastic-net tends to select more variables than LASSO, although the difference is not big. Different censoring rates do not have a significant impact on the results. Lin and Lv (2013) generated $\mathbf{X}$ in the same way as Example 3. Their results didn't show any obvious difference between elastic-net and LASSO. They concluded that the elastic-net improved the performance of the LASSO only marginally in their case. Table 7 shows the selected variables in the highest ranks in terms of the percentages being selected in all the simulation runs. We notice that under high correlation ($\rho = 0.8$), the fifth variable with the smallest absolute value 0.5 of coefficient is not selected in any methods or censoring rates.

4 Model misspecification study

If the model and the model assumptions are violated, it is natural to ask whether the proposed method is still reliable. We conduct simulations to assess the effect of model misspecification on the inference. In Examples 1, 2 and 3, we generate data from a log-linear Weibull regression model. Specifically, the error follows a centred extreme value distribution, that is, $\varepsilon = \sigma(W - 0.5771)/\sqrt{\pi^2/6}$; here W has an extreme-value density function, $f(w) = \exp(w - \exp(-w))$, $-\infty < w < \infty$, $\sigma > 0$ is a constant, and 0.5771 is Euler constant, so that $E(\varepsilon) = 0$ and $Var(\varepsilon) = \sigma^2$. In the BS model, the error ε follows the sinh-normal distribution $SN(\alpha, 0, 2)$. Its mean is zero, the variance does not admit a closed form. By simulation, we obtain its variance as approximately 0.236 and 1.645 for $\alpha = 0.5$ and $\alpha = 1.5$, respectively. Hence, in the simulated log-linear Weibull model, we generate data for $\sigma^2 = 0.236$ and 1.645 respectively such that the two models have the

Table 6: Example 3: Results for variable selection in censored correlated data simulation.

Numbers in parentheses are standard deviations.

ρ	cr (%)	Method	$\#S$	$\#FN$	L_1 -loss	L_2 -loss	Median of MSE
0.1	17	LASSO	18.93 (5.80)	0.02 (0.00)	1.41 (0.39)	0.46 (0.11)	0.00
		Enet	18.38 (6.37)	0.02 (0.14)	1.36(0.39)	0.46 (0.11)	0.00
	27	LASSO	19.81 (5.25)	0.06 (0.14)	1.66 (0.43)	0.54 (0.15)	0.00
		Enet	18.39 (6.01)	0.05 (0.22)	1.56 (0.43)	0.54 (0.16)	0.00
	43	LASSO	20.57 (5.32)	0.14 (0.17)	2.10 (0.67)	0.70 (0.24)	0.00
		Enet	17.64 (6.14)	0.12 (0.33)	1.96 (0.64)	0.70 (0.23)	0.00
0.5	17	LASSO	15.1 (7.15)	0.38 (0.49)	1.59 (0.47)	0.62 (0.18)	0.00
		Enet	15.4 (7.05)	0.37 (0.49)	1.59 (0.47)	0.62 (0.18)	0.00
	27	LASSO	15.69 (6.28)	0.46 (0.50)	1.82 (0.53)	0.70 (0.19)	0.00
		Enet	15.12 (6.68)	0.50 (0.51)	1.85 (0.42)	0.72 (0.16)	0.00
	43	LASSO	15.03 (6.08)	0.62 (0.49)	2.11(0.61)	0.82 (0.21)	0.00
		Enet	15.05 (6.21)	0.65 (0.48)	2.11 (0.64)	0.82 (0.21)	0.00
0.8	17	LASSO	15.47 (5.63)	0.98 (0.14)	2.45 (0.56)	0.98 (0.17)	0.00
		Enet	16.32 (5.01)	0.97 (0.39)	2.53 (0.56)	0.98 (0.17)	0.00
	27	LASSO	14.67 (5.49)	0.98 (0.14)	2.66 (0.68)	1.08 (0.25)	0.01
		Enet	17.08 (4.75)	0.98 (0.38)	2.81 (0.70)	1.08 (0.25)	0.01
	43	LASSO	14.87 (5.22)	0.98 (0.14)	3.03 (0.83)	1.24 (0.34)	0.01
		Enet	18.42 (4.55)	0.98 (0.44)	3.23 (0.85)	1.24 (0.34)	0.01

Table 7: Example 3: Selected variables in censored correlated data simulation under different settings.

ρ	cr (%)	Method	Top S ranked variables																				
0.1	17	LASSO	2	3	4	5	96	149	201	134	103	153	89	28	97	13	198	19	84	144	142		
		Enet	2	3	4	5	97	141	16	84	110	149	55	64	86	127	143	158	23	33			
	27	LASSO	2	3	4	5	89	134	13	28	178	149	153	198	201	103	96	97	147	84	163	204	
		Enet	2	3	4	5	141	86	96	97	137	149	10	16	113	140	11	48	100	110			
	43	LASSO	2	3	4	5	84	134	37	201	19	97	103	13	153	163	198	89	45	149	158	204	28
		Enet	2	3	4	5	140	59	86	13	39	89	96	127	149	36	48	52	54	70			
0.5	17	LASSO	2	3	4	5	6	42	57	68	146	158	34	50	62	65	82						
		Enet	2	3	4	5	68	179	6	42	57	158	34	50	51	53	62						
	27	LASSO	2	3	4	5	57	80	135	146	14	42	53	62	75	144	8	24					
		Enet	2	3	4	5	42	60	80	203	57	62	63	136	14	32	66						
	43	LASSO	2	3	4	5	92	57	31	37	51	53	112	118	129	152	204						
		Enet	2	3	4	5	179	51	92	53	57	82	12	14	49	66	80						
0.8	17	LASSO	2	3	4	200	24	53	176	1178	108	198	21	32	33	93	124						
		Enet	2	3	4	111	37	16	119	133	167	81	192	191	112	134	33	45					
	27	LASSO	2	3	4	10	24	156	178	200	56	108	110	152	172	22	43						
		Enet	2	3	4	111	16	66	106	81	63	141	171	15	145	191	17	26	107				
	43	LASSO	2	3	4	43	54	37	56	178	203	24	32	42	45	69	83						
		Enet	2	3	4	16	111	107	141	17	102	66	152	42	121	171	63	104	64	49			

same variance in error. We apply the proposed method to this log-linear Weibull model and report the results in Tables 8 and 9. Comparing these results to those shown in Tables 2 and 4 in the paper for Examples 1, 2 and 3 from the correctly specified BS model, we see that the results are quite close, although the results for the misspecified model have slight larger L_1 and L_2 losses. This indicates that when the model assumptions are violated, the proposed method is still reliable.

Table 8: Compared to Example 1, the true model is log-linear Weibull: Results for variable selection in uncensored and censored data simulation. Numbers in parentheses are standard deviations.

α	cr (%)	$\#S$	$\#FN$	L_1 -loss	L_2 -loss	Median of MSE
0.5	0	11.69 (4.36)	0.00 (0.00)	0.90 (0.20)	0.49 (0.08)	0.00
	16	11.91 (4.75)	0.00 (0.00)	1.06 (0.33)	0.54 (0.12)	0.00
	29	11.80 (4.42)	0.02 (0.14)	1.34 (0.68)	0.66 (0.24)	0.00
	41	11.98 (4.09)	0.08 (0.27)	1.84 (0.94)	0.89 (0.36)	0.00
1.5	0	16.00 (7.60)	0.17 (0.40)	3.13 (1.74)	1.46 (0.57)	0.01
	16	16.24 (7.02)	0.12 (0.33)	3.06 (0.87)	1.45 (0.35)	0.01
	29	17.06 (6.95)	0.16 (0.37)	3.44 (0.98)	1.60 (0.42)	0.01
	41	16.69 (6.66)	0.24 (0.45)	4.20 (1.73)	1.98 (0.73)	0.02

Table 9: Compared to Example 2, the true model is log-linear Weibull: Comparisons between LASSO and elastic-net for variable selection in censored data simulation with grouping effects. Numbers in parentheses are standard deviations.

cr (%)	Method	$\#S$	$\#FN$	L_1 -loss	L_2 -loss	Median of MSE
33	LASSO	18.97(5.74)	3.27 (2.12)	36.60(28.18)	11.18 (11.26)	0.40
	Enet	22.05(7.20)	1.92 (2.69)	24.15(18.36)	7.68 (5.23)	0.13
43	LASSO	20.02 (5.20)	5.49 (2.86)	45.64 (18.85)	14.26 (11.17)	0.63
	Enet	22.78 (6.76)	3.65 (3.58)	39.34 (24.00)	12.96 (12.50)	0.46

5 R code for simulation studies

All the R codes for the simulation studies and data analysis are included in separated files in the Online Supplementary Material.

The R code for Example 1 is given in `Example1-uncensored.R` and `Example1-censored.R` respectively for the uncensored and censored case.

The R code for Example 2 is given in `Example2-Lasso.R` and `Example2-EN.R` respectively for the LASSO and the elastic-net method.

The R code for Example 3 is given in `Example3-Lasso.R` and `Example3-EN.R` respectively for the LASSO and the elastic-net method. .

6 R code for data analysis

6.1 Vessel data

The data set was kindly provided by Professor Maronna. Following Maronna (2011), we consider this data set corresponding to electron-probe X ray microanalysis of archaeological glass vessels (Janssens et al., 1998), where each of $n = 180$ vessels is represented by a spectrum of 1920 frequencies. For each vessel the contents of 13 chemical compounds are registered. Since the frequencies below 15 and above 500 have almost null values of x_{ij} , we keep frequencies 15 to 500, so that we have $p = 486$. This data set does not contain censored observations. We choose the content of the 11th chemical compound as the response T . We can treat T as an event time since it is positive. According to our model, the response Y is a logarithmic transformation of the original chemical compound T .

The original data can be found in the supplementary files `X1.csv` and `Y1.csv` for frequencies and chemical compounds.

The R code for analysis of this dataset is given in `vessel.R`.

6.2 Breast cancer gene expression data

This data set was studied by Sørbye et al. (2003) and contains 549 gene expression measurements, exit time and exit status in a study of breast cancer among 115 women. 77 observa-

tions are censored, resulting in about 67% censoring rate. Using the proposed methodology, we implement SEM algorithm and variable selection by elastic-net simultaneously.

The data set can be found in the R library `ahaz` by the following commands:

```
library(ahaz)
data(sorlie)
```

The R code for analysis of this dataset is given in `breastcancer.R`.

References

- Donoho, D. L. and Johnstone, J. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, **81**(3):452–455.
- Janssens, K. H., Deraedt, I., Schalm, O., and Veeckman, J. (1998). *Composition of 15–17th Century Archaeological Glass Vessels Excavated in Antwerp*. Belgium: Springer.
- Lin, W. and Lv, J. (2013). High-dimensional sparse additive hazards regression. *Journal of the American Statistical Association*, **108**(501):247–264.
- Maronna, R. A. (2011). Robust ridge regression for high-dimensional data. *Technometrics*, **53**(1):44–53.
- Simon, N., Friedman, J. H., Hastie, T., and Tibshirani, R. (2011). Regularization Paths for Cox’s Proportional Hazards Model via Coordinate Descent. *Journal of Statistical Software*, **39**(5):1–13.
- Sørlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., et al. (2003). Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proceedings of the National Academy of Sciences*, **100**(14):8418–8423.