Learning Survival Distributions with the Asymmetric Laplace Distribution

Deming Sheng, Ricardo Henao
Duke University
https://github.com/demingsheng/ALD

❖ Introduction and Background

Survival Data. A survival dataset $\mathcal{D}$ is represented as a set of triplets $\{(x_n, y_n, e_n)\}_{n=1}^{N}$, where $x_n \in \mathbb{R}^d$ denotes the set of covariates in d dimensions, $y_n = \min(o_n, c_n) \in \mathbb{R}_+$ represents the observed time, and e_n is the event indicator. If the event of interest is observed, *e.g.* death, then $o_n < c_n$ and the event indicator is set to $e_n = 1$, otherwise, the event is *censored* and $e_n = 0$.

❖ Introduction and Background

Survival Models. Survival models can be broadly classified into three main categories: *parametric*, *semiparametric* and *nonparametric* models.

- *Parametric* models assume that the survival PDF follows a specific probability distribution, *e.g.*, exponential (*Feigl & Zelen, 1965*), log-normal (*Royston, 2001*) or Weibull distribution (*Scholz & Works, 1996*).
- *Semiparametric* methods, such as the Cox proportional hazards model (*Cox, 1972*), assume a proportional hazards structure without specifying a baseline hazard distribution, which offers robustness and interpretability.
- *Nonparametric* models, such as DeepHit (*Lee et al., 2018*) and CQRNN (*Pearce et al., 2022*) directly modeling conditional distributions by discretizing survival CDFs.

❖ Motivation and Contribution

- *Parametric* and *semiparametric* models deliver smooth survival CDFs but falter when data deviates from their strict assumptions, compromising accuracy.
- *Nonparametric* models show their strong performance at quantile regression tasks, but their discrete outputs (point- or step-wise survival CDFs) restrict flexibility.

Thus, our **goal** is to introduce a simple but efficient and flexible method for survival modeling.

❖ Motivation and Contribution

In this paper, our contributions are listed below:

- ➢ We introduce a flexible parametric survival model based on the Asymmetric Laplace Distribution (ALD), which offers superior flexibility in capturing diverse survival patterns compared to other distributions (*parametric* methods).
- ➢ The continuous nature of the ALD-based approach offers great flexibility in summarizing distribution-based predictions, thus addressing the limitations of existing discretized *nonparametric* methods.
- ➢ Experiments on 14 synthetic datasets and 7 real-world datasets in terms of 9 performance metrics demonstrate that our proposed framework consistently outperforms both *parametric* and *nonparametric* approaches in terms of both discrimination and calibration.

❖ Methods

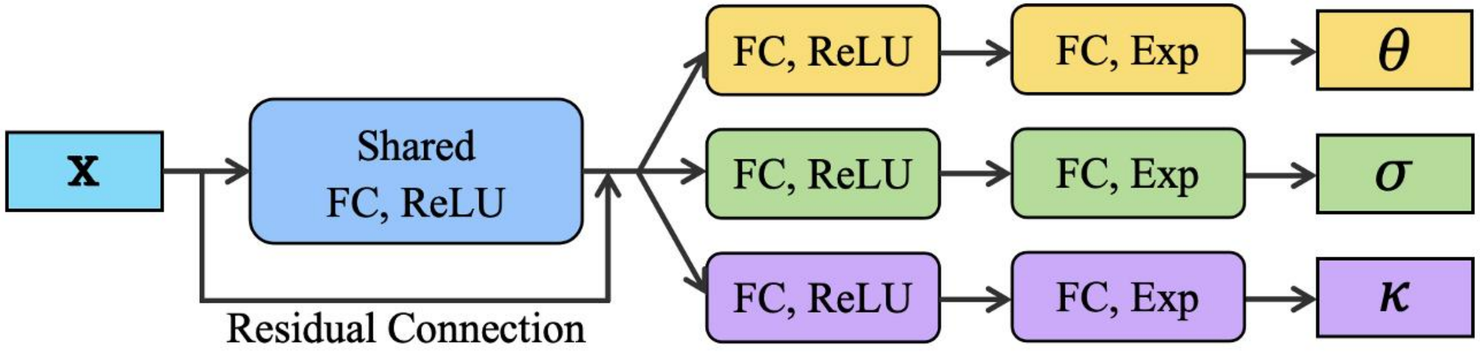
Asymmetric Laplace Distribution (ALD, *Kotz et al, 2012*). A random variable Y is said to have an asymmetric Laplace distribution with parameters (θ, σ, κ), if its PDF is:

$$f_{\text{ALD}}(y;\theta,\sigma,\kappa) = \frac{\sqrt{2}}{\sigma}\frac{\kappa}{1+\kappa^2}\begin{cases}\exp(\frac{\sqrt{2}\kappa}{\sigma}(\theta - y)) & \text{if } y \geq \theta \\ \exp(\frac{\sqrt{2}}{\kappa\sigma}(y-\theta)) & \text{if } y < \theta\end{cases}$$

where θ, $\sigma > 0$, and $\kappa > 0$, are the location, scale and asymmetry parameters. Moreover, its CDF can be expressed as:

$$F_{\text{ALD}}(y;\theta,\sigma,\kappa) = \begin{cases}1 - \frac{\kappa}{1+\kappa^2}\exp(\frac{\sqrt{2}\kappa}{\sigma}(\theta - y)) & \text{if } y \geq \theta \\ \frac{\kappa}{1+\kappa^2}\exp(\frac{\sqrt{2}}{\kappa\sigma}(y-\theta)) & \text{if } y < \theta\end{cases}$$

Network Architecture.



Loss Function.

$$\mathcal{L}_{\text{ALD}} = -\sum_{n \in \mathcal{D}_o} f_{\text{ALD}}(y_n|x_n) - \sum_{n \in \mathcal{D}_c} S_{\text{ALD}}(y_n|x_n)$$

where $f_{\text{ALD}}(\bullet)$ and $S_{\text{ALD}}(\bullet)$ are the PDF and survival function of Asymmetric Laplace Distribution, $\mathcal{D}_o$ and $\mathcal{D}_c$ are the subsets of $\mathcal{D} = \mathcal{D}_o \cup \mathcal{D}_c$ for which $e = 1$ and $e = 0$.

❖ Comparison between our Method and CQRNN

Corollary 3.2. The Asymmetric Laplace Distribution, denoted as $\mathcal{AL}(\theta, \sigma, \kappa)$, can be reparameterized as $\mathcal{AL}(\theta, \sigma, q)$ to facilitate quantile regression (*Yu & Moyeed, 2001*), where $q \in (0, 1)$ is the percentile parameter that represents the desired quantile. The relationship between q and κ is given by $q = \kappa^2/(\kappa^2 + 1)$.

The widely used *pinball* or *checkmark* loss (*Koenker & Bassett Jr, 1978*) in the quantile regression literature (*e.g.*, CQRNN, *Pearce et al., 2022*) is essentially the maximum likelihood estimation of $\mathcal{AL}(\theta, \sigma, q)$ up to a constant. CQRNN optimizes a model with the *pinball* loss to predict θ_q for a predefined collection of quantile values, *e.g.*, $q = \{0.1, 0.2, \ldots, 0.9\}$:

$$\mathcal{L}_{\mathrm{QR}}(y; \theta_q, q) = \begin{cases} q(y - \theta_q) & \text{if } y \geq \theta_q \\ (1-q)(\theta_q - y) & \text{if } y < \theta_q \end{cases} = (y - \theta_q)(q - \mathbb{1}[\theta_q > y])$$

The maximum likelihood estimation of $\mathcal{AL}(\theta_q, \sigma, q)$ is:

$$\log\sigma - \log[\mathrm{q}(1-\mathrm{q})] + \frac{1}{\sigma}\begin{cases} q(y - \theta_q) & \text{if } y \geq \theta_q \\ (1-q)(\theta_q - y) & \text{if } y < \theta_q \end{cases}$$

❖ Experimental Results

Table 2. Summary of benchmarking results across 21 datasets. Each column group shows three figures: the number of datasets where our method significantly outperforms, underperforms or is comparable with the baseline indicated. The last two rows summarize the column totals and proportions to simplify the comparisons. For reference, the total number of comparisons is 189.

Metric	***vs.* CQRNN**			***vs.* LogNorm**			***vs.* DeepSurv**			***vs.* DeepHit**		
	Better	**Worse**	**Same**	**Better**	**Worse**	**Same**	**Better**	**Worse**	**Same**	**Better**	**Worse**	**Same**
MAE	6	8	7	10	3	8	6	8	7	12	6	3
IBS	19	1	1	21	0	0	21	0	0	21	0	0
Harrell's C-Index	4	2	15	10	3	8	6	2	13	15	0	6
Uno's C-Index	2	3	16	9	2	10	6	1	14	15	0	6
CensDcal	8	4	9	10	1	10	8	5	8	15	1	5
Cal $[S(t\|\mathbf{x})]$**(Slope)**	0	0	21	15	0	6	13	0	8	12	0	9
Cal $[S(t\|\mathbf{x})]$**(Intercept)**	0	0	21	14	0	7	0	11	10	16	0	5
Cal $[f(t\|\mathbf{x})]$**(Slope)**	4	0	17	14	0	7	9	0	12	14	0	7
Cal $[f(t\|\mathbf{x})]$**(Intercept)**	0	4	17	10	0	11	8	0	13	18	0	3
Total	43 / 189	22 / 189	124 / 189	113 / 189	9 / 189	67 / 189	77 / 189	27 / 189	85 / 189	138 / 189	7 / 189	44 / 189
Proportion	0.228	0.116	0.656	0.598	0.048	0.354	0.407	0.143	0.450	0.730	0.037	0.233

Metric	***vs.* GBM**			***vs.* RSF**			***vs.* DSM (LogNorm)**			***vs.* DSM (Weibull)**		
	Better	**Worse**	**Same**	**Better**	**Worse**	**Same**	**Better**	**Worse**	**Same**	**Better**	**Worse**	**Same**
MAE	11	7	3	9	6	6	11	6	4	11	5	5
IBS	17	1	3	14	2	5	19	1	1	19	1	1
Harrell's C-Index	14	2	5	16	2	3	17	0	4	15	0	6
Uno's C-Index	13	1	7	14	2	5	16	0	5	14	0	7
CensDcal	0	2	19	6	4	11	21	0	0	21	0	0
Cal $[S(t\|\mathbf{x})]$**(Slope)**	6	0	15	12	0	9	12	0	9	16	0	5
Cal $[S(t\|\mathbf{x})]$**(Intercept)**	12	0	9	10	0	11	13	0	8	17	0	4
Cal $[f(t\|\mathbf{x})]$**(Slope)**	14	0	7	15	0	6	13	0	8	11	0	10
Cal $[f(t\|\mathbf{x})]$**(Intercept)**	11	0	10	8	0	13	12	0	9	6	0	15
Total	98 / 189	13 / 189	78 / 189	104 / 189	16 / 189	69 / 189	134 / 189	7 / 189	48 / 189	130 / 189	6 / 189	53 / 189
Proportion	0.519	0.069	0.413	0.550	0.085	0.365	0.709	0.037	0.254	0.688	0.032	0.280

❖ Code and Reference

Thank you for your attention!

Code: https://github.com/demingsheng/ALD

References

Feigl, P. and Zelen, M. Estimation of exponential survival probabilities with concomitant information. Biometrics, pp. 826–838, 1965.

Royston, P. The lognormal distribution as a model for survival time in cancer, with an emphasis on prognostic factors. Statistica Neerlandica, 55(1):89–104, 2001.

Scholz, F. and Works, B. P. Maximum likelihood estimation for type i censored weibull data including covariates. Technical report, ISSTECH-96-022, Boeing Information and Support Services, 1996.

Cox, D. R. Regression models and life-tables. Journal of the Royal Statistical Society: Series B (Methodological), 34(2):187–202, 1972.

Lee, C., Zame, W., Yoon, J., and Van Der Schaar, M. Deephit: A deep learning approach to survival analysis with competing risks. In Proceedings of the AAAI conference on artificial intelligence, volume 32, 2018.

Pearce, T., Jeong, J.-H., Zhu, J., et al. Censored quantile regression neural networks for distribution-free survival analysis. *Advances in Neural Information Processing Systems*, 35:7450–7461, 2022.

Kotz, S., Kozubowski, T., and Podgorski, K. The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance. Springer Science & Business Media, 2012.

Yu, K. and Moyeed, R. A. Bayesian quantile regression. Statistics & Probability Letters, 54(4):437–447, 2001.

Koenker, R. and Bassett Jr, G. Regression quantiles. Econometrica: journal of the Econometric Society, pp. 33–50, 1978.