Objective Bayesian Analysis of Recall-based Observations with Application to Breastfeeding Data

Vikas Barnwal[1], C P Yadav[2], Dr. M S Panwar[1]

[1] Banaras Hindu University, India
[2] National University of Singapore

Abbreviated abstract: The recall-based studies provide the status of an event of interest at monitoring time along with the time to event when participants are able to recall it. In this work, objective Bayesian analysis has been performed for the recall-based data as it provides additional information related to the population characteristics than current status data. The Bayesian estimation of unknown parameters is obtained under derived reference priors for different loss functions. To elucidate the performance of Bayesian estimators, breastfeeding data from NFHS IV, India is utilized and the established methodology is illustrated.

Related publications:

– Berger, J. O. and Bernardo, J. M. (1992a). "On the development of the reference prior method." Bayesian Statistics, 4(4): 35–60.

– S.M. Salehabadi and D. Sengupta, Nonparametric estimation of time-to-event distribution based on recall data in observational studies, Lifetime Data Anal. 22 (2016), pp. 473–503.

काशी हिन्दू विश्वविद्यालय BANARAS HINDU UNIVERSITY

vikas.barnwal5@bhu.ac.in

4th Conference on Statistics and Data Science
Salvador, Brazil (online)
December 1-3, 2022

Problem under Consideration

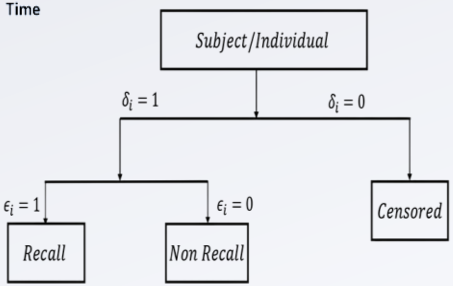
T → time to event, S → monitoring time

The problem arises when

- The respondent is not able to recall time "T" at monitoring time "S"
- The respondent has not been experienced the event at "S"
- Incomplete data leads to computational challenges while estimating the parameters of the population of interest.
- Objective Bayesian is preferred in case no prior information is available for parameters.

Consider the likelihood function for observed data $(t_i, s_i, \delta_i, \epsilon_i)$

$$L(\theta|d) = \prod_{i=1}^{n} \left[f(t_i;\theta)\Big(1-\psi(s_i,t_i)\Big)\right]^{\delta_i\epsilon_i} \left[\int_0^{s_i} f(u;\theta)\psi(s_i,u)du\right]^{\delta_i(1-\epsilon_i)} \left[\bar{F}(s_i;\theta)\right]^{(1-\delta_i)}$$



- The second term containing integral and the last term poses difficulty while optimizing the log-likelihood function to obtain the maximum likelihood (ML) estimator of the parameters θ.
- Previously, Yadav et al. (2022) analyzed the recall-based data with nested E-M algorithm and under Bayesian paradigm with informative priors.

vikas.barnwal5@bhu.ac.in

Established Methodology

- Assume $T \sim W(\alpha, \beta)$ and
- Recall probability, $\varphi(.)$ → exponential function of difference between S and T.

The likelihood function for Θ given complete data Y=(D, D*), can be rewritten as

$$L_Y(\Theta|d,d^*) = \prod_{i=1}^{n_r}\left[\alpha\beta t_i^{\alpha-1}\exp\{-\beta t_i^{\alpha}\}\exp\{-\lambda(s_i - t_i)\}\right]\prod_{i=1}^{n_{nr}}\left[\alpha\beta {t_{li}^*}^{\alpha-1}\exp\{-\beta {t_{li}^*}^{\alpha}\}\lambda\exp\{-\lambda w_{li}^*\}\right]\prod_{i=1}^{n_c}\left[\alpha\beta {t_{ri}^*}^{\alpha-1}\exp\{-\beta {t_{ri}^*}^{\alpha}\}\right]$$

Reference prior: $\pi^R(\Theta) \propto (\alpha\beta\lambda)^{-1} T(\beta)^c$

Posterior density: $\Pi^R(\Theta \mid d, d^*) \propto T(\beta)^c \beta^{n-1}\alpha^{n-1}\lambda^{n_{nr}-1}\left(\prod_{i=1}^{n_r} t_i^{\alpha-1}\right)\left(\prod_{i=1}^{n_{nr}} {t_{li}^*}^{\alpha-1}\right)\left(\prod_{i=1}^{n_c} {t_{ri}^*}^{\alpha-1}\right)$

$$\exp\left\{-\beta\left(\sum_{i=1}^{n_r} t_i^{\alpha} + \sum_{i=1}^{n_{nr}} {t_{li}^*}^{\alpha} + \sum_{i=1}^{n_c} {t_{ri}^*}^{\alpha}\right)\right\}\exp\left\{-\lambda\left(\sum_{i=1}^{n_r}(s_i - t_i) + \sum_{i=1}^{n_{nr}} w_{li}^*\right)\right\}$$

Grouping order	Reference prior	c
$\{\alpha,\beta,\lambda\}$	$\pi_1^R(\Theta)$	0
$\{(\alpha,\beta),\lambda\}$	$\pi_1^R(\Theta)$	0
$\{\lambda,(\beta,\alpha)\}$	$\pi_1^R(\Theta)$	0
$\{\beta,\alpha,\lambda\}$	$\pi_2^R(\Theta)$	-1/2
$\{(\beta,\alpha),\lambda\}$	$\pi_3^R(\Theta)$	1

- The samples are generated using MCMC techniques from the conditional posterior of the respective parameter.

Data Analysis & Results

- National family of health survey round four (NFHS-IV) data is utilized.
- The duration of breastfeeding for recent child is estimated using the established methodology under the objective Bayesian paradigm.
- The estimated median duration is 25.47 months when we fixed the non-recall observations at 5%.
- The estimation is carried out for different level of non-recall and censored proportion.

$\pi_1^R(\theta)$					SELF	LINEX		95% HPD	
Level	n_r	n_{nr}	n_c			c=1.2	c=-1.5	Lower	Upper
I	45	5	50	α	1.8043	1.8032	1.8058	1.7166	1.8907
				β	0.6219	0.6209	0.6231	0.5434	0.7040
				λ	0.1204	0.1201	0.1209	0.0737	0.1680
II	40	10	50	α	1.8002	1.7990	1.8016	1.7116	1.8841
				β	0.6404	0.6393	0.6417	0.5616	0.7240
				λ	0.2606	0.2598	0.2616	0.1950	0.3381
III	35	15	50	α	1.7914	1.7903	1.7929	1.7068	1.8798
				β	0.6453	0.6442	0.6466	0.5609	0.7237
				λ	0.4240	0.4226	0.4258	0.3273	0.5167

$\pi_2^R(\theta)$					SELF	LINEX		95% HPD	
Level	n_r	n_{nr}	n_c			c=1.2	c=-1.5	Lower	Upper
I	45	5	50	α	1.8059	1.8048	1.8074	1.7216	1.8902
				β	0.6217	0.6207	0.6229	0.5448	0.7019
				λ	0.1209	0.1206	0.1213	0.0751	0.1660
II	40	10	50	α	1.7987	1.7976	1.8001	1.7062	1.8806
				β	0.6400	0.6390	0.6413	0.5638	0.7251
				λ	0.2606	0.2598	0.2616	0.1922	0.3346
III	35	15	50	α	1.7913	1.7901	1.7928	1.7031	1.8760
				β	0.6447	0.6436	0.6461	0.5616	0.7279
				λ	0.4226	0.4212	0.4244	0.3334	0.5240

- If interest lies in the estimates of median time to breastfeeding of the recent child, one can go with $\pi_1^R(\theta)$ which gives more weight to the shape and scale parameters equally.

काशी हिन्दू विश्वविद्यालय BANARAS HINDU UNIVERSITY

vikas.barnwal5@bhu.ac.in