Robust Bayesian Inference for Postpartum Amenorrhea Duration Using Weibull Models

C P Yadav[1], Sachin Kumar[2], Anup Kumar[2]

[1]Pfizer Healthcare India Pvt.Ltd., Chennai, Tamil Nadu, India
[2]Department of Biostatistics and Health Informatics, SGPGIMS, Lucknow, India

Introduction

Background:

- The duration of postpartum amenorrhea (PPA) plays a critical role in reproductive health and family planning, directly impacting birth spacing decisions.
- Recall bias in PPA studies occurs due to retrospective reporting, with durations often clustered at intervals like 3, 6, or 12 months, leading to inaccuracies.
- Current status data addresses this limitation by utilizing real-time information, enhancing the precision and reliability of PPA duration estimates.

Objective:

- To develop and apply robust statistical methods, informative and non-informative Bayesian estimation, for analyzing PPA duration using current status data.

Methodology

Data Structure and Likelihood:

- Each individual has a random monitoring time C and event status δ where $\delta = 1$ (left-censored, $T_i \leq C_i$) if the event occurs before C, and $\delta = 0$ (right-censored, $T_i > C_i$) if the event occurs after C.
- The likelihood function for the parameter θ under this data setup is:

$$L(\theta) = \prod_{i=1}^{n} [F(C_i, \theta)]^{\delta_i} [S(C_i, \theta)]^{1-\delta_i} \qquad (1)$$

- The observed dataset for n individuals is $\{(C_i, \delta_i)\}_{i=1}^{n}$, where $n_l = \sum_{i=1}^{n} \delta_i$ are left-censored and $(n - n_l)$ right-censored observations.
- Assume $T \sim W(\alpha, \beta)$ with density and survival functions as $f(t|\alpha, \beta) = \beta\alpha t^{\alpha-1} e^{-\beta t^{\alpha}}$, and $S(t|\alpha, \beta) = e^{-\beta t^{\alpha}}$ respectively.

Bayesian Inference

- We assume the following independent gamma priors on the model parameters to obtain the Bayes estimator:

$$\pi_1(\alpha) = \frac{\nu_1^{\mu_1}}{\Gamma(\mu_1)} \alpha^{\mu_1 - 1} e^{-\nu_1 \alpha}; \quad \nu_1, \mu_1, \alpha > 0,$$
$$\pi_2(\beta) = \frac{\nu_2^{\mu_2}}{\Gamma(\mu_2)} \beta^{\mu_2 - 1} e^{-\nu_2 \alpha}; \quad \nu_2, \mu_2, \beta > 0$$

- We applied the augmentation approach that simplifies posterior inference by introducing complete likelihood based on latent variables, enabling efficient sampling with Gibbs sampling.
- We introduce two latent variables for the left-censored and right-censored categories as T_l^* and T_r^*.
- The full conditionals of α and β are obtained as:

$$\pi_1(\alpha|\beta) \propto \alpha^{n+\mu_1-1} \left(\prod_{i=1}^{n_l} t_l^{*\alpha-1} \right) \left(\prod_{i=n_l+1}^{n} t_r^{*\alpha-1} \right) \times \exp\left\{ -\beta \left(\sum_{i=1}^{n_l} t_l^* + \sum_{i=n_l+1}^{n} t_r^* + \nu_1 \right) \right\}, \qquad (2)$$

$$\pi_2(\beta|\alpha) \sim G\left(n + \mu_2, \sum_{i=1}^{n_l} t_l^* + \sum_{i=n_l+1}^{n} t_r^* + \nu_2 \right), \qquad (3)$$

- Posterior samples were generated using nested Gibbs sampling, incorporating Metropolis-Hastings steps, as described by [3] and [1].
- The sampling process utilized observed data along with given values of parameters and observations on latent variables to estimate the posterior distribution.
- Informative and non-informative scenarios were addressed in the simulation study.

Simulation Study

- T follows a Weibull ($\alpha = 1.35, \beta = 1.85$) and C are generated from uniform ($U[0.2, 2]$) distribution.
- We simulate observations for $n = 30, 50, 80$ and maintain different proportions of left- and right-censored observations at 50% each.
- For informative prior, we set $\mu_1 = 2, \nu_1 = 3, \mu_2 = 2, \nu_2 = 3$ and non-informative prior, we set $\mu_1 = 0.001, \nu_1 = 0.001, \mu_2 = 0.001, \nu_2 = 0.001$.
- Generated 10^5 posterior samples, with a thinning interval of 20 and an initial burn-in of 20,000 samples.

Table 1. Parameter estimates under informative prior for different sample sizes.

	n=30	n=50	n=80
Point estimates			
$\alpha\ (SELF)$	1.26	1.26	1.26
$\beta\ (SELF)$	1.54	1.62	1.63
$\alpha\ (a=1)$	1.24	1.24	1.25
$\beta\ (a=1)$	1.46	1.56	1.59
$\alpha\ (a=-1)$	1.27	1.27	1.28
$\beta\ (a=-1)$	1.62	1.68	1.67
Mean Squared Error (MSE)			
$\alpha\ (SELF)$	0.01	0.01	0.01
$\beta\ (SELF)$	0.20	0.13	0.12
$\alpha\ (a=1)$	0.01	0.01	0.01
$\beta\ (a=1)$	0.24	0.15	0.13
$\alpha\ (a=-1)$	0.01	0.01	0.01
$\beta\ (a=-1)$	0.18	0.12	0.11
Bias			
$\alpha\ (SELF)$	-0.09	-0.09	-0.09
$\beta\ (SELF)$	-0.31	-0.23	-0.22
$\alpha\ (a=1)$	-0.11	-0.11	-0.10
$\beta\ (a=1)$	-0.39	-0.29	-0.26
$\alpha\ (a=-1)$	-0.08	-0.08	-0.07
$\beta\ (a=-1)$	-0.23	-0.17	-0.18
HPD interval			
$\alpha\ (AL)$	0.69	0.68	0.66
$\beta\ (AL)$	1.48	1.30	1.11
$\alpha\ (CP)$	1.00	1.00	1.00
$\beta\ (CP)$	0.83	0.89	0.86

Table 2. Parameter estimates under non-informative prior for different sample sizes.

	n=30	n=50	n=80
Point estimates			
$\alpha\ (SELF)$	1.30	1.29	1.30
$\beta\ (SELF)$	1.91	1.81	1.86
$\alpha\ (a=1)$	1.28	1.28	1.29
$\beta\ (a=1)$	1.77	1.73	1.80
$\alpha\ (a=-1)$	1.32	1.31	1.32
$\beta\ (a=-1)$	2.11	1.91	1.93
Mean Squared Error (MSE)			
$\alpha\ (SELF)$	0.00	0.01	0.01
$\beta\ (SELF)$	0.33	0.13	0.08
$\alpha\ (a=1)$	0.01	0.01	0.01
$\beta\ (a=1)$	0.26	0.13	0.08
$\alpha\ (a=-1)$	0.00	0.00	0.01
$\beta\ (a=-1)$	0.56	0.17	0.10
Bias			
$\alpha\ (SELF)$	0.06	0.07	0.07
$\beta\ (SELF)$	0.47	0.29	0.23
$\alpha\ (a=1)$	-0.07	-0.07	-0.06
$\beta\ (a=1)$	-0.08	-0.12	-0.05
$\alpha\ (a=-1)$	-0.03	-0.04	-0.03
$\beta\ (a=-1)$	0.26	0.06	0.08
HPD interval			
$\alpha\ (AL)$	0.74	0.72	0.70
$\beta\ (AL)$	2.06	1.58	1.35
$\alpha\ (CP)$	1.00	1.00	1.00
$\beta\ (CP)$	0.93	0.97	0.97

- Informative priors reduce bias, MSE, and uncertainty, improving estimates, especially with smaller sample sizes.
- Non-informative priors increase bias, MSE, and uncertainty, but their impact lessens as sample size grows.
- SELF shows stable estimates with low MSE and bias, improving with sample size.
- LINEX (with a=1 and a=-1) benefits from informative priors, which reduce MSE and bias compared to non-informative priors.

Real Data Analysis

- Data Source: A subset of the NFHS-V (2020-21) dataset, specifically focusing on the duration of postpartum amenorrhea (PPA) in the district of Budaun, Uttar Pradesh, is used for the analysis [2].
- Women who had not experienced PPA at the interview were treated as right-censored, while those who had were treated as left-censored, aligning with the analysis framework.
- Key Variables: Interview date, date of last birth, and PPA indicator, with analysis restricted to women having an open birth interval (OBI) of less than 60 months for relevant and recent PPA observations.
- Out of 380 women, 327 were in the left-censored category while 53 were in the right-censored category.
- Hyperparameters: We calculated informative hyperparameters using the mean and sd of the data while in case of non-informative we set the same as in simulation.

Table 3. Point and interval estimates using NFHS data.

	Non-Informative	Informative
Point estimates		
$\alpha\ (SELF)$	0.50	0.51
$\beta\ (SELF)$	0.56	0.56
$\alpha\ (a=1)$	0.50	0.51
$\beta\ (a=1)$	0.56	0.56
$\alpha\ (a=-1)$	0.50	0.51
$\beta\ (a=-1)$	0.56	0.56
HPD		
α	(0.46, 0.55)	(0.46, 0.55)
β	(0.45, 0.68)	(0.45, 0.68)
$mean\ duration(mo)$	3.28 (2.53, 4.42)	3.28 (2.58, 4.46)

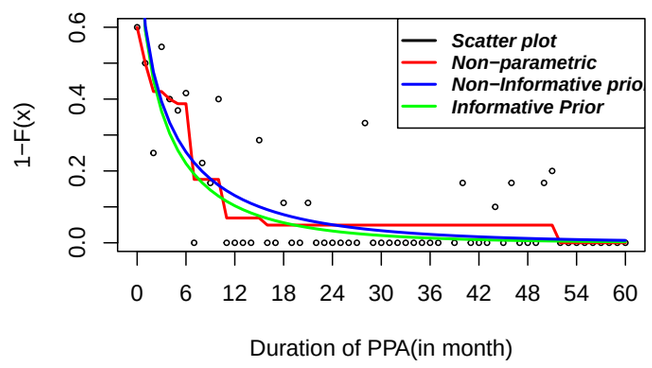

Figure 1. Survival plots for the duration of PPA using different methods.

- The results show consistent estimates and intervals across both priors, indicating robust parameter estimation.
- The survival plot (Figure 1) shows smoother estimates for Bayesian methods, with slight regularization under the informative prior.

Conclusion

This study demonstrates that Bayesian estimation with current status data provides reliable and precise estimates of PPA duration, offering valuable insights for reproductive health and family planning.

References

[1] W Keith Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

[2] International Institute for Population Sciences (IIPS) and ICF. National Family Health Survey (NFHS-5). **https://dhsprogram.com/publications/publication-fr347-dhs-final-reports.cfm**, 2020. Accessed: [date accessed].

[3] Nicholas Metropolis and Stanislaw Ulam. The Monte Carlo method. *Journal of the American Statistical Association*, 44(247):335–341, 1949.