BayesPPDSurv: an R Package for Bayesian Sample Size Determination Using the Power Prior for Survival Data

Joseph Ibrahim

September 21, 2023

University of North Carolina at Chapel Hill

*Joint work with Yueqi Shen and Matthew Psioda

Table of contents

Introduction

Introduction: Bayesian Survival Analysis

- One of the most convenient and popular models for semiparametric survival analysis is the piecewise constant hazard model (Ibrahim et al., 2001).
- Semiparametric Bayesian survival analysis first involves a discretization of the time axis.
- Then over each time interval, a parametric model is specified.
- In BayesPPDSurv, we use the stratified Cox model with piecewise constant baseline hazard, which is a common approach for Bayesian analysis of time-to-event data.

Introduction: Power Prior

- There has been increasing interest over the past few decades in incorporating historical data in clinical trials, particularly on controls.
- Bayesian methods provide a natural mechanism for information borrowing through the use of informative priors.
- The power prior (Chen and Ibrahim (2000)) has become popular due to its easy construction, its natural way of incorporating historical data and its intuitive interpretation.

Introduction: Normalized Power Prior

- Duan et al. (2006) propose a modification of the power prior, the normalized power prior, which adds a normalizing constant component when the power parameter is modeled as random.
- In this talk, we will focus on the case where a_0 is fixed.

Introduction: Bayesian Sample Size Determination

- There is a growing literature on Bayesian sample size determination, including the works of Wang and Gelfand (2002) and Joseph et al. (2008).
- We consider the simulation-based method developed in Chen et al. (2011) and Psioda and Ibrahim (2019), which extends the the fitting and sampling priors of Wang and Gelfand (2002) with a focus on controlling the type I error rate and calculating power.

Introduction: BayesPPD

- We developed R package BayesPPD (Shen et al., 2023), published on CRAN in 2021.
- BayesPPD (Bayesian Power Prior Design) supports Bayesian clinical trial design after incorporating historical data with the power prior and the normalized power prior for Bernoulli, normal, Poisson and Exponential outcomes.
- We aim to develop an extension package that accommodates time-to-event outcomes.

Methods Implemented

Power Prior

- Let D denote data from the current study and D_0 denote data from a historical study. Let θ denote model parameters.
- The power prior (Chen and Ibrahim (2000)) is formulated as

$$\pi(\theta|D_0, a_0) \propto L(\theta|D_0)^{a_0}\pi_0(\theta),$$

  where $0 \leq a_0 \leq 1$ is a discounting parameter for the historical data likelihood, and $\pi_0(\theta)$ is the initial prior for θ.
- Suppose there are K historical datasets denoted by D_{0k} for $k = 1, \cdots, K$ and let $D_0 = (D_{01}, \cdots, D_{0K})$. The power prior becomes

$$\pi(\theta|D_0, a_0) \propto \prod_{k=1}^{K} L(\theta|D_{0k})^{a_{0k}}\pi_0(\theta)$$

  where $a_0 = (a_{01}, \cdots, a_{0K})'$ and $0 \leq a_{0k} \leq 1$ for $k = 1, \cdots, K$.

Piecewise Constant Hazard Model

- We implement the stratified Cox model with piecewise constant baseline hazard.
- Let $s = 1, 2, \ldots, S$ be the stratum index.
- For stratum s, let $\lambda_s(t)$ denote the piecewise constant baseline hazard and $\Lambda_s(t)$ denote the baseline cumulative hazard.

Piecewise Constant Hazard Model

- Let t_i denote the time to event and c_i denote the time to censorship for subject i, $i = 1, \ldots, n$.
- Let $\nu_i = I[t_i \leq c_i]$ denote the indicator that an event is observed for subject i.
- Let $y_i = min(t_i, c_i)$ denote the observed survival time for subject i.
- Let x_i denote the covariates for subject i and β denote the regression coefficients.
- If there are no covariates, x_i is simply the treatment indicator.
- Then the likelihood of a stratified Cox model is

$$\prod_{i=1}^{n} \{\lambda_{si}(y_i) \exp(x_i'\beta)\}^{\nu_i} \exp\{-\Lambda_{si}(y_i) \exp(x_i'\beta)\}.$$

Piecewise Constant Hazard Model

- We allow the baseline hazard and time interval partition to vary across S levels of a stratification variable.
- For stratum s, we partition time into $k = 1, \ldots, K_s$ intervals with change points $0 = t_{s,0} < t_{s,1} < \cdots < t_{s,K_s} = \infty$.
- Let $\lambda_{sk} > 0$ denote the constant hazard over interval $I_{s,k} = (t_{s,k-1}, t_{s,k}]$.

Piecewise Constant Hazard Model

- Let ν_{ik} denote the indicator that the event occurred in interval $I_{s_i,k}$.
- Let r_{ik} denote the subject's time at risk in interval $I_{s_i,k}$.
- Let G_s denote the set of indices corresponding to subjects from stratum s.
- Then the likelihood of a piecewise constant model is

$$\prod_{s=1}^{S}\prod_{k=1}^{K_s}\lambda_{sk}^{\sum_{i\in G_s}\nu_{ik}}\exp\left\{-\lambda_{sk}\left(\sum_{i\in G_s}\exp(x_i'\beta)r_{ik}\right)\right\}\times\prod_{i=1}^{n}\exp(x_i'\beta)^{\nu_i}.$$

Piecewise Constant Hazard Model

Analogously, the likelihood for historical datasets $j = 1, \ldots, J$, each raised to a power of a_{0j}, is

$$\prod_{j=1}^{J} L(\beta, \lambda | D_{0j})^{a_{0j}}$$

$$= \prod_{j=1}^{J} \prod_{s=1}^{S} \prod_{k=1}^{K_s} \lambda_{sk}^{a_{0j}\left(\sum_{i \in G_{sj}} \nu_{ik}\right)} \exp\left\{ -a_{0j} \lambda_{sk} \left(\sum_{i \in G_{sj}} \exp(x_i' \beta) r_{ik} \right) \right\}$$

$$\times \prod_{j=1}^{J} \prod_{i=1}^{n_j} \exp(x_i' \beta)^{\nu_i a_{0j}}$$

- Note that in this formulation, the baseline line hazard parameters for the current and historical data are the same. If one does not want to borrow information on the baseline hazards, then λ_{sk} is replaced by λ_{0sk}.

Piecewise Constant Hazard Model

- For the prior on β, we allow choices of
  - Uniform improper priors
  - Independent normal priors
- For the prior on λ_{sk}, we allow choices of
  - Independent Gamma priors
  - Independent normal priors on $\log(\lambda_{sk})$
  - Improper priors $\pi(\lambda_{sk}) \sim \lambda_{sk}^{-1}$

Bayesian Sample Size Determination

- We assume the parameter for the treatment indicator is β_1. The power / type I error calculation algorithm by default assumes the null and alternative hypotheses are given by

$$H_0 : \beta_1 \geq \delta \text{ and } H_1 : \beta_1 < \delta,$$

  where δ is a prespecified constant.

Bayesian Sample Size Determination

- Following Chen et al. (2011), let Θ_0 and Θ_1 denote the parameter spaces corresponding to H_0 and H_1.
- Let $y^{(n)}$ denote the simulated current data associated with a sample size of n and let θ denote the model parameters.
- Let $\pi^{(s)}(\theta)$ denote the sampling prior and let $\pi^{(f)}(\theta)$ denote the fitting prior. The sampling prior is used to generate the hypothetical data while the fitting prior is used to fit the model after the data is generated.
- Let $\pi_0^{(s)}(\theta)$ denote a sampling prior that only puts mass in the null region, i.e., $\theta \subset \Theta_0$. Let $\pi_1^{(s)}(\theta)$ denote a sampling prior that only puts mass in the alternative region, i.e., $\theta \subset \Theta_1$.

Bayesian Sample Size Determination

- To determine Bayesian sample size, we estimate the quantity

$$q_{sj}^{(n)} = E_s[I\{P(\beta_1 < \delta | y^{(n)}, \pi^{(f)}) \geq \gamma\}]$$

  where $j = 0$ or 1, corresponding to the expectation taken with respect to $\pi_0^{(s)}(\theta)$ or $\pi_1^{(s)}(\theta)$.
- The constant $\gamma > 0$ is a prespecified posterior probability threshold for rejecting the null hypothesis (e.g., 0.975).
- The probability is computed with respect to the posterior distribution given the simulated data $y^{(n)}$ and the fitting prior $\pi^{(f)}(\theta)$, and the expectation is taken with respect to the marginal distribution of $y^{(n)}$ defined based on the sampling prior $\pi^{(s)}(\theta)$.
- Then $q_{s0}^{(n)}$ corresponding to $\pi^{(s)}(\theta) = \pi_0^{(s)}(\theta)$ is the Bayesian type I error rate, while $q^{(n)}$ corresponding to $\pi^{(s)}(\theta) = \pi_1^{(s)}(\theta)$ is the Bayesian power.

Bayesian Sample Size Determination

- For given $\alpha_0 > 0$ and $\alpha_1 > 0$, we can compute $n_{\alpha_0} = \min\{n : q_{s0}^{(n)} \leq \alpha_0\}$ and $n_{\alpha_1} = \min\{n : q_{s1}^{(n)} \geq 1 - \alpha_1\}$. Then, the sample size is taken to be $\max\{n_{\alpha_0}, n_{\alpha_1}\}$.
- Common choices of α_0 and α_1 include $\alpha_0 = 0.05$ and $\alpha_1 = 0.2$. These choices guarantee that the Bayesian type I error rate is at most 0.05 and the Bayesian power is at least 0.8.

Bayesian Sample Size Determination

- A simulation-based procedure is used to estimate the Bayesian type I error rate and power.
- Let N denote the number of simulated trials. To compute $q_{sj}^{(n)}$, the following algorithm is used for each simulated trial b:
  - Step 1: Generate $\theta^{(b)} \sim \pi_j^{(s)}(\theta)$.
  - Step 2: Generate $y^{(b)} \sim f(y^{(b)}|\theta^{(b)})$.
  - Step 3: Estimate the posterior distribution $\pi(\theta|y^{(b)}, D_0, a_0)$ and the posterior probability $P(\beta_1 < \delta|y^{(b)}, \pi^{(f)}, D_0, a_0)$.
  - Step 4: Compute the indicator $r^{(b)} = I\{P(\beta_1 < \delta|y^{(b)}, \pi^{(f)}, D_0, a_0) \geq \gamma\}$.
- Then the estimate of $q_{sj}^{(n)}$ is $\frac{1}{N}\sum_{b=1}^{N} r^{(b)}$.

Sampling Priors

- Our implementation does not assume any particular distribution for the sampling priors.
- The user specifies discrete approximations of the sampling priors by providing a vector or a matrix of sample values and the algorithm samples with replacement from the vector or the matrix as the first step of data generation.
- For example, suppose one wants to compute the power for the hypotheses $H_0 : \beta_1 \geq 0$ and $H_1 : \beta_1 < 0$.
- To approximate the sampling prior for β_1, one can simply sample from a truncated normal distribution with negative mean, so that the mass of the prior falls in the alternative space.
- Conversely, to compute the type I error rate, one can sample from a truncated normal distribution with positive mean, so that the mass of the prior falls in the null space.

Using the Package

Functions Available

- pch.fixed.a0()
  - returns posterior samples of the parameters
- power.pch.fixed.a0()
  - returns Bayesian power/type I error
- Functions that allow a_0 to be modeled as random using the normalized power prior are under development.

Features

- The user can specify the change points. By default, we assign the change points so that the same number of events are observed in all the intervals.
- The baseline hazard parameters can be shared or unshared between the current and historical data. If shared, historical information is borrowed for the baseline hazard parameters as well.
- The distribution of enrollment times can be uniform or exponential.
- The distribution of censorship times can be uniform, exponential or constant.
- The user can specify the probability of subjects dropping out of the study (non-administrative censoring).
- The user can specify the minimum and maximum amount of time that subjects are followed up for.

Case Study: Melanoma Trial Design

Case Study: Melanoma Trial Design

- We replicate the melanoma trial design application in Psioda and Ibrahim (2019).
- Interferon Alpha-2b (IFN) is an adjuvant chemotherapy for deep primary or regionally metastatic melanoma.
- The E1684 trial was a randomized controlled trial conducted to assess the utility of INF as an adjuvant therapy following surgery for melanoma.
- A detailed analysis of the trial was given by Kirkwood and others (1996).

Case Study: Melanoma Trial Design

- The primary outcome is relapse-free survival.
- The covariate is receiving the INF treatment vs. the standard of care (SOC).
- The number of positive lymph nodes was used as a stratification variable (≤ 1 vs. ≥ 2) due to its prognostic value.

Case Study: Melanoma Trial Design

Treatment	# Nodes	Sample Size	# Events	Risk Time
SOC	≤ 1	37	26	88.4
	≥ 2	47	36	105.8
INF	≤ 1	43	21	176.3
	≥ 2	39	31	81.1

Table 1: Relapse-free survival data by treatment group and number of positive nodes at lymphadenectomy. The number of positive lymph nodes was used as a stratification variable.

Case Study: Melanoma Trial Design

- Following Psioda and Ibrahim (2019), we assume a model with a common treatment effect and stratum-specific baseline hazard.
- For stratum 1 (# nodes ≤ 1), we use four time intervals (selected using the deviance information criterion (DIC)). The baseline hazard parameters are $\lambda_{1,1}$, $\lambda_{1,2}$, $\lambda_{1,3}$, $\lambda_{1,4}$. For stratum 2 (# nodes ≥ 2), we use three time intervals. The baseline hazard parameters are $\lambda_{2,1}$, $\lambda_{2,2}$, $\lambda_{2,3}$.
- The change points are chosen so that equal number of events are observed in each time interval.
- A non-informative normal prior is used for the treatment effect β (mean zero and variance 10^5).
- An independent non-informative gamma prior is used for each baseline hazard parameter (shape and rate parameters equal to 10^{-5}).

Case Study: Melanoma Trial Design

- We first obtain the posterior summary of the E1684 trial using the pch.fixed.a0() function.
- We can see from the HPD interval that the treatment is efficacious but the evidence is not strong by common criteria.

Parameter	Mean	SD	95% HPD	$t_{s,k}$
β	-0.27	0.19	(-0.64, 0.10)	n/a
$\lambda_{1,1}$	0.48	0.22	(0.11, 0.90)	0.15
$\lambda_{1,2}$	1.08	0.27	(0.60, 1.62)	0.46
$\lambda_{1,3}$	0.27	0.07	(0.14, 0.40)	2.30
$\lambda_{1,4}$	0.03	0.02	(0.01, 0.07)	∞
$\lambda_{2,1}$	1.23	0.21	(0.83, 1.66)	0.60
$\lambda_{2,2}$	0.30	0.07	(0.17, 0.44)	2.87
$\lambda_{2,3}$	0.09	0.04	(0.03, 0.16)	∞

Table 2: Posterior mean, SD, and 95% highest posterior density (HPD) interval. $\lambda_{s,k}$ = baseline hazard for time interval k for stratum s. $t_{s,k}$ = right endpoint for time interval k for stratum s.

Case Study: Melanoma Trial Design

- Next, we design a new trial using the E1684 data.
- Let ν be the number of events at which the new trial will stop and n be the total number of subjects enrolled.
- For each ν, we take $n = 3\nu$.
- We assume uniform accrual of subjects over a 4-year period with no censoring other than administrative censoring that occur when the planned number of events is reached.
- Subjects are allocated to strata in proportions similar to the historical trial (i.e., around 50% to stratum one) and balanced randomization is used.

Case Study: Melanoma Trial Design

- Next, we design a new trial using the E1684 data.
- To obtain type I error rate, we use a point mass sampling prior at $\beta_1 = 0$. To obtain power, we use a point mass sampling prior at $\beta_1 = -0.27$ (posterior mean of β_1 of the E1684 data).
- We use the power.pch.fixed.a0() function to obtain the following table of tyep I error rates and powers for a_0 values of 0, 0.2 and 0.6.

Case Study: Melanoma Trial Design

a_0	Number of events	Type I error rate	Power
0	350	0.0239	0.3145
	710	0.0226	0.5682
0.2	350	0.0298	0.3545
	710	0.0349	0.5983
0.6	350	0.0508	0.4435
	710	0.0474	0.6851

Table 3: Bayesian type I error rate and power using point mass sampling priors.

Discussion

- BayesPPDSurv implements the piecewise constant hazard model for survival data using the power prior for Bayesian clinical trial design.
- It includes functions for both trial analysis and design.
- The final package will implement the normalized power prior when a_0 is modeled as random.
- The beta version is available on Github at https://github.com/angieshen6/bayesppdsurv

Thank you !


We're on it!
UNC GILLINGS SCHOOL OF GLOBAL PUBLIC HEALTH
EST. 1940

References

1. Ibrahim, J. G. and Chen, M.-H. (2000). "Power Prior Distributions for Regression Models." Statistical Science, 15(1), 46–60. doi:10.1214/ss/1009212673.

2. Shen, Y., Psioda, M. A., and Ibrahim, J. G. (2023). "BayesPPD: An R Package for Bayesian Sample Size Determination Using the Power and Normalized Power Prior for Generalized Linear Models." The R Journal, 14(4), 335-351. doi:10.32614/RJ-2023-016.

3. Psioda, M. A. and Ibrahim, J. G. (2019). "Bayesian clinical trial design using historical data that inform the treatment effect." Biostatistics, 20(3), 400–415.