Survival Analysis

General Principles

Survival analysis studies the time until an event of interest (e.g., death, recovery, information acquisition) occurs. When analyzing binary survival outcomes (e.g., alive or dead), we can use models such as Cox proportional hazards to evaluate the effect of predictors on survival probabilities.

Key concepts include:

1. **Hazard Function**: The instantaneous risk of the event occurring at a given time.
2. **Survival Function**: The probability of surviving beyond a given time.
3. **Covariates**: Variables (e.g., age, treatment) that may affect survival probabilities.
4. **Baseline Hazard**: The hazard when all covariates are zero, which forms the reference for comparing different conditions.

Considerations

> 🔥 Caution
>
> - Bayesian models provide a framework to account for uncertainty in parameter estimates through posterior distributions. You will need to define prior distributions for all model parameters, such as baseline hazard, covariate effects, and variance terms.
>
> - In survival analysis:
>
>   - The **baseline hazard** can follow distributions like Exponential, Weibull, or Gompertz, depending on the data.
>   - Censoring (when the event is not observed for some subjects) must be accounted for in the likelihood function. Proper handling is essential for unbiased results.
>
> - Bayesian survival models allow flexible handling of time-dependent covariates, ran-

> dom effects, and incorporate uncertainty more naturally than Frequentist methods.

Example

Here's an example of a Bayesian survival analysis using the **Bayesian Inference (BI)** package. The data come from a clinical trial of mastectomy for breast cancer. The goal is to estimate the effect of the `metastasized` covariate, coded as 0 (no metastasis) and 1 (metastasis), on the survival outcome `event` for each patient. Time is continuous and censoring is indicated by the event variable.

Python

```python
from BI import bi

# Setup device------------------------------------------
m = bi(platform='cpu')

# Import Data & Data Manipulation ---------------------------------------------
# Import
from importlib.resources import files
data_path = files('BI.resources.data') / 'mastectomy.csv'
m.data(data_path, sep=',')

m.df.metastasized = (m.df.metastasized == "yes").astype(np.int64)
m.df.event = jnp.array(m.df.event.values, dtype=jnp.int32)

## Create survival object
m.surv_object(time='time', event='event', cov='metastasized', interval_length=3)

# Plot censoring -----------------------------------------------
m.plot_censoring(cov='metastasized')

# Model ------------------------------------------------
def model(intervals, death, metastasized, exposure):
    # Parameter prior distributions-------------------------
    ## Base hazard distribution
    lambda0 = m.dist.gamma(0.01, 0.01, shape= intervals.shape, name = 'lambda0')
    ## Covariate effect distribution
    beta = m.dist.normal(0, 1000, shape = (1,),  name='beta')
    ### Likelihood
```

```
    #### Compute hazard rate based on covariate effect
    lambda_ = m.hazard_rate(cov = metastasized, beta = beta, lambda0 = lambda0)
    #### Compute exposure rates
    mu = exposure * lambda_

    # Likelihood calculation
    y = m.poisson(mu + jnp.finfo(mu.dtype).tiny, obs = death)

# Run mcmc -------------------------------------------------
m.fit(model, num_samples=500)

# Summary --------------------------------------------------
print(m.summary())

# Plot hazards and survival function --------------------------------------------------------
m.plot_surv()
```

R

Mathematical Details

Frequentist formulation

The Cox proportional hazards model can be expressed as:

$$h(t|X) = h_0(t) \exp(\beta^T X)$$

- Where:
    - $h(t|X)$ is the hazard at time t for covariates X.
    - $h_0(t)$ is the baseline hazard function (e.g., exponential, Weibull).
    - X represents the covariates (such as age, treatment).
    - β are the regression coefficients to be estimated.
- Censoring is accounted for by multiplying the hazard function by a factor that depends on the censoring distribution, usually modeled as independent censoring with a rate (t):
    - $Y_i(t) = Poisson(h(t|X) * \delta(t))$

3

Bayesian formulation

In Bayesian survival analysis, we define priors for each parameter:

1. **Hazard Function**: The hazard rate at time t for an individual is given by:

$$Y_i(t) = \text{Poisson}(\lambda(t) * censoring(t))$$

$$\lambda(t) = \lambda_0(t) \exp(x\beta)$$

$$\beta \sim \text{Normal}(\mu_\beta, \sigma_\beta^2)$$

$$\mu_\beta \sim \text{Normal}(0, 10^2)$$

$$\sigma_\beta^2 \sim \text{Uniform}(0, 10)$$

Where:

- $Y_i(t)$ is the status of the i-th subject at time t coded as a binary variable:

$$Y_i(t) = \begin{cases} 1 & \text{if subject } i \text{ died at time } t \text{ ,} \\ 0 & \text{otherwise.} \end{cases}$$

- $\lambda(t)$: Hazard function at time t.
- $\lambda_0(t)$: Baseline hazard function (e.g., *Exponential* or *Weibull*).
- x: Covariates (e.g., age, treatment).
- β: Regression coefficients capturing the effect of x on the hazard are assigned a prior with a *Normal* distribution.
- μ_β: Mean of the *Normal* distribution.
- σ_β^2: Variance of the *Normal* distribution.

Reference(s)