Tuesday 9:00-12:30

Part 3
Robust Bayesian statistics & applications in reliability networks

by Gero Walter

Robust Bayesian statistics & applications in reliability networks

Outline

- Robust Bayesian Analysis (9am)
  - Why
  - The Imprecise Dirichlet Model
  - General Framework for Canonical Exponential Families
- Exercises I (9:30am)
- System Reliability Application (10am)
- Break (10:30am)
- Exercises II (11am)

Robust Bayesian statistics & applications in reliability networks

Outline

Robust Bayesian Analysis: Why

- choice of prior can severely affect inferences
  even if your prior is 'non-informative'

Robust Bayesian Analysis: Why

- choice of prior can severely affect inferences
  even if your prior is 'non-informative'
- solution: systematic sensitivity analysis over prior parameters

Robust Bayesian Analysis: Why

- choice of prior can severely affect inferences
  even if your prior is 'non-informative'
- solution: systematic sensitivity analysis over prior parameters
- models from canonical exponential family make this easy to do [18]

Robust Bayesian Analysis: Why

- choice of prior can severely affect inferences
  even if your prior is 'non-informative'
- solution: systematic sensitivity analysis over prior parameters
- models from canonical exponential family make this easy to do [18]
- close relations to robust Bayes literature, e.g. [7, 19, 20]

Robust Bayesian Analysis: Why

- choice of prior can severely affect inferences
  even if your prior is 'non-informative'
- solution: systematic sensitivity analysis over prior parameters
- models from canonical exponential family make this easy to do [18]
- close relations to robust Bayes literature, e.g. [7, 19, 20]
- concerns uncertainty in the prior
  (uncertainty in data generating process: imprecise sampling models)

Robust Bayesian Analysis: Why

- choice of prior can severely affect inferences
  even if your prior is ‘non-informative’
- solution: systematic sensitivity analysis over prior parameters
- models from canonical exponential family make this easy to do [18]
- close relations to robust Bayes literature, e.g. [7, 19, 20]
- concerns uncertainty in the prior
  (uncertainty in data generating process: imprecise sampling models)
- here: focus on imprecise Dirichlet model

Robust Bayesian Analysis: Why

- choice of prior can severely affect inferences
  even if your prior is ‘non-informative’
- solution: systematic sensitivity analysis over prior parameters
- models from canonical exponential family make this easy to do [18]
- close relations to robust Bayes literature, e.g. [7, 19, 20]
- concerns uncertainty in the prior
  (uncertainty in data generating process: imprecise sampling models)
- here: focus on imprecise Dirichlet model
- if your prior is informative then prior-data conflict can be an issue [31, 29]
  (we’ll come back to this in the system reliability application)

Robust Bayesian Analysis: Principle of Indifference

How to construct a prior if we do not have a lot of information?

Laplace: Principle of Indifference

Use the uniform distribution.

Obvious issue: this depends on the parametrisation!

Robust Bayesian Analysis: Principle of Indifference

How to construct a prior if we do not have a lot of information?

Laplace: Principle of Indifference

Use the uniform distribution.

Obvious issue: this depends on the parametrisation!

Example

An object of 1kg has uncertain volume V between 1ℓ and 2ℓ.

- Uniform distribution over volume $V \implies E(V) = 1.5\ell$.
- Uniform distribution over density $\rho = 1/V \implies$ $E(V) = E(1/\rho) = \int_{0.5}^{1} 2/\rho d\rho = 2(\ln 1 - \ln 0.5) = 1.39\ell$

Robust Bayesian Analysis: Principle of Indifference

How to construct a prior if we do not have a lot of information?

Laplace: Principle of Indifference

Use the uniform distribution.

Obvious issue: this depends on the parametrisation!

Example

An object of 1kg has uncertain volume V between 1ℓ and 2ℓ.

- Uniform distribution over volume $V \implies E(V) = 1.5\ell$.
- Uniform distribution over density $\rho = 1/V \implies$ $E(V) = E(1/\rho) = \int_{0.5}^{1} 2/\rho d\rho = 2(\ln 1 - \ln 0.5) = 1.39\ell$

The uniform distribution does not really model prior ignorance.
(Jeffreys prior is transformation-invariant, but depends on the sample space and can break decision making!)

Robust Bayesian Analysis: Prior Ignorance via Sets of Probabilities

How to construct prior if we do not have a lot of information?

Boole: Probability Bounding

Use the set of all probability distributions (vacuous model).

Results no longer depend on parametrisation!

Robust Bayesian Analysis: Prior Ignorance via Sets of Probabilities

How to construct prior if we do not have a lot of information?

Boole: Probability Bounding

Use the set of all probability distributions (vacuous model).

Results no longer depend on parametrisation!

Example

An object of 1kg has uncertain volume V between 1ℓ and 2ℓ.

- Set of all distributions over volume $V \implies E(V) \in [1, 2]$.
- Set of all distribution over density $\rho = 1/V \implies E(V) = E(1/\rho) \in [1, 2]$

Robust Bayesian Analysis: Prior Ignorance via Sets of Probabilities

Theorem
The set of posterior distributions resulting from a vacuous set of prior distributions is again vacuous, regardless of the likelihood.

We can never learn anything when starting from a vacuous set of priors.

Robust Bayesian Analysis: Prior Ignorance via Sets of Probabilities

Theorem

The set of posterior distributions resulting from a vacuous set of prior distributions is again vacuous, regardless of the likelihood.

We can never learn anything when starting from a vacuous set of priors.

Solution: Near-Vacuous Sets of Priors

Only insist that the prior predictive, or other classes of inferences, are vacuous.

This can be done using sets of conjugate priors [4, 5].

Robust Bayesian statistics & applications in reliability networks

Outline

- Robust Bayesian Analysis (9am)
  - Why
  - The Imprecise Dirichlet Model
  - General Framework for Canonical Exponential Families
- Exercises I (9:30am)
- System Reliability Application (10am)
- Break (10:30am)
- Exercises II (11am)

The Imprecise Dirichlet Model: Definition

- introduced by Peter Walley [27, 28]
- for multinomial sampling, k categories 1, 2, ..., k
- Bayesian conjugate analysis
  - multinomial likelihood (sample $\boldsymbol{n} = (n_1, \ldots, n_k)$, $\sum n_i = n$)

$$f(\boldsymbol{n} \mid \boldsymbol{\theta}) = \frac{n!}{n_1! \cdots n_k!} \prod_{i=1}^{k} \theta_i^{n_i}$$

  - conjugate Dirichlet prior
    - with mean $\boldsymbol{t} = (t_1, \ldots, t_k)$ = prior expected proportions
    - and parameter $s > 0$

$$f(\boldsymbol{\theta}) = \frac{\Gamma(s)}{\prod_{i=1}^{k} \Gamma(st_i)} \prod_{i=1}^{k} \theta_i^{st_i - 1}$$

The Imprecise Dirichlet Model: Definition

- introduced by Peter Walley [27, 28]
- for multinomial sampling, k categories $1, 2, \ldots, k$
- Bayesian conjugate analysis
  - multinomial likelihood (sample $\boldsymbol{n} = (n_1, \ldots, n_k)$, $\sum n_i = n$)

$$f(\boldsymbol{n} \mid \boldsymbol{\theta}) = \frac{n!}{n_1! \cdots n_k!} \prod_{i=1}^{k} \theta_i^{n_i}$$

  - conjugate Dirichlet prior
    - with mean $\boldsymbol{t} = (t_1, \ldots, t_k)$ = prior expected proportions
    - and parameter $s > 0$

$$f(\boldsymbol{\theta}) = \frac{\Gamma(s)}{\prod_{i=1}^{k} \Gamma(st_i)} \prod_{i=1}^{k} \theta_i^{st_i - 1}$$

Definition (Imprecise Dirichlet Model)

Use the set $\mathcal{M}^{(0)}$ of all Dirichlet priors, for a fixed $s > 0$, and take the infimum/supremum over $\boldsymbol{t}$ of the posterior to get lower/upper predictive probabilities/expectations.

The Imprecise Dirichlet Model: Properties

- conjugacy: $f(\boldsymbol{\theta} \mid \boldsymbol{n})$ again Dirichlet with parameters

$$t_i^* = \frac{st_i + n_i}{s + n} = \frac{s}{s + n} t_i + \frac{n}{s + n} \frac{n_i}{n},$$
$$s^* = s + n$$

- $t_i^* = \mathrm{E}(\theta_i \mid \boldsymbol{n}) = \mathrm{P}(i \mid \boldsymbol{n})$ is a weighted average of t_i and n_i/n, with weights proportional to s and n, respectively
- s can be interpreted as a prior strength or pseudocount
- lower and upper expectations / probabilities by min and max over $\boldsymbol{t} \in \Delta$ (unit simplex)

The Imprecise Dirichlet Model: Properties

Posterior predictive probabilities

- for observing a particular category

$$\underline{\mathrm{P}}\,(i \mid \boldsymbol{n}) = \frac{n_i}{s+n}, \qquad \overline{\mathrm{P}}\,(i \mid \boldsymbol{n}) = \frac{s+n_i}{s+n}$$

- for observing a non-trivial event $A \subseteq \{1, \ldots, k\}$

$$\underline{\mathrm{P}}\,(A \mid \boldsymbol{n}) = \frac{n_A}{s+n}, \qquad \overline{\mathrm{P}}\,(A \mid \boldsymbol{n}) = \frac{s+n_A}{s+n},$$

with $n_A = \sum_{i \in A} n_i$

Satisfies prior near ignorance:
vacuous for prior predictive $\underline{P}(A) = 0, \overline{P}(A) = 1$
Inferences are independent of categorisation
('Representation Invariance Principle').

The Imprecise Dirichlet Model: Why A ***Set*** of Priors?

- single prior $\implies$ dependence on categorisation
- for example, single Dirichlet prior (with $t_A = \sum_{i \in A} t_i$, $s = 2$)

$$\mathrm{P}(A \mid \boldsymbol{n}) = \frac{2t_A + n_A}{n+2}$$

one red marble observed

- two categories red (R) and other (O):
  prior ignorance $\implies t_R = t_O = \frac{1}{2} \implies \mathrm{P}(R \mid \boldsymbol{n}) = \frac{2}{3}$
- three categories red (R), green (G), and blue (B):
  prior ignorance $\implies t_R = t_G = t_B = \frac{1}{3} \implies \mathrm{P}(R \mid \boldsymbol{n}) = \frac{5}{9}$

The Imprecise Dirichlet Model: Why A ***Set*** of Priors?

- single prior $\implies$ dependence on categorisation
- for example, single Dirichlet prior (with $t_A = \sum_{i \in A} t_i$, $s = 2$)

$$\mathrm{P}(A \mid \boldsymbol{n}) = \frac{2t_A + n_A}{n+2}$$

one red marble observed

- two categories red (R) and other (O):
  prior ignorance $\implies$ $t_R = t_O = \frac{1}{2}$ $\implies$ $\mathrm{P}(R \mid \boldsymbol{n}) = \frac{2}{3}$
- three categories red (R), green (G), and blue (B):
  prior ignorance $\implies$ $t_R = t_G = t_B = \frac{1}{3}$ $\implies$ $\mathrm{P}(R \mid \boldsymbol{n}) = \frac{5}{9}$

prior ignorance + representation invariance principle
$\implies$ must use set of priors

The Imprecise Dirichlet Model: The s Parameter

- s can be interpreted as a prior strength or pseudocount
- s determines learning speed:

$$\overline{\mathrm{P}}\,(A \mid \boldsymbol{n}) - \underline{\mathrm{P}}\,(A \mid \boldsymbol{n}) = \frac{s}{s+n}$$

- no objective way of choosing s, but $s = 2$ covers most Bayesian and frequentist inferences
- for $s = n$ posterior imprecision is half the prior imprecision
- for informative t_i bounds, using a range of s values allows the set of posteriors to reflect *prior-data conflict* (see system reliability application)

Robust Bayesian statistics & applications in reliability networks

Outline

General Framework for Canonical Exponential Families

Conjugate priors like the Dirichlet can be constructed for sample distributions (likelihood) from:

Definition (Canonical exponential family)

$$f(x \mid \psi) = h(x) \exp\left\{\psi^T \tau(x) - b(\psi)\right\}$$

- includes multinomial, normal, Poisson, exponential, ...
- ψ generally a transformation of original parameter θ

General Framework for Canonical Exponential Families

Conjugate priors like the Dirichlet can be constructed for sample distributions (likelihood) from:

Definition (Canonical exponential family)

$$f(x \mid \psi) = h(x) \exp\left\{\psi^T \tau(x) - b(\psi)\right\}$$

- includes multinomial, normal, Poisson, exponential, ...
- ψ generally a transformation of original parameter θ

Definition (Family of conjugate priors)

A family of priors for i.i.d. sampling from the can. exp. family:

$$f(\psi \mid n^{(0)}, y^{(0)}) \propto \exp\left\{n^{(0)}\left[\psi^T y^{(0)} - b(\psi)\right]\right\}$$

with hyper-parameters $n^{(0)}$ ($\leftrightarrow s$) and $y^{(0)}$ ($\leftrightarrow \boldsymbol{t}$ from IDM).

General Framework for Canonical Exponential Families

Theorem (Conjugacy)

Posterior is of the same form:

$$f(\psi \mid n^{(0)}, y^{(0)}, \boldsymbol{x}) \propto \exp\left\{ n^{(n)} \left[\psi^T y^{(n)} - b(\psi) \right] \right\}$$

where

$$\boldsymbol{x} = (x_1, \ldots, x_n)$$

$$(s^* \leftrightarrow) \qquad n^{(n)} = n^{(0)} + n$$

$$(\boldsymbol{t}^* \leftrightarrow) \qquad y^{(n)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot y^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{\tau(\boldsymbol{x})}{n}$$

$$(n_i \leftrightarrow) \qquad \tau(\boldsymbol{x}) = \sum_{i=1}^{n} \tau(x_i)$$

General Framework for Canonical Exponential Families

- $y^{(0)}$ $(\leftrightarrow \boldsymbol{t})$ = prior expectation of $\tau(\boldsymbol{x})/n$
- $n^{(0)}$ $(\leftrightarrow s)$ determines spread and learning speed

General Framework for Canonical Exponential Families

- $y^{(0)}$ $(\leftrightarrow \boldsymbol{t})$ = prior expectation of $\tau(\boldsymbol{x})/n$
- $n^{(0)}$ $(\leftrightarrow s)$ determines spread and learning speed

- usefulness of this framework for IP / robust Bayes discovered by Quaghebeur & de Cooman [18]
- near-noninformative sets of priors developed by Benavoli & Zaffalon [4, 5]
- for informative sets of priors, Walter & Augustin [29, 31] suggest to use parameter sets $\Pi^{(0)} = [\underline{n}^{(0)}, \overline{n}^{(0)}] \times [\underline{y}^{(0)}, \overline{y}^{(0)}]$

General Framework: Why vary $n^{(0)} (\leftrightarrow s)$?

What if prior assumptions and data tell different stories?

Prior-Data Conflict

- *informative prior beliefs* and *trusted data* (sampling model correct, no outliers, etc.) are in conflict
- "the prior [places] its mass primarily on distributions in the sampling model for which the observed data is surprising" [10]
- there are not enough data to overrule the prior

Example: IDM with $k = 2 \implies$ Imprecise Beta Model

Imprecise Beta Model (IBM)

- binomial likelihood (observing x successes in n trials)

$$f(x \mid \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

- conjugate Beta prior
  - with mean $y^{(0)}$ = prior expected probability of success
  - and prior strength parameter $n^{(0)} > 0$

$$f(\theta) \propto \theta^{n^{(0)} y^{(0)} - 1} (1-\theta)^{n^{(0)}(1-y^{(0)})-1}$$

Imprecise Beta Model (IBM)

- binomial likelihood (observing x successes in n trials)

$$f(x \mid \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

- conjugate Beta prior
  - with mean $y^{(0)}$ = prior expected probability of success
  - and prior strength parameter $n^{(0)} > 0$

$$f(\theta) \propto \theta^{n^{(0)} y^{(0)} - 1} (1-\theta)^{n^{(0)}(1-y^{(0)})-1}$$

- informative set of priors: Use the set $\mathcal{M}^{(0)}$ of Beta priors with $y^{(0)} \in [\underline{y}^{(0)}, \overline{y}^{(0)}]$ and
  - $n^{(0)} > 0$ fixed, or
  - $n^{(0)} \in [\underline{n}^{(0)}, \overline{n}^{(0)}]$

Imprecise Beta Model (IBM)

- binomial likelihood (observing x successes in n trials)

$$f(x \mid \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

- conjugate Beta prior
  - with mean $y^{(0)}$ = prior expected probability of success
  - and prior strength parameter $n^{(0)} > 0$

$$f(\theta) \propto \theta^{n^{(0)} y^{(0)} - 1} (1-\theta)^{n^{(0)}(1-y^{(0)})-1}$$

- informative set of priors: Use the set $\mathcal{M}^{(0)}$ of Beta priors with $y^{(0)} \in [\underline{y}^{(0)}, \overline{y}^{(0)}]$ and
  - $n^{(0)} > 0$ fixed, or
  - $n^{(0)} \in [\underline{n}^{(0)}, \overline{n}^{(0)}]$
- $\mathrm{E}(\theta \mid x) = y^{(n)}$ is a weighted average of $\mathrm{E}(\theta) = y^{(0)}$ and $\frac{x}{n}$!

Imprecise Beta Model (IBM)

- binomial likelihood (observing x successes in n trials)

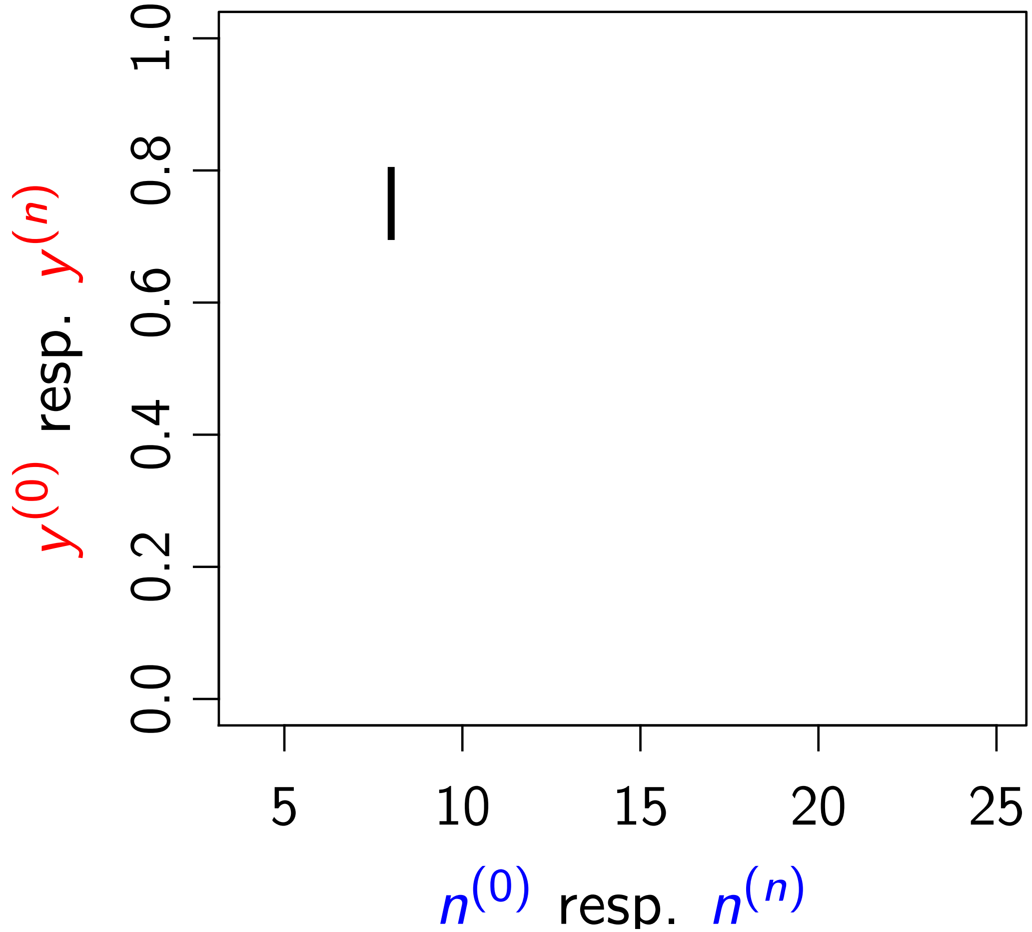
$$f(x \mid \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

- conjugate Beta prior
  - with mean $y^{(0)}$ = prior expected probability of success
  - and prior strength parameter $n^{(0)} > 0$

$$f(\theta) \propto \theta^{n^{(0)} y^{(0)} - 1} (1-\theta)^{n^{(0)}(1-y^{(0)})-1}$$

- informative set of priors: Use the set $\mathcal{M}^{(0)}$ of Beta priors with $y^{(0)} \in [\underline{y}^{(0)}, \overline{y}^{(0)}]$ and
  - $n^{(0)} > 0$ fixed, or
  - $n^{(0)} \in [\underline{n}^{(0)}, \overline{n}^{(0)}]$
- $\mathrm{E}(\theta \mid x) = y^{(n)}$ is a weighted average of $\mathrm{E}(\theta) = y^{(0)}$ and $\dfrac{x}{n}$!
- $\mathrm{Var}(\theta \mid x) = \dfrac{y^{(n)}(1-y^{(n)})}{n^{(n)}+1}$ decreases with n!

Imprecise Beta Model with $n^{(0)}$ fixed

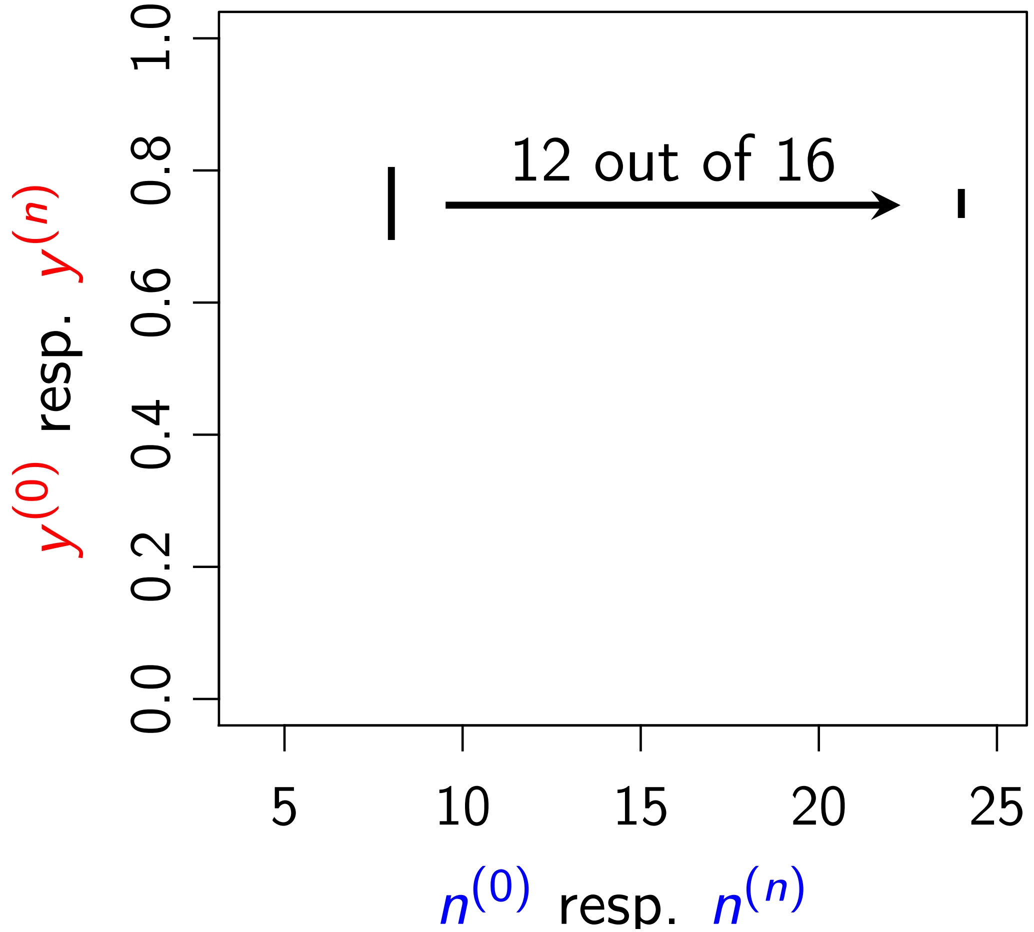


no conflict:

prior $n^{(0)} = 8$, $y^{(0)} \in [0.7, 0.8]$

data $s/n = 12/16 = 0.75$

Imprecise Beta Model with $n^{(0)}$ fixed



no conflict:

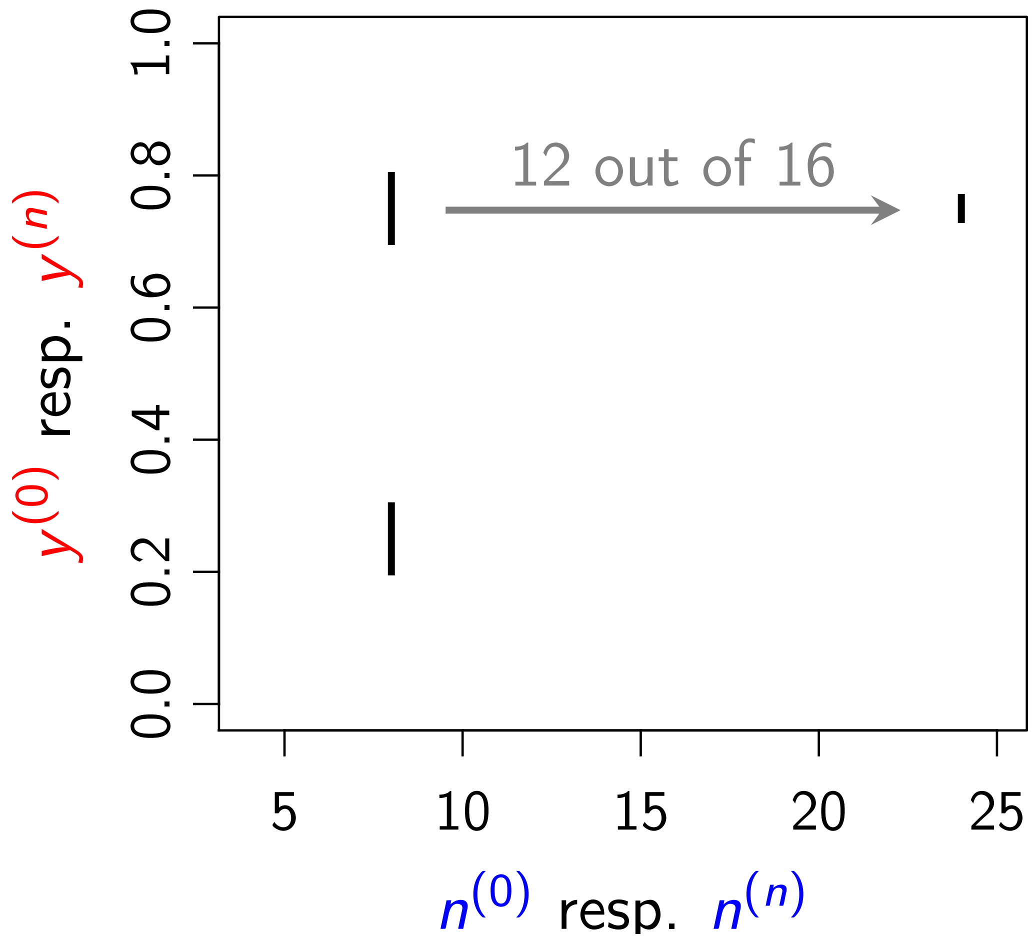
prior $n^{(0)} = 8$, $y^{(0)} \in [0.7, 0.8]$

data $s/n = 12/16 = 0.75$

▼

$n^{(n)} = 24$, $y^{(n)} \in [0.73, 0.77]$

Imprecise Beta Model with $n^{(0)}$ fixed



no conflict:

prior $n^{(0)} = 8$, $y^{(0)} \in [0.7, 0.8]$

data $s/n = 12/16 = 0.75$

▼

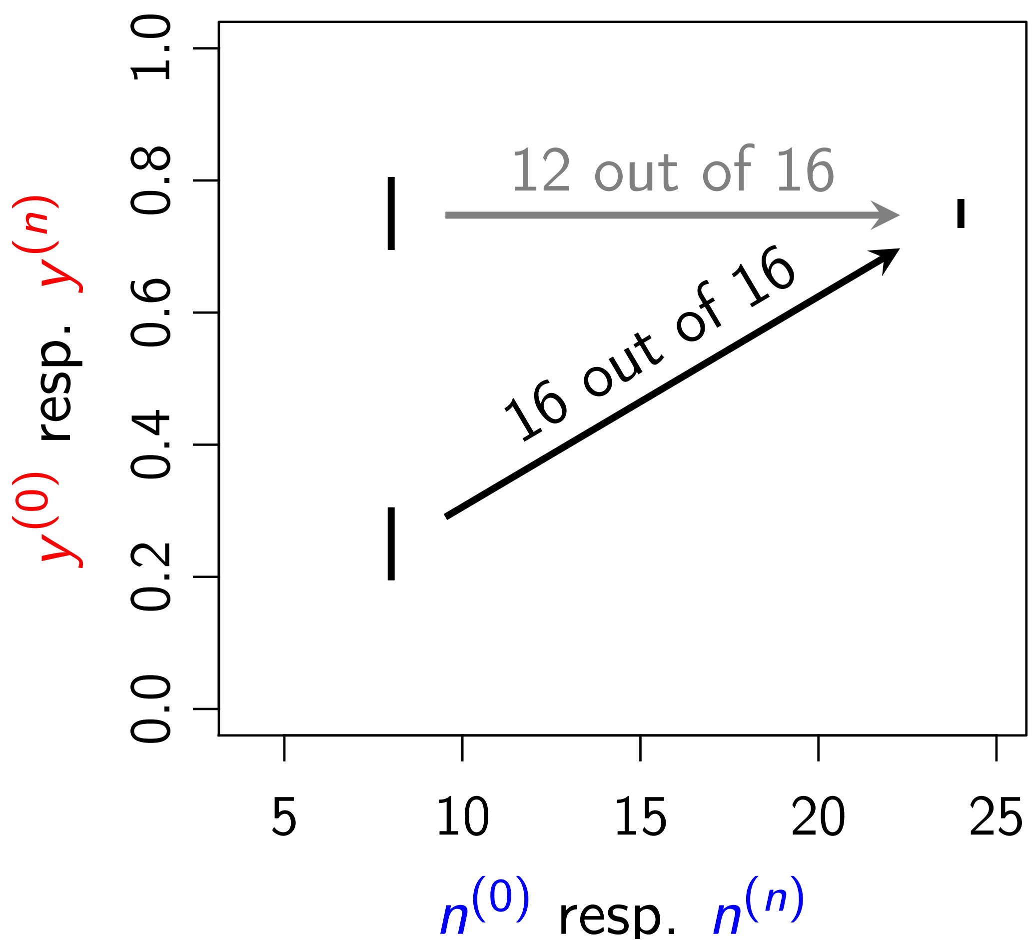
$n^{(n)} = 24$, $y^{(n)} \in [0.73, 0.77]$

prior data conflict:

prior $n^{(0)} = 8$, $y^{(0)} \in [0.2, 0.3]$

data $s/n = 16/16 = 1$

Imprecise Beta Model with $n^{(0)}$ fixed



no conflict:

prior $n^{(0)} = 8$, $y^{(0)} \in [0.7, 0.8]$

data $s/n = 12/16 = 0.75$

▼

$n^{(n)} = 24$, $y^{(n)} \in [0.73, 0.77]$

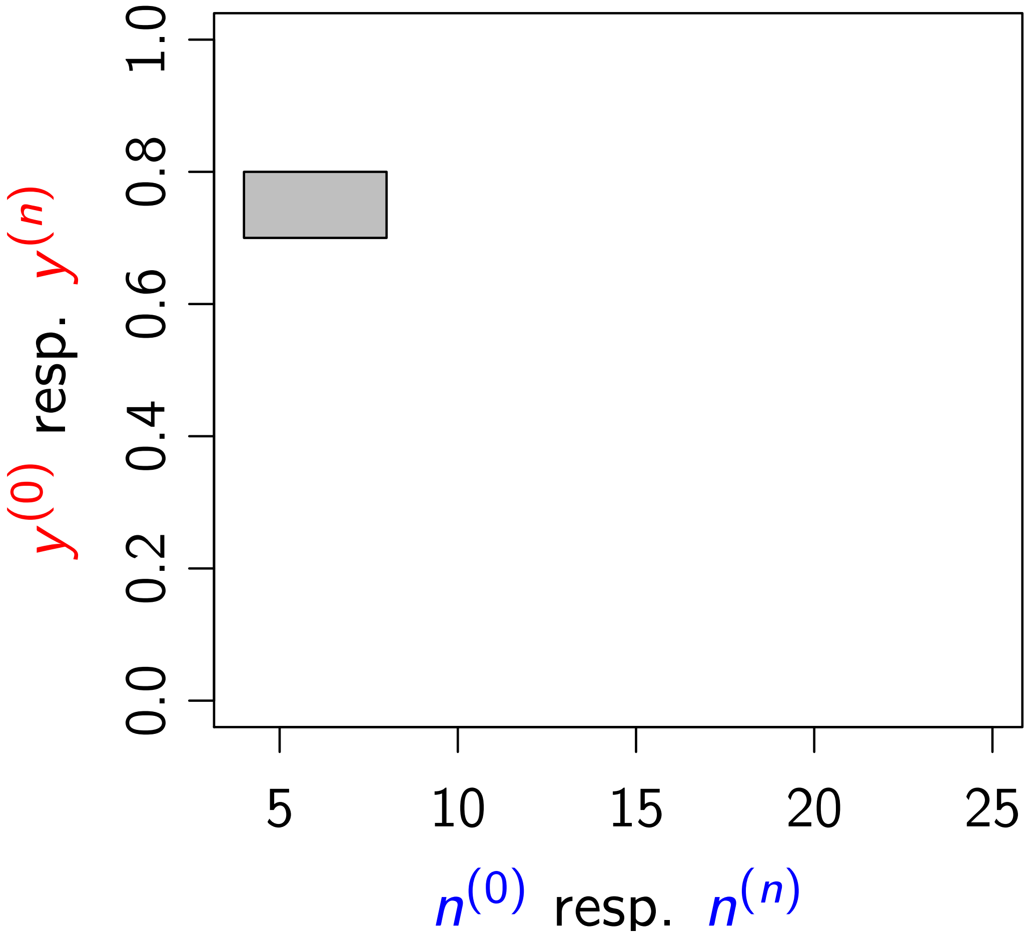
▲

prior data conflict:

prior $n^{(0)} = 8$, $y^{(0)} \in [0.2, 0.3]$

data $s/n = 16/16 = 1$

Imprecise Beta Model with $n^{(0)}$ interval

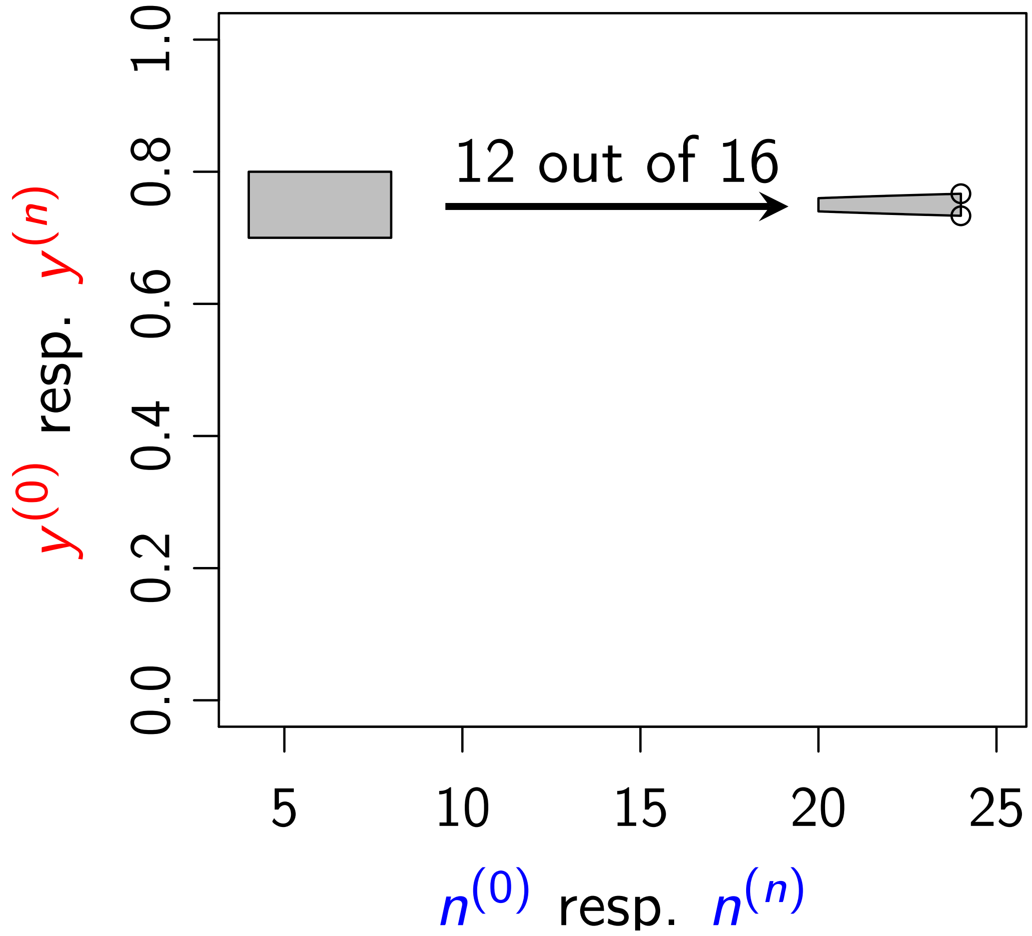


no conflict:

prior $n^{(0)} \in [4, 8]$, $y^{(0)} \in [0.7, 0.8]$

data $s/n = 12/16 = 0.75$

Imprecise Beta Model with $n^{(0)}$ interval



no conflict:

prior $n^{(0)} \in [4, 8]$, $y^{(0)} \in [0.7, 0.8]$

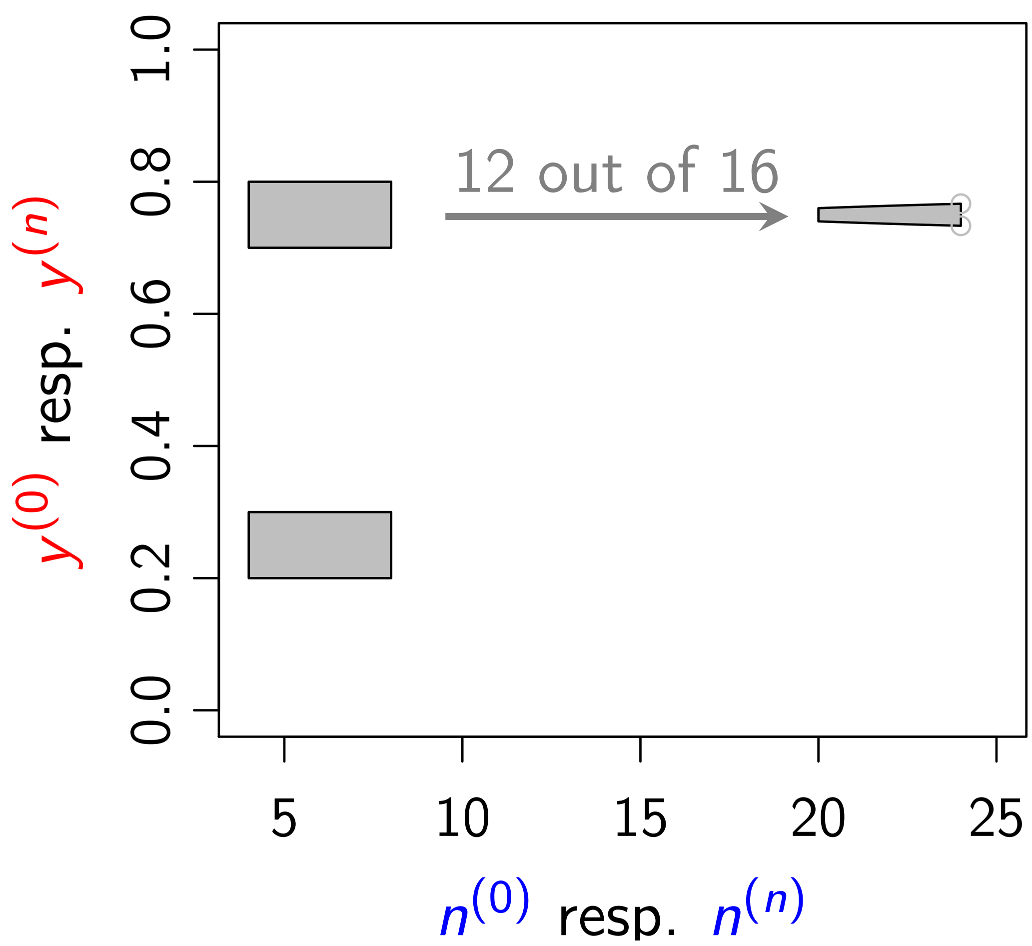
data $s/n = 12/16 = 0.75$

▼

$y^{(n)} \in [0.73, 0.77]$

"spotlight" shape

Imprecise Beta Model with $n^{(0)}$ interval



no conflict:

prior $n^{(0)} \in [4, 8]$, $y^{(0)} \in [0.7, 0.8]$

data $s/n = 12/16 = 0.75$

▼

$y^{(n)} \in [0.73, 0.77]$

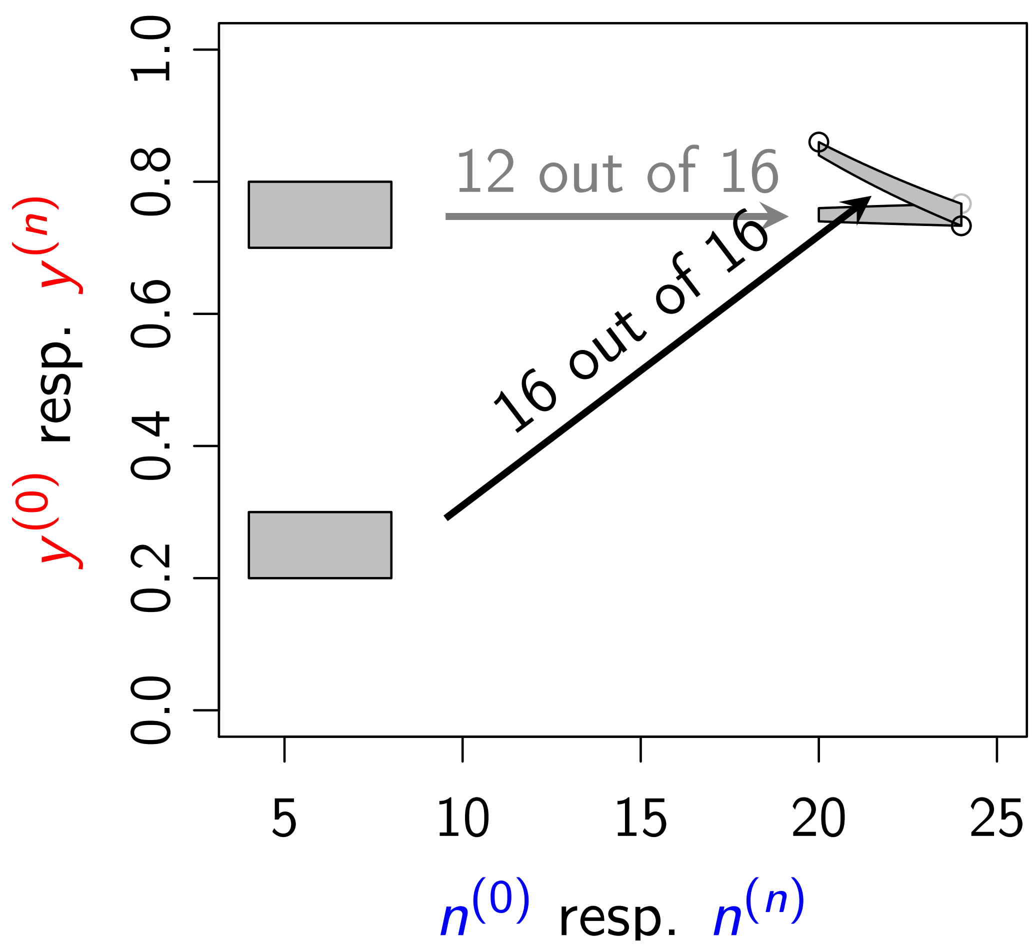
"spotlight" shape

prior-data conflict:

prior $n^{(0)} \in [4, 8]$, $y^{(0)} \in [0.2, 0.3]$

data $s/n = 16/16 = 1$

Imprecise Beta Model with $n^{(0)}$ interval



no conflict:

prior $n^{(0)} \in [4, 8]$, $y^{(0)} \in [0.7, 0.8]$

data $s/n = 12/16 = 0.75$

▼

$y^{(n)} \in [0.73, 0.77]$

“spotlight” shape

prior-data conflict:

prior $n^{(0)} \in [4, 8]$, $y^{(0)} \in [0.2, 0.3]$

data $s/n = 16/16 = 1$

▼

$y^{(n)} \in [0.73, 0.86]$

“banana” shape

Robust Bayesian Analysis: Other Models

- How to define sets of priors $\mathcal{M}^{(0)}$ is a crucial modeling choice
- Sets $\mathcal{M}^{(0)}$ via parameter sets $\Pi^{(0)}$ seem to work better than other models discussed in the robust Bayes literature:
  - Neighbourhood models
    - set of distributions 'close to' a central distribution P_0
    - example: ε-contamination class: $\{P : P = (1-\varepsilon)P_0 + \varepsilon Q, Q \in \mathcal{Q}\}$
      - not necessarily closed under Bayesian updating

Robust Bayesian Analysis: Other Models

- How to define sets of priors $\mathcal{M}^{(0)}$ is a crucial modeling choice
- Sets $\mathcal{M}^{(0)}$ via parameter sets $\Pi^{(0)}$ seem to work better than other models discussed in the robust Bayes literature:
  - Neighbourhood models
    - set of distributions 'close to' a central distribution P_0
    - example: ε-contamination class: $\{P : P = (1-\varepsilon)P_0 + \varepsilon Q, Q \in \mathcal{Q}\}$
      - not necessarily closed under Bayesian updating
  - Density ratio class / interval of measures
    - set of distributions by bounds for the density function $f(\vartheta)$:

$$\mathcal{M}^{(0)}_{l,u} = \big\{f(\theta) : \exists c \in \mathbb{R}_{>0} : l(\theta) \leq cf(\theta) \leq u(\theta)\big\}$$

    - posterior set is bounded by updated $l(\theta)$ and $u(\theta)$
    - $u(\theta)/l(\theta)$ is constant under updating
      - size of the set does not decrease with n
      - very vague posterior inferences

Robust Bayesian statistics & applications in reliability networks

Outline

Exercise: Update step in the IBM

Consider first a single Beta prior with parameters $n^{(0)}$ and $y^{(0)}$.

Discuss how $\text{E}[\theta \mid x] = y^{(n)}$ *and* $\text{Var}(\theta \mid x) = \frac{y^{(n)}(1-y^{(n)})}{n^{(n)}+1}$ *behave when*

(i) $n^{(0)} \to 0$, (ii) $n^{(0)} \to \infty$, (iii) $n \to \infty$ when $x/n = \text{const.}$

What do the results for $\text{E}[\theta \mid x]$ *and* $\text{Var}(\theta \mid x)$ *imply for the shape of* $f(\theta \mid x)$*?*

When considering a set of Beta priors with

$$(n^{(0)}, y^{(0)}) \in \Pi^{(0)} = [\underline{n}^{(0)}, \overline{n}^{(0)}] \times [\underline{y}^{(0)}, \overline{y}^{(0)}],$$

what does the weighted average structure of $y^{(n)}$ *tell you about the updated set of parameters?* (Remember the prior-data conflict example!)

Exercise: Update step for canonically constructed priors

The canonically constructed prior for an i.i.d. sample distribution from the canonical exponential family

$$f(\mathbf{x} \mid \psi) = \left(\prod_{i=1}^{n} h(x_i) \right) \exp\left\{ \psi^T \tau(\mathbf{x}) - n b(\psi) \right\}$$

is given by

$$f(\psi \mid n^{(0)}, y^{(0)}) \propto \exp\left\{ n^{(0)} \left[\psi^T y^{(0)} - b(\psi) \right] \right\},$$

and the corresponding posterior by

$$f(\psi \mid n^{(0)}, y^{(0)}, \mathbf{x}) \propto \exp\left\{ n^{(n)} \left[\psi^T y^{(n)} - b(\psi) \right] \right\},$$

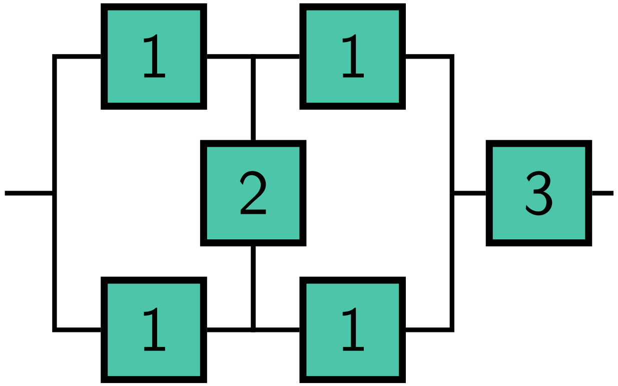
where $y^{(n)} = \frac{n^{(0)}}{n^{(0)} + n} \cdot y^{(0)} + \frac{n}{n^{(0)} + n} \cdot \frac{\tau(\mathbf{x})}{n}$ and $n^{(n)} = n^{(0)} + n.$

Confirm the expressions for $y^{(n)}$ and $n^{(n)}$.

Robust Bayesian statistics & applications in reliability networks

Outline

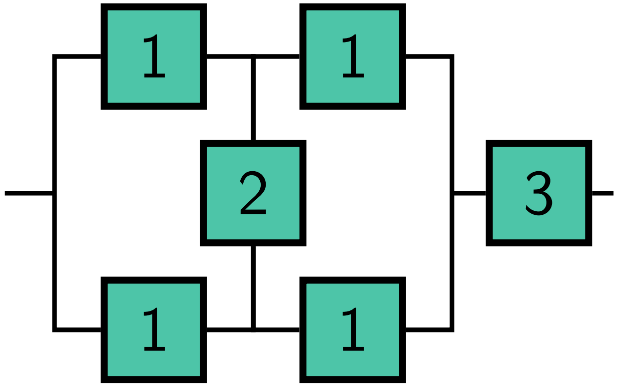
System Reliability Application: Reliability Block Diagrams



Reliability block diagrams:

- system consists of components k (different types $k = 1, \ldots, K$)
- each k either works or not
- system works when there is a path using only working components

System Reliability Application: Reliability Block Diagrams



Reliability block diagrams:

- system consists of components [k] (different types $k = 1, \dots, K$)
- each [k] either works or not
- system works when there is a path using only working components

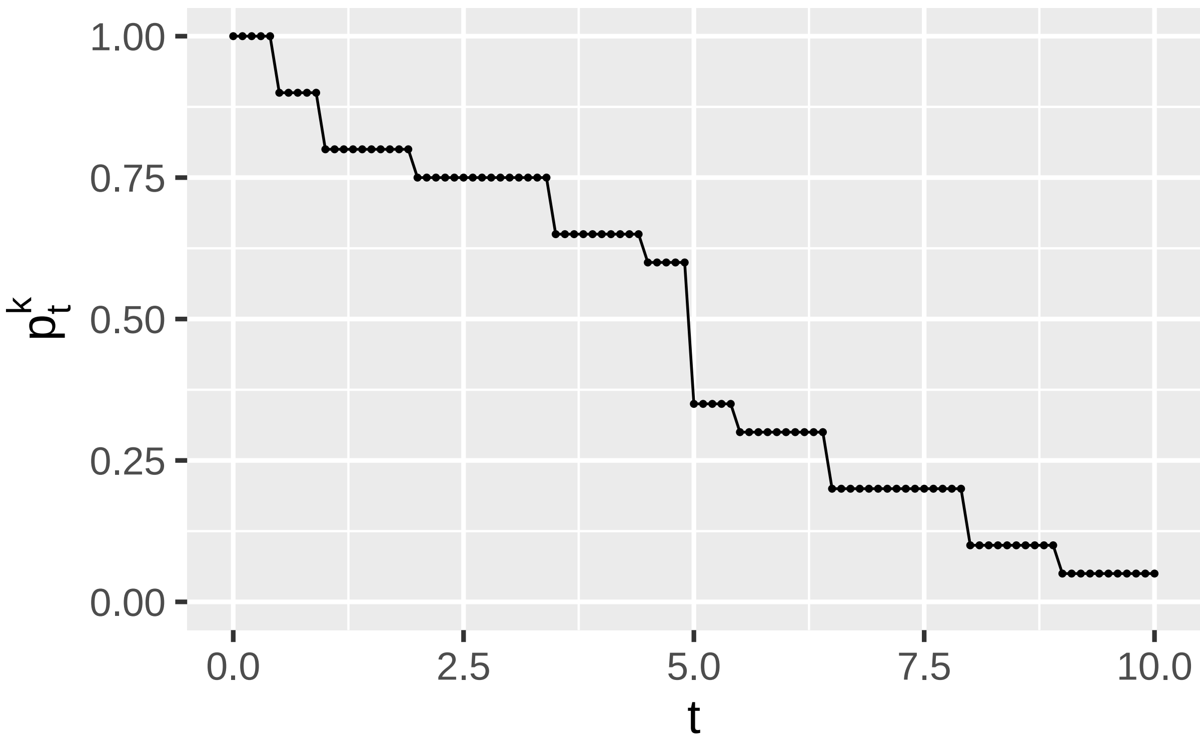
We want to learn about the system reliability $R_{sys}(t) = P(T_{sys} > t)$
(system survival function)
based on

- component test data:
  n_k failure times for components of type k, $k = 1, \dots, K$
- cautious assumptions on component reliability:
  expert information, e.g. from maintenance managers and staff

Nonparametric Component Reliability

Functioning probability p_t^k of k for each time $t \in \mathcal{T} = \{t'_1, t'_2, \ldots\}$

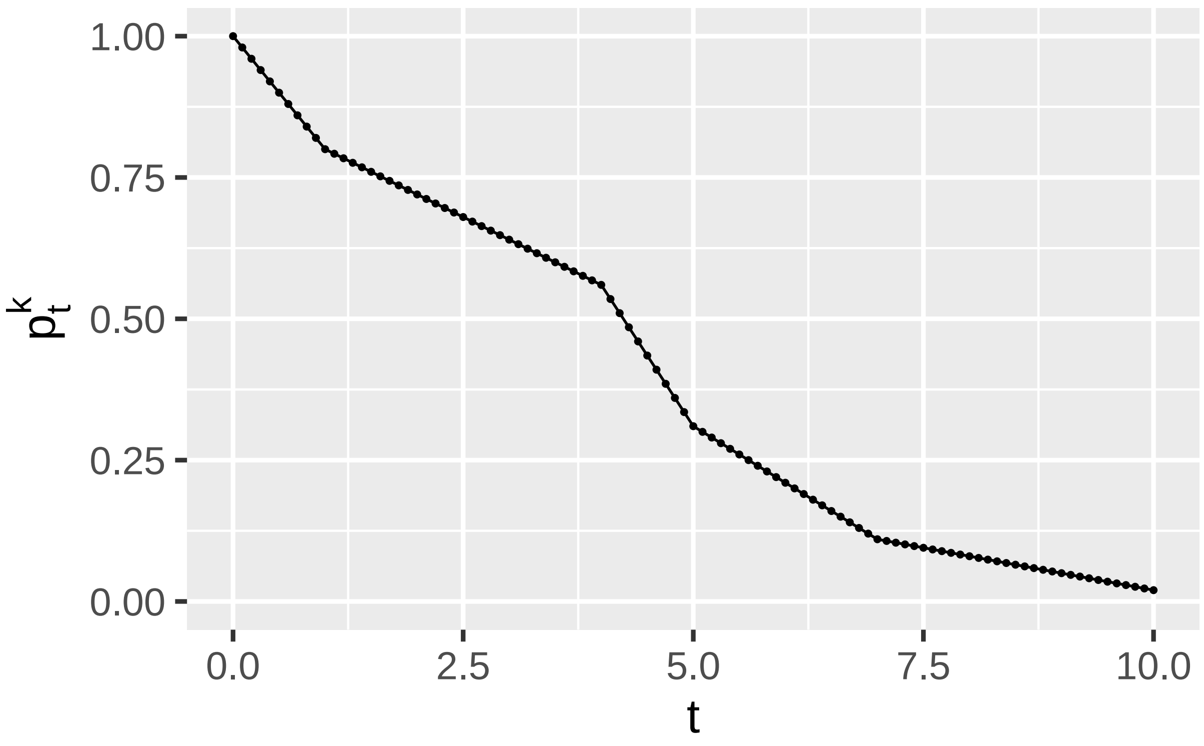
► discrete component reliability function $R^k(t) = p_t^k, \; t \in \mathcal{T}$.



Nonparametric Component Reliability

Functioning probability p_t^k of k for each time $t \in \mathcal{T} = \{t'_1, t'_2, \ldots\}$

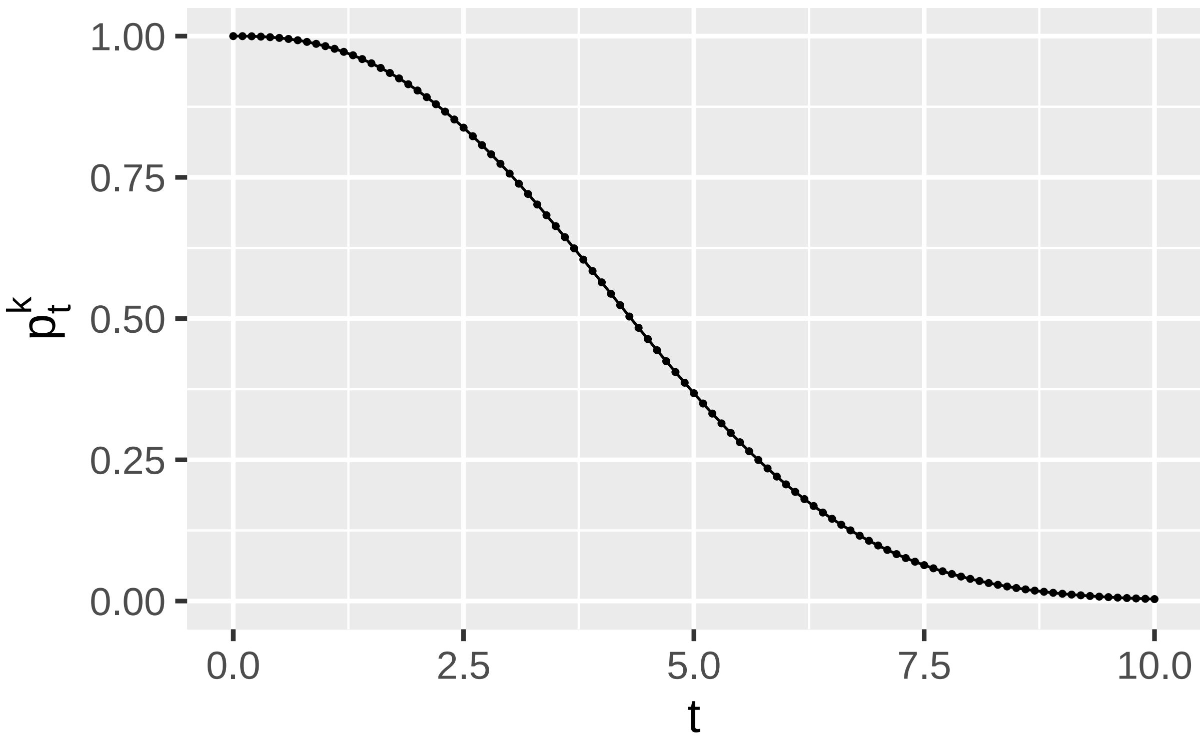
▶ discrete component reliability function $R^k(t) = p_t^k, \; t \in \mathcal{T}$.



Nonparametric Component Reliability

Functioning probability p_t^k of k for each time $t \in \mathcal{T} = \{t'_1, t'_2, \ldots\}$

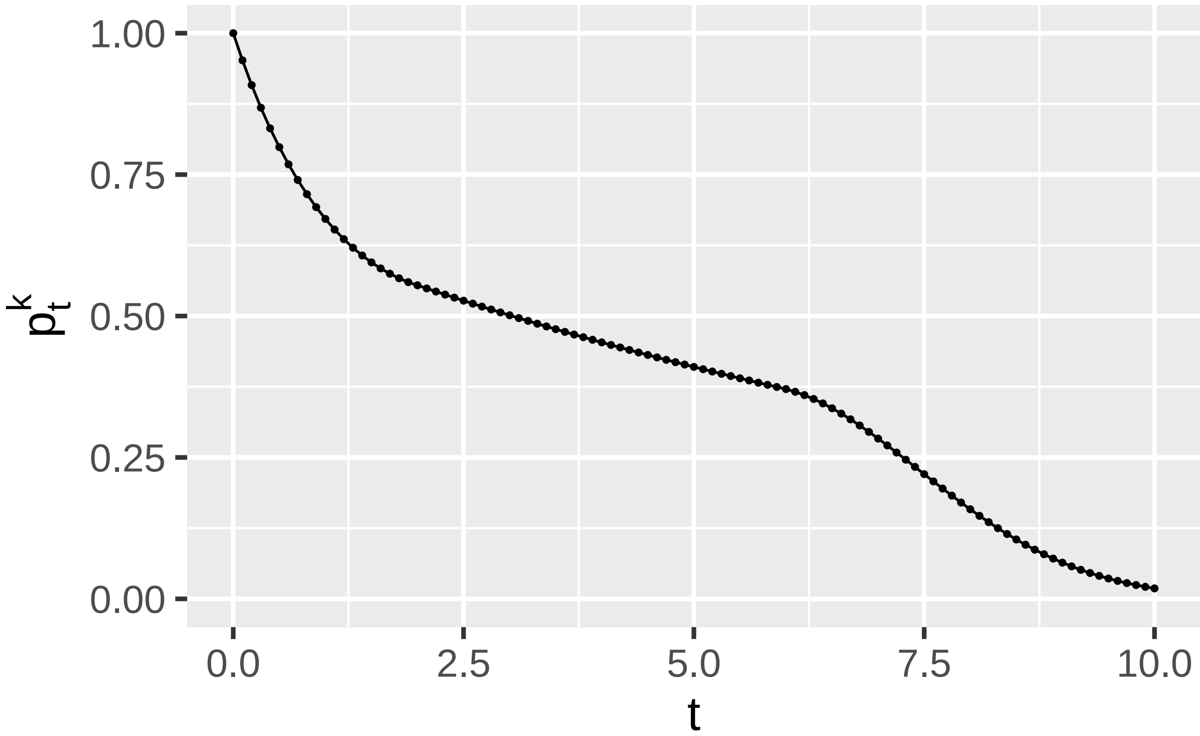
- discrete component reliability function $R^k(t) = p_t^k, \; t \in \mathcal{T}$.



Nonparametric Component Reliability

Functioning probability p_t^k of k for each time $t \in \mathcal{T} = \{t'_1, t'_2, \ldots\}$

▶ discrete component reliability function $R^k(t) = p_t^k,\ t \in \mathcal{T}$.



Nonparametric Component Reliability

Functioning probability p_t^k of k for each time $t \in \mathcal{T} = \{t'_1, t'_2, \ldots\}$
▶ discrete component reliability function $R^k(t) = p_t^k, \; t \in \mathcal{T}$.

use Imprecise Beta Model to estimate p_t^k's:

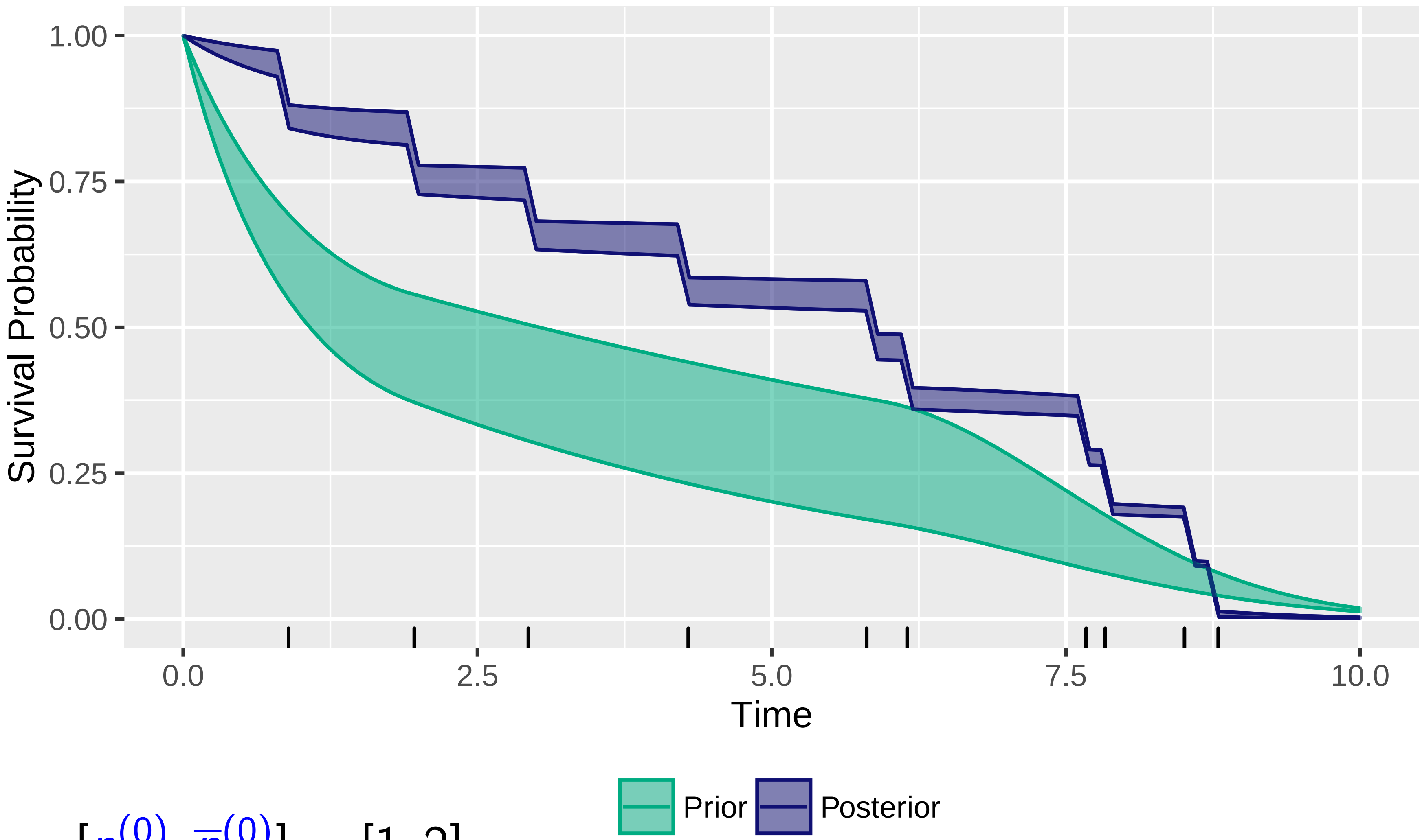
- Set of Beta priors for each p_t^k:

  $p_t^k \sim \text{Beta}(n_{k,t}^{(0)}, y_{k,t}^{(0)})$ with
  $(n_{k,t}^{(0)}, y_{k,t}^{(0)}) \in \Pi_{k,t}^{(0)} = [\underline{n}_{k,t}^{(0)}, \overline{n}_{k,t}^{(0)}] \times [\underline{y}_{k,t}^{(0)}, \overline{y}_{k,t}^{(0)}]$
  $\mathcal{M}_{k,t}^{(0)}$ can be near-vacuous by setting $[\underline{y}_{k,t}^{(0)}, \overline{y}_{k,t}^{(0)}] = [0, 1]$
  $\mathcal{M}_{k,t}^{(n)}$ will reflect prior-data conflict for informative $\mathcal{M}_{k,t}^{(0)}$

- failure times $\boldsymbol{t}^k = (t_1^k, \ldots, t_{n_k}^k)$ from component test data:

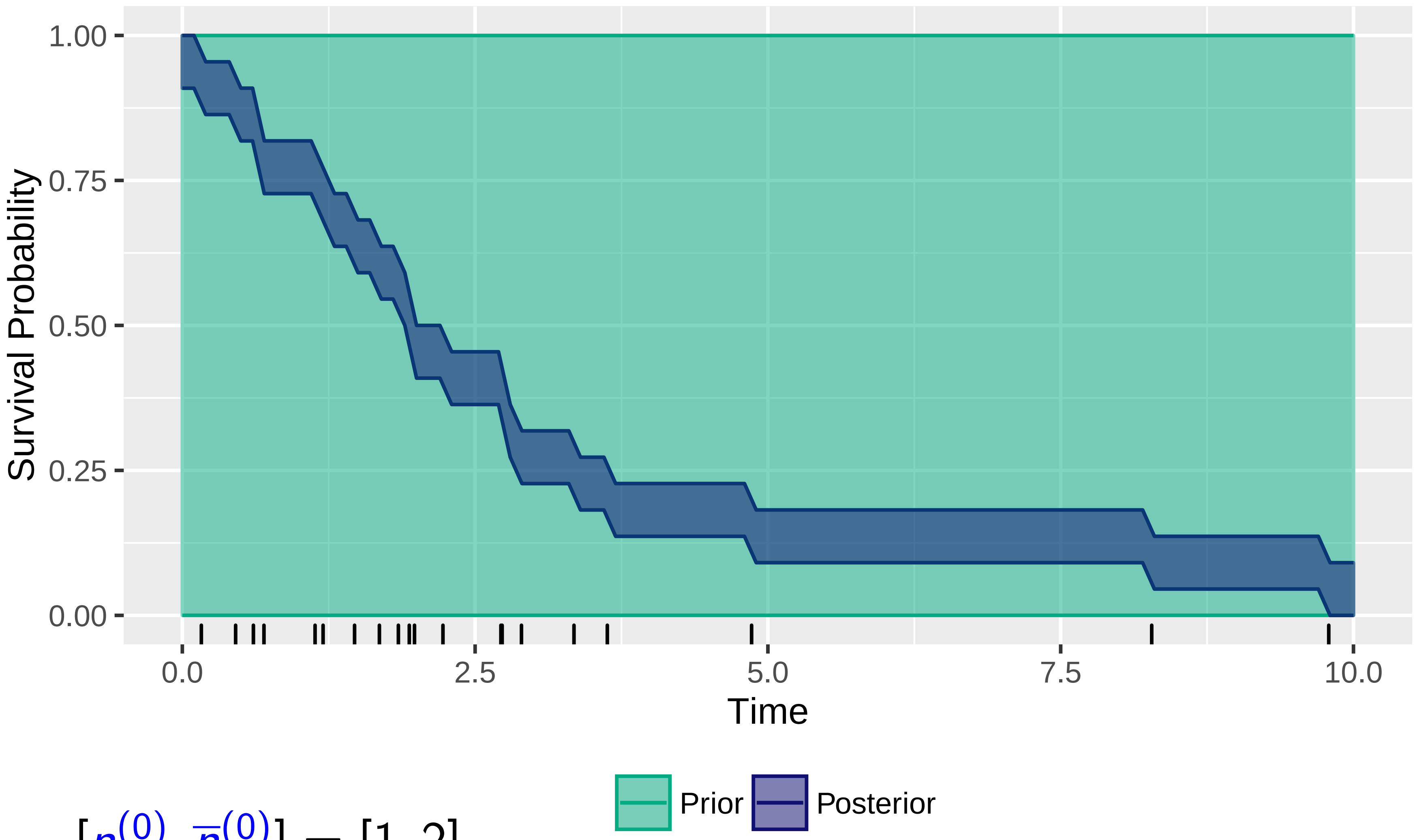
  number of type k components functioning at t:
  $S_t^k \mid p_t^k \sim \text{Binomial}(p_t^k, n_k)$

Component Reliability with Sets of Priors



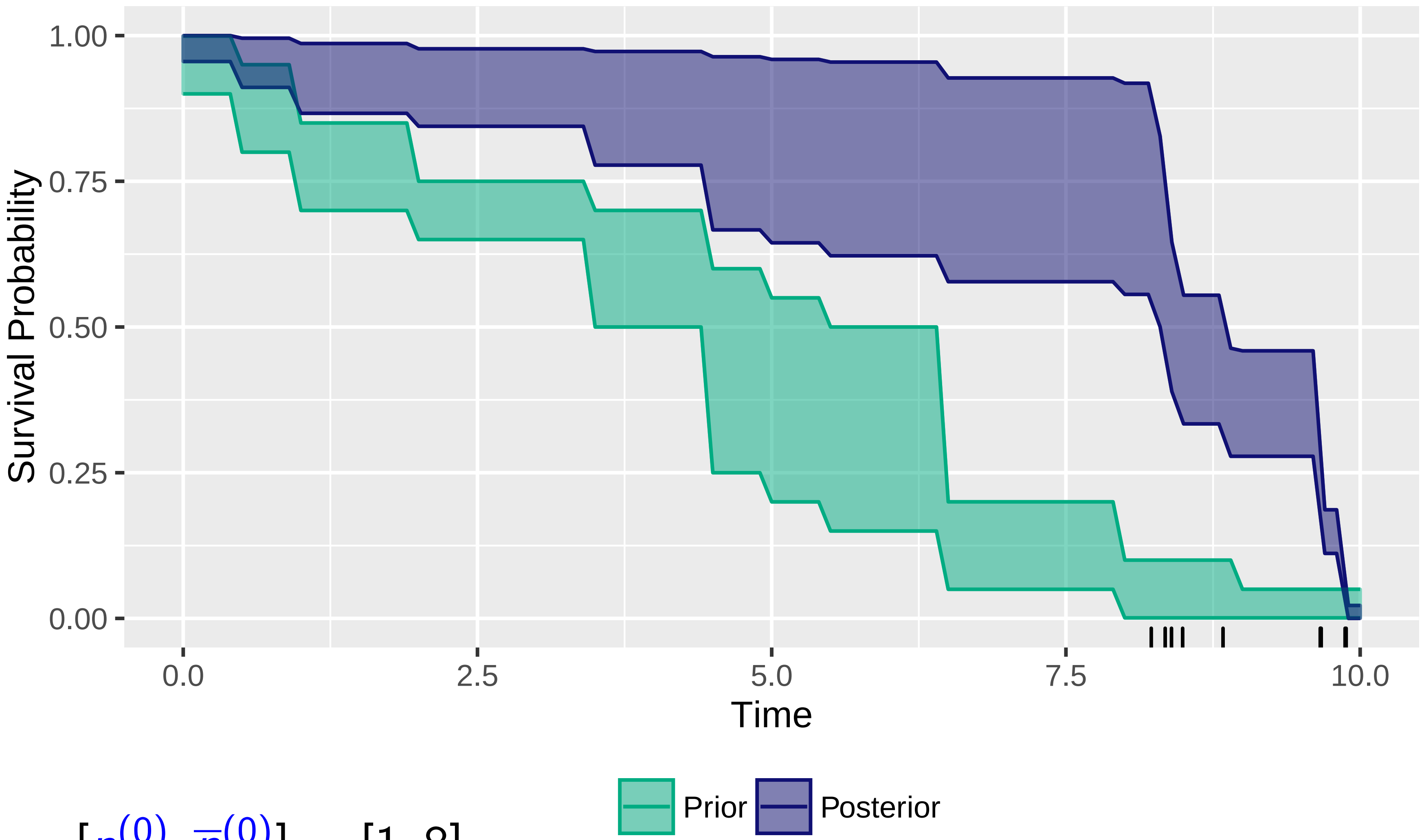
$$[\underline{n}^{(0)}, \overline{n}^{(0)}] = [1, 2]$$
$$[\underline{y}^{(0)}, \overline{y}^{(0)}] = \updownarrow$$

Component Reliability with Sets of Priors



$$[\underline{n}^{(0)}, \overline{n}^{(0)}] = [1, 2]$$
$$[\underline{y}^{(0)}, \overline{y}^{(0)}] = (0, 1)$$

Component Reliability with Sets of Priors



$[\underline{n}^{(0)}, \overline{n}^{(0)}] = [1, 8]$

$[\underline{y}^{(0)}, \overline{y}^{(0)}] = \updownarrow$

System Reliability

- Closed form for the system reliability via the survival signature:

$$R_{\text{sys}}\left(t \mid \bigcup_{k=1}^{K}\{n_{k,t}^{(0)}, y_{k,t}^{(0)}, \boldsymbol{t}^k\}\right) = P(T_{\text{sys}} > t \mid \cdots)$$

$$= \sum_{l_1=0}^{m_1} \cdots \sum_{l_K=0}^{m_K} \Phi(l_1, \ldots, l_K) \prod_{k=1}^{K} P(C_t^k = l_k \mid n_{k,t}^{(0)}, y_{k,t}^{(0)}, \boldsymbol{t}^k)$$

System Reliability

- Closed form for the system reliability via the survival signature:

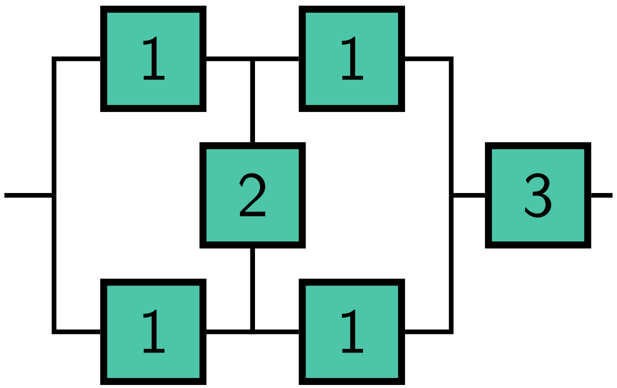
$$R_{\text{sys}}\Big(t \mid \cup_{k=1}^{K}\{n_{k,t}^{(0)}, y_{k,t}^{(0)}, \boldsymbol{t}^k\}\Big) = P(T_{\text{sys}} > t \mid \cdots)$$

$$= \sum_{l_1=0}^{m_1} \cdots \sum_{l_K=0}^{m_K} \Phi(l_1, \ldots, l_K) \prod_{k=1}^{K} P(C_t^k = l_k \mid n_{k,t}^{(0)}, y_{k,t}^{(0)}, \boldsymbol{t}^k)$$

Survival signature [8] $\Phi(l_1, \ldots, l_K) =$
$P(\text{system functions} \mid \{l_k \text{ k's function}\}^{1:K})$

l_1	l_2	l_3	Φ
0	0	1	0
1	0	1	0
2	0	1	1/3
3	0	1	1
4	0	1	1

l_1	l_2	l_3	Φ
0	1	1	0
1	1	1	0
2	1	1	2/3
3	1	1	1
4	1	1	1



System Reliability

- Closed form for the system reliability via the survival signature:

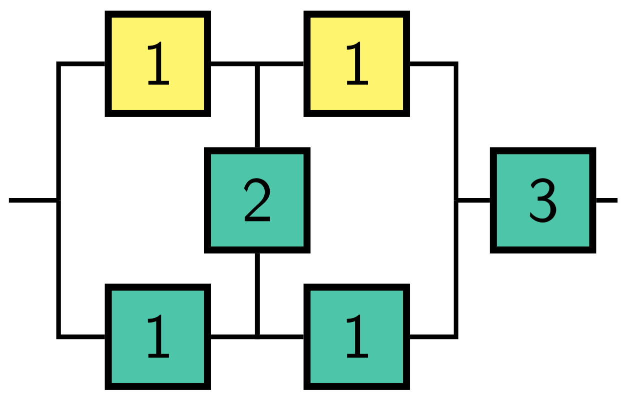
$$R_{\text{sys}}\Big(t \mid \bigcup_{k=1}^{K}\{n_{k,t}^{(0)}, y_{k,t}^{(0)}, \boldsymbol{t}^k\}\Big) = P(T_{\text{sys}} > t \mid \cdots)$$

$$= \sum_{l_1=0}^{m_1} \cdots \sum_{l_K=0}^{m_K} \Phi(l_1, \ldots, l_K) \prod_{k=1}^{K} P(C_t^k = l_k \mid n_{k,t}^{(0)}, y_{k,t}^{(0)}, \boldsymbol{t}^k)$$

Survival signature [8] $\Phi(l_1, \ldots, l_K) =$
$P(\text{system functions} \mid \{l_k \text{ k's function}\}^{1:K})$

l_1	l_2	l_3	Φ	l_1	l_2	l_3	Φ
0	0	1	0	0	1	1	0
1	0	1	0	1	1	1	0
2	0	1	1/3	2	1	1	2/3
3	0	1	1	3	1	1	1
4	0	1	1	4	1	1	1

✓



System Reliability

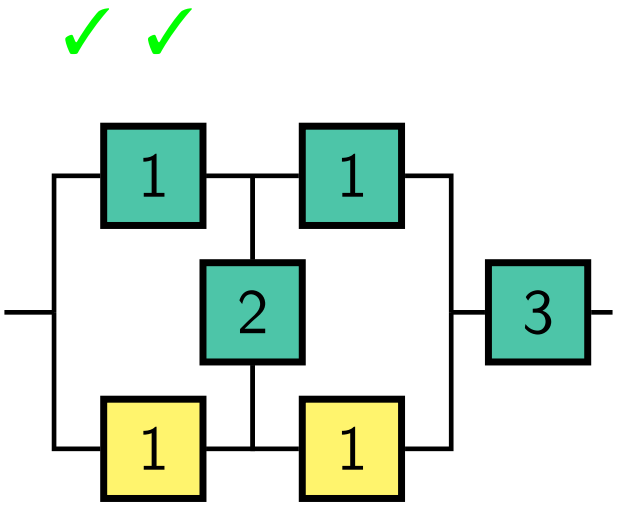
- Closed form for the system reliability via the survival signature:

$$R_{\text{sys}}\Big(t \mid \textstyle\bigcup_{k=1}^{K}\{n_{k,t}^{(0)}, y_{k,t}^{(0)}, \boldsymbol{t}^k\}\Big) = P(T_{\text{sys}} > t \mid \cdots)$$

$$= \sum_{l_1=0}^{m_1} \cdots \sum_{l_K=0}^{m_K} \Phi(l_1, \ldots, l_K) \prod_{k=1}^{K} P(C_t^k = l_k \mid n_{k,t}^{(0)}, y_{k,t}^{(0)}, \boldsymbol{t}^k)$$

Survival signature [8] $\Phi(l_1, \ldots, l_K) =$
$P(\text{system functions} \mid \{l_k \text{ k's function}\}^{1:K})$

l_1	l_2	l_3	Φ	l_1	l_2	l_3	Φ
0	0	1	0	0	1	1	0
1	0	1	0	1	1	1	0
2	0	1	1/3	2	1	1	2/3
3	0	1	1	3	1	1	1
4	0	1	1	4	1	1	1



System Reliability

- Closed form for the system reliability via the survival signature:

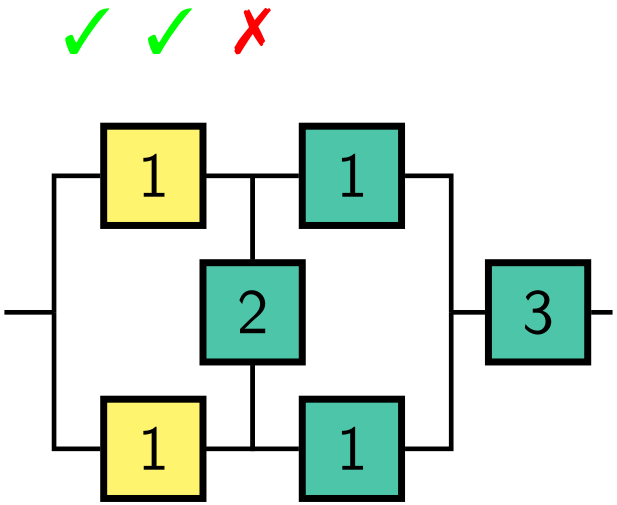
$$R_{\text{sys}}\left(t \mid \bigcup_{k=1}^{K}\{n_{k,t}^{(0)}, y_{k,t}^{(0)}, \boldsymbol{t}^k\}\right) = P(T_{\text{sys}} > t \mid \cdots)$$

$$= \sum_{l_1=0}^{m_1} \cdots \sum_{l_K=0}^{m_K} \Phi(l_1, \ldots, l_K) \prod_{k=1}^{K} P(C_t^k = l_k \mid n_{k,t}^{(0)}, y_{k,t}^{(0)}, \boldsymbol{t}^k)$$

Survival signature [8] $\Phi(l_1, \ldots, l_K) =$
$P(\text{system functions} \mid \{l_k$ k 's function$\}^{1:K})$

l_1	l_2	l_3	Φ
0	0	1	0
1	0	1	0
2	0	1	1/3
3	0	1	1
4	0	1	1

l_1	l_2	l_3	Φ
0	1	1	0
1	1	1	0
2	1	1	2/3
3	1	1	1
4	1	1	1



System Reliability

- Closed form for the system reliability via the survival signature:

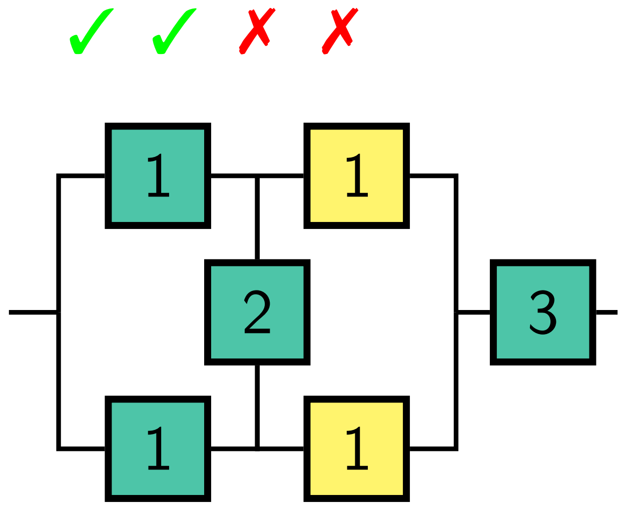
$$R_{\text{sys}}\Big(t \mid \textstyle\bigcup_{k=1}^{K}\{n_{k,t}^{(0)}, y_{k,t}^{(0)}, \boldsymbol{t}^k\}\Big) = P(T_{\text{sys}} > t \mid \cdots)$$

$$= \sum_{l_1=0}^{m_1} \cdots \sum_{l_K=0}^{m_K} \Phi(l_1, \ldots, l_K) \prod_{k=1}^{K} P(C_t^k = l_k \mid n_{k,t}^{(0)}, y_{k,t}^{(0)}, \boldsymbol{t}^k)$$

Survival signature [8] $\Phi(l_1, \ldots, l_K) =$
$P(\text{system functions} \mid \{l_k \text{ k's function}\}^{1:K})$

l_1	l_2	l_3	Φ
0	0	1	0
1	0	1	0
2	0	1	1/3
3	0	1	1
4	0	1	1

l_1	l_2	l_3	Φ
0	1	1	0
1	1	1	0
2	1	1	2/3
3	1	1	1
4	1	1	1



System Reliability

- Closed form for the system reliability via the survival signature:

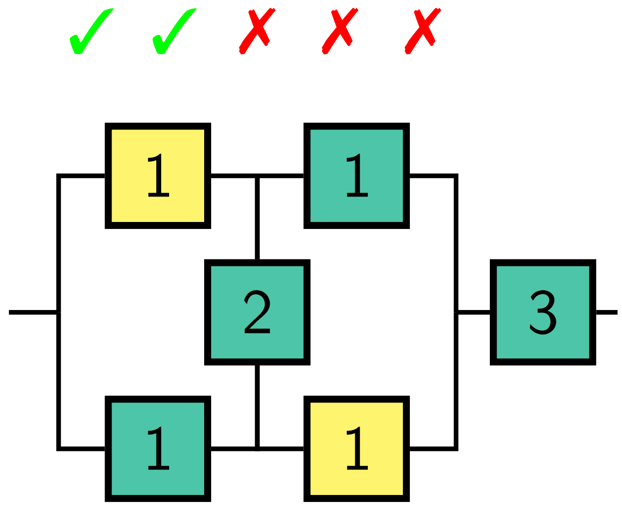
$$R_{\text{sys}}\left(t \mid \bigcup_{k=1}^{K}\{n_{k,t}^{(0)}, y_{k,t}^{(0)}, \boldsymbol{t}^k\}\right) = P(T_{\text{sys}} > t \mid \cdots)$$

$$= \sum_{l_1=0}^{m_1} \cdots \sum_{l_K=0}^{m_K} \Phi(l_1, \ldots, l_K) \prod_{k=1}^{K} P(C_t^k = l_k \mid n_{k,t}^{(0)}, y_{k,t}^{(0)}, \boldsymbol{t}^k)$$

Survival signature [8] $\Phi(l_1, \ldots, l_K) =$
$P(\text{system functions} \mid \{l_k \text{ k's function}\}^{1:K})$

l_1	l_2	l_3	Φ
0	0	1	0
1	0	1	0
2	0	1	1/3
3	0	1	1
4	0	1	1

l_1	l_2	l_3	Φ
0	1	1	0
1	1	1	0
2	1	1	2/3
3	1	1	1
4	1	1	1



System Reliability

- Closed form for the system reliability via the survival signature:

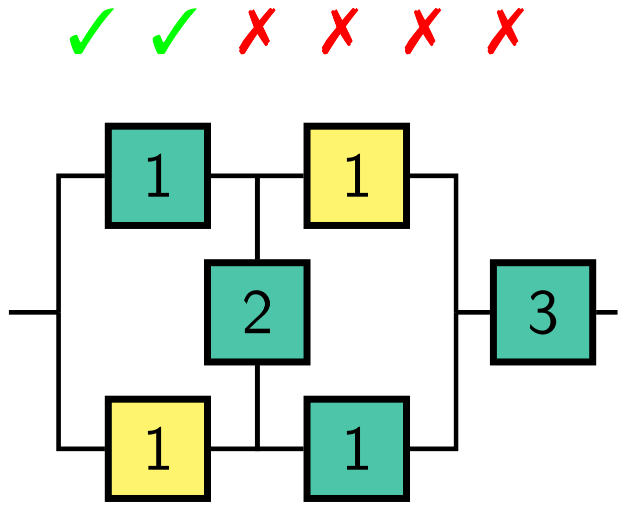
$$R_{\text{sys}}\Big(t \mid \bigcup_{k=1}^{K}\{n_{k,t}^{(0)}, y_{k,t}^{(0)}, \boldsymbol{t}^{k}\}\Big) = P(T_{\text{sys}} > t \mid \cdots)$$

$$= \sum_{l_1=0}^{m_1} \cdots \sum_{l_K=0}^{m_K} \Phi(l_1, \ldots, l_K) \prod_{k=1}^{K} P(C_t^k = l_k \mid n_{k,t}^{(0)}, y_{k,t}^{(0)}, \boldsymbol{t}^{k})$$

Survival signature [8] $\Phi(l_1, \ldots, l_K) =$
$P(\text{system functions} \mid \{l_k \text{ k's function}\}^{1:K})$

l_1	l_2	l_3	Φ
0	0	1	0
1	0	1	0
2	0	1	1/3
3	0	1	1
4	0	1	1

l_1	l_2	l_3	Φ
0	1	1	0
1	1	1	0
2	1	1	2/3
3	1	1	1
4	1	1	1



System Reliability

- Closed form for the system reliability via the survival signature:

$$R_{sys}\Big(t \mid \cup_{k=1}^{K}\{n^{(0)}_{k,t}, y^{(0)}_{k,t}, \boldsymbol{t}^k\}\Big) = P(T_{sys} > t \mid \cdots)$$

$$= \sum_{l_1=0}^{m_1} \cdots \sum_{l_K=0}^{m_K} \Phi(l_1, \ldots, l_K) \prod_{k=1}^{K} P(C_t^k = l_k \mid n^{(0)}_{k,t}, y^{(0)}_{k,t}, \boldsymbol{t}^k)$$

Survival signature [8] $\Phi(l_1, \ldots, l_K) = P(\text{system functions} \mid \{l_k \text{ k's function}\}^{1:K})$

l_1	l_2	l_3	Φ
0	0	1	0
1	0	1	0
2	0	1	1/3
3	0	1	1
4	0	1	1

l_1	l_2	l_3	Φ
0	1	1	0
1	1	1	0
2	1	1	2/3
3	1	1	1
4	1	1	1

Post. pred. probability that in a new system, l_k of the m_k k's function at time t:

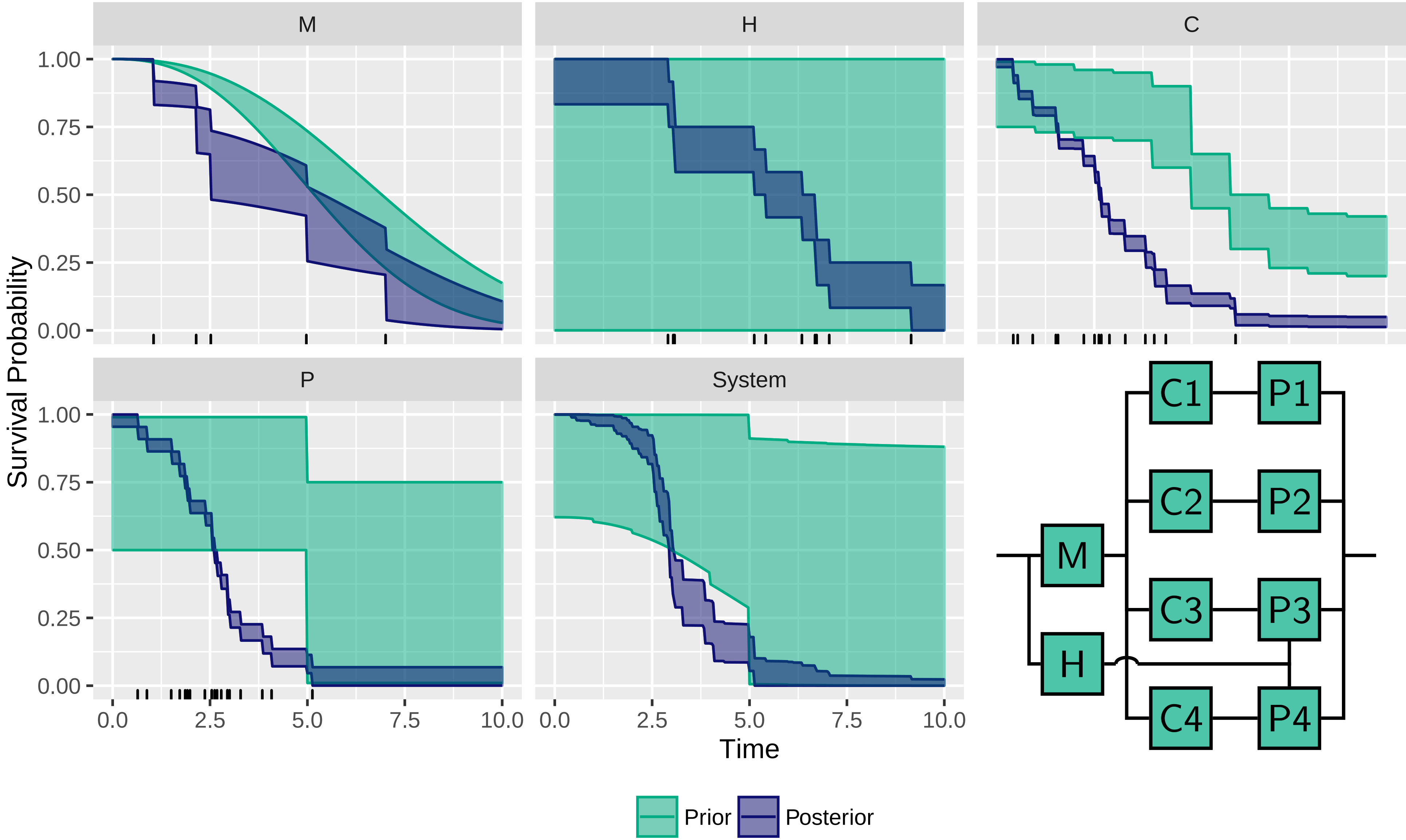
$$\binom{m_k}{l_k} \int [P(T < t \mid p_t^k)]^{l_k} [P(T \geq t \mid p_t^k)]^{m_k - l_k} f(p_t^k \mid n^{(0)}_{k,t}, y^{(0)}_{k,t}, \boldsymbol{t}^k)\, dp_t^k$$

- analytical solution for integral: $C_t^k \mid n^{(0)}_{k,t}, y^{(0)}_{k,t}, \vec{t}^k \sim$Beta-Binom.

System Reliability Bounds

- Bounds for $R_{\text{sys}}\Big(t \mid \bigcup_{k=1}^{K} \big\{n_{k,t}^{(0)}, y_{k,t}^{(0)}, \boldsymbol{t}^k\big\}\Big)$ over $\Pi_{k,t}^{(0)}$, $k = 1, \ldots, K$:
  - min $R_{\text{sys}}(\cdot)$ by $y_{k,t}^{(0)} = \underline{y}_{k,t}^{(0)}$ for any $n_{k,t}^{(0)}$ [30, Theorem 1]
  - min $R_{\text{sys}}(\cdot)$ for $\underline{n}_{k,t}^{(0)}$ or $\overline{n}_{k,t}^{(0)}$ according to simple conditions [30, Theorem 2 & Lemma 3]
  - numeric optimization over $[\underline{n}_{k,t}^{(0)}, \overline{n}_{k,t}^{(0)}]$ in the very few cases where Theorem 2 & Lemma 3 do not apply
  - implemented in **R** package `ReliabilityTheory` [1]

System Reliability Bounds



Robust Bayesian statistics & applications in reliability networks

Outline

Robust Bayesian statistics & applications in reliability networks

Outline

Exercise: Try it yourself!

- Download the development version of the package `ReliabilityTheory`:

  ```
  library("devtools")
  install_github("louisaslett/ReliabilityTheory")
  ```

  (The author is present!)
- You can find a How-To in Appendix B (p. 32) of the preprint paper at `https://arxiv.org/abs/1602.01650`.
- Exercise 1 considers a single component only.
- Exercise 2 considers a system with several types of components.

Exercise: Technical Hints

- Keeping the number of time points low (e.g., $|\mathcal{T}| = 50$) keeps computation time short.
- Setting lower and upper $y^{(0)}_{k,t}$ bounds to exactly 0 or 1 can lead to errors, use, e.g., `1e-5` and `1-1e-5` instead.
- To use `nonParBayesSystemInferencePriorSets()` for a single component, you can set up a system with one component only:

  ```
  onecomp <- graph.formula(s -- 1 -- t)
  onecomp <- setCompTypes(onecomp, list("A" = 1))
  onecompss <- computeSystemSurvivalSignature(onecomp)
  ```

  The component type name is arbitrary, but must be used in the `test.data` argument (as `test.data` is a named list).
- You can use `nonParBayesSystemInferencePriorSets()` to calculate the set of *prior* system reliability functions by setting

  ```
  test.data = list(A = NULL, B = NULL, C = NULL)
  ```

Exercise 1: Single Component Reliability Analysis

- Simulate 10 failure times from Gamma$(5, 1)$.
- Calculate and graph the (sets of) prior and posterior reliability functions for...
  (a) a (discrete) precise prior, where the prior expected reliability values follow `1 - plnorm(t, 1.5, 0.25)`, and the prior strength is 5 for all t.
  (b) a vacuous set of priors, with prior strength interval $[1, 5]$. (Try out other prior strength intervals! Why is the lower bound irrelevant here?)
  (c) a set of priors based on an expert's opinion that the functioning probability is, at time 2, between 0.9 and 0.8, and between 0.6 and 0.3 at time 5. Use $[1, 2]$ and $[10, 20]$ for the prior strength interval.
  (d) a set of priors, where $\underline{y}_{k,t}^{(0)}$ follows `1-pgamma(t, 3, 4)`, $\overline{y}_{k,t}^{(0)}$ follows `1-pgamma(t, 3, 2)`, and the prior strength interval is $[1, 10]$.

Exercise 2: System Reliability Analysis

- Define a system with multiple types of components using `graph.formula()` and `setCompTypes()`. You can take the 'bridge' system, the braking system, or invent your own system.
- For prior sets of component reliability functions, you can use the sets from the previous exercise, or think of your own.
- Simulate test data for the components such that they are, from the viewpoint of the component prior, . . .
  - as expected,
  - surprisingly early,
  - surprisingly late.

  What is the effect on the posterior set of system reliability functions?
- Vary the sample size and the prior strength interval. What is the effect on the posterior set of system reliability functions?