Imprecise probability foundations – Extra material

SIPTA School 2026, Munich, Germany

Erik Quaeghebeur

Eindhoven University of Technology

27 July 2026

Overview

Kick-off

Classical probability theory

Multivariate probability — exercise

Interpretation of probability

Frequentist & Bayesian approaches

Limitations of probability theory

Fréchet's bounds & independence
Aversions of real agents

Probability intervals

Credal sets

The robust Bayesian approach
The completely independent product is not convex

Acceptability & Desirability

Changing what shape $\mathcal{D}$ can take

Interval expectation & probability

Learning: categorical prediction

Overview

Kick-off

Classical probability theory

Interpretation of probability

Limitations of probability theory

Probability intervals

Credal sets

Acceptability & Desirability

Interval expectation & probability

Overview

Multivariate probability — exercise

Consider the (joint) random variable $\boldsymbol{X} = (X_1, X_2, X_3)$ with $\mathcal{X} = \{0, 1\}^3$.
The joint probabilities are determined by the following information:

- X_1 and X_2 are independent
- It holds that $P(X_1 = 0) = 1/2$ and $P(X_2 = 0) = 1/5$
- The conditional probabilities $P(X_3 \mid X_1, X_2)$ are determined by the following table:

x_1	0	0	1	1
x_2	0	1	0	1
$P(X_3 = 0 \mid X_1 = x_1, X_2 = x_2)$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{5}$	$\frac{1}{6}$

Derive that $P(X_1 = 0 \mid X_2 = 0, X_3 = 0) = \frac{5}{8}$.

As part of your calculation, you should derive and write a general expression for $P(X_1 \mid X_2, X_3)$ in terms of the (symbolic) probabilities that are given—before filling in the specific numbers.

Overview

The frequentist approach

- Associated to the frequentist interpretation
- Used in classical statistics

- The **likelihood function** describes the data-generating process:

$$L_{\boldsymbol{x}}(\theta) = P(\boldsymbol{X} = \boldsymbol{x} | \Theta = \theta),$$

where θ represents the process-specific parameters.

The frequentist approach

- Associated to the frequentist interpretation
- Used in classical statistics

- The **likelihood function** describes the data-generating process:

$$L_{\boldsymbol{x}}(\theta) = P(\boldsymbol{X} = \boldsymbol{x} | \Theta = \theta),$$

  where θ represents the process-specific parameters.

- Typical reasoning techniques & tools:
  - Maximum likelihood estimation
  - Confidence intervals
  - Hypothesis tests

The Bayesian approach

- Associated to the subjective interpretation
- Used in Bayesian statistics and uncertainty quantification

- Centered around the use of Bayes's rule for updating belief:

$$\underbrace{P(\Theta = \theta | \boldsymbol{X} = \boldsymbol{x})}_{\textbf{posterior}} \propto L_{\boldsymbol{x}}(\theta) \underbrace{P(\Theta = \theta)}_{\textbf{prior}},$$

The Bayesian approach

- Associated to the subjective interpretation
- Used in Bayesian statistics and uncertainty quantification

- Centered around the use of Bayes's rule for updating belief:

$$\underbrace{P(\Theta = \theta | \boldsymbol{X} = \boldsymbol{x})}_{\textbf{posterior}} \propto L_{\boldsymbol{x}}(\theta)\, \underbrace{P(\Theta = \theta)}_{\textbf{prior}},$$

- Typical reasoning techniques & tools:
  - Posterior mean and maximum a posteriori estimation
  - Credible intervals
  - Probabilities for hypotheses

Overview

Kick-off

Classical probability theory

Interpretation of probability

Limitations of probability theory

- Fréchet's bounds & independence
- Aversions of real agents

Probability intervals

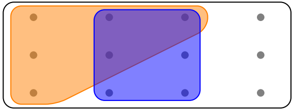
Credal sets

Acceptability & Desirability

Interval expectation & probability

Unknown dependence and Fréchet's bounds

- Nature of dependence between two events A and B often not known



independence

- Fréchet's bounds:

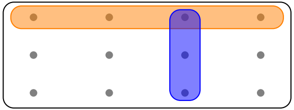
$$P(A)P(B) = P(A \cap B) \in \Big[\max\{0, P(A) + P(B) - 1\}, \min\{P(A), P(B)\}\Big]$$

$$P(A \cup B) \in \Big[\max\{P(A), P(B)\}, \min\{1, P(A) + P(B)\}\Big]$$

- How does independence fit in here?

Unknown dependence and Fréchet's bounds

- Nature of dependence between two events A and B often not known



independence

- Fréchet's bounds:

$$P(A)P(B) = P(A \cap B) \in \left[\max\{0, P(A) + P(B) - 1\}, \min\{P(A), P(B)\}\right]$$

$$P(A \cup B) \in \left[\max\{P(A), P(B)\}, \min\{1, P(A) + P(B)\}\right]$$

- How does independence fit in here?

Are real agents irrational?

Risk aversion

- Urn with 20 red, 20 gray balls
- Two gambles:

	R	G
f_{RG}	\$50	\$50
f_{R}	\$100	\$0

Are real agents irrational?

Risk aversion

- Urn with 20 red, 20 gray balls
- Two gambles:

	R	G
f_{RG}	\$50	\$50
f_{R}	\$100	\$0

- Rational agents are *indifferent* between f_{RG} and f_{R}
- Real agents *strictly prefer* f_{RG} over f_{R}

Are real agents irrational?

Risk aversion

- Urn with 20 red, 20 gray balls
- Two gambles:

	R	G
f_{RG}	\$50	\$50
f_{R}	\$100	\$0

- Rational agents are *indifferent* between f_{RG} and f_{R}
- Real agents *strictly prefer* f_{RG} over f_{R}

Ambiguity aversion

- Two urns; the agent must choose one:
  - E 20 black balls, 20 white balls
  - U unknown proportion of black and white balls
- One gamble:

	B	W
f_{B}	\$100	\$0

Are real agents irrational?

Risk aversion

- Urn with 20 red, 20 gray balls
- Two gambles:

	R	G
f_{RG}	\$50	\$50
f_{R}	\$100	\$0

- Rational agents are *indifferent* between f_{RG} and f_{R}
- Real agents *strictly prefer* f_{RG} over f_{R}

Ambiguity aversion

- Two urns; the agent must choose one:
  - E 20 black balls, 20 white balls
  - U unknown proportion of black and white balls
- One gamble:

	B	W
f_{B}	\$100	\$0

- Rational agents must choose a pmf for U; in case they choose $p_{\mathrm{B}} = p_{\mathrm{W}} = 1/2$ they are *indifferent* between urns E and U
- Real agents *strictly prefer* urn E over urn U

Overview

Overview

- Kick-off
- Classical probability theory
- Interpretation of probability
- Limitations of probability theory
- Probability intervals
- Credal sets
  - The robust Bayesian approach
  - The completely independent product is not convex
- Acceptability & Desirability
- Interval expectation & probability

The robust Bayesian approach

- Generalization of the classical Bayesian approach
- Tries to overcome criticisms of precision associated to the subjective interpretation
- Still centered around the use of Bayes's rule for updating belief:

$$\mathcal{C}^{\Theta|\boldsymbol{x}} = \left\{ P(\Theta|\boldsymbol{X} = \boldsymbol{x}) \propto L_{\boldsymbol{x}}(\Theta)P(\Theta) : P(\Theta) \in \mathcal{C}^{\Theta} \right\},$$

  where
  - $\mathcal{C}^{\Theta}$ is called the *set of priors* (many classes exist)
  - $\mathcal{C}^{\Theta|\boldsymbol{x}}$ is called the *set of posteriors*
- Typical inferences:
  - Lower and upper posterior means
  - Credible intervals (with lower bound guarantees)
  - Probability bounds for hypotheses

The completely independent product is not convex

Problem setup

- Random variables X_1 and X_2 with outcome spaces $\mathcal{X}_1 = \{0, 1\}$ and $\mathcal{X}_2 = \{-, +\}$
- Linear-vacuous marginal credal sets

$$\mathcal{C}^{X_1} := \mathcal{C}^{p^{X_1}, \frac{1}{5}} \text{ with } p^{X_1} = \left(\tfrac{1}{2}, \tfrac{1}{2}\right) \quad \text{and} \quad \mathcal{C}^{X_2} := \mathcal{C}^{p^{X_2}, \frac{1}{4}} \text{ with } p^{X_2} = \left(\tfrac{1}{3}, \tfrac{2}{3}\right)$$

The completely independent product is not convex

Problem setup

- Random variables X_1 and X_2 with outcome spaces $\mathcal{X}_1 = \{0, 1\}$ and $\mathcal{X}_2 = \{-, +\}$
- Linear-vacuous marginal credal sets

$$\mathcal{C}^{X_1} := \mathcal{C}^{p^{X_1}, \frac{1}{5}} \text{ with } p^{X_1} = (\tfrac{1}{2}, \tfrac{1}{2}) \quad \text{and} \quad \mathcal{C}^{X_2} := \mathcal{C}^{p^{X_2}, \frac{1}{4}} \text{ with } p^{X_2} = (\tfrac{1}{3}, \tfrac{2}{3})$$

Goal

Show that their completely independent joint credal set is not convex

The completely independent product is not convex

Problem setup

- Random variables X_1 and X_2 with outcome spaces $\mathcal{X}_1 = \{0, 1\}$ and $\mathcal{X}_2 = \{-, +\}$
- Linear-vacuous marginal credal sets

$$\mathcal{C}^{X_1} := \mathcal{C}^{p^{X_1}, \frac{1}{5}} \text{ with } p^{X_1} = \left(\tfrac{1}{2}, \tfrac{1}{2}\right) \quad \text{and} \quad \mathcal{C}^{X_2} := \mathcal{C}^{p^{X_2}, \frac{1}{4}} \text{ with } p^{X_2} = \left(\tfrac{1}{3}, \tfrac{2}{3}\right)$$

Goal

Show that their completely independent joint credal set is not convex

Demonstration

- We construct a counterexample to convexity, i.e., we construct a convex mixture of elements of the joint that lies outside the joint because it does not factorize

The completely independent product is not convex

Problem setup

- Random variables X_1 and X_2 with outcome spaces $\mathcal{X}_1 = \{0, 1\}$ and $\mathcal{X}_2 = \{-, +\}$
- Linear-vacuous marginal credal sets

$$\mathcal{C}^{X_1} := \mathcal{C}^{p^{X_1}, \frac{1}{5}} \text{ with } p^{X_1} = (\tfrac{1}{2}, \tfrac{1}{2}) \quad \text{and} \quad \mathcal{C}^{X_2} := \mathcal{C}^{p^{X_2}, \frac{1}{4}} \text{ with } p^{X_2} = (\tfrac{1}{3}, \tfrac{2}{3})$$

Goal

Show that their completely independent joint credal set is not convex

Demonstration

- We construct a counterexample to convexity, i.e., we construct a convex mixture of elements of the joint that lies outside the joint because it does not factorize
- An element q of $\mathcal{C}^{\boldsymbol{X}} = \mathcal{C}^{X_1} \otimes \mathcal{C}^{X_2}$ is defined for every $r^{X_1} \in \mathcal{P}_{X_1}$ and $r^{X_2} \in \mathcal{P}_{X_2}$ by

$$q_{(x_1,x_2)} = (q^{X_1} \otimes q^{X_2})_{(x_1,x_2)} = q^{X_1}_{x_1} q^{X_2}_{x_2} = (\tfrac{4}{5} p^{X_1}_{x_1} + \tfrac{1}{5} r^{X_1}_{x_1})(\tfrac{3}{4} p^{X_2}_{x_2} + \tfrac{1}{4} r^{X_2}_{x_2})$$

The completely independent product is not convex

Problem setup

- Random variables X_1 and X_2 with outcome spaces $\mathcal{X}_1 = \{0, 1\}$ and $\mathcal{X}_2 = \{-, +\}$
- Linear-vacuous marginal credal sets

$$\mathcal{C}^{X_1} := \mathcal{C}^{p^{X_1}, \frac{1}{5}} \text{ with } p^{X_1} = (\tfrac{1}{2}, \tfrac{1}{2}) \quad \text{and} \quad \mathcal{C}^{X_2} := \mathcal{C}^{p^{X_2}, \frac{1}{4}} \text{ with } p^{X_2} = (\tfrac{1}{3}, \tfrac{2}{3})$$

Goal

Show that their completely independent joint credal set is not convex

Demonstration

$r_0^{X_1} = r_-^{X_2} = 0$

	0	1	q^{X_2}
$-$			$1/4$
$+$			$3/4$
q^{X_1}	$2/5$	$3/5$	

$r_0^{X_1} = r_-^{X_2} = 1$

	0	1	q^{X_2}
$-$			$1/2$
$+$			$1/2$
q^{X_1}	$3/5$	$2/5$	

The completely independent product is not convex

Problem setup

- Random variables X_1 and X_2 with outcome spaces $\mathcal{X}_1 = \{0, 1\}$ and $\mathcal{X}_2 = \{-, +\}$
- Linear-vacuous marginal credal sets

$$\mathcal{C}^{X_1} := \mathcal{C}^{p^{X_1}, \frac{1}{5}} \text{ with } p^{X_1} = (\tfrac{1}{2}, \tfrac{1}{2}) \quad \text{and} \quad \mathcal{C}^{X_2} := \mathcal{C}^{p^{X_2}, \frac{1}{4}} \text{ with } p^{X_2} = (\tfrac{1}{3}, \tfrac{2}{3})$$

Goal

Show that their completely independent joint credal set is not convex

Demonstration

$r_0^{X_1} = r_-^{X_2} = 0$

	0	1	q^{X_2}
$-$	$^2/_{20}$	$^3/_{20}$	$^1/_4$
$+$	$^6/_{20}$	$^9/_{20}$	$^3/_4$
q^{X_1}	$^2/_5$	$^3/_5$	

$r_0^{X_1} = r_-^{X_2} = 1$

	0	1	q^{X_2}
$-$	$^3/_{10}$	$^2/_{10}$	$^1/_2$
$+$	$^3/_{10}$	$^2/_{10}$	$^1/_2$
q^{X_1}	$^3/_5$	$^2/_5$	

The completely independent product is not convex

Problem setup

- Random variables X_1 and X_2 with outcome spaces $\mathcal{X}_1 = \{0, 1\}$ and $\mathcal{X}_2 = \{-, +\}$
- Linear-vacuous marginal credal sets

$$\mathcal{C}^{X_1} := \mathcal{C}^{p^{X_1}, \frac{1}{5}} \text{ with } p^{X_1} = \left(\tfrac{1}{2}, \tfrac{1}{2}\right) \quad \text{and} \quad \mathcal{C}^{X_2} := \mathcal{C}^{p^{X_2}, \frac{1}{4}} \text{ with } p^{X_2} = \left(\tfrac{1}{3}, \tfrac{2}{3}\right)$$

Goal

Show that their completely independent joint credal set is not convex

Demonstration

$$r_0^{X_1} = r_-^{X_2} = 0 \qquad\qquad r_0^{X_1} = r_-^{X_2} = 1$$

$$\frac{1}{2}\begin{array}{c|cc|c} & 0 & 1 & q^{X_2} \\ \hline - & 2/20 & 3/20 & 1/4 \\ + & 6/20 & 9/20 & 3/4 \\ \hline q^{X_1} & 2/5 & 3/5 & \end{array} + \frac{1}{2}\begin{array}{c|cc|c} & 0 & 1 & q^{X_2} \\ \hline - & 3/10 & 2/10 & 1/2 \\ + & 3/10 & 2/10 & 1/2 \\ \hline q^{X_1} & 3/5 & 2/5 & \end{array} = \begin{array}{c|cc|c} & 0 & 1 & q^{X_2} \\ \hline - & 8/40 & 7/40 & \\ + & 12/40 & 13/40 & \\ \hline q^{X_1} & & & \end{array}$$

The completely independent product is not convex

Problem setup

- Random variables X_1 and X_2 with outcome spaces $\mathcal{X}_1 = \{0, 1\}$ and $\mathcal{X}_2 = \{-, +\}$
- Linear-vacuous marginal credal sets

$$\mathcal{C}^{X_1} := \mathcal{C}^{p^{X_1}, \frac{1}{5}} \text{ with } p^{X_1} = (\tfrac{1}{2}, \tfrac{1}{2}) \quad \text{and} \quad \mathcal{C}^{X_2} := \mathcal{C}^{p^{X_2}, \frac{1}{4}} \text{ with } p^{X_2} = (\tfrac{1}{3}, \tfrac{2}{3})$$

Goal

Show that their completely independent joint credal set is not convex

Demonstration

$r_0^{X_1} = r_-^{X_2} = 0$

$\frac{1}{2}$

	0	1	q^{X_2}
$-$	$2/20$	$3/20$	$1/4$
$+$	$6/20$	$9/20$	$3/4$
q^{X_1}	$2/5$	$3/5$	

$+ \frac{1}{2}$

$r_0^{X_1} = r_-^{X_2} = 1$

	0	1	q^{X_2}
$-$	$3/10$	$2/10$	$1/2$
$+$	$3/10$	$2/10$	$1/2$
q^{X_1}	$3/5$	$2/5$	

$=$

not factorizing

	0	1	q^{X_2}
$-$	$8/40$	$7/40$	$3/8$
$+$	$12/40$	$13/40$	$5/8$
q^{X_1}	$1/2$	$1/2$	

Overview

Changing what shape $\mathcal{D}$ can take

- Replacing the generating rules in Axioms D3 and D4
- Adding additional generating rules

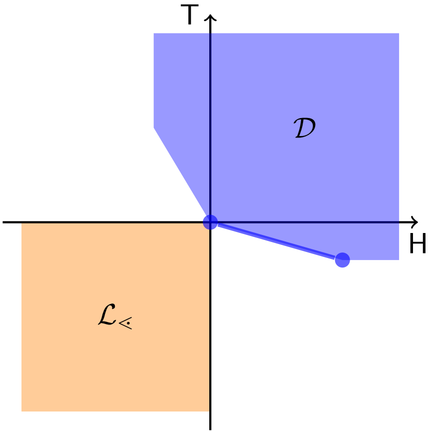
Changing what shape $\mathcal{D}$ can take

- Replacing the generating rules in Axioms D3 and D4
- Adding additional generating rules

- *Example*: replace positive scaling and combination by *convexity*,

  if $f, g \in \mathcal{D}$ and $\mu \in [0, 1]$,
  then $\mu \cdot f + (1 - \mu) \cdot g \in \mathcal{D}$

  This means replacing the *positive linear hull* posi by the *convex hull* co



Overview

Kick-off

Classical probability theory

Interpretation of probability

Limitations of probability theory

Probability intervals

Credal sets

Acceptability & Desirability

Interval expectation & probability

Learning: categorical prediction

The categorical prediction task

What is the color of the next ball to be drawn from an urn?

The categorical prediction task

What is the color of the next ball to be drawn from an urn?

Setup

- Outcome space $\mathcal{X}$ (of ball colors: Red, Green, Black, White,...)
- We have $n \in \mathbb{N}$ observations $\boldsymbol{x} = (x_1, \ldots, x_n)$
  (sequence of colors of previous balls drawn)
- Draw inferences for or make decisions related to the next observation X_{n+1}
  (color of next ball drawn)

The categorical prediction task

What is the color of the next ball to be drawn from an urn?

Setup

- Outcome space $\mathcal{X}$ (of ball colors: Red, Green, Black, White,...)
- We have $n \in \mathbb{N}$ observations $\boldsymbol{x} = (x_1, \dots, x_n)$ (sequence of colors of previous balls drawn)
- Draw inferences for or make decisions related to the next observation X_{n+1} (color of next ball drawn)

Exchangeability assumption

- *The order of the observations is irrelevant* (inferences should remain the same when $(\mathrm{R}, \mathrm{G}, \mathrm{R})$ or $(\mathrm{R}, \mathrm{R}, \mathrm{G})$ has been observed)
- Observed occurrence vector $\boldsymbol{n} \in \mathbb{N}^{\mathcal{X}}$ with $n_z := \sum_{k=1}^{n} \delta_{zx_k}$ and $n_S := \sum_{x \in S} n_x$ (if $\boldsymbol{x} = (\mathrm{R}, \mathrm{G}, \mathrm{R}, \mathrm{W})$ then $n_\mathrm{R} = 2$, $n_\mathrm{G} = 1$, $n_\mathrm{W} = 1$, and $n_{\neg\mathrm{W}} = 3$)

Learning lower expectations for categorical prediction

Idea

Add epistemic uncertainty using $s > 0$ pseudo-observations (of unknown color) and predict according to the 'observed' frequency

Learning lower expectations for categorical prediction

Idea

Add epistemic uncertainty using $s > 0$ pseudo-observations (of unknown color) and predict according to the 'observed' frequency

Predictive inference

- Outcome prediction: $\underline{P}_s(\{x\}|\boldsymbol{n}) = \frac{n_x}{s+n}$ and $\overline{P}_s(\{x\}|\boldsymbol{n}) = \frac{s+n_x}{s+n}$

Learning lower expectations for categorical prediction

Idea

Add epistemic uncertainty using $s > 0$ pseudo-observations (of unknown color) and predict according to the 'observed' frequency

Predictive inference

- Event prediction: $\underline{P}_s(S|\boldsymbol{n}) = \frac{n_S}{s+n}$ and $\overline{P}_s(S|\boldsymbol{n}) = \frac{s+n_S}{s+n}$

Learning lower expectations for categorical prediction

Idea

Add epistemic uncertainty using $s > 0$ pseudo-observations (of unknown color) and predict according to the 'observed' frequency

Predictive inference

- Event prediction: $\underline{P}_s(S|\boldsymbol{n}) = \frac{n_S}{s+n}$ and $\overline{P}_s(S|\boldsymbol{n}) = \frac{s+n_S}{s+n}$
- Lower expectation for $f \in \mathcal{L}$:

$$\underline{E}_s(f|\boldsymbol{n}) = \frac{n}{s+n} E_p(f) + \frac{s}{s+n} \min_{x\in\mathcal{X}} f(x) \quad \text{with } p_x := \frac{n_x}{n}$$

Learning lower expectations for categorical prediction

Idea

Add epistemic uncertainty using $s > 0$ pseudo-observations (of unknown color) and predict according to the ‘observed’ frequency

Predictive inference

- Event prediction: $\underline{P}_s(S|\boldsymbol{n}) = \frac{n_S}{s+n}$ and $\overline{P}_s(S|\boldsymbol{n}) = \frac{s+n_S}{s+n}$
- Lower expectation for $f \in \mathcal{L}$:

$$\underline{E}_s(f|\boldsymbol{n}) = \frac{n}{s+n} E_p(f) + \frac{s}{s+n} \min_{x \in \mathcal{X}} f(x) \quad \text{with } p_x := \frac{n_x}{n}$$

Properties of $\underline{E}_s(\cdot|\boldsymbol{n})$

- linear-vacuous model; vacuous for $n = 0$; more precise with more observations
- does not depend on a specific categorization $\mathcal{X}$
- immediate prediction model of the *imprecise Dirichlet-multinomial model* ID(M)M

Learning lower expectations for categorical prediction

Idea

Add epistemic uncertainty using $s > 0$ pseudo-observations (of unknown color) and predict according to the 'observed' frequency

Predictive inference

- Event prediction: $\underline{P}_s(S|\boldsymbol{n}) = \frac{n_S}{s+n}$ and $\overline{P}_s(S|\boldsymbol{n}) = \frac{s+n_S}{s+n}$
- Lower expectation for $f \in \mathcal{L}$:

$$\underline{E}_s(f|\boldsymbol{n}) = \frac{n}{s+n} E_p(f) + \frac{s}{s+n} \min_{x\in\mathcal{X}} f(x) \quad \text{with } p_x := \frac{n_x}{n}$$

Example setup

- $\boldsymbol{x} = (\mathrm{R}, \mathrm{G}, \mathrm{R}, \mathrm{W}, \mathrm{W})$
- $s = 2$

Inferences

- $\underline{P}_2(\{B, W\}|\boldsymbol{n}) = ?$
- $\overline{P}_2(\{B, W\}|\boldsymbol{n}) = ?$

Learning lower expectations for categorical prediction

Idea

Add epistemic uncertainty using $s > 0$ pseudo-observations (of unknown color) and predict according to the 'observed' frequency

Predictive inference

- Event prediction: $\underline{P}_s(S|\boldsymbol{n}) = \frac{n_S}{s+n}$ and $\overline{P}_s(S|\boldsymbol{n}) = \frac{s+n_S}{s+n}$
- Lower expectation for $f \in \mathcal{L}$:

$$\underline{E}_s(f|\boldsymbol{n}) = \frac{n}{s+n} E_p(f) + \frac{s}{s+n} \min_{x \in \mathcal{X}} f(x) \quad \text{with } p_x := \frac{n_x}{n}$$

Example setup

- $\boldsymbol{x} = (\mathrm{R}, \mathrm{G}, \mathrm{R}, \mathrm{W}, \mathrm{W})$
- $s = 2$

Inferences

- $\underline{P}_2(\{B, W\}|\boldsymbol{n}) = 2/7$
- $\overline{P}_2(\{B, W\}|\boldsymbol{n}) = 4/7$