

Frontmatter

inference

Erik Quaeghebeur and Gert de Cooman

CWI, Amsterdam, the Netherlands and UGent, Ghent, Belgium

July 24, 2014

Introduction

The meaning of *inference* in this lecture

- ▶ *Learning from data*

The meaning of *inference* in this lecture

- ▶ *Learning from data*
- ▶ Learning a *probabilistic model* from data

The meaning of *inference* in this lecture

- ▶ *Learning from data*
- ▶ Learning a *probabilistic model* from data
- ▶ Learning an *imprecise* probabilistic model from data

The meaning of *inference* in this lecture

- ▶ *Learning from data*
- ▶ Learning *a probabilistic model* from data
- ▶ Learning an *imprecise* probabilistic model from data
- ▶ Learning *and updating* an imprecise probabilistic model from data

The meaning of *inference* in this lecture

- ▶ *Learning from data*
- ▶ Learning *a probabilistic model* from data
- ▶ Learning an *imprecise* probabilistic model from data
- ▶ Learning *and updating* an imprecise probabilistic model from data
- ▶ Learning and updating an imprecise probabilistic model from *samples*

The nature of *inference models*

► Induction

Let men be once fully perswaded of these two principles, *that there is nothing in any object, consider'd in itself, which can afford us a reason for drawing a conclusion beyond it; and, that even after the observation of the frequent or constant conjunction of objects, we have no reason to draw any inference concerning any object beyond those of which we have had experience*; I say, let men be once fully convinc'd of these two principles, and this will throw them so loose from all common systems, that they will make no difficulty in receiving any, which may appear the most extraordinary.

Hume in *A treatise of human nature* [1739, §1.3.12, ¶20]

The nature of *inference models*

▶ Induction

- ▶ There is nothing in any object, considered in itself, which can afford us a reason for drawing a conclusion beyond it.
- ▶ Even after the observation of the frequent or constant conjunction of objects, we have no reason to draw any inference concerning any object beyond those of which we have had experience.

▶ So anything goes?

- ▶ *Not* in practice, experience tells us.
- ▶ But yes, we should not dogmatically stick to a given system.
- ▶ And evaluate different systems using the actual data.

The nature of *inference models*

- ▶ Induction
 - ▶ There is nothing in any object, considered in itself, which can afford us a reason for drawing a conclusion beyond it.
 - ▶ Even after the observation of the frequent or constant conjunction of objects, we have no reason to draw any inference concerning any object beyond those of which we have had experience.
- ▶ So anything goes?
 - ▶ *Not* in practice, experience tells us.
 - ▶ But yes, we should not dogmatically stick to a given system.
 - ▶ And evaluate different systems using the actual data.
- ▶ The importance of *assumptions, principles, convenience, and interpretability*
 - ▶ sampling model
 - ▶ exchangeability, iid
 - ▶ invariance under symmetry operations
 - ▶ simplicity, mathematical tractability
 - ▶ specificity, partition exchangeability, . . .

The *plan*

1. Inference for exponential families
 - 1.1 An example to introduce the theory
 - 1.2 An exercise to practice the theory
2. Symmetry considerations
3. Exchangeability and its consequences

Example:
Inference for Poisson samples

Example: the data

Observed numbers of emails arriving in Gert's mailbox between 9am and 10am on ten consecutive Mondays:

2 7 5 3 3 3 1 5 1 2.

Example: the data

Observed numbers of emails arriving in Gert's mailbox between 9am and 10am on ten consecutive Mondays:

2 7 5 3 3 3 1 5 1 2.

Number of samples: $N = 10$.

(Write this down!)

The sample mean: $\bar{x} = 3.2$.

(Write this down!)

Order statistics $x_{(i)}$ with $0 \leq i \leq 10$:

(Write this down!)

1 1 2 2 3 3 3 5 5 7.

Example: the data

Observed numbers of emails arriving in Gert's mailbox between 9am and 10am on ten consecutive Mondays:

2 7 5 3 3 3 1 5 1 2.

Number of samples: $N = 10$. (Write this down!)

The sample mean: $\bar{x} = 3.2$. (Write this down!)

Order statistics $x_{(i)}$ with $0 \leq i \leq 10$: (Write this down!)

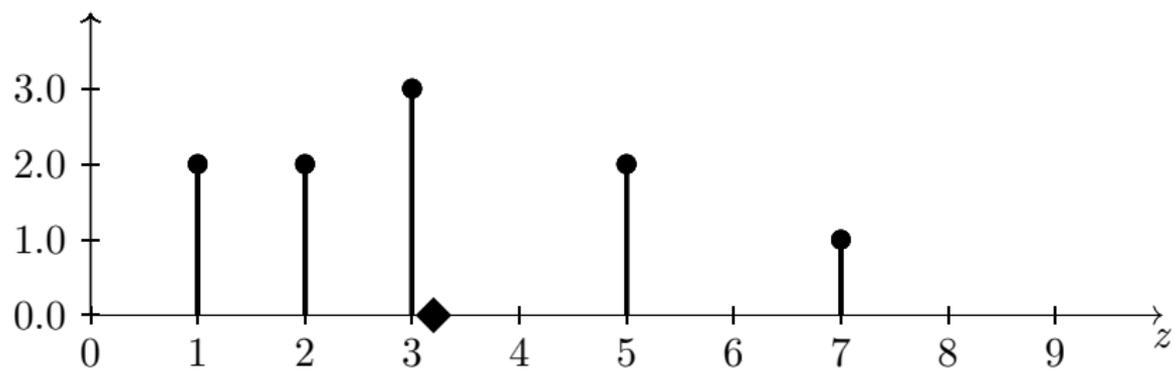
1 1 2 2 3 3 3 5 5 7.

Count and frequency vectors: (Write this down!)

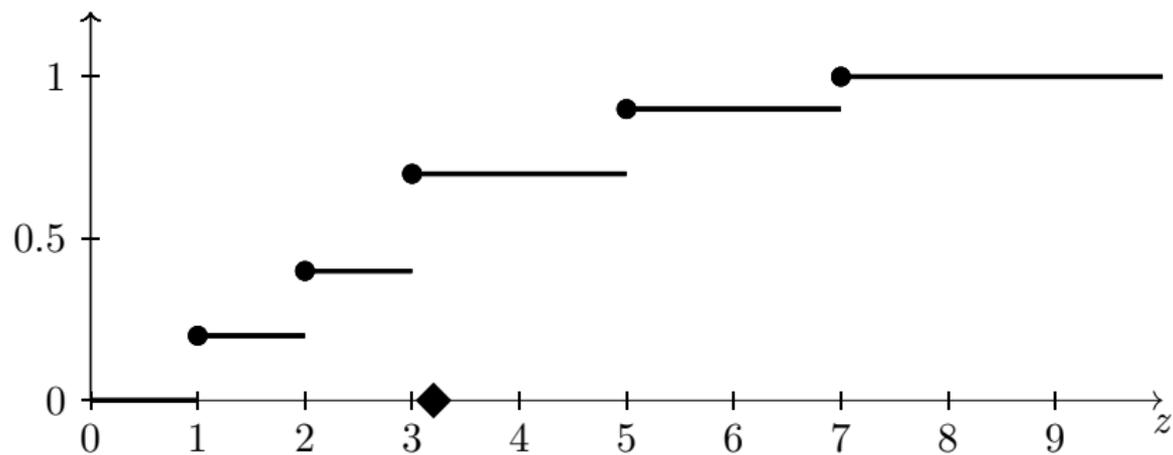
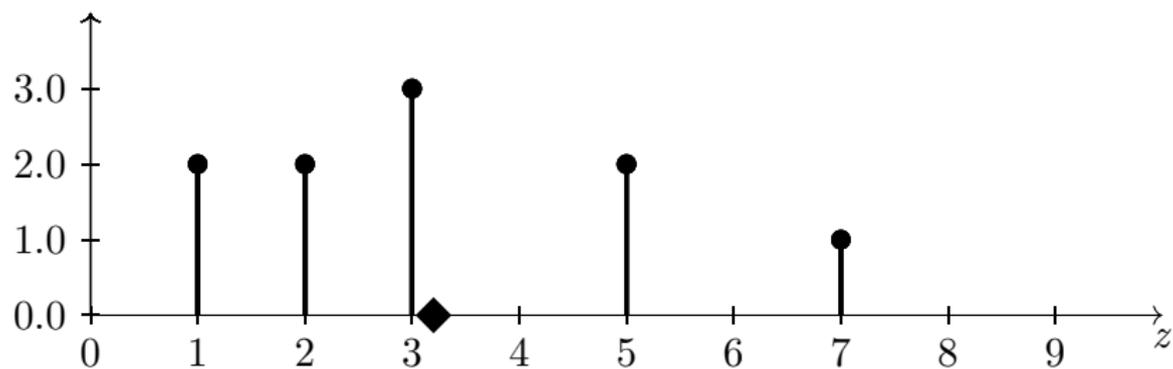
Value z	1	2	3	5	7
Counts n_z	2	2	3	2	1
Frequencies f_z	0.2	0.2	0.3	0.2	0.1

(Assumptions: order in sequence does not matter, ...)

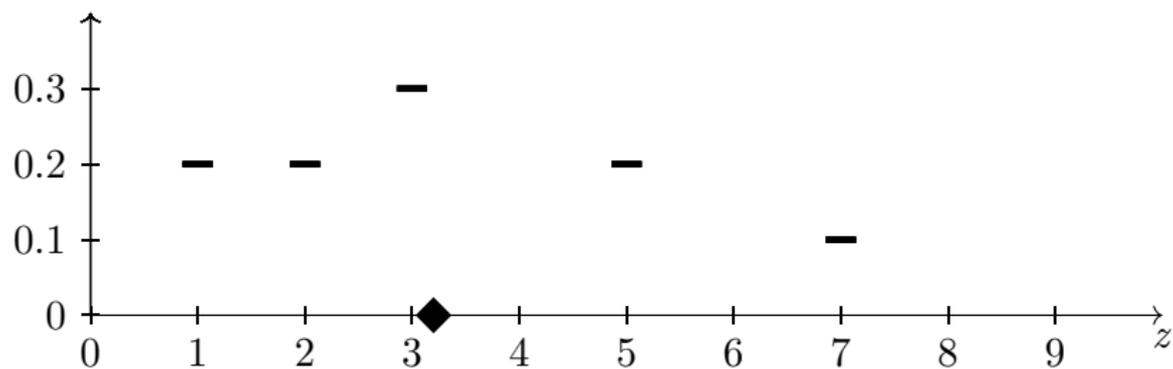
Example: visualizing the data



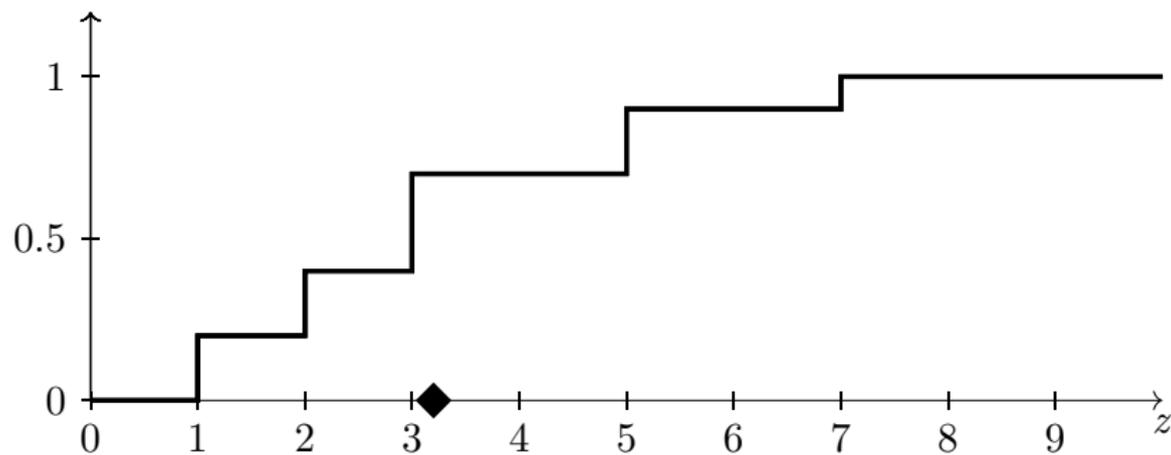
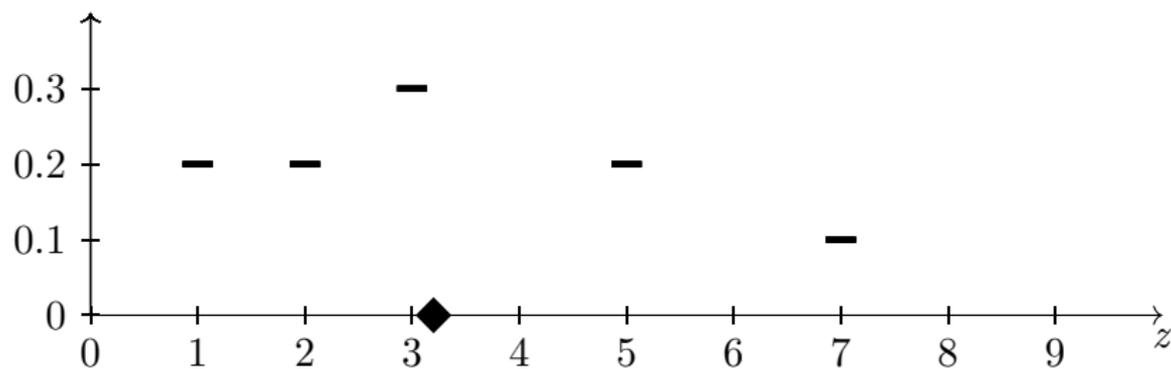
Example: visualizing the data



Example: visualizing the data slightly differently



Example: visualizing the data slightly differently



Distribution-free *immediate predictive* inference

- ▶ Just use the empirical probability mass function f ?

Distribution-free *immediate predictive* inference

- ▶ Just use the empirical probability mass function f ?
- ▶ *Nonparametric Predictive Inference* approach:
Assign equal probability mass between each pair of observations in the order statistics.

Distribution-free *immediate predictive* inference

- ▶ Just use the empirical probability mass function f ?
- ▶ *Nonparametric Predictive Inference* approach:

Assign equal probability mass between each pair of observations in the order statistics.

So assign mass $\frac{1}{N+1}$ to $[x_{(i)}, x_{(i+1)}]$ for $0 \leq i \leq N$, with $x_{(0)} = 0$ and $x_{(N+1)} = +\infty$ (*assumed range*).

Distribution-free *immediate predictive* inference

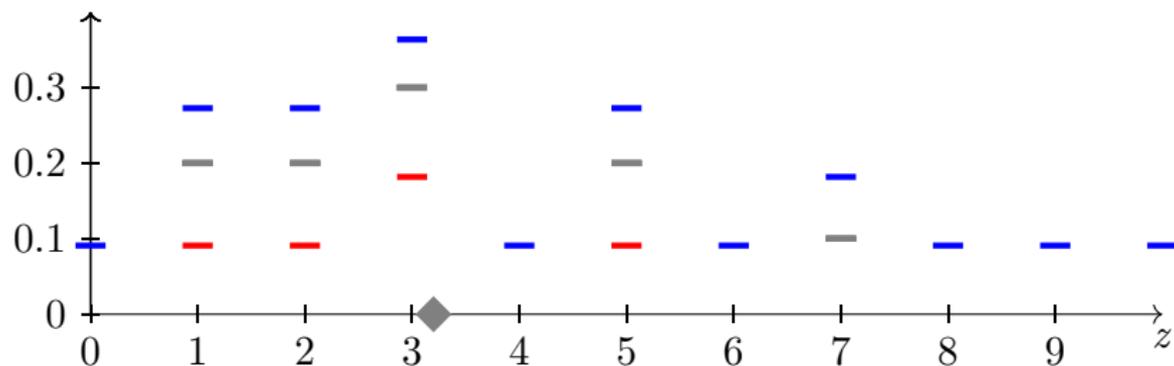
- ▶ Just use the empirical probability mass function f ?
- ▶ *Nonparametric Predictive Inference* approach:

Assign equal probability mass between each pair of observations in the order statistics.

So assign mass $\frac{1}{N+1}$ to $[x_{(i)}, x_{(i+1)}]$ for $0 \leq i \leq N$, with $x_{(0)} = 0$ and $x_{(N+1)} = +\infty$ (*assumed range*).

So, for example, a mass in $[\max\{0, \frac{n_z-1}{N+1}\}, \frac{n_z+1}{N+1}]$ for every value z .

(Do you see why? Can you calculate some values?)



Distribution-free *immediate predictive* inference

- ▶ Just use the empirical probability mass function f ?

Distribution-free *immediate predictive* inference

- ▶ Just use the empirical probability mass function f ?
- ▶ Linear-vacuous model obtained by adding a number $s > 0$ of pseudo-counts or pseudo-observations:

Consider all probability mass functions that result by considering all possible distributions of the pseudo-counts over the possible observation values.

Distribution-free *immediate predictive* inference

- ▶ Just use the empirical probability mass function f ?
- ▶ Linear-vacuous model obtained by adding a number $s > 0$ of pseudo-counts or pseudo-observations:

Consider all probability mass functions that result by considering all possible distributions of the pseudo-counts over the possible observation values.

So we assign a mass in $[\frac{n_z}{N+s}, \frac{n_z+s}{N+s}]$ for every value z .

(Do you see why? Can you calculate some values?)

Distribution-free *immediate predictive* inference

- ▶ Just use the empirical probability mass function f ?
- ▶ Linear-vacuous model obtained by adding a number $s > 0$ of pseudo-counts or pseudo-observations:

Consider all probability mass functions that result by considering all possible distributions of the pseudo-counts over the possible observation values.

So we assign a mass in $[\frac{n_z}{N+s}, \frac{n_z+s}{N+s}]$ for every value z .

(Do you see why? Can you calculate some values?)

Also called ε -contaminated model: $[(1 - \varepsilon)f_z, (1 - \varepsilon)f_z + \varepsilon]$
for every value z , with $\varepsilon = \frac{s}{N+s}$.

Distribution-free *immediate predictive* inference

- ▶ Just use the empirical probability mass function f ?
- ▶ Linear-vacuous model obtained by adding a number $s > 0$ of pseudo-counts or pseudo-observations:

Consider all probability mass functions that result by considering all possible distributions of the pseudo-counts over the possible observation values.

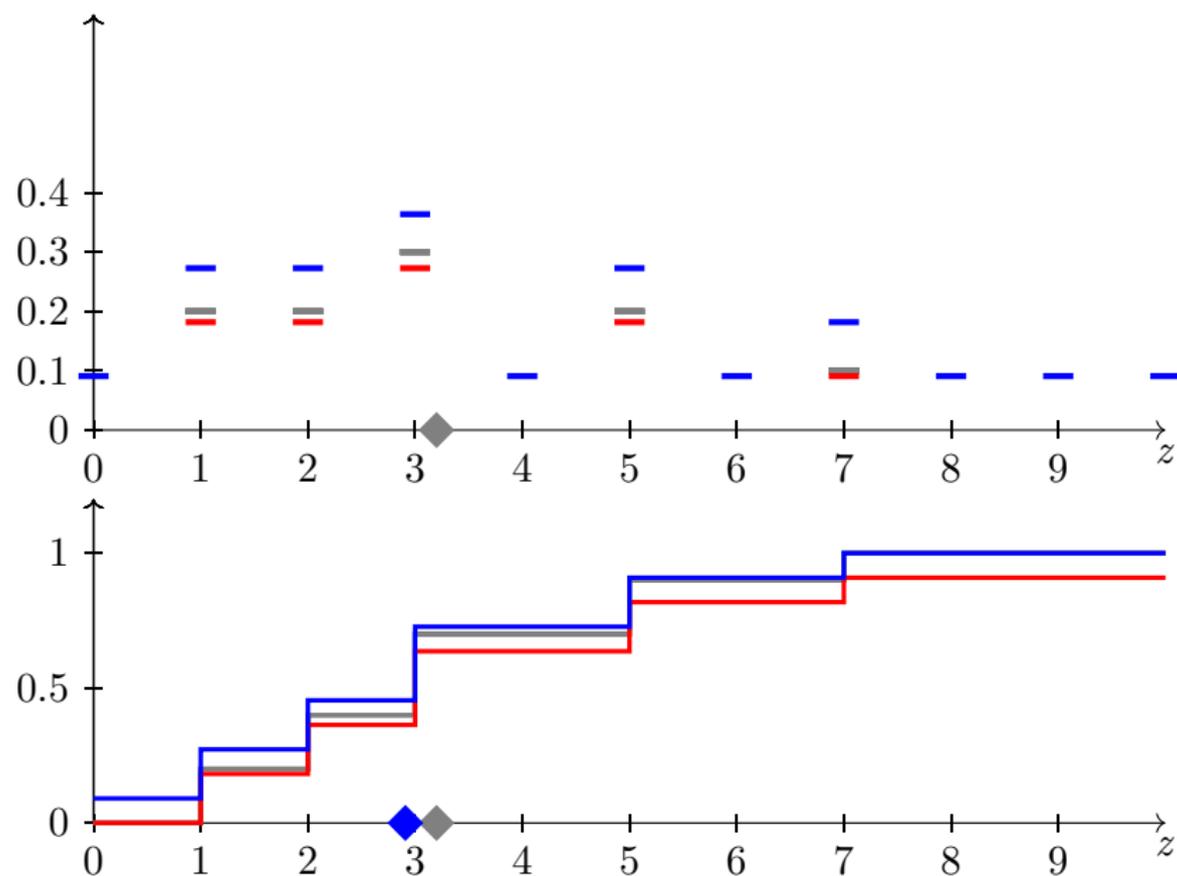
So we assign a mass in $[\frac{n_z}{N+s}, \frac{n_z+s}{N+s}]$ for every value z .

(Do you see why? Can you calculate some values?)

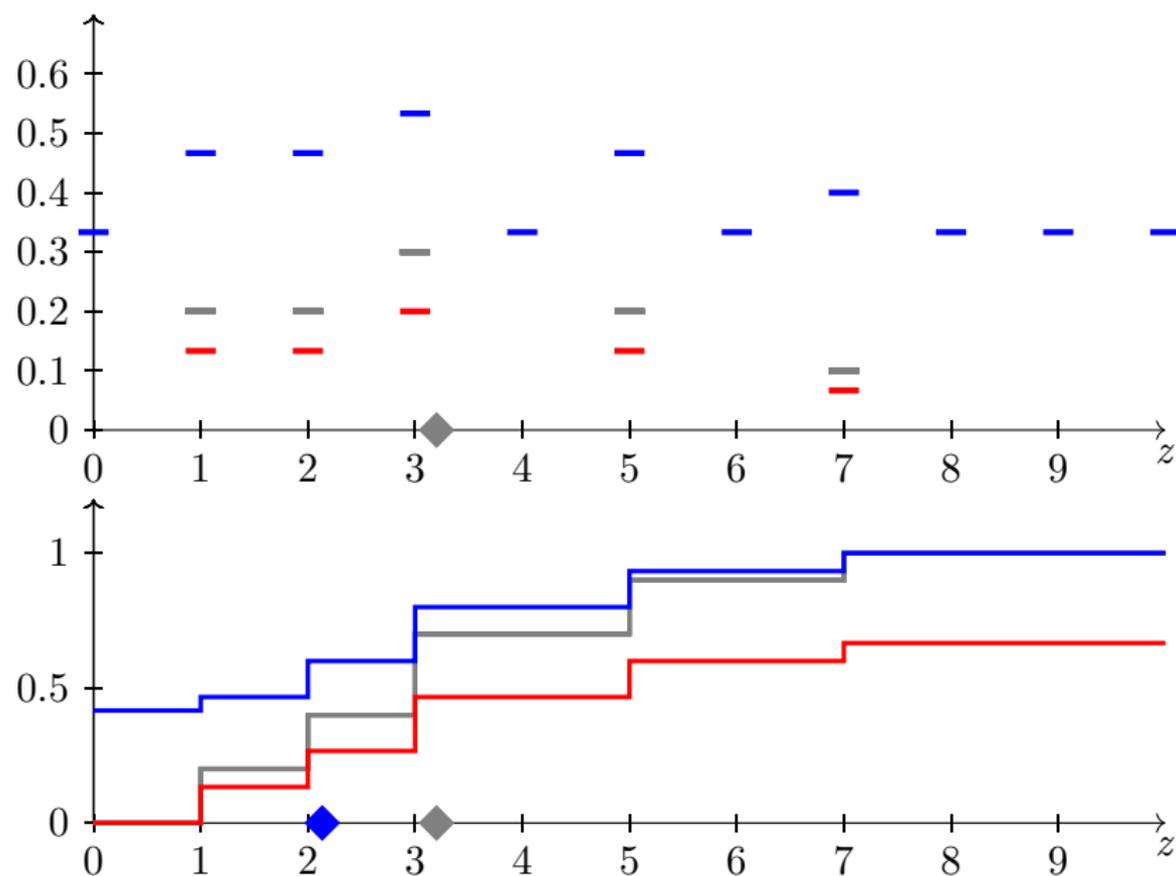
Also called ε -contaminated model: $[(1 - \varepsilon)f_z, (1 - \varepsilon)f_z + \varepsilon]$
for every value z , with $\varepsilon = \frac{s}{N+s}$.

- ▶ Different from NPI, we now use a parameter, s , that regulates the learning rate.

Example: visualizing the linear-vacuous model for $s = 1$



Example: visualizing the linear-vacuous model for $s = 5$



Example: assuming a *Poisson sampling model*

- ▶ A probabilistic model for the number of events occurring in a fixed interval $\Delta t > 0$ (here one hour)
- ▶ Underlying assumptions about event occurrence:
 - ▶ fixed average rate $\dot{\lambda} \geq 0$ (here unknown),
 - ▶ independent of the time since the last event.

Example: assuming a *Poisson sampling model*

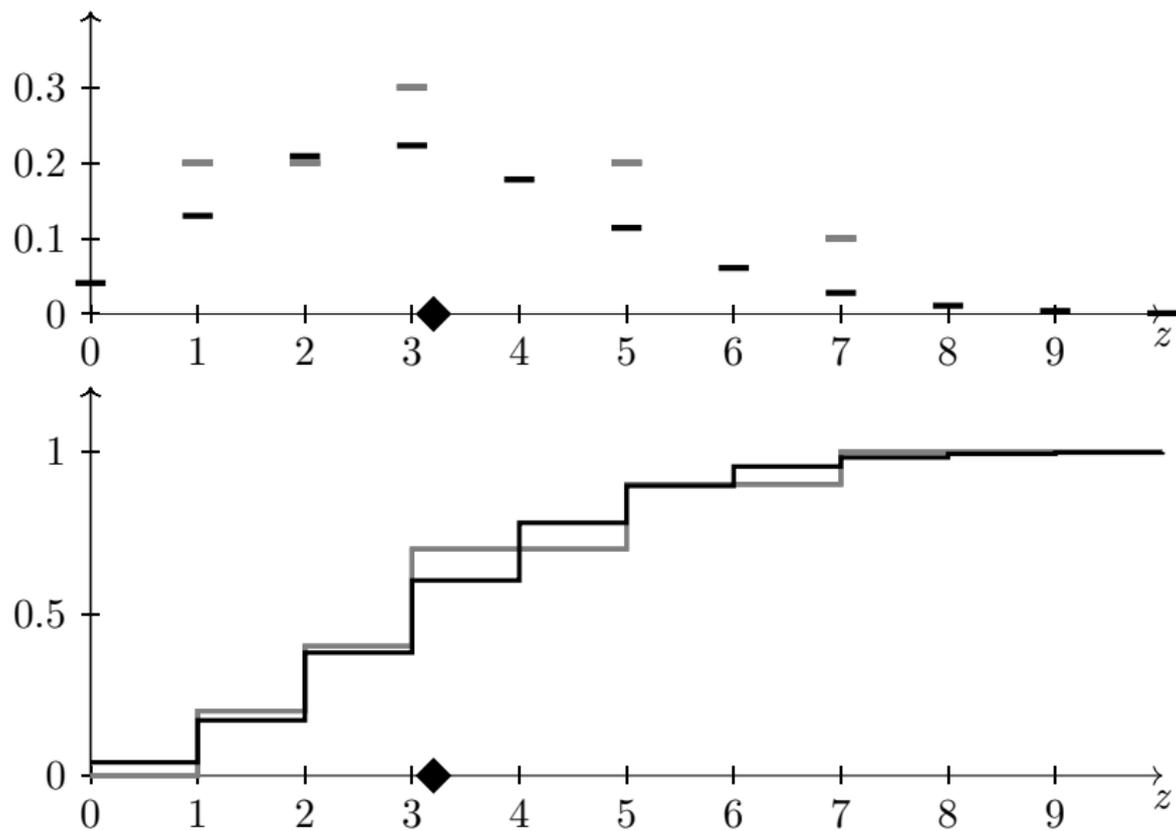
- ▶ A probabilistic model for the number of events occurring in a fixed interval $\Delta t > 0$ (here one hour)
- ▶ Underlying assumptions about event occurrence:
 - ▶ fixed average rate $\dot{\lambda} \geq 0$ (here unknown),
 - ▶ independent of the time since the last event.
- ▶ Probability mass assigned by the *Poisson distribution* with parameter $\lambda := \dot{\lambda}\Delta t$ to any possible number of occurrences $z \geq 0$ is

$$p_{\text{Ps}}(z | \lambda) := \frac{1}{z!} \lambda^z \exp(-\lambda).$$

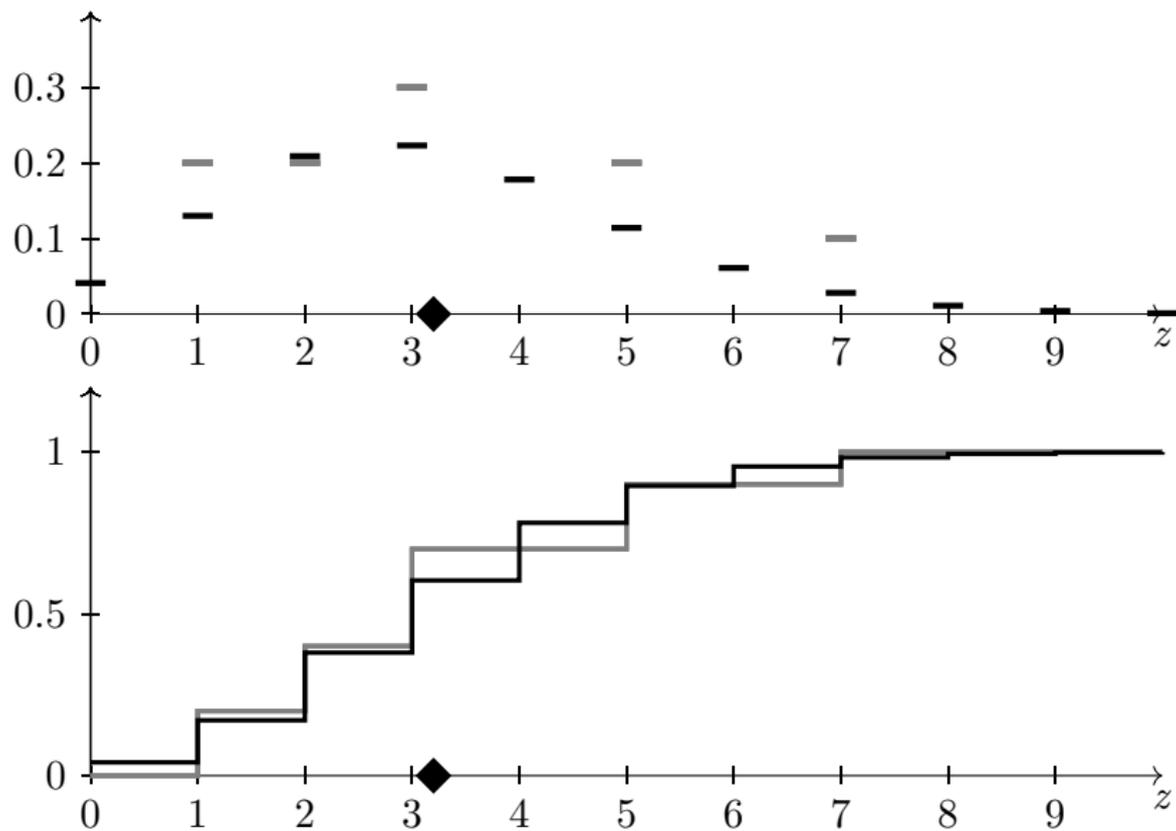
- ▶ The expected number of events in this interval is equal to λ .

(Easiest way to see this?)

Example: visualizing the Poisson distribution ($\lambda = \bar{x} = 3.2$)



Example: visualizing the Poisson distribution ($\lambda = \hat{\lambda} = 3.2$)



Example: the Poisson *likelihood*

- ▶ *Likelihood function*, reversing the roles of the parameter and values in the expression of the probability mass function:

$$L_z(\lambda) := p_{\text{Ps}}(z | \lambda) = \frac{1}{z!} \lambda^z \exp(-\lambda).$$

Example: the Poisson *likelihood*

- ▶ *Likelihood function*, reversing the roles of the parameter and values in the expression of the probability mass function:

$$L_z(\lambda) := p_{\text{Ps}}(z | \lambda) = \frac{1}{z!} \lambda^z \exp(-\lambda).$$

- ▶ *iid assumption*: samples are independently and identically distributed

Example: the Poisson *likelihood*

- ▶ *Likelihood function*, reversing the roles of the parameter and values in the expression of the probability mass function:

$$L_z(\lambda) := p_{\text{Ps}}(z | \lambda) = \frac{1}{z!} \lambda^z \exp(-\lambda).$$

- ▶ *iid assumption*: samples are independently and identically distributed
- ▶ Likelihood function for multi-sample sequence: (Explicit expression?)

$$L_x(\lambda) := p(x | \lambda) = \prod_{i=1}^N p_{\text{Ps}}(x_i | \lambda)$$

Example: the Poisson *likelihood*

- ▶ *Likelihood function*, reversing the roles of the parameter and values in the expression of the probability mass function:

$$L_z(\lambda) := p_{\text{Ps}}(z | \lambda) = \frac{1}{z!} \lambda^z \exp(-\lambda).$$

- ▶ *iid assumption*: samples are independently and identically distributed
- ▶ Likelihood function for multi-sample sequence: (Explicit expression?)

$$\begin{aligned} L_x(\lambda) &:= p(x | \lambda) = \prod_{i=1}^N p_{\text{Ps}}(x_i | \lambda) \\ &= \prod_{i=1}^N L_{x_i}(\lambda) = \frac{1}{\prod_{i=1}^N x_i!} \lambda^{N\bar{x}} \exp(-N\lambda). \end{aligned}$$

Example: the Poisson *likelihood*

- ▶ *Likelihood function*, reversing the roles of the parameter and values in the expression of the probability mass function:

$$L_z(\lambda) := p_{\text{Ps}}(z | \lambda) = \frac{1}{z!} \lambda^z \exp(-\lambda).$$

- ▶ *iid assumption*: samples are independently and identically distributed
- ▶ Likelihood function for multi-sample sequence: (Explicit expression?)

$$\begin{aligned} L_x(\lambda) &:= p(x | \lambda) = \prod_{i=1}^N p_{\text{Ps}}(x_i | \lambda) \\ &= \prod_{i=1}^N L_{x_i}(\lambda) = \frac{1}{\prod_{i=1}^N x_i!} \lambda^{N\bar{x}} \exp(-N\lambda). \end{aligned}$$

- ▶ *Maximum likelihood estimate* $\hat{\lambda} = \bar{x}$. (Can you quickly derive this?)

Example: the Poisson *likelihood*

- ▶ *Likelihood function*, reversing the roles of the parameter and values in the expression of the probability mass function:

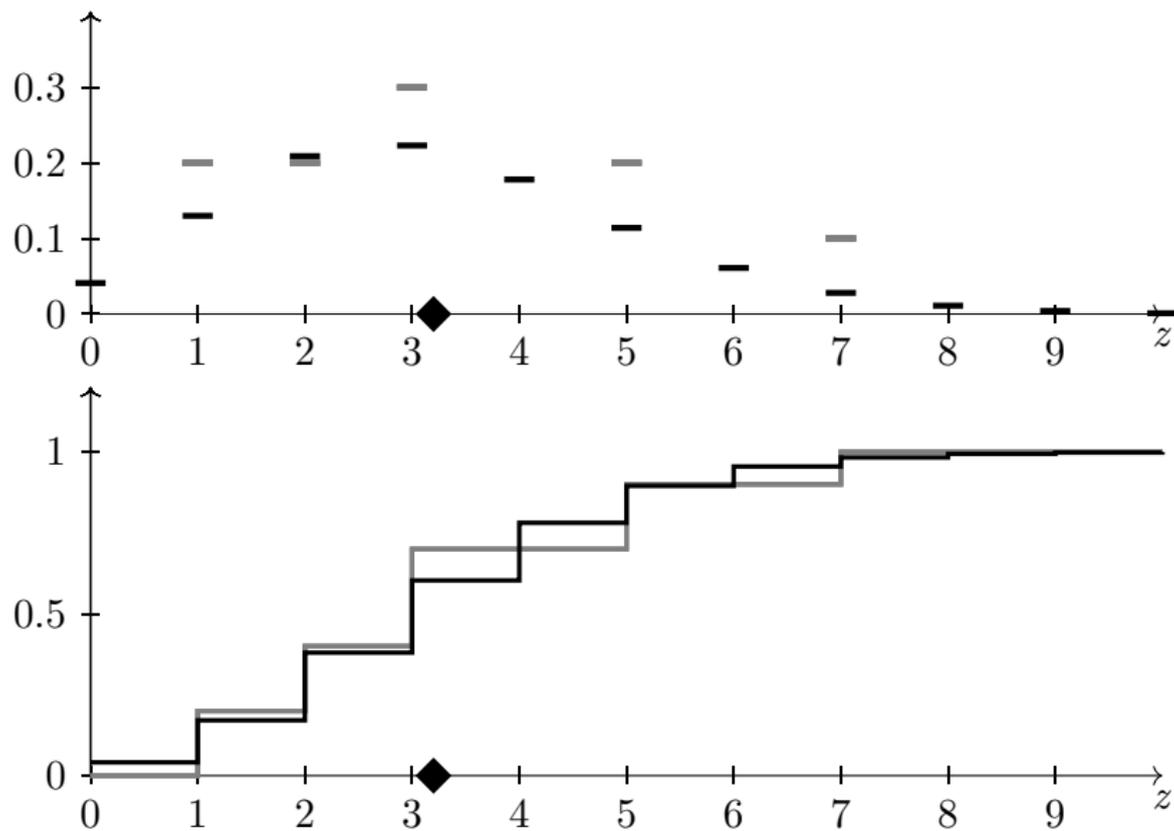
$$L_z(\lambda) := p_{\text{Ps}}(z | \lambda) = \frac{1}{z!} \lambda^z \exp(-\lambda).$$

- ▶ *iid assumption*: samples are independently and identically distributed
- ▶ Likelihood function for multi-sample sequence: (Explicit expression?)

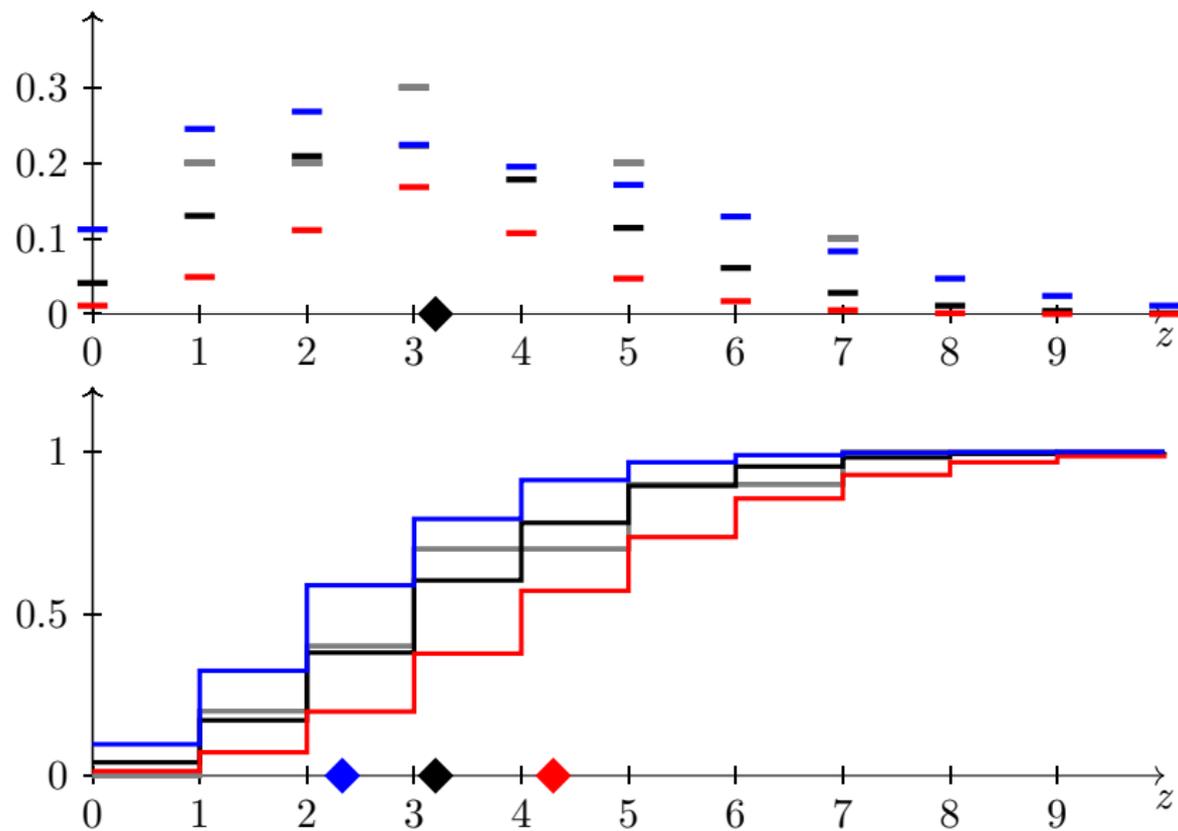
$$\begin{aligned} L_x(\lambda) &:= p(x | \lambda) = \prod_{i=1}^N p_{\text{Ps}}(x_i | \lambda) \\ &= \prod_{i=1}^N L_{x_i}(\lambda) = \frac{1}{\prod_{i=1}^N x_i!} \lambda^{N\bar{x}} \exp(-N\lambda). \end{aligned}$$

- ▶ *Maximum likelihood estimate* $\hat{\lambda} = \bar{x}$. (Can you quickly derive this?)
- ▶ Fixing N , the sample mean \bar{x} is a *sufficient statistic* ($L_x(\lambda) = f_N(\lambda, \bar{x})h(x)$).

Example: visualizing the Poisson distribution ($\lambda = \hat{\lambda} = 3.2$)



Example: heuristic 90% confidence-based imprecise model



Example: Bayesian *parametric* inference

- ▶ Assume given, before observing any samples, a *prior* probability distribution over the possible *parameter* values, with probability density $p(\lambda)$ for all $\lambda \geq 0$.

Example: Bayesian *parametric* inference

- ▶ Assume given, before observing any samples, a *prior* probability distribution over the possible *parameter* values, with probability density $p(\lambda)$ for all $\lambda \geq 0$.
- ▶ Given an observation x , we can *update* the prior by *conditioning* using Bayes's rule for density functions:

$$p(\lambda | x) \propto L_x(\lambda)p(\lambda).$$

We obtain the (parametric) *posterior* density function $p(\cdot | x)$.

Example: Bayesian *parametric* inference

- ▶ Assume given, before observing any samples, a *prior* probability distribution over the possible *parameter* values, with probability density $p(\lambda)$ for all $\lambda \geq 0$.
- ▶ Given an observation x , we can *update* the prior by *conditioning* using Bayes's rule for density functions:

$$p(\lambda | x) \propto L_x(\lambda)p(\lambda).$$

We obtain the (parametric) *posterior* density function $p(\cdot | x)$.

Refresher exercise for (standard) Bayes's rule: Given a uniform prior probability *mass* function for $\lambda \in \{2, 4\}$, the expression for the posterior probability of $\lambda = 2$ is of the form $\frac{1}{1+g(z)}$. Give the expression for g .

Example: Bayesian *parametric* inference

- ▶ Assume given, before observing any samples, a *prior* probability distribution over the possible *parameter* values, with probability density $p(\lambda)$ for all $\lambda \geq 0$.
- ▶ Given an observation x , we can *update* the prior by *conditioning* using Bayes's rule for density functions:

$$p(\lambda | x) \propto L_x(\lambda)p(\lambda).$$

We obtain the (parametric) *posterior* density function $p(\cdot | x)$.

Refresher exercise for (standard) Bayes's rule: Given a uniform prior probability *mass* function for $\lambda \in \{2, 4\}$, the expression for the posterior probability of $\lambda = 2$ is of the form $\frac{1}{1+g(z)}$. Give the expression for g .

Answer: $g(z) = 2^z \exp(-2)$.

Example: Conjugate Bayesian inference — the prior

- ▶ General priors: posterior difficult to obtain in closed analytic form.
- ▶ Mathematical *convenience*: use *conjugate* priors.

Example: Conjugate Bayesian inference — the prior

- ▶ General priors: posterior difficult to obtain in closed analytic form.
- ▶ Mathematical *convenience*: use *conjugate* priors.

Derivation of conjugate prior for the Poisson likelihood:

$$p(\lambda | x) \propto L_x(\lambda)p(\lambda) \propto \lambda^{N\bar{x}} \exp(-N\lambda)p(\lambda).$$

Example: Conjugate Bayesian inference — the prior

- ▶ General priors: posterior difficult to obtain in closed analytic form.
- ▶ Mathematical *convenience*: use *conjugate* priors.

Derivation of conjugate prior for the Poisson likelihood:

$$p(\lambda | x) \propto L_x(\lambda)p(\lambda) \propto \lambda^{N\bar{x}} \exp(-N\lambda)p(\lambda).$$

So take $p(\lambda) \propto \lambda^a \exp(-b\lambda)$.

Example: Conjugate Bayesian inference — the prior

- ▶ General priors: posterior difficult to obtain in closed analytic form.
- ▶ Mathematical *convenience*: use *conjugate* priors.

Derivation of conjugate prior for the Poisson likelihood:

$$p(\lambda | x) \propto L_x(\lambda)p(\lambda) \propto \lambda^{N\bar{x}} \exp(-N\lambda)p(\lambda).$$

So take $p(\lambda) \propto \lambda^a \exp(-b\lambda)$.

Normalization? Gamma integral for $r > 0$ and $k \in \mathbb{N}$: (Write this down!)

$$\Gamma(r) = \int_0^\infty t^{r-1} \exp(-t) dt, \quad \frac{\Gamma(r+1)}{\Gamma(r)} = r, \quad k! = \Gamma(k+1).$$

Example: Conjugate Bayesian inference — the prior

- ▶ General priors: posterior difficult to obtain in closed analytic form.
- ▶ Mathematical *convenience*: use *conjugate* priors.

Derivation of conjugate prior for the Poisson likelihood:

$$p(\lambda | x) \propto L_x(\lambda)p(\lambda) \propto \lambda^{N\bar{x}} \exp(-N\lambda)p(\lambda).$$

So take $p(\lambda) \propto \lambda^a \exp(-b\lambda)$.

Normalization? Gamma integral for $r > 0$ and $k \in \mathbb{N}$: (Write this down!)

$$\Gamma(r) = \int_0^\infty t^{r-1} \exp(-t) dt, \quad \frac{\Gamma(r+1)}{\Gamma(r)} = r, \quad k! = \Gamma(k+1).$$

So, for notational convenience, take $\alpha = a + 1 > 0$ and $\beta = b > 0$.

(Now calculate the normalization factor.)

Example: Conjugate Bayesian inference — the prior

- ▶ General priors: posterior difficult to obtain in closed analytic form.
- ▶ Mathematical *convenience*: use *conjugate* priors.

Derivation of conjugate prior for the Poisson likelihood:

$$p(\lambda | x) \propto L_x(\lambda)p(\lambda) \propto \lambda^{N\bar{x}} \exp(-N\lambda)p(\lambda).$$

So take $p(\lambda) \propto \lambda^a \exp(-b\lambda)$.

Normalization? Gamma integral for $r > 0$ and $k \in \mathbb{N}$: (Write this down!)

$$\Gamma(r) = \int_0^\infty t^{r-1} \exp(-t) dt, \quad \frac{\Gamma(r+1)}{\Gamma(r)} = r, \quad k! = \Gamma(k+1).$$

So, for notational convenience, take $\alpha = a + 1 > 0$ and $\beta = b > 0$.

(Now calculate the normalization factor.)

The *Gamma distribution* is conjugate to the Poisson likelihood; its density:

$$p_{\text{Ga}}(\lambda | \alpha, \beta) := \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} \exp(-\beta\lambda).$$

Example: Conjugate Bayesian inference — the prior

- ▶ General priors: posterior difficult to obtain in closed analytic form.
- ▶ Mathematical *convenience*: use *conjugate* priors.

Derivation of conjugate prior for the Poisson likelihood:

$$p(\lambda | x) \propto L_x(\lambda)p(\lambda) \propto \lambda^{N\bar{x}} \exp(-N\lambda)p(\lambda).$$

So take $p(\lambda) \propto \lambda^a \exp(-b\lambda)$.

Normalization? Gamma integral for $r > 0$ and $k \in \mathbb{N}$: (Write this down!)

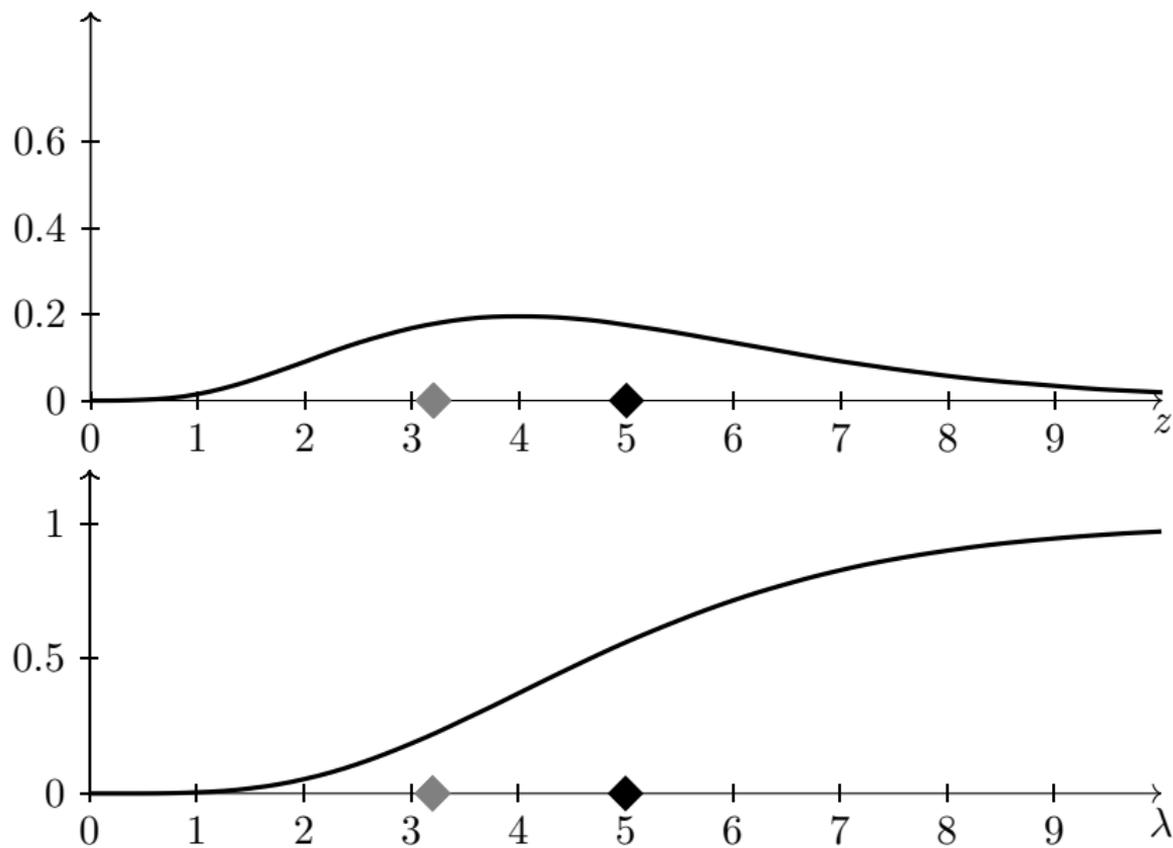
$$\Gamma(r) = \int_0^\infty t^{r-1} \exp(-t) dt, \quad \frac{\Gamma(r+1)}{\Gamma(r)} = r, \quad k! = \Gamma(k+1).$$

Now, for notational convenience, take $s = \beta > 0$ and *mean* $t = \frac{\alpha}{\beta} > 0$.

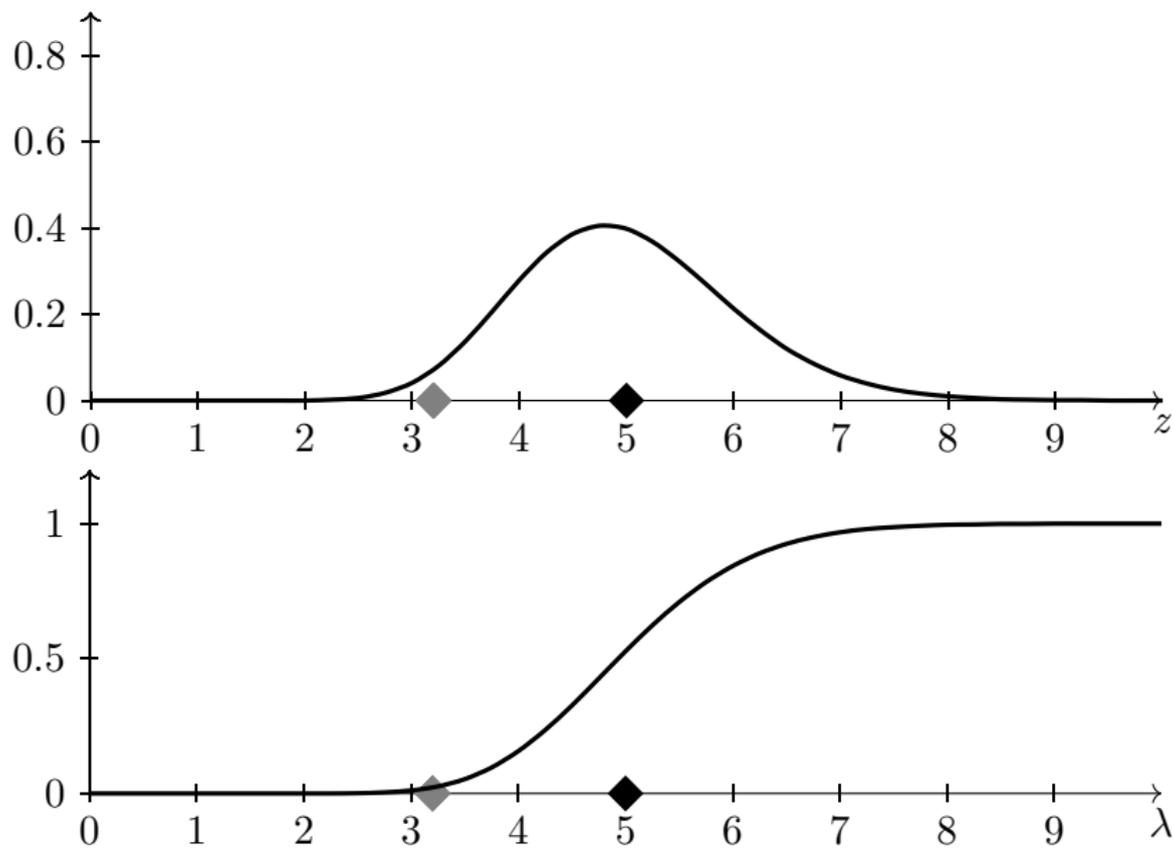
The *Gamma distribution* is conjugate to the Poisson likelihood; its density:

$$p_{\text{Ga}}(\lambda | s, t) := \frac{s^{st}}{\Gamma(st)} \lambda^{st-1} \exp(-s\lambda).$$

Example: visualization of Gamma prior ($s = 1, t = 5$)



Example: visualization of Gamma prior ($s = 5, t = 5$)



Example: Conjugate Bayesian inference — the posterior

Derivation of the posterior probability density function's expression:

$$p(\lambda | x) \propto L_x(\lambda)p_{\text{Ga}}(\lambda | s, t)$$

Example: Conjugate Bayesian inference — the posterior

Derivation of the posterior probability density function's expression:

$$\begin{aligned} p(\lambda | x) &\propto L_x(\lambda) p_{\text{Ga}}(\lambda | s, t) \\ &\propto \lambda^{N\bar{x}} \exp(-N\lambda) \lambda^{st-1} \exp(-s\lambda) \end{aligned}$$

Example: Conjugate Bayesian inference — the posterior

Derivation of the posterior probability density function's expression:

$$\begin{aligned} p(\lambda | x) &\propto L_x(\lambda) p_{\text{Ga}}(\lambda | s, t) \\ &\propto \lambda^{N\bar{x}} \exp(-N\lambda) \lambda^{st-1} \exp(-s\lambda) \\ &= \lambda^{N\bar{x}+st-1} \exp(-(N+s)\lambda) \end{aligned}$$

Example: Conjugate Bayesian inference — the posterior

Derivation of the posterior probability density function's expression:

$$\begin{aligned} p(\lambda | x) &\propto L_x(\lambda) p_{\text{Ga}}(\lambda | s, t) \\ &\propto \lambda^{N\bar{x}} \exp(-N\lambda) \lambda^{st-1} \exp(-s\lambda) \\ &= \lambda^{N\bar{x}+st-1} \exp(-(N+s)\lambda) \propto p_{\text{Ga}}\left(\lambda \mid N+s, \frac{N\bar{x}+st}{N+s}\right). \end{aligned}$$

Example: Conjugate Bayesian inference — the posterior

Derivation of the posterior probability density function's expression:

$$\begin{aligned} p(\lambda | x) &\propto L_x(\lambda) p_{\text{Ga}}(\lambda | s, t) \\ &\propto \lambda^{N\bar{x}} \exp(-N\lambda) \lambda^{st-1} \exp(-s\lambda) \\ &= \lambda^{N\bar{x}+st-1} \exp(-(N+s)\lambda) \propto p_{\text{Ga}}\left(\lambda \mid N+s, \frac{N\bar{x}+st}{N+s}\right). \end{aligned}$$

Conjugate nature:

- ▶ Prior and posterior distribution belong to the same family.
- ▶ Parameters are easy to update: sum of pseudocounts and sample counts, mixture of prior mean and sufficient statistic.
- ▶ Posterior mean $\frac{N\bar{x}+st}{N+s}$.

Example: Conjugate Bayesian inference — the posterior

Derivation of the posterior probability density function's expression:

$$\begin{aligned} p(\lambda | x) &\propto L_x(\lambda) p_{\text{Ga}}(\lambda | s, t) \\ &\propto \lambda^{N\bar{x}} \exp(-N\lambda) \lambda^{st-1} \exp(-s\lambda) \\ &= \lambda^{N\bar{x}+st-1} \exp(-(N+s)\lambda) \propto p_{\text{Ga}}\left(\lambda \mid N+s, \frac{N\bar{x}+st}{N+s}\right). \end{aligned}$$

Conjugate nature:

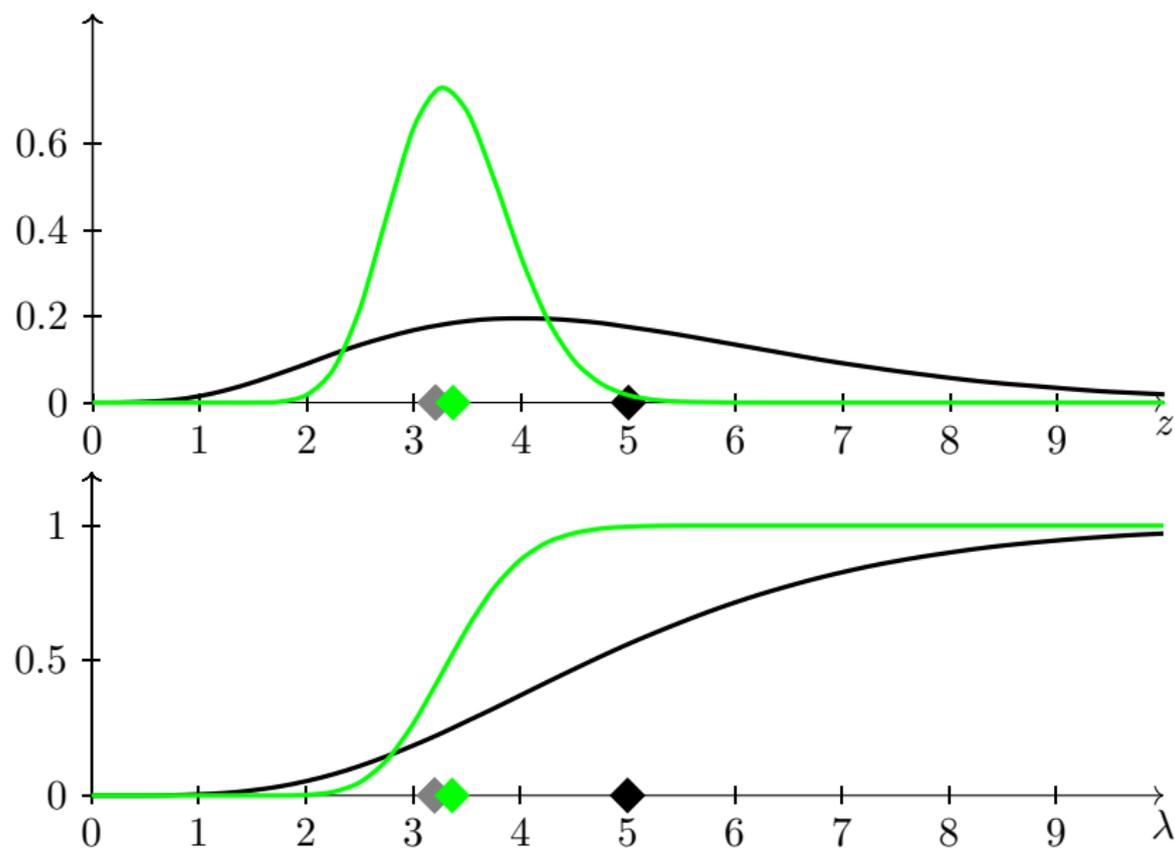
- ▶ Prior and posterior distribution belong to the same family.
- ▶ Parameters are easy to update: sum of pseudocounts and sample counts, mixture of prior mean and sufficient statistic.
- ▶ Posterior mean $\frac{N\bar{x}+st}{N+s}$.

Expectation of a function f of the parameter λ ?

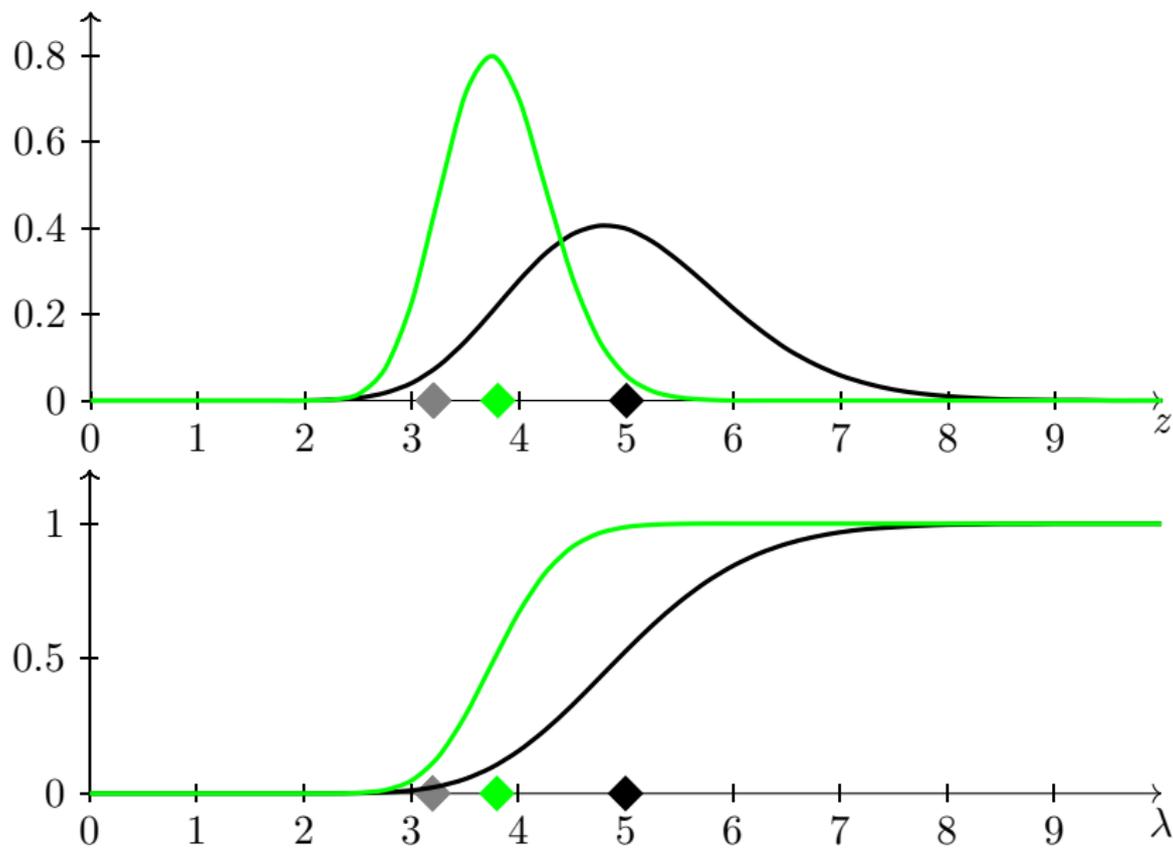
$$\int_0^\infty f(\lambda) p_{\text{Ga}}(\lambda | \tilde{s}, \tilde{t}) d\lambda = \frac{\tilde{s}^{\tilde{t}}}{\Gamma(\tilde{s}\tilde{t})} \int_0^\infty f(\lambda) \lambda^{N\bar{x}+\tilde{s}\tilde{t}-1} \exp(-(N+\tilde{s})\lambda) d\lambda$$

(For which f is this 'straightforward'?)

Example: visualization of Gamma posterior ($s = 1, t = 5$)



Example: visualization of Gamma posterior ($s = 5, t = 5$)



Example: Conjugate imprecise probabilistic inference

- ▶ Hard to justify a single prior, so use a *set of priors*.
- ▶ Mathematical *convenience*: use conjugate priors.

Example: Conjugate imprecise probabilistic inference

- ▶ Hard to justify a single prior, so use a *set of priors*.
- ▶ Mathematical *convenience*: use conjugate priors.
- ▶ How do we generate such a set: varying parameters s and t in some subset of the possible values.
- ▶ Here:
 - ▶ keep s fixed — interpret as a *learning rate*,
 - ▶ vary t — so uncertainty about the prior *mean*.

Example: Conjugate imprecise probabilistic inference

- ▶ Hard to justify a single prior, so use a *set of priors*.
- ▶ Mathematical *convenience*: use conjugate priors.
- ▶ How do we generate such a set: varying parameters s and t in some subset of the possible values.
- ▶ Here:
 - ▶ keep s fixed — interpret as a *learning rate*,
 - ▶ vary t — so uncertainty about the prior *mean*.

Set of probability density functions:

$$\{p_{\text{Ga}}(\lambda \mid \tilde{s}, \tilde{t}) : t \in [\underline{t}, \bar{t}]\}$$

(Can you easily calculate the lower and upper probability density functions?)

(Which *deductive* inferences are 'straightforward' to calculate?)

Example: Conjugate imprecise probabilistic inference

- ▶ Hard to justify a single prior, so use a *set of priors*.
- ▶ Mathematical *convenience*: use conjugate priors.
- ▶ How do we generate such a set: varying parameters s and t in some subset of the possible values.
- ▶ Here:
 - ▶ keep s fixed — interpret as a *learning rate*,
 - ▶ vary t — so uncertainty about the prior *mean*.

Set of probability density functions:

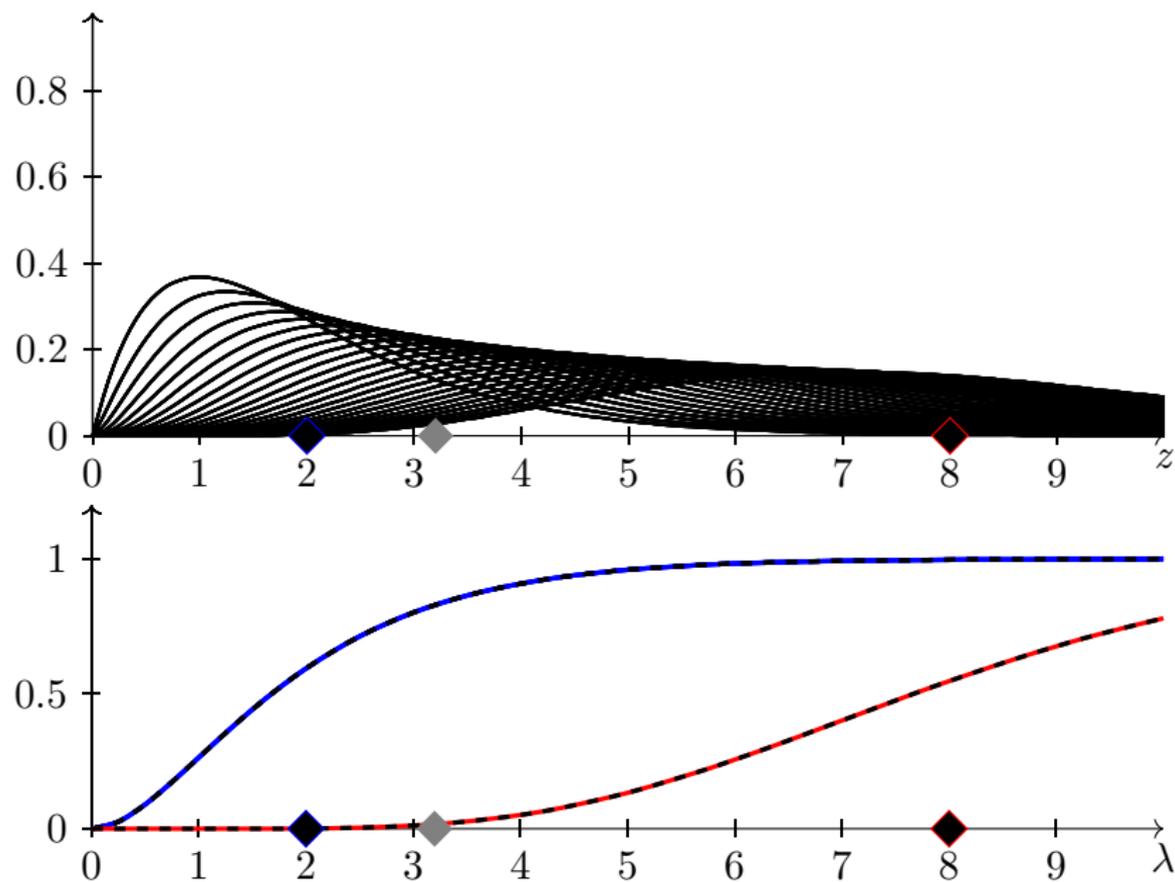
$$\{p_{\text{Ga}}(\lambda \mid \tilde{s}, \tilde{t}) : t \in [\underline{t}, \bar{t}]\}$$

(Can you easily calculate the lower and upper probability density functions?)

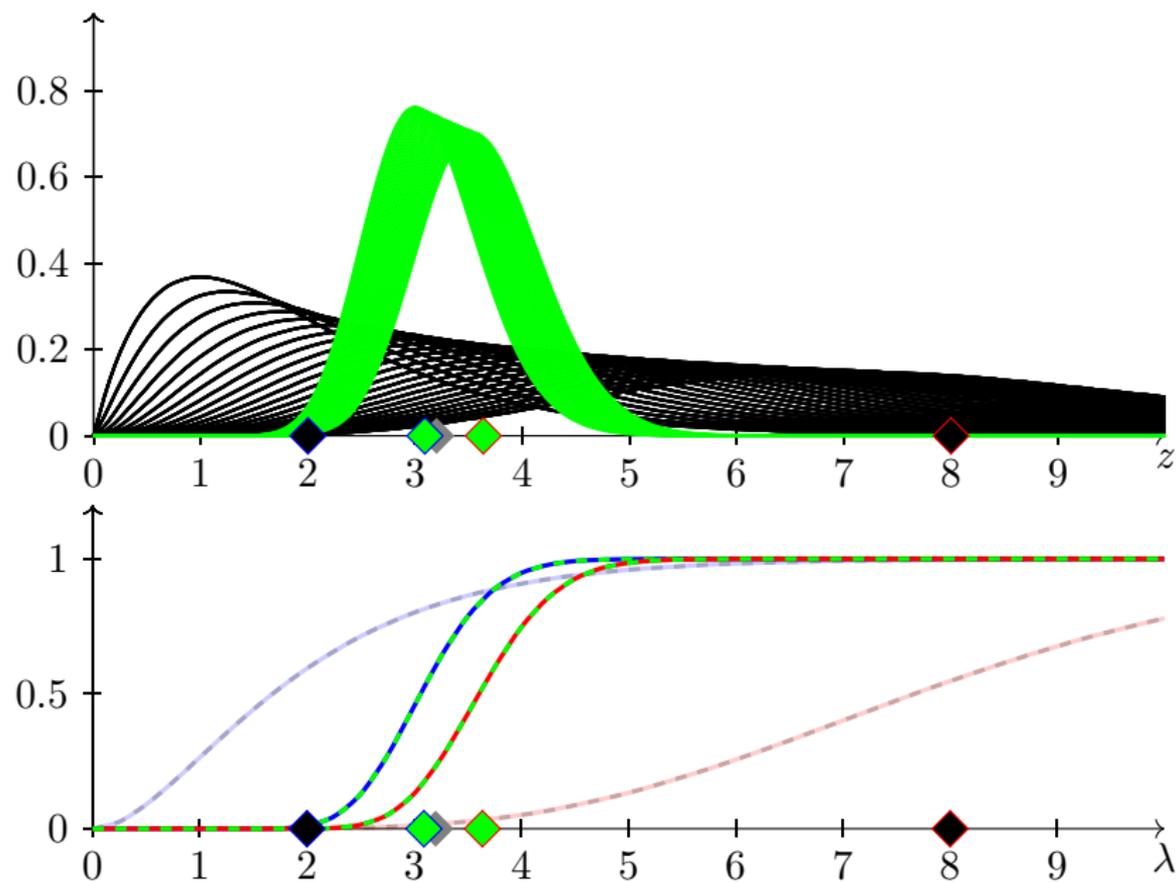
(Which *deductive* inferences are 'straightforward' to calculate?)

- ▶ Set of values for t must be bounded in the direction to learn in.

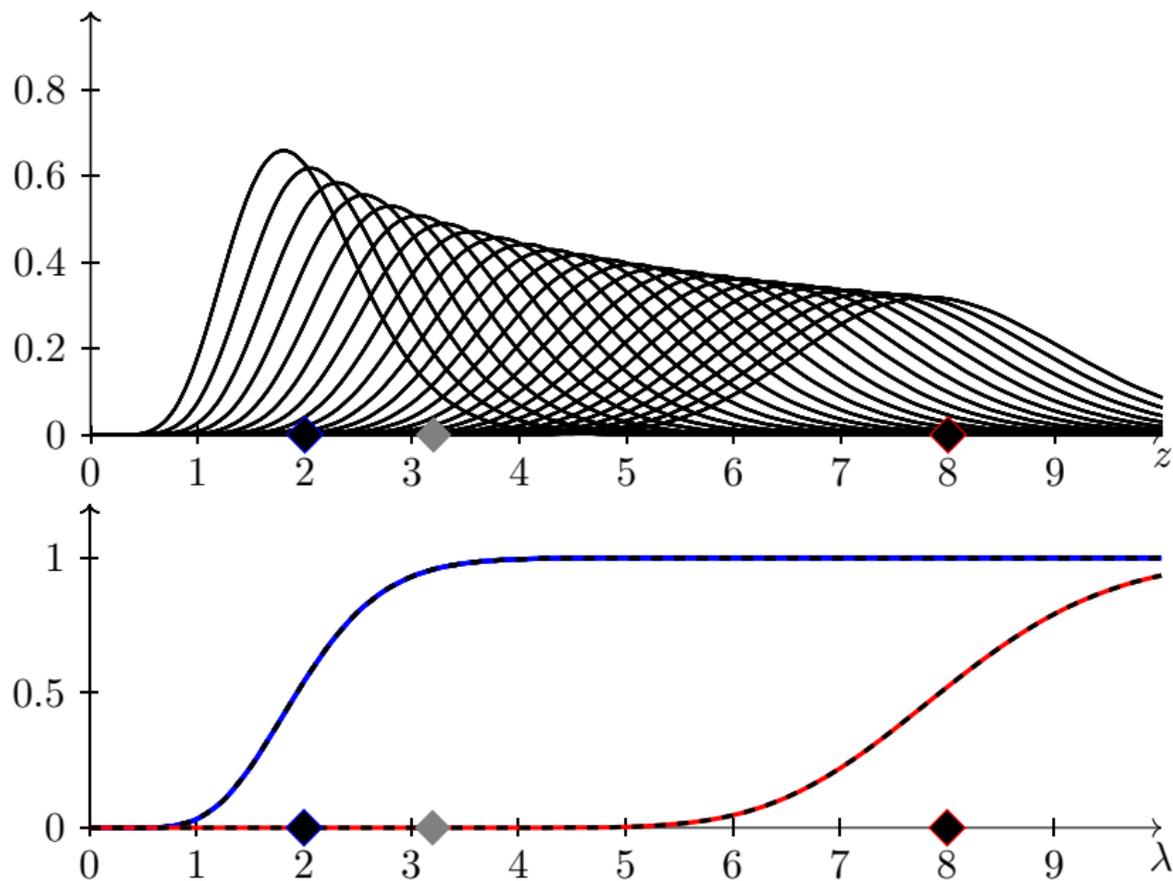
Example: Visualiz. of Gamma priors ($s = 1, t \in [2, 8]$)



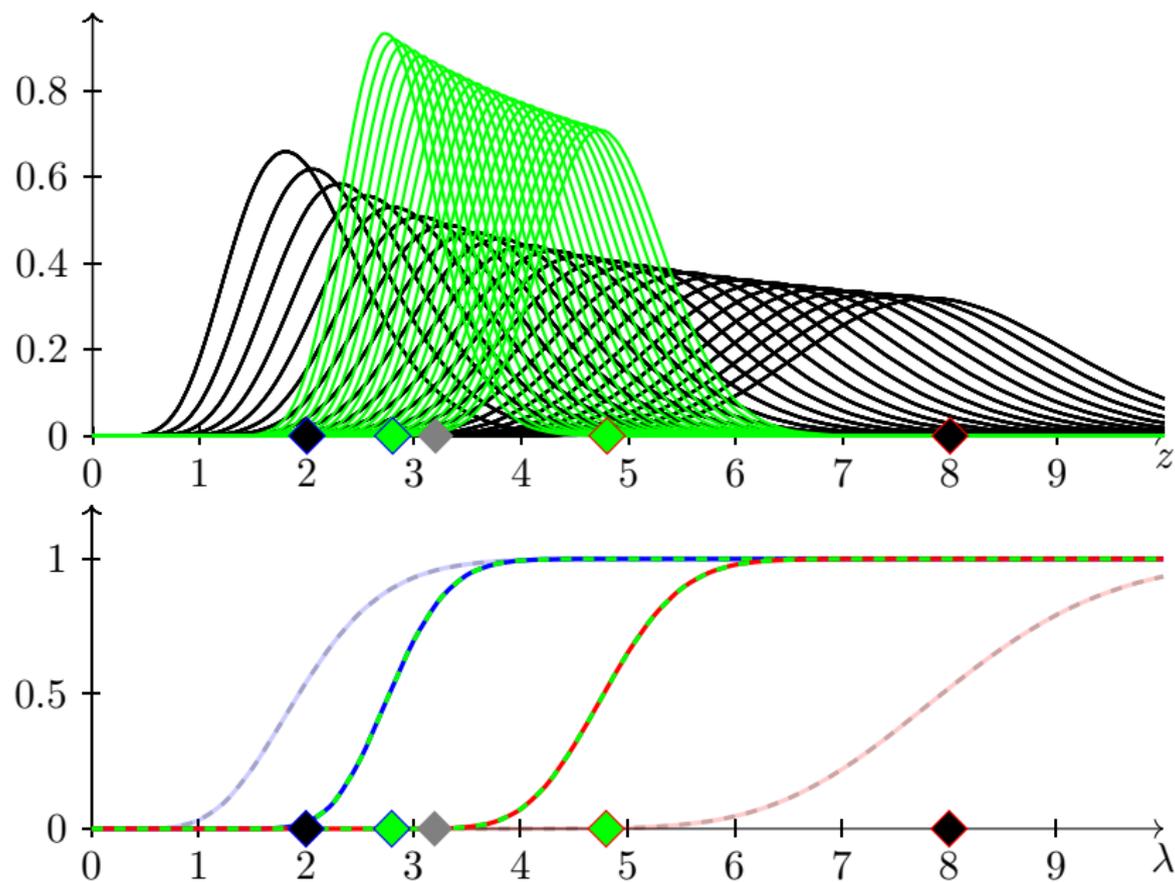
Example: Visualiz. of Gamma posteriors ($s = 1, t \in [2, 8]$)



Example: Visualiz. of Gamma priors ($s = 5$, $t \in [2, 8]$)



Example: Visualiz. of Gamma posteriors ($s = 5, t \in [2, 8]$)



Example: Parametric to predictive inference — joint

- ▶ Given is a conjugate parametric model.
- ▶ We wish to do *predictive inference*:
inference about a sequence y of M unseen observations.

Example: Parametric to predictive inference — joint

- ▶ Given is a conjugate parametric model.
- ▶ We wish to do *predictive inference*: inference about a sequence y of M unseen observations.
- ▶ In case the parameter λ is known, then by the iid assumption we know the joint probability mass function

$$p(y | \lambda) := \prod_{i=1}^M p_{\mathcal{P}_s}(y_i | \lambda).$$

Example: Parametric to predictive inference — joint

- ▶ Given is a conjugate parametric model.
- ▶ We wish to do *predictive inference*: inference about a sequence y of M unseen observations.
- ▶ In case the parameter λ is known, then by the iid assumption we know the joint probability mass function

$$p(y | \lambda) := \prod_{i=1}^M p_{\mathcal{P}_s}(y_i | \lambda).$$

- ▶ But we are uncertain about λ .
- ▶ Assume a precise (prior or posterior) conjugate model, i.e., a density $p_{\text{Ga}}(\lambda | \tilde{s}, \tilde{t})$.
- ▶ The joint density's expression is

$$p(y, \lambda | \tilde{s}, \tilde{t}) = p(y | \lambda) p_{\text{Ga}}(\lambda | \tilde{s}, \tilde{t})$$

Example: Parametric to predictive inference — marginal

Deriving the predictive mass function: marginalize to y , so integrate out λ :

$$p(y | \tilde{s}, \tilde{t}) = \int_0^{\infty} p(y | \lambda) p(\lambda | \tilde{s}, \tilde{t}) d\lambda$$

Example: Parametric to predictive inference — marginal

Deriving the predictive mass function: marginalize to y , so integrate out λ :

$$\begin{aligned} p(y | \tilde{s}, \tilde{t}) &= \int_0^{\infty} p(y | \lambda) p(\lambda | \tilde{s}, \tilde{t}) d\lambda \\ &= \int_0^{\infty} L_y(\lambda) p(\lambda | \tilde{s}, \tilde{t}) d\lambda \end{aligned}$$

Example: Parametric to predictive inference — marginal

Deriving the predictive mass function: marginalize to y , so integrate out λ :

$$\begin{aligned} p(y | \tilde{s}, \tilde{t}) &= \int_0^\infty p(y | \lambda) p(\lambda | \tilde{s}, \tilde{t}) d\lambda \\ &= \int_0^\infty L_y(\lambda) p(\lambda | \tilde{s}, \tilde{t}) d\lambda \\ &= \frac{1}{\prod_{i=1}^M y_i!} \frac{\tilde{s}^{\tilde{s}\tilde{t}}}{\Gamma(\tilde{s}\tilde{t})} \frac{\Gamma(M\bar{y} + \tilde{s}\tilde{t})}{(M + \tilde{s})^{M\bar{y} + \tilde{s}\tilde{t}}} \int_0^\infty p\left(\lambda \mid M + \tilde{s}, \frac{M\bar{y} + \tilde{s}\tilde{t}}{M + \tilde{s}}\right) d\lambda \end{aligned}$$

Example: Parametric to predictive inference — marginal

Deriving the predictive mass function: marginalize to y , so integrate out λ :

$$\begin{aligned} p(y | \tilde{s}, \tilde{t}) &= \int_0^{\infty} p(y | \lambda) p(\lambda | \tilde{s}, \tilde{t}) d\lambda \\ &= \int_0^{\infty} L_y(\lambda) p(\lambda | \tilde{s}, \tilde{t}) d\lambda \\ &= \frac{1}{\prod_{i=1}^M y_i!} \frac{\tilde{s}^{\tilde{s}\tilde{t}}}{\Gamma(\tilde{s}\tilde{t})} \frac{\Gamma(M\bar{y} + \tilde{s}\tilde{t})}{(M + \tilde{s})^{M\bar{y} + \tilde{s}\tilde{t}}} \int_0^{\infty} p\left(\lambda \mid M + \tilde{s}, \frac{M\bar{y} + \tilde{s}\tilde{t}}{M + \tilde{s}}\right) d\lambda \\ &= \frac{\Gamma(M\bar{y} + \tilde{s}\tilde{t})}{\Gamma(\tilde{s}\tilde{t}) \prod_{i=1}^M y_i!} \left(\frac{\tilde{s}}{M + \tilde{s}}\right)^{\tilde{s}\tilde{t}} \frac{1}{(M + \tilde{s})^{M\bar{y}}}. \end{aligned}$$

Example: Parametric to predictive inference — marginal

Deriving the predictive mass function: marginalize to y , so integrate out λ :

$$\begin{aligned} p(y | \tilde{s}, \tilde{t}) &= \int_0^\infty p(y | \lambda) p(\lambda | \tilde{s}, \tilde{t}) d\lambda \\ &= \int_0^\infty L_y(\lambda) p(\lambda | \tilde{s}, \tilde{t}) d\lambda \\ &= \frac{1}{\prod_{i=1}^M y_i!} \frac{\tilde{s}^{\tilde{s}\tilde{t}}}{\Gamma(\tilde{s}\tilde{t})} \frac{\Gamma(M\bar{y} + \tilde{s}\tilde{t})}{(M + \tilde{s})^{M\bar{y} + \tilde{s}\tilde{t}}} \int_0^\infty p\left(\lambda \mid M + \tilde{s}, \frac{M\bar{y} + \tilde{s}\tilde{t}}{M + \tilde{s}}\right) d\lambda \\ &= \frac{\Gamma(M\bar{y} + \tilde{s}\tilde{t})}{\Gamma(\tilde{s}\tilde{t}) \prod_{i=1}^M y_i!} \left(\frac{\tilde{s}}{M + \tilde{s}}\right)^{\tilde{s}\tilde{t}} \frac{1}{(M + \tilde{s})^{M\bar{y}}}. \end{aligned}$$

(What is the explicit expression for the posterior?)

Example: Parametric to predictive inference — marginal

Deriving the predictive mass function: marginalize to y , so integrate out λ :

$$\begin{aligned} p(y | \tilde{s}, \tilde{t}) &= \int_0^\infty p(y | \lambda) p(\lambda | \tilde{s}, \tilde{t}) d\lambda \\ &= \int_0^\infty L_y(\lambda) p(\lambda | \tilde{s}, \tilde{t}) d\lambda \\ &= \frac{1}{\prod_{i=1}^M y_i!} \frac{\tilde{s}^{\tilde{s}\tilde{t}}}{\Gamma(\tilde{s}\tilde{t})} \frac{\Gamma(M\bar{y} + \tilde{s}\tilde{t})}{(M + \tilde{s})^{M\bar{y} + \tilde{s}\tilde{t}}} \int_0^\infty p\left(\lambda \mid M + \tilde{s}, \frac{M\bar{y} + \tilde{s}\tilde{t}}{M + \tilde{s}}\right) d\lambda \\ &= \frac{\Gamma(M\bar{y} + \tilde{s}\tilde{t})}{\Gamma(\tilde{s}\tilde{t}) \prod_{i=1}^M y_i!} \left(\frac{\tilde{s}}{M + \tilde{s}}\right)^{\tilde{s}\tilde{t}} \frac{1}{(M + \tilde{s})^{M\bar{y}}}. \end{aligned}$$

(What is the explicit expression for the posterior?)

$$p\left(y \mid N + s, \frac{N\bar{x} + st}{N + s}\right) = \frac{\Gamma(M\bar{y} + N\bar{x} + st)}{\Gamma(N\bar{x} + st) \prod_{i=1}^M y_i!} \left(\frac{N + s}{M + N + s}\right)^{N\bar{x} + st} \frac{1}{(M + N + s)^{M\bar{y}}}$$

Example: Parametric to predictive inference — marginal

Deriving the predictive mass function: marginalize to y , so integrate out λ :

$$\begin{aligned} p(y | \tilde{s}, \tilde{t}) &= \int_0^\infty p(y | \lambda) p(\lambda | \tilde{s}, \tilde{t}) d\lambda \\ &= \int_0^\infty L_y(\lambda) p(\lambda | \tilde{s}, \tilde{t}) d\lambda \\ &= \frac{1}{\prod_{i=1}^M y_i!} \frac{\tilde{s}^{\tilde{s}\tilde{t}}}{\Gamma(\tilde{s}\tilde{t})} \frac{\Gamma(M\bar{y} + \tilde{s}\tilde{t})}{(M + \tilde{s})^{M\bar{y} + \tilde{s}\tilde{t}}} \int_0^\infty p\left(\lambda \mid M + \tilde{s}, \frac{M\bar{y} + \tilde{s}\tilde{t}}{M + \tilde{s}}\right) d\lambda \\ &= \frac{\Gamma(M\bar{y} + \tilde{s}\tilde{t})}{\Gamma(\tilde{s}\tilde{t}) \prod_{i=1}^M y_i!} \left(\frac{\tilde{s}}{M + \tilde{s}}\right)^{\tilde{s}\tilde{t}} \frac{1}{(M + \tilde{s})^{M\bar{y}}}. \end{aligned}$$

(What is the explicit expression for the posterior?)

$$p\left(y \mid N + s, \frac{N\bar{x} + st}{N + s}\right) = \frac{\Gamma(M\bar{y} + N\bar{x} + st)}{\Gamma(N\bar{x} + st) \prod_{i=1}^M y_i!} \left(\frac{N + s}{M + N + s}\right)^{N\bar{x} + st} \frac{1}{(M + N + s)^{M\bar{y}}}$$

Using this mass function requires numerical computations in general.

Example: Parametric to predictive inference — immediate

Immediate prediction is inference about one unknown observation (e.g., the next one), so $M := 1$:

$$p(z | \tilde{s}, \tilde{t}) = \frac{\Gamma(z + \tilde{s}\tilde{t})}{\Gamma(\tilde{s}\tilde{t})z!} \left(\frac{\tilde{s}}{1 + \tilde{s}} \right)^{\tilde{s}\tilde{t}} \frac{1}{(1 + \tilde{s})^z}$$

Example: Parametric to predictive inference — immediate

Immediate prediction is inference about one unknown observation (e.g., the next one), so $M := 1$:

$$\begin{aligned} p(z | \tilde{s}, \tilde{t}) &= \frac{\Gamma(z + \tilde{s}\tilde{t})}{\Gamma(\tilde{s}\tilde{t})z!} \left(\frac{\tilde{s}}{1 + \tilde{s}}\right)^{\tilde{s}\tilde{t}} \frac{1}{(1 + \tilde{s})^z} \\ &= \binom{z + \tilde{s}\tilde{t} - 1}{z} \left(1 - \frac{1}{1 + \tilde{s}}\right)^{\tilde{s}\tilde{t}} \left(\frac{1}{1 + \tilde{s}}\right)^z. \end{aligned}$$

Example: Parametric to predictive inference — immediate

Immediate prediction is inference about one unknown observation (e.g., the next one), so $M := 1$:

$$\begin{aligned} p(z | \tilde{s}, \tilde{t}) &= \frac{\Gamma(z + \tilde{s}\tilde{t})}{\Gamma(\tilde{s}\tilde{t})z!} \left(\frac{\tilde{s}}{1 + \tilde{s}}\right)^{\tilde{s}\tilde{t}} \frac{1}{(1 + \tilde{s})^z} \\ &= \binom{z + \tilde{s}\tilde{t} - 1}{z} \left(1 - \frac{1}{1 + \tilde{s}}\right)^{\tilde{s}\tilde{t}} \left(\frac{1}{1 + \tilde{s}}\right)^z. \end{aligned}$$

This is the negative binomial distribution:

- ▶ the probability of 'success': $\frac{1}{1 + \tilde{s}}$
- ▶ the possibly non-integer number of 'failures' $\tilde{s}\tilde{t}$ that determine when sampling is stopped

Example: Parametric to predictive inference — immediate

Immediate prediction is inference about one unknown observation (e.g., the next one), so $M := 1$:

$$\begin{aligned} p(z | \tilde{s}, \tilde{t}) &= \frac{\Gamma(z + \tilde{s}\tilde{t})}{\Gamma(\tilde{s}\tilde{t})z!} \left(\frac{\tilde{s}}{1 + \tilde{s}}\right)^{\tilde{s}\tilde{t}} \frac{1}{(1 + \tilde{s})^z} \\ &= \binom{z + \tilde{s}\tilde{t} - 1}{z} \left(1 - \frac{1}{1 + \tilde{s}}\right)^{\tilde{s}\tilde{t}} \left(\frac{1}{1 + \tilde{s}}\right)^z. \end{aligned}$$

This is the negative binomial distribution:

- ▶ the probability of 'success': $\frac{1}{1 + \tilde{s}}$
- ▶ the possibly non-integer number of 'failures' $\tilde{s}\tilde{t}$ that determine when sampling is stopped

(Interpretation in terms of number of arriving mails?)

Example: Parametric to predictive inference — immediate

Immediate prediction is inference about one unknown observation (e.g., the next one), so $M := 1$:

$$\begin{aligned} p(z | \tilde{s}, \tilde{t}) &= \frac{\Gamma(z + \tilde{s}\tilde{t})}{\Gamma(\tilde{s}\tilde{t})z!} \left(\frac{\tilde{s}}{1 + \tilde{s}}\right)^{\tilde{s}\tilde{t}} \frac{1}{(1 + \tilde{s})^z} \\ &= \binom{z + \tilde{s}\tilde{t} - 1}{z} \left(1 - \frac{1}{1 + \tilde{s}}\right)^{\tilde{s}\tilde{t}} \left(\frac{1}{1 + \tilde{s}}\right)^z. \end{aligned}$$

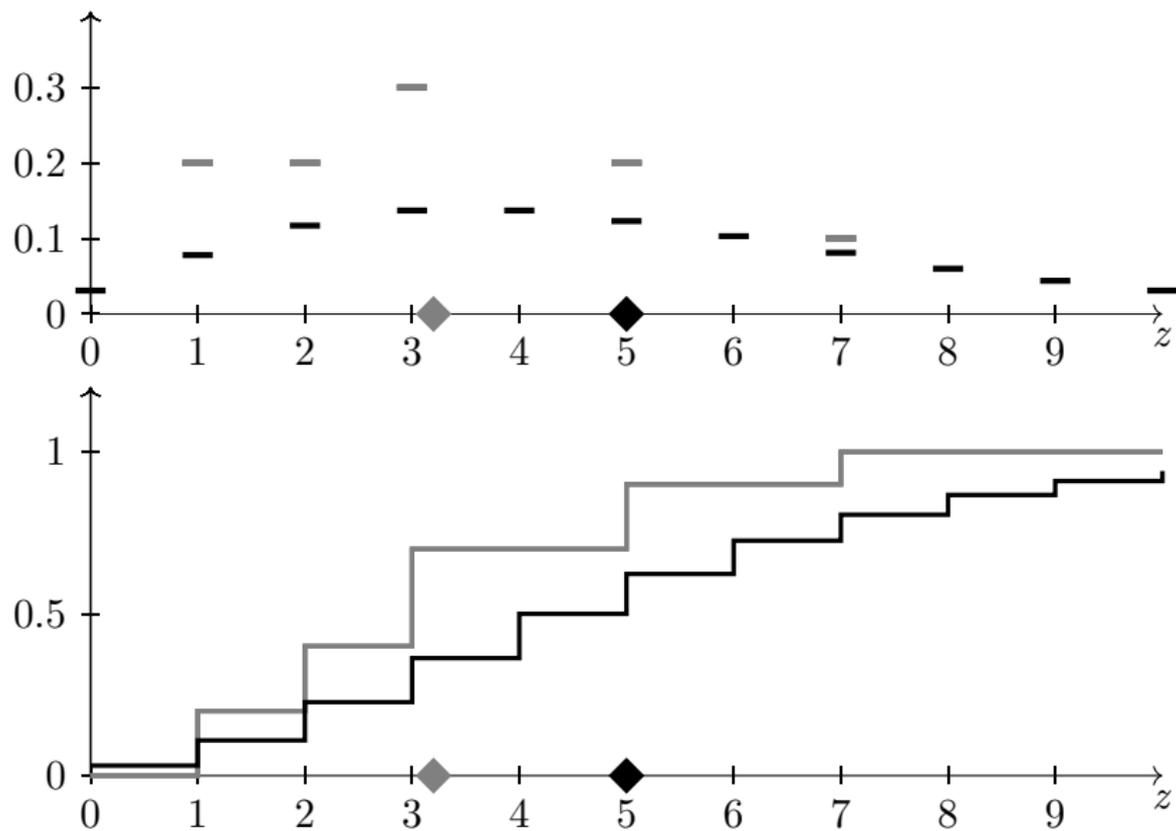
This is the negative binomial distribution:

- ▶ the probability of 'success': $\frac{1}{1 + \tilde{s}}$
- ▶ the possibly non-integer number of 'failures' $\tilde{s}\tilde{t}$ that determine when sampling is stopped

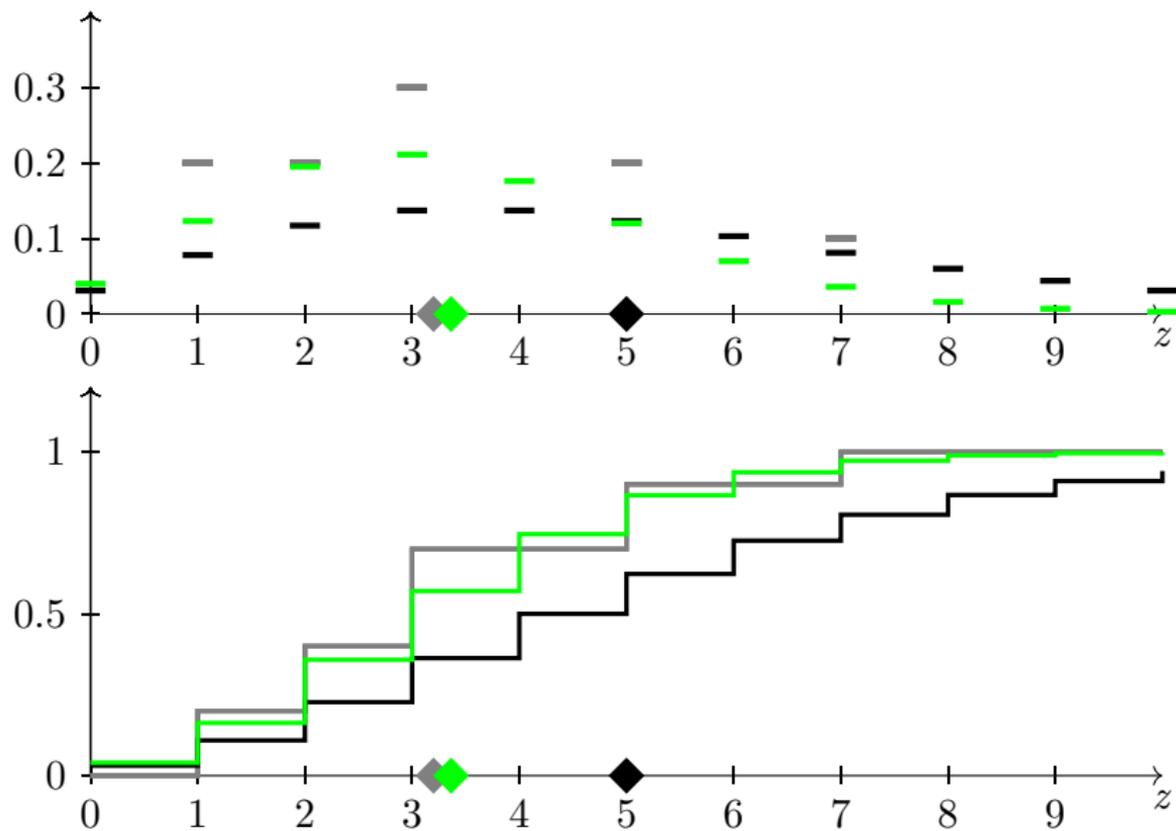
(Interpretation in terms of number of arriving mails?)

(What is the mean?)

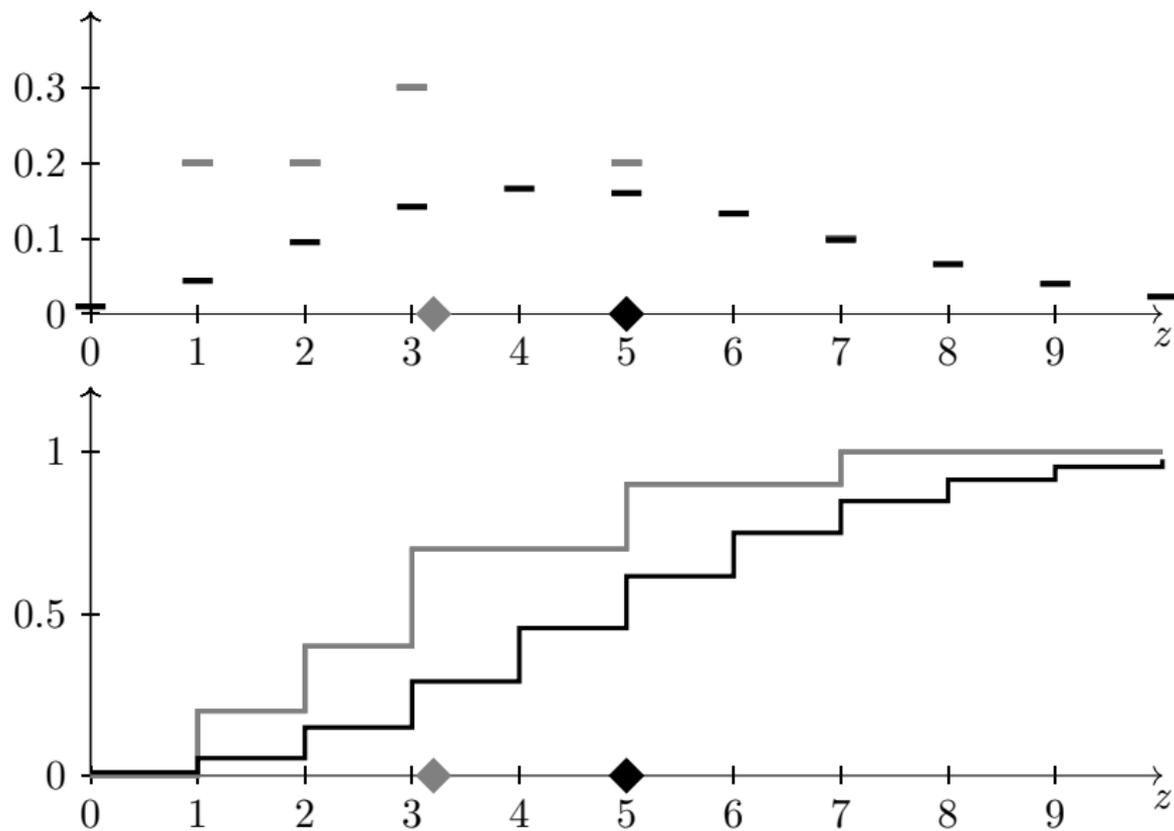
Example: Visualiz. of negative binomial prior ($s = 1, t = 5$)



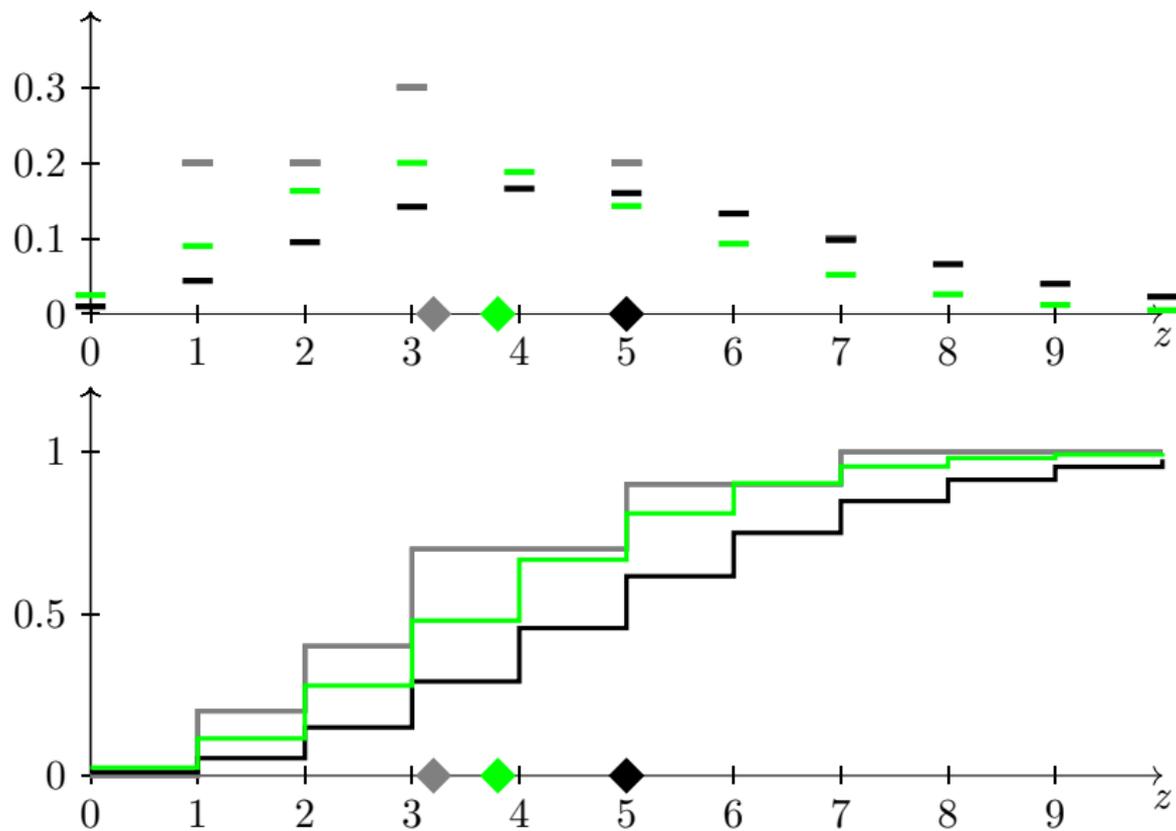
Example: Visualiz. of neg. binomial posterior ($s = 1, t = 5$)



Example: Visualiz. of negative binomial prior ($s = 5, t = 5$)



Example: Visualiz. of neg. binomial posterior ($s = 5, t = 5$)



Example: imprecise-probabilistic predictive inference

Set of probability mass functions:

$$\{p(y | \tilde{s}, \tilde{t}) : t \in [\underline{t}, \bar{t}]\}$$

Example: imprecise-probabilistic predictive inference

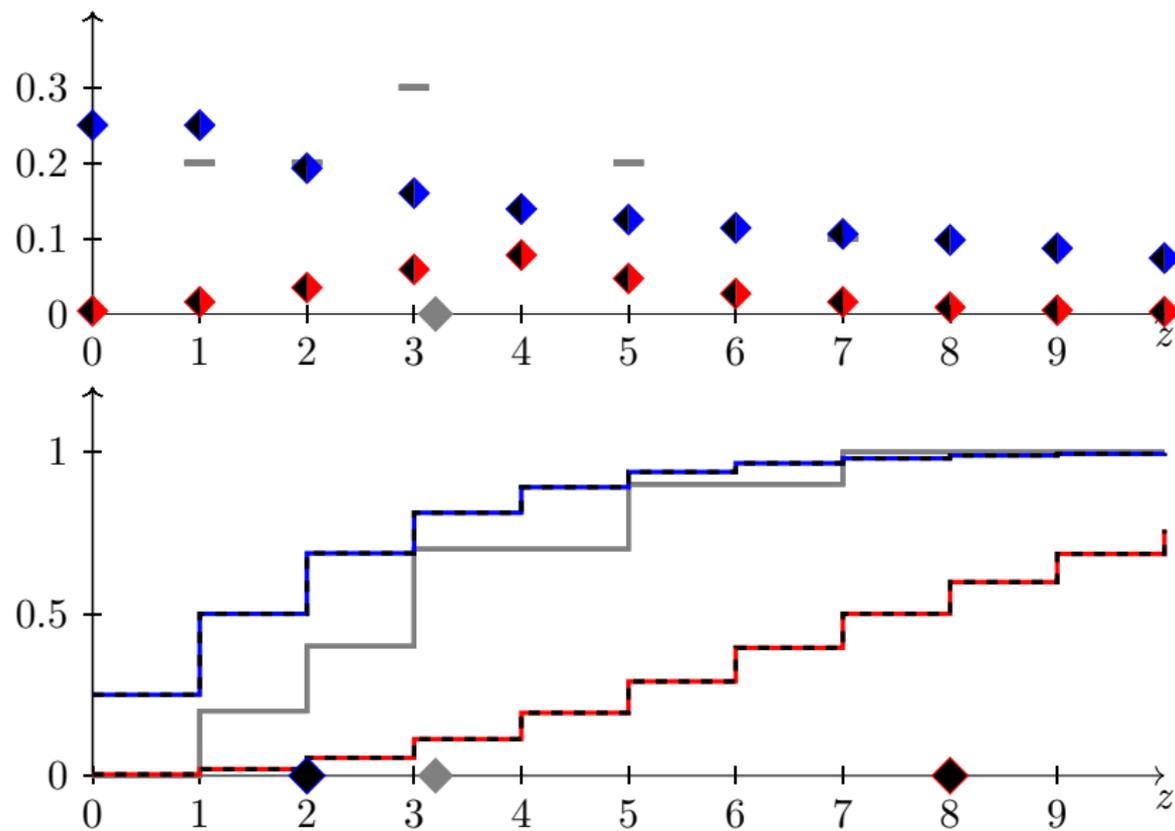
Set of probability mass functions:

$$\{p(y | \tilde{s}, \tilde{t}) : t \in [\underline{t}, \bar{t}]\}$$

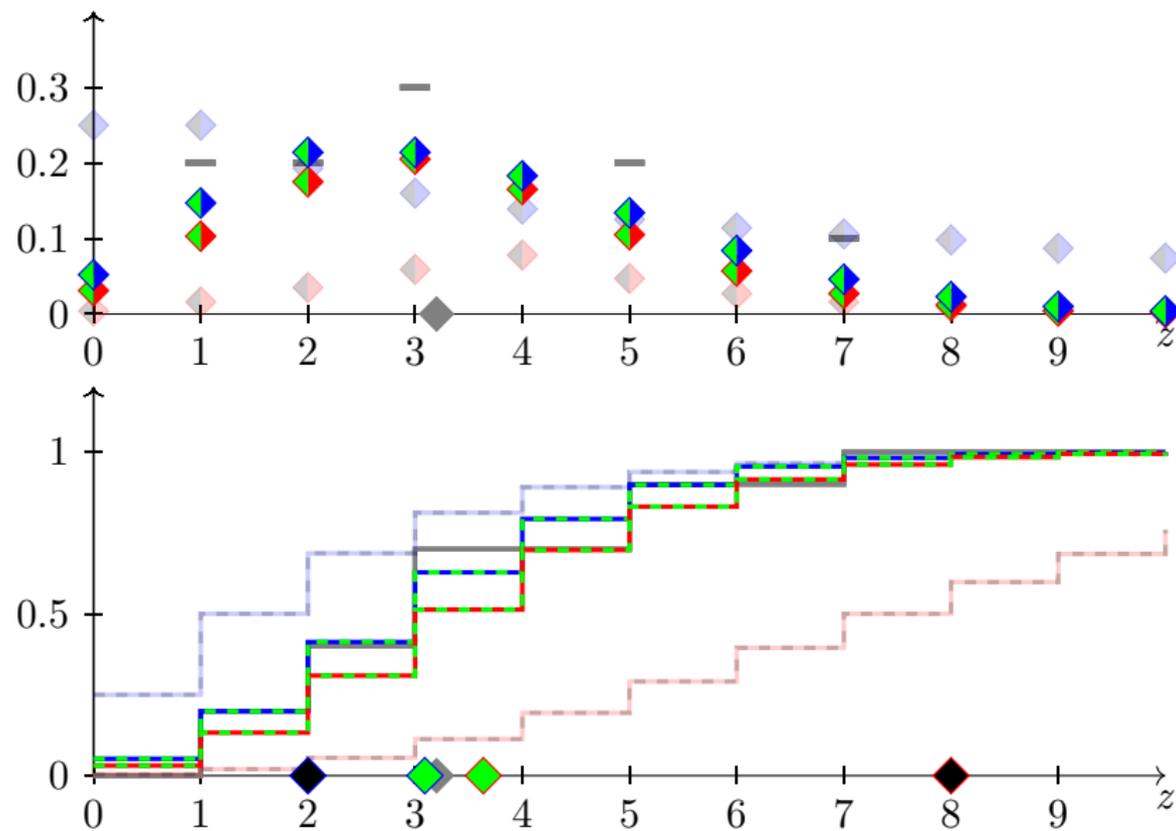
(Can you easily calculate the lower and upper probability mass functions?)

(Which *deductive* inferences are 'straightforward' to calculate?)

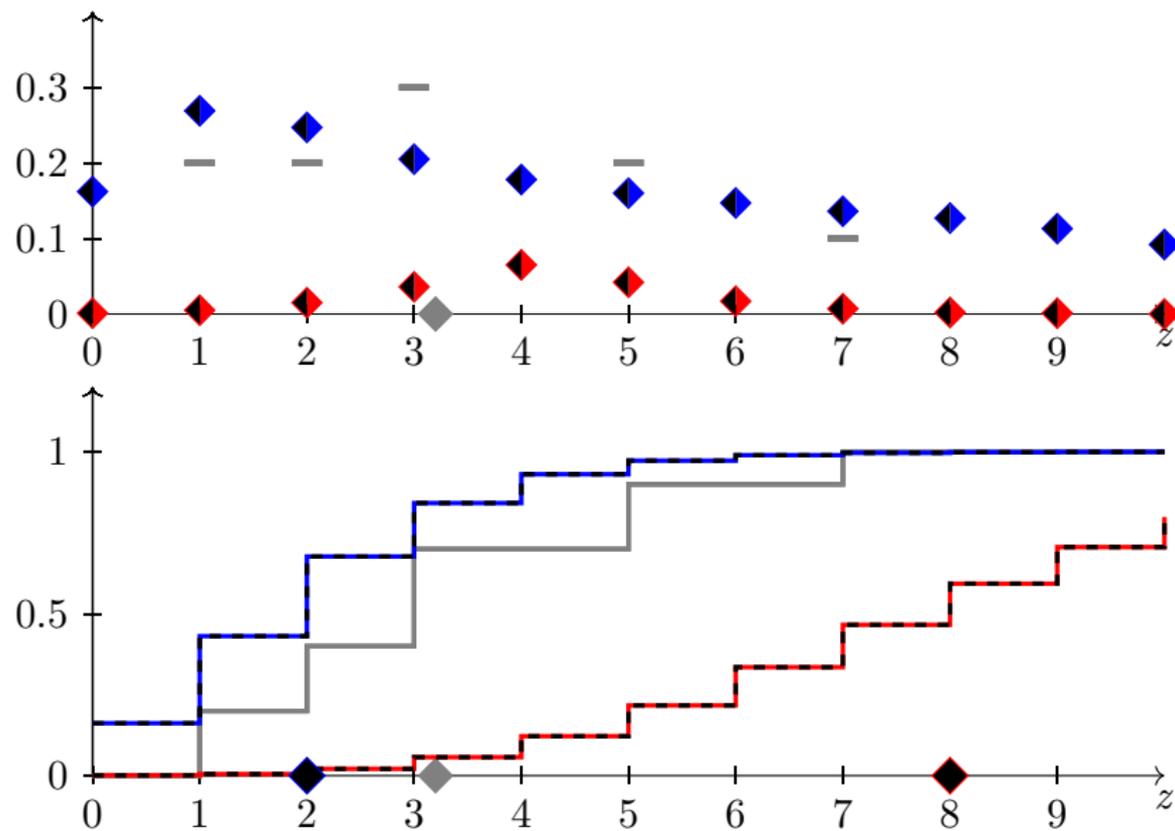
Example: Visualiz. of neg. binomial priors ($s = 1$, $t \in [2, 8]$)



Example: Visualiz. of neg. binom. post's ($s = 1$, $t \in [2, 8]$)



Example: Visualiz. of neg. binomial priors ($s = 5$, $t \in [2, 8]$)



Exercise

Exercise: Data

Gert labels some emails he receives as interesting (I). We label the others with a U .

Within a busy hour, he gets

I U U U I I I U U I

Exercise: Data

Gert labels some emails he receives as interesting (I). We label the others with a U .

Within a busy hour, he gets

$I \quad U \quad U \quad U \quad I \quad I \quad I \quad U \quad U \quad I$

1. Number of samples N ? Order statistics? Count and frequency vectors $n = (n_I, n_U)$ and $f = (f_I, f_U)$?
2. Write down the lower and upper probability values for I (and U) resulting from the NPI reasoning.
3. Write down the expressions for the lower and upper probability values for I (and U) produced by ε -contaminating the frequency vector. What are the values obtained by choosing ε values that correspond to 1 and 5 pseudocounts, respectively. Compare with the NPI model.

Exercise: Sampling model

Bernoulli process, iid 'flips of a coin' with chance ϑ of receiving an interesting mail:

$$p_{\text{Br}}(z | \vartheta) = \begin{cases} \vartheta, & z = I, \\ 1 - \vartheta, & z = U. \end{cases}$$

(To symmetrize, use $(\theta_I, \theta_U) = (\vartheta, 1 - \vartheta)$.)

Exercise: Sampling model

Bernoulli process, iid 'flips of a coin' with chance ϑ of receiving an interesting mail:

$$p_{\text{Br}}(z | \vartheta) = \begin{cases} \vartheta, & z = I, \\ 1 - \vartheta, & z = U. \end{cases}$$

(To symmetrize, use $(\theta_I, \theta_U) = (\vartheta, 1 - \vartheta)$.)

1. Qualitative difference in expressiveness as compared to Poisson case?
2. Likelihood function? For sequence of N samples?
3. Maximum likelihood estimate $\hat{\vartheta}$?
4. Fixing N , what possibly vectorial quantities are sufficient statistics?
Which of these are of minimal dimension?

Exercise: the conjugate distribution

The conjugate distribution for binomial sampling is the Beta distribution, for which the the following expression (with $u > 0$, $v > 0$) allows us to compute the normalization factor:

$$\int_0^1 t^{u-1}(1-t)^{v-1}dt = \frac{\Gamma(u)\Gamma(v)}{\Gamma(u+v)}.$$

Exercise: the conjugate distribution

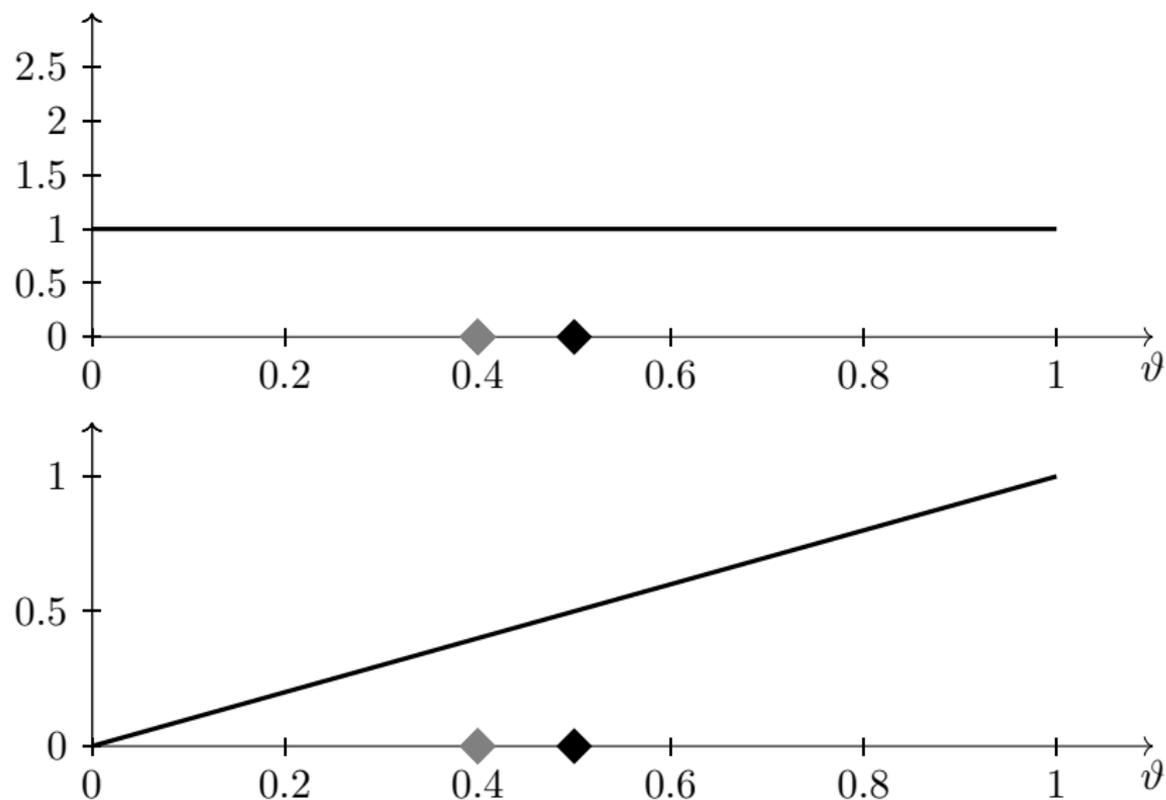
The conjugate distribution for binomial sampling is the Beta distribution, for which the the following expression (with $u > 0$, $v > 0$) allows us to compute the normalization factor:

$$\int_0^1 t^{u-1}(1-t)^{v-1} dt = \frac{\Gamma(u)\Gamma(v)}{\Gamma(u+v)}.$$

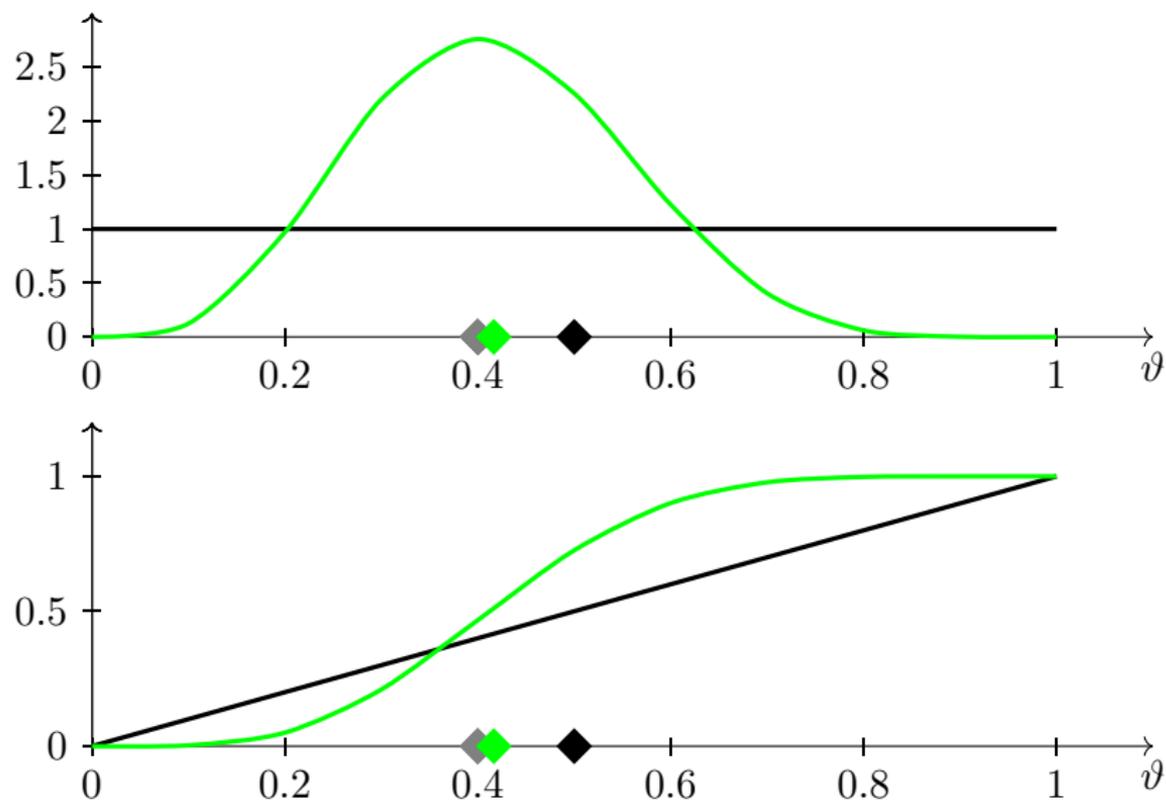
1. Derive the expression for the density function of the conjugate distribution, both in terms of 'mathematical' parameters α and β and in terms of counts s and mean vector $t = (t_I, t_U)$, with $t_I = 1 - t_U$.
2. What are the posterior parameters as a function of s , t , N , and f ?
3. Visualize the priors and posteriors for:
 - ▶ Laplace prior: $s = 2$, $t_I = \frac{1}{2}$;
 - ▶ Haldane prior: $s = 0$;
 - ▶ Jeffrey prior: $s = 1$, $t = \frac{1}{2}$;
 - ▶ Imprecise prior: $s = 1$, $t_I \in [0, 1]$;
 - ▶ Imprecise prior: $s = 5$, $t_I \in [0, 1]$.

Discuss what you see, especially the impact of s .

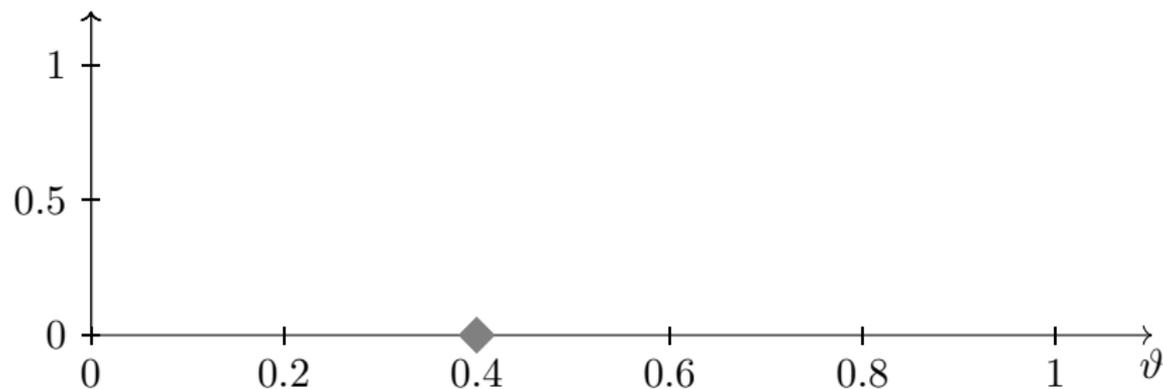
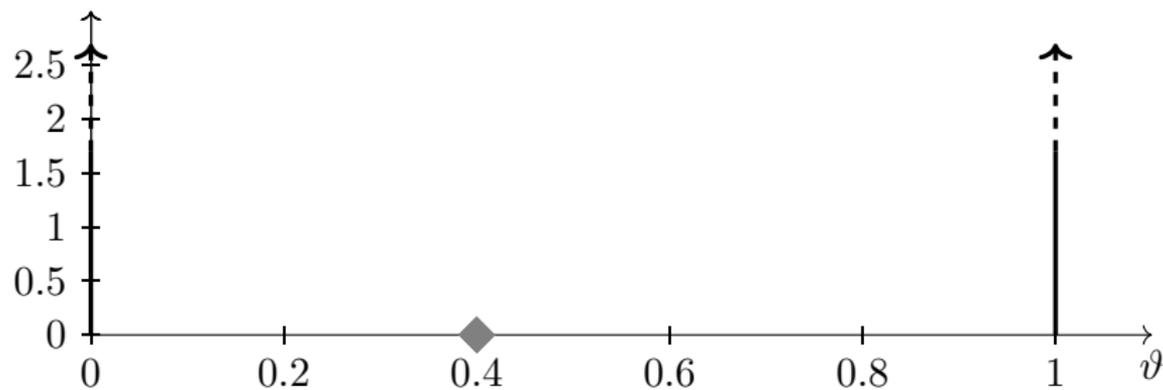
Example: visualization of Laplace prior ($s = 2$, $t_I = \frac{1}{2}$)



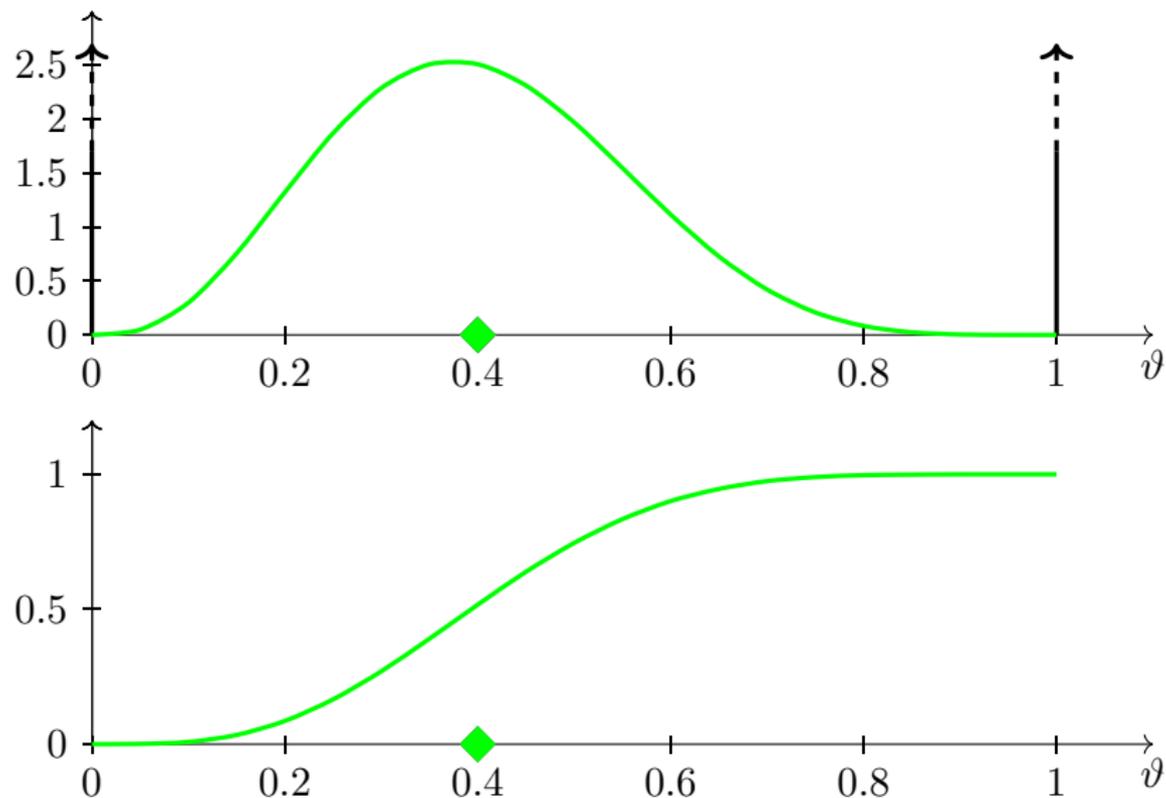
Example: visualization of Laplace posterior ($s = 2$, $t_I = \frac{1}{2}$)



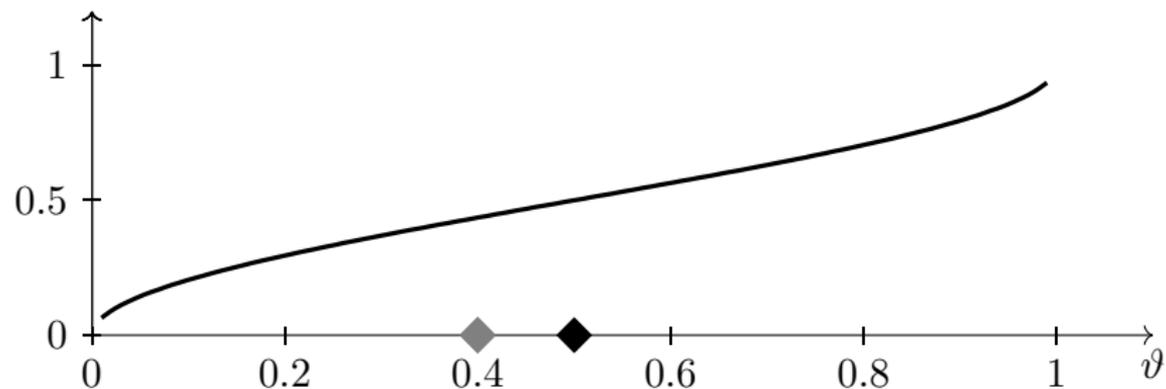
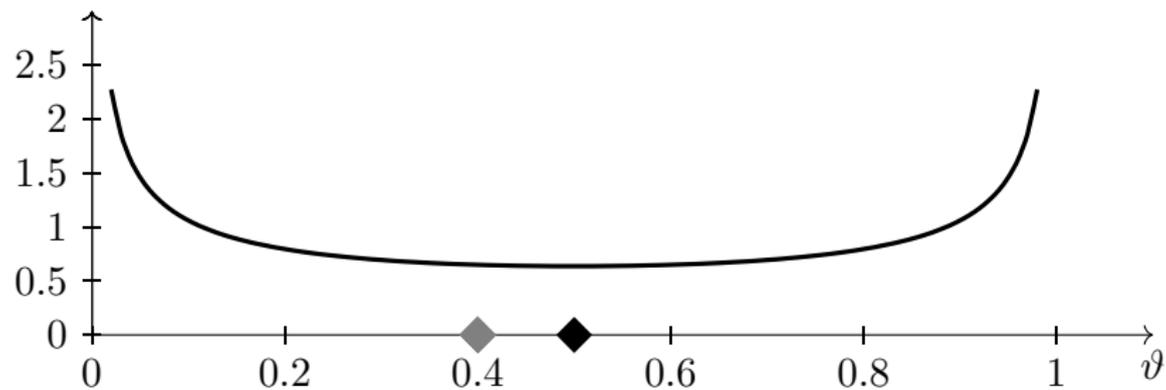
Example: visualization of Haldane prior ($s = 0$)



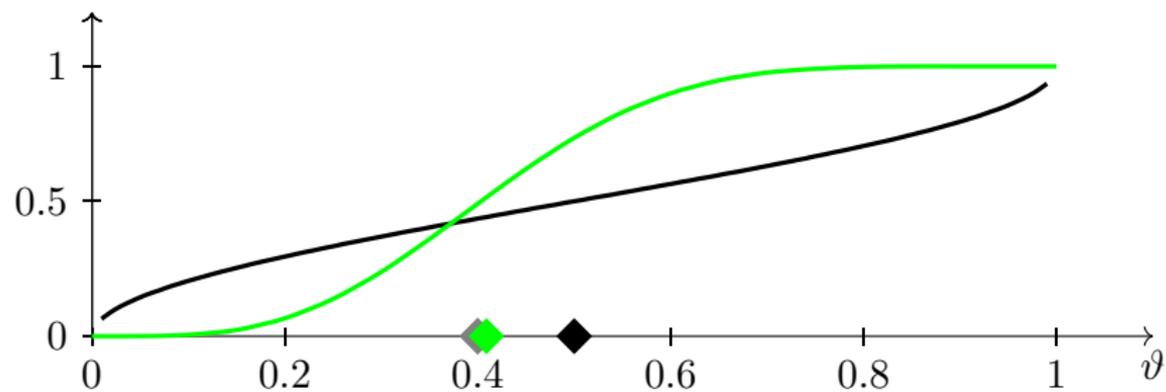
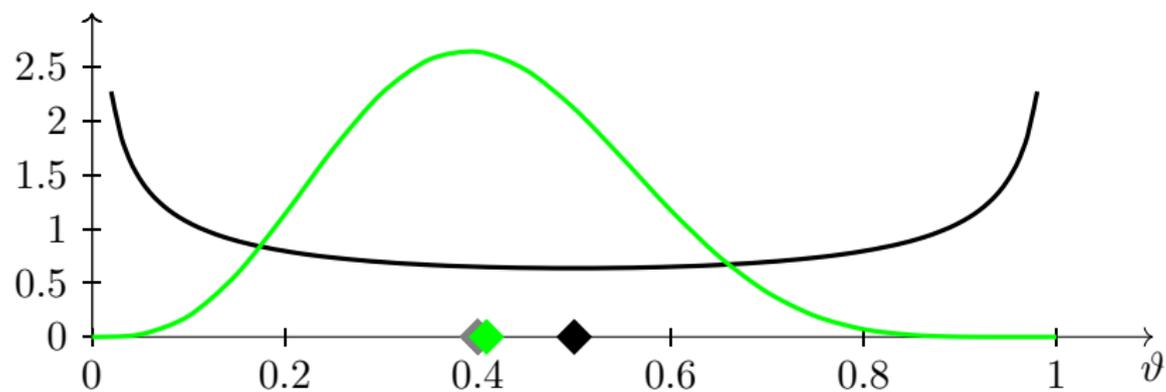
Example: visualization of Haldane posterior ($s = 0$)



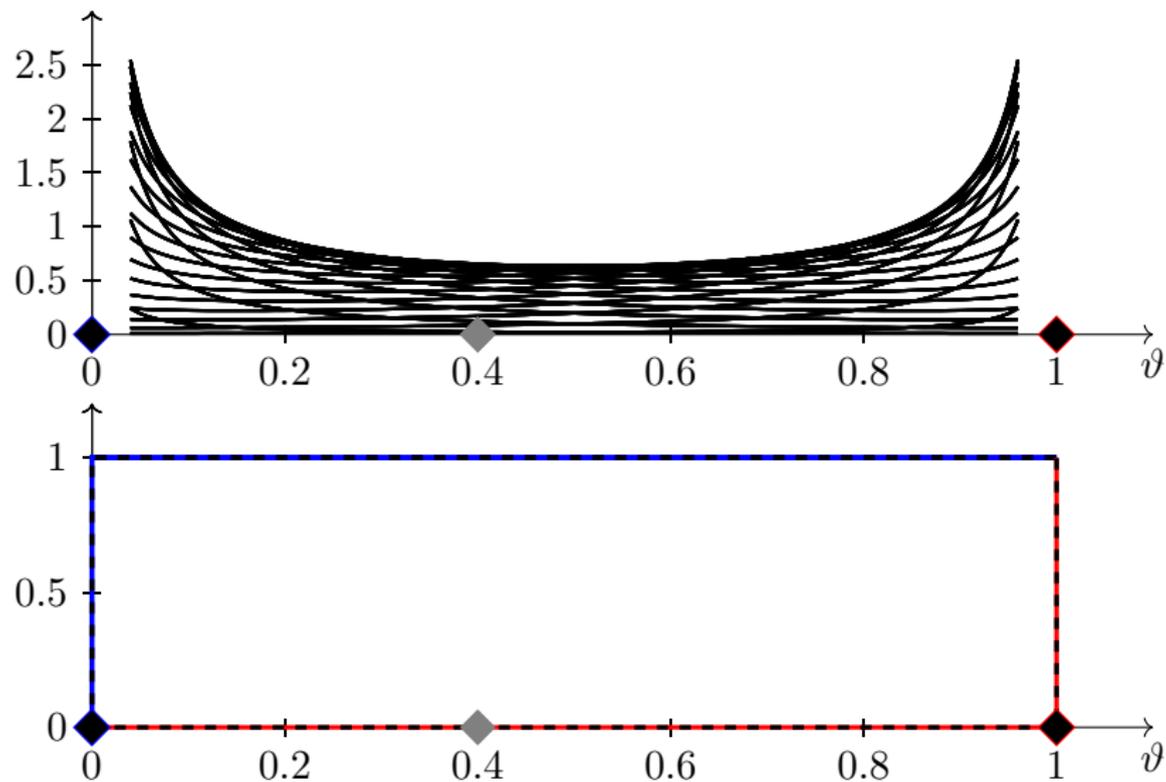
Example: visualization of Jeffrey prior ($s = 1$, $t_I = \frac{1}{2}$)



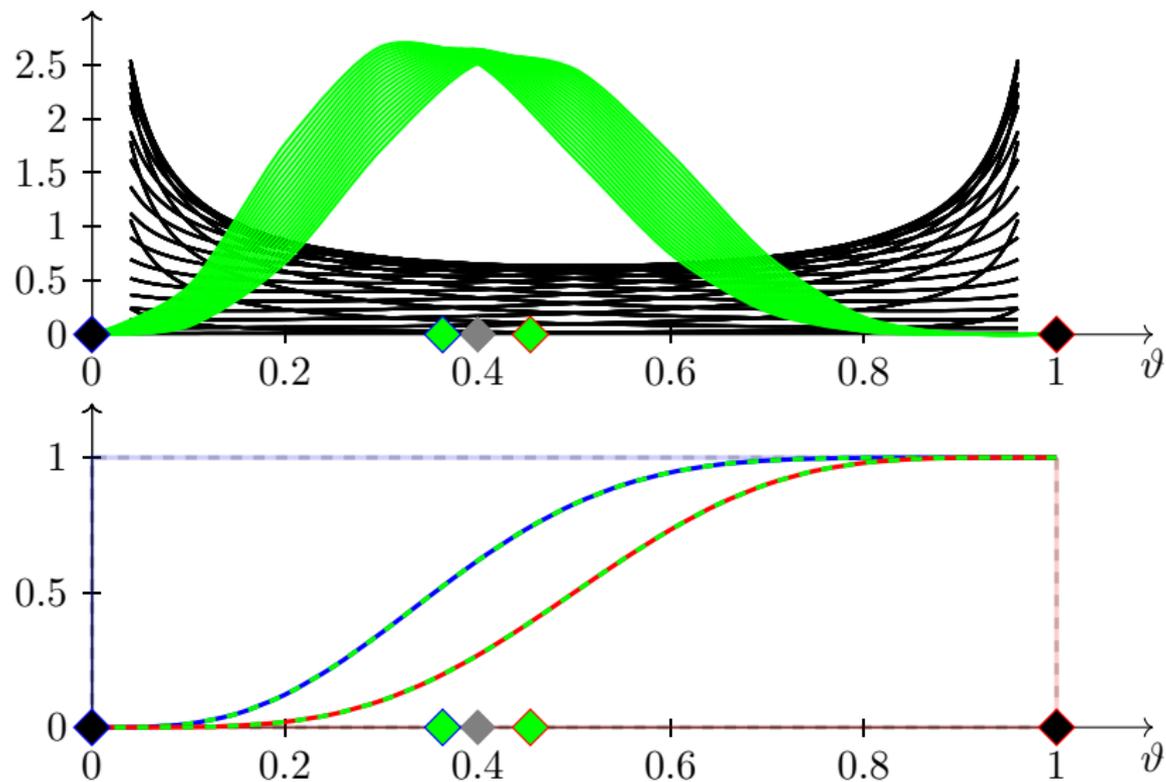
Example: visualization of Jeffrey posterior ($s = 1$, $t_I = \frac{1}{2}$)



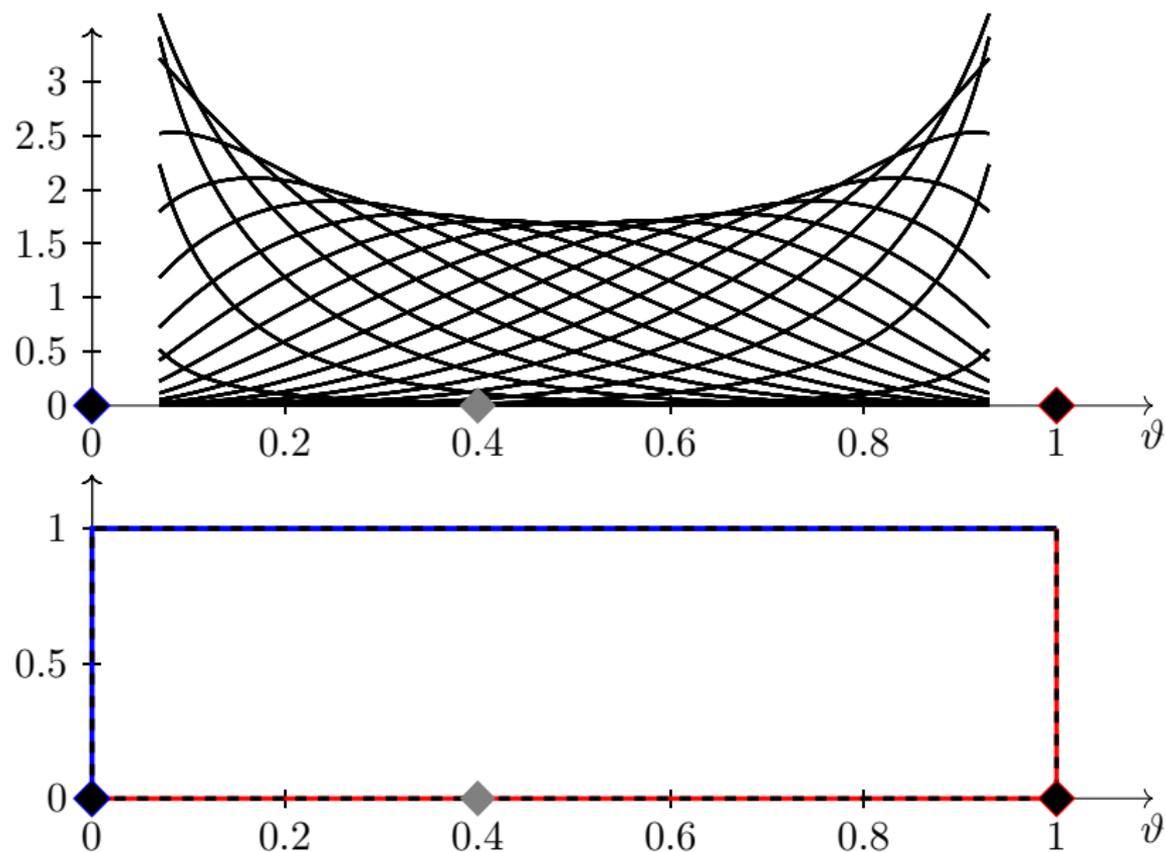
Example: Visualization of Beta priors ($s = 1, t_I \in [0, 1]$)



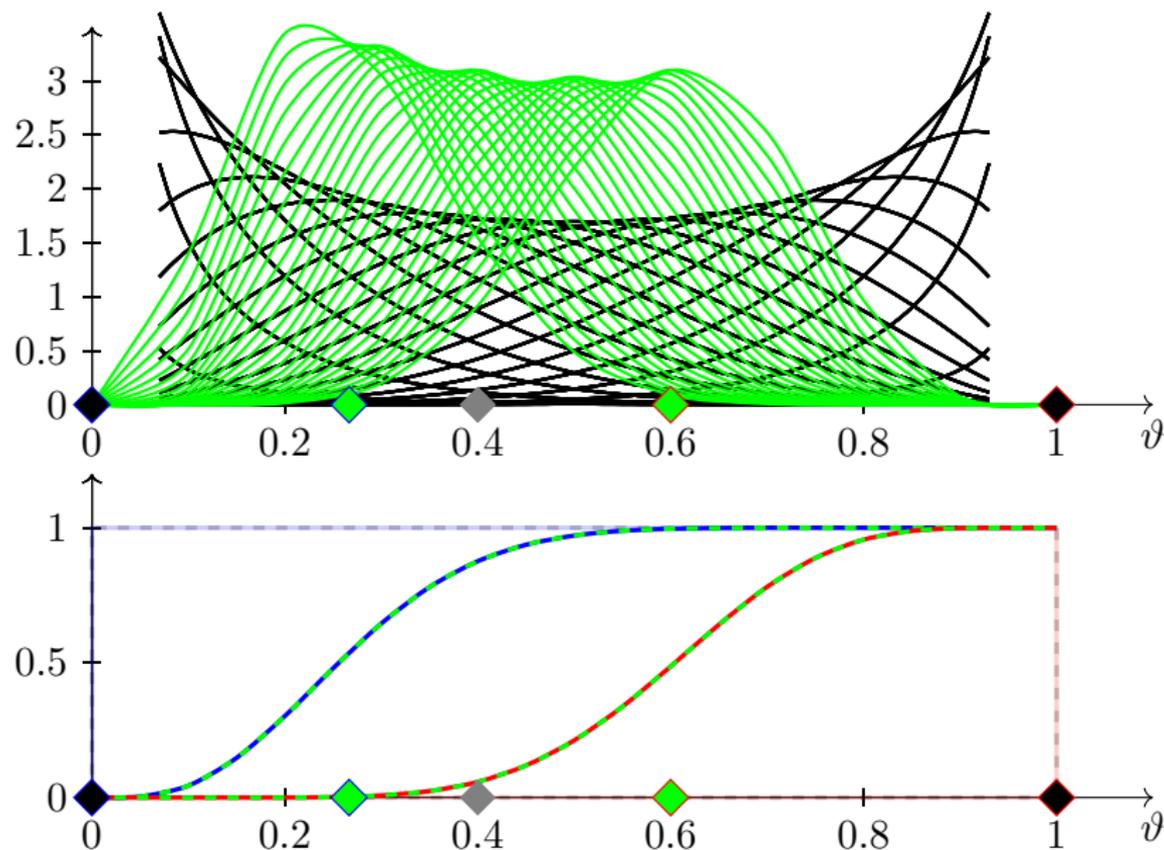
Example: Visualization of Beta post's ($s = 1, t_I \in [0, 1]$)



Example: Visualization of Beta priors ($s = 5, t_I \in [0, 1]$)



Example: Visualization of Beta post's ($s = 5, t_I \in [0, 1]$)



Exercise: from parametric to predictive inference

The conjugate predictive distribution for binomial sampling is the Beta-binomial distribution, whose expression can be written somewhat compactly using generalized binomial coefficients ($r > 0$ and $k \geq 0$):

$$\binom{k+r-1}{r} = \frac{\Gamma(k+r)}{k!\Gamma(r)}.$$

Exercise: from parametric to predictive inference

The conjugate predictive distribution for binomial sampling is the Beta-binomial distribution, whose expression can be written somewhat compactly using generalized binomial coefficients ($r > 0$ and $k \geq 0$):

$$\binom{k+r-1}{r} = \frac{\Gamma(k+r)}{k!\Gamma(r)}.$$

1. Derive the expression of the Beta-binomial probability mass function, both in prior and posterior versions.
2. Write down the expression for the posterior Beta-binomial probability mass function for immediate prediction and simplify it using properties of the Gamma function. Compare the result with the expressions obtained for the NPI model and linear-vacuous model.
3. If $s = 5$ and $t \in [0, 1]$, what are the prior and posterior probability intervals for the event that the next two observations are both interesting?

Symmetry considerations

The framework

We are uncertain about something: the value that a variable X assumes in a **finite** set \mathcal{X} .

The models we will use are **coherent lower previsions** \underline{P} , or equivalently, convex closed sets \mathcal{M} of mass functions p —also called **credal sets**.

Mass functions

A (probability) **mass function** p on \mathcal{X} is a real-valued map on \mathcal{X} such that

$$(\forall x \in \mathcal{X}) p(x) \geq 0 \text{ and } \sum_{x \in \mathcal{X}} p(x) = 1.$$

We denote the set (simplex) of all mass functions on \mathcal{X} by $\Sigma_{\mathcal{X}}$.

With a mass function p there corresponds an **expectation operator** E_p defined on the set $\mathcal{L}(\mathcal{X})$ of all gambles on \mathcal{X} :

$$E_p(f) := \sum_{x \in \mathcal{X}} p(x)f(x) \text{ for all } f: \mathcal{X} \rightarrow \mathbb{R}.$$

Lower and upper previsions

A **coherent lower prevision** on $\mathcal{L}(\mathcal{X})$ is a map $\mathcal{L}(\mathcal{X}) \rightarrow \mathbb{R}$ with the following properties:

1. $\underline{P}(f) \geq \min f$ for all $f \in \mathcal{L}(\mathcal{X})$ [bounds]
2. $\underline{P}(f + g) \geq \underline{P}(f) + \underline{P}(g)$ for all $f, g \in \mathcal{L}(\mathcal{X})$ [super-additivity]
3. $\underline{P}(\lambda f) = \lambda \underline{P}(f)$ for all $f \in \mathcal{L}(\mathcal{X})$ and all real $\lambda \geq 0$ [non-negative homogeneity]

Its **conjugate upper prevision** is defined by

$$\overline{P}(f) := -\underline{P}(-f) \text{ for all } f \in \mathcal{L}(\mathcal{X}).$$

Credal sets

With a convex closed set \mathcal{M} of mass functions, we can construct a coherent lower prevision and the conjugate upper prevision on $\mathcal{L}(\mathcal{X})$ by

$$\underline{P}(f) := \min\{E_p(f) : p \in \mathcal{M}\} \text{ for all } f: \mathcal{X} \rightarrow \mathbb{R}$$

$$\overline{P}(f) := \max\{E_p(f) : p \in \mathcal{M}\} \text{ for all } f: \mathcal{X} \rightarrow \mathbb{R}.$$

Conversely, with a coherent lower prevision \underline{P} there corresponds a convex closed set of mass functions, given by

$$\mathcal{M} := \{p \in \Sigma_{\mathcal{X}} : (\forall f \in \mathcal{L}(\mathcal{X})) E_p(f) \geq \underline{P}(f)\}.$$

A precise model is a singleton $\mathcal{M} = \{p\}$ and the associated lower prevision is the (self-conjugate) expectation operator E_p :

$$E_p(-f) = -E_p(f) \text{ for all } f \in \mathcal{L}(\mathcal{X}).$$

Symmetry

Symmetry is typically modelled by considering a collection of transformations of the space of interest.

Something is considered to be symmetrical when it is left unchanged by these transformations.

Symmetry groups

We will focus on a **group** \mathcal{P} of *permutations* π of \mathcal{X} , meaning that:

1. $\pi_1 \circ \pi_2 \in \mathcal{P}$ for all $\pi_1, \pi_2 \in \mathcal{P}$ [internality]
2. $\pi_1 \circ (\pi_2 \circ \pi_3) = (\pi_1 \circ \pi_2) \circ \pi_3$ for all $\pi_1, \pi_2, \pi_3 \in \mathcal{P}$ [associativity]
3. $\pi \circ \text{id} = \text{id} \circ \pi$ for all $\pi \in \mathcal{P}$ [neutral element]
4. For all $\pi \in \mathcal{P}$ there is some $\pi^{-1} \in \mathcal{P}$ such that
 $\pi \circ \pi^{-1} = \pi^{-1} \circ \pi = \text{id}$ [inverse]

Running example: permutations

We flip a coin twice, and the uncertain outcome X is an element of the finite set $\mathcal{X} = \{HH, HT, TH, TT\}$.

The symmetry we consider is that the order of the observations does not matter, which leads us to identify HT with TH .

$$\mathcal{P} = \{\text{id}, \varpi\},$$

where ϖ is the permutation defined by

$$\begin{pmatrix} HH & HT & TH & TT \\ HH & TH & HT & TT \end{pmatrix}.$$

This a group, with $\varpi^2 = \varpi \circ \varpi = \text{id}$, so $\varpi^{-1} = \varpi$.

Lifting

Our uncertainty models involve gambles, so we need a way to let permutations act on gambles.

This is done by **lifting**:

$$\pi^t f := f \circ \pi \text{ for any } \pi \in \mathcal{P},$$

meaning that

$$(\pi^t f)(x) := f(\pi x) \text{ for all } x \in \mathcal{X}.$$

Running example: lifting

For the gamble

$$f = 2I_{\{HH, TT\}} - I_{\{HT, TH\}},$$

and for the gamble

$$g = I_{\{HH, HT\}} - 3I_{\{TH, TT\}},$$

Running example: lifting

For the gamble

$$f = 2I_{\{HH, TT\}} - I_{\{HT, TH\}},$$

we see that $\varpi^t f = f$ —we call this gamble permutation invariant.

and for the gamble

$$g = I_{\{HH, HT\}} - 3I_{\{TH, TT\}},$$

the permuted gamble is

$$\varpi^t g = I_{\{HH, TH\}} - 3I_{\{HT, TT\}}.$$

Invariant gambles and events

A gamble f is **permutation invariant** if it is left unchanged by the permutations, so

$$\pi^t f = f \text{ or equivalently } f \circ \pi = f \text{ for all } \pi \in \mathcal{P}.$$

An event $A \subseteq \mathcal{X}$ is permutation invariant if its indicator I_A is:

$$(\forall x \in A) \pi x \in A \text{ or equivalently } \pi(A) = A, \text{ for all } \pi \in \mathcal{P}.$$

Invariant atoms

The smallest invariant sets are the so-called **invariant atoms**

$$[x] := \{\pi x : \pi \in \mathcal{P}\},$$

which constitute a partition of \mathcal{X} .

We denote the set of all invariant atoms by $\mathcal{A}_{\mathcal{P}}$.

A gamble is permutation invariant if and only if it is **constant on the invariant atoms**.

So a permutation invariant gamble is completely determined by the values it assumes on the invariant atoms.

Running example: invariant atoms and gambles

The invariant atoms are

Running example: invariant atoms and gambles

The invariant atoms are

$$[HH] = \{HH\} \text{ and } [HT] = [TH] = \{HT, TH\} \text{ and } [TT] = \{TT\}.$$

The permutation invariant gambles are

Running example: invariant atoms and gambles

The invariant atoms are

$$[HH] = \{HH\} \text{ and } [HT] = [TH] = \{HT, TH\} \text{ and } [TT] = \{TT\}.$$

The permutation invariant gambles are the ones that are constant on $[HT] = [TH] = \{HT, TH\}$ and therefore give the same value to HT and TH .

Symmetrical models: (weak) invariance

The uncertainty models \mathcal{M} and \underline{P} are symmetrical: (weakly) **invariant** under the permutations in \mathcal{P} :

$$(\forall f \in \mathcal{L}(\mathcal{X})) \underline{P}(f) = \underline{P}(\pi^t f) \text{ or equivalently } \underline{P} = \underline{P} \circ \pi^t, \text{ for all } \pi \in \mathcal{P}$$

and this is equivalent to

$$(\forall p \in \mathcal{M}) \pi^t p \in \mathcal{M} \text{ for all } \pi \in \mathcal{P},$$

where $\pi^t p := p \circ \pi$.

Symmetrical models: (weak) invariance

The uncertainty models \mathcal{M} and \underline{P} are symmetrical: (weakly) **invariant** under the permutations in \mathcal{P} :

$$(\forall f \in \mathcal{L}(\mathcal{X})) \underline{P}(f) = \underline{P}(\pi^t f) \text{ or equivalently } \underline{P} = \underline{P} \circ \pi^t, \text{ for all } \pi \in \mathcal{P}$$

and this is equivalent to

$$(\forall p \in \mathcal{M}) \pi^t p \in \mathcal{M} \text{ for all } \pi \in \mathcal{P},$$

where $\pi^t p := p \circ \pi$.

A precise model $\mathcal{M} = \{p\}$, or equivalently, a self-conjugate lower prevision (expectation operator) E_p , is therefore invariant if and only if

$$\pi^t p = p \text{ or equivalently } E_p \circ \pi^t = E_p, \text{ for all } \pi \in \mathcal{P}.$$

Running example: weak invariance

Consider the mass functions

$$p_1 := \frac{1}{3}I_{\{HH,HT\}} + \frac{1}{6}I_{\{TH,TT\}} \quad \text{and} \quad p_2 := \frac{1}{3}I_{\{TH,HH\}} + \frac{1}{6}I_{\{HT,TT\}}$$

Running example: weak invariance

Consider the mass functions

$$p_1 := \frac{1}{3}I_{\{HH,HT\}} + \frac{1}{6}I_{\{TH,TT\}} \quad \text{and} \quad p_2 := \frac{1}{3}I_{\{TH,HH\}} + \frac{1}{6}I_{\{HT,TT\}}$$

Observe that $\varpi^t p_1 = p_2$ and consequently also $\varpi^t p_2 = p_1$, so these precise models are *not* permutation invariant with respect to \mathcal{P} .

The credal set

$$\mathcal{M}_1 := \{\alpha p_1 + (1 - \alpha)p_2 : \alpha \in [0, 1]\}$$

Running example: weak invariance

Consider the mass functions

$$p_1 := \frac{1}{3}I_{\{HH,HT\}} + \frac{1}{6}I_{\{TH,TT\}} \text{ and } p_2 := \frac{1}{3}I_{\{TH,HH\}} + \frac{1}{6}I_{\{HT,TT\}}$$

Observe that $\varpi^t p_1 = p_2$ and consequently also $\varpi^t p_2 = p_1$, so these precise models are *not* permutation invariant with respect to \mathcal{P} .

The credal set

$$\mathcal{M}_1 := \{\alpha p_1 + (1 - \alpha)p_2 : \alpha \in [0, 1]\}$$

is permutation invariant, because

$$\varpi^t[\alpha p_1 + (1 - \alpha)p_2] = \alpha \varpi^t p_1 + (1 - \alpha) \varpi^t p_2 = \alpha p_2 + (1 - \alpha)p_1 \in \mathcal{M}_1.$$

Strong invariance: motivation

The subject believes that the 'mechanism generating the observations of the variable X is symmetrical'.

Consider any gamble f .

If the subject believes there is this symmetry, he will be **indifferent** between f and its permutations $\pi^t f$, for all $\pi \in \mathcal{P}$.

He is willing to exchange f for $\pi^t f$ when paid any positive amount of utility $\epsilon > 0$:

$$\underline{P}(\pi^t f - f + \epsilon) \geq 0 \text{ for all } \epsilon > 0,$$

Strong invariance: criterion

Requirement for **strong invariance** (with respect to \mathcal{P}):

$$\underline{P}(\pi^t f - f) = \overline{P}(\pi^t f - f) = 0 \text{ for all } f \in \mathcal{L}(\mathcal{X}) \text{ and all } \pi \in \mathcal{P}.$$

Equivalent to the following requirement for credal sets \mathcal{M} :

$$(\forall p \in \mathcal{M}) \pi^t p = p, \text{ for all } \pi \in \mathcal{P}.$$

So a lower prevision is strongly invariant if and only if it is a lower envelope of (weakly and therefore strongly) invariant precise expectations.

Running example: strong invariance

The mass functions

$$p_3 := \frac{1}{3}I_{\{HH,TT\}} + \frac{1}{6}I_{\{HT,TH\}} \quad \text{and} \quad p_4 := \frac{1}{6}I_{\{HH,TT\}} + \frac{1}{3}I_{\{HT,TH\}}$$

Running example: strong invariance

The mass functions

$$p_3 := \frac{1}{3}I_{\{HH,TT\}} + \frac{1}{6}I_{\{HT,TH\}} \quad \text{and} \quad p_4 := \frac{1}{6}I_{\{HH,TT\}} + \frac{1}{3}I_{\{HT,TH\}}$$

are permutation invariant: $\varpi^t p_3 = p_3$ and $\varpi^t p_4 = p_4$.

The credal set

$$\mathcal{M}_2 := \{\alpha p_3 + (1 - \alpha)p_4 : \alpha \in [0, 1]\}$$

Running example: strong invariance

The mass functions

$$p_3 := \frac{1}{3}I_{\{HH,TT\}} + \frac{1}{6}I_{\{HT,TH\}} \quad \text{and} \quad p_4 := \frac{1}{6}I_{\{HH,TT\}} + \frac{1}{3}I_{\{HT,TH\}}$$

are permutation invariant: $\varpi^t p_3 = p_3$ and $\varpi^t p_4 = p_4$.

The credal set

$$\mathcal{M}_2 := \{\alpha p_3 + (1 - \alpha)p_4 : \alpha \in [0, 1]\}$$

is *strongly* permutation invariant, because

$$\varpi^t[\alpha p_3 + (1 - \alpha)p_4] = \alpha \varpi^t p_3 + (1 - \alpha) \varpi^t p_4 = \alpha p_3 + (1 - \alpha)p_4.$$

Running example: independent strongly permutation invariant models

The independent and permutation invariant precise models are given by

Running example: independent strongly permutation invariant models

The independent and permutation invariant precise models are given by

$$p_r := r^2 I_{\{HH\}} + r(1-r) I_{\{HT, TH\}} + (1-r)^2 I_{\{TT\}}, \text{ for } r \in [0, 1].$$

This implies that, for instance, the credal set

$$\mathcal{M}_{r_1, r_2} := \{\alpha p_{r_1} + (1-\alpha) p_{r_2} : \alpha \in [0, 1]\}$$

is strongly independent and permutation invariant, for any choice of $r_1, r_2 \in [0, 1]$.

The indifferent gambles

Consider the following linear subspace:

$$\mathcal{I}_{\mathcal{P}} := \text{span}(\{\pi^t f - f : f \in \mathcal{L}(\mathcal{X}) \text{ and } \pi \in \mathcal{P}\}).$$

\underline{P} is strongly invariant if and only if

$$\underline{P}(g) = \overline{P}(g) = 0 \text{ for all } g \in \mathcal{I}_{\mathcal{P}},$$

so we can see $\mathcal{I}_{\mathcal{P}}$ as the linear subspace of **indifferent** gambles: the gambles that the subject judges to be equivalent to the zero gamble.

The permutation invariant gambles

The linear subspace of permutation invariant gambles

$$\mathcal{L}_{\mathcal{P}}(\mathcal{X}) := \{f \in \mathcal{L}(\mathcal{X}) : (\forall \pi \in \mathcal{P}) \pi^t f = f\}$$

is the set of all gambles that are **constant on the invariant atoms**.

They are completely determined by the values that they assume on these invariant atoms, and the **dimension** of this space $\mathcal{L}_{\mathcal{P}}(\mathcal{X})$ is therefore the same as the dimension of the linear space $\mathcal{L}(\mathcal{A}_{\mathcal{P}})$ that is linearly isomorphic to it, and therefore **equal to the number of invariant atoms**.

This is generally smaller than the dimension $|\mathcal{X}|$ of the original space $\mathcal{L}(\mathcal{X})$: typically, the more permutations there are in \mathcal{P} , the fewer invariant atoms there are.

Running example: indifferent and invariant gambles

The subspace of indifferent gambles is given by:

Running example: indifferent and invariant gambles

The subspace of indifferent gambles is given by:

$$\mathcal{I}_{\mathcal{P}} = \{\lambda(I_{\{HT\}} - I_{\{TH\}}) : \lambda \in \mathbb{R}\},$$

and is one-dimensional.

The subspace of permutation invariant gambles is given by:

Running example: indifferent and invariant gambles

The subspace of indifferent gambles is given by:

$$\mathcal{I}_{\mathcal{P}} = \{\lambda(I_{\{HT\}} - I_{\{TH\}}) : \lambda \in \mathbb{R}\},$$

and is one-dimensional.

The subspace of permutation invariant gambles is given by:

$$\mathcal{L}_{\mathcal{P}}(\mathcal{X}) = \{\lambda_1 I_{\{HH\}} + \lambda_2 I_{\{TH, HT\}} + \lambda_3 I_{\{TT\}} : \lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}\},$$

and has dimension 3.

Running example: indifferent and invariant gambles

The subspace of indifferent gambles is given by:

$$\mathcal{I}_{\mathcal{P}} = \{\lambda(I_{\{HT\}} - I_{\{TH\}}) : \lambda \in \mathbb{R}\},$$

and is one-dimensional.

The subspace of permutation invariant gambles is given by:

$$\mathcal{L}_{\mathcal{P}}(\mathcal{X}) = \{\lambda_1 I_{\{HH\}} + \lambda_2 I_{\{TH, HT\}} + \lambda_3 I_{\{TT\}} : \lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}\},$$

and has dimension 3.

Observe that $\mathcal{I}_{\mathcal{P}} \cap \mathcal{L}_{\mathcal{P}}(\mathcal{X}) = \{0\}$, and that

$$\mathcal{L}_{\mathcal{P}}(\mathcal{X}) + \mathcal{I}_{\mathcal{P}}$$

Running example: indifferent and invariant gambles

The subspace of indifferent gambles is given by:

$$\mathcal{I}_{\mathcal{P}} = \{\lambda(I_{\{HT\}} - I_{\{TH\}}) : \lambda \in \mathbb{R}\},$$

and is one-dimensional.

The subspace of permutation invariant gambles is given by:

$$\mathcal{L}_{\mathcal{P}}(\mathcal{X}) = \{\lambda_1 I_{\{HH\}} + \lambda_2 I_{\{TH,HT\}} + \lambda_3 I_{\{TT\}} : \lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}\},$$

and has dimension 3.

Observe that $\mathcal{I}_{\mathcal{P}} \cap \mathcal{L}_{\mathcal{P}}(\mathcal{X}) = \{0\}$, and that

$$\mathcal{L}_{\mathcal{P}}(\mathcal{X}) + \mathcal{I}_{\mathcal{P}}$$

$$= \{\lambda_1 I_{\{HH\}} + \lambda_2 I_{\{TH,HT\}} + \lambda_3 I_{\{TT\}} + \lambda(I_{\{HT\}} - I_{\{TH\}}) : \lambda, \lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}\}$$

$$= \{\lambda_1 I_{\{HH\}} + (\lambda_2 - \lambda) I_{\{TH\}} + (\lambda_2 + \lambda) I_{\{HT\}} + \lambda_3 I_{\{TT\}} : \lambda, \lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}\}$$

$$= \mathcal{L}(\mathcal{X}).$$

The projection operator

Consider the following operator $\text{inv}_{\mathcal{P}}$, which maps any gamble to the uniform average of all its permutations:

$$\text{inv}_{\mathcal{P}} f := \frac{1}{|\mathcal{P}|} \sum_{\pi \in \mathcal{P}} \pi^t f.$$

This is a linear transformation of $\mathcal{L}(\mathcal{X})$ that satisfies the following properties:

1. $\text{inv}_{\mathcal{P}} \circ \pi^t = \text{inv}_{\mathcal{P}} = \pi^t \circ \text{inv}_{\mathcal{P}}$ for all $\pi \in \mathcal{P}$ [permutation invariance]
2. $\text{inv}_{\mathcal{P}} \circ \text{inv}_{\mathcal{P}} = \text{inv}_{\mathcal{P}}$ [projection]
3. $\text{kern}(\text{inv}_{\mathcal{P}}) = \mathcal{I}_{\mathcal{P}}$ [kernel]
4. $\text{rng}(\text{inv}_{\mathcal{P}}) = \mathcal{L}_{\mathcal{P}}(\mathcal{X})$ [range]

The uniform distributions over the atoms

The permutation invariant gamble $\text{inv}_{\mathcal{P}} f$ is constant on the invariant atoms.

The constant value it assumes there can also be written as:

$$(\text{inv}_{\mathcal{P}} f)(x) = \frac{1}{|\mathcal{P}|} \sum_{\pi \in \mathcal{P}} f(\pi x) = \frac{1}{|[x]|} \sum_{y \in [x]} f(y) =: U(f|[x]) \text{ for all } x \in \mathcal{X},$$

which is the expectation associated with the **uniform distribution** over the atom $[x]$.

We can see U as a linear map taking gambles f on \mathcal{X} to the corresponding gambles $U(f|\cdot)$ on $\mathcal{A}_{\mathcal{P}}$:

$$U: \mathcal{L}(\mathcal{X}) \rightarrow \mathcal{L}(\mathcal{A}_{\mathcal{P}}), \text{ where } U(f)([x]) := U(f|[x]) = \frac{1}{|[x]|} \sum_{y \in [x]} f(y),$$

Strong invariance representation theorem

Any gamble f can be decomposed uniquely into a permutation invariant part and an indifferent part:

$$f = \underbrace{\text{inv}_{\mathcal{P}} f}_{\in \mathcal{L}_{\mathcal{P}}(\mathcal{X})} + \underbrace{f - \text{inv}_{\mathcal{P}} f}_{\in \mathcal{I}_{\mathcal{P}}}$$

and using coherence, we derive from this that

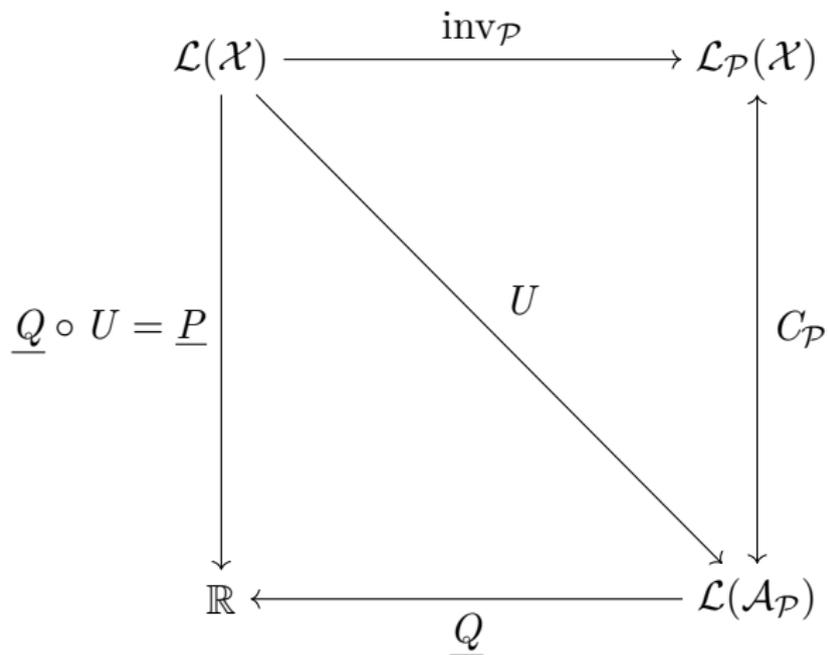
$$\underline{P}(\text{inv}_{\mathcal{P}} f) + \underbrace{\underline{P}(f - \text{inv}_{\mathcal{P}} f)}_{=0} \leq \underline{P}(f) \leq \underline{P}(\text{inv}_{\mathcal{P}} f) + \underbrace{\bar{P}(f - \text{inv}_{\mathcal{P}} f)}_{=0}.$$

Theorem (Strong Invariance Representation Theorem)

A coherent lower prevision \underline{P} is strongly invariant with respect to \mathcal{P} if and only if any (and hence all) of the following equivalent statements holds:

1. $\underline{P} = \underline{P} \circ \text{inv}_{\mathcal{P}}$;
2. There is a coherent lower prevision \underline{Q} on $\mathcal{L}(\mathcal{A}_{\mathcal{P}})$ such that $\underline{P} = \underline{Q} \circ U$.

Strong invariance representation theorem



Running example: representation

With $\mathcal{P} = \{\text{id}, \varpi\}$, we have $\text{inv}_{\mathcal{P}} = \frac{1}{2}(\text{id} + \varpi^t)$, and therefore

$$\text{inv}_{\mathcal{P}} h(x) =$$

Running example: representation

With $\mathcal{P} = \{\text{id}, \varpi\}$, we have $\text{inv}_{\mathcal{P}} = \frac{1}{2}(\text{id} + \varpi^t)$, and therefore

$$\text{inv}_{\mathcal{P}} h(x) = \begin{cases} h(HH) & \text{if } x = HH \\ h(TT) & \text{if } x = TT \\ \frac{h(HT) + h(TH)}{2} & \text{if } x = HT \text{ or } x = TH. \end{cases}$$

All permutation invariant expectation operators have the following form:

$$E_q(h) = h(HH)q([HH]) + \frac{h(HT) + h(TH)}{2}q([HT]) + h(TT)q([TT]),$$

where q is any mass function on $\mathcal{A}_{\mathcal{P}}$.

Exchangeability and its consequences

What is exchangeability?

We consider a **process**:

$$X_1, X_2, X_3, \dots, X_n, \dots \in \mathcal{X}$$

A subject calls these variables **exchangeable** if he decides that the inferences and decisions he makes about these variables will not depend on the order in which these variables are observed.

Here: special case that $\mathcal{X} = \{H, T\}$, but the extension to more general cases is straightforward.

The set of possible outcomes

Let us first look at a **finite** number n of coin flips.

The uncertain outcomes in this situation are now sequences of H and T of length n .

The set of possible outcomes $x = (x_1, \dots, x_n)$ is

$$\mathcal{X}^n = \{(x_1, \dots, x_n) : x_k \in \{H, T\}\}$$

Running example: possible outcomes

Let us look at the case $n = 3$.

The possible outcomes x are:

HHH HHT HTH HTT THH THT TTH TTT,

and all these outcomes make up the set of possible outcomes $\{H, T\}^3$.

The group of permutations

The subject's assumption that the order of observations will not matter leads us to consider the following type of **symmetry**.

Consider all permutations π of the index set $\{1, 2, \dots, n\}$.

We can use these permutations to turn a sequence of observations $x = (x_1, x_2, \dots, x_n)$ into a permuted sequence of observations:

$$\pi x := (x_{\pi(1)}, x_{\pi(2)}, \dots, x_{\pi(n)}).$$

This allows us to define a permutation group \mathcal{P}_n on \mathcal{X}^n .

Running example: the permutations

The $3! = 6$ possible permutations of the index set $\{1, 2, 3\}$, and their actions on the sequence HTH are:

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix} \quad HTH \rightarrow$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} \quad HTH \rightarrow$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \quad HTH \rightarrow$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \quad HTH \rightarrow$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} \quad HTH \rightarrow$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \quad HTH \rightarrow$$

Running example: the permutations

The $3! = 6$ possible permutations of the index set $\{1, 2, 3\}$, and their actions on the sequence HTH are:

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix} \quad HTH \rightarrow HTH$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} \quad HTH \rightarrow$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \quad HTH \rightarrow$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \quad HTH \rightarrow$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} \quad HTH \rightarrow$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \quad HTH \rightarrow$$

Running example: the permutations

The $3! = 6$ possible permutations of the index set $\{1, 2, 3\}$, and their actions on the sequence HTH are:

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix} \quad HTH \rightarrow HTH$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} \quad HTH \rightarrow HHT$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \quad HTH \rightarrow$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \quad HTH \rightarrow$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} \quad HTH \rightarrow$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \quad HTH \rightarrow$$

Running example: the permutations

The $3! = 6$ possible permutations of the index set $\{1, 2, 3\}$, and their actions on the sequence HTH are:

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix} \quad HTH \rightarrow HTH$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} \quad HTH \rightarrow HHT$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \quad HTH \rightarrow THH$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \quad HTH \rightarrow$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} \quad HTH \rightarrow$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \quad HTH \rightarrow$$

Running example: the permutations

The $3! = 6$ possible permutations of the index set $\{1, 2, 3\}$, and their actions on the sequence HTH are:

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix} \quad HTH \rightarrow HTH$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} \quad HTH \rightarrow HHT$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \quad HTH \rightarrow THH$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \quad HTH \rightarrow HHT$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} \quad HTH \rightarrow$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \quad HTH \rightarrow$$

Running example: the permutations

The $3! = 6$ possible permutations of the index set $\{1, 2, 3\}$, and their actions on the sequence HTH are:

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix} \quad HTH \rightarrow HTH$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} \quad HTH \rightarrow HHT$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \quad HTH \rightarrow THH$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \quad HTH \rightarrow HHT$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} \quad HTH \rightarrow THH$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \quad HTH \rightarrow$$

Running example: the permutations

The $3! = 6$ possible permutations of the index set $\{1, 2, 3\}$, and their actions on the sequence HTH are:

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix} \quad HTH \rightarrow HTH$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} \quad HTH \rightarrow HHT$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \quad HTH \rightarrow THH$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \quad HTH \rightarrow HHT$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} \quad HTH \rightarrow THH$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \quad HTH \rightarrow HTH$$

Running example: the permutations

With the permutation $\pi = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$ of the index set $\{1, 2, 3\}$ there corresponds the following permutation of the set of possible observations $\{H, T\}^3$:

$$\begin{aligned} HHH &\rightarrow \pi(HHH) = \\ HHT &\rightarrow \pi(HHT) = \\ HTH &\rightarrow \pi(HTH) = \\ HTT &\rightarrow \pi(HTT) = \\ THH &\rightarrow \pi(THH) = \\ THT &\rightarrow \pi(THT) = \\ TTH &\rightarrow \pi(TTH) = \\ TTT &\rightarrow \pi(TTT) = \end{aligned} ,$$

and we will also denote this permutation by π .

Running example: the permutations

With the permutation $\pi = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$ of the index set $\{1, 2, 3\}$ there corresponds the following permutation of the set of possible observations $\{H, T\}^3$:

$$HHH \rightarrow \pi(HHH) = HHH$$

$$HHT \rightarrow \pi(HHT) =$$

$$HTH \rightarrow \pi(HTH) =$$

$$HTT \rightarrow \pi(HTT) =$$

$$THH \rightarrow \pi(THH) =$$

$$THT \rightarrow \pi(THT) =$$

$$TTH \rightarrow \pi(TTH) =$$

$$TTT \rightarrow \pi(TTT) = \quad ,$$

and we will also denote this permutation by π .

Running example: the permutations

With the permutation $\pi = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$ of the index set $\{1, 2, 3\}$ there corresponds the following permutation of the set of possible observations $\{H, T\}^3$:

$$HHH \rightarrow \pi(HHH) = HHH$$

$$HHT \rightarrow \pi(HHT) = THH$$

$$HTH \rightarrow \pi(HTH) =$$

$$HTT \rightarrow \pi(HTT) =$$

$$THH \rightarrow \pi(THH) =$$

$$THT \rightarrow \pi(THT) =$$

$$TTH \rightarrow \pi(TTH) =$$

$$TTT \rightarrow \pi(TTT) = \quad ,$$

and we will also denote this permutation by π .

Running example: the permutations

With the permutation $\pi = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$ of the index set $\{1, 2, 3\}$ there corresponds the following permutation of the set of possible observations $\{H, T\}^3$:

$$HHH \rightarrow \pi(HHH) = HHH$$

$$HHT \rightarrow \pi(HHT) = THH$$

$$HTH \rightarrow \pi(HTH) = HTH$$

$$HTT \rightarrow \pi(HTT) =$$

$$THH \rightarrow \pi(THH) =$$

$$THT \rightarrow \pi(THT) =$$

$$TTH \rightarrow \pi(TTH) =$$

$$TTT \rightarrow \pi(TTT) = \quad ,$$

and we will also denote this permutation by π .

Running example: the permutations

With the permutation $\pi = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$ of the index set $\{1, 2, 3\}$ there corresponds the following permutation of the set of possible observations $\{H, T\}^3$:

$$HHH \rightarrow \pi(HHH) = HHH$$

$$HHT \rightarrow \pi(HHT) = THH$$

$$HTH \rightarrow \pi(HTH) = HTH$$

$$HTT \rightarrow \pi(HTT) = TTH$$

$$THH \rightarrow \pi(THH) =$$

$$THT \rightarrow \pi(THT) =$$

$$TTH \rightarrow \pi(TTH) =$$

$$TTT \rightarrow \pi(TTT) = \quad ,$$

and we will also denote this permutation by π .

Running example: the permutations

With the permutation $\pi = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$ of the index set $\{1, 2, 3\}$ there corresponds the following permutation of the set of possible observations $\{H, T\}^3$:

$$HHH \rightarrow \pi(HHH) = HHH$$

$$HHT \rightarrow \pi(HHT) = THH$$

$$HTH \rightarrow \pi(HTH) = HTH$$

$$HTT \rightarrow \pi(HTT) = TTH$$

$$THH \rightarrow \pi(THH) = HHT$$

$$THT \rightarrow \pi(THT) =$$

$$TTH \rightarrow \pi(TTH) =$$

$$TTT \rightarrow \pi(TTT) = \quad ,$$

and we will also denote this permutation by π .

Running example: the permutations

With the permutation $\pi = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$ of the index set $\{1, 2, 3\}$ there corresponds the following permutation of the set of possible observations $\{H, T\}^3$:

$$HHH \rightarrow \pi(HHH) = HHH$$

$$HHT \rightarrow \pi(HHT) = THH$$

$$HTH \rightarrow \pi(HTH) = HTH$$

$$HTT \rightarrow \pi(HTT) = TTH$$

$$THH \rightarrow \pi(THH) = HHT$$

$$THT \rightarrow \pi(THT) = THT$$

$$TTH \rightarrow \pi(TTH) =$$

$$TTT \rightarrow \pi(TTT) = ,$$

and we will also denote this permutation by π .

Running example: the permutations

With the permutation $\pi = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$ of the index set $\{1, 2, 3\}$ there corresponds the following permutation of the set of possible observations $\{H, T\}^3$:

$$HHH \rightarrow \pi(HHH) = HHH$$

$$HHT \rightarrow \pi(HHT) = THH$$

$$HTH \rightarrow \pi(HTH) = HTH$$

$$HTT \rightarrow \pi(HTT) = TTH$$

$$THH \rightarrow \pi(THH) = HHT$$

$$THT \rightarrow \pi(THT) = THT$$

$$TTH \rightarrow \pi(TTH) = HTT$$

$$TTT \rightarrow \pi(TTT) = \quad ,$$

and we will also denote this permutation by π .

Running example: the permutations

With the permutation $\pi = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$ of the index set $\{1, 2, 3\}$ there corresponds the following permutation of the set of possible observations $\{H, T\}^3$:

$$HHH \rightarrow \pi(HHH) = HHH$$

$$HHT \rightarrow \pi(HHT) = THH$$

$$HTH \rightarrow \pi(HTH) = HTH$$

$$HTT \rightarrow \pi(HTT) = TTH$$

$$THH \rightarrow \pi(THH) = HHT$$

$$THT \rightarrow \pi(THT) = THT$$

$$TTH \rightarrow \pi(TTH) = HTT$$

$$TTT \rightarrow \pi(TTT) = TTT,$$

and we will also denote this permutation by π .

Running example: invariant atoms

The invariant atoms are

Running example: invariant atoms

The invariant atoms are

$$[HHH] = \{HHH\}$$

$$[HHT] = [HTH] = [THH] = \{HHT, HTH, THH\}$$

$$[HTT] = [THT] = [TTH] = \{HTT, THT, TTH\}$$

$$[TTT] = \{TTT\}$$

so there are four invariant atoms.

Running example: invariant atoms

The invariant atoms are

$$[HHH] = \{HHH\} \quad m_H = 3 \text{ and } m_T = 0$$

$$[HHT] = [HTH] = [THH] = \{HHT, HTH, THH\} \quad m_H = 2 \text{ and } m_T = 1$$

$$[HTT] = [THT] = [TTH] = \{HTT, THT, TTH\} \quad m_H = 1 \text{ and } m_T = 2$$

$$[TTT] = \{TTT\} \quad m_H = 0 \text{ and } m_T = 3$$

so there are four invariant atoms.

Counting

The invariant atoms are completely determined by their **count vectors** $m = (m_H, m_T)$.

$$C: \mathcal{X}^n \rightarrow \mathcal{N}_n$$

with

$$C(x)_H = |\{k : x_k = H\}| = \text{number of heads in the sequence } x$$

$$C(x)_T = |\{k : x_k = T\}| = \text{number of tails in the sequence } x$$

and

$$\mathcal{N}_n := \{(m_H, m_T) \in \mathbb{N}_0^2 : m_H + m_T = n\}.$$

All elements of an atom $[x]$ **have the same count vector** $m = C(x)$, and are completely determined by it:

$$[x] = \{y \in \mathcal{X}^n : C(y) = C(x)\}$$

so we also use the notation $[m]$.

Counting

The number of elements in this atom is equal to the number of possible permutations of x , and therefore given by

$$|[m]| = \binom{n}{m_H \quad m_T} = \binom{n}{m_H} = \frac{n!}{m_H!m_T!}.$$

Running example: counting

The invariant atoms are

$$[(3, 0)] = \{HHH\} \quad \text{has } \binom{3}{3} = 1 \text{ element}$$

$$[(2, 1)] = \{HHT, HTH, THH\} \quad \text{has } \binom{3}{2} = 3 \text{ elements}$$

$$[(1, 2)] = \{HTT, THT, TTH\} \quad \text{has } \binom{3}{1} = 3 \text{ elements}$$

$$[(0, 3)] = \{TTT\} \quad \text{has } \binom{3}{0} = 1 \text{ element}$$

and the set of invariant atoms is in a one-to-one correspondence with the set of count vectors

$$\mathcal{N}_3 = \{(3, 0), (2, 1), (1, 2), (0, 3)\}.$$

The projection operator

What are, in this case, the projection operator $\text{inv}_{\mathcal{P}_n}$ and the related uniform average expectation operator U ?

$$U(f|m) = \frac{1}{\binom{n}{m}} \sum_{y \in [m]} f(y)$$

The projection operator

What are, in this case, the projection operator $\text{inv}_{\mathcal{P}_n}$ and the related uniform average expectation operator U ?

$$U(f|m) = \frac{1}{\binom{n}{m}} \sum_{y \in [m]} f(y) = \text{Hy}(f|m).$$

This is the expectation operator associated with the **hypergeometric distribution**:

Independently taking balls, **without replacement**, from an urn whose composition is determined by the count vector m ; so there are in total n balls, m_H of which are of type 'heads' and m_T of which of type 'tails'.

Finite exchangeability representation theorem

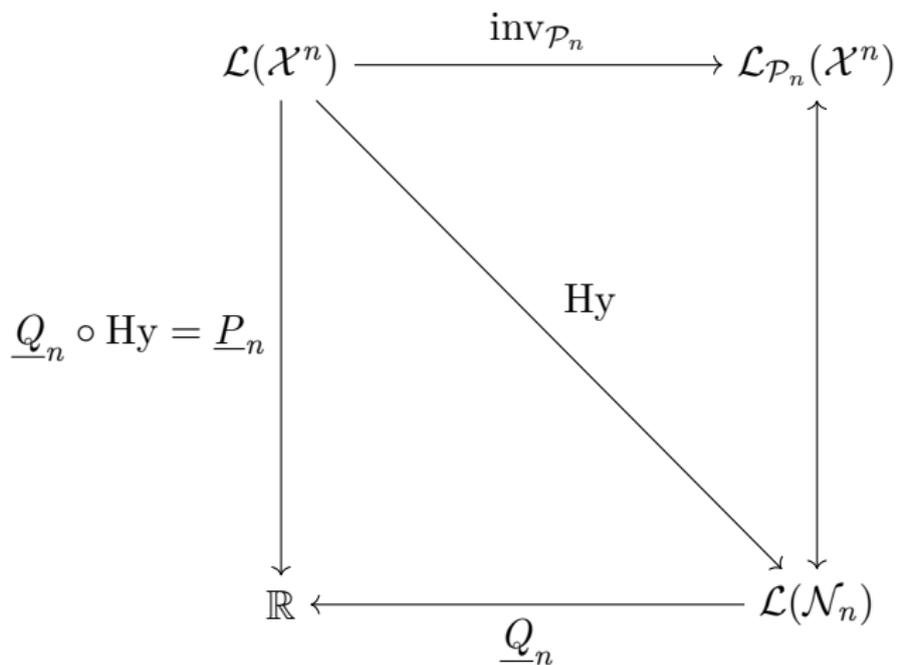
We call a lower prevision \underline{P} on the sequences in \mathcal{X}^n **exchangeable** if it is strongly invariant with respect to the permutations in \mathcal{P}_n .

Our permutation invariance representation theorem now turns into Bruno de Finetti's Representation Theorem for finite exchangeable sequences.

Theorem (de Finetti's Finite Representation Theorem)

*A coherent lower prevision \underline{P}_n on $\mathcal{L}(\mathcal{X}^n)$ is exchangeable if and only if there is some coherent **count** lower prevision \underline{Q}_n on $\mathcal{L}(\mathcal{N}_n)$ such that $\underline{P}_n = \underline{Q}_n \circ \text{Hy}$.*

Finite exchangeability representation theorem



Running example: inference

What can we say about the probability that in a sequence of three exchangeable coin flips, heads is followed by tails?

$$A = \{HTH, HTT, THT, HHT\}.$$

$$\text{Hy}(I_A|(3, 0)) =$$

$$\text{Hy}(I_A|(2, 1)) =$$

$$\text{Hy}(I_A|(1, 2)) =$$

$$\text{Hy}(I_A|(0, 3)) = \quad .$$

Running example: inference

What can we say about the probability that in a sequence of three exchangeable coin flips, heads is followed by tails?

$$A = \{HTH, HTT, THT, HHT\}.$$

$$\text{Hy}(I_A|(3, 0)) = I_A(HHH) = 0$$

$$\text{Hy}(I_A|(2, 1)) =$$

$$\text{Hy}(I_A|(1, 2)) =$$

$$\text{Hy}(I_A|(0, 3)) = \quad .$$

Running example: inference

What can we say about the probability that in a sequence of three exchangeable coin flips, heads is followed by tails?

$$A = \{HTH, HTT, THT, HHT\}.$$

$$\text{Hy}(I_A|(3, 0)) = I_A(HHH) = 0$$

$$\text{Hy}(I_A|(2, 1)) = \frac{1}{3} \left(I_A(HHT) + I_A(HTH) + I_A(THH) \right) = \frac{2}{3}$$

$$\text{Hy}(I_A|(1, 2)) =$$

$$\text{Hy}(I_A|(0, 3)) = \quad .$$

Running example: inference

What can we say about the probability that in a sequence of three exchangeable coin flips, heads is followed by tails?

$$A = \{HTH, HTT, THT, HHT\}.$$

$$\text{Hy}(I_A|(3, 0)) = I_A(HHH) = 0$$

$$\text{Hy}(I_A|(2, 1)) = \frac{1}{3} \left(I_A(HHT) + I_A(HTH) + I_A(THH) \right) = \frac{2}{3}$$

$$\text{Hy}(I_A|(1, 2)) = \frac{1}{3} \left(I_A(TTH) + I_A(THT) + I_A(HTT) \right) = \frac{2}{3}$$

$$\text{Hy}(I_A|(0, 3)) = \quad .$$

Running example: inference

What can we say about the probability that in a sequence of three exchangeable coin flips, heads is followed by tails?

$$A = \{HTH, HTT, THT, HHT\}.$$

$$\text{Hy}(I_A|(3, 0)) = I_A(HHH) = 0$$

$$\text{Hy}(I_A|(2, 1)) = \frac{1}{3} \left(I_A(HHT) + I_A(HTH) + I_A(THH) \right) = \frac{2}{3}$$

$$\text{Hy}(I_A|(1, 2)) = \frac{1}{3} \left(I_A(TTH) + I_A(THT) + I_A(HTT) \right) = \frac{2}{3}$$

$$\text{Hy}(I_A|(0, 3)) = I_A(TTT) = 0.$$

Running example: inference

What can we say about the probability that in a sequence of three exchangeable coin flips, heads is followed by tails?

$$A = \{HTH, HTT, THT, HHT\}.$$

$$\text{Hy}(I_A|(3, 0)) = I_A(HHH) = 0$$

$$\text{Hy}(I_A|(2, 1)) = \frac{1}{3} \left(I_A(HHT) + I_A(HTH) + I_A(THH) \right) = \frac{2}{3}$$

$$\text{Hy}(I_A|(1, 2)) = \frac{1}{3} \left(I_A(TTH) + I_A(THT) + I_A(HTT) \right) = \frac{2}{3}$$

$$\text{Hy}(I_A|(0, 3)) = I_A(TTT) = 0.$$

We only know that the sequence is exchangeable, so the representing lower prevision is $\underline{Q}_3 = \min$, and it follows from the representation theorem that

$$\underline{P}_3(A) = \min \text{Hy}(I_A) = \min \left\{ 0, \frac{2}{3} \right\} = 0$$

$$\overline{P}_3(A) = \max \text{Hy}(I_A) = \max \left\{ 0, \frac{2}{3} \right\} = \frac{2}{3}.$$

Infinite exchangeability

Let us now consider an **infinite** sequence $X_1, X_2, \dots, X_n \dots$ which our subject assumes to be exchangeable, which means that all its finite subsequences are assumed to be exchangeable.

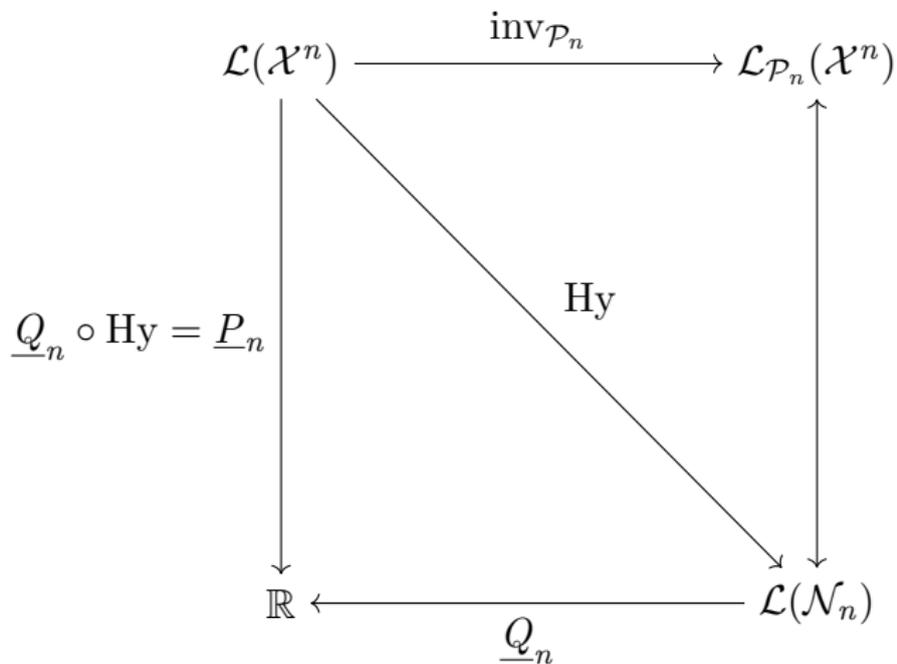
In effect, this means that all the finite sequences

$$X_1, X_2, \dots, X_n \text{ for all } n \in \mathbb{N}$$

are assumed to be exchangeable.

Finite exchangeability representation theorem

For all $n \in \mathbb{N}$:



Binomial expectations and Bernstein polynomials

Consider flipping the same coin independently n times, where the probability of heads on each coin flip is $\theta \in [0, 1]$.

The probability of observing a sequence with count vector $m \in \mathcal{N}_n$ is

$$B_m(\theta) = \binom{n}{m_H} \theta^{m_H} (1 - \theta)^{m_T}.$$

B_m is a polynomial on $[0, 1]$ of degree n , called a **Bernstein basis polynomial** of degree n .

Binomial expectations and Bernstein polynomials

Consider flipping the same coin independently n times, where the probability of heads on each coin flip is $\theta \in [0, 1]$.

The probability of observing a sequence with count vector $m \in \mathcal{N}_n$ is

$$B_m(\theta) = \binom{n}{m_H} \theta^{m_H} (1 - \theta)^{m_T}.$$

B_m is a polynomial on $[0, 1]$ of degree n , called a **Bernstein basis polynomial** of degree n .

For a gamble $g \in \mathcal{L}(\mathcal{N}_n)$ on the counts m , its **binomial expectation** is

$$\text{Mn}(g|\theta) = \sum_{m \in \mathcal{N}_n} g(m) B_m(\theta),$$

a polynomial of degree at most n on $[0, 1]$.

Running example: Bernstein polynomials

The probability of observing m_H heads and m_T tails in a run of $n = 3$ coin flips, is

$$B_{(3,0)}(\theta) =$$

$$B_{(2,1)}(\theta) =$$

$$B_{(1,2)}(\theta) =$$

$$B_{(0,3)}(\theta) =$$

Running example: Bernstein polynomials

The probability of observing m_H heads and m_T tails in a run of $n = 3$ coin flips, is

$$B_{(3,0)}(\theta) = \binom{3}{3} \theta^3 (1 - \theta)^0 = \theta^3$$

$$B_{(2,1)}(\theta) =$$

$$B_{(1,2)}(\theta) =$$

$$B_{(0,3)}(\theta) =$$

Running example: Bernstein polynomials

The probability of observing m_H heads and m_T tails in a run of $n = 3$ coin flips, is

$$B_{(3,0)}(\theta) = \binom{3}{3} \theta^3 (1 - \theta)^0 = \theta^3$$

$$B_{(2,1)}(\theta) = \binom{3}{2} \theta^2 (1 - \theta)^1 = 3\theta^2(1 - \theta) = 3\theta^2 - 3\theta^3$$

$$B_{(1,2)}(\theta) =$$

$$B_{(0,3)}(\theta) =$$

Running example: Bernstein polynomials

The probability of observing m_H heads and m_T tails in a run of $n = 3$ coin flips, is

$$B_{(3,0)}(\theta) = \binom{3}{3} \theta^3 (1 - \theta)^0 = \theta^3$$

$$B_{(2,1)}(\theta) = \binom{3}{2} \theta^2 (1 - \theta)^1 = 3\theta^2(1 - \theta) = 3\theta^2 - 3\theta^3$$

$$B_{(1,2)}(\theta) = \binom{3}{1} \theta^1 (1 - \theta)^2 = 3\theta(1 - 2\theta + \theta^2) = 3\theta - 6\theta^2 + 3\theta^3$$

$$B_{(0,3)}(\theta) =$$

Running example: Bernstein polynomials

The probability of observing m_H heads and m_T tails in a run of $n = 3$ coin flips, is

$$B_{(3,0)}(\theta) = \binom{3}{3} \theta^3 (1 - \theta)^0 = \theta^3$$

$$B_{(2,1)}(\theta) = \binom{3}{2} \theta^2 (1 - \theta)^1 = 3\theta^2(1 - \theta) = 3\theta^2 - 3\theta^3$$

$$B_{(1,2)}(\theta) = \binom{3}{1} \theta^1 (1 - \theta)^2 = 3\theta(1 - 2\theta + \theta^2) = 3\theta - 6\theta^2 + 3\theta^3$$

$$B_{(0,3)}(\theta) = \binom{3}{0} \theta^0 (1 - \theta)^3 = (1 - \theta)^3 = 1 - 3\theta + 3\theta^2 - \theta^3,$$

which are all Bernstein basis polynomials of degree 3.

Running example: Bernstein polynomials

Any polynomial of degree 3

$$p(\theta) = a + b\theta + c\theta^2 + d\theta^3$$

can be written uniquely as a linear combination of these basis polynomials:

$$p = (a + b + c + d)B_{(3,0)} + \left(a + \frac{2b}{3} + \frac{c}{3}\right)B_{(2,1)} + \left(a + \frac{b}{3}\right)B_{(1,2)} + aB_{(0,3)}.$$

Using polynomials rather than counts

For general n , the Bernstein basis polynomials B_m , $m \in \mathcal{N}_n$ of degree n constitute a basis for the linear space $\mathcal{V}_n([0, 1])$ of all polynomials of degree at most n .

Using polynomials rather than counts

For general n , the Bernstein basis polynomials B_m , $m \in \mathcal{N}_n$ of degree n constitute a basis for the linear space $\mathcal{V}_n([0, 1])$ of all polynomials of degree at most n .

The multinomial expectation operator Mn turns any gamble g on the counts in \mathcal{N}_n into a polynomial $\text{Mn}(g)$ on $[0, 1]$, where

$$\text{Mn}(g|\theta) = \sum_{m \in \mathcal{N}_n} g(m) B_m(\theta).$$

Using polynomials rather than counts

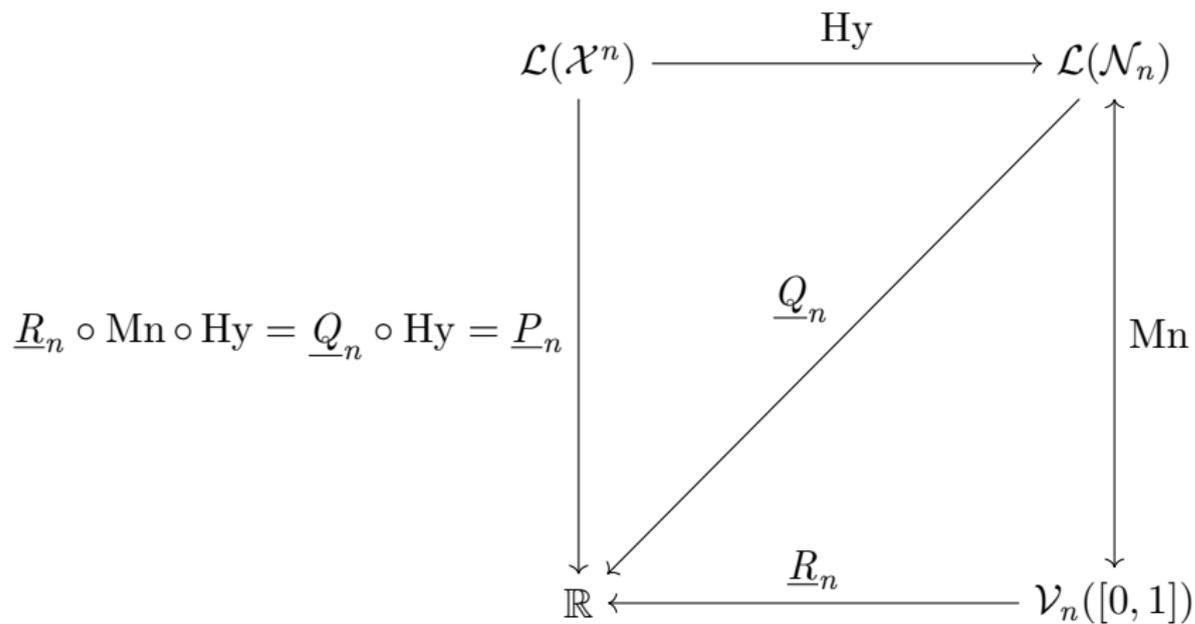
For general n , the Bernstein basis polynomials B_m , $m \in \mathcal{N}_n$ of degree n constitute a basis for the linear space $\mathcal{V}_n([0, 1])$ of all polynomials of degree at most n .

The multinomial expectation operator Mn turns any gamble g on the counts in \mathcal{N}_n into a polynomial $\text{Mn}(g)$ on $[0, 1]$, where

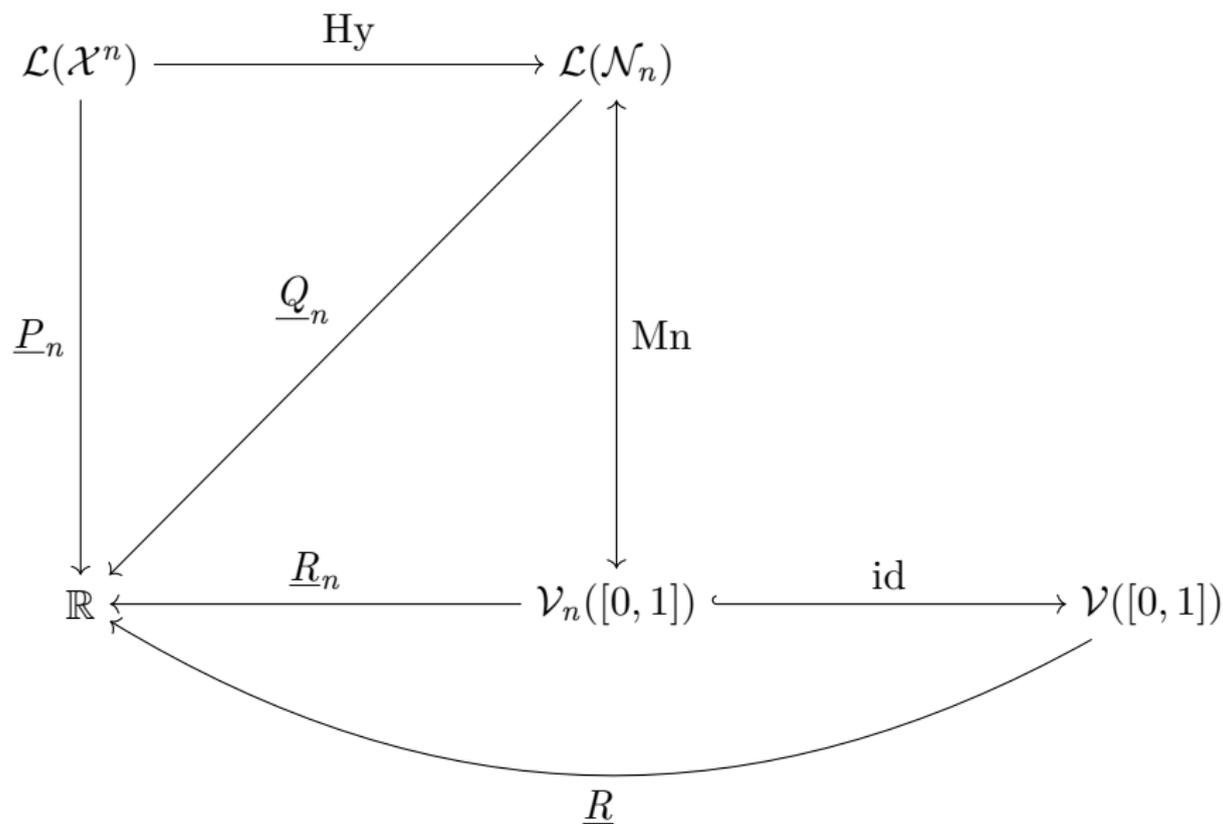
$$\text{Mn}(g|\theta) = \sum_{m \in \mathcal{N}_n} g(m) B_m(\theta).$$

Because the Bernstein basis polynomials of degree n constitute a basis for all polynomials of degree at most n , the map Mn is *one-to-one*—a linear **isomorphism**, which preserves dimension.

Infinite exchangeability representation theorem



Infinite exchangeability representation theorem



Infinite exchangeability representation theorem

Theorem (de Finetti's Infinite Representation Theorem)

A coherent lower prevision \underline{P} on $\bigcup_{n \in \mathbb{N}} \mathcal{L}(\mathcal{X}^n)$ is exchangeable if and only if there is some coherent *frequency* lower prevision \underline{R} on $\mathcal{V}([0, 1])$ such that $\underline{P} = \underline{R} \circ \text{Hy} \circ \text{Mn}$.

Running example: inference

We want to find out about the probability of the event A —‘heads followed by tails in a run of three coin flips’—using the polynomial representation.

Recall that, with $g = \text{Hy}(I_A)$,

$$g(3, 0) = 0 \text{ and } g(2, 1) = \frac{2}{3} \text{ and } g(1, 2) = \frac{2}{3} \text{ and } g(0, 3) = 0$$

and therefore the count gamble g corresponds to the polynomial

$$\text{Mn}(g|\theta) =$$

Running example: inference

We want to find out about the probability of the event A —‘heads followed by tails in a run of three coin flips’—using the polynomial representation.

Recall that, with $g = \text{Hy}(I_A)$,

$$g(3, 0) = 0 \text{ and } g(2, 1) = \frac{2}{3} \text{ and } g(1, 2) = \frac{2}{3} \text{ and } g(0, 3) = 0$$

and therefore the count gamble g corresponds to the polynomial

$$\text{Mn}(g|\theta) = \frac{2}{3}B_{(2,1)}(\theta) + \frac{2}{3}B_{(1,2)}(\theta) = \frac{2}{3}[3\theta^2(1-\theta) + 3\theta(1-\theta)^2] = 2\theta(1-\theta).$$

Running example: inference

We want to find out about the probability of the event A —‘heads followed by tails in a run of three coin flips’—using the polynomial representation.

Recall that, with $g = \text{Hy}(I_A)$,

$$g(3, 0) = 0 \text{ and } g(2, 1) = \frac{2}{3} \text{ and } g(1, 2) = \frac{2}{3} \text{ and } g(0, 3) = 0$$

and therefore the count gamble g corresponds to the polynomial

$$\text{Mn}(g|\theta) = \frac{2}{3}B_{(2,1)}(\theta) + \frac{2}{3}B_{(1,2)}(\theta) = \frac{2}{3}[3\theta^2(1-\theta) + 3\theta(1-\theta)^2] = 2\theta(1-\theta).$$

Because **infinite** exchangeability is all we have assumed, $\underline{P} = \inf$, and therefore

$$\underline{P}(A) = \inf_{\theta \in [0,1]} 2\theta(1-\theta) = 0 \text{ and } \overline{P}(A) = \sup_{\theta \in [0,1]} 2\theta(1-\theta) = \frac{1}{2}.$$