Tutorial 3

Statistical Computation and Analysis

Spring 2025

Tutorial Outline

- Bayes theorem
- Likelihood
- Posterior
- Prior
- Preliz
- Model visualization

Bayes Theorem

- $P(\text{model}|\text{data}) = \dfrac{P(\text{data}|\text{model})P(\text{model})}{\sum_{\text{models}} P(\text{data}|\text{model})P(\text{model})}$

- Prior distribution: what do we know about the values of the model before seeing the data.
- Likelihood: how we will introduce our data.
- Posterior distribution: the result of Bayesian analysis -> reflects all we know about our question given our data and model.
  - Probability distribution for the parameters in our model
  - Not a single value
- Marginal likelihood: probability of observing the data averaged over all the possible values the parameter can take
  - Normalization factor

Bayes Theorem

- Let's calculate the probability of having a disease given a positive test result.
- $P(disease|\text{positive } test) = \frac{P(\text{positive test}|\text{disease})P(\text{disease})}{\sum_{\text{condition}} P(\text{positive test}|\text{condition})P(\text{condition})}$

- $P(positive\ test|disease) = sensitivity = \frac{\#true\ positive}{\#\ sick}$
- $P(disease) = disease\ prevelence$
- $\sum_{condition} P(positive\ test|condition)P(condition) =$

$= P(positive\ test|disease)P(disease) +$
$= P(positive\ test|not\ disease)P(not\ disease) =$
$= sensitivity \cdot prevelance + \frac{\#false\ positive}{\#\ not\ sick} \cdot (1 - prevelance)$

- $\frac{\#false\ positive}{\#\ not\ sick} = 1 - \text{specificity}$

Bayes Theorem

- Sensitivity = 0.7
- Specificity = 0.95
- Prevalence = 0.05
- $P(disease|\text{positive } test) = \frac{P(\text{positive test}|\text{disease})P(\text{disease})}{\sum_{\text{condition}} P(\text{positive test}|\text{condition})P(\text{condition})}$

$$= \frac{0.7 \cdot 0.05}{0.7 \cdot 0.05 + (1 - 0.95) \cdot (1 - 0.05)} = 0.42$$

Bayes Theorem

Effect of likelihood:

- Sensitivity = 0.9
- Specificity = 0.95
- Prevalence = 0.05
- $P(disease|\text{positive } test) = \frac{P(\text{positive test}|\text{disease})P(\text{disease})}{\sum_{\text{condition}} P(\text{positive test}|\text{condition})P(\text{condition})}$

$$= \frac{0.9 \cdot 0.05}{0.9 \cdot 0.05 + (1-0.95) \cdot (1-0.05)} = 0.48$$

- Why does this result make sense?
  - The sensitivity of the test increased -> the test got better -> the chance of being sick given a positive result is higher.

Bayes Theorem

Effect of prior:

- Sensitivity = 0.7
- Specificity = 0.95
- Prevalence = 0.20
- $P(disease|\text{positive } test) = \frac{P(\text{positive test}|\text{disease})P(\text{disease})}{\sum_{\text{condition}} P(\text{positive test}|\text{condition})P(\text{condition})}$

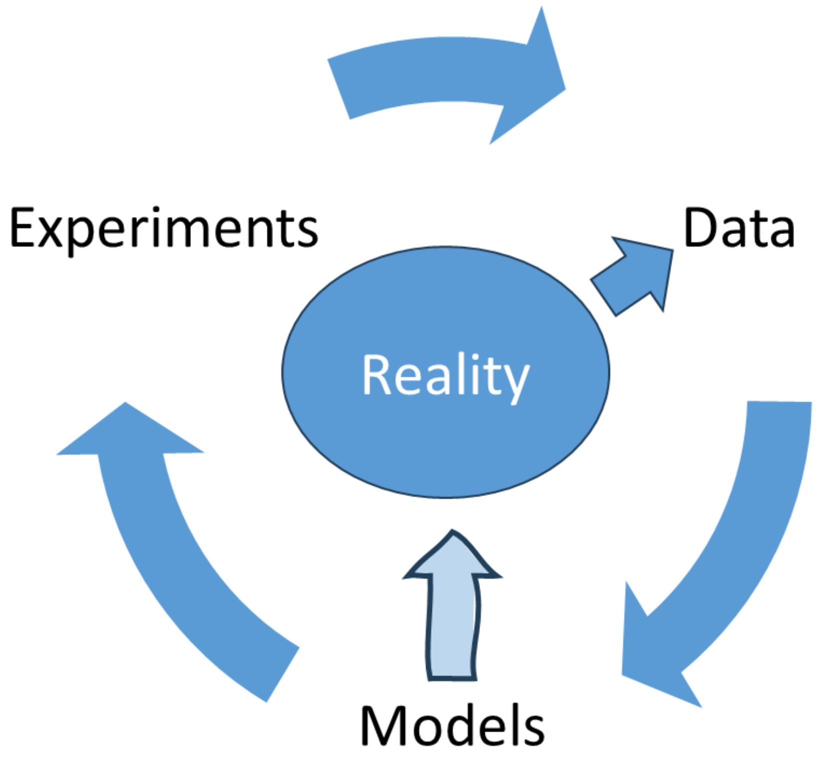
$$= \frac{0.7 \cdot 0.2}{0.7 \cdot 0.2 + (1 - 0.95) \cdot (1 - 0.2)} = 0.77$$

- Why does this result make sense?
  - Even if the test's sensitivity does not change, the fact that there's a higher prior probability to have the disease makes the posterior higher.

Models

- A model is a specific idea about how data comes from reality
  - A particular distribution
    - With specific parameters

- $P(\text{model}|\text{data}) = \frac{P(\text{data}|\text{model})P(\text{model})}{\sum_{\text{models}} P(\text{data}|\text{model})P(\text{model})}$
- $P(\text{model}|\text{data}) \propto P(\text{data}|\text{model})P(\text{model})$
- We will use data to update our model



Likelihood

- Let's model the number of emails a 3rd year biomedical engineering student receives on average per day.
- What distribution is appropriate?
  - Poisson

$$f(k, \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

- The probability of a given number of events occurring in a fixed interval
  - Events occur at a known constant mean rate - λ
  - Events occur independently of the time since the last event

Likelihood

- $P(\lambda|\text{data}) \propto \boldsymbol{P}(\mathbf{data}|\boldsymbol{\lambda})P(\lambda)$

- $Poisson\ P(k,\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$

- For n iid observations:

  - $P(data|\lambda) = \prod_{i=1}^{n} \frac{\lambda^{k_i} e^{-\lambda}}{k_i!} = \frac{\lambda^{\sum_{i=1}^{n} ki} e^{-n\lambda}}{\prod_{i=1}^{n} k_i!}$

  - $S = \sum_{i=1}^{n} ki$ (total number of observed events)

  - $P(data|\lambda) = \prod_{i=1}^{n} \frac{\lambda^{k_i} e^{-\lambda}}{k_i!} = \frac{\lambda^{S} e^{-n\lambda}}{\prod_{i=1}^{n} k_i!}$

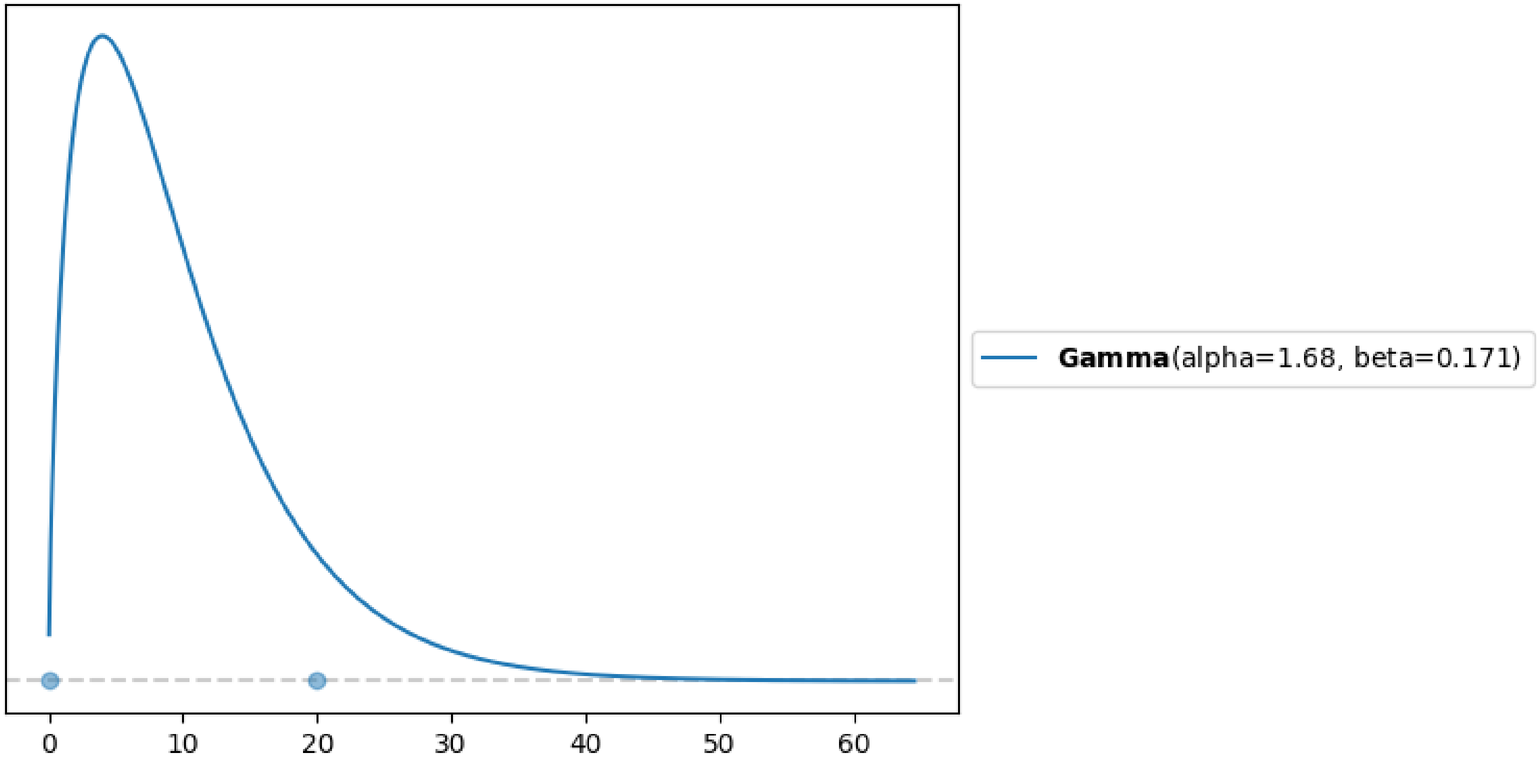
- https://docs.google.com/forms/d/e/1FAIpQLSft4eVg_WuREK27fZdXFnpXCdgbbKLlxVSRyDpXuuOTIfuHng/viewform?usp=dialog

Bayes Theorem

- $P(\lambda|\text{data}) \propto P(\text{data}|\lambda)\boldsymbol{P}(\boldsymbol{\lambda})$

Prior

- $P(\lambda) \sim Gamma(\alpha, \beta) = \dfrac{\beta^{\alpha}\lambda^{\alpha-1}e^{-\beta\lambda}}{\Gamma(\alpha)}$



Bayes Theorem

- $P(\lambda|\text{data}) \propto P(\text{data}|\lambda)P(\lambda)$

$$= \frac{\lambda^S e^{-n\lambda}}{\prod_{i=1}^{n} k_i!} \cdot \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\beta\lambda}}{\Gamma(\alpha)}$$

- Leaving out the constants:

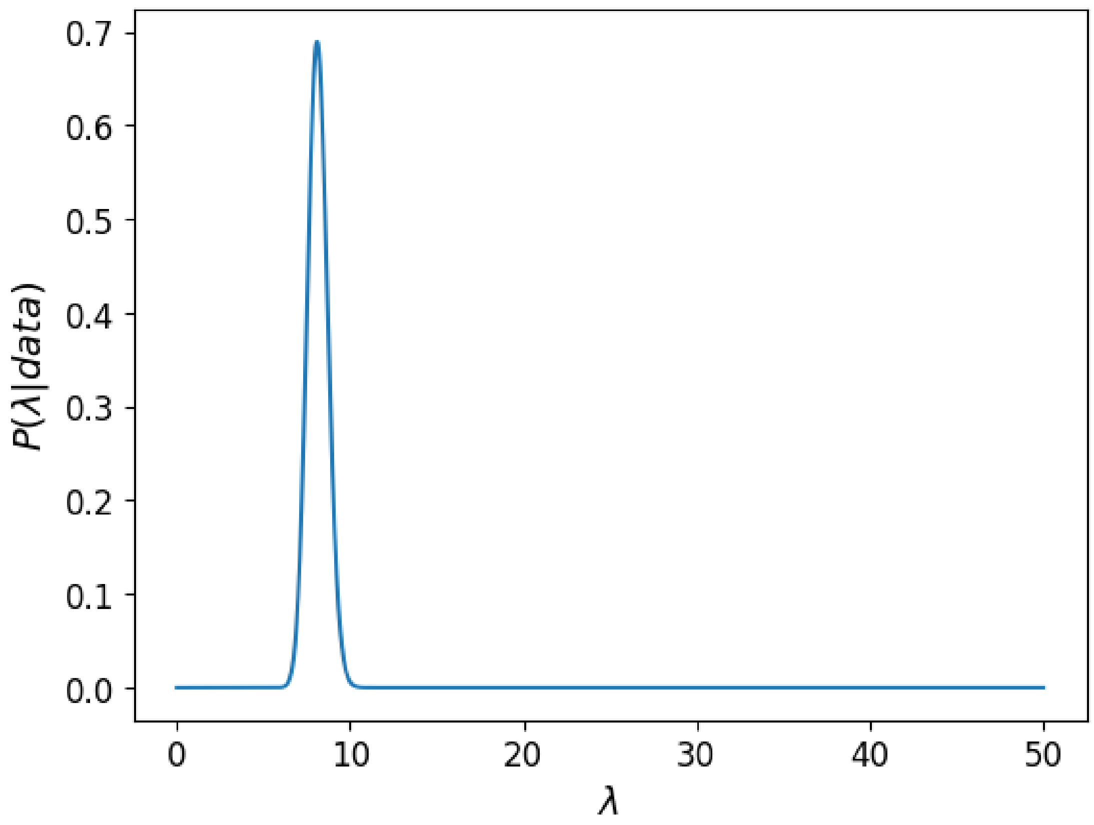
$$P(\lambda|\text{data}) \propto \lambda^{\alpha-1+S} e^{-(\beta+n)\lambda}$$

- Therefore:

$$P(\lambda|\text{data}) \sim Gamma(\alpha + S, n + \beta)$$

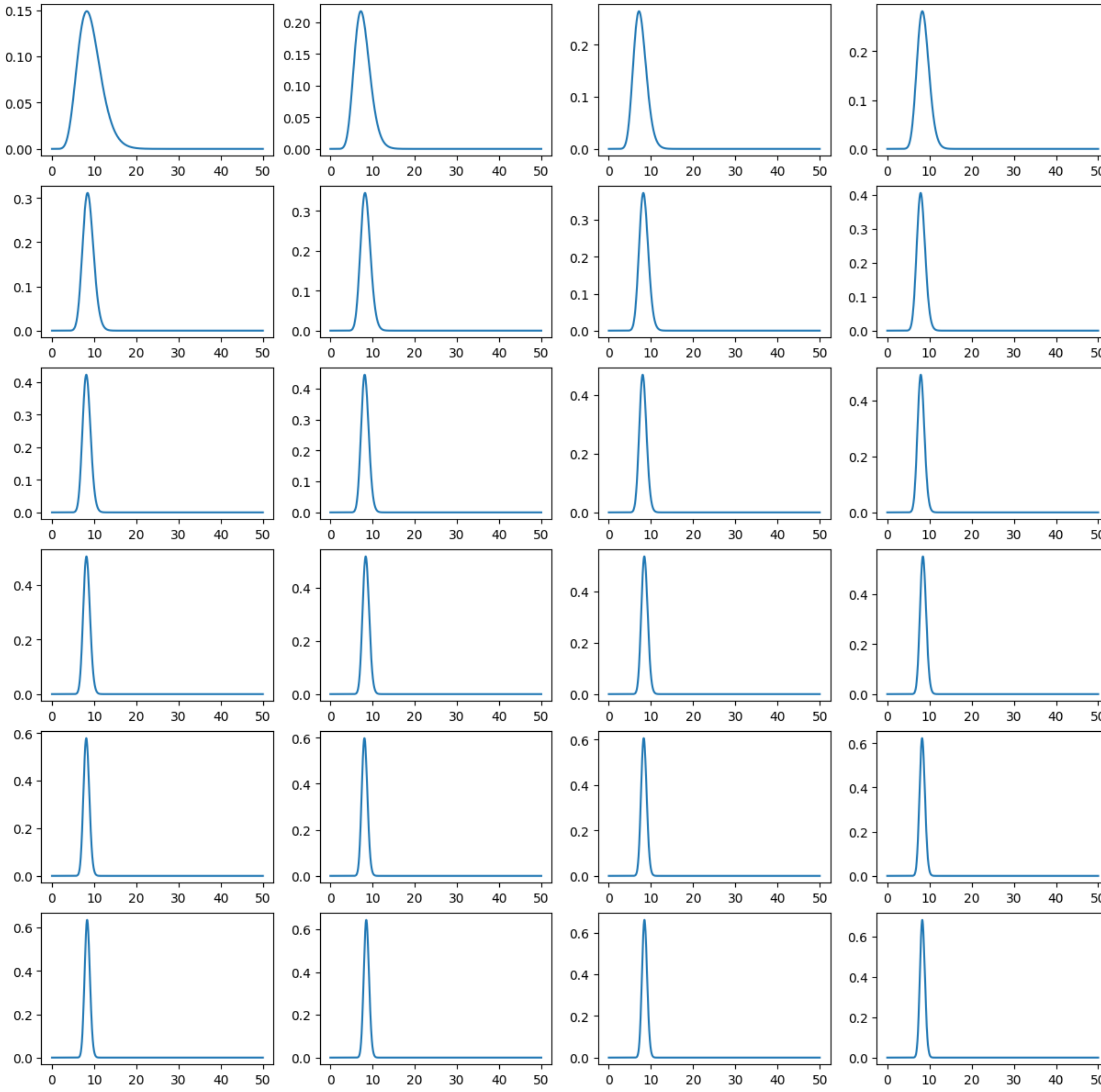
Posterior

- $P(\lambda|\text{data}) \sim Gamma(\alpha + S, n + \beta)$
  - S = total number of emails received
  - n = Number of students
  - With simulated data ($\lambda = 8$)



Posterior

- We can see the effect of each data point



Prior

- Probability distribution
- Describes our prior beliefs on the values of the model
- If we have no prior beliefs, what distribution would we use?
  - Uniform
- If we know for a fact what the values of the model are, what function would we use (hypothetical)?
  - Delta
- In reality, we generally use something in between.

Prior

- Choose a distribution with values that make sense.
- The gamma distribution is defined for non-negative values.
  - The average number of emails received has to be non-negative.
- The gamma distribution is a **conjugate prior** for the Poisson distribution.
- We chose the shape and scale parameters based on our belief of how many emails students receive each day.

Conjugate Priors

- The prior and posterior have the same shape distribution with updated parameters.

Prior	Likelihood
Gamma	Poisson
Gamma	Exponential
Beta	Binomial
Beta	Bernoulli
Gaussian	Mean: gaussian Variance: Inverse Gamma

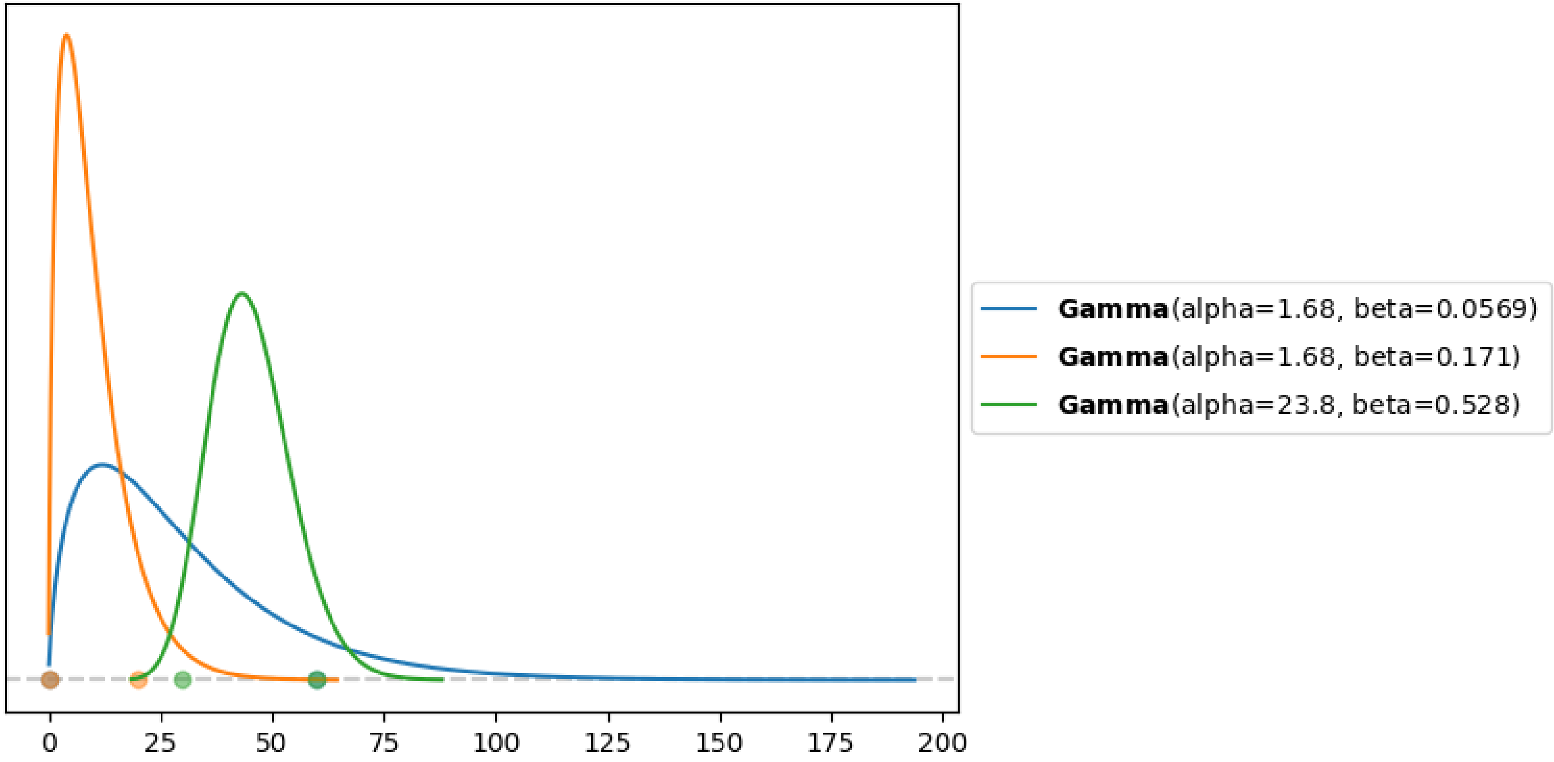
Effect of the Prior

- Preliz distributions documentation:

https://preliz.readthedocs.io/en/latest/distributions.html

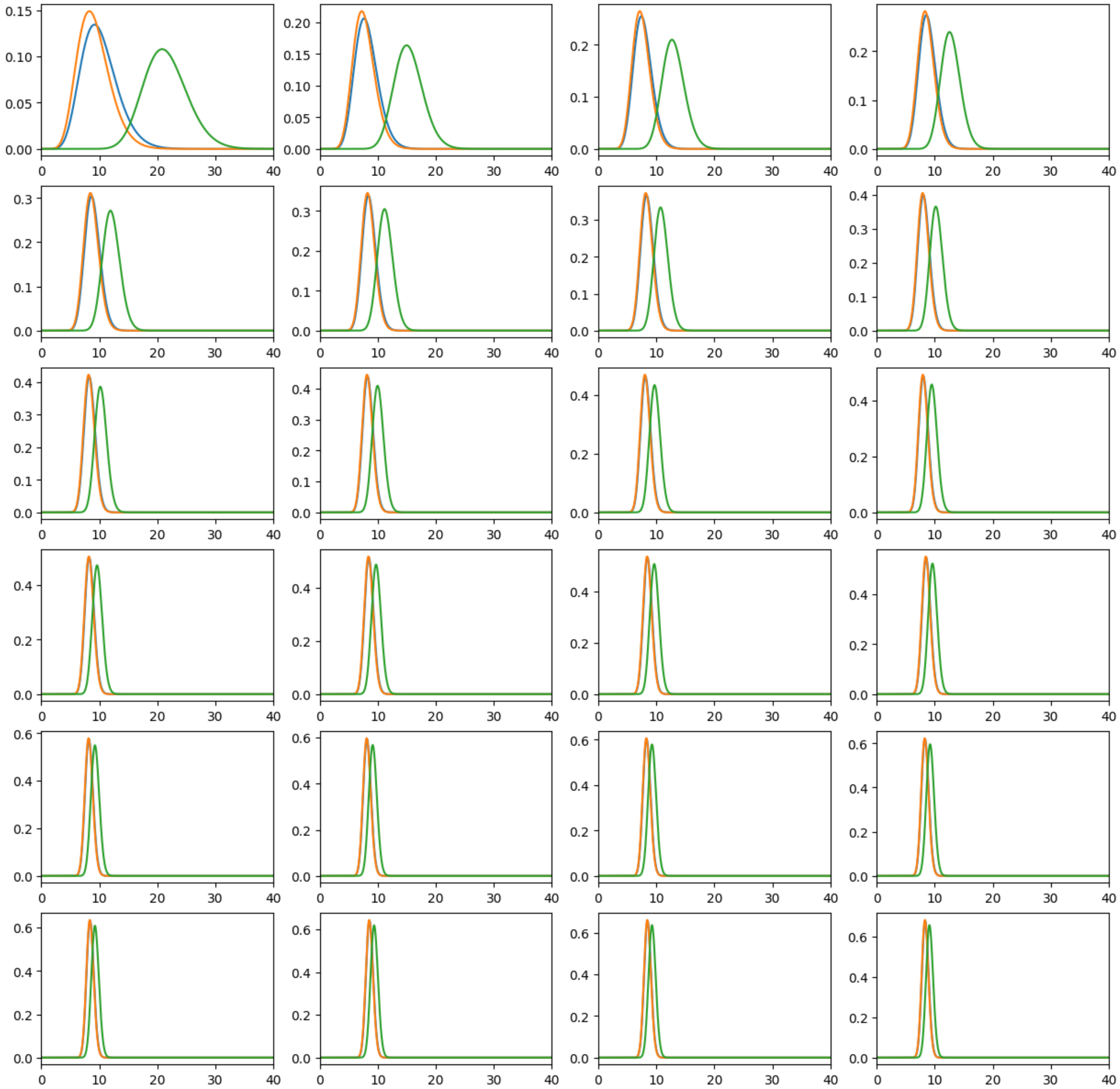
Prior

- Let’s look at the effect of different priors
  - Which is the least informative?
  - Which is likely wrongly informative?



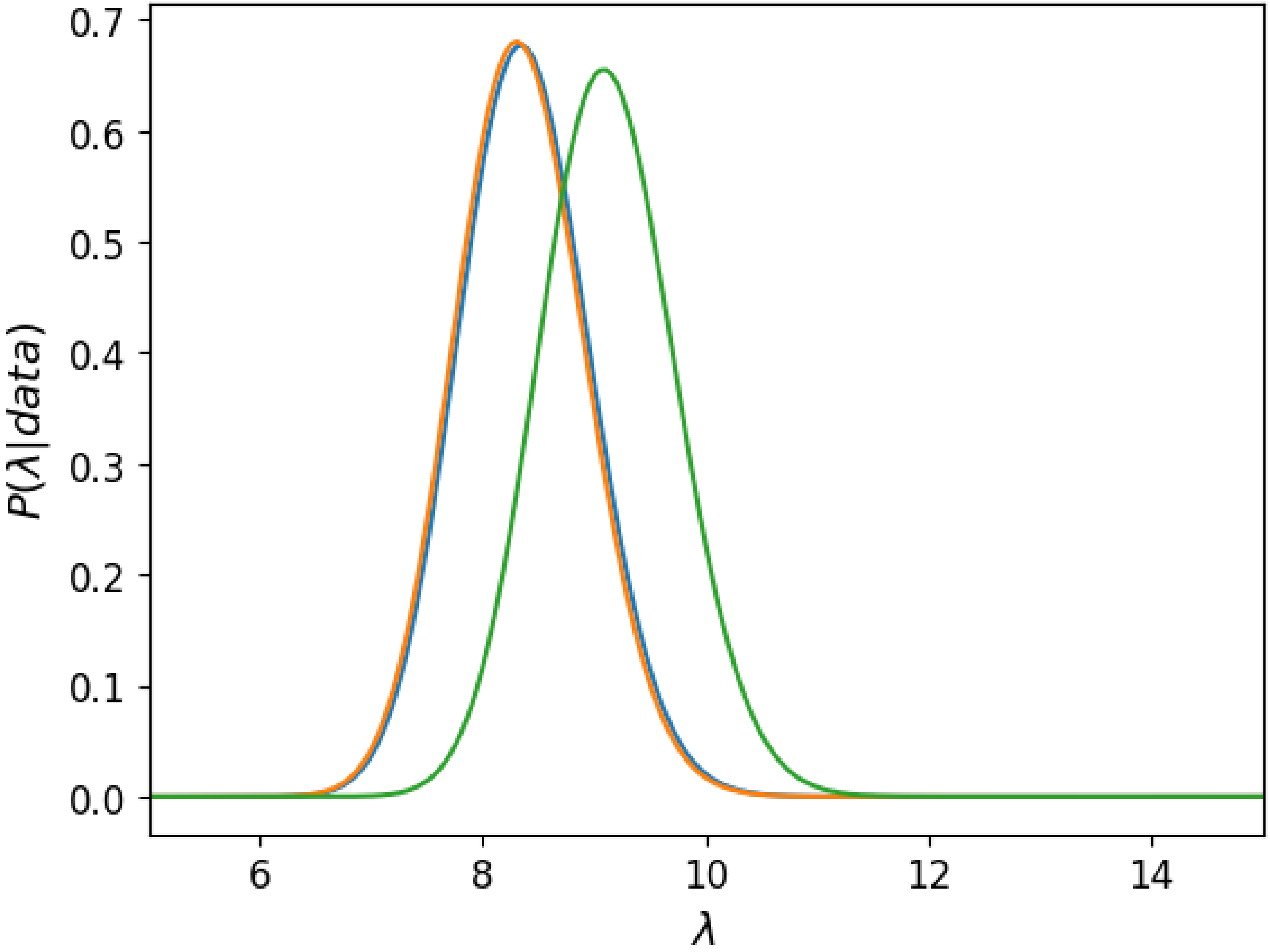
Posterior

- Simulated data ($\lambda = 8$)
- The weakly informative prior converged to the informative correct one quickly.
- Even the wrong prior gives a final similar result after only 24 data points.



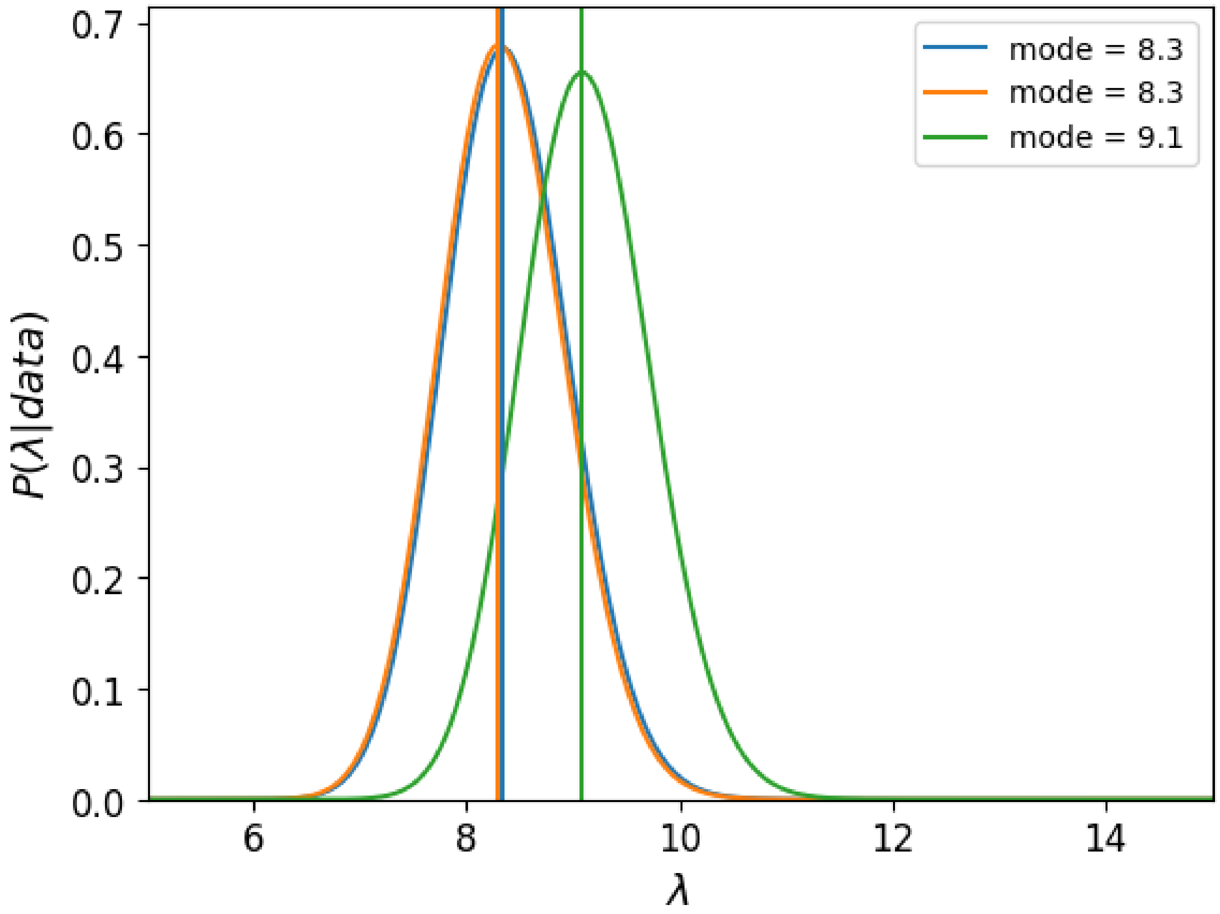
Posterior

- Final posterior



Posterior

- Final posterior
  - Let's compute the mode of each posterior



Models and Bayes

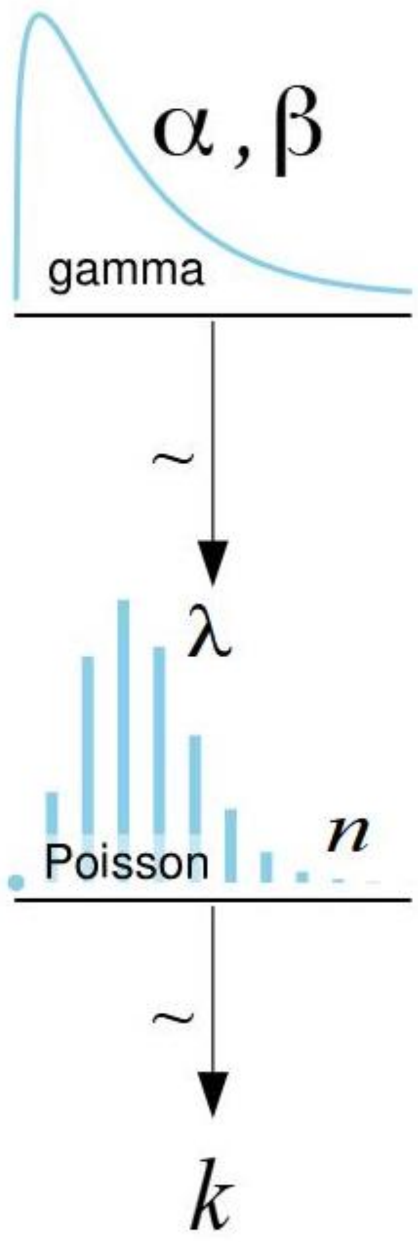
- A model has
  - Probability distributions
  - Parameters
- Probabilities are used to measure the uncertainty we have about parameters
- Bayes' theorem is a mechanism to correctly update those probabilities in light of new data

Bayesian Inference

- The result of a Bayesian analysis is a posterior distribution – not a single value but a distribution of plausible values given the data and our model.
- The most probable value is the mode of the posterior.
- The spread of the posterior reflects the uncertainty about the value of the parameter.
- The more data we have, the less the prior effects the posterior.
- Given a sufficiently large amount of data, two or more Bayesian models with different priors will tend to converge to the same result. In the limit of infinite data, no matter which prior we use, all of them will provide the same posterior.

Model Visualization

- We will represent our models with diagrams
  - Data at the bottom
    - k emails
  - Likelihood in the middle
    - This is our model of how the data was produced
    - n students with an average of λ emails
  - Prior at the top
    - This reflects our uncertainty about λ when the experiment began.



Model Visualization - downloads

1. Download LibreOffice:

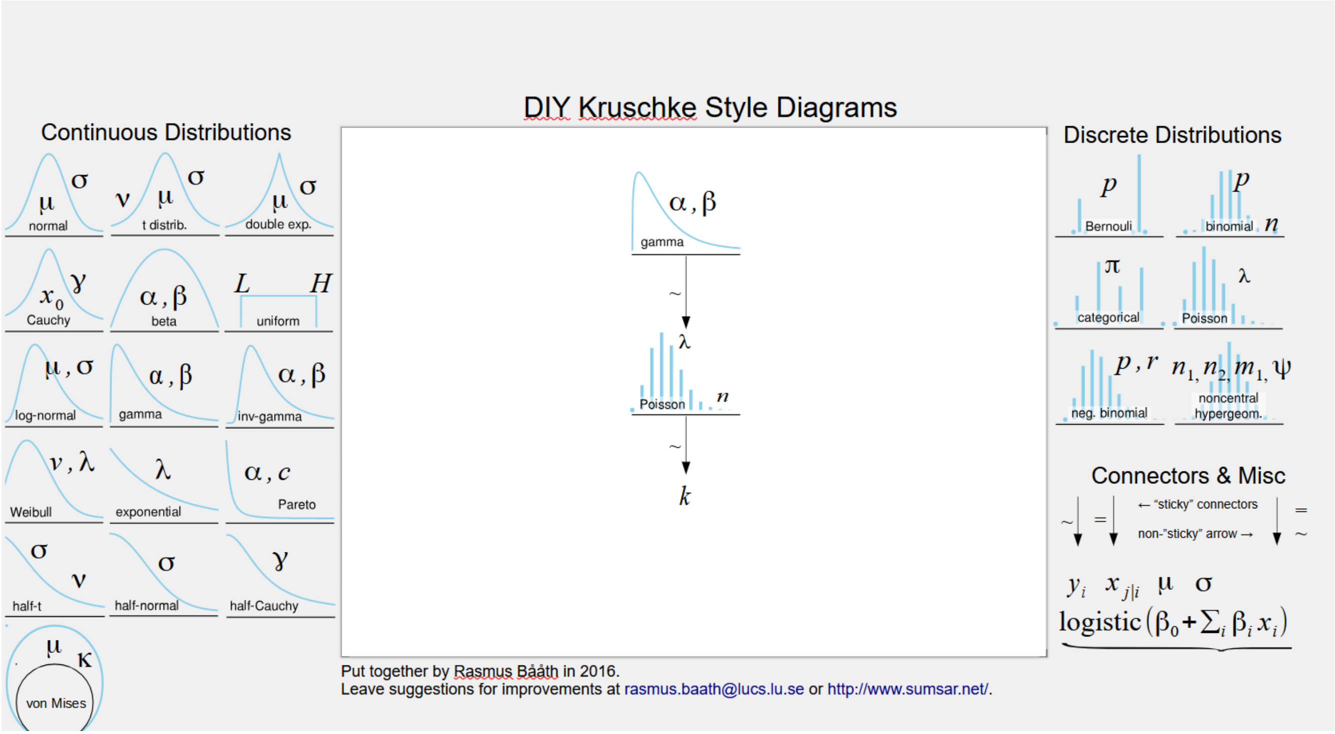
[Download LibreOffice | LibreOffice - Free and private office suite - Based on OpenOffice - Compatible with Microsoft]

2. Download diagrams template:

https://www.sumsar.net/blog/2013/10/diy-kruschke-style-diagrams/

Model Visualization - Explanations

1. Copy-paste and drag
2. Double click to edit
3. To save: export + selection -> as jpeg

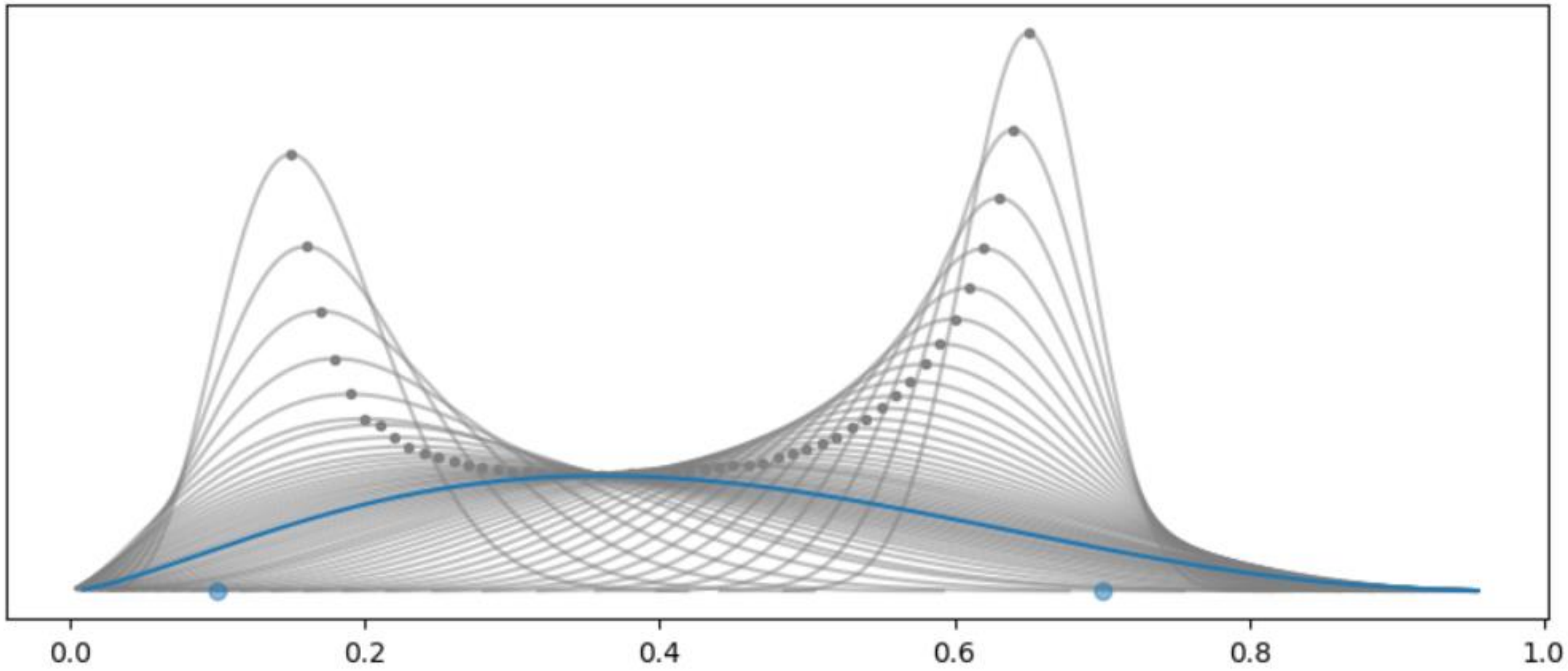


More About Priors

- We used the maxent function
  - There are a lot of priors that would give X% of the range in a given distribution.
  - Which should we choose?
  - Maximizing the entropy means choosing the probability distribution that is the most uncertain (least biased) while still satisfying given constraints.
  - This guarantees that we have the less informative distribution, given a set of constraints.

More About Priors

- Visualizes this for a beta prior with 90% of the mass between 0.1-0.7
  - The blue maximizes the entropy



More About Priors

- In our example, we used a gamma prior, where all we set was the range and the percentage of the distribution.

```
dist = pz.Gamma()
pz.maxent(dist, 0, 20, 0.9)
```

- We can also add constraints, such as the mean.

```
pz.maxent(pz.Gamma(mu=4), 1, 10, 0.9);
```

  - Here, the mean has to be 4.
  - If you pass a distribution with all the parameters specified, like pz.Gamma(mu=4, sigma=1), you will get an error saying “All parameters are fixed, at least one should be free”.

More About Priors

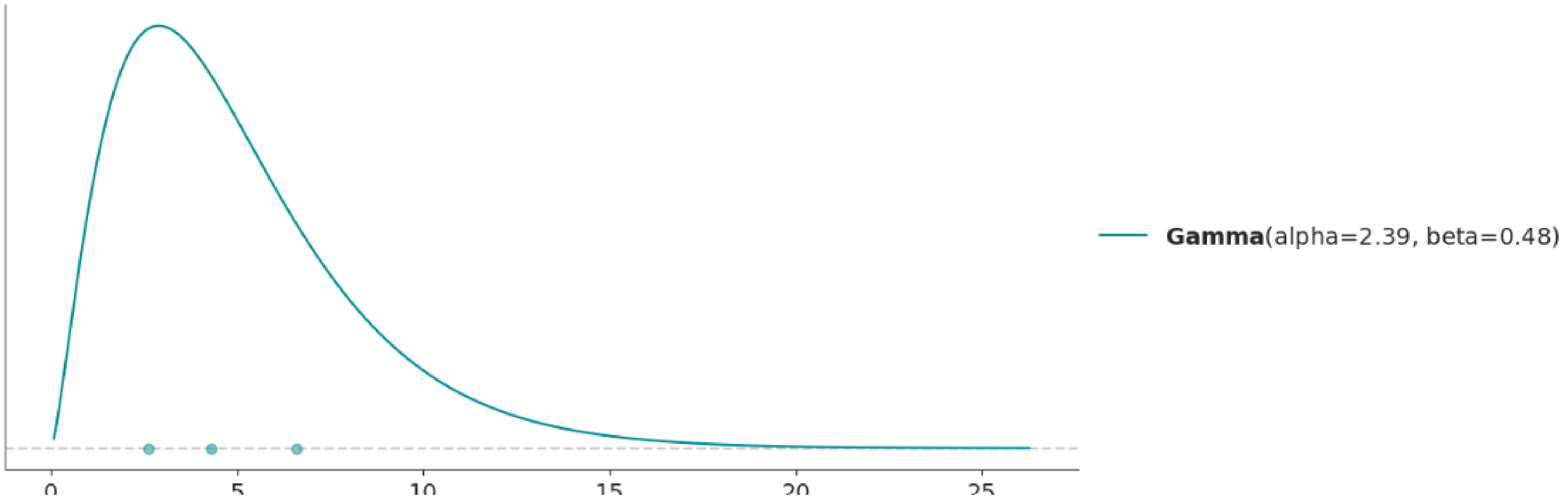
- We can also fix other properties, such as the mode.

```
dist = pz.Beta()
pz.maxent(dist, 0.1, 0.7, 0.94, fixed_stat=("mode", 0.3))
```

More About Priors

- We can also use methods other than maximizing the entropy.
  - We can define a distribution by its quartiles, that is by the 3 points which divides the distribution into 4 parts each with 25% of the total mass.

```
pz.quartile(pz.Gamma(), 2.6, 4.3, 6.6);
```



More About Priors

- A few more options can be found:

https://preliz.readthedocs.io/en/latest/examples/gallery/direct_elicitation_1D.html

- Reminder: we can also plot interactively using:
  - pz.Gamma(mu=2, sigma=1).plot_interactive()