POLITECNICO DI MILANO

Estimation of components failure rates from statistical data: Bayesian approach

Non-Information Priors

Masoud Naseri
masoud.Naseri@polimi.it

Lack of knowledge (Ignorance)

and

Non-Informative Priors

Introduction – Example 1

Consider the parameter estimation using Bayesian approach:

- p is the probability of failure on demand of a new type of pump used in an emergency system of an energy production plant.
- An expert is required to expert express his/her opinion on the failure probability of this pump (i.e., prior).
- Later, some tests have been conducted, and over 1000 demands, 2 failures have been observed (i.e., evidence).

The expert is expected to use Bayesian approach to update prior probability $P(p)$ given this new evidence.

However, the expert does not have considerable experience from past as he/she did not have the opportunity to observe similar equipment in the past.

How would you suggest the expert should express his/her failure prior failure probabilities, and then update it given the new evidence?

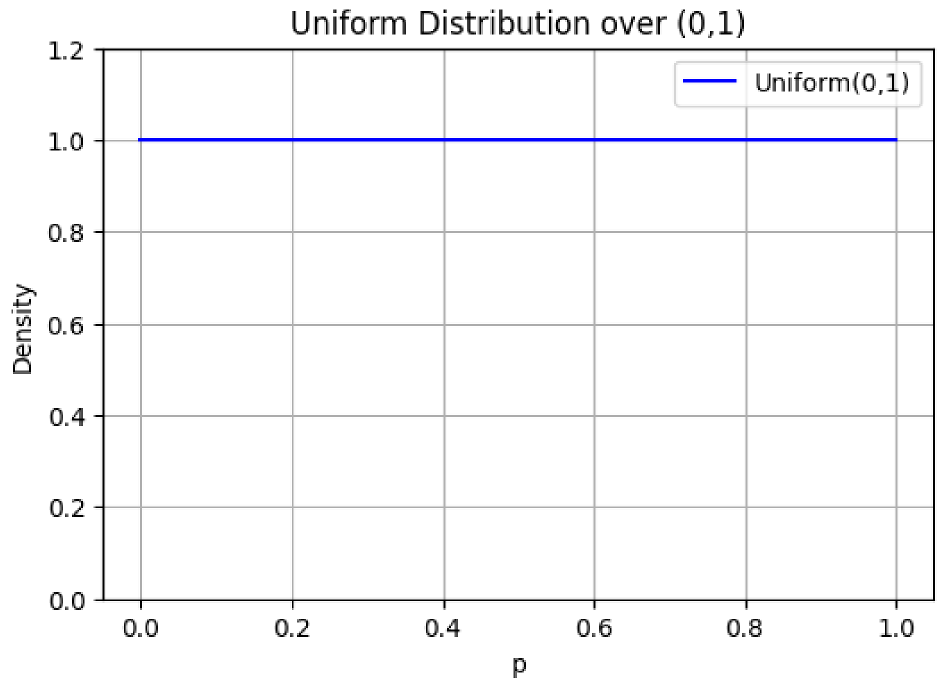
Introduction – Example 1

A suggested approach

Likelihood

Binomial distribution: $$P(x|n,p) = \binom{n}{x} p^x (1-p)^{n-x}$$

One may choose a **uniform distribution** as a **prior distribution** to express p, admitting the lack of knowledge:



Introduction – Example 1

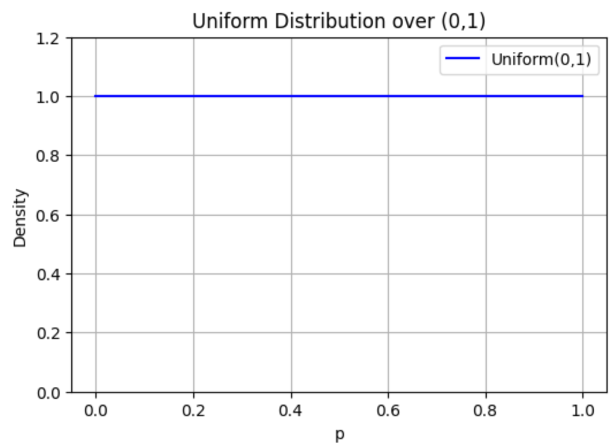
A suggested approach

Evidence: Over 1000 demands, 2 failures have been observed

Likelihood

Binomial distribution: $$P(x|n,p) = \binom{n}{x} p^x (1-p)^{n-x}$$

One may choose a **uniform distribution** as a **prior distribution** to express p, admitting the lack of knowledge:



$$f(p) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)} = \frac{p^0(1-p)^0}{B(1,1)} = 1$$

Uniform distribution $U(0,1)$ is equivalent to $\boldsymbol{Beta(\alpha = 1, \beta = 1)}$

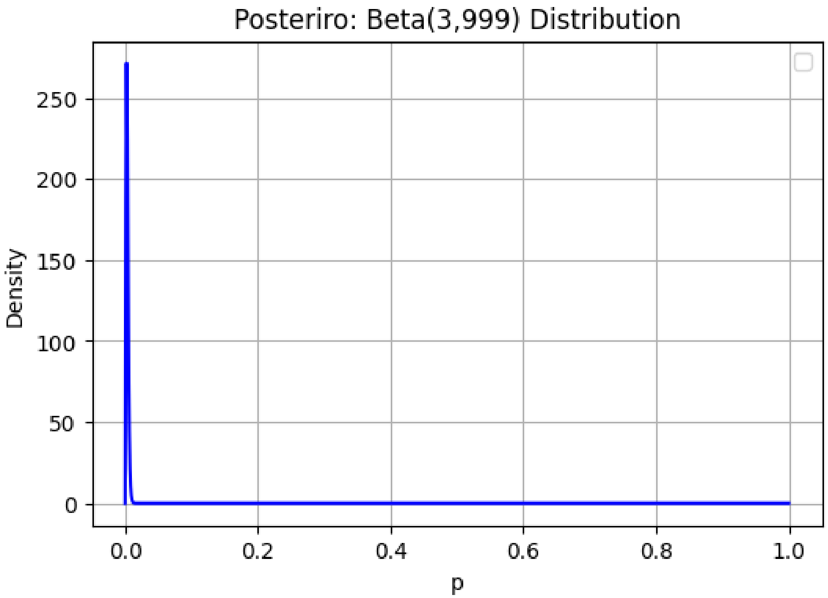
Prior: $\mathrm{Beta}(\alpha = 1, \beta = 1)$ → Posterior: $Beta(\alpha + x, \beta + n - x)$

$E = \{x = 2 | n = 1000\}$

$$Beta(3{,}999)$$

Introduction – Example 1

Posterior: $Beta(\alpha + x, \beta + n - x)$



$$E(p) = \frac{\alpha + x}{\alpha + \beta + n}$$

$$E(p) = 2.99 \cdot 10^{-3}$$

$$Var(p) = \frac{(\alpha + x)(\beta + n - x)}{(\alpha + \beta + n)^2(\alpha + \beta + n + 1)}$$

$$Var(p) = 2.98 \cdot 10^{-6}$$

$$\hat{p}_{MLE} = \frac{2}{1000}$$

Introduction – Example 1

$$E(p) = \frac{\alpha + x}{\alpha + \beta + n} \xrightarrow{n, x \to \infty} E(p) = \frac{x}{n} = \hat{p}_{MLE} = \frac{2}{1000}$$

- **Strong evidence**
- **Huge amount of observations**
- **The effect of prior vanishes out…**

$$E(p) = \frac{\alpha + x}{\alpha + \beta + n} \xrightarrow{\alpha, \beta \to 0} E(p) = \frac{x}{n} = \hat{p}_{MLE} = \frac{2}{1000}$$

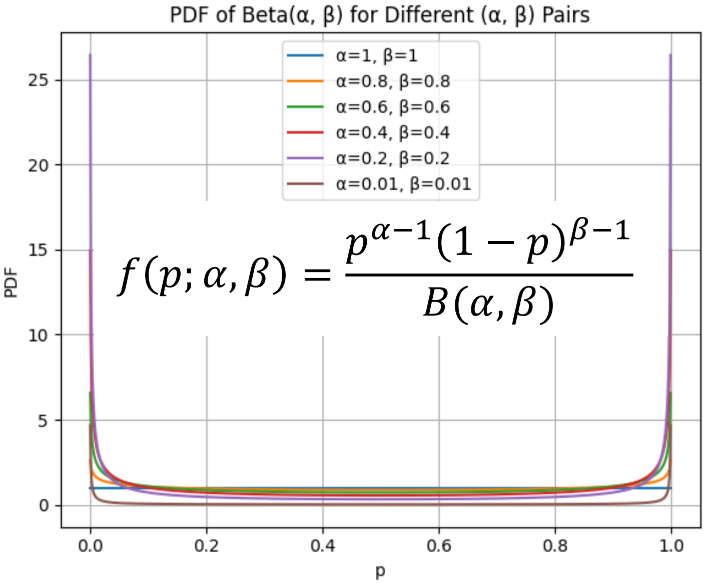
- **No information (i.e., Very weak knowledge)**
- **No strong evidence necessarily, but "prior itself contributes nothing" regardless of evidence strength**

Introduction – No information

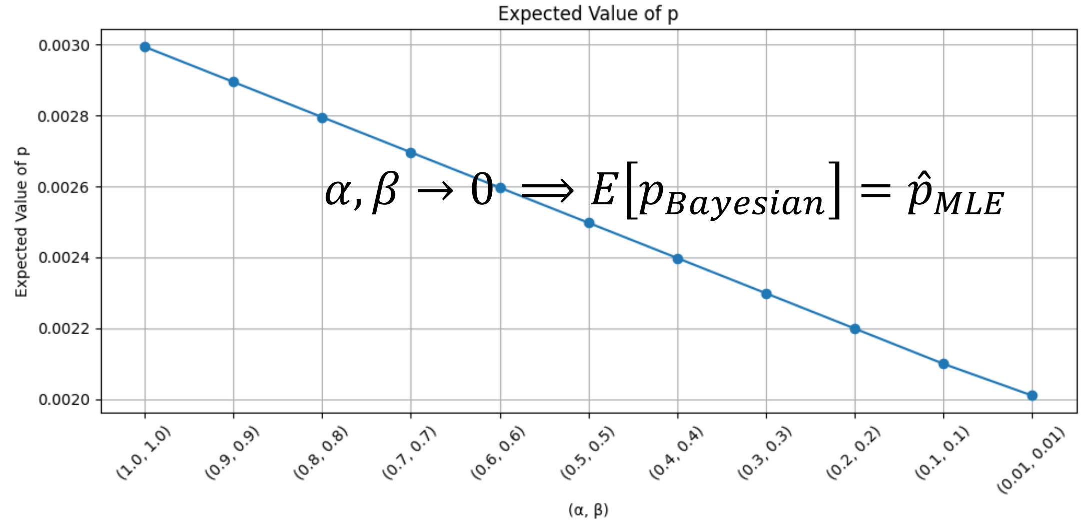
Posterior $\mathrm{Beta}(\alpha + x, \beta + n - x)$

$x = 2, n = 1000$

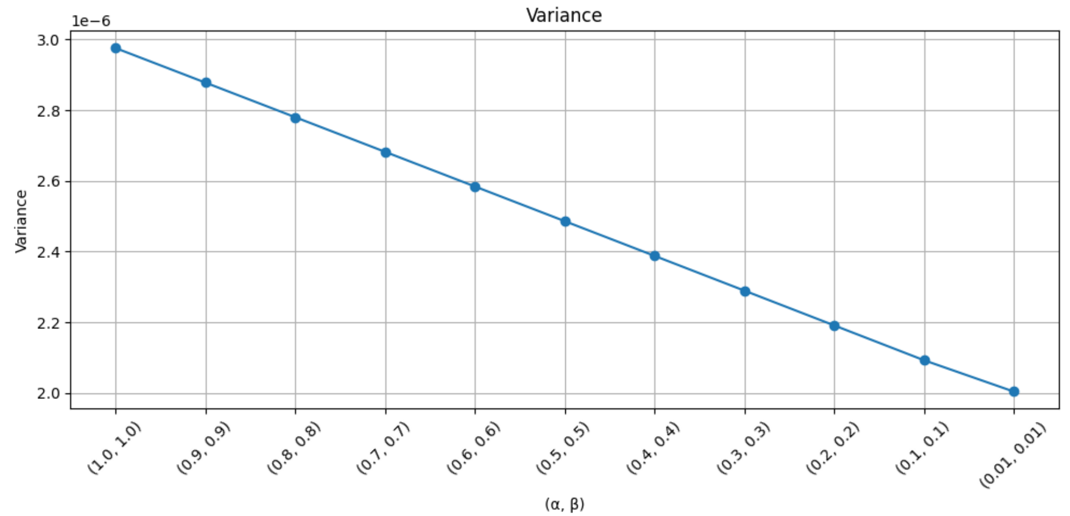
Behaviour of Prior $\mathrm{Beta}(\alpha, \beta)$



$$f(p; \alpha, \beta) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha, \beta)}$$



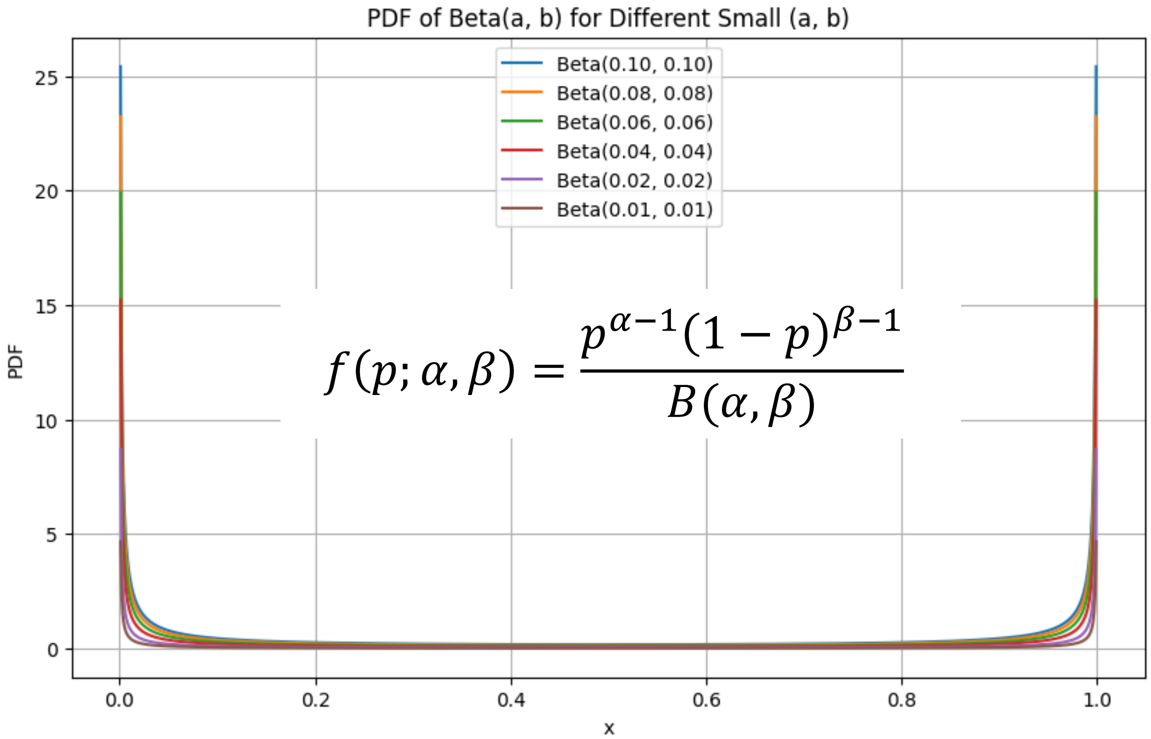
$$\alpha, \beta \to 0 \implies E\left[p_{Bayesian}\right] = \hat{p}_{MLE}$$



$\alpha, \beta \to 0$

No information (i.e., Very weak knowledge)

Introduction – Example 1



$$f(p;\alpha,\beta) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}$$

For $\alpha, \beta \to 0$ ⇨ $p^{\alpha-1}$ and $(1-p)^{\beta-1}$ approach extreme behaviours

Beta distribution tends to a **Dirac delta function** at both endpoints (0 and 1) when $\alpha, \beta \to 0$.

i.e., All probability is concentrated at the extremes of the interval [0, 1] (each with a value of ∞, and with no probability mass in the middle.

Introduction – Example 1

Prior **Beta(0,0)** (**Haldane prior**) is an **improper** prior as it does not integrate to 1.

Yet,

Haldane prior results in the posterior $\boldsymbol{Beta}(0 + \boldsymbol{x}, 0 + \boldsymbol{n} - \boldsymbol{x})$ which will be proper as long as $\boldsymbol{x} \neq \boldsymbol{0}$ and $\boldsymbol{n} - \boldsymbol{x} \neq \boldsymbol{0}$.

Introduction – Example 2

Recall - Consider

- The failure rate of the pump is constant, λ
- We have available the following three sources of information:
  - E1: engineering knowledge (description of the design and construction of the pump)
  - **E2: past performance of similar pumps in similar plants**

$$\boxed{\begin{aligned} \bar{\lambda}' &= 3\cdot 10^{-5}\,\mathrm{h}^{-1} \\ \sigma'_{\lambda} &= 7.4\cdot 10^{-5}\,\mathrm{h}^{-1} \end{aligned}}$$

  - E3: performance of the specific machine = 0 failures in t = 1000h

Questions:

1) Use E1 and E2 to build the prior of the failure rate distribution, $P(\lambda)$
2) Update the prior using the information in E3. Determine the point estimator of λ, its variance and its 95 percentile

Introduction – Example 2

Recall

- To choose the family of the prior $P(\lambda)$, we observe that E3 can be seen as the result of a **Poisson experiment: 0 occurrence in 1000*h*:**

$$P\left[k \text{ failures in } [0,t]\right] = \frac{(\lambda t)^k}{k!} e^{-\lambda t} \text{ Poisson process}$$

- The conjugate distribution of the Poisson Distribution is the Gamma distribution

$$\bar{\lambda}' = 3 \cdot 10^{-5}\,\text{h}^{-1}$$
$$\sigma_\lambda' = 7.4 \cdot 10^{-5}\,\text{h}^{-1}$$

Prior: $P(\lambda|E_1) = \Gamma(\alpha', \beta') = \dfrac{\beta'^{\alpha'} \lambda^{\alpha'-1}}{\Gamma(\alpha')} e^{-\beta' \lambda'}$

Expert knowledge

$$\bar{\lambda}' = \frac{\alpha'}{\beta'} \qquad \sigma_\lambda' = \frac{\sqrt{\alpha'}}{\beta'}$$

- Prior = Gamma with parameters (α', β') obtained as:

$$\frac{\alpha\prime}{\beta\prime} = \overline{\lambda\prime} = 3 \cdot 10^{-5}$$
$$\frac{\sqrt{\alpha\prime}}{\beta\prime} = \sigma_{\lambda\prime} = 7.4 \cdot 10^{-5}$$

$$\alpha' = \frac{\overline{\lambda'}^2}{\sigma^2_{\lambda'}} = 0.1644$$
$$\beta' = \frac{\overline{\lambda'}}{\sigma^2_{\lambda'}} = 5478$$

$$\alpha'' = \alpha' + k = 0.1644$$
$$\beta'' = \beta' + t = 6478$$
$$\bar{\lambda}'' = \frac{\alpha''}{\beta''} = 2.5 \cdot 10^{-5}\,\text{h}^{-1}$$

$$\overline{\lambda''} = \frac{\alpha}{\beta} = 2.5 \cdot 10^{-5}$$
$$\sigma_{\lambda''} = \frac{\sqrt{\alpha}}{\beta} = 6.2 \cdot 10^{-5}$$
$$\lambda''_{95} = 1.4 \cdot 10^{-4}$$

Introduction – Example 2

Consider

- The failure rate of the pump is constant, λ
- We have available the following three sources of information:
  - E1: engineering knowledge (description of the design and construction of the pump)
  - ~~E2: past performance of similar pumps in similar plants~~

$$\cancel{\boxed{\begin{aligned} \bar{\lambda}' &= 3\cdot 10^{-5}\,\mathrm{h}^{-1} \\ \sigma_{\lambda}' &= 7.4\cdot 10^{-5}\,\mathrm{h}^{-1} \end{aligned}}}$$

Suppose, there is no past experience i.e., No expert knowledge

How should one proceed with choosing the prior and updating it given the evidence?

We **cannot** choose a **uniform** distribution as a prior distribution to express $\lambda \in [0, \infty]$

Non-informative Prior

Examples 1 and 2 (where there were was **no past experience**), illustrate cases where there is **minimal prior knowledge** about parameters

If we do not have strong beliefs about what p should be, we may use a **non-informative prior** to

- let the data speak for itself
- avoid introducing subjective beliefs or biases into analyses

Various types of non-informative priors exist, such as **uniform**, **Jeffreys** and **reference** priors.

- Each has unique properties and applications in different statistical scenarios.

Non-informative Prior (Uniform): U(0,1)

Uniform U(0,1) - Laplace and the Principle of Insufficient Reason:

Principle of insufficient reason: If the parameter space is finite, use a uniform prior that assigns equal probability to each point in the parameter space

Example

Pump failure probability, p (Example 1)

Non-informative prior for p: $\boldsymbol{U(0,1)}$ which is equivalent to $\boldsymbol{Beta(1,1)}$

Prior: $Beta(\alpha = 1, \beta = 1)$

$E = \{x = 2|n = 1000\}$

Posterior: $Beta(\alpha + x, \beta + n - x)$

$Beta(3,999)$

Non-informative Prior (Uniform)

Some issues with the Laplace's **Principle of insufficient reason,** including

A) Variance under reparameterization flaw

Recall: Monotone transformation
If $x_1 < x_2$, then after applying a monotone function g, it will still hold that:
- If g is an increasing function: $g(x_1) < g(x_2)$
- If g is a decreasing function: $g(x_1) > g(x_2)$

Monotone transformation of p using Logit (Log-Odds) transformation:

$$\varphi = log\left(\frac{p}{1-p}\right)$$

Principle of insufficient reason

Ignorance or lack of information $\Rightarrow$ No preference between $p = p_1$ and $p = p_2$

$$f_p(p) = 1$$

Log-Odds monotone transformation of p: $\varphi = log\left(\frac{p}{1-p}\right)$

Principle of insufficient reason should also hold here $\Rightarrow$ i.e., there should be no preference between $\varphi = \varphi_1 = log\left(\frac{p_1}{1-p_1}\right)$ and $\varphi = \varphi_2 = log\left(\frac{p_2}{1-p_2}\right)$

$$\boxed{pdf(\varphi) = Uniform\ ?}$$

See Appendix

Using change of variables ($\varphi = log\left(\frac{p}{1-p}\right)$) applied to PDF: $f_\varphi(\varphi) = \frac{e^\varphi}{(1+e^\varphi)^2}$

Contradicting the Laplace's Principle of insufficient reason

Some issues with the Laplace's **Principle of insufficient reason,** including:

A) Variance under reparameterization

B) When the space parameter is not finite (e.g., pump's constant failure rate problem) $\lambda \in (0, \infty)$

Non-informative Prior (Jeffry's)

Jeffry argued that a **non-informative prior** should be **invariant to the parametrisation** (i.e., if $f(\theta)$ is non-informative, then any reparameterisation of the prior, should be also be non-informative)

If the likelihood is either Bernoulli or binomial distribution

Jeffry's prior

$$\xi^I(\theta) \propto \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}} \propto \mathbf{Beta}\left(\frac{1}{2},\frac{1}{2}\right)$$

Proper prior Beta distribution

$$\mathrm{Beta}\left(\frac{1}{2},\frac{1}{2}\right)$$

Proper Posterior: Consider the conjugacy of the Beta distribution relative to the Bernoulli and binomial likelihood.

If the likelihood is Poisson distribution

Jeffry's prior

$$\xi^I(\theta) \propto \theta^{-\frac{1}{2}} \propto \mathbf{Gamma}\left(\frac{1}{2},0\right)$$

Improper prior Gamma distribution

$$\mathrm{Gamma}\left(\frac{1}{2},0\right)$$

Proper Posterior: Consider the conjugacy of the Gamma distribution relative to the Poisson likelihood.

See Appendix for proofs

Example 1 using Jeffry's prior

Jeffry's prior as Non-informative Prior

Prior: $\text{Beta}(\alpha = \frac{1}{2}, \beta = \frac{1}{2})$

$$E = \{x = 2 | n = 1000\}$$

Posterior: $\text{Beta}(\alpha + x, \beta + n - x)$

$$\text{Beta}(2.5, 998.5)$$

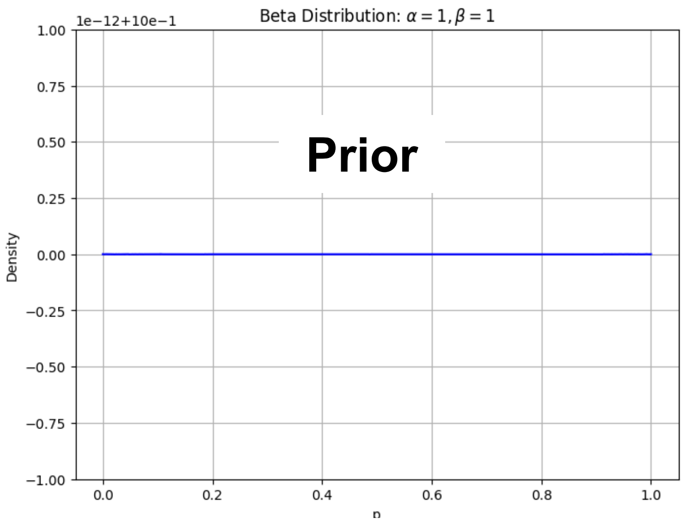
$E(p) = 2.50 \cdot 10^{-3}$ $\quad Var(p) = 2.49 \cdot 10^{-6}$

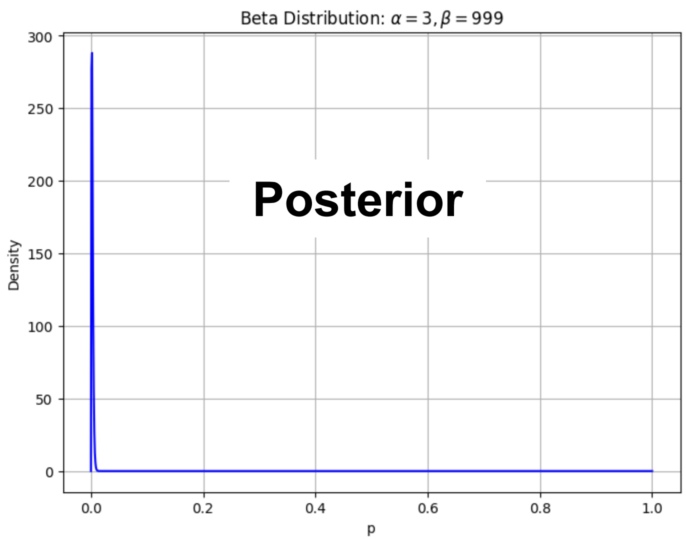
Versus Uniform prior:

$E(p) = 2.99 \cdot 10^{-3}$ $\quad Var(p) = 2.98 \cdot 10^{-6}$

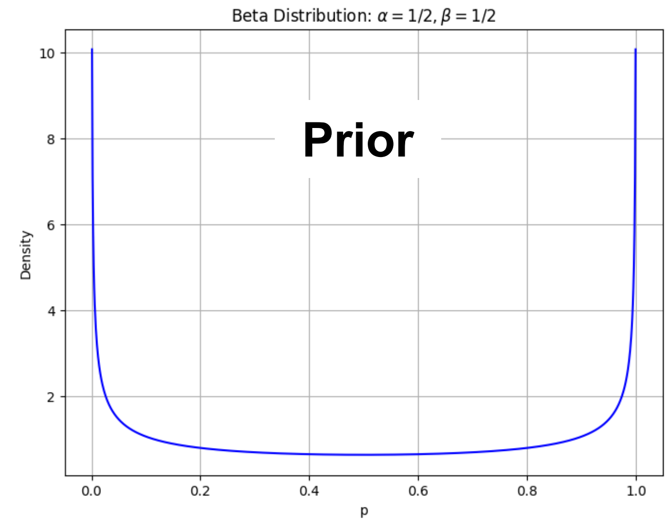
Comparing Uniform $U(0,1)$ prior with Jeffry's prior

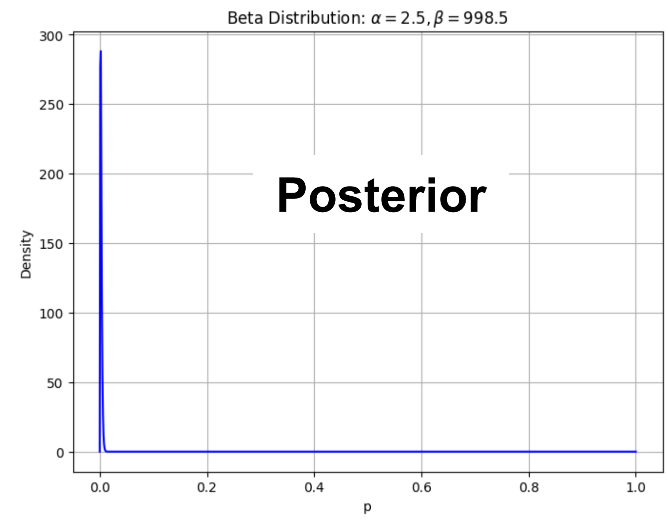
Uniform distribution as Non-informative Prior





Jeffry's prior as Non-informative Prior



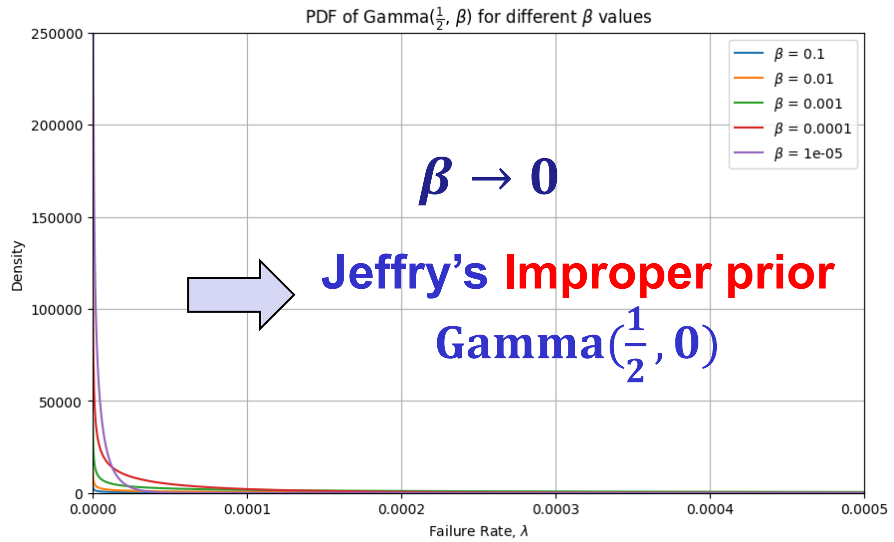


Example 2 using Jeffry's prior

If the likelihood is Poisson distribution

Jeffry's prior

$$\xi^{I}(\theta) \propto \theta^{-\frac{1}{2}} \propto \mathbf{Gamma}\left(\frac{1}{2}, 0\right)$$



Proper Posterior: Consider the conjugacy of the Gamma distribution relative to the Poisson likelihood.

E = {0 failures in t = 1000h}

Posterior: $\mathrm{Gamma}(\frac{1}{2} + k, 0 + t)$

$$\mathrm{Gamma}(\frac{1}{2}, 1000)$$

$$E(\lambda'') = \frac{\alpha}{\beta} = 5 \cdot 10^{-4}$$

$$\sigma_{\lambda''} = \frac{\sqrt{\alpha}}{\beta} = 7.07 \cdot 10^{-4}$$

Example 2 Some observations

The case, where **there were some past experience:**

$$\bar{\lambda}' = 3 \cdot 10^{-5} \text{h}^{-1}$$

$$\sigma'_{\lambda} = 7.4 \cdot 10^{-5} \text{h}^{-1}$$

Prior distribution

$$\alpha' = \frac{\overline{\lambda'}^2}{\sigma^2_{\lambda'}} = 0.1644$$

$$\beta' = \frac{\overline{\lambda'}}{\sigma^2_{\lambda'}} = 5478$$

Posterior distribution

$$\alpha'' = \alpha' + k = 0.1644$$

$$\beta'' = \beta' + t = 6478$$

$$\overline{\lambda''} = \frac{\alpha}{\beta} = 2.5 \cdot 10^{-5}$$

$$\sigma_{\lambda''} = \frac{\sqrt{\alpha}}{\beta} = 6.2 \cdot 10^{-5}$$

The case, where **there were no past experience**:

Using Jeffry's Prior: $\text{Gamma}(\frac{1}{2}, 0)$

Posterior: $\text{Gamma}(\frac{1}{2}, 1000)$

$$E(\lambda'') = \frac{\alpha}{\beta} = 5 \cdot 10^{-4}$$

$$\sigma_{\lambda''} = \frac{\sqrt{\alpha}}{\beta} = 7.07 \cdot 10^{-4}$$

Using Uniform Prior :$Beta(\alpha = 1, \beta = 1)$

Cannot be used over infinite space parameter

$$\lambda \in (0, \infty)$$

Appendix

Monotone transformation of p using Logit transformation

We want to find $f_\varphi(\varphi)$ given that $p \sim U(0,1)$, i.e., $f_p(p) = 1$ under the Log-Odds transformation $\varphi = log\left(\frac{p}{1-p}\right)$:

Probability mass must be conserved under transformation. That is, the probability of p falling in a small interval $[p, p + dp]$ must equal the probability of φ falling in the corresponding interval $[\varphi, \varphi + d\varphi]$

$$f_\varphi(\varphi)|d\varphi| = f_p(p) \cdot |dp| \rightarrow f_\varphi(\varphi) = f_p(p) \cdot \left|\frac{dp}{d\varphi}\right|$$

$$\varphi = log\left(\frac{p}{1-p}\right) \rightarrow e^\varphi = \frac{p}{1-p} \rightarrow p = \frac{e^\varphi}{1+e^\varphi}$$

$$\left|\frac{dp}{d\varphi}\right| = \frac{e^\varphi}{(1+e^\varphi)^2}$$

$$\rightarrow f_\varphi(\varphi) = f_p(p)\left(\frac{e^\varphi}{(1+e^\varphi)^2}\right) = 1 \cdot \left(\frac{e^\varphi}{(1+e^\varphi)^2}\right)$$

$f_\varphi(\varphi) = \frac{e^\varphi}{(1+e^\varphi)^2}$ is **not** uniform; that means there **is** a preference between different values of φ. Thus, **contradicting** the **Laplace's Principle of insufficient reason**

Non-informative Prior (Jeffry's)

Let $X = (X_1, \dots, X_n)$ (iid) and $X_i \sim g(x_i|\theta)$, then $f(x|\theta) = \prod_{i=1}^{n} g(x_i|\theta)$

$$I^F(\theta) = \sum_{i=1}^{n} \left[-E_{[X_i|\theta]} \frac{\partial^2}{\partial \theta^2} \log g(X_i|\theta) \right] = n I_1^F(\theta)$$

where $I_1^F(\theta)$ is the single observation Fisher information of $X_i \sim g(x_i|\theta)$ at θ

Jeffry's $I_1^F(\theta)$ is the **single observation Fisher information** of $X \sim f(x|\theta)$ is

$$\xi^I(\theta) = \text{const} \cdot \sqrt{I^F(\theta)}$$

If $\int_\Theta \sqrt{I^F(\theta)}\, d\theta$ is finite number,

$$\xi^I(\theta) = \frac{\sqrt{I^F(\theta)}}{\int_\Theta \sqrt{I^F(\theta)}}$$

If $\int_\Theta \sqrt{I^F(\theta)}\, d\theta$ is infinite, then $\xi^I(\theta)$ is an **improper prior** pdf of $\theta \in \Theta$.

An **improper prior** pdf is accepted as long as it can produce a proper posterior pdf.

Non-informative Prior (Jeffry's)

$X_i \sim \mathbf{Bern}(\boldsymbol{\theta})$

Likelihood for n observation:

$$f(x|\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{n\bar{x}}(1-\theta)^{n-n\bar{x}}$$

$$\frac{\partial^2}{\partial\theta^2}\log f(x|\theta) = \frac{\partial^2}{\partial\theta^2}[n\bar{x}\log\theta + (n-n\bar{x})\log(1-\theta)] = -\frac{n\bar{x}}{\theta^2} - \frac{n-n\bar{x}}{(1-\theta)^2}$$

$$I^F(\theta) = -E\left\{-\frac{n\bar{x}}{\theta^2} - \frac{n-n\bar{x}}{(1-\theta)^2}\middle|\,\theta\right\} = \frac{nE(\bar{X}|\theta)}{\theta^2} + \frac{n-nE(\bar{X}|\theta)}{(1-\theta)^2} = \frac{n/\theta}{\theta^2} + \frac{n-n/\theta}{(1-\theta)^2}$$

$$I^F(\theta) = \frac{n}{\theta} + \frac{n}{1-\theta} = \frac{n}{\theta(1-\theta)}$$

$$\boldsymbol{\xi^I(\theta)} \propto \sqrt{\frac{n}{\theta(1-\theta)}} \propto \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}} \propto \mathbf{Beta}\left(\frac{\mathbf{1}}{\mathbf{2}},\frac{\mathbf{1}}{\mathbf{2}}\right)$$

$X_i \sim \mathbf{Binomial}(\boldsymbol{x}|\boldsymbol{n},\boldsymbol{p})$ Similar results will be obtained

Non-informative Prior (Jeffry's)

$\boldsymbol{X_i \sim \mathbf{Possion}(\theta)}$

Likelihood

$$f(x|\theta) = \prod_{i=1}^{n} \frac{\theta^{x_i} e^{-\theta}}{x_i!} = \frac{\theta^{n\bar{x}} e^{-n\theta}}{\prod_{i=1}^{n} x_i!}$$

$$\frac{\partial^2}{\partial\theta^2} \log f(x|\theta) = \frac{\partial^2}{\partial\theta^2}\left[n\bar{x}\log\theta - n\theta - \sum_{i=1}^{n} \log x_i! \right] = -\frac{n\bar{x}}{\theta^2}$$

$$I^F(\theta) = -E\left\{ -\frac{n\bar{x}}{\theta^2} \middle| \theta \right\} = \frac{nE(\bar{X}|\theta)}{\theta^2} = \frac{n/\theta}{\theta^2} = \frac{n}{\theta}$$

$$\boldsymbol{\xi^I(\theta)} \propto \sqrt{\frac{n}{\theta}} \propto \theta^{-\frac{1}{2}} \propto \mathbf{Gamma}\left(\frac{\mathbf{1}}{\mathbf{2}}, \mathbf{0}\right)$$

Improper Gamma distribution $\mathrm{Gamma}\left(\frac{1}{2}, 0\right)$

Posterior: Consider the conjugacy of the Gamma distribution relative to the Poisson likelihood.