statstutor

community project

encouraging academics to share statistics support resources

Bayesian Statistics

In this summary sheet, let us assume that we have a model with a parameter θ, that we want to estimate. In Bayesian statistics or inference, we estimate a distribution (see resource "Probability Distribution Functions") for that parameter θ rather than just a single point estimate.

The first distribution we define for θ is called a ***prior distribution*** or ***prior*** and is represented by the probability density function $f(\theta)$. Then, we can apply Bayes' theorem to find the ***posterior distribution*** of θ: $f(\theta|x)$. This posterior distribution of θ is the distribution of θ, that is updated by the **information** $f(x|\theta)$ provided by the **likelihood of the model**. For that, Bayes' theorem is used:

$$f(\theta \mid x) = \frac{f(\theta, x)}{f(x)} = \frac{f(\theta)\, f(x \mid \theta)}{f(x)}$$

$$f(\theta \mid x) \quad \propto \quad f(\theta)\, f(x \mid \theta)$$

<u>What kind of prior distribution should be used?</u> The next page presents several ways of choosing a prior. In general, you will decide with your supervisor or based on previous papers what prior distribution you need to define for your parameter θ.

<u>What is a hyperparameter?</u> A ***hyperparameter*** (from ancient Greek ὑπέρ meaning "over" or "above") is a parameter of the prior distribution of the parameter we want to estimate. If, for instance, θ has a normal distribution with mean μ and standard deviation σ, then μ and σ will be considered as hyperparameters.

<u>What is conjugacy?</u> The property of ***conjugacy*** occurs when the prior distribution is of the same family as the posterior distribution. For instance, if the prior distribution of θ is a normal distribution and if we obtain a posterior distribution for θ that is also a normal distribution, then the property of conjugacy will be respected.

The table on the next page presents some priors where the posteriors obtained are of the same family as their respective priors. The hyperparameter(s) of the prior are updated thanks to the likelihood in order to give the hyperparameter(s) of the posterior distribution.

Priors

Choice

- Subjective Bayesianism: prior should incorporate as much detail as possible the research's a priori knowledge — via *prior elicitation.*
- Objective Bayesianism: prior should incorporate as little detail as possible (*non-informative* prior).
- Robust Bayesianism: consider various priors and determine *sensitivity* of our inferences to changes in the prior.

Types

- Flat: $f(\theta) \propto constant$
- Proper: $\int_{-\infty}^{\infty} f(\theta)\, d\theta = 1$
- Improper: $\int_{-\infty}^{\infty} f(\theta)\, d\theta = \infty$
- JEFFREYS' prior (transformation-invariant):

$$f(\theta) \propto \sqrt{I(\theta)} \qquad f(\theta) \propto \sqrt{\det(I(\theta))}$$

- Conjugate: $f(\theta)$ and $f(\theta \mid x^n)$ belong to the same parametric family

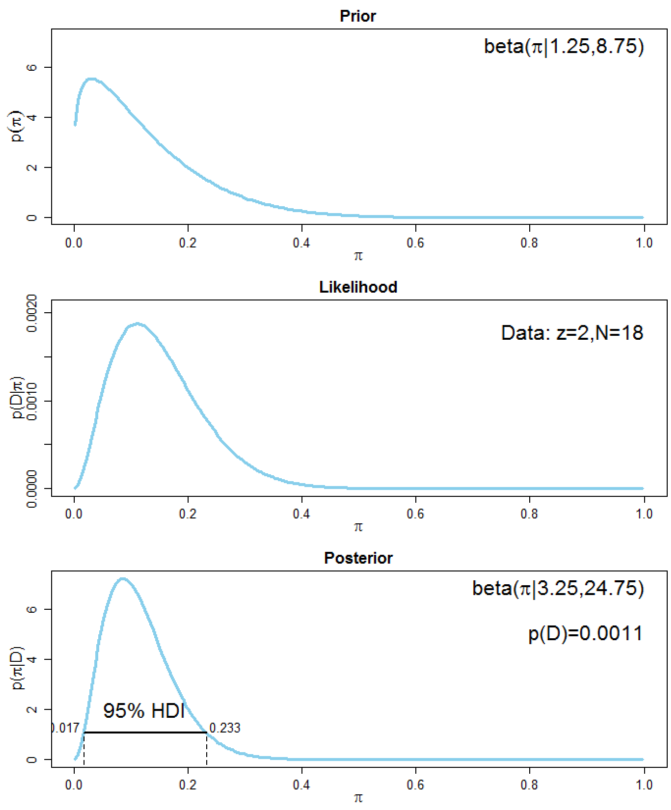
Conjugate Priors

Discrete likelihood		
Likelihood	Conjugate Prior	Posterior hyperparameters
Bernoulli(p)	Beta(α,β)	$\alpha + \sum_{i=1}^{n} x_i, \beta + n - \sum_{i=1}^{n} x_i$
Binomial(p)	Beta(α,β)	$\alpha + \sum_{i=1}^{n} x_i, \beta + \sum_{i=1}^{n} N_i - \sum_{i=1}^{n} x_i$
Negative Binomial(p)	Beta(α,β)	$\alpha + rn, \beta + \sum_{i=1}^{n} x_i$
Poisson(λ)	Gamma(α,β)	$\alpha + \sum_{i=1}^{n} x_i, \beta + n$
Multinomial($\mathbf{p}$)	Dirichlet(α)	$\alpha + \sum_{i=1}^{n} \mathbf{x}^{(i)}$
Geometric(p)	Beta(α,β)	$\alpha + n, \beta + \sum_{i=1}^{n} x_i$

Continuous likelihood (subscript c denotes constant)		
Likelihood	Conjugate Prior	Posterior hyperparameters
Uniform($0,\theta$)	Pareto(x_m, k)	$\max\{x_{(n)}, x_m\}, k + n$
Exponential(λ)	Gamma(α,β)	$\alpha + n, \beta + \sum_{i=1}^{n} x_i$
Normal(μ, σ_c^2)	Normal(μ_0, σ_0^2)	$\left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i=1}^n x_i}{\sigma_c^2}\right) / \left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma_c^2}\right)$, $\left(\frac{1}{\sigma_0^2} + \frac{n}{\sigma_c^2}\right)^{-1}$
Normal(μ_c, σ^2)	Scaled Inverse Chi-square(ν, σ_0^2)	$\nu + n, \frac{\nu\sigma_0^2 + \sum_{i=1}^n (x_i - \mu)^2}{\nu + n}$
Normal(μ, σ^2)	Normal-scaled Inverse Gamma($\lambda, \nu, \alpha, \beta$)	$\frac{\nu\lambda + n\bar{x}}{\nu + n}, \quad \nu + n, \quad \alpha + \frac{n}{2}, \beta + \frac{1}{2}\sum_{i=1}^{n}(x_i - \bar{x})^2 + \frac{\gamma(\bar{x} - \lambda)^2}{2(n+\gamma)}$
MVN(μ, Σ_c)	MVN(μ_0, Σ_0)	$\left(\Sigma_0^{-1} + n\Sigma_c^{-1}\right)^{-1}\left(\Sigma_0^{-1}\mu_0 + n\Sigma^{-1}\bar{x}\right)$, $\left(\Sigma_0^{-1} + n\Sigma_c^{-1}\right)^{-1}$
MVN(μ_c, Σ)	Inverse-Wishart(κ, Ψ)	$n + \kappa, \Psi + \sum_{i=1}^{n}(x_i - \mu_c)(x_i - \mu_c)^T$
Pareto(x_{m_c}, k)	Gamma(α,β)	$\alpha + n, \beta + \sum_{i=1}^{n} \log \frac{x_i}{x_{m_c}}$
Pareto(x_m, k_c)	Pareto(x_0, k_0)	$x_0, k_0 - kn$ where $k_0 > kn$
Gamma(α_c, β)	Gamma(α_0, β_0)	$\alpha_0 + n\alpha_c, \beta_0 + \sum_{i=1}^{n} x_i$

APPLICATION

In the 3 graphs on the right hand side, we can see the update from a prior distribution to its posterior distribution thanks to the Likelihood.



In this case, we have an unknown parameter π that follows a Beta(*) prior distribution (with parameters 1.25 and 8.75 chosen by us). The Likelihood of the model, using the data D, is proportional to a Binomial distribution. The posterior distribution of the unknown parameter π is also a Beta distribution due to the property of conjugacy (see Table of Conjugate Priors on Page 2 and Section on the Priors) and is shown on the third graph at the bottom of the figure on the right.

From the posterior distribution of π, we can pick up any kind of information we want: for instance the **mode**, the **mean** and the **median**. We can also capture **the 95% Highest Density Interval** (HDI) for π, often called **credible interval** that is, the interval where pi has 95% chance of being.

We can see that rather than obtaining just one value to estimate the parameter π**,** Bayesian statistics can offer a distribution to estimate it, giving thus more information about π (credible intervals, shape of the distribution, mode, mean etc.)

(*) For more detail, MASH resources provide a table of “Probability Distribution Functions” where the Beta and Binomial Distributions are defined.