Random Algorithms for Random Things

Benedetto Jacopo Buratti - Homework I

January 12, 2017

Abstract

This first homework covers the main concepts behind bayesian inference and some related issues. In the **first** and **second** exercises we focus on conjugate analysis, a classical approach to bayesian inference, used in the special occasion where the posterior distribution has the same distribution of the likelihood function. In the **third** and **fourth** exercise we introduce the Integral Transformation Formula, which will be the core, with unifrom pseudo-random numer generator of exercises **fifth** and **sixth** Finally in the **seventh**, **eighth** and **nineth** exerices we introduce and use the acceptance and rejection algorithm and apply in in a variety of situations.

Contents

Glossary **1**

Conjugate Analysis on Bernulli Sample **3**
- Parametric Space of Intrest . . . 3
- Ingredients for Bayesian Inference . . . 3
  - Likelihood Function . . . 3
  - Prior Distribution . . . 3
  - Posterior Distribution . . . 3

Conjugate Analysis on the Dugongs Sample **4**
- Likelihood . . . 4
- Maximum Likelihood Estimate . . . 4
- Vector of parameters of Interest . . . 4
- Maximum a Posteriori Estimates . . . 4

Integral Transform Method **5**
- Parametric Space of Intrest . . . 5

Beta-Bernoulli Conjugate analysis **6**
- Parametric Space of Intrest . . . 6

Simulating a Pareto Distribution **7**
- First Moment with Monte Carlo . . . 7
  - Empirical Distribution vs Desired Target Distribution . . . 8

Beta-Bernoulli Conjugate analysis **9**
- Parametric Space of Intrest . . . 9

Beta-Bernoulli Conjugate analysis **10**
- Parametric Space of Intrest . . . 10

Beta-Bernoulli Conjugate analysis **11**
- Parametric Space of Intrest . . . 11

Beta-Bernoulli Conjugate analysis **12**
- Parametric Space of Intrest . . . 12

Glossary

1. p for ”probability”, the cumulative distribution function (c. d. f.)
2. q for ”quantile”, the inverse c. d. f.

3. d for ”density”, the density function (p. f. or p. d. f.)
4. r for ”random”, a random variable having the specified distribution

Conjugate Analysis on Bernulli Sample

Parametric Space of Intrest

We define the Parametric Space of Intrest as:

Definition 1.1 A *parametric space* is the set of all possible combinations of values for all the different parameters contained in a particular mathematical model.

Given the previous definition the parametric space of interest, which is the probability of success of a single event, is $\Theta \in [0, 1]$

Ingredients for Bayesian Inference

Likelihood Function

The likelihood function for a series of Bernoulli experiment could be considered as a Binomial distribution, hence given the definition of likelihood:

Definition 1.2 The *likelihood function* of a set of parameter values, , given outcomes x, is equal to the probability of those observed outcomes given those parameter values. More specifically we define the likelihoodFor a iid sample in the following way:

$$\mathcal{L}_{x_1,\dots X_n}(\theta|x) = P(x|\theta) = \prod_{i=1}^{10} f(x_i|\theta) \tag{1}$$

Since our idd sample is generated by $X_i \sim Ber(p)$ which $f(x|\theta) = p$, therefore our 10 Bernoulli Trials' likelihood is:

$$\mathcal{L}_{\mathbf{x}}(\theta) = \prod_{i=1}^{10} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^3(1-\theta)^7 \tag{2}$$

since in our case we have had 3 successess and 7 failures.

Prior Distribution

Since we have no particular information about the experiment we should use a non-informative prior, which ever distribution we use in the modelling.

Another important point to take into account before choosing the prior is that our process is generated by $X_i \sim Ber(p)$.This means that we are able to use conjugate analysis in order to choose a suitable prior.

Given those two element we know that given a $X_i \sim Ber(p)$ likelihood function and $X_i \sim Beta(s_1, s_2)$ as a prior, our postetior distribution would be distributed as a prior. Moreover, taking into account the requirement of non-informativeness of the prior we should use a Beta(0.5, 0.5).

Posterior Distribution

The core of Bayesian Inference is knowledge update via Bayes Theorem

Theorem 1.1 The*Bayes Theorem* states posterior probability of an event is the consequence of two antecedents, a prior probability and a “likelihood function” derived from a statistical model for the observed data.

Thanks to this theorem we are able to update knowledge about the parameter space θ.

$$\pi(\theta|x) \propto \theta^3(1-\theta)^7 \cdot \theta^{-0.5}(1-\theta)^{-0.5} = \theta^{3.5}\theta^{7.5} \tag{3}$$

which is, as expected, $\pi(\theta|\mathbf{x}) \sim Beta(3.5, 7.5)$

\subsection(Bayesian Inference)

Now that we have a computed our posterior distribution we are able to perform bayesian inference using Beta(3.5, 7.5). Just to demosntrare the possibility we compute the following

Conjugate Analysis on the Dugongs Sample

Likelihood

First of all we to load the Dugongs into R:

```
## $x
##  [1]  1.0  1.5  1.5  1.5  2.5  4.0  5.0  5.0  7.0  8.0  8.5  9.0  9.5  9.5
## [15] 10.0 12.0 12.0 13.0 13.0 14.5 15.5 15.5 16.5 17.0 22.5 29.0 31.5
##
## $Y
##  [1] 1.80 1.85 1.87 1.77 2.02 2.27 2.15 2.26 2.47 2.19 2.26 2.40 2.39 2.41
## [15] 2.50 2.32 2.32 2.43 2.47 2.56 2.65 2.47 2.64 2.56 2.70 2.72 2.57
##
## $N
## [1] 27
```

As reported on dataset website $Y_i \sim N(m_i, t)$, with:

1. $i = 1, ..., 27$
2. $m_i = a - bg^{X_i}$
3. $a, b > 0$
4. $0 < g < 1$

Hence, according to definition 1.1 the likelihood for the given dataset is:

$$\mathcal{L}_{\mathbf{x}}(\theta) = \prod_{i=1}^{27} \theta^{x_i} \frac{1}{\sigma\sqrt{2\pi}} exp\frac{1}{2}(\frac{x_i - \alpha - s\beta\gamma^{x_i}}{\tau}) \tag{4}$$

Maximum Likelihood Estimate

Now that we have the maximum likelihood estimate, we are able to compute the MLE using opt

Vector of parameters of Interest

Maximum a Posteriori Estimates

Integral Transform Method

Parametric Space of Intrest

Definition 1.1 A *parametric space* is the set of all possible combinations of values for all the different parameters contained in a particular mathematical model

Given the previous definition the parametric space of interest, which is the probability of success of a single event, is $\Theta \in [0, 1]$

Beta-Bernoulli Conjugate analysis

Parametric Space of Intrest

Definition 1.1 A *parametric space* is the set of all possible combinations of values for all the different parameters contained in a particular mathematical model

Given the previous definition the parametric space of interest, which is the probability of success of a single event, is $\Theta \in [0, 1]$

Simulating a Pareto Distribution

```r
library("VGAM")
```

```
## Loading required package: stats4
## Loading required package: splines
```

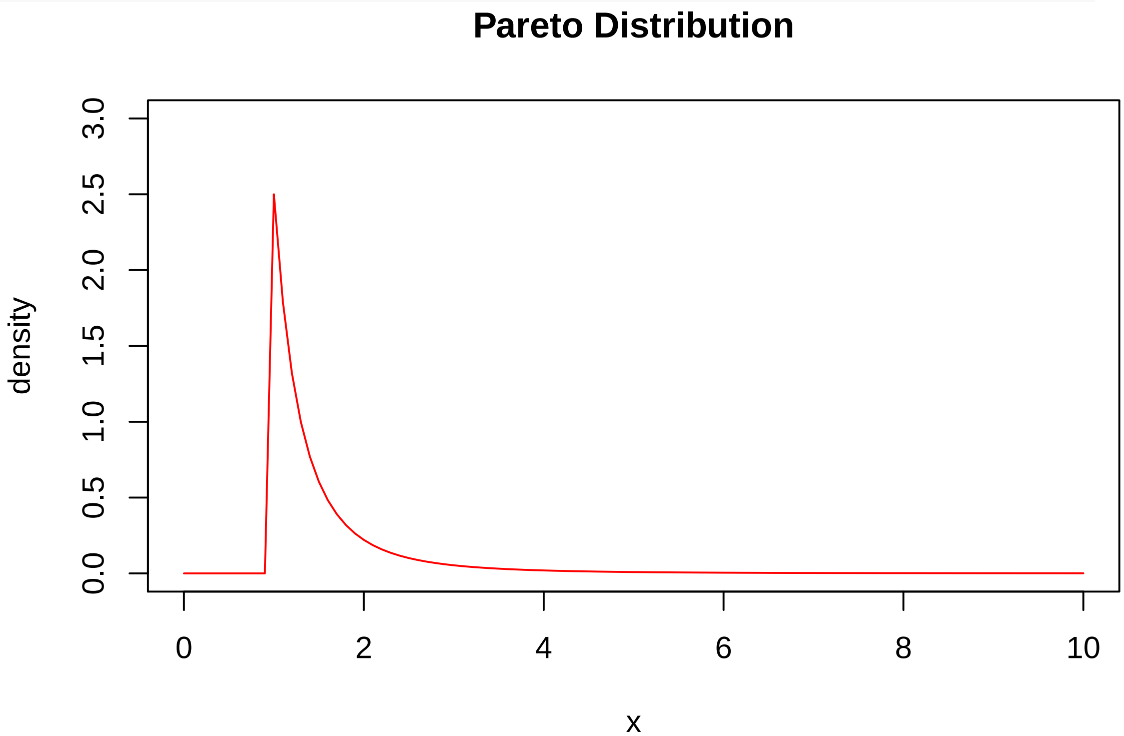
First Moment with Monte Carlo

Given a a random variable distributed $X \sim Par(2.5, 1)$ which density is:

$$F_X(x) = \begin{cases} \frac{5}{2}x^{-\frac{7}{2}} & \text{if } x \geq 1 \\ 0 & \text{if } x < 1 \end{cases} \tag{5}$$

Which graph is:

```r
x <- seq(from=1, to=100, by=0.01)
curve(dpareto(x, shape=2.5, scale=1), xlim = c(0,10), ylim=c(0,3), type="l", main="Pareto Distribution",
     ylab="density", col="red")
```



Since this is a well known distribution, of which first moment is easily computable in closed form, is intresting to check how closely a Monte Carlo simulation is able to estimate the exact parameter.

In order to do this we have two options:

1. Generate a random sample from R-Cran Built in Pareto distribution
2. Use the Inverse Transform Method, using a Uniform Distribution as instrumental function

However, in advance to any techinque we need to know, exploiting the LLN, how many samples we need in order to estimate with a confidence of $10^{-}2$ the mean.

The first option is pretty straightfoward since we can use R built in function:

```r
first <- rpareto(85000, shape=2.5, scale=1)
I_hat = mean(first)
I_hat
```

```
## [1] 1.662475
```

The second one consist in the Integral Transform Method which using the inverse function of the Pareto and a Uniform RNG allows us to sample from a pareto:

Since $Par(2.5, 2)$ first moment is given by, we can see how the two approaches compares:

Empirical Distribution vs Desired Target Distribution

Now that can also check two different EDF and see that the while qualitatively there is no much difference between the two approaches, from a efficiency point of view there two order of magnitude slower.

Beta-Bernoulli Conjugate analysis

Parametric Space of Intrest

Definition 1.1 A *parametric space* is the set of all possible combinations of values for all the different parameters contained in a particular mathematical model

Given the previous definition the parametric space of interest, which is the probability of success of a single event, is $\Theta \in [0, 1]$

Beta-Bernoulli Conjugate analysis

Parametric Space of Intrest

Definition 1.1 A *parametric space* is the set of all possible combinations of values for all the different parameters contained in a particular mathematical model

Given the previous definition the parametric space of interest, which is the probability of success of a single event, is $\Theta \in [0, 1]$

Beta-Bernoulli Conjugate analysis

Parametric Space of Intrest

Definition 1.1 A *parametric space* is the set of all possible combinations of values for all the different parameters contained in a particular mathematical model

Given the previous definition the parametric space of interest, which is the probability of success of a single event, is $\Theta \in [0, 1]$

Beta-Bernoulli Conjugate analysis

Parametric Space of Intrest

Definition 1.1 A *parametric space* is the set of all possible combinations of values for all the different parameters contained in a particular mathematical model

Given the previous definition the parametric space of interest, which is the probability of success of a single event, is $\Theta \in [0, 1]$