Université de Liège | Academic year 2024–2025

MATH2022 – Monte Carlo Methods in Statistics
Examination project

Instructions:

- Please hand in the following **two documents**:
  - A pdf file containing your answers (including all necessary plots), calculations and a copy of your code (in appendix), written as a report.
  - An annotated `R`-file with the executable code (preference goes to `R`, but `Python` is also acceptable).

  Write all code in a single file, clearly indicating which part of the code goes with each of the questions. The answers and calculations can be either handwritten (make sure your handwriting is legible) or typeset (preferable) using, for example, Overleaf or LaTeX.
- Send both files to `arnout.vanmessem@uliege.be` at the latest on **Wednesday, June 4**.
- By June 16, you will receive your score after which you can decide for an oral continuation of the exam. This oral continuation is optional and will only take place if you specifically request for it; it will contain one theory question and one question related to your project. It can influence your final score with up to 2 marks (both positive or negative). A suitable moment will be sought together.
- The project makes use of the beta distribution. Should you suddenly have an urge to solve the project based on another, not common, distribution, feel free to contact me for its suitability.

The beta distribution is used to, for example, model the probability of probabilities. This continuous distribution is defined on the interval $[0,1]$ (one of the very few distributions that is defined on a finite interval, next to the uniform distribution) and is characterized by 2 parameters α and β, which control the shape of the distribution. The pdf of a random variable $X \sim Be(\alpha,\beta)$ is given by:

$$f(x|\alpha,\beta) = \frac{1}{\mathrm{B}(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}\,, \quad 0 < x < 1\,, \alpha > 0\,, \beta > 0\,,$$

where the normalizing constant $\mathrm{B}(\alpha,\beta) := \dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ corresponds to the beta function. The gamma function is defined as $\Gamma(\alpha) := \int_0^\infty t^{\alpha-1}e^{-t}dt$.

1. Starting from the general form $\eta x^{\alpha-1}(1-x)^{\beta-1}$, derive the normalizing constant for the beta distribution.
2. Plot the distribution $Be(\alpha,\beta)$ for different values of its shape parameters, covering all different possibilities. Describe the shape and skewness of the distribution in function of α and β. What happens if both shape parameters converge to $+\infty$? Can we recover the uniform distribution $\mathcal{U}[0,1]$ using the beta distribution?

Let us now take a look at generating random values for the $Be(\alpha,\beta)$ distribution.

3. Can we apply the inverse transform method to sample random values of a beta distribution? If yes, explain and implement the algorithm. Simulate $15,000$ values of a $Be(3,9)$ distribution. If no, explain why not.

4. In Chapter 3, a formula is given to simulate $Be(\alpha, \beta)$ from the uniform distribution, given that both parameters α and β are natural numbers (and not equal to 0). Implement this transformation method algorithm and apply it to simulate 15,000 values of a $Be(3, 9)$ distribution.

5. It is also possible to generate a beta distribution using gamma distributions. A gamma distribution $Ga(\alpha, \beta)$ is defined by the density function $f(x|\alpha, \beta) := \dfrac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \mathbb{I}_{[0,+\infty[}(x)$ for $\alpha, \beta > 0$.

    (a) Show formally that if $Y_1 \sim Ga(\alpha, 1)$ and $Y_2 \sim Ga(\beta, 1)$ then $X = \frac{Y_1}{Y_1+Y_2} \sim Be(\alpha, \beta)$.

    (b) Use this relation to construct an algorithm to generate a beta random variable and simulate 15,000 values from a $Be(3, 9)$.

6. Compare the implemented procedures from questions 4 and 5 and the beta distribution generator incorporated in `R` in terms of their runtime (keep $n = 15,000$) as well as their precision (use different values of n).

Since the transformation method given in question 4 is only valid for natural values of the shape parameters, we will often rely on an Accept-Reject algorithm to sample values from a more general beta distribution.

7. In the case where α and β are larger than 1 (but not necessarily natural numbers), it is possible to use the uniform distribution $\mathcal{U}[0, 1]$ as instrumental distribution.

    (a) Explain why this is possible and how to determine the limiting constant C.

    (b) Determine (using `R`) this limiting constant C used in the Accept-Reject algorithm if the target distribution is $Be(3.3, 9.5)$.

    (c) Implement the algorithm and perform 15,000 simulations to obtain values from a $Be(3.3, 9.5)$ distribution.

    (d) Create a plot that shows all simulated values, distinguishing between the accepted (in green) and rejected values (in red). Also indicate the instrumental density and the true density function of the $Be(3.3, 9.5)$ distribution. Calculate the acceptance rate.

    (e) Show, in a general setting, that the probability of acceptance in an Accept-Reject algorithm with limiting constant C on the density ratio f/g is equal to $1/C$. Compare this result with the calculated value in the previous question.

    (f) Rewrite the algorithm to obtain $n = 15,000$ simulated (accepted) values from the $Be(3.3, 9.5)$ distribution. Does this influence the acceptance rate obtained by the Accept-Reject algorithm? Plot the simulated values and compare with the true distribution. Discuss your findings.

8. Another option for the instrumental distribution in the Accept-Reject algorithm is to use another beta distribution. Since we saw before how to simulate random values from a beta distribution for which the shape parameters take on natural values, we will use this kind of beta distribution.

    (a) Show formally that, for the ratio f/g to be bounded when f is a $Be(\alpha, \beta)$ density and g is a $Be(a, b)$ density, we must have both $a \leq \alpha$ and $b \leq \beta$. Determine the maximal ratio in terms of α, β, a, and b.

    We can therefore choose $a = \lfloor \alpha \rfloor$ and $b = \lfloor \beta \rfloor$.

    (b) Redo questions 7(b) to (d) with the appropriate beta distribution as instrumental distribution. Compare both methods.

In Bayesian theory, we assume that a variable X follows a certain distribution $f(X|\theta)$ in which θ represent the (unknown) parameters of the distribution. As such, these parameters are assumed to have their own distribution. Prior knowledge, before gathering the data x, is specified in the prior distribution with density $\pi(\theta)$. The information contained in the data x is then combined with this prior knowledge on the parameters to obtain the posterior distribution $\pi(\theta|x)$ using Bayes' theorem. If the distribution of the posterior is the same as that of the prior, the prior is called a *conjugate prior*.

9. Show that, for X an observation from the negative binomial distribution $NB(r, p)$, the family of beta distributions $Be(\alpha, \beta)$ is a family of conjugate priors.

Next, let us apply Monte Carlo for the estimation of the shape parameters of a beta distribution.

10. In Chapter 2, we estimated the shape parameters of the beta distribution using the maximum likelihood estimator. The problem, however, has no explicit solution. Another option is to use the Method of Moments (MOM) to determine $\hat{\alpha}$ and $\hat{\beta}$. As the name implies, the moments up to order k will be used to estimate the k parameters of a distribution.
    (a) Show that for the gamma function, the following property holds: $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$, $\alpha > 0$. As such, the gamma function can be seen as a generalization of factorials.
    (b) Determine the moment of order t of a $Be(\alpha, \beta)$ distribution.
    (c) Use this formula to obtain the expected value $\mathbb{E}[X]$ and variance $\mathrm{Var}(X)$ for $X \sim Be(\alpha, \beta)$.
    (d) Use the Monte Carlo approach to estimate the shape parameters of a $Be(9, 3)$ distribution (set $n = 15,000$).

11. Finally, let us compare both estimation methods (MLE and MOM). To this end, set the number of simulations equal to $Nsim = 1500$.
    (a) Use the previous implementation for the MOM and use the function `ebeta` contained in the package `EnvStats` for the MLE estimator for the beta distribution to create histograms illustrating the distribution of each of the shape parameters
    (b) Evaluate the efficiency of each method in terms of bias and variance. Compare the results.
    (c) Illustrate the asymptotic normality of the obtained MLE estimators.