Summary of TMA4300 - Spring 2023

Summary of TMA4300

What did we do?

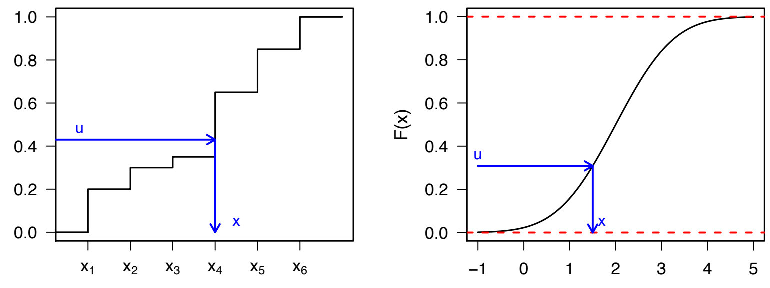
Summary of TMA4300

What did we do?

We had three blocks:

- Simulation
- Markov chain Monte Carlo and INLA
- Bootstrap and EM-algorithm

Block 1



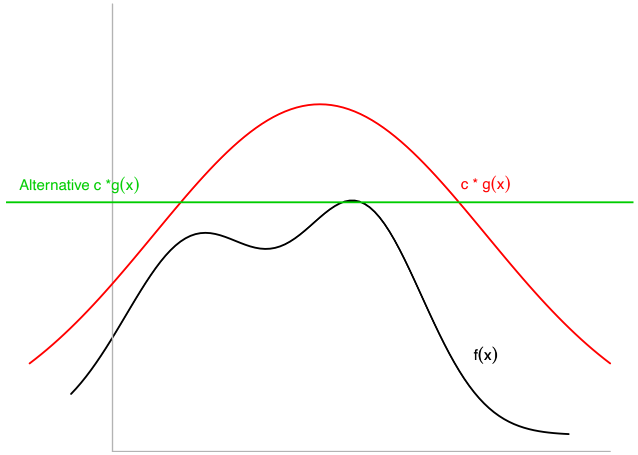
One important theorem to remeber:

If $X \sim F(x)$ what is the distribution of $F(X)$?

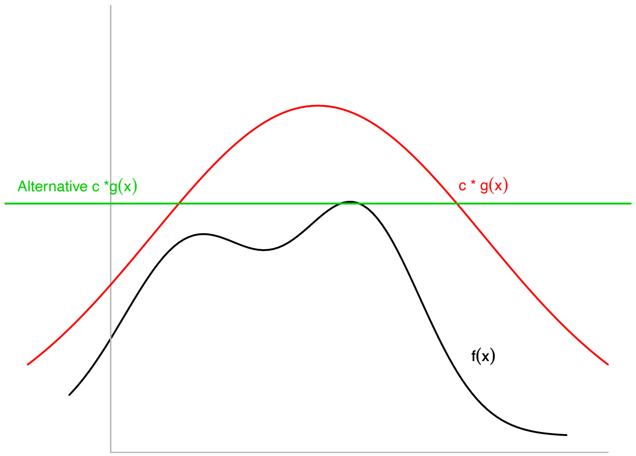
What else . . . ?

- Variable transformation
  - Univariate (e.g. location and scale transformation)
  - Bivariate (e.g. the Box-Muller algorithm)
  - Multivariate (e.g. multivariate Gaussian)
- Ratio-of-uniforms method
- Methods based on mixtures

Rejection Sampling: Do you remember this figure?

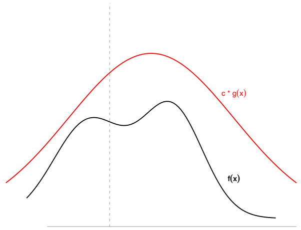


Rejection Sampling: Do you remember this figure?



Refinements: Make the envelope adaptive (different approaches)

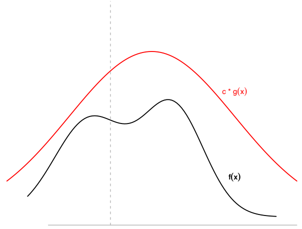
Rejection Sampling: When do we accept?



Let $x \sim g(x)$ and $u \sim \text{Unif}(0, 1)$ then

$$\begin{cases} u \leq \frac{1}{c}\frac{f(x)}{g(x)} & \text{keep} \\ u > \frac{1}{c}\frac{f(x)}{g(x)} & \text{reject} \end{cases}$$

Rejection Sampling: When do we accept?

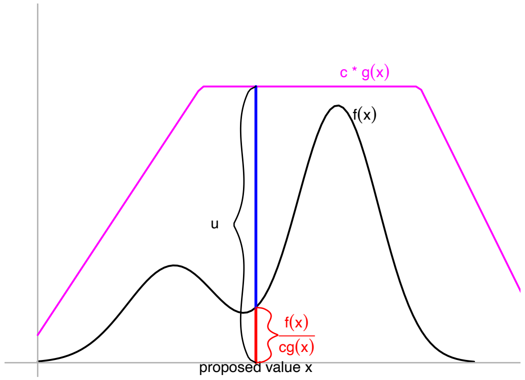


Let $x \sim g(x)$ and $u \sim \text{Unif}(0,1)$ then

$$\begin{cases} u \leq \frac{1}{c}\frac{f(x)}{g(x)} & \text{keep} \\ u > \frac{1}{c}\frac{f(x)}{g(x)} & \text{reject} \end{cases}$$

- What is the overall acceptance probability?
- Do we need to know the normalizing constant of $f(\cdot)$?

Rejection sampling



Why do we want samples?

Often we would like to approximate a statistic that is difficult to compute directly.

Keywords:

- Monte Carlo integration
- Importance sampling
  - What is it?
  - When is it particularly usefull?

Monte Carlo approximation

We are interested in

$$\mu = E(h(x)) = \int h(x)f(x)dx; \qquad x \sim f(x)$$

if we can sample $x_i \sim f(x)$, $i = 1, \ldots, n$ (iid) then

$$\hat{\mu} = \sum_{i=1}^{n} h(x_i)$$

- Do you know how to compute the mean and variance of this estimator?

Importance sampling

We are interested in

$$\mu = E_f(h(x)) = \int h(x)f(x)dx$$

- If possible compute it analytically!
- If we can sample from $f(x)$ we can use Monte Carlo integration
- Possible alternative: Importance sampling
  - sample from auxiliary distribution $g(x)$ and re-weight
  - can be used as variance-reduction technique

Importance sampling

Let $x_1, \dots, x_n \sim g(x)$, and let $w(x_i) = \frac{f(x_i)}{g(x_i)}$, $i = 1, \dots, n$ then

$$\hat{\mu}_{IS} = \frac{\sum h(x_i)w(x_i)}{n}$$

- Unbiased
- Consistent
- Need to know the normalizing constant

$$\tilde{\mu}_{IS} = \frac{\sum h(x_i)w(x_i)}{\sum w(x_i)}$$

- Biased for finite n
- Consistent
- Self-normalizing

Bayesian inference

Basics:

- Posterior $\propto$ Likelihood $\times$ Prior
- Prior Choice
- Bayesian hierarchical models:
- Full-conditional distributions

Bayesian inference

Basics:

- Posterior $\propto$ Likelihood $\times$ Prior
- Prior Choice
- Bayesian hierarchical models:
  - Observation $\pi(\mathbf{y}|\mathbf{x})$
  - Latent process $\pi(\mathbf{x}|\theta)$
  - Hyperpriors $\pi(\theta)$
- Full-conditional distributions

Block 2: Two big topics

- Markov chain Monte Carlo
- Integrated Nested Laplace Approximations (INLA)

Markov chain Monte Carlo:

- What is the idea? What kind of Markov chain do we create?

Markov chain Monte Carlo:

- What is the idea? What kind of Markov chain do we create?
  - Construct a Markov chain that converges to the distribution of interest.

Markov chain Monte Carlo:

- What is the idea? What kind of Markov chain do we create?
- Why do we not use an approach from block 1?
- What kind of different MCMC techniques have we seen?

Markov chain Monte Carlo:

- What is the idea? What kind of Markov chain do we create?
- Why do we not use an approach from block 1?
- What kind of different MCMC techniques have we seen?
- Is the algorithm working at all?

Elements of a MCMC algorithm

- Target distribution $\pi(x)$: Given by the problem
- Proposal distribution $Q(y|x)$: Chosen by the user
- Acceptance probability $\alpha(y|x)$: Derived in order to fullfill the detailed balance condition

Some keywords

detailed balance condition, Metropolis-within-Gibbs, random-walk proposal, burn-in, convergence diagnostics, mixing, effective sample size, ...

Integrated nested Laplace approximations

- What is the idea?
- For which models does it work?
- What are the main “ingredients”
- Potential advantages over MCMC

Integrated nested Laplace approximations

- What is the idea?
  - Deterministic approximation instead of simulations
  - Focus on marginal posterior instead of joint posterior
- For which models does it work?
- What are the main "ingredients"
- Potential advantages over MCMC

Integrated nested Laplace approximations

- What is the idea?
- For which models does it work?
  - Latent Gaussian Models (with Markov dependence structure)
- What are the main "ingredients"
- Potential advantages over MCMC

Integrated nested Laplace approximations

- What is the idea?
- For which models does it work?
- What are the main "ingredients"
  - Conditional probability

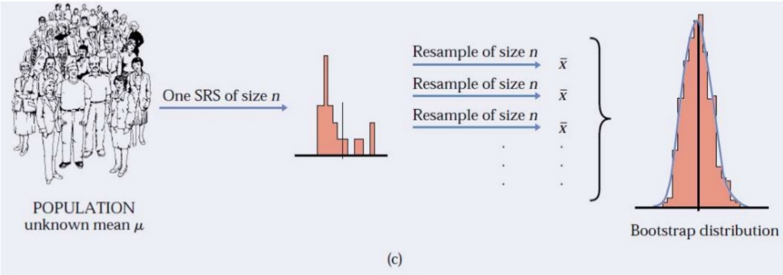
$$\pi(x|z) = \frac{\pi(x,z)}{\pi(z)} \Rightarrow \pi(z) = \frac{\pi(x,z)}{\pi(x|z)}$$

  - Laplace Approximation
    - 2nd order Taylor approximation of non-Gaussian responses
- Potential advantages over MCMC

Block 3

- Bootstrap
- EM algorithm

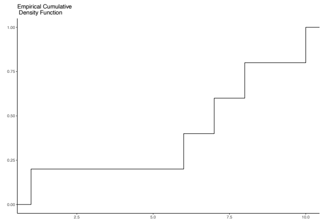
Bootstrap



Bootstrap - Some keywords

- Empirical distribution function

  Data: 1, 7, 8, 10, 6
  Empirical distribution function:



Bootstrap - Some keywords

- Empirical distribution function
- Plug-in principle
- Bootstrap sample
  - Non-parametric
  - Parametric
- Bootstrapping more complex data structures - Regression
  - Bootstrap residuals
  - Paired-Bootstrap - Bootstrap time series
  - Model based bootstrap
  - Block-bootstrap

EM-algorithm

- Goal? Basic idea? What are the steps?
- Field of applications: mixture models, censored data, missing data, hidden models, . . .

The exam - June 8, 2023

- Digital School Exam

The exam - June 8, 2023

- Digital School Exam
- To pass you need to have all 3 approved projects and pass the exam (at least 40% correct)

The exam - June 8, 2023

- Digital School Exam
- To pass you need to have all 3 approved projects and pass the exam (at least 40% correct)
- 10 subproblems worth 10 points each

The exam - June 8, 2023

- Digital School Exam
- To pass you need to have all 3 approved projects and pass the exam (at least 40% correct)
- 10 subproblems worth 10 points each
- 21 minutes per subproblem + 30 minutes extra

The exam - June 8, 2023

- Digital School Exam
- To pass you need to have all 3 approved projects and pass the exam (at least 40% correct)
- 10 subproblems worth 10 points each
- 21 minutes per subproblem + 30 minutes extra
- You must bring your laptop and charger

The exam - June 8, 2023

- Digital School Exam
- To pass you need to have all 3 approved projects and pass the exam (at least 40% correct)
- 10 subproblems worth 10 points each
- 21 minutes per subproblem + 30 minutes extra
- You must bring your laptop and charger
- You must have safe exam browser installed on your laptop

The exam - June 8, 2023

- Digital School Exam
- To pass you need to have all 3 approved projects and pass the exam (at least 40% correct)
- 10 subproblems worth 10 points each
- 21 minutes per subproblem + 30 minutes extra
- You must bring your laptop and charger
- You must have safe exam browser installed on your laptop
- Pseudocoding may be involved

The exam - June 8, 2023

- Digital School Exam
- To pass you need to have all 3 approved projects and pass the exam (at least 40% correct)
- 10 subproblems worth 10 points each
- 21 minutes per subproblem + 30 minutes extra
- You must bring your laptop and charger
- You must have safe exam browser installed on your laptop
- Pseudocoding may be involved
- Justify your answers, and show all your work

The exam - June 8, 2023

- Digital School Exam
- To pass you need to have all 3 approved projects and pass the exam (at least 40% correct)
- 10 subproblems worth 10 points each
- 21 minutes per subproblem + 30 minutes extra
- You must bring your laptop and charger
- You must have safe exam browser installed on your laptop
- Pseudocoding may be involved
- Justify your answers, and show all your work
- https://i.ntnu.no/wiki/-/wiki/English/Digital+school+exam+-+for+students

Permitted exam support materials: C

- Specific basic calculator with empty memory

Permitted exam support materials: C

- Specific basic calculator with empty memory
- Tabeller og formler i statistikk, Tapir forlag

Permitted exam support materials: C

- Specific basic calculator with empty memory
- Tabeller og formler i statistikk, Tapir forlag
- K. Rottman: Matematisk forelsamling

Permitted exam support materials: C

- Specific basic calculator with empty memory
- Tabeller og formler i statistikk, Tapir forlag
- K. Rottman: Matematisk forelsamling
- A bilingual dictionary

Permitted exam support materials: C

- Specific basic calculator with empty memory
- Tabeller og formler i statistikk, Tapir forlag
- K. Rottman: Matematisk forelsamling
- A bilingual dictionary
- 1 yellow stamped A5 sheet with handwritten notes and equations (both sides, 7th floor of Sentralbygg 2)