MT4531/5731: (Advanced) Bayesian Inference

Introduction to Bayesian Computation (Monte Carlo sampling)

Nicolò Margaritella

School of Mathematics and Statistics, University of St Andrews

–

University of St Andrews

Outline

Outline

Bayesian inference - univariate case

- Data $\boldsymbol{x} = (x_1, \ldots, x_n)'$ with likelihood $f(\boldsymbol{x}|\theta)$.
- Prior $p(\theta)$ for the single parameter θ.
- Posterior

$$\pi(\theta|\boldsymbol{x}) = \frac{f(\boldsymbol{x}|\theta)p(\theta)}{f(\boldsymbol{x})}$$

where $f(\boldsymbol{x}) = \int f(\boldsymbol{x}|\theta)p(\theta)d\theta$

Bayesian inference - multivariate case (1)

- Data $\boldsymbol{x} = (x_1, \ldots, x_n)'$ with likelihood $f(\boldsymbol{x}|\boldsymbol{\theta})$.
- Prior $p(\boldsymbol{\theta})$, now for a vector $\boldsymbol{\theta}$ of parameters.
- Posterior

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{f(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{f(\boldsymbol{x})}$$

where $f(\boldsymbol{x}) = \int \ldots \int f(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$

Bayesian inference - multivariate case (2)

- Often the parameters $\boldsymbol{\theta} = \{\theta_1, \theta_2, \theta_3, \dots\}$ are assumed to be independent of each other, *a priori*, so that,

$$p(\boldsymbol{\theta}) = \prod_{i=1}^{n} p(\theta_i).$$

- It is also common to introduce prior dependence between the parameters, by introducing another hyper-parameter ϕ, so that,

$$p(\boldsymbol{\theta}|\phi) = \prod_{i=1}^{n} p(\theta_i|\phi), \qquad p(\phi) = \dots.$$

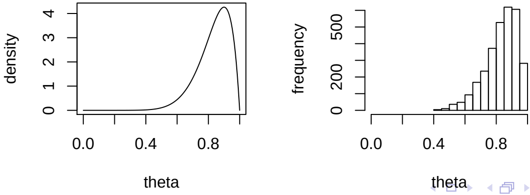
- In more than one dimensions the posterior distribution significantly increases in complexity.

Intractable posteriors

- Data: $x_1, \ldots, x_n$ with $x_i \sim N(\mu, \sigma^2)$. μ and σ^2 unknown.
- Priors: $\mu \sim N(\phi, \tau^2)$ and $\sigma^2 \sim \Gamma^{-1}(\alpha, \beta)$
- $$\pi(\mu, \sigma^2 | \boldsymbol{x}) \propto (\sigma^2)^{-(n/2+\alpha+1)} \exp\left(-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}\right) \exp\left(-\frac{\beta}{\sigma^2}\right) \exp\left(-\frac{(\mu-\phi)^2}{2\tau^2}\right)$$
- $$f(\boldsymbol{x}) = \int_{\mu=-\infty}^{\infty} \int_{\sigma^2=0}^{\infty} (\sigma^2)^{-(n/2+\alpha+1)} \exp\left(-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}\right) \exp\left(-\frac{\beta}{\sigma^2}\right) \exp\left(-\frac{(\mu-\phi)^2}{2\tau^2}\right) d\mu d\sigma^2$$

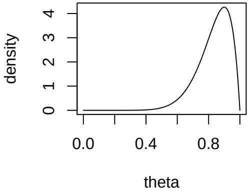
Bayesian computation

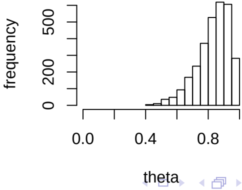
- Q: How do we do Bayesian analysis when we have posterior distributions of non-standard form, possibly of high dimension?
- A: Use a computer to generate random samples from the posterior.
- Use these samples to provide empirical estimates of the quantities we want



Bayesian computation

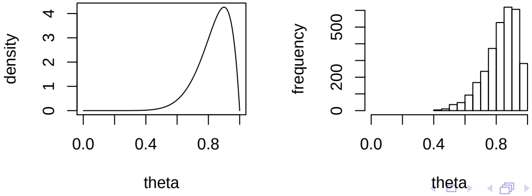
- Q: How do we do Bayesian analysis when we have posterior distributions of non-standard form, possibly of high dimension?
- A: Use a computer to generate random samples from the posterior.
- Use these samples to provide empirical estimates of the quantities we want





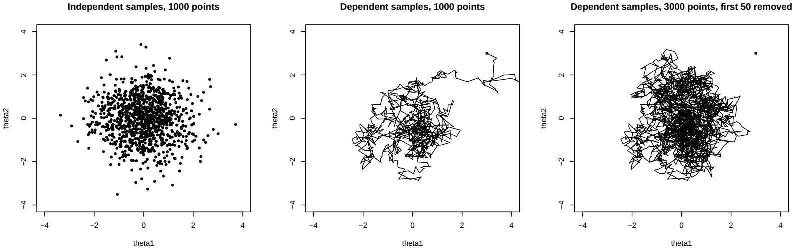
Bayesian computation

- Q: How do we do Bayesian analysis when we have posterior distributions of non-standard form, possibly of high dimension?
- A: Use a computer to generate random samples from the posterior.
- Use these samples to provide empirical estimates of the quantities we want



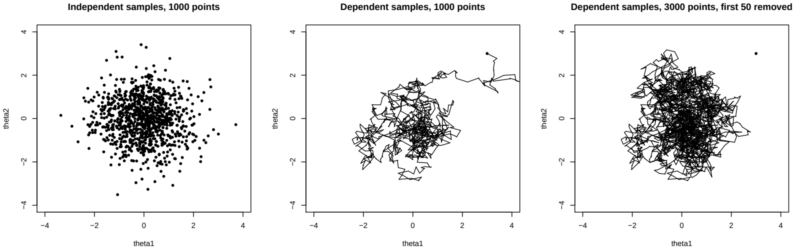
Samples can be independent or dependent

- Two ways to generate the samples: independent or dependent (i.e., value of one sample depends on the value of the previous one)
- Both give the same distribution in the end, but with dependent samples:
  - Start point matters for a while – i.e., it takes a while to "burn-in"
  - You need more samples for the same level of accuracy



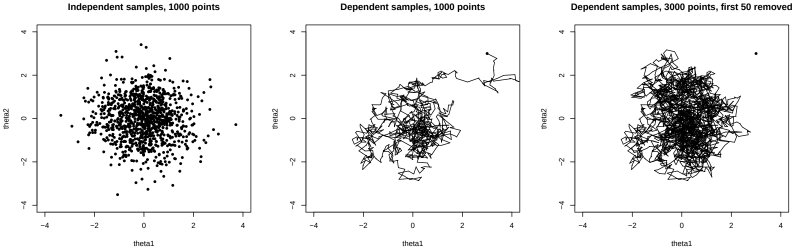
Samples can be independent or dependent

- Two ways to generate the samples: independent or dependent (i.e., value of one sample depends on the value of the previous one)
- Both give the same distribution in the end, but with dependent samples:
  - Start point matters for a while – i.e., it takes a while to "burn-in"
  - You need more samples for the same level of accuracy



Samples can be independent or dependent

- Two ways to generate the samples: independent or dependent (i.e., value of one sample depends on the value of the previous one)
- Both give the same distribution in the end, but with dependent samples:
  - Start point matters for a while – i.e., it takes a while to “burn-in”
  - You need more samples for the same level of accuracy



Outline

Monte Carlo Integration

- Example: Estimate the posterior (marginal) expectation of θ given $\boldsymbol{x}$

$$\mathbb{E}_\pi(\theta) = \int \theta \pi(\theta|\boldsymbol{x}) d\theta$$

- Say we have a sample $\theta^1, \ldots, \theta^n$ from the $\pi(\theta|\boldsymbol{x})$
- We can estimate the expectation using

$$\hat{\mathbb{E}}_\pi(\theta) = \frac{1}{n} \sum_{i=1}^{n} \theta^i$$

- This idea extends to any function of θ, $f(\theta)$. For example, the posterior mean of $f(\theta)$:

$$\hat{\mathbb{E}}_\pi(f(\theta)) = \frac{1}{n} \sum_{i=1}^{n} f(\theta^i)$$

Monte Carlo Integration

- For 100 samples θ^i, the 5% quantile of $\pi(\theta|\boldsymbol{x})$, is obtained by considering the fifth sample $\theta^{(5)}$, after sorting the sample to $\theta^{(1)}, \theta^{(2)}, \theta^{(3)}, \dots$ in increasing order:

$$P_\pi(\theta \leq \theta^{(5)}|\boldsymbol{x}) \simeq 0.05.$$

- The posterior variance of $f(\theta)$?
- The probability $f(\theta)$ is more than some value c?

See Exercise 2 in tutorial 4.

Example

- Suppose $\pi(\theta|\boldsymbol{x}) \sim N(0,1)$.
- Hence $\mathbb{E}(\theta) = 0$ and $\mathrm{sd}(\theta) = 1$.
- Simulate values from `rnorm` in `R`:

Samples	1		2		3	
n	Sample Mean	Sample SD	S. Mean	S. SD	S. Mean	S. SD
10	-0.39	1.26	-0.36	1.32	-0.09	0.87
100	0.36	0.88	-0.23	0.94	0.10	0.95
1000	-0.10	1.01	-0.01	1.01	0.02	0.98
10000	-0.01	1.00	0.00	0.99	0.00	1.00
100000	0.00	1.00	0.00	1.00	0.00	1.00

Monte Carlo Error

- Measures the between-simulation variability in estimates
- For n independent θ^i samples, for estimating the expected value of θ:

$$\overline{\theta} \sim N\left(\mathbb{E}_\pi(\theta), \frac{\sigma^2}{n}\right).$$

- σ^2/n is the Monte Carlo variance
- $\sqrt{\sigma^2/n}$ is the Monte Carlo error.
- See later in module for dependent samples...

Aside: Calculating any definite integral using MC integration

- MC integration can in fact be used to calculate any definite integral

$$\int_a^b f(x)dx$$

- you can introduce the Uniform $U(a, b)$ within the integral

$$(b-a)\int_a^b f(x)\frac{1}{b-a}dx$$

- so that the integral is calculated by approximating the expectation of $f(x)$ multiplied by $(b-a)$.

$$\int_a^b f(x)dx \simeq (b-a)\hat{\mathbb{E}}(f(x)) = \frac{(b-a)}{n}\sum_{i=1}^{n} f(x^i)$$

- where the x^i are sampled from the uniform $U(a, b)$.

Conclusion

- Bayesian problems frequently lead to intractable integrals
- If we can sample from the posterior, we can use Monte Carlo integration to obtain estimates of quantities of interest
- So, we have replaced an integration problem with a sampling problem
- How do we obtain samples from the posterior?
- ... see future lectures.