Mattias Villani
Department of Statistics
Stockholm University
Dept of Computer and Information Science
Linköping University

Advanced Bayesian Learning

Computer Lab 3 - Variational Inference

The labs are the only examination, so you should do the labs **individually**.
You can use any programming language you prefer, but do **submit the code**.
Submit a readable report in **PDF** (no Word documents!) or a **JuPyteR notebook**

---

1. **AR process with steady state parametrization**. Consider the autoregressive process of first order
$$y_t = \mu + \phi(y_{t-1} - \mu) + \varepsilon_t, \quad \text{where } \varepsilon_t \overset{iid}{\sim} N(0, \sigma^2)$$
for $t = 1, \ldots, T$ and μ is the unconditional mean of the process $\mathbb{E}(y_t) = \mu$. Assume the priors $\mu \sim N(0, \sigma_\mu^2)$, $\phi|\sigma^2 \sim N(0, \sigma^2/\kappa_0)$ and $\sigma^2 \sim \text{Inv}-\chi^2(\nu_0, \sigma_0^2)$. For simplicity we do not impose the stationarity restriction $-1 < \phi < 1$, but instead set $\sigma_0^2 = 1$, $\kappa_0 = 8$ and $\nu_0 = 4$, so the process is stationary with prior probability ≈ 0.95 [Note that the marginal prior for ϕ is $t_{\nu_0}(0, \sigma_0^2/\kappa_0)$ so these values gives $\Pr(-1 < \phi < 1) \approx 0.95$].

   (a) Derive the three full conditional posteriors $p(\phi|\mu, \sigma^2, \boldsymbol{y})$, $p(\sigma^2|\phi, \mu, \boldsymbol{y})$ and $p(\mu|\phi, \sigma^2, \boldsymbol{y})$ where $\boldsymbol{y}$ is a time series of length T. Assume that $y_0 = 0$ so that the likelihood can be set up conditional on this pre-sample value. It is OK to refer to results from these slides `https://github.com/mattiasvillani/Talks/raw/master/GuestLectureKTH2020.pdf` instead of doing boring completions of squares.
   [Hint 1: when the pre-sample value y_0 is known, AR processes are estimated just like linear regressions].
   [Hint 2: The joint posterior is anatytically intractable because of the product of parameters, $\phi\mu$, but the full conditional posteriors are tractable].
   [Hint 3: when deriving full conditional posteriors it is helpful to think about the *model* when the conditioning parameters are known. One can then often rewrite the model so that the full conditional posterior is immediate from the results in `https://github.com/mattiasvillani/Talks/raw/master/GuestLectureKTH2020.pdf`].

   (b) Code up a Gibbs sampler using these full conditionals and simulate from the posterior for the time series in file `timeseries.csv`. Set $\sigma_\mu = 2$.

   (c) Derive a mean-field variational approximation for the posterior $p(\mu, \phi, \sigma^2|\boldsymbol{y})$, code it up, and compare the VI approximation of the posterior to the results in 1b).

   (d) Code up the Fixed Form VI algorithm with control variate and adaptive learning rate (Algorithm 3 in Tran's notes) to approximate the posterior distribution $p(\mu, \phi, \sigma^2|\boldsymbol{y})$ for the data in file `timeseries.csv`. Use a multivariate normal as approximating density for $p(\mu, \phi, \psi|\boldsymbol{y})$ where $\psi = \log \sigma^2$. Compare the accuracy of this approximation to 1b) and 1c).

Good luck! Remember that sometimes an approximate answer is the correct answer!