Lab 8.5: Review for Exam II

Olivier Binette

Friday October 16, 2020

Agenda

- Annoucements
- Review of gaussian mixture models
- Gibbs sampling exercise
- Appendix: general review of sampling methods

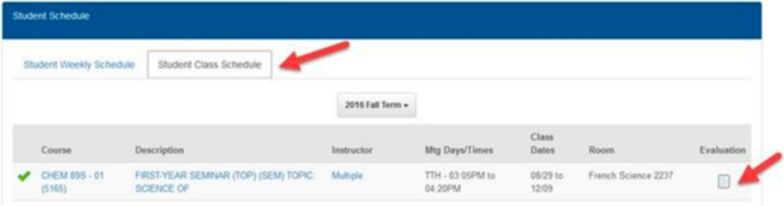
Announcements

Announcements

- Please vote!
  - Forgot to register? Not a problem! Bring an ID and proof of residence (e.g. bank statement or utility bill) to vote **before October 31** at an early voting site.
- Please fill out the TA section of your class evaluation!
  - Your evaluations are very important to me.

Announcements

On DukeHub:



After filling out the course evaluation, you should have the option to add an evaluation for your TAs.

Review of gaussian mixture models

Review of gaussian mixture models

Let's consider the **two-component** gaussian mixture model from Module 7 (part 3).

We have height data X_i, $i = 1, 2, \ldots, n$, corresponding to males ($Z_i = 0$) and females ($Z_i = 1$). Here we assume that the variable Z_i are unobserved.

Model: If $Z_i = 0$, then $X_i \sim N(\mu_1, \lambda^{-1})$. If $Z_i = 1$, then $X_i \sim N(\mu_2, \lambda^{-1})$. We assume that the X_i are conditionally independent given the other variables.

Priors:

- $Z_i \mid \pi \sim^{i.i.d.}$ Bernouilli(π)
- $\pi \sim$ Beta(a, b)
- $\mu_j \sim^{i.i.d.} N(m, \ell^{-1})$
- $\lambda \sim$ Gamma(c, d)

Review of gaussian mixture models

Task 1: Write down the likelihood of the data $X_{1:n}$.

Review of gaussian mixture models

Task 2: Write down the joint posterior distribution (up to a proportionality constant).

Review of gaussian mixture models

Task 3: Derive the full conditional distributions for all of the parameters:

1. $Z_i \mid - \sim ?$
2. $\pi \mid - \sim ?$
3. $\mu_j \mid - \sim ?$
4. $\lambda \mid - \sim ?$

Gibbs sampling exercise

From Lab 7:

Consider the following Exponential model for observations $x = (x_1, \ldots, x_n)$:

$$p(x|a, b) = ab \exp(-abx) I(x > 0)$$

and suppose the prior is

$$p(a, b) = \exp(-a - b) I(a, b > 0).$$

You want to sample from the posterior $p(a, b|x)$.

Gibbs sampling exercise

Task 1: Write down the joint posterior distribution, up to a normalization constant.

Task 2: Derive the full conditional distributions.

Task 3: Implement a Gibbs sampler.

Supplementary exercices

- modern-bayes/exercises/exercises-exam-two/practice-exercises-examII.pdf

Hoff book:

- Exercise 6.1
- Exercise 6.2

Appendix: review of sampling methods

Review of sampling methods

1. Inverse CDF method
2. Rejection sampling
3. MCMC methods
   - Metropolis-Hastings
   - **Gibbs sampling**

1. Inverse CDF method

Goal: Generate samples $X_1, X_2, \ldots, X_n$ from a distribution on $\mathbb{R}$ with CDF F.

The trick: If F is invertible and $U \sim \text{Unif}(0,1)$, then $X = F^{-1}(U)$ has the correct distribution.

When is it used? - Works only for *univariate* distributions. - You need to be able to evaluate F^{-1}.

1. Inverse CDF method

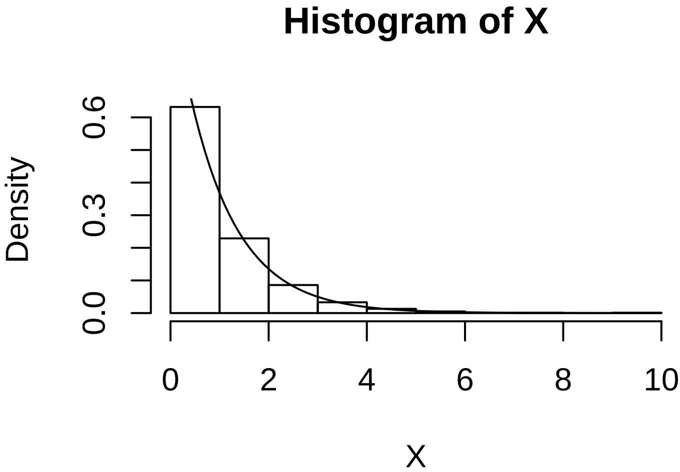
Example: Sampling from an $\text{Exp}(\lambda)$ distribution

1. The CDF of $X \sim \text{Exp}(\lambda)$ is $F(x) = 1 - e{-}\lambda x$.
2. Its inverse is $F^{-1}(u) = -\log(1-u)/\lambda$.

```r
F.inv <- function(u, lambda=1) -log(1-u)/lambda

n = 1000
X = F.inv(runif(n))
```

1. Inverse CDF method

```
hist(X, prob=TRUE)
curve(dexp(x), add=TRUE)
```



2. Rejection sampling

Goal: Generate samples $X_1, X_2, \ldots, X_n$ from a distribution with density (proportional to) $p(x)$.

The trick: Try to find a density $q(x)$ which you can sample from and such that $cq(x) \geq p(x)$ for some c.

Algorithm:

1. Generate $X \sim q(x)$ and $Y \sim \text{Unif}(0, cq(X))$.
2. If $Y < p(X)$, then return X. Otherwise go back to step 1.

2. Rejection sampling

Example:

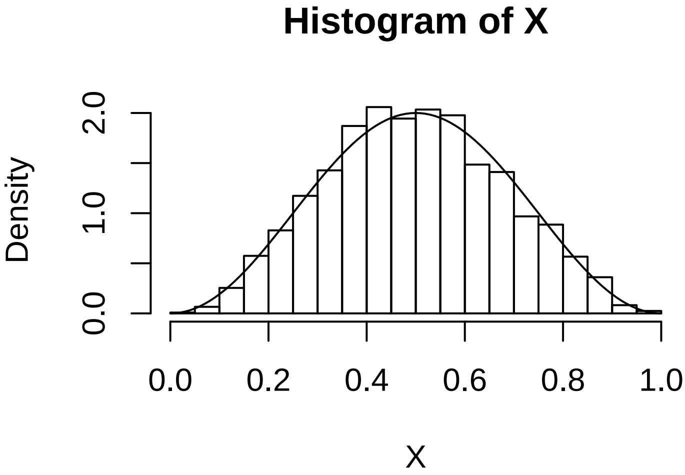
Let $p(x) = \sin^2(\pi x)$ be defined on $[0, 1]$ and let $q(x) = 1$ for all x. Take $c = 1$ since $p(x) \leq 1$.

```r
p <- function(x) sin(pi*x)^2
q <- Vectorize(function(x) 1)

# Vectorized form of rejection sampling:
k = 5000
X = runif(k) # Samples from q
Y = runif(k) # Samples uniform between 0 and cq(X)
X = X[Y < p(X)] # Only keep the X for which Y < p(X).

length(X)/5000 # Acceptance rate
```

```
## [1] 0.4876
```

2. Rejection sampling

```
hist(X, prob=TRUE, breaks=20)
curve(2*p(x), add=TRUE)
```



2. Rejection sampling

When is rejection sampling used?

- Works great for *univariate* densities (just like the inverse CDF method).
- You don't even need a normalizing constant for p (e.g. posterior distributions!).
- Trickier for higher-dimensional distributions (that's where Gibbs sampling comes in).

3. Metropolis-Hastings

Goal: Generate a Markov Chain $X^{(1)}, X^{(2)}, \dots, X^{(n)}$ with stationary distribution (proportional to) $p(x)$.

- In practice the $X^{(s)}$ are seen as correlated samples from the density proportional to $p(x)$.

The trick:

- Given $X^{(s)} = x$, propose $X^{(s+1)} = x^*$ following some distribution $J(x^* \mid x)$.
- Accept the proposal with probability

$$\alpha = \min\left\{1, \frac{p(x^*)J(x \mid x^*)}{p(x)J(x^* \mid x)}\right\},$$

- Otherwise set $X^{(s+1)} = X^{(s)} = x$.

Metropolis-Hastings

https://gfycat.com/relievedglossyhowlermonkey

3. Metropolis-Hastings

When is it used?

- To sample from high-dimensional distributions
- No need to know a normalizing constant for $p(x)$ (e.g. posterior distributions!).

What to watch out for?

- Convergence issues: you want your samples to be a good approximation to p and to not be too correlated with one another.
- The acceptance rate of the proposals can help diagnose issues, but it doen't tell you about convergence.
- You need to look at convergence diagnostics.

4. Gibbs sampling

Goal: Generate a Markov Chain $X^{(1)}, X^{(2)}, \ldots, X^{(n)}$ with stationary distribution (proportional to) $p(x)$, where $x = (x_1, x_2, \ldots, x_k)$.

The trick: Reduce to sampling from the *full conditional distributions* $p(x_i \mid x_{(-i)})$.

Algorithm:

1. Initialize $X^{(1)} = (X_1^{(1)}, X_2^{(1)}, \ldots, X_k^{(1)})$ to fixed values.
2. For $s = 2, 3, \ldots, n$, do:

- $X_1^{(s)} \sim p(x_1 \mid X_2^{(s-1)}, X_2^{(s-1)}, \ldots, X_k^{(s-1)})$
- $X_2^{(s)} \sim p(x_2 \mid X_1^{(s)}, X_3^{(s-1)}, \ldots, X_k^{(s-1)})$
- $X_3^{(s)} \sim p(x_3 \mid X_1^{(s)}, X_2^{(s)}, X_4^{(s-1)}, \ldots, X_k^{(s-1)})$
- $\vdots$
- $X_k^{(s)} \sim p(x_k \mid X_1^{(s)}, X_2^{(s)}, \ldots, X_{k-1}^{(s)})$

4. Gibbs sampling

Example: Go back to the gaussian mixture model example.

When is it used?:

- To sample from high-dimensional distributions
- No need to know a normalizing constant for $p(x)$ (e.g. posterior distributions!).
- You need to derive the full-posterior distributions.

What to watch out for:

- Convergence issues: you want your samples to be a good approximation to p and to not be too correlated with one another.
- You need to look at convergence diagnostics.