ABSTAT17

IGC, April 10-13, 2017

Challenge: Gibbs Sampling

At the end of this challenge, you should be able to (1) understand the Gibbs Sampler mechanism, (2) how to run it for the Multinomial–Dirichlet distributions with known $\mathbf{x}$, and (3) implement the main functions in `R`.

Step 1. Simulate the data - Multinomial Distribution

(a) Learn about Multinomial ditribution. For the three-dimensional case we have:

$$\mathbf{X} = (X_1, X_2, X_3) \frown Multinomial(n; \theta_1, \theta_2, \theta_3)$$

$$\text{where } X_i \in \{0, 1, 2, \ldots\} \text{ and } \theta_1 + \theta_2 + \theta_3 = 1 \quad (\theta_i > 0)$$

$$P[(X_1, X_2, X_3) = (x_1, x_2, x_3) \mid \theta_1, \theta_2, \theta_3] = \frac{n!}{x_1!x_2!x_3!}\ \theta_1^{x_1}\theta_2^{x_2}\theta_3^{x_3}$$

(b) Search for the `R` function which generates multinomially distributed random number vectors and computes multinomial probabilities.

(c) Simulate 1 random vector, $\mathbf{x} = (x_1, x_2, x_3)$, following a Multinomial distribution with parameters $n = 1000$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3) =$ $= (0.2, 0.3, 0.5)$. Store the simulated data in an object named `data`.

(d) Calculate the probability of observing the vector (220,350,430), that is, $P[(X_1, X_2, X_3) = (220, 350, 430) \mid \boldsymbol{\theta}]$.

Step 2. Dirichlet Distribution - Prior / Posterior Distribution

(a) Learn about Dirichlet ditribution. For the three-dimensional case we have:

$\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3) \frown Dirichlet(a_1, a_2, a_3)$

where $a_1, a_1, a_3 > 0$ and $a = a_1 + a_2 + a_3$

$$p_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \frac{\Gamma(a)}{\Gamma(a_1)\Gamma(a_2)\Gamma(a_3)}\ \theta_1^{a_1-1}\theta_2^{a_2-1}\theta_3^{a_3-1}$$

(b) Search for the `R` function which generates Dirichlet distributed random number vectors and computes Dirichlet probabilities.

(c) Simulate a random vector $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \theta_3^{(0)})$, from a Dirichlet distribution with hyperparameters $\mathbf{a}^{(0)} = (a_1^{(0)}, a_2^{(0)}, a_3^{(0)}) = = (1, 1, 1)$. Store the simulated vector in `theta.0` and vector $\mathbf{a}^{(0)}$ in `a.0` .

(d) Calculate the probability density for the vector (0.15,0.25,0.6), that is, $p_{\boldsymbol{\theta}}[(0.15, 0.25, 0.6)]$.

(e) **Optional challenge**: Show that the Dirichlet distribution is the conjugate prior of the Multinomial distribution, by achiving the following result: $p_{\boldsymbol{\theta}|\mathbf{x}}(\boldsymbol{\theta}) \propto \theta_1^{a_1+x_1-1}\theta_2^{a_2+x_2-1}\theta_3^{a_3+x_3-1}$

What are the parameters of the posterior distribution?

Step 3. The Gibbs Sampler.

(a) Develop a function in `R` for the Gibbs Sampler. Consider 5000 iterations ($nr.iter = 5000$).

(b) Store the updated parameters for each iteration in a matrix of order $nr.iter \times 3$.

Step 4. Trace / Parameters Estimation

(a) Represent graphically the trace for each parameter along the 5000 iterations.

(b) Evaluate the need of setting a period of burn-in.

(c) Find estimates of the parameters according to the decisions made in (b).