Master in Computer Vision Barcelona

Module: Optimization methods in CV

Inference algorithms II: Gibbs sampling

Lecturer: O. Ramos Terrades

Outline

Main concepts

Random samples

Let be x a set of random variables, $x_1, \ldots, x_N$ sorted in a topological order. A *random sampling* is a set of instances $\xi[1], \ldots, \xi[M]$ sampled from a distribution $p(x)$.

Particle

Instantiation to all, or some of, the variables in the graphical model. It is designed to provide good representations of the overall joint probability distribution.

- *Full*: complete assignments to all of graphical model variables
- *Collapsed*: a partial assignment w only to some subset of variables.

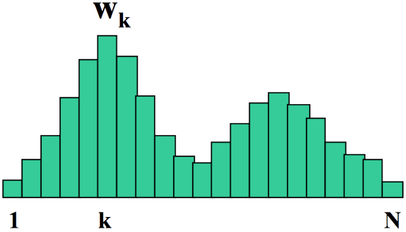
Forward sampling

For each r.v. x_i

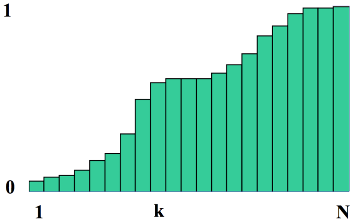
1. $u_i \leftarrow$ Assignment to Pa_{x_i} in $x_1, \ldots, x_{i-1}$
2. Sample x_i from $p(x_i | u_i)$

return $\xi[i] = x_1, \ldots, x_N$

Distribution:
$p(x = l) = \theta_l$



Cumulative function:
$F_p(l) = \sum_{k=1}^{l} \theta_l$



Interval $[0, 1)$ is uniformly sampled: $[0, \theta_1)$; $[\theta_1; \theta_1 + \theta_2); \ldots$

$$\hat{p}_{\mathcal{D}}(y) = \frac{1}{M} \sum_m \mathbb{1}\{\xi[m] = y\}$$

Markov chain

Markov chain

A sequence $x_1, x_2, \ldots$ of random elements of some set is a *Markov chain* if the conditional distribution of x_{n+1} given $x_1 \ldots, x_n$, depends on x_n only:

- x_1 is called the *initial transition*.
- $p(x_{n+1}|x_n)$ is called the *transition probability distribution*.

Stationary transition probabilities

A Markov chain has *stationary transition probabilities* if the conditional distribution of x_{n+1} given x_n does not depend on n.

Stationary

An initial distribution is said to be *stationary* or *invariant* or *equilibrium* for some transition probability distribution if the Markov chain specified by this initial distribution and transition probability distribution is stationary.

Reversibility

Reversible transition probability

A transition probability distribution is *reversible* with respect to an initial distribution if, for the Markov chain $x_1, x_2, \ldots$ they specify, the distribution of pairs (x_i, x_{i+1}) is exchangeable.

Reversible MC

A Markov chain is *reversible* if its transition probability is reversible with respect to its initial distribution. Reversibility implies stationarity, but not vice-versa.

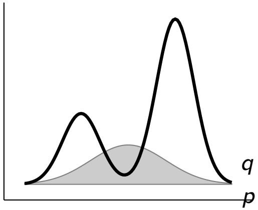
Importance sampling

Estimates the expectation of $f(x)$ relative to some *target distribution* $p(x)$. Samples $\xi[1], \ldots, \xi[M]$ from the *sampling distribution* q are generated and then:

$$E_p[f] \approx \frac{1}{M} \sum_m f(\xi[m]) \frac{p(\xi[m])}{q(\xi[m])} \tag{1}$$

- The estimator, and its variance, are *unbiased*
- p usually is unknown
- reject samples such that:

$$q(\xi[m]) \ll f(\xi[m])p(\xi[m])$$



Normalized importance sampling

Estimates the expectation of $f(x)$ relative to some distribution $p(x)$ is known only up to a normalizing constant Z. Samples $\xi[1], \dots, \xi[M]$ from the *sampling distribution* q are generated and then:

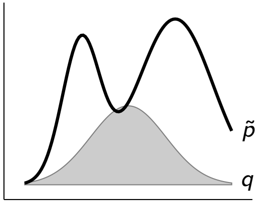
$$E_{\tilde{p}}[f] \approx \frac{1}{M} \sum_m f(\xi[m]) \frac{\tilde{p}(\xi[m])}{q(\xi[m])} \tag{2}$$

- $\tilde{p}(x) = Z \cdot p(x)$,
- The estimator, and its variance, are *biased*
- Reject samples such that:

$$q(\xi[m]) \ll f(\xi[m])\tilde{p}(\xi[m])$$

- Estimate Z:

$$Z \approx \frac{1}{M} \sum_m \frac{\tilde{p}(\xi[m])}{q(\xi[m])} \tag{3}$$



The Metropolis-Hasting algorithm

- Let be $h(x)$ an *unnormalized density*,
- Current state x propose to move to y, having conditional probability $q(x, \cdot)$.
- *Hastings ratio* is:

$$r(x, y) = \frac{h(y)q(y, x)}{h(x)q(x, y)} \quad (4)$$

- Acceptance probability:

$$a(x, y) = \min\{1, r(x, y)\} \quad (5)$$

- Draw $u \sim U(0, 1)$ and accept y if $u < a(x, y)$

Example of q

$$q(x, y|\sigma) = e^{-\frac{(x-y)^2}{2\sigma^2}}$$

Gibbs sampling

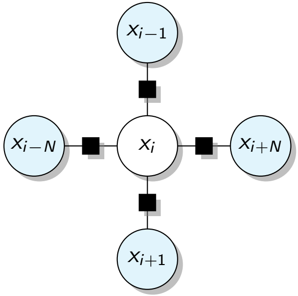
- Define a conditional probability q as:

$$p(x_i|x_{-i}) = p(x_i|x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_N) \quad \forall x_i \tag{6}$$

- The new sample y satisfies the acceptance ratio.
- Gibbs sampling produces Markov chain whose *stationary* distribution is the *posterior* distribution

Conditional probability on the Markov blanket

$$p(x_i|x_{-i}) = \frac{\prod_{x_i \in D} \phi_D(x)}{\sum \prod_{x_i \in D} \phi_D(x)} \tag{7}$$



Gibbs sampling. Final remarks

- Gibbs samples are highly correlated
- The *burn-in time* is the number of steps to wait until the state distribution is reasonably close to p
- The *mixing time* is the number of in-between steps required so the samples are enough uncorrelated to be considered *independent*.