Final exam
Probabilistic graphical models
Master MVA 2015-2016

December 16th 2015

A - Short problems

1. What is the family of distribution $(p_\alpha)_{\alpha \in \mathbb{R}_+}$ with p_α the distribution of maximal entropy on $\mathbb{N}$, the set of natural integers $\{0, 1, 2, \ldots\}$, such that $\mathbb{E}_p[X] = \alpha$?

2. Propose a sampling scheme to sample exactly from the distribution $\mathbb{P}(X \in \cdot \mid \|X - y\|_2 \leq 1)$ where $y \in \mathbb{R}^d$ and X is a multivariate Gaussian random variable $\mathcal{N}(0, I_d)$. Prove that the proposed sampling scheme indeed yields a variable that has exactly the desired distribution.

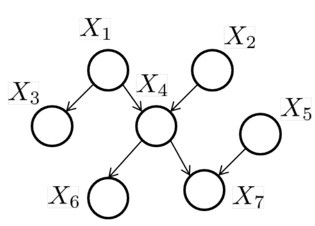
B - Factorization and Markov properties

1. Find an undirected graphical model on four random variables X_1, X_2, X_3, X_4 that satisfies simultaneous the following conditions (a) and (b)

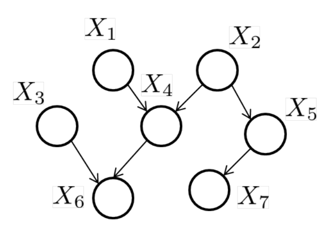
   (a) All distributions that factorize according to the model satisfy the following conditional independence statements:
   $$X_1 \perp\!\!\!\perp X_2 \mid (X_3, X_4), \quad X_3 \perp\!\!\!\perp X_4 \mid (X_1, X_2),$$

   (b) There exist distributions that factorize according to the model and such that
   $$X_1 \not\perp\!\!\!\perp X_3 \mid X_2, \quad X_1 \not\perp\!\!\!\perp X_3, \quad X_2 \not\perp\!\!\!\perp X_3, \quad X_1 \not\perp\!\!\!\perp X_4, \quad X_2 \not\perp\!\!\!\perp X_4.$$

2. Consider the two directed graphical models below. Is it possible to have a distribution p on $X_1, \ldots, X_7$ such that $p \in \mathcal{L}(G_1)$ and $p \in \mathcal{L}(G_2)$? Justify your answer.



G_1



G_2

C - EM algorithm

Consider (X, Y) a pair of correlated Bernoulli random variables which we parameterize by $(\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11})$ with $\pi_{kj} = \mathbb{P}(X = k, Y = j)$. Assume that we observe the three following independent samples

- an i.i.d. sample $(X_1, \ldots, X_{n_x})$ from the marginal distribution of X, together with
- an i.i.d. sample $(Y_1, \ldots, Y_{n_y})$ from the marginal distribution of Y, and
- an i.i.d. sample $\big((X'_1, Y'_1), \ldots, (X'_m, Y'_m)\big)$ from the joint distribution of (X, Y)

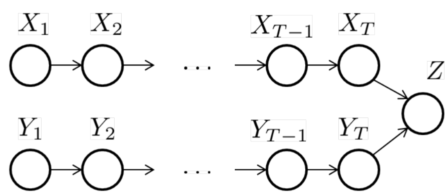
Consider the notations:

- $N_x = \sum_{i=1}^{n_x} X_i,\ N_y = \sum_{i=1}^{n_y} Y_i,$
- $M_{jk} = \sum_{i=1}^{m} \delta(X'_i, k)\, \delta(Y'_i, j)$

Propose an EM algorithm to estimate $(\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11})$. In particular, specify which quantities are computed for the E-step and which quantities are computed for the M-step.

D - Parallel chains

Consider the directed graphical model G below:



Suppose that the variables X_t and Y_t are discrete K-valued for $t = 1, \ldots, T$, and that Z is a binary random variable. Consider the following form for the factors for a specific distribution p in $\mathcal{L}(G)$:

- $p(X_1 = l) = p(Y_1 = l) = \pi_l$ for $l = 1, \ldots, K$.
- $p(X_t = i|X_{t-1} = j) = p(Y_t = i|Y_{t-1} = j) = A_{ij}$ for $i, j = 1, \ldots, K$ and $t = 2, \ldots, T$.
- $p(Z = 1|X_T = k, Y_T = l)$ takes the value p whenever $k = l$, and q otherwise (i.e. when $k \neq l$).

By using the graph eliminate algorithm (or just clever use of distributivity), give a simple formula for the marginal $p(Z = 1)$ in terms of the matrix A, vector π and scalars q, p and T. Provide brief explanations of how to obtain the result.

E - Metropolized Gibbs sampler

We consider the Gibbs distribution p of the random variable $X = (X_1, \ldots, X_n)$ (which we can think of as a distribution from a graphical model of interest). For simplicity we assume that the X_i are K-valued random variables. We will use the notation X_{-i} to refer to $X_{\{1,\ldots,n\}\setminus\{i\}}$. We will denote by p_i the conditional distribution of the ith variable given all the others, as induced from the Gibbs distribution so that

$$p_i(\, z_i \mid x^t_{-i}) := \mathbb{P}(X_i = z_i \mid X_{-i} = x^t_{-i}).$$

The Metropolized Gibbs sampler is a variant of Gibbs sampling, which takes the form of a Metropolis-Hasting algorithm that updates a single variable X_i at a time: to update the variable X_i, instead of just sampling this variable conditionally on the other variables, it makes a proposal drawn from the transition T_i with

$$T_i\big((x^t_i, x^t_{-i}), (z_i, x^t_{-i})\big) := \begin{cases} \dfrac{p_i(z_i \mid x^t_{-i})}{1 - p_i(x^t_i \mid x^t_{-i})} & \text{for } z_i \neq x^t_i \\ 0 & \text{for } z_i = x^t_i \end{cases} \tag{1}$$

and accepts the move with probability

$$\alpha_i\big((x_i^t, x_{-i}^t), (z_i, x_{-i}^t)\big) := \min\left\{1, \frac{1 - p_i(x_i^t \mid x_{-i}^t)}{1 - p_i(z_i \mid x_{-i}^t)}\right\}.$$

You may use the notations $\alpha_i(x_i^t, z_i) := \alpha_i\big((x_i^t, x_{-i}^t), (z_i, x_{-i}^t)\big)$ and $T_i(x_i^t, z_i) := T_i\big((x_i^t, x_{-i}^t), (z_i, x_{-i}^t)\big)$ in you answers to simplify the notations.

1. Show that the transition corresponding to an update of variable X_i of the Metropolized Gibbs sampler satisfies detailed balance with the Gibbs distribution.

2. Does the Markov Chain produced by the Metropolized Gibbs sampler converge to the Gibbs distribution? Explain why. To answer this question you can either consider a random scan Gibbs sampler that picks the variable i uniformly at random at each iteration or a cyclic scan Gibbs sampler that samples each variable X_i in a cycle.

F - Bayesian estimation: posterior mean *vs* MAP

Assume we observe an i.i.d. sample $(x_1, \dots, x_n)$ of realizations of a Bernoulli random variable whose unknown moment parameter we denote by $\mu = \mathbb{E}_\mu[X] = \mathbb{P}_\mu(X = 1)$.

We consider first the Bayesian estimation of μ and the corresponding MAP estimator, based on a uniform prior distribution of μ.

1. Express the posterior mean estimate $\hat{\mu}_{\mathrm{PM}}$ for μ as a function of $\bar{x} := \frac{1}{n}\sum_{i=1}^n x_i$.

2. What is the maximum a posteriori (MAP) estimate $\hat{\mu}_{\mathrm{MAP}}$ for μ?

We now consider again the same estimators, but we instead work with η, the canonical parameter of the exponential family.

(c) If a uniform prior distribution $p_\mu(\mu)$ is placed on μ what is the induced prior distribution on η? Provide the density of that prior $p_\eta(\eta)$ when p_μ is the uniform prior on the interval $[0, 1]$.

(d) What is the value of the posterior mean estimate $\int \mu(\eta)\, p(\eta | x_1, \dots, x_n)\, d\eta$ under this prior p_η on η, for $\mu(\eta) = (1 + e^{-\eta})^{-1}$ the usual moment mapping?

(e) For this prior p_η, what is the MAP estimator $\hat{\eta}_{\mathrm{MAP}}$? What is the corresponding moment parameter $\mu(\hat{\eta}_{\mathrm{MAP}})$?

(f) Comment on the different estimators obtained.