Homework 12

Reading

- Section 8.3 covers undirected graphical models.
- Section 1.2.3 provides a general discussion of Bayesian probability. Bishop uses Bayesian models throughout the text, we have just not emphasized them previously.

1. The file `HW12_problem1.txt` contains 50 iid samples, $\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_{50}$, from a one-dimensional random variable X.. Assume that $X \sim \mathcal{N}(\mu, \sigma^2)$ where σ^2 is known with $\sigma^2 = 1$. Our goal is to estimate μ. Let $\bar{x}$ be the sample mean and $N = 50$.

   (a) Show that the maximum likelihood estimate (MLE) of μ is given by the sample mean $\bar{x}$.

   (b) Taking a Bayesian approach, assume a normal prior on μ, $\mu \sim \mathcal{N}(0, \beta^2)$ with $\beta = 10$. Let $f(\mu)$ be the pdf of the prior. Let $p(\mu)$ be the posterior. Show,

   $$p(\mu) = \frac{1}{Z} P(\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_{50} \mid \mu) f(\mu), \tag{1}$$

   where

   $$Z = \int_{-\infty}^{\infty} P(\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_{50} \mid \mu) f(\mu) d\mu. \tag{2}$$

   Then show

   $$p(\mu) \sim \mathcal{N}\left(\frac{\bar{x}}{1 + \frac{\sigma^2}{\beta^2 N}}, \frac{\sigma^2 \beta^2}{N\beta^2 + \sigma^2}\right) \tag{3}$$

   Graph $p(\mu)$ using the data. Hint: use completing the squares to combine a product of exponentials into a single exponential. (Here you're showing that for random mean and fixed variance, the normal distribution is a conjugate prior to itself)

   (c) Again take a Bayesian approach, but this time assume a prior $f(\mu)$,

   $$f(\mu) = \begin{cases} \frac{1}{10} & \text{if } x \in [5, 7] \\ 4 & \text{if } x \in [9, 9.2] \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

   In this case the posterior $p(\mu)$ is not normally distributed. Graph $p(\mu)$ and compare to the graph in (b). (Hint: you can compute Z using a numerical integration function. In R, **integrate**).

2. This problem relates to the figure below, which is Figure 17.3 in the book Elements of Statistical Learning. Let W be a 3-dimensional r.v. Typically we write $W = (W_1, W_2, W_3)$, but to match the figure, let $W = (X, Y, Z)$.

Suppose that X, Y, Z are each in $\{-1, 1\}$. Consider three parametrizations of a probabilities distrubution for W, i.e. a joint probability distribution for X, Y, Z. Let $\eta = (\eta_1, \eta_2, \eta_3)$.

$$P(X = x, Y = y, Z = z) = \alpha \exp[\eta_1 x + \eta_2 y + \eta_3 z - w_{12}xy - w_{13}xz] \quad (5)$$
$$P(X = x, Y = y, Z = z) = \alpha \exp[\eta_1 x + \eta_2 y + \eta_3 z - w_{12}xy - w_{13}xz - w_{23}yz] \quad (6)$$
$$P(X = x, Y = y, Z = z) = \alpha \exp[\eta_1 x + \eta_2 y + \eta_3 z - w_{12}xy - w_{13}xz - w_{23}yz - w_{123}xyz], \quad (7)$$

where the α is a normalizing constant that differs between the three distributions. (In the lecture I used $1/Z$ for the normalization, but here Z is one of the random variables.)

(a) Show that for distribution (5), $Y \perp Z | X$ so that the edge between Y and Z in the graph is not consistent with the distribution. Show that for distribution (5), we can write

$$P(X = x, Y = y, Z = z) = \alpha \psi_1(x, y)\psi_2(x, z) \quad (8)$$

corresponding to the maximal clique form of the Hammersley-Clifford Theorem. What are the cliques in this case?

(b) Show that for distributions (6) and (7), no pair of the three r.v. X, Y, Z is conditionally independent given the third r.v., making each edge in the graph necessary for consistency with the probability distribution. (Note: this reflects the comment directly below the figure: *A graphical model does not always uniquely specify the higher-order structure of a joint probability distribution.*)

(c) Consider distribution (5). Let $\eta_i = 1/2$ for $i = 1, 2, 3$. Let $w_{12} = 1$, $w_{13} = -1$. Use a Monte-Carlo approach based on a Metropolis-Hastings sampler to estimate the correlation between X and Y. (Since the graph is so small, in this case we could directly compute the correlation by summing through all possible values of W, but this would not be the case if the dimension of W was, say, 20.)

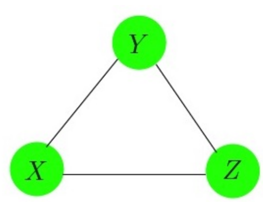

FIGURE 17.3. *A complete graph does not uniquely specify the higher-order dependence structure in the joint distribution of the variables.*