



English

Contact during exam:

Håkon Tjelmeland 73 59 35 38

EXAM IN TMA4300 COMPUTER INTENSIVE STATISTICAL METHODS

Thursday June 10th 2010

Time: 09:00–13:00

Aids: Calculator HP30S or Citizen SR-270X with empty memory.

Statistiske tabeller og formler, Tapir forlag.

K. Rottman: Matematisk formelsamling.

One yellow paper (A4 with stamp) with own formulas and notes.

Grading: July 1st 2010.

Problem 1

Consider a random sample $X_1, \dots, X_n$ where $X_i \in \{0, 1\}$ for $i = 1, \dots, n$, $P(X_i = 1) = p$ and $P(X_i = 0) = 1 - p$, where $p \in [0, 1]$ is a parameter. Note that we alternatively may write this as

$$P(X_i = x) = p^x(1 - p)^{1-x} \quad \text{for } x = 0, 1.$$

a) Write the distribution $P(X_i = x)$ on the form of a one-parameter exponential family,

$$P(X_i = x) = a(x)e^{\phi(p)t(x)+b(p)}$$

i.e. identify the functions $a(x)$, $\phi(p)$, $t(x)$ and $b(p)$.

Use this to write down a formula for the conjugate prior distribution for p . Show that this prior distribution can be expressed as

$$\pi(p) \propto p^{\alpha-1}(1 - p)^{\beta-1} \quad \text{for } p \in [0, 1].$$

What is this distribution called?

In the following we assume that we adopt this conjugate prior distribution for p and that we use $\alpha \geq 1$ and $\beta \geq 1$. As you know, it then follows from general theory that the posterior distribution $\pi(p|x_1, \dots, x_n)$ is of the same form as the prior, but with updated parameters, i.e.

$$\pi(p|x_1, \dots, x_n) \propto p^{\tilde{\alpha}-1}(1-p)^{\tilde{\beta}-1} \quad \text{for } p \in [0, 1].$$

- b) Find expressions for $\tilde{\alpha}$ and $\tilde{\beta}$ as a function of α , β , n and the observed values $x_1, \dots, x_n$.

Using a uniform distribution on $[0, 1]$ as proposal distribution, describe how one can generate samples from $\pi(p|x_1, \dots, x_n)$ by rejection sampling. In particular find formula for the acceptance probability expressed by $\tilde{\alpha}$, $\tilde{\beta}$ and the proposed value, and write pseudo code for the simulation algorithm.

Problem 2

In this problem we will consider a classification problem in a Bayesian setting. Assume we have only two classes, which we denote as class 0 and class 1, respectively. Corresponding to class 0 we have n observations $x_1, \dots, x_n$, which we assume to come from a normal distribution with mean μ_0 and variance θ_0 . Corresponding to class 1 we have m observations $y_1, \dots, y_m$, which we assume to come from a normal distribution with mean μ_1 and variance θ_1 . Finally, we have k observations $z_1, \dots, z_k$ with unknown classes, so our goal is to classify each of these to one of the two classes. Let $c_i \in \{0, 1\}$ denote the (unknown) class of z_i , and assume

$$z_i|c_i = 0 \sim N(\mu_0, \theta_0)$$

and

$$z_i|c_i = 1 \sim N(\mu_1, \theta_1).$$

We also assume all $x_1, \dots, x_n$, $y_1, \dots, y_m$ and $z_1, \dots, z_k$ to be independent of each other.

Following a Bayesian strategy we assign prior distributions to the unknown quantities $\mu_0, \theta_0, \mu_1, \theta_1$ and $c_1, \dots, c_k$. Apriori we assume μ_0, μ_1, θ_0 and θ_1 to be independent of each other and independent of $c_1, \dots, c_k$. As priors for μ_0 and μ_1 we adopt improper uniform distributions on $(-\infty, \infty)$, and for θ_0 and θ_1 we assume inverse gamma prior distributions with (known) parameters α and β . Given a hyper-parameter $p \in [0, 1]$ we assume $z_1, \dots, z_k$ to be apriori independent with

$$P(c_i = 1|p) = p, \quad P(c_i = 0|p) = 1 - p.$$

Finally, as prior for p we assume a uniform distribution on the interval $[0, 1]$.

In Figure 1 the specified model is illustrated as a graphical model when $n = m = k = 3$. To explore the resulting posterior distribution

$$\pi(\mu_0, \mu_1, \theta_0, \theta_1, p, c_1, \dots, c_k | x_1, \dots, x_n, y_1, \dots, y_m, z_1, \dots, z_k)$$

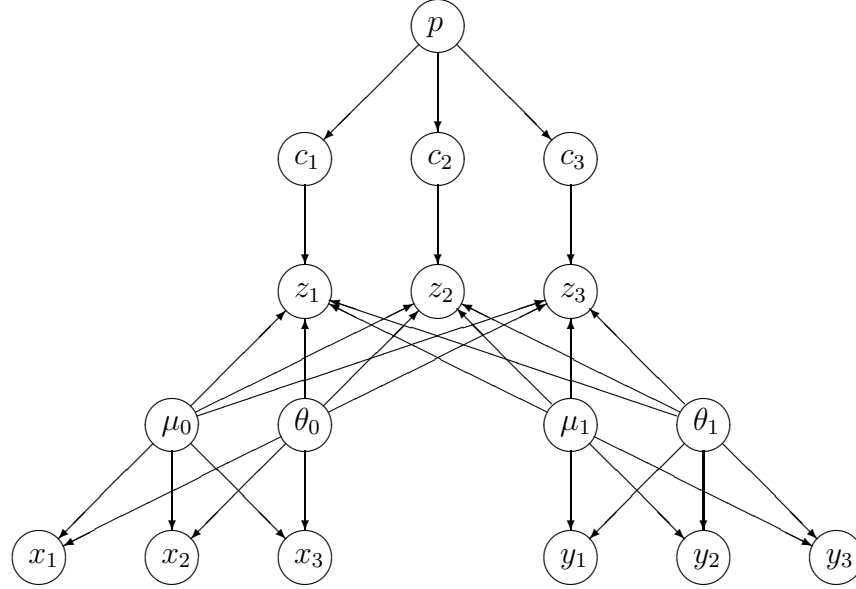


Figure 1: A graphical model representation of the Bayesian model considered in Problem 2 when $n = m = k = 3$.

we will define and simulate a Gibbs sampler algorithm.

- a) Discuss how the full conditional distribution for p can be found from your results in Problem 1. Conclude by writing down a formula for the full conditional distribution for p .

It can be shown (you do not need to do this!) that the full conditional distributions for μ_0 , μ_1 , θ_0 and θ_1 are

$$\mu_0 | \text{everything else} \sim N \left(\frac{\sum_{i=1}^n x_i + \sum_{i=1}^k (1 - c_i) z_i}{n + \sum_{i=1}^k (1 - c_i)}, \frac{\theta_0}{n + \sum_{i=1}^k (1 - c_i)} \right),$$

$$\mu_1 | \text{everything else} \sim N \left(\frac{\sum_{i=1}^m y_i + \sum_{i=1}^k c_i z_i}{m + \sum_{i=1}^k c_i}, \frac{\theta_1}{m + \sum_{i=1}^k c_i} \right),$$

$$\theta_0 | \text{everything else} \sim \text{IG} \left(\alpha + \frac{1}{2} \left[n + \sum_{i=1}^k (1 - c_i) \right], \beta + \frac{1}{2} \sum_{i=1}^n (x_i - \mu_0)^2 + \frac{1}{2} \sum_{i=1}^k (1 - c_i) (z_i - \mu_0)^2 \right)$$

and

$$\theta_1 | \text{everything else} \sim \text{IG} \left(\alpha + \frac{1}{2} \left[m + \sum_{i=1}^k c_i \right], \beta + \frac{1}{2} \sum_{i=1}^m (y_i - \mu_1)^2 + \frac{1}{2} \sum_{i=1}^k c_i (z_i - \mu_1)^2 \right),$$

respectively.

- b) Find the full conditional distribution for c_i . Give pseudo-code for simulating from this distribution. *Note: Remember that $c_i \in \{0, 1\}$*

Assume we have run a Gibbs sampler algorithm for S iterations and denote the generated states by $\{\mu_0^s, \mu_1^s, \theta_0^s, \theta_1^s, p^s, c_1^s, \dots, c_k^s\}_{s=0}^S$. In particular $\mu_0^0, \mu_1^0, \theta_0^0, \theta_1^0, p^0, c_1^0, \dots, c_k^0$ are the initial values.

- c) How would you estimate $P(c_i = 1 | x_1, \dots, x_n, y_1, \dots, y_m, z_1, \dots, z_k)$ from the simulated values? If necessary, specify additional assumptions you are doing.

Specify how you would use the simulated values to classify each z_i to one of the two classes. If necessary, specify additional assumptions you are doing.

Discuss how you from the simulated values would estimate the error rate of your classification rule. If necessary, specify additional assumptions you are doing.

In the model discussed above we consider all $z_1, \dots, z_k$ at the same time. An alternative is to consider only one of the z_i 's at a time with prior distribution $P(c_i = 0) = P(c_i = 1) = 1/2$, i.e. without the hyper-parameter p . Thus, for each $i = 1, \dots, k$ in turn we should then simulate from $\pi(\mu_0, \mu_1, \theta_0, \theta_1, c_i, | x_1, \dots, x_n, y_1, \dots, y_m, z_i)$ and use the result to classify z_i .

- d) Discuss how this would change the classification results relative to the above procedure. Give reasons for your answer.

Problem 3

Assume we have observations $x_1, \dots, x_n$ which are independent realisations from some (unknown) distribution F . We are interested in the variance of the distribution F , which we denote by θ , i.e.

$$\theta = t(F) = \text{Var}_F(X).$$

To estimate θ we use the sample variance, i.e.

$$\hat{\theta} = s(x) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

We know of course that $\hat{\theta}$ is an unbiased estimator of θ , but in this problem we will pretend we do not know this fact and therefore consider how we can estimate the bias of $\hat{\theta}$ by (non-parametric) bootstrapping.

- a) Define the bias of $\hat{\theta}$ as a function of F .

Define the ideal bootstrap estimator for the bias of $\hat{\theta}$.

Write pseudo-code for estimating the ideal bootstrap estimator for the bias of $\hat{\theta}$ by stochastic simulation.

The situation we consider here is sufficiently simple that it is also possible to evaluate analytically (i.e. without simulation) the ideal bootstrap estimator for the bias of $\hat{\theta}$.

- b) Show that the ideal bootstrap estimator for the bias of $\hat{\theta}$ is in fact equal to zero.