MATEMATIIKAN JA TILASTOTIETEEN LAITOS
Bayes-päättely
Loppukoe 8.3.2012

1. Explain the following terms or concepts by using an example:

a) Conjugate priors.

A specific type of prior distribution which leads to a posterior distribution that belongs to the same family of distributions as the prior. For example: prior $\pi(p) = \mathrm{Beta}(\alpha, \beta)$ together with the binomial model of observable data X, $\pi(X) = \mathrm{Bin}(N, p)$, leads to posterior $\pi(p \mid X) = \mathrm{Beta}(X + \alpha, N - X + \beta)$.

b) Combining individual expert knowledge: sum of prior densities versus product of prior densities.

Expert knowledge, when expressed in the form of probability density, can be combined with other such densities as a weighted sum (mixture of probability distributions) with equal weights (if there is no reason to believe one expert is better than another), or as a product of densities in which case the result needs to be normalized to make it a proper probability density. The simplest case is to have uniform densities for each expert. Mixture density is then $\pi(x) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{U}(x \mid a_i, b_i)$ where the density of the ith expert is $\mathrm{U}(x \mid a_i, b_i) = \frac{1}{b_i - a_i}$ when $x \in [a_i, b_i]$ and 0 otherwise. In contrast, the product of the individual uniform densities would give a uniform density over an interval that is intersection of all the intervals of the individual densities ($\cap_{i=1}^{n}[a_i, b_i]$).

c) Improper prior distribution.

A prior distribution whose integral is infinite, and thus cannot be normalized to a proper probability distribution. (These can still be used, **if** the resulting posterior density will be a proper probability distribution, but that depends on the data model, i.e. the likelihood function).

d) HPD interval.

Highest posterior density interval. Credible interval (probability interval) of a posterior distribution, containing a chosen probability (e.g. 95%), so that it is the shortest such interval. (Posterior density outside the interval is everywhere smaller than inside the interval).

e) Posterior predictive distribution.

Predictive distribution of new data, conditional to the existing data, but not conditional to specific parameter values of the model, because they are integrated out:

$$\pi(X_{\text{new}} \mid X) = \int_{\Theta} \pi(X_{\text{new}}, \theta \mid X)\mathrm{d}\theta = \int_{\Theta} \pi(X_{\text{new}} \mid \theta, X)\pi(\theta \mid X)\mathrm{d}\theta$$

$$= \int_{\Theta} \pi(X_{\text{new}} \mid \theta)\pi(\theta \mid X)\mathrm{d}\theta \ \ \text{(when } X_{\text{new}} \text{ is conditionally independent of } X\text{, given } \theta\text{).}$$

For example, this could be beta-binomial distribution (resulting from $X \sim \text{binomial}(n, p)$ and $p \sim$ beta), or negative binomial distribution (resulting from $X \sim \text{Poisson}(\lambda)$ and $\lambda \sim$ Gamma).
f) Bayes factor.
Ratio of posterior odds to prior odds. It shows how much the prior odds are changed by the data. This can be used in hypothesis testing problems. In the simplest case, the hypotheses involve only two points: A null hypothesis $H_0 : \theta = \theta_0$ against alternative hypothesis $H_1 : \theta = \theta_1$. The prior odds (for H_0) would be $\pi(\theta_0)/\pi(\theta_1)$ and the posterior odds $\pi(\theta_0 \mid X)/\pi(\theta_1 \mid X)$. With these point-hypotheses, terms of the prior distributions cancel out from the Bayes factor so that it will be just the ratio of the likelihood functions: $\pi(X \mid \theta_0)/\pi(X \mid \theta_1)$, to be calculated according to the chosen model (density-, or probability point mass function) of X.

2. There are two balls in a bag, one is known to be black, and the other is known to be either black or white. One ball is picked out and observed black. What is the probability that the remaining ball is then black, assuming you originally had no idea whether the unknown ball was more likely to be black or white.

$$P(\bullet\bullet \mid \bullet) = \frac{P(\bullet \mid \bullet\bullet)P(\bullet\bullet)}{P(\bullet \mid \bullet\bullet)P(\bullet\bullet) + P(\bullet \mid \bullet\circ)P(\bullet\circ)} = \frac{P(\bullet\bullet)}{P(\bullet\bullet) + 0.5P(\bullet\circ)}$$

$$= \frac{0.5}{0.5 + 0.5 * 0.5} = \frac{2}{3} \text{ , prior based on principle of 'insufficient reason'}$$

3. You have obtained two samples of measurements $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$. Assume the model for X_i is $N(\mu_1, \sigma_1^2)$ and for Y_i it is $N(\mu_2, \sigma_2^2)$ with known means μ_1 and μ_2. You need to study the hypothesis $H_0 : \sigma_1^2 > \sigma_2^2$ based on the samples and conjugate priors which are as 'uninformative' as possible. (Equivalently, $H_0 : \tau_1 < \tau_2$ for precisions $\tau_i = 1/\sigma_i^2$). Explain how to compute the posterior probability of H_0 with Monte Carlo method. Sketch the algorithm so it could be implemented based on your instructions. [This may be useful to remember: Gamma(α, β)-density for some variable θ is $\frac{\beta^\alpha}{\Gamma(\alpha)}\theta^{\alpha-1}\exp(-\beta\theta)$, and $N(\mu, \sigma^2)$- density for some variable z is $\frac{1}{\sigma\sqrt{2\pi}}\exp(-0.5\frac{(z-\mu)^2}{\sigma^2})$].

The Monte Carlo algorithm:
1. Sample a large number (k) of independent random values from the posterior distributions $\pi(\tau_1 \mid X_1, \ldots, X_n)$ and $\pi(\tau_2 \mid Y_1, \ldots, Y_m)$.
2. To approximate the posterior probability $P(\tau_1 < \tau_2 \mid X_1, \ldots, X_n, Y_1, \ldots, Y_n)$, calculate from the random sample the proportion of such pairs where $\tau_1 < \tau_2$. In other words: $\frac{1}{k}\sum_{i=1}^{k} 1_{\{\tau_1<\tau_2\}}(\tau_1^{(k)}, \tau_2^{(k)})$, where $1_{\{\}}$ is the indicator function.

What remains is to find out the posterior distributions. Below is solution for τ_1 (for τ_2 it is similar).

$$\pi(\tau_1 \mid X_1, \ldots, X_n) \propto \overbrace{\underbrace{\tau_1^{\alpha-1} \exp(-\beta\tau_1)}_{\text{from conjugate gamma prior}} \prod_{i=1}^{n} \underbrace{\tau_1^{1/2} \exp(-0.5(X_i - \mu)^2 \tau_1)}_{\text{from normal model of } X_i}}^{\text{from Bayes formula: '}\pi(\tau_1 \mid \text{data}) \propto \pi(\tau_1)\pi(\text{data} \mid \tau_1)\text{'}}$$

$$= \tau_1^{\alpha-1} \exp(-\beta\tau_1)\tau_1^{n/2} \exp(-0.5 \sum_{i=1}^{n} (X_i - \mu)^2 \tau_1)$$

$$= \tau_1^{n/2+\alpha-1} \exp(-(\beta + 0.5 \sum_{i=1}^{n} (X_i - \mu)^2)\tau_1)$$

Which is proportional to a Gamma$(n/2 + \alpha, \beta + 0.5ns_0^2)$ distribution, where (for simplicity of notations) $s_0^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$. The uninformative prior is obtained by setting α and β equal to zero, so the posterior is then Gamma$(n/2, 0.5ns_0^2)$.

4. Let $\pi(X \mid \lambda) = \text{Poisson}(\lambda) = \exp(-\lambda)\lambda^X/X!$ and $\pi(\lambda) = \text{Gamma}(0.1, 0.1)$. Write a Gibbs sampler algorithm for sampling the joint distribution of X, λ.

The Gibbs-sampler is based on sampling sequentially from $\pi(X \mid \lambda)$ and $\pi(\lambda \mid X)$, which should be based on the joint distribution $\pi(X, \lambda)$ (which should be a proper distribution). The joint distribution can be written either as $\pi(X \mid \lambda)\pi(\lambda)$ or as $\pi(\lambda \mid X)\pi(X)$. In the former case, it is clear from the question that $\pi(X \mid \lambda)$ is Poisson(λ). In the latter case, we need $\pi(\lambda \mid X)$ which is the same as posterior distribution of λ, given X. With the Poisson model and (conjugate) Gamma-prior, we can solve the posterior as Gamma$(X + \alpha, 1 + \beta)$, (by calculating from Bayes formula or simply remembering this from the notes), where $\alpha = \beta = 0.1$. The Gibbs sampler is initiated by selecting starting values X_0, λ_0 and then we keep sampling in a sequence X_{i+1} from Poisson(λ_i) and λ_{i+1} from Gamma$(X_{i+1} + \alpha, 1 + \beta)$, for large number of iterations $i = 1, 2, 3, \ldots$. [Note, that the parameters α, β need to be positive, otherwise the joint distribution is not proper].

5. Assume diagnostic test is applied to $N = 300$ patients and $X = 49$ of them are detected as positive. An informative prior distribution Beta$(90, 10)$ is chosen for the sensitivity of the test $q_1 = P(\text{test '+'} \mid \text{disease})$, and Beta$(99, 1)$ for specificity $q_2 = P(\text{test '-'} \mid \text{no disease})$. Population prevalence of the disease is p, and uniform prior U$(0, 1)$ is chosen for this. Write a BUGS code for computing posterior distribution $\pi(p, q_1, q_2 \mid N, X)$ and consequently the posterior distribution for the expected proportion of test negatives $\psi = p(1 - q_1) + (1 - p)q_2$.

Model code (there can be several variations of this):

```
model{
X ~ dbin(pr,N)
pr <- p*q[1]+(1-p)*(1-q[2])
q[1] ~  dbeta(90,10)
q[2] ~  dbeta(99,1)
p ~ dunif(0,1)
psi <- p*(1-q[1])+(1-p)*q[2]
}
list(X=49,N=300)
```