MATEMATIIKAN JA TILASTOTIETEEN LAITOS
Bayes-päättely
Loppukoe 15.5.2012

1. Explain the following terms or concepts by using an example:
a) Elicitation of prior probability.
Transforming expert's knowledge into the form of a numerical probability or distribution. Personal probabilities P(A) for some statement "A"could be elicited e.g. by offering a choice between a lottery ticket with known probability of win - and a similar win if "A"proves to be true. The expert would prefer the choice that he/she believes to be more likely winning. Therefore, by adjusting the known lottery probabilities we can find out what is a value for which both choices become equally attractive. This gives a numerical value for the personal probability P(A). Personal probability distributions are harder to elicit because several numbers need to be specified, starting e.g. with a mode and some percentiles (e.g. 95%). Also psychological effects can bias the elicitation process and care should be taken to avoid typical biases. Elicitation should result into a fair probabilistic representation of the experts actual rational beliefs that are based on his/her relevant knowledge.

b) Uninformative prior.
A prior distribution that represents 'as little information as possible'. Typical example is uniform prior over a range of all possible values. However, it is not possible to define universally completely uninformative priors because all distributions are somewhat informative in some aspects. Typical example of this is to calculate some parameter transformation, e.g. $g(\theta) = \theta^2$ when θ has uniform prior. The transformation θ^2 obviously does not have uniform distribution. Jeffreys' prior addresses this problem but has some other disadvantages.

c) Law of total expectation and total variance.
$E(X) = E(E(X \mid Y))$ and $V(X) = V(E(X \mid Y)) + E(V(X \mid Y))$. These are useful for computing means and variances for e.g. posterior (or prior) predictive distributions where the conditional means and conditional variances are easy to know. Example: $X \sim \text{Bin}(n, p)$ with $p \sim \text{Beta}(\alpha, \beta)$ leads to $E(X) = E(E(X \mid n, p)) = E(np) = nE(p) = n\alpha/(\alpha + \beta)$, etc.

d) Normal approximation of posterior distribution.
Way of approximating posterior distribution by a normal density. For example, if posterior mean and variance can be known, then the approximation could be simply N(mean,var). Another option would be to use posterior mode $\hat{\theta}$ so that the approximation is $\text{N}(\hat{\theta}, I(\hat{\theta})^{-1})$ where I is 'observed information'.

e) Marginal posterior distribution.
When the posterior distribution is a joint distribution of several parameters, e.g. $\pi(\theta_1, \theta_2 \mid \text{data})$, the marginal is e.g. $\pi(\theta_1 \mid \text{data}) = \int \pi(\theta_1, \theta_2 \mid \text{data})\mathbf{d}\theta_2$.

f) Burn-in period.
The initial run period of an MCMC sampler, during which the sampler has not converged to the target distribution. Therefore, the sampled values from this period do not represent the target distribution and should be thrown away from final analysis.

2. In a tea-tasting experiment, tea is served in 10 cups. In some cups, milk was added before the tea, but in some cups milk was added after the tea was poured in. Mr 'A' is a very experienced tea taster and he claims to be able to tell which was the case in each cup. After he tasted, he got the right answer with all 10 cups! Mr 'B' claims to have *extra sensory perception*. He says he can tell the difference even without touching any of the cups. He also got the right answer for all 10 cups! Let θ_A and θ_B be the probability of a correct answer for Mr 'A' and Mr 'B'. Should the conclusion from Bayesian inference about θ_A and θ_B be the same for both? If yes, explain why. If not, explain why.

In both cases, the likelihood function (probability of observations) is the same: $\theta^{10}(1-\theta)^0$. Therefore, estimates based on these data only would be necessarily the same for both θ_A and θ_B. We should then conclude the same for both. It would seem that the best estimate is that both have the ability to be correct with 100%. This would be the best estimate from a posterior distribution if the prior was uniform, because the posterior density is: $\pi(\theta \mid \text{data}) = \theta^{10}(1-\theta)^0 \times 1 \times$ constant which is simply proportional to the likelihood function, if prior is uniform. With this prior, the conclusion is that Mr A and Mr B are equally good. However, this may not be a credible answer because we also know that Mr A used some knowledge and skills which could be relevant and helpful whereas Mr B claims to have supernatural powers. It is the question of how much we believe in such powers compared to practical tasting experience. Based on our background information about the plausibility of supernatural powers, our prior $\pi(\theta_B)$ should perhaps be very much shifted towards $\theta_B = 0.5$ (=random guessing), instead of a uniform distribution. With this background information included in the prior, the conclusions would be very different, regardless of the same data in this isolated and rather small experiment. This is an example where all the relevant information is not included in the plain data. Our background knowledge also counts when analyzing the problem instead of analyzing the limited data only.

3. Assume observed data $X_1, \ldots, X_n$ are exponentially distributed $X_i \sim \exp(\lambda)$. Use a conjugate prior and solve the posterior distribution $\pi(\lambda \mid X_1, \ldots, X_n)$.

Density function of exponential distribution is $\pi(X_i \mid \lambda) = \lambda \exp(-\lambda X_i)$ for $X_i > 0$. Therefore, probability of the observed data (likelihood function) is $\prod_{i=1}^{n} \pi(X_i \mid \lambda) = \lambda^n \exp(-\lambda n \bar{X})$, where $\bar{X} = \frac{1}{n} \sum X_i$. We calculate posterior distribution $\pi(\lambda \mid X_1, \ldots, X_n) = \text{const} \times \lambda^n \exp(-\lambda n \bar{X}) \times \pi(\lambda)$. A conjugate prior is gamma-density $\pi(\lambda) = \text{const} \times \lambda^{\alpha-1} \exp(-\beta\lambda)$. Therefore, posterior is of the form: const $\times \lambda^{n+\alpha-1}$ $\exp(-(\beta + n\bar{X})\lambda)$, which is $\text{Gamma}(n+\alpha, \beta + n\bar{X})$.

4. Write a Gibbs sampler algorithm to sample from a uniform density in a triangle defined within a unit square, $(x, y) \in [0,1] \times [0,1]$, where the density function is $\pi(x,y) = 2\ 1_{\{0<x<1, 0<y<1-x\}}(x,y)$. Write also an alternative Monte Carlo sampler (which is not MCMC).

A Gibbs sampler is to sample repeatedly from the two conditional distributions: $\pi(x \mid y) = \text{U}(0, 1-y)$ and $\pi(y \mid x) = \text{U}(0, 1-x)$. Alternatively, we could sample independent random variables from the unit square and accept only those pairs (x, y) that fall within the triangle, as our final sample. (This would be a 'rejection sampler'). Alternatively, we could repeat sampling the first variable, e.g. x, from U(0,1), and the second from $\text{U}(0, 1-x)$. (This would be independent Monte Carlo sampling, because each pair x, y is independent of others).

5. Let $X_1, \ldots, X_n$ be conditionally independent normal variables $X_i \sim \text{N}(\mu, \sigma^2)$. The priors are $\mu \sim \text{N}(\mu_0, \sigma_0^2)$ and $\sigma \sim \text{Gamma}(\gamma\theta, \theta)$. Assuming the sum $Y = \sum X_i$ is observed as data, and $\mu_0, \sigma_0, \gamma, \theta$ are given constants, write a BUGS model code for computing the posterior distribution $\pi(\mu, \sigma \mid Y)$.

```
model{
Y ~ dnorm(m,tau)
m <- mu*n; tau <- 1/(n*sigma*sigma)
mu ~ dnorm(mu0,tau0); tau0 <- 1/(sigma0*sigma0)
sigma ~ dgamma(a,b); a <- gamma*theta; b <- theta
}
list(Y=...,mu0=...,sigma0=...,gamma=...,theta=...,n=...)
# example data and fixed values
```

The model has two unknown parameters μ and σ, but only one data point Y, so the model is not identifiable. Therefore, informative priors would be needed. (But this was not asked. Extra points given for noting this).