Summary of main results

Bayesiansk statistik
MS2505

Bruna G. Palm

1. Beta(α, β) distribution:

$$
\begin{aligned}
f_Y(y) =& \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-\theta)^{\beta-1} \\
\mathrm{E}(Y) =& \frac{\alpha}{\alpha+\beta} \\
var(Y) =& \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}
\end{aligned}
$$

2. mean and variance of conditional distributions

$$
\begin{aligned}
\mathrm{E}(u) =& \mathrm{E}(\mathrm{E}(u|v)) = \int\int up(u,v)dudv = \int\int up(u|v)dup(v) = \int \mathrm{E}(u|v)p(v)dv \\
var(u) =& 4\,\mathrm{E}(var(u|v)) + var(\mathrm{E}(u|v))
\end{aligned}
$$

3. Bayes' rule:

$$
p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta}
$$

4. The unnormalized posterior density:

$$
p(\theta|y) \propto p(\theta)p(y|\theta)
$$

5. Normalization term Z (constant given y):

$$
Z = \int_0^1 \theta^y(1-\theta)^{n-y}d\theta = \frac{\Gamma(y+1)\Gamma(n-y+1)}{\Gamma(n+2)}
$$

6. Binomial with Uniform prior
   - Posterior is

   $$
   p(\theta|y,n,M) = \frac{\Gamma(n+2)}{\Gamma(y+1)\Gamma(n-y+1)}\theta^y(1-\theta)^{n-y},
   $$

   which is called Beta distribution

   $$
   \theta|y,n \sim \mathrm{Beta}(y+1, n-y+1)
   $$

   - Expected value

   $$
   \mathrm{E}[\theta|y] = \frac{y+1}{n+2}
   $$

   - Prior predictive distribution for new $\tilde{y}$ (discrete)

   $$
   p(\tilde{y}=1|M) = \int_0^1 p(\tilde{y}=1|\theta,y,n,M)p(\theta|M)d\theta
   $$

   - Posterior predictive distribution for new $\tilde{y}$ (discrete)

   $$
   p(\tilde{y}=1|y,n,M) = \int_0^1 p(\tilde{y}=1|\theta,y,n,M)p(\theta|y,n,M)d\theta
   $$

7. Binomial with Beta prior
   - Prior

   $$
   \mathrm{Beta}(\theta|\alpha,\beta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}
   $$

- Posterior

$$p(\theta|y,n,M) \propto \theta^{y}(1-\theta)^{n-y}\theta^{\alpha-1}(1-\theta)^{\beta-1}$$
$$\propto \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}$$

after normalization

$$p(\theta|y,n,M) = \mathrm{Beta}(\theta|\alpha+y,\beta+n-y)$$

- Posterior mean

$$\mathrm{E}[\theta|y] = \frac{\alpha+y}{\alpha+\beta+n}$$

- Posterior variance

$$\mathrm{Var}[\theta|y] = \frac{\mathrm{E}[\theta|y](1-\mathrm{E}[\theta|y])}{\alpha+\beta+n+1}$$

8. Normal distribution with conjugate prior (Normal) for θ

- Assume σ^2 known
  - Likelihood

$$p(y|\theta) \propto \exp\left(-\frac{1}{2\sigma^2}(y-\theta)^2\right)$$

  - Prior

$$p(\theta) \propto \exp\left(-\frac{1}{2\tau_0^2}(\theta-\mu_0)^2\right)$$

  - Posterior

$$p(\theta|y) \propto \exp\left(-\frac{1}{2}\left[\frac{(y-\theta)^2}{\sigma^2}+\frac{(\theta-\mu_0)^2}{\tau_0^2}\right]\right)$$
$$\propto \exp\left(-\frac{1}{2\tau_1^2}(\theta-\mu_1)^2\right)$$
$$\theta|y \sim \mathrm{N}(\mu_1,\tau_1^2), \quad \text{where} \quad \mu_1 = \frac{\frac{1}{\tau_0^2}\mu_0+\frac{1}{\sigma^2}y}{\frac{1}{\tau_0^2}+\frac{1}{\sigma^2}} \quad \text{and} \quad \frac{1}{\tau_1^2} = \frac{1}{\tau_0^2}+\frac{1}{\sigma^2}$$

  - Several observations $y = (y_1,\ldots,y_n)$

$$p(\theta|y) = \mathrm{N}(\theta|\mu_n,\tau_n^2)$$
$$\text{where} \quad \mu_n = \frac{\frac{1}{\tau_0^2}\mu_0+\frac{n}{\sigma^2}\bar{y}}{\frac{1}{\tau_0^2}+\frac{n}{\sigma^2}} \quad \text{and} \quad \frac{1}{\tau_n^2} = \frac{1}{\tau_0^2}+\frac{n}{\sigma^2}$$

  - Posterior predictive distribution

$$p(\tilde{y}|y) = \int p(\tilde{y}|\theta)p(\theta|y)d\theta$$
$$p(\tilde{y}|y) \propto \int \exp\left(-\frac{1}{2\sigma^2}(\tilde{y}-\theta)^2\right)\exp\left(-\frac{1}{2\tau_1^2}(\theta-\mu_1)^2\right)d\theta$$

$$\tilde{y}|y \sim \mathrm{N}(\mu_1,\sigma^2+\tau_1^2), \quad \text{where} \quad \tau_1^2 \quad \text{is the posterior variance}$$

9. Monte Carlo and posterior draws

$$E_{p(\theta|y)}[\theta] = \int \theta p(\theta \mid y) \approx \frac{1}{S}\sum_{s=1}^{S}\theta^{(s)}$$

10. Joint distribution of parameters

$$p(\theta_1,\theta_2 \mid y) \propto p(y \mid \theta_1,\theta_2)p(\theta_1,\theta_2)$$

11. Marginalization

$$p(\theta_1 \mid y) = \int p(\theta_1,\theta_2 \mid y)d\theta_2$$

$p(\theta_1 \mid y)$ is a marginal distribution

12. Monte Carlo approximation

$$p(\theta_1 \mid y) \approx \frac{1}{S}\sum_{s=1}^{S}p(\theta_1,\theta_2^{(s)} \mid y),$$

where $\theta_2^{(s)}$ are draws from $p(\theta_2 \mid y)$

13. Marginalization over posterior distribution

$$
\begin{aligned}
p(\tilde{y} \mid y) &= \int p(\tilde{y} \mid \theta) p(\theta \mid y) d\theta \\
&= \int p(\tilde{y}, \theta \mid y) d\theta
\end{aligned}
$$

$p(\tilde{y} \mid y)$ is a predictive distribution

14. Gaussian non-informative prior (Uniform)

$$
\begin{aligned}
p(\mu, \sigma^2) &\propto \sigma^{-2} \\
p(\mu, \sigma^2 \mid y) &\propto \sigma^{-2} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y_i - \mu)^2\right) \\
p(\mu, \sigma^2 \mid y) &\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2\right) \\
&= \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} \left[\sum_{i=1}^{n} (y_i - \bar{y})^2 + n(\bar{y} - \mu)^2\right]\right) \\
&= \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} \left[(n-1)s^2 + n(\bar{y} - \mu)^2\right]\right)
\end{aligned}
$$

- Marginal posterior $p(\sigma^2 \mid y)$
  - Known mean

$$
\begin{aligned}
&\sigma^2 \mid y \sim \mathrm{Inv} - \chi^2(n, v) \\
&\text{where} \quad v = \frac{1}{n} \sum_{i=1}^{n} (y_i - \theta)^2
\end{aligned}
$$

  - Unknown mean

$$
\begin{aligned}
&\sigma^2 \mid y \sim \mathrm{Inv} - \chi^2(n-1, s^2) \\
&\text{where} \quad s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2
\end{aligned}
$$

- Marginal posterior $p(\mu \mid y)$

$$
\begin{aligned}
p(\mu \mid y) &\propto \left[1 + \frac{n(\mu - \bar{y})^2}{(n-1)s^2}\right]^{-n/2} \\
p(\mu \mid y) &= t_{n-1}(\mu \mid \bar{y}, s^2/n)
\end{aligned}
$$

- Posterior predictive distribution given known variance

$$
\begin{aligned}
p(\tilde{y} \mid \sigma^2, y) &= \int p(\tilde{y} \mid \mu, \sigma^2) p(\mu \mid \sigma^2, y) d\mu \\
&= \int \mathrm{N}(\tilde{y} \mid \mu, \sigma^2) \, \mathrm{N}(\mu \mid \bar{y}, \sigma^2/n) d\mu \\
&= \mathrm{N}(\tilde{y} \mid \bar{y}, (1 + \tfrac{1}{n})\sigma^2)
\end{aligned}
$$

- Posterior predictive distribution given unknown variance

$$
p(\tilde{y} \mid y) = t_{n-1}(\tilde{y} \mid \bar{y}, (1 + \tfrac{1}{n})s^2)
$$

15. Gaussian conjugate prior

- Conjugate prior has to have a form $p(\sigma^2)p(\mu \mid \sigma^2)$

$$
\begin{aligned}
\mu \mid \sigma^2 &\sim \mathrm{N}(\mu_0, \sigma^2/\kappa_0) \\
\sigma^2 &\sim \mathrm{Inv} - \chi^2(\nu_0, \sigma_0^2) \\
p(\mu, \sigma^2) &= \mathrm{N\,Inv} - \chi^2(\mu_0, \sigma_0^2/\kappa_0; \nu_0, \sigma_0^2)
\end{aligned}
$$

- Joint posterior

$$
\begin{aligned}
p(\mu, \sigma^2 \mid y) \approx &\sigma^{-1}(\sigma^2)^{(-\nu_0/2+1)} \exp\{\frac{-1}{2\sigma^2}[\nu_0\sigma_0^2 + \kappa_0(\mu - \mu_0)^2]\} \\
&\times (\sigma^2)^{-n/2} \exp\{\frac{-1}{2\sigma^2}[(n-1)s^2 + n(\bar{y} - \mu)^2]\} \\
p(\mu, \sigma^2 \mid y) = &\mathrm{N\,Inv} - \chi^2(\mu_n, \sigma_n^2/\kappa_n; \nu_n, \sigma_n^2)
\end{aligned}
$$

where

$$\begin{aligned}
\mu_n &= \frac{\kappa_0}{\kappa_0 + n}\mu_0 + \frac{n}{\kappa_0 + n}\bar{y} \\
\kappa_n &= \kappa_0 + n \\
\nu_n &= \nu_0 + n \\
\nu_n\sigma_n^2 &= \nu_0\sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_0 + n}(\bar{y} - \mu_0)^2
\end{aligned}$$

16. Normal approximation (large sample) Normal approximation

$$p(\theta|y) \approx \mathrm{N}(\hat{\theta}, [I(\hat{\theta})]^{-1})$$

where

$$I(\theta) = -\frac{d^2}{d\theta^2}\log p(\theta|y)$$

17. Posterior odds

$$\frac{p(\theta_1|y)}{p(\theta_2|y)} = \frac{p(\theta_1)p(y|\theta_1)}{p(\theta_2)p(y|\theta_2)}$$