Exercises
Applied Bayesian Analysis and Numerical Methods
(STK 9021)

Boris Simovski

29.09.2016

Exercise 3: The multivariate normal distribution

The bivariate Normal distribution for two dependent variables should be known. If x_1 has mean μ_1 variance σ_1^2, and x_2 has mean μ_2, variance σ_2^2, and the correlation between x_1 and x_2 is ρ, then(x_1, x_2) has the bivariate normal distribution if the joint density of x1 and x2 is

$$f(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{ -\frac{\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}}{2(1-\rho^2)} \right\} \tag{1}$$

Now let $x = (x_1, x_2)^T$, $\mu = (\mu_1, \mu_2)^T$ and

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

Show that we can write

$$f(x) = f(x_1, x_1) = \frac{1}{\sqrt{(2\pi)^2|\Sigma|}} \exp\left\{ -\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) \right\}$$

First we calculate the determinant of Σ

$$\left|\Sigma\right| = \sigma_1^2\sigma_2^2 - \rho^2\sigma_1^2\sigma_2^2 = \sigma_1^2\sigma_2^2\left(1-\rho^2\right)$$

We can rewrite (1) now.

For clarity, let $f(x_1, x_2) = A \exp\{B\}$ where $A = 1/2\pi\sigma_1\sigma_2\sqrt{(1-\rho^2)}$
and B is the term inside the exp{} part of (1)
Then, A can be rewritten as

$$A = \frac{1}{2\pi\sqrt{\sigma_1^2\sigma_2^2(1-\rho^2)}} = \frac{1}{2\pi\sqrt{\left|\Sigma\right|}}$$

And the inverse of Σ is

$$\Sigma^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)}\begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix} = \frac{1}{1-\rho^2}\begin{bmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2} \\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{bmatrix}$$

Next we rewrite B

$$B=-\frac{1}{2(1-\rho^2)}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2}-2\frac{\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)=$$

$$=-\frac{1}{2(1-\rho^2)}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2}-\frac{\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}-\frac{\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}\right)=$$

$$=-\frac{1}{2(1-\rho^2)}\left((x_1-\mu_1)\left(\frac{(x_1-\mu_1)}{\sigma_1^2}-\frac{\rho(x_2-\mu_2)}{\sigma_1\sigma_2}\right)+(x_2-\mu_2)\left(\frac{(x_2-\mu_2)}{\sigma_2^2}-\frac{\rho(x_1-\mu_1)}{\sigma_1\sigma_2}\right)\right)=$$

$$=-\frac{1}{2(1-\rho^2)}\left[\begin{matrix}\frac{(x_1-\mu_1)}{\sigma_1^2}-\frac{\rho(x_2-\mu_2)}{\sigma_1\sigma_2} & -\frac{\rho(x_1-\mu_1)}{\sigma_1\sigma_2}+\frac{(x_2-\mu_2)}{\sigma_2^2}\end{matrix}\right]\left[\begin{matrix}(x_1-\mu_1)\\(x_2-\mu_2)\end{matrix}\right]=$$

$$= -\frac{1}{2(1-\rho^2)} \begin{bmatrix} x_1 - \mu_1 & x_2 - \mu_2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_1^2} & \frac{-\rho}{\sigma_1 \sigma_2} \\ \frac{-\rho}{\sigma_1 \sigma_2} & \frac{1}{\sigma_2^2} \end{bmatrix} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix} =$$

$$= -\frac{1}{2} (\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)$$

$$f(\mathbf{x}) = f(x_1, x_2) = A \exp(B) = \frac{1}{2\pi \sqrt{|\Sigma|}} \exp\left(-\frac{1}{2} (\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu) \right)$$

Exercise 2.10

Discrete sample spaces: suppose there are N cable cars in San Francisco, numbered sequentially from 1 to N. You see cable car at random; it is numbered 203. You wish to estimate N.

a) Assume your prior distribution on N is geometric with mean 100. What is your posterior distribution for N?

$$p(N) = (1/100)(99/100)^{N-1}$$

Since we observed car 203 we can only conclude that there are at least 203 cars. For any N>202 we can assume it has the same probability to be the number of the last car. So the likelihood is given by:

$$p(y \mid N) = \frac{1}{N}$$ for N>202, else 0

$$p(N \mid y) \propto p(N)p(y \mid N) = \frac{1}{N}(0.01)(0.99)^{N-1}$$ For N>202

$$\propto \frac{1}{N}(0.99)^{N}$$ For N>202

b) What are the posterior mean and standard deviation of N?

In order to be a proper probability distribution the posterior probability needs to sum to 1. We need to find a normalizing constant such that:

$$\sum_{N=203}^{\infty} p(N \mid y) = \sum_{N=203}^{\infty} C \frac{1}{N} (0.99)^N = 1$$

$$C = \frac{1}{\sum_{N=203}^{\infty} \frac{1}{N} (0.99)^N}$$

We will approximate numerically the series in the denominator with an error err < 10^{-7}. We iteratively sum the elements in the series and for each N we calculate an upper limit of the error with the formula:

$$err = \sum_{N=n}^{\infty} \frac{1}{N} (0.99)^N < \frac{1}{n+1} \sum_{N=n+1}^{\infty} (0.99)^N = \frac{1}{n+1} \frac{(0.99)^{n+1}}{1-0.99}$$

The calculation is derived from the formula for finite and infinite geometric series:

$$\sum_{k=n+1}^{\infty} ar^k = \sum_{k=0}^{\infty} ar^k - \sum_{k=0}^{n} ar^k = \frac{a}{1-r} - \frac{a - ar^{n+1}}{1-r} = \frac{ar^{n+1}}{1-r}$$

The python script that does the calculation is:

```
s=0
n=203
while True:
    s=s+(1.0/n)*(0.99**n)
    err = 1.0/(n+1)*((0.99**(n+1))/(1-0.99))
    if err < 0.0000001:
        print "n: ", n
        print "s: ", s
        break
    n = n+1
```

Resulting in the output:

```
n:  1345
s:  0.0465802355607
```

C=1/0.0465802355607

C=21.46833

The expected value then is:

$$E(N / y) = \sum_{N=203}^{\infty} Np(N \mid y) = 21.46833 \sum_{N=203}^{\infty} (0.99)^N$$

$$= 21.46833 \frac{0.99^{203}}{1 - 0.99} = 279.089$$

For the standard deviation we calculate the variance first:

$$\operatorname{var}(N \mid y) = \sum_{N=203}^{\infty} (N - 279.089)^2 \frac{C}{N} (0.99)^N$$

$$\approx \sum_{N=203}^{1345} (N - 279.089)^2 \frac{21.46833}{N} (0.99)^N = 6391.6271$$

$$sd(N \mid y) = \sqrt{\operatorname{var}(N \mid y)} = 79.9477$$

Additionally to the previous observation y1 = 203, we also observe y2 = 157 and y3 = 222

$$p(y_2 \mid N) = \frac{1}{N}$$ for N>156 else 0

$$p(N \mid y_1, y_2) \propto p(N \mid y_1) p(y_2 \mid N) =$$
$$= \frac{1}{N^2}(0.99)^N$$ For N>202 else 0

$$p(y_3 \mid N) = \frac{1}{N}$$ for N>221 else 0

$$p(N \mid y_1, y_2, y_3) \propto p(N \mid y_1, y_2) p(y_3 \mid N) =$$
$$= \frac{1}{N^3}(0.99)^N$$ For N>221 else 0

Output from the python script:
n: 330
s: 4.08361192509e-07

C_3=2448812.5178

$E(N|y_1, y_2, y_3) = 255.7372$

$Var(N|y_1, y_2, y_3) = 770.3968$

$sd(N|y_1, y_2, y_3) = 27.756$

Exercise 3.7 Poison and binomial distributions: a student sits on a street corner for an hour and records the number of bicycles b and the number of other vehicles v that go by. Two models are considered:

- The outcomes b and v have independent Poisson distributions, with unknown means θ_b and θ_v.
- The outcome b has a binomial distribution, with unknown probability p and sample size b + v.

Show that the two models have the same likelihood if we define $p = \theta_b/(\theta_b+\theta_v)$.

Model 1:

$$p(b=y \mid \theta_b,\theta_v,b+v=n) = \frac{p(b=y,b+v=n \mid \theta_b,\theta_v)}{p(b+v=n \mid \theta_b,\theta_v)}$$

$$= \frac{p(b=y \mid \theta_b,\theta_v)p(b+v=n \mid \theta_b,\theta_v)}{p(b+v=n \mid \theta_b,\theta_v)} = \frac{p(b=y \mid \theta_b,\theta_v)p(v=n-b \mid \theta_b,\theta_v)}{p(b+v=n \mid \theta_b,\theta_v)}$$

Since the sum of two independent Poisson distributed variables is also Poisson distributed, i.e. $b+v \sim Po(\theta_b+\theta_v)$, we have

$$= \frac{\dfrac{\theta_b^y e^{-\theta_b}}{y!} \dfrac{\theta_v^{n-y} e^{-\theta_v}}{(n-y)!}}{\dfrac{\left(\theta_b + \theta_v\right)^n e^{-(\theta_b+\theta_v)}}{n!}} = \frac{n!}{y!(n-y)!} \frac{\theta_b^y \theta_v^{n-y}}{\left(\theta_b + \theta_v\right)^n}$$

Model 2:

$$p(b \mid p, b+v) = \begin{pmatrix} b+v \\ b \end{pmatrix} p^b \left(1-p\right)^v =$$

$$= \frac{(b+v)!}{b!v!} \left(\frac{\theta_b}{\theta_b + \theta_v}\right)^b \left(1 - \frac{\theta_b}{\theta_b + \theta_v}\right)^v = \frac{(b+v)!}{b!v!} \left(\frac{\theta_b}{\theta_b + \theta_v}\right)^b \left(\frac{\theta_b + \theta_v - \theta_b}{\theta_b + \theta_v}\right)^v$$

$$= \frac{(b+v)!}{b!v!} \left(\frac{\theta_b}{\theta_b + \theta_v}\right)^b \left(\frac{\theta_v}{\theta_b + \theta_v}\right)^v = \frac{(b+v)!}{b!v!} \frac{\theta_b^b}{\left(\theta_b + \theta_v\right)^b} \frac{\theta_v^v}{\left(\theta_b + \theta_v\right)^v} =$$

$$= \frac{(b+v)!}{b!v!} \frac{\theta_b^b \theta_v^v}{\left(\theta_b + \theta_v\right)^{b+v}}$$

Exercise 3.15: Joint distributions: The autoregressive time-series model $y_1, y_2, \ldots$ with mean level 0, autocorrelation 0.8, residual standard deviation 1, and normal errors can be written as $(y_t | y_{t-1}, y_{t-2}, \ldots) \sim N(0.8y_{t-1}, 1)$ for all t.

(a) Prove that the distribution of y_t, given the observations at all other integer time points t, depends only on y_{t-1} and y_{t+1}.

(b) What is the distribution of y_t given y_{t-1} and y_{t+1}?

3.15

$$P(y_t \mid y_1, y_2 \ldots) =$$

$$= \frac{P(y_1, y_2, \ldots, y_t, \ldots)}{\int P(y_1, y_2, \ldots, y_t, \ldots)\, dy_t} = (*)$$

From the given condition that

$$P(y_t \mid y_{t-1}, \ldots, y_1) = N(0.8\, y_{t-1}, 1)$$

we see that $P(y_t \mid y_{t-1}, \ldots, y_1) = P(y_t \mid y_{t-1})$

$$(*) = \frac{P(y_1)\,P(y_2|y_1)\,P(y_3|y_1,y_2)\dots P(y_t|y_{t-1},\dots y_1)\dots}{\int P(y_1)\,P(y_2|y_1)\dots P(y_t|y_{t-1},\dots,y_1)\dots\,dy_t} =$$

$$= \frac{P(y_1)\,P(y_2|y_1)\,P(y_3|y_2)\dots P(y_t|y_{t-1})\dots}{P(y_1)\,P(y_2|y)\dots P(y_{t-1}|y_{t-2})\,P(y_{t+2}|y_{t+1})\dots \int P(y_{t+1}|y_t)\cdot P(y_t|y_{t-1})\,dy_t}$$

$$= \frac{P(y_t|y_{t-1})\,P(y_{t+1}|y_t)}{\int P(y_t|y_{t-1})\,P(y_{t+1}|y_t)\,dy_t}$$