2017. 08. 04

$23.8b = 22.17778 \cdot b^2$

$b = \frac{23.8}{22.17778} = 1.073146$

$a = 25.54087$

1 Exercise - Bayesian modelling of discrete data

The table below gives the number of fatal accidents and deaths on scheduled airline flights per year over a ten-year period.

① "uninformative prior"

prior not very informative,

Gamma(1,1) ⇒ proper

Gamma(1,0) ⇒ improper

dominated by likelihood

$a+y = 1+238 = 239$

$b+n = 1+10 = 11$

(19.05979, 24.56691)

Year	Fatal accidents	Passenger Deaths	Death Rate
1976	24	734	0.19
1977	25	516	0.12
1978	31	754	0.15
1979	31	877	0.16
1980	22	814	0.14
1981	21	362	0.06
1982	26	764	0.13
1983	20	809	0.13
1984	16	223	0.03
1985	22	1066	0.15

qgamma(c(0.025, 0.975), a+y, b+n)

② another way. ⇒ this way (21, 26.8)

"Empirical Bayes"

$E(\theta) = \frac{a}{b} = \bar{y} = 23.8$

$Var(\theta) = \frac{a}{b^2} = s^2 = 22.17778$

solve for a & b.

use them as hyperparameters

plug back into Gamma( )

(Death rate is passenger deaths per 100 million passenger miles).

(a) Assume that the numbers of fatal accidents in each year are independent with a Poisson(θ) distribution. Set a prior distribution for θ and determine the posterior distribution based on the data from 1976 through 1985. Under this model, give a 95% predictive interval for the number of fatal accidents in 1986.

$Y_i \sim Pois(x_i\theta)$ x_i: exposure variable

likelihood.

now

$p(y_i|\theta) = \frac{(x_i\theta)^{y} e^{-x_i\theta}}{y!}$

$\propto \theta^{y_i} e^{-\theta x_i}$

$\theta \propto \theta^{a-1} e^{-b\theta}$

(n data pts)

⇓

posterior:

$Gamma(a+\sum y_i, b+\sum x_i)$

instead of n.

(b) Assume that the numbers of fatal accidents in each year follow independent Poisson distributions with a constant rate and an exposure in each year proportional to the number of passenger miles flown. Set a prior distribution for θ and determine the posterior distribution based on the data 1976-1985. (Estimate the number of passenger miles flown in each year by dividing the appropriate columns of the table above and ignoring round-off errors). Give a 95% predictive interval for the number of fatal accidents in 1986 under the assumption that 8×10^{11} passenger miles are flown in that year.

(c) Repeat (a) above, replacing 'fatal accidents' with 'passenger deaths'.

(d) Repeat (b) above, replacing 'fatal accidents' with 'passenger deaths'.

(e) In which of the cases (a)-(d) above does the Poisson model seem more or less reasonable? Why? Discuss based on general principles, without specific reference to the numbers in the table.

probably not.

b/c multiple deaths per plane crash.

~~not~~ violating the assumption of independence.

(a) will be ideal one.