Outline of the chapter 2

- 2.1 Binomial model (e.g. biased coin flipping)
- 2.2 Posterior as compromise between data and prior information
- 2.3 Posterior summaries
- 2.4 Informative prior distributions (skip exponential families and sufficient statistics)
- 2.5 Gaussian model with known variance
- 2.6 Other single parameter models
  - the normal distribution with known mean but unknwon variance is the most important
  - glance through Poisson and exponential
- 2.7 glance through this example, which illustrates benefits of prior information, no need to read all the details (it's quite long example)
- 2.8 Noninformative and weakly informative priors

Binomial: unknwon θ

- Bayes rule

$$p(\theta|y,n,M) = \frac{p(y|\theta,n,M)p(\theta|n,M)}{p(y|n,M)}$$

- Start with uniform prior

$$p(\theta|n,M) = p(\theta|M) = 1, \text{ kun } 0 \leq \theta \leq 1$$

- Then

$$p(\theta|y,n,M) = \frac{p(y|\theta,n,M)}{p(y|n,M)} = \frac{\binom{n}{y}\theta^y(1-\theta)^{n-y}}{\int_0^1 \binom{n}{y}\theta^y(1-\theta)^{n-y} d\theta}$$
$$= \frac{1}{Z}\theta^y(1-\theta)^{n-y}$$

Binomial: unknwon θ

- Normalization term Z

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

- Normalisation term has Beta function form
  - when integarted over $(0, 1)$ the result can presented with Gamma functions
  - with integers $\Gamma(n) = (n-1)!$
  - for large integers even this is challenging and usually $\log \Gamma(\cdot)$ is computed instead of $\Gamma(\cdot)$

Binomial: unknwon θ

- 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

Binomial: unknwon θ

- 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

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

Binomial: unknwon θ

- 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

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

disttool

Binomial: computation*

- Beta CDF not trivial to compute
- For example, Matlab uses a continued fraction expression, and if that does not converge another approximation is used
- Laplace developed normal approximation (Laplace approximation), because he didn't know how to compute Beta CDF

Justification for uniform prior*

- $p(\theta|M) = 1$ if
  - we want the prior predictive distribution to be uniform

$$p(y|n) = \frac{1}{n+1}, \quad y = 0, \dots, n$$

    - nice justification as it is based on observables y and n
  - we think all values of θ are equally likely

Priors

- Conjugate prior (BDA3 p. 35)
- Noninformative prior (BDA3 p. 51)
- Proper and improper prior (BDA3 p. 52)
- Weakly informative prior (BDA3 p. 55)
- Informative prior (BDA3 p. 55)
- Prior sensitivity (BDA3 p. 38)

Sufficient statistics

- The quantity $t(y)$ is said to be a *sufficient statistic* for θ, because the likelihood for θ depends on the data y only through the value of $t(y)$.

Effect of integration

- Binomial example

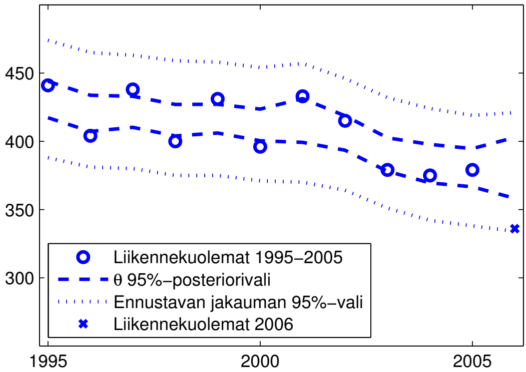
Central limit theorem*

- De Moivre, Laplace, Gauss, Chebysev, Liapounov, Markov, et al.
- Given certain conditions sum (and mean) of random variables approach Gaussian distribution as d $n \to \infty$
- Problems
  - does not hold for all distributions, e.g., Cauchy
  - may require large n,
    e.g. Binomial, when θ close to 0 or 1
  - does not hold if one the variables has much larger scale

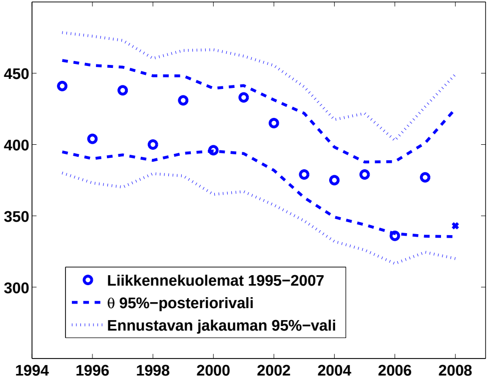
Poisson distribution
example

- Liikenneturva (Finnish traffic safety organization) reports about 400 traffic deaths per year
- In 2006 Suomen 336 traffic deaths
- Was year 2006 exceptional?
  - $p(y_{2006} \leq 336 | \theta = 400) \approx 10^{-4}$
  - $p(y_{2006} \leq 336 | y_{1995,\ldots,2005}, \text{constant risk}) \approx 4 \times 10^{-4}$

- Assuming that traffic safety changes slowly
  - time series model (Gaussian process prior for log risk in time)
    $p(y_{2006} \leq 336 | y_{1995,\ldots,2005}, \text{changing risk}) \approx 0.03$



- Add deaths of years 2006 and 2007
  - In year 2007 377 deaths
- In 2008 343 deaths
  - $p(y_{2008} \geq 343 | y_{1995,\ldots,2007}, \text{changing risk riski}) \approx 0.87$



Algae status is monitored in 274 sites at Finnish lakes and rivers. The observations for the 2008 algae status at each site are presented in file algae.mat ('0': no algae, '1': algae present). Let π be the probability of a monitoring site having detectable blue-green algae levels.

- Use a binomial model for observations and a beta(2,10) prior.
- What can you say about the value of the unknown π?
- Experiment how the result changes if you change the prior.

R demonstrations

- demo2_1: 437 girls and 543 boys have been observed. Calculate and plot the posterior distribution of the proportion of girls θ, using uniform prior on θ.
- demo2_2: Comparison of posterior distributions with different parameter values for the beta prior distribution.
- demo2_3: Simulating samples from Beta(438,544), drawing a histogram with quantiles, and doing the same for a transformed variable.
- demo2_4: Calculating the posterior distribution on a discrete grid of points (by multiplying the likelihood and a non-conjugate prior at each point, and normalizing over the points). Simulating samples from the resulting non-standard posterior distribution using the inverse-cdf method.