Why the decomposition?

- Objective reasons due to the **modelling of observations**. Examples are:

  - single- or multi-stage cluster samples studies (–> multilevel, random effects, mixed effects models);
  - meta-analysis studies;
  - repeated measurements or growth curve analyses

  —–> *The basic idea is that of **borrowing strength, pooling information** from related sources of data*

  *as well as the idea of **allowing for random effect variation** between units to account for data extra-variation or overdispersion*

Bayesian Methods

LABORATORY

Lesson 4: Feb 14 2002

Software: **BUGS and R**

Hierarchical Priors for Pooling Strength: a meta-analysis study

The general situation is as follows. Consider a set of survey units $j = 1, \ldots, J$, with unit j having data y_j modelled as $p(y_j|\theta_j)$ where θ_j is the parameter of interest.

The **2-stage hierarchical model** decomposed in

$$p(\mathbf{y}|\boldsymbol{\theta}) = \prod_{j=1}^{J} p(y_j|\theta_j) \quad (1),$$
$$p(\boldsymbol{\theta}|\boldsymbol{\phi}) = \prod_{j=1}^{J} p(\theta_j|\boldsymbol{\phi}) \quad (2)$$

and $p(\boldsymbol{\phi})$, **allows for jointly modelling the** J **experiments**.

Pooling over all the units requires the assumption of **exchangeability** ($\sim$ simmetry among the θ_j's at prior) to be satisfied.
The iid mixture for $\boldsymbol{\theta}$, $p(\boldsymbol{\theta}) = \int \prod_{j=1}^{J} p(\theta_j|\boldsymbol{\phi}) p(\boldsymbol{\phi}) d\boldsymbol{\phi}$, is an exchangeable distribution.

HIERARCHICAL *priors*

A Bayesian statistical model $(p(y|\theta_1), p(\theta_1))$ defines a **hierarchical Bayesian model** if the prior distribution $p(\theta_1)$ is decomposed in **conditional** distributions

$$p_1(\theta_1|\theta_2) \ldots p_{k-1}(\theta_{k-1}|\theta_k)$$

and a marginal distribution $p(\theta_k)$ such that

$$p(\theta_1) = \int_{\Theta_2 \times \ldots \times \Theta_k} p_1(\theta_1|\theta_2) \ldots p_{k-1}(\theta_{k-1}|\theta_k) p(\theta_k) d\theta_2 \ldots d\theta_k.$$

For the generic $p(\theta_i|\theta_{i+1})$, θ_i are the i-th **stage** parameters and θ_{i+1} are called **hyperparameters** of the i-th stage prior specification.

a look to the data

```
Results of 22 clinical trials of beta-blockers for reduc
mortality after myocardial infarction.  Data from Yusuf
(1985).  In Table 5.4 of Bayesian Data Analysis.

study control.deaths control.total treated.deaths treate
 1   3 39            3 38
 2   14 116         7 114
 3   11 93           5 69
 4   127 1520    102 1533
 5   27 365        28 355
 6   6 52            4 59
 7   152 939       98 945
 8   48 471        60 632
 9   37 282        25 278
 11   52 583        64 873
 12   47 266        45 263
 13   16 293         9 291
```


but also for ...

- **uncertainty and robustness** issues
  In a non informative setting, adding levels at prior increases the degree of uncertainty and of robustness of the prior distribution.
  especially in conjugate structures, otherwise too restrictive.
- **simplifying Bayesian computations**,
  e.g. computation of full conditional distributions MCMC and providing **'bricks' for graphical modelling**

Original data are realizations of 2 independent binomial models,

$$y_{1j} \sim Bin(n_{1j}, p_{1j}) \quad y_{0j} \sim Bin(n_{0j}, p_{0j}),$$

where $j = 1, \ldots, 22$ indexes clinical trial, $(y_{1j},\, n_{1j},\, p_{1j})$ and $(y_{0j},\, n_{0j},\, p_{0j})$ are deaths, subjects, probability of death in the treatment group and, respectively, in the control group for trial j.

Estimand of interest: probability difference, ratio, OR (odds ratio), ...
We choose LOR (logarithm of the odds ratio)

$$log(p_{1j}/(1-p_{1j})) - log(p_{0j}/(1-p_{0j}))$$

because normal approximation is close to its posterior distribution, even for small sample sizes.

Hierarchical modelling applied to a meta-analysis

The *beta-blockers* example

Gelman's book, pag. 148, sec. 5.6

On the course web page:

- *beta-blockers.data*, Table 5.4
- *N-N.r* R program
  Posterior simulation under the N-N hierarchical model via the factorization
  $p(\boldsymbol{\theta}, \mu, \tau|y) \propto p(\tau|y)p(\mu|\tau, y)p(\boldsymbol{\theta}|\mu, \tau, y)$
- *N-N.b* BUGS program
  Posterior simulation under the N-N hierarchical model via (Gibbs sampling) BUGS implementation (user is required only to specify the model!)

It is the possibility (iii), in which the studies are exchangeable (not either completely unrelated or identical), that we call **meta-analysis**.

–> *meaningful* parameters are:

- μ, the mean of the distribution of the effect sizes, which represents the overall 'average' effect across all studies (that could be regarded as exchangeable with the observed studies);
- the effect size in any of the observed studies;
- the effect size in another, comparable (exchangeable) study.

Besides, the fully Bayesian treatment of the hierarchical model requires $\phi = (\mu, \tau)$ unknown. In beta-blockers analysis $p(\mu|\tau) \propto 1$, $p(\tau) \propto 1$.

We estimate the LOR=θ_j by the *empirical logit*

$$y_j = log(y_{1j}/(n_{1j} - y_{1j})) - log(y_{0j}/(n_{0j} - y_{0j})),$$

with approximate sampling variance

$$\sigma_j^2 = 1/y_{1j} + 1/(n_{1j} - y_{1j}) + 1/y_{0j} + 1/(n_{0j} - y_{0j})$$

By the transformed data y_j and σ_j^2, we turn back to a **Normal-Normal hierarchical model** setting:

$$\begin{aligned} y_j|\theta_j &\overset{ind.}{\sim} N(\theta_j, \sigma_j^2), && (1) \\ \theta_j|\mu,\, \tau &\overset{i.i.d.}{\sim} N(\mu, \tau^2), && (2) \\ \mu,\, \tau &\sim p(\mu|\tau)\, p(\tau), \end{aligned}$$

where, w.r.t. the general setting (slide 4), $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{22})$ $\boldsymbol{\phi} = (\mu, \tau)$

(i) Graph: $\hat{\theta}_j = y_j$ and 95% confidence intervals, $\hat{\theta}_j \pm 1.96 * \sigma_j$, reflecting the precision of $\hat{\theta}_j$.
The posterior s.d of $\hat{\theta}_j \to \sigma_j$ as $\tau^2 \to \infty$.

(ii) Graph: $\hat{\mu} = \frac{\sum_j \frac{1}{\sigma_j^2} y_j}{\sum_j \frac{1}{\sigma_j^2}}$ and 95% confidence interval, $\hat{\mu} \pm 1.96 * V_\mu^{1/2}$, where $V_\mu^{-1} = \sum_j \frac{1}{\sigma_j^2}$ (in pink color). Moreover, 95% confidence interval, $\hat{\mu} \pm 1.96 * \sigma$, to compare observed data with the variation we would expect if assumption (ii) were true (in blue color)

(iii) Graph: posterior $2.5\%, 50\%, 97.5\%$ quantiles of θ_j (in red color), and of μ (in blue color).
$\hat{\mu}$ and 95% ci in (ii) are superimposed in green color.

The following 3 graphs show the two **limit cases** and the *continuum* between them of the N-N model:

(i) **Separate analyses**: $\tau^2 \to \infty$
N-N model reduces to

$$y_j|\theta_j \overset{ind.}{\sim} N(\theta_j, \sigma_j^2), \quad \boldsymbol{\theta} \sim p(\boldsymbol{\theta})$$

–> *meaningful* parameters are the effect sizes θ_j

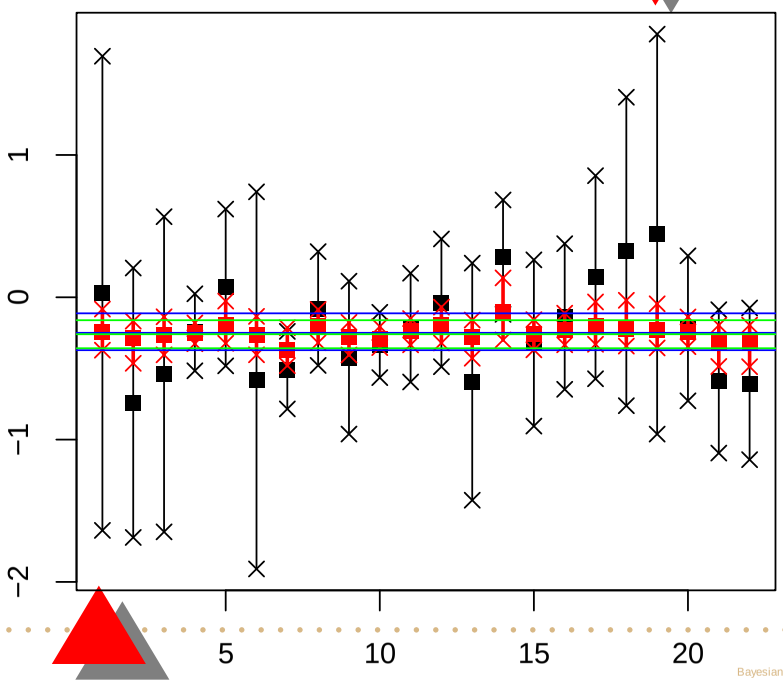
(ii) **Complete pooling**: $\tau^2 = 0$
N-N model reduces to

$$y_j|\mu, \sigma^2 \overset{i.i.d}{\sim} N(\mu, \sigma^2), \quad \mu \sim p(\mu)$$

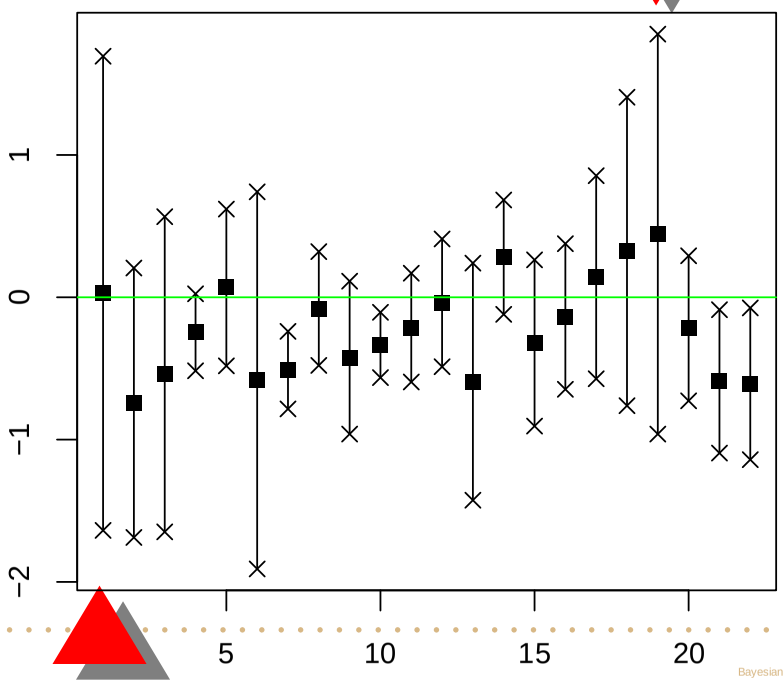
where $\sigma^2 = \frac{1}{22} \sum_j \sigma_j^2$
–> *meaningful* parameter is μ

(iii) **Exchangeability**: $0 < \tau^2 < \infty$

Exchangeable studies: pooling strength



Separate analyses



The following graphs show:

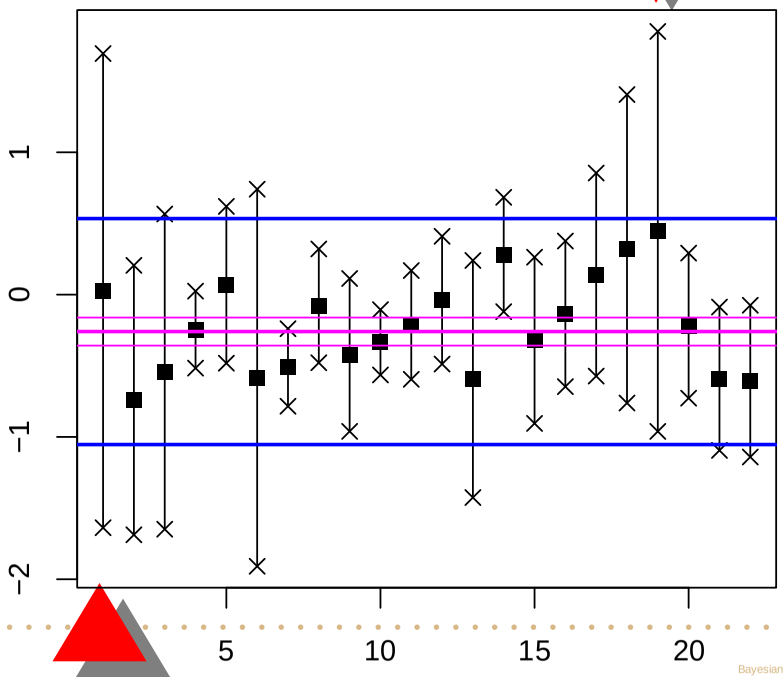
- Histograms of the simulated posterior densities for each effect size θ_j

  Note the skewness away from the central value of the overall mean, and the longer-tailed distributions in more *imprecise* trials.

- Marginal posterior densities of τ and μ

Since (at posterior) τ is concentrated around values that are small relative to the σ_j, considerable shrinkage is evident in the θ_j, especially in more *imprecise* studies; besides, that (which reflects the substantial homogeneity between the studies) produces a large reduction in their posterior variability.

or complete pooling?

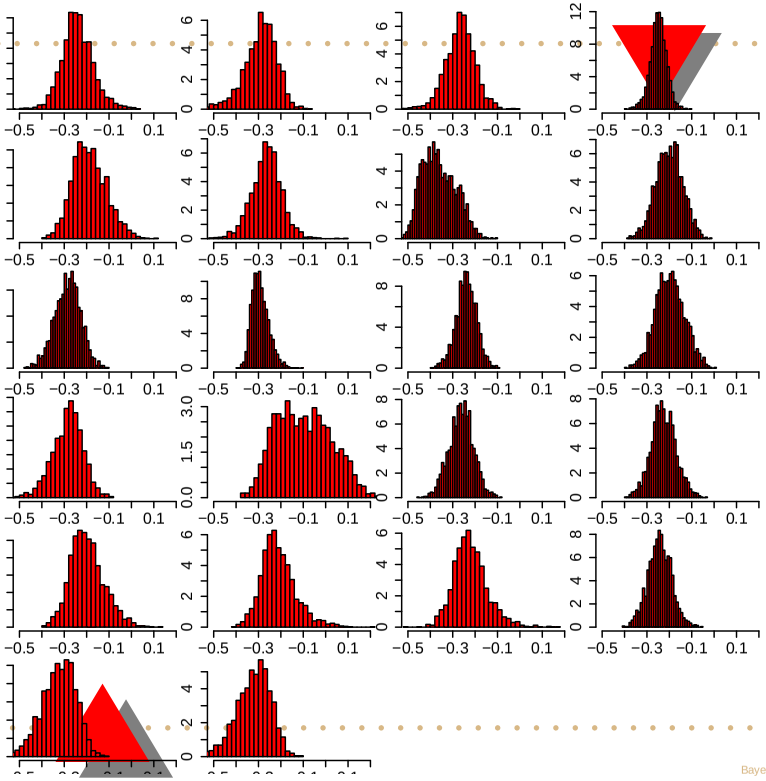


Below, for each τ value (taken in the plausible range of values of τ, according to its marginal posterior), $E(\mu|\tau, y)$ and $sd(\mu|\tau, y)$ are given by row.

Note the almost doubling of the sd. That shows the importance of averaging over τ (and not fixing τ as in a not full Bayes approach) in order to account adequately for uncertainty in the estimation of μ.

```
 tau: 0.00 0.05 0.10 0.15 0.20 0.25 0.30

mu|tau -0.26  0.05
       -0.256 0.05
       -0.245 0.06
       -0.244 0.07
       -0.240 0.07
       -0.239 0.08
       -0.238 0.09
```

Results of the hierarchical analysis in comparison to study specific or complete pooling analysis:

- *effect sizes*: shrinkage and reduction of posterior variance –> Bayesian estimates of θ_js borrow strength from each other
- *overall mean*: averaging over τ, uncertainty of μ is not underestimated
- *prediction*: uncertainty about the treatment effect in a not observed study may be more reasonable estimated by inference for a new study effect (exchangeable with the observed ones), rather than for the overall mean (see the 95% variation (-.58,.11) for $\tilde{\theta}$ w.r.t. the one (-.37,-.11) for μ in Table 5.5)