

Random effects ordinal/discrete time models

Application to grouped toxicological data

M.J. Martinez*

*Joint work with J. Hinde, NUIG, Ireland

25 January 2011

Motivating Application

Biological control of *Heterotermes tenuis* termites using the fungus *Beauveria bassiana*.

Experiment :

- 142 different isolates of the fungus
- One solution of each isolate is applied to 5 groups of 30 insects
↔ 5 replicates
- Cumulative mortality is measured daily during an 8-day period
↔ Multinomial observations of length eight

Objectives :

- Comparaison of isolates and selection of the effective ones
- Determination of the lethal time LT_p - time for p % mortality

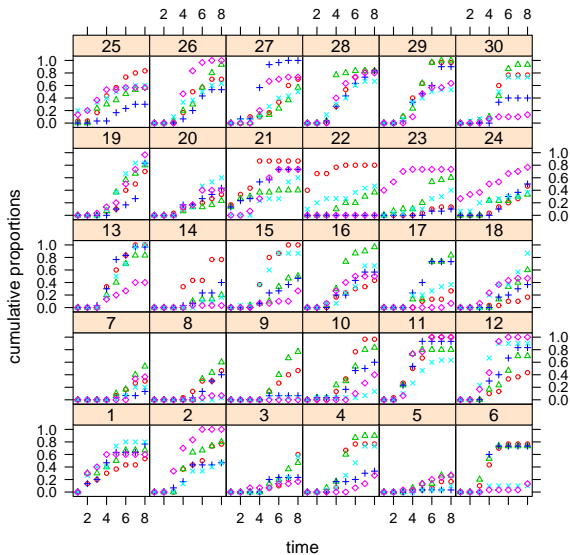


Figure: Cumulative proportions of dead termites for 30 isolates in the 5 replicates

Model specification

Model specification

General idea

- The cumulative mortalities measured over time result in multinomial responses.
- Because of natural time ordering, classical approaches consider models for the cumulative proportions.
- Alternative approach : Continuation-ratio models

Model specification

General idea

- The cumulative mortalities measured over time result in multinomial responses.
- Because of natural time ordering, classical approaches consider models for the cumulative proportions.
- Alternative approach : Continuation-ratio models

Notation

- The cumulative mortality is measured over D consecutive days
- n_{ik} : initial number of insects for replicate k of isolate i , $k = 1, \dots, K$ and $i = 1, \dots, I$
- Y_{jik} : number of dead insects on day j
 $Y_{D+1ik} = n_{ik} - \sum_{j=1}^D Y_{jik}$: number of insects still alive on day D
- π_{jik} : probability of an insect dying on day j for isolate i and replicate k
- For each replicate, we treat the counts in the $D + 1$ categories, $Y_{ik} = (Y_{1ik}, \dots, Y_{D+1ik})$, as multinomial with probabilities $(\pi_{1ik}, \dots, \pi_{D+1ik})$ where $\sum_{j=1}^{D+1} \pi_{jik} = 1$.

The model : a random effects continuation-ratio logit model

- Model the number of deaths on a day conditional on the number at risk, i.e. those surviving up to that day
- w_{jik} = probability an insect dies on day j given that it has survived up to that day =
$$\frac{\pi_{jik}}{\sum_{j'=j}^{D+1} \pi_{j'ik}}$$

The model : a random effects continuation-ratio logit model

- Model the number of deaths on a day conditional on the number at risk, i.e. those surviving up to that day
- w_{jik} = probability an insect dies on day j given that it has survived up to that day = $\frac{\pi_{jik}}{\sum_{j'=j}^{D+1} \pi_{j'ik}}$
- The multinomial probability of $p(y_{1ik}, \dots, y_{D+1ik})$ can be easily expressed in the form

$$b(n_{ik}; y_{1ik}; w_{1ik}) \times b(n_{ik} - y_{1ik}; y_{2ik}; w_{2ik}) \times \dots \times b(n_{ik} - y_{1ik} - \dots - y_{D-1ik}; y_{Dik}; w_{Dik})$$

where $b(n; y; w) =$ binomial probability of y successes out of n trials with probability w

- The continuation-ratio logits are ordinary logits of the conditional probabilities w_{jik} :

$$\eta_{jik} = \text{logit}(w_{jik}) = \log\left(\frac{w_{jik}}{1 - w_{jik}}\right) = \log\left(\frac{\pi_{jik}}{\pi_{j+1ik} + \dots + \pi_{D+1ik}}\right)$$

Main advantage : The continuation-ratio model can be fitted using methods for binomial logit models merely by a rearrangement of the data

Day	No. at risk	No. of deaths	Proba. of death
1	n_{ik}	y_{1ik}	π_{1ik}
2	$n_{ik} - y_{1ik}$	y_{2ik}	$w_{2ik} = \frac{\pi_{2ik}}{1 - \pi_{1ik}}$
3	$n_{ik} - y_{1ik} - y_{2ik}$	y_{3ik}	$w_{3ik} = \frac{\pi_{3ik}}{1 - \pi_{1ik} - \pi_{2ik}}$
⋮			
D	$n_{ik} - y_{1ik} - \cdots - y_{D-1ik}$	y_{Dik}	$w_{Dik} = \frac{\pi_{Dik}}{1 - \sum_{j'=1}^{D-1} \pi_{j'ik}}$

Linear predictor η_{jik} :

Variability observed among the replicates \leftrightarrow Introduction of an additive random intercept

Linear predictor η_{jik} :

Variability observed among the replicates \leftrightarrow Introduction of an additive random intercept

Model I

$$\eta_{jik} = \alpha_i + \beta_j + \xi_{ik}$$

where α_i = isolate effect of isolate i , β_j = time effect of day j , $\xi_{ik} \sim \mathcal{N}(0, \sigma^2)$ and the ξ_{ik} 's are assumed independent.

Linear predictor η_{jik} :

Variability observed among the replicates \leftrightarrow Introduction of an additive random intercept

Model I

$$\eta_{jik} = \alpha_i + \beta_j + \xi_{ik}$$

where α_i = isolate effect of isolate i , β_j = time effect of day j , $\xi_{ik} \sim \mathcal{N}(0, \sigma^2)$ and the ξ_{ik} 's are assumed independent.

Model II

$$\eta_{jik} = \alpha_i + \beta_j + \gamma_i t_j + \xi_{ik}$$

where $t_j = j$ is a quantitative variable for day j and γ_i is the time effect on isolate i .

Linear predictor η_{jik} :

Variability observed among the replicates \leftrightarrow Introduction of an additive random intercept

Model I

$$\eta_{jik} = \alpha_i + \beta_j + \xi_{ik}$$

where α_i = isolate effect of isolate i , β_j = time effect of day j , $\xi_{ik} \sim \mathcal{N}(0, \sigma^2)$ and the ξ_{ik} 's are assumed independent.

Model II

$$\eta_{jik} = \alpha_i + \beta_j + \gamma_i t_j + \xi_{ik}$$

where $t_j = j$ is a quantitative variable for day j and γ_i is the time effect on isolate i .

Model III

$$\eta_{jik} = \alpha_i + \gamma_i t_j + \xi_{ik}$$

Parameter estimation

The likelihood of the k th replicate of the i th isolate is given by

$$\begin{aligned} L_{ik}(\beta, \sigma^2) &= \int_{-\infty}^{+\infty} \prod_{j=1}^D f(y_{jik} | \beta, \xi_{ik}) \varphi(\xi_{ik}; 0, \sigma^2) d\xi_{ik} \\ &= \int_{-\infty}^{+\infty} \prod_{j=1}^D w_{jik}^{y_{jik}} (1 - w_{jik})^{n_{ik} - \sum_{j'=1}^j y_{jik}} \varphi(\xi_{ik}; 0, \sigma^2) d\xi_{ik} \end{aligned}$$

This likelihood function has no closed form and has to be approximated numerically

Parameter estimation

The likelihood of the k th replicate of the i th isolate is given by

$$\begin{aligned} L_{ik}(\beta, \sigma^2) &= \int_{-\infty}^{+\infty} \prod_{j=1}^D f(y_{jik} | \beta, \xi_{ik}) \varphi(\xi_{ik}; 0, \sigma^2) d\xi_{ik} \\ &= \int_{-\infty}^{+\infty} \prod_{j=1}^D w_{jik}^{y_{jik}} (1 - w_{jik})^{n_{ik} - \sum_{j'=1}^j y_{jik}} \varphi(\xi_{ik}; 0, \sigma^2) d\xi_{ik} \end{aligned}$$

This likelihood function has no closed form and has to be approximated numerically

Considered approach

- Numerical evaluation of the marginal likelihood using Gauss-Hermite quadrature (GQ/AGQ methods)

Ordinary Gaussian quadrature methods (GQ)

- GQ methods replace the integral by a finite weighted sum :

$$L_{ik}(\beta, \sigma^2) = \int_{-\infty}^{+\infty} \prod_{j=1}^D f(y_{jik} | \beta, \xi_{ik}) \varphi(\xi_{ik}; 0, \sigma^2) d\xi_{ik}$$
$$\approx \sum_{r=1}^R \pi_r \left\{ \prod_{j=1}^D f(y_{jik} | \beta, \sigma^2, z_r) \right\}$$

- The weights π_r and quadrature points z_r depend only on the order of the approximation R and the normal density.
- The approximation improves as the number R of quadrature points increases.
- In practice, a large number of quadrature points is often required to approximate correctly the likelihood.
- The approximation can be poor for large random effects variances or small cluster sizes.

Adaptive Gaussian quadrature methods (AGQ)

- With AGQ, the quadrature points and weights are adapted to the 'support' of $\prod_{j=1}^D f(y_{jik}|\beta, \xi_{ik}) \varphi(\xi_{ik}; 0, \sigma^2)$
- The integrand is proportional to the posterior density of ξ_{ik} given the responses and can often be well approximated by a normal density $\varphi(\xi_{ik}; \mu_{ik}, \tau_{ik}^2)$ with mean μ_{ik} and variance τ_{ik}^2 .

Adaptive Gaussian quadrature methods (AGQ)

- With AGQ, the quadrature points and weights are adapted to the 'support' of $\prod_{j=1}^D f(y_{jik}|\beta, \xi_{ik}) \varphi(\xi_{ik}; 0, \sigma^2)$
- The integrand is proportional to the posterior density of ξ_{ik} given the responses and can often be well approximated by a normal density $\varphi(\xi_{ik}; \mu_{ik}, \tau_{ik}^2)$ with mean μ_{ik} and variance τ_{ik}^2 .
- Re-write the integral as

$$L_{ik}(\beta, \sigma^2) = \int_{-\infty}^{+\infty} \varphi(\xi_{ik}, \mu_{ik}, \tau_{ik}^2) \frac{\prod_{j=1}^D f(y_{jik}|\beta, \sigma^2, \xi_{ik}) \varphi(\xi_{ik}; 0, 1)}{\varphi(\xi_{ik}; \mu_{ik}, \tau_{ik}^2)} d\xi_{ik}$$

and treat the normal density approximating the posterior density as the weight function.

Adaptive Gaussian quadrature methods (AGQ)

- Applying the standard quadrature rules yields

$$L_{ik}(\beta, \sigma^2) \approx \sum_{r=1}^R \pi_{ikr} \left\{ \prod_{j=1}^D f(y_{jik} | \beta, \sigma^2, z_{ikr}) \right\}.$$

$$\text{with } \begin{cases} z_{ikr} = \mu_{ik} + \tau_{ik} z_r \\ \pi_{ikr} = \sqrt{2\pi\tau_{ik}} \exp(z_r^2/2) \varphi(z_{ikr}) \pi_r \end{cases}$$

Adaptive Gaussian quadrature methods (AGQ)

- Applying the standard quadrature rules yields

$$L_{ik}(\beta, \sigma^2) \approx \sum_{r=1}^R \pi_{ikr} \left\{ \prod_{j=1}^D f(y_{jik} | \beta, \sigma^2, z_{ikr}) \right\}.$$

$$\text{with } \begin{cases} z_{ikr} = \mu_{ik} + \tau_{ik} z_r \\ \pi_{ikr} = \sqrt{2\pi\tau_{ik}} \exp(z_r^2/2) \varphi(z_{ikr}) \pi_r \end{cases}$$

- We take here μ_{ik} and τ_{ik}^2 to be the posterior moments which are not known exactly but must be themselves obtained using adaptive quadrature (as in Rabe-Hesketh et al. 2005).

Adaptive Gaussian quadrature methods (AGQ)

- Applying the standard quadrature rules yields

$$L_{ik}(\beta, \sigma^2) \approx \sum_{r=1}^R \pi_{ikr} \left\{ \prod_{j=1}^D f(y_{jik} | \beta, \sigma^2, z_{ikr}) \right\}.$$

$$\text{with } \begin{cases} z_{ikr} = \mu_{ik} + \tau_{ik} z_r \\ \pi_{ikr} = \sqrt{2\pi\tau_{ik}} \exp(z_r^2/2) \varphi(z_{ikr}) \pi_r \end{cases}$$

- We take here μ_{ik} and τ_{ik}^2 to be the posterior moments which are not known exactly but must be themselves obtained using adaptive quadrature (as in Rabe-Hesketh et al. 2005).

↔ The integration is iterative

Parameter estimation

Method considered in this work

Adaptive Gaussian quadrature within an EM algorithm

- To evaluate the marginal likelihood for given parameter values θ
↔ Adaptive Gaussian quadrature
- To maximize this marginal likelihood with respect to θ
↔ EM algorithm

Algorithm principle

The algorithm alternates between

- a step of the EM algorithm to update the parameter values
- the set of previous iterations to update the quadrature points and weights

Remarks

- A similar iterative algorithm is considered by Rabe-Hesketh et al. (2005) within a Newton-Raphson procedure
- An alternative way to approximate μ_{ik} and τ_{ik}^2 is to use the mode and the curvature at the mode (Liu and Pierce, 1994).
- AGQ is essentially a deterministic version of importance sampling with $\varphi(\xi_{ik}, \mu_{ik}, \tau_{ik}^2)$ as importance density.
- In general, the higher the order K , the better the approximation will be.
- Typically, adaptive Gaussian quadrature needs less quadrature points than ordinary Gaussian quadrature.
- On the other hand, adaptive Gaussian quadrature is much more time consuming .

Some results

- Subset of 30 isolates
- Simple GLMs are fitted to the data set as well
- Disparity, deviance and σ^2 values for models I, II and III using ordinary and adaptive Gaussian quadrature

Some results

		GQ(3)	GQ(5)	GQ(10)	GQ(20)	GQ(40)	GQ(60)
Model I	-2 logL	4937.8	4899.9	4924.4	4933.6	4940.0	4940.2
	Deviance	3327.5	3289.6	3314.2	3323.4	3329.7	3329.9
	σ	0.972	0.976	1.084	0.971	0.839	0.836
Model II	-2 logL	4269.2	4269.7	4274.7	4287.8	4294.8	4294.9
	Deviance	2658.9	2659.4	2664.4	2677.6	2684.5	2684.6
	σ	0.931	1.014	1.143	1.027	0.852	0.850
Model III	-2 logL	5202.6	5197.8	5189.2	5195.0	5204.8	5205.0
	Deviance	3592.3	3587.5	3578.9	3584.7	3594.6	3594.8
	σ	0.930	1.012	1.066	1.077	0.908	0.893
		AGQ(3)	AGQ(5)	AGQ(10)	Fixed model		
Model I	-2 logL	4941.0	4940.2	4940.2	5651.8		
	Deviance	3330.8	3329.9	3329.9	4041.6		
	σ	0.831	0.836	0.836	-		
Model II	-2 logL	4295.7	4294.9	4294.9	5014.5		
	Deviance	2685.4	2684.6	2684.6	3404.2		
	σ	0.846	0.850	0.850	-		
Model III	-2 logL	5206.0	5205.1	5205.1	6029.7		
	Deviance	3595.7	3594.8	3594.8	4419.4		
	σ	0.887	0.893	0.893			

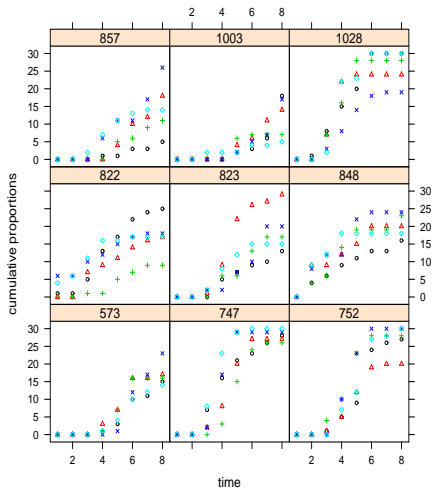
Some results

- Subset of 30 isolates
- Simple GLMs are fitted to the data set as well
- Disparity, deviance and σ^2 values for models I, II and III using ordinary and adaptive Gaussian quadrature

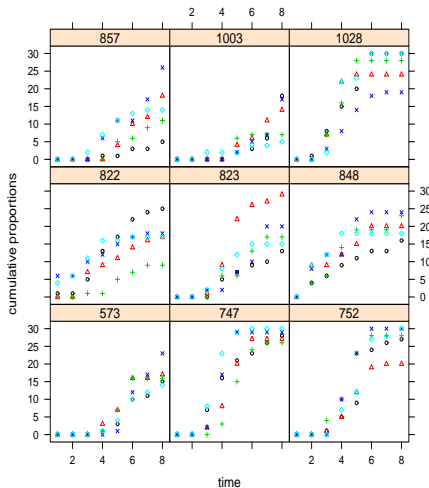
Conclusions

- The large variability is captured by these random effects models
- Clearly, we may be able to achieve stable results with a smaller number of quadrature points using AGQ instead of GQ for Models I, II and III
- Model II with a deviance of 2684.6 seems to be the model that best fits the data

▷ 9 isolates with "homogeneous" replicates

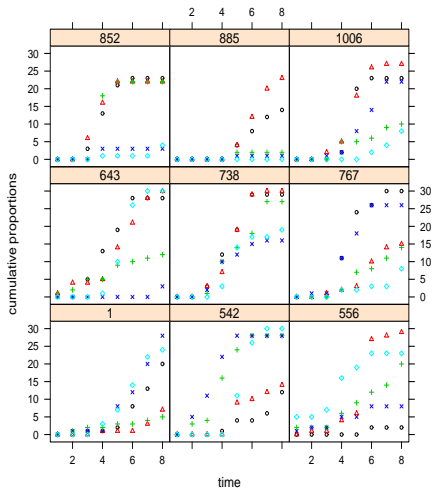


▷ 9 isolates with "homogeneous" replicates

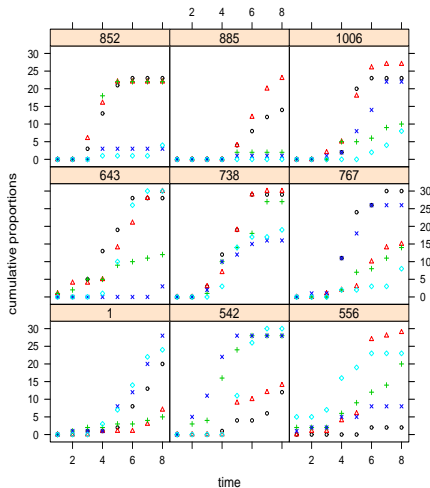


	Model II		
	$-2 \log L$	Deviance	σ
GQ(3)	1325.2	789.1	0.671
GQ(5)	1330.5	794.4	0.660
GQ(10)	1332.4	796.3	0.488
GQ(20)	1332.6	796.4	0.468
AGQ(3)	1332.7	796.5	0.467
AGQ(5)	1332.6	796.4	0.469
AGQ(10)	1332.6	796.4	0.469

▷ 9 isolates with "heterogeneous" replicates

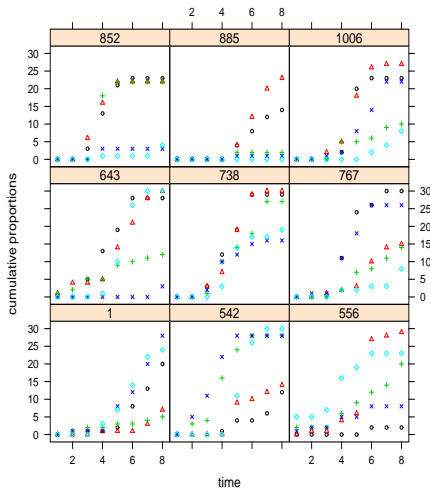


▷ 9 isolates with "heterogeneous" replicates



	Model II		
	-2 logL	Deviance	σ
GQ(3)	1358.9	889.9	0.999
GQ(5)	1340.1	871.1	0.994
GQ(10)	1333.6	864.7	1.302
GQ(20)	1331.4	862.4	1.176
GQ(40)	1335.3	866.3	1.517
GQ(60)	1340.7	871.8	1.228
GQ(80)	1340.8	871.8	1.194
AGQ(3)	1341.2	872.2	1.180
AGQ(5)	1340.9	871.9	1.186
AGQ(10)	1340.9	871.9	1.187

▷ 9 isolates with "heterogeneous" replicates



	Model II		
	-2 logL	Deviance	σ
GQ(3)	1358.9	889.9	0.999
GQ(5)	1340.1	871.1	0.994
GQ(10)	1333.6	864.7	1.302
GQ(20)	1331.4	862.4	1.176
GQ(40)	1335.3	866.3	1.517
GQ(60)	1340.7	871.8	1.228
GQ(80)	1340.8	871.8	1.194
AGQ(3)	1341.2	872.2	1.180
AGQ(5)	1340.9	871.9	1.186
AGQ(10)	1340.9	871.9	1.187

↪ Much worse behaviour

Predicted replicate-specific evolutions and the average evolution

- Posterior probability p_{ikr} that the unobserved ξ_{ik} takes the value z_r are obtained at the convergence of the EM-algorithm and provide the posterior distribution of the ξ_{ik} by replacing the unknown parameters by their ML estimates.

$$\hat{p}_{ikr} = \frac{\pi_r \prod_{j=1}^D f(y_{jik} | \hat{\beta}, \hat{\sigma}, z_r)}{\sum_{l=1}^R \pi_l \prod_{j=1}^D f(y_{jik} | \hat{\beta}, \hat{\sigma}, z_l)}$$

- Posterior estimates of the random effects :

$$\tilde{\xi}_{ik} = \sum_{r=1}^R \hat{p}_{ikr} z_r$$

↪ Empirical Bayes predictions $\hat{\eta}_{jik}$ and fitted probabilities of real interest $\hat{\pi}_{jik}$

- Marginal average evolution is derived from averaging the conditional means over the random effects ξ_{ik} .

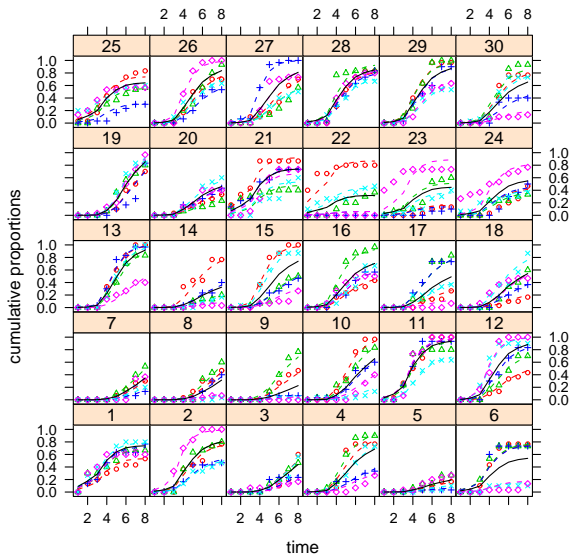
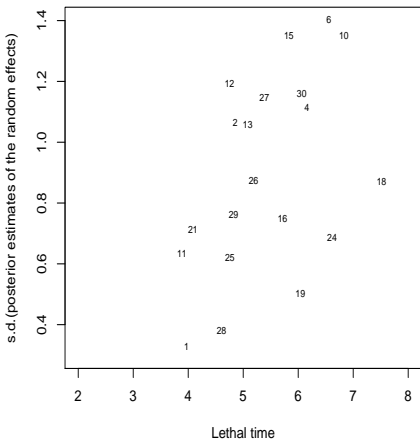


Figure: Predicted replicate-specific evolutions and marginal average evolution

Median lethal time LT_{50}

- LT_{50} = time for 50% mortality
- Marginal median lethal time for 30 isolates
- Plot LT_{50} against Variability of the posterior estimates of the random effects



Discussion

- In the previous model, the random effects ξ_{ik} are assumed to be sampled from a normal distribution, i.e. $\xi_{ik} \sim \mathcal{N}(0, \sigma^2)$
- This assumption reflects the prior believe that the random effects are drawn from one homogeneous population of random effects

▷ Extension : **Heterogeneity model**

- The normality assumption can be replaced by a mixture of R normal distributions with mean μ_r and variance σ_r

$$\xi_{ik} \sim \sum_{r=1}^R p_r \mathcal{N}(\mu_r, \sigma_r^2)$$

with $\sum_{r=1}^R p_r = 1$ and $\sum_{r=1}^R \mu_r p_r = 0$.

- Prior believe of presence of unobserved heterogeneity among the replicates
- Also used for classification purposes