



Training course on animal breeding data analysis and genomics prediction

Vincent Ducrocq

► To cite this version:

Vincent Ducrocq. Training course on animal breeding data analysis and genomics prediction. 3rd cycle. Animal breeding Data Analysis and Genomics Prediction (Training course on Animal breeding Data Analysis and Genomics Prediction), 2017. hal-02786129

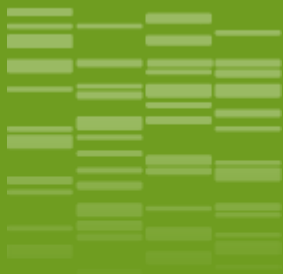
HAL Id: hal-02786129

<https://hal.inrae.fr/hal-02786129>

Submitted on 4 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



What is a genetic evaluation ?

Vincent Ducrocq

GABI, Jouy-en-Josas
France



Notion of genetic value

- A farmer chooses the animals which are going to produce progeny. These offspring will replace older animals...
- This farmer usually wants them to be better, i.e., they should have a phenotype different from the one obtained without selection of the parents.
- The difference between the two represents a « **genetic gain** »
- The **genetic value** or **breeding value** of an animal is (twice) what it **transmits to its progeny ... on average**

(twice because only half of its genes are transmitted)

- So the « genetic value » is a mathematical notion it cannot be observed, it cannot be measured ...

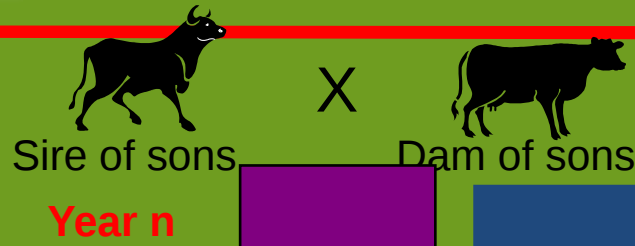
But it can be **predicted / estimated** using

Genetic values : progeny test

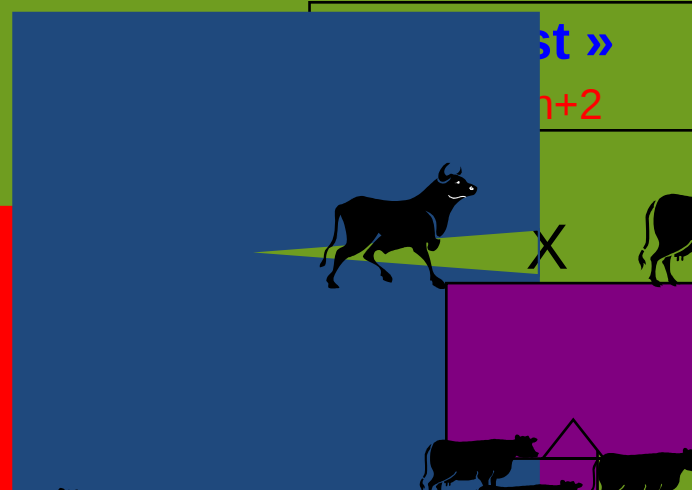
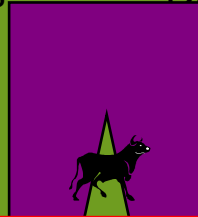
- The average « performance » of a large number of progeny gives a precise estimation of its genetic value ...
- This is how bulls of dairy breeds have been selected during 50 years:
 - ✓ a representative sample of daughters is created ☐ we « test » each bull via the characteristics of 50-100 daughters
 - ✓ Then we keep the best bulls using the information on these characteristics but also on other related animals (mother, sisters,...)
 - ☐ these sources of information are combined into an

estimated genetic value = estimated breeding value

Genetic evaluation



Genetic evaluation



Genetic evaluation



daughters born

evaluation?



November 20-24 2017



The traits of interest are often complex traits under the influence of **genetic** effects and **environmental** effects

The **phenotype** (= the measure of a trait) is misleading : for example, it depends strongly on the management of the herd.

We want to avoid a confusion between the **genetic value** and the **phenotype** influenced by the environment

Example :

cow A produces **3000 kg** of milk in a **herd** with a herd average of **4000 kg**

cow B, from the same breed, produces **2500 kg** of milk

in a **herd** with a herd average of **2000 kg**

Which one has the highest genetic value for production?



Data :

- Phenotypes
(mesure of a trait, P)
- environment
(herd, season, age...)
- Pedigree



Models:

- Description of the data

$$P = a + M + e$$

- Genetic model of « transmission » of

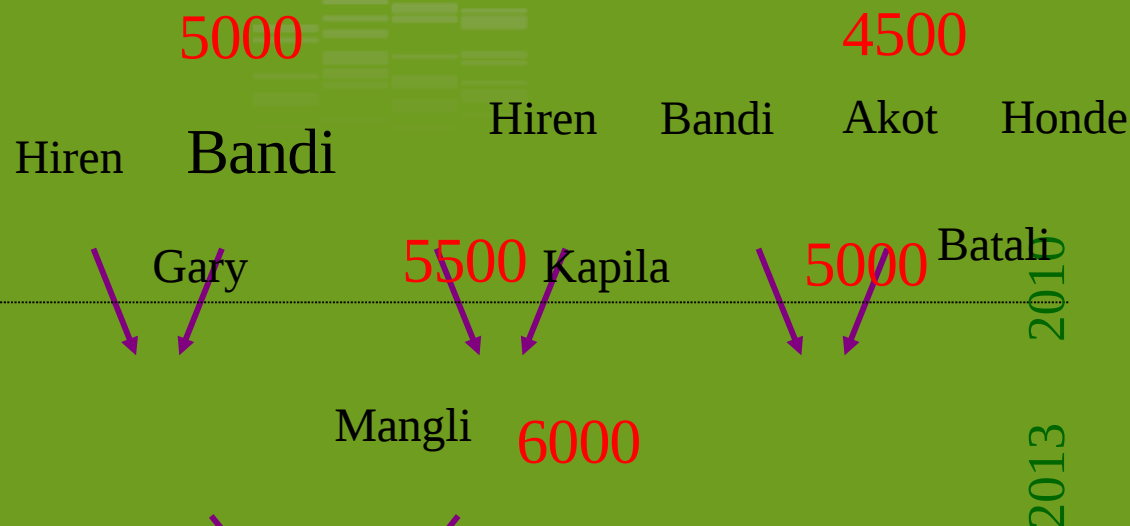
$$a_i = \frac{1}{2}a_s + \frac{1}{2}a_d + \phi_i$$

genetic value





cow	sire	dam	Sire of sire	Dam of sire	Sire of dam	Dam of dam	305d milk yield	Calving date
Bandi	-	-	-	-	-	-	5000	14/10/2010
Honde	-	-	-	-	-	-	4500	23/11/2010
Kapila	Hiren	Bandi	-	-	-	-	5500	6/7/2013
Batali	Akot	Honde	-	-	-	-	5000	15/9/2013
Mangli	Gary	Kapila	Hiren	Bandi	Hiren	Bandi	6000	25/04/2016



Choose a **model** :

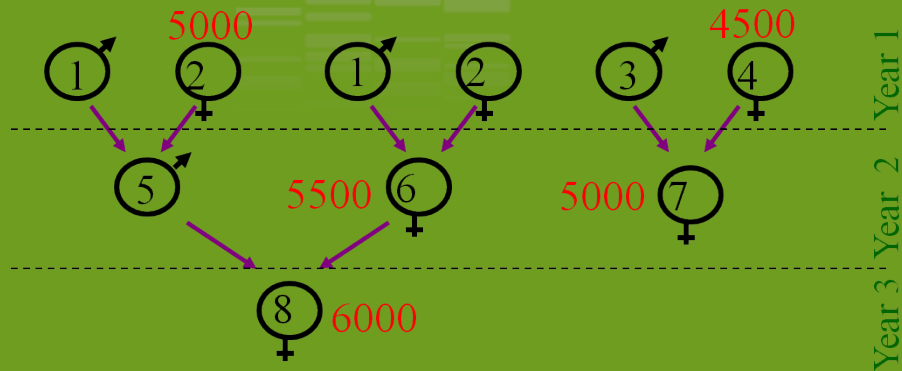
milk yield of cow i =
effect of **year j**
+ **genetic effect** of i
+ an unexplained
“**residual**”

➤ **Data file**

Cow (i)	Year (j)	305d milk yield
Bandi	2010	5000
Honde	2010	4500
Kapila	2013	5500
Batali	2013	5000
Mangli	2016	6000

➤ **Pedigree file**

Animal	Sire	dam
Hiren	?	?
Bandi	?	?
Akot	?	?
Honde	?	?
Gary	Hiren	Bandi
Kapila	Hiren	Bandi
Batali	Akot	Honde
Mangli	Gary	Kapila



- The chosen model :

$$y_i = \beta_j + a_i + e_i$$

We would like to estimate:

- the year effects ($\beta_1, \beta_2, \beta_3$)
- the genetic values (a_1 to a_8)

➤ **Data file:** animal n°, year, production

2	1	5000
4	1	4500
6	2	5500
7	2	5000
8	3	6000

➤ **Pedigree file :** animal n°, sire n°, dam n°

1	0	0
2	0	0
3	0	0
4	0	0
5	1	2
6	1	2
7	3	4
8	5	6

- “BLUP” theory $\square$ a system of equations, in matrix form :

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

< mixed model equations >

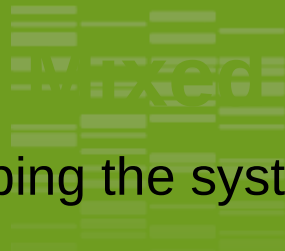
$$\begin{pmatrix} 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ \hline 0 & 0 & 0 & 0+2\alpha & 0+\alpha & 0 & 0 & 0 & 0-\alpha & 0-\alpha & 0 & 0 \\ 1 & 0 & 0 & 0+\alpha & 1+2\alpha & 0 & 0 & 0 & 0-\alpha & 0-\alpha & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0+1.5\alpha & 0+0.5\alpha & 0 & 0 & 0 & 0-\alpha & 0 \\ 1 & 0 & 0 & 0 & 0 & 0+0.5\alpha & 1+1.5\alpha & 0 & 0 & 0 & 0-\alpha & 0 \\ 0 & 0 & 0 & 0-\alpha & 0-\alpha & 0 & 0 & 0 & 0+2.5\alpha & 0+0.5\alpha & 0 & 0-\alpha \\ 0 & 1 & 0 & 0-\alpha & 0-\alpha & 0 & 0 & 0 & 0+0.5\alpha & 1+2.5\alpha & 0 & 0-\alpha \\ 0 & 1 & 0 & 0 & 0 & 0-\alpha & 0-\alpha & 0 & 0 & 0 & 1+2\alpha & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0-\alpha & 0-\alpha & 0 & 1+2\alpha \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \hline a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \\ a_8 \end{pmatrix} = \begin{pmatrix} 5000+4500 \\ 5500+5000 \\ 6000 \\ \hline 0 \\ 5000 \\ 0 \\ 4500 \\ 0 \\ 5500 \\ 5000 \\ 6000 \end{pmatrix}$$

- in red : contribution of pedigree
- in black: contribution for records
- $\hat{a}_i$: genetic value



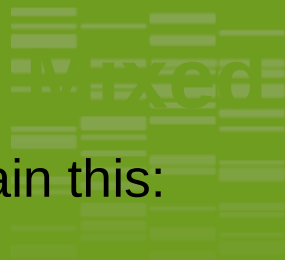
$$\begin{pmatrix}
 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 2 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
 0 & 0 & 0 & 0+2\alpha & 0+\alpha & 0 & 0 & 0 & 0 \\
 1 & 0 & 0 & 0+\alpha & 1+2\alpha & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0+1.5\alpha & 0+0.5\alpha & 0 & 0 \\
 1 & 0 & 0 & 0 & 0 & 0+0.5\alpha & 1+1.5\alpha & 0 & 0 \\
 0 & 0 & 0 & 0-\alpha & 0-\alpha & 0 & 0 & 0+2.5\alpha & 0+0.5\alpha \\
 0 & 1 & 0 & 0-\alpha & 0-\alpha & 0 & 0 & 0+0.5\alpha & 1+2.5\alpha \\
 0 & 1 & 0 & 0 & 0 & 0-\alpha & 0-\alpha & 0 & 1+2\alpha \\
 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0-\alpha & 0-\alpha & 0 & 1+2\alpha
 \end{pmatrix}
 \begin{pmatrix}
 \beta_1 \\
 \beta_2 \\
 \beta_3 \\
 a_1 \\
 a_2 \\
 a_3 \\
 a_4 \\
 a_5 \\
 a_6 \\
 a_7 \\
 a_8
 \end{pmatrix}
 =
 \begin{pmatrix}
 5000 + 4500 \\
 5500 + 5000 \\
 6000 \\
 0 \\
 5000 \\
 0 \\
 4500 \\
 0 \\
 5500 \\
 5000 \\
 6000
 \end{pmatrix}$$

- characteristics: a lot of 0's, simple numbers easy to interpret and easy to “construct” just by reading two files (data + pedigree)
- the elements in red representing the links between related animals are easy to construct from a pedigree file
- α = a simple function of the heritability h^2 of the trait
- Also, the equations can be interpreted intuitively



- developing the system of equations:

$$\begin{cases}
 2\beta_1 + 0\beta_2 + 0\beta_3 + 0a_1 + 1a_2 + 0a_3 + 1a_4 + 0a_5 + 0a_6 + 0a_7 + 0a_8 = 5000 + 4500 \\
 0\beta_1 + 2\beta_2 + 0\beta_3 + 0a_1 + 0a_2 + 0a_3 + 0a_4 + 0a_5 + 1a_6 + 1a_7 + 0a_8 = 5500 + 5000 \\
 0\beta_1 + 0\beta_2 + 1\beta_3 + 0a_1 + 0a_2 + 0a_3 + 0a_4 + 0a_5 + 0a_6 + 0a_7 + 1a_8 = 6000 \\
 0\beta_1 + 0\beta_2 + 0\beta_3 + 2\alpha a_1 + \alpha a_2 + 0a_3 + 0a_4 - \alpha a_5 - \alpha a_6 + 0a_7 + 0a_8 = 0 \\
 1\beta_1 + 0\beta_2 + 0\beta_3 + \alpha a_1 + (1+2\alpha)a_2 + 0a_3 + 0a_4 - \alpha a_5 - \alpha a_6 + 0a_7 + 0a_8 = 5000 \\
 0\beta_1 + 0\beta_2 + 0\beta_3 + 0a_1 + 0a_2 + 1.5\alpha a_3 + 0.5\alpha a_4 + 0a_5 + 0a_6 - \alpha a_7 + 0a_8 = 0 \\
 1\beta_1 + 0\beta_2 + 0\beta_3 + 0a_1 + 0a_2 + 0.5\alpha a_3 + (1+1.5\alpha)a_4 + 0a_5 + 0a_6 - \alpha a_7 + 0a_8 = 4500 \\
 0\beta_1 + 0\beta_2 + 0\beta_3 - \alpha a_1 - \alpha a_2 + 0a_3 + 0a_4 + 2.5\alpha a_5 + 0.5\alpha a_6 + 0a_7 + \alpha a_8 = 0 \\
 0\beta_1 + 1\beta_2 + 0\beta_3 - \alpha a_1 - \alpha a_2 + 0a_3 + 0a_4 + 0.5\alpha a_5 + (1+2.5\alpha)a_6 + 0a_7 - \alpha a_8 = 5500 \\
 0\beta_1 + 1\beta_2 + 0\beta_3 + 0a_1 + 0a_2 - \alpha a_3 - \alpha a_4 + 0a_5 + 0a_6 + 2\alpha a_7 + 0a_8 = 5000 \\
 0\beta_1 + 0\beta_2 + 1\beta_3 + 0a_1 + 0a_2 + 0a_3 + 0a_4 - \alpha a_5 - \alpha a_6 + 0a_7 + (1+2\alpha)a_8 = 6000
 \end{cases}$$



• We obtain this:

$$2\beta_1 + 1a_2 + 1a_4 = y_2 + y_4$$

$$2\beta_2 + 1a_6 + 1a_7 = y_6 + y_7$$

$$1\beta_3 + 1a_8 = y_8$$

} equations to estimate
the “year” effects β

$$2\alpha a_1 + \alpha a_2 - \alpha a_5 - \alpha a_6 = 0$$

$$1\beta_1 + \alpha a_1 + (1 + 2\alpha) a_2 - \alpha a_5 - \alpha a_6 = y_2$$

$$1.5\alpha a_3 + 0.5\alpha a_4 - \alpha a_7 = 0$$

$$1\beta_1 + 0.5\alpha a_3 + (1 + 1.5\alpha) a_4 - \alpha a_7 = y_4$$

$$-\alpha a_1 - \alpha a_2 + 2.5\alpha a_5 + 0.5\alpha a_6 - \alpha a_8 = 0$$

$$1\beta_2 - \alpha a_1 - \alpha a_2 + 0.5\alpha a_5 + (1 + 2.5\alpha) a_6 - \alpha a_8 = y_6$$

$$1\beta_2 - \alpha a_3 - \alpha a_4 + 2\alpha a_7 = y_7$$

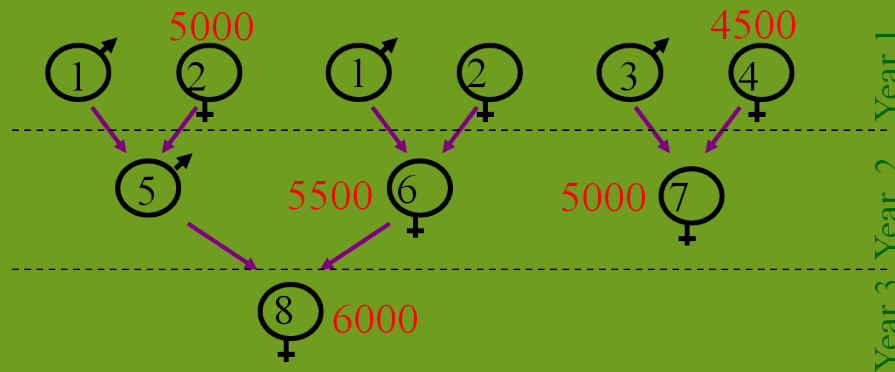
$$1\beta_3 - \alpha a_5 - \alpha a_6 + (1 + 2\alpha) a_8 = y_8$$

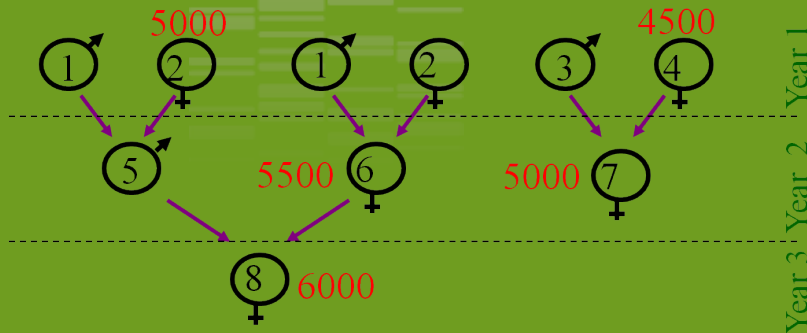
} equations to estimate
the genetic values a_i



$$\left\{ \begin{array}{l} 2\beta_1 + 1a_2 + 1a_4 = y_2 + y_4 \longrightarrow \hat{\beta}_1 = \frac{1}{2}[(y_2 - \hat{a}_2) + (y_4 - \hat{a}_4)] \\ 2\beta_2 + 1a_6 + 1a_7 = y_6 + y_7 \longrightarrow \hat{\beta}_2 = \frac{1}{2}[(y_6 - \hat{a}_6) + (y_7 - \hat{a}_7)] \\ 1\beta_3 + 1a_8 = y_8 \longrightarrow \hat{\beta}_3 = (y_8 - \hat{a}_8) \end{array} \right.$$

= average of performances corrected for the genetic effect





of
= weighted average
and accounting for the
average genetic values
of parents and of progeny

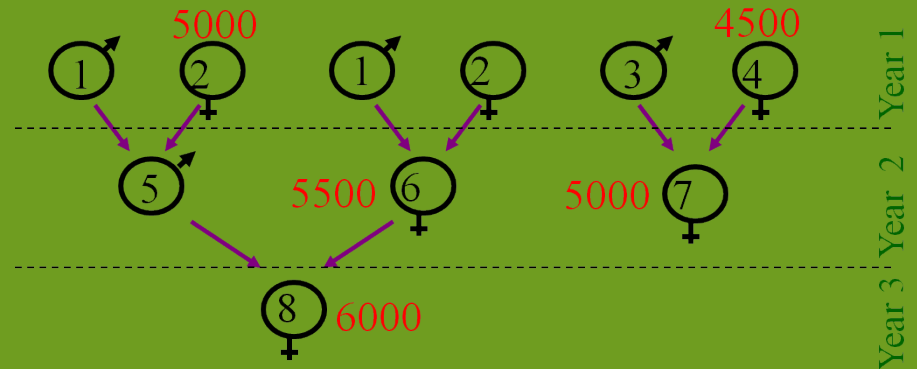
$$\begin{cases}
 2\alpha a_1 + \alpha a_2 - \alpha a_5 - \alpha a_6 = 0 & \Rightarrow \hat{a}_1 = \frac{1}{2} \left[\left(\hat{a}_5 - \frac{1}{2} \hat{a}_2 \right) + \left(\hat{a}_6 - \frac{1}{2} \hat{a}_2 \right) \right] \\
 1\beta_3 - \alpha a_5 - \alpha a_6 + (1 + 2\alpha) a_8 = y_8 & \Rightarrow \hat{a}_8 = \frac{2\alpha}{(1 + 2\alpha)} \left(\frac{\hat{a}_5 + \hat{a}_6}{2} \right) + \frac{1}{(1 + 2\alpha)} (y_8 - \hat{\beta}_3) \\
 1\beta_2 - \alpha a_1 - \alpha a_2 + 0.5\alpha a_5 + (1 + 2.5\alpha) a_6 - \alpha a_8 = y_6 & \Rightarrow \hat{a}_6 = \frac{2.5\alpha}{(1 + 2.5\alpha)} \underbrace{\left(\frac{\hat{a}_1 + \hat{a}_2}{2} \right)}_{\text{information from parents}} + \frac{1}{(1 + 2.5\alpha)} \underbrace{(y_6 - \hat{\beta}_2)}_{\text{corrected own performance}} + \frac{0.5\alpha}{(1 + 2.5\alpha)} \underbrace{\left(2 \left(\hat{a}_8 - \frac{1}{2} \hat{a}_5 \right) \right)}_{\text{information on progeny}}
 \end{cases}$$

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \\ \hat{a}_5 \\ \hat{a}_6 \\ \hat{a}_7 \\ \hat{a}_8 \end{pmatrix} = \begin{pmatrix} 4750 \\ 5250 \\ 5930 \\ 28 \\ 83 \\ -28 \\ -83 \\ 56 \\ 83 \\ -83 \\ 70 \end{pmatrix}$$

Sum for
«base animals»
=0

average = 18.6

average = 70



Genetic trend
/ generation

$$y_8 = 6000 = 5930 + 70 \quad \square \text{ connexion through pedigree}$$

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \\ \mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}$$

(BLUP) Mixed Model Equations : univariate case

$$\begin{bmatrix} \mathbf{X}_1'\mathbf{R}^{11}\mathbf{X}_1 & \mathbf{X}_1'\mathbf{R}^{12}\mathbf{X}_2 & & \\ \mathbf{X}_2'\mathbf{R}^{21}\mathbf{X}_1 & \mathbf{X}_2'\mathbf{R}^{22}\mathbf{X}_2 & & \\ \hline & \text{sym.} & & \\ & & \mathbf{Z}_1'\mathbf{R}^{11}\mathbf{Z}_1 + \mathbf{G}^{11} & \mathbf{Z}_1'\mathbf{R}^{12}\mathbf{Z}_2 + \mathbf{G}^{12} \\ & & \mathbf{Z}_2'\mathbf{R}^{21}\mathbf{Z}_1 + \mathbf{G}^{21} & \mathbf{Z}_2'\mathbf{R}^{22}\mathbf{Z}_2 + \mathbf{G}^{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \mathbf{a}_1 \\ \mathbf{a}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1'\mathbf{R}^{11}\mathbf{y}_1 + \mathbf{X}_1'\mathbf{R}^{12}\mathbf{y}_2 \\ \mathbf{X}_2'\mathbf{R}^{21}\mathbf{y}_1 + \mathbf{X}_2'\mathbf{R}^{22}\mathbf{y}_2 \\ \hline \mathbf{Z}_1'\mathbf{R}^{11}\mathbf{y}_1 + \mathbf{Z}_1'\mathbf{R}^{12}\mathbf{y}_2 \\ \mathbf{Z}_2'\mathbf{R}^{21}\mathbf{y}_1 + \mathbf{Z}_2'\mathbf{R}^{22}\mathbf{y}_2 \end{bmatrix}$$

(BLUP) Mixed Model Equations : multivariate case

➤ $P = G + E$

➤ **P:** Phenotype

- may be standardized, corrected, combined ...

➤ **G:** Genetic component

- depends on the objective of the analysis
- may be incomplete (e.g., just « sire » or « additive » component)

$$\left[\begin{array}{c} \text{genetic value} \\ \text{of } i \end{array} \right] = \left[\begin{array}{c} \text{gamete of} \\ \text{sire } s \text{ of } i \end{array} \right] + \left[\begin{array}{c} \text{gamete of} \\ \text{dam } d \text{ of } i \end{array} \right]$$

$$\left[\begin{array}{c} \text{Mean effect} \\ \text{of a gamete} \\ \text{of } s \end{array} \right] + \left[\begin{array}{c} \text{Deviation with respect} \\ \text{to mean effect of} \\ \text{a gamete of } s \end{array} \right] + \left[\begin{array}{c} \text{Mean effect} \\ \text{of a gamete} \\ \text{of } d \end{array} \right] + \left[\begin{array}{c} \text{Deviation with respect} \\ \text{to mean effect of} \\ \text{a gamete of } s \end{array} \right]$$

$$a_i = \frac{1}{2} a_s + \frac{1}{2} a_d + \phi_i$$

ϕ_i = mendelian sampling



- $P = G + E$
- **P: Phenotype**
 - may be standardized, corrected, combined ...
- **G: Genotype**
 - depends on the objective of the analysis
 - may be incomplete (e.g., just « sire » or « additive » component)
- **E: Environment**
 - « $P - E$ »
 - may contain a genetic component (which was ignored in G)
 - includes what is not explained (residual)

Environment

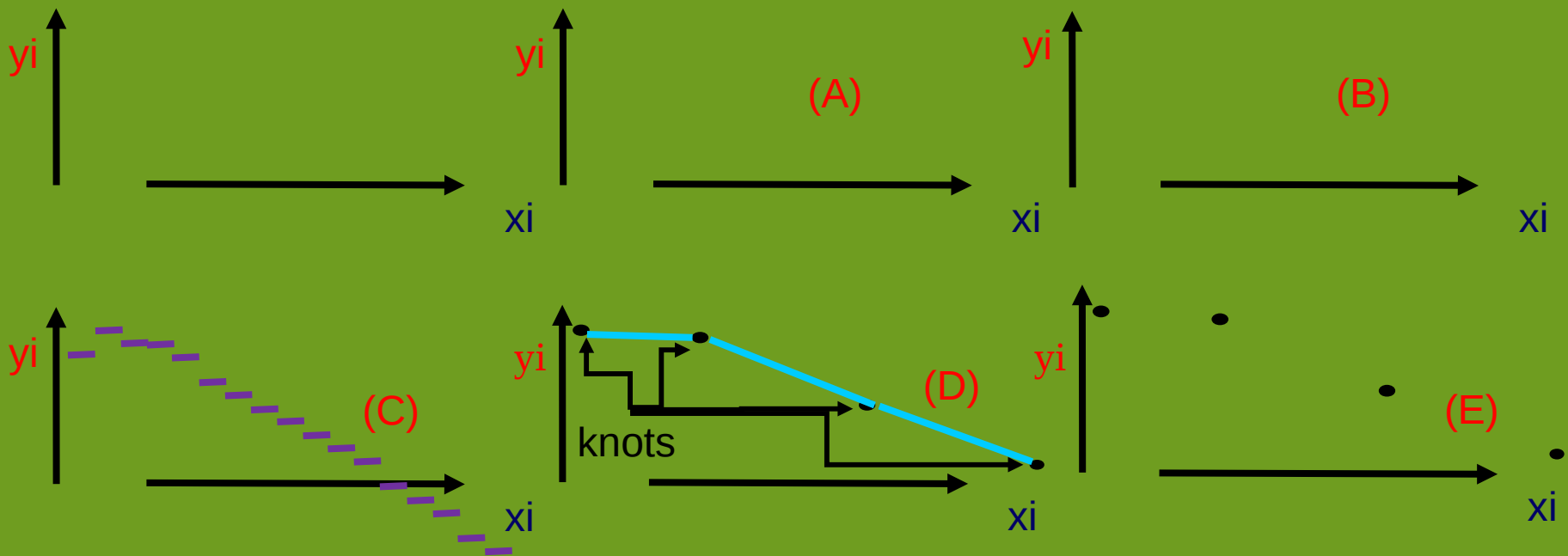
- **E**: « = $P - G$ »
 - may contain a genetic component
 - includes what is not explained (residual)
- **E** = an overall mean
 - + **effects that are identified and recorded**
 - + effects that are identified and not recorded
 - + effects that are not identified
 - + inaccuracies, errors
- By definition: **residual** = $e = y - E(y)$

residual

How to represent an environmental

effect?

- **Covariables:** example: age, stage of lactation
 - define a relation between the observation (y_i) and age (x_i)
 - parametric: linear (A), quadratic (B) polynomial
 - non parametric: define classes (C)
 - semi-parametric: splines (linear (D), quadratic (E), cubic)



How to represent an environmental effect 2 (2)

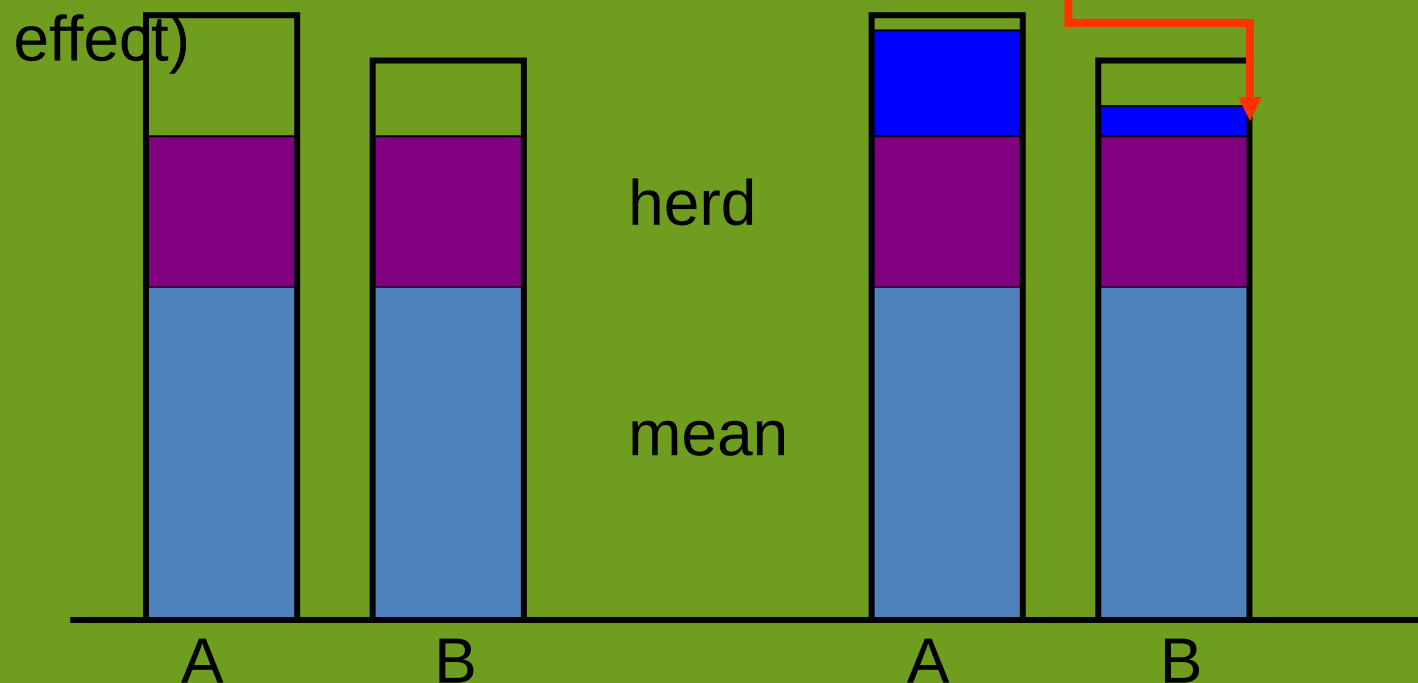
➤ Factors:

example: herd, year

- use a 0/1 covariable
- $y_i = \dots + 0^* h_1 + 0^* h_2 + \dots + 1^* h_i + \dots 0^* h_n + \dots$
- $y_i = \dots + h_i + \dots$ n levels

If we ignore important effects ...

- **Biased estimation** (eg, season



- **Accuracy overestimated**

If we include useless effects ...

➤ example: season effect

- semester
- trimester
- month
- week
- day ...

?

➤ no bias

➤ less precise estimation

➤ risk of disconnection

Disconnection ...

- = absence of information for a fair estimation

	herd 1	herd 2
bull A	2000	
bull B		3000

disconnected

	herd 1	herd 2
bull A	2000	3500
bull B		3000

connected

➤ Interactions

- example: herd x year x season x lactation number

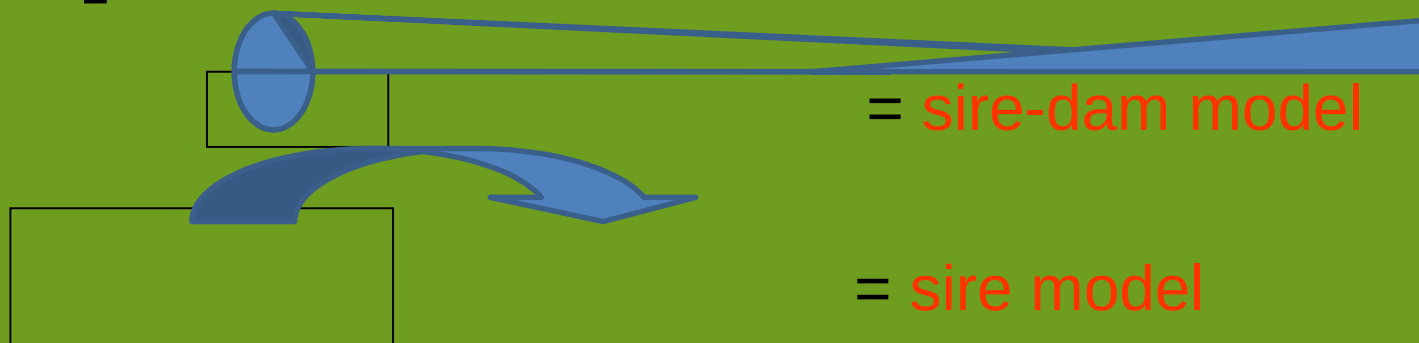
- balance between risk of bias, disconnection, loss of accuracy ...

Description of the genetic part

- (individual) **animal model** $y_i = \dots + a_i + e_i$

many effects to estimate (often more animals than observations)

- but: $a_i = \frac{1}{2}a_s + \frac{1}{2}a_d + \phi_i$



- interest: fewer parameters to estimate (e.g., only BV of males)
- but stricter assumptions !!! (eg: $E(\text{ud}) = 0$)

If G x E interaction...

- 1 trait per environment (1 G for each E group
e.g., production in each state = 1 trait)
 - **multivariate approach**
- Included in the model
(scale factors, heterogeneous variances, random regression models
« reaction norms »)

Random regression

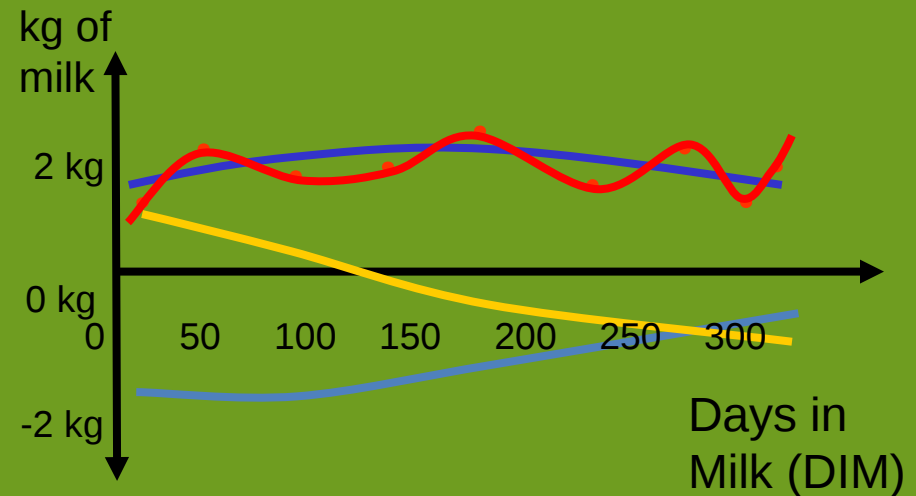
- Milk Yield varies with DIM

$$a(t) = \lambda_0 + \lambda_1 t + \lambda_2 t^2 + \lambda_3 t^3 + \lambda_4 t^4$$

$$a(t) = \alpha_1(t) \textcircled{a_1} + \alpha_2(t) \textcircled{a_2}$$

Mean
production

Persistency : explain 96.5-99.6% of total genetic variance



Fixed and random effects

- Assumption: **Polygenic model**:
many (infinite?) genes, each one with a small effect
sum of effects = a_i = additive genetic effect
Law of large numbers $\square a_i \sim N(\mu_a, \sigma_a^2)$ (assume $\mu_a = 0$)
 a_i is a **random** effect
- Similarly, residual = sum of many small effects = **random effect**
 $e_i \sim N(0, \sigma_e^2)$
- but « Year » or « age at calving » ? No distribution $\square$ **fixed** effect
- Quite arbitrary distinction (eg, herd effect?)
... unless you are a Bayesian...

Bayesian perspective (cf D. Gianola)

Bayes (1763), Laplace (1774)

$$p(A, B) = p(A|B) p(B) = p(B|A) p(A)$$

$$p(y, \theta) = p(y|\theta) p(\theta) = p(\theta|y) p(y)$$

Likelihood function:
Information coming
from the sample

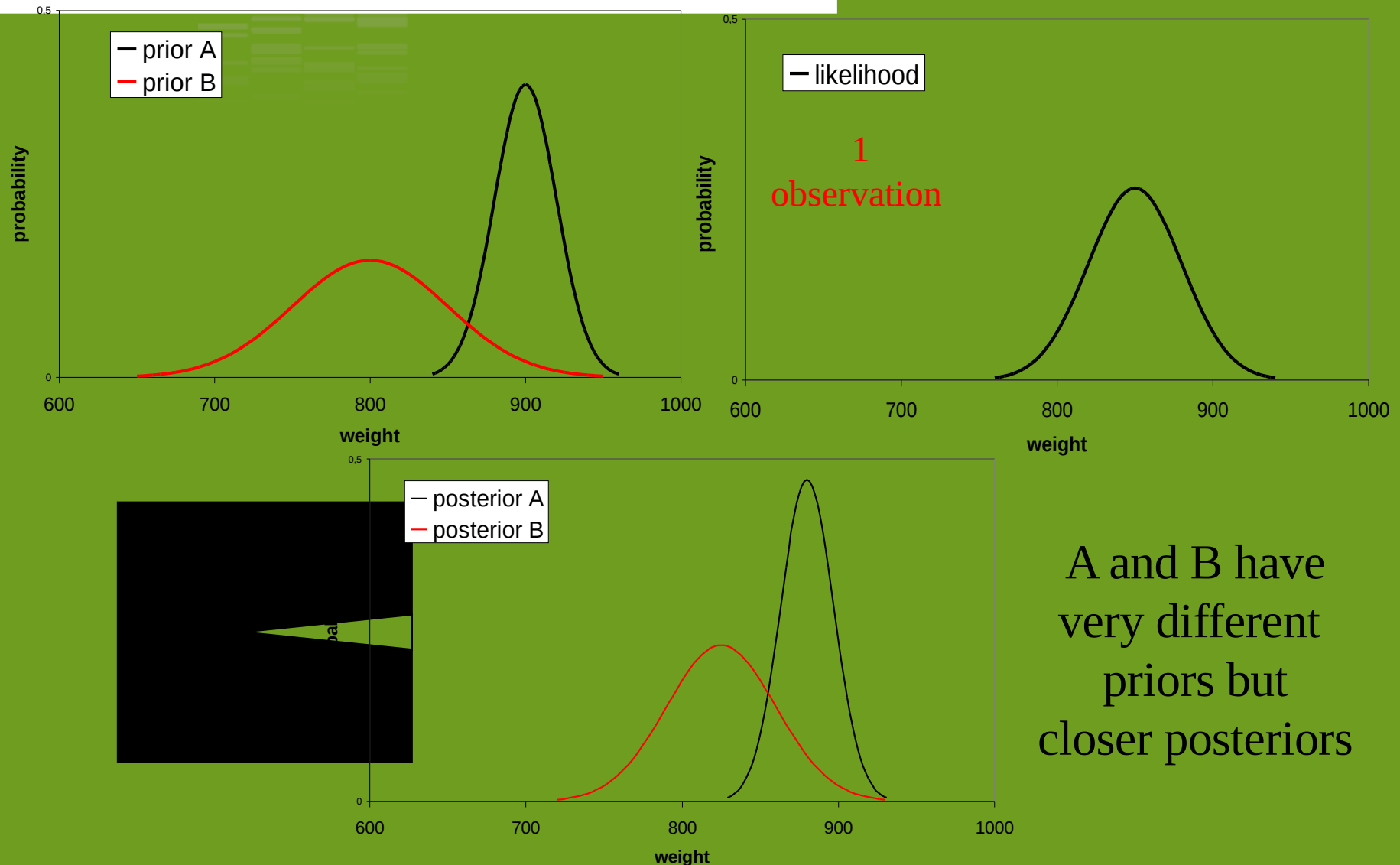
Marginal density of the data:
No information about θ

Prior distribution : what we know
about θ before looking
at the sample

Posterior density:
a summary of sample
and non sample information
about θ

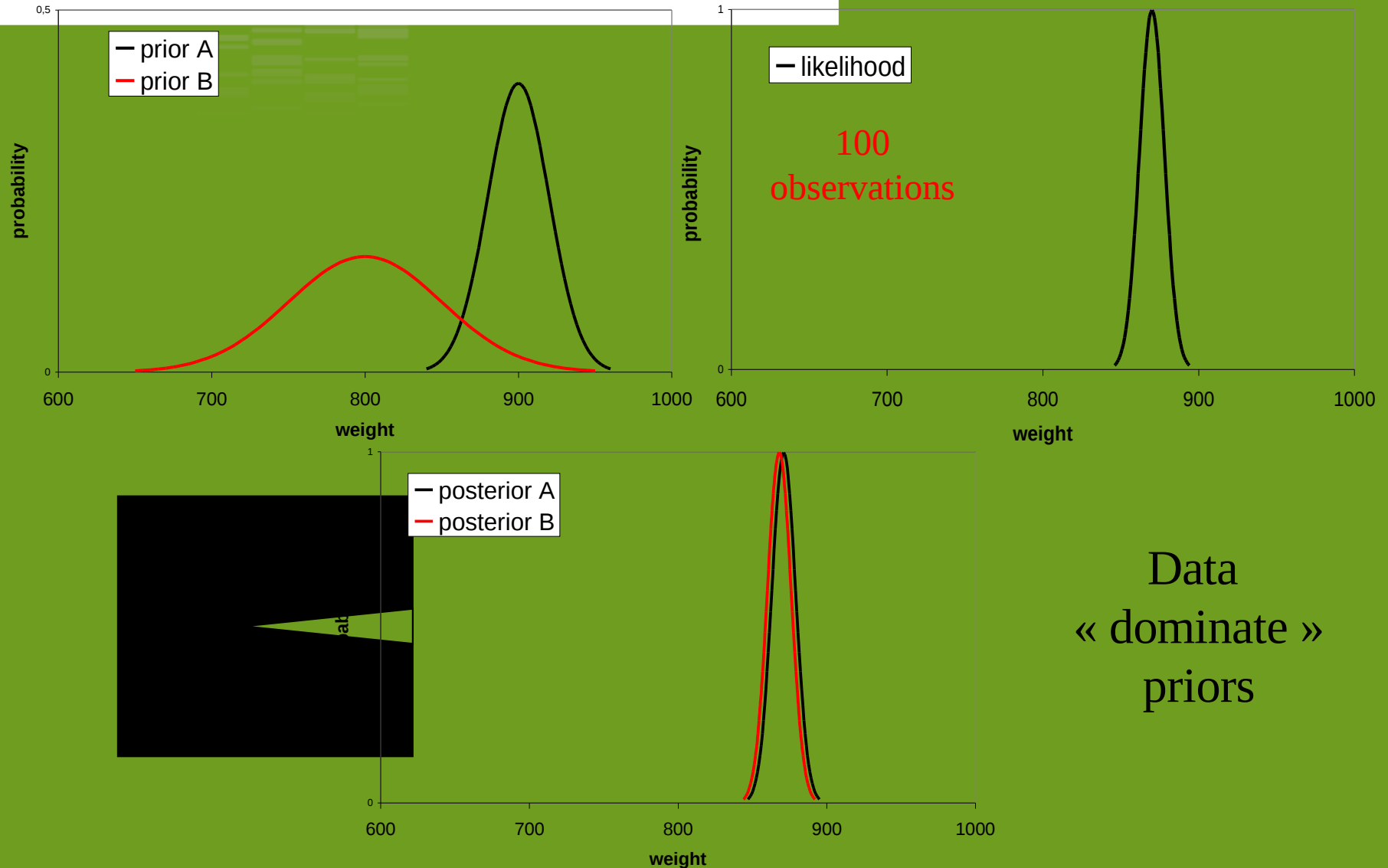
Remark: the formulas are complex but only the « philosophy » is important here: see next example

an example



A and B have
very different
priors but
closer posteriors

an example



Data
« dominate »
priors

« Bayesian » fixed and random effects

- What we are trying to do in a Bayesian analysis is to describe our « knowledge » through a distribution (posterior density)
 - For a Bayesian, all effects are **random**
- In many instances, there exists a Bayesian «point estimate» similar
 - with what we get with a frequentist approach (e.g., BLUP)
- To mimic the frequentist estimate of a fixed effect: assume no prior

My practical definition of fixed and random effects:

fixed effect: no prior information (or we don't use it)

random effect: we use prior information

Estimation