# Modelli Bayesiani Gerarchici

## Un’introduzione e qualche applicazione

Matilde Trevisani

Inferenza statistica Bayesiana

Trieste, 2022-2025

# Schema della lezione I

# Indice della sezione I

# Nella sezione

In questa lezione daremo dei cenni su come estendere la struttura dei modelli Bayesiani per costruire modelli gerarchici e rendere più flessibile la modellazione Bayesiana standard.

# Modelli gerarchici Bayesiani

## *Aliases ("Also Known As")*

I modelli gerarchici Bayesiani (*Bayesian Hierarchical Models*, HBM) costituiscono un'ampia classe di modelli e comprendono, anzi talvolta—in un'accezione più ristretta—sono "sinonimi" di

# Modelli gerarchici Bayesiani

## *Aliases ("Also Known As")*

I modelli gerarchici Bayesiani (*Bayesian Hierarchical Models*, HBM) costituiscono un'ampia classe di modelli e comprendono, anzi talvolta—in un'accezione più ristretta—sono "sinonimi" di

- modelli multilivello (*Multi-level models*)

# Modelli gerarchici Bayesiani
## *Aliases ("Also Known As")*

I modelli gerarchici Bayesiani (*Bayesian Hierarchical Models*, HBM) costituiscono un'ampia classe di modelli e comprendono, anzi talvolta—in un'accezione più ristretta—sono "sinonimi" di

- modelli multilivello (*Multi-level models*)
- modelli a coefficienti casuali (*Random Coefficients Models*) tra cui la ben nota tipologia dei

## Modelli gerarchici Bayesiani

### *Aliases ("Also Known As")*

I modelli gerarchici Bayesiani (*Bayesian Hierarchical Models*, HBM) costituiscono un'ampia classe di modelli e comprendono, anzi talvolta—in un'accezione più ristretta—sono "sinonimi" di

- modelli multilivello (*Multi-level models*)
- modelli a coefficienti casuali (*Random Coefficients Models*) tra cui la ben nota tipologia dei
  - modelli lineari generalizzati misti (*Generalized Linear Mixed Models*, GLMM)

# Modelli gerarchici Bayesiani

## Quando sono utili?

In genere, anche in approccio frequentista

# Modelli gerarchici Bayesiani

## Quando sono utili?

In genere, anche in approccio frequentista

- per caratteristiche oggettive o strutturali dei dati:

# Modelli gerarchici Bayesiani

## Quando sono utili?

In genere, anche in approccio frequentista

- per caratteristiche oggettive o strutturali dei dati: ogni volta che i dati sono multi-livello o si possono riconoscere diversi strati di variabilità;

# Modelli gerarchici Bayesiani

## Quando sono utili?

In genere, anche in approccio frequentista

- per caratteristiche **oggettive** o strutturali dei dati: ogni volta che i dati sono multi-livello o si possono riconoscere diversi strati di variabilità;
  - struttura gerarchica intrinseca

# Modelli gerarchici Bayesiani

## Quando sono utili?

In genere, anche in approccio frequentista

- per caratteristiche **oggettive** o strutturali dei dati: ogni volta che i dati sono multi-livello o si possono riconoscere diversi strati di variabilità;
  - struttura gerarchica intrinseca (e.g. studenti-scuole; misure ripetute; campioni multi-stadio, stratificati: sotto-popolazioni di una popolazione globale)

# Modelli gerarchici Bayesiani

## Quando sono utili?

In genere, anche in approccio frequentista

- per caratteristiche oggettive o **strutturali** dei dati: ogni volta che i dati sono multi-livello o si possono riconoscere diversi strati di variabilità;
  - struttura gerarchica intrinseca (e.g. studenti-scuole; misure ripetute; campioni multi-stadio, stratificati: sotto-popolazioni di una popolazione globale)
  - analisi molteplici (*multiple*)

# Modelli gerarchici Bayesiani

## Quando sono utili?

In genere, anche in approccio frequentista

- per caratteristiche oggettive o **strutturali** dei dati: ogni volta che i dati sono multi-livello o si possono riconoscere diversi strati di variabilità;
    - struttura gerarchica intrinseca (e.g. studenti-scuole; misure ripetute; campioni multi-stadio, stratificati: sotto-popolazioni di una popolazione globale)
    - analisi molteplici (*multiple*) (e.g. meta-analisi: sotto-popolazioni di una meta-popolazione).

# Modelli gerarchici Bayesiani

## Quando sono utili?

In genere, anche in approccio frequentista

- per caratteristiche oggettive o strutturali dei dati: ogni volta che i dati sono multi-livello o si possono riconoscere diversi strati di variabilità;
  - struttura gerarchica intrinseca (e.g. studenti-scuole; misure ripetute; campioni multi-stadio, stratificati: sotto-popolazioni di una popolazione globale)
  - analisi molteplici (*multiple*) (e.g. meta-analisi: sotto-popolazioni di una meta-popolazione).

Più specificatamente in approccio Bayesiano

# Modelli gerarchici Bayesiani

## Quando sono utili?

In genere, anche in approccio frequentista

- per caratteristiche oggettive o strutturali dei dati: ogni volta che i dati sono multi-livello o si possono riconoscere diversi strati di variabilità;
  - struttura gerarchica intrinseca (e.g. studenti-scuole; misure ripetute; campioni multi-stadio, stratificati: sotto-popolazioni di una popolazione globale)
  - analisi molteplici (*multiple*) (e.g. meta-analisi: sotto-popolazioni di una meta-popolazione).

Più specificatamente in approccio Bayesiano

- per ragioni soggettive o di modellazione

# Modelli gerarchici Bayesiani
## Quando sono utili?

In genere, anche in approccio frequentista

- per caratteristiche oggettive o strutturali dei dati: ogni volta che i dati sono multi-livello o si possono riconoscere diversi strati di variabilità;
  - struttura gerarchica intrinseca (e.g. studenti-scuole; misure ripetute; campioni multi-stadio, stratificati: sotto-popolazioni di una popolazione globale)
  - analisi molteplici (*multiple*) (e.g. meta-analisi: sotto-popolazioni di una meta-popolazione).

Più specificatamente in approccio Bayesiano

- per ragioni **soggettive** o di modellazione
  - incertezza nella informazione *a priori*

# Modelli gerarchici Bayesiani

## Quando sono utili?

In genere, anche in approccio frequentista

- per caratteristiche oggettive o strutturali dei dati: ogni volta che i dati sono multi-livello o si possono riconoscere diversi strati di variabilità;
  - struttura gerarchica intrinseca (e.g. studenti-scuole; misure ripetute; campioni multi-stadio, stratificati: sotto-popolazioni di una popolazione globale)
  - analisi molteplici (*multiple*) (e.g. meta-analisi: sotto-popolazioni di una meta-popolazione).

Più specificatamente in approccio Bayesiano

- per ragioni soggettive o di **modellazione**
  - incertezza nella informazione *a priori*
  - espansione del modello per renderlo più flessibile (e.g. modelli a variabili latenti, modelli mistura, modelli semiparametrici)

# Modelli gerarchici Bayesiani

## Quando sono utili?

In genere, anche in approccio frequentista

- per caratteristiche oggettive o strutturali dei dati: ogni volta che i dati sono multi-livello o si possono riconoscere diversi strati di variabilità;
  - struttura gerarchica intrinseca (e.g. studenti-scuole; misure ripetute; campioni multi-stadio, stratificati: sotto-popolazioni di una popolazione globale)
  - analisi molteplici (*multiple*) (e.g. meta-analisi: sotto-popolazioni di una meta-popolazione).

Più specificatamente in approccio Bayesiano

- per ragioni soggettive o di **modellazione**
  - incertezza nella informazione *a priori*
  - espansione del modello per renderlo più flessibile (e.g. modelli a variabili latenti, modelli mistura, modelli semiparametrici)
    - problemi di eterogeneità dei dati

## Modelli gerarchici Bayesiani

### Quando sono utili?

In genere, anche in approccio frequentista

- per caratteristiche oggettive o strutturali dei dati: ogni volta che i dati sono multi-livello o si possono riconoscere diversi strati di variabilità;
  - struttura gerarchica intrinseca (e.g. studenti-scuole; misure ripetute; campioni multi-stadio, stratificati: sotto-popolazioni di una popolazione globale)
  - analisi molteplici (*multiple*) (e.g. meta-analisi: sotto-popolazioni di una meta-popolazione).

Più specificatamente in approccio Bayesiano

- per ragioni soggettive o di **modellazione**
  - incertezza nella informazione *a priori*
  - espansione del modello per renderlo più flessibile (e.g. modelli a variabili latenti, modelli mistura, modelli semiparametrici)
    - problemi di eterogeneità dei dati
    - *model uncertainty*

# Modelli gerarchici Bayesiani

## Quando sono utili?

In genere, anche in approccio frequentista

- per caratteristiche oggettive o strutturali dei dati: ogni volta che i dati sono multi-livello o si possono riconoscere diversi strati di variabilità;
  - struttura gerarchica intrinseca (e.g. studenti-scuole; misure ripetute; campioni multi-stadio, stratificati: sotto-popolazioni di una popolazione globale)
  - analisi molteplici (*multiple*) (e.g. meta-analisi: sotto-popolazioni di una meta-popolazione).

Più specificatamente in approccio Bayesiano

- per ragioni soggettive o di modellazione
  - incertezza nella informazione *a priori*
  - espansione del modello per renderlo più flessibile (e.g. modelli a variabili latenti, modelli mistura, modelli semiparametrici)
    - problemi di eterogeneità dei dati
    - *model uncertainty*

# Modelli gerarchici Bayesiani
## Quando sono utili?

L'idea di base è quella del **prendere in prestito forza**, **mettendo in comune le informazioni** da fonti di dati correlate,

così come l'idea di consentire la variazione tramite effetti casuali tra le unità per tenere conto della extra-variazione o **sovradispersione** dei dati

ma anche per problemi di **incertezza** e **robustezza**: in un ambiente non-informativo, aggiungere livelli alla a priori aumenta il grado di incertezza e di robustezza della distribuzione a priori (soprattutto nelle strutture coniugate, altrimenti troppo restrittive)

e per **semplificare il calcolo** bayesiano, e.g.. per il calcolo delle distribuzioni condizionali complete (MCMC), *mattoni* per la modellazione grafica

# Nella sezione

# Definizione di modelli gerarchici Bayesiani

## Una *a priori* specificata in modo gerarchico

Da qui in avanti, costruiremo un HBM, i.e. un modello avente diversi livelli di distribuzioni condizionali a priori.

## Definizione di modelli gerarchici Bayesiani

### Una *a priori* specificata in modo gerarchico

Da qui in avanti, costruiremo un HBM, i.e. un modello avente diversi livelli di distribuzioni condizionali a priori.

# Definizione di modelli gerarchici Bayesiani

## Una *a priori* specificata in modo gerarchico

Da qui in avanti, costruiremo un HBM, i.e. un modello avente diversi livelli di distribuzioni condizionali a priori.

Una definizione formale è la seguente Robert (2007):

## Definizione

Un modello gerarchico Bayesiano

## Definizione

Un modello gerarchico Bayesiano è un modello statistico Bayesiano

$$p(\mathbf{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})$$

## Definizione

Un modello gerarchico Bayesiano è un modello statistico Bayesiano

$$p(\mathbf{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})$$

dove la distribuzione a priori $p(\boldsymbol{\theta})$ viene scomposta in distribuzioni condizionali

$$p_1(\boldsymbol{\theta}|\boldsymbol{\theta}_1) \;\; p_2(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) \quad \dots \;\; \dots \quad p_m(\boldsymbol{\theta}_{m-1}|\boldsymbol{\theta}_m)$$

## Definizione

Un modello gerarchico Bayesiano è un modello statistico Bayesiano

$$p(\mathbf{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})$$

dove la distribuzione a priori $p(\boldsymbol{\theta})$ viene scomposta in distribuzioni condizionali

$$p_1(\boldsymbol{\theta}|\boldsymbol{\theta}_1) \;\; p_2(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) \quad \ldots \;\; \ldots \quad p_m(\boldsymbol{\theta}_{m-1}|\boldsymbol{\theta}_m)$$

e una distribuzione marginale $p_{m+1}(\boldsymbol{\theta}_m)$

## Definizione

Un modello gerarchico Bayesiano è un modello statistico Bayesiano

$$p(\mathbf{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})$$

dove la distribuzione a priori $p(\boldsymbol{\theta})$ viene scomposta in distribuzioni condizionali

$$p_1(\boldsymbol{\theta}|\boldsymbol{\theta}_1) \;\; p_2(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) \quad \ldots \;\; \ldots \quad p_m(\boldsymbol{\theta}_{m-1}|\boldsymbol{\theta}_m)$$

e una distribuzione marginale $p_{m+1}(\boldsymbol{\theta}_m)$ di modo che

$$p(\boldsymbol{\theta}) = \int_{\Theta_1 \times \ldots \Theta_m} p_1(\boldsymbol{\theta}|\boldsymbol{\theta}_1)\, p_2(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) \ldots p_m(\boldsymbol{\theta}_{m-1}|\boldsymbol{\theta}_m)\, p_{m+1}(\boldsymbol{\theta}_m)\, d\boldsymbol{\theta}_1 \ldots \boldsymbol{\theta}_m.$$

## Definizione

Un modello gerarchico Bayesiano è un modello statistico Bayesiano

$$p(\mathbf{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})$$

dove la distribuzione a priori $p(\boldsymbol{\theta})$ viene scomposta in distribuzioni condizionali

$$p_1(\boldsymbol{\theta}|\boldsymbol{\theta}_1) \;\; p_2(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) \quad \ldots \;\; \ldots \quad p_m(\boldsymbol{\theta}_{m-1}|\boldsymbol{\theta}_m)$$

e una distribuzione marginale $p_{m+1}(\boldsymbol{\theta}_m)$ di modo che

$$p(\boldsymbol{\theta}) = \int_{\Theta_1\times\ldots\Theta_m} p_1(\boldsymbol{\theta}|\boldsymbol{\theta}_1)\, p_2(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2)\, \ldots\, p_m(\boldsymbol{\theta}_{m-1}|\boldsymbol{\theta}_m)\, p_{m+1}(\boldsymbol{\theta}_m)\, d\boldsymbol{\theta}_1 \ldots \boldsymbol{\theta}_m.$$

La a priori $p(\boldsymbol{\theta})$ si ottiene integrando sugli *iperparametri* $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m$.

## Definizione

Un modello gerarchico Bayesiano è un modello statistico Bayesiano

$$p(\mathbf{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})$$

dove la distribuzione a priori $p(\boldsymbol{\theta})$ viene scomposta in distribuzioni condizionali

$$p_1(\boldsymbol{\theta}|\boldsymbol{\theta}_1) \;\; p_2(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) \quad \ldots \;\; \ldots \quad p_m(\boldsymbol{\theta}_{m-1}|\boldsymbol{\theta}_m)$$

e una distribuzione marginale $p_{m+1}(\boldsymbol{\theta}_m)$ di modo che

$$p(\boldsymbol{\theta}) = \int_{\Theta_1 \times \ldots \Theta_m} p_1(\boldsymbol{\theta}|\boldsymbol{\theta}_1)\, p_2(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2) \, \ldots \, p_m(\boldsymbol{\theta}_{m-1}|\boldsymbol{\theta}_m)\, p_{m+1}(\boldsymbol{\theta}_m)\, d\boldsymbol{\theta}_1 \ldots \boldsymbol{\theta}_m.$$

I parametri $\boldsymbol{\theta}_i$ sono detti iperparametri di livello $i$.

## Essence of the hierarchical modeling

It can be pragmatic to view modeling problems in terms of three entities (all of which have stochastic elements).

- First is the **data** which is presumed to be drawn from some facet(s) of the underlying process.

## Essence of the hierarchical modeling

It can be pragmatic to view modeling problems in terms of three entities (all of which have stochastic elements).

I First is the **data** which is presumed to be drawn from some facet(s) of the underlying process.

II Second is the **process** specification itself which involves unknowns that will be estimated as parameters.

## Essence of the hierarchical modeling

It can be pragmatic to view modeling problems in terms of three entities (all of which have stochastic elements).

I. First is the **data** which is presumed to be drawn from some facet(s) of the underlying process.
II. Second is the **process** specification itself which involves unknowns that will be estimated as parameters.
III. Third we have **parameters** that will be expected to vary depending how and where the data were obtained.

## Essence of the hierarchical modeling

It can be pragmatic to view modeling problems in terms of three entities (all of which have stochastic elements).

I First is the **data** which is presumed to be drawn from some facet(s) of the underlying process.

II Second is the **process** specification itself which involves unknowns that will be estimated as parameters.

III Third we have **parameters** that will be expected to vary depending how and where the data were obtained.

With this three-part structure in mind, we are prepared to extend the basic version of a bayesian model to more levels in a general and flexible way.

## Essence of the hierarchical modeling

$$
\begin{aligned}
f(data, process, parameters) \\
&\propto f(data|process, parameters) \\
&\quad \times f(process|parameters) \\
&\quad \times f(parameters)
\end{aligned}
$$

# Essence of the hierarchical modeling

$$
\begin{aligned}
f(data, process, parameters) \\
&\propto f(data|process, parameters) \\
&\quad \times f(process|parameters) \\
&\quad \times f(parameters)
\end{aligned}
$$

$$
\begin{aligned}
&[data, process, parameters] \propto \\
&\qquad [data|process, parameters][process|parameters][parameters]
\end{aligned}
$$

## Essence of the hierarchical modeling

Advantages of this hierarchical perspective

- the ability to construct complex models from simple conditional relationships (we need not think about the entire joint distribution for the problem, only the components)
- we can relax customary requirements for independent data (conditional independence is enough)
- by attaching randomness to what we observe as well as to what we do not observe, we build a fully Bayesian specification and look at the posterior distribution of every unobservable given every observable. (though the posterior may be high domensional and analytically intractable, we can take advantage of the Bayesian computational tools)

# Nella sezione

# Un modello gerarchico non Bayesiano

## Un modello gerarchico non Bayesiano

$$x|N \sim \ Bin(N, p)$$

- Il numero $x$ di uova (gattini, etc.) che riescono a sopravvivere

## Un modello gerarchico non Bayesiano

$$
\begin{aligned}
x|N &\sim Bin(N, p) \\
N &\sim Pois(\lambda)
\end{aligned}
$$

- Il numero $x$ di uova (gattini, etc.) che riescono a sopravvivere
- in una nidiata la cui numerosità $N$ è sconosciuta

## Un modello gerarchico non Bayesiano

$$
\begin{aligned}
x|N &\sim Bin(N, p) \\
N &\sim Pois(\lambda)
\end{aligned}
$$

- Il numero $x$ di uova (gattini, etc.) che riescono a sopravvivere
- in una nidiata la cui numerosità $N$ è sconosciuta

La distribuzione del numero $x$ di sopravvissuti è dunque gerarchica.

## Un modello gerarchico non Bayesiano

$$
\begin{aligned}
x|N &\sim Bin(N, p) \\
N &\sim Pois(\lambda)
\end{aligned}
$$

- Il numero $x$ di uova (gattini, etc.) che riescono a sopravvivere
- in una nidiata la cui numerosità $N$ è sconosciuta

entrambe le variabili sono osservabili e entrambi i livelli sono strutturali

# Modello (lineare) a coefficienti casuali

## Modello (lineare) a coefficienti casuali

$$y|\beta \quad \sim \quad N_n(X\boldsymbol{\beta}, \Sigma_1)$$

- un modello (lineare) di regressione standard

## Modello (lineare) a coefficienti casuali

$$
\begin{aligned}
y|\boldsymbol{\theta} &\sim N_n(\boldsymbol{\theta}, \Sigma_1) \\
\boldsymbol{\theta}|\boldsymbol{\beta} &\sim N_n(X\boldsymbol{\beta}, \Sigma_2)
\end{aligned}
$$

- un modello (lineare) di regressione standard
- diventa un modello a coefficienti casuali se $\boldsymbol{\theta} = X\boldsymbol{\beta} + Z\boldsymbol{\eta}$, $Z\boldsymbol{\eta}$ effetti casuali, $X\boldsymbol{\beta}$ effetti fissi

La media $\theta$ di $y$ viene decomposta in effetti fissi, $X\beta$, e in effetti casuali, $Z\eta$, dove $\eta$ è normale con media $0$.

## Modello (lineare) a coefficienti casuali

$$
\begin{aligned}
y|\boldsymbol{\theta} &\sim N_n(\boldsymbol{\theta}, \Sigma_1) \\
\boldsymbol{\theta}|\boldsymbol{\beta} &\sim N_n(X\boldsymbol{\beta}, \Sigma_2)
\end{aligned}
$$

- un modello a coefficienti casuali

## Modello (lineare) a coefficienti casuali

$$
\begin{aligned}
y|\boldsymbol{\theta} &\sim N_n(\boldsymbol{\theta}, \Sigma_1) \\
\boldsymbol{\theta}|\boldsymbol{\beta} &\sim N_n(X\boldsymbol{\beta}, \Sigma_2) \\
\boldsymbol{\beta} &\sim N_p(Z\xi, \Sigma_3)
\end{aligned}
$$

- un modello a coefficienti casuali
- che diviene propriamente un modello gerarchico Bayesiano a 2 stadi se viene completamente specificato, cioè se viene assegnata una distribuzione a priori agli iperparametri $\beta$.

## Un modello di regressione Bayesiano a 2 stadi

$$y|\beta \quad \sim \quad N_n(X\beta, \sigma^2 I_n)$$

## Un modello di regressione Bayesiano a 2 stadi

$$
\begin{aligned}
y|\beta &\sim N_n(X\beta, \sigma^2 I_n) \\
\beta|\eta &\sim N_p(\eta\mathbf{1}, \sigma_\beta^2 I_p)
\end{aligned}
$$

- Per ragioni strutturali, i coefficienti di regressione sono simili, quindi un modello scambiabile per i $\beta$'s viene specificato

Ad esempio, i $\beta$'s possono descrivere i tassi di investimento di differenti case automobilistiche Europee, per le quali i tassi sono altamente simili.

## Un modello di regressione Bayesiano a 2 stadi

$$\begin{aligned} y|\beta &\sim N_n(X\beta, \sigma^2 I_n) \\ \beta|\eta &\sim N_p(\eta\mathbf{1}, \sigma_\beta^2 I_p) \\ \eta &\sim N_1(\eta_0, \sigma_0^2) \end{aligned}$$

- Per ragioni strutturali, i coefficienti di regressione sono simili, quindi un modello scambiabile per i $\beta$'s viene specificato
- Una a priori sull'iperparametro (scalare) $\eta$ può assumere una forma propria come sopra, quando qualche informazion aggiuntiva sia disponibile

## Un modello di regressione Bayesiano a 2 stadi

$$\begin{aligned} y|\beta &\sim N_n(X\beta, \sigma^2 I_n) \\ \beta|\eta &\sim N_p(\eta\mathbf{1}, \sigma^2_\beta I_p) \\ \eta &\sim N_1(\eta_0, \sigma_0^2) \quad \text{(altrimenti)} \quad \pi_2(\eta) = 1 \end{aligned}$$

- Per ragioni strutturali, i coefficienti di regressione sono simili, quindi un modello scambiabile per i $\beta$'s viene specificato
- Una a priori sull'iperparametro (scalare) $\eta$ può assumere una forma propria come sopra, quando qualche informazion aggiuntiva sia disponibile, altrimenti, può essere non-informativa

## Il più semplice HBM

Il modello scambiabile, 1-way anova, ...

## Il più semplice HBM

Il modello scambiabile, 1-way anova, ...

$$\mathbf{y}_j|\theta_j \sim N_{n_j}(\theta_j, \sigma^2 I_{n_j}) \qquad\qquad j = 1, \ldots, J$$

- modello scambiabile per le $y$'s

## Il più semplice HBM

Il modello scambiabile, 1-way anova, ...

$$
\begin{aligned}
\mathbf{y}_j|\theta_j &\sim N_{n_j}(\theta_j, \sigma^2 I_{n_j}) && j = 1, \ldots, J \\
\boldsymbol{\theta}|\eta &\sim N_J(\eta \mathbf{1}_J, \sigma_\eta^2 I_J) && \boldsymbol{\theta} = (\theta_1, \ldots, \theta_J)^T
\end{aligned}
$$

- modello scambiabile per le $y$'s
- modello scambiabile per i $\theta$'s

## Il più semplice HBM

Il modello scambiabile, 1-way anova, ...

$$
\begin{aligned}
\mathbf{y}_j|\theta_j &\sim N_{n_j}(\theta_j, \sigma^2 I_{n_j}) & j &= 1, \ldots, J \\
\boldsymbol{\theta}|\eta &\sim N_J(\eta \mathbf{1}_J, \sigma_\eta^2 I_J) & \boldsymbol{\theta} &= (\theta_1, \ldots, \theta_J)^T \\
\eta &\sim N_1(\eta_0, \sigma_0^2)
\end{aligned}
$$

- modello scambiabile per le $y$'s
- modello scambiabile per i $\theta$'s
- una a priori per $\eta$ (ad es. come sopra) che consenta di modellare congiuntamente i $J$ esperimenti.

## Il più semplice HBM

Il modello scambiabile, 1-way anova, ...

$$
\begin{aligned}
\mathbf{y}_j|\theta_j &\sim N_{n_j}(\theta_j, \sigma^2 I_{n_j}) & \qquad j &= 1, \ldots, J \\
\boldsymbol{\theta}|\eta &\sim N_J(\eta \mathbf{1}_J, \sigma^2_\eta I_J) & \qquad \boldsymbol{\theta} &= (\theta_1, \ldots, \theta_J)^T \\
\eta &\sim N_1(\eta_0, \sigma^2_0)
\end{aligned}
$$

Mettere insieme (*pooling*) tutte le unità richiede un'assunzione di **scambiabilità** ($\sim$ simmetria degli $\theta_j$'s a priori)

## Il più semplice HBM

Il modello scambiabile, 1-way anova, ...

$$
\begin{aligned}
\mathbf{y}_j|\theta_j &\sim N_{n_j}(\theta_j, \sigma^2 I_{n_j}) & \qquad j &= 1, \ldots, J \\
\boldsymbol{\theta}|\eta &\sim N_J(\eta \mathbf{1}_J, \sigma^2_\eta I_J) & \qquad \boldsymbol{\theta} &= (\theta_1, \ldots, \theta_J)^T \\
\eta &\sim N_1(\eta_0, \sigma^2_0)
\end{aligned}
$$

Si consideri $J$ esperimenti indipendenti (e.g. gruppi di pazienti soggetti a differenti trattamenti), con l'esperimento $j$ che stima il parametro $\theta_j$ (effetto del trattamento) da $n_j$ osservazioni di dati indipendenti normalmente distribuiti (misurazioni di una variabile osservabile $Y$), $y_{ij}$, ognuno con varianza dell'errore $\sigma^2$ nota.

## Il più semplice HBM

Il modello scambiabile, 1-way anova, ...

$$\begin{aligned} \mathbf{y}_j|\theta_j &\sim N_{n_j}(\theta_j, \sigma^2 I_{n_j}) & j &= 1, \ldots, J \\ \boldsymbol{\theta}|\eta &\sim N_J(\eta \mathbf{1}_J, \sigma^2_\eta I_J) & \boldsymbol{\theta} &= (\theta_1, \ldots, \theta_J)^T \\ \eta &\sim N_1(\eta_0, \sigma^2_0) \end{aligned}$$

La mistura iid per $\boldsymbol{\theta}$, $p(\boldsymbol{\theta}) = \int \prod_{j=1}^J p(\theta_j|\eta)p(\eta)d\eta$, è una distribuzione scambiabile.

# Nella sezione

# Lo scheletro essenziale del più semplice HBM

# Lo scheletro essenziale del più semplice HBM

$$\mathbf{Y}|\boldsymbol{\theta} \quad \sim \quad p(\mathbf{y}|\boldsymbol{\theta})$$

## Lo scheletro essenziale del più semplice HBM

$$
\begin{aligned}
\mathbf{Y}|\boldsymbol{\theta} &\sim p(\mathbf{y}|\boldsymbol{\theta}) \\
\boldsymbol{\theta}|\boldsymbol{\eta} &\sim p(\boldsymbol{\theta}|\boldsymbol{\eta})
\end{aligned}
$$

# Lo scheletro essenziale del più semplice HBM

$$
\begin{aligned}
\mathbf{Y}|\boldsymbol{\theta} &\sim p(\mathbf{y}|\boldsymbol{\theta}) \\
\boldsymbol{\theta}|\boldsymbol{\eta} &\sim p(\boldsymbol{\theta}|\boldsymbol{\eta}) \\
\boldsymbol{\eta} &\sim p(\boldsymbol{\eta})
\end{aligned}
$$

## Lo scheletro essenziale del più semplice HBM

$$\begin{aligned} \mathbf{Y}|\boldsymbol{\theta} &\sim p(\mathbf{y}|\boldsymbol{\theta}) \\ \boldsymbol{\theta}|\boldsymbol{\eta} &\sim p(\boldsymbol{\theta}|\boldsymbol{\eta}) \\ \boldsymbol{\eta} &\sim p(\boldsymbol{\eta}) \end{aligned}$$

- Se gli iperparametri $\eta$ sono noti,

## Lo scheletro essenziale del più semplice HBM

$$
\begin{aligned}
\mathbf{Y}|\boldsymbol{\theta} &\sim p(\mathbf{y}|\boldsymbol{\theta}) \\
\boldsymbol{\theta}|\eta &\sim p(\boldsymbol{\theta}|\eta)
\end{aligned}
$$

- Se gli iperparametri $\eta$ sono noti, allora $\eta$ sono soppressi e torniamo al modello Bayesiano di base;

## Lo scheletro essenziale del più semplice HBM

$$
\begin{aligned}
\mathbf{Y}|\boldsymbol{\theta} &\sim p(\mathbf{y}|\boldsymbol{\theta}) \\
\boldsymbol{\theta}|\boldsymbol{\eta} &\sim p(\boldsymbol{\theta}|\boldsymbol{\eta}) \\
\boldsymbol{\eta} &\sim p(\boldsymbol{\eta})
\end{aligned}
$$

- Se gli iperparametri $\eta$ sono noti, allora $\eta$ sono soppressi e torniamo al modello Bayesiano di base;
- se $\eta$ sono non-noti, allora sono possibili due approcci:

## Lo scheletro essenziale del più semplice HBM

$$
\begin{aligned}
\mathbf{Y}|\boldsymbol{\theta} &\sim p(\mathbf{y}|\boldsymbol{\theta}) \\
\boldsymbol{\theta}|\boldsymbol{\eta} &\sim p(\boldsymbol{\theta}|\boldsymbol{\eta}) \\
\boldsymbol{\eta} &\sim p(\boldsymbol{\eta})
\end{aligned}
$$

- Se gli iperparametri $\eta$ sono noti, allora $\eta$ sono soppressi e torniamo al modello Bayesiano di base;
- se $\eta$ sono non-noti, allora sono possibili due approcci:
  - Empirical Bayes (EB);

## Lo scheletro essenziale del più semplice HBM

$$
\begin{aligned}
\mathbf{Y}|\boldsymbol{\theta} &\sim p(\mathbf{y}|\boldsymbol{\theta}) \\
\boldsymbol{\theta}|\boldsymbol{\eta} &\sim p(\boldsymbol{\theta}|\boldsymbol{\eta}) \\
\boldsymbol{\eta} &\sim p(\boldsymbol{\eta})
\end{aligned}
$$

- Se gli iperparametri $\eta$ sono noti, allora $\eta$ sono soppressi e torniamo al modello Bayesiano di base;
- se $\eta$ sono non-noti, allora sono possibili due approcci:
  - Empirical Bayes (EB);
  - Full Bayes ((full) HB).

## Full Bayes o Empirical Bayes ?

- L'analisi Bayesiana completa specifica un modello probabilistico per ogni variabile (sia osservabile che non-osservabile),

## Full Bayes o Empirical Bayes ?

- L'analisi Bayesiana completa specifica un modello probabilistico per ogni variabile (sia osservabile che non-osservabile), quindi studia la distribuzione congiunta a posteriori della totalità dei parametri (o variabili non osservabili).

## Full Bayes o Empirical Bayes ?

- L'analisi Bayesiana completa specifica un modello probabilistico per ogni variabile (sia osservabile che non-osservabile), quindi studia la distribuzione congiunta a posteriori della totalità dei parametri (o variabili non osservabili).
- L'analisi Bayesiana Empirica ottiene delle stime per gli iperparametri dai dati,

## Full Bayes o Empirical Bayes ?

- L’analisi Bayesiana completa specifica un modello probabilistico per ogni variabile (sia osservabile che non-osservabile), quindi studia la distribuzione congiunta a posteriori della totalità dei parametri (o variabili non osservabili).
- L’analisi Bayesiana Empirica ottiene delle stime per gli iperparametri dai dati, quindi studia la distribuzione congiunta a posteriori dei rimanenti parametri condizionatamente a tali stime;

## Full Bayes o Empirical Bayes ?

- L’analisi Bayesiana completa specifica un modello probabilistico per ogni variabile (sia osservabile che non-osservabile), quindi studia la distribuzione congiunta a posteriori della totalità dei parametri (o variabili non osservabili).
- L’analisi Bayesiana Empirica ottiene delle stime per gli iperparametri dai dati, quindi studia la distribuzione congiunta a posteriori dei rimanenti parametri condizionatamente a tali stime;
  - non integra pienamente tutta l’incertezza

## Full Bayes o Empirical Bayes ?

- L'analisi Bayesiana completa specifica un modello probabilistico per ogni variabile (sia osservabile che non-osservabile), quindi studia la distribuzione congiunta a posteriori della totalità dei parametri (o variabili non osservabili).
- L'analisi Bayesiana Empirica ottiene delle stime per gli iperparametri dai dati, quindi studia la distribuzione congiunta a posteriori dei rimanenti parametri condizionatamente a tali stime;
  - non integra pienamente tutta l'incertezza
  - usa i dati due volte $\rightarrow$ la precisione risulta sovrastimata

## Full Bayes o Empirical Bayes ?

- L'analisi Bayesiana completa specifica un modello probabilistico per ogni variabile (sia osservabile che non-osservabile), quindi studia la distribuzione congiunta a posteriori della totalità dei parametri (o variabili non osservabili).
- L'analisi Bayesiana Empirica ottiene delle stime per gli iperparametri dai dati, quindi studia la distribuzione congiunta a posteriori dei rimanenti parametri condizionatamente a tali stime;
  - non integra pienamente tutta l'incertezza
  - usa i dati due volte $\rightarrow$ la precisione risulta sovrastimata
  - confonde i concetti "a priori" e "empirico" (che sono epistemologicamente ben distinti !)

## Distribuzione a posteriori

- se gli iperparametri $\eta$ sono noti, l'inferenza su $\theta$ si basa sulla sua distribuzione a posteriori come si ottiene dal Teorema di Bayes

## Distribuzione a posteriori

- se gli iperparametri $\eta$ sono noti, l'inferenza su $\boldsymbol{\theta}$ si basa sulla sua distribuzione a posteriori come si ottiene dal Teorema di Bayes

**Bayes Theorem**

$$p(\boldsymbol{\theta}|\mathbf{y},\eta) = \frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\eta)}{p(\mathbf{y}|\eta)}$$

## Distribuzione a posteriori

- se gli iperparametri $\eta$ sono noti, l'inferenza su $\boldsymbol{\theta}$ si basa sulla sua distribuzione a posteriori come si ottiene dal Teorema di Bayes

**Bayes Theorem**

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})}$$

# Distribuzione a posteriori

- Con $\eta$ non-noti

## Distribuzione a posteriori

- Con $\boldsymbol{\eta}$ non-noti
  - il metodo EB “plugs-in” una stima di $\boldsymbol{\eta}$, $\hat{\eta} \equiv \hat{\boldsymbol{\eta}}(\mathbf{y})$ in $p(\boldsymbol{\theta}|\mathbf{y}, \boldsymbol{\eta})$

## Distribuzione a posteriori

- Con $\boldsymbol{\eta}$ non-noti
  - il metodo EB "plugs-in" una stima di $\boldsymbol{\eta}$, $\hat{\eta} \equiv \hat{\boldsymbol{\eta}}(\mathbf{y})$ in $p(\boldsymbol{\theta}|\mathbf{y}, \boldsymbol{\eta})$ (e.g., MLE dalla distribuzione marginale $p(\mathbf{y}|\boldsymbol{\eta})$),

## Distribuzione a posteriori

- Con $\boldsymbol{\eta}$ non-noti
  - il metodo EB "plugs-in" una stima di $\boldsymbol{\eta}$, $\hat{\eta} \equiv \hat{\boldsymbol{\eta}}(\mathbf{y})$ in $p(\boldsymbol{\theta}|\mathbf{y}, \boldsymbol{\eta})$ i.e.

**il metodo EB**

usa la "distribuzione a posteriori stimata"

$$p(\boldsymbol{\theta}|\mathbf{y}, \hat{\eta})$$

## Distribuzione a posteriori

- Con $\eta$ non-noti
  - il metodo EB “plugs-in” una stima di $\boldsymbol{\eta}$, $\hat{\boldsymbol{\eta}} \equiv \hat{\boldsymbol{\eta}}(\mathbf{y})$ in $p(\boldsymbol{\theta}|\mathbf{y}, \boldsymbol{\eta})$ i.e.

**il metodo EB**

usa la “distribuzione a posteriori stimata”

$$p(\boldsymbol{\theta}|\mathbf{y}, \hat{\boldsymbol{\eta}})$$

($\rightarrow$ approssima una analisi Bayesiana piena (*full*)).

## Distribuzione a posteriori

- Con $\boldsymbol{\eta}$ non-noti
  - il metodo EB "plugs-in" una stima di $\boldsymbol{\eta}$, $\hat{\boldsymbol{\eta}} \equiv \hat{\boldsymbol{\eta}}(\mathbf{y})$ in $p(\boldsymbol{\theta}|\mathbf{y}, \boldsymbol{\eta})$ i.e.

    **il metodo EB**

    usa la "distribuzione a posteriori stimata"
    $$p(\boldsymbol{\theta}|\mathbf{y}, \hat{\boldsymbol{\eta}})$$

  - il metodo HB specifica una a priori $p(\boldsymbol{\eta})$

## Distribuzione a posteriori

- Con $\boldsymbol{\eta}$ non-noti
  - il metodo EB "plugs-in" una stima di $\boldsymbol{\eta}$, $\hat{\boldsymbol{\eta}} \equiv \hat{\boldsymbol{\eta}}(\mathbf{y})$ in $p(\boldsymbol{\theta}|\mathbf{y},\boldsymbol{\eta})$ i.e.

**il metodo EB**

usa la "distribuzione a posteriori stimata"

$$p(\boldsymbol{\theta}|\mathbf{y},\hat{\boldsymbol{\eta}})$$

  - il metodo HB specifica una a priori $p(\boldsymbol{\eta})$ e

**Il metodo (full) HB**

calcola

$$p(\boldsymbol{\theta}|\mathbf{y}) = \int p(\boldsymbol{\theta}|\mathbf{y},\boldsymbol{\eta})p(\boldsymbol{\eta}|\mathbf{y})d\boldsymbol{\eta}$$

## Distribuzione a posteriori

- Con $\eta$ non-noti
  - il metodo EB “plugs-in” una stima di $\eta$, $\hat{\eta} \equiv \hat{\eta}(\mathbf{y})$ in $p(\theta|\mathbf{y}, \eta)$ i.e.

**il metodo EB**

usa la “distribuzione a posteriori stimata”

$$p(\theta|\mathbf{y}, \hat{\eta})$$

  - il metodo HB specifica una a priori $p(\eta)$ e

**Il metodo (full) HB**

calcola

$$p(\theta|\mathbf{y}) = \int p(\theta|\mathbf{y}, \eta) p(\eta|\mathbf{y}) d\eta$$

cioè, HB fa un’integrazione mentre EB fa e.g. una massimizzazione come sopra.

# Indice della sezione I

## Modelli gerarchici Bayesiani

I modelli Bayesiani gerarchici o multi-stadio sono un modo naturale di pensare a come modellare l'informazione di unità parzialmente scambiabili

- essi posso essere adatti per modellare sia le proprietà delle unità stesse

$$y_i^s|\theta^s \sim f(\theta^s) \quad i = 1, \ldots, n^s$$

## Modelli gerarchici Bayesiani

I modelli Bayesiani gerarchici o multi-stadio sono un modo naturale di pensare a come modellare l'informazione di unità parzialmente scambiabili

- essi posso essere adatti per modellare sia le proprietà delle unità stesse
$$y_i^s|\theta^s \sim f(\theta^s) \quad i = 1, \ldots, n^s$$
- sia come queste proprietà varino tra le unità
$$\theta^s|\theta^* \sim g(\theta^*) \quad s = 1, \ldots, S$$

## Modelli gerarchici Bayesiani

I modelli Bayesiani gerarchici o multi-stadio sono un modo naturale di pensare a come modellare l'informazione di unità parzialmente scambiabili

- essi posso essere adatti per modellare sia le proprietà delle unità stesse

$$y_i^s|\theta^s \sim f(\theta^s) \quad i = 1, \ldots, n^s$$

- sia come queste proprietà varino tra le unità

$$\theta^s|\theta^* \sim g(\theta^*) \quad s = 1, \ldots, S$$

- insieme ad una specificazione delle distr a priori per gli iperparametri nell'ultimo stadio

$$\theta^* \sim h(\theta_0)$$

# Modelli gerarchici

## Esempio: studio dell'efficacia di un trattamento (cardiaco)

- nell'ospedale $j$ la probabilità di sopravvivenza è $\theta_j$

## Modelli gerarchici

### Esempio: studio dell'efficacia di un trattamento (cardiaco)

- nell'ospedale $j$ la probabilità di sopravvivenza è $\theta_j$
- le osservazioni $y_{ij}$ dicono se il paziente $i$ sia sopravvisuto nell'ospedale $j$

$$
\begin{tikzcd}
\theta_1 \arrow[d] & \theta_2 \arrow[d] & \cdots & \theta_n \arrow[d] \\
y_{i1} & y_{i2} & & y_{in}
\end{tikzcd}
$$

# Modelli gerarchici

## Esempio: studio dell'efficacia di un trattamento (cardiaco)

- nell'ospedale $j$ la probabilità di sopravvivenza è $\theta_j$
- le osservazioni $y_{ij}$ dicono se il paziente $i$ sia sopravvisuto nell'ospedale $j$

$$
\begin{tikzcd}
\theta_1 \arrow[d] & \theta_2 \arrow[d] & \cdots & \theta_n \arrow[d] \\
y_{i1} & y_{i2} & & y_{in}
\end{tikzcd}
$$

# Modelli gerarchici

## Esempio: studio dell’efficacia di un trattamento (cardiaco)

- nell’ospedale $j$ la probabilità di sopravvivenza è $\theta_j$
- le osservazioni $y_{ij}$ dicono se il paziente $i$ sia sopravvisuto nell’ospedale $j$

$$\begin{tikzcd}
\theta_1 \arrow[d] & \theta_2 \arrow[d] & \cdots & \theta_n \arrow[d] \\
y_{i1} & y_{i2} & & y_{in}
\end{tikzcd}$$

- è ragionevole assumere che i $\theta_j$s - che rappresentano un gruppo di ospedali - siano collegati tra loro, simili

$$\begin{tikzcd}
& \tau \arrow[dl] \arrow[d] \arrow[drr] & & \\
\theta_1 \arrow[d] & \theta_2 \arrow[d] & \cdots & \theta_n \arrow[d] \\
y_{i1} & y_{i2} & & y_{in}
\end{tikzcd}$$

# Modelli gerarchici per meta-analisi

## Esempio: studio dell'efficacia di un trattamento (cardiaco)

- e che i $\theta_j$ abbiano una distribuzione comune, $\theta_j \sim g(\tau)$, con $\tau$ probabilità di sopravvivenza della popolazione generale

## Modelli gerarchici per meta-analisi

### Esempio: studio dell'efficacia di un trattamento (cardiaco)

- e che i $\theta_j$ abbiano una distribuzione comune, $\theta_j \sim g(\tau)$, con $\tau$ probabilità di sopravvivenza della popolazione generale
- $\theta_j$ non è direttamente osservato e la distribuzione della popolazione non è nota

# Modelli gerarchici per meta-analisi

## Esempio: studio dell'efficacia di un trattamento (cardiaco)

- e che i $\theta_j$ abbiano una distribuzione comune, $\theta_j \sim g(\tau)$, con $\tau$ probabilità di sopravvivenza della popolazione generale
- $\theta_j$ non è direttamente osservato e la distribuzione della popolazione non è nota
  1. stima di ogni $\theta_j$ prende a prestito informazione da tutti gli altri ospedali, $p(\theta_j | y_1, \ldots, y_n)$

# Modelli gerarchici per meta-analisi

## Esempio: studio dell'efficacia di un trattamento (cardiaco)

- e che i $\theta_j$ abbiano una distribuzione comune, $\theta_j \sim g(\tau)$, con $\tau$ probabilità di sopravvivenza della popolazione generale
- $\theta_j$ non è direttamente osservato e la distribuzione della popolazione non è nota
  1. stima di ogni $\theta_j$ prende a prestito informazione da tutti gli altri ospedali, $p(\theta_j | y_1, \ldots, y_n)$
  2. stima di $\tau$ tiene conto della variabilità tra gli ospedali, $p(\tau | y_1, \ldots, y_n)$

# Modelli gerarchici per meta-analisi

## Esempio: studio dell'efficacia di un trattamento (cardiaco)

- e che i $\theta_j$ abbiano una distribuzione comune, $\theta_j \sim g(\tau)$, con $\tau$ probabilità di sopravvivenza della popolazione generale
- $\theta_j$ non è direttamente osservato e la distribuzione della popolazione non è nota
    1. stima di ogni $\theta_j$ prende a prestito informazione da tutti gli altri ospedali, $p(\theta_j | y_1, \dots, y_n)$
    2. stima di $\tau$ tiene conto della variabilità tra gli ospedali, $p(\tau | y_1, \dots, y_n)$
- Se ci sono diversi studi che considerano lo stesso problema di ricerca, si può pensare di combinare le informazioni da tutti i singoli studi al fine di giungere ad una conclusione complessiva riguardo al problema d'interesse

# Modelli gerarchici per meta-analisi

## Esempio: studio dell'efficacia di un trattamento (cardiaco)

- e che i $\theta_j$ abbiano una distribuzione comune, $\theta_j \sim g(\tau)$, con $\tau$ probabilità di sopravvivenza della popolazione generale
- $\theta_j$ non è direttamente osservato e la distribuzione della popolazione non è nota
  1. stima di ogni $\theta_j$ prende a prestito informazione da tutti gli altri ospedali, $p(\theta_j|y_1, \ldots, y_n)$
  2. stima di $\tau$ tiene conto della variabilità tra gli ospedali, $p(\tau|y_1, \ldots, y_n)$
- Se ci sono diversi studi che considerano lo stesso problema di ricerca, si può pensare di combinare le informazioni da tutti i singoli studi al fine di giungere ad una conclusione complessiva riguardo al problema d'interesse
- Gli studi possono essere pensati come appartenenti ad una popolazione di studi rivolti allo stesso problema di ricerca, e la messa insieme di singoli studi per trarre conclusioni sulla totalità è nota in letteratura come **meta-analisi**

# Modello gerarchico: terminologia

**Level 1:** observations given parameters $p(y_{ij} \mid \theta_j)$

$$\begin{array}{cccccc} p(\theta_j \mid \tau) & \theta_1 & \theta_2 & \cdots & \theta_n & \text{parameters} \\ & \downarrow & \downarrow & & \downarrow & \\ p(y_{ij} \mid \theta_j) & y_{i1} & y_{i2} & & y_{in} & \text{observations} \end{array}$$

# Modello gerarchico: terminologia

**Level 1:** observations given parameters $p(y_{ij} \mid \theta_j)$

**Level 2:** parameters given hyperparameters $p(\theta_j \mid \tau)$

$$
\begin{tikzcd}
p(\tau) & & \tau \arrow[dl] \arrow[d] \arrow[drr] & & & \text{hyperparameter} \\
p(\theta_j \mid \tau) & \theta_1 \arrow[d] & \theta_2 \arrow[d] & \cdots & \theta_n \arrow[d] & \text{parameters} \\
p(y_{ij} \mid \theta_j) & y_{i1} & y_{i2} & & y_{in} & \text{observations}
\end{tikzcd}
$$

## Modello gerarchico: terminologia

**Level 1:** observations given parameters $p(y_{ij} \mid \theta_j)$

**Level 2:** parameters given hyperparameters $p(\theta_j \mid \tau)$

$$
\begin{tikzcd}
p(\tau) & & \tau \arrow[dl] \arrow[d] \arrow[drr] & & & \text{hyperparameter} \\
p(\theta_j \mid \tau) & \theta_1 \arrow[d] & \theta_2 \arrow[d] & \cdots & \theta_n \arrow[d] & \text{parameters} \\
p(y_{ij} \mid \theta_j) & y_{i1} & y_{i2} & & y_{in} & \text{observations}
\end{tikzcd}
$$

$$
\begin{aligned}
p(\theta, \tau \mid y) &\propto p(y \mid \theta, \tau) p(\theta, \tau) \\
&\propto p(y \mid \theta) p(\theta \mid \tau) p(\tau)
\end{aligned}
$$

# Confronto tra modelli

- "Separate model" (model with separate/independent effects)

$$\begin{array}{cccc} \theta_1 & \theta_2 & \cdots & \theta_n \\ \downarrow & \downarrow & & \downarrow \\ y_1 & y_2 & & y_n \end{array}$$

# Confronto tra modelli

- "Separate model" (model with separate/independent effects)

$$\begin{tikzcd} \theta_1 \arrow[d] & \theta_2 \arrow[d] & \cdots & \theta_n \arrow[d] \\ y_1 & y_2 & & y_n \end{tikzcd}$$

- "Joint model" (model with a common effect / pooled model)

$$\begin{tikzcd} & \theta \arrow[dl] \arrow[d] \arrow[drr] & & \\ y_1 & y_2 & \cdots & y_n \end{tikzcd}$$

## Confronto tra modelli

- "Separate model" (model with separate/independent effects)

$$
\begin{tikzcd}
\theta_1 \arrow[d] & \theta_2 \arrow[d] & \cdots & \theta_n \arrow[d] \\
y_1 & y_2 & & y_n
\end{tikzcd}
$$

- "Joint model" (model with a common effect / pooled model)

$$
\begin{tikzcd}
& \theta \arrow[dl] \arrow[d] \arrow[drr] & & \\
y_1 & y_2 & \cdots & y_n
\end{tikzcd}
$$

- Hierarchical model

$$
\begin{tikzcd}
& \tau \arrow[dl] \arrow[d] \arrow[drr] & & \\
\theta_1 \arrow[d] & \theta_2 \arrow[d] & \cdots & \theta_n \arrow[d] \\
y_1 & y_2 & & y_n
\end{tikzcd}
$$

# Nella sezione

# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

- Medicine testing

# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

- Medicine testing
- Type F344 female rats in control group given placebo

# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

- Medicine testing
- Type F344 female rats in control group given placebo
  - count how many get endometrial stromal polyps

# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

- Medicine testing
- Type F344 female rats in control group given placebo
  - count how many get endometrial stromal polyps
  - familiar binomial model example

# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

- Medicine testing
- Type F344 female rats in control group given placebo
  - count how many get endometrial stromal polyps
  - familiar binomial model example
- Experiment has been repeated 71 times

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0/20 | 0/20 | 0/20 | 0/20 | 0/20 | 0/20 | 0/20 | 0/19 | 0/19 | 0/19 |
| 0/19 | 0/18 | 0/18 | 0/17 | 1/20 | 1/20 | 1/20 | 1/20 | 1/19 | 1/19 |
| 1/18 | 1/18 | 2/25 | 2/24 | 2/23 | 2/20 | 2/20 | 2/20 | 2/20 | 2/20 |
| 2/20 | 1/10 | 5/49 | 2/19 | 5/46 | 3/27 | 2/17 | 7/49 | 7/47 | 3/20 |
| 3/20 | 2/13 | 9/48 | 10/50 | 4/20 | 4/20 | 4/20 | 4/20 | 4/20 | 4/20 |
| 4/20 | 10/48 | 4/19 | 4/19 | 4/19 | 5/22 | 11/46 | 12/49 | 5/20 | 5/20 |
| 6/23 | 5/19 | 6/22 | 6/20 | 6/20 | 6/20 | 16/52 | 15/46 | 15/47 | 9/24 |
| 4/14 | | | | | | | | | |

# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

- Medicine testing
- Type F344 female rats in control group given placebo
  - count how many get endometrial stromal polyps
  - familiar binomial model example
- Experiment has been repeated 71 times

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0/20 | 0/20 | 0/20 | 0/20 | 0/20 | 0/20 | 0/20 | 0/19 | 0/19 | 0/19 |
| 0/19 | 0/18 | 0/18 | 0/17 | 1/20 | 1/20 | 1/20 | 1/20 | 1/19 | 1/19 |
| 1/18 | 1/18 | 2/25 | 2/24 | 2/23 | 2/20 | 2/20 | 2/20 | 2/20 | 2/20 |
| 2/20 | 1/10 | 5/49 | 2/19 | 5/46 | 3/27 | 2/17 | 7/49 | 7/47 | 3/20 |
| 3/20 | 2/13 | 9/48 | 10/50 | 4/20 | 4/20 | 4/20 | 4/20 | 4/20 | 4/20 |
| 4/20 | 10/48 | 4/19 | 4/19 | 4/19 | 5/22 | 11/46 | 12/49 | 5/20 | 5/20 |
| 6/23 | 5/19 | 6/22 | 6/20 | 6/20 | 6/20 | 16/52 | 15/46 | 15/47 | 9/24 |
| 4/14 | | | | | | | | | |

- $\theta$ probabilià di tumore nei topi (femmina) di laboratorio che ricevono dose $0$ del farmaco

# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

- Medicine testing
- Type F344 female rats in control group given placebo
  - count how many get endometrial stromal polyps
  - familiar binomial model example
- Experiment has been repeated 71 times

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0/20 | 0/20 | 0/20 | 0/20 | 0/20 | 0/20 | 0/20 | 0/19 | 0/19 | 0/19 |
| 0/19 | 0/18 | 0/18 | 0/17 | 1/20 | 1/20 | 1/20 | 1/20 | 1/19 | 1/19 |
| 1/18 | 1/18 | 2/25 | 2/24 | 2/23 | 2/20 | 2/20 | 2/20 | 2/20 | 2/20 |
| 2/20 | 1/10 | 5/49 | 2/19 | 5/46 | 3/27 | 2/17 | 7/49 | 7/47 | 3/20 |
| 3/20 | 2/13 | 9/48 | 10/50 | 4/20 | 4/20 | 4/20 | 4/20 | 4/20 | 4/20 |
| 4/20 | 10/48 | 4/19 | 4/19 | 4/19 | 5/22 | 11/46 | 12/49 | 5/20 | 5/20 |
| 6/23 | 5/19 | 6/22 | 6/20 | 6/20 | 6/20 | 16/52 | 15/46 | 15/47 | 9/24 |
| 4/14 | | | | | | | | | |

- $\theta$ probabilià di tumore nei topi (femmina) di laboratorio che ricevono dose $0$ del farmaco
- $y/n$ proporzione osservata di cavie con tumore

## Modello binomiale gerarchico

### Esempio: Incidenza di tumore nei topi

- Medicine testing
- Type F344 female rats in control group given placebo
  - count how many get endometrial stromal polyps
  - familiar binomial model example
- Experiment has been repeated 71 times

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0/20 | 0/20 | 0/20 | 0/20 | 0/20 | 0/20 | 0/20 | 0/19 | 0/19 | 0/19 |
| 0/19 | 0/18 | 0/18 | 0/17 | 1/20 | 1/20 | 1/20 | 1/20 | 1/19 | 1/19 |
| 1/18 | 1/18 | 2/25 | 2/24 | 2/23 | 2/20 | 2/20 | 2/20 | 2/20 | 2/20 |
| 2/20 | 1/10 | 5/49 | 2/19 | 5/46 | 3/27 | 2/17 | 7/49 | 7/47 | 3/20 |
| 3/20 | 2/13 | 9/48 | 10/50 | 4/20 | 4/20 | 4/20 | 4/20 | 4/20 | 4/20 |
| 4/20 | 10/48 | 4/19 | 4/19 | 4/19 | 5/22 | 11/46 | 12/49 | 5/20 | 5/20 |
| 6/23 | 5/19 | 6/22 | 6/20 | 6/20 | 6/20 | 16/52 | 15/46 | 15/47 | 9/24 |
| 4/14 | | | | | | | | | |

- $\theta$ probabilià di tumore nei topi (femmina) di laboratorio che ricevono dose $0$ del farmaco
- $y/n$ proporzione osservata di cavie con tumore
- esperimento corrente $(71^o)$: $y/n = 4/14$

# Costruzione di una priori: un singolo esperimento

## Esempio: Incidenza di tumore nei topi

Analisi di un singolo esperimento nel contesto di dati storici

1. Analisi con una a priori fissata

## Costruzione di una priori: un singolo esperimento

### Esempio: Incidenza di tumore nei topi

Analisi di un singolo esperimento nel contesto di dati storici

1. Analisi con una a priori fissata

- Supponiamo di determinare i parametri di una Beta conoscendo media e varianza della distribuzione di $\theta$ (che varia per le differenze che ci sono nelle cavie e nelle condizioni sperimentali tra gli studi)

$$\begin{aligned} y|n,\theta &\sim \mathrm{Bin}(n,\theta) \\ \theta &\sim \mathrm{Beta}(\alpha,\beta) \\ \theta|y,n &\sim \mathrm{Beta}(\alpha+4,\beta+10) \end{aligned}$$

# Costruzione di una priori: un singolo esperimento

## Esempio: Incidenza di tumore nei topi

2. Dati storici su simili studi: i $70$ studi precedenti

## Costruzione di una priori: un singolo esperimento

### Esempio: Incidenza di tumore nei topi

2. Dati storici su simili studi: i $70$ studi precedenti

- media e sd campionaria dei $y_j/n_j$ precedenti sono, rispettivamente, 0.136 e 5.1.

# Costruzione di una priori: un singolo esperimento

## Esempio: Incidenza di tumore nei topi

2. Dati storici su simili studi: i $70$ studi precedenti

- media e sd campionaria dei $y_j/n_j$ precedenti sono, rispettivamente, 0.136 e 5.1.
- dalle relazioni note con i parametri di una beta, troviamo che una stima di $(\alpha, \beta)$ è $(1.4, 8.6)$

# Costruzione di una priori: un singolo esperimento

## Esempio: Incidenza di tumore nei topi

2. Dati storici su simili studi: i $70$ studi precedenti

- media e sd campionaria dei $y_j/n_j$ precedenti sono, rispettivamente, 0.136 e 5.1.
- dalle relazioni note con i parametri di una beta, troviamo che una stima di $(\alpha, \beta)$ è $(1.4, 8.6)$
- $$y_j|n_j, \theta_j \sim \mathrm{Bin}(n_j, \theta_j) \quad j = 1, \ldots, 70, 71$$
$$\theta_j \sim \mathrm{Beta}(\hat{\alpha}, \hat{\beta})$$
$$\theta_{71}|y_{71}, n_{71}, \mathrm{datistorici} \sim \mathrm{Beta}(\hat{\alpha} + 4, \hat{\beta} + 10)$$

# Costruzione di una priori: un singolo esperimento

## Esempio: Incidenza di tumore nei topi

2. Dati storici su simili studi: i $70$ studi precedenti

- media e sd campionaria dei $y_j/n_j$ precedenti sono, rispettivamente, 0.136 e 5.1.
- dalle relazioni note con i parametri di una beta, troviamo che una stima di $(\alpha, \beta)$ è $(1.4, 8.6)$
-
$$
\begin{aligned}
y_j | n_j, \theta_j &\sim \mathrm{Bin}(n_j, \theta_j) \quad j = 1, \ldots, 70, 71 \\
\theta_j &\sim \mathrm{Beta}(\hat{\alpha}, \hat{\beta}) \\
\theta_{71} | y_{71}, n_{71}, \mathrm{datistorici} &\sim \mathrm{Beta}(\hat{\alpha} + 4, \hat{\beta} + 10)
\end{aligned}
$$

  - $E(\theta_{71} | \mathrm{datistorici}) = \frac{\hat{\alpha}}{\hat{\alpha} + \hat{\beta}} = 0.136$

# Costruzione di una priori: un singolo esperimento

## Esempio: Incidenza di tumore nei topi

2. Dati storici su simili studi: i $70$ studi precedenti

- media e sd campionaria dei $y_j/n_j$ precedenti sono, rispettivamente, 0.136 e 5.1.
- dalle relazioni note con i parametri di una beta, troviamo che una stima di $(\alpha, \beta)$ è $(1.4, 8.6)$
-
$$y_j | n_j, \theta_j \sim \mathrm{Bin}(n_j, \theta_j) \quad j = 1, \ldots, 70, 71$$
$$\theta_j \sim \mathrm{Beta}(\hat{\alpha}, \hat{\beta})$$
$$\theta_{71} | y_{71}, n_{71}, \mathrm{datistorici} \sim \mathrm{Beta}(\hat{\alpha} + 4, \hat{\beta} + 10)$$

  - $E(\theta_{71} | \mathrm{datistorici}) = \frac{\hat{\alpha}}{\hat{\alpha}+\hat{\beta}} = 0.136$
  - $\hat{\theta}_{71} = \frac{4}{14} = 0.286$

# Costruzione di una priori: un singolo esperimento

## Esempio: Incidenza di tumore nei topi

2. Dati storici su simili studi: i $70$ studi precedenti

- media e sd campionaria dei $y_j/n_j$ precedenti sono, rispettivamente, 0.136 e 5.1.
- dalle relazioni note con i parametri di una beta, troviamo che una stima di $(\alpha, \beta)$ è $(1.4, 8.6)$
- $$y_j|n_j, \theta_j \sim \mathrm{Bin}(n_j, \theta_j) \quad j = 1, \ldots, 70, 71$$
  $$\theta_j \sim \mathrm{Beta}(\hat{\alpha}, \hat{\beta})$$
  $$\theta_{71}|y_{71}, n_{71}, \mathrm{datistorici} \sim \mathrm{Beta}(\hat{\alpha} + 4, \hat{\beta} + 10)$$
  - $E(\theta_{71}|\mathrm{datistorici}) = \frac{\hat{\alpha}}{\hat{\alpha}+\hat{\beta}} = 0.136$
  - $\hat{\theta}_{71} = \frac{4}{14} = 0.286$
  - $E(\theta_{71}|y_{71}, n_{71}, , \mathrm{datistorici}) = 0.223$ (perchè dal confronto dello studio corrente con l'esperienza pregressa ...)

## Costruzione di una priori: un singolo esperimento

### Esempio: Incidenza di tumore nei topi

2. Dati storici su simili studi: i $70$ studi precedenti

- media e sd campionaria dei $y_j/n_j$ precedenti sono, rispettivamente, 0.136 e 5.1.
- dalle relazioni note con i parametri di una beta, troviamo che una stima di $(\alpha, \beta)$ è $(1.4, 8.6)$
-
$$y_j|n_j, \theta_j \sim \mathrm{Bin}(n_j, \theta_j) \quad j = 1, \ldots, 70, 71$$
$$\theta_j \sim \mathrm{Beta}(\hat{\alpha}, \hat{\beta})$$
$$\theta_{71}|y_{71}, n_{71}, \mathrm{datistorici} \sim \mathrm{Beta}(\hat{\alpha} + 4, \hat{\beta} + 10)$$
  - $E(\theta_{71}|\mathrm{datistorici}) = \frac{\hat{\alpha}}{\hat{\alpha}+\hat{\beta}} = 0.136$
  - $\hat{\theta}_{71} = \frac{4}{14} = 0.286$
  - $E(\theta_{71}|y_{71}, n_{71}, , \mathrm{datistorici}) = 0.223$ (perchè dal confronto dello studio corrente con l'esperienza pregressa ...)
- NB: Non abbiamo adottato fino a qui un approccio Bayesiano *full*, abbiamo semplicemente usato dati storici per ottenere delle stime per i parametri della popolazione (iperparametri)

# Costruzione di una priori: un singolo esperimento

## Esempio: Incidenza di tumore nei topi

- E se usassimo questo modello per fare inferenza su $\theta_j$ per $j = 1 - 70$? Sarebbe lecito?

# Costruzione di una priori: un singolo esperimento

## Esempio: Incidenza di tumore nei topi

- E se usassimo questo modello per fare inferenza su $\theta_j$ per $j = 1 - 70$? Sarebbe lecito?
  - sovraprecisione ...

# Costruzione di una priori: un singolo esperimento

## Esempio: Incidenza di tumore nei topi

- E se usassimo questo modello per fare inferenza su $\theta_j$ per $j = 1 - 70$? Sarebbe lecito?
  - sovraprecisione ...
  - stima comporta sempre non integrazione dell'incertezza

## Costruzione di una priori: un singolo esperimento

### Esempio: Incidenza di tumore nei topi

- E se usassimo questo modello per fare inferenza su $\theta_j$ per $j = 1 - 70$? Sarebbe lecito?
  - sovraprecisione ...
  - stima comporta sempre non integrazione dell'incertezza
- Ma forse, nella logica dell'inferenza Bayesiana, non ha proprio senso *stimare* $(\alpha, \beta)$ (per trattarli come noti prima che i dati siano raccolti) ...

# Costruzione di una priori: un singolo esperimento
## Esempio: Incidenza di tumore nei topi

- E se usassimo questo modello per fare inferenza su $\theta_j$ per $j = 1 - 70$? Sarebbe lecito?
  - sovraprecisione ...
  - stima comporta sempre non integrazione dell'incertezza
- Ma forse, nella logica dell'inferenza Bayesiana, non ha proprio senso *stimare* $(\alpha, \beta)$ (per trattarli come noti prima che i dati siano raccolti) ...
- (Stiamo richiamando il concetto di EB contrapposto a FB)

## Costruzione di una priori: combinare informazioni

### Esempio: Incidenza di tumore nei topi

Logica del combinare le informazioni

- Ha più senso stimare la distribuzione della popolazione da tutti i dati, e ciò ‘aiuta’ anche la stima dei $\theta_j$, piuttosto che stimare separatamente ognuno dei $71$ $\theta_j$

# Costruzione di una priori: combinare informazioni

## Esempio: Incidenza di tumore nei topi

Logica del combinare le informazioni

- Ha più senso stimare la distribuzione della popolazione da tutti i dati, e ciò 'aiuta' anche la stima dei $\theta_j$, piuttosto che stimare separatamente ognuno dei $71\ \theta_j$
- Evitiamo i problemi connessi con un approccio EB, specificando un modello probabilistico per l'insieme complessivo dei parametri e degli esperimenti ed effettuando un'analisi Bayesiana completa (*full*)

## Costruzione di una priori: combinare informazioni

### Esempio: Incidenza di tumore nei topi

Logica del combinare le informazioni

- Ha più senso stimare la distribuzione della popolazione da tutti i dati, e ciò 'aiuta' anche la stima dei $\theta_j$, piuttosto che stimare separatamente ognuno dei $71\ \theta_j$
- Evitiamo i problemi connessi con un approccio EB, specificando un modello probabilistico per l'insieme complessivo dei parametri e degli esperimenti ed effettuando un'analisi Bayesiana completa (*full*)

$\Rightarrow$ scambiabilità e modellazione gerarchica

## Scambiabilità

- $j = 1, \ldots .J$, un insieme di unità (eg esperimenti) ove

## Scambiabilità

- $j = 1, \ldots .J$, un insieme di unità (eg esperimenti) ove
  - $y_1, \ldots, y_J$, dati

## Scambiabilità

- $j = 1, \ldots .J$, un insieme di unità (eg esperimenti) ove
  - $y_1, \ldots, y_J$, dati
  - $\theta_1, \ldots, \theta_J$, parametri specifici per unità

# Scambiabilità

- $j = 1, \ldots .J$, un insieme di unità (eg esperimenti) ove
  - $y_1, \ldots, y_J$, dati
  - $\theta_1, \ldots, \theta_J$, parametri specifici per unità
  - $p(y_1|\theta_1), \ldots, p(y_J|\theta_J)$, likelihood specifiche per unità

## Scambiabilità

- $j = 1, \dots .J$, un insieme di unità (eg esperimenti) ove
  - $y_1, \dots, y_J$, dati
  - $\theta_1, \dots, \theta_J$, parametri specifici per unità
  - $p(y_1|\theta_1), \dots, p(y_J|\theta_J)$, likelihood specifiche per unità

Se non si può assumere alcun ordine o raggruppamento dei parametri $\theta_j$, se non abbiamo alcuna informazione (a parte $\mathbf{y}$) per distinguerli, allora dobbiamo assumere nella distribuzione a priori simmetria tra i parametri, rappresentata probabilisticamente dalla **scambiabilità**:

$\theta_1, \dots, \theta_J$ sono scambiabili se $p(\theta_1, \dots, \theta_J)$ è invariante per permutazioni degli indici $j = 1, \dots, J$

## Scambiabilità

- $j = 1, \ldots .J$, un insieme di unità (eg esperimenti) ove
  - $y_1, \ldots, y_J$, dati
  - $\theta_1, \ldots, \theta_J$, parametri specifici per unità
  - $p(y_1|\theta_1), \ldots, p(y_J|\theta_J)$, likelihood specifiche per unità

Se non si può assumere alcun ordine o raggruppamento dei parametri $\theta_j$, se non abbiamo alcuna informazione (a parte $\mathbf{y}$) per distinguerli, allora dobbiamo assumere nella distribuzione a priori simmetria tra i parametri, rappresentata probabilisticamente dalla **scambiabilità**:

$\theta_1, \ldots, \theta_J$ sono scambiabili se $p(\theta_1, \ldots, \theta_J)$ è invariante per permutazioni degli indici $j = 1, \ldots, J$

**Ignoranza $\rightarrow$ Scambiabilità**

## Scambiabilità come IID condizionale

La forma più semplice di una distr scambiabile (ma non l'unica) è l'indipendenza condizionale

- $\theta_j$ è un'estrazione indipendente da una comune distribuzione (popolazione) governata da qualche parametro $\phi$

$$p(\boldsymbol{\theta}|\phi) = \prod_{j=1}^{J} p(\theta_j|\phi)$$

## Scambiabilità come IID condizionale

La forma più semplice di una distr scambiabile (ma non l'unica) è l'indipendenza condizionale

- $\theta_j$ è un'estrazione indipendente da una comune distribuzione (popolazione) governata da qualche parametro $\phi$

$$p(\boldsymbol{\theta}|\phi) = \prod_{j=1}^{J} p(\theta_j|\phi)$$

- da cui, se $\phi$ è non noto (come tipicamente è), la distribuzione per $\theta$ deve essere mediata rispetto all'incertezza in $\phi$

$$p(\boldsymbol{\theta}) = \int \left[\prod_{j=1}^{J} p(\theta_j|\phi)\right] p(\phi)d\phi$$

i.e. $p(\boldsymbol{\theta})$ è una mistura iid.

## Scambiabilità: Teorema di de Finetti

- **Teorema di de Finetti**
  In the limit as $J \to \infty$, any suitably well-behaved exchangeable distribution on $(\theta_1, \ldots, \theta_J)$ can be expressed as a mixture of independent and identical distributions

## Scambiabilità: Teorema di de Finetti

- **Teorema di de Finetti**
  In the limit as $J \to \infty$, any suitably well-behaved exchangeable distribution on $(\theta_1, \ldots, \theta_J)$ can be expressed as a mixture of independent and identical distributions
- The theorem does not hold when $J$ is finite

# Scambiabilità: Teorema di de Finetti

- **Teorema di de Finetti**
  In the limit as $J \to \infty$, any suitably well-behaved exchangeable distribution on $(\theta_1, \ldots, \theta_J)$ can be expressed as a mixture of independent and identical distributions
- The theorem does not hold when $J$ is finite
- Statistically, the mixture model characterizes parameters $\theta$ as drawn from a common superpopulation that is determined by the unknown hyperparameters, $\phi$.

# Scambiabilità - Counter example

- A six sided die with probabilities $\theta_1, \ldots, \theta_6$

## Scambiabilità - Counter example

- A six sided die with probabilities $\theta_1, \ldots, \theta_6$
  - without additional knowledge $\theta_1, \ldots, \theta_6$ exchangeable

# Scambiabilità - Counter example

- A six sided die with probabilities $\theta_1, \ldots, \theta_6$
  - without additional knowledge $\theta_1, \ldots, \theta_6$ exchangeable
  - due to the constraint $\sum_{j=1}^{6} \theta_j$, parameters are not independent and thus joint distribution can not be presented as iid

## Scambiabilità

- Justifies why we can use

## Scambiabilità

- Justifies why we can use
  - a joint model for data

# Scambiabilità

- Justifies why we can use
  - a joint model for data
  - a joint prior for a set of parameters

# Scambiabilità

- Justifies why we can use
  - a joint model for data
  - a joint prior for a set of parameters
- Less strict than independence

# Scambiabilità

- Justifies why we can use
  - a joint model for data
  - a joint prior for a set of parameters
- Less strict than independence
- Exchangeability: Parameters $\theta_1, \ldots, \theta_J$ (or observations $y_1, \ldots, y_J$) are exchangeable if the joint distribution $p$ is invariant to the permutation of indices $(1, \ldots, J)$

## Scambiabilità

- Justifies why we can use
  - a joint model for data
  - a joint prior for a set of parameters
- Less strict than independence
- Exchangeability: Parameters $\theta_1, \ldots, \theta_J$ (or observations $y_1, \ldots, y_J$) are exchangeable if the joint distribution $p$ is invariant to the permutation of indices $(1, \ldots, J)$
  - E.g., $p(\theta_1, \theta_2, \theta_3) = p(\theta_3, \theta_1, \theta_2)$

## Scambiabilità

- Justifies why we can use
  - a joint model for data
  - a joint prior for a set of parameters
- Less strict than independence
- Exchangeability: Parameters $\theta_1, \ldots, \theta_J$ (or observations $y_1, \ldots, y_J$) are exchangeable if the joint distribution $p$ is invariant to the permutation of indices $(1, \ldots, J)$
  - E.g., $p(\theta_1, \theta_2, \theta_3) = p(\theta_3, \theta_1, \theta_2)$
- *Ignorance implies exchangeability*: If there is no information which can be used *a priori* to separate $\theta_j$ form each other, we can assume exchangeability ($\sim$ simmetry).

## Scambiabilità: obiezioni

- In virtually any statistical application, it is natural to object to exchangeability on the grounds that the units actually differ. (For example, the 71 rat tumor experiments were performed at different times, on different rats, and presumably in different laboratories.)

## Scambiabilità: obiezioni

- In virtually any statistical application, it is natural to object to exchangeability on the grounds that the units actually differ. (For example, the 71 rat tumor experiments were performed at different times, on different rats, and presumably in different laboratories.)
- That the experiments differ implies that the $\theta_j$'s differ, but it might be perfectly acceptable to consider them as if drawn from a common distribution.

## Scambiabilità: obiezioni

- In virtually any statistical application, it is natural to object to exchangeability on the grounds that the units actually differ. (For example, the 71 rat tumor experiments were performed at different times, on different rats, and presumably in different laboratories.)
- That the experiments differ implies that the $\theta_j$'s differ, but it might be perfectly acceptable to consider them as if drawn from a common distribution.
- With no information available to distinguish them, we have no logical choice but to model the $\theta_j$'s exchangeably.

## Scambiabilità: obiezioni

- In virtually any statistical application, it is natural to object to exchangeability on the grounds that the units actually differ. (For example, the 71 rat tumor experiments were performed at different times, on different rats, and presumably in different laboratories.)
- That the experiments differ implies that the $\theta_j$'s differ, but it might be perfectly acceptable to consider them as if drawn from a common distribution.
- With no information available to distinguish them, we have no logical choice but to model the $\theta_j$'s exchangeably.
  - e.g. if we know that the experiments have been in two different laboratories, and we know that the other laboratory has better conditions for the rats, but we do not know which experiments have been made in which laboratory

## Scambiabilità: obiezioni

- In virtually any statistical application, it is natural to object to exchangeability on the grounds that the units actually differ. (For example, the 71 rat tumor experiments were performed at different times, on different rats, and presumably in different laboratories.)
- That the experiments differ implies that the $\theta_j$'s differ, but it might be perfectly acceptable to consider them as if drawn from a common distribution.
- With no information available to distinguish them, we have no logical choice but to model the $\theta_j$'s exchangeably.
  - e.g. if we know that the experiments have been in two different laboratories, and we know that the other laboratory has better conditions for the rats, but we do not know which experiments have been made in which laboratory
- A priori experiments are exchangeable

## Scambiabilità e informazione addizionale

- Example: bioassay

## Scambiabilità e informazione addizionale

- Example: bioassay
  - $y_i$ number of dead animals are not exchangeable alone

# Scambiabilità e informazione addizionale

- Example: bioassay
  - $y_i$ number of dead animals are not exchangeable alone
  - $x_i$ dose is additional information

## Scambiabilità e informazione addizionale

- Example: bioassay
  - $y_i$ number of dead animals are not exchangeable alone
  - $x_i$ dose is additional information
  - $(x_i, y_i)$ exchangeable and logistic regression was used

$$p(\alpha, \beta \mid y, n, x) \propto \prod_{i=1}^{n} p(y_i \mid \alpha, \beta, n_i, x_i) p(\alpha, \beta)$$

# Scambiabilità gerarchica

- Example: hierarchical rats example

# Scambiabilità gerarchica

- Example: hierarchical rats example
  - all rats not exchangeable

## Scambiabilità gerarchica

- Example: hierarchical rats example
  - all rats not exchangeable
  - in a single laboratory rats exchangeable

# Scambiabilità gerarchica

- Example: hierarchical rats example
  - all rats not exchangeable
  - in a single laboratory rats exchangeable
  - laboratories exchangeable

## Scambiabilità gerarchica

- Example: hierarchical rats example
  - all rats not exchangeable
  - in a single laboratory rats exchangeable
  - laboratories exchangeable
  - $\rightarrow$ hierarchical model

## Scambiabilità parziale o condizionale

Often observations are not fully exchangeable, but are conditionally or partially exchangeable:

- If $y_i$ has additional information $x_i$ so that $y_i$ are not exchangeable but $(y_i, x_i)$ still are exchangeable, then we can make a joint model for $(y_i, x_i)$ or a conditional model for $(y_i|x_i)$.

## Scambiabilità parziale o condizionale

Often observations are not fully exchangeable, but are conditionally or partially exchangeable:

- If $y_i$ has additional information $x_i$ so that $y_i$ are not exchangeable but $(y_i, x_i)$ still are exchangeable, then we can make a joint model for $(y_i, x_i)$ or a conditional model for $(y_i|x_i)$.
- If observations can be grouped (a priori), we may make a hierarchical model, where each group has its own submodel, but the group properties are unknown. If we assume that group properties are exchangeable, we can use a common prior distribution for the group properties.

  Partial exchangeability allows us to model variation within each group and between groups

# Scambiabilità condizionale

- $x = (x_1, \ldots, x_J)$, covariate specifiche per unità

## Scambiabilità condizionale

- $x = (x_1, \ldots, x_J)$, covariate specifiche per unità
- $\theta = (\theta_1, \ldots, \theta_J)$, parametri specifici per unità

## Scambiabilità condizionale

- $x = (x_1, \ldots, x_J)$, covariate specifiche per unità
- $\theta = (\theta_1, \ldots, \theta_J)$, parametri specifici per unità
- $\theta_1 | x_1, \ldots, \theta_J | x_J$ sono scambiabili

# Scambiabilità condizionale

- $x = (x_1, \ldots, x_J)$, covariate specifiche per unità
- $\theta = (\theta_1, \ldots, \theta_J)$, parametri specifici per unità
- $\theta_1|x_1, \ldots, \theta_J|x_J$ sono scambiabili
- $p(\theta|x) = \int \left[\prod_j p(\theta_j|\phi, x_j)\right] p(\phi|x) d\phi$

## Scambiabilità condizionale

- $x = (x_1, \ldots, x_J)$, covariate specifiche per unità
- $\theta = (\theta_1, \ldots, \theta_J)$, parametri specifici per unità
- $\theta_1|x_1, \ldots, \theta_J|x_J$ sono scambiabili
- $p(\theta|x) = \int \left[\prod_j p(\theta_j|\phi, x_j)\right] p(\phi|x) d\phi$

In questo modo, i modelli scambiabili diventano quasi universalmente applicabili, perché qualsiasi informazione disponibile per distinguere le diverse unità dovrebbe essere codificata nelle variabili $x$ e $y$.

## Analisi bayesiana completa del modello gerarchico

L'elemento "gerarchico" chiave di questi modelli è che $\phi$ non è noto e quindi ha una propria distribuzione a priori, $p(\phi)$. La distribuzione a posteriori appropriata è del vettore $(\phi, \theta)$.

- Distribuzione a priori

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

## Analisi bayesiana completa del modello gerarchico

L'elemento "gerarchico" chiave di questi modelli è che $\phi$ non è noto e quindi ha una propria distribuzione a priori, $p(\phi)$. La distribuzione a posteriori appropriata è del vettore $(\phi, \theta)$.

- Distribuzione a priori

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

- Distribuzione a posteriori congiunta

$$\begin{aligned} p(\phi, \theta|y) &\propto p(\phi, \theta)p(y|\phi, \theta) \\ &= p(\phi, \theta)p(y|\theta) \end{aligned}$$

con quest'ultima semplificazione che vale perché la distribuzione dei dati, $p(y|\phi, \theta)$, dipende solo da $\theta$; gli iperparametri $\phi$ influenzano $y$ solo attraverso $\theta$.

## Distribuzione *hyperprior*

Se si sa poco di $\phi$, possiamo assegnare una distribuzione a priori diffusa, ma dobbiamo stare attenti quando si utilizza una a priori non propria per

- verificare che la a posteriori risultante sia propria,

## Distribuzione *hyperprior*

Se si sa poco di $\phi$, possiamo assegnare una distribuzione a priori diffusa, ma dobbiamo stare attenti quando si utilizza una a priori non propria per

- verificare che la a posteriori risultante sia propria,
- valutare se le nostre conclusioni sono sensibili a questa ipotesi semplificatrice.

## Distribuzione *hyperprior*

Se si sa poco di $\phi$, possiamo assegnare una distribuzione a priori diffusa, ma dobbiamo stare attenti quando si utilizza una a priori non propria per

- verificare che la a posteriori risultante sia propria,
- valutare se le nostre conclusioni sono sensibili a questa ipotesi semplificatrice.

Nella maggior parte dei problemi reali, si dovrebbe avere una conoscenza sostanziale dei parametri

- almeno per vincolare gli iperparametri in una regione finita,

## Distribuzione *hyperprior*

Se si sa poco di $\phi$, possiamo assegnare una distribuzione a priori diffusa, ma dobbiamo stare attenti quando si utilizza una a priori non propria per

- verificare che la a posteriori risultante sia propria,
- valutare se le nostre conclusioni sono sensibili a questa ipotesi semplificatrice.

Nella maggior parte dei problemi reali, si dovrebbe avere una conoscenza sostanziale dei parametri

- almeno per vincolare gli iperparametri in una regione finita,
- se non per assegnare una distribuzione *hyperprior* sostanziale.

## Distribuzione *hyperprior*

Se si sa poco di $\phi$, possiamo assegnare una distribuzione a priori diffusa, ma dobbiamo stare attenti quando si utilizza una a priori non propria per

- verificare che la a posteriori risultante sia propria,
- valutare se le nostre conclusioni sono sensibili a questa ipotesi semplificatrice.

Nella maggior parte dei problemi reali, si dovrebbe avere una conoscenza sostanziale dei parametri

- almeno per vincolare gli iperparametri in una regione finita,
- se non per assegnare una distribuzione *hyperprior* sostanziale.

Come nei modelli non gerarchici, è spesso sensato iniziare con una distribuzione a priori semplice, relativamente non informativa su $\phi$ e cercare di aggiungere ulteriori informazioni a priori se rimane troppa variabilità nella distribuzione a posteriori.

## Distribuzioni predittive a posteriori

Ci sono due distribuzioni predittive a posteriori che potrebbero interessare nei modelli gerarchici:

1. la distribuzione di osservazioni future $\tilde{y}$ corrispondenti a un $\theta_j$ esistente (*cavie aggiuntive da un esperimento esistente*), o

## Distribuzioni predittive a posteriori

Ci sono due distribuzioni predittive a posteriori che potrebbero interessare nei modelli gerarchici:

1. la distribuzione di osservazioni future $\tilde{y}$ corrispondenti a un $\theta_j$ esistente (*cavie aggiuntive da un esperimento esistente*), o
2. la distribuzione di osservazioni $\tilde{y}$ corrispondenti a futuri $\theta_j$'s tratti dalla stessa superpopolazione (*risultati da un esperimento futuro*). Etichettiamo $\theta_j$'s futuri come $\tilde{\theta}_j$.

## Distribuzioni predittive a posteriori

Ci sono due distribuzioni predittive a posteriori che potrebbero interessare nei modelli gerarchici:

1. la distribuzione di osservazioni future $\tilde{y}$ corrispondenti a un $\theta_j$ esistente (*cavie aggiuntive da un esperimento esistente*), o
2. la distribuzione di osservazioni $\tilde{y}$ corrispondenti a futuri $\theta_j$'s tratti dalla stessa superpopolazione (*risultati da un esperimento futuro*). Etichettiamo $\theta_j$'s futuri come $\tilde{\theta}_j$.

Nei due casi, le estrazioni predittive a posteriori $\tilde{y}$ si basano su

1. estrazioni a posteriori di $\theta_j$ per l'esperimento esistente.

## Distribuzioni predittive a posteriori

Ci sono due distribuzioni predittive a posteriori che potrebbero interessare nei modelli gerarchici:

1. la distribuzione di osservazioni future $\tilde{y}$ corrispondenti a un $\theta_j$ esistente (*cavie aggiuntive da un esperimento esistente*), o
2. la distribuzione di osservazioni $\tilde{y}$ corrispondenti a futuri $\theta_j$'s tratti dalla stessa superpopolazione (*risultati da un esperimento futuro*). Etichettiamo $\theta_j$'s futuri come $\tilde{\theta}_j$.

Nei due casi, le estrazioni predittive a posteriori $\tilde{y}$ si basano su

1. estrazioni a posteriori di $\theta_j$ per l'esperimento esistente.
2. $\tilde{\theta}_j$ simulati:

# Distribuzioni predittive a posteriori

Ci sono due distribuzioni predittive a posteriori che potrebbero interessare nei modelli gerarchici:

1. la distribuzione di osservazioni future $\tilde{y}$ corrispondenti a un $\theta_j$ esistente (*cavie aggiuntive da un esperimento esistente*), o
2. la distribuzione di osservazioni $\tilde{y}$ corrispondenti a futuri $\theta_j$'s tratti dalla stessa superpopolazione (*risultati da un esperimento futuro*). Etichettiamo $\theta_j$'s futuri come $\tilde{\theta}_j$.

Nei due casi, le estrazioni predittive a posteriori $\tilde{y}$ si basano su

1. estrazioni a posteriori di $\theta_j$ per l'esperimento esistente.
2. $\tilde{\theta}_j$ simulati:
   (a) si estrae $\tilde{\phi}$ da $p(\phi|y)$

## Distribuzioni predittive a posteriori

Ci sono due distribuzioni predittive a posteriori che potrebbero interessare nei modelli gerarchici:

1. la distribuzione di osservazioni future $\tilde{y}$ corrispondenti a un $\theta_j$ esistente (*cavie aggiuntive da un esperimento esistente*), o
2. la distribuzione di osservazioni $\tilde{y}$ corrispondenti a futuri $\theta_j$'s tratti dalla stessa superpopolazione (*risultati da un esperimento futuro*). Etichettiamo $\theta_j$'s futuri come $\tilde{\theta}_j$.

Nei due casi, le estrazioni predittive a posteriori $\tilde{y}$ si basano su

1. estrazioni a posteriori di $\theta_j$ per l'esperimento esistente.
2. $\tilde{\theta}_j$ simulati:
   - **(a)** si estrae $\tilde{\phi}$ da $p(\phi|y)$
   - **(b)** si estrae $\tilde{\theta}$ da $p(\theta|\tilde{\phi})$

## Distribuzioni predittive a posteriori

Ci sono due distribuzioni predittive a posteriori che potrebbero interessare nei modelli gerarchici:

1. la distribuzione di osservazioni future $\tilde{y}$ corrispondenti a un $\theta_j$ esistente (*cavie aggiuntive da un esperimento esistente*), o
2. la distribuzione di osservazioni $\tilde{y}$ corrispondenti a futuri $\theta_j$'s tratti dalla stessa superpopolazione (*risultati da un esperimento futuro*). Etichettiamo $\theta_j$'s futuri come $\tilde{\theta}_j$.

Nei due casi, le estrazioni predittive a posteriori $\tilde{y}$ si basano su

1. estrazioni a posteriori di $\theta_j$ per l'esperimento esistente.
2. $\tilde{\theta}_j$ simulati:
   (a) si estrae $\tilde{\phi}$ da $p(\phi|y)$
   (b) si estrae $\tilde{\theta}$ da $p(\theta|\tilde{\phi})$
   (c) si estrae $\tilde{y}$ da $p(y|\tilde{\theta})$.

# Nella sezione

## Calcolo nei modelli gerarchici coniugati

La strategia inferenziale per i modelli gerarchici segue l'approccio generale adottato con i problemi multiparametrici

- $(\theta, \phi)$ con $\theta$ 'parametri d'interesse' e $\phi$ 'parametri di disturbo'

# Calcolo nei modelli gerarchici coniugati

La strategia inferenziale per i modelli gerarchici segue l'approccio generale adottato con i problemi multiparametrici

- $(\theta, \phi)$ con $\theta$ 'parametri d'interesse' e $\phi$ 'parametri di disturbo'
- $p(\theta|\phi)$ coniugata alla varosimiglianza $p(y|\theta)$

## Calcolo nei modelli gerarchici coniugati

La strategia inferenziale per i modelli gerarchici segue l'approccio generale adottato con i problemi multiparametrici

- $(\theta, \phi)$ con $\theta$ 'parametri d'interesse' e $\phi$ 'parametri di disturbo'
- $p(\theta|\phi)$ coniugata alla varosimiglianza $p(y|\theta)$

Derivazione analitica delle distribuzioni condizionali e marginali

1. si scriva la distr congiunta a posteriori in forma non-normalizzata (passo immediato): $p(\theta, \phi|y) \propto p(\phi)p(\theta|\phi)p(y|\theta)$

## Calcolo nei modelli gerarchici coniugati

La strategia inferenziale per i modelli gerarchici segue l'approccio generale adottato con i problemi multiparametrici

- $(\theta, \phi)$ con $\theta$ 'parametri d'interesse' e $\phi$ 'parametri di disturbo'
- $p(\theta|\phi)$ coniugata alla varosimiglianza $p(y|\theta)$

Derivazione analitica delle distribuzioni condizionali e marginali

1. si scriva la distr congiunta a posteriori in forma non-normalizzata (passo immediato): $p(\theta, \phi|y) \propto p(\phi)p(\theta|\phi)p(y|\theta)$
2. si determini analiticamente $p(\theta|\phi, y)$ (semplice in modelli coniugati dato che $p(\boldsymbol{\theta}|\phi) = \prod_{j=1}^{J} p(\theta_j|\phi)$, quindi $p(\theta|\phi, y)$ è il prodotto di densità a posteriori coniugate per le componenti $\theta_j$)

## Calcolo nei modelli gerarchici coniugati

La strategia inferenziale per i modelli gerarchici segue l'approccio generale adottato con i problemi multiparametrici

- $(\theta, \phi)$ con $\theta$ 'parametri d'interesse' e $\phi$ 'parametri di disturbo'
- $p(\theta|\phi)$ coniugata alla varosimiglianza $p(y|\theta)$

Derivazione analitica delle distribuzioni condizionali e marginali

1. si scriva la distr congiunta a posteriori in forma non-normalizzata (passo immediato): $p(\theta, \phi|y) \propto p(\phi)p(\theta|\phi)p(y|\theta)$
2. si determini analiticamente $p(\theta|\phi, y)$ (semplice in modelli coniugati dato che $p(\boldsymbol{\theta}|\phi) = \prod_{j=1}^{J} p(\theta_j|\phi)$, quindi $p(\theta|\phi, y)$ è il prodotto di densità a posteriori coniugate per le componenti $\theta_j$)
3. si determini $p(\phi|y)$, due modi

# Calcolo nei modelli gerarchici coniugati

La strategia inferenziale per i modelli gerarchici segue l'approccio generale adottato con i problemi multiparametrici

- $(\theta, \phi)$ con $\theta$ 'parametri d'interesse' e $\phi$ 'parametri di disturbo'
- $p(\theta|\phi)$ coniugata alla varosimiglianza $p(y|\theta)$

Derivazione analitica delle distribuzioni condizionali e marginali

1. si scriva la distr congiunta a posteriori in forma non-normalizzata (passo immediato): $p(\theta, \phi|y) \propto p(\phi)p(\theta|\phi)p(y|\theta)$
2. si determini analiticamente $p(\theta|\phi, y)$ (semplice in modelli coniugati dato che $p(\boldsymbol{\theta}|\phi) = \prod_{j=1}^{J} p(\theta_j|\phi)$, quindi $p(\theta|\phi, y)$ è il prodotto di densità a posteriori coniugate per le componenti $\theta_j$)
3. si determini $p(\phi|y)$, due modi
   - $p(\phi|y) = \int p(\theta, \phi|y) d\theta$

## Calcolo nei modelli gerarchici coniugati

La strategia inferenziale per i modelli gerarchici segue l'approccio generale adottato con i problemi multiparametrici

- $(\theta, \phi)$ con $\theta$ 'parametri d'interesse' e $\phi$ 'parametri di disturbo'
- $p(\theta|\phi)$ coniugata alla varosimiglianza $p(y|\theta)$

Derivazione analitica delle distribuzioni condizionali e marginali

1. si scriva la distr congiunta a posteriori in forma non-normalizzata (passo immediato): $p(\theta, \phi|y) \propto p(\phi)p(\theta|\phi)p(y|\theta)$
2. si determini analiticamente $p(\theta|\phi, y)$ (semplice in modelli coniugati dato che $p(\boldsymbol{\theta}|\phi) = \prod_{j=1}^{J} p(\theta_j|\phi)$, quindi $p(\theta|\phi, y)$ è il prodotto di densità a posteriori coniugate per le componenti $\theta_j$)
3. si determini $p(\phi|y)$, due modi
   - $p(\phi|y) = \int p(\theta, \phi|y) d\theta$
   - $p(\phi|y) = \frac{p(\theta,\phi|y)}{p(\theta|\phi,y)}$ (delicata per la determinaz della cost di norm di $p(\theta|\phi, y)$; non è d'aiuto quando l'int non ha sol in forma chiusa)

# Simulazione dalla distribuzione a posteriori

La seguente strategia è utile per simulare un'estrazione dalla distribuzione congiunta a posteriori $p(\theta, \phi|y)$

1. si estrae $\phi^*$ da $p(\phi|y)$

# Simulazione dalla distribuzione a posteriori

La seguente strategia è utile per simulare un'estrazione dalla distribuzione congiunta a posteriori $p(\theta, \phi|y)$

1. si estrae $\phi^*$ da $p(\phi|y)$
2. si estrae $\theta^*$ da $p(\theta|\phi^*, y)$

## Simulazione dalla distribuzione a posteriori

La seguente strategia è utile per simulare un'estrazione dalla distribuzione congiunta a posteriori $p(\theta, \phi|y)$

1. si estrae $\phi^*$ da $p(\phi|y)$
2. si estrae $\theta^*$ da $p(\theta|\phi^*, y)$
   - se vale la fattorizzazione $p(\boldsymbol{\theta}|\phi, y) = \prod_{j=1}^{J} p(\theta_j|\phi, y)$ allora si possono estrarre le componenti $\theta_j$ indipendentemente una alla volta

# Simulazione dalla distribuzione a posteriori

La seguente strategia è utile per simulare un'estrazione dalla distribuzione congiunta a posteriori $p(\theta, \phi|y)$

1. si estrae $\phi^*$ da $p(\phi|y)$
2. si estrae $\theta^*$ da $p(\theta|\phi^*, y)$
   - se vale la fattorizzazione $p(\boldsymbol{\theta}|\phi, y) = \prod_{j=1}^{J} p(\theta_j|\phi, y)$ allora si possono estrarre le componenti $\theta_j$ indipendentemente una alla volta
3. in caso, si estrae $\tilde{y}$ da $p(y|\theta^*)$

## Simulazione dalla distribuzione a posteriori

La seguente strategia è utile per simulare un'estrazione dalla distribuzione congiunta a posteriori $p(\theta, \phi|y)$

1. si estrae $\phi^*$ da $p(\phi|y)$
2. si estrae $\theta^*$ da $p(\theta|\phi^*, y)$
   - se vale la fattorizzazione $p(\boldsymbol{\theta}|\phi, y) = \prod_{j=1}^{J} p(\theta_j|\phi, y)$ allora si possono estrarre le componenti $\theta_j$ indipendentemente una alla volta
3. in caso, si estrae $\tilde{y}$ da $p(y|\theta^*)$
4. Si ripete la procedura $L$ volte per ottenere un insieme di $L$ estrazioni

# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

- $j = 1, \ldots, J, J = 71$, esperimenti

$$
\begin{aligned}
y_j | n_j, \theta_j &\sim \mathrm{Bin}(n_j, \theta_j) \quad \text{ind.}, n_j \,\text{noto} \\
\theta &\sim \mathrm{Beta}(\alpha, \beta) \quad \text{ind.} \\
(\alpha, \beta) &\sim \text{noninformative}
\end{aligned}
$$

# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

- $j = 1, \ldots, J$, $J = 71$, esperimenti

$$y_j | n_j, \theta_j \sim \text{Bin}(n_j, \theta_j) \quad \text{ind.}, n_j \text{ noto}$$
$$\theta \sim \text{Beta}(\alpha, \beta) \quad \text{ind.}$$
$$(\alpha, \beta) \sim \text{noninformative}$$

- Eseguiamo i tre passi per determinare la forma analitica della distribuzione a posteriori: distribuzioni congiunta, condizionale e marginale a posteriori

# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

- $j = 1, \ldots, J, J = 71$, esperimenti

$$
\begin{aligned}
y_j | n_j, \theta_j &\sim \mathrm{Bin}(n_j, \theta_j) \quad \text{ind.}, n_j \text{ noto} \\
\theta &\sim \mathrm{Beta}(\alpha, \beta) \quad \text{ind.} \\
(\alpha, \beta) &\sim \text{noninformative}
\end{aligned}
$$

- Eseguiamo i tre passi per determinare la forma analitica della distribuzione a posteriori: distribuzioni congiunta, condizionale e marginale a posteriori

1. $p(\theta, \alpha, \beta | y) \propto p(\alpha, \beta) p(\theta | \alpha, \beta) p(y | \theta)$

# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

- $j = 1, \ldots, J$, $J = 71$, esperimenti

$$\begin{aligned} y_j | n_j, \theta_j &\sim \mathrm{Bin}(n_j, \theta_j) \quad \text{ind.}, n_j \text{ noto} \\ \theta &\sim \mathrm{Beta}(\alpha, \beta) \quad \text{ind.} \\ (\alpha, \beta) &\sim \text{noninformative} \end{aligned}$$

- Eseguiamo i tre passi per determinare la forma analitica della distribuzione a posteriori: distribuzioni congiunta, condizionale e marginale a posteriori

1. $p(\theta, \alpha, \beta | y) \propto p(\alpha, \beta) p(\theta | \alpha, \beta) p(y | \theta)$
2. $p(\theta | \alpha, \beta, y) = \prod_j \mathrm{Beta}(\alpha + y_j, \beta + n_j - y_j)$

# Modello binomiale gerarchico
## Esempio: Incidenza di tumore nei topi

- $j = 1, \ldots, J$, $J = 71$, esperimenti

$$y_j | n_j, \theta_j \sim \mathrm{Bin}(n_j, \theta_j) \quad \text{ind.}, n_j \text{ noto}$$
$$\theta \sim \mathrm{Beta}(\alpha, \beta) \quad \text{ind.}$$
$$(\alpha, \beta) \sim \text{noninformative}$$

- Eseguiamo i tre passi per determinare la forma analitica della distribuzione a posteriori: distribuzioni congiunta, condizionale e marginale a posteriori

1. $p(\theta, \alpha, \beta | y) \propto p(\alpha, \beta) p(\theta | \alpha, \beta) p(y | \theta)$
2. $p(\theta | \alpha, \beta, y) = \prod_j \mathrm{Beta}(\alpha + y_j, \beta + n_j - y_j)$
3. $p(\alpha, \beta | y) = \frac{p(\theta, \alpha, \beta | y)}{p(\theta | \alpha, \beta, y)}$ (BDA3, p. 110) equazione che non può essere semplificata analiticamente ma è facile da calcolare per qualsiasi valore specificato di $(\alpha, \beta)$

# Modello binomiale gerarchico
## Esempio: Incidenza di tumore nei topi

3. Simulazione da $p(\alpha, \beta|y)$

# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

3. Simulazione da $p(\alpha, \beta|y)$
   3.1 Valutiamo $p(\alpha, \beta|y)$ su una griglia di valori

# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

3. Simulazione da $p(\alpha, \beta|y)$
   3.1 Valutiamo $p(\alpha, \beta|y)$ su una griglia di valori
   3.2 Approssimiamo secondo una funzione a gradini e imponiamo somma prob a 1

# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

3. Simulazione da $p(\alpha, \beta | y)$
   - 3.1 Valutiamo $p(\alpha, \beta | y)$ su una griglia di valori
   - 3.2 Approssimiamo secondo una funzione a gradini e imponiamo somma prob a 1
   - 3.3 Campioniamo $\alpha_l \sim p(\alpha | y)$ e $\beta_l \sim p(\beta | \alpha_l, y)$ usando il metodo *inverse-cdf*

# Campionamento con il metodo *inverse-cdf*

$$F(v^*) = P(v \leq v^*) = \begin{cases} \sum v < v^* \, p(v) \\ \int_{-\infty}^{v^*} p(v) dv \end{cases}$$

1. Estrai $u \sim U(0,1)$

# Campionamento con il metodo *inverse-cdf*

$$F(v^*) = P(v \leq v^*) = \begin{cases} \sum v < v^* \, p(v) \\ \int_{-\infty}^{v^*} p(v) dv \end{cases}$$

1. Estrai $u \sim U(0, 1)$
2. Calcola $v = F^{-1}(u)$, quindi $v \sim p$

# Campionamento con il metodo *inverse-cdf*

$$F(v^*) = P(v \leq v^*) = \begin{cases} \sum v < v^* \, p(v) \\ \int_{-\infty}^{v^*} p(v) dv \end{cases}$$

1. Estrai $u \sim U(0,1)$
2. Calcola $v = F^{-1}(u)$, quindi $v \sim p$

Esempio

- $v \sim \text{Exp}(\lambda)$

# Campionamento con il metodo *inverse-cdf*

$$F(v^*) = P(v \leq v^*) = \begin{cases} \sum v < v^* \, p(v) \\ \int_{-\infty}^{v^*} p(v) dv \end{cases}$$

1. Estrai $u \sim U(0,1)$
2. Calcola $v = F^{-1}(u)$, quindi $v \sim p$

Esempio

- $v \sim \text{Exp}(\lambda)$
- $F(v) = 1 - e^{-\lambda v)}$

# Campionamento con il metodo *inverse-cdf*

$$F(v^*) = P(v \leq v^*) = \begin{cases} \sum v < v^* \, p(v) \\ \int_{-\infty}^{v^*} p(v) dv \end{cases}$$

1. Estrai $u \sim U(0,1)$
2. Calcola $v = F^{-1}(u)$, quindi $v \sim p$

Esempio

- $v \sim \text{Exp}(\lambda)$
- $F(v) = 1 - e^{-\lambda v)}$
- $v = F^{-1}(u)$ iff $-\log(1-u)/\lambda$

# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

Specifichiamo una distribuzione *hyperprior* non-informativa per $(\alpha, \beta)$

- Sia $\gamma = \text{logit}\frac{\alpha}{\alpha+\beta} = \log\frac{\alpha}{\beta}$ e $\delta = \log(\alpha + \beta)$

# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

Specifichiamo una distribuzione *hyperprior* non-informativa per $(\alpha, \beta)$

- Sia $\gamma = \text{logit} \frac{\alpha}{\alpha+\beta} = \log \frac{\alpha}{\beta}$ e $\delta = \log(\alpha + \beta)$

  $p(\gamma, \delta) \propto \text{constanton}(-\infty, \infty)$

# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

Specifichiamo una distribuzione *hyperprior* non-informativa per $(\alpha, \beta)$

- Sia $\gamma = \text{logit}\frac{\alpha}{\alpha+\beta} = \log\frac{\alpha}{\beta}$ e $\delta = \log(\alpha+\beta)$

  $p(\gamma, \delta) \propto \text{constant on} (-\infty, \infty)$

  Questa a priori porta ad una a post impropria

## Modello binomiale gerarchico

### Esempio: Incidenza di tumore nei topi

Specifichiamo una distribuzione *hyperprior* non-informativa per $(\alpha, \beta)$

- Sia $\gamma = \text{logit}\frac{\alpha}{\alpha+\beta} = \log\frac{\alpha}{\beta}$ e $\delta = \log(\alpha+\beta)$
  $p(\gamma, \delta) \propto \text{constanton}(-\infty, \infty)$
  Questa a priori porta ad una a post impropria
- Con una numerosità campionaria ragionevolmente alta, possiamo specificare una iperpriori 'noninformativa' dominata dalla verosimiglianza e che conduce ad una distr a post propria:

# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

Specifichiamo una distribuzione *hyperprior* non-informativa per $(\alpha, \beta)$

- Sia $\gamma = \text{logit}\frac{\alpha}{\alpha+\beta} = \log\frac{\alpha}{\beta}$ e $\delta = \log(\alpha+\beta)$
  $p(\gamma, \delta) \propto \text{constanton}(-\infty, \infty)$
  Questa a priori porta ad una a post impropria
- Con una numerosità campionaria ragionevolmente alta, possiamo specificare una iperpriori 'noninformativa' dominata dalla verosimiglianza e che conduce ad una distr a post propria:
  $(\gamma_1, \delta_1)(\frac{\alpha}{\alpha+\beta}, (\alpha+\beta)^{-1/2})$ uniforme, che moltiplicata per l'opportuno Jacobiano ha densità

# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

Specifichiamo una distribuzione *hyperprior* non-informativa per $(\alpha, \beta)$

- Sia $\gamma = \text{logit}\frac{\alpha}{\alpha+\beta} = \log\frac{\alpha}{\beta}$ e $\delta = \log(\alpha+\beta)$
  $p(\gamma, \delta) \propto \text{constanton}(-\infty, \infty)$
  Questa a priori porta ad una a post impropria
- Con una numerosità campionaria ragionevolmente alta, possiamo specificare una iperpriori ‘noninformativa’ dominata dalla verosimiglianza e che conduce ad una distr a post propria:
  $(\gamma_1, \delta_1)(\frac{\alpha}{\alpha+\beta}, (\alpha+\beta)^{-1/2})$ uniforme, che moltiplicata per l’opportuno Jacobiano ha densità
  - nella scala originaria

$$p(\alpha, \beta) \propto (\alpha+\beta)^{-5/2}$$

# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

Specifichiamo una distribuzione *hyperprior* non-informativa per $(\alpha, \beta)$

- Sia $\gamma = \text{logit}\frac{\alpha}{\alpha+\beta} = \log\frac{\alpha}{\beta}$ e $\delta = \log(\alpha+\beta)$

  $p(\gamma, \delta) \propto \text{constanton}(-\infty, \infty)$

  Questa a priori porta ad una a post impropria
- Con una numerosità campionaria ragionevolmente alta, possiamo specificare una iperpriori 'noninformativa' dominata dalla verosimiglianza e che conduce ad una distr a post propria:

  $(\gamma_1, \delta_1)(\frac{\alpha}{\alpha+\beta}, (\alpha+\beta)^{-1/2})$ uniforme, che moltiplicata per l'opportuno Jacobiano ha densità
  - nella scala originaria

    $$p(\alpha, \beta) \propto (\alpha+\beta)^{-5/2}$$
  - nella scala della trasformata naturale sopra

    $$p(\gamma, \delta) \propto \alpha\beta(\alpha+\beta)^{-5/2}$$

**Calcolo della densità marginale a posteriori**

- Calcoliamo $p(\gamma, \delta|y)$ nonnormalizzata su una griglia di valori (curve di livello)

## Calcolo della densità marginale a posteriori

- Calcoliamo $p(\gamma, \delta|y)$ nonnormalizzata su una griglia di valori (curve di livello)
  - calcolo del $\log$ della fdp $p(\alpha, \beta|y)$ con a priori $p(\alpha, \beta) \propto (\alpha + \beta)^{-5/2}$, moltiplicato per lo Jacobiano per fornire $p(\gamma, \delta|y)$

### Calcolo della densità marginale a posteriori

- Calcoliamo $p(\gamma, \delta|y)$ nonnormalizzata su una griglia di valori (curve di livello)
  - calcolo del $\log$ della fdp $p(\alpha, \beta|y)$ con a priori $p(\alpha, \beta) \propto (\alpha + \beta)^{-5/2}$, moltiplicato per lo Jacobiano per fornire $p(\gamma, \delta|y)$
  - griglia $(\gamma, \delta) \in [-2., -1] \times [1.5, 3]$ centrata nella stima empirica basata sui dati storici, $(\hat{\alpha}, \hat{\beta}) = (1.4, 8.6)$, che in termini di $(\gamma, \delta)$ è $(-2.5, -1)$ e che ha una copertura con fattore $4$ per ogni parametro

## Calcolo della densità marginale a posteriori

- Calcoliamo $p(\gamma, \delta|y)$ nonnormalizzata su una griglia di valori (curve di livello)
  - calcolo del $\log$ della fdp $p(\alpha, \beta|y)$ con a priori $p(\alpha, \beta) \propto (\alpha + \beta)^{-5/2}$, moltiplicato per lo Jacobiano per fornire $p(\gamma, \delta|y)$
  - griglia $(\gamma, \delta) \in [-2., -1] \times [1.5, 3]$ centrata nella stima empirica basata sui dati storici, $(\hat{\alpha}, \hat{\beta}) = (1.4, 8.6)$, che in termini di $(\gamma, \delta)$ è $(-2.5, -1)$ e che ha una copertura con fattore $4$ per ogni parametro
  - si sottrae il $\max$ della log densità ed infine si esponenzia per ottenere la $p(\gamma, \delta|y)$ nonnormalizzata

## Calcolo della densità marginale a posteriori

- Calcoliamo $p(\gamma, \delta|y)$ nonnormalizzata su una griglia di valori (curve di livello)
  - calcolo del $\log$ della fdp $p(\alpha, \beta|y)$ con a priori $p(\alpha, \beta) \propto (\alpha + \beta)^{-5/2}$, moltiplicato per lo Jacobiano per fornire $p(\gamma, \delta|y)$
  - griglia $(\gamma, \delta) \in [-2., -1] \times [1.5, 3]$ centrata nella stima empirica basata sui dati storici, $(\hat{\alpha}, \hat{\beta}) = (1.4, 8.6)$, che in termini di $(\gamma, \delta)$ è $(-2.5, -1)$ e che ha una copertura con fattore $4$ per ogni parametro
  - si sottrae il $\max$ della log densità ed infine si esponenzia per ottenere la $p(\gamma, \delta|y)$ nonnormalizzata
  - poichè parti importanti della marg a post cadono fuori il range, si ricalcola su $[-2.3, -1.3] \times [1, 5]$

### Calcolo della densità marginale a posteriori

- Calcoliamo $p(\gamma, \delta|y)$ nonnormalizzata su una griglia di valori (curve di livello)
  - calcolo del $\log$ della fdp $p(\alpha, \beta|y)$ con a priori $p(\alpha, \beta) \propto (\alpha + \beta)^{-5/2}$, moltiplicato per lo Jacobiano per fornire $p(\gamma, \delta|y)$
  - griglia $(\gamma, \delta) \in [-2., -1] \times [1.5, 3]$ centrata nella stima empirica basata sui dati storici, $(\hat{\alpha}, \hat{\beta}) = (1.4, 8.6)$, che in termini di $(\gamma, \delta)$ è $(-2.5, -1)$ e che ha una copertura con fattore $4$ per ogni parametro
  - si sottrae il $\max$ della log densità ed infine si esponenzia per ottenere la $p(\gamma, \delta|y)$ nonnormalizzata
  - poichè parti importanti della marg a post cadono fuori il range, si ricalcola su $[-2.3, -1.3] \times [1, 5]$
- Normalizziamo la densità a post relativa approssimandola attraverso una funzione a gradini e imponendo che la prob totale sulla griglia sia 1

**Calcolo della densità marginale a posteriori**

- Calcoliamo $p(\gamma, \delta|y)$ nonnormalizzata su una griglia di valori (curve di livello)
  - calcolo del $\log$ della fdp $p(\alpha, \beta|y)$ con a priori $p(\alpha, \beta) \propto (\alpha + \beta)^{-5/2}$, moltiplicato per lo Jacobiano per fornire $p(\gamma, \delta|y)$
  - griglia $(\gamma, \delta) \in [-2., -1] \times [1.5, 3]$ centrata nella stima empirica basata sui dati storici, $(\hat{\alpha}, \hat{\beta}) = (1.4, 8.6)$, che in termini di $(\gamma, \delta)$ è $(-2.5, -1)$ e che ha una copertura con fattore $4$ per ogni parametro
  - si sottrae il $\max$ della log densità ed infine si esponenzia per ottenere la $p(\gamma, \delta|y)$ nonnormalizzata
  - poichè parti importanti della marg a post cadono fuori il range, si ricalcola su $[-2.3, -1.3] \times [1, 5]$
- Normalizziamo la densità a post relativa approssimandola attraverso una funzione a gradini e imponendo che la prob totale sulla griglia sia 1
- Possiamo calcolare momenti a posteriori, ad es $E(\alpha|y)$ e $E(\beta|y)$ (che risultano $(2.4, 14.3)$ vicini alla moda per la simmetria della distr nella scala $(\gamma, \delta)$)

# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

### Campionamento dalla densità marginale a posteriori

- Estrai $1000$ valori dalla congiunta a post $p(\alpha, \beta, \boldsymbol{\theta}|y)$ in questo modo

# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

### Campionamento dalla densità marginale a posteriori

- Estrai $1000$ valori dalla congiunta a post $p(\alpha, \beta, \boldsymbol{\theta}|y)$ in questo modo
  1. estrai $\gamma^l, \delta^l$ dalla loro a post $p(\gamma, \delta|y)$ con la procedura del campionamento *discrete-grid*

# Modello binomiale gerarchico
## Esempio: Incidenza di tumore nei topi

### Campionamento dalla densità marginale a posteriori

- Estrai $1000$ valori dalla congiunta a post $p(\alpha, \beta, \boldsymbol{\theta}|y)$ in questo modo
  1. estrai $\gamma^l, \delta^l$ dalla loro a post $p(\gamma, \delta|y)$ con la procedura del campionamento *discrete-grid*
  2. per $l = 1, \ldots, 1000$:

# Modello binomiale gerarchico
## Esempio: Incidenza di tumore nei topi

### Campionamento dalla densità marginale a posteriori

- Estrai $1000$ valori dalla congiunta a post $p(\alpha, \beta, \boldsymbol{\theta}|y)$ in questo modo
  1. estrai $\gamma^l, \delta^l$ dalla loro a post $p(\gamma, \delta|y)$ con la procedura del campionamento *discrete-grid*
  2. per $l = 1, \ldots, 1000$:
     - trasforma $\gamma^l, \delta^l \rightarrow \alpha^l, \beta^l$

# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

### Campionamento dalla densità marginale a posteriori

- Estrai $1000$ valori dalla congiunta a post $p(\alpha, \beta, \boldsymbol{\theta}|y)$ in questo modo
  1. estrai $\gamma^l, \delta^l$ dalla loro a post $p(\gamma, \delta|y)$ con la procedura del campionamento *discrete-grid*
  2. per $l = 1, \ldots, 1000$:
     - trasforma $\gamma^l, \delta^l \rightarrow \alpha^l, \beta^l$
     - per ogni $j$ estrai $\theta_j^l$ dalla cond a post $\mathrm{Beta}(\alpha^l + y_j, \beta^l + n_j - y_j)$

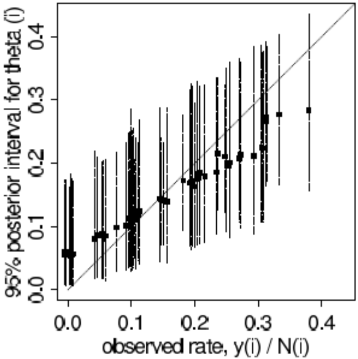
# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

### Campionamento dalla densità marginale a posteriori

- Estrai $1000$ valori dalla congiunta a post $p(\alpha, \beta, \boldsymbol{\theta}|y)$ in questo modo
  1. estrai $\gamma^l, \delta^l$ dalla loro a post $p(\gamma, \delta|y)$ con la procedura del campionamento *discrete-grid*
  2. per $l = 1, \ldots, 1000$:
     - trasforma $\gamma^l, \delta^l \rightarrow \alpha^l, \beta^l$
     - per ogni $j$ estrai $\theta_j^l$ dalla cond a post $\mathrm{Beta}(\alpha^l + y_j, \beta^l + n_j - y_j)$
- mostra i risultati

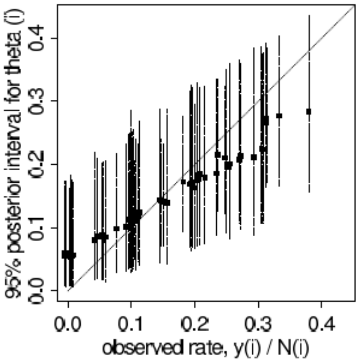
# Modello binomiale gerarchico
## Esempio: Incidenza di tumore nei topi




- $\theta_j$: *shrinkage* verso la media a posteriori $0.14$

# Modello binomiale gerarchico

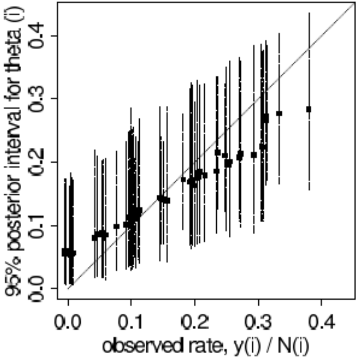
## Esempio: Incidenza di tumore nei topi




- $\theta_j$: *shrinkage* verso la media a posteriori $0.14$
  - esperimenti con meno osservazioni mostrano uno *shrinkage* più evidente e varianza a post maggiore

# Modello binomiale gerarchico

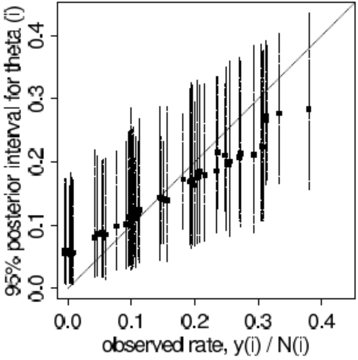
## Esempio: Incidenza di tumore nei topi




- $\theta_j$: *shrinkage* verso la media a posteriori $0.14$
  - esperimenti con meno osservazioni mostrano uno *shrinkage* più evidente e varianza a post maggiore
- risultati simili a quanto si ottiene con EB (per il numero piuttosto alto degli esperimenti)

# Modello binomiale gerarchico

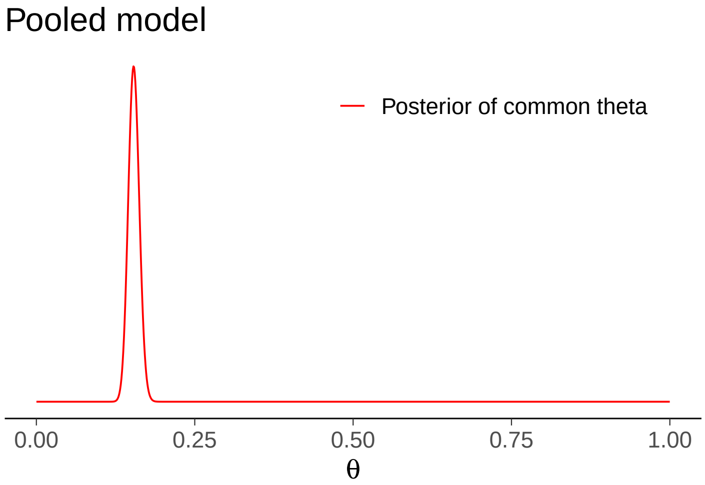
## Esempio: Incidenza di tumore nei topi




- $\theta_j$: *shrinkage* verso la media a posteriori $0.14$
  - esperimenti con meno osservazioni mostrano uno *shrinkage* più evidente e varianza a post maggiore
- risultati simili a quanto si ottiene con EB (per il numero piuttosto alto degli esperimenti)
- variabilità a posteriori maggiore che con EB (per incertezza a post degli iperparametri) (Nota: nel laboratorio, questo non si verifica!)

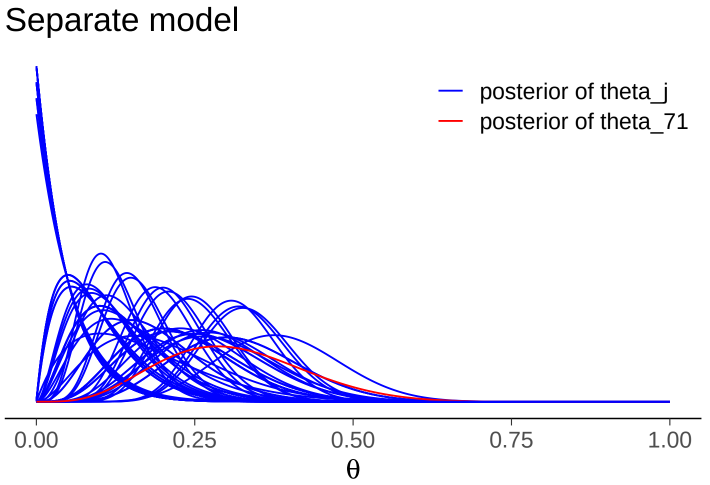
# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi



# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi



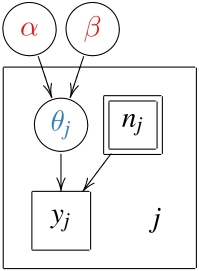
# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

- Hierarchical binomial model for rats
  prior parameters $\alpha$ and $\beta$ are unknown

$$\theta_j \mid \alpha, \beta \sim \mathrm{Beta}(\theta_j \mid \alpha, \beta)$$

$$y_j \mid n_j, \theta_j \sim \mathrm{Bin}(y_j \mid n_j, \theta_j)$$



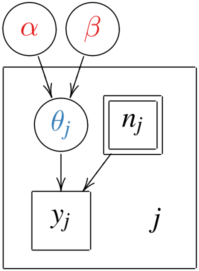
# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

- Hierarchical binomial model for rats
  prior parameters $\alpha$ and $\beta$ are unknown

$$\theta_j \mid \alpha, \beta \sim \mathrm{Beta}(\theta_j \mid \alpha, \beta)$$

$$y_j \mid n_j, \theta_j \sim \mathrm{Bin}(y_j \mid n_j, \theta_j)$$

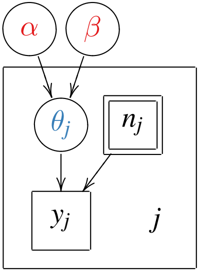



- Joint posterior $p(\theta_1, \ldots, \theta_J, \alpha, \beta \mid y)$

# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

- Hierarchical binomial model for rats
  prior parameters $\alpha$ and $\beta$ are unknown




$$\theta_j \mid \alpha, \beta \sim \mathrm{Beta}(\theta_j \mid \alpha, \beta)$$

$$y_j \mid n_j, \theta_j \sim \mathrm{Bin}(y_j \mid n_j, \theta_j)$$

- Joint posterior $p(\theta_1, \ldots, \theta_J, \alpha, \beta \mid y)$
  - multiple parameters

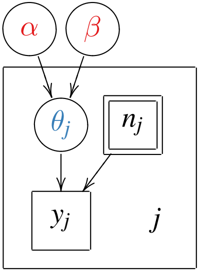
# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

- Hierarchical binomial model for rats
  prior parameters $\alpha$ and $\beta$ are unknown

$$\theta_j \mid \alpha, \beta \sim \mathrm{Beta}(\theta_j \mid \alpha, \beta)$$

$$y_j \mid n_j, \theta_j \sim \mathrm{Bin}(y_j \mid n_j, \theta_j)$$




- Joint posterior $p(\theta_1, \ldots, \theta_J, \alpha, \beta \mid y)$
  - multiple parameters
  - factorize $\prod_{j=1}^{J} p(\theta_j \mid \alpha, \beta, y) p(\alpha, \beta \mid y)$

# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

- Population prior $\mathrm{Beta}(\theta_j \mid \alpha, \beta)$

# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

- Population prior $\mathrm{Beta}(\theta_j \mid \alpha, \beta)$
- Hyperprior $p(\alpha, \beta)$?

# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

- Population prior $\mathrm{Beta}(\theta_j \mid \alpha, \beta)$
- Hyperprior $p(\alpha, \beta)$?
  - $\alpha, \beta$ both affect the location and scale

# Modello binomiale gerarchico
## Esempio: Incidenza di tumore nei topi

- Population prior $\mathrm{Beta}(\theta_j \mid \alpha, \beta)$
- Hyperprior $p(\alpha, \beta)$?
  - $\alpha, \beta$ both affect the location and scale
  - BDA3 has $p(\alpha, \beta) \propto (\alpha + \beta)^{-5/2}$

# Modello binomiale gerarchico

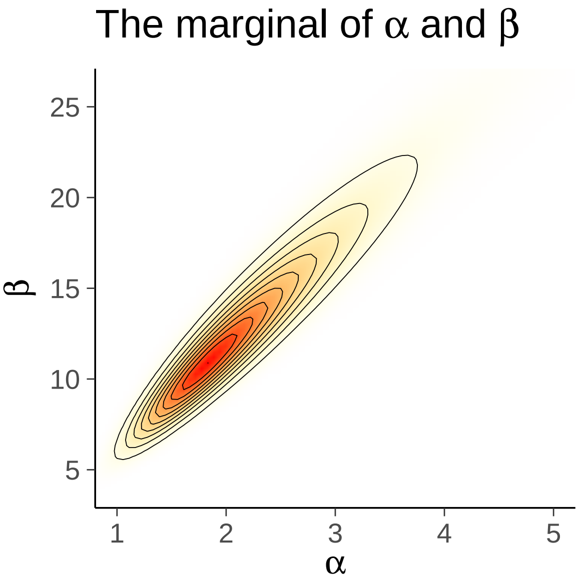
## Esempio: Incidenza di tumore nei topi

- Population prior $\mathrm{Beta}(\theta_j \mid \alpha, \beta)$
- Hyperprior $p(\alpha, \beta)$?
  - $\alpha, \beta$ both affect the location and scale
  - BDA3 has $p(\alpha, \beta) \propto (\alpha + \beta)^{-5/2}$
    - diffuse prior for location and scale (BDA3 p. 110)

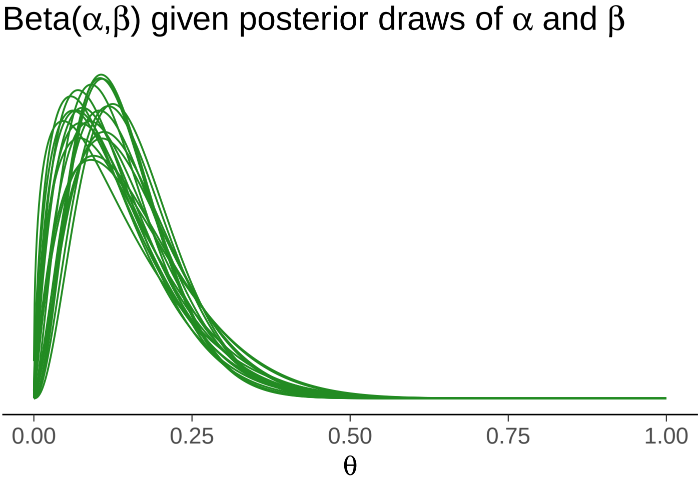
# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

- Population prior $\mathrm{Beta}(\theta_j \mid \alpha, \beta)$
- Hyperprior $p(\alpha, \beta)$?
  - $\alpha, \beta$ both affect the location and scale
  - BDA3 has $p(\alpha, \beta) \propto (\alpha + \beta)^{-5/2}$
    - diffuse prior for location and scale (BDA3 p. 110)
- lab


The marginal of α and β
β
25
20
15
10
5
1
2
3
4
5
α

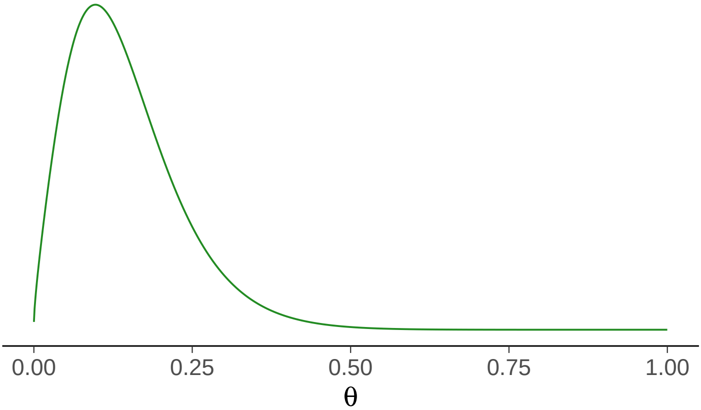
# Modello binomiale gerarchico
## Esempio: Incidenza di tumore nei topi



## Modello binomiale gerarchico

### Esempio: Incidenza di tumore nei topi

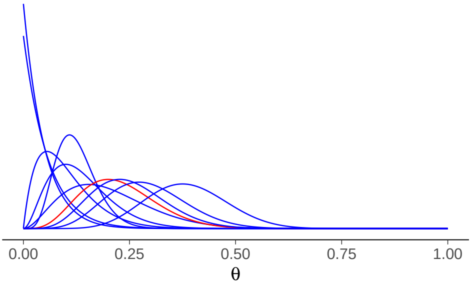
Population distribution (prior) for $\theta_j$



# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

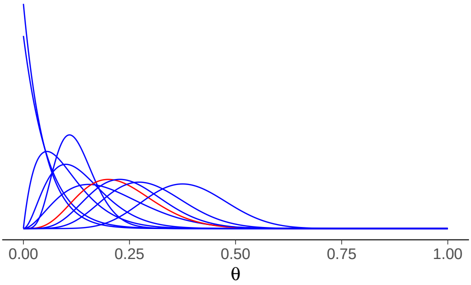
Separate model



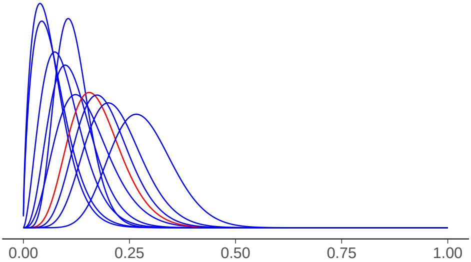
# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

Separate model



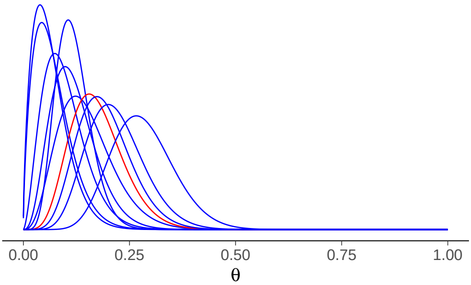

Hierarchical model



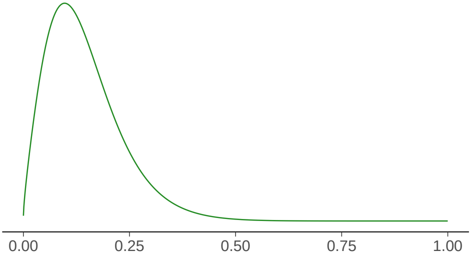
# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi

Hierarchical model

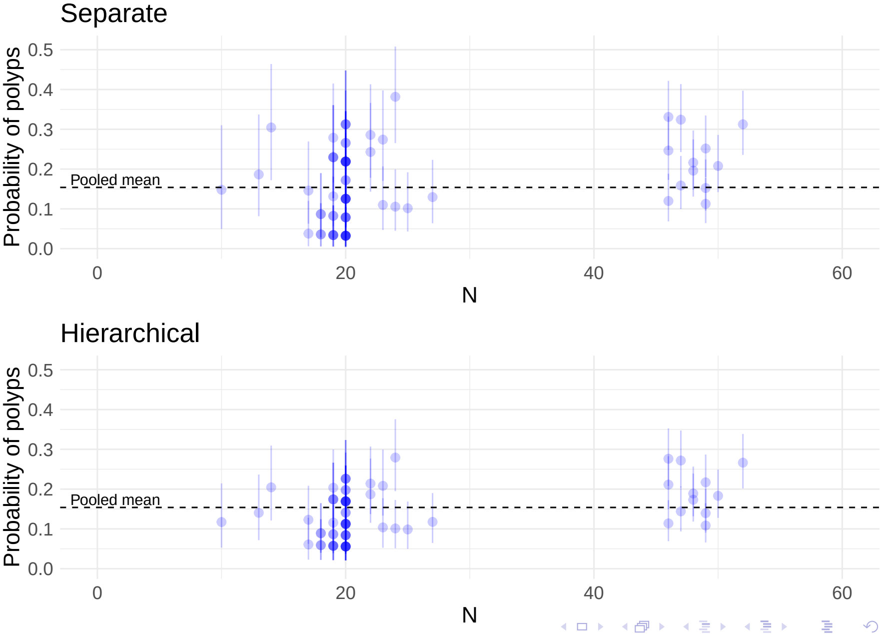



Population distribution (prior) for $\theta_j$



# Modello binomiale gerarchico

## Esempio: Incidenza di tumore nei topi. Effetto della numerosità dei gruppi



# Indice della sezione I

# Alcuni riferimenti

Robert C. (2007) *The Bayesian Choice* 2nd ed. Springer def. di HBM

## Alcuni riferimenti


Robert C. (2007) *The Bayesian Choice* 2nd ed. Springer ◂ def. di HBM

Gelman *et al.* (2004) *Bayesian Data Analysis* 2nd ed. London: Chapman & Hall ◂ esempio NN

# Alcuni riferimenti


Robert C. (2007) *The Bayesian Choice* 2nd ed. Springer ◂ def. di HBM

Gelman *et al.* (2004) *Bayesian Data Analysis* 2nd ed. London: Chapman & Hall ◂ esempio NN

Carlin B. and Louis T. (2000) *Bayes and Empirical Bayes Methods for Data Analysis* 2nd ed. London: Chapman & Hall

## Alcuni riferimenti

Robert C. (2007) *The Bayesian Choice* 2nd ed. Springer ◂ def. di HBM

Gelman *et al.* (2004) *Bayesian Data Analysis* 2nd ed. London: Chapman & Hall ◂ esempio NN

Carlin B. and Louis T. (2000) *Bayes and Empirical Bayes Methods for Data Analysis* 2nd ed. London: Chapman & Hall

Carlin B. and Louis T. (2008) *Bayesian Methods for Data Analysis* 3rd ed. Chapman & Hall/CRC Texts in Statistical Science

## Alcuni riferimenti


Robert C. (2007) *The Bayesian Choice* 2nd ed. Springer ◂ def. di HBM

Gelman *et al.* (2004) *Bayesian Data Analysis* 2nd ed. London: Chapman & Hall ◂ esempio NN

Carlin B. and Louis T. (2000) *Bayes and Empirical Bayes Methods for Data Analysis* 2nd ed. London: Chapman & Hall

Carlin B. and Louis T. (2008) *Bayesian Methods for Data Analysis* 3rd ed. Chapman & Hall/CRC Texts in Statistical Science

Gelman A. and Hill J. (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models* Cambridge University Press

## Alcuni riferimenti


Robert C. (2007) *The Bayesian Choice* 2nd ed. Springer ◂ def. di HBM

Gelman *et al.* (2004) *Bayesian Data Analysis* 2nd ed. London: Chapman & Hall ◂ esempio NN

Carlin B. and Louis T. (2000) *Bayes and Empirical Bayes Methods for Data Analysis* 2nd ed. London: Chapman & Hall

Carlin B. and Louis T. (2008) *Bayesian Methods for Data Analysis* 3rd ed. Chapman & Hall/CRC Texts in Statistical Science

Gelman A. and Hill J. (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models* Cambridge University Press

Gilks, Richardson, and Spiegelhalter (1995) *Markov Chain Monte Carlo in Practice*

## Alcuni riferimenti


Robert C. (2007) *The Bayesian Choice* 2nd ed. Springer ◂ def. di HBM

Gelman *et al.* (2004) *Bayesian Data Analysis* 2nd ed. London: Chapman & Hall ◂ esempio NN

Carlin B. and Louis T. (2000) *Bayes and Empirical Bayes Methods for Data Analysis* 2nd ed. London: Chapman & Hall

Carlin B. and Louis T. (2008) *Bayesian Methods for Data Analysis* 3rd ed. Chapman & Hall/CRC Texts in Statistical Science

Gelman A. and Hill J. (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models* Cambridge University Press

Gilks, Richardson, and Spiegelhalter (1995) *Markov Chain Monte Carlo in Practice*

Congdon P. (2006) *Bayesian Statistical Modelling* 2nd ed. Wiley

## Alcuni riferimenti


Robert C. (2007) *The Bayesian Choice* 2nd ed. Springer def. di HBM

Gelman *et al.* (2004) *Bayesian Data Analysis* 2nd ed. London: Chapman & Hall esempio NN

Carlin B. and Louis T. (2000) *Bayes and Empirical Bayes Methods for Data Analysis* 2nd ed. London: Chapman & Hall

Carlin B. and Louis T. (2008) *Bayesian Methods for Data Analysis* 3rd ed. Chapman & Hall/CRC Texts in Statistical Science

Gelman A. and Hill J. (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models* Cambridge University Press

Gilks, Richardson, and Spiegelhalter (1995) *Markov Chain Monte Carlo in Practice*

Congdon P. (2006) *Bayesian Statistical Modelling* 2nd ed. Wiley

Congdon P. (2003) *Applied Bayesian Modelling* Wiley

## Alcuni riferimenti


Robert C. (2007) *The Bayesian Choice* 2nd ed. Springer ◂ def. di HBM

Gelman *et al.* (2004) *Bayesian Data Analysis* 2nd ed. London: Chapman & Hall ◂ esempio NN

Carlin B. and Louis T. (2000) *Bayes and Empirical Bayes Methods for Data Analysis* 2nd ed. London: Chapman & Hall

Carlin B. and Louis T. (2008) *Bayesian Methods for Data Analysis* 3rd ed. Chapman & Hall/CRC Texts in Statistical Science

Gelman A. and Hill J. (2007) *Data Analysis Using Regression and Multilevel/Hierarchical Models* Cambridge University Press

Gilks, Richardson, and Spiegelhalter (1995) *Markov Chain Monte Carlo in Practice*

Congdon P. (2006) *Bayesian Statistical Modelling* 2nd ed. Wiley

Congdon P. (2003) *Applied Bayesian Modelling* Wiley

Congdon P. (2005) *Bayesian Models for Categorical Data* Wiley