- Modelli multiparametrici
  - marginalizzazione
- Modello Normale con media e varianza non note
  - a priori non-informativa
  - a priori coniugata e semi-coniugata
- Integrazione Monte Carlo (es per dati normali)
- Modello Multinomiale con a priori coniugata Dirichlet
- (Modello Normale multivariato con media non nota e covarianza nota/non nota)
- Applicazione: analisi di un esperimento biologico

- θ_1, θ_2
- θ_1 é il parametro d'interesse
- θ_2 é parametro di disturbo
- Distribuzione congiunta dei parametri (Teorema di Bayes)

$$p(\theta_1, \theta_2 \mid y) \propto p(y \mid \theta_1, \theta_2) p(\theta_1, \theta_2)$$

- Marginalizzazione della congiunta a posteriori per trovare la distribuzione marginale a posteriori di θ_1 (*averaging over* θ_2)

$$p(\theta_1 \mid y) = \int p(\theta_1, \theta_2 \mid y) d\theta_2$$

- θ_1, θ_2
- θ_1 é il parametro d'interesse
- θ_2 é parametro di disturbo
- Distribuzione congiunta dei parametri (Teorema di Bayes)

$$p(\theta_1, \theta_2 \mid y) \propto p(y \mid \theta_1, \theta_2) p(\theta_1, \theta_2)$$

- Marginalizzazione della congiunta a posteriori per trovare la distribuzione marginale a posteriori di θ_1 (*averaging over* θ_2)

$$p(\theta_1 \mid y) = \int p(\theta_1, \theta_2 \mid y) d\theta_2$$

  - Una forma alternativa della fattorizzazione della distribuzione congiunta

$$p(\theta_1 \mid y) = \int p(\theta_1 \mid \theta_2, y) p(\theta_2 \mid y) d\theta_2$$

  che mostra come la $p(\theta_1 \mid y)$ sia una mistura di distribuzioni condizionali a posteriori dato θ_2 dove $p(\theta_2 \mid y)$ é una funzione peso per i possibili valori di θ_2.

- L'obiettivo é trovare la marginale a posteriori di una quantitá d'interesse
  - un parametro del modello
  - un evento futuro, ad es, la distribuzione predittiva a posteriori che é ottenuta *marginalizzando* la distribuzione a posteriori

$$p(\tilde{y} \mid y) = \int p(\tilde{y}, \theta \mid y) d\theta$$

- L'obiettivo é trovare la marginale a posteriori di una quantitá d'interesse
    - un parametro del modello
    - un evento futuro, ad es, la distribuzione predittiva a posteriori che é ottenuta *marginalizzando* la distribuzione a posteriori

$$\begin{aligned} p(\tilde{y} \mid y) &= \int p(\tilde{y}, \theta \mid y) d\theta \\ &= \int p(\tilde{y} \mid \theta) p(\theta \mid y) d\theta \end{aligned}$$

Monte Carlo e estrazioni dall'a posteriori

- Raramente gli integrali vengono calcolati esplicitamente.
- O si usa il metodo Monte Carlo
  - $\theta^{(s)}$ estrazioni da $p(\theta \mid y)$ possono essere usate per approssimare valori attesi (integrali)

$$E_{p(\theta|y)}[\theta] = \int \theta p(\theta \mid y) \approx \frac{1}{S} \sum_{s=1}^{S} \theta^{(s)}$$

  da cui anche l'approssimazione di valori attesi di funzioni

$$E_{p(g(\theta)|y)}[g(\theta)] = \int g(\theta) p(\theta \mid y) \approx \frac{1}{S} \sum_{s=1}^{S} g(\theta^{(s)})$$

- Raramente gli integrali vengono calcolati esplicitamente.
- O si usa il metodo Monte Carlo
  - $\theta^{(s)}$ estrazioni da $p(\theta \mid y)$ possono essere usate per approssimare valori attesi (integrali)

$$E_{p(\theta|y)}[\theta] = \int \theta p(\theta \mid y) \approx \frac{1}{S} \sum_{s=1}^{S} \theta^{(s)}$$

    da cui anche l'approssimazione di valori attesi di funzioni

$$E_{p(g(\theta)|y)}[g(\theta)] = \int g(\theta) p(\theta \mid y) \approx \frac{1}{S} \sum_{s=1}^{S} g(\theta^{(s)})$$

- come, ad es., l'approssimazione Monte Carlo della distribuzione marginale

$$p(\theta_1 \mid y) \approx \frac{1}{S} \sum_{s=1}^{S} p(\theta_1 | \theta_2^{(s)}, y),$$

dove $\theta_2^{(s)}$ sono estrazioni da $p(\theta_2 \mid y)$

- Raramente gli integrali vengono calcolati esplicitamente.
- O si usa il metodo Monte Carlo
  - $\theta^{(s)}$ estrazioni da $p(\theta \mid y)$ possono essere usate per approssimare valori attesi (integrali)

$$E_{p(\theta|y)}[\theta] = \int \theta p(\theta \mid y) \approx \frac{1}{S}\sum_{s=1}^{S} \theta^{(s)}$$

    da cui anche l'approssimazione di valori attesi di funzioni

$$E_{p(g(\theta)|y)}[g(\theta)] = \int g(\theta) p(\theta \mid y) \approx \frac{1}{S}\sum_{s=1}^{S} g(\theta^{(s)})$$

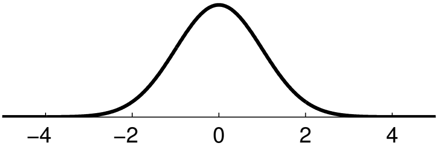
- come, ad es., l'approssimazione Monte Carlo della distribuzione marginale

$$p(\theta_1 \mid y) \approx \frac{1}{S}\sum_{s=1}^{S} p(\theta_1 | \theta_2^{(s)}, y),$$

  dove $\theta_2^{(s)}$ sono estrazioni da $p(\theta_2 \mid y)$
- oppure essa viene simulata procedendo in questo modo:

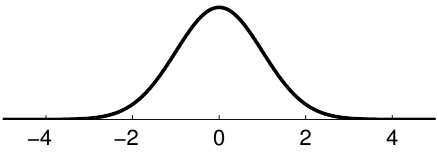
- estrai $\theta_2^{(m)} \sim p(\theta_2 \mid y)$
- estrai $\theta_1^{(m)} \sim p(\theta_1 \mid \theta_2^{(m)}, y)$
- $\theta_1^{(m)}$, $\theta_2^{(m)}$ per $m = 1, \ldots, M$

$$p(y|\mu) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y-\mu)^2\right)$$



Modello Normale / Gaussiano

$$p(y|\mu) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y-\mu)^2\right)$$



- Notazione abbreviata
  - $y \sim \mathrm{N}(\mu, \sigma^2)$ con varianza σ^2 (utile nelle derivazioni analitiche)
  - $y \sim \mathrm{N}(\mu, \sigma)$ con deviazione σ (utile per interpretare le scale a priori e a posteriori, usato in Stan)

Modello Normale / Gaussiano

- Regressione lineare normale
  connessione alla regressione *least squares* via $(y - \mu)^2$

Modello Normale / Gaussiano

- Regressione lineare normale
  connessione alla regressione *least squares* via $(y-\mu)^2$
- Analisi di osservazioni a valori reali

Modello Normale / Gaussiano

- Regressione lineare normale
  connessione alla regressione *least squares* via $(y - \mu)^2$
- Analisi di osservazioni a valori reali
- Talvolta approssimazione conveniente per osservazioni discrete

Modello Normale / Gaussiano

- Regressione lineare normale
  connessione alla regressione *least squares* via $(y - \mu)^2$
- Analisi di osservazioni a valori reali
- Talvolta approssimazione conveniente per osservazioni discrete
- Conveniente come distribuzione a priori

Modello Normale / Gaussiano

- Regressione lineare normale
  connessione alla regressione *least squares* via $(y - \mu)^2$
- Analisi di osservazioni a valori reali
- Talvolta approssimazione conveniente per osservazioni discrete
- Conveniente come distribuzione a priori
- Processi Gaussiani sono in pratica normali multivariate

Modello Normale / Gaussiano

- Regressione lineare normale
  connessione alla regressione *least squares* via $(y - \mu)^2$
- Analisi di osservazioni a valori reali
- Talvolta approssimazione conveniente per osservazioni discrete
- Conveniente come distribuzione a priori
- Processi Gaussiani sono in pratica normali multivariate
- filtri di Kalman sono normali piú *chain rule*

Modello Normale / Gaussiano

- Regressione lineare normale
  connessione alla regressione *least squares* via $(y - \mu)^2$
- Analisi di osservazioni a valori reali
- Talvolta approssimazione conveniente per osservazioni discrete
- Conveniente come distribuzione a priori
- Processi Gaussiani sono in pratica normali multivariate
- filtri di Kalman sono normali piú *chain rule*
- Come approssimazione della distribuzione a posteriori con Laplace, inferenza variazionale, *expectation propagation*

- μ é il parametro d'interesse
- σ_2^2 é parametro di disturbo

$$p(\mu, \sigma^2 \mid y) \propto p(y \mid \mu, \sigma^2) p(\mu, \sigma^2)$$
$$p(\mu \mid y) = \int p(\mu \mid \sigma^2, y) p(\sigma^2 \mid y) d\sigma^2$$

$p(\mu \mid y)$ é una mistura di distribuzioni condizionali a posteriori dato σ^2 dove $p(\sigma^2 \mid y)$ é una funzione peso per i possibili valori di σ^2.

Iniziamo con una a priori non-informativa e poi passsiamo ad una informativa.

Una a priori non-informativa convenzionale per $p(\mu, \sigma^2)$ é - assumendo indipendenza a priori tra μ e σ^2 - uniforme su $(\mu, \log \sigma)$ ovvero

$$p(\mu, \sigma^2) \propto \frac{1}{\sigma^2}$$

a priori impropria.

Modello Normale, $p(\mu, \sigma^2) \propto \frac{1}{\sigma^2}$

- Modello osservazionale
  $\mathbf{y} = (y_1, \ldots, y_n)$, si assuma scambiabilitá:
  $y_i | \mu, \sigma^2 \overset{iid}{\sim} N(\mu, \sigma^2)$,

$$y_i \sim \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y_i - \mu)^2\right)$$

- A priori non-informativa

$$p(\mu, \sigma^2) \propto \sigma^{-2}$$

Modello Normale, $p(\mu, \sigma^2) \propto \frac{1}{\sigma^2}$

- Modello osservazionale
  $\mathbf{y} = (y_1, \ldots, y_n)$, si assuma scambiabilitá:
  $y_i | \mu, \sigma^2 \overset{iid}{\sim} N(\mu, \sigma^2)$,
  $$y_i \sim \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y_i - \mu)^2\right)$$
- A priori non-informativa
  $$p(\mu, \sigma^2) \propto \sigma^{-2}$$
- Congiunta a posteriori

$$p(\mu, \sigma^2 \mid y) \propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{y} - \mu)^2\right]\right)$$

dove $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$

($\bar{y}$ e s^2 statistiche sufficienti)

Modello Normale, $p(\mu, \sigma^2) \propto \frac{1}{\sigma^2}$

Per fattorizzare la distr congiunta a posteriori come visto sopra deriviamo:

- distribuzione condizionale a posteriori per μ

$$p(\mu \mid \sigma^2, y) = \mathrm{N}(\mu \mid \bar{y}, \sigma^2/n) \propto \exp\left(-\frac{n}{2\sigma^2}(\bar{y}-\mu)^2\right)$$

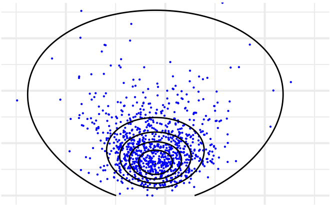
  che ricalca quanto ottenuto per l'a posteriori di μ quando σ^2 é noto e l'a priori per μ é uniforme.
- distribuzione marginale a posteriori per σ^2

$$p(\sigma^2 \mid y) = \mathrm{IG}((n-1)/2, s^2(n-1)/2)$$

ove s^2 é la varianza campionaria

Calcolo della congiunta a posteriori

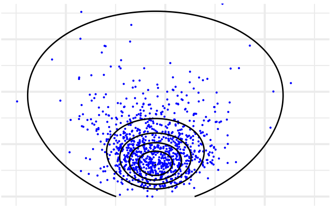
$$\mu^{(s)}, \sigma^{(s)} \sim p(\mu, \sigma \mid y)$$



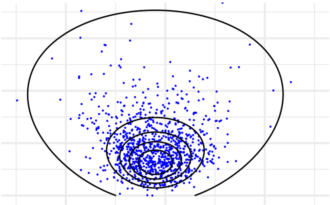
Calcolo della congiunta a posteriori

$$\mu^{(s)}, \sigma^{(s)} \sim p(\mu, \sigma \mid y)$$

con $p(\mu, \sigma^2) \propto \sigma^{-2}$



Calcolo della congiunta a posteriori

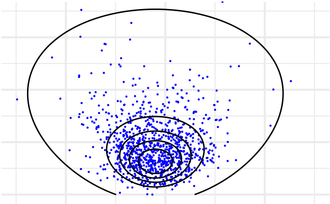


$$\mu^{(s)}, \sigma^{(s)} \sim p(\mu, \sigma \mid y)$$

con $p(\mu, \sigma^2) \propto \sigma^{-2}$

con $p(\mu, \sigma) \propto \sigma^{-1}$ (vedi BDA3 p. 21 cambio di variabile)

Calcolo della congiunta a posteriori

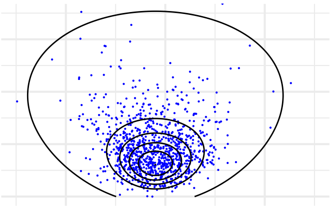


$$\mu^{(s)}, \sigma^{(s)} \sim p(\mu, \sigma \mid y)$$

con $p(\mu, \sigma^2) \propto \sigma^{-2}$

$$p(\mu, \sigma^2 \mid y) \propto \sigma^{-2} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y_i - \mu)^2\right)$$

Calcolo della congiunta a posteriori

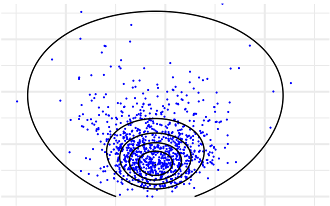


$$\mu^{(s)}, \sigma^{(s)} \sim p(\mu, \sigma \mid y)$$

con $p(\mu, \sigma^2) \propto \sigma^{-2}$

$$p(\mu, \sigma^2 \mid y) \propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right)$$

Calcolo della congiunta a posteriori



$$\mu^{(s)}, \sigma^{(s)} \sim p(\mu, \sigma \mid y)$$

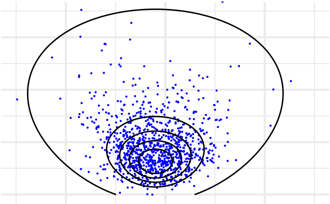
$$\text{con } p(\mu, \sigma^2) \propto \sigma^{-2}$$

$$p(\mu, \sigma^2 \mid y) \propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2\right)$$

$$= \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} \left[\sum_{i=1}^{n} (y_i - \bar{y})^2 + n(\bar{y} - \mu)^2\right]\right)$$

$$\text{dove } \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

Calcolo della congiunta a posteriori



$$\mu^{(s)}, \sigma^{(s)} \sim p(\mu, \sigma \mid y)$$

$$\text{con } p(\mu, \sigma^2) \propto \sigma^{-2}$$

$$p(\mu, \sigma^2 \mid y) \propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2\right)$$

$$= \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} \left[\sum_{i=1}^{n} (y_i - \bar{y})^2 + n(\bar{y} - \mu)^2\right]\right)$$

$$\text{dove } \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$$

$$= \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} \left[(n-1)s^2 + n(\bar{y} - \mu)^2\right]\right)$$

$$\text{dove } s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2$$

Come esprimere in termini delle statistiche sufficienti

$$\sum_{i=1}^{n}(y_i - \mu)^2$$

Come esprimere in termini delle statistiche sufficienti

$$\sum_{i=1}^{n}(y_i - \mu)^2$$

$$\sum_{i=1}^{n}(y_i^2 - 2y_i\mu + \mu^2)$$

Come esprimere in termini delle statistiche sufficienti

$$\sum_{i=1}^{n}(y_i - \mu)^2$$

$$\sum_{i=1}^{n}(y_i^2 - 2y_i\mu + \mu^2)$$

$$\sum_{i=1}^{n}(y_i^2 - 2y_i\mu + \mu^2 - \bar{y}^2 + \bar{y}^2 - 2y_i\bar{y} + 2y_i\bar{y})$$

$$\sum_{i=1}^{n}(y_i - \mu)^2$$

$$\sum_{i=1}^{n}(y_i^2 - 2y_i\mu + \mu^2)$$

$$\sum_{i=1}^{n}(y_i^2 - 2y_i\mu + \mu^2 - \bar{y}^2 + \bar{y}^2 - 2y_i\bar{y} + 2y_i\bar{y})$$

$$\sum_{i=1}^{n}(y_i^2 - 2y_i\bar{y} + \bar{y}^2) + \sum_{i=1}^{n}(\mu^2 - 2y_i\mu - \bar{y}^2 + 2y_i\bar{y})$$

$$\sum_{i=1}^{n}(y_i - \mu)^2$$

$$\sum_{i=1}^{n}(y_i^2 - 2y_i\mu + \mu^2)$$

$$\sum_{i=1}^{n}(y_i^2 - 2y_i\mu + \mu^2 - \bar{y}^2 + \bar{y}^2 - 2y_i\bar{y} + 2y_i\bar{y})$$

$$\sum_{i=1}^{n}(y_i^2 - 2y_i\bar{y} + \bar{y}^2) + \sum_{i=1}^{n}(\mu^2 - 2y_i\mu - \bar{y}^2 + 2y_i\bar{y})$$

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 + n(\mu^2 - 2\bar{y}\mu - \bar{y}^2 + 2\bar{y}\bar{y})$$

$$\sum_{i=1}^{n}(y_i - \mu)^2$$

$$\sum_{i=1}^{n}(y_i^2 - 2y_i\mu + \mu^2)$$

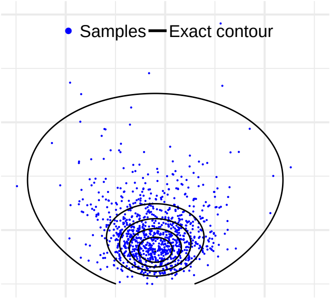
$$\sum_{i=1}^{n}(y_i^2 - 2y_i\mu + \mu^2 - \bar{y}^2 + \bar{y}^2 - 2y_i\bar{y} + 2y_i\bar{y})$$

$$\sum_{i=1}^{n}(y_i^2 - 2y_i\bar{y} + \bar{y}^2) + \sum_{i=1}^{n}(\mu^2 - 2y_i\mu - \bar{y}^2 + 2y_i\bar{y})$$

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 + n(\mu^2 - 2\bar{y}\mu - \bar{y}^2 + 2\bar{y}\bar{y})$$

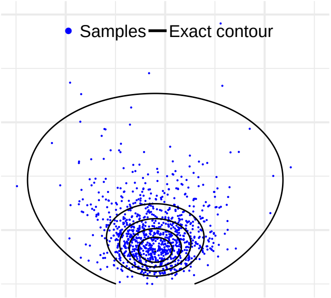
$$\sum_{i=1}^{n}(y_i - \bar{y})^2 + n(\bar{y} - \mu)^2$$

Joint posterior

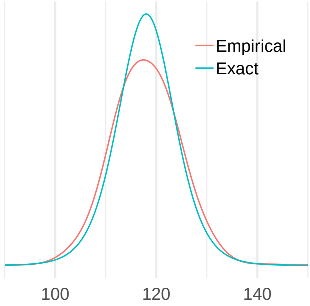


$$\mu^{(s)}, \sigma^{(s)} \sim p(\mu, \sigma \mid y)$$

Joint posterior



Marginal of mu

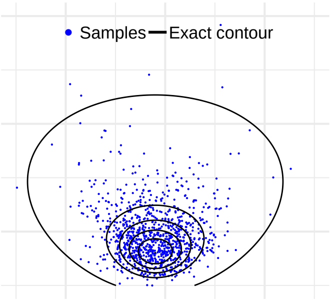


$$\color{blue}{\mu^{(s)}, \sigma^{(s)}} \sim p(\mu, \sigma \mid y)$$

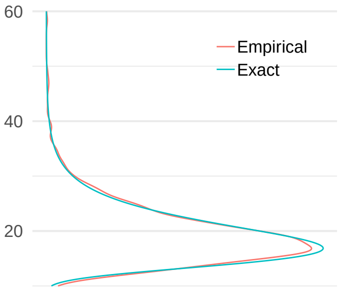
marginali

$$p(\mu \mid y) = \int p(\mu, \sigma \mid y) d\sigma$$

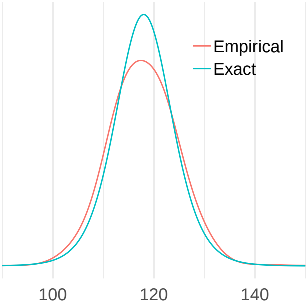
Joint posterior



Marginal of sigma



Marginal of mu



$$\mu^{(s)}, \sigma^{(s)} \sim p(\mu, \sigma \mid y)$$

marginali

$$p(\mu \mid y) = \int p(\mu, \sigma \mid y) d\sigma$$

$$p(\sigma \mid y) = \int p(\mu, \sigma \mid y) d\mu$$

Media nota

$$\sigma^2 \mid y \sim \text{Inv-}\chi^2(n, v)$$
$$\text{ove} \quad v = \frac{1}{n}\sum_{i=1}^{n}(y_i - \theta)^2$$

Media non nota

$$\sigma^2 \mid y \sim \text{Inv-}\chi^2(n-1, s^2)$$
$$\text{ove} \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

Marginale a posteriori $p(\sigma^2 \mid y)$ (piú facile per σ^2 che per σ)

$$p(\sigma^2 \mid y) \quad \propto \quad \int p(\mu, \sigma^2 \mid y) d\mu$$

Marginale a posteriori $p(\sigma^2 \mid y)$ (piú facile per σ^2 che per σ)

$$\begin{aligned} p(\sigma^2 \mid y) &\propto \int p(\mu, \sigma^2 \mid y) d\mu \\ &\propto \int \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2} \left[(n-1)s^2 + n(\bar{y} - \mu)^2 \right] \right) d\mu \end{aligned}$$

$$\begin{aligned}
p(\sigma^2 \mid y) &\propto \int p(\mu, \sigma^2 \mid y) d\mu \\
&\propto \int \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{y}-\mu)^2\right]\right) d\mu \\
&\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}(n-1)s^2\right) \cdot \\
&\quad \int \exp\left(-\frac{n}{2\sigma^2}(\bar{y}-\mu)^2\right) d\mu
\end{aligned}$$

$$
\begin{aligned}
p(\sigma^2 \mid y) &\propto \int p(\mu, \sigma^2 \mid y) d\mu \\
&\propto \int \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{y}-\mu)^2\right]\right) d\mu \\
&\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}(n-1)s^2\right) \cdot \\
&\quad \int \exp\left(-\frac{n}{2\sigma^2}(\bar{y}-\mu)^2\right) d\mu
\end{aligned}
$$

$$\int \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y-\theta)^2\right) d\theta = 1$$

$$
\begin{aligned}
p(\sigma^2 \mid y) &\propto \int p(\mu, \sigma^2 \mid y) d\mu \\
&\propto \int \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{y}-\mu)^2\right]\right) d\mu \\
&\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}(n-1)s^2\right) \cdot \\
&\quad \int \exp\left(-\frac{n}{2\sigma^2}(\bar{y}-\mu)^2\right) d\mu
\end{aligned}
$$

$$
\int \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y-\theta)^2\right) d\theta = 1
$$

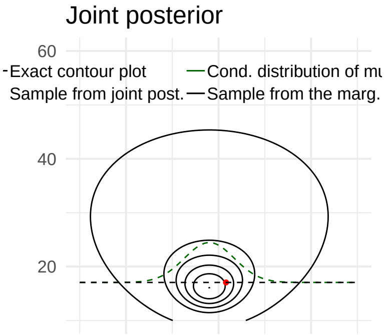
$$
\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}(n-1)s^2\right) \sqrt{2\pi\sigma^2/n}
$$

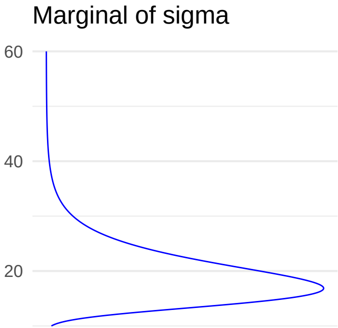
$$
\begin{aligned}
p(\sigma^2 \mid y) &\propto \int p(\mu, \sigma^2 \mid y) d\mu \\
&\propto \int \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{y}-\mu)^2\right]\right) d\mu \\
&\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}(n-1)s^2\right) \cdot \\
&\qquad \int \exp\left(-\frac{n}{2\sigma^2}(\bar{y}-\mu)^2\right) d\mu \\
&\qquad\qquad \int \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y-\theta)^2\right) d\theta = 1 \\
&\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}(n-1)s^2\right) \sqrt{2\pi\sigma^2/n} \\
&\propto (\sigma^2)^{-(n+1)/2} \exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right)
\end{aligned}
$$

$$
\begin{aligned}
p(\sigma^2 \mid y) &\propto \int p(\mu, \sigma^2 \mid y) d\mu \\
&\propto \int \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{y}-\mu)^2\right]\right) d\mu \\
&\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}(n-1)s^2\right) \cdot \\
&\qquad \int \exp\left(-\frac{n}{2\sigma^2}(\bar{y}-\mu)^2\right) d\mu \\
&\qquad\qquad \int \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y-\theta)^2\right) d\theta = 1 \\
&\propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}(n-1)s^2\right) \sqrt{2\pi\sigma^2/n} \\
&\propto (\sigma^2)^{-(n+1)/2} \exp\left(-\frac{(n-1)s^2}{2\sigma^2}\right) \\
p(\sigma^2 \mid y) &= \text{Inv-}\chi^2(\sigma^2 \mid n-1, s^2) = IG((n-1)/2, s^2(n-1)/2)
\end{aligned}
$$

Normale - a priori non-informativa

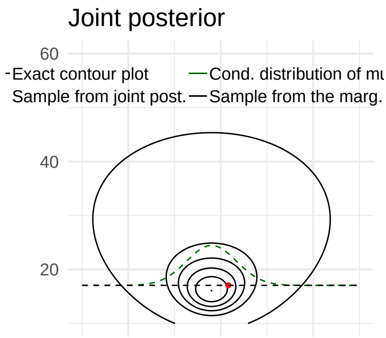
Possiamo quindi facilmente campionare dalla distr cong a post: prima estraiamo σ^2 dalla marg a post, quindi estraiamo μ dalla condiz a post ...

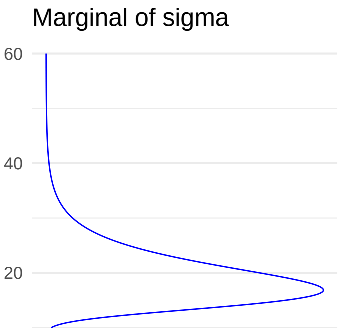




Factorization

$$p(\mu, \sigma^2 \mid y) = {\color{green} p(\mu \mid \sigma^2, y)}{\color{blue} p(\sigma^2 \mid y)}$$



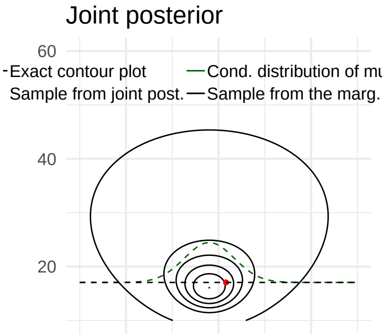


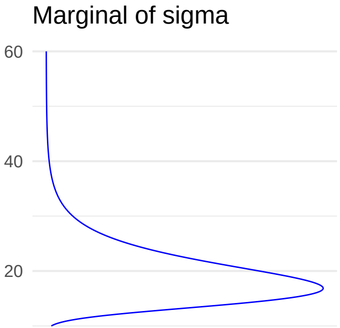
Factorization

$$p(\mu, \sigma^2 \mid y) = p(\mu \mid \sigma^2, y) p(\sigma^2 \mid y)$$

$$p(\sigma^2 \mid y) = \text{Inv-}\chi^2(\sigma^2 \mid n-1, s^2)$$

$$(\sigma^2)^{(s)} \sim p(\sigma^2 \mid y)$$





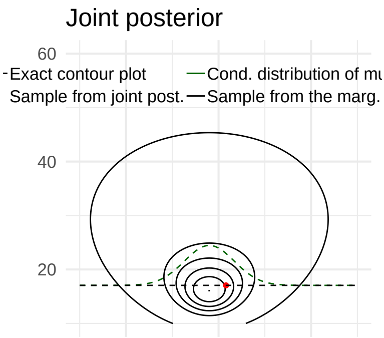
Factorization

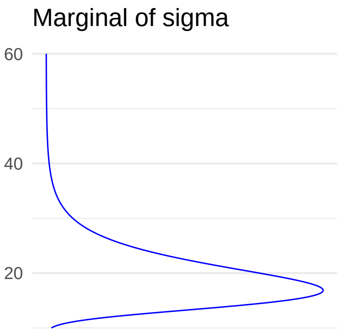
$$p(\mu, \sigma^2 \mid y) = p(\mu \mid \sigma^2, y) p(\sigma^2 \mid y)$$

$$p(\sigma^2 \mid y) = \text{Inv-}\chi^2(\sigma^2 \mid n-1, s^2)$$

$$(\sigma^2)^{(s)} \sim p(\sigma^2 \mid y)$$

$$p(\mu \mid \sigma^2, y) = \mathrm{N}(\mu \mid \bar{y}, \sigma^2/n)$$





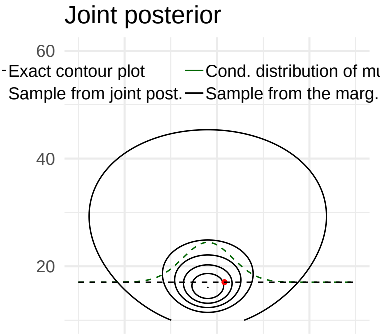
Factorization

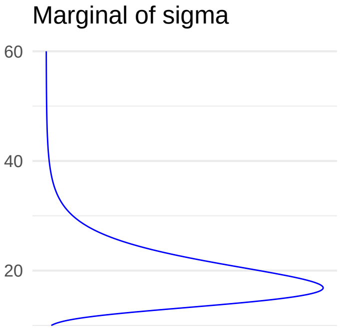
$$p(\mu, \sigma^2 \mid y) = p(\mu \mid \sigma^2, y) p(\sigma^2 \mid y)$$

$$p(\sigma^2 \mid y) = \text{Inv-}\chi^2(\sigma^2 \mid n-1, s^2)$$

$$(\sigma^2)^{(s)} \sim p(\sigma^2 \mid y)$$

$$p(\mu \mid \sigma^2, y) = \mathrm{N}(\mu \mid \bar{y}, \sigma^2/n) \propto \exp\left(-\frac{n}{2\sigma^2}(\bar{y}-\mu)^2\right)$$





Factorization

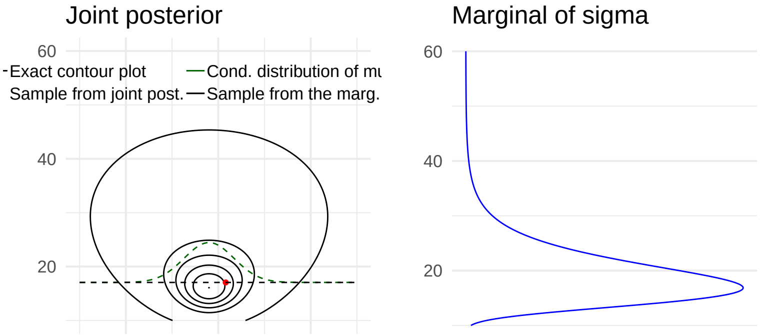
$$p(\mu, \sigma^2 \mid y) = p(\mu \mid \sigma^2, y) p(\sigma^2 \mid y)$$

$$p(\sigma^2 \mid y) = \text{Inv-}\chi^2(\sigma^2 \mid n-1, s^2)$$

$$(\sigma^2)^{(s)} \sim p(\sigma^2 \mid y)$$

$$p(\mu \mid \sigma^2, y) = \mathrm{N}(\mu \mid \bar{y}, \sigma^2/n)$$

$$\mu^{(s)} \sim p(\mu \mid \sigma^2, y)$$



Factorization

$$p(\mu,\sigma^2 \mid y) = p(\mu \mid \sigma^2, y)p(\sigma^2 \mid y)$$

$$p(\sigma^2 \mid y) = \text{Inv-}\chi^2(\sigma^2 \mid n-1, s^2)$$

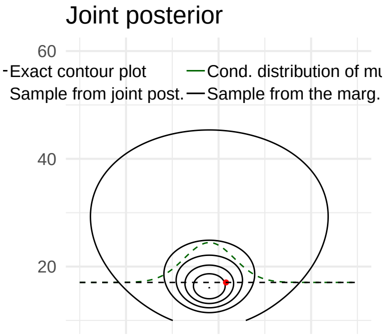
$$(\sigma^2)^{(s)} \sim p(\sigma^2 \mid y)$$

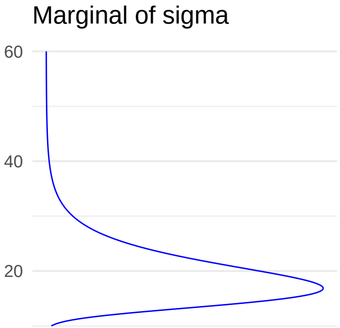
$$p(\mu \mid \sigma^2, y) = \mathrm{N}(\mu \mid \bar{y}, \sigma^2/n)$$

$$\mu^{(s)} \sim p(\mu \mid \sigma^2, y)$$

$$\mu^{(s)}, \sigma^{(s)} \sim p(\mu, \sigma \mid y)$$

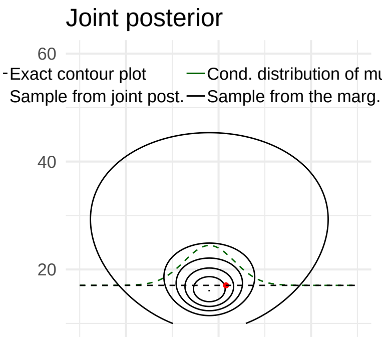
Possiamo anche derivare la marginale a posteriori per μ come media delle condizionali ...

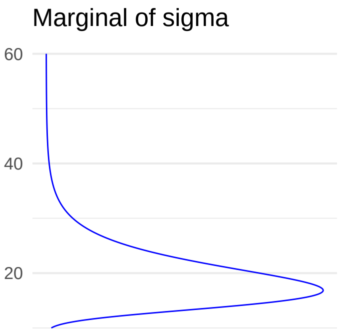




Factorization

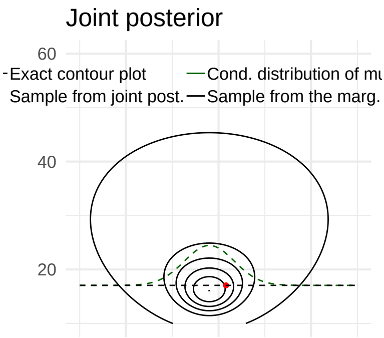
$$p(\mu, \sigma^2 \mid y) = p(\mu \mid \sigma^2, y) p(\sigma^2 \mid y)$$

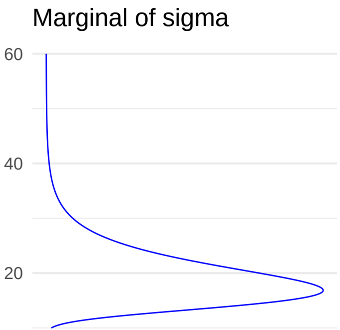




Factorization

$$p(\mu, \sigma^2 \mid y) = p(\mu \mid \sigma^2, y) p(\sigma^2 \mid y)$$
$$(\sigma^2)^{(s)} \sim p(\sigma^2 \mid y)$$





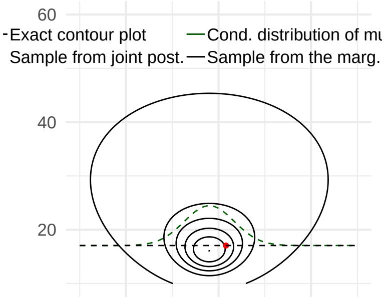
Factorization

$$p(\mu, \sigma^2 \mid y) = p(\mu \mid \sigma^2, y) p(\sigma^2 \mid y)$$

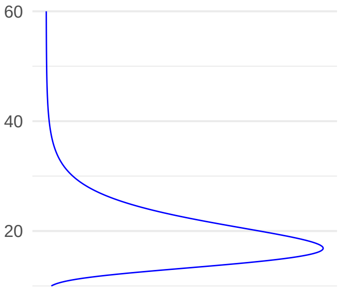
$$(\sigma^2)^{(s)} \sim p(\sigma^2 \mid y)$$

$$p(\mu \mid (\sigma^2)^{(s)}, y) = \mathrm{N}(\mu \mid \bar{y}, (\sigma^2)^{(s)}/n)$$

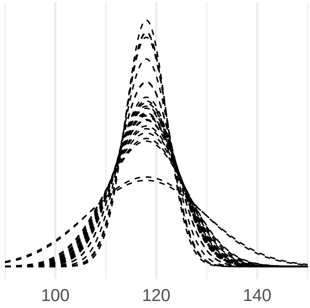
Joint posterior



Marginal of sigma



Cond distr of mu for 25 draws



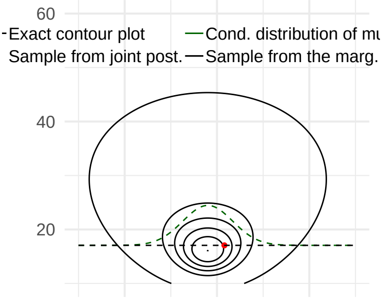
Factorization

$$p(\mu, \sigma^2 \mid y) = p(\mu \mid \sigma^2, y) p(\sigma^2 \mid y)$$

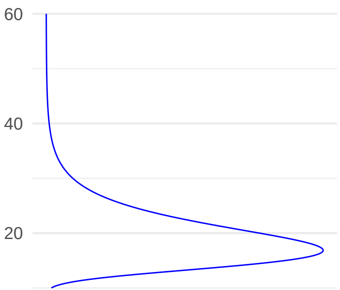
$$(\sigma^2)^{(s)} \sim p(\sigma^2 \mid y)$$

$$p(\mu \mid (\sigma^2)^{(s)}, y) = \mathrm{N}(\mu \mid \bar{y}, (\sigma^2)^{(s)}/n)$$

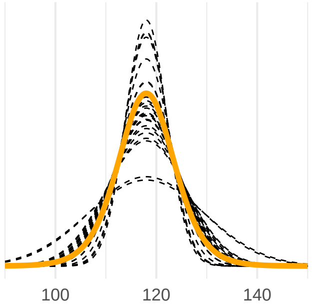
Joint posterior



Marginal of sigma



Cond distr of mu for 25 draws



Factorization

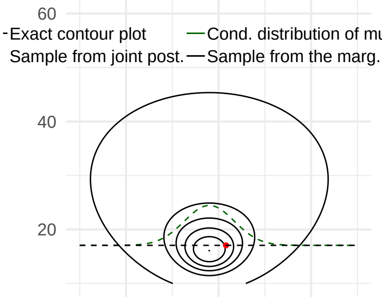
$$p(\mu, \sigma^2 \mid y) = p(\mu \mid \sigma^2, y) p(\sigma^2 \mid y)$$

$$(\sigma^2)^{(s)} \sim p(\sigma^2 \mid y)$$

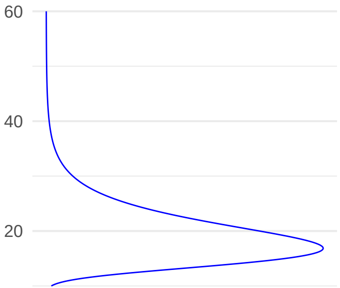
$$p(\mu \mid (\sigma^2)^{(s)}, y) = \mathrm{N}(\mu \mid \bar{y}, (\sigma^2)^{(s)}/n)$$

$$p(\mu \mid y) \approx \frac{1}{S} \sum_{s=1}^{S} \mathrm{N}(\mu \mid \bar{y}, (\sigma^2)^{(s)}/n)$$

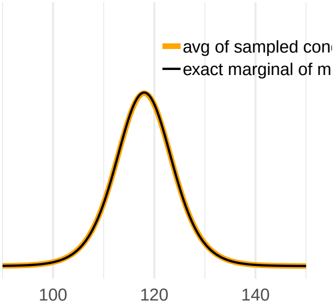
Joint posterior



Marginal of sigma



Cond. distr of mu



Factorization

$$p(\mu, \sigma^2 \mid y) = p(\mu \mid \sigma^2, y) p(\sigma^2 \mid y)$$

$$(\sigma^2)^{(s)} \sim p(\sigma^2 \mid y)$$

$$p(\mu \mid (\sigma^2)^{(s)}, y) = \mathrm{N}(\mu \mid \bar{y}, (\sigma^2)^{(s)}/n)$$

$$p(\mu \mid y) \approx \frac{1}{S} \sum_{s=1}^{S} \mathrm{N}(\mu \mid \bar{y}, (\sigma^2)^{(s)}/n)$$

Marginale a posteriori $p(\mu \mid y)$

$$p(\mu \mid y) = \int_0^\infty p(\mu, \sigma^2 \mid y) d\sigma^2$$

Marginale a posteriori $p(\mu \mid y)$

$$
\begin{aligned}
p(\mu \mid y) &= \int_0^\infty p(\mu, \sigma^2 \mid y) d\sigma^2 \\
&\propto \int_0^\infty \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{y}-\mu)^2\right]\right) d\sigma^2
\end{aligned}
$$

Marginale a posteriori $p(\mu \mid y)$

$$p(\mu \mid y) = \int_0^\infty p(\mu, \sigma^2 \mid y) d\sigma^2$$
$$\propto \int_0^\infty \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{y}-\mu)^2\right]\right) d\sigma^2$$

Per sostituzione

$$A = (n-1)s^2 + n(\mu - \bar{y})^2$$

Marginale a posteriori $p(\mu \mid y)$

$$p(\mu \mid y) = \int_0^\infty p(\mu, \sigma^2 \mid y) d\sigma^2$$
$$\propto \int_0^\infty \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{y}-\mu)^2\right]\right) d\sigma^2$$

Per sostituzione

$$A = (n-1)s^2 + n(\mu - \bar{y})^2 \quad \text{e} \quad z = \frac{A}{2\sigma^2}$$

Marginale a posteriori $p(\mu \mid y)$

$$p(\mu \mid y) = \int_0^\infty p(\mu, \sigma^2 \mid y) d\sigma^2$$
$$\propto \int_0^\infty \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{y}-\mu)^2\right]\right) d\sigma^2$$

Per sostituzione

$$A = (n-1)s^2 + n(\mu - \bar{y})^2 \quad \text{e} \quad z = \frac{A}{2\sigma^2}$$

$$p(\mu \mid y) \propto A^{-n/2} \int_0^\infty z^{(n-2)/2} \exp(-z) dz$$

Marginale a posteriori $p(\mu \mid y)$

$$p(\mu \mid y) = \int_0^\infty p(\mu, \sigma^2 \mid y) d\sigma^2$$
$$\propto \int_0^\infty \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{y}-\mu)^2\right]\right) d\sigma^2$$

Per sostituzione

$$A = (n-1)s^2 + n(\mu - \bar{y})^2 \quad \text{e} \quad z = \frac{A}{2\sigma^2}$$

$$p(\mu \mid y) \propto A^{-n/2} \int_0^\infty z^{(n-2)/2} \exp(-z) dz$$

E riconoscendo un integrale gamma non-normalizzato.
$\Gamma(u) = \int_0^\infty x^{u-1} \exp(-x) dx$

Marginale a posteriori $p(\mu \mid y)$

$$p(\mu \mid y) = \int_0^\infty p(\mu, \sigma^2 \mid y) d\sigma^2$$

$$\propto \int_0^\infty \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{y}-\mu)^2\right]\right) d\sigma^2$$

Per sostituzione

$$A = (n-1)s^2 + n(\mu - \bar{y})^2 \quad \text{e} \quad z = \frac{A}{2\sigma^2}$$

$$p(\mu \mid y) \propto A^{-n/2} \int_0^\infty z^{(n-2)/2} \exp(-z) dz$$

E riconoscendo un integrale gamma non-normalizzato.
$\Gamma(u) = \int_0^\infty x^{u-1} \exp(-x) dx$

$$\propto [(n-1)s^2 + n(\mu - \bar{y})^2]^{-n/2}$$

Marginale a posteriori $p(\mu \mid y)$

$$p(\mu \mid y) = \int_0^\infty p(\mu, \sigma^2 \mid y) d\sigma^2$$

$$\propto \int_0^\infty \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{y}-\mu)^2\right]\right) d\sigma^2$$

Per sostituzione

$$A = (n-1)s^2 + n(\mu - \bar{y})^2 \quad \text{e} \quad z = \frac{A}{2\sigma^2}$$

$$p(\mu \mid y) \propto A^{-n/2} \int_0^\infty z^{(n-2)/2} \exp(-z) dz$$

E riconoscendo un integrale gamma non-normalizzato.
$\Gamma(u) = \int_0^\infty x^{u-1} \exp(-x) dx$

$$\propto [(n-1)s^2 + n(\mu - \bar{y})^2]^{-n/2}$$

$$\propto \left[1 + \frac{n(\mu - \bar{y})^2}{(n-1)s^2}\right]^{-n/2}$$

Marginale a posteriori $p(\mu \mid y)$

$$p(\mu \mid y) = \int_0^\infty p(\mu, \sigma^2 \mid y) d\sigma^2$$
$$\propto \int_0^\infty \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{y}-\mu)^2\right]\right) d\sigma^2$$

Per sostituzione

$$A = (n-1)s^2 + n(\mu - \bar{y})^2 \quad \text{e} \quad z = \frac{A}{2\sigma^2}$$

$$p(\mu \mid y) \propto A^{-n/2} \int_0^\infty z^{(n-2)/2} \exp(-z) dz$$

E riconoscendo un integrale gamma non-normalizzato.
$\Gamma(u) = \int_0^\infty x^{u-1} \exp(-x) dx$

$$\propto [(n-1)s^2 + n(\mu - \bar{y})^2]^{-n/2}$$
$$\propto \left[1 + \frac{n(\mu - \bar{y})^2}{(n-1)s^2}\right]^{-n/2}$$

$$p(\mu \mid y) = t_{n-1}(\mu \mid \bar{y}, s^2/n) \quad \text{Student's } t$$

Marginale a posteriori $p(\mu \mid y)$

$$p(\mu \mid y) = t_{n-1}(\mu \mid \bar{y}, s^2/n)$$

Notiamo che:

1. $\frac{\mu-\bar{y}}{s/\sqrt{n}}|y \ \sim \ t_{n-1}$
2. $\frac{\bar{y}-\mu}{s/\sqrt{n}}|\mu, \sigma^2 \ \sim \ t_{n-1}$

(1) la distr a post della quantitá pivotale non dipende dai dati y, e (2) la distr campionaria della quantitá pivotale non dipende dai parametri μ e σ^2

In generale, una quantitá pivotale per il parametro da stimare si definisce come una funzione non banale dei dati e del parametro da stimare la cui distr campionaria é indipendente da tutti i parametri e dai dati.

Normale - a priori non-informativa

Teniamo a mente che la

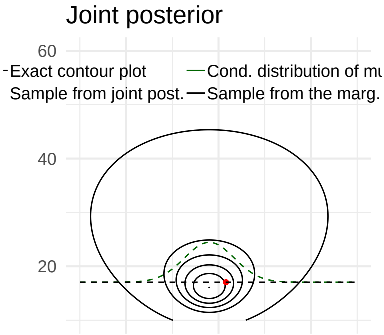
- marginale a posteriori $p(\mu \mid y)$

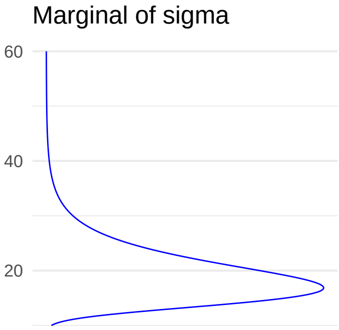
$$p(\mu \mid y) = \int_0^{\infty} p(\mu \mid \sigma^2, y) p(\sigma^2 \mid y) d\sigma^2$$

- é una mistura di distribuzioni normali dove la densitá mixante é la marginale a posteriori di σ^2

per derivare la distr pred a post analiticamente.

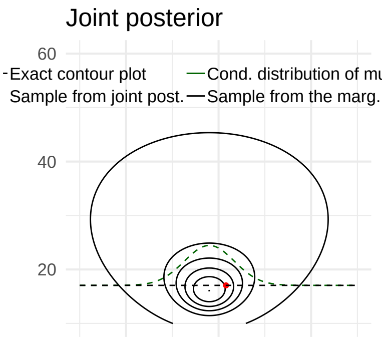
Prima la otteniamo via simulazione, poi analiticamente. $\rightarrow$

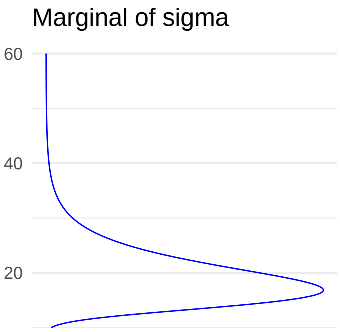




Distribuzione predittiva per un nuovo $\tilde{y}$

$$p(\tilde{y} \mid y) = \int p(\tilde{y} \mid \mu, \sigma) p(\mu, \sigma \mid y) d\mu\sigma$$

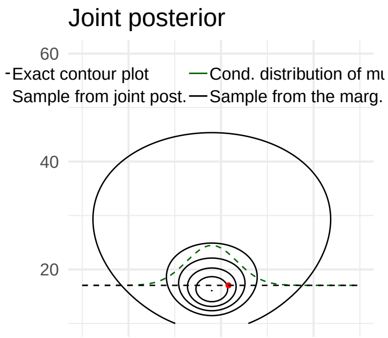


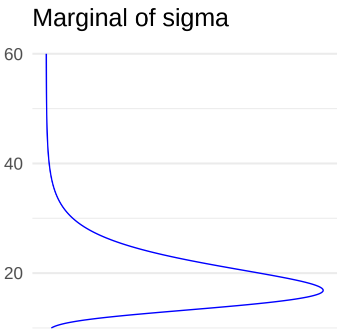


Distribuzione predittiva per un nuovo $\tilde{y}$

$$p(\tilde{y} \mid y) = \int p(\tilde{y} \mid \mu, \sigma) p(\mu, \sigma \mid y) d\mu\sigma$$

$$\mu^{(s)}, \sigma^{(s)} \sim p(\mu, \sigma \mid y)$$



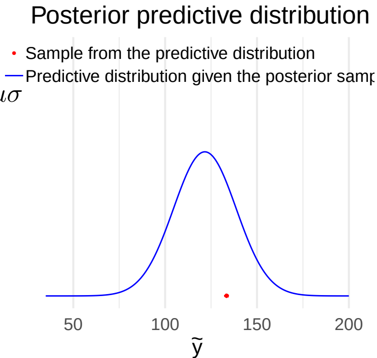


Distribuzione predittiva per un nuovo $\tilde{y}$

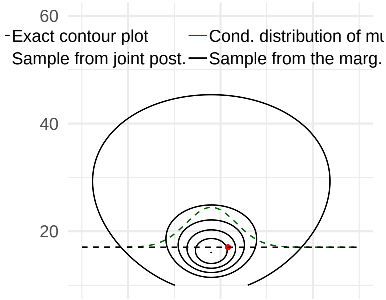
$$p(\tilde{y} \mid y) = \int p(\tilde{y} \mid \mu, \sigma) p(\mu, \sigma \mid y) d\mu\sigma$$

$$\mu^{(s)}, \sigma^{(s)} \sim p(\mu, \sigma \mid y)$$

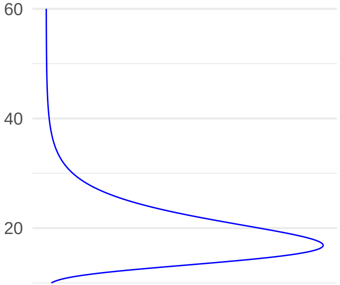
$$\tilde{y}^{(s)} \sim p(\tilde{y} \mid \mu^{(s)}, \sigma^{(s)})$$



Joint posterior



Marginal of sigma



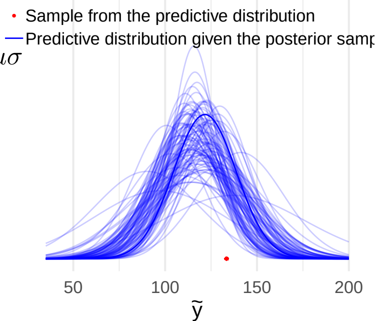
Distribuzione predittiva per un nuovo $\tilde{y}$

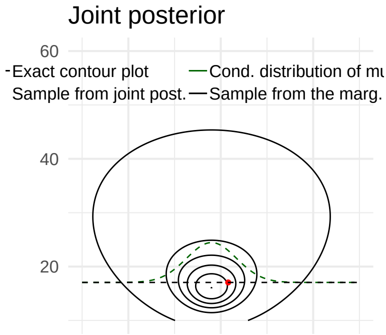
$$p(\tilde{y} \mid y) = \int p(\tilde{y} \mid \mu, \sigma) p(\mu, \sigma \mid y) d\mu\sigma$$

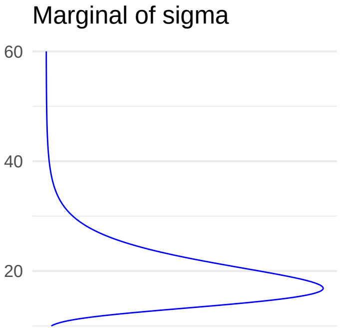
$$\mu^{(s)}, \sigma^{(s)} \sim p(\mu, \sigma \mid y)$$

$$\tilde{y}^{(s)} \sim p(\tilde{y} \mid \mu^{(s)}, \sigma^{(s)})$$

Posterior predictive distribution





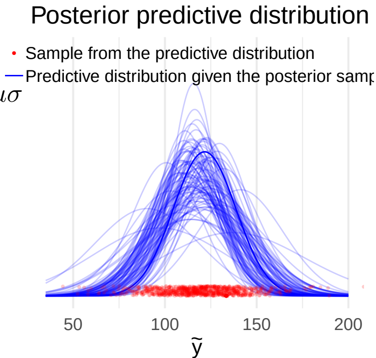


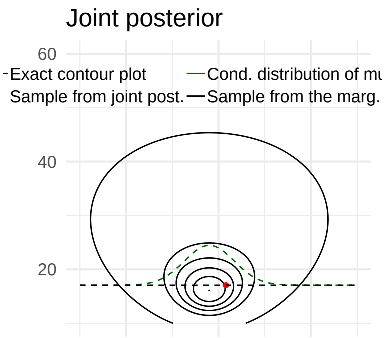
Distribuzione predittiva per un nuovo $\tilde{y}$

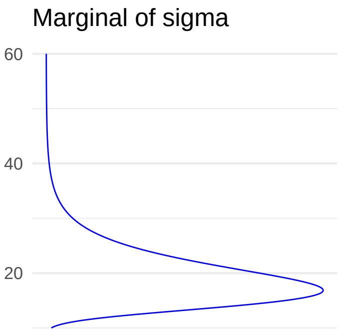
$$p(\tilde{y} \mid y) = \int p(\tilde{y} \mid \mu, \sigma) p(\mu, \sigma \mid y) d\mu\sigma$$

$$\mu^{(s)}, \sigma^{(s)} \sim p(\mu, \sigma \mid y)$$

$$\tilde{y}^{(s)} \sim p(\tilde{y} \mid \mu^{(s)}, \sigma^{(s)})$$





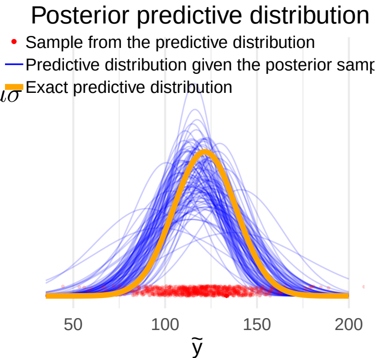


Distribuzione predittiva per un nuovo $\tilde{y}$

$$p(\tilde{y} \mid y) = \int p(\tilde{y} \mid \mu, \sigma) p(\mu, \sigma \mid y) d\mu\sigma$$

$$\mu^{(s)}, \sigma^{(s)} \sim p(\mu, \sigma \mid y)$$

$$\tilde{y}^{(s)} \sim p(\tilde{y} \mid \mu^{(s)}, \sigma^{(s)})$$



Normale - distribuzione predittiva a posteriori

Distribuzione predittiva a posteriori quando la varianza é nota

$$p(\tilde{y} \mid \sigma^2, y) = \int p(\tilde{y} \mid \mu, \sigma^2) p(\mu \mid \sigma^2, y) d\mu$$

Normale - distribuzione predittiva a posteriori

Distribuzione predittiva a posteriori quando la varianza é nota

$$
\begin{aligned}
p(\tilde{y} \mid \sigma^2, y) &= \int p(\tilde{y} \mid \mu, \sigma^2) p(\mu \mid \sigma^2, y) d\mu \\
&= \int \mathrm{N}(\tilde{y} \mid \mu, \sigma^2)\, \mathrm{N}(\mu \mid \bar{y}, \sigma^2/n) d\mu
\end{aligned}
$$

Normale - distribuzione predittiva a posteriori

Distribuzione predittiva a posteriori quando la varianza é nota

$$
\begin{aligned}
p(\tilde{y} \mid \sigma^2, y) &= \int p(\tilde{y} \mid \mu, \sigma^2) p(\mu \mid \sigma^2, y) d\mu \\
&= \int \mathrm{N}(\tilde{y} \mid \mu, \sigma^2)\, \mathrm{N}(\mu \mid \bar{y}, \sigma^2/n) d\mu \\
&= \mathrm{N}(\tilde{y} \mid \bar{y}, (1 + \tfrac{1}{n})\sigma^2)
\end{aligned}
$$

Normale - distribuzione predittiva a posteriori

Distribuzione predittiva a posteriori quando la varianza é nota

$$
\begin{aligned}
p(\tilde{y} \mid \sigma^2, y) &= \int p(\tilde{y} \mid \mu, \sigma^2) p(\mu \mid \sigma^2, y) d\mu \\
&= \int \mathrm{N}(\tilde{y} \mid \mu, \sigma^2)\, \mathrm{N}(\mu \mid \bar{y}, \sigma^2/n) d\mu \\
&= \mathrm{N}(\tilde{y} \mid \bar{y}, (1 + \tfrac{1}{n})\sigma^2)
\end{aligned}
$$

se come per $p(\mu \mid \sigma^2, y)$ consideriamo la mistura con distribuzione mixante la marginale a posteriori per σ^2

Normale - distribuzione predittiva a posteriori

Distribuzione predittiva a posteriori quando la varianza é nota

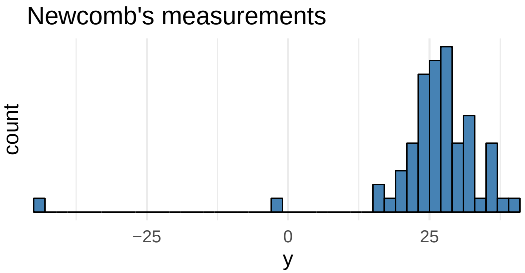
$$
\begin{aligned}
p(\tilde{y} \mid \sigma^2, y) &= \int p(\tilde{y} \mid \mu, \sigma^2) p(\mu \mid \sigma^2, y) d\mu \\
&= \int \mathrm{N}(\tilde{y} \mid \mu, \sigma^2)\, \mathrm{N}(\mu \mid \bar{y}, \sigma^2/n) d\mu \\
&= \mathrm{N}(\tilde{y} \mid \bar{y}, (1 + \tfrac{1}{n})\sigma^2)
\end{aligned}
$$

se come per $p(\mu \mid \sigma^2, y)$ consideriamo la mistura con distribuzione mixante la marginale a posteriori per σ^2

$$
p(\tilde{y} \mid y) = t_{n-1}(\tilde{y} \mid \bar{y}, (1 + \tfrac{1}{n})s^2)
$$

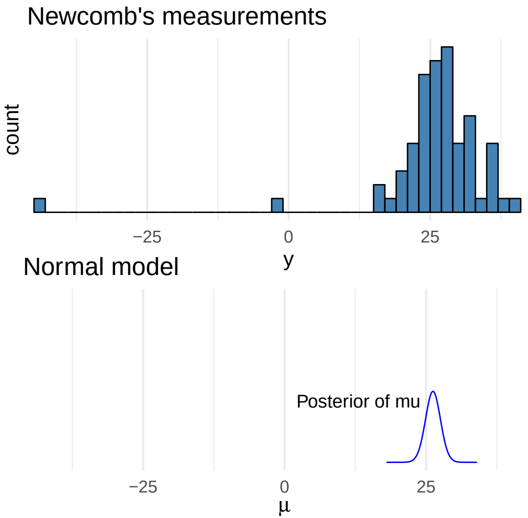
Esperimento di Simon Newcomb (1882): velocitá della luce

Newcomb misuró ($n = 66$) il tempo necessario alla luce per viaggiare dal suo laboratorio sul fiume Potomac a uno specchio alla base del Washington Monument e ritorno, una distanza totale di 7422 metri.



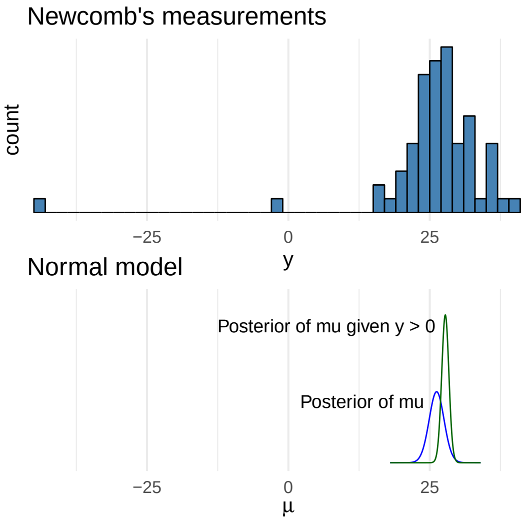
Esperimento di Simon Newcomb (1882): velocitá della luce

Newcomb misuró ($n = 66$) il tempo necessario alla luce per viaggiare dal suo laboratorio sul fiume Potomac a uno specchio alla base del Washington Monument e ritorno, una distanza totale di 7422 metri.



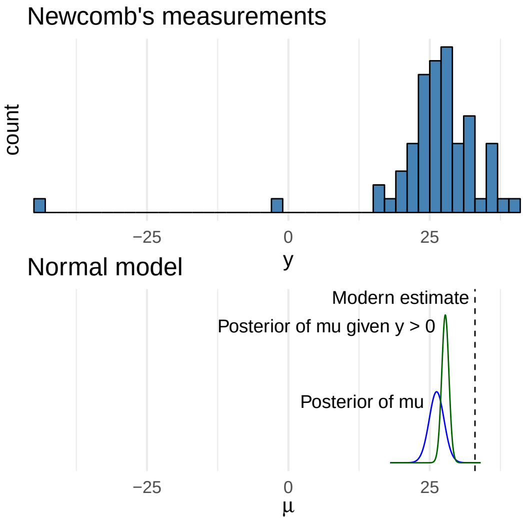
Esperimento di Simon Newcomb (1882): velocitá della luce

Newcomb misuró ($n = 66$) il tempo necessario alla luce per viaggiare dal suo laboratorio sul fiume Potomac a uno specchio alla base del Washington Monument e ritorno, una distanza totale di 7422 metri.



Esperimento di Simon Newcomb (1882): velocitá della luce

Newcomb misuró ($n = 66$) il tempo necessario alla luce per viaggiare dal suo laboratorio sul fiume Potomac a uno specchio alla base del Washington Monument e ritorno, una distanza totale di 7422 metri.



Normale - a priori coniugata

- L'a priori conjugata deve avere la forma $p(\sigma^2)p(\mu \mid \sigma^2)$ (si veda la forma della verosimiglianza nella sez precedente)

Normale - a priori coniugata

- L'a priori conjugata deve avere la forma $p(\sigma^2)p(\mu \mid \sigma^2)$ (si veda la forma della verosimiglianza nella sez precedente)
- Una conveniente parametrizzazione é

$$\mu \mid \sigma^2 \sim \mathrm{N}(\mu_0, \sigma^2/\kappa_0)$$
$$\sigma^2 \sim \text{Inv-}\chi^2(\nu_0, \sigma_0^2)$$

che puó essere scritta come

$$p(\mu, \sigma^2) = \text{N-Inv-}\chi^2(\mu_0, \sigma_0^2/\kappa_0; \nu_0, \sigma_0^2)$$

Normale - a priori coniugata

- L'a priori conjugata deve avere la forma $p(\sigma^2)p(\mu \mid \sigma^2)$ (si veda la forma della verosimiglianza nella sez precedente)
- Una conveniente parametrizzazione é

$$\begin{aligned} \mu \mid \sigma^2 &\sim \mathrm{N}(\mu_0, \sigma^2/\kappa_0) \\ \sigma^2 &\sim \text{Inv-}\chi^2(\nu_0, \sigma_0^2) \end{aligned}$$

  che puó essere scritta come

$$p(\mu, \sigma^2) = \text{N-Inv-}\chi^2(\mu_0, \sigma_0^2/\kappa_0; \nu_0, \sigma_0^2)$$

- μ e σ^2 sono dipendenti a priori
  - se σ^2 is grande, allora μ ha una a priori estesa

Normale - a priori coniugata

- L'a priori conjugata deve avere la forma $p(\sigma^2)p(\mu \mid \sigma^2)$ (si veda la forma della verosimiglianza nella sez precedente)
- Una conveniente parametrizzazione é

$$\begin{aligned} \mu \mid \sigma^2 &\sim \mathrm{N}(\mu_0, \sigma^2/\kappa_0) \\ \sigma^2 &\sim \text{Inv-}\chi^2(\nu_0, \sigma_0^2) \end{aligned}$$

che puó essere scritta come

$$p(\mu, \sigma^2) = \text{N-Inv-}\chi^2(\mu_0, \sigma_0^2/\kappa_0; \nu_0, \sigma_0^2)$$

- μ e σ^2 sono dipendenti a priori
    - se σ^2 is grande, allora μ ha una a priori estesa
- Marginalmente, $p(\mu) = t_{\nu_0}(\mu_0, \sigma_0^2/\kappa_0)$

Normale - a priori coniugata

Congiunta a posteriori

$$p(\mu, \sigma^2 \mid y) = \text{N-Inv-}\chi^2(\mu_n, \sigma_n^2/\kappa_n; \nu_n, \sigma_n^2)$$

dove

$$\begin{aligned}
\mu_n &= \frac{\kappa_0}{\kappa_0 + n}\mu_0 + \frac{n}{\kappa_0 + n}\bar{y} \\
\kappa_n &= \kappa_0 + n \\
\nu_n &= \nu_0 + n \\
\nu_n\sigma_n^2 &= \nu_0\sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_0 + n}(\bar{y} - \mu_0)^2
\end{aligned}$$

Normale - a priori coniugata

Congiunta a posteriori

$$p(\mu, \sigma^2 \mid y) = \text{N-Inv-}\chi^2(\mu_n, \sigma_n^2/\kappa_n; \nu_n, \sigma_n^2)$$

dove

$$\begin{aligned}
\mu_n &= \frac{\kappa_0}{\kappa_0 + n}\mu_0 + \frac{n}{\kappa_0 + n}\bar{y} \\
\kappa_n &= \kappa_0 + n \\
\nu_n &= \nu_0 + n \\
\nu_n\sigma_n^2 &= \nu_0\sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_0 + n}(\bar{y} - \mu_0)^2
\end{aligned}$$

Si noti che

$$E(\sigma^2|y) = \frac{\nu_n\sigma_n^2}{\nu_n - 2} = \frac{\nu_0\sigma_0^2 + (n-1)s^2 + \frac{\kappa_0 n}{\kappa_0+n}(\bar{y} - \mu_0)^2}{\nu_0 + n - 2}$$

Normale - a priori coniugata

- Condizionale $p(\mu \mid \sigma^2, y)$

$$
\begin{aligned}
\mu \mid \sigma^2, y &\sim \mathrm{N}(\mu_n, \sigma^2/\kappa_n) \\
&= \mathrm{N}\left(\frac{\frac{\kappa_0}{\sigma^2}\mu_0 + \frac{n}{\sigma^2}\bar{y}}{\frac{\kappa_0}{\sigma^2} + \frac{n}{\sigma^2}}, \frac{1}{\frac{\kappa_0}{\sigma^2} + \frac{n}{\sigma^2}} \right) \\
&= \mathrm{N}\left(\frac{\kappa_0\mu_0 + n\bar{y}}{\kappa_0 + n}, \frac{\sigma^2}{\kappa_0 + n} \right)
\end{aligned}
$$

- Marginale $p(\sigma^2 \mid y)$

$$\sigma^2 \mid y \sim \text{Inv-}\chi^2(\nu_n, \sigma_n^2)$$

- Marginale $p(\mu \mid y)$

$$\mu \mid y \sim t_{\nu_n}(\mu \mid \mu_n, \sigma_n^2/\kappa_n)$$

Normale - a priori semiconiugata

- μ e σ^2 sono indipendenti a priori

$$\begin{aligned} \mu \mid \sigma^2 &\sim \mathrm{N}(\mu_0, \tau_0^2) \\ \sigma^2 &\sim \text{Inv-}\chi^2(\nu_0, \sigma_0^2) \end{aligned}$$

- a priori noninformativa per σ^2 se $\nu_0 = 0$ (che corrisponde a $p(\sigma^2) \propto \frac{1}{\sigma^2}$)
- Condizionale $p(\mu \mid \sigma^2, y)$

$$\mu \mid \sigma^2, y \sim \mathrm{N}(\mu_n, \tau_n^2)$$

ove

$$\mu_n = \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{n}{\sigma^2}\bar{y}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}} \quad \text{e} \quad \tau_n^2 = \frac{1}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}$$

- Marginale $p(\sigma^2 \mid y)$ non in forma chiusa

- BDA3 p. 69

  - Estensione di binomiale
  - y_j numero di osservazioni della categoria j-esima
  - $y = (y_1, \dots, y_k)$ vettore dei conteggi del numero di osservazioni per ogni categoria, con $\sum y_j = n$
  - Modello di osservazione

$$p(y \mid \theta) \propto \prod_{j=1}^{k} \theta_j^{y_j},$$

    con $\sum \theta_j = 1$
  - (Come per la binomiale, si assume che la distribuzione sia condizionata al numero n delle osservazioni.)
  - A priori coniugata Ã¨ la Dirichlet

$$p(\theta \mid \alpha) \propto \prod_{j=1}^{k} \theta_j^{\alpha_j - 1} \propto \mathcal{D}(\alpha)$$

- La Dirichlet, $\mathcal{D}(\alpha)$, é un'estensione multivariata della Beta.
- I θ_j sono nonnegativi e con somma 1. More formally, the support of a k-dimensional Dirichlet is the $(k-1$-)simplex of R^k
$$\left\{\theta \in \mathbb{R}^k : \theta_j > 0,\ \sum_{j=1}^{k} \theta_j = 1\right\}$$
- $\mathcal{D}(\alpha)$ é definita per $\alpha_j > 0$ e se $\alpha_0 = \sum_j \alpha_j$, $E(\theta_j) = \alpha_j/\alpha_0$
- L'a priori Dirichlet contiene un informazione equivalente a $\sum_j(\alpha_j - 1)$ osservazioni con $\alpha_j - 1$ osservazioni della categoria j-esima
- $\alpha_0 = \sum_j \alpha_j$ determines the *concentration* of a Dirichlet, that is, how much the distribution is dense (high α_0) or sparse (low α_0)

- A posteriori

$$p(\theta \mid \alpha, y) \propto \prod_{j=1}^{k} \theta_j^{y_j+\alpha_j-1} \propto \mathcal{D}(y+\alpha)$$

- Si noti che le medie a posteriori delle probabilitá multinomiali sono

$$E(\theta_j|y) = \frac{y_j+\alpha_j}{n+\alpha_0} = \frac{y_j}{n} \cdot \frac{n}{n+\alpha_0} + \frac{\alpha_j}{\alpha_0} \cdot \frac{\alpha_j}{n+\alpha_0}$$

  che esprime ...

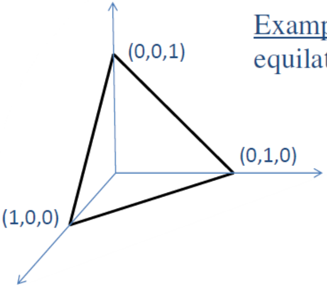
- Ci sono varie noninformative plausibili:
  - a priori uniforme: $\alpha_j = 1 \forall j$
  - a priori impropria: $\alpha_j = 0 \forall j$, i.e., uniforme su $\log(\theta_j)$
    - a posteriori propria se c'é almeno un'osservazione in ogni categoria cosí che ogni componente di y sia positiva.

Example of simplex

The support Ω_k is the space of all the probability vectors of k probabilities that sum to 1.

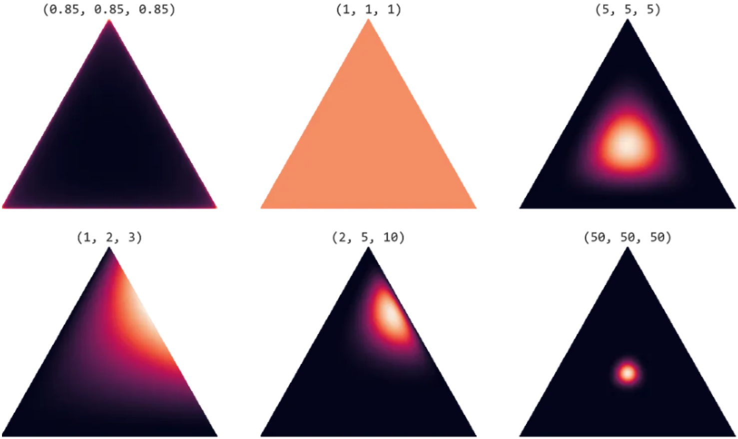
Because of the constraint on the components of a probability vector, Ω_k is $k-1$ dimensional and is the **standard or probability $K-1$-simplex.**

For $k=3$ the support is the triangle with vertices $(1,0,0)$, $(0,1,0)$, $(0,0,1)$



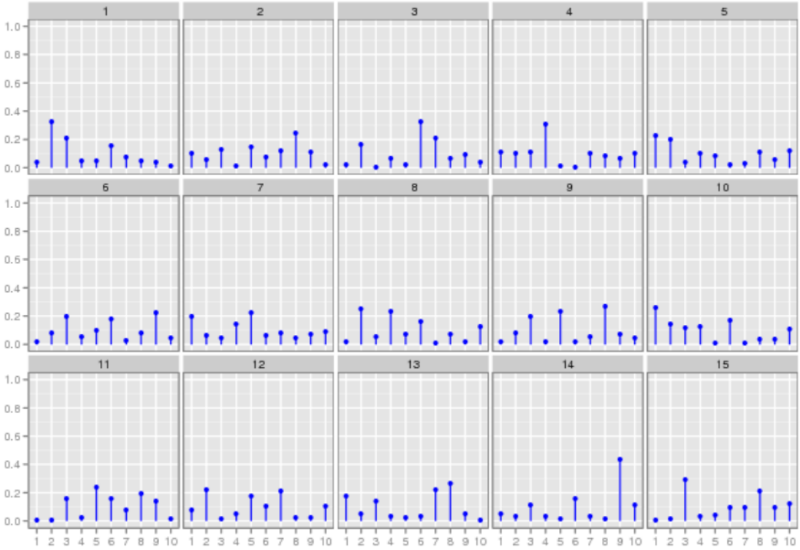
Example: Ω_3 is a 2-dimensional equilateral triangle.

3-dimensional Dirichlet distribution on a 2-simplex for different values of α.

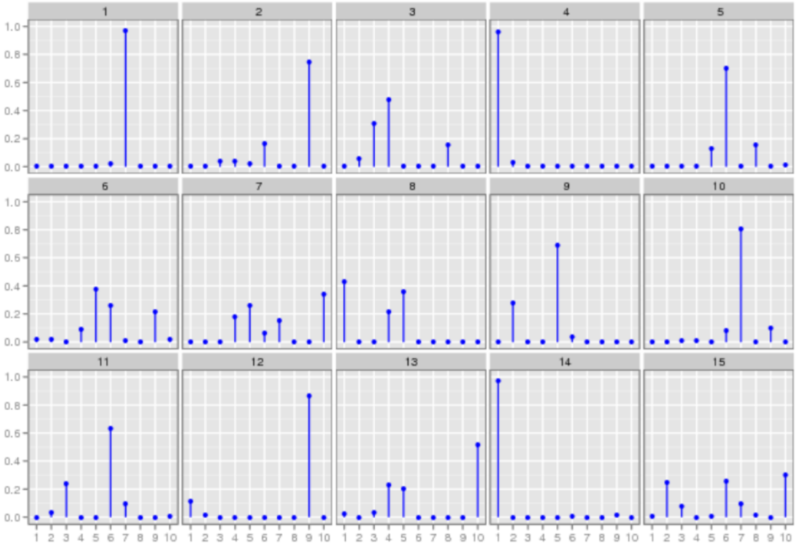


Properties: Symmetric, flat Dirichlet; concentration par. $\sum_k \alpha_k$.

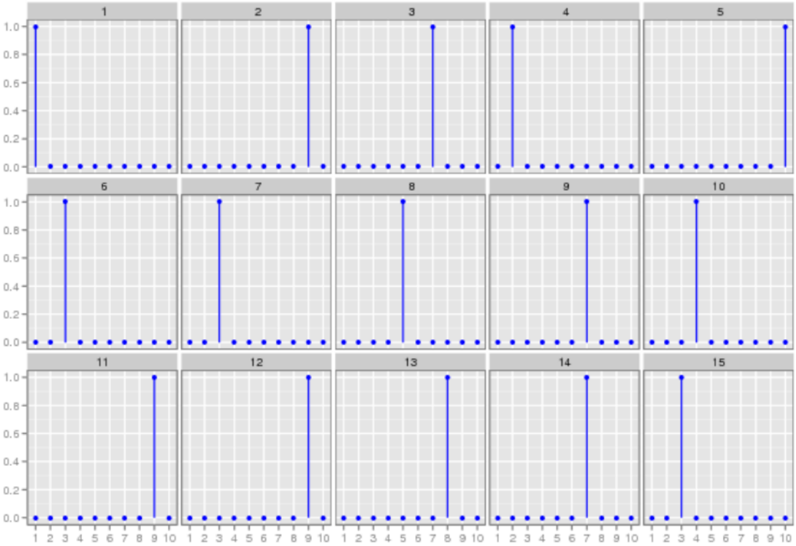
$\alpha = 1/0.1/0.001/10/100$



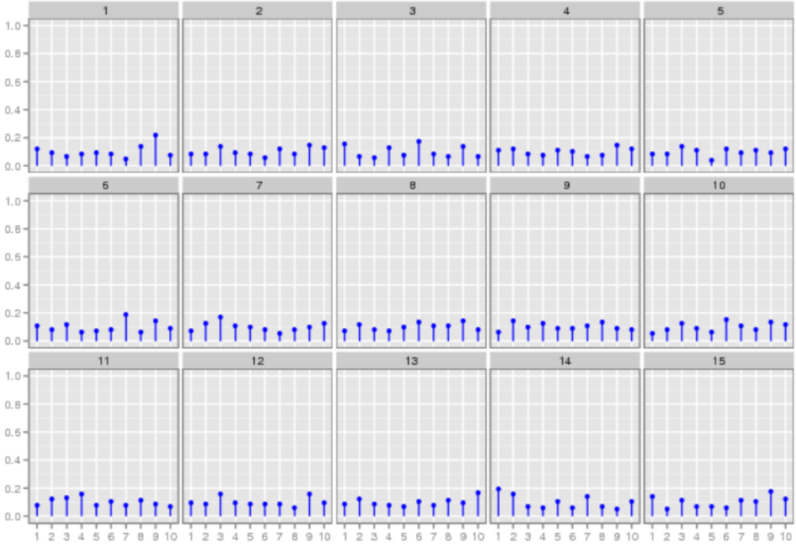
$\alpha = 1/0.1/0.001/10/100$



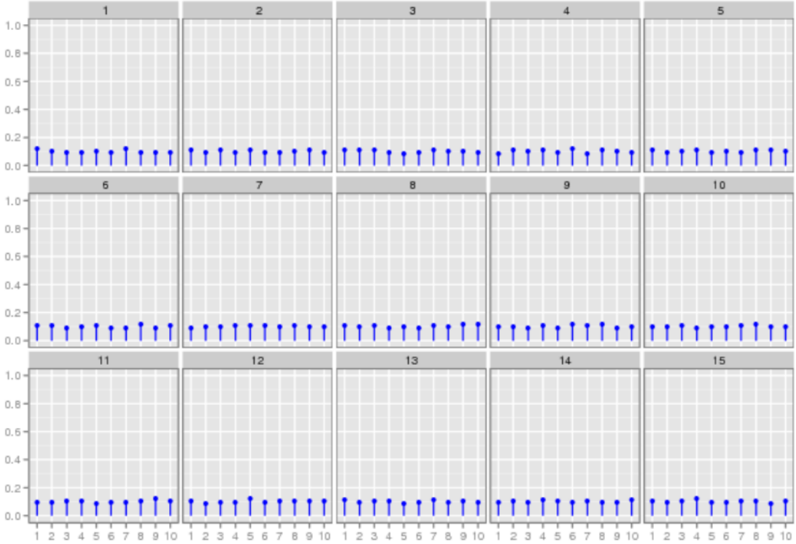
$\alpha = 1/0.1/0.001/10/100$



$\alpha = 1/0.1/0.001/10/100$



$\alpha = 1/0.1/0.001/10/100$



Normale multivariata

- y é un vettore di d componenti
- $y \mid \mu, \Sigma \sim \mathrm{N}_d(y \mid \mu, \Sigma)$
- Modello di osservazione

$$p(y \mid \mu, \Sigma) \propto \mid \Sigma \mid^{-1/2} \exp\left(-\frac{1}{2}(y-\mu)^T \Sigma^{-1}(y-\mu)\right),$$

- Multivariate normal with known variance p. 70-71
- Multivariate normal with unknown mean and variance p. 72-73
  The two cases follow the development seen in the univariate context but with matrix expressions since the distributions are here multivariate