Legge della Probabilità Totale

1. $P(A) = P(A \mid B)P(B) + P(A \mid B^c)P(B^c)$
2. $P(A) = \sum_j P(A \mid B_j)P(B_j)$
3. $f_X(x) = \int f_{X|Y}(x \mid y) f_Y(y)\, dy$

Regola di Bayes

1. $P(B \mid A) = \dfrac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid B^c)P(B^c)}$
2. $P(B_i \mid A) = \dfrac{P(A \mid B_i)P(B_i)}{\sum_j P(A \mid B_j)P(B_j)}$
3. $f_{Y|X}(y \mid x) = \dfrac{f_{X|Y}(x \mid y) f_Y(y)}{\int f_{X|Y}(x \mid y) f_Y(y)\, dy}$
4. $p(\theta \mid y) = \dfrac{p(y \mid \theta)\, p(\theta)}{\int_\Theta p(y \mid \tilde{\theta})\, p(\tilde{\theta})\, d\tilde{\theta}}$

Mixture Process

Siano $f_1(Y), f_2(Y), \ldots, f_n(Y)$ delle pdf/pmf. Allora:

$$f(Y) = \sum_{h=1}^{n} c_h\, f_h(Y) \quad \text{con} \quad \sum_{h=1}^{n} c_h = 1$$

Sia $Y \mid \theta$ una famiglia di variabili casuali indicizzate da θ, e sia θ una variabile casuale con pdf/pmf $G(\theta)$. Allora:

$$f(Y) = \int_{-\infty}^{\infty} f_\theta(Y)\, dG(\theta)$$

Indipendenza Condizionata e Scambiabilità

1. Indipendenza Condizionata

Siano $y_1, \ldots, y_n$ delle osservazioni condizionatamente indipendenti dato θ. Allora:

$$p(y_1, \ldots, y_n \mid \theta) = \prod_{i=1}^{n} p(y_i \mid \theta)$$

2. Scambiabilità (Exchangeability)

Siano $y_1, \ldots, y_{10}$ scambiabili. Allora:

$$p(y_1, \ldots, y_{10}) = \int_0^1 p(y_1, \ldots, y_{10} \mid \theta)\, p(\theta)\, d\theta$$

Nel caso Bernoulliano (ad esempio $y_i \in \{0, 1\}$), si può scrivere:

$$p(y_1, \ldots, y_{10}) = \int_0^1 \theta^{\sum y_i} (1-\theta)^{10 - \sum y_i}\, p(\theta)\, d\theta$$

Modelli a un parametro

1. Modello Binomiale

$$p(Y = y \mid \theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}$$

2. Distribuzione Beta

$$p(x;a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$$

Proprietà:

$$\text{Mode}(x) = \frac{a-1}{(a-1)+(b-1)} = \frac{a-1}{a+b-2}$$
$$\mathbb{E}[x] = \frac{a}{a+b}$$
$$\mathbb{V}(x) = \frac{ab}{(a+b)^2(a+b+1)} = \mathbb{E}[x](1-\mathbb{E}[x]) \cdot \frac{1}{a+b+1}$$

3. Modello Beta-Binomiale

Una prior Beta e un modello binomiale portano a una posterior Beta:

$$p(\theta \mid y) = \text{Beta}(a+y, b+n-y)$$

Implica:

$$\mathbb{E}[\theta \mid y] = \frac{a+y}{a+b+n}$$
$$\text{Mode}[\theta \mid y] = \frac{a+y-1}{a+b+n-2}$$
$$\mathbb{V}(\theta \mid y) = \mathbb{E}[\theta \mid y](1-\mathbb{E}[\theta \mid y]) \cdot \frac{1}{a+b+n+1}$$

4. Previsione (Predizione)

$$P(\tilde{Y}=1 \mid y_1,\dots,y_n) = \int_\Theta P(\tilde{Y}=1,\theta \mid y_1,\dots,y_n)\, d\theta = \mathbb{E}[\theta \mid y_1,\dots,y_n]$$
$$= \frac{a+\sum_i y_i}{a+b+n}$$
$$P(\tilde{Y}=0 \mid y_1,\dots,y_n) = \frac{b+n-\sum_i y_i}{a+b+n}$$

5. Modello di Poisson

$$P(Y_1=y_1,\dots,Y_n=y_n \mid \theta) = \prod_{i=1}^n p(y_i \mid \theta) = \prod_{i=1}^n \frac{\theta^{y_i} e^{-\theta}}{y_i!}$$
$$= \theta^{\sum_i y_i} e^{-n\theta} \prod_{i=1}^n \frac{1}{y_i!}$$

6. Prior Coniugata: Distribuzione Gamma

$$p(\theta) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta}$$

Posterior:

$$\theta \mid y_1,\dots,y_n \sim \text{Gamma}\left(a+\sum_i y_i,\, b+n\right)$$

Valori attesi e varianze:

$$\mathbb{E}[\theta \mid y_1,\dots,y_n] = \frac{a+\sum_i y_i}{b+n} = \frac{b}{b+n}\cdot\frac{a}{b} + \frac{n}{b+n}\cdot\frac{\sum_i y_i}{n}$$
$$\mathbb{V}[\theta \mid y_1,\dots,y_n] = \frac{a+\sum_i y_i}{(b+n)^2} = \frac{\mathbb{E}[\theta \mid y]}{b+n}$$

Quando $n \gg b$:

$$\mathbb{E}[\theta \mid y] \approx \bar{y}, \quad \mathbb{V}[\theta \mid y] \approx \frac{\bar{y}}{n}$$

7. Valori Predittivi Attesi e Varianza

$$\mathbb{E}[\tilde{Y} \mid y_1, \dots, y_n] = \frac{a + \sum_i y_i}{b+n} = \mathbb{E}[\theta \mid y]$$
$$\mathbb{V}[\tilde{Y} \mid y_1, \dots, y_n] = \frac{a + \sum_i y_i}{b+n} \cdot \frac{b+n+1}{b+n} = \mathbb{E}[\theta \mid y] \cdot \frac{b+n+1}{b+n}$$

Modello Normale

Assumiamo di avere un campione $y = (y_1, \dots, y_n)$ di osservazioni indipendenti e identicamente distribuite, tali che

$$y_i \sim \mathcal{N}(\mu, \sigma_y^2), \quad i = 1, \dots, n,$$

dove μ è incognito e σ_y^2 è un parametro noto. Scegliamo una distribuzione a priori normale per μ:

$$\mu \sim \mathcal{N}(\mu_0, \sigma_0^2),$$

con μ_0 e σ_0^2 iperparametri fissati.

Distribuzione Posteriore

La distribuzione posteriore di μ risulta essere anch'essa normale:

$$\mu \mid \sigma_y^2, y \sim \mathcal{N}(\mu_n, \sigma_n^2),$$

con

$$\mu_n = \frac{\frac{n}{\sigma_y^2}\,\overline{y} + \frac{1}{\sigma_0^2}\,\mu_0}{\frac{n}{\sigma_y^2} + \frac{1}{\sigma_0^2}} \quad \text{e} \quad \sigma_n^2 = \frac{1}{\frac{n}{\sigma_y^2} + \frac{1}{\sigma_0^2}},$$

dove $\overline{y}$ è la media campionaria.

Precisioni

Spesso si utilizza la precisione, definita come l'inverso della varianza:

$$\tau_0^2 = \frac{1}{\sigma_0^2} \quad \text{(precisione a priori)},$$

$$\tau_y^2 = \frac{1}{\sigma_y^2} \quad \text{(precisione del campionamento)},$$

$$\tau_n^2 = \frac{1}{\sigma_n^2} \quad \text{(precisione posteriore)}.$$

Nel modello, la precisione posteriore risulta essere la somma della precisione a priori e della precisione derivante dai dati:

$$\tau_n^2 = \tau_0^2 + n\,\tau_y^2.$$

Di conseguenza, la media posteriore si esprime anche in termini di precisioni:

$$\mu_n = \frac{n\,\tau_y^2}{n\,\tau_y^2 + \tau_0^2}\,\overline{y} + \frac{\tau_0^2}{n\,\tau_y^2 + \tau_0^2}\,\mu_0.$$

Modello Normale con Parametri Incerti: Prior Normal–Gamma

Assumiamo di avere un campione

$$y_1, \dots, y_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2),$$

con μ e σ^2 incerti.

Dobbiamo definire un modello a priori per la coppia (μ, σ^2). Utilizziamo la regola della moltiplicazione:

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

Scegliamo:

- $\mu \mid \sigma^2 \sim \mathcal{N}\Big(\mu_0,\, \frac{\sigma^2}{\kappa_0}\Big)$, dove μ_0 è il valore a priori per la media e κ_0 rappresenta la forza (o affidabilità) del prior.
- Definiamo la precisione $\tau^2 = 1/\sigma^2$ e poniamo:

$$\tau^2 \sim \text{Gamma}\left(\frac{\nu_0}{2},\, \frac{\nu_0\sigma_0^2}{2}\right),$$

dove ν_0 e σ_0^2 sono iperparametri che esprimono la quantità di informazione a priori sulla varianza.

Questa scelta porta ad un prior congiunto di tipo **Normal–Gamma**, indicato con:

$$\text{NG}(\mu_0,\, \kappa_0,\, \tfrac{\nu_0}{2},\, \tfrac{\nu_0\sigma_0^2}{2}).$$

Distribuzione Posteriore Coniugata

Il joint posterior di μ e σ^2 è dato da:

$$p(\mu, \sigma^2 \mid y_1, \ldots, y_n) = p(\mu \mid \sigma^2, y_1, \ldots, y_n)\, p(\sigma^2 \mid y_1, \ldots, y_n).$$

1. Posterior per μ condizionato a σ^2:
Dall'aggiornamento con i dati si ottiene:

$$\mu \mid \sigma^2, y \sim \mathcal{N}\Big(\mu_n,\, \frac{\sigma^2}{\kappa_n}\Big),$$

dove

$$\kappa_n = \kappa_0 + n, \quad \mu_n = \frac{\kappa_0\mu_0 + n\overline{y}}{\kappa_0 + n}.$$

2. Posterior per σ^2 (o $\tau^2 = 1/\sigma^2$):
Utilizzando

$$p(\sigma^2 \mid y) \propto p(\sigma^2) p(y_1, \ldots, y_n \mid \sigma^2)\, \frac{1}{\sigma^2},$$

si ottiene

$$\frac{1}{\sigma^2} \mid y \sim \text{Gamma}\left(\frac{\nu_n}{2},\, \frac{\nu_n\sigma_n^2}{2}\right),$$

con

$$\nu_n = \nu_0 + n,$$

e

$$\sigma_n^2 = \frac{1}{\nu_n}\left[\nu_0\sigma_0^2 + (n-1)s^2 + \frac{n\kappa_0}{\kappa_0 + n}(\overline{y} - \mu_0)^2\right],$$

dove s^2 è la varianza campionaria e $\overline{y}$ la media campionaria.

Pertanto, il joint posterior è un $\text{NG}(\mu_n, \kappa_n, \frac{\nu_n}{2}, \frac{\nu_n\sigma_n^2}{2})$.

Distribuzioni Marginali Post.

- $\mu \mid \sigma^2, y \sim \mathcal{N}\Big(\mu_n,\, \frac{\sigma^2}{\kappa_n}\Big)$.
- $\frac{1}{\sigma^2} \mid y \sim \text{Gamma}\left(\frac{\nu_n}{2},\, \frac{\nu_n\sigma_n^2}{2}\right)$.
- Marginalizzando σ^2, si ottiene la distribuzione a posteriori per μ che risulta essere una distribuzione t:

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

Stima della Media e Confronto degli Stimatori

Un possibile stimatore della media è il *biasn*, definito come la media pesata:

$$\hat{\mu}_b = \frac{n\,\overline{y} + \kappa_0\mu_0}{n + \kappa_0} = w\,\overline{y} + (1-w)\mu_0, \quad \text{con } w = \frac{n}{n + \kappa_0}.$$

Lo stimatore classico per la media, invece, è semplicemente:

$$\hat{\mu}_e = \overline{y}.$$

Le proprietà attese sono:

$$\mathbb{E}\big[\hat{\mu}_b \mid \mu = \mu^*\big] = w\mu^* + (1-w)\mu_0, \quad \mathbb{E}\big[\hat{\mu}_e \mid \mu = \mu^*\big] = \mu^*,$$

mentre i rispettivi errori quadratici medi (MSE) sono:

$$\text{MSE}\left(\hat{\mu}_e \mid \mu^*\right) = \sigma_n^2, \quad \text{MSE}\left(\hat{\mu}_b \mid \mu^*\right) = w^2\sigma_n^2 + (1-w)^2(\mu_0 - \mu^*)^2.$$

Prior Non-Informativi di Jeffreys

Sia una trasformazione del parametro data da

$$\eta = g(\theta),$$

con inversa

$$\theta = g^{-1}(\eta),$$

ammettendo che entrambe esistano. La densità corrispondente per η si ottiene tramite la formula di cambio variabile:

$$p(\eta) \propto p\big(g^{-1}(\eta)\big) \left| \frac{\partial g^{-1}(\eta)}{\partial \eta} \right|.$$

Per scegliere un prior non-informativo per θ si utilizza il criterio di Jeffreys, cioè:

$$p(\theta) \propto \sqrt{I(\theta)},$$

dove $I(\theta)$ è l'informazione di Fisher definita da

$$I(\theta) = -\mathbb{E}\left[\frac{\partial^2 \log l(y \mid \theta)}{\partial \theta^2} \middle| \theta \right].$$

Pertanto, anche per il parametro trasformato η si ha:

$$p(\eta) \propto \sqrt{I(\eta)}.$$