Massima verosimiglianza

I. Frosio
AIS Lab.
frosio@dsi.unimi.it
http:\\homes.dsi.unimi.it\~frosio\

Overview

- Nozioni di base
- Funzione di verosimiglianza
- Stima alla massima verosimiglianza
- Il caso Gaussiano
- Sistemi ai minimi quadrati
- Il caso Poissoniano
- Stima di due rette

Nozioni di base

- Variabile casuale: variabile che può assumere un valore secondo una densità di probabilità.
- Es. distribuzione gaussiana, poissoniana, binomiale, uniforme, …
- Gaussiana (media μ, std σ):

$$p(x \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left[-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2 \right]$$

Nozioni di base

- Gaussiana: quale è la probabilità che la variabile x assuma il valore y?

$$p\left(x = y \mid \mu, \sigma\right) = \frac{1}{\sqrt{2\pi\sigma}} \cdot \exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right]$$

Nozioni di base

- Gaussiana: siano date due realizzazioni **indipendenti** della stessa variabile casuale x… Quale è la probabilità di misurare y1 nella prima realizzazione e y2 nella seconda realizzazione?

$$p(y1 \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left[-\frac{1}{2}\left(\frac{y1-\mu}{\sigma}\right)^2\right]$$

$$p(y2 \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left[-\frac{1}{2}\left(\frac{y2-\mu}{\sigma}\right)^2\right]$$

$$p(y1, y2 \mid \mu, \sigma) = p(y1 \mid \mu, \sigma) \cdot p(y2 \mid \mu, \sigma) =$$

$$= \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left[-\frac{1}{2}\left(\frac{y1-\mu}{\sigma}\right)^2\right] \cdot \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left[-\frac{1}{2}\left(\frac{y2-\mu}{\sigma}\right)^2\right]$$

Funzione di verosimiglianza

- Siano date **N variabili casuali** **indipendenti**… Quale è la **probabilità di misurare il vettore [y1, …, yN]**?

$$p(y1, y2, \ldots, yN) = p(y1) \cdot p(y2) \cdot \ldots p(yN) = L(y1, y2, \ldots, yN)$$

- Questa è la FUNZIONE DI VEROSIMIGLIANZA (Likelihood, L).

Funzione di verosimiglianza (riassunto)

- Data una serie di misure yi i=1…N di variabili casuali…
- … Note le densità di probabilità di ciascuna variabile casuale…
- … Sotto l'ipotesi che le variabili siano tra loro indipendenti…
- … E' possibile scrivere la funzione di verosimiglianza come il prodotto delle probabilità di ciascuna misura yi i=1…N.

Stima alla massima verosimiglianza

- Supponiamo il vettore **y** corrisponda a N realizzazioni di una variabile gaussiana a media μ, std σ.
- La funzione verosimiglianza dipende da μ e σ.

$$L(y1, y2, ..., yN) = p(y1) \cdot p(y2) \cdot ... p(yN) =$$

$$= \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left[-\frac{1}{2}\left(\frac{y1-\mu}{\sigma}\right)^2\right] \cdot ... \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left[-\frac{1}{2}\left(\frac{yN-\mu}{\sigma}\right)^2\right] =$$

$$= L(y1, y2, ..., yN \mid \mu, \sigma)$$

Stima alla massima verosimiglianza

- Se **massimizziamo** $L=L(\mathbf{y}|\mu,\sigma)$...
- Troviamo i parametri μ,σ tali per cui è massima la probabilità di misurare il vettore di dati $\mathbf{y}$.

- STIMA ALLA MASSIMA VEROSIMIGLIANZA.

- Att.ne! In generale, le variabili possono avere densità di probabilità diverse, ciascuna descritta da un set di parametri stimabili con l'approccio alla massima verosimigianza...

Stima alla massima verosimiglianza (riassunto)

- La funzione di verosimiglianza dipende dai parametri che definiscono le densità di probabilità delle variabili casuali che entrano nella verosimiglianza…
- Massimizzando la funzione di verosimiglianza rispetto a tali parametri se ne effettua la stima in modo tale che il vettore osservato yi i=1…N sia massimamente probabile (massima verosimiglianza).

Il caso gaussiano

- Nella maggior parte dei casi viene assunto rumore gaussiano sui dati.
- N misurazioni di una variabile casuale distribuita in modo gaussiano.
- Si vogliono stimare media e varianza della gaussiana.

$$L(y1, y2, \ldots, yN \mid \mu, \sigma) =$$

$$= \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left[-\frac{1}{2}\left(\frac{y1-\mu}{\sigma} \right)^2 \right] \cdot \ldots \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left[-\frac{1}{2}\left(\frac{yN-\mu}{\sigma} \right)^2 \right]$$

Il caso gaussiano

- E' solitamente più facile minimizzare il logaritmo negativo della verosimiglianza (prodotto → sommatoria)

$$
\begin{aligned}
f(\mu,\sigma) &= -\ln\left[L(y1, y2, \dots, yN \mid \mu,\sigma)\right] = \\
&= -\ln\left\{\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left[-\frac{1}{2}\left(\frac{yi-\mu}{\sigma}\right)^2\right]\right\} = \\
&= -\sum_{i=1}^{N} \ln\left\{\frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left[-\frac{1}{2}\left(\frac{yi-\mu}{\sigma}\right)^2\right]\right\} = \\
&= N \cdot \ln\left(\sqrt{2\pi}\right) + N \cdot \ln(\sigma) + \sum_{i=1}^{N}\left[\frac{1}{2}\left(\frac{yi-\mu}{\sigma}\right)^2\right]
\end{aligned}
$$

Il caso gaussiano

- Per trovare il minimo, poniamo a zero le derivate:

$$\frac{\partial f(\mu,\sigma)}{\partial \mu} = \frac{\partial}{\partial \mu}\left\{ N \cdot \ln\left(\sqrt{2\pi}\right) + N \cdot \ln(\sigma) + \sum_{i=1}^{N}\left[\frac{1}{2}\cdot\left(\frac{yi-\mu}{\sigma}\right)^2\right]\right\} =$$

$$= 0 + 0 + \sum_{i=1}^{N}\frac{1}{2}\cdot\left(\frac{yi-\mu}{\sigma}\right)\cdot 2\cdot\left(-\frac{1}{\sigma}\right) = \frac{1}{\sigma^2}\cdot\sum_{i=1}^{N}(yi-\mu) = 0 \Rightarrow$$

$$\sum_{i=1}^{N}(yi-\mu) = 0 \Rightarrow \sum_{i=1}^{N} yi = N\cdot\mu \Rightarrow \mu = \frac{\sum_{i=1}^{N} yi}{N}$$ ***Media campionaria!***

Il caso gaussiano

- Per trovare il minimo, poniamo a zero le derivate:

$$\frac{\partial f(\mu,\sigma)}{\partial \sigma} = \frac{\partial}{\partial \sigma}\left\{ N \cdot \ln\left(\sqrt{2\pi}\right) + N \cdot \ln(\sigma) + \sum_{i=1}^{N}\left[\frac{1}{2}\cdot\left(\frac{yi-\mu}{\sigma}\right)^2\right]\right\} =$$

$$= 0 + \frac{N}{\sigma} + \sum_{i=1}^{N}\frac{1}{2}\cdot\left(\frac{yi-\mu}{\sigma}\right)\cdot(-2)\cdot(yi-\mu)\cdot\left(\frac{1}{\sigma^2}\right) =$$

$$= \frac{N}{\sigma} - \frac{1}{\sigma^3}\sum_{i=1}^{N}(yi-\mu)^2 = 0 \Rightarrow N - \frac{1}{\sigma^2}\sum_{i=1}^{N}(yi-\mu)^2 = 0 \Rightarrow$$

$$N \cdot \sigma^2 = \sum_{i=1}^{N}(yi-\mu)^2 \Rightarrow \sigma^2 = \frac{\sum_{i=1}^{N}(yi-\mu)^2}{N}$$

Varianza campionaria!

Il caso gaussiano (riassunto)

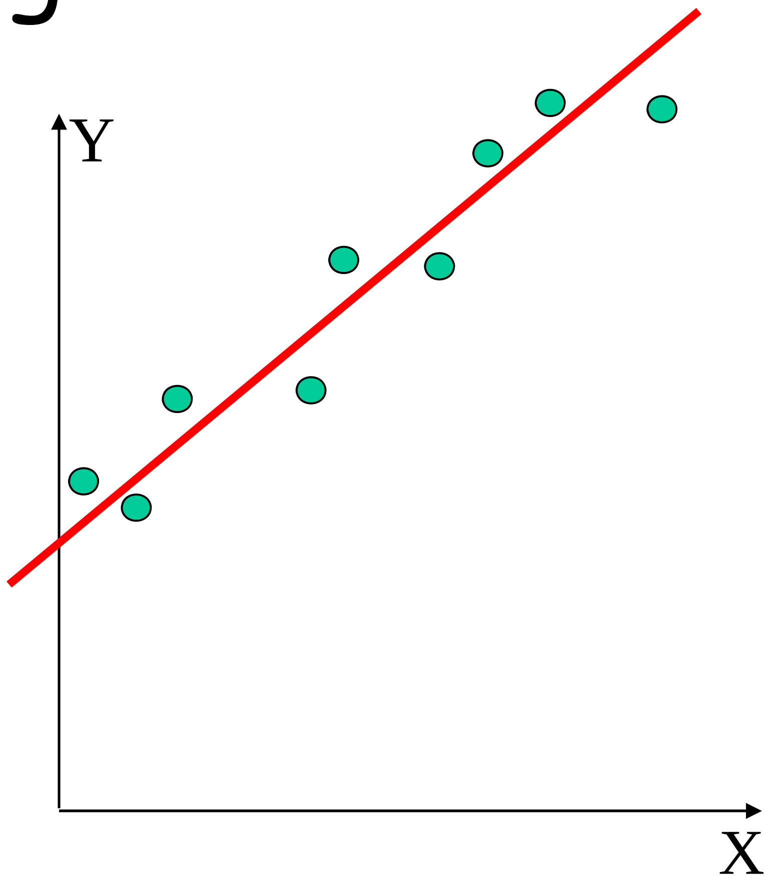
- Date N misurazioni di una variabile casuale con distribuzione gaussiana, posso scrivere la funzione di verosimiglianza;
- Massimizzando la funzione di verosimiglianza (minimizzando il logaritmo negativo della verosimiglianza) rispetto a μ e σ, ottengo una stima della media e varianza della distribuzione;
- Tali stime coincidono con la media e la varianza campionarie.

Stima ai minimi quadrati e verosimiglianza

- Si consideri il problema di stima di una retta:
- Sia $y = mx + b$ una retta, con m e b incogniti;
- Siano $y_i = 1 \ldots N$ una serie di N misure effettuate per $x_i = x_1 \ldots x_N$.
- Le misure y_i siano affette da rumore gaussiano a media nulla.
- In pratica: $y_i = G(mx_i + b, \sigma^2)$, dove $G(\mu, \sigma^2)$ indica una distribuzione gaussiana a media m e varianza σ^2.

Stima ai minimi quadrati e verosimiglianza

Il problema di stima della retta:



- Date le misure y_i $i=1\ldots N$ per le posizioni x_i $i=1\ldots N$…
- … Trovare i parametri m e b della retta $y=mx+b$ cui appartengono i dati.

Stima ai minimi quadrati e verosimiglianza

- Impostiamo il problema scrivendo la funzione di verosimiglianza e massimizzando tale funzione rispetto a m e b…
- Scriviamo prima di tutto la densità di probabilità per ciascun dato:

$$p(yi \mid m, b, xi) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left[-\frac{1}{2} \left(\frac{yi - mxi - b}{\sigma} \right)^2 \right]$$

Stima ai minimi quadrati e verosimiglianza

- Scriviamo allora il logaritmo negativo della verosimiglianza:

$$f(m,b) = -\sum_{i=1}^{N} \ln\left\{ \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left[-\frac{1}{2}\left(\frac{yi - mxi - b}{\sigma} \right)^2 \right] \right\} =$$

$$= -\sum_{i=1}^{N} \ln\left(\frac{1}{\sqrt{2\pi}\sigma} \right) - \sum_{i=1}^{N} \left[-\frac{1}{2}\left(\frac{yi - mxi - b}{\sigma} \right)^2 \right] =$$

$$= -\sum_{i=1}^{N} \ln\left(\frac{1}{\sqrt{2\pi}\sigma} \right) + \frac{1}{2\sigma^2} \sum_{i=1}^{N} (yi - mxi - b)^2$$

Stima ai minimi quadrati e verosimiglianza

- E massimizziamolo ponendo a zero le derivate rispetto a m e b:

$$\frac{\partial f(m,b)}{\partial m} = \frac{\partial}{\partial m}\left[-\sum_{i=1}^{N} \ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right) + \frac{1}{2\sigma^2}\sum_{i=1}^{N}(yi - m\cdot xi - b)^2 \right] =$$

$$= 0 + \frac{1}{2\sigma^2}\sum_{i=1}^{N}(yi - m\cdot xi - b)\cdot 2\cdot(-xi) =$$

$$= \frac{1}{\sigma^2}\sum_{i=1}^{N}(yi - m\cdot xi - b)\cdot xi = 0 \Rightarrow \sum_{i=1}^{N}(yi - m\cdot xi - b)\cdot xi = 0 \Rightarrow$$

$$\left[\sum_{i=1}^{N}(yi\cdot xi)\right] - m\cdot\left[\sum_{i=1}^{N}(xi^2)\right] - b\cdot\left[\sum_{i=1}^{N}(xi)\right] = 0 \Rightarrow$$

$$m\cdot\left[\sum_{i=1}^{N}(xi^2)\right] + b\cdot\left[\sum_{i=1}^{N}(xi)\right] = \left[\sum_{i=1}^{N}(yi\cdot xi)\right]$$

Stima ai minimi quadrati e verosimiglianza

- E massimizziamolo ponendo a zero le derivate rispetto a m e b:

$$\frac{\partial f(m,b)}{\partial b}=\frac{\partial}{\partial b}\left[-\sum_{i=1}^{N}\ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right)+\frac{1}{2\sigma^2}\sum_{i=1}^{N}(yi-m\cdot xi-b)^2\right]=$$

$$=0+\frac{1}{2\sigma^2}\sum_{i=1}^{N}(yi-m\cdot xi-b)\cdot 2\cdot(-1)=$$

$$=\frac{1}{\sigma^2}\sum_{i=1}^{N}(yi-m\cdot xi-b)=0\Rightarrow\sum_{i=1}^{N}(yi-m\cdot xi-b)=0\Rightarrow$$

$$\left[\sum_{i=1}^{N}(yi)\right]-m\cdot\left[\sum_{i=1}^{N}(xi)\right]-b\cdot\left[\sum_{i=1}^{N}(1)\right]=0\Rightarrow$$

$$m\cdot\left[\sum_{i=1}^{N}(xi)\right]+b\cdot\left[\sum_{i=1}^{N}(1)\right]=\left[\sum_{i=1}^{N}(yi)\right]$$

Stima ai minimi quadrati e verosimiglianza

- Ponendo a zero le due derivate abbiamo dunque ottenuto un sistema lineare che permette di calcolare m e b*:

$$\begin{bmatrix} \sum_{i=1}^{N}\left(xi^2\right) & \sum_{i=1}^{N}\left(xi\right) \\ \sum_{i=1}^{N}\left(xi\right) & \sum_{i=1}^{N}\left(1\right) \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N}\left(yi \cdot xi\right) \\ \sum_{i=1}^{N}\left(yi\right) \end{bmatrix}$$

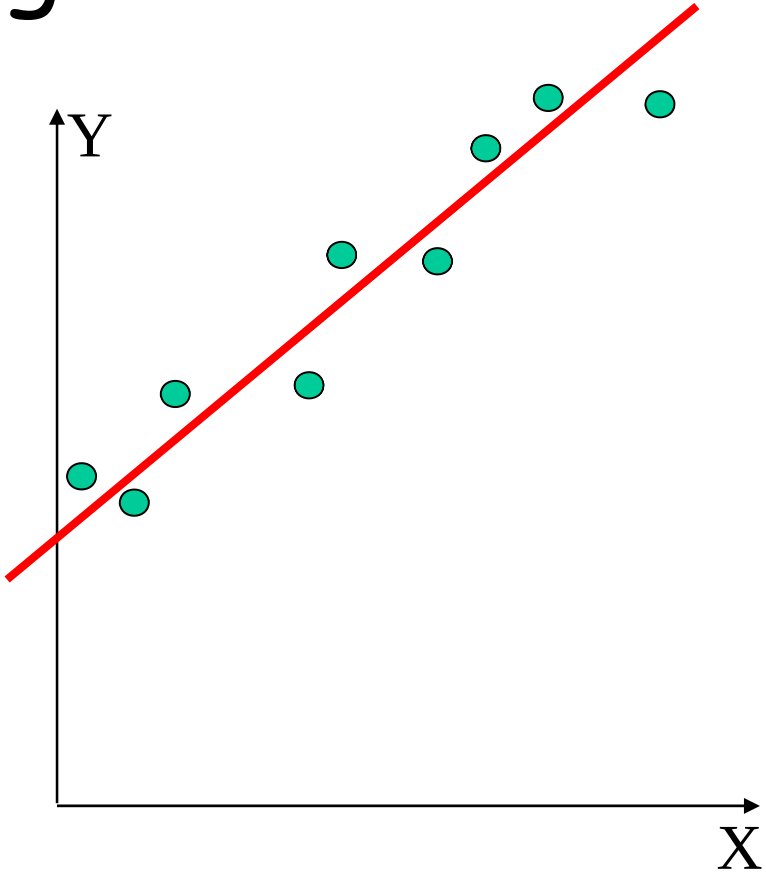
- *Ax=b → x=A^{-1}b

Stima ai minimi quadrati e verosimiglianza

Riformuliamo il problema della retta come un problema ai minimi quadrati.

Per ogni punto, dovrebbe valere $y_i = m x_i + b$.

Cerchiamo i parametri m e b tali per cui questa condizione è verificata "al meglio" nel senso dei minimi quadrati.



Stima ai minimi quadrati e verosimiglianza

- Scriviamo l'equazione della retta per tutti i punti in forma matriciale (sistema lineare Ax=b, N equazioni, 2 incognite):

$$\begin{bmatrix} x1 & 1 \\ x2 & 1 \\ \dots & \dots \\ xN & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} y1 \\ y2 \\ \dots \\ yN \end{bmatrix}$$

- Vogliamo trovare x t.x. $||(Ax-b)^T(Ax-b)||$ è minima (minimizzazione dei quadrati dei residui).

Stima ai minimi quadrati e verosimiglianza

- Nella soluzione ai minimi quadrati del sistema lineare $Ax=b$ si definisce un vettore errore $e=Ax-b$;
- Nel caso di soluzione "perfetta" $e=0$;
- Dal momento che abbiamo un numero di equazioni maggiore rispetto al numero di incognite, cerchiamo il vettore e a norma minima;
- In pratica cerchiamo x t.c. $e^T e=\Sigma_i ei^2$ è minimo.

Stima ai minimi quadrati e verosimiglianza

- La soluzione ai minimi quadrati del sistema Ax=b si ottiene da:
- $A^T A x = A^T b \rightarrow x = (A^T A)^{-1} A^T b$.
- Scriviamo allora $A^T A$ e $A^T b$ per il nostro sistema:

$$A^T A \rightarrow \begin{bmatrix} x1 & x2 & \dots & xN \\ 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} x1 & 1 \\ x2 & 1 \\ \dots & \dots \\ xN & 1 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} \left(xi^2\right) & \sum_{i=1}^{N} (xi) \\ \sum_{i=1}^{N} (xi) & \sum_{i=1}^{N} (1) \end{bmatrix}$$

$$A^T b \rightarrow \begin{bmatrix} x1 & x2 & \dots & xN \\ 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} y1 \\ y2 \\ \dots \\ yN \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} (yi \cdot xi) \\ \sum_{i=1}^{N} (yi) \end{bmatrix}$$

Stima ai minimi quadrati e verosimiglianza

- Se confrontiamo il sistema lineare ottenuto dalla massimizzazione della verosimiglianza (slide 22) con il sistema lineare $A^TAx=A^Tb$ per la soluzione del sistema ai minimi quadrati (slide 27) ci accorgiamo che… Sono uguali!

$$\begin{bmatrix} \sum_{i=1}^{N}\left(xi^2\right) & \sum_{i=1}^{N}\left(xi\right) \\ \sum_{i=1}^{N}\left(xi\right) & \sum_{i=1}^{N}\left(1\right) \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N}\left(yi \cdot xi\right) \\ \sum_{i=1}^{N}\left(yi\right) \end{bmatrix}$$

- In generale: **<u>LA SOLUZIONE DEL PROBLEMA DI MASSIMIZZAZIONE DELLA VEROSMIGLIANZA E' EQUIVALENTE ALLA SOLUZIONE DEL PROBLEMA AI MINIMI QUADRATI IN CASO DI DISTRIBUZIONE GAUSSIANA (RUMORE GAUSSIANO SUL VETTORE DEI DATI MISURATI Y).</u>**

Stima ai minimi quadrati e verosimiglianza

- Anche nella formulazione di f(m,b) è evidente l'analogia con i minimi quadrati (slide 18):

$$f(m,b) = -\sum_{i=1}^{N} \ln\left(\frac{1}{\sqrt{2\pi\sigma}}\right) + \frac{1}{2\sigma^2}\sum_{i=1}^{N}(yi - mxi - b)^2$$

$$\min_{m,b} f(m,b) = \min_{m,b} \frac{1}{2\sigma^2}\sum_{i=1}^{N}(yi - mxi - b)^2 =$$

$$\min_{m,b} \sum_{i=1}^{N}(yi - mxi - b)^2$$

Minizzazione dei quadrati dei residui!

Stima ai minimi quadrati e verosimiglianza (riassunto)

- La formulazione di un problema nei termini dei minimi quadrati rappresenta un caso particolare di problema di massimizzazione della verosimiglianza.
- In particolare, l'approccio alla verosimiglianza e ai minimi quadrati sono equivalenti nel caso di distribuzione gaussiana della variabili misurate.

Il caso poissoniano

- La formulazione di un problema di verosimiglianza permette di trattare casi in cui la variabili misurate abbiano distribuzione diversa da quella gaussiana.
- Consideriamo ad esempio una variabile poissoniana (es. conteggio di un numero limitato di fotoni, es. radiografia).
- Media della variabile casuale (=varianza nel caso della poisson) = λ.

$$p(yi \mid \lambda) = \frac{\lambda^{yi} \cdot e^{-\lambda}}{yi!}$$

Il caso poissoniano

- Sia y un vettore di misure di una variabile poissoniana. Si vuole stimare la media della variabile.
- Scriviamo il logaritmo negativo della funzione di verosimiglianza:

$$
-\ln(L) = -\sum_{i=1}^{N} \ln\left[p(yi \mid \lambda)\right] = -\sum_{i=1}^{N} \ln\left[\frac{\lambda^{yi} \cdot e^{-\lambda}}{yi!}\right] =
$$

$$
= -\sum_{i=1}^{N} \ln\left[\lambda^{yi}\right] - \sum_{i=1}^{N} \ln\left[e^{-\lambda}\right] - \sum_{i=1}^{N} \ln\left[\frac{1}{yi!}\right] =
$$

$$
-\sum_{i=1}^{N} yi \cdot \ln[\lambda] + \sum_{i=1}^{N} \lambda + \sum_{i=1}^{N} \ln[yi!] =
$$

$$
= -\ln(\lambda) \cdot \sum_{i=1}^{N} yi + N \cdot \lambda + \sum_{i=1}^{N} \ln[yi!]
$$

Il caso poissoniano

- Massimizziamo la verosimiglianza rispetto a λ (ponendo a zero la derivata):

$$\frac{\partial}{\partial\lambda}\left[-\ln(L)\right]=\frac{\partial}{\partial\lambda}\left[-\ln(\lambda)\cdot\sum_{i=1}^{N} yi+N\cdot\lambda+\sum_{i=1}^{N}\ln[yi!]\right]=$$

$$=-\frac{1}{\lambda}\cdot\sum_{i=1}^{N} yi+N+0=0\Rightarrow\lambda=\frac{\sum_{i=1}^{N} yi}{N}$$

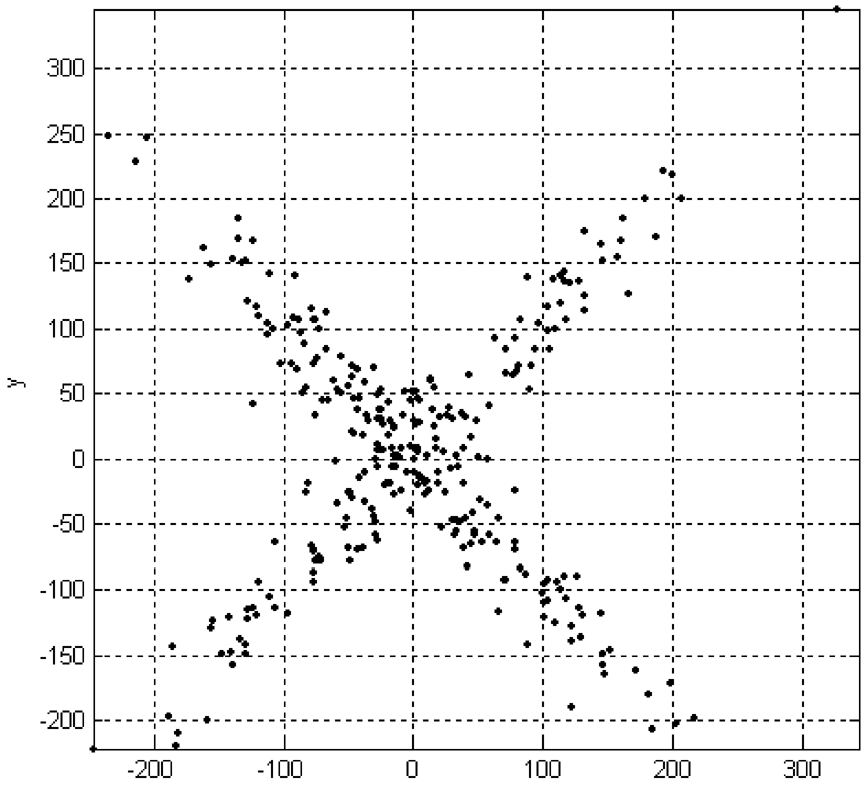
- Otteniamo anche in questo caso la media campionaria come stima della media della distribuzione.

Il caso poissoniano (riassunto)

- L'approccio alla massima verosimiglianza permette di effettuare stime di parametri, data una qualsiasi densità di probabilità dei dati misurati (non solo gaussiana, poisson!)

Stima di due rette

- Proviamo ora a complicare il problema: si vogliono stimare i coefficienti angolari di due rette passanti per l'origine.
- I dati misurati yi possono provenire dall'una o dall'altra retta con la stessa probabilità.
- Sui dati misurati è presente rumore gaussiano con varianza σ^2.



Stima di due rette

- Scriviamo la funzione di verosimiglianza:

$$p(yi) = P1 \cdot G\left(m1 \cdot xi, \sigma^2\right) + P2 \cdot G\left(m2 \cdot xi, \sigma^2\right)$$

$$G\left(\mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

- In pratica un punto yi può provenire dalla retta 1 con probabilità P1 o dalla retta 2 con probabilità P2. In ciascuno dei due casi il punto misurato ha una distribuzione gaussiana "centrata" sulla retta stessa.

Stima di due rette

- Calcolo il logaritmo negativo della verosimiglianza:

$$f(m1, m2) = -\sum_{i=1}^{N} \ln[p(yi)] = -\sum_{i=1}^{N} \ln\left[P1 \cdot G\left(m1 \cdot xi, \sigma^2\right) + P2 \cdot G\left(m2 \cdot xi, \sigma^2\right)\right]$$

$$= -\sum_{i=1}^{N} \ln\left[P1 \cdot p_1\left(xi, \sigma^2\right) + P2 \cdot p_2\left(xi, \sigma^2\right)\right]$$

$$p_j\left(yi, \sigma^2\right) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{1}{2}\left(\frac{x - m_j \cdot xi}{\sigma}\right)^2}$$

Stima di due rette

- Provo a minimizzare il logaritmo negativo ponendo a zero le derivate rispetto a m1 e m2:

$$\frac{\partial}{\partial mj}\left[f\left(m1,m2\right)\right]=\frac{\partial}{\partial mj}\left\{-\sum_{i=1}^{N}\ln\left[p\left(yi\right)\right]\right\}=-\sum_{i=1}^{N}\frac{\partial}{\partial mj}\left\{\ln\left[p\left(yi\right)\right]\right\}=$$

$$=-\sum_{i=1}^{N}\frac{1}{p\left(yi\right)}\cdot\frac{\partial}{\partial mj}\left\{p\left(yi\right)\right\}=-\sum_{i=1}^{N}\frac{1}{p\left(yi\right)}\cdot\frac{\partial}{\partial mj}\left\{\sum_{k=1}^{2}P_k\cdot p_k\left(yi\right)\right\}=$$

$$=-\sum_{i=1}^{N}\frac{1}{p\left(yi\right)}\cdot\frac{\partial}{\partial mj}\left\{Pj\cdot p_j\left(yi\right)\right\}=-\sum_{i=1}^{N}\frac{Pj}{p\left(yi\right)}\cdot\frac{\partial}{\partial mj}\left\{p_j\left(yi\right)\right\}$$

Stima di due rette

$$\frac{\partial}{\partial mj}\{p_j(yi)\}=\frac{\partial}{\partial mj}\left\{\frac{1}{\sqrt{2\pi}\sigma}\cdot e^{-\frac{1}{2}\left(\frac{x-m_j\cdot xi}{\sigma}\right)^2}\right\}=\frac{1}{\sqrt{2\pi}\sigma}\cdot\frac{\partial}{\partial mj}\left\{e^{-\frac{1}{2}\left(\frac{x-m_j\cdot xi}{\sigma}\right)^2}\right\}=$$

$$=\frac{1}{\sqrt{2\pi}\sigma}\cdot e^{-\frac{1}{2}\left(\frac{x-m_j\cdot xi}{\sigma}\right)^2}\cdot\frac{\partial}{\partial mj}\left[-\frac{1}{2}\left(\frac{x-m_j\cdot xi}{\sigma}\right)^2\right]=$$

$$=p_j(yi)\cdot\left(-\frac{1}{2\sigma^2}\right)\cdot 2\cdot(x-m_j\cdot xi)\cdot(-xi)=p_j(yi)\cdot\frac{(x-m_j\cdot xi)\cdot xi}{\sigma^2}$$

Stima di due rette

- Riprendendo la derivata:

$$\frac{\partial}{\partial mj}\left[f\left(m1,m2\right)\right]=-\sum_{i=1}^{N}\frac{Pj}{p\left(xi\right)}\cdot\frac{\partial}{\partial mj}\left\{p_j\left(yi\right)\right\}=$$

$$=-\sum_{i=1}^{N}\frac{Pj}{p\left(yi\right)}\cdot p_j\left(yi\right)\cdot\frac{\left(x-m_j\cdot xi\right)\cdot xi}{\sigma^2}$$

Stima di due rette

- Se cerchiamo di porre a zero le due derivate contemporaneamente otteniamo il seguente sistema (non lineare in due incognite – non può essere risolto in maniera analitica!):

$$
\begin{cases}
-\sum_{i=1}^{N} \dfrac{P1 \cdot \dfrac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{1}{2}\left(\frac{x - m_1 \cdot xi}{\sigma}\right)^2}}{P1 \cdot \dfrac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{1}{2}\left(\frac{x - m_1 \cdot xi}{\sigma}\right)^2} + P2 \cdot \dfrac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{1}{2}\left(\frac{x - m_2 \cdot xi}{\sigma}\right)^2}} \cdot \dfrac{(x - m_1 \cdot xi) \cdot xi}{\sigma^2} = 0 \\
-\sum_{i=1}^{N} \dfrac{P1 \cdot \dfrac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{1}{2}\left(\frac{x - m_2 \cdot xi}{\sigma}\right)^2}}{P1 \cdot \dfrac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{1}{2}\left(\frac{x - m_1 \cdot xi}{\sigma}\right)^2} + P2 \cdot \dfrac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{1}{2}\left(\frac{x - m_2 \cdot xi}{\sigma}\right)^2}} \cdot \dfrac{(x - m_2 \cdot xi) \cdot xi}{\sigma^2} = 0
\end{cases}
$$

Stima di due rette (riassunto)

- In generale, la formulazione di un problema alla massima verosimiglianza porta ad un sistema non lineare di equazioni non risolvibile analiticamente.
- Per la minimizzazione del logaritmo negativo della verosimiglianza è allora necessario ricorrere ad un **algoritmo di ottimizzazione iterativo**.

Bibliografia

- Cristopher M. Bishop, Pattern Recognition and Machine Learning, Capitoli: 1.2.3 (verosimiglianza), 1.2.4 (gaussiana), 2.3.4 (likelihood e gaussiana), 1.2.5 (fitting di curve con la verosimiglianza).