LABORATORIO STATISTICO INFORMATICO

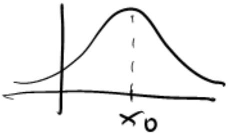
Pietro Millossovich

Corso di Laurea Triennale in Statistica e Informatica per l'Azienda, la Finanza e l'Assicurazione
DEAMS, Universita di Trieste

OTTIMIZZAZIONE

- Obbiettivo: risolvere in **R** problemi di massimizzazione/minimizzazione di una funzione
- Applicazioni
  - economia: max del profitto/utilità attesa
  - finanza matematica: selezione di portafoglio
  - fitting: adattamento di una funzione a un insieme di punti
  - risoluzione di un (sistema) di equazioni
  - Statistica inferenziale parametrica: massima verosimiglianza
  - ...

OTTIMIZZAZIONE

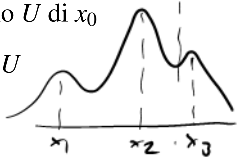


- Sia $f : D(\subset \mathbb{R}^n) \to \mathbb{R}$, con $D \neq \varnothing$
  - ▷ x_0 punto di max (min) assoluto se

$$f(x_0) \geq (\leq) f(x) \text{ per ogni } x \in D$$

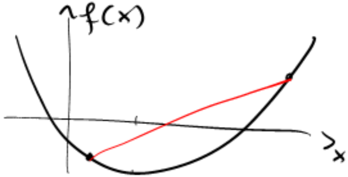
  - ▷ x_0 punto di max (min) relativo se esiste un intorno U di x_0

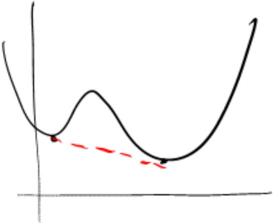
$$f(x_0) \geq (\leq) f(x) \text{ per ogni } x \in D \cap U$$



- Osservazioni:
  - ▷ Max/min relativi e/o assoluti possono non esistere o esistere (ma non essere unici)
  - ▷ x_0 max (min) assoluto $\Rightarrow$ x_0 max (min) relativo; $\Leftarrow$ vale se f concava (convessa)
  - ▷ Problema: trovare max (min) assoluto di una funzione non concava (convessa) con molti punti di max (min) relativi
  - ▷ ottimizzazione vincolata: $D = \{x \in \mathbb{R}^n : g(x) \leq 0, h(x) = 0\}$,
  - ▷ massimizzare f equivale a minimizzare $-f$

f CONVESSA

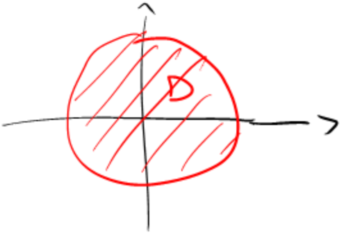




f STRETTAMENTE CONVESSA

$\Rightarrow$ PUNTI DI MIN (SE ESISTONO) SONO UNICI

ESEMPIO : OTTIMIZZAZIONE VINCOLATA



$$\min \; x - y$$
$$x^2 + y^2 \leq 1$$

SOLUZIONE $\longrightarrow$ MOLTIPLICATORI DI LAGRANGE (VINCOLI DI UGUAGLIANZA)

KHUN-TUCKER (VINCOLI DI UGUAGLIANZA E DISUGUAGLIANZA)

ESEMPIO

$$f(x) = \begin{cases} e^{-\frac{1}{|x-1|}} & -1 < x < 4,\ x \neq 1 \\ 0 & x = 1 \\ 10 & \text{ALTRIMENTI} \end{cases}$$

$$\lim_{x \to 1} f(x) = 0 = f(1) \Rightarrow f \text{ CONTINUA IN } 1$$

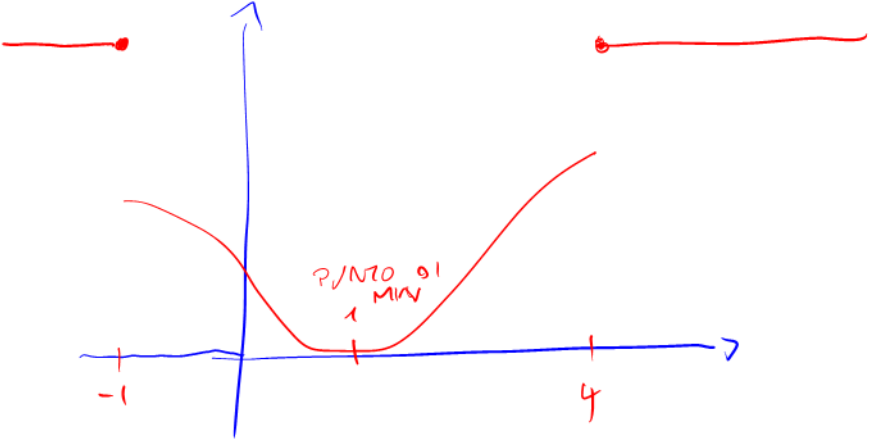
$$\lim_{x \to -1} f(x) < 10, \quad \lim_{x \to 4} f(x) < 10 \Rightarrow f \text{ DISCONTINUA IN } -1, 4$$

$$f'(x) = \begin{cases} e^{-\frac{1}{x-1}} \dfrac{1}{(x-1)^2} & x > 1 \\ -e^{\frac{1}{x-1}} \dfrac{1}{(x-1)^2} & x < 1 \end{cases}$$

$f'(x) \approx 0$ PER x VICINO A 1

$\Rightarrow f$ "PIATTA" VICINO A $x = 1$

— ESEMPIO



Selezione di Portafoglio

- Problema come ripartire l'investimento tra attività rischiose (e non rischiose)?
- Siano $R_1, \ldots, R_n$ i rendimenti (semplici) aleatori di n attività rischiose
  - ▷ $E[R_i] = \mu_i$ rendimento atteso, $\sigma_{i,j} = \mathrm{COV}[R_i, R_j]$ covarianza tra rendimenti, $i, j = 1, \ldots, n$
  - ▷ μ vettore dei rendimenti attesi, Σ matrice varianza-covarianza

$$\begin{bmatrix} R_1 \\ \vdots \\ R_n \end{bmatrix}$$

$$\mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \ldots & \sigma_{1n} \\ \vdots & \vdots & \ldots, \ldots & \\ \sigma_{n1} & \sigma_{n2} & \ldots & \sigma_{nn} \end{bmatrix}$$

. . . SELEZIONE DI PORTAFOGLIO

$x_1 \to R_1$

$x_n \to R_n$

- Obbiettivo: ripartire la ricchezza fra le attività rischiose in maniera ottimale (secondo un certo criterio)

  $1 - \sum x_i \to r$

  - ▷ x_i = percentuale di ricchezza investita nell'attività i
  - ▷ $1 - \sum_{i=1}^{n} x_i$ = percentuale di ricchezza investita nell'attività priva di rischio, con rendimento certo r
- rendimento di portafoglio $R_P = \sum_{i=1}^{n} x_i R_i + (1 - \sum_{i=1}^{n} x_i) r$
- media e varianza del rendimento di portafoglio:

T = TRASPOSTA

$$E[R_P] = x^{\mathsf{T}} \mu + (1 - x^{\mathsf{T}} 1) r$$
$$\mathrm{VAR}[R_P] = x^{\mathsf{T}} \Sigma x$$

dove $1^{\mathsf{T}} = [1, \ldots, 1]$

. . . Selezione di Portafoglio

- Modello Media-Varianza di Markowitz (1952): fissato il rendimento atteso $\overline{E}$, scegliere il portafoglio con varianza (rischio?) minima

$$\min_{x \in \mathbb{R}^n} \mathrm{VAR}[R_P], \qquad E[R_P] = \overline{E}$$

cioè

$$\min_{x \in \mathbb{R}^n} x^\top \Sigma x, \qquad x^\top \mu + (1 - x^\top 1) r = \overline{E}$$

problema di ottimizzazione quadratica (FUNZIONE QUADRATICA VINCOLI LINEARI)

- al variare di $\overline{E} \geq r$ si descrive la frontiera efficiente dei portafogli ottimi
- possibili vincoli aggiuntivi:
  - ▷ $x_i \geq 0$: no vendita allo scoperto dell'attività i
  - ▷ $\sum_{i=1}^n x_i \leq 1$: non ci si può indebitare
  - ▷ $\sum_{i=1}^n x_i = 1$: non c'è attività priva di rischio

FITTING DI UNA CURVA

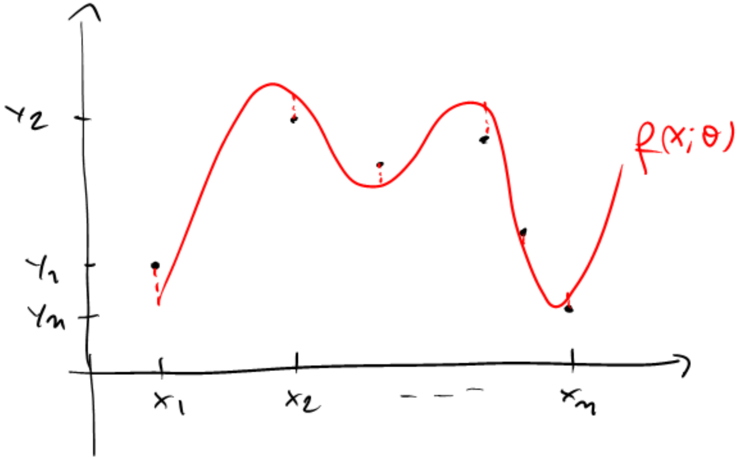
- Dati $(x_i, y_i) \in \mathbb{R}^{h+1}$, per $i = 1, \ldots, k$, trovare una funzione $f : \mathbb{R}^h \to \mathbb{R}$ tale che $f(x_i)$ sia "vicino" a y_i, per ogni $i \rightsquigarrow$ perequazione, interpolazione, estrapolazione
- la funzione f viene scelta in una famiglia parametrica: $f(x) \equiv f(x;\theta)$ con $\theta \in \Theta \subset \mathbb{R}^n$
- misurando la "vicinanza" con la distanza Euclidea, il problema è

$$\min_{\theta\in\Theta} \sum_{i=1}^{h} (y_i - f(x_i;\theta))^2 \quad \text{o} \quad \min_{\theta\in\Theta} \sum_{i=1}^{h} w_i (y_i - f(x_i;\theta))^2$$

dove w_i sono pesi

  - ▷ $f(x_i;\theta) = \sum_{i=1}^{k} \theta_i x_i \rightsquigarrow$ regressione lineare
  - ▷ funzioni non lineari (in θ)
  - ▷ splines polinomiali o di altro tipo
  - ▷ …

FITTING DI UNA CURVA



... FITTING DI UNA CURVA

- Esempio 1: fitting della curva dei tassi - Modello di Nelson-Siegel
  - ▷ $f(0,t)$ tasso forward istantaneo;

$$f(0,t) = \beta_0 + \beta_1 \mathrm{e}^{-t/a} + \beta_2 \frac{t}{a} \mathrm{e}^{-t/a}$$

FATTORE DI SCONTO

$v(0,t) = e^{-\int_0^t f(0,u)du}$

con $\theta^\mathsf{T} = [\beta_0, \beta_1, \beta_2, a] \in \mathbb{R}^3 \times \mathbb{R}_+$

  - ▷ Dati $\widehat{P}_i$ prezzi di mercato di h obbligazioni (con o senza cedola) e $P_i^{NS}(\theta)$ i corrispondenti prezzi teorici secondo Nelson-Siegel, $i = 1, \dots, h$, si risolve il problema

$$\min_{\theta \in \Theta} \sum_{i=1}^{h} w_i (\widehat{P}_i - P_i^{NS}(\theta))^2$$

- altri modelli (Svensson), strumenti, ...

. . . FITTING DI UNA CURVA

- Esempio 2: smoothing di tassi di mortalità grezzi - modello di Heligman-Pollard
  - ▷ q_x/p_x "odds ratio"; il modello è

$$\frac{q_x}{p_x} = \underbrace{A^{(x+B)^C}}_{\text{INFANTILE}} + \overbrace{De^{-E(\log x - \log F)^2}}^{\text{MORTALITÀ ACCIDENTALE}} + \underbrace{GH^x}_{\text{ETÀ ADULTA}},$$

    con $\theta^\mathsf{T} = [A, B, C, D, E, F, G, H] \in \Theta = \mathbb{R}^8_+$
  - ▷ i tre termini riflettono le tre componenti della mortalità
  - ▷ Dati i tassi grezzi $\widehat{q}_x = \frac{D_x}{E_x}$, $x = 0, 1, \ldots$ probabilità annua di morte a età x (D_x decessi, E_x esposti al rischio), e q_x^{HP} la corrispondente probabilità secondo Heligman-Pollard, $x = 0, 1, \ldots$, si risolve il problema

$$\min_{\theta \in \Theta} \sum_x w_x (\widehat{q}_x - \widehat{q}_x^{HP})^2$$

- altri modelli (Gompertz, Makeham, Thiele, . . .)

(Sistema di) Equazioni

- La risoluzione di un'equazione o un sistema di equazioni si può impostare come problema di ottimizzazione
  - ▷ $f : D(\subset \mathbb{R}) \to \mathbb{R}$. Le soluzioni di

    $$f(x) = 0, \; x \in D$$

    sono i punti di minimo di

    $$\min_{x \in D} f(x)^2$$

  - ▷ $f : D(\subset \mathbb{R}^n) \to \mathbb{R}^n$, cioè $f^\mathsf{T} = [f_1, \ldots, f_n]$ è una funzione vettoriale. Le soluzioni di

    $$\begin{cases} f_1(x) = 0 \\ \vdots \\ f_n(x) = 0, \end{cases} \qquad x \in D$$

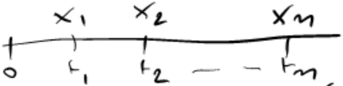
    sono i punti di minimo di

    $$\min_{x \in D} \left(f_1(x)^2 + \ldots + f_n(x)^2 \right)$$

ESEMPIO: TIR DI UNA OPERAZIONE FINANZIARIA

$x_1, \ldots, x_m$ FLUSSI FINANZIARI

IN $t_1, t_2, \ldots, t_m$

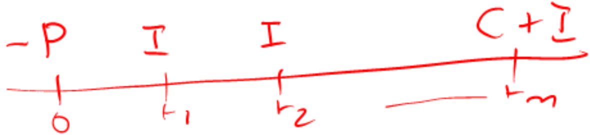


i^* TASSO INTERNO DI RENDIMENTO (TIR)

$$\underbrace{\sum_{i=1}^{n} x_i (1+i)^{-t_i}}_{VAN(i)} = 0$$

ESEMPIO: TIR

CASO PARTICOLARE: COUPON BOND

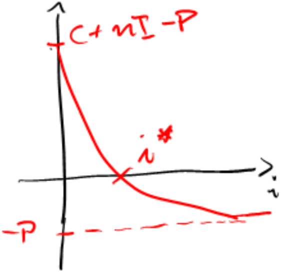


$t_i - t_{i-1} = \Delta$ (COSTANTE)

P = PREZZO DI ACQUISTO

I = CEDOLA

C = VALORE NOMINALE



SE $P < C + nI \Rightarrow$ TIR ESISTE UNICO

$$VAN(i) = -P + I \sum_{i=1}^{n} (1+i)^{-i\Delta} + C(1+i)^{-n\Delta}$$

CONDIZIONI DI OTTIMALITÀ

- $f : D(\subset \mathbb{R}^n) \to \mathbb{R}$ e x_0 interno a D
  - ▷ Gradiente di f in x_0: vettore delle derivate parziali di f (se esistono in x_0)

$$\nabla f(x_0) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x_0) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x_0) \end{bmatrix}$$

  - ▷ Hessiano di f in x_0: matrice delle derivate parziali seconde di f (se esistono in x_0)

$$H_f(x_0) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x_0) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(x_0) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x_0) \\ \vdots & \vdots & \dots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(x_0) & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x_0) & \dots & \frac{\partial^2 f}{\partial x_n^2}(x_0) \end{bmatrix}$$

Condizioni di Ottimalità

- Condizioni necessarie per punti ottimo: se x_0 interno a D è max/min relativo allora
  - primo ordine: $\nabla f(x_0) = 0$
  - secondo ordine: $H_f(x_0)$ è semi-definita negativa/positiva

$$x^\mathsf{T} H_f(x_0) x \leq / \geq 0 \ \text{ per ogni } \ x \in \mathbb{R}^n$$

- Condizioni sufficienti per punti ottimo: se x_0 interno a D e
  - primo ordine: $\nabla f(x_0) = 0$
  - secondo ordine: $H_f(x_0)$ è definita negativa/positiva

$$x^\mathsf{T} H_f(x_0) x < / > 0 \ \text{ per ogni } \ x \in \mathbb{R}^n - \{0\}$$

allora x_0 è max/min relativo

ALGORITMI PER L'OTTIMIZZAZIONE

- Molti algoritmi cercano di risolvere il sistema di equazioni

$$\nabla f(x) = 0 \Leftrightarrow \begin{cases} \frac{\partial f}{\partial x_1}(x) = 0 \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) = 0 \end{cases}$$

- Metodi basati sul gradiente $\rightsquigarrow$ approssimano ∇f per individuare la direzione di massima ascesa/discesa
- altri metodi usano solo la funzione valutata in certi punti $\rightsquigarrow$ funzioni non differenziabili
- alcuni metodi richiedono di indicare limitazioni inferiori e superiori per ognuna delle variabili

. . . ALGORITMI PER L'OTTIMIZZAZIONE

- Gli algoritmi iterativi si basano su un

INIZIO punto iniziale $y_0 \in D$ (in input)

ITERAZIONE dato $y_n \in D$, l'algoritmo genera

$$y_{n+1} = T_n(y_n) \in D$$

che "migliora" il punto precedente ($f(y_{n+1}) < f(y_n)$ per un problema di minimo)

ARRESTO quando una data condizione è verificata

- Condizione di arresto:
  - ▷ incremento relativo/assoluto rispetto al punto precedente inferiore a un dato valore
  - ▷ numero di iterazioni
  - ▷ . . .
- l'algoritmo può convergere o no, o convergere a un max/min locale

ESEMPIO: ROSENBROCK "BANANA" FUNCTION

$$f(x,y) = 100\,(x_2 - x_1^2)^2 + (1-x_1)^2$$

$f: \mathbb{R}^2 \to \mathbb{R}$ OSS: $f(x,y) \geq 0 \quad \forall (x,y)$

$$f(x,y) = 0 \iff \begin{cases} 1 - x_1 = 0 \\ x_2 = x_1^2 \end{cases} \quad \begin{cases} x_1 = 1 \\ x_2 = 1 \end{cases}$$

$(1,1)$ PUNTO DI MINIMO ASSOLUTO

GRADIENTE: $\nabla f(x,y) = \begin{bmatrix} -400 \cdot x_1 \cdot (x_2 - x_1^2) - 2(1-x_1) \\ 200(x_2 - x_1^2) \end{bmatrix}$

Statistica Inferenziale Parametrica

- X v.a., continua o discreta (o mistura): ammontare dei sinistri, numeri di sinistri, frequenze,...
  - ▷ si osservano $X_1, \ldots, X_n$ indipendenti e identicamente distribuite (iid) con la stessa distribuzione di X $\rightsquigarrow$ campione aleatorio di dimensione n
  - ▷ si sceglie una famiglia parametrica $F_X(x;\theta)$: densità $f_X(x;\theta)$ se X è continua, massa $p_X(x_i;\theta)$ se X discreta
  - ▷ $$\theta = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_d \end{bmatrix} \in \Theta$$ è il (vettore) di parametri, $\Theta \subset \mathbb{R}^d$ è lo spazio dei parametri
  - ▷ Esempio: $X \sim \mathrm{Gamma}(\alpha, \lambda)$, $\theta = \begin{bmatrix} \alpha \\ \lambda \end{bmatrix}$ e $\Theta = \{(\alpha, \lambda) | \alpha > 0, \lambda > 0\}$

. . . STIMATORI

- Statistiche e stimatori
  - ▷ Statistica: una funzione $T = T(X_1, \ldots, X_n)$; stimatore di θ: ogni statistica tale che T prende valori in Θ
  - ▷ Esempio: $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ media campionaria, uno stimatore di $E[X]$; $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ varianza campionaria, uno stimatore di $\mathrm{VAR}[X]$
- ricordiamo che:
  - ▷ prima di osservare il campione aleatorio,

$$X_1, \ldots, X_n \text{ e ogni statistica } T = T(X_1, \ldots, X_n)$$

    sono variabili aleatorie
  - ▷ dopo che l'osservazione ha avuto luogo,

$$X_1 = x_1, \ldots, X_n = x_n \text{ e } T(x_1, \ldots, x_n) = t$$

    sono numeri; $T = t$ stima t di θ
- Come trovare degli stimatori di θ?

METODO DEI MOMENTI

- uguagliare momenti empirici a momenti della popolazione $\rightsquigarrow$ risolvere per il parametro θ
  - ▷ momenti empirici: $m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ for $k = 1, 2, \ldots$ $\rightsquigarrow$ statistica
  - ▷ momenti della popolazione: $\mu_k = E[X^k]$ per $k = 1, 2, \ldots$ $\rightsquigarrow$ funzione del parametro θ, $\mu_k \equiv \mu_k(\theta)$
  - ▷ se ci sono d parametri da trovare, $\theta^\top = [\theta_1, \ldots, \theta_d]$, si uguagliano i primi d momenti empirici con i corrispondenti momenti della popolazione:

$$\begin{cases} \mu_1(\theta_1, \ldots, \theta_d) = m_1 \\ \mu_2(\theta_1, \ldots, \theta_d) = m_2 \\ \ldots \\ \mu_d(\theta_1, \ldots, \theta_d) = m_d \end{cases}$$

    $\rightsquigarrow$ sistema di d equazioni in d incognite
  - ▷ risolvendo il sistema si arriva a uno stimatore $\widehat{\theta}^\top = [\widehat{\theta}_1, \ldots, \widehat{\theta}_d]$ dipendente dai termini noti $m_1, \ldots, m_d$

ESEMPIO: GAMMA (α, λ)

$$f(x; \alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$$

$$E(X) = \frac{\alpha}{\lambda}$$

$$VAR(X) = \frac{\alpha}{\lambda^2} \qquad E(X^2) = \frac{\alpha}{\lambda^2} + \frac{\alpha^2}{\lambda^2} = \frac{\alpha(\alpha+1)}{\lambda^2}$$

METODO DEI MOMENTI

$$\begin{cases} \frac{\alpha}{\lambda} = m_1 \\ \frac{\alpha(\alpha+1)}{\lambda^2} = m_2 \end{cases} \qquad \alpha = \lambda m_1$$

$$\frac{\cancel{\lambda} m_1 (\lambda m_1 + 1)}{\lambda^{\cancel{2}}} = m_2 \Rightarrow m_1 + \frac{1}{\lambda} = \frac{m_2}{m_1}$$

$$\Rightarrow \frac{1}{\lambda} = \frac{m_2}{m_1} - m_1 = \frac{m_2 - m_1^2}{m_1} \Rightarrow \boxed{\hat{\lambda} = \frac{m_1}{m_2 - m_1^2} \qquad \hat{\alpha} = \frac{m_1^2}{m_2 - m_1^2}}$$

METODO DEI QUANTILI

- Uguagliare quantili empirici a quelli della popolazione
  - ▷ quantili empirici: statistica d'ordine di $x_1, \dots, x_n$

$$x_{(1)} \le x_{(2)} \le \dots \le x_{(n)}$$

per $0 < p < 1$, il p-esimo quantile empirico è la statistica

$$\boxed{\widehat{x}_p = (1-w)x_{(j)} + wx_{(j+1)}}$$

dove $j = \lfloor np \rfloor =$ parte intera di np e $0 \le w \le 1$; scelta classica: $w = np - j =$ parte frazionaria di np

  - ▷ quantile della popolazione: per $0 < p < 1$ è definito come

$$\boxed{x_p = F_X^{-1}(p) = \inf\{x \in \mathbb{R} : F_X(x) \ge p\}}$$

$\rightsquigarrow$ una funzione del parametro θ, $F_X^{-1}(p) \equiv F_X^{-1}(p;\theta)$; se F_X crescente allora F_X^{-1} è l'inversa di F_X e $F_X(x_p) = p$ definisce il quantile

. . . Metodo dei Quantili

- Uguagliare quantili empirici a quelli della popolazione
  - ▷ i quantili più importanti corrispondono a $p = 0.5$ (mediana), $p = 0.25$ (primo quartile), $p = 0.75$ (terzo quartile)
  - ▷ si scelgono d quantili corrispondenti a $p_1 < p_2 < \ldots < p_d$ poi si uguagliano i quantili della popolazione ai corrispondenti quantili empirici

$$\begin{cases} F_X^{-1}(p_1;\theta_1,\ldots,\theta_d) = \widehat{x}_{p_1} \\ F_X^{-1}(p_2;\theta_1,\ldots,\theta_d) = \widehat{x}_{p_2} \\ \ldots \\ F_X^{-1}(p_d;\theta_1,\ldots,\theta_d) = \widehat{x}_{p_d} \end{cases}$$

    $\leadsto$ un sistema di d equazioni in d incognite
  - ▷ risolvendo il sistema si arriva a uno stimatore $\widehat{\theta}^{\mathsf{T}} = [\widehat{\theta}_1,\ldots,\widehat{\theta}_d]$ dipendente dal termine noto $\widehat{x}_{p_1},\ldots,\widehat{x}_{p_d}$
  - ▷ se F_X è crescente allora le equazioni del sistema possono essere scritte come

$$F_X(\widehat{x}_{p_j};\theta) = p_j, j = 1\ldots,d$$

METODO DELLA MASSIMA VEROSIMIGLIANZA

- Idea: scegliere i parametri in modo che la verosimiglianza ($\equiv$ probabilità) di ottenere i valori effettivamente osservati è massima
  - ▷ X è continua con densità f_X: la verosimiglianza = densità congiunta di $X_1, \ldots, X_n$ (iid!) in funzione di θ:

$$L(\theta) \equiv L(\theta; x_1, \ldots, x_n) = f_{X_1,\ldots,X_n}(x_1, \ldots, x_n; \theta) = \prod_{i=1}^{n} f_X(x_i; \theta)$$

    si vuole risolvere $\max_{\theta \in \Theta} L(\theta) \rightsquigarrow$ difficile
  - ▷ è equivalente e più semplice massimizzare il logaritmo della verosimiglianza

$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f_X(x_i; \theta)$$

    cioè risolvere

$$\max_{\theta \in \Theta} \ell(\theta)$$

  - ▷ quando X è discreta con f_X sostituita da p_X

Confronto tra i Metodi di Stima

- il metodo dei momenti e dei quantili sono più semplici da implementare rispetto a max verosimiglianza
  - ▷ max verosimiglianza può essere complicato analiticamente quando più di un parametro è coinvolto
  - ▷ tuttavia, i computer moderni possono risolvere problemi di max verosimiglianza anche in spazi di parametri di dimensione elevata
- i metodi dei momenti e dei quantili non usano tutta l'informazione contenuta nel campione, a differenza di max verosimiglianza
- max verosimiglianza è un metodo più robusto e gode di proprietà asintotiche (campioni elevati):
  - ▷ non distorto
  - ▷ consistente: lo stimatore "si avvicina" al (vero) parametro della popolazione
  - ▷ distribuzione normale $\rightsquigarrow$ intervalli di confidenza
  - ▷ ha varianza minima $\rightsquigarrow$ miglior stimatore tra quelli non distorti

CONFRONTO TRA MODELLI STIMATI

- Avendo stimato diverse distribuzioni parametriche su un dato campione tramite max verosimiglianza, come si può scegliere quella migliore?
- Criteri di informazione: se n è il numero di osservazioni (dimensione campionaria), P è una distribuzione (modello) con d_P parametri e max verosimiglianza $\widehat{\ell}_P$
  - ▷ Criterio di Informazione di Akaike

$$\mathrm{AIC}(P) = 2d_P - 2\widehat{\ell}_P$$

  - ▷ Criterio di Informazione di Bayes

$$\mathrm{BIC}(P) = d_P \log n - 2\widehat{\ell}_P$$

  - ▷ Idea: trade-off tra bontà di adattamento e parsimonia $\rightsquigarrow$ BIC penalizza più per il numero di parametri
- Si sceglie il modello con criterio di informazione più basso