

Interpolazione in base alla teoria di Wiener-Kolmogorov

I metodi classici di interpolazione si basano sull'uso di particolari famiglie di funzioni, e i risultati dipendono dalla distribuzione dei punti di campionamento e dai valori rilevati in questi punti; in questi dati è contenuta tutta la conoscenza disponibile sulla grandezza da interpolare. Può però accadere che si vogliano utilizzare conoscenze aggiuntive a priori sull'influenza che i valori noti della funzione possono esercitare sulla variazione della funzione stessa in intorno dei punti di campionamento, per adottare procedure di interpolazione in qualche modo adattate al fenomeno in esame. Un esempio elementare già trattato è quello di medie mobili pesate con pesi che sono funzioni (in generale decrescenti) della distanza dai punti di applicazione.

Il metodo esaminato in questo paragrafo prende spunto dal metodo di stima di Wiener-Kolmogorov, che può essere enunciato nel modo seguente:

Sia $Y(x)$ un *processo stocastico* (ossia una funzione i cui valori per x fissato sono variabili aleatorie) con valore atteso $E\{Y(x)\} = 0$ e funzione di covarianza $E\{Y(x)Y(x')\} = C(x, x')$. Allora, data un'estrazione di $Y(x)$ nei punti x_1, \dots, x_n con valori y_1, \dots, y_n , lo stimatore lineare di Y in un generico punto x in funzione delle variabili $Y(x_1), \dots, Y(x_n)$ che minimizza la varianza dello scarto $E\left\{Y(x) - \hat{Y}(x)\right\}^2$ è data da $\hat{Y}(x) = \sum_{j,k} C(x, x_j) (C^{-1})_{jk} Y(x_k)$, dove la matrice C è definita da $C_{jk} = C(x_j, x_k)$. Quindi, se sono noti i valori y_1, \dots, y_n nei punti x_1, \dots, x_n , la stima $\hat{y}(x) = \sum_{j,k} C(x, x_j) (C^{-1})_{jk} y_k$ può essere interpretata come un'interpolazione dei valori y_1, \dots, y_n .

Se il processo fosse un rumore bianco, se cioè le variabili $Y(x)$ fossero incorrelate, e quindi $C(x, x') = 0$ per $x \neq x'$, $Y(x)$ non sarebbe influenzata dai valori assunti nei punti x_1, \dots, x_n , e la stima non potrebbe essere che $\hat{y}(x) = 0$, uguale al suo valore atteso. La possibilità di avere una stima non nulla è quindi legata ad una qualche influenza dei valori assunti in altri punti, che in questo caso è espressa in termini probabilistici.

La procedura sopra descritta fornisce un'interpolazione esatta dei valori campionati. Infatti $\hat{y}(x_\ell) = \sum_{j,k} C(x_\ell, x_j) (C^{-1})_{jk} y_k = y_\ell$. Se si vuole tenere conto degli errori di misura sugli y_j , si può descrivere il processo come $Z(x) = Y(x) + \nu(x)$, dove ν è un rumore bianco incorrelato con Y , con varianza σ_ν^2 . Allora, per $x \neq x'$, $C^{(ZZ)}(x, x') = C^{(ZY)}(x, x') = C^{(YY)}(x, x')$, mentre $C^{(ZZ)}(x, x) = C^{(YY)}(x, x) + \sigma_\nu^2$. La stima diventa quindi

$$\hat{y}(x) = \sum_{j,k} C^{(YZ)}(x, x_j) \left((C^{(ZZ)})^{-1} \right)_{jk} y_k = \sum_{j,k} C^{(YY)}(x, x_j) \left((C^{(YY)} + \sigma_\nu^2 \mathbf{I})^{-1} \right)_{jk} y_k,$$

che fornisce un'interpolazione approssimata.

Stima empirica della funzione di covarianza

Naturalmente il vero problema è quello di trovare un'espressione per la funzione $C(x, x')$. Infatti, a rigore, per poterne fare una stima bisognerebbe disporre di un numero, possibilmente elevato, di estrazioni del processo, mentre in realtà si ha a disposizione soltanto un campione in un numero finito, in molti casi piuttosto piccolo, di punti.

In pratica, un modo per affrontare il problema consiste nell'assumere una qualche forma di regolarità spaziale di $C(x, x')$, in modo da poter sfruttare la numerosità del campione. Una tipica assunzione è quella di *stazionarietà*, che comporta $E\{Y(x)\} = \text{cost.}$, $C(x, x') = C(|x - x'|)$. La prima condizione è in questo caso verificata, dato che $E\{Y(x)\} = 0$; la seconda è molto forte, e impone fra l'altro che la correlazione fra due punti sia isotropa, ossia non dipenda dalla direzione del vettore congiungente i due punti. E' evidente che si possono immaginare tanti casi concreti di quantità distribuite geograficamente che non rispettano questa condizione.

In ogni caso, assumendo la condizione di stazionarietà, si può dare la seguente stima di $C(r)$ per valori discreti $r_j = j\Delta r$:

$$\hat{C}(r_j) = \frac{1}{N_j} \sum_{r_j \leq |x-x'| < r_{j+1}} y(x)y(x')$$
, dove x, x' sono punti campionati e N_j è il numero di coppie la cui distanza è nella fascia $r_j \leq r < r_{j+1}$.

Per stimare poi la funzione $C(r)$, bisogna tenere conto che essa deve essere definita positiva, ossia la matrice \mathbf{C} introdotta sopra deve essere definita positiva per ogni scelta di x_1, \dots, x_n . Questa condizione è verificata da alcune famiglie di funzioni, fra le quali va cercata quella che dà la migliore interpolazione approssimata dei valori $\hat{C}(r_j)$.

Kriging

Se $E\{Z(x)\} = \mu(x) \neq 0$, ma $\mu(x)$ è una funzione nota, il problema si riconduce a dare una stima di $Y(x) = Z(x) - \mu(x)$ con il metodo sopra descritto, e porre poi $\hat{z}(x) = \hat{y}(x) + \mu(x)$.

Se invece $\mu(x)$, pur essendo trattata come una quantità deterministica, non è nota, occorre darne una stima; inoltre non è possibile dai valori misurati z_i ricavare i valori y_i corrispondenti al processo $Y(x)$, e quindi bisogna eseguire la stima direttamente su $Z(x)$.

Si consideri inizialmente il caso con μ costante. Si cerca una stima lineare $\hat{Z}(x) = \sum \lambda_i Z(x_i)$ *corretta* (o *non deviata*), ossia si richiede $E\{\hat{Z}(x)\} = E\{Z(x)\}$, da cui segue $\sum \lambda_i = 1$.

La minimizzazione di $E\left\{Z(x) - \hat{Z}(x)\right\}^2 = \sigma_Z^2(x) - 2\sum \lambda_i C(x, x_i) + \sum \lambda_i \lambda_j C(x_i, x_j)$ questa volta va fatta sotto la condizione $\sum \lambda_i = 1$; occorre quindi introdurre un moltiplicatore di Lagrange k .

Si ottiene l'equazione $\sum_i \lambda_i C(x_i, x_j) = C(x, x_j) + k$, che, unita alla condizione $\sum \lambda_i = 1$, consente di ricavare i λ_i .

Si procede poi alla ricerca della stima lineare corretta di μ , $\hat{\mu} = \sum \lambda_i^{(\mu)} Z_i$ con $\sum \lambda_i^{(\mu)} = 1$, che minimizza $E\left\{(\mu - \hat{\mu})^2\right\} = E\left\{\left(\sum \lambda_i^{(\mu)} (Z(x_i) - \mu)\right)^2\right\} = \sum \lambda_i^{(\mu)} \lambda_j^{(\mu)} C(x_i, x_j)$. Anche in questo caso si introduce un moltiplicatore di Lagrange $k^{(\mu)}$, e si ottiene l'equazione $\sum_i \lambda_i^{(\mu)} C(x_i, x_j) = k^{(\mu)}$.

Se non si assume che μ sia costante, è necessario cercare la stima in uno spazio funzionale a dimensione finita, generato da una base $\{\varphi_1, \dots, \varphi_K\}$: $\hat{\mu}(x) = \sum c_k \varphi_k(x)$; inoltre, sia per la stima

di $Z(x)$ sia per la stima di $\mu(x)$, la condizione di correttezza diventa $\mu(x) = \sum \lambda_i \mu(x_i)$, che comporta le K condizioni $\varphi_k(x) = \sum \lambda_i \varphi_k(x_i)$. Ovviamente, il numero K deve essere più piccolo del numero di punti x_i campionati, per lasciare dei gradi di libertà ai coefficienti λ_i da stimare. Per il resto, si procede come nel caso precedente, salvo che bisogna introdurre K moltiplicatori di Lagrange.

NOTA – In geostatistica vengono usate tecniche analoghe a quelle sopra illustrate, note con il nome di *kriging*, che si applicano a situazioni più generali, in cui le statistiche non sono definite per le singole variabili $Z(x)$, ma soltanto per i loro incrementi $Z(x) - Z(x')$.

Nel kriging non viene usata la funzione di covarianza, ma il *variogramma* $\gamma(x, x') = \frac{1}{2} E\{(Z(x) - Z(x'))^2\}$, che, in caso di stazionarietà, dipende solo da $h = x - x'$, o anche, in presenza di isotropia, solo da $|h|$. In questo caso, in analogia a quanto visto per la funzione di covarianza, può essere data per il variogramma una stima campionaria discretizzata:

$$\hat{\gamma}(r_j) = \frac{1}{2N_j} \sum_{r_j \leq |x-x'| < r_{j+1}} (Z(x) - Z(x'))^2.$$

In realtà c'è una relazione fra variogramma e funzione di covarianza. Se $Z(x)$ è esso stesso stazionario, e non solo i suoi incrementi, allora

$$E\{Z(x)\} = \mu = \text{costante}, \quad E\{(Z(x) - \mu)(Z(x') - \mu)\} = C(|x - x'|),$$

$$\text{e quindi } E\{(Z(x) - \mu)^2\} = E\{(Z(x') - \mu)^2\} = C(0).$$

$$\text{Pertanto } E\{(Z(x) - Z(x'))^2\} = E\{(Z(x) - \mu) - (Z(x') - \mu)\}^2 = 2C(0) - 2C(|x - x'|).$$

Di conseguenza, in questo caso $\gamma(h) = C(0) - C(h)$.

Wiener-Kolmogorov come metodo statistico di stima

Il procedimento di stima di Wiener-Kolmogorov può essere formulato in maniera più generale di quella vista sopra, in un contesto non necessariamente legato all'interpolazione. Si assuma di dover stimare una quantità s avendo a disposizione un vettore \mathbf{x} di dati, e supponendo note \mathbf{C}_{sx} , \mathbf{C}_{xx} .

Per semplicità, si assume $E\{s\} = E\{\mathbf{x}\} = 0$. Si cerca una stima lineare, $\hat{s} = \sum h_i x_i = \mathbf{h}^T \mathbf{x}$, che minimizza l'errore quadratico medio di stima, $E\{(s - \hat{s})^2\}$. Si ottiene

$$E\{(s - \mathbf{h}^T \mathbf{x})^2\} = \sigma_s^2 - 2\mathbf{C}_{sx}^T \mathbf{h} + \mathbf{h}^T \mathbf{C}_{xx} \mathbf{h}.$$

Annullando le derivate rispetto alle componenti di \mathbf{h} , si ottiene $\mathbf{C}_{xx} \mathbf{h} = \mathbf{C}_{sx}$, ossia $\mathbf{h} = \mathbf{C}_{xx}^{-1} \mathbf{C}_{sx}$, e quindi $\hat{s} = \mathbf{C}_{sx}^T \mathbf{C}_{xx}^{-1} \mathbf{x}$, che corrisponde alla formula di interpolazione vista sopra.

Si noti che, con il valore di \mathbf{h} ottenuto, si ha

$$E\{(s - \mathbf{h}^T \mathbf{x})^2\} = \sigma_s^2 - \mathbf{C}_{sx}^T \mathbf{C}_{xx}^{-1} \mathbf{C}_{sx} < \sigma_s^2 \quad (\mathbf{C}_{sx}^T \mathbf{C}_{xx}^{-1} \mathbf{C}_{sx} > 0, \text{ dato che } \mathbf{C}_{xx}^{-1} \text{ è definita positiva}).$$

Questo significa che, se s è correlata con \mathbf{x} , l'errore quadratico medio di stima è inferiore all'incertezza a priori di s , e quindi in effetti l'uso dei dati \mathbf{x} fornisce un'informazione utile per la stima di s .

ESERCIZIO: sia data la funzione di covarianza (stazionaria) $C(\tau) = e^{-|\tau|}$. Dati i valori campionati $f(0) = 1$, $f(1) = -2$, $f(3) = -1$, rappresentare la funzione interpolante $f(t)$ nell'intervallo $[-2,5]$; verificare che la matrice di covarianza è effettivamente definita positiva.

L'esercizio può essere esteso al caso di un rettangolo nel piano xy , campionando un certo numero di valori in alcuni punti del rettangolo, e considerando la funzione di covarianza

$$C\left(\begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} x' \\ y' \end{pmatrix}\right) = \exp\left(-\sqrt{(x-x')^2 + (y-y')^2}\right).$$

Una possibile applicazione del metodo di Wiener-Kolmogorov si ha nello studio delle *serie temporali*, che sono sequenze discrete di variabili aleatorie, sulle quali possono essere introdotti diversi tipi di modellizzazione lineare a scopo predittivo, suggeriti in generale da un'analisi statistica del fenomeno in esame. Un esempio sono i *modelli autoregressivi* (AR), la cui forma è

$$y_N = \sum_{k=1}^p a_k y_{N-k} + \varepsilon_N, \text{ dove } \{\varepsilon_N\} \text{ è un rumore incorrelato:}$$

$E\{\varepsilon_j\} = 0$; $E\{\varepsilon_j \varepsilon_k\} = \sigma_\varepsilon^2 \delta_{jk}$; $E\{\varepsilon_j y_k\} = 0$ per $k < j$ (come in precedenza si assume per semplicità $E\{y_k\} = 0$). p è detto *ordine* del modello. Di particolare interesse sono le serie stazionarie, per le quali, oltre alla condizione $E\{y_k\} = \mu$ costante, si richiede anche $E\{y_j y_k\} = C_{|j-k|}$. La condizione di stazionarietà impone dei vincoli sul comportamento della sequenza C_k .

Ad esempio, se si considera un modello AR del I ordine: $y_N = a y_{N-1} + \varepsilon_N$, si ottiene:

$$E\{y_N^2\} = a^2 E\{y_{N-1}^2\} + \sigma_\varepsilon^2, \text{ (dato che } E\{\varepsilon_N y_{N-1}\} = 0), \text{ ossia } (1-a^2)C_0 = \sigma_\varepsilon^2; \text{ inoltre}$$

$E\{y_N y_{N-k}\} = a E\{y_{N-1} y_{N-k}\} + E\{\varepsilon_N y_{N-k}\}$, ossia $C_k = a C_{k-1} \Rightarrow C_k = a^k C_0$. Quindi, dati a e σ_ε , si può ricavare C_k per ogni k .

Viceversa, se $\{y_j | j = 1, \dots, N\}$ è modellizzabile come AR1, si può stimare

$$\hat{C}_0 = \frac{1}{N} \sum_{k=1}^N y_k^2, \hat{C}_1 = \frac{1}{N-1} \sum_{k=1}^{N-1} y_k y_{k+1}, \hat{a} = \hat{C}_1 / \hat{C}_0, \hat{\sigma}_\varepsilon^2 = (1 - \hat{a}^2) \hat{C}_0.$$

Noti i C_k , è possibile utilizzarli per una stima predittiva di y_{N+1} dati y_1, \dots, y_n :

$$\begin{aligned} \hat{y}_{N+1} &= C_0 \begin{pmatrix} a & \dots & \dots & \dots & a^N \end{pmatrix} \cdot C_0^{-1} \begin{pmatrix} 1 & a & \dots & \dots & a^{N-1} \\ a & 1 & a & \dots & a^{N-2} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ a^{N-1} & a^{N-2} & \dots & \dots & 1 \end{pmatrix}^{-1} \begin{pmatrix} y_N \\ \vdots \\ \vdots \\ \vdots \\ y_1 \end{pmatrix} = \\ &= a \begin{pmatrix} 1 & \dots & \dots & \dots & a^{N-1} \end{pmatrix} \cdot \begin{pmatrix} 1 & a & \dots & \dots & a^{N-1} \\ a & 1 & a & \dots & a^{N-2} \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ a^{N-1} & a^{N-2} & \dots & \dots & 1 \end{pmatrix}^{-1} \begin{pmatrix} y_N \\ \vdots \\ \vdots \\ \vdots \\ y_1 \end{pmatrix} = a \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} y_N \\ \vdots \\ \vdots \\ \vdots \\ y_1 \end{pmatrix} = a y_N \end{aligned}$$

Quindi la predizione di y_{N+1} segue immediatamente dalla forma del modello AR1, e dipende soltanto dal dato immediatamente precedente y_N .

Analogamente, per un modello AR del II ordine stazionario, $y_N = a_1 y_{N-1} + a_2 y_{N-2} + \varepsilon_N$, si ha

$E\{y_N y_{N-k}\} = a_1 E\{y_{N-1} y_{N-k}\} + a_2 E\{y_{N-2} y_{N-k}\} + E\{\varepsilon_N y_{N-k}\}$, da cui

$$C_1 = a_1 C_0 + a_2 C_1$$

$$C_2 = a_1 C_1 + a_2 C_0$$

$$C_{N-1} = a_1 C_{N-2} + a_2 C_{N-3}$$

Quindi

$$\hat{y}_{N+1} = (C_1 \ \dots \ \dots \ \dots \ C_N) \cdot \begin{pmatrix} C_0 & C_1 & \dots & \dots & C_{N-1} \\ C_1 & C_0 & C_1 & \dots & C_{N-1} \\ \dots & & & & \dots \\ \dots & & & & \dots \\ C_{N-1} & C_{N-2} & \dots & \dots & C_0 \end{pmatrix}^{-1} \begin{pmatrix} y_N \\ \vdots \\ \vdots \\ \vdots \\ y_1 \end{pmatrix} =$$

$$= (a_1 \ a_2) \begin{pmatrix} C_0 & C_1 & \dots & \dots & C_{N-1} \\ C_1 & C_0 & C_1 & \dots & C_{N-1} \\ \dots & & & & \dots \\ \dots & & & & \dots \\ C_{N-1} & C_{N-2} & \dots & \dots & C_0 \end{pmatrix}^{-1} \begin{pmatrix} y_N \\ \vdots \\ \vdots \\ \vdots \\ y_1 \end{pmatrix} = (a_1 \ a_2) \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} y_N \\ \vdots \\ \vdots \\ \vdots \\ y_1 \end{pmatrix} = a_1 y_N + a_2 y_{N-1}$$

Aggiornamento

Si consideri ora, oltre al vettore di dati \mathbf{x} , che si suppone di dimensione N , un'ulteriore quantità misurata x_{N+1} , che si suppone correlata sia con \mathbf{x} , sia con s , e quindi fornisce un'ulteriore informazione per la stima di s . Naturalmente è possibile costituire con \mathbf{x} e x_{N+1} un vettore $\tilde{\mathbf{x}}$ di dimensione $N+1$, da utilizzare per una nuova stima di s con il metodo di Wiener-Kolmogorov. Questo modo di procedere non è però conveniente, dato che l'inversione della matrice di covarianza $\tilde{\mathbf{C}}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}$ va eseguita ex-novo, senza poter utilizzare $\mathbf{C}_{\mathbf{x}\mathbf{x}}^{-1}$ calcolata in precedenza.

In alternativa, si può calcolare $\hat{x}_{N+1} = \mathbf{C}_{x_{N+1}\mathbf{x}}^T \mathbf{C}_{\mathbf{x}\mathbf{x}}^{-1} \mathbf{x}$, e aggiungere al vettore \mathbf{x} dei dati, anziché la componente x_{N+1} , la quantità $e_{N+1} = x_{N+1} - \hat{x}_{N+1}$ (innovazione). Il vantaggio è che

$$\mathbf{C}_{e_{N+1}\mathbf{x}} = \mathbf{C}_{x_{N+1}\mathbf{x}} - \mathbf{C}_{\hat{x}_{N+1}\mathbf{x}} = \mathbf{C}_{x_{N+1}\mathbf{x}} - \mathbf{C}_{x_{N+1}\mathbf{x}}^T \mathbf{C}_{\mathbf{x}\mathbf{x}}^{-1} \mathbf{C}_{\mathbf{x}\mathbf{x}} = 0$$

(*principio di ortogonalità*: l'innovazione è incorrelata con tutti i dati disponibili in precedenza), e quindi la nuova matrice $\tilde{\mathbf{C}}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}$ ha la forma

$$\tilde{\mathbf{C}}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}} = \begin{pmatrix} \mathbf{C}_{\mathbf{x}\mathbf{x}} & 0 \\ 0 & \sigma_{e_{N+1}}^2 \end{pmatrix}, \text{ da cui } \tilde{\mathbf{C}}_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}^{-1} = \begin{pmatrix} \mathbf{C}_{\mathbf{x}\mathbf{x}}^{-1} & 0 \\ 0 & 1/\sigma_{e_{N+1}}^2 \end{pmatrix}, \text{ dove } \sigma_{e_{N+1}}^2 = \sigma_{x_{N+1}}^2 - \mathbf{C}_{x_{N+1}\mathbf{x}}^T \mathbf{C}_{\mathbf{x}\mathbf{x}}^{-1} \mathbf{C}_{x_{N+1}\mathbf{x}}.$$

Di conseguenza la nuova stima di s è

$$\hat{s}^{(N+1)} = \mathbf{C}_{x_{N+1}x}^T \mathbf{C}_{xx}^{-1} \mathbf{x} + \frac{\sigma_{se_{N+1}} e_{N+1}}{\sigma_{e_{N+1}}^2} = \hat{s}^{(N)} + \frac{\sigma_{se_{N+1}} e_{N+1}}{\sigma_{e_{N+1}}^2}$$

(dove $\sigma_{se_{N+1}} = \sigma_{sx_{N+1}} - \sigma_{s\hat{x}_{N+1}} = \sigma_{sx_{N+1}} - \mathbf{C}_{sx}^T \mathbf{C}_{xx}^{-1} \mathbf{C}_{sx}$).

Si vede che la nuova stima si ottiene dalla vecchia semplicemente aggiungendo un termine dipendente dal nuovo dato (*aggiornamento*).

In modo analogo si può vedere che il nuovo errore quadratico medio di stima si ottiene da quello vecchio semplicemente sottraendo il termine $\sigma_{se_{N+1}}^2 / \sigma_{e_{N+1}}^2$ (e quindi riducendolo ulteriormente).

Filtro di Kalman

Un caso tipico in cui può essere applicato il metodo di stima di Wiener-Kolmogorov è quello in cui il vettore dei dati è costituito da una sequenza temporale di misure (affette da errore) di una certa quantità, e, istante per istante, si vuole stimare una grandezza di cui è noto a priori un modello approssimato di evoluzione temporale. Si pensi, ad esempio, ad un aereo che viaggia con un moto approssimativamente rettilineo uniforme, con fluttuazioni causate dalla presenza del vento, di cui non è possibile fornire un modello deterministico. Questo esempio ha una rilevanza anche in ambito topografico, se si pensa, ad esempio, alla ricostruzione della traiettoria dell'aereo da una sequenza di misure eseguite con un ricevitore GPS a bordo. Alla modellizzazione statistica dell'errore di misura si affianca quella dell'incertezza nel modello di evoluzione. Inoltre, se si vuole calcolare le stime in tempo reale, è utile applicare una procedura di aggiornamento come quella sopra descritta.

ESEMPIO: per descrivere un moto approssimativamente uniforme: $s(t + \Delta t) = s(t) + u\Delta t + \varepsilon(t + \Delta t)$, dove $s(t)$ è lo spazio percorso, u la velocità (costante) e $\varepsilon(t)$ rappresenta una fluttuazione non descrivibile in termini deterministici, si può introdurre un vettore

di stato $\begin{pmatrix} s(t) \\ u(t) \end{pmatrix}$ e esprimere l'evoluzione temporale mediante un modello lineare di propagazione:

$$\begin{pmatrix} s(t + \Delta t) \\ u(t + \Delta t) \end{pmatrix} = \begin{pmatrix} 1 & \Delta t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} s(t) \\ u(t) \end{pmatrix} + \begin{pmatrix} \varepsilon^{(s)}(t + \Delta t) \\ \varepsilon^{(u)}(t + \Delta t) \end{pmatrix}. \text{ Se si misura direttamente lo spazio percorso, il}$$

modello osservativo, che esprime una relazione lineare fra l'osservabile e il vettore di stato, è

$$x(t) = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} s(t) \\ u(t) \end{pmatrix} + \varepsilon^{(x)}(t), \text{ dove } \varepsilon^{(x)}(t) \text{ è l'errore di misura.}$$

Il metodo di calcolo adottato, detto *filtro di Kalman*, è stato elaborato intorno al 1960, in relazione allo studio del moto dei satelliti artificiali e dei missili balistici. Esso consiste in una procedura ricorsiva che, utilizzando ad ogni passo sia le conoscenze a priori sull'evoluzione temporale del fenomeno sia i nuovi dati acquisiti, porta al calcolo delle covarianze e all'aggiornamento delle stime.

La deduzione delle formule, che è piuttosto elaborata, viene qui schematicamente fornita in un caso semplice: la quantità $s(t)$ è scalare, e viene osservata direttamente. Ossia

$$s(N+1) = F(N)s(N) + v^{(s)}(N+1) \\ x(N) = s(N) + v^{(x)}(N)$$

dove tutte le quantità sono scalari. Si assume per semplicità che abbiano tutte valore atteso nullo, e inoltre si assume che i $\nu^{(s)}, \nu^{(x)}$ siano incorrelati fra di loro, e incorrelati con gli s a tutti gli istanti precedenti, e che $\nu^{(x)}$ sia incorrelato con s allo stesso istante.

Si parte da N dati, $x(1), \dots, x(N)$ (misurazioni di x a N diversi istanti in successione)

Stime Wiener-Kolmogorov di s agli istanti N e $N+1$:

$$\hat{s}(N) = \sum_{j=1}^N C_{s(N)x(j)} (C_{xx}^{-1})_{ji} x(i)$$

$$\hat{s}(N+1|N) = \sum_{j=1}^N C_{s(N+1)x(j)} (C_{xx}^{-1})_{ji} x(i)$$

NOTA: dato che $E\{\nu^{(x)}(N+1)x(j)\} = 0$, segue che $C_{x(N+1)x(j)} = C_{s(N+1)x(j)}$, e quindi $\hat{x}(N+1|N) = \hat{s}(N+1|N)$

Propagazione

$$E\{s(N+1)x(j)\} = F(N)E\{s(N)x(j)\} + E\{\nu^{(s)}(N+1)x(j)\}. \quad \text{Ma} \quad E\{\nu^{(s)}(N+1)x(j)\} = 0$$

$$\Rightarrow C_{s(N+1)x(j)} = F(N)C_{s(N)x(j)} \quad \Rightarrow \hat{s}(N+1|N) = F(N)\hat{s}(N)$$

Aggiornamento (si aggiunge il dato $x(N+1)$)

$$\hat{s}(N+1) = \hat{s}(N+1|N) + K(N+1)e(N+1), \quad \text{dove} \quad e(N+1) = x(N+1) - \hat{x}(N+1|N),$$

$$K(N+1) = \frac{1}{\sigma_{e(N+1)}^2} C_{s(N+1)e(N+1)}$$

BISOGNA CALCOLARE $C_{s(N+1)e(N+1)}$, $\sigma_{e(N+1)}^2$

NOTA poiché $\hat{s}(N+1|N)$ è combinazione lineare degli $x_j, j=1, \dots, N$, e $E\{e(N+1)x(j)\} = 0$ per il principio di ortogonalità, segue che $E\{e(N+1)\hat{s}(N+1|N)\} = 0$.

$$\begin{aligned} \text{Quindi } C_{s(N+1)e(N+1)} &= E\{s(N+1)e(N+1)\} = E\{(s(N+1) - \hat{s}(N+1|N))e(N+1)\} \equiv \\ &\equiv E\{(s(N+1) - \hat{s}(N+1|N))(x(N+1) - \hat{x}(N+1|N))\} = \\ &= E\{(s(N+1) - \hat{s}(N+1|N))(s(N+1) + \nu^{(s)}(N+1) - \hat{s}(N+1|N))\} = E\{(s(N+1) - \hat{s}(N+1|N))^2\}, \end{aligned}$$

dato che, come già visto, $\hat{x}(N+1|N) = \hat{s}(N+1|N)$, e $\nu^{(x)}(N+1)$ è incorrelato sia con $s(N+1)$, sia con $x(i)$, $i=1, \dots, N$, e quindi con $\hat{s}(N+1|N)$.

$$\begin{aligned} \text{D'altra parte, } \sigma_{e(N+1)}^2 &= E\{e(N+1)^2\} = E\{(x(N+1) - \hat{x}(N+1|N))^2\} = \\ &= E\{(s(N+1) + \nu^{(x)}(N+1) - \hat{s}(N+1|N))^2\} = E\{(s(N+1) - \hat{s}(N+1|N))^2\} + \sigma_{\nu^{(x)}(N+1)}^2 \end{aligned}$$

Bisogna quindi concentrare l'attenzione sul termine

$$\begin{aligned}
E\{(s(N+1) - \hat{s}(N+1|N))^2\} &= E\{[F(N)(s(N) - \hat{s}(N)) + v^{(s)}(N+1)]^2\} = \\
&= F(N)^2 E\{(s(N) - \hat{s}(N))^2\} + \sigma_{v^{(s)}(N+1)}^2 \quad (\text{dato che } v^{(s)}(N+1) \text{ è incorrelato sia con } s(N), \text{ sia con} \\
&\text{gli } x(i), i=1, \dots, N, \text{ e quindi con } \hat{s}(N)).
\end{aligned}$$

Ora, $s(N) - \hat{s}(N)$, per il principio di ortogonalità, è incorrelato con gli $x(i)$, $i=1, \dots, N$, e quindi con $\hat{s}(N)$. Di conseguenza

$$\begin{aligned}
E\{(s(N) - \hat{s}(N))^2\} &= E\{(s(N) - \hat{s}(N))s(N)\} = E\{(s(N) - \hat{s}(N|N-1) - K(N)e(N))s(N)\} = \\
&= E\{(s(N) - \hat{s}(N|N-1))s(N)\} - K(N)E\{e(N)s(N)\} = E\{(s(N) - \hat{s}(N|N-1))^2\} - K(N)E\{e(N)s(N)\}
\end{aligned}$$

(nell'ultimo passaggio si è ancora fatto uso del principio di ortogonalità).

Mettendo insieme i vari passaggi, si giunge alla formula di ricorrenza

$$E\{(s(N+1) - \hat{s}(N+1|N))^2\} = F(N)^2 (E\{(s(N) - \hat{s}(N|N-1))^2\} - K(N)E\{e(N)s(N)\}) + \sigma_{v^{(s)}(N+1)}^2$$

che consente immediatamente di calcolare $C_{s(N+1)e(N+1)}$, $\sigma_{e(N+1)}^2$ (quest'ultima richiede anche la conoscenza di $\sigma_{v^{(x)}(N+1)}^2$), essendo noto $C_{s(N)e(N)}$, e infine $K(N+1)$, che permette l'aggiornamento; inoltre sono calcolate le varianze degli scarti $s(N+1) - \hat{s}(N+1|N)$, $s(N+1) - \hat{s}(N+1)$.

Per poter attivare il processo di ricorrenza, bisogna che siano noti $\sigma_{v^{(s)}(i)}^2$, $\sigma_{v^{(x)}(i)}^2$ per tutti gli i , e inoltre $E\{s(0)^2\} \equiv E\{(s(0) - \hat{s}(0|-1))^2\}$, dove si è posto $\hat{s}(0|-1) = 0$, che significa che la stima di s al tempo 0, prima che venga fatta qualsiasi misura, è nulla (cosa ragionevole, dato che il valore atteso di s è nullo).