## MMMPG 2006/07 Registro delle lezioni

**Lezione 1** (4/10). Introduzione al corso:

a) materiale delle prime 20 ore circa alla pag. di Flandoli: http://www2.ing.unipi.it/~d848 Materiale per esercizi: tutti i compiti d'esame degli anni precedenti, alla stessa pagina ed a quella di Gubinelli:

http://www2.ing.unipi.it/~d9615/didattica/didattica.html

b) Lo scritto verte unicamente su questo materiale. Salvo inconvenienti, si terrà un pre-scritto verso i 3/4 del corso, che sostituisce lo scritto degli appelli ufficiali.

Oggetti del calcolo delle probabilità: eventi, eventi elementari, universo degli eventi elementari; probabilità di eventi (regole minime); probabilità condizionale; formula di fattorizzazione, esemplificata su un esempio che anticipa le catene di Markov; formula di Bayes, calcolo delle probabilità a posteriori date quelle a priori e le probabilità degli effetti note le cause.

**Lezione 2** (6/10). Indipendenza tra eventi.

Variabili aleatorie discrete, alcuni simboli e formule (es.  $p_k = P(X = k)$ ,  $P(X > a) = \sum_{k > a} p_k$ ); valor medio e varianza e loro proprietà (linearità del valor medio, regola del prodotto per v.a. indipendenti, regola della varianza della somma per v.a. indipendenti,  $Var\left[aX + b\right] = a^2Var\left[X\right]$ ); deviazione standard; v.a. di Bernoulli, loro media e varianza; v.a. binomiali B(n, p), teorema di legame con le Bernoulli, interpretazione come numero di successi in n esperimenti con probabilità di successo p, esempio della banca (pag. 12, calcolo nel numero medio di clienti giornalieri e della probabilità che il numero di clienti giornalieri superi 300), grafico delle  $p_k$  per la  $B(1000, \frac{1}{5})$ , sua "concentrazione" attorno alla media 200, apprezzata anche tramite il fatto che  $\sigma \sim 12, 6$  (quindi anche  $\mu + 3\sigma \sim 240$  è molto vicino a 200 rispetto alla scala delle centinaia).

**Lezione 3** (11/10). V.a. di Poisson: legame tra binomiale e Poisson con gli eventi rari, media e varianza di una Poisson.

V.a. continue, densità di probabilità, funzione di distribuzione cumulativa, valor medio e varianza. V.a. esponenziali, media e varianza, assenza di memoria, interpretazione come tempi d'attesa o tempi di rottura senza logoramento. Somma di v.a. esponenziali indipendenti e v.a. di Erlang; dimostrazione della formula per la densità di Erlang nel caso n=2.

**Lezione 4** (13/10). Legame tra v.a. esponenziali - o Erlang - e v.a. di Poisson; esempi ed esercizi. Teorema sul minimo di v.a. esponenziali; esempi

ed esercizi. Per casa: es. 1 e 3 del 2/2/04, es. 1.1 del 25/6/04, es 1 parte 1 del 4/11/04, es. 2 del 12/12/2002.

**Lezione 5** (18/10). Introduzione alle catene di Markov: grafo (stati, transizioni), probabilità di transizione, traiettorie (storie) e loro probabilità (condizionata e non alla posizione iniziale), regole per il calcolo delle probabilità dello stato al tempo n (come somma delle probabilità delle traiettorie). Probabilità, o misura o distribuzione, *invariante* (come limite delle probabilità di stato a tempo infinito), equazioni per le probabilità invarianti

$$\pi_j = \sum_i p_{ij} \pi_i, \quad \sum_j \pi_j = 1.$$

**Lezione 6** (20/10). Bilancio di flusso:

$$(1 - p_{jj}) \pi_j = \sum_{i \neq j} p_{ij} \pi_i$$

(la massa uscente da j deve essere uguale alla massa entrante in j). Tre semplici esempi: 1) stati 1 e 2,  $p_{12} = 1$ ,  $p_{21} = 1/2$ ,  $p_{22} = 1/2$ ; 2) stati 1, 2 e 3,  $p_{12} = 1/2$ ,  $p_{13} = 1/2$ ;  $p_{23} = 1$ ,  $p_{32} = 1$ ; 3) stati 1, 2 e 3,  $p_{12} = 1/2$ ,  $p_{13} = 1/2$ ;  $p_{23} = 1/2$ ,  $p_{32} = 1/2$ .

Classificazione degli stati: definizioni di stati comunicanti, classe chiusa, classe irriducibile, stato assorbente; stato transitorio e stato ricorrente (solo per catene con un numero finito di stati). Esemplificazione nei tre esempi.

Conseguenze sulle misure invarianti: a regime, la probabilità di uno stato transitorio è zero; in ogni classe irriducibile c'è un'unica misura invariante (va estesa uguale a zero fuori); se ci sono più classi irriducibili, tutte le combinazioni convesse delle misure invarianti delle singole classi irriducibili sono anch'esse invarianti, e sono tutte. Esercizi per casa dai compiti del 2004.

**Lezione 7** (25/10). Catene di nascita e morte; caso finito; caso infinito con alcuni richiami sulle serie nimeriche a termini positivi, in particolare la serie geometrica:  $\sum_{n=0}^{\infty} \rho^n$  diverge a  $+\infty$  per  $\rho \geq 1$ , converge a  $\frac{1}{1-\rho}$  per  $0 < \rho < 1$ . Applicazione al caso di catene con  $\lambda_i = \lambda$ ,  $\mu_i = \mu$ ,  $\rho = \frac{\lambda}{\mu}$ .

Esempio n.8 ed esercizi riassuntivi sulle catene di Markov (c'è un errore di stampa, vedi l'elenco degli errori nel materiale 2005/06).

**Lezione 8** (27/10). Un esempio sulla catena di nascita e morte: coda ad uno sportello, tempo discreto (1 minuto), probabilità  $p_s$  di servizio in un minuto, probabilità  $p_a$  di un nuovo arrivo in un minuto, non più di un servizio o un arrivo in un minuto. Stati: interi non negativi, rappresentanti

il numero di persone in coda (inclusa quella in fase di servizio). Probabilità di transizione:  $\lambda_i = p_a (1 - p_s)$  per i > 0,  $\lambda_0 = p_a$ ,  $\mu_i = p_s (1 - p_a)$  per ogni i > 0. Condizioni su  $p_a$  e  $p_s$  tali che la catena raggiunga il regime stazionario.

Elementi teorici sulle catene di Markov: processi stocastici a tempo discreto, tentativo di calcolo delle probabilità delle storie tramite condizionamento interativo; definizione di proprietà di Markov e di omogeneità, ricongiungimento con lo schema del grafo e le sue regole.

Processi di Markov a salti: definizione di processo stocastico a tempo continuo, tentativo di calcolo delle probabilità delle storie tramite condizionamento interativo, emergenza di probabilità di transizione dipendenti dall'intervallo temporale.

Esercizi per casa: n. 2 e n. 5 della lista; es. 1 del 13/1/06; es. 1,i,ii del 23/6/06; cominciare a ragionare sul testo dell'es. 1 del 14/1/04 e 8/11/03.

Lezione 9 (3/11). Viene richiamata la coda a tempo discreto e si inizia lo studio di un modello alternativo, a tempo continuo. Si immagina che il servente, ogni volta che inizia un servizio, accenda un olorogio aleatorio. Quando questo squilla il servente completa il servizio e ne inizia un'altro accendendo unnuovo orologio, indipendente dal precedente (se c'è qualcuno in coda, altrimenti aspetta). Similmente, ogni volta che arriva un nuovo utente in coda, si accende un orologio aleatorio; quando squilla, arriva un nuovo utente e si accende un nuovo orologio, e così via. Tutti gli orologi sono supposti indipendenti. Se gli orologi degli arrivi sono esponenziali di parametro  $\lambda$ , per un teorema visto in precedenza si ha che il numero di utenti arrivati entro un tempo t è una v.a. di Poisson di parametro  $\lambda t$  e si dice quindi che il processo degli arrivi è un processo di Poisson di intensità  $\lambda$ .

Detto  $T_s$  il tempo aleatorio di attesa per un servizio e  $T_a$  l'intertempo di attesa tra un arrivo ed il successivo, bisogna specificare la distribuzione di probabilità di  $T_s$  e  $T_a$ . Ipotizziamo che siano esponenziali di parametri  $\lambda$  e  $\mu$  rispettivamente. Questo significa ipotizzare che tutte le attese siano senza memoria: quindi se ci poniamo in un istante qualsiasi, possiamo far ripartire gli orologi che sono attivi senza modificare la casualità del sistema. Ora tutto è specificato. Come possiamo studiare cosa avviene per tempi grandi (se si raggiunge un regime stazionario o meno, ed in caso positivo che probabilità hanno i diversi stati - numero pdi persone in coda)?

Cerchiamo di ricondurci approssimativamente ad un modello a tempo discreto, così da applicare la teoria delle catene di Markov. Prendiamo un passo temporale  $\tau$  piccolo e consideriamo gli istanti discreti  $0, \tau, 2\tau, 3\tau, \dots$ 

ecc. Il tempo ora è discreto. Indichiamo con  $X_n$  il numero di persone in coda all'istante  $n\tau$ . Gli **stati** della catena sono gli interi  $k \geq 0$ : in coda possono esserci 0,1,2,... utenti. Ricordiamo però che il concetto di stato prevede di essere esaustivo: se conosciamo lo stato al tempo n possiamo effettuare senza ambiguità le previsioni circa la successiva transizione. Sapere solo k (numero di utenti in coda) esaurisce le informazioni di cui abbiamo bisogno? In generale no: servirebbe sapere da quanto tempo si attende un nuovo arrivo o un nuovo servizio. Ma nell'ipotesi di tempi  $T_s$  e  $T_a$  esponenziali, il tempo trascorso non conta: possiamo in ogni momento riaggiornare gli orologi. Quindi basta sapere k.

Capiti gli stati, bisogna individuare le transizioni. Tutte le transizioni  $k_1 \to k_2$  sono possibili. Infatti, in un lasso di tempo  $\tau$  possono essere servite o possono arrivare un numero arbitrario di persone. Tuttavia vedremo che si possono considerare trascurabili le transizioni con salti maggiori di uno.

Calcoliamo approssimativamente le probabilità di transizione da k a k+1 e da k a k-1. La transizione da k a k+1 in un tempo  $\tau$  può avvenire in infiniti modi: il più semplice è che ci sia uno ed un solo nuovo arrivo e non ci siano servizi, cioè che si verifichi l'evento

$$A := \left\{ T_a^{(1)} < \tau, T_a^{(1)} + T_a^{(2)} > \tau, T_s > \tau \right\}.$$

Ma ci sono infiniti altri modi. Tuttavia gli altri richiedono che squillino più orologi entro il tempo  $\tau$ . Quando  $\tau$  è molto piccolo, vedremo tra un attimo che la probabilità p che un orologio squilli entro il tempo  $\tau$  è molto piccola, quindi la probabilità che squillino due orologi indipendenti è dell'ordine di  $p \cdot p$ , cioè enormemente più piccola di p. Per questo trascuriamo tutte le eventualità complesse (cioè che richiedono più di uno squillo) che portano k a k+1 in un tempo  $\tau$ . Quindi, approssimativamente,

$$P(X_{n+1} = k+1 | X_n = k) \sim P(A).$$

Per evitare argomentazioni inutilmente complicate, accettiamo a livello intuitivo, con un piccolo grado di approssimazione, che

$$P(A) \sim P\left(T_a < \tau, T_s > \tau\right)$$
.

Essendo  $T_a$  e  $T_s$  indipendenti, questa vale  $P\left(T_a < \tau\right) \cdot P\left(T_s > \tau\right)$ . Inoltre  $P\left(T_s > \tau\right) = e^{-\mu\tau}$  che è circa uguale ad 1, per  $\tau$  molto piccolo. Analogamente

$$P\left(T_a < \tau\right) = 1 - e^{-\lambda \tau} \sim \lambda \tau$$

per  $\tau$  molto piccolo (infatti  $e^{-\lambda \tau} = 1 + \frac{(-\lambda \tau)^1}{1!} + \frac{(-\lambda \tau)^2}{2!} + ...$ ). In conclusione, per  $\tau$  molto piccolo,

$$p_{k,k+1} \sim \lambda \tau$$
.

Analogamente  $p_{k,k-1} \sim \mu \tau$ . Per quanto detto sopra sulle transizioni che richiedono più di uno squillo,  $p_{k,k+2} \sim 0$ ,  $p_{k,k-2} \sim 0$  e così via. Quindi, approssimativamente, possiamo descrivere la nostra coda (originariamente a tempo continuo) con una catenza di Markov (a tempo discreto) di tipo nascita-morte, con probabilità di transizione  $p_{k,k+1} = \lambda \tau$ ,  $p_{k,k-1} = \mu \tau$ . A questo punto si vede che se facessimo il bilancio di flusso, il fattore  $\tau$  si semplificherebbe in ogni equazione. Lo stesso se usiamo le formule finali di nascita-morte:

$$\rho = \frac{\lambda \tau}{\mu \tau} = \frac{\lambda}{\mu}.$$

In tutti i calcoli che riguardano il comportamento a regime  $\tau$  sparisce. Per questo fatto, se usiamo la rappresentazione grafica in cui scriviamo  $\lambda$  e  $\mu$  sulle frecce di transizione (invece che  $\lambda \tau$  e  $\mu \tau$ ), otteniamo alla fine i risultati giusti.

Si noti che  $\lambda$  e  $\mu$  non sono probabilità, in particolare possono essere numeri maggiori di 1. Invece  $\lambda \tau$  e  $\mu \tau$ , per  $\tau$  sufficientemente piccolo, sono probabilità.

La lezione si conclude con la risoluzione della parte iniziale dell'esercizio 1 dell'8/11/03.

Esercizi per casa: es. 1,3,4 della lista; es. 2 del 14/1/03; es. 1 del 14/1/05; es. 1 del 14/1/05; es. 1 del 22/7/03 domande a-c (ragionare anche su d-f).

**Lezione 10** (10/11). Si definisce la matrice di transizione P per le catene di Markov, si mostra che vale  $p^{(n)} = p^{(0)}P^n$  e si interpretano le misure invarianti tramite P ( $\pi = \pi P$ , o equivalentemente  $\pi = \pi P^n$  per ogni n = 1, 2, ...).

Si riprende lo studio dei processi a tempo continuo. Si rammenta che nel tentativo di calcolare probabilità del tipo  $P(X_{t_n} = x_n, ..., X_{t_0} = x_0)$  emergevano, sotto l'ipotesi di markovianità, le probabilità di transizione a tempo continuo  $P(X_t = x | X_{t_0} = x_0)$ . Ipotizziamo l'omogenerità temporale, cioè che queste siano uguali a  $P(X_{t-t_0} = x | X_0 = x_0)$ . Vogliamo allora esaminare le probabilità di transizione (a lezione indicate con  $p_{x_0x}^{(t)}$ )

$$p_{ij}(t) := P(X_t = j | X_0 = i).$$

Introdiciamo anche la relativa matrice di transizione P(t). Il problema difficile, rispetto al tempo discreto, è la mancanza di una matrice P o di probabilità  $p_{x_0x}$ , tramite le quali calcolare ad es. le misure invarianti. La definizione

di misura invariante è sempre  $\pi = \pi P(t)$  per ogni  $t \ge 0$ , ma non è più riformulabile in un modo semplice computaizonalmente del tipo  $\pi = \pi P$ .

Si introduce allora lo sviluppo di Taylor

$$p_{ij}(t) = p_{ij}(0) + \lambda_{ij}t + o(t)$$

dove  $p_{ij}(0) = 1$  se j = i, = 0 altrimenti, o in forma matriciale

$$P(t) = I + At + o(t)$$

dove A è la matrice di elementi  $\lambda_{ij}$ . Vale, per  $j \neq i$ ,

$$\lim_{t \to 0} \frac{p_{ij}\left(t\right)}{t} = \lambda_{ij}$$

quindi chiamiamo  $\lambda_{ij}$  tasso di transizione da i a j, e chiamiamo A la matrice dei tassi di transizione. A viene anche detto generatore infinitesimale.

Si dimostra che se vale  $\pi = \pi P(t)$  per ogni  $t \ge 0$ , allora vale

$$0 = \pi A$$
.

Si enuncia (senza dimostrazione) la validità anche della affermazione inversa. Quindi il modo semplice che cercavamo di trovare  $\pi$  è l'equazione  $0 = \pi A$ .

Le equazioni di bilancio di flusso, equivalenti al sistema  $0 = \pi A$ , sono ora

$$\pi_i \sum_{j \neq i} \lambda_{ij} = \sum_{j \neq i} \pi_j \lambda_{ji}.$$

Viene esemplificato il loro uso nell'es. 1 della lista.

Viene infine esunciato il fatto che i tassi  $\lambda_{ij}$  sono i parametri degli orologi esponenziali introdotti per altra via in una lezione precedente, e che di solito si calcolano negli esempi come inversi dei tempi medi di attesa.

**Lezione 11** (15/11). Introduzione sulla filosofia della previsione (ripetizione di elementi strutturali presenti nei dati noti).

Premesso il metodo elementare di pura ripetizione e la generalizzazione astratta alla regressione multipla, si individuano due prime metodologie: la media mobile e lo smorzamento esponenziale.

Sul secondo, si discute l'espressione dei coefficienti che tendono a zero geometricamente, si introduce la riformulazione iterativa e si discute il ruolo del parametro  $\alpha$ . Il ruolo viene esemplificato nel caso di una serie fatta di

un velor medio più white noise (in cui è conveniente approssimare il valor medio, quindi prendere  $\alpha$  piccolo), ed una serie fatta da una retta più white noise (in cui conviene prendere  $\alpha$  pari circa ad uno).

Si sottolinea il fatto che la bontà della previsione si misura tramite la differenza tra i dati reali e le loro previsioni, differenza che a livello grafico si percepisce verticalmente; pertanto, due grafici identici, uno traslato dell'altro in orizzontale ( $\alpha=1$ ), somigliano molto ad occhio ma possono dare pessimi risultati di previsione (sono vicini in orizzontale ma non necessariamente in verticale).

**Lezione 12** (17/11). Equivalenza tra le due formulazioni dello smorzamento esponenziale.

Residui r(n) = p(n) - d(n). Indicatori della bontà di un metodo:  $MSE = \frac{1}{N-N_0+1} \sum_{k=N_0}^{N} r(k)^2$ ,  $\hat{\sigma}_e = \sqrt{MSE}$ ,  $MAPE = \frac{100}{N-N_0+1} \sum_{k=N_0}^{N} \left| \frac{r(k)}{d(k)} \right|$ , meno sensibile a singoli residui eccezionali. Loro uso per la determinazione automatica di  $\alpha$  (si cerca il minimo dell'indicatore scelto tra tutti gli  $\alpha \in [0,1]$ , o almeno su alcuni).

Implementazione su Excel del metodo di smorzamento esponenziale semplice: colonna dei dati, colonna delle previsioni, inizializzata alla seconda casella, colonna dei residui, indicatore; tabella dell'indicatore al variare di  $\alpha$ ; grafici dei dati e previsioni.

Media mobile: ruolo di k, complementare a quello di  $\alpha$ . Implementazione su Excel.

Metodo di smorzamento esponenziale con trend: variabili ausiliarie s(n) e m(n), interpretate geometricamente come intercetta e coefficiente angolare di una retta che descrive i dati, nel sistema di assi in cui l'asse delle ordinate è collocato al tempo n. Dati  $\rightarrow$  coppia  $(s,m) \rightarrow$  previsione. Formule (non ancora giustificate o interpretate) per l'aggiornamento iterativo di s(n) e m(n), e per il calcolo di p(n+1). Schema dell'implementazione su Excel. Bisogna determinare su parametri,  $\alpha \in \beta$ .

**Lezione 13** (22/11). Smorzamento esponenziale con trend.

Spiegazione grafica delle formule iterative e della formula per la previsione.

Tre tipi di inizializzazione: a) s(1) = d(1), m(1) = 0; b) s(2) = d(2), m(2) = d(2) - d(1); c) regressione lineare sui primi k dati d(1), ..., d(k), trovando una retta y = at + b, poi si pone s(k) = ak + b, m(k) = a. Pregi e difetti dei tre metodi.

Ruolo dei parametri  $\alpha, \beta$ . Sommariamente, è simile al caso di un parametro,

ma l'interazione non è sempre prevedibile.  $\beta$  grande offre più flessibilità se si deve permettere al metodo di cambiare inclinazione del trend. La ricerca automatica ha senso se si ritiene che nel futuro si ripeteranno le cose viste in passato; se invece ci aspettiamo determinate cose sulla base di ragionamenti esterni ai dati noti - es. economici -, può essere migliore una scelta a priori, soggettiva.

Implementazione Excel: colonne ausiliarie per s, m, attenzione a non commettere errori di allineamento orizzontale:

$$d(n) \rightarrow (s(n), m(n))$$
  
 $\rightarrow p(n+1).$ 

Tabella a due entrate,  $\alpha$ ,  $\beta$ , per un indicatore di prestazione ( $\sigma$  o altro), compilata a mano, con ricerca prima sommaria poi più precisa nelle zone migliori.

Viene sottolineata la soggettività e la necessità di spirito critico, da applicarsi in molte fasi. Viene inoltre illustrata sommariamente la filosofia di cercare soluzioni personalizzate, al di là dei metodi automatici. Come esempio, un'idea è di cercare il trend globale con un metodo di regressione, sottrarre tale trend ed applicare il metodo SE semplice. Questa idea non è iterativa, cioè non si adatta ai nuovi dati, ma la si può rendere iterativa con ulteriori correzioni.

Per casa: preparare fogli Excel sui vari metodi visti fino ad ora.

Lezione 14 (24/11). Smorzamento esponenziale con trend e stagionalità (Winters): descrizione della genesi del metodo; implementazione su Excel (in particolare, fase di inizializzazione del termine periodico tramite regressione, ed attenzione all'allineamento).

Esercizio su Markov (1/7/05), che completeremo.

**Lezione 15** (29/11). Varie precisazioni sui metodi di previsione visti, dal punto di vista pratico. Tra esse:

a) Per confrontare metodi diversi si può usare MSE o MAPE, ma con le due possibili strategie (finestra temporale naturale del metodo o finestra uniforme tra i metodi), entrambe con pregi e difetti. Ad esempio, l'inizializzazione banale del metodo SET produce sempre un cattivo risultato nei primi passi; se si valutano gli indicatori includendo tali passi, ne viene un valore ingiustamente alto dell'indicatore rispetto all'indicatore calcolato per un altro metodo in cui la finestra esclude i primi passi (es. l'inizializzazione per regressione, che non ha bisogno di aggiustamento iniziale).

b) Viene spiegata meglio l'inizializzazione di Winters. Già era stato spiegato come calcolare F sul primo periodo: si scrive una colonna col numero d'ordine dei dati, 1, 2, 3, ..., si scrive la colonna dei dati d(n), si introduce una colonna in cui si calcola la retta r(n) di regressione sul primo periodo  $(r(n) = \text{INTERCETTA}(X,Y) + \text{PENDENZA}(X,Y)^*n, n = 1, 2, ..., P$ , dove X è la colonna dei numeri d'ordine, Y dei dati), ed infine si introduce una colonna coi valori  $F(n) = \frac{d(n)}{r(n)}$ . Oggi si è spiegato come inizializzare s ed m:

$$s(P) = r(P)$$
  
 $m(P) = PENDENZA(X,Y).$ 

Avendo in possesso i dati fino al tempo n=P ed avendo effettuato l'inizializzazione suddetta, si può cominciare a prevedere il futuro, ovvero al tempo P+1, con la formula

$$p(P+1) = (s(P) + m(P)) * F(1).$$

Arrivati al tempo P+1, quindi in possesso anche del dato d(P+1), comincia l'iterazione delle variabile nascoste s, m, F: si calcola prima s(P+1) (usando d(P+1), F(1), s(P), m(P)); poi si calcola m(P+1) (usando s(P+1), s(P), m(P)) ed infine si calcola F(P+1) (usando d(P+1), s(P+1), F(1)); con tutto questo si può prevedere il valore futuro al tempo P+2:

$$p(P+2) = (s(P+1) + m(P+1)) * F(2).$$

E così via.

- c) Circa il problema della ricerca automatica di  $\alpha$ ,  $\beta$ ,  $\gamma$ , non è pensabile di esplorare in modo fine il range di tutte le variabili (se ciascuna viene divisa in dieci valori, ne risulterebbero 1000 prove). Si può ad esempio esplorare un basso insieme di terne, es. 8 terne del tipo (0.2, 0.2, 0.2), (0.2, 0.2, 0.2), (0.2, 0.8, 0.2), (0.2, 0.8, 0.8), e così via. Trovata la terna migliore, si può esplorare un basso insieme di terne attorno ad essa; se ad esempio è uscita (0.2, 0.8, 0.2), si possono esplorare le 8 terne ottenute aumentando e diminuendo di 1 ogni numero della terna. In questo modo dopo 2-3 iterazioni ci si può accontantare. Naturalmente un lavoro più lungo produce risultati più precisi e sicuri.
- d) Viene osservato che, anche se teoricamente Winters contiene in sé i petodi precedenti nel limite dei parametri che tendono a zero, in pratica questo non accade. Quindi se non c'è periodicità bisogna evitare l'uso di Winters, non basta cercare i parametri migliori confidando che escano parametri

che la escludono automaticamente. In alcuni casi l'uso di Winters forza la periodicità replicando la forma del primo periodo e producendo risultati molto peggiori che con altri metodi.

Per casa: prevedere il valore successivo delle serie storiche in rete, denominate Esempio 2, Esempio 3, Esempio 4 (file "parte 1"). Sono dati mensili di due anni, presi da rete. Usare invece l'Esempio 1 per verificare la correttezza della propria pagina Excel.

I valori successivi (presi da rete) verranno messi sulla pag. web uno ogni mezza settimana circa.

Esercizio 1 del 22/7/03: viene descritto come si trovano gli stati, le transizioni ed i tassi di transizione.

Lezione 16 (1/12, 1 ora). Esercizio 2 del 13/1/06, punto 1 (stati, transizioni e loro tassi), osservando che alle specifiche del testo va aggiunta la seguente precisazione: i pezzi che arrivando al macchinario A lo trovano occupato, vengono eliminati, cioè non esiste una coda di attesa a monte di A. Punto 3: viene data la seguente regola: per un processo a salti, se ci troviamo in uno stato j, da cui escono le frecce verso gli stati  $k_1, k_2$  con tassi  $\lambda_1, \lambda_2$ , la probabilità che avvenga la transizione a  $k_1$  (invece che a  $k_2$ ) è

$$\frac{\lambda_1}{\lambda_1 + \lambda_2}$$
.

Detti  $T_1, T_2$  i tempi di attesa dei rispettivi orologi aleatori, Questa è la probabilità  $P(T_1 < T_2)$ .

Esercizio 1 del 13/1/06, punto 2. Per svolgere questo esercizio (lasciato per casa) bisogna conoscere la seguente formula

$$\sum_{k=0}^{N} \rho^k = \frac{1 - \rho^{N+1}}{1 - \rho}.$$

Essa, tra l'altro, è alla base della formula  $\sum_{k=0}^{\infty} \rho^k = \frac{1}{1-\rho}$  per  $-1 < \rho < 1$ . Si osserva che questa formula viene usata nelle catene di nascita-morte con un numero finito di stati.

**Lezione 17** (6/12). Dimostrazione della formula della scorsa lezione,

tramite fattorizzazione nel continuo  $(f_{T_1}(t))$  indica la densità di  $T_1$ ):

$$P(T_{1} < T_{2}) = \int_{0}^{\infty} P(T_{1} < T_{2} | T_{1} = t) f_{T_{1}}(t) dt$$
$$= \int_{0}^{\infty} P(t < T_{2}) f_{T_{1}}(t) dt$$
$$= \int_{0}^{\infty} e^{-\lambda_{2} t} \lambda_{1} e^{-\lambda_{1} t} dt = \frac{\lambda_{1}}{\lambda_{1} + \lambda_{2}}.$$

Valor medio di  $X_n$  (o  $X_t$  a tempo continuo), in regime stazionario (cioè quando  $P(X_n = k) = \pi_k$ ), per modelli di nascita e morte con tassi  $\lambda$  e  $\mu$ : se i valori (stati) di  $X_n$  sono i numeri interi non negativi, vale ( $\rho = \lambda/\mu$ ):

$$E[X_n] = \sum_{k=0}^{\infty} kP(X_n = k) = \sum_{k=0}^{\infty} k\pi_k = (1 - \rho) \sum_{k=0}^{\infty} k\rho^k = \frac{\rho}{1 - \rho}.$$

Infatti si dimostra (usando il fatto che  $D\rho^k=k\rho^{k-1}$ ) che

$$\sum_{k=0}^{\infty} k \rho^k = \frac{\rho}{(1-\rho)^2}, \quad 0 < \rho < 1.$$

Esempio: numero medio di persone in un sistema di servizio (con tasso di arrivo  $\lambda$  ed un servitore con tasso di servizio  $\mu$ ); se  $\lambda << \mu$  allora  $\rho$  è piccolo e di conseguenza  $E[X_t]$  è piccolo (basso numero medio di persone in coda).

Code con c serventi (prese dalle dispense denominate "dimensionamento con processi a salti"), di tipo M/M/c. Non serve il dettaglio delle formule finali o di tutti i calcoli, ma le idee principali:

- i) quando k serventi sono attivi, con tempi  $T_1, ..., T_k \sim Exp(\mu)$  di servizio, l'istante in cui viene completato il primo dei loro servizi è descritto dalla v.a.  $T = \min(T_1, ..., T_k) \sim Exp(k\mu)$ ;
  - ii) la condizione di non esplosione diventa  $\lambda < c \cdot \mu$ .

Tempo medio di permanenza di un utente nel sistema (caso M/M/1):

$$T_{glob} = T_1 + \dots + T_N + T_{utente}$$

ma N, il numero di utenti davanti a lui già presenti, è aleatorio. Se fosse deterministico N=k, varrebbe  $E\left[T_{glob}\right]=\frac{k+1}{\mu}$ ; più precisamente,

$$E\left[T_{glob}|N=k\right] = \frac{k+1}{\mu}.$$

Fattorizzando il valor medio:

$$E[T_{glob}] = \sum_{k=0}^{\infty} E[T_{glob}|N=k]P(N=k) = \sum_{k=0}^{\infty} \frac{k+1}{\mu} \pi_k.$$

Si lascia la conclusione per esercizio.

Lezione 18 (13/12). Interpretazione del tasso  $\lambda$  di una v.a. esponenziale: numero medio di eventi nell'unità di tempo. Infatti, dette  $T_1, ..., T_n, ...$  delle copie indipendenti del tempo esponenziale di parametro  $\lambda$ , il numero  $N_t$  di eventi entro il tempo t è una v.a. di Poisson di parametro  $\lambda t$ , che ha media  $\lambda t$ ; quindi il numero medio di eventi nel tempo t è  $\lambda t$  (in un tempo unitario è  $\lambda$ ).

Questo offre la percezione immediata che  $\lambda' > \lambda''$  significa che gli eventi relativi a  $\lambda'$  sono più frequenti di quelli relativi a  $\lambda''$ . Ad esempio, se il tasso degli arrivi è  $\lambda$  e quello dei servizi è  $\mu$ , la condizione  $\lambda < \mu$  significa che nell'unità di tempo ci sono più servizi che arrivi (e questo è alla base del raggiungimento dell'equilibrio).

Esercizio. Unica coda, due serventi con tassi di servizio  $\mu' < \mu''$ . Se sono entrambi liberi, l'utente va dal primo. Descrivere con un processo di Markov a salti. Stati (numero utenti nel sistema): interi non negativi (attenzione, si legga il seguito). Transizioni: ai primi vicini. Tasso da k=0 a k=1:  $\lambda$ ; e così da ogni k a k+1. Da k=1 a k=0: non si può decidere: dipende se chi sta servendo è il primo o il secondo. Conclusione: l'impostazione non va bene. Invece andava bene se i tassi di servizio erano gli stessi.

Stati: 0, 1A, 1B, 2, 3, 4, ... Transizioni:  $0 \to 1A$ ,  $1A \to 0$ ,  $1A \to 2$ ,  $1B \to 0$ ,  $1B \to 2$ ,  $2 \to 1A$ ,  $2 \to 1B$ ,  $2 \to 3$ , e gli altri sono standard. Ora i tassi sono chiari.

Esercizio su una catena a tempo discreto con probabilità uguali da 1 in poi, diversa tra 0 ed 1. Viene precisata la soluzione esatta. Si osserva peraltro che, con un certo margine di approssimazione, è ragionevole semplificare il problema e prendere  $p_{0,1}$  uguale alle altre; si vedano ad esempio le risoluzioni dell'es. 1 del 30/1/06, es. 1 del 23/6/06.

La soluzione esatta è la seguente. Supponiamo che sia  $p_{0,1} = \lambda_0$ ,  $p_{k,k+1} = \lambda$  per ogni  $k \geq 1$ ,  $p_{k,k-1} = \mu$  per ogni  $k \geq 1$ . Vale

$$a_0 = 1$$

$$a_n = \frac{\lambda_0 \lambda^{n-1}}{\mu^n} = \frac{\lambda_0}{\lambda} \rho^n \text{ per } n \ge 1$$

quindi

$$\sum_{n=0}^{\infty} a_n = 1 + \frac{\lambda_0}{\lambda} \sum_{n=1}^{\infty} \rho^n = 1 - \frac{\lambda_0}{\lambda} + \frac{\lambda_0}{\lambda} \sum_{n=0}^{\infty} \rho^n$$
$$= 1 - \frac{\lambda_0}{\lambda} + \frac{\lambda_0}{\lambda} \frac{1}{1 - \rho}$$

da cui si possono calcolare  $\pi_0$  ecc. La condizione di stazionarietà non cambia  $(\lambda < \mu)$  ma i valori esatti di  $\pi_0$  ecc. cambiano.

Esercizio sul magazzino illimitato.

**Lezione 19** (15/12). Esercizi di preparazione per il compitino, presi dagli esercizi d'esame.

Lezion1 20 e 21 (20 e 22/12). Lezioni finali sulla previsione.

Per comodità di lettura, dividiamo in punti l'elenco degli argomenti descritti.

- Nelle lezioni precedenti abbiamo appreso l'uso dei metodi automatici MM (Media Mobile), SE (Smorzamento Esponenziale), SET (SE con Trend), W (Winters, ovvero SET con anche stagionalità). Inoltre abbiamo appreso l'uso degli indicatori MSE e MAPE.
- Teoricamente parlando, il metodo W include SE e SET se si potessero esplorare tutte le inizializzazioni possibili (nel caso di W ci riferiamo anche ad inizializzazioni non descritte nel corso, come quella banale di porre F=1 in tutto il periodo iniziale) e tutti i valori dei parametri  $\alpha, \beta, \gamma$ . Siccome questo non è praticamente operativo, conviene pensare che SET e W sono due metodi alternativi e vanno provati entrambi. In alcuni casi si verifica appunto che SET è migliore di W: questo accade tipicamente quando la periodicità è troppo lieve o assente, per cui W è obbligato a replicare la forma del primo periodo, quando questa invece non deve essere replicata. Per quanto riguarda MM e SE, si osserva negli esempi che usualmente danno risultati peggiori, salvo situazioni particolarmente stazionarie e vicine al White Noise (WN, di cui parleremo dopo). Quindi MM ed SE vanno provati ma di solito vengono poi esclusi. La competizione avviene di solito tra SET e W.
- Circa W, aggiungiamo la seguente osservazione. Se la serie data ha solo due periodi (es. 24 mesi =  $12 \cdot 2$ ), allora il parametro  $\gamma$  è ininfluente,

come si può verificare dalle formule (principalmente dal fatto che usiamo per la previsione la formula p(n+1) = (s(n) + m(n)) F(n+1-P), dove F(n+1-P) viene calcolata sulla base della sola inizializzazione e non viene mai aggiornata). Questo da un lato semplifica la ricerca dei parametri ottimali, dall'altro però conferisce meno elasticità al metodo. Questa è anche una delle ragioni del fallimento di W per serie con solo 2 periodi e poca periodicità.

- Circa l'uso degli indicatori MSE e MAPE si suggerisce di usare una finestra comune a tutti i metodi provati, ad es. la finestra degli ultimi 12 dati nel caso di serie di 24 mesi =  $12 \cdot 2$ . Circa la preferenza tra l'uno e l'altro, fermo restando che MAPE è meno sensibile agli outliers, e che però MSE è più legato a grandezze tradizionali come la deviazione standard, non c'è un criterio generale di scelta. Può naturalmente accadere che MSE sia ottimizzato da certi parametri  $(\alpha_0, \beta_0, \gamma_0)$  mentre MAPE sia ottimizzato da altri parametri  $(\alpha_1, \beta_1, \gamma_1)$ . Quando però accade che uno dei due indicatori è poco sensibile a lievi cambiamenti dei paramtri, es. il valore di MSE attorno ai parametri ottimali  $(\alpha_0, \beta_0, \gamma_0)$  varia poco, mentre l'altro, ad es. MAPE, è più sensibile e la sua terna ottimale  $(\alpha_1, \beta_1, \gamma_1)$  è vicina a  $(\alpha_0, \beta_0, \gamma_0)$ , conviene scegliere  $(\alpha_1, \beta_1, \gamma_1)$ .
- Per WN (White Noise) intendiamo qui un processo  $\varepsilon_1, \varepsilon_2, ...$  a tempo discreto fatto di v.a. indipendenti con la stessa distribuzione. A volte si specifica che deve avere media nulla e distribuzione gaussiana, ma questo non è necessario. Se una serie storica è un WN, è inutile cercare di prevederne i valori futuri accuratamente, a causa dell'indipendenza: la miglior previsione è il valor medio. Quindi MM e SE vanno bene, ripsettivamente con n elevato o  $\alpha$  basso.
- Il WN interviene concettualmente nella seguente procedura. Data Una serie  $X_1, X_2, ...$ , invece che (o meglio, oltre ad) applicare ad essa i metodi automatici visti sopra, può essere utile eseguire delle trasformazioni per ricondurla ad una serie migliore.
- Una tipica trasformazione è quella delle differenze prime:

$$D_n = X_n - X_{n-1}.$$

Questa trasfromazione ha tanti effetti, uno dei quali è di eliminare o smorzare il trend. Se ad esempio accade che la serie trovata  $D_n$  è un

WN, abbiamo scoperto che la serie originaria  $X_n$  verifica il modello ricorsivo

$$X_n = X_{n-1} + D_n$$

dove l'unica informazione utile, a livello previsivo, per  $D_n$  è la sua media  $\mu$ . Ne scaturisce il metodo di previsione

$$p_n = X_{n-1} + \mu.$$

• Questo è naturalmente un metodo piuttosto rozzo. Ma elaborando questa idea, tramite ad esempio il concetto di differenze stagionali

$$Y_n = X_n - X_{n-P}$$

ed eventualmente eseguendo in cascata (l'ordine non conta) le differenze prime e quelle stagionali

$$D_n = X_n - X_{n-1}$$
$$Y_n = D_n - D_{n-P}$$

si arriva con alcuni calcoli al modello

$$X_n = X_{n-1} + X_{n-P} - X_{n-P-1} + Y_n$$

da cui scaturisce, se si decide di considerare  $Y_n$  come un WN, il metodo previsivo

$$p_n = X_{n-1} + X_{n-P} - X_{n-P-1} + \mu.$$

• Questi metodi si possono generalizzare: essi dicono che la previsione è legata ai valori immediatamente precedenti ed a quelli intorno al periodo precedente. Quindi più in generale potremmo considerare un metodo previsivo della forma

$$p_n = \alpha X_{n-1} + \beta X_{n-2} + \gamma X_{n-P} + \delta X_{n-P-1} + \mu$$

dove i parametri  $\alpha, ..., \delta, \mu$  andrebbero cercati ottimizzando MSE o MAPE. Naturalmente si possono immaginare modelli ancora più complessi, che includono i valori precedenti ad altri istanti o addirittura variabili aleatorie diverse dalla serie  $X_n$ , dette predittori. Si arriva in questo modo ai processi AR (una classe particolare dei processi ARIMA) ed

alla regressione lineare multipla, che però non discutiamo. Limitiamoci al modello precedente. Anche così, la minimizzazione degli indicatori al variare di 5 parametri è eccessiva, quindi conviene fare tentativi con sottomodelli più leggeri, come

$$p_n = \alpha X_{n-1} + \beta X_{n-2} + \mu \tag{1}$$

oppure

$$p_n = \alpha X_{n-1} + \gamma X_{n-P} + \mu. \tag{2}$$

• Si noti che, almeno approssimativamente, se si esclude la presenza di un trend molto marcato, i valori della serie ai diversi istanti dovrebbero essere numeri dello stesso ordine di grandezza. Per questa ragione è naturale esplorare i possibili parametri  $\alpha, ..., \delta$  nelle vicinanze del vincolo

$$\alpha + \dots + \delta = 1$$
.

In particolare, per il metodo (1) il vincolo è  $\alpha + \beta = 1$ , per cui, se lo si imone esattamente, abbiamo un solo parametro invece che due. In altre parole, il metodo (1) può essere riscritto

$$p_n = \alpha X_{n-1} + (1 - \alpha) X_{n-2} + \mu.$$

Analogamente, il metodo (2) può essere riscritto

$$p_n = \alpha X_{n-1} + (1 - \alpha) X_{n-P} + \mu.$$

Va però osservato che il vincolo di somma = 1 non è una regola necessaria e stretta, per cui conviene provare anche valori vicini al vincolo.

• In tutti questi metodi che genericamente chiameremo AR (Auto Regressive) in quanto sono sottoclassi degli AR, abbiamo sostituito μ al posto di un processo più complicato Y<sub>n</sub>. Se si volessero fare tentativi più complessi, ad esempio per il fatto che nessun metodo (SET, W, AR) fornisce risultati convincenti (es. un MAPE del 50%), si può tentare la strada assai complicata di capire la struttura del processo Y<sub>n</sub> e sfruttarla. Discuteremo questo approccio solo relativamente ad un esempio.

- Come ultimo elemento pratico e teorico di appoggio alle varie indagini, descriviamo la ACF ( $Auto\ Correlation\ Function$ ). Intanto si ricordino i concetti di covarianza  $Cov\ (X,Y)$  tra due variabili aleatorie e di coefficiente di correlazione  $\rho\ (X,Y)$ , che chiameremo per brevità "correlazione". La correlazione tra due v.a. è un numero compreso tra -1 ed 1, indipendente dall'unità di misura, uguale a zero se le v.a. sono indipendenti (e se sono gaussiane vale anche il viceversa,: correlazione nulla implica indipendenza). Se delle v.a. X ed Y abbiamo dei campioni sperimentali, quindi se abbiamo due stringhe di numeri  $x_1, ..., x_n$  e  $y_1, ..., y_n$  della stessa lunghezza, si può calcolare la correlazione (empirica) tramite il comando CORRELAZIONE di Excel.
- Fatte queste premesse, descriviamo la ACF. Data una serie  $X_1, X_2, ..., X_N$ , si pone ACF(0) uguale alla correlazione tra la stringa  $X_1, X_2, ..., X_N$ e se stessa: il risultato è sempre uguale ad 1. Questo serve per porre l'unità di misura nel grafico della funzione  $k \to ACF(k)$ . Poi si pone ACF(1) uguale alla correlazione tra la stringa  $X_1, X_2, ..., X_{N-1}$  e la stringa  $X_2, X_3, ..., X_N$ . Poi si pone ACF(2) uguale alla correlazione tra la stringa  $X_1, X_2, ..., X_{N-2}$  e la stringa  $X_3, X_4, ..., X_N$ . Si prosegue così per ACF(3) ecc.. In teoria, si potrebbe proseguire fino a ACF(N-1), che però sarebbe la correlazione tra la stringa di un solo elemento  $X_1$  e la stringa  $X_N!$  Chiaramente un tale valore non ha nessuna credibilità statistica. E' quindi evidente che i valori di ACF(k) sono statisticamente significativi solo per i primi valori di k, es.  $k = 1, 2, 3, \dots$  ma non per gli ultimi  $k = N - 1, N - 2, \dots$  Non si può prescrivere in modo netto dove fermarsi, forse si può suggerire di limitarsi a k = N/2o ancor meno (alcune prescrizioni molto restrittive dicono addirittura  $k = \log N$ ). Se si traccia il grafico della funzione  $k \to ACF(k)$  (che è appunto la funzione di autocorrelazione), si nota di solito che essa scende dal valore 1 (per k=0) a valori bassi, può oscillare; ma ad un certo punto ritorna genericamente grande: questo è il segno che stiamo entrando nel range di valori di k per cui ACF(k) non è più statisticamente significativo. Si suggerisce di predisporre un forglio Excel che traccia ACF.
- La ACF si può usare per due scopi almeno. Il primo è riconoscere i potenziali periodi, magari solo vaghi: basta vedere se ci sono valori  $k_0$  per cui ACF ( $k_0$ ) è piuttosto elevato, abbastanza vicino ad 1,

o comunque più vicino ad 1 degli altri valori di k (sempre però nel range ancora statisticamente significativo). Infatti, un elevato valore di ACF(P) dichiara che le stringhe  $X_1, X_2, ..., X_{N-P}$  e la stringa  $X_{P+1}, X_3, ..., X_N$  sono molto correlate, cioè somiglianti e questa è un'indicazione di periodicità di periodo P.

- Il secondo uso della ACF è di testare se una serie è un WN. Infatti, il WN puro ha come ACF la funzione che vale 1 per k=0 e zero per ogni altro k. Sperimentalmente non verrà mai prorio zero, però un graqfico di ACF che da 1 scende repentimanente verso zero e resta molto vicino a zero (oscillando magari tra positivi e negativi), è una indicazione che la serie è approssimativamente un WN. Questo tipo di verifica serve quando si fanno differenze prime o stagionali e si vuole capire se si è trovato un WN.
- Esemplifichiamo varie cose dette con gli esempi messi in rete. Ci riferiamo alle 4 serie di 24 mesi. L'esempio 1 serve come verifica del software. L'esempio 2 si studia benissimo con W ed esemplifica il successo incredibile di W in problemi fortemente periodici.
- L'esempio 4 è molto più difficile (sono le importazioni di alimentari una certa categoria, misurate in milioni di euro, relative al periodo 1995-1996, da sito Eurostat). La visualizzazione del grafico ed una sua attenta osservazione mostrano un trend lievemente crescente e pochissima periodicità (di periodo 12, quella naturale trattandosi di dati mensili). La discussione della percezione ad occhio della periodicità si può articolare maggiormente: da un lato, non c'è un evidente disegno (pattern) che si ripete ed anzi sembra che le oscillazioni ampie dei mesi 1-5 non corrispondano alla quasi costanza dei dati 13-17, mentre la quasi costanza dei mesi 9-12 non corrisponde all'ampia variabilità dei mesi 21-24. Dall'altro, se si osserva il disegno dei punti 1-4 e quello dei punti 13-16 sono simili in forma, pur diversi in ampiezza; ed il disegno dei punti 9-12 somiglia come forma a quello dei punti 21-24. Insomma, una traccia di periodicità c'è ma non è così appariscente. L'uso meccanico del metodo SET produce tra i valori buoni dei parametri la coppia  $\alpha = 0.1, \beta = 0.1 \text{ con } MSE = 345.4 \text{ e } MAPE = 3.6\%.$  Il grafico della predizione corrispondente è una curva abbastanza rettilinea che passa in mezzo ai dati. Il valore previsto per il mese 25 è 8137.8. Circa W con periodo 12, che non dipende da  $\gamma$  per quanto detto sopra, una

rapida ricerca di buoni parametri conduce ad esempio ai valori  $\alpha = 0.1$ ,  $\beta = 0.3 \text{ con } MSE = 379.9 \text{ e } MAPE = 3.8\%$ . Quindi è meglio SET. Il grafico della previsione di W è ottimo nella parte finale ma oscilla in modo sproporzionato nella parte iniziale. Nel complesso, perde rispetto alla strategia media di SET. La previsione del mese 25 è 7915.4. E' stato poi provato un semplice modello del tipo (2), in quanto le differenze stagionali potrebbero catturare la lieve periodicità, mediano però con il valore precedente  $X_{n-1}$  per non dare alla periodicità lo stesso ruolo eccessivo dato da W (errore che è alla base della cattiva predizione dei primi mesi del secondo periodo). Le prestazioni di questo modellino AR sono buone. Addirittura sono migliori di SET se si prende la situazione paradossalmente banale  $\alpha = 0, \gamma = 1$ , in cui si deve prendere ad esempio  $\mu = 400$ : MSE = 313, MAPE = 3.6% (sempre calcolato sugli ultimi 12 valori). Il grafico della predizione ha in effetti lo stesso difeto di W, cioè oscilla troppo nella prima parte del secondo periodo, ma l'aggiustamente tramite la costante  $\mu$  permette di porre questo grafico in una posizione più media rispetto ai dati veri di quanto non facesse W. In conclusione, AR, ispirato alle differenze stagionali, è il metodo vincente. Si suggerisce, in ogni esempio, di vedere anche il grafico dei residui, nel caso questo suggerisca ulteriore struttura non ancora messa in evidenza. Tra le combinazioni di parametri interessanti si noti  $\alpha = 0.55$ ,  $\gamma = 0.45$ ,  $\mu = 0$ : MSE = 369, MAPE = 4% che compete con alcune altre precedenti. E' interessante in quanto non serve l'aggiustamento con la costante  $\mu$ , ma corrisponde più marcatamente all'idea di mediare tra la ripetizione periodica e l'aggiornamento con l'ultimo dato. Basta però reinserire  $\mu$  per fare molto meglio, ad esempio:  $\alpha = 0.55$ ,  $\gamma = 0.45$ ,  $\mu = 100$ , MSE = 315, MAPE = 3.3%. Può essere interessante approfondire questo esempio aggiungendo il termine  $\beta X_{n-2}$ . Il grafico di questa previsione è anche piuttosto istruttivo.

• L'esempio 3 è invece difficilissimo. Ad occhio, si nota un lieve trend decrescente ma non si nota alcuna forma di periodicità di periodo 12, nonostante siano dati mensili. Senza pretesa di ottimizzazione, uno dei valori migliori sempbra SET,  $\alpha=0.2$ ,  $\beta=0.3$ , MSE=2.77, MAPE=50%. Un simile MAPE è un pessimo risultato ma non è facile far meglio. Con W si ottiene ad esempio:  $\alpha=0.2$ ,  $\beta=0.4$ , MSE=4.03, MAPE=69% ed il grafico della previsione è pessimo, con ampie oscillazioni in ritardo. Con AR basato sul metodo (1),

 $\alpha=0.6,\ \beta=0.4,\ \mu=-1$  si ottiene  $MSE=3.3,\ MAPE=45\%,\ un$  po' meglio e un po' peggio di SET, ma il grafico non è incoraggiante. Tutto sommato il grafico stile regressione del SET ha un'aria più solida. L'unica cosa non ancora completamente esplorata è la struttura delle serie che si ottengono per differenza. In particolare, si osservi il grafico delle differenze terze:  $Y_n=X_n-X_{n-3}$ . Si notano delle strutture a "montagna", sempre più nitide al crescere del tempo. Il loro punto debole è che hanno a volte periodo 4 a volte 5, quindi non sono facilmente catturabili con metodi stagionali come W o differenze stagionali. Se ad esempio avessero avuto periodo 5, si poteva innescare il seguente procedimento: dai dati si calcolano le differenze terze, poi si applica W con periodo 5, si ottiene la previsoone di  $Y_n$  e da questa si risale alla previsione di  $X_n=Y_n+X_{n-3}$ . Essendo purtroppo il periodo variabile, non c'è una soluzione facile. Il problema viene lasciato come problema aperto per chi volesse cimentarsi in vista dell'esame.

- Infine, riguardo al fatto che spesso ci sono diversi metodi che forniscono risultati abbastanza vicini e, pur avendo ottimizzato gli indici non siamo così certi che il metodo che li ottimizza sia davvero migliore degli altri, si può tentare la seguente strategia, utilizzata nell'ambito delle previsioni metereologiche. Si considerano i metodi  $M_1, \ldots, M_N$  che si ritengono ragionevoli (questi possono essere metodi realmente diversi, come SET e W, o anche lo stesso metodo con diversi parametri, ad es. i parametri ( $\alpha_0, \beta_0, \gamma_0$ ) e ( $\alpha_1, \beta_1, \gamma_1$ ) che ottimizzano MSE e MAPE, oppure anche insiemi di parametri abbastanza buoni per una ragione o un'altra); per ciascuno di essi si calcola la previsione,  $p_1, \ldots, p_N$ ; si prende infine come previone finale la media  $\frac{p_1+\ldots+p_N}{N}$  (o una media pesata se sembra più opportuna). Si può anche, dalla stringa di previsioni  $p_1, \ldots, p_N$ , estrarre un istogramma di previsione e delle informazioni su intervalli di confidenza delle previsioni.
- Citiamo per concludere due argomenti facoltativi (non svolti a lezione) ma non difficili. Il primo è che la previsione dei valori futuri può essere fatta non solo per il primo valore futuro ma anche per più valori (naturalmente perdendo via via di attendibilità). Le formule per far questo si trovano sulle dispense.
- Il secondo è che la previsione di un valore futuro è un singolo numero, secco, che ovviamente differirà sempre dal valore vero; in analogia con

la teoria della stima e degli intervalli di confidenza, tale previsone è la cosidetta stima puntuale. Essa può essere arricchita di intervalli di confidenza o soglie unilatere di confidenza, cioè di dichiarazioni del tipo: il valore previsto è il seguente, e con confidenza del 95% possiamo dire che starà nell'intervallo ... Per calcolare intervalli di confidenza e soglie, supponendo per semplicità che gli errori di previsione siano gaussiani, è sufficiente conoscere la deviazione standard  $\sigma$  di questi errori. Una sua stima è l'indicatore MSE (inteso nel senso della radice quadrata della media aritmetica dei quadrati dei residui). Quindi, avendo già MSE, è molto semplice calcolare ciò che serve, tramite quantili gaussiani che Excel calcola con INV.NORM. Le formule teoriche si trovano sulle dispense.