



Balancing of Food Balance Sheets (FBSs)

Marco Garieri, Natalia Golini, Luca Pozzi

[name.cognome]@fao.org

November 8, 2013

Simulation on sample table

Testiamo l'algoritmo proposto (quindi dobbiamo modificare quello che abbiamo scritto nel paper) partendo da un FBS verosimile (sulla falsa riga di alcune righe di Italia 2010)

Item	Food	Feed	Losses	Seed	IndUse	StVar	Tot
Cereals	9230	12950	130	630	860	-350	24150
Starchy Roots	2300	190	135	155	0	-195	2975
Oilcrops	180	310	26	24	5169	258	5451
Vegetable Oils	1530	12	402	0	3	65	1882
Vegetables	12500	895	0	16	0	0	13411
Fruits	8990	0	4	0	7000	120	15874
Meat	5218	0	0	0	20	0	5239
Tot Col	39948	14357	697	825	13052	-102	68981

where

$$Tot = Food + Feed + Losses + Seed + IndUse - StVar$$

Ora supponiamo di non conoscere i valori "veri" all'interno di ogni cella, ma solo quelli relativi ai totali di riga.

L'obiettivo È quello di imputare i valori mancanti usando le informazioni che lo staff FAO sarà in grado di fornire. Queste informazioni saranno date in termini di valori attesi per ogni cella e il grado di incertezza su tale valore verrà espresso in termini di scostamento percentuale dal valore atteso.

Per valutare il comportamento dell'algoritmo proposto sono stati considerati 3 possibili scenari che rispecchiano il grado di accuratezza delle prior che lo staff FAO metter  a disposizione.

I valori attesi che fornir  lo staff FAO dovranno essere tali che per ogni riga (item) il $Tot = Food + Feed + Losses + Seed + IndUse - StVar$ sia uguale o molto prossimo a $Tot2 = Production + Imports - Exports$ (termini consolidati). La definizione di "molto prossimo"   ancora oggetto di studio.

per poter convergere i totali di riga della tabella "muTab" devo essere molto vicini ai totali di riga "veri". Quindi potremmo pensare ad un primo controllo in cui chiediamo all'algoritmo di controllare che questa assunzione sia valida prima di iniziare il running. Questo mi sembra ragionevole nella realt . La FAO non pu  darci dei valori di riga (expected values) la cui somma si discosta troppo dalla somma dei valori veri. La definizione di "molto vicini" richiede ancora qualche simulazione.

In the following table are shown the expected values hypothesized for each cell. Note how they do not differ greatly from the true value taken by each individual cell.

Expected Value	Food	Feed	Losses	Seed	IndUse	StVar	Tot	Tot2
Cereals	9210	12940	122	624	833	-344	24073	24150
Starchy Roots	2274	191	129	150	0	-175	2919	2975
Oilcrops	177	310	26	24	5169	277	5429	5451
Vegetable Oils	1527	12	402	0	4	65	1880	1882
Vegetables	12430	930	0	12	0	0	13372	13411
Fruits	9000	0	6	0	6965	90	15881	15874
Meat	5218	0	0	0	16	0	5234	5238
Tot Col	39836	14383	685	810	12987	-87	68788	68981

Scenario 1

In questo scenario le prior hanno dei bounds molto stretti. Nella tabella vengono riportati gli scostamenti dal valore atteso per ogni cella sia in termini percentuali che assoluti, ad eccezione fatta per i totali di colonna.

$\pm\%$ (absolute)	Food	Feed	Losses	Seed	IndUse	StVar
Cereals	2 (184)	5 (647)	10 (12)	2 (12)	2 (17)	10 (-34)
Starchy Roots	2 (45)	5 (10)	10 (13)	2 (3)	0	10 (-18)
Oilcrops	2 (4)	5 (16)	10 (3)	10 (2)	2 (103)	10 (28)
Vegetable Oils	2 (31)	5 (1)	10 (40)	0	10 (0)	10 (7)
Vegetables	2 (249)	2 (19)	0	10 (1)	0	0
Fruits	2 (180)	0	10 (1)	0	2 (139)	10 (9)
Meat	2 (104)	0	0	0	10 (2)	0
Tot Col	20 (7967)	20 (2877)	20 (137)	20 (162)	20 (2597)	20 (-17)

Scenario 2

In questo scenario sono state utilizzate delle prior con bounds con ampiezza quasi doppia rispetto a quelli utilizzati nello Scenario 1.

$\pm\%$ (absolute)	Food	Feed	Losses	Seed	IndUse	StVar
Cereals	5 (461)	10 (1294)	20 (24)	5 (31)	2 (17)	20 (-69)
Starchy Roots	5 (114)	10 (19)	20 (26)	5 (8)	0	20 (-35)
Oilcrops	5 (9)	10 (31)	20 (5)	20 (5)	2 (103)	20 (55)
Vegetable Oils	5 (76)	10 (1)	20 (80)	0	20 (1)	20 (13)
Vegetables	5 (622)	5 (47)	0	20 (2)	0	0
Fruits	5 (450)	0	20 (1)	0	2 (139)	20 (18)
Meat	5 (261)	0	0	0	20 (3)	0
Tot Col	20 (7967)	20 (2877)	20 (137)	20 (162)	20 (2597)	20 (-17)

Scenario 3

In this scenario, the prior limits have much more extensive than those of the Scenario 2. (TABELLA CON VALORI DA MODIFICARE)

$\pm\%$ (absolute)	Food	Feed	Losses	Seed	IndUse	StVar
Cereals	10 (921)	10 (1294)	30 (37)	5 (31)	5 (42)	30 (-103)
Starchy Roots	10 (227)	10 (19)	30 (39)	5 (8)	0	30 (-58)
Oilcrops	10 (18)	10 (31)	30 (8)	30 (7)	5 (258)	30 (83)
Vegetable Oils	10 (153)	10 (1)	30 (121)	0	30 (1)	30 (20)
Vegetables	10 (1243)	5 (47)	0	30 (4)	0	0
Fruits	10 (900)	0	30 (2)	0	5 (348)	30 (27)
Meat	10 (522)	0	0	0	30 (5)	0
Tot Col	20 (7967)	20 (2877)	20 (137)	20 (162)	20 (2597)	20 (-17)

Results

Nella tabella che segue vengono riportati i tempi di esecuzione dell'algoritmo per i 3 Scenari definiti nella precedente section. (SONO DA RICALCOLARE DOPO AVER INTEGRATO IL CODICE DI LUCA A QUELLO DEL 131108.)

Scenario	user	system	elapsed
1	84.88	2.80	127.20
2	33.49	1.62	1239.19
3	19.55	1.01	26.44

Per sintetizzare le diverse soluzioni generate dall'algoritmo di \ddot{E} pensato in prima battuta di calcolare una distribuzione di frequenza per le tabelle simulate e considerare come sintesi quella (o quelle) che veniva campionata più molte (moda della distribuzione). Purtroppo un successivo controllo sulle tabelle campionate ha mostrata la non presenza di una moda nella distribuzione: ogni tabella campionata \ddot{E} unica, non \ddot{E} uguale a nessuna delle tabelle campionate. Quindi per 100 iterazioni abbiamo 100 tabelle simulate tutte diverse tra loro. E questo vale per tutti e 3 gli scenari anche con un numero di iterazioni pari a 1000 e 10000.

Per valutare la performance dell'algoritmo proposto \ddot{E} stato calcolato il RMSE (METTEREI IL RELATIVE RMSE, ALMENO POSSIAMO CONFORNTARLO) tra i valori "veri" e quelli campionati per ciascuna tabella. Ovviamente questo indice di performance può essere calcolato solo nel caso in cui si conoscano i valori veri delle celle, e quindi solo durante uno studio di simulazione.

Appunti

- per velocizzare l'algoritmo si potrebbe pensare di ordinare le colonne della tabella "muTab" in ordine crescente di sd. Rimane esclusa da questo ordinamento "varStock" che resta

ad occupare l'ultima colonna della tabella.(FATTO DA LUCA E DA INTEGRARE CON L'ALGORIMO DEL 131108.)

- lascerei la sdev sempre pari a 100 per questi primi risultati. Una sdev così alta sostanzialmente obbliga l'algoritmo a campionare da un'uniforme discreta limitata dagli upper e lower bound che calcoliamo. In questo modo si considera un grado di incertezza alto per il valore medio che ci fornisce lo staff FAO per ogni cella. CiÚ garantisce all'algoritmo di poter campionare il "vero" valore della cella anche se quest'ultimo È molto distante al valore atteso che È stato fornito. Praticamente in questo modo si campiona anche sulle code della normale troncata.

Cose da fare

- Individuare la tabella (se esiste o esistono) che viene piú spesso campionata. Potrebbe rappresentare un indice di sintesi per le diverse tabelle campionate e quindi una soluzione unica da offrire a FAO. (FATTO E NON SONO STATE TROVATE DELLE TABELLE UGUALE NEANCHE PER 10000 ITERAZIONI, IL CHE FA SUPPORRE CHE NON LE TROVEREMO MAI, NEANCHE AUMENTANDO IL NUMERO DI ITERAZIONI. BISOGNEREBBE PENSARE AD UN MODO, CHIEDENDO ANCHE DIRETTAMENTE AD ADAM E JOSEPH,PER SINTETIZZARE I RISULTATI CHE ABBIAMO TROVATO. IN QUESTO STUDIO DI SIMULAZIONE POSSIAMO CALCOLARE IL RMSE TRA I VALORI VERI E QUELLI CAMPIONATI PER OGNI TABELLA E QUINDI FORNIRE COME SINTESI LA TABELLA (O LE PRIME 10 TABELLE) CHE HA IL RMSE PIU' BASSO. QUANDO PERÚ SIMULEREMO SENZA CONOSCERE I DATI VERI DOVREMO INVENTARCI ALTRO...)
- Calcolare un indice di bontà di adattamento quale ad esempio il RMSE tra i valori "veri" e quelli campionati.(FATTO DA LUCA)
- Giocare con la sd del totale di colonne. Tempo di esecuzione? Bontà di campionamento? (... ci sto pensando ...) (FATTO. AL MOMENTO IN TUTTI E TRE GLI SCENARI TUTTI I TOTALI DI COLONNA POSSONO VARIARE PER UN VALORE PARI AL $\pm 20\%$ DEL LORO VALORE ATTESO, DOVE QUEST'ULTIMO È CALCOLATO COME SOMMA DEI VALORI ATTESI PER CIASCUNA COLONNA. IN REALTÀ NEL CODICE HO FORZATO L'UPPER BOUND DELLA TOT DI VARSTOCK AD ESSERE PARI A ZERO E L'ALGORITMO NON FATICA MOLTO AD ACCETTARE)