Università Politecnica delle Marche

Facoltà di Economia “Giorgio Fuà ”

Corso di Laurea Triennale in Economia e Commercio

Analisi della concentrazione dei consumi mediante l'indice di Theil

Analysis of the consumptions concentration through Theil's index

Relatore: Prof. Elvio Mattioli

Rapporto Finale di: Michael Marchetti

Anno Accademico 2019-2020

Indice

Introduzione

Nel primo capitolo si farà un breve richiamo agli aspetti sociologici riguardanti le origini del consumo, ponendo attenzione alle teorie dei principali autori dell'epoca nel corso dei secoli addietro. Si andrà quindi a "toccare" la componente più astratta del consumo, riflettendo sui diversi pensieri, a volte contrastanti, degli autori dell'epoca. Nel secondo capitolo, invece, verrà fatto un excursus generale sulle condizioni ottimali del consumatore e sulla forma generale delle equazioni della domanda. Qua, ci occuperemo di studiare, in modo generale, le diverse variabili che influenzano il consumo e la domanda dei consumatori, adottando un approccio più matematico rispetto a quello discorsivo del capitolo 1. Nel terzo capitolo, dopo aver fatto un richiamo sull'eterogeneità e sull'entropia, si andrà a delineare l'indice di Theil (e le rispettive proprietà), fondamentale per misurare la concentrazione (totale, entro e tra i gruppi) dei consumi. Nel quarto ed ultimo capitolo andremo analiticamente a fare un'analisi della concentrazione dei consumi delle famiglie nelle diverse regioni italiane.

1.Le teorie sulle origini del consumo

La cultura del consumo risale alla rivoluzione commerciale. Secondo C.Mukerji è nel periodo rinascimentale che si sono formati nel commercio europeo ed internazionale dei modelli culturali di tipo materialistico per l'utilizzo dei beni che condizionano fortemente anche i nostri attuali comportamenti di consumo e che sono invece solitamente fatti risalire allo sviluppo del XVIII secolo della prima Rivoluzione industriale. Per Mukerji, tale segna il punto di passaggio tra una forma di economia ancora preindustriale, orientata prevalentemente verso l'autoconsumo da parte dei produttori, e un'economia invece industriale e capitalistica caratterizzata dalla produzione in serie di oggetti standardizzati. I modelli di consumo che tale economia utilizza provengono però dall'epoca precedente della rivoluzione commerciale. D'altronde, Braudel mostra come la nascita nel Rinascimento di una cultura del consumo abbia esercitato un ruolo fondamentale nel determinare lo sviluppo economico e sociale che i paesi europei hanno avuto nel corso degli ultimi secoli. Per lui, i cambiamenti avvenuti a partire dal Rinascimento nell'ambito dei consumi non hanno soltanto portato alla produzione di nuove merci, ma favorito lo svilupparsi di una nuova cultura sociale che sostiene lo sviluppo economico stimolando la mobilità personale e l'attitudine al cambiamento. Secondo McCracken, invece, la nascita di quella cultura materialistica su cui si regge il modello di consumo occidentale risale nell'Inghilterra della seconda metà del XVI secolo. In quest'epoca, per farsi notare dalla regina, ogni nobile ha alimentato una spirale competitiva che portò alla necessità di un elevato impiego di consumi di prestigio. Il consumo di beni di lusso era un fenomeno diffuso anche in Italia. Per McCracken, i nobili acquistarono i beni di lusso per affermare sia il prestigio e la reputazione dei loro familiari, che quello dei discendenti futuri.

Da un certo punto in poi, però, è nata l'esigenza di spendere per sé stessi per affermare la propria personalità. Così il singolo individuo sostituì il nucleo familiare come unità di consumo di base. Anche McKendrick colloca la nascita della cultura del consumo in Inghilterra, ma per lui risale alla rivoluzione industriale. Questa nascita è conseguenza del grande processo di trasformazione prodotto dalla Rivoluzione industriale. Infatti, è soltanto con l'arrivo dell'industrializzazione che il consumo è diventato un fenomeno sociale significativo perché riservato a grandi masse di persone, costrette a scegliere autonomamente i prodotti da acquistare. Dunque, per l'autore, le origini del consumo risalgono a ragioni di tipo economico, e non politico come sostiene McCracken. La nascita della cultura del consumo nell'Inghilterra del XVIII secolo è dimostrata dall'impiego per la prima volta da parte delle imprese degli strumenti del marketing e della pubblicità. Nasce, quindi, un nuovo soggetto sociale, l'impresa. In precedenza, le mode e i beni di consumo si diffondevano a partire dalle scelte effettuate dalla ristretta cerchia dell'élite aristocratica. Ora, invece, anche le donne hanno iniziato ad essere protagoniste del consumo. Per Sombart la cultura del consumo risale alla domanda di beni di lusso. Lui ha sostenuto che il capitalismo ha attraversato due distinte fasi. La prima tra il 1200 e il 1750, in cui il consumo dei beni di lusso ha permesso alla produzione industriale di avere ampi margini di profitto ed ha stimolato lo sviluppo del capitalismo. Secondo Sombart il consumo di beni di lusso era riservato al ceto aristocratico, ma progressivamente è diventato un modello da imitare per la borghesia. La seconda fase del capitalismo (capitalismo maturo), sviluppatasi a partire dalla metà del Settecento, è stata caratterizzata dall'arrivo al potere della borghesia che ha portato con sé un allargamento della base di consumo del lusso che ha abbassato il livello di raffinatezza estetica dei beni consumati. Vanni Codeluppi, invece, ha affermato che è nella logica comunicativa della vetrina, comparsa sulla scena sociale di tutti i paesi più industrializzati durante il XVIII secolo e basata sulla messa in scena spettacolare dei prodotti, che va rintracciata la nascita della cultura del consumo, mentre per Campbell è decisivo da questo punto di vista il ruolo svolto in Inghilterra dalla cultura romantica nel XVIII e nel XIX secolo. Per Jackson Lears il consumo è nato tra la fine del XIX secolo e l'inizio del secolo successivo attraverso la sostituzione dell'etica protestante con un nuovo tipo di etica che spinge gli individui a ricercare la propria soddisfazione e la propria realizzazione personale nel mondo terreno: l'etica terapeutica dell'autorealizzazione.

2.Le condizioni ottimali del consumatore e la forma generale delle equazioni della domanda

Supponiamo che ci sia un consumatore per il quale m è l'importo (positivo) dato che deve spendere per n merci; chiameremo m il suo "reddito" per comodità. I prezzi $p_1 \dots p_n$ delle n merci sono anche dati(positivi). La scelta del consumatore è quindi soggetta al vincolo di bilancio: $\Sigma\, p_i q_i = m$ (1.1)

dove le q sono le quantità delle n merci che decide di acquistare. Poiché p e m sono dati, la conoscenza delle quote di valore w_i è pienamente equivalente alla conoscenza delle quantità q_i. Anche se si suppone che il reddito e i prezzi siano dati dal punto di vista del consumatore, essi non sono realmente costanti nel tempo, poiché sono soggetti a fattori di cambiamento che sfuggono al controllo dei consumatori. Quando i prezzi e il reddito cambiano, il consumatore vorrà rivedere la sua posizione modificando le q e quindi le w. Questo significa che le quantità diventano funzioni di reddito e di prezzo, chiamate funzioni di domanda. A questo proposito è istruttivo considerare la variazione infinitesimale della quota di valore i^{th}:

$$dw_i = d(p_iq_i/m) = (q_i/m)dp_i + (p_i/m)dq_i – (p_iq_i/m^2)dm$$

che può anche essere scritta nella seguente forma più conveniente:

$$dw_i = w_id(\log p_i) + w_id(\log q_i) - w_id(\log m) \qquad (1.2)$$

La variazione della quota dell'i-esimo valore viene quindi scritta come somma ponderata di tre variazioni. La prima e la terza riguardano le variazioni di prezzo e di reddito, mentre il secondo termine è la componente quantitativa della variazione della quota dell'i-esimo valore. Infatti, formuleremo le nostre equazioni della domanda in modo tale che questo termine sia la variabile dipendente. Cominciamo con il caso in cui il reddito e i prezzi non cambiano affatto. Il problema di scelta del consumatore riguarda le n quantità q_i soggette al vincolo *(1.1)*. L'approccio convenzionale implica che il consumatore deve massimizzare una funzione di utilità:

$$u = u(q_1, q_2, \ldots, q_n) \qquad (1.3)$$

che misura la sua soddisfazione quando acquista q_1 unità della prima merce, q_2 unità della seconda, e così via.

In primo luogo, si suppone che le q possano essere variate continuamente (il che è restrittivo rispetto ai beni indivisibili).

In secondo luogo, si suppone che la funzione di utilità abbia derivate positive di primo ordine nella regione interessata:

$$\partial u / \partial q_i > 0 \qquad i = 1...n \qquad (1.4)$$

*i*n modo che non ci sia saturazione rispetto a nessuna delle n merci.

[La derivata $\partial u / \partial q_i$ è conosciuta come l'utilità marginale dell'i-esima merce.]

In terzo luogo, si assume che la funzione di utilità abbia derivate continue del secondo ordine, in modo che le derivate incrociate siano uguali secondo il teorema di Young:

$$\partial^2 u / \partial q_i \partial q_j = \partial^2 u / \partial q_j \partial q_i \qquad i,j = 1...n \qquad (1.5)$$

In quarto luogo, la funzione di utilità dovrebbe essere tale che, quando è massimizzata e soggetta al vincolo di bilancio, la soluzione consiste in un insieme unico di valori positivi delle q, per tutti i valori di reddito e di prezzo che considereremo di interesse. La condizione di massimizzazione *(1.3)* soggetta a *(1.1)* viene effettuata opportunamente per mezzo di un moltiplicatore lagrangiano λ. Pertanto, consideriamo:

$$u(q_1, q_2, \dots, q_n) - \lambda(\Sigma\, p_i q_i - m)$$

che viene differenziato rispetto a ciascuna delle q e poi messo uguale a zero. Il risultato è:

$$\partial u / \partial q_i = \lambda p_i \qquad i = 1 \dots n \qquad (1.6)$$

A sinistra abbiamo l'utilità marginale dell'i-esimo bene, che risulta così essere proporzionale al prezzo i-esimo nel punto ottimale del consumatore, essendo λ la costante di proporzionalità.

Per interpretare λ notiamo che la *(1.6)* può anche essere scritta come $\partial u / \partial(p_i q_i) = \lambda$ con $i = 1...n$, poiché p_i è una costante dal punto di vista del consumatore. In altre parole, se il reddito del consumatore aumenta di un centesimo, può aumentare il livello di utilità per unità di λ (spendendo quel centesimo per una qualsiasi delle n merci). Pertanto, il moltiplicatore Lagrangiano λ è noto come "utilità marginale del reddito". Si noti che il suo valore nel punto ottimale per il consumatore è necessariamente positivo: $\lambda > 0$ in vista della *(1,6), (1,4*), e del segno positivo di p. Combinando le n equazioni *(1.6)* con il vincolo di bilancio *(1.1)* si ottengono le n+1 condizioni ottimali, che sono esattamente sufficienti in numero per determinare le n+1 incognite del problema: le n quantità ottimali e la relativa utilità marginale del reddito. La forma generale delle equazioni della domanda è:

$$q^0_i = q_i(m, p_1, \dots, p_n) \qquad i = 1 \dots n \qquad (1.7)$$

o in notazione vettoriale: $q^0 = q(m,p)$ *(1.8)* dove q^0 è il vettore colonna delle n quantità ottimali, mentre $q(\ )$ indica la sua dipendenza dai fattori determinanti. Questi fattori sono il reddito m e i prezzi $p_1,\ldots,p_n$ che sono i coefficienti del vincolo di bilancio a cui è sottoposta la funzione di utilità. Inoltre, vale:

$$q(km,kp) = q(m,p) \qquad \text{per ogni } k > 0 \qquad (1.9)$$

Ciò deriva direttamente dal fatto che le funzioni di domanda sono ottenute massimizzando la funzione di utilità *(1.3)*, che è completamente indipendente dal reddito e dai prezzi, e quindi da k, ma è soggetta al vincolo di bilancio *(1.1)*, che non viene influenzato quando i suoi coefficienti a sinistra e a destra del segno di uguaglianza sono tutti moltiplicati per k. In generale, ogni q_i^0 dipende da tutti i prezzi $p_1,p_2,\ldots,p_n$, non solo dal prezzo p_i della stessa merce.

L'utilità marginale del reddito nel punto ottimale del consumatore è anche funzione del reddito e dei prezzi:

$$\lambda^0 = \lambda(m,p) \qquad (1.10)$$

Quindi, è omogenea di grado -1 in queste variabili, il che è di nuovo generalmente vero:

$$\lambda(km,kp) = (1/k)^* \lambda(m,p) \qquad \text{per ogni } k > 0 \qquad (1.11)$$

Questo è il risultato da *(1.14)* e *(1.7)*. Quando il reddito e i prezzi sono tutti moltiplicati per k, le quantità ottimali non vengono influenzate, lo stesso vale per la parte sinistra della *(1.7)*. Poiché p_i a destra diventa kp_i, il moltiplicatore λ (da scrivere λ^0 nella presente notazione) diventa λ^0/k.

3.Misure di diversità e concentrazione di entità economiche

3.1 Eterogeneità ed entropia

Per mettere a punto le nostre misure ci serviremo della strumentazione fornita dalla teoria dell'informazione, che richiamiamo brevemente nel seguito.

Consideriamo il problema di misurare l'**eterogeneità** di un carattere X, anche qualitativo, avente la seguente distribuzione di frequenza relativa:

$$X = \begin{cases} x_1, \ldots, x_k \\ p_1, \ldots, p_k \end{cases} \;, \; p_i \geq 0 \;, \quad i = 1, \ldots, k; \; \sum_{i=1}^{k} p_i = 1 \qquad [3.1]$$

Si ritiene ragionevole richiedere ad un indice che misuri tale entità di soddisfare le seguenti proprietà:

1) Assumere un valore pari a zero in assenza di eterogeneità, ossia nel caso di omogeneità, nel quale tutte le unità presentano la stessa modalità del carattere;

2) Assumere il suo valore massimo quando nessuna delle modalità è più frequente delle altre,

cioè massima eterogeneità.

Esistono diversi indici che soddisfano di tali proprietà, ma nel seguito, tenendo conto delle ulteriori proprietà di aggregazione e scomposizione delle quali gode, ci limiteremo al seguente:

$$H(X) = \sum_{i=1}^{k} p_i \ln \frac{1}{p_i} = -\sum_{i=1}^{k} p_i \ln p_i \qquad [3.2]$$

noto con il nome di <u>entropia</u>.

Verifichiamo che l'entropia gode delle proprietà sopra riportate.

Per quanto riguarda la prima proprietà, osserviamo che se tutte le unità presentano la stessa modalità,

ad esempio, x_i, allora $p_i = 1$ mentre $p_j = 0$ per $j \neq i$.

In tal caso, $p_i \ln p_i = 0$ e, per completamento per continuità, assegnando a $p_j \ln p_j$ il valore del limite per p_j che tende a 0,

$p_j \ln p_j = 0$

Dunque, nel caso di assenza di eterogeneità l'entropia è pari zero:

$H(X) = 0$

Consideriamo la seconda proprietà.

Nel caso di massima eterogeneità, nessuna delle modalità è più frequente delle altre:

$$p_1 = p_2 = \dots p_k = 1/k \qquad [3.3]$$

In tal caso si ha:

$$H(X) = -\sum_{i=1}^{k} \ln p_i = -\sum_{i=1}^{k} \frac{1}{k} \ln \frac{1}{k} = \sum_{i=1}^{k} \frac{1}{k} \ln k = \ln k$$

Dimostriamo che la distribuzione uniforme [3.3] massimizza <u>effettivamente</u> la [3.2].

Si tratta di un problema di massimo vincolato:

$$\begin{cases} massimzzare f(\underline{p}) = -\sum_{i=1}^{k} p_i \ln p_i \\ con\, il\, vincolo \sum_{i=1}^{k} p_i = 1 \end{cases}$$

Utilizzando il metodo dei moltiplicatori di Lagrange ed esprimendo il vincolo nella forma implicita

$$g(\underline{p}) = \sum_{i=1}^{k} p_i - 1 = 0,$$

la funzione di Lagrange è:

$$L(\underline{p}, \lambda) = f(\underline{p}) + \lambda g(\underline{p}) = -\sum_{i=1}^{k} p_i \ln p_i + \lambda\left(\sum_{i=1}^{k} p_i - 1\right)$$

Le condizioni del primo ordine sono:

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}p_i} L(\underline{p}, \lambda) = f_i'(\underline{p}) + \lambda g_i'(\underline{p}) = -\ln p_i - p_i \frac{1}{p_i} + \lambda = 0 \; i = 1, \ldots, k \\ \frac{\mathrm{d}}{\mathrm{d}\lambda} L(\underline{p}, \lambda) = \sum_{i=1}^{k} p_i - 1 = 0 \end{cases}$$

Dalle prime *k* equazioni abbiamo:

$$\ln p_i = \lambda - 1 \;\; i = 1, \ldots, k$$

quindi:

$$p_i = \exp(\lambda - 1) \qquad i = 1, \ldots, k$$

[3.6]

Sommando i primi membri della [3.6] e tenendo conto dell’ultima equazione della [3.5], cioè del vincolo:

$$\sum_{i=1}^{k} p_i = \sum_{i=1}^{k} \exp(\lambda - 1) = k \exp(\lambda - 1) = 1 \quad [3.7]$$

Dalla quale si ottiene:

$$\exp(\lambda - 1) = \frac{1}{k}; \quad \hat{\lambda} = 1 + \ln \frac{1}{k};$$

che sostituita nella [3.6] fornisce la soluzione:

$$\hat{p_i} = \exp(\hat{\lambda} - 1) = \frac{1}{k} \qquad i = 1, \ldots, k \quad [3.8]$$

I valori [3.8] soddisfano le condizioni del primo ordine, ma affinché corrispondano ad un punto di massimo occorre considerare le

condizioni del secondo ordine nel punto $\underline{\hat{p}} = \left[\frac{1}{k}\,\frac{1}{k}\,\cdots\cdot\frac{1}{k}\right]'$
Al riguardo, osserviamo che:

$$f''_{ij}(\underline{p}) = \frac{\mathrm{d}}{\mathrm{d}p_i} f'_j(\underline{p}) = \frac{\mathrm{d}}{\mathrm{d}p_i}(-\ln p_j - 1) = \begin{cases} -1/p_i \; se\, i = j \\ 0 \qquad se\, i \neq j \end{cases}$$

mentre

$$g''_{ij} = \frac{\mathrm{d}}{\mathrm{d}p_i} g'_j(\underline{p}) = 0 \qquad i, j = 1, \ldots, k$$

Pertanto la matrice della forma quadratica:

$$\sum_{i=1}^{k} \sum_{j=1}^{k} \left[f''_{ij}(\underline{\hat{p}}) + \hat{\lambda} g''_{ij}(\underline{\hat{p}}) \right] h_i h_j$$

corrisponde alla matrice hessiana della funzione *f(p):*

$$H(\underline{p}) = \left[\frac{\mathrm{d}}{\mathrm{d}p_i}\frac{\mathrm{d}}{\mathrm{d}p_j} f(\underline{p})\right] = \begin{bmatrix} -1/p_1 & 0 & \ldots & 0 \\ 0 & -1/p_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & -1/p_k \end{bmatrix}$$

che nel punto $\underline{\hat{p}}$ assume i valori:

$$H(\underline{\hat{p}}) = \begin{bmatrix} -k & 0 & \ldots & 0 \\ 0 & -k & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & -k \end{bmatrix} \quad [3.9]$$

e che, come è immediato accertare, è definita negativa.

Pertanto le [3.8] individuano il punto di massimo dell’entropia.

3.2 L'eterogeneità nei dati elementari

Consideriamo un carattere trasferibile (ad es. reddito, numero di addetti), cioè che può essere trasferito da un'unità statistica (ad es. individuo, famiglia, impresa) all'altra.

Per fissare le idee, immaginiamo che siano stati rilevati i redditi (non negativi) di un insieme di *n* inviduidi:

$X = \{x_1, \ldots, x_i, \ldots x_n\}, \ \sum_{i=1}^{n} x_i = T$ e denotiamo con:

$$Y = \{y_1, \ldots, y_i, \ldots y_n\} = \left\{\frac{x_1}{T}, \ldots, \frac{x_i}{T}, \ldots, \frac{x_n}{T}\right\} \quad [3.10]$$

la distribuzione delle **quote** di reddito detenute dai singoli individui.

La [3.10], diversamente dalla distribuzione di frequenza [3.1] (rappresentazione"modalitàfrequenze"), è una distribuzione di dati individuali o elementari (rappresentazione "modalità-unità).

Tenendo conto che:

$$y_i \geq 0, \quad i = 1, \ldots, n \quad e \quad \sum_{i=1}^{n} y_i = 1 \qquad [3.11]$$

possiamo calcolare l'entropia della distribuzione delle quote di reddito:

$$H(Y) = \sum_{i=1}^{n} y_i \ln \frac{1}{y_i} = -\sum_{i=1}^{n} y_i \ln y_i \qquad [3.12]$$

La [10.12] può quindi essere considerata come una misura dell'entropia dei redditi.

Consideriamo ora un risultato che si dimostrerà molto utile nelle elaborazioni che faremo successivamente.

3.3 Scomposizione dell'entropia

Supponiamo che il collettivo U, $|U| = n$, delle unità sia partizionato in g sottoinsiemi : $U_1 \dots U_g$.

Ciò vuol dire che ogni individuo appartiene soltanto a uno di tali sottoinsiemi. Indichiamo con: $n_j = |U_j|$ il numero di unità del sottoinsieme U_j, $j=1 \dots g$, pertanto:

$$\sum_{j=1}^{g} n_j = n \qquad [3.13]$$

E indichiamo con Y_j la quota complessivamente detenuta dalle unità che appartengono al gruppo U_j:

$$Y_j = \sum_{i \in U_j} y_i; \quad \sum_{i \in U_j} \frac{y_i}{Y_j} = 1, \qquad j = 1, \dots, g$$

In questo caso l'entropia [10.12] può essere scritta come segue:

$$H(Y) = \sum_{j=1}^{g} \left[\sum_{i \in U_j} y_i \ln \frac{1}{y_i} \right] \qquad [3.14]$$

Per quanto ci proponiamo di ottenere, ci conviene riscrivere l'espressione tra parentesi come segue:

$$\sum_{i \in U_j} \ln \frac{1}{y_i} = Y_j \sum_{i \in U_j} \frac{y_i}{Y_j} \left(\ln \frac{1/Y_j}{y_i/Y_j} \right) = Y_j \sum_{i \in U_j} \frac{y_i}{Y_j} \left(\ln \frac{1}{y_i/Y_j} + \ln \frac{1}{Y_j} \right) = Y_j \sum_{i \in U_j} \frac{y_i}{Y_j} \left(\ln \frac{1}{y_i/Y_j} \right) + Y_j \ln \frac{1}{Y_j} \sum_{i \in U_j} \frac{y_i}{Y_j} = Y_j H_j(Y) + Y_j \ln \frac{1}{Y_j}$$

[3.15]

Dove:

$$H_j(Y) = \sum_{i \in U_j} \frac{y_i}{Y_j} \left(\ln \frac{1}{y_i/Y_j} \right)$$

[3.16]

Sostituendo la [3.15] nella
[3.14]:

$$H(Y) = \sum_{j=1}^{g} Y_j H_j(Y) + \sum_{j=1}^{g} Y_j \ln \frac{1}{Y_j} = H_W(Y) + H_B(Y)$$

[3.17]

Il primo addendo del membro di destra è una media, ponderata con pesi Y_j, delle entropie **entro** i gruppi $H_j(Y)$ mentre il secondo addendo rappresenta l'entropia **tra** gruppi, essendo Y_j la quota di reddito del gruppo U_j.

3.4 Sensibilità dell'entropia ai trasferimenti.

Supponiamo che l'individuo *r* sia più povero dell'individuo *s*, $y_r < y_s$ e supponiamo che il reddito del più povero si incrementi a spese del più ricco ma in modo che il reddito totale dei due resti costante, cioè che: $y_r + y_s$ = costante.

Ci proponiamo di dimostrare che l'entropia [3.12] si incrementa sino al punto in cui $y_r = y_s$ e poi si decrementa.

Per accertare questa proprietà utilizziamo la [3.17] e supponiamo che il gruppo U_1 sia costituito dall'individuo *r* e dall'individuo *s*,

$y_r + y_s = Y_1 = costante$.

L'entropia dentro questo gruppo è:

$$H_1(Y) = \frac{y_r}{Y_1}\ln\frac{Y_1}{y_r} + \frac{y_s}{Y_1}\ln\frac{Y_1}{y_s} = x\ln\frac{1}{x} + (1-x)\ln\frac{1}{1-x} = -x\ln x - (1-x)\ln(1-x) = f(x)$$

dove per semplicità di notazione, abbiamo posto:

$$\frac{y_r}{Y_1} = x \;\; e \;\; \frac{y_s}{Y_1} = 1 - x$$

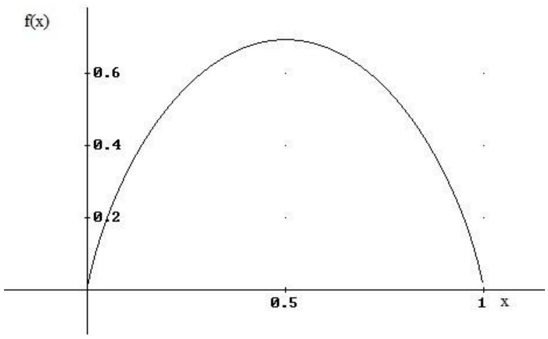
Considerando l'andamento della funzione *f(x)* osserviamo innanzitutto che: *f(0) = f(1) = 0; f(0.5) = ln2.* Tenendo poi conto che:

$$f'(x) = -\ln x - x\frac{1}{x} - \ln(1-x) - (1-x)\frac{1}{1-x}(-1) = -\ln x - 1 + \ln(1-x) + 1 =$$
$$= \ln(1-x) - \ln x = \ln\frac{1-x}{x}$$ e pertanto:

se $0<x<0.5$, $f'(x)> 0$;

se $x = 0.5$, $f'(x)= 0$

se $0.5<x<1$, $f'(x)< 0$ e quindi l'entropia ha l'andamento riportato nella figura seguente:



Inizialmente abbiamo:

$$\frac{y_r}{Y_1} = x < 0.5$$

dopo di che y_r aumenta a spese di y_s e $H_1(Y) = f(x)$

aumenta sino al punto in cui:

$$\frac{y_r}{Y_1} = x = 0.5$$

e $H_1(Y) = f(x)$ raggiunge il valore *ln 2*.

Per ulteriori aumenti di y_r l'entropia diminuisce sino ad annullarsi per

$$\frac{y_r}{Y_1} = 1 \;\; e \;\; \frac{y_s}{Y_1} = 0$$

Poiché gli altri termini a destra della [3.17] non vengono influenzati da queste variazioni, risulta che le variazioni di *H(Y)* sono le stesse di $H_1(Y)$, a meno della costante moltiplicativa Y_1.

3.5 L'indice di concentrazione di Theil e le relative proprietà

Se si vuole misurare la concentrazione occorre mettere a punto un indice che abbia un andamento opposto a quello dell'entropia. È piuttosto facile ottenere un indice di questo tipo, basta sottrarre *H(Y)* al valore massimo che può raggiungere:

$$T(Y) = \ln n - H(Y) = \sum_{i=1}^{n} y_i \ln n + \sum_{i=1}^{n} y_i \ln y_i = \sum_{i=1}^{n} y_i (\ln n + \ln y_i) =$$
$$= \sum_{i=1}^{n} y_i \ln (n y_i) = \sum_{i=1}^{n} y_i \ln \frac{y_i}{1/n} \qquad [3.19]$$

si ottiene così una misura nota come indice di concentrazione di Theil. Chiaramente, questa misura varia tra zero, nel caso di equidistribuzione e *ln n* nel caso di massima concentrazione.

La prima proprietà è l'indipendenza dall'unità di misura del carattere. In altri termini, se tutti i valori del carattere sono moltiplicati per una costante positiva il valore dell'indice deve risultare invariato. Si tratta di un requisito fondamentale di ogni indice di concentrazione, in questo caso banalmente verificato in quanto le quote sono indipendenti dall'unità di misura.

La seconda proprietà è la sensibilità ai trasferimenti. Si tratta di una conseguenza della sensibilità ai trasferimenti dell'entropia, chiaramente con andamento opposto.

La terza proprietà è che il limite superiore dell'indice, *ln n*, aumenta all'aumentare del numero *n* delle unità. Questa proprietà può destare qualche perplessità se si tiene conto che abitualmente gli gli indici di concentrazione vengono normalizzati così che il loro valore massimo è pari ad 1, quale che sia il numero delle unità considerate.

Tuttavia, non è difficile rendersi conto che tale proprietà risulta ragionevole. Vediamone la motivazione.

Quando il collettivo è formato da due sole unità, si verifica la massima disuguaglianza (o concentrazione) quando un'unità detiene la totalità del carattere e l'altra nulla. In tal caso il valore di $T(Y)$ è $\ln 2$. Quando il collettivo è formato da 2 milioni di individui, la disuguaglianza (o concentrazione) è massima quando un'unità detiene la totalità del carattere e le altre 1˙999˙999 ne detengono una quota nulla. Si tratta di una situazione nella quale la concentrazione è molto più elevata della precedente. Ciò è registrato dall'indice che assume il valore $\ln 2\dot{}000\dot{}000$.

Vediamo, più precisamente, che ragionamento conviene fare per condurre un corretto confronto.

Nel primo collettivo metà degli individui possiedono la totalità del carattere e l'altra metà nulla. È ragionevole congetturare che la concentrazione nel secondo collettivo sarebbe uguale a quella del primo quando la metà degli individui detiene la totalità del carattere e ciascuno di questi ne detiene la stessa quantità. Mostriamo che l'indice *T(Y)* soddisfa questa condizione.

Indichiamo con U l'intero collettivo di n unità: $n = |U|$ e con U_1 quello costituito dalle m unità che detengono la totalità del carattere: $m = |U_1|$ e supponiamo <u>le loro quote</u> di possesso del carattere <u>tutte</u> <u>uguali</u> e quindi pari a $1/m$. Allora:

$$T(Y) = \sum_{i=1}^{n} y_i \ln(ny_i) = \sum_{i \in U_1} y_i \ln(ny_i) = \sum_{i \in U_1} \frac{1}{m} \ln\left(n\frac{1}{m}\right) = \underbrace{=}_{\text{essendo gli addendi} =} m\frac{1}{m}\ln\left(\frac{n}{m}\right) = \ln\frac{1}{m/n} = \ln\frac{1}{\gamma} \qquad [3.20]$$

dove $\gamma = m/n$ *è la* <u>frazione di individui che detengono la totalità del carattere</u>. Possiamo concludere che quando la totalità del carattere è distribuita uniformemente su un sottoinsieme U_1 dell'intero collettivo mentre le restanti unità ne possiedono una quota nulla, l'indice *T(Y)* è univocamente determinato dalla frazione γ di unità che appartengono a U_1.

Riprendendo il nostro esempio, la situazione nella quale si hanno *2* unità una delle quali detiene la totalità del carattere e l'altra nulla può essere assimilata a quella nella quale si hanno *2˙000˙000* unità delle quali *1˙000˙000* detengono quote tutte uguali del carattere e il restante *1˙000˙000* unità nulla. In entrambi i casi $\gamma = \frac{1}{2}$ e dunque *T(Y)* è *ln 2*.

3.6 Scomposizione dell'indice di concentrazione di Theil

Tenendo conto della scomposizione [3.17], l'indice [3.19] può essere scritto:

$$T(Y) = \ln n - \sum_{j=1}^{g} Y_j \ln \frac{1}{Y_j} - \sum_{j=1}^{g} Y_j H_j(Y) = \ln n + \sum_{j=1}^{g} Y_j \ln Y_j - \sum_{j=1}^{g} Y_j H_j(Y) \quad [3.21]$$

Togliendo e aggiungendo $\sum_{j=1}^{g} Y_j \ln n_j$ e tenendo conto che:

$$\ln n = \ln n \sum_{j=1}^{g} Y_j = \sum_{j=1}^{g} Y_j \ln n$$

$$\begin{aligned} T(Y) &= \sum_{j=1}^{g} Y_j \ln n + \sum_{j=1}^{g} Y_j \ln n_j - \sum_{j=1}^{g} Y_j \ln n_j + \sum_{j=1}^{g} Y_j \ln n_j - \sum_{j=1}^{g} Y_j H_j(Y) = \\ &= \sum_{j=1}^{g} Y_j (\ln n + \ln Y_j - \ln n_j) + \sum_{j=1}^{g} Y_j [\ln n_j - H_j(Y)] = \sum_{j=1}^{g} Y_j \ln \frac{nY_j}{n_j} + \sum_{j=1}^{g} Y_j T_j(Y) = \\ &= \sum_{j=1}^{g} Y_j \ln \frac{Y_j}{n_j/n} + \sum_{j=1}^{g} Y_j T_j(Y) \end{aligned} \quad [3.22]$$

Dove il primo addendo al secondo membro misura la <u>concentrazione **tra** i gruppi</u> mentre il secondo addendo è la media delle <u>concentrazioni **entro** i gruppi</u>.

Osserviamo che l'indice di concentrazione entro il gruppo j-mo, $T_j(Y)$, può anche essere scritto:

$$\begin{aligned} T_j(Y) &= \ln n_j - \sum_{i \in U_j} \frac{y_i}{Y_j} \left(\ln \frac{1}{y_i/Y_j} \right) = \sum_{i \in U_j} \frac{y_i}{Y_j} \left(\ln \frac{y_i}{Y_j} + \ln n_j \right) = \sum_{i \in U_j} \frac{y_i}{Y_j} \ln \left(\frac{y_i}{Y_j} n_j \right) = \\ &= \sum_{i \in U_j} \frac{y_i}{Y_j} \ln \left(\frac{y_i/Y_j}{1/n_j} \right) \end{aligned} \quad [3.23]$$

3.7 Distribuzione per classi del carattere

Consideriamo la situazione nella quale non si disponga dei dati individuali (elementari) ma soltanto della distribuzione per classi del carattere.

Sia, ad esempio, per $h = 1 \ldots m$:

$I_h = [x_h; x_{h+1})$: *h-ma* classe di reddito;

n_h : numero di individui che hanno un reddito entro tale classe;

T_h: reddito complessivamente detenuto da tali individui;

$Y_h = t_h / \sum_{h=1}^{m} t_h$: quota di reddito complessivamente detenuta da coloro che si trovano nell'*h-ma* classe.

L'ipotesi di lavoro che viene adottata in questo caso è quella di sintetizzare i valori nell'intervallo con il valore medio $x_h^* = t_h/n_h$ e poi di supporre che ogni individuo $u_i \in U_h$, dove con U_h indichiamo l'insieme di unita con reddito nella classe I_h, possieda lo stesso reddito x_h^*. Pertanto, la quota posseduta da ciascuno di tali individui è $y_i = x_h^* / \sum_{h=1}^{m} t_h$ e quindi:

$$\frac{y_i}{Y_h} = \frac{x_h^* / \sum_{h=1}^{m} t_h}{t_h / \sum_{h=1}^{m} t_h} = \frac{x_h^*}{t_h} = \frac{t_h/n_h}{t_h} = \frac{1}{n_h} \qquad [3.24]$$

Se la distribuzione per classi non riporta le quantità t_h, si conviene, normalmente, di sintetizzare i valori dell'intervallo con il valore centrale dello stesso: $x_h^* = (x_h+x_{h+1})/2$ e quindi supporre che ogni individuo $u_i \in U_h$ possieda lo stesso reddito x_h^*. La stima del reddito complessivo della classe è allora:

$t_h = x_h^* n_h$ [3.25] e quindi vale ancora la [3.24].

Tenendo conto delle considerazioni appena esposte, nel caso di una distribuzione per classi, l'indice di Theil può essere desunto come caso particolare della scomposizione [3.22] con i gruppi corrispondenti alle classi di reddito:

$$T(Y) = \sum_{h=1}^{m} Y_h \ln \frac{Y_h}{n_h/n} + \sum_{h=1}^{m} Y_h T_h(Y) = \sum_{h=1}^{m} Y_h \ln \frac{Y_h}{n_h/n} \quad [3.26]$$

in quanto: $$T_h(Y) = \sum_{i \in U_h} \frac{y_i}{Y_h} \ln \frac{y_i/Y_h}{1/n_h} = 0 \quad [3.27]$$

Poiché per la [3.24] $\frac{y_i}{Y_h} = \frac{1}{n_h} \quad i \in U_h \quad h = 1, \ldots, m$

3.8 Distribuzione del carattere in classi e gruppi

Consideriamo ancora la situazione più frequente nella quale non si disponga dei dati elementari. Ripetendo le ipotesi di lavoro richiamate nel precedente paragrafo e supponendo data una distribuzione in *m* classi e *g* gruppi, l'indice [3.26] assume la seguente espressione:

$$T(Y) = \sum_{j=1}^{g} \sum_{h=1}^{m} Y_{jh} \ln \frac{Y_{jh}}{n_{jh}/n} \quad [3.28]$$

dove:

Y_{jh} = quota di reddito della classe *h* del gruppo *j* ;

n_{jh} = numero di unità della classe *h* del gruppo *j* ;

$Y_j = \sum_{h=1}^{m} Y_{jh} = quota\, di\, reddito\, del\, gruppo\, j;\ \sum_{j=1}^{g} Y_j = \sum_{j=1}^{g} \sum_{h=1}^{m} Y_{jh} = 1$

$n_j = \sum_{h=1}^{m} n_{jh} =$ numero di unità del gruppo *j* ;

$\sum_{j=1}^{g} n_j = \sum_{j=1}^{g} \sum_{h=1}^{m} n_{jh} = n$ numero di unità del collettivo

Anche l'indice di concentrazione [3.28] può essere scisso in termini di concentrazione tra gruppi ed entro i gruppi.

Riscriviamo la [10.28] nella forma equivalente:

$$T(Y) = \sum_{j=1}^{g} Y_j \sum_{h=1}^{m} \frac{Y_{jh}}{Y_j} \ln \frac{Y_{jh}}{n_{jh/n}} \qquad [3.29]$$

e consideriamo gli addendi della seconda sommatoria.

$$\sum_{h=1}^{m} \frac{Y_{jh}}{Y_j} \ln \frac{Y_{jh}}{n_{jh}/n} = \sum_{h=1}^{m} \frac{Y_{jh}}{Y_j} \ln \left(\frac{Y_{jh} Y_j n_j}{n_{jh}/n Y_j n_j} \right) = \sum_{h=1}^{m} \frac{Y_{jh}}{Y_j} \ln \left(\frac{Y_{jh}/Y_j}{n_{jh}/n_j} \frac{Y_j}{n_j/n} \right) =$$

$$= \sum_{h=1}^{m} \frac{Y_{jh}}{Y_j} \left(\ln \frac{Y_{jh}/Y_j}{n_{jh}/n_j} + \ln \frac{Y_j}{n_j/n} \right) = \sum_{h=1}^{m} \frac{Y_{jh}}{Y_j} \ln \frac{Y_{jh}/Y_j}{n_{jh}/n_j} + \ln \frac{Y_j}{n_j/n} =$$

$$= T_j(Y) + \ln \frac{Y_j}{n_j/n} .\, Sostituendo\, nella\, [3.29] si\, ottiene:$$

$$T(Y) = \sum_{j=1}^{g} Y_j T_j(Y) + \sum_{j=1}^{g} Y_j \ln \frac{Y_j}{n_j/n} \qquad [3.30]$$

Dove il primo addendo al secondo membro è la media delle concentrazioni **entro** i gruppi mentre il secondo addendo misura la concentrazione **tra** i gruppi.

4.Analisi della concentrazione dei consumi delle famiglie italiane

Dopo esserci soffermati sull'aspetto teorico riguardante la misura della concentrazione, ora andiamo a spiegare i risultati ottenuti dall'applicazione economica effettuata. L'obiettivo era quello di analizzare la concentrazione dei consumi delle famiglie nelle diverse regioni italiane. Prima di tutto, abbiamo reperito dall'Istat i dati riguardanti la spesa media mensile familiare per ogni regione, considerando famiglie composte da 1, 2, 3, 4, 5 e più persone.

	spesa media mensile familiare (in euro correnti)				
	2018				
	1	2	3	4	5 e più
Territorio					
Piemonte	1933,07	2696,96	3175,26	3705,56	3295,96
Valle d'Aosta	2079,17	3027,32	3672,42	5137,37	3260,89
Liguria	1840,12	2798,74	3179,12	3315,49	3624,31
Lombardia	2077,85	3182,22	3615,76	3996,14	3996,82
Trentino Alto Adige	1989,71	3070,21	3389,59	3652,43	4671,75
Veneto	1850,46	2686,86	3330,51	3458,21	3244,26
Friuli-Venezia Giulia	1774,46	2706,51	3040,17	3759,08	2708,02
Emilia-Romagna	1818,93	3113,89	3651,49	3889,84	3649,08
Toscana	1932,40	3159,10	3213,16	3815,07	3968,39
Umbria	1666,32	2289,12	2458,35	2975,93	3769,53
Marche	1520,04	2297,68	2825,05	3177,98	3544,94
Lazio	1983,24	2998,93	3221,72	3613,35	3630,52
Abruzzo	1553,28	2142,46	2606,55	3299,68	3265,20
Molise	1385,02	2250,71	2459,14	3511,72	2660,35
Campania	1396,68	1986,59	2397,09	2748,54	2465,20
Puglia	1266,44	1933,77	2493,98	2571,32	2961,44
Basilicata	1380,83	2015,05	2415,53	2790,20	2867,30
Calabria	1314,32	1880,62	2105,24	2534,31	2509,12
Sicilia	1346,21	1933,39	2226,09	2831,64	2317,38
Sardegna	1540,96	2371,39	2434,77	2800,85	2317,77

Dopo di che, abbiamo reperito dall'Istat i dati riguardanti il numero totale di famiglie in ogni regione e le percentuali, per ogni regione, di tutte le famiglie composte sempre da 1, 2, 3, 4, 5 e più persone.

	Numero di famiglie (composizioni percentuali e totali in migliaia)					
	2018					
	1	2	3	4	5 e più	
Territorio						**Totale**
Piemonte	36,2	29,3	19,2	11,9	3,4	1996
Valle d'Aosta	49,1	26,1	13,6	7,8	3,5	66
Liguria	40,9	29,9	15,9	10,7	2,6	756
Lombardia	34,1	28,1	20,2	13,3	4,3	4389
Trentino Alto Adige	33,1	27,9	16,9	15,7	6,4	446
Veneto	30,2	29,6	19,2	15,0	6,0	2038
Friuli-Venezia Giulia	38,9	28,3	17,1	11,9	3,8	564
Emilia-Romagna	36,5	28,0	18,6	12,2	4,6	1995
Toscana	34,8	28,2	19,7	13,0	4,3	1648
Umbria	35,4	28,7	17,0	13,7	5,3	389
Marche	30,2	27,5	22,7	14,3	5,3	641
Lazio	36,1	25,6	18,9	15,0	4,4	2589
Abruzzo	29,6	26,8	21,5	17,2	4,9	542
Molise	35,0	26,1	18,4	15,7	4,7	134
Campania	25,8	22,1	21,1	20,8	10,2	2147
Puglia	26,4	26,6	21,2	20,1	5,7	1595
Basilicata	32,5	26,7	19,3	16,2	5,3	240
Calabria	31,3	25,5	18,2	18,3	6,7	796
Sicilia	30,1	25,3	19,4	18,4	6,7	2022
Sardegna	35,0	26,6	19,7	14,3	4,4	723
Italia						**25716**

L’Istat, quindi, non ci ha fornito quanto era l’ammontare di famiglie per ogni singola categoria di famiglia in ogni regione. Per calcolare ciò, abbiamo ripartito per ogni singola categoria familiare e per ogni regione la percentuale di famiglie rispetto al totale di esse per ogni regione, e ci siamo andati a costruire la corrispettiva tabella.

	Numero di famiglie (valori assoluti in migliaia)					
	2018					
	1	2	3	4	5 e più	Totale
Territorio						
Piemonte	723	585	383	238	68	1996
Valle d'Aosta	32	17	9	5	2	66
Liguria	309	226	120	81	20	756
Lombardia	1497	1233	887	584	189	4389
Trentino Alto Adige	148	124	75	70	29	446
Veneto	615	603	391	306	122	2038
Friuli-Venezia Giulia	219	160	96	67	21	564
Emilia-Romagna	728	559	371	243	92	1993
Toscana	574	465	325	214	71	1648
Umbria	138	112	66	53	21	389
Marche	194	176	146	92	34	641
Lazio	935	663	489	388	114	2589
Abruzzo	160	145	117	93	27	542
Molise	47	35	25	21	6	134
Campania	554	474	453	447	219	2147
Puglia	421	424	338	321	91	1595
Basilicata	78	64	46	39	13	240
Calabria	249	203	145	146	53	796
Sicilia	609	512	392	372	135	2020
Sardegna	253	192	142	103	32	723
Italia						**25712**

Fatto ciò, ci siamo andati a costruire la tabella contenente la spesa totale mensile di ogni categoria familiare per ogni regione. Ovviamente, poi, ci siamo anche calcolati la spesa totale mensile di ogni regione e la spesa totale in Italia. Come si evince, abbiamo sempre considerato i dati riguardanti il 2018.

		spesa totale mensile familiare (in euro correnti)					
		2018					
		1	2	3	4	5 e più	**Totale**
Territorio							
Piemonte		1396744	1577258	1216861	880159	223677	5294699
Valle d'Aosta		67378	52149	32964	26447	7533	186470
Liguria		568972	632638	382143	268197	71239	1923190
Lombardia		3109812	3924661	3205653	2332695	754308	13327129
Trentino Alto Adige		293733	382039	255487	255750	133350	1320359
Veneto		1138914	1620843	1303215	1057175	396708	5516855
Friuli-Venezia Giulia		389309	431991	293206	252294	58038	1424840
Emilia-Romagna		1324499	1739419	1354958	946748	334876	5700501
Toscana		1108239	1468147	1043172	817341	281216	4718115
Umbria		229462	255564	162571	158596	77716	883910
Marche		294252	405024	411065	291303	120432	1522076
Lazio		1853594	1987643	1576455	1403244	413574	7234510
Abruzzo		249196	311205	303741	307609	86717	1258469
Molise		64957	78716	60633	73880	16755	294941
Campania		773657	942611	1085923	1227432	539864	4569487
Puglia		533273	820441	843314	824352	269239	3290619
Basilicata		107705	129124	111887	108483	36472	493672
Calabria		327460	381728	304990	369168	133816	1517163
Sicilia		819333	989057	873224	1053506	313945	4049064
Sardegna		389940	456061	346787	289577	73733	1556098
Italia							**66082165**

Ora che abbiamo tutto ciò che di cui avevamo bisogno, iniziamo con il calcolare qual è la quota di ogni classe di famiglie per ogni regione sul totale delle famiglie italiane, cioè calcoliamo gli n_{jh}/n. Successivamente, facciamo la somma per righe (per ogni riga) per andarci a trovare la quota di famiglie in ciascuna regione sul totale, cioè stiamo calcolando gli $n_j/n = \Sigma\, n_{jh}/n$ con h = 1...m, nel nostro caso m = 5 (n° classi di famiglie).

Essendo delle quote, logicamente $\Sigma\, n_j/n = 1$ (a meno di approssimazioni), con j = 1...20, dove 20 è il numero delle regioni.

	Frequenze relative del numero di famiglie per regioni							
Territorio						nj/n		
Piemonte	0,02810	0,02274	0,01490	0,00924	0,00264	0,07762		
Valle d'Aosta	0,00126	0,00067	0,00035	0,00020	0,00009	0,00257		
Liguria	0,01202	0,00879	0,00467	0,00315	0,00076	0,02940		
Lombardia	0,05820	0,04796	0,03448	0,02270	0,00734	0,17067		
Trentino Alto Adige	0,00574	0,00484	0,00293	0,00272	0,00111	0,01734		
Veneto	0,02393	0,02346	0,01522	0,01189	0,00476	0,07925		
Friuli-Venezia Giulia	0,00853	0,00621	0,00375	0,00261	0,00083	0,02193		
Emilia-Romagna	0,02832	0,02172	0,01443	0,00946	0,00357	0,07750		
Toscana	0,02230	0,01807	0,01262	0,00833	0,00276	0,06408		njh/n
Umbria	0,00535	0,00434	0,00257	0,00207	0,00080	0,01514		
Marche	0,00753	0,00685	0,00566	0,00356	0,00132	0,02493		
Lazio	0,03634	0,02577	0,01903	0,01510	0,00443	0,10068		
Abruzzo	0,00624	0,00565	0,00453	0,00363	0,00103	0,02108		
Molise	0,00182	0,00136	0,00096	0,00082	0,00024	0,00521		
Campania	0,02154	0,01845	0,01762	0,01737	0,00852	0,08349		
Puglia	0,01637	0,01650	0,01315	0,01247	0,00354	0,06202		
Basilicata	0,00303	0,00249	0,00180	0,00151	0,00049	0,00933		
Calabria	0,00969	0,00789	0,00563	0,00566	0,00207	0,03095		
Sicilia	0,02367	0,01989	0,01525	0,01447	0,00527	0,07855		
Sardegna	0,00984	0,00748	0,00554	0,00402	0,00124	0,02811		
						1		

Fatto ciò, abbiamo calcolato la quota di ogni classe di famiglie di ogni regione sul totale delle famiglie in ogni regione, cioè le n_{jh}/n_j. Poi siamo passati a verificare che la somma per righe portasse all'incirca 1, sempre per il fatto che stiamo sommando delle quote.

	Frequenze relative del numero di famiglie per dimensione							
Territorio								
Piemonte	0,36200	0,29300	0,19200	0,11900	0,03400	1		
Valle d'Aosta	0,49100	0,26100	0,13600	0,07800	0,03500	1		
Liguria	0,40900	0,29900	0,15900	0,10700	0,02600	1		
Lombardia	0,34100	0,28100	0,20200	0,13300	0,04300	1		
Trentino Alto Adige	0,33100	0,27900	0,16900	0,15700	0,06400	1		
Veneto	0,30200	0,29600	0,19200	0,15000	0,06000	1		
Friuli-Venezia Giulia	0,38900	0,28300	0,17100	0,11900	0,03800	1		
Emilia-Romagna	0,36500	0,28000	0,18600	0,12200	0,04600	1		
Toscana	0,34800	0,28200	0,19700	0,13000	0,04300	1		
Umbria	0,35400	0,28700	0,17000	0,13700	0,05300	1		njh/nj
Marche	0,30200	0,27500	0,22700	0,14300	0,05300	1		
Lazio	0,36100	0,25600	0,18900	0,15000	0,04400	1		
Abruzzo	0,29600	0,26800	0,21500	0,17200	0,04900	1		
Molise	0,35000	0,26100	0,18400	0,15700	0,04700	1		
Campania	0,25800	0,22100	0,21100	0,20800	0,10200	1		
Puglia	0,26400	0,26600	0,21200	0,20100	0,05700	1		
Basilicata	0,32500	0,26700	0,19300	0,16200	0,05300	1		
Calabria	0,31300	0,25500	0,18200	0,18300	0,06700	1		
Sicilia	0,30100	0,25300	0,19400	0,18400	0,06700	1		
Sardegna	0,35000	0,26600	0,19700	0,14300	0,04400	1		

Il prossimo step è quello di andarci a calcolare la quota di spesa totale mensile di ogni classe di famiglia in ogni regione sul totale della spesa italiana, cioè gli Y_{jh}. A seguito di ciò, facendo la somma per righe troviamo gli $Y_j = \Sigma\, Y_{jh}$ con h = 1…5.

Poi, facendo la $\Sigma\ Y_j$ con j = 1...20, dobbiamo sempre avere 1 come risultato, come prima. A questo punto, abbiamo già tutto ciò che ci serve per arrivare al primo risultato, cioè trovare quanto è la concentrazione “tra i gruppi”. Per fare ciò, dobbiamo utilizzare la seguente formula:

$T_b(Y) = \Sigma\ Y_j * \ln (Y_j/(n_j/n))$ con j = 1...20

		Intensità relative dei consumi per regione								
Territorio							Yj			
Piemonte		0,02114	0,02387	0,01841	0,01332	0,00338	0,08012	0,00255		
Valle d'Aosta		0,00102	0,00079	0,00050	0,00040	0,00011	0,00282	0,00026		
Liguria		0,00861	0,00957	0,00578	0,00406	0,00108	0,02910	-0,00029		
Lombardia		0,04706	0,05939	0,04851	0,03530	0,01141	0,20168	0,03366		
Trentino Alto Adige		0,00444	0,00578	0,00387	0,00387	0,00202	0,01998	0,00283		
Veneto		0,01723	0,02453	0,01972	0,01600	0,00600	0,08348	0,00435		
Friuli-Venezia Giulia		0,00589	0,00654	0,00444	0,00382	0,00088	0,02156	-0,00037		
Emilia-Romagna		0,02004	0,02632	0,02050	0,01433	0,00507	0,08626	0,00924		
Toscana		0,01677	0,02222	0,01579	0,01237	0,00426	0,07140	0,00772		
Umbria		0,00347	0,00387	0,00246	0,00240	0,00118	0,01338	-0,00166		Yjh
Marche		0,00445	0,00613	0,00622	0,00441	0,00182	0,02303	-0,00182		
Lazio		0,02805	0,03008	0,02386	0,02123	0,00626	0,10948	0,00917		
Abruzzo		0,00377	0,00471	0,00460	0,00465	0,00131	0,01904	-0,00193		
Molise		0,00098	0,00119	0,00092	0,00112	0,00025	0,00446	-0,00069		
Campania		0,01171	0,01426	0,01643	0,01857	0,00817	0,06915	-0,01303		
Puglia		0,00807	0,01242	0,01276	0,01247	0,00407	0,04980	-0,01093		
Basilicata		0,00163	0,00195	0,00169	0,00164	0,00055	0,00747	-0,00166		
Calabria		0,00496	0,00578	0,00462	0,00559	0,00203	0,02296	-0,00686		
Sicilia		0,01240	0,01497	0,01321	0,01594	0,00475	0,06127	-0,01522		
Sardegna		0,00590	0,00690	0,00525	0,00438	0,00112	0,02355	-0,00417		
							1	0,01114	Tb(Y)	

Per arrivare al secondo risultato, cioè quello di calcolare la concentrazione “entro i gruppi”, dobbiamo innanzitutto andarci a calcolare la quota di spesa totale mensile di ogni classe di famiglia in ogni regione sulla spesa totale di ogni regione, cioè le cosiddette Y_{jh}/Y_j. Facendo la solita somma per righe (per ogni riga), verifichiamo che la somma di ogni quota sia pari a 1.

		Intensità relative dei consumi per dimensione delle famiglie							
Territorio									
Piemonte		0,26380	0,29789	0,22983	0,16623	0,04225	1		
Valle d'Aosta		0,36133	0,27966	0,17678	0,14183	0,04040	1		
Liguria		0,29585	0,32895	0,19870	0,13945	0,03704	1		
Lombardia		0,23334	0,29449	0,24054	0,17503	0,05660	1		
Trentino Alto Adige		0,22246	0,28934	0,19350	0,19370	0,10100	1		
Veneto		0,20644	0,29380	0,23622	0,19163	0,07191	1		
Friuli-Venezia Giulia		0,27323	0,30319	0,20578	0,17707	0,04073	1		
Emilia-Romagna		0,23235	0,30513	0,23769	0,16608	0,05875	1		
Toscana		0,23489	0,31117	0,22110	0,17323	0,05960	1		
Umbria		0,25960	0,28913	0,18392	0,17943	0,08792	1		
Marche		0,19332	0,26610	0,27007	0,19139	0,07912	1		
Lazio		0,25622	0,27474	0,21791	0,19397	0,05717	1		Yjh/Yj
Abruzzo		0,19802	0,24729	0,24136	0,24443	0,06891	1		
Molise		0,22024	0,26689	0,20558	0,25049	0,05681	1		
Campania		0,16931	0,20628	0,23765	0,26861	0,11815	1		
Puglia		0,16206	0,24933	0,25628	0,25052	0,08182	1		
Basilicata		0,21817	0,26156	0,22664	0,21975	0,07388	1		
Calabria		0,21584	0,25161	0,20103	0,24333	0,08820	1		
Sicilia		0,20235	0,24427	0,21566	0,26019	0,07754	1		
Sardegna		0,25059	0,29308	0,22286	0,18609	0,04738	1		

Ora, abbiamo tutto ciò che ci serve per arrivare al secondo risultato, e iniziamo calcolandoci i $T_j(Y)$ tramite la seguente formula:

$T_j(Y) = \Sigma\, Y_{jh}/Y_j * \ln(Y_{jh}/Y_j)/(n_{jh}/n)$ con $h=1...5$

Fatto ciò, per trovare la concentrazione "entro i gruppi" moltiplichiamo ogni $T_j(Y)$ per ogni Y_j e ne facciamo la somma, cioè $T_W(Y) = \Sigma\, T_j(Y) * Y_j$ con $j=1...20$. A questo punto, possiamo anche andarci a calcolare qual è la percentuale dei contributi delle singole regioni, semplicemente attraverso questa formula:

$T_j(Y) * Y_j / T_W(Y)$, ovviamente la somma per $j = 1...20$ ci dà sempre 1.

	Calcolo degli indici di concentrazione entro le regioni									
Territorio						Tj(Y)			Tj(Y)*Yj/Tw(Y)	
Piemonte	-0,08348	0,00493	0,04133	0,05557	0,00917	0,02752	0,00221		0,06814	
Valle d'Aosta	-0,11080	0,01931	0,04636	0,08480	0,00579	0,04547	0,00013		0,00396	
Liguria	-0,09582	0,03141	0,04429	0,03694	0,01311	0,02994	0,00087		0,02692	
Lombardia	-0,08852	0,01381	0,04200	0,04807	0,01555	0,03090	0,00623		0,19256	
Trentino Alto Adige	-0,08840	0,01053	0,02619	0,04069	0,04607	0,03509	0,00070		0,02166	
Veneto	-0,07853	-0,00219	0,04897	0,04693	0,01302	0,02819	0,00235		0,07272	percentuale dei contributi delle singole regioni
Friuli-Venezia Giulia	-0,09652	0,02089	0,03810	0,07037	0,00283	0,03567	0,00077		0,02376	
Emilia-Romagna	-0,10494	0,02623	0,05829	0,05123	0,01437	0,04517	0,00390		0,12039	
Toscana	-0,09233	0,03063	0,02552	0,04974	0,01946	0,03302	0,00236		0,07283	
Umbria	-0,08052	0,00214	0,01448	0,04841	0,04450	0,02901	0,00039		0,01199	
Marche	-0,08623	-0,00875	0,04692	0,05578	0,03171	0,03941	0,00091		0,02805	
Lazio	-0,08785	0,01941	0,03101	0,04986	0,01497	0,02741	0,00300		0,09270	
Abruzzo	-0,07961	-0,01989	0,02791	0,08590	0,02349	0,03781	0,00072		0,02225	
Molise	-0,10202	0,00595	0,02279	0,11702	0,01077	0,05452	0,00024		0,00752	
Campania	-0,07132	-0,01421	0,02826	0,06870	0,01736	0,02879	0,00199		0,06150	
Puglia	-0,07908	-0,01614	0,04861	0,05517	0,02958	0,03813	0,00190		0,05867	
Basilicata	-0,08695	-0,00538	0,03642	0,06700	0,02454	0,03562	0,00027		0,00822	
Calabria	-0,08022	-0,00337	0,01999	0,06933	0,02425	0,02997	0,00069		0,02126	
Sicilia	-0,08035	-0,00858	0,02283	0,09014	0,01132	0,03536	0,00217		0,06694	
Sardegna	-0,08373	0,02841	0,02748	0,04902	0,00351	0,02470	0,00058		0,01797	
							0,03237	Tw(Y)	**1**	

Ultimo step è quello di calcolare la concentrazione totale, e lo possiamo fare attraverso questa formula:

$T(Y) = \Sigma\ Y_j * \ln(Y_j/(n_j/n))$ con $j = 1...20$,

dopo aver fatto la $\Sigma\ Y_{jh} * \ln(Y_{jh}/(n_{jh}/n))$ con $h = 1...5$.

Sostanzialmente abbiamo scorporato la concentrazione nelle due parti illustrate dalla [3.28].

	Calcolo diretto dell'indice di concentrazione totale						
Territorio							
Piemonte	-0,00602	0,00115	0,00390	0,00488	0,00084		
Valle d'Aosta	-0,00022	0,00013	0,00018	0,00028	0,00003		
Liguria	-0,00288	0,00082	0,00123	0,00103	0,00037		
Lombardia	-0,01000	0,01270	0,01657	0,01559	0,00504		
Trentino Alto Adige	-0,00114	0,00103	0,00107	0,00136	0,00121		
Veneto	-0,00566	0,00109	0,00511	0,00475	0,00140		
Friuli-Venezia Giulia	-0,00218	0,00034	0,00075	0,00145	0,00005		
Emilia-Romagna	-0,00693	0,00506	0,00720	0,00594	0,00178		
Toscana	-0,00478	0,00459	0,00353	0,00489	0,00185		
Umbria	-0,00150	-0,00045	-0,00011	0,00035	0,00045		
Marche	-0,00234	-0,00069	0,00059	0,00094	0,00059		
Lazio	-0,00727	0,00465	0,00539	0,00724	0,00216		
Abruzzo	-0,00190	-0,00086	0,00007	0,00116	0,00031		
Molise	-0,00061	-0,00016	-0,00004	0,00035	0,00001		
Campania	-0,00714	-0,00367	-0,00114	0,00125	-0,00034		
Puglia	-0,00571	-0,00353	-0,00038	0,00001	0,00058		
Basilicata	-0,00101	-0,00048	-0,00010	0,00014	0,00006		
Calabria	-0,00332	-0,00180	-0,00092	-0,00008	-0,00005		
Sicilia	-0,00802	-0,00426	-0,00190	0,00155	-0,00049		
Sardegna	-0,00302	-0,00055	-0,00028	0,00038	-0,00012		
						0,04337	T(Y)

Conclusioni

L'indice di Theil è stato scelto per l'applicazione economica oggetto di studio per vari motivi. L'indice di Theil, facendo parte delle Generalised Entropy Inequality Measures, soddisfa gli assiomi di simmetria, indipendenza dalla media, indipendenza dalla popolazione, di coerenza con il principio del trasferimento di Pigou-Dalton, di coerenza con il principio di anonimità e il principio della scomponibilità per gruppi. Una delle caratteristiche salienti dell'indice di Theil, e degli indici di entropia più in generale, è quella di essere perfettamente scomponibile nelle componenti "entro" e "tra" gruppi; questa caratteristica non è esattamente applicabile all'indice di Gini a meno della non esistenza di valori di sovrapposizione. Questo fa sì che tale tipologia di indici sia molto usata per lo studio della spesa totale mensile all'interno di una data regione e tra regioni diverse, considerando quindi la componente "tra" come la media ponderata delle distanze tra le spese medie dei vari gruppi e la componente "entro" come la media ponderata delle disuguaglianze interne ad ogni singolo gruppo. Per quanto concerne, invece, i risultati della nostra applicazione, ne abbiamo ottenuti alcuni molto interessanti da analizzare. Un aspetto interessante è quello della percentuale dei contributi delle singole regioni. Dall'elaborazione, si evince che la Lombardia è la regione con la percentuale più alta, di circa il 19,26%, cioè da sola contribuisce a circa 1/5 di tutta la nazione. Questo risultato, si giustifica principalmente con il fatto che questa regione è quella con il maggior numero di famiglie (4389, valore assoluto in migliaia), mentre la spesa media mensile è sì tra le più alte, ma non la più alta. Inoltre, possiamo anche notare che generalmente le regioni del centro-nord sono quelle con le percentuali più alte. Per quanto riguarda le concentrazioni "tra i gruppi", notiamo che queste sono molto basse, e le componenti che concorrono a formare la misura della concentrazione "tra i gruppi" presentano anche valori negativi.

Questo succede perché ci sono molte disuguaglianze tra le regioni, sia in termini di spesa media mensile sia in termini di diversa composizione numerica delle famiglie. Infatti, la concentrazione “tra i gruppi” è inferiore alla concentrazione “entro i gruppi”, per il fatto che la disuguaglianza tra le diverse regioni è molto più ampia rispetto a quella all’interno delle stesse. Da qui, per l’appunto, si evince che sulla concentrazione totale pesa molto di più la concentrazione “entro” rispetto a quella “tra” i gruppi; ricordandosi che la concentrazione totale è data dalla somma delle altre due.

Bibliografia

Istat, Annuario statistico italiano 2019

Istat, Statbase, https://www.istat.it/it/dati-analisi-e-prodotti/banche-dati/statbase

Mattioli E., Dispense del corso di Statistica economica, a.a. 2019/2020, https://learn.univpm.it/

Orazi F., Dispense del corso di Sociologia Economica, a.a. 2017/2018, CLUA

Theil H., Economics and Information Theory, North-Holland, Amsterdam,1967