

Storia di un caso: intervalli di confidenza per una proporzione per la regolazione della qualità del servizio nel settore energetico nazionale *

Emanuela Olgiati, Luca Paglieri, Simonetta Salvati, Piercesare Secchi

26 ottobre 2005

MOX– Modellistica e Calcolo Scientifico
Dipartimento di Matematica “F. Brioschi”
Politecnico di Milano
via Bonardi 9, 20133 Milano, Italy
emanuela.olgiati, luca.paglieri, simonetta.salvati,
piercesare.secchi@mate.polimi.it

Keywords: qualità del servizio, metodologia di auditing, campionamento da popolazione finita, intervallo di confidenza per una proporzione.

Sommario

Si descrive il processo di costruzione di un metodo di controllo campionario della qualità dei servizi di distribuzione misura e vendita dell'energia elettrica e del gas attualmente adottato dall'Autorità per l'energia elettrica e il gas italiana.

1 L'Autorità per l'energia elettrica e il gas

La regolazione della qualità del servizio nella distribuzione dell'energia elettrica e del gas ha ricevuto particolare attenzione negli ultimi anni, sia in Europa che negli Stati Uniti. In Italia ciò si è concretizzato nell'istituzione della “Autorità per l'energia elettrica e il gas” operativa dal 1997, un organismo indipendente con il compito di perseguire le finalità indicate dalla legge istitutiva n.481 del 1995, “garantire la promozione della concorrenza e dell'efficienza” nei settori dell'energia elettrica e del gas, nonché “assicurare adeguati livelli di qualità” dei servizi. In ottemperanza a tali funzioni, l'Autorità ha di recente emanato le

*Lavoro svolto nell'ambito del progetto AEEG2004, finanziato dalla *Autorità per l'energia elettrica e il gas*. Si ringraziano l'Ing.A.Grossi, l'Ing.L.Lo Schiavo ed il loro staff per la proficua e attiva collaborazione.

direttive sui livelli generali e specifici di qualità dei servizi erogati dagli esercenti e sui meccanismi di rimborso ai consumatori in caso del loro mancato rispetto. I livelli di qualità investono sia aspetti di natura contrattuale (come tempestività di intervento e risposta a reclami), che aspetti di carattere tecnico (come la continuità dei servizi e la sicurezza). Tra i compiti dell’Autorità vi è anche quello di verifica nei confronti degli esercenti, allo scopo di tutelare gli utenti finali e i consumatori ma anche di ingenerare un meccanismo virtuoso di miglioramento dell’efficienza e della sicurezza del servizio. Si è resa così necessaria per Aeg la definizione di una metodologia di controllo dell’operato degli esercenti. A questo proposito bisogna sottolineare che la situazione che l’Autorità si trova a fronteggiare è estremamente varia sia in relazione alla dimensione delle aziende coinvolte, sia in relazione alla loro posizione geografica. Questa considerazione ha motivato la scelta di costruire una metodologia di controllo flessibile, basata sui dati di qualità in possesso della stessa azienda controllata. Contemporaneamente Aeg ha promosso un processo di standardizzazione della registrazione dei dati. Ciò ha reso possibile pensare ad una loro ispezione tramite campionamento da parte di un ispettore dotato di strumenti adeguati per l’analisi contestuale dei dati estratti. L’Autorità si è pertanto rivolta al MOX, laboratorio di Modellistica e di Calcolo Scientifico del Dipartimento di Matematica del Politecnico di Milano, in collaborazione col quale è stato messo a punto un metodo di controllo che è ora legge dello Stato pubblicata sulla Gazzetta Ufficiale n.245/04. In questo lavoro viene descritto il processo che ha determinato la scelta del metodo adottato, evidenziando come la sua generazione sia stata dovuta alla costante interazione di considerazioni legate alla modellazione statistica del problema e di vincoli di fattibilità e di opportunità. Infatti, molte delle proposte elaborate durante il processo di costruzione del metodo di controllo sono state sottoposte al vaglio degli esercenti attraverso lo strumento della “consultazione” che ha generato importanti approfondimenti delle tematiche fino ad arrivare alla adozione di un metodo che ha convinto poi sul campo sulla base della sua praticabilità e del controllo statistico della qualità del servizio che esso permette.

2 Genesi del problema

L’esigenza iniziale espressa dall’Autorità durante il primo incontro nell’Ottobre 2003 è quella di “definire una metodologia di effettuazione dei controlli e di valutazione degli esiti ai fini di verificare l’attuazione da parte degli esercenti delle deliberazioni emanate dall’Autorità e, in caso di esito negativo, dell’eventuale comminazione di sanzioni”. Le deliberazioni cui si fa riferimento sono relative alla qualità commerciale

- del servizio di distribuzione e vendita dell’energia elettrica (delibera n.220/02);
- del servizio di distribuzione e vendita del gas (delibera n.47/00);

e alla qualità tecnica (sicurezza e continuità) del servizio di distribuzione del gas (delibera n.236/00 e n.47/00, limitatamente al pronto intervento). Le deliberazioni citate prevedono che ogni operazione compiuta dall'esercente debba essere adeguatamente operata e documentata, secondo la disciplina imposta dall'Autorità. Nel caso della qualità commerciale del gas le operazioni vengono indicate come "prestazioni"; prestazione è, ad esempio, la stesura di un preventivo per lavori o l'attivazione di una fornitura. Una prestazione deve essere eseguita secondo precise modalità ed entro termini temporali stabiliti nelle delibere, inoltre tutto il percorso di una prestazione deve essere memorizzato su un registro elettronico di formato standard. Compito di un ispettore dell'Autorità è dunque quello di esaminare un campione di prestazioni verificandone la corretta esecuzione e registrazione anche attraverso riscontri cartacei. Sulla base degli esiti dell'ispezione l'esercente ritenuto inadempiente rispetto alla normativa viene sanzionato in termini economici. Il problema posto al MOX consiste nella definizione di un metodo inferenziale, basato su dati campionari, che permetta di quantificare il livello di inadempienza dell'esercente; il metodo deve essere valido da un punto di vista statistico e allo stesso tempo deve tener conto di fattori di carattere logistico, economico e temporale, quali ad esempio i tempi di solito stretti entro i quali un'ispezione avviene. Inoltre il metodo dovrà essere sottoposto al vaglio degli stessi esercenti: è prassi dell'Autorità lavorare in collaborazione con i controllati attraverso la pubblicazione dei "Documenti di consultazione", in cui l'Autorità sottopone alcune proposte agli esercenti che hanno la possibilità di replicare con controproposte; in questo modo si instaura un processo dinamico di revisione e ridefinizione del metodo di controllo, anche sulla base delle osservazioni degli esercenti.

I dati da campionare, cioè le prestazioni, si presentano suddivisi in numerose classi la cui dimensione tipicamente varia da poche decine a svariate migliaia di unità; in questa prima fase l'Autorità richiede che tutte le classi vengano adeguatamente rappresentate nel campione. A fine Novembre 2003 il MOX propone di accertare l'eventuale inadempienza dell'esercente per mezzo di una batteria di prove delle ipotesi che verifichino se, nelle diverse classi, le proporzioni di prestazioni non coerenti con le direttive siano maggiori di valori soglia stabiliti dall'Autorità. Questa prima proposta non incontra il favore dell'Autorità sostanzialmente per due motivi: il primo riguarda il fatto che, trattandosi della prima esperienza del genere, l'Autorità non ha ancora la sensibilità al problema per ipotizzare valori di soglia ragionevoli da utilizzare nella formulazione dell'ipotesi nulla dei test. Ma ancora più che questo, emerge la forte preoccupazione dell'Autorità in merito all'inevitabile rischio di errore (di primo tipo) a carico dell'esercente, ossia al rischio di sanzionare un esercente che invece si dovrebbe considerare "virtuoso". Emerge cioè il fatto che, perlomeno in questa fase iniziale di controlli, l'Autorità preferisce adottare un atteggiamento molto "garantista" nei confronti dell'esercente. Inoltre, dalle prime discussioni, risulta chiaro che la dimensione campionaria è sottoposta a limiti logistici invalicabili; il campione sarà spesso piuttosto piccolo rispetto alla popolazione di riferimento. Durante il

mezzo di Marzo 2004, una volta acquisiti i vincoli sulla dimensione del campione e soprattutto lo spirito con cui l’Autorità intende muoversi, il MOX propone di stimare un “limite inferiore” per la proporzione p di prestazioni non coerenti con le direttive per mezzo dell’estremo inferiore di un intervallo di confidenza destro per p del tipo $[p_L, 1]$. Inoltre, visti i risultati delle prime simulazioni numeriche e acclarato che la dimensione campionaria è fortemente vincolata, l’Autorità rinuncia all’esigenza che tutti i tipi di prestazione vengano rappresentati nel campione e ipotizza ispezioni in cui è esaminata una sola classe di prestazione, fatto questo che riduce ulteriormente la dimensione della popolazione di riferimento.

3 Il problema da un punto di vista statistico

Allo scopo di fissare la notazione, si riformula matematicamente il problema. Da una popolazione composta da un numero noto $N > 0$ di unità (le prestazioni appartenenti ad una data classe), è estratto senza reimmissione un campione casuale semplice di n unità, $1 \leq n \leq N$. Un numero ignoto M , $0 \leq M \leq N$, di unità della popolazione presentano una certa caratteristica (ossia sono prestazioni non coerenti con le direttive); sulla base del numero X di unità del campione che presentano la caratteristica in questione, si vuole stimare il numero M . Poiché N è noto, stimare la quantità M è problema equivalente a stimare la proporzione $p = M/N$.

In letteratura statistica (si veda, per esempio, [4], [7]), si trovano molti metodi per la costruzione di un intervallo di confidenza del tipo $[p_L, 1]$ per p , quasi sempre trattati nel caso di popolazioni infinite, ossia nell’ambito del modello binomiale; talora viene utilizzata l’approssimazione normale della distribuzione binomiale. E’ chiaro che una costruzione basata sulla distribuzione gaussiana ha maggiore possibilità di essere accettata da un interlocutore con poche nozioni di statistica rispetto ad una basata, ad esempio, sulla distribuzione ipergeometrica. D’altro canto in molti casi concreti esibiti da AeeG, la dimensione N della popolazione delle prestazioni è piccola (dell’ordine di poche decine di unità), e di conseguenza è anche piccola la dimensione n del campione. Compito del MOX è quello di elaborare metodi inferenziali opportuni sia in relazione al modello statistico utilizzato sia in relazione alla facilità con cui essi possono essere spiegati ad un pubblico di non esperti. Conformemente a tali esigenze il MOX elabora tre metodi, che chiama “metodo Wald”, “metodo Wilson” e “metodo Buonaccorsi” in quanto ispirati a procedure di analogo nome per la costruzione di intervalli di confidenza per una proporzione. I primi due metodi fanno uso dell’ approssimazione normale per la distribuzione dello stimatore X/n di p ; essi vengono adattati al caso di popolazione finita utilizzando il fattore di correzione per popolazione finita $f = n/N$ nell’espressione della varianza dello stimatore. Il terzo metodo utilizza l’esatta distribuzione ipergeometrica per X . Gli intervalli di confidenza per M (o equivalentemente per p) che i tre metodi generano sono presentati dal MOX ad AeeG agli inizi di Marzo del 2004. Essi sono costruiti

mediante inversione della famiglia di test di livello $\alpha \in (0, 1)$ per verificare le ipotesi

$$H_0 : M = M_0 \quad vs \quad H_1 : M > M_0$$

con $M_0 \in \{0, N\}$, aventi regione di rifiuto definita da

$$R = \{x \in \{0, n\} : x > \min\{j \in \{0, 1, \dots, n\} : \mathbb{P}_{M_0}(X \geq j) \leq \alpha\}\}.$$

L'intervallo di confidenza generato dall'inversione di questa famiglia di test è del tipo $[M_L, N]$, dove, detto x_{obs} il valore assunto da X in corrispondenza del campione osservato,

$$M_L = \min\{M \in \{0, N\} : \mathbb{P}_M(X \geq x_{obs}) > \alpha\}. \quad (1)$$

Ciò per cui i tre metodi si differenziano è la distribuzione che essi assumono per la variabile X .

Il “metodo Buonaccorsi” [in [6] si trova il caso bilatero] considera per X la distribuzione ipergeometrica $Ipg(N, M, n)$; in questo caso l'estremo inferiore dell'intervallo di confidenza per M si calcola direttamente utilizzando la formula in (1) per mezzo di un algoritmo a terminazione finita facilmente implementabile su calcolatore.

Il “metodo Wald” [4] assume per X la distribuzione Gaussiana di media np e varianza $n(1-f)\hat{p}(1-\hat{p})$ dove $f = n/N$, $p = M/N$ e $\hat{p} = x_{obs}/n$. L'estremo inferiore dell'intervallo di confidenza destro per M di livello $1 - \alpha$ è espresso da:

$$M_L = \text{Int}\left(\left[\hat{p} - z_{1-\alpha}\sqrt{(1-f)\frac{\hat{p}(1-\hat{p})}{n}}\right]N\right)$$

dove $\text{Int}(r)$ indica la parte intera di un numero reale e $z_{1-\alpha}$ è il quantile di ordine $1 - \alpha$ della Gaussiana standard.

Il “metodo Wilson” ([11], [4]) assume che X abbia distribuzione Gaussiana di media np e varianza $n(1-f)p(1-p)$. In questo caso l'estremo inferiore dell'intervallo di confidenza destro per M di livello $1 - \alpha$ è dato da:

$$M_L = \text{Int}\left(\frac{\hat{p} + \frac{(1-f)z_{1-\alpha}^2}{2n} - (z_{1-\alpha}\sqrt{1-f})\sqrt{[\hat{p}(1-\hat{p}) + (1-f)z_{1-\alpha}^2/4n]/n}}{1 + (1-f)z_{1-\alpha}^2/n}N\right)$$

In analogia alla nomenclatura utilizzata nel caso di popolazioni infinite, si può dire che l'intervallo costruito con il metodo Buonaccorsi è quello esatto in quanto utilizza la distribuzione esatta della variabile aleatoria X . Si tratta invero di un metodo conservativo, che genera intervalli di confidenza la cui probabilità di copertura effettiva è sempre maggiore o uguale a quella nominale dichiarata dal coefficiente $1 - \alpha$. I metodi Wald e Wilson approssimano la distribuzione ipergeometrica di X con una Gaussiana; tuttavia, mentre il metodo Wilson ha

origine da una corretta valutazione della varianza di X , il metodo Wald stima questa quantità introducendo un ulteriore fattore di variabilità. Numerosi lavori (si veda, ad esempio, [3], [5]) hanno confrontato tra loro questi e altri tipi di intervalli nel caso di popolazioni infinite, soprattutto in relazione alla loro probabilità effettiva di copertura e alla loro lunghezza. Mentre è ormai largamente condivisa l'opinione che il classico intervallo di Wald risulta insoddisfacente soprattutto in merito alla probabilità di copertura effettiva, l'intervallo di Wilson è ritenuto interessante e consigliabile per alcune situazioni. In [3], in cui si affronta il caso di stima intervallare per una proporzione binomiale, esso viene consigliato anche per valori piccoli della dimensione campionaria, in quanto la sua probabilità di copertura risulta un ragionevole compromesso tra quella troppo alta degli intervalli esatti e quella troppo bassa dell'intervallo di Wald. Da una lettura generale si evince comunque che non esiste un intervallo che possa ritenersi "il migliore" per tutte le dimensioni (campionarie e della popolazione) e per tutti i valori di $p = M/N$. Pertanto il MOX decide di esaminare in maggiore dettaglio i tre metodi per mezzo di simulazioni numeriche che restringano l'attenzione a quei valori dei parametri di particolare interesse per l'Autorità.

4 Prime analisi numeriche

4.1 Probabilità di copertura

Se $C(X_1, \dots, X_n)$ è un intervallo di confidenza per un parametro θ , la probabilità di copertura (dipendente da θ) è definita da $\mathbb{P}_\theta(\theta \in C(X_1, \dots, X_n))$. Se l'intervallo di confidenza è dichiarato essere di livello $1 - \alpha$, l'utilizzatore di questa procedura di stima assume spesso inconsciamente che, per qualunque valore ammissibile del parametro θ , la probabilità di copertura dell'intervallo sia almeno $1 - \alpha$. I metodi illustrati nella sezione precedente consentono di costruire intervalli con un livello di confidenza "nominale" pari a $1 - \alpha$. In effetti la probabilità di copertura degli intervalli da essi generati è spesso diversa da quella dichiarata dal livello $1 - \alpha$ e varia al variare di n e M , ovvero di $p = M/N$. Per stimare la probabilità di copertura effettiva degli intervalli si procede mediante simulazione: fissati N , M ed n , si genera un numero elevato r di realizzazioni di $X \sim Ipg(N, M, n)$. Scelto un metodo tra i tre esaminati, per ciascuna delle r realizzazioni campionarie di X si costruisce il corrispondente intervallo di confidenza unilatero destro per $p = M/N$ di livello nominale pari a $1 - \alpha$ e, posto S_i uguale a 1 oppure a 0 rispettivamente se l'intervallo i -esimo contiene p oppure no, si stima la probabilità di copertura effettiva con la media campionaria delle S_i , ovvero per mezzo di $\frac{1}{r} \sum_{i=1}^r S_i$. Nelle simulazioni implementate dal MOX si considerano $\alpha = 0.05$, $r = 1000$; per gli altri parametri N , n e p vengono esaminati valori ritenuti di interesse per il problema specifico. In particolare:

- $N=100$, p variabile da 0 a 0.2 con passo 0.01, n variabile da 1 a 10 con passo 1;

- $N=500$, p variabile da 0 a 0.2 con passo 0.01, n variabile da 1 a 10 con passo 1;
- $N=1000$, p variabile da 0 a 0.2 con passo 0.01, n variabile da 1 a 10 con passo 1;
- $N=10000$, p variabile da 0 a 0.2 con passo 0.01, n variabile da 1 a 40 con passo 1.

Nel seguito, per brevità, si discutono solo i casi relativi ad $N = 100$ ed $N = 1000$. Per la descrizione di tutte le simulazioni condotte si veda [9].

Si riportano anzitutto i grafici delle probabilità di copertura stimate in funzione della dimensione campionaria n nei due casi $p = 0.05$ e $p = 0.15$. Sia $N = 100$; in Figura 1 si nota che la probabilità di copertura degli intervalli generati coi metodi Wald e Buonaccorsi è quasi sempre almeno uguale a 0.95, diversamente da quella dell'intervallo generato col metodo Wilson, la quale oscilla molto al variare di n . Inoltre, se per p "piccolo", $p = 0.05$, è l'intervallo costruito con il metodo Wald ad avere la più alta probabilità di copertura, per p "grande", $p = 0.15$, l'intervallo di Buonaccorsi si comporta meglio degli altri due. Entrambi i metodi Wald e Buonaccorsi si rivelano comunque conservativi per questi valori dei parametri.

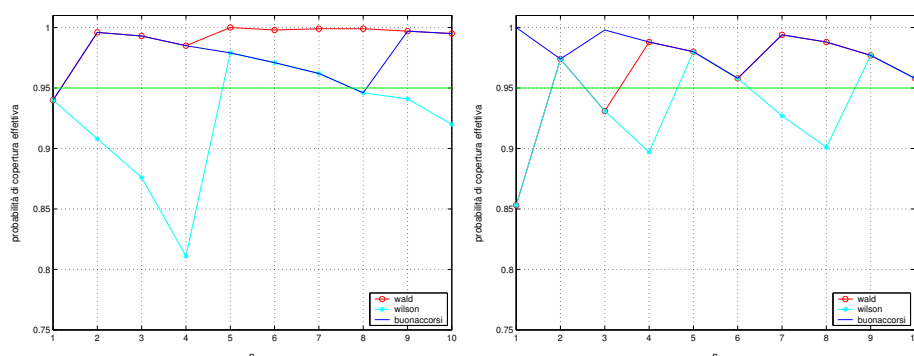


Figura 1: *Probabilità di copertura effettiva degli intervalli generati dai metodi di Wald, Wilson, Buonaccorsi con $N = 100$, $\alpha=0.05$, $1 \leq n \leq 10$. A sinistra: $p = 0.05$. A destra: $p = 0.15$.*

Per $N = 500$, $N = 1000$ (Figura 2) e $N = 10000$ si ottengono risultati molto simili al caso appena considerato. L'intervallo generato con il metodo Wilson mostra sempre una probabilità di copertura oscillante nell'intervallo $[0.8, 0.98]$. Per quanto riguarda gli altri due intervalli, quello generato dal metodo Buonaccorsi tende ad avere probabilità di copertura maggiore di quello generato col metodo Wald per quasi ogni n , quando p tende a 0.2. Si osservi che in alcune figure la probabilità di copertura dell'intervallo generato dal metodo Buonaccorsi risulta inferiore a 0.95. Questo fenomeno, che contrasta con la teoria, è dovuto

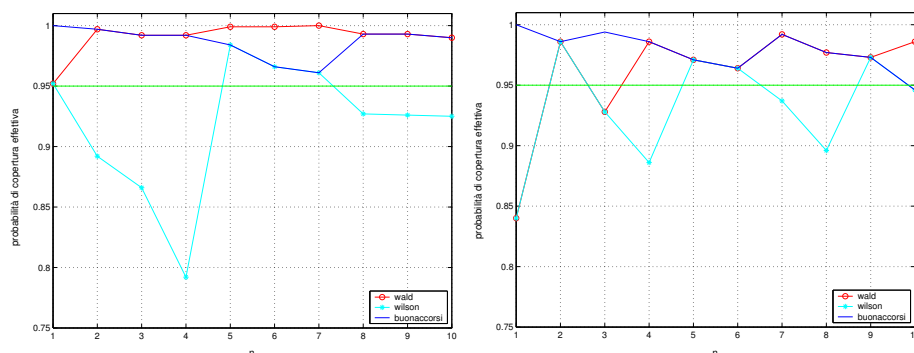


Figura 2: *Probabilità di copertura effettiva degli intervalli generati dai metodi di Wald, Wilson, Buonaccorsi con $N = 1000$, $\alpha=0.05$, $1 \leq n \leq 10$. A sinistra: $p = 0.05$. A destra: $p = 0.15$.*

ovviamente alla variabilità dello stimatore della probabilità di copertura che utilizza i dati generati dalla simulazione.

Si confrontano ora le probabilità di copertura degli intervalli costruiti con i tre metodi al variare di p tra 0 e 0.2 con passo 0.01, per n fissato. Le simulazioni mostrano che i risultati ottenuti per popolazioni di dimensione $N = 100$ e $N = 1000$ sono molto simili e conducono alle stesse conclusioni. Osservando le Figure 3 e 4 si nota che per valori “piccoli” di $p \in (0.01, \dots, 0.1)$ l’intervallo generato dal metodo Wald ha una probabilità di copertura maggiore o uguale a quella dell’intervallo generato dal metodo Buonaccorsi, mentre per valori “grandi” di $p \in (0.12, \dots, 0.2)$ è in generale il secondo a garantire la probabilità di copertura maggiore. Il metodo Wilson genera intervalli con probabilità di copertura che risulta sempre oscillante con ampiezza dell’oscillazione decrescente all’aumentare dell’ampiezza del campione.

Le probabilità di copertura, studiate nelle righe precedenti prima al variare di n e poi al variare di p , possono essere riassunte attraverso una media rispetto a p calcolata assumendo per p una distribuzione uniforme sull’intervallo $[0, 0.2]$. I risultati per una popolazione di dimensione $N = 100$ e $N = 1000$ sono riportati in Figura 5. Da una prima osservazione dei grafici si nota come l’andamento della media delle probabilità di copertura dei tre intervalli sia quasi identico nei due casi esposti (e ciò vale anche nei casi $N = 500$ ed $N = 10000$). Infatti l’intervallo generato dal metodo Wilson ha una media sempre inferiore a quella degli altri due intervalli e al livello di confidenza nominale; il metodo Buonaccorsi consente invece di costruire intervalli molto conservativi: la media delle probabilità di copertura è sempre maggiore o uguale a 0.98. L’intervallo generato dal metodo Wald ha per $n = 1$ una media delle probabilità di copertura piuttosto bassa, pari a 0.9, e molto inferiore rispetto ai valori che assume per $n \geq 2$. Per $n \geq 4$ l’andamento segue quello dell’intervallo generato dal metodo Buonaccorsi, assumendo in particolare valori leggermente superiori.

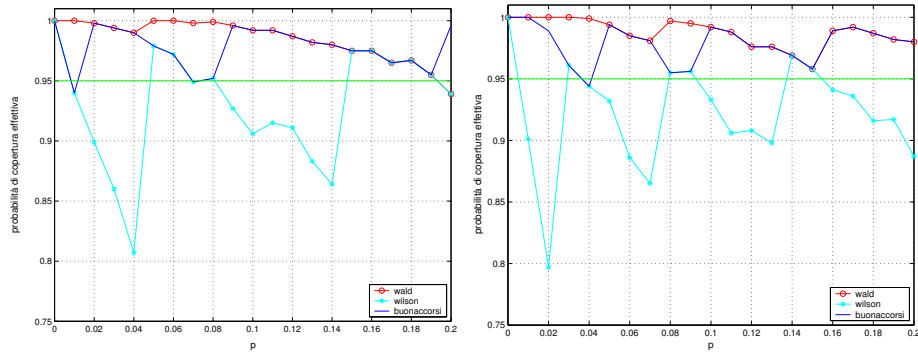


Figura 3: *Probabilità di copertura effettiva degli intervalli generati dai metodi di Wald, Wilson, Buonaccorsi con $N = 100$, $\alpha = 0.05$, $p_{max} = 0.2$, passo di $p = 0.01$. A sinistra: $n = 5$. A destra: $n = 10$.*

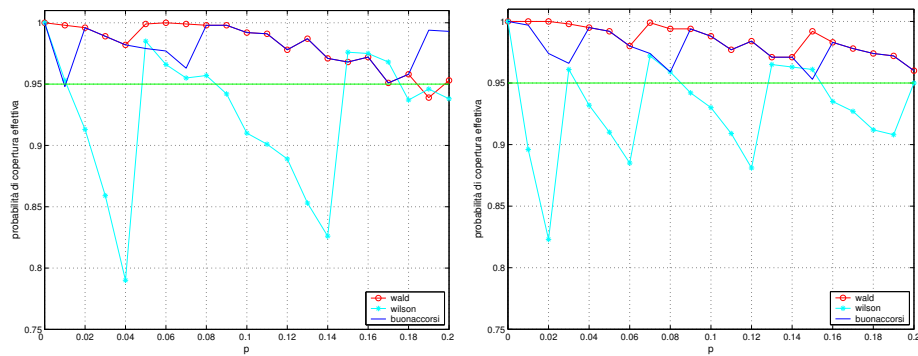


Figura 4: *Probabilità di copertura effettiva degli intervalli generati dai metodi di Wald, Wilson, Buonaccorsi con $N = 1000$, $\alpha = 0.05$, $p_{max} = 0.2$, passo di $p = 0.01$. A sinistra: $n = 5$. A destra: $n = 10$.*

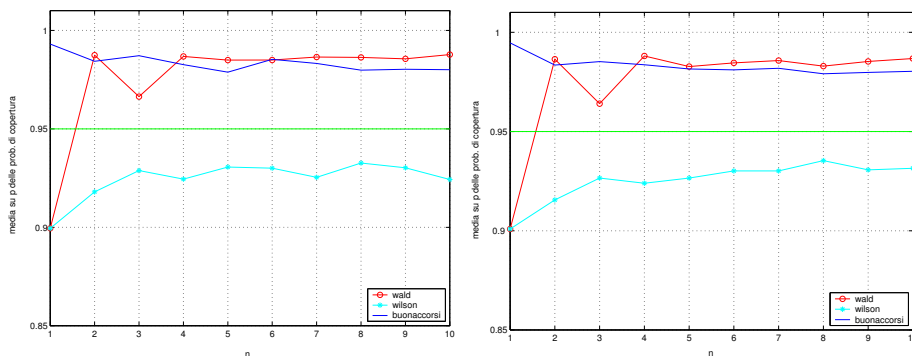


Figura 5: *Media su p delle probabilità di copertura degli intervalli generati dai metodi di Wald, Wilson e Buonaccorsi con $\alpha=0.05$, $1 \leq n \leq 10$, $p_{max}=0.2$, passo di $p = 0.01$. A sinistra: $N = 100$. A destra: $N = 1000$.*

A conclusione di questa analisi sembrerebbe che l'intervallo di Wald sia il migliore tra quelli considerati; tuttavia, accanto alla probabilità di copertura, si deve tener conto di un altro fattore di non minore importanza nella valutazione della bontà di un intervallo di confidenza, ovvero la sua lunghezza.

4.2 Lunghezza attesa degli intervalli

L'intervallo "ottimo" per la stima del parametro p di interesse deve presentare la massima probabilità di copertura e, nello stesso tempo, la minima lunghezza possibile. Fin qui si è presa in considerazione la probabilità di copertura; si esamina ora il comportamento dei tre metodi in termini di lunghezza degli intervalli costruiti.

Il procedimento seguito per costruire la lunghezza attesa è il seguente: per ogni possibile coppia (n, p) si generano $r = 1000$ realizzazioni di X campionando dalla distribuzione ipergeometrica $Ipg(N, M = pN, n)$, si costruiscono gli r intervalli corrispondenti (con i metodi Wald, Wilson e Buonaccorsi) e si calcolano le lunghezze degli r intervalli, pari a $1 - p_L$. La lunghezza attesa dell'intervallo per n e p fissati è stimata dalla media delle lunghezze degli r intervalli costruiti.

Si esaminino le Figure 6 e 7. Mentre, come ci si aspetta, la lunghezza attesa è minore o al più uguale a 1 per gli intervalli di Wilson e Buonaccorsi, nel caso dell'intervallo generato dal metodo Wald essa è maggiore di 1 per tutti i valori di p . Ciò accade perchè, per quei valori di n e p , il metodo Wald fornisce spesso stime negative dell'estremo inferiore p_L dell'intervallo di confidenza unilatero, fatto che genera le alte probabilità di copertura registrate nella sezione precedente. A parte questa anomalia si riscontra che l'intervallo generato dal metodo Wilson è caratterizzato dalla lunghezza media minima, seguito da quello di Buonaccorsi e infine da quello di Wald. Eseguendo le simulazioni per $N=500$ si ottengono risultati molto simili al caso $N=1000$; in particolare la lunghezza media dell'in-

tervallo generato dal metodo Buonaccorsi è sempre maggiore o uguale di quella dell'intervallo generato dal metodo Wilson.

Nel caso $N=10000$ con $n=40$ il comportamento del metodo Wald migliora decisamente: infatti a partire da $p=0.05$ la lunghezza media dell'intervallo di Wald si mantiene inferiore o pari all'unità; gli altri due metodi, e in particolare il metodo Wilson, generano degli intervalli più precisi.

4.3 Le prime conclusioni

L'analisi condotta rispetto alla lunghezza attesa degli intervalli costruiti con i tre metodi, porta dunque a escludere l'intervallo generato dal metodo Wald come possibile metodo risolutivo per il problema dell'Autorità in quanto per buona parte dei possibili valori di interesse per n e p esso non restringe l'ovvio insieme di valori possibili per p , ossia l'intervallo $[0,1]$.

Terminate le analisi descritte, il MOX sottopone i risultati all'attenzione dell'Autorità, la quale stabilisce di proporre agli esercenti, nel Documento di consultazione [1] del Luglio 2004, gli intervalli generati dai metodi Buonaccorsi e Wilson.

Gli esercenti accolgono inizialmente il metodo Buonaccorsi con una certa diffidenza poichè la formula che definisce M_L , nella sua formulazione (1), è di difficile interpretazione. D'altro canto gli esempi numerici mostrano come le stime di M_L ottenute col metodo Buonaccorsi siano sempre più piccole di quelle ottenute col metodo Wilson. Questo aspetto interessa particolarmente gli esercenti, dal momento che, nello stesso Documento di consultazione, si propone di sanzionare economicamente gli esercenti inadempienti in termini proporzionali alla stima M_L .

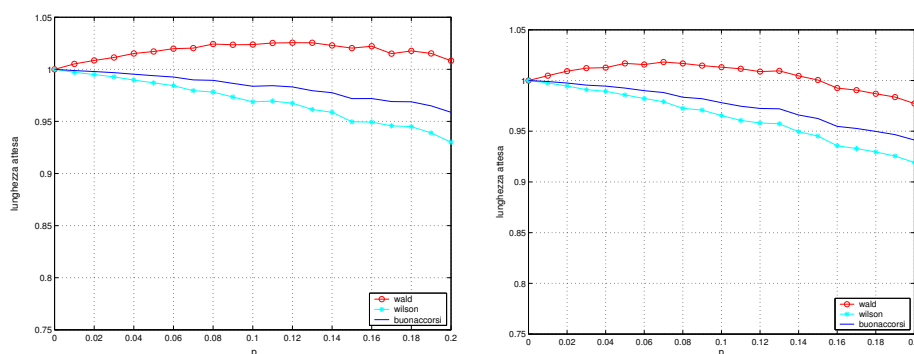


Figura 6: *Stima della lunghezza attesa degli intervalli generati dai metodi di Wald, Wilson e Buonaccorsi con $N=100$, $\alpha=0.05$, $p_{max}=0.2$, passo di $p=0.01$. A sinistra: $n = 5$. A destra: $n = 10$.*

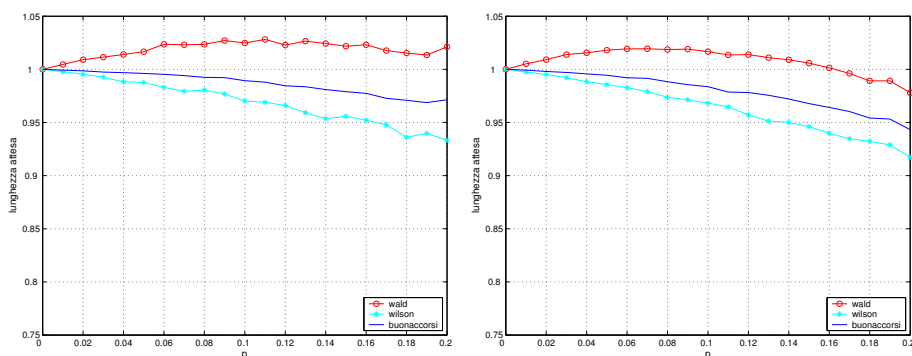


Figura 7: Stima della lunghezza attesa degli intervalli generati dai metodi di Wald, Wilson e Buonaccorsi con $N=1000$, $\alpha=0.05$, $p_{max}=0.2$, passo di $p=0.01$. A sinistra: $n = 5$. A destra: $n = 10$.

5 Il problema diventa più complesso: un metodo di controllo a due passi

Nel periodo di elaborazione delle analisi sin qui esposte, l'Autorità realizza che, in effetti, per alcune prestazioni, può risultare impossibile per l'ispettore effettuare il controllo di ottemperanza alle direttive, come ad esempio nel caso in cui per la prestazione "preventivazione" manchi la copia del preventivo, ossia l'oggetto stesso della prestazione. Aeeg decide pertanto di contemplare due gradi di non coerenza con le direttive, l'uno più grave, di "non validità", come nel caso della preventivazione in assenza di preventivo, l'altro, di "non conformità", di entità più lieve (ad esempio un preventivo privo di data). Si delinea dunque una metodologia di controllo in due passi successivi. Si ipotizza che un ispettore estragga un campione casuale dall'universo di prestazioni fornite dall' esercente. In primo luogo le prestazioni del campione vengono osservate rispetto alla variabile dicotomica *validità*: ogni prestazione del campione viene dichiarata *valida* o *non-valida* sulla base dei criteri che l'Autorità definirà a seguito del processo di consultazione. Obiettivo di questa analisi è la stima del numero di prestazioni non valide nella popolazione di prestazioni dalla quale è stato estratto il campione. Le prestazioni del campione classificate come non valide sono scartate dal campione; ciò che rimane è considerato campione della popolazione delle prestazioni valide. A questo punto una seconda analisi stima il numero di prestazioni *non conformi* tra quelle valide.

La definizione del problema in questi termini avviene durante il mese di Luglio del 2004, quando la metodologia di controllo in due tempi viene proposta nel Documento di consultazione [1] in cui viene anche definita la notazione di seguito adottata. Il primo passo dell'ispezione ha come obiettivo il calcolo della stima M_v del numero N_1 di prestazioni non valide nella popolazione attraverso

il numero di prestazioni non valide osservate nel campione. Il secondo passo del controllo prevede che si escludano dal campione le prestazioni non valide e le restanti prestazioni si considerino campione per la stima M_c del numero N_2 di prestazioni non conformi nella popolazione. In questo secondo passo, la dimensione della popolazione di riferimento viene ridotta a $N - M_v$, che rappresenta una stima del numero $N - N_1$ di prestazioni valide nella popolazione globale.

Alla fine del processo di consultazione Aeeg decide di adottare solo l'intervallo generato dal metodo Buonaccorsi sia per la stima di M_v che per la stima di M_c . Trattandosi di un metodo conservativo, la stima M_v di N_1 ottenuta con (1) sarà minore o uguale al vero valore di N_1 almeno nel $(1 - \alpha) * 100\%$ delle ispezioni. Questa proprietà è goduta anche dalla stima fornita, nel secondo passo della metodologia, per il calcolo di M_c . Ma i due passi della metodologia sono susseguenti e il risultato del primo condiziona quello del secondo in quanto modifica la dimensione della popolazione di riferimento nel secondo passo. Emerge dunque l'esigenza di valutare non più solo la legge di M_v o M_c separatamente, quanto piuttosto la distribuzione congiunta del vettore aleatorio (M_v, M_c) , esito finale dell'ispezione, ossia di valutare le probabilità di errore di un'ispezione condotta utilizzando il metodo Buonaccorsi sia per la stima di M_v che per la stima di M_c . L'Autorità fornisce quindi alcuni set di dati di riferimento per ciascuno dei quali si ipotizza di conoscere la dimensione della popolazione N , il numero di prestazioni non valide N_1 e il numero di prestazioni non conformi N_2 . La scelta dei dati è fatta in modo da rappresentare le situazioni che l'Autorità ritiene possibili sia in merito alla dimensione della popolazione N che in relazione ad N_1 ed N_2 (cioè al grado di scorrettezza dell' esercente). In particolare N varia, nelle simulazioni effettuate, da un minimo di 250 ad un massimo di 4000 unità. Per ciascuna popolazione si ipotizzano tre situazioni: la prima rappresenta un comportamento sostanzialmente corretto dell' esercente ($N_1 \leq 1\%$ della popolazione e $N_2 \leq 3\%$), la seconda un non rispetto di lieve entità ($1\% < N_1 \leq 3\%$ della popolazione e $3\% < N_2 \leq 9\%$), infine la terza un non rispetto di grave entità ($N_1 > 3\%$ della popolazione e $N_2 > 9\%$). Analogamente, le simulazioni vengono condotte immaginando varie dimensioni possibili per il campione. Si illustra in Figura 8 la sintesi grafica di una delle analisi effettuate. Il grafico tridimensionale in basso a sinistra mostra la densità congiunta di (M_v, M_c) , il grafico in alto a destra fornisce le stesse informazioni proiettate sul piano M_v, M_c : i colori sono indicativi della densità di probabilità, quanto più chiaro è il colore, tanto maggiore è la probabilità di osservare la coppia (M_v, M_c) nell'area colorata, e viceversa. Le linee azzurre corrispondono ai veri valori di N_1 ed N_2 (noti). Le due linee suddividono il piano in quattro aree, in corrispondenza di ciascuna delle quali è scritta la probabilità complessiva della regione. Ad esempio, la regione in basso a sinistra è costituita da coppie in cui sia M_v che M_c sono minori dei veri valori N_1 ed N_2 , di cui sono stima. Viceversa, la regione in alto a destra contiene coppie (M_v, M_c) in cui entrambi i valori sono sovrastimati. I due istogrammi rappresentano le densità marginali di M_v (in basso a destra) e di M_c (in alto a

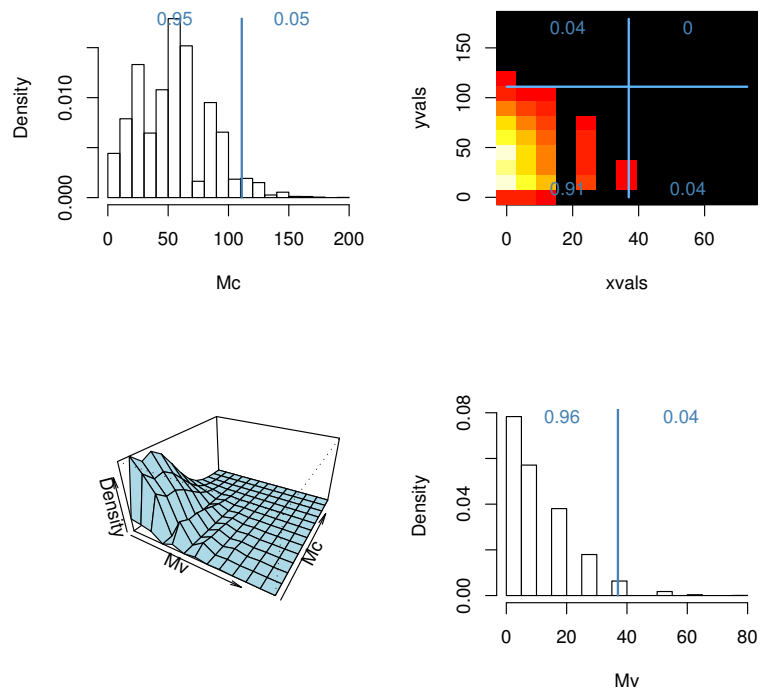


Figura 8: Distribuzione congiunta di (M_v, M_c) e distribuzioni marginali di M_v ed M_c .

sinistra). Una analisi qualitativa dei numerosi grafici prodotti evidenzia che la coppia di stime esito del controllo sottostima contemporaneamente entrambi i valori di N_1 ed N_2 (penalizzando quindi l'esercente meno di quanto si dovrebbe) con una probabilità oscillante tra 0.9 e 0.93, mentre la probabilità di ottenere una coppia (M_v, M_c) in cui entrambe le stime sono superiori ai veri valori è prossima allo 0 in tutti gli esempi considerati. Infine, la probabilità di ottenere una coppia in cui una sola delle stime M_v o M_c è superiore al vero valore N_1 o N_2 rispettivamente, oscilla da un minimo di 0.06 ad un massimo di 0.1.

Le analisi illustrate, ed una serie di altre analisi che non si riportano per brevità, convincono l'Autorità dell'adeguatezza del metodo Buonaccorsi agli scopi che l'Autorità stessa si era prefissata. La Delibera n.168/04 sancisce nel Titolo III, Art.70, l'adozione della metodologia di controllo in due passi basata sul campionamento casuale semplice senza reimmissione ed il ricorso alla distribuzione ipergeometrica per la stima di N_1 e N_2 . La stessa Delibera n.168/04 specifica che la metodologia messa a punto riguarda i dati relativi alla qualità commerciale del servizio di vendita e del servizio di distribuzione, nonché quelli relativi alla sicurezza e continuità del servizio di distribuzione del gas.

6 Ingegnerizzazione della metodologia

Nella Delibera n.168/04, contestualmente alla adozione della metodologia in due passi, l'Autorità riconosce ai distributori la necessità di un periodo di sperimentazione che confermi l'adeguatezza della disciplina (anche in senso più generale) e l'applicabilità dei criteri di validazione e di conformità. L'Autorità incarica il MOX di elaborare un codice software che permetta agli ispettori di ottenere in tempo reale le stime M_v ed M_c a partire dai dati campionari. Inoltre Aeeg richiede che il software venga reso disponibile anche agli esercenti in modo che questi ultimi possano simulare una visita ispettiva ed esercitarsi sul funzionamento del metodo.

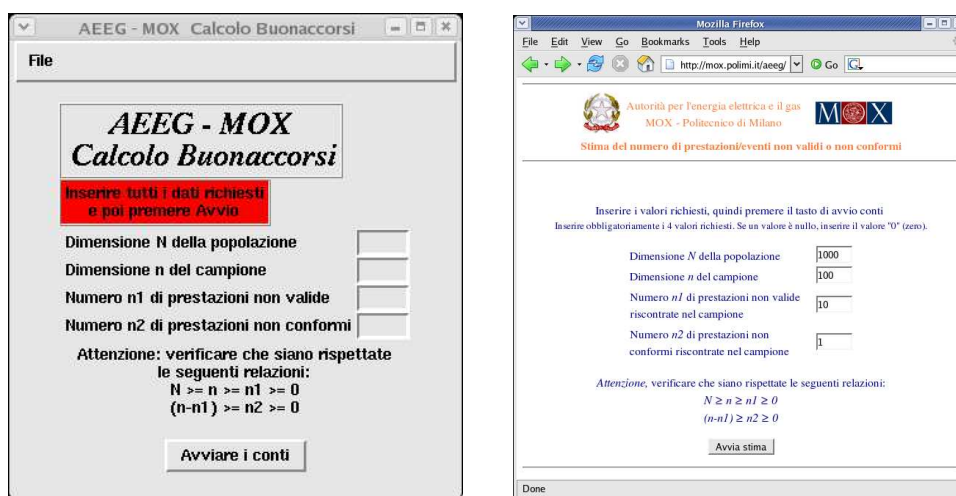


Figura 9: *Ingegnerizzazione del software: interfacce d'uso. A sinistra, interfaccia nativa. A destra, interfaccia web.*

Il MOX sceglie di utilizzare a tal scopo il programma R, distribuito sotto licenza GPL, liberamente scaricabile dal sito www.r-project.org. Tale programma è specializzato per la statistica e quindi particolarmente efficiente per problemi di campionamento. Viene elaborato un programma che, a partire dalla dimensione N della popolazione e dai dati campionari (numerosità n del campione, numero n_1 di prestazioni non valide e numero n_2 di prestazioni non conformi riscontrate nel campione), restituisce in uscita le stime M_v ed M_c di N_1 ed N_2 rispettivamente. Per rendere il programma utilizzabile da parte degli ispettori e degli esercenti, sono state sviluppate due interfacce grafiche, l'una installabile su calcolatore, l'altra di tipo web. L'interfaccia per PC, che verrà utilizzata dagli ispettori durante le prossime campagne ispettive, utilizza le estensioni Tcl/Tk del software R, automaticamente installate assieme ad R stesso e anch'esse distribuite sotto licenza GPL. La pagina web, visibile all'indirizzo <http://mox.polimi.it/aeeeg>, è

scritta in linguaggio PHP, interfacciato con R tramite il pacchetto CGIwithR. Sulla pagina web è anche visibile la function utilizzata scritta in linguaggio R.

7 Conclusioni

In questo lavoro si racconta brevemente il processo di costruzione della metodologia di auditing sviluppata in collaborazione tra l’Autorità per l’energia elettrica e il gas e il MOX del Politecnico di Milano nel periodo Ottobre 2003/Marzo 2005. L’elaborazione del metodo, finalizzato al controllo della qualità commerciale e tecnica (sicurezza e continuità) dei servizi erogati dagli esercenti nei settori dell’energia elettrica e del gas, è avvenuta attraverso raffinamenti successivi indotti dal costante contributo delle parti interessate. Il sistema elaborato è stato implementato attraverso un programma scritto in linguaggio R installato sui calcolatori utilizzati dagli ispettori. La metodologia elaborata, pubblicata sulla Gazzetta Ufficiale n.245/04, è del tutto nuova nel settore e pertanto è attualmente considerata in fase di sperimentazione.

Riferimenti bibliografici

- [1] AUTORITÀ PER L’ENERGIA ELETTRICA E IL GAS Documento di consultazione “Regolazione della qualità dei servizi di distribuzione, misura e vendita del gas” consultabile all’indirizzo www.autorita.energia.it/docs/dc/dc-040715.pdf.
- [2] AUTORITÀ PER L’ENERGIA ELETTRICA E IL GAS Delibera n. 168/04 del 29/09/04:
 - Allegato A “Testo integrato delle disposizioni dell’Autorità per l’Energia Elettrica e il Gas in materia di qualità dei servizi di distribuzione, misura e vendita del gas”, Parte IV - Verifica dei dati di qualità, Titolo III, pag.68;
 - Relazione tecnica, Sezione 10, pag.24;consultabili all’indirizzo www.autorita.energia.it/docs/04/168-04all.pdf.
- [3] AGRESTI, A. e COULL, B. A. (1998). Approximate is better than “exact” for interval estimation of binomial proportions. *American Statistician*, 52, 119–126.
- [4] BARNETT, V. (2002). *Sample Survey. Principles and Methods*. Arnold, London.
- [5] BROWN, L.D., CAI, T.T. e DASGUPTA, A. (2001). Interval estimation for binomial proportions. *Statistical Science*, 16, 101–133.
- [6] BUONACCORSI, J. P. (1987). A note on confidence interval for proportions in finite populations. *American Statistician*, 41, 215–218.

- [7] COCHRAN, W.G. (1977). Sampling Techniques. Third edition. Wiley, New York.
- [8] HENDERSON, M., MEYER, M.C. (2001). Exploring the confidence interval for a binomial parameter in a first course in statistical computing. *American Statistician* 55, 337-344.
- [9] OLGATI E. Intervalli di confidenza per una proporzione generati da campionamento di una popolazione finita. Tesi di laurea di primo livello, MOX, Politecnico di Milano, a.a. 2003 - 2004.
- [10] RAO, J.N.K. (1968). Some small sample results in ratio and regression estimation. *Journ. Ind. Stat. Assoc.*, 6, 160–168.
- [11] WILSON, E.B. (1927). Probable Inference, the Law of Succession and Statistical Inference. *Journal of the American Statistical Association*, 22, 209-212.