

UNIVERSITÀ DEGLI STUDI DI MILANO–BICOCCA  
SCUOLA DI ECONOMIA E STATISTICA

CORSO DI LAUREA IN  
SCIENZE STATISTICHE ED ECONOMICHE



# METODI DI STIMA INTERVALLARE IN UN MODELLO BINOMIALE

RELATORE: Dott. Tommaso Rigon

TESI DI LAUREA DI:  
Stefano Andreoli  
MATRICOLA N. 880053

ANNO ACCADEMICO 2023/2024



# Indice

<b>1</b>	<b>Verosimiglianza e stima intervallare</b>	<b>3</b>
1.1	Il modello statistico . . . . .	3
1.1.1	Modelli parametrici . . . . .	3
1.2	La verosimiglianza statistica . . . . .	4
1.2.1	La funzione di verosimiglianza . . . . .	4
1.2.2	Famiglie Esponenziali . . . . .	5
1.2.3	Stima di massima verosimiglianza . . . . .	5
1.2.4	Informazione di Fisher . . . . .	7
1.2.5	Teoria asintotica della verosimiglianza . . . . .	9
1.3	Verifica d'ipotesi . . . . .	13
1.3.1	Il sistema d'ipotesi . . . . .	13
1.3.2	Test statistico e statistica test . . . . .	13
1.3.3	Errori di I e II tipo . . . . .	14
1.3.4	Livello di significatività del test . . . . .	14
1.3.5	Livello di significatività osservato . . . . .	15
1.3.6	Tre test connessi alla verosimiglianza . . . . .	15
1.4	Stima intervallare . . . . .	19
1.4.1	Quantità-Pivot . . . . .	19
1.4.2	Test e regioni di confidenza, impostazione di Neyman . . . . .	20
1.4.3	Procedimenti approssimati . . . . .	21
1.4.4	Valutazione dei procedimenti di stima intervallare . . . . .	21
<b>2</b>	<b>Stima intervallare nel modello binomiale</b>	<b>23</b>
2.1	La variabile casuale binomiale . . . . .	23
2.1.1	Momenti . . . . .	23
2.1.2	Verosimiglianza e modello binomiale . . . . .	24
2.2	Tre test connessi alla verosimiglianza . . . . .	25
2.2.1	Log-rapporto di verosimiglianza nel caso binomiale . . . . .	25

2.2.2	Test di Wald nel caso binomiale . . . . .	25
2.2.3	Test di Wilson nel caso binomiale . . . . .	31
2.3	Intervalli di confidenza . . . . .	33
2.3.1	Intervallo di Clopper-Pearson . . . . .	33
2.3.2	Intervallo di Blaker . . . . .	38
2.3.3	Intervallo di confidenza dal Log-rapporto di verosimiglianza	40
2.3.4	Intervallo di confidenza di Wald . . . . .	43
2.3.5	Intervallo di confidenza di Wilson . . . . .	45
2.3.6	Intervallo di credibilità . . . . .	48
2.3.7	Confronto pratico tra intervalli . . . . .	49
<b>3</b>	<b>Copertura e ampiezza</b>	<b>51</b>
3.1	Confronto con il coefficiente di confidenza . . . . .	51
3.1.1	Probabilità di copertura . . . . .	52
3.1.2	Probabilità media di copertura . . . . .	56
3.2	Ampiezza dell'intervallo . . . . .	58
<b>4</b>	<b>Riparametrizzazioni</b>	<b>61</b>
4.1	Quote . . . . .	61
4.1.1	Tre test di verosimiglianza per le quote . . . . .	62
4.2	Intervalli di confidenza per le quote . . . . .	63
4.2.1	Clopper-Pearson . . . . .	64
4.2.2	Blaker . . . . .	64
4.2.3	Log-rapporto di verosimiglianza . . . . .	65
4.2.4	Wald . . . . .	65
4.2.5	Wilson . . . . .	66
4.2.6	Bayesiano . . . . .	68
4.2.7	Applicazione ai dati . . . . .	69
4.2.8	Valutazione degli intervalli . . . . .	70
4.3	Log-quote . . . . .	76
4.3.1	Tre test di verosimiglianza per le log-quote . . . . .	77
4.4	Intervalli di confidenza per le log-quote . . . . .	78
4.4.1	Clopper-Pearson . . . . .	78
4.4.2	Blaker . . . . .	78
4.4.3	Log-rapporto di verosimiglianza . . . . .	78
4.4.4	Wald . . . . .	79

4.4.5	Wilson . . . . .	79
4.4.6	Bayesiano . . . . .	81
4.4.7	Applicazione ai dati . . . . .	81
4.4.8	Valutazione degli intervalli . . . . .	82
<b>Bibliografia</b>		<b>91</b>
<b>A Codice R</b>		<b>93</b>



## **Ringraziamenti**

Ringrazio la mia famiglia e i miei amici. Ringrazio anche il mio relatore, Tommaso Rigon.



# Introduzione

L'inferenza statistica è un ambito cardine della statistica e si articola in diverse branche, essa verte sull'ottenimento di informazioni riguardante una popolazione d'indagine per via induttiva, ovvero sfruttando un campione derivante dalla suddetta popolazione. Uno dei problemi d'interesse dell'inferenza è la stima intervallare, la cui fondazione teorica verrà approfondita nel primo capitolo. L'idea di questa tesi parte da un articolo, [Agresti & Coull \(1998\)](#), il quale confronta due classi di metodi di stima intervallare: esatti ed approssimati. L'obiettivo dell'articolo è legittimare l'uso di questi ultimi. Nell'elaborato ci si concentra su un modello binomiale, in questo contesto la questione è stata ampliata risultando in un'analisi approfondita di diversi metodi di stima intervallare, mirata ad individuare una gerarchia plausibile per questi. Dei suddetti metodi verranno mostrati la loro costruzione, la forma analitica degli estremi (se presente) e gli intervalli risultanti dalla loro applicazione ad un insieme di dati. Si procederà poi con una valutazione dei metodi al fine di categorizzarli in base alla loro precisione nella stima, ma anche alla loro capacità di adattamento; essa sarà esaminata nell'ultimo capitolo, applicando i medesimi metodi a modelli con parametrizzazioni diverse ed esaminando la performance dei vari approcci. Come spesso accade, non si potrà rilevare un metodo perfetto o costantemente superiore agli altri, ma si potranno delineare le caratteristiche di ogni approccio presente, per effettuare una scelta consapevole tra questi.



## Capitolo 1

# Verosimiglianza e stima intervallare

Questo capitolo contiene un'introduzione teorica ai concetti di verosimiglianza, test d'ipotesi, stima intervallare e inferenza statistica. L'argomento su cui verteranno i capitoli successivi sarà la stima intervallare, un ambito della statistica inferenziale. Essa consiste nell'individuazione di un insieme di valori considerati plausibili per un parametro di interesse. Tale procedura sarà applicata ad un modello binomiale, che assume grande importanza in ambiti come quello medico e biologico.

### 1.1 Il modello statistico

Si suppone di avere un insieme di dati  $y$  detto *campione* e si considera questo campione come una determinazione di una certa v.c.  $Y$ . L'obiettivo dell'inferenza statistica è quello di determinare alcune caratteristiche della distribuzione  $F_0$  di  $Y$  tramite le informazioni fornite dal campione  $y$ . Il processo di generazione del campione è detto *esperimento casuale*. Nello studio della distribuzione della v.c.  $Y$  le affermazioni fatte saranno inevitabilmente affette da incertezza. L'approccio inferenziale tenta di ridurre il più possibile questa incertezza ma soprattutto di quantificarla. Il *modello statistico*  $\mathcal{F}$  contiene le caratteristiche principali riguardanti il contesto di campionamento ed il fenomeno che ha generato il campione  $y$ . Per semplicità si ipotizza che  $Y$  sia una v.c. a componenti indipendenti e identicamente distribuite ovvero che  $y$  sia un *campione casuale semplice*.

#### 1.1.1 Modelli parametrici

Il modello statistico  $\mathcal{F}$  può contenere funzioni di ripartizioni diverse, ma il modello d'interesse è quello per cui le funzioni contenute in  $\mathcal{F}$  differiscono tra di

loro per il valore di un parametro  $\theta$  il cui supporto è  $\Theta$ . Nel caso discreto  $\mathcal{F}$  sarà un insieme di funzioni di probabilità, nel caso continuo di funzioni di densità di probabilità

$$\mathcal{F} = \{f(\cdot; \theta) : \theta \in \Theta \subseteq \mathbb{R}^k\}, \quad (1.1)$$

dove  $f$  è detta *funzione del modello*,  $\theta$  è il *parametro* e  $\Theta$  è lo *spazio parametrico*.

Se la distribuzione  $F_0$  della v.c.  $Y$  appartiene al modello  $\mathcal{F}$  si dirà che il modello è *correttamente specificato*. Il valore del parametro  $\theta$  associato a  $F_0$  è il vero valore del parametro, detto  $\theta_0$ .

Lo *spazio campionario*  $\mathcal{Y}$  è l'insieme di tutti i possibili campioni ottenibili dato il modello statistico. Si può inoltre utilizzare una funzione biunivoca  $h(\cdot)$  da  $\Theta$  in  $\Psi$  per effettuare una *riparametrizzazione*. Il modello  $\mathcal{F}$  diventerà  $\mathcal{F} = \{f(\cdot; \psi) : \psi = h(\theta), \theta \in \Theta\} = \{f(\cdot; \psi) : \psi \in \Psi\}$ , definendo  $\Psi = \{\psi : \psi = h(\theta), \theta \in \Theta\}$ . Per un modello è possibile scegliere tra parametrizzazioni diverse, la scelta fra queste può essere dettata da necessità di ricerca oppure per semplicità di calcolo. Si definisce adesso la proprietà di *identificabilità* di un modello statistico. Un modello statistico si dice identificabile se a ogni coppia  $\theta_1 \neq \theta_2$  corrispondono distribuzioni di probabilità diverse, ovvero  $\mathbb{P}\{Y \in B \mid \theta_1\} \neq \mathbb{P}\{Y \in B \mid \theta_2\}$ , dove  $B \subseteq \mathcal{Y}$ .

## 1.2 La verosimiglianza statistica

### 1.2.1 La funzione di verosimiglianza

Dato un modello del tipo (1.1), è possibile considerare il campione  $y$  come fissato e vedere la funzione del modello come una funzione del parametro  $\theta$ . La funzione così concepita è chiamata *funzione di verosimiglianza* e rappresenta la probabilità che un campione  $y$  sia stato generato da una distribuzione con parametro  $\theta$ .

La funzione di verosimiglianza è una funzione  $L: \Theta \rightarrow \mathbb{R}^+$ ; dato  $y \in \mathcal{Y}$ , viene definita da [Azzalini \(2008\)](#) come

$$L(\theta) = L(\theta; y) = c(y)f(y; \theta),$$

con  $c(y)$  costante positiva non dipendente da  $\theta$ . Poiché al variare di  $c(y)$  si ottengono funzioni diverse, la verosimiglianza è tecnicamente una *classe di equivalenza* di funzioni i cui elementi differiscono per una costante moltiplicativa. Essendo una funzione solo di  $\theta$ , l'interpretazione è la seguente: avendo osservato il campione  $y$ , considerando due valori  $\theta_1 \in \Theta$ ,  $\theta_2 \in \Theta$ ; come valore del parametro è più verosimile  $\theta_1$  se  $L(\theta_1) > L(\theta_2)$ . La verosimiglianza  $L(\theta)$  ha valori in  $\mathbb{R}^+$ , quindi risulta possibile applicare una trasformazione logaritmica

$$\ell(\theta) = \log L(\theta),$$

$\ell(\theta)$  viene chiamata *funzione di log-verosimiglianza*, con la convenzione  $\ell(\theta) = -\infty$  quando  $L(\theta) = 0$ .

### 1.2.2 Famiglie Esponenziali

Una *classe parametrica* costituisce una famiglia esponenziale se i suoi elementi sono del tipo:

$$f(y; \theta) = q(y) \exp \left( \sum_{i=1}^r \psi_i(\theta) t_i(y) - \tau(\theta) \right).$$

Quando le  $\{1, \psi_1(\theta), \dots, \psi_r(\theta)\}$  sono linearmente indipendenti in  $\theta$  (se non lo fossero ci si può ricondurre ad un numero minore di  $\psi$  che soddisfino questa condizione) la famiglia esponenziale è scritta in forma ridotta ed  $r$  sarà chiamato *ordine* della famiglia.

### 1.2.3 Stima di massima verosimiglianza

In un problema di stima puntuale l'obiettivo è scegliere il valore di  $\theta$  più credibile tenendo conto dei dati osservati  $y$ ; questo si chiamerà *stima puntuale* di  $\theta$ . Ad ogni elemento  $y \in \mathcal{Y}$  dovrà corrispondere un valore  $\hat{\theta}$  in  $\Theta$  perciò il procedimento di stima sarà una funzione da  $\mathcal{Y}$  in  $\Theta \subseteq \mathbb{R}^k$ , esso prende il nome di *stimatore*.

Esistono vari criteri, uno di questi è il *metodo dei momenti*, il quale non verrà approfondito. A questo viene solitamente affiancato il *metodo della massima verosimiglianza* data la sua ampia applicabilità ed efficienza.

### Stime di massima verosimiglianza

Il valore  $\hat{\theta}$  tale che

$$L(\hat{\theta}) = \sup_{\theta \in \Theta} L(\theta)$$

è definito come *stimatore di massima verosimiglianza* o SMV. Alcune note sulla SMV:

- Non è detto che la SMV esista,
- La SMV può non essere unica,
- Spesso  $\hat{\theta}$  non è esprimibile esplicitamente come funzione dei dati campionari (in quel caso andrà ricavato per via numerica).

Tuttavia, in molti casi, la SMV esiste ed è unica.

### Proprietà di equivarianza

**Teorema 1.1.** Sia  $\psi: \Theta \rightarrow \Psi$  una funzione biunivoca. Allora la SMV di  $\psi(\theta)$  è  $\psi(\hat{\theta})$ , dove  $\hat{\theta}$  è la SMV di  $\theta$  relativa a  $L(\theta)$ .

Questa proprietà garantisce la coerenza dei risultati di stima con parametrizzazioni diverse. La funzione  $\psi(\cdot)$  è biunivoca, se non lo fosse bisognerebbe introdurre la *verosimiglianza indotta*, definita come:

$$L_{\Psi}(\psi_0) = \sup_{\{\theta: \psi(\theta) = \psi_0\}} L(\theta).$$

### Equazioni di verosimiglianza

Per determinare la SMV si richiede di massimizzare  $L(\theta)$  o, alternativamente,  $\ell(\theta)$ . Posto che  $L(\theta)$  sia derivabile e quindi anche  $\ell(\theta)$ , solitamente  $\hat{\theta}$  si otterrà risolvendo l'equazione (o sistema di equazioni) di verosimiglianza:

$$\frac{d}{d\theta} \ell(\theta) = 0,$$

il sistema si ha nel caso di  $\theta$  multidimensionale. La risoluzione di questa non porta necessariamente al punto di massimo globale di  $L(\theta)$ .

### 1.2.4 Informazione di Fisher

#### Informazione osservata di Fisher

Nell'ipotesi che la funzione di log-verosimiglianza sia sufficientemente regolare, attuando uno sviluppo di Taylor attorno a  $\hat{\theta}$  si ottiene:

$$\ell(\theta) = \ell(\hat{\theta}) + \ell'(\hat{\theta})(\theta - \hat{\theta}) + \frac{1}{2}\ell''(\hat{\theta})(\theta - \hat{\theta})^2 + \dots$$

Poiché  $\hat{\theta}$  è stima SMV,  $\ell'(\hat{\theta}) = 0$ , di conseguenza

$$\ell(\theta) = \ell(\hat{\theta}) + \frac{1}{2}\ell''(\hat{\theta})(\theta - \hat{\theta})^2 + \dots$$

dove  $\ell(\hat{\theta}), \ell''(\hat{\theta})$  sono costanti rispetto a  $\theta$ . Il comportamento di  $\ell(\theta)$  in un intorno di  $\hat{\theta}$  è pertanto determinato da  $\ell''(\hat{\theta})$  che misura la curvatura locale della log-verosimiglianza. Questa grandezza rappresenterà dunque quanto si distinguono nello spazio parametrico i valori con alta verosimiglianza (la stima SMV nel nostro caso) dai valori con verosimiglianza inferiore. Per questo motivo, si introduce la quantità non negativa  $\mathcal{J}(\hat{\theta}) = -\ell''(\hat{\theta})$  che prende il nome di *informazione osservata di Fisher*. Considerato quanto appena detto, questo indice può essere visto come una misura della preferenza che la verosimiglianza assegna a  $\hat{\theta}$  rispetto ad altri valori di  $\theta$ .

#### Distorsione e SMV

Si vuole valutare quanto la SMV sia valida nell'affrontare problemi di stima e derivare alcune proprietà formali di questa relative alla sua accuratezza. Il campione  $y$  è una determinazione di una v.c.  $Y$ , ragion per cui  $\hat{\theta}$  sarà anch'esso stocastico in quanto funzione di  $y$ . Si immagini di poter replicare il processo di estrazione del campione  $y$  ciascuno con il suo valore di  $\hat{\theta}$  e si consideri l'intera distribuzione di  $\hat{\theta}$  ottenuta. Nella valutazione di uno stimatore questa distribuzione assume grande importanza. La variabile casuale che tuttavia assume maggiore interesse è  $\hat{\theta} - \theta_0$  ovvero la differenza tra la stima e il valore vero del parametro. Le valutazioni basate sull'analisi di questa distribuzione saranno a priori rispetto all'esperimento.

Tra le proprietà desiderabili che può avere questa distribuzione vi sono:

- Il fatto che sia centrata sul valore  $\theta$ , ovvero

$$\mathbb{E} [\hat{\theta} | \theta] = \theta \quad \text{per ogni } \theta \in \Theta,$$

con l'assunzione che il valore medio esista. Quando questa proprietà viene soddisfatta lo stimatore viene detto *non distorto* in media. La *distorsione* è il valore  $\mathbb{E} \{ \hat{\theta} | \theta \} - \theta$ .

- La *non distorsione asintotica*, ovvero

$$\lim_{n \rightarrow \infty} \mathbb{E} \{ \hat{\theta} | \theta \} = \theta \quad \text{per ogni } \theta \in \Theta,$$

dove  $n$  è la dimensione campionaria.

La SMV però è spesso una funzione complessa dei dati e il valor medio della distribuzione  $\hat{\theta} - \theta_0$  non permette una completa valutazione dello stimatore. Al fine di individuare alcune proprietà generali della SMV, è opportuno introdurre determinate assunzioni.

### Problemi regolari di stima

Si definisce *problema regolare di stima* un contesto in cui si verificano le seguenti condizioni:

- Il modello statistico è *identificabile*,
- Lo spazio parametrico  $\Theta$  è un intervallo aperto di  $\mathbb{R}^k$ ,
- Le funzioni di densità (o probabilità) specificate dal modello hanno tutte lo stesso supporto,
- Per la funzione  $f$  si può scambiare due volte il segno di integrale con quello di derivata rispetto a  $\theta$

$$\int_{\mathbf{y}} \frac{\partial}{\partial \theta} f(\mathbf{y}; \theta) \, d\nu(\mathbf{y}) = \frac{d}{d\theta} \int_{\mathbf{y}} f(\mathbf{y}; \theta) \, d\nu(\mathbf{y})$$

$$\int_{\mathbf{y}} \frac{\partial^2}{\partial \theta^2} f(\mathbf{y}; \theta) \, d\nu(\mathbf{y}) = \frac{d^2}{d\theta^2} \int_{\mathbf{y}} f(\mathbf{y}; \theta) \, d\nu(\mathbf{y}).$$

Queste assunzioni non sono eccessivamente restrittive, in caso di problemi di non identificabilità un restringimento dello spazio  $\Theta$  può risolverla.

### Informazione attesa di Fisher

La funzione punteggio di Fisher è:

$$u(\theta) = u(\theta; y) = \frac{\partial}{\partial \theta} \ell(\theta; y).$$

Si può dimostrare che il suo valore atteso è pari a zero, perciò la varianza coincide con il momento secondo, che prende il nome di *informazione attesa di Fisher*

$$I(\theta) = \mathbb{E} \left\{ u(\theta) u(\theta)^T \mid \theta \right\}.$$

Essa è una matrice semidefinita positiva (se  $k = 1$  è uno scalare) e si ha che

$$I(\theta) = \mathbb{E} \left\{ u(\theta)^2 \mid \theta \right\} = -\mathbb{E} \left\{ \frac{d}{d\theta} u(\theta) \mid \theta \right\}.$$

L'informazione attesa di Fisher gode di alcune proprietà, come la proprietà additiva in caso di v.c.  $Y_1, Y_2$  indipendenti:  $I(\theta) = I_1(\theta) + I_2(\theta)$ . Questo comporta che nel caso di  $n$  osservazioni da variabili casuali indipendenti  $I(\theta) = ni(\theta)$ , dove  $i(\theta)$  è l'informazione attesa della singola osservazione. Per una riparametrizzazione da  $\theta$  a  $\psi(\theta)$  con  $\psi(\cdot)$  funzione monotona derivabile, l'informazione attesa di Fisher diventa

$$\{\psi'(\theta)\}^{-2} I(\theta) \Big|_{\theta=\theta(\psi)}$$

dove  $\theta(\cdot)$  è la funzione inversa di  $\psi(\cdot)$ .

#### 1.2.5 Teoria asintotica della verosimiglianza

In questa sezione verranno discusse le proprietà asintotiche della SMV sotto specifiche assunzioni.

Le proprietà analizzate sono dette asintotiche in quanto valgono al divergere della numerosità campionaria  $n$ . Questa situazione è chiaramente impossibile a livello pratico, ciononostante si auspica che queste caratteristiche possano essere un'approssimazione adatta a problemi in cui  $n$  è finito.

Nonostante questo approccio sia molto diffuso e spesso necessario per problemi di inferenza, l'imprecisione introdotta dall'utilizzo di proprietà asintotiche in

contesti reali è molto spesso sottovalutata se non ignorata in toto.

Le assunzioni necessarie sono: la corretta specificazione del modello parametrico, la regolarità della verosimiglianza e la presenza di un campionamento casuale semplice alla base dell'esperimento. Inoltre si supponga che il supporto di  $Y$  non dipenda da  $\theta \in \Theta \subseteq \mathbb{R}^k$ .

### Consistenza

Sapendo che  $\mathbb{E}[\ell(\theta, Y) \mid \theta_0] = n\mathbb{E}[\ell(\theta; Y_1) \mid \theta_0]$  si introduce il Teorema:

**Teorema 1.2.** *Sotto le seguenti assunzioni:*

1)

$$\sup_{\theta \in \Theta} \left| \frac{1}{n} \ell(\theta) - \mathbb{E}[\ell(\theta; Y_1) \mid \theta_0] \right| \xrightarrow{P} 0,$$

2)

$$\sup_{\theta \in \Theta: |\theta - \theta_0| \geq \epsilon} \mathbb{E}[\ell(\theta; Y) \mid \theta_0] < \mathbb{E}[\ell(\theta_0; Y_1) \mid \theta_0].$$

La successione  $\hat{\theta}_n$  converge in probabilità al vero valore del parametro  $\theta_0$ .

La notazione  $\hat{\theta}_n$  viene utilizzata per sottolineare la dipendenza della stima SMV dal valore di  $n$ .

### Distribuzione asintotica di $\hat{\theta}$ e $\ell'(\theta)$

Con lo scopo di individuare la distribuzione asintotica della SMV nel caso  $k = 1$ , si introducono le seguenti assunzioni:

1. ci si trova in un problema regolare di stima,
2. il vettore  $Y = (Y_1, \dots, Y_n)^T$  è a componenti indipendenti e identicamente distribuite con funzione di densità marginale  $g(y; \theta)$ ,
3. l'informazione attesa di Fisher  $i(\theta)$  per una singola osservazione  $Y_i$  è positiva e finita,
4. la SMV  $\hat{\theta}$  è consistente,
5. esiste una funzione  $M(y; \theta)$  tale che

$$\left| \frac{\partial^3}{\partial \theta^3} \log g(y; \theta) \right| < M(y; \theta)$$

e  $M(y; \theta)$ , è integrabile rispetto a  $g$  con integrale limitato  $\mathbb{E}[M((Y; \theta) | \theta] < M_0 < \infty$ .

Si inizia dallo sviluppo in serie di Taylor della quantità  $\ell'(\hat{\theta})$  nel punto  $\theta_0$

$$0 = \ell'(\hat{\theta}) = \ell'(\theta_0) + \ell''(\theta_0)(\hat{\theta} - \theta_0) + \frac{1}{2}\ell'''(\tilde{\theta})(\hat{\theta} - \theta_0)$$

con  $\tilde{\theta} \in (\hat{\theta}, \theta_0)$ . Moltiplicando tutti i termini per  $\frac{1}{\sqrt{n}}$  e isolando il termine  $\sqrt{n}(\hat{\theta} - \theta_0)$

$$\sqrt{n}(\hat{\theta} - \theta_0) = \frac{-\frac{1}{\sqrt{n}}\ell'(\theta_0)}{\frac{1}{n}\ell''(\theta_0) + \frac{\ell'''(\tilde{\theta})}{2n}(\hat{\theta} - \theta_0)}. \quad (1.2)$$

Si consideri la (1.2):

- Per il Teorema del limite centrale il numeratore

$$\frac{1}{\sqrt{n}}\ell'(\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial}{\partial \theta} \log g(Y_i; \theta) \xrightarrow{d} I \sim N(0, i(\theta))$$

al divergere di  $n$ .

- al denominatore

$$\frac{\ell''(\theta_0)}{n} \xrightarrow{q.c.} -i(\theta_0)$$

per la legge forte dei grandi numeri,

- il termine  $(\hat{\theta} - \theta_0) = o_p(1)$  per la consistenza della SMV,

- sfruttando l'assunzione 5. si può provare che

$$\frac{1}{n}|\ell'''(\tilde{\theta})| = O_p(1),$$

ed è moltiplicata per un  $o_p(1)$  risultando in un  $o_p(1)$ .

In conclusione quindi

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} I' \sim N(0, i(\theta_0)^{-1}).$$

Solitamente da ciò si conclude che

$$\hat{\theta} \sim N(\theta_0; (ni(\theta_0))^{-1}) \quad (1.3)$$

in maniera approssimata, se  $n$  sufficientemente elevato.

Da questa dizione sorgono spesso molti problemi, innanzitutto non esiste una soglia oggettiva per  $n$  per cui la distribuzione approssimata è sensata. Spesso inoltre non viene eseguita una valutazione qualitativa dell'approssimazione né una presentazione dei casi in cui questa può essere troppo imprecisa.

Tuttavia rimane vero che la SMV, vista come stimatore del parametro  $\theta$ , risulta:

- asintoticamente non distorta,
- asintoticamente efficiente, in quanto la varianza della sua distribuzione asintotica raggiunge il limite inferiore di Cramér-Rao.

### Riparametrizzazioni

Si ipotizzi di aver effettuato una riparametrizzazione di  $\theta$  tramite una funzione biunivoca  $\psi(\cdot)$ . Se essa è dotata di derivate parziali miste continue, con  $\Delta$  matrice delle derivate parziali prime delle componenti di  $\theta$  rispetto a  $\psi = \psi(\theta)$ , si ha

$$\sqrt{n}(\hat{\psi} - \psi_0) \xrightarrow{d} I'' \sim N_k \left( 0, (\Delta^T i(\theta_0) \Delta)^{-1} \right)$$

### Informazione osservata

La valutazione dell'accuratezza della SMV a livello asintotico non termina con l'analisi della sua distorsione, si pone quindi l'attenzione sulla variabilità dello stimatore in questione. Dalla (2.13) è chiaro che la quantità d'interesse è  $(ni(\theta_0))^{-1} = I(\theta_0)^{-1}$ , che dipende dal vero valore del parametro. Il prossimo passo è naturalmente quello di stimare questa quantità, ma la soluzione non è univoca. Ci si trova difatti davanti a due possibilità: la prima è quella più intuitiva, ovvero utilizzare  $I(\hat{\theta})^{-1}$ . Questo si rivela uno stimatore consistente se  $\hat{\theta}$  lo è e se  $I(\theta)$  è continua, due condizioni tutto sommato plausibili. Un metodo alternativo sarebbe fare uso del reciproco dell'informazione osservata di Fisher  $\mathcal{J}(\hat{\theta})$ :

$$\mathcal{J}(\hat{\theta}) = -\ell''(\hat{\theta}) = -\frac{d^2}{d\theta^2} \ell(\theta) \Big|_{\theta=\hat{\theta}}.$$

La scelta tra le due è irrilevante a livello asintotico, in quanto le due quantità sono equivalenti, in applicazioni reali però la questione è interessante. In [Azzalini \(2008\)](#) l'informazione osservata viene considerata la scelta migliore in quanto essa, oltre a far parte dell'ambito del principio della verosimiglianza, fornisce un resoconto migliore dell'informazione contenuta nel campione osservato.

### 1.3 Verifica d'ipotesi

#### 1.3.1 Il sistema d'ipotesi

L'ambito della verifica delle ipotesi si concentra sull'elaborazione di una procedura in grado di valutare la conformità dei dati ad un sottomodello  $\mathcal{F}_0$  di  $\mathcal{F}$  espresso da una *ipotesi nulla*  $H_0$ . Facendo riferimento a (1.1), è chiaro che una restrizione sul modello implica una restrizione sullo spazio parametrico  $\Theta$ , pertanto si avrà  $H_0 : \theta \in \Theta_0$ , con  $\Theta_0 \subset \Theta$ . L'ipotesi nulla  $H_0$  è solitamente accompagnata dalla cosiddetta *ipotesi alternativa*  $H_1$  il cui sottomodello è complementare a  $\mathcal{F}_0$ . Il sistema di ipotesi generico si articola nel seguente modo:

$$\begin{cases} H_0 : \theta \in \Theta_0, \\ H_1 : \theta \in \Theta_1, \end{cases}$$

con  $\Theta_0 \cap \Theta_1 = \emptyset$  e  $\Theta_0 \cup \Theta_1 = \Theta$ . Di seguito si delinea la procedura di risoluzione di un problema di ipotesi.

#### 1.3.2 Test statistico e statistica test

Azzalini (2008) definisce *test statistico* il processo di decisione riguardo un sistema di ipotesi, ovvero una funzione  $T: \mathcal{Y} \rightarrow \{\Theta_0, \Theta_1\}$ , la decisione sarà ovviamente corrispondente all'insieme individuato dal test. A livello pratico, ai fini del test si utilizza solitamente una *statistica test*, ovvero una funzione dei dati  $t: \mathcal{Y} \rightarrow \mathbb{R}$ . Essendo  $t(Y)$  una funzione dei dati, essa può assumere valori più o meno conformi all'ipotesi  $H_0$  proprio perché i dati stessi riflettono, con un certo grado di incertezza, il valore del parametro  $\theta$  e quindi il sottoinsieme a cui appartiene. La partizione dello spazio parametrico si ripercuote sullo spazio campionario  $\mathcal{Y}$  il quale è caratterizzato da una divisione in due regioni.

La *regione di rifiuto* o regione critica  $R$

$$R = \{y \in \mathcal{Y} : T(y) = \Theta_1\}.$$

La *regione di accettazione*  $A$

$$A = \{y \in \mathcal{Y} : T(y) = \Theta_0\}.$$

Da un punto di vista procedurale la divisione dello spazio campionario verrà vista in relazione non al test  $T(y)$ , ma alla statistica test  $t(y)$ , che sarà a tutti gli effetti il nostro strumento decisionale.

L'allontanamento dei dati da  $H_0$  si può riflettere su  $t(y)$  in vari modi. Se all'aumentare di  $t(y)$  l'ipotesi nulla diventa più inadeguata, allora il test ha regione critica *unilaterale destra*. Nel caso contrario, con regione di rifiuto composta da valori piccoli di  $t(y)$ , si parla di regione critica *unilaterale sinistra*. Infine, se la regione critica contiene sia valori piccoli che grandi di  $t$ , essa è chiamata *bilaterale*.

### 1.3.3 Errori di I e II tipo

Il criterio di verifica delle ipotesi non è perfetto, l'incertezza intrinseca nella casualità del campione rende possibile incorrere in due tipi di errore.

L'*errore di primo tipo* consiste nel rifiutare l'ipotesi nulla quando essa è vera. L'*errore di secondo tipo* rappresenta il caso opposto, ovvero l'accettazione dell'ipotesi nulla in situazioni in cui essa è falsa. A partire da questi si può definire la *funzione di potenza*  $\gamma(\theta)$

$$\gamma(\theta) = \mathbb{P}(T(Y) = \Theta_1 \mid \theta),$$

questa è una funzione caratteristica del test  $T(y)$ . Chiaramente si desidera una funzione di potenza più alta possibile quando l'ipotesi alternativa è vera e più bassa possibile altrimenti. Non è possibile soddisfare entrambe le richieste insieme, per cui la questione di ottimalità di un test non è affatto banale. Nonostante ciò non verrà approfondita la questione di ricerca dei test ottimali ma si adotterà l'impostazione di Neyman-Pearson sintetizzata in seguito.

### 1.3.4 Livello di significatività del test

Il *livello di significatività* del test è la quantità  $\alpha$

$$\alpha = \sup_{\theta \in \Theta_0} \gamma(\theta) = \sup_{\theta \in \Theta_0} \mathbb{P}(Y \in \mathbf{R} \mid \theta),$$

ovvero la massima probabilità di errore del I tipo. L'*impostazione di Neyman-Pearson* assegna un'importanza maggiore all'errore di I tipo, perciò fissa un valore  $\alpha$  di significatività e minimizza la probabilità di errore di secondo tipo dato  $\alpha$ . La consuetudine è di scegliere  $\alpha = 0.05$  tuttavia è frequente l'uso di valori alternativi, in base alle necessità di ricerca.

### 1.3.5 Livello di significatività osservato

Un approccio molto comune è la determinazione del livello di significatività osservato  $\alpha_{\text{oss}}$ . Il processo di calcolo di questo valore è un modo di valutare l'evidenza empirica contro  $H_0$ , il risultato è infatti il livello di significatività più basso per il quale si accetterebbe l'ipotesi nulla. Utilizzando la notazione  $t = t(Y)$  che è una v.c. e  $t_{\text{oss}} = t(y_{\text{oss}})$  si ha che, per come sono stati definiti i differenti tipi di regioni critiche,  $\alpha_{\text{oss}}$  può essere espresso tramite tre formule differenti. Pace & Salvan (2001) definisce questa grandezza in base al test che tratta: se il test ha regione di rifiuto unilaterale sinistra

$$\alpha_{\text{oss}} = \sup_{\theta \in \Theta_0} \mathbb{P}(t \geq t_{\text{oss}} \mid \theta).$$

Per quanto riguarda le altre due categorie di regioni critiche: nella unilaterale destra

$$\alpha_{\text{oss}} = \sup_{\theta \in \Theta_0} \mathbb{P}(t \leq t_{\text{oss}} \mid \theta),$$

in quella bilaterale

$$\alpha_{\text{oss}} = 2 \sup_{\theta \in \Theta_0} \min\{\mathbb{P}(t \leq t_{\text{oss}} \mid \theta), \mathbb{P}(t \geq t_{\text{oss}} \mid \theta)\};$$

di conseguenza  $\alpha_{\text{oss}} \in [0, 1]$ . L'indicatore ottenuto è chiaramente una funzione del campione  $y$ . Nell'eventualità che il livello di significatività del test ci sia stato fornito in precedenza, la decisione potrà essere svolta semplicemente considerando la seguente formula alternativa per le regioni del test:  $R_\alpha = \{y \in \mathcal{Y} : \alpha_{\text{oss}}(y) < \alpha\}$ , e  $A_\alpha = \{y \in \mathcal{Y} : \alpha_{\text{oss}}(y) \geq \alpha\}$ .

### 1.3.6 Tre test connessi alla verosimiglianza

Si consideri il sistema d'ipotesi

$$\begin{cases} H_0 : \theta = \theta_0, \\ H_1 : \theta \neq \theta_0, \end{cases} \quad (1.4)$$

una ragionevole statistica test potrebbe essere il *rapporto di verosimiglianza*

$$\lambda(y) = \frac{L(\theta_0; y)}{\sup_{\theta \neq \theta_0} L(\theta; y)} = \frac{L(\theta_0; y)}{\sup_{\theta \in \Theta} L(\theta; y)} = \frac{L(\theta_0; y)}{L(\hat{\theta}; y)}.$$

Data la descrizione della verosimiglianza come funzione che valuta la credibilità di un valore del parametro, questa statistica test sembra ragionevole. Tuttavia, solitamente si fa uso di un'altra quantità, del tutto equivalente a  $\lambda(\mathbf{y})$  ai fini del test statistico, ovvero  $W(\mathbf{y})$ :

$$W(\mathbf{y}) = -2 \log \lambda(\mathbf{y}) = -2 \log \frac{L(\theta_0; \mathbf{y})}{L(\hat{\theta}; \mathbf{y})}, \quad (1.5)$$

detto *log-rapporto di verosimiglianza* con  $W(\mathbf{y}) \in [0, \infty)$  poiché  $0 < \lambda(\mathbf{y}) \leq 1$ . Il test basato su questa è chiamato TRV.

Nel seguito, saranno considerate valide le assunzioni in 1.2.5. Si può dimostrare che, sotto  $H_0$ :  $W(\mathbf{y}) = W_e(\mathbf{y}) + o_p(1)$ , dove

$$W_e(\mathbf{y}) = n(\hat{\theta} - \theta_0)^2 i(\hat{\theta}), \quad (1.6)$$

chiamato *test di Wald*. Inoltre vale  $W(\mathbf{y}) = W_u(\mathbf{y}) + o_p(1)$ , con

$$W_u(\mathbf{y}) = \frac{\ell'(\theta_0)^2}{ni(\theta_0)} \quad (1.7)$$

che prende il nome di *funzione test di punteggio*.

### Distribuzione asintotica

La determinazione della distribuzione esatta della statistica test del TRV è possibile in determinati casi, tuttavia solitamente bisogna ricorrere alla distribuzione approssimata.

Si può provare che, al divergere di  $n$ , la distribuzione in questione è  $\chi_k^2$ , dove  $k$  è il numero di componenti del parametro  $\theta$  tenendo a mente il sistema di ipotesi (1.4) considerato. I gradi di libertà della Chi-quadrato sono pari al numero di restrizioni poste sul parametro dall'ipotesi nulla. La differenza tra i tre test consiste in  $o_p(1)$ , rendendoli equivalenti al divergere di  $n$ .

Il fatto che questa non sia una distribuzione esatta implica che il *livello reale* del test sia differente dal *livello nominale*, ovvero quello prefissato. Questa discrepanza tra i due livelli viene solitamente trattata con un approccio *conservatore* preferendo un test con livello reale inferiore o al più pari a quello nominale. La sensatezza di questo approccio verrà discussa in seguito.

**Distribuzione asintotica di  $W_e$** 

La statistica test approssimata  $W_e$  converge in distribuzione ad una v.c. con distribuzione  $\chi^2_1$ , si dimostra adesso perché ciò avviene quando sono valide le assunzioni in 1.2.5.

*Dimostrazione.* Conviene tenere a mente la formula di  $W_e$  ovvero

$$W_e(\mathbf{y}) = n(\hat{\theta} - \theta_0)^2 i(\hat{\theta})$$

dove  $\theta_0$  è il valore del parametro assegnato a  $\theta$  dall'ipotesi nulla  $H_0$  mentre  $\hat{\theta}$  è la stima SMV di  $\theta$ .

Date queste assunzioni vale, sotto  $H_0$ ,

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} Z' \sim N(0, i(\theta_0)^{-1});$$

per ricondurci ad una distribuzione normale standard

$$\frac{\sqrt{n}(\hat{\theta} - \theta_0)}{\sqrt{i(\theta_0)^{-1}}} = \sqrt{n}(\hat{\theta} - \theta_0) \sqrt{i(\theta_0)} \xrightarrow{d} Z \sim N(0, 1).$$

A causa della relazione tra la distribuzione normale standard e quella chi-quadrato  $\left(\sqrt{n}(\hat{\theta} - \theta_0) \sqrt{i(\theta_0)}\right)^2 = n(\hat{\theta} - \theta_0)^2 i(\theta_0) \xrightarrow{d} Z^2 \sim \chi^2_1$ . La convergenza in distribuzione, insieme alla convergenza di  $\hat{\theta}$  e a

$$\frac{i(\theta_0)}{i(\hat{\theta})} \xrightarrow{p} 1,$$

ci portano a scrivere

$$W_e(\mathbf{y}) = n(\hat{\theta} - \theta_0)^2 i(\hat{\theta}) = n(\hat{\theta} - \theta_0)^2 i(\hat{\theta}) \frac{i(\theta_0)}{i(\hat{\theta})} \xrightarrow{d} Z^2 \sim \chi^2_1,$$

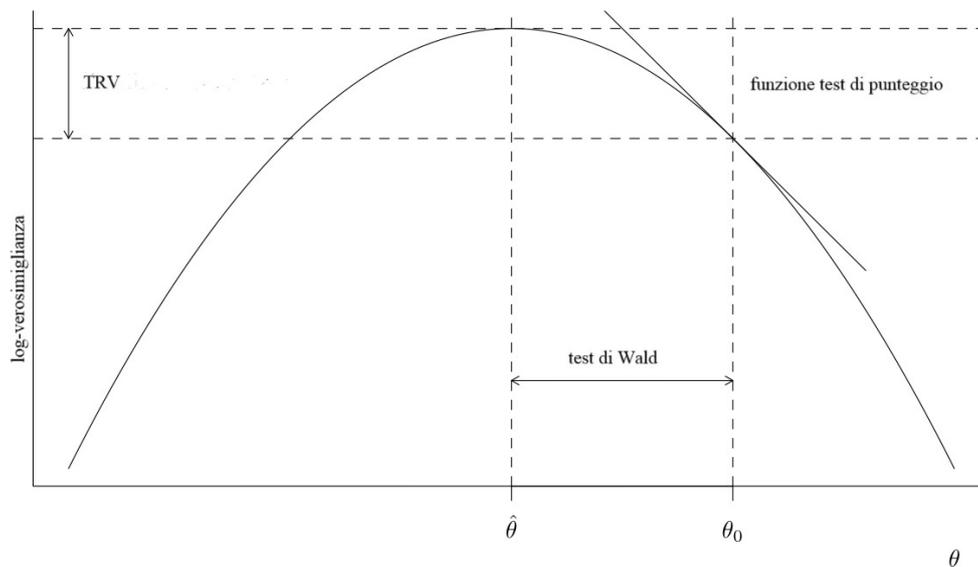
per il Teorema di Slutsky. □

È stato dichiarato, anche se non dimostrato, che questi tre test sono asintoticamente equivalenti tra loro, perciò la distribuzione appena provata risulta valida anche per le altre due statistiche test:  $W_e(\mathbf{y})$  e  $W_u(\mathbf{y})$ .

### Interpretazione grafica

Si intende chiarire il significato di questi test e il motivo per cui essi sono così simili. Essi misurano tre aspetti della log-verosimiglianza esaminata dal punto di vista grafico.

- il TRV  $W(y)$  misura la differenza tra le ordinate della log-verosimiglianza, ovvero tra i valori della funzione stessa in  $\hat{\theta}$  e  $\theta_0$ ,
- il test di Wald  $W_e(y)$  misura la differenza tra ascisse in  $\hat{\theta}$  e  $\theta_0$  effettuando una standardizzazione,
- il test di punteggio  $W_u(y)$  misura la pendenza di  $\ell(\theta)$  in  $\theta_0$  con una standardizzazione opportuna.



**Figura 1.1:** Rappresentazione grafica dai tre test basati sulla verosimiglianza

Tra questi tre test non ne esiste uno migliore, almeno in termini generali. Il test di Wald è però l'unico a non essere invariante rispetto a riparametrazioni.

## 1.4 Stima intervallare

A livello formale: si dice *regione di confidenza* per  $\theta$ , basata sui dati  $y$  e sull'assunzione del modello con spazio parametrico  $\Theta$ , un sottoinsieme  $C(y) \subset \Theta$ . Se il parametro è scalare si dirà *intervallo di confidenza*. Si definisce *probabilità di copertura* della regione  $C(y)$  la probabilità  $\mathbb{P}(\theta \in C(Y) \mid \theta)$ . Il *livello di confidenza* o coefficiente di confidenza è il valore più basso ottenibile di probabilità di copertura:

$$\inf_{\theta} \mathbb{P}(\theta \in C(Y) \mid \theta) = 1 - \alpha, \quad (1.8)$$

dove  $1 - \alpha$  è il valore scelto solitamente per il coefficiente di confidenza. In seguito si illustrerà la strada da percorrere per la costruzione di un intervallo di confidenza, definito in [Casella & Berger \(2008\)](#) come

$$C(Y) : \mathbb{P}(\theta \in C(Y) \mid \theta) \geq 1 - \alpha, \quad (1.9)$$

con livello di confidenza pari ad  $1 - \alpha$ . Un elemento fondamentale nella costruzione di intervalli di confidenza è la *Quantità-Pivot*.

### 1.4.1 Quantità-Pivot

La *Quantità-Pivot*, se esiste, è una funzione  $T(y, \theta) : \mathcal{Y} \times \Theta \rightarrow \mathbb{R}^m$  per qualche  $m$ . La quantità  $T(y, \theta)$  deve avere una distribuzione che non dipende dal vero valore del parametro  $\theta$ . Per un generico  $B \in \mathbb{R}^m$  è possibile definire  $\mathbb{P}(T(Y, \theta) \in B \mid \theta)$ . L'appartenenza di  $T(y, \theta)$  a  $B$  equivale a  $\{y : C(y) \ni \theta\}$  dove  $C(y)$  è un sottoinsieme di  $\Theta$ . Ponendo la probabilità appena definita almeno uguale a  $1 - \alpha$  e sostituendo l'insieme  $B$ :

$$\mathbb{P}(C(Y) \ni \theta \mid \theta) \geq 1 - \alpha.$$

Di conseguenza si dirà che la probabilità che  $C(Y)$  contenga il vero valore del parametro  $\theta$  è almeno pari ad  $1 - \alpha$  per ogni  $\theta \in \Theta$ . Pertanto  $C(y)$  è un *insieme di confidenza* di livello  $1 - \alpha$  per  $\theta$ .

Il ruolo chiave della quantità-pivot sta nel fatto che essa, avendo una distribuzione nota, permette di identificare la regione  $B$  con dei precisi quantili della distribuzione di  $T(y; \theta)$ . Una volta trovati questi quantili, basterà isolare il parametro  $\theta$  per ottenere una regione della forma appena scritta. Se  $Y$  è una v.c. scalare

continua allora una quantità-pivot esiste sempre.

Rimangono irrisolte alcune questioni: come si determina una quantità-pivot? Come si seleziona la migliore se ve ne molteplici? Manca quindi un procedimento univoco da poter seguire in un problema di stima intervallare.

#### 1.4.2 Test e regioni di confidenza, impostazione di Neyman

Per la selezione tra diverse quantità pivot possibili, Neyman fornisce un criterio per definire un insieme di confidenza *ottimo*. L'insieme di confidenza  $C(y)$  è ottimo se:

$$\mathbb{P}(C(Y) \ni \theta \mid \theta) \geq 1 - \alpha,$$

$$\mathbb{P}(C(Y) \ni \theta' \mid \theta) = \text{minimo} \quad \forall \theta' \neq \theta \quad \forall \theta \in \Theta.$$

Una via per trovare questo tipo di regione possono essere i test per la verifica di ipotesi, vi è infatti una dualità tra regioni di confidenza e regioni di accettazione esaminata in [Casella & Berger \(2008\)](#). Non è detto che il test ottimo esista.

**Teorema 1.3.** Per ogni  $\theta_0 \in \Theta$  sia  $A(\theta_0)$  la regione di accettazione di livello  $\alpha$  del test con ipotesi  $H_0 : \theta = \theta_0$ . Per ogni  $x \in \mathcal{X}$  si definisce  $C(x)$  nello spazio parametrico con

$$C(x) = \{\theta_0 : x \in A(\theta_0)\}.$$

Allora  $C(X)$  è un intervallo di confidenza con livello  $1 - \alpha$  per  $\theta$ .

Al fine di costruire un intervallo di confidenza sarà quindi sufficiente trovare un opportuno test per un sistema d'ipotesi del tipo (1.4). La quantità pivot verrà fornita dal test dal momento che la funzione indicatrice dell'insieme  $\{y \in A(\theta_0)\}$  è proprio una quantità pivot. A questo punto per soddisfare il criterio di Neyman si sfrutta il test per scrivere  $\mathbb{P}(Y \in A(\theta_0) \mid \theta) = \mathbb{P}(\theta_0 \in C(Y) \mid \theta) = 1 - \pi(\theta)$  con  $\theta \neq \theta_0$  e  $\pi(\theta)$  funzione di potenza del test sopra definito. Questa è la quantità da minimizzare, dunque per creare un insieme di confidenza  $C(Y)$  ottimo esso deve essere associato ad una regione di accettazione  $A(\theta_0)$  con funzione di potenza massima quando vale l'ipotesi alternativa.

Purtroppo non è sempre possibile ricavare il test ottimo facilmente, ragion per cui ci si accontenta spesso di un test "buono". Il procedimento da seguire

in un problema di stima intervallare sarà: trovare un test "buono" per il sistema d'ipotesi (1.4) e livello di significatività  $\alpha$  (è necessaria l'alternativa bilaterale); a questo punto la regione di confidenza sarà data dall'insieme di tutti i valori di  $\theta_0$  per cui si accetterebbe l'ipotesi nulla.

**Procedimenti con variabili casuali discrete** La definizione dell'intervallo di confidenza con (1.9) è necessaria perché nel caso di v.c. discrete, non è sempre possibile determinare estremi dell'intervallo la cui probabilità di copertura è esattamente pari ad  $1 - \alpha$ . Ciò accade perché la funzione di ripartizione è "a scalini" quindi non sono ottenibili tutti i valori nell'intervallo  $(0, 1)$  per la probabilità di appartenenza all'IC.

### 1.4.3 Procedimenti approssimati

Spesso non è facile trovare quantità-pivotali o test che sfruttano distribuzioni esatte, ragion per cui si ricorre a distribuzioni asintotiche. In problemi reali però,  $n$  è finito, di conseguenza si ha una perdita più o meno significativa di informazioni. Una procedura esatta non è necessariamente migliore di una asintotica, per valutarle vi sono delle quantità apposite.

### 1.4.4 Valutazione dei procedimenti di stima intervallare

Nella valutazione di un intervallo di confidenza ci sono vari aspetti di cui tener conto.

#### Probabilità di copertura

La *probabilità di copertura* è la probabilità che la stima intervallare contenga il vero valore del parametro, valutata a priori, cioè

$$\mathbb{P}(\theta \in C(Y) \mid \theta).$$

Il coefficiente di confidenza (1.8) corrisponde al valore minimo che può assumere la probabilità di copertura, rispetto al valore del parametro. È bene che un intervallo abbia quindi probabilità di copertura che si mantiene sempre vicina al livello di confidenza. Da notare come, nel caso di un test esatto, la probabilità di copertura può essere solo maggiore del coefficiente di confidenza.

Questo indice da solo non è una metrica adatta a valutare un intervallo di confidenza, in quanto l'intervallo  $(-\infty, \infty)$  ha sempre probabilità di copertura pari ad 1. Un buon principio è quello di valutarla in concomitanza con un indicatore di grandezza dell'intervallo, banalmente la lunghezza di questo.

### **Ampiezza attesa**

L'*ampiezza attesa* equivale al valore atteso della lunghezza dell'intervallo  $(t_1(y), t_2(y))$ , valutata su tutti i possibili campioni:

$$\mathbb{E}[t_2(Y) - t_1(Y) \mid \theta].$$

Nel contesto di stima intervallare è apprezzabile un intervallo che riesce ad avere una buona probabilità di copertura mantenendo una lunghezza contenuta. Risulta infatti poco utile un intervallo molto ampio, che di conseguenza fornirà poche informazioni sul parametro d'interesse.

## Capitolo 2

# Stima intervallare nel modello binomiale

### 2.1 La variabile casuale binomiale

La distribuzione binomiale è una distribuzione discreta basata su un esperimento Bernoulliano ripetuto un determinato numero di volte. Un esperimento Bernoulliano è un esperimento con due esiti possibili, codificati con 0 (insuccesso) e 1 (successo), ciascuno con probabilità associata pari a  $1 - p$  e  $p$  rispettivamente. Più precisamente la v.c. binomiale  $X$  è il numero totale di successi in  $n$  esperimenti Bernoulliani indipendenti e identicamente distribuiti.

Questa variabile casuale sarà caratterizzata da due parametri: il numero di prove effettuate  $n$  e la probabilità di successo della singola prova  $p$ . Se  $X \sim \text{Bin}(n, p)$  allora  $\text{Supp}(X) = \{0, 1, 2, \dots, n\}$ , con  $n \in \mathbb{N}$  e  $p \in (0, 1)$ . La funzione di probabilità è

$$\mathbb{P}(X = x | p) = \binom{n}{x} p^x (1 - p)^{n-x},$$

quindi la funzione di ripartizione sarà

$$\mathbb{P}(X \leq x | p) = \sum_{k=0}^x \binom{n}{k} p^k (1 - p)^{n-k}.$$

#### 2.1.1 Momenti

Il valore atteso di una v.c. binomiale è  $\mathbb{E}[X] = np$ , la varianza  $\text{Var}(X) = np(1 - p)$ , mentre la funzione generatrice dei momenti  $M_X(t) = [pe^t + (1 - p)]^n$ .

### 2.1.2 Verosimiglianza e modello binomiale

Nel contesto di un *modello binomiale* la funzione del modello è

$$f(x; p) = \binom{n}{x} p^x (1-p)^{n-x}, \quad (2.1)$$

con supporto per il parametro  $\Theta = (0, 1)$ .

La verosimiglianza verrà scritta come

$$L(p) = L(X; p) = c(x) \binom{n}{x} p^x (1-p)^{n-x},$$

la log-verosimiglianza

$$\ell(p) = \log c(x) + \log \binom{n}{x} + x \log p + (n-x) \log(1-p).$$

La funzione di probabilità (2.1) può essere riscritta come

$$f(x; p) = \binom{n}{x} \exp \left( x \log \frac{p}{1-p} - (-n \log(1-p)) \right),$$

quindi la v.c. binomiale appartiene alla famiglia esponenziale (regolare) di ordine 1 con  $\tau(p) = -n \log(1-p)$ ;  $\psi(p) = \log(p/(1-p))$ ;  $t(x) = x$ .

### Funzione di Punteggio e Informazione di Fisher

Tenendo a mente che  $X \sim \text{Bin}(n, p)$ , le principali quantità direttamente connesse alla verosimiglianza saranno (prendendo  $c(x) = 1$ ):

$$L(p) = \binom{n}{x} p^x (1-p)^{n-x}, \quad (2.2)$$

$$\ell(p) = \log \binom{n}{x} + x \log p + (n-x) \log(1-p), \quad (2.3)$$

$$u(p) = \frac{\partial}{\partial p} \ell(p) = \frac{x}{p} - \frac{n-x}{1-p}, \quad (2.4)$$

$$\frac{\partial^2}{\partial p^2} \ell(p) = -\frac{x}{p^2} - \frac{n-x}{(1-p)^2},$$

$$I(p) = -\mathbb{E} \left[ \frac{\partial^2}{\partial p^2} \ell(p) \right] = \mathbb{E} \left[ \frac{x}{p^2} + \frac{n-x}{(1-p)^2} \right] = \frac{np}{p^2} + \frac{n-np}{(1-p)^2} = \frac{n}{p(1-p)} \quad (2.5)$$

sapendo che  $\mathbb{E}[x] = np$ . Poiché  $I(p) = ni(p)$

$$i(p) = \frac{1}{p(1-p)}. \quad (2.6)$$

## 2.2 Tre test connessi alla verosimiglianza

Si riprendono i tre test connessi alla verosimiglianza presentati nel 1.3.6 ma nel caso binomiale. Si farà riferimento ad un sistema d'ipotesi del tipo (1.4) utilizzando  $p$  invece di  $\theta$ . Come visto nella sezione 1.3.6 i 3 test sono asintoticamente equivalenti, adesso si dimostra ciò nel caso binomiale, presentandone uno alla volta.

### 2.2.1 Log-rapporto di verosimiglianza nel caso binomiale

Il test di log-rapporto di verosimiglianza in (1.5) viene definito come

$$W(x) = -2 \log \frac{L(\theta_0, x)}{\sup_{\theta \neq \theta_0} L(\theta, x)} = -2 \log \frac{L(\theta_0, x)}{L(\hat{\theta}, x)}$$

con  $\theta \in \Theta$ ;  $\hat{\theta}$  è la stima SMV di  $\theta$ . Utilizzando (2.2),  $W(x)$  diventerà

$$W(x) = -2 \log \frac{\binom{n}{x} p_0^x (1-p_0)^{n-x}}{\binom{n}{x} \hat{p}^x (1-\hat{p})^{n-x}};$$

semplificando i coefficienti binomiali e utilizzando le proprietà del logaritmo:

$$W(x) = -2 [x \log p_0 + (n-x) \log(1-p_0) - x \log \hat{p} - (n-x) \log(1-\hat{p})],$$

sommando i termini a fattori comuni e distribuendo il  $-1$

$$W(x) = 2 \left[ x \log \frac{\hat{p}}{p_0} + (n-x) \log \frac{1-\hat{p}}{1-p_0} \right].$$

### 2.2.2 Test di Wald nel caso binomiale

Si prende il test di Wald come in (1.6)

$$W_e(x) = n(\hat{\theta} - \theta_0)^2 i(\hat{\theta}),$$

dato (2.6), si sostituisce la SMV:

$$W_e(x) = n \frac{(\hat{p} - p_0)^2}{\hat{p}(1 - \hat{p})}. \quad (2.7)$$

Si descrive adesso il legame con TRV e due caratteristiche poco apprezzabili di questa statistica test.

### Log-rapporto di verosimiglianza e Test di Wald

L'equivalenza asintotica tra i due test si può provare facilmente nel caso binomiale.

*Dimostrazione.* Iniziando dalla formula di  $W(x)$  nel caso binomiale

$$2 \left[ x \log \frac{\hat{p}}{p_0} + (n - x) \log \frac{1 - \hat{p}}{1 - p_0} \right],$$

si riscrive sfruttando le proprietà del logaritmo

$$2 [x \log \hat{p} - x \log p_0 + (n - x) \log(1 - \hat{p}) - (n - x) \log(1 - p_0)].$$

Si sviluppa in Taylor al secondo ordine in  $p_0 = \hat{p}$ , con i due logaritmi visti in funzione di  $p_0$ . Ovviamente non sarà più un'uguaglianza ma un'approssimazione della quantità vista in precedenza. Si riscrivono:

$$\log(p_0) \approx \left( \log \hat{p} + (p_0 - \hat{p}) \frac{1}{\hat{p}} - (p_0 - \hat{p})^2 \frac{1}{2\hat{p}^2} \right),$$

$$\log(1 - p_0) \approx \left( \log(1 - \hat{p}) - (p_0 - \hat{p}) \frac{1}{1 - \hat{p}} - (p_0 - \hat{p})^2 \frac{1}{2(1 - \hat{p})^2} \right),$$

i termini in funzione logaritmica rimasti si semplificano, risultando in:

$$2 \left[ x \left( (p_0 - \hat{p}) \frac{1}{\hat{p}} - (p_0 - \hat{p})^2 \frac{1}{2\hat{p}^2} \right) - (n - x) \left( -(p_0 - \hat{p}) \frac{1}{1 - \hat{p}} - (p_0 - \hat{p})^2 \frac{1}{2(1 - \hat{p})^2} \right) \right].$$

Raccogliendo i termini comuni

$$2 \left[ (\hat{p} - p_0) \left( \frac{x}{\hat{p}} - \frac{n - x}{1 - \hat{p}} \right) + \frac{1}{2} (\hat{p} - p_0)^2 \left( \frac{x}{\hat{p}^2} + \frac{n - x}{(1 - \hat{p})^2} \right) \right].$$

Lavorando sul termine:

$$\frac{x}{\hat{p}} - \frac{n-x}{(1-\hat{p})} = \frac{x(1-\hat{p}) - (n-x)\hat{p}}{\hat{p}(1-\hat{p})} = \frac{x - x\hat{p} - n\hat{p} + x\hat{p}}{\hat{p}(1-\hat{p})},$$

semplificando  $x\hat{p}$  e tenendo a mente che  $\hat{p} = \frac{x}{n}$

$$\frac{x - n\frac{x}{n}}{\hat{p}(1-\hat{p})} = 0.$$

Il primo addendo nella parentesi è quindi nullo, perciò si tratterà

$$2 \left[ \frac{1}{2}(\hat{p} - p_0)^2 \left( \frac{x}{\hat{p}^2} + \frac{n-x}{(1-\hat{p})^2} \right) \right].$$

Anche in questo caso si pone l'attenzione su

$$\frac{x}{\hat{p}^2} + \frac{n-x}{(1-\hat{p})^2} = \frac{x + x\hat{p}^2 - 2x\hat{p} + n\hat{p}^2 - x\hat{p}^2}{\hat{p}^2(1-\hat{p}^2)},$$

seguendo gli stessi passaggi di prima

$$\frac{x - 2\frac{x^2}{n} + \frac{x^2}{n}}{\hat{p}^2(1-\hat{p}^2)} = n \frac{\frac{x}{n} - \frac{x^2}{n^2}}{\hat{p}^2(1-\hat{p}^2)} = n \frac{\hat{p}(1-\hat{p})}{\hat{p}^2(1-\hat{p}^2)}.$$

Tornando alla quantità iniziale

$$W(x) \approx 2 \left[ \frac{1}{2}(\hat{p} - p_0)^2 \frac{n}{\hat{p}(1-\hat{p})} \right] = n \frac{(\hat{p} - p_0)^2}{\hat{p}(1-\hat{p})} = W_e(x).$$

□

### Dipendenza all'interno della statistica t

Andersson (2023) fa riferimento all'applicazione della formula per la *statistica t* nel caso binomiale e a come in questo caso sorga un problema per la dipendenza tra numeratore e denominatore di questa quantità.

*Dimostrazione.* La statistica test basata sulla distribuzione t viene definita come  $\sqrt{n}(\bar{X} - \mu)/S$ . Nel caso  $X \sim \text{Bin}(n, p)$ ,

$$\frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \quad \text{dove} \quad \hat{p} = \frac{X}{n},$$

che coincide con la (2.7).

Si può dimostrare che il numeratore e il denominatore non sono indipendenti, in quanto:

$$\text{cov} \left( \hat{p}; \frac{\hat{p}(1-\hat{p})}{n-1} \right) = \frac{p(1-p)(1-2p)}{n^2}.$$

É noto che per due v.c.  $X$  e  $Y$  vale che  $\text{cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ .

In questo caso

$$\text{cov} \left( \hat{p}, \frac{\hat{p}(1-\hat{p})}{n-1} \right) = \mathbb{E} \left[ \hat{p} \frac{\hat{p}(1-\hat{p})}{n-1} \right] - \mathbb{E}[\hat{p}] \mathbb{E} \left[ \frac{\hat{p}(1-\hat{p})}{n-1} \right],$$

$\hat{p}$  è uno stimatore non distorto perciò si può scrivere

$$\mathbb{E} \left[ \hat{p} \frac{\hat{p}(1-\hat{p})}{n-1} \right] - \mathbb{E}[\hat{p}] \mathbb{E} \left[ \frac{\hat{p}(1-\hat{p})}{n-1} \right] = \mathbb{E} \left[ \frac{\hat{p}^2(1-\hat{p})}{n-1} \right] - p \mathbb{E} \left[ \frac{\hat{p}(1-\hat{p})}{n-1} \right].$$

Si utilizza la proprietà di linearità del valore atteso per raccogliere i denominatori che non sono stocastici e per "distribuire" il valore atteso all'interno dei numeratori

$$\mathbb{E} \left[ \frac{\hat{p}^2(1-\hat{p})}{n-1} \right] - p \mathbb{E} \left[ \frac{\hat{p}(1-\hat{p})}{n-1} \right] = \frac{1}{n-1} \left( \mathbb{E}[\hat{p}^2] - \mathbb{E}[\hat{p}^3] - p \mathbb{E}[\hat{p}] + p \mathbb{E}[\hat{p}^2] \right),$$

utilizzando ancora la non distorsione di  $\hat{p}$ :

$$\frac{1}{n-1} \left( \mathbb{E}[\hat{p}^2] - \mathbb{E}[\hat{p}^3] - p^2 + p \mathbb{E}[\hat{p}^2] \right).$$

Si sostituisce  $\hat{p}$  per ottenere i momenti secondi e terzi di cui è nota la formula

$$\frac{1}{n-1} \left( \mathbb{E} \left[ \frac{X^2}{n^2} \right] - \mathbb{E} \left[ \frac{X^3}{n^3} \right] - p^2 + p \mathbb{E} \left[ \frac{X^2}{n^2} \right] \right)$$

i momenti di una v.c. binomiale sono:  $\mathbb{E}[X^2] = np(1-p) + n^2p^2$  e  $\mathbb{E}[X^3] = n(n-1)(n-2)p^3 + 3n(n-1)p^2 + np$ . Adesso si utilizzano queste formule e, semplificando e raccogliendo  $n^2$ , si giunge a

$$\frac{[np(1-p) + n^2p^2 - (n-1)(n-2)p^3 - 3(n-1)p^2 - p - n^2p^2 + np^2(1-p) + n^2p^3]}{n^2(n-1)},$$

si procede raccogliendo i termini in modo da ottenere  $n - 1$  come fattore comune a tutti gli addendi:

$$\frac{[p(n-1) + np^2(n-1) - (n-1)(n-2)p^3 - 3(n-1)p^2 - np^2(n-1) + np^3(n-1)]}{n^2(n-1)},$$

si semplifica  $n - 1$

$$\frac{1}{n^2} [p - (n-2)p^3 - 3p^2 + np^3] = \frac{1}{n^2} [p + 2p^3 - 3p^2].$$

Dopo un raccoglimento di  $p$  e una fattorizzazione

$$\frac{1}{n^2} p(1-p)(1-2p) = \text{cov} \left( \hat{p}; \frac{\hat{p}(1-\hat{p})}{n-1} \right),$$

che coincide con la formula presentata prima. □

### Distorsione della statistica test di Wald

La dimostrazione ha il fine di mostrare che

$$\mathbb{E} \left[ \frac{\sqrt{n}(\hat{p} - p)}{\sqrt{\hat{p}(1-\hat{p})}} \right] \approx -\frac{1-2p}{2\sqrt{np(1-p)}},$$

ovvero che la statistica di Wald è distorta, proprietà mostrata sempre da [Andersson \(2023\)](#).

*Dimostrazione.* Si riscrive la statistica come

$$\frac{\sqrt{n}(\hat{p} - p)}{\sqrt{\hat{p}(1-\hat{p})}} = \frac{\sqrt{n}(\hat{p} - p)}{\sqrt{p(1-p)}} \sqrt{\frac{p(1-p)}{\hat{p}(1-\hat{p})}}.$$

Applicando una linearizzazione di Taylor al secondo fattore in  $\hat{p} = p$

$$\sqrt{\frac{p(1-p)}{\hat{p}(1-\hat{p})}} \approx \sqrt{\frac{p(1-p)}{p(1-p)}} + (\hat{p} - p) \frac{-\frac{1}{2}[\hat{p}(1-\hat{p})]^{-\frac{1}{2}}(1-2\hat{p})\sqrt{p(1-p)}}{\hat{p}(1-\hat{p})} \Big|_{\hat{p}=p};$$

il primo addendo è uguale a 1 mentre nel secondo, valutando  $\hat{p}$  in  $p$ , si semplificano due fattori ottenendo

$$1 - (\hat{p} - p) \frac{1-2p}{2p(1-p)}.$$

Sapendo ciò si riscrive la t-statistic come

$$\frac{\sqrt{n}(\hat{p} - p)}{\sqrt{\hat{p}(1 - \hat{p})}} \approx \frac{\sqrt{n}(\hat{p} - p)}{\sqrt{p(1 - p)}} \left[ 1 - (\hat{p} - p) \frac{1 - 2p}{2p(1 - p)} \right],$$

distribuendo nel secondo termine

$$\frac{\sqrt{n}(\hat{p} - p)}{\sqrt{p(1 - p)}} - (\hat{p} - p)^2 \frac{\sqrt{n}(1 - 2p)}{2[p(1 - p)]^{3/2}}.$$

Si considera il valore atteso di questa quantità

$$\mathbb{E} \left[ \frac{\sqrt{n}(\hat{p} - p)}{\sqrt{p(1 - p)}} - (\hat{p} - p)^2 \frac{\sqrt{n}(1 - 2p)}{2[p(1 - p)]^{3/2}} \right],$$

il primo addendo ha una variabile casuale solo al numeratore

$$\mathbb{E} \left[ \frac{\sqrt{n}(\hat{p} - p)}{\sqrt{p(1 - p)}} \right] = \frac{\sqrt{n}\mathbb{E}[\hat{p} - p]}{\sqrt{p(1 - p)}} = 0$$

in quanto  $\hat{p}$  è uno stimatore non distorto di  $p$ . Passando al secondo

$$\mathbb{E} \left[ (\hat{p} - p)^2 \frac{\sqrt{n}(1 - 2p)}{2[p(1 - p)]^{3/2}} \right] = \mathbb{E}[(\hat{p} - p)^2] \frac{\sqrt{n}(1 - 2p)}{2[p(1 - p)]^{3/2}} = \frac{p(1 - p)}{n} \frac{\sqrt{n}(1 - 2p)}{2[p(1 - p)]^{3/2}}$$

perché

$$\mathbb{E}[(\hat{p} - p)^2] = \mathbb{E}[\hat{p}^2 + p^2 - 2p\hat{p}] = \mathbb{E}[\hat{p}^2] + p^2 - 2p^2 = \frac{p(1 - p)}{n};$$

sfruttando la formula  $\hat{p} = \frac{X}{n}$  così da avere

$$\mathbb{E}[\hat{p}^2] = \frac{\mathbb{E}[X^2]}{n^2} = \frac{\text{var}(X) + \mathbb{E}[X]^2}{n^2} = \frac{p(1 - p)}{n} + p^2.$$

Ritornando al valore atteso iniziale

$$\mathbb{E} \left[ \frac{\sqrt{n}(\hat{p} - p)}{\sqrt{\hat{p}(1 - \hat{p})}} \right] \approx \frac{p(1 - p)}{n} \frac{\sqrt{n}(1 - 2p)}{2[p(1 - p)]^{3/2}} = -\frac{1 - 2p}{2\sqrt{np(1 - p)}}$$

□

### 2.2.3 Test di Wilson nel caso binomiale

Il test di Wilson in (1.7)

$$W_u(x) = \frac{\ell'(\theta_0)^2}{ni(\theta_0)},$$

si può riscrivere come

$$W_u(x) = n \frac{(\hat{p} - p_0)^2}{p_0(1 - p_0)}, \quad (2.8)$$

dati (2.6) e (2.4), sostituendo  $\hat{p} = \frac{x}{n}$  in (2.4).

### Log-rapporto di verosimiglianza e Test di Wilson

Anche per il test di Wilson si può dimostrare la sua connessione con il test di rapporto di verosimiglianza nel caso binomiale.

*Dimostrazione.* Si inizia direttamente da

$$W(x) = 2 [x \log \hat{p} - x \log p_0 + (n - x) \log(1 - \hat{p}) - (n - x) \log(1 - p_0)].$$

Considerando i due logaritmi come funzioni di  $\hat{p}$ , si applica uno sviluppo di Taylor al secondo ordine in  $\hat{p} = p_0$ . Anche in questo non sarà più un'uguaglianza ma un'approssimazione della quantità vista in precedenza. Le quantità in questione sono:

$$\log(\hat{p}) \approx \left( \log p_0 + (\hat{p} - p_0) \frac{1}{p_0} - (\hat{p} - p_0)^2 \frac{1}{2p_0^2} \right),$$

$$\log(1 - \hat{p}) \approx \left( \log(1 - p_0) - (\hat{p} - p_0) \frac{1}{1 - p_0} - (\hat{p} - p_0)^2 \frac{1}{2(1 - p_0)^2} \right).$$

Sostituendo queste quantità, tutti i termini con il log si annullano, giungendo a

$$2 \left[ x \left( (\hat{p} - p_0) \frac{1}{p_0} - (\hat{p} - p_0)^2 \frac{1}{2p_0^2} \right) - (n - x) \left( -(\hat{p} - p_0) \frac{1}{1 - p_0} - (\hat{p} - p_0)^2 \frac{1}{2(1 - p_0)^2} \right) \right].$$

Dopo un raccoglimento parziale

$$2 \left[ (\hat{p} - p_0) \left( \frac{x}{p_0} - \frac{n - x}{1 - p_0} \right) + \frac{1}{2} (\hat{p} - p_0)^2 \left( -\frac{x}{p_0^2} - \frac{n - x}{(1 - p_0)^2} \right) \right].$$

Concentrandosi su:

$$\frac{x}{p_0} - \frac{n-x}{1-p_0} = \frac{x(1-p_0) - (n-x)p_0}{p_0(1-p_0)} = \frac{x - xp_0 - np_0 + xp_0}{p_0(1-p_0)},$$

adesso è possibile raccogliere  $n$  e, dato che  $\hat{p} = \frac{x}{n}$ ,

$$n \frac{\frac{x}{n} - p_0}{p_0(1-p_0)} = n \frac{\hat{p} - p_0}{p_0(1-p_0)}.$$

Il primo addendo risulterà quindi

$$(\hat{p} - p_0) \left( \frac{x}{p_0} - \frac{n-x}{1-p_0} \right) = n \frac{(\hat{p} - p_0)^2}{p_0(1-p_0)}.$$

Passando al secondo addendo

$$\frac{1}{2} (\hat{p} - p_0)^2 \left( -\frac{x}{p_0^2} - \frac{n-x}{(1-p_0)^2} \right),$$

analizzando sempre

$$-\frac{x}{p_0^2} - \frac{n-x}{(1-p_0)^2} = \frac{-x - xp_0^2 + 2xp_0 - np_0^2 + xp_0^2}{p_0^2(1-p_0)^2} = \frac{-x + 2xp_0 - np_0^2}{p_0^2(1-p_0)^2}.$$

Tornando alla quantità completa

$$2 \left[ n \frac{(\hat{p} - p_0)^2}{p_0(1-p_0)} + \frac{1}{2} (\hat{p} - p_0)^2 \frac{-x + 2xp_0 - np_0^2}{p_0^2(1-p_0)^2} \right],$$

si raccolgono i termini comuni

$$n \frac{(\hat{p} - p_0)^2}{p_0(1-p_0)} \left[ 2 + \frac{1}{n} \frac{-x + 2xp_0 - np_0^2}{p_0^2(1-p_0)^2} \right] = n \frac{(\hat{p} - p_0)^2}{p_0(1-p_0)} \left[ 2 + \frac{-\hat{p} + 2\hat{p}p_0 - p_0^2}{p_0^2(1-p_0)^2} \right].$$

Si sposta l'attenzione sul secondo fattore

$$2 + \frac{-\hat{p} + 2\hat{p}p_0 - p_0^2}{p_0^2(1-p_0)^2} = \frac{2p_0 - 2p_0^2 - \hat{p} + 2\hat{p}p_0 - p_0^2}{p_0^2(1-p_0)^2} = \frac{2p_0 - 3p_0^2 - \hat{p} + 2\hat{p}p_0}{p_0^2(1-p_0)^2}.$$

Grazie alla consistenza della SMV, si può sostituire in questa quantità  $\hat{p} = p_0 + o_p(1)$

$$\frac{2p_0 - 3p_0^2 - p_0 + 2p_0^2}{p_0^2(1 - p_0)^2} + o_p(1) = \frac{p_0 - p_0^2}{p_0^2(1 - p_0)^2} + o_p(1) = 1 + o_p(1).$$

Finalmente è stato dimostrato che

$$W(x) = n \frac{(\hat{p} - p_0)^2}{p_0(1 - p_0)} (1 + o_p(1)) = W_u(x) + o_p(1).$$

□

## 2.3 Intervalli di confidenza

Di seguito verranno presentati diversi metodi di stima intervallare per la probabilità di successo in un modello binomiale. Oltre alla caratterizzazione teorica di questi, si presenterà un'applicazione a due insiemi di dati con numerosità diverse. Il primo dataset presente in [dat \(2024\)](#) appartiene all'ambito medico, in particolare riguarda la probabilità che un paziente muoia dopo un attacco cardiaco, quando è stato sorvegliato per un periodo di 30 giorni. I dati contenuti sono:

$$n_1 = 5 \quad x_1 = 4 \quad \hat{p}_1 = 0.8. \quad (2.9)$$

Il secondo insieme di dati è stato ricavato da [dat \(2017\)](#) e riguarda la lega di basket americana NBA, in particolare la probabilità che un *rookie*, ovvero un giocatore appena approdato nella lega, riesca a continuare a giocare per almeno 5 anni all'interno di questa. Questo dataset è più ampio, in particolare contiene:

$$n_2 = 100 \quad x_2 = 59 \quad \hat{p}_2 = 0.59. \quad (2.10)$$

### 2.3.1 Intervallo di Clopper-Pearson

L'intervallo di confidenza per una v.c. binomiale  $X$  di parametri  $n$  e  $p$  ha estremi dati dalle soluzioni a  $p_0$  nelle equazioni:

$$\sum_{k=0}^x \binom{n}{k} p_0^k (1 - p_0)^{n-k} = \frac{\alpha}{2}, \quad (2.11)$$

$$\sum_{k=x}^n \binom{n}{k} p_0^k (1-p_0)^{n-k} = \frac{\alpha}{2}. \quad (2.12)$$

Queste equazioni hanno soluzione in forma chiusa

$$p_1 = \left[ 1 + \frac{n-x+1}{x q_{\alpha/2, 2x, 2(n-x-1)}^F} \right]^{-1} = q_{\alpha/2, x, n-x+1}^B \quad (2.13)$$

$$p_2 = \left[ 1 + \frac{n-x}{(x+1) q_{1-\alpha/2, 2(x+1), 2(n-x)}^F} \right]^{-1} = q_{1-\alpha/2, x+1, n-x}^B \quad (2.14)$$

dove  $q_{a,b,c}^F$  è il quantile di ordine  $a$  di una distribuzione  $F$  con gradi di libertà  $b$  e  $c$  e  $q_{a,b,c}^B$  è il quantile di ordine  $a$  di una distribuzione Beta con parametri  $b$  e  $c$ .

### Test di Clopper-Pearson

Partendo dai test bilaterali esatti con  $H_0 : p = p_0$  si può giungere precisamente all'intervallo appena mostrato.

*Dimostrazione.* Si ha  $X \sim \text{Bin}(n, p)$  e il sistema di ipotesi

$$\begin{cases} H_0 : p = p_0 \\ H_1 : p \neq p_0 \end{cases} \quad (2.15)$$

si consideri come statistica test la v.c.  $X$  creando così una regione di rifiuto  $R$  :  $\mathbb{P}(X \in R \mid p = p_0) \leq \alpha$ . Essendo l'ipotesi bilaterale,  $R$  sarà del tipo  $R = [0, r_1] \cup [r_2, n]$ , di conseguenza

$$\mathbb{P}(X < r_1 \mid p = p_0) + \mathbb{P}(X > r_2 \mid p = p_0) \leq \alpha,$$

in casi bilaterali si sceglie, per convenzione, di avere le due zone che compongono la regione di rifiuto "simmetriche" per quanto riguarda la probabilità di errore di primo tipo. Ciò implica che

$$r_1 = \sup\{r \in (0, \dots, n) : \mathbb{P}(X < r \mid p = p_0) \leq \alpha/2\},$$

$$r_2 = \inf\{r \in (0, \dots, n) : \mathbb{P}(X > r \mid p = p_0) \leq \alpha/2\}.$$

Questo approccio è di tipo conservativo, in quanto viene richiesta una probabilità di errore del primo tipo che sia al più pari ad  $\alpha$ . Ovviamente tra tutti gli estremi che soddisfano questa condizione si opta per quello più ampio (queste due condizioni giustificano la presenza dei sup e inf). Conoscendo la distribuzione di  $X$  si deduce che

$$\mathbb{P}(X \leq r_1 \mid p = p_0) = \sum_{k=0}^{r_1} \binom{n}{k} p_0^k (1-p_0)^{n-k} \leq \alpha/2,$$

$$\mathbb{P}(X \geq r_2 \mid p = p_0) = \sum_{k=r_2}^n \binom{n}{k} p_0^k (1-p_0)^{n-k} \leq \alpha/2.$$

Questo test avrà quindi regione di accettazione  $A(p_0) = (r_1(p_0), r_2(p_0))$ . Al fine di passare all'intervallo di confidenza, si fa uso del Teorema 1.3. Per fare ciò si scrivono gli estremi in funzione di  $p_0$ .

Si considera  $C(X)$  come  $C(X) = \{p_0 : x \in A(p_0)\} = \{p_0 : x \in (r_1(p_0), r_2(p_0))\}$ , ciò avviene se e solo se  $x > r_1 \cap x < r_2$ , quindi per come sono stati definiti  $r_1$  e  $r_2$ :

$$x > r_1 \implies \sum_{k=0}^x \binom{n}{k} p_0^k (1-p_0)^{n-k} \geq \alpha/2,$$

$$x < r_2 \implies \sum_{k=x}^n \binom{n}{k} p_0^k (1-p_0)^{n-k} \geq \alpha/2$$

Poiché  $x \in \mathbb{N} \cap [0, n]$  e la funzione di ripartizione della binomiale ha ogni elemento della sommatoria  $\geq 0$  con  $p_0 \in [0, 1]$  è possibile riscrivere  $C(X)$  come

$$C(X) = \{p_0 : \sum_{k=0}^x \binom{n}{k} p_0^k (1-p_0)^{n-k} \geq \alpha/2 \cap \sum_{k=x}^n \binom{n}{k} p_0^k (1-p_0)^{n-k} \geq \alpha/2\},$$

$$C(X) = \{p_0 : \mathbb{P}(X \geq r_1 \mid p) \geq \alpha/2 \cap \mathbb{P}(X \leq r_2 \mid p) \geq \alpha/2\}.$$

Questa regione dal punto di vista di  $p$  è effettivamente un intervallo perché la funzione di verosimiglianza per la v.c. binomiale si ha che  $\mathbb{P}(X \leq x \mid p) \quad \mathbb{P}(X \geq x \mid p)$  sono entrambe monotone in  $p$ . Gli estremi di questo intervallo saranno quindi i valori di  $p_0$  per cui  $\mathbb{P}(X \leq x \mid p_0) = \alpha/2 \quad \mathbb{P}(X \geq x \mid p_0) = \alpha/2$ , riscritto come

$$\sum_{k=0}^x \binom{n}{k} p_0^k (1-p_0)^{n-k} = \alpha/2$$

$$\sum_{k=x}^n \binom{n}{k} p_0^k (1-p_0)^{n-k} = \alpha/2.$$

□

### Estremi dell'intervallo

Le soluzioni alle due equazioni appena scritte sono  $p_1 = q_{\alpha/2, x, n-x+1}^B$  e  $p_2 = q_{1-\alpha/2, x+1, n-x}^B$ . La seguente proposizione servirà a spiegare questi due estremi.

**Teorema 2.1.** *Siano  $X \sim \text{Bin}(n, p)$  e  $Y \sim \text{Beta}(j, n-j+1)$ , allora*

$$\mathbb{P}(X \geq j) = \mathbb{P}(Y \leq p).$$

*Dimostrazione.* Siano  $U_1, \dots, U_n$  delle variabili aleatorie uniformi in  $(0, 1)$ . Allora la statistica d'ordine  $U_{(j)}$  è tale per cui

$$\mathbb{P}(U_{(j)} \leq p) = \mathbb{P}(X \geq j) = \sum_{k=j}^n \binom{n}{k} p^k (1-p)^{n-k}.$$

Si può infatti considerare una v.c.  $X$  che conta il numero di volte in cui  $U_1, \dots, U_n$  sono minori o uguali di  $p$ . Si considerano quindi due eventi:  $\{U_j \leq p\}$  preso come successo e  $\{U_j > p\}$  preso come insuccesso. Dunque  $X$  sarà il numero di successi in  $n$  prove che, data la natura del campione  $U_1, \dots, U_n$ , sono indipendenti.

Questa è l'esatta definizione di una v.c. binomiale, quindi  $X \sim \text{Bin}(n, p)$  essendo  $p$  la probabilità che  $U_j \leq p$ . A questo punto guardando l'evento  $\{U_{(j)} \leq p\}$ , esso è equivalente a  $\{X \geq j\}$ . Se la statistica di ordine  $j$  è minore di  $p$ , allora lo saranno almeno le prime  $j$  osservazioni. È stato quindi mostrato che  $\mathbb{P}(U_{(j)} \leq p) = \mathbb{P}(X \geq j)$ .

Per dimostrare l'affermazione iniziale è necessario individuare il legame con la v.c. Beta, per farlo, si fa uso del seguente Teorema presente in [Casella & Berger \(2008\)](#).

**Teorema 2.2.** *Si considerano  $X_{(1)}, \dots, X_{(n)}$  statistiche d'ordine di un campione casuale  $X_1, \dots, X_n$  da una popolazione continua con funzione di densità  $f_X$  e funzione di*

ripartizione  $F_X$ . Allora la funzione di densità di  $X_{(j)}$  è

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} f_X(x) [F_X(x)]^{j-1} [1 - F_X(x)]^{n-j}.$$

Nel caso in esame  $X_{(j)} = U_{(j)}$ ,  $f_{U_j}(x) = 1$  e  $F_{U_j}(u) = u$ , di conseguenza

$$f_{U_{(j)}}(u) = \frac{n!}{(j-1)!(n-j)!} [u]^{j-1} [1-u]^{n-j} = \frac{u^{j-1} (1-u)^{n-j}}{B(j, n-j+1)},$$

ciò implica che  $U_{(j)} \sim \text{Beta}(j, n-j+1)$ , stessa distribuzione di  $Y$ . Per concludere è sufficiente considerare

$$\mathbb{P}(U_{(j)} \leq p) = \mathbb{P}(X \geq j) = \mathbb{P}(Y \leq p),$$

□

Adesso gli estremi dell'intervallo di confidenza si possono ottenere con una certa semplicità. Quello inferiore sarà dato da  $\mathbb{P}(X \geq x | p) = \frac{\alpha}{2} = \mathbb{P}(Y \leq p)$ , risolvendo per  $p$   $p_1 = q_{\alpha/2, x, n-x+1}^B$ .

Per quanto riguarda l'estremo superiore:

$$\mathbb{P}(X \leq x | p) = 1 - \mathbb{P}(X \geq x+1 | p) = \frac{\alpha}{2} = 1 - \mathbb{P}(Y' \leq p)$$

con  $Y' \sim \text{Beta}(x+1, n-x+1)$ , risolvendo sempre per  $p$ :  $p_2 = q_{1-\alpha/2, x+1, n-x}^B$ .

**Un approccio conservativo** Il legame diretto con la legge di distribuzione binomiale rende questo intervallo il più intuitivo se si vogliono evitare procedimenti approssimati. La sua esattezza non implica però un comportamento perfetto, secondo [Blaker \(1998\)](#) e [Agresti & Coull \(1998\)](#) esso tende ad essere molto conservativo: privilegia una copertura molto elevata spesso troppo superiore al livello nominale.

Un'alta probabilità di copertura è certamente apprezzata, ma questa viene accompagnata da una stima più "generica", ovvero un intervallo più ampio. L'intervallo in questione è anche caratterizzato da estremi poco intuitivi e il cui legame con la regione di rifiuto è quello maggiormente complesso da dimostrare tra quelli che verranno analizzati. Il comportamento per valori limite di  $X$  è interessante, infatti per  $x = 0$  l'estremo inferiore dell'intervallo si azzerava, per  $x = n$  l'estremo superiore diventa 1.

### Applicazione sui dati

Il sottoinsieme di dati di dimensioni ridotte (2.9) produce estremi dell'intervallo di confidenza di Clopper-Pearson al 95%  $(q_{\alpha/2,4,2}^B, q_{1-\alpha/2,5,1}^B) = (0.284, 0.995)$ . Per quanto riguarda (2.10) invece  $(q_{\alpha/2,59,42}^B, q_{1-\alpha/2,53,41}^B) = (0.487, 0.687)$ .

### 2.3.2 Intervallo di Blaker

L'intervallo di Clopper-Pearson ha assunto, nel corso degli anni, il ruolo di *gold standard* tra i metodi di stima intervallare per una v.c. binomiale. Il suddetto titolo è dovuto interamente al fatto che esso sia un approccio esatto, ovvero che garantisce una copertura sempre almeno pari al coefficiente di confidenza. Nonostante ciò, è noto che l'intervallo di Clopper-Pearson tenda ad un'eccessiva conservatività.

Blaker (1998) introduce un intervallo che sostiene debba detronizzare il Clopper-Pearson come gold-standard nell'ambito di stima intervallare in un modello binomiale. Esso infatti riesce ad essere un intervallo esatto, ma non dovrebbe eccedere nella copertura. Di conseguenza la probabilità di copertura sarà sempre maggiore del livello nominale, ma rimarrà vicina ad esso.

### Accettabilità

La creazione dell'intervallo passa attraverso una funzione chiamata *funzione di accettabilità*. Con l'obiettivo di migliorare l'intervallo di Clopper-Pearson, Blaker cerca difatti una funzione che possa essere una curva di confidenza migliore da cui costruire un intervallo esatto. Per farlo, si impiega il seguente Teorema:

**Teorema 2.3.** *Sia  $X$  una variabile casuale con distribuzione indicizzata da  $p$  e  $T = T(X)$  statistica qualsiasi. Si definisca la funzione  $\gamma(p, t) = \min\{\mathbb{P}(T \geq t | p), \mathbb{P}(T \leq t | p)\}$  e sia  $\alpha(p, t) = \mathbb{P}[\gamma(p, T) \leq \gamma(p, t) | p]$ . Allora:*

- *L'insieme  $S_\alpha(t) = \{p : \alpha(p, t) > \alpha\}$  è un insieme di confidenza di livello  $1 - \alpha$  per  $p$ ,*

*Definendo  $\beta(p, t) = \min[2 \min\{\mathbb{P}(T \geq t | p), \mathbb{P}(T \leq t | p)\}, 1]$  e  $U_\alpha(t)$  l'intervallo derivante da  $\beta(p, t)$ :*

- *$\alpha(p, t) \leq \beta(p, t)$  quindi  $S_\alpha(t) \subseteq U_\alpha(t)$ ,*
- *$1 - \alpha \leq \mathbb{P}[p \in S_\alpha(T) | p] \leq \mathbb{P}[p \in U_\alpha(T) | p]$ .*

Ne consegue che la scelta di  $\alpha(p, t)$  permette la creazione di un intervallo sempre esatto ma più corto di quello di Clopper-Pearson, che sarebbe  $U_\alpha(T)$ . Nel caso binomiale la statistica a cui viene fatto riferimento è  $T = X$ .

Avendo osservato il campione  $X = x$ , per verificare se il generico valore  $p'$  del parametro  $p$  è compreso nell'intervallo si svolgono i seguenti passaggi:

1. Si calcola  $\gamma(p', x) = \min\{\mathbb{P}(X \geq x | p'), \mathbb{P}(X \leq x | p')\}$  per tutti i possibili valori di  $X$ ,
2. Si considerano tutte le  $x_i$  per cui  $\gamma(p', x_i) \leq \gamma(p', x)$ ,
3. Per trovare  $\alpha(p', x) = \mathbb{P}[\gamma(p', X) \leq \gamma(p', x) | p']$  si utilizza

$$\sum_{x_i: \gamma(p', x_i) \leq \gamma(p', x)} \mathbb{P}(X = x_i | p') = \sum_{x_i: \gamma(p', x_i) \leq \gamma(p', x)} \binom{n}{x_i} (p')^{x_i} (1 - p')^{n - x_i},$$

4. Se  $\alpha(p', x) > \alpha$ ,  $p'$  è incluso in  $S_\alpha(x)$ .

### Costruzione dell'intervallo

A livello pratico è necessario l'utilizzo di un software, infatti verrà sfruttata una funzione che possa trovare per quali valori di  $p$  la quantità  $\alpha(p, x) - \alpha$  si annulla. Il software svolge in maniera implicita il procedimento descritto per  $p'$  per tutti i valori di  $p$ , perciò riuscirà ad individuare i valori limite dell'intervallo.

**Il nuovo gold standard** L'intervallo di Blaker rappresenta di fatto il metodo da seguire se si desidera un intervallo esatto. Esso marginalizza completamente l'applicazione del Clopper-Pearson, in quanto le sue performance sono oggettivamente peggiori di quello di Blaker.

La ricerca di metodi di stima intervallare risulta tuttavia ancora aperta, poiché gli approcci esatti non costituiscono necessariamente la strada da seguire. La filosofia dietro questi metodi richiede infatti una copertura che sia sempre almeno pari al livello nominale, tuttavia spesso risulta desiderabile una stima che rispetti questo livello in media. È evidente che l'esigenza di un supporto computazionale renda meno appetibile questo metodo in contesti poco specializzati.

### Applicazione ai dati

Analizzando il dataset riguardante la sopravvivenza post-attacco cardiaco, l'intervallo di confidenza ottenuto è

$$(0.343, 0.99).$$

L'applicazione del metodo di Blaker all'insieme di dati riguardanti la lega cestistica risulta in un intervallo:

$$(0.49, 0.687).$$

### 2.3.3 Intervallo di confidenza dal Log-rapporto di verosimiglianza

L'IC si ottiene partendo dal test che sfrutta come statistica  $W(x)$ , esso avrà regione di rifiuto  $R = \{x : \mathbb{P}(W(x) \geq \lambda_\alpha | p_0) = \alpha\}$  con  $x \in (0, \dots, n)$ .

La regione risulta essere unilaterale destra perché il rapporto di verosimiglianza è tanto maggiore quanto più il valore ipotizzato da  $H_0$  è verosimile in relazione al campione osservato  $x$ .

A livello pratico questa probabilità viene calcolata sommando tutte le probabilità che  $x$  assuma valori per i quali  $W(x)$  supera la soglia prevista.

L'obiettivo tuttavia è costruire un intervallo di confidenza per il parametro  $p$ , quindi si considera la regione di accettazione  $A(p_0)$  definita come  $A(p_0) = \{x : W(x) \leq \lambda_\alpha | p_0\}$ . A questa regione corrisponderà l'intervallo di confidenza

$$C(x) = \{p_0 : x \in A(p_0)\} = \{p_0 : W(x) \leq \lambda_\alpha\} = \{p_0 : W(x) - \lambda_\alpha \leq 0\}$$

la regione è effettivamente un intervallo perché, come mostrato in [Azzalini \(2008\)](#), l'andamento di  $W(x)$ , al quale è stato già sottratto  $\lambda_\alpha$ , è il seguente

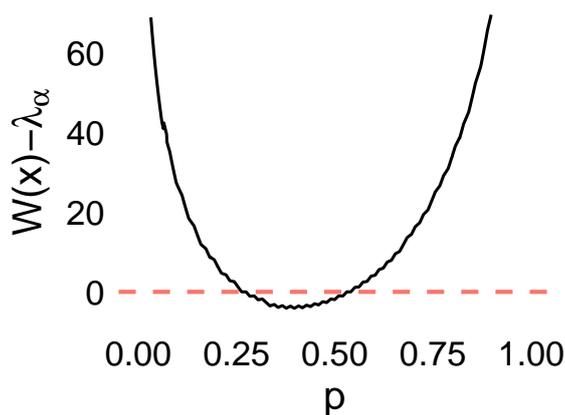


Figura 2.1: Andamento di  $W(x) - \lambda_\alpha$  al variare del parametro  $p$ .

è evidente che imponendo una soglia massima per  $W(x)$  (e quindi anche per  $W(x) - \lambda_\alpha$ ) si ottiene un intervallo per  $p$ . La formula per questo intervallo non è esprimibile in forma chiusa, questo lascia come unica opzione la via numerica. Gli estremi dell'intervallo sono le soluzioni all'equazione  $W(x) - \lambda_\alpha = 0$  che come si vede dalla figura sono due e sono distinte. Prima di risolvere l'equazione è necessario conoscere  $\lambda_\alpha$  ovvero la soglia per  $W(x)$ . Per calcolarla si determina  $W(x)$  per ogni valore di  $x \in (0, \dots, n)$  e la corrispondente  $\mathbb{P}(X = x \mid p = p_0)$ ; siccome il test ha probabilità di errore del primo tipo pari ad  $\alpha$  si prenderà come soglia per  $W(x)$  un valore  $\lambda_\alpha$  tale che  $\mathbb{P}(W(x) \geq \lambda_\alpha \mid p = p_0) = \alpha$ . Ritrovandosi in un caso discreto basterà  $\mathbb{P}(W(x) \geq \lambda_\alpha \mid p = p_0) \leq \alpha$  quindi  $\lambda_\alpha = W^*(x)$  sarà il valore più piccolo tra i  $W(x)$  che soddisfano questa disuguaglianza

$$\sum_{k:W(k) \geq W^*(x)} \mathbb{P}(X = k \mid p = p_0) \leq \alpha.$$

**Esempio:** Tabella con i valori di  $W(x)$  e  $\mathbb{P}(X = x \mid p = p_0)$  nel caso di  $x = 3$   $p_0 = 0.45$   $n = 10$

$x$	$W(x)$	$P(x = x \mid p = p_0)$
0	11.96	0.003
1	5.86	0.021
2	2.75	0.076
3	0.94	0.166
4	0.10	0.238
5	0.10	0.234
6	0.90	0.160
7	2.55	0.075
8	5.16	0.023
9	9.07	0.004
10	15.97	0.000

se  $\alpha = 0.05$  il valore scelto sarà  $W^*(x) = 5.16$ .

Determinato  $\lambda_\alpha = W^*(x)$ , le soluzioni all'equazione saranno determinate tramite il comando *uniroot* sul software R. Nell'esempio l'IC sarà (0.135, 0.632).

**Difficoltà procedurali** Come è evidente dall'esempio appena descritto, la derivazione di un intervallo di confidenza dal log-rapporto di verosimiglianza è un processo lungo e dispendioso o addirittura proibitivo se non si ha a disposizione un software che possa trovare le soluzioni a  $W(x) - \lambda_\alpha = 0$ . A livello computazionale, l'algoritmo che permette di arrivare all'intervallo risulta molto intenso quando la numerosità campionaria è elevata. Il fatto che non portasse sostanziali miglioramenti rispetto ad approcci più semplici e veloci ha fatto in modo che questo metodo venisse accantonato.

### Applicazione sui dati

Nel primo insieme di dati (2.9) l'intervallo al 95% risulta

$$(0.376, 0.987).$$

Riguardo al secondo insieme (2.10) si avrà una stima intervallare data da

$$(0.493, 0.683).$$

### 2.3.4 Intervallo di confidenza di Wald

L'intervallo di confidenza per la v.c. binomiale a cui si farà riferimento è un intervallo approssimato, con estremi

$$\hat{p} \pm z_{\alpha/2} \sqrt{\hat{p}(1-\hat{p})/n},$$

dove  $z_\alpha$  è il quantile di ordine  $\alpha$  di una distribuzione normale standard.

#### Test di Wald

L'intervallo appena definito deriva direttamente dall'omonimo test.

*Dimostrazione.* Da (2.7) si apprende che il test di Wald, con  $X \sim \text{Bin}(n, p)$  ha espressione

$$W_e(x) = n \frac{(\hat{p} - p_0)^2}{\hat{p}(1-\hat{p})},$$

riscrivibile come

$$W_e(x) = \left( \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \right)^2,$$

questa statistica test ha distribuzione asintotica  $\chi_1^2$  supponendo  $H_0$  vera. La relazione tra normale e chi-quadrato implica che sotto le stesse condizioni

$$\text{sign}(\hat{p} - p_0) \sqrt{W_e(x)} = \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \xrightarrow{d} Z \sim N(0, 1).$$

Si costruiscono adesso le regioni del test, in particolare quella di accettazione che permetterà di giungere ad un intervallo di confidenza per  $p$ . Come statistica test si impiega ovviamente  $W_e(x)$ .

$$A(p_0) : \mathbb{P}(X \in A(p_0) \mid p = p_0) = 1 - \alpha,$$

$$\mathbb{P}(X \in A(p_0) \mid p = p_0) = \mathbb{P} \left( z_{\alpha/2} \leq \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \leq z_{1-\alpha/2} \mid p = p_0 \right)$$

questa probabilità, al divergere di  $n$ , è proprio pari ad  $1 - \alpha$ . Utilizzando il Teorema che lega intervalli di confidenza a test d'ipotesi:

$$C(x) = \{p_0 : x \in A(p_0)\} = \left\{ p_0 : z_{\alpha/2} \leq \frac{\hat{p} - p_0}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \leq z_{1-\alpha/2} \right\}$$

perciò isolando  $p_0$  e riscrivendolo come  $p$

$$C(x) = \left\{ p : \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right\}.$$

□

**Un approccio liberale** Nonostante sia il metodo presentato più spesso nei problemi di stima intervallare di una v.c. binomiale, questo intervallo è notoriamente poco adatto e sicuramente rappresenta un'approssimazione valida solo in contesti di campioni estremamente numerosi. Come approccio è inoltre fortemente liberale, ovvero tende ad avere probabilità di copertura al di sotto del livello nominale, specialmente quando  $p$  si avvicina agli estremi del suo supporto. Chiaramente una proprietà come questa è poco apprezzabile in un contesto come quello di stima intervallare. Gli ultimi importanti difetti di questo metodo sono l'impossibilità di applicarlo per valori estremi di  $x$ , ovvero  $x \in \{0, 1\}$ , e la possibilità di avere estremi esterni al supporto del parametro d'interesse. Tuttavia il metodo ha sicuramente dalla sua parte una grande semplicità e interpretabilità, che lo rendono il più diffuso e conosciuto in contesti esterni alla statistica.

### Applicazione ai dati

Il valore della statistica test di Wald in corrispondenza dei dati (2.9) è  $(0.8 - p)/0.179$ . L'intervallo di confidenza al 95% sarà invece

$$\left( 0.8 - 1.96 \sqrt{\frac{0.8 \cdot 0.2}{5}}, 0.8 + 1.96 \sqrt{\frac{0.8 \cdot 0.2}{5}} \right) = (0.449, 1.150).$$

Con l'altro sottoinsieme di dati (2.10) si riscontra una quantità pivotale pari a  $(0.59 - p)/0.049$  e un intervallo

$$\left( 0.59 - 1.96 \sqrt{\frac{0.59 \cdot 0.41}{100}}, 0.59 + 1.96 \sqrt{\frac{0.59 \cdot 0.41}{100}} \right) = (0.494, 0.686).$$

### Intervallo di Wald corretto

L'intervallo di Wald è notoriamente un intervallo con performance poco desiderabili, è possibile ottenere un miglioramento sostanziale per questo con l'aggiunta (fittizia) di due successi e due insuccessi al risultato dell'esperimento. Questa "aggiunta" si riflette nel punto centrale del nuovo intervallo  $\tilde{p} = (X + 2)/(n + 4)$  creando l'*intervallo di Wald corretto*, esaminato in [Agresti & Coull \(1998\)](#).

Il cambiamento apportato sposta l'intervallo di Wald verso quello di Wilson (2.3.5) per quanto riguarda il punto centrale, esso inoltre permette di passare da un approccio fortemente liberale a uno leggermente conservativo, per cui la probabilità di copertura sarà tendenzialmente più alta del coefficiente di confidenza.

Il miglioramento portato in termini di copertura verrà approfondito nel Capitolo 3.

### Applicazione ai dati

La costruzione dell'intervallo è sostanzialmente identica, ma verrà utilizzata  $\tilde{p}$  al posto di  $\hat{p}$ :  $n_1 = 5$   $x_1 = 4$   $\hat{p}_1 = 0.8$   $\tilde{p} = 6/9 = 0.67$ . Questo cambiamento dà origine ad una statistica test pari a  $(0.67 - p)/0.157$  e relativo intervallo

$$\left( 0.67 - 1.96\sqrt{\frac{0.67 \cdot 0.33}{9}}, 0.67 + 1.96\sqrt{\frac{0.67 \cdot 0.33}{9}} \right) = (0.359, 0.975).$$

Per il secondo sottoinsieme di dati:  $n_2 = 100$   $x_2 = 59$   $\hat{p}_2 = 0.59$   $\tilde{p} = 61/104 = 0.587$ , la quantità pivotale  $(0.587 - p)/0.048$  e relativo intervallo

$$\left( 0.587 - 1.96\sqrt{\frac{0.587 \cdot 0.413}{104}}, 0.587 + 1.96\sqrt{\frac{0.587 \cdot 0.413}{104}} \right) = (0.492, 0.681).$$

### 2.3.5 Intervallo di confidenza di Wilson

Anche in questo caso si tratta un intervallo approssimato: l'intervallo di Wilson, i cui estremi sono

$$\frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n} \pm z_{\alpha/2} \sqrt{\frac{1}{n} \left[ \hat{p}(1 - \hat{p}) + \frac{z_{\alpha/2}^2}{4n} \right]}}{1 + \frac{z_{\alpha/2}^2}{n}}$$

questo si può ottenere dal test del punteggio.

### Test di Wilson

Gli estremi per l'intervallo sono ricavabili dal test di Wilson (2.8)

$$W_u(x) = n \frac{(\hat{p} - p_0)^2}{p_0(1-p_0)} = \left( \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \right)^2 \stackrel{d}{H_0} Z^2 \sim \chi_1,$$

*Dimostrazione.* Gli estremi si ricavano partendo dalla q.p. approssimata

$$\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \stackrel{d}{H_0} Z \sim N(0, 1).$$

Si segue lo stesso procedimento svolto col test di Wald, ovvero si ricava la regione di accettazione  $A(p_0)$  e l'intervallo di confidenza  $C(x)$ :

$$C(x) = \{p_0 : x \in A(p_0)\} = \left\{ p_0 : z_{\alpha/2} \leq \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \leq z_{1-\alpha/2} \right\}.$$

Stavolta, nell'inversione per  $p_0$ , si considera anche il denominatore,

$$z_{\alpha/2} \leq \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \leq z_{1-\alpha/2}.$$

Si prendono in esame solo gli estremi, per ricavarli si risolve quindi l'equazione

$$\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \pm z_{\alpha/2}.$$

I passaggi per la risoluzione sono:

$$\frac{(\hat{p} - p_0)^2}{p_0(1-p_0)} = z_{\alpha/2}^2,$$

$$[\hat{p}^2 + p_0^2 - 2p_0\hat{p}] = z_{\alpha/2}^2 \frac{p_0(1-p_0)}{n},$$

a questo punto si raccoglie

$$p_0^2 \left[ 1 + \frac{z_{\alpha/2}^2}{n} \right] - 2p_0 \left[ \hat{p} + \frac{z_{\alpha/2}^2}{2n} \right] + \hat{p}^2 = 0.$$

Applicando la formula per le equazioni di secondo grado:

$$p_0 = \frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n} \pm \sqrt{\left[ \hat{p} + \frac{z_{\alpha/2}^2}{2n} \right]^2 - \hat{p}^2 \left[ 1 + \frac{z_{\alpha/2}^2}{n} \right]}}{1 + \frac{z_{\alpha/2}^2}{n}},$$

$$p_0 = \frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n} \pm \sqrt{\hat{p}^2 + \frac{z_{\alpha/2}^4}{4n^2} + \frac{z_{\alpha/2}}{n} \hat{p} - \hat{p}^2 - \hat{p} \frac{z_{\alpha/2}^2}{n}}}{1 + \frac{z_{\alpha/2}^2}{n}}.$$

Si semplifica e si raccoglie nella radice  $z_{\alpha/2}^2$  e  $\frac{1}{n}$ , giungendo così agli estremi presentati all'inizio, ovvero:

$$p_0 = \frac{\hat{p} + \frac{z_{\alpha/2}^2}{2n} \pm z_{\alpha/2} \sqrt{\frac{1}{n} \left[ \hat{p}(1 - \hat{p}) + \frac{z_{\alpha/2}^2}{4n} \right]}}{1 + \frac{z_{\alpha/2}^2}{n}}.$$

□

**Un'approssimazione eccellente** L'intervallo basato sulla funzione di punteggio è uno dei metodi più validi di stima intervallate nel contesto binomiale, nonostante la sua natura asintotica ha performance sorprendenti anche lavorando con campioni esigui. Esso sembra rappresentare un giusto compromesso tra la conservatività dell'intervallo di Clopper-Pearson e la liberalità di quello di Wald. Ad ogni modo, la formula per ricavare gli estremi di tale intervallo è certamente la più complicata da presentare e l'inversione della quantità pivotale è un procedimento leggermente più complesso rispetto ad altri metodi. Il comportamento per valori limite di  $X$  è lo stesso di quello di Clopper-Pearson, per quanto riguarda l'estremo inferiore, se  $x = 0$ , o superiore, se  $x = n$ .

### Applicazione ai dati

Con riferimento a (2.9) avremo un IC:

$$\left( \frac{0.8 + \frac{1.96^2}{10} \mp 1.96 \sqrt{\frac{0.8 \cdot 0.2}{5} + \frac{1.96^2}{20}}}{1 + \frac{1.96^2}{5}} \right) = (0.376, 0.964).$$

Avendo osservato (2.10) invece:

$$\left( \frac{0.59 + \frac{1.96^2}{200} \mp 1.96 \sqrt{\frac{0.59 \cdot 0.41}{100} + \frac{1.96^2}{400}}}{1 + \frac{1.96^2}{100}} \right) = (0.492, 0.681).$$

**Equivalenza asintotica** I tre intervalli appena presentati (TRV, Wald e Wilson), sono asintoticamente equivalenti, in quanto i tre test da cui derivano lo sono. A livello pratico però, sono estremamente diversi nel loro comportamento quando  $n$  è finito.

### 2.3.6 Intervallo di credibilità

L'ultimo approccio presentato appartiene al mondo della statistica bayesiana, la cui particolarità è la considerazione del parametro come una variabile casuale. Esso ha una distribuzione a priori e una a posteriori, i cui fondamenti teorici non verranno discussi.

Il concetto di intervallo di confidenza è sostituito da quello di intervallo di credibilità, la cui definizione secondo Casella & Berger (2008) è:

**Intervallo di credibilità** Sia  $\pi(\theta | x)$  la distribuzione a posteriori di  $\theta$  dato  $X = x$ ,  $A \subset \Theta$  si definisce un *intervallo di credibilità* con probabilità di credibilità  $1 - \alpha$  se:

$$\mathbb{P}(\theta \in A | x) = \int_A \pi(\theta | x) d\theta = 1 - \alpha,$$

dove l'integrale viene sostituito da una sommatoria in caso di distribuzioni a posteriori discrete.

Si ricorda che nel contesto Bayesiano la distribuzione a posteriori è influenzata dalla scelta, più o meno soggettiva, di una distribuzione a priori per il parametro. Si illustra un esempio con una distribuzione a priori che possiede proprietà asintotiche ideali.

### Priori di Jeffreys

La distribuzione a priori ideata da Jeffreys  $\pi(\theta) \propto |I(\theta)|^{1/2}$ , nel caso binomiale  $\pi(p) \propto \text{Beta}(1/2, 1/2)$ . La distribuzione a posteriori ricavata da questa priori è

$$\pi(p | x) \propto \text{Beta}\left(x + \frac{1}{2}, n - x + \frac{1}{2}\right),$$

l'intervallo di credibilità perciò sarà dato da

$$\left(q_{\alpha/2, x+1/2, n-x+1/2}^B, q_{1-\alpha/2, x+1/2, n-x+1/2}^B\right).$$

### Applicazione ai dati

L'osservazione dei dati (2.9) porterebbe ad un intervallo di credibilità:

$$\left(q_{\alpha/2, 4.5, 1.5}^B, q_{1-\alpha/2, 4.5, 1.5}^B\right) = (0.371, 0.977).$$

Invece, se il campione estratto fosse (2.10), si avrebbe

$$\left(q_{\alpha/2, 59.5, 41.5}^B, q_{1-\alpha/2, 59.5, 41.5}^B\right) = (0.492, 0.683).$$

### 2.3.7 Confronto pratico tra intervalli

Si utilizza un confronto su dati reali per esaminare le differenze tra i metodi di stima intervallare presentati. I dati in questione sono descritti in (2.9) e (2.10).

Metodo	n=5 $\hat{p} = 0.8$		n=100 $\hat{p} = 0.59$	
	Inf	Sup	Inf	Sup
Clopper-Pearson	0.284	0.995	0.487	0.687
Blaker	0.343	0.99	0.49	0.687
TRV	0.376	0.987	0.493	0.683
Wald	0.449	1.150	0.493	0.686
Wald corretto	0.359	0.975	0.492	0.681
Wilson	0.376	0.964	0.492	0.681
Bayesiano	0.371	0.977	0.492	0.683

**Tabella 2.1:** Estremi dei vari intervalli di confidenza al 95% per i dati (2.9) e (2.10).

Dai risultati in 4.2 si può dedurre innanzitutto che gli estremi dell'intervallo di Wald possono eccedere il supporto del parametro  $p \in [0, 1]$ . La conservatività dell'intervallo di Clopper-Pearson si riflette nella sua ampiezza, rimanendo quello

di dimensioni maggiori anche quando  $n = 100$ , mentre la differenza col metodo di Blaker risulta più evidente quando  $n = 5$ . Nonostante la somiglianza delle formule degli estremi degli intervalli di credibilità e di Clopper-Pearson, il primo sembra allinearsi nei risultati con il metodo corretto di Wald e quello di Wilson e il secondo con quello di Blaker. Il log-rapporto di verosimiglianza non sembra allinearsi agli intervalli basati su distribuzioni asintotiche, per il dataset con 5 osservazioni. Come ci si aspetterebbe, tutti gli intervalli contengono la stima di massima verosimiglianza. La presenza ridotta di osservazioni nel primo set di dati fa in modo che tutti i metodi rimangano piuttosto generici nella loro stima, optando per intervalli decisamente ampi. Le differenze nelle stime tra i vari metodi tendono ad assottigliarsi all'aumentare della numerosità campionaria.

## Capitolo 3

# Copertura e ampiezza

Questo capitolo sarà dedicato alla valutazione dei metodi di stima intervallare presentati in precedenza, tramite varie grandezze. Tutte le quantità utilizzate sono prese da [Agresti & Coull \(1998\)](#), espandendo le analisi ad ulteriori intervalli rispetto a questo.

### 3.1 Confronto con il coefficiente di confidenza

Il coefficiente di confidenza è stato descritto come il minimo tra i valori di probabilità di copertura al variare del valore del parametro. Tutti gli intervalli costruiti dovrebbero essere equivalenti per quanto riguarda questo coefficiente, ma ciò non implica che lo siano dal punto di vista della probabilità di copertura.

**Metodi approssimati** Quanto detto non è esattamente corretto per i metodi asintotici, in quanto essi rispettano queste condizioni al divergere di  $n$ . In problemi reali non è garantito il livello minimo di copertura.

A causa di questa discrepanza, risulta interessante valutare i metodi di stima intervallare utilizzando la probabilità di copertura e vedendo quanto si distacca dal livello di confidenza  $1 - \alpha$ . Nella costruzione di un insieme di confidenza è desiderabile che esso abbia una probabilità di copertura vicina al livello nominale, se un intervallo di livello  $1 - \alpha$  ha una probabilità di copertura che tende ad essere molto maggiore del suo minimo, questo renderà il metodo troppo conservativo e quindi poco ideale.

Perché un intervallo con copertura troppo elevata non è ideale? Se l'intervallo ha probabilità di copertura più alte di quanto richiesto, esso dovrà essere

necessariamente più ampio rispetto al caso in cui dovesse mantenere quella predeterminata. La probabilità di copertura verrà dunque valutata anche in riferimento alla lunghezza dell'intervallo.

### 3.1.1 Probabilità di copertura

La probabilità di copertura per un parametro  $p$  di una v.c. binomiale è

$$C_n(p) = \sum_{k=0}^n I(k, p) \binom{n}{k} p^k (1-p)^{n-k},$$

dove  $I(k, p)$  è uguale a 1 se l'intervallo contiene  $p$  quando  $X = k$  e 0 altrimenti. Concentrandosi su questa formula, la scrittura è equivalente a

$$C_n(p) = \sum_{k=0}^n \mathbb{P}(X = k | p) * I(C(k) \ni p),$$

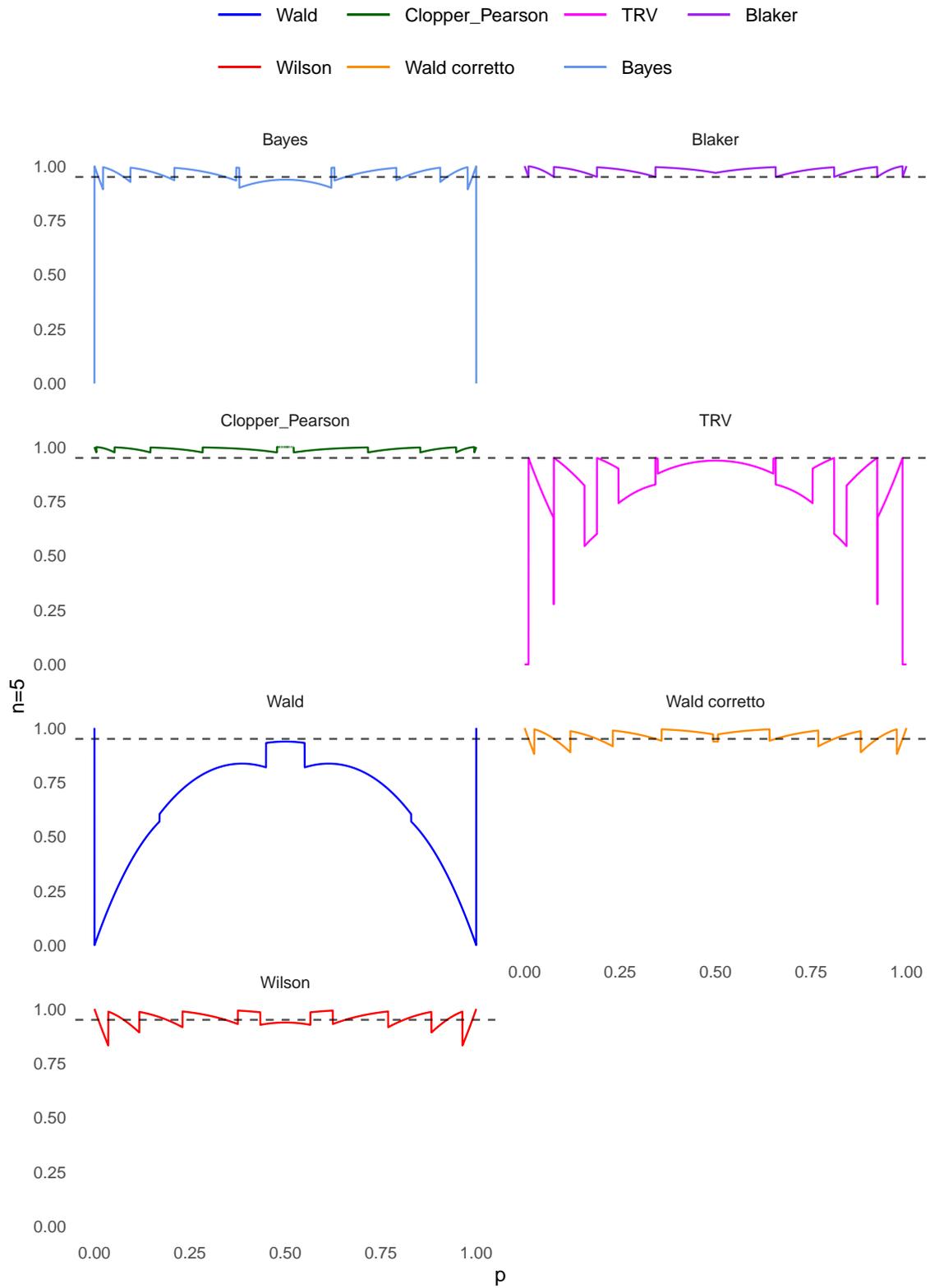
dove  $C(k)$  è l'intervallo di confidenza creato per  $X = k$ . Avendo una sommatoria su tutto il supporto di  $X$  e la funzione di probabilità della variabile casuale, si può scrivere  $C_n(p) = \mathbb{E}[I(C(X) \ni p) | p]$ .

Questa quantità valuta quindi la copertura dell'intervallo di confidenza per ogni possibile valore di  $X$  con un  $p$  fissato, essa di conseguenza dipende necessariamente dal valore scelto  $p$ .

Si rappresentano di seguito le probabilità di copertura dei diversi metodi di stima intervallare presentati, prima per  $n = 5$  poi per  $n = 100$ .

Il grafico 3.1 rende evidente l'insoddisfacente copertura degli intervalli di Wald e del log-rapporto di verosimiglianza. In particolare, è chiaro che l'approssimazione derivante dall'intervallo di Wald risulta totalmente inapplicabile per campioni di numerosità così esigua. La correzione applicata sul metodo di Wald comporta un netto miglioramento nella copertura anche per  $n = 5$ . Il comportamento agli estremi cambia molto tra i diversi metodi: il bayesiano e il TRV hanno copertura nulla ad entrambi gli estremi, Wald invece arriva ad una copertura pari a 1 per  $p \in \{0, 1\}$ .

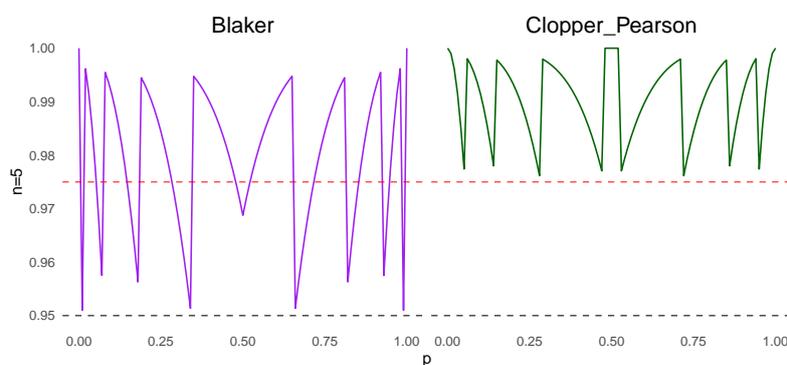
Agresti & Coull (1998) fa notare una particolarità riguardo l'intervallo di Wilson: esso possiede due regioni simmetriche, agli estremi del supporto di  $p$ , in



**Figura 3.1:** Probabilità di copertura al variare di  $p$  per  $n = 5$  per gli intervalli di confidenza al 95%.

cui la probabilità di copertura cade considerevolmente sotto il livello nominale. Prendendo in esame la regione di sinistra, il calo avviene perché sotto una certa soglia, il vero valore di  $p$  viene accettato solo se  $X = 0$ . Lo stesso vale per la regione di destra, ovviamente con  $p$  maggiore di un certo valore che viene compreso nell'IC solo se  $X = n$ . Nel caso di  $n = 5$  si ha che a  $p = 0.036$  e  $p = 0.964$  la probabilità di copertura è pari a 0.832. Le regioni in questione diventano più piccole all'aumentare di  $n$ , infatti per  $n = 100$  si ha una copertura di 0.838 a  $p = 0.00177$  e  $p = 0.99823$ .

L'intervallo di Blaker dovrebbe sostituire il Clopper-Pearson come l'intervallo esatto di riferimento, per veder se ciò è confermato dalla copertura, la si esamina più dettagliatamente.

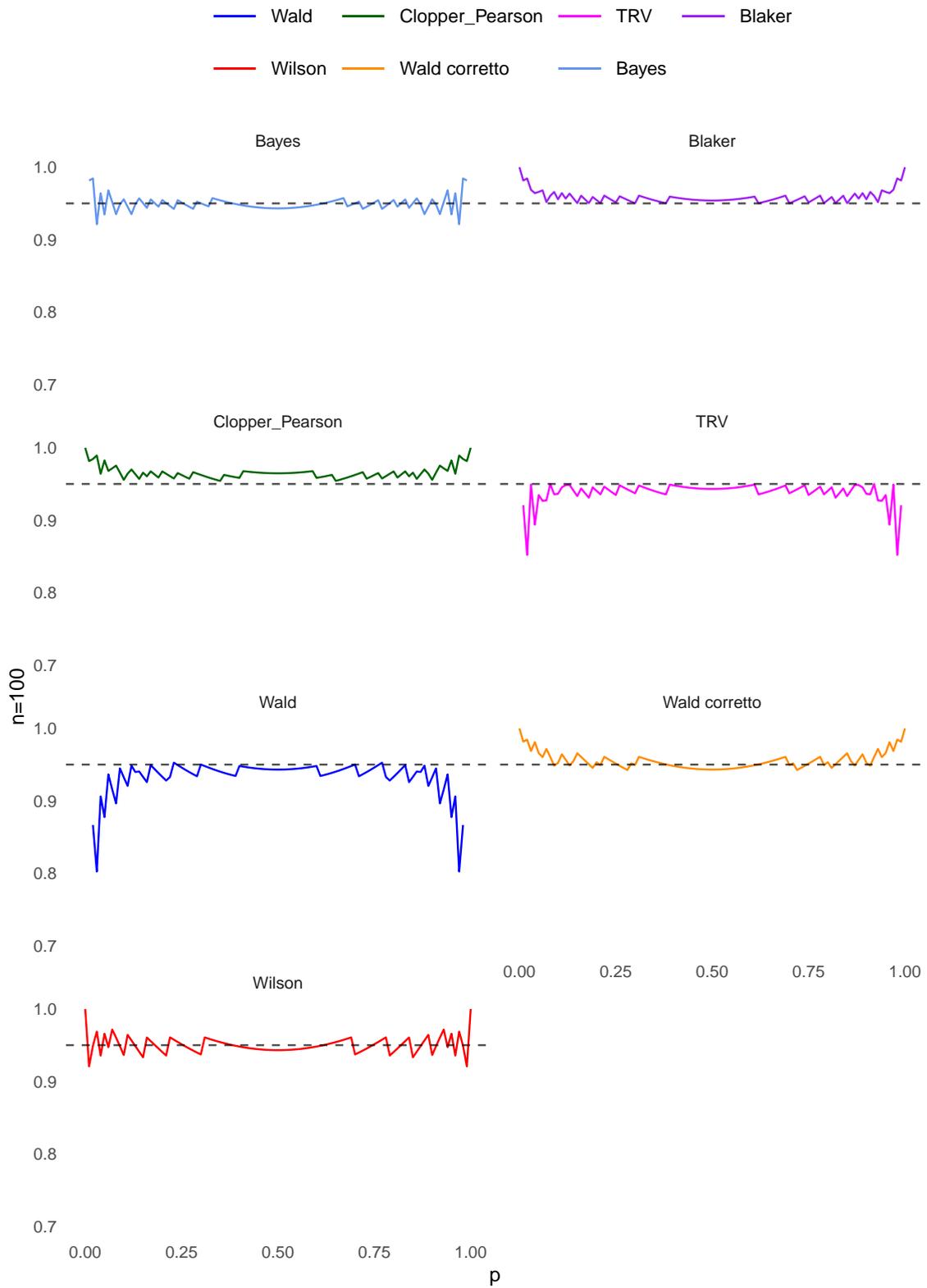


**Figura 3.2:** Confronto tra le probabilità di copertura degli intervalli di confidenza esatti al 95% con  $n = 5$ .

Come fa notare [Blaker \(1998\)](#), il Clopper-Pearson non scende sotto una copertura pari ad  $1 - \alpha/2 = 97.5\%$ , confermandosi esageratamente conservativo. Il comportamento del Blaker è molto più ragionevole, toccando spesso la probabilità desiderata.

Per verificare l'efficacia delle approssimazioni, si esaminano le coperture quando la numerosità campionaria è pari a 100.

Da [3.3](#) si può apprezzare il miglioramento portato dalla correzione all'intervallo di Wald, la copertura infatti si mantiene tendenzialmente più vicina al livello nominale e subisce lievi cali verso gli estremi. Il calo di prestazioni quando  $p$  tende ai suoi valori limite si ha anche nell'intervallo basato sul log-rapporto di verosimiglianza. L'aumentare della numerosità campionaria ha diminuito la differenza di prestazioni tra i due intervalli esatti e ha logicamente portato



**Figura 3.3:** Probabilità di copertura al variare di  $p$  per  $n = 100$  per gli intervalli di confidenza al 95%

a coperture più adatte per quelli approssimati. In una situazione simile, dove si potrebbe dire che  $n$  è *sufficientemente elevato*, non tutte le approssimazioni garantiscono una probabilità di copertura vicina a 0.95. Risulta perciò importante contestualizzare l'affermazione *valido per  $n$  sufficientemente grande*, soprattutto considerando che le performance di approssimazioni come il metodo di Wald dipendono fortemente da dove si trova il vero valore del parametro. Anche se non visibili in 3.3, i metodi bayesiano e TRV hanno copertura pari a zero agli estremi, ovvero a  $p \in \{0, 1\}$ .

**Copertura e appartenenza** Si precisa che in un problema di stima intervallare, in ambito frequentista, la quantità casuale è l'insieme di confidenza, poiché il parametro è fissato. È quindi imprecisa la frase "il parametro appartiene all'intervallo", in quanto è l'intervallo che ha l'obiettivo di coprire il parametro. Per questo motivo si parla di probabilità di copertura e non, per esempio, di probabilità di appartenenza.

### 3.1.2 Probabilità media di copertura

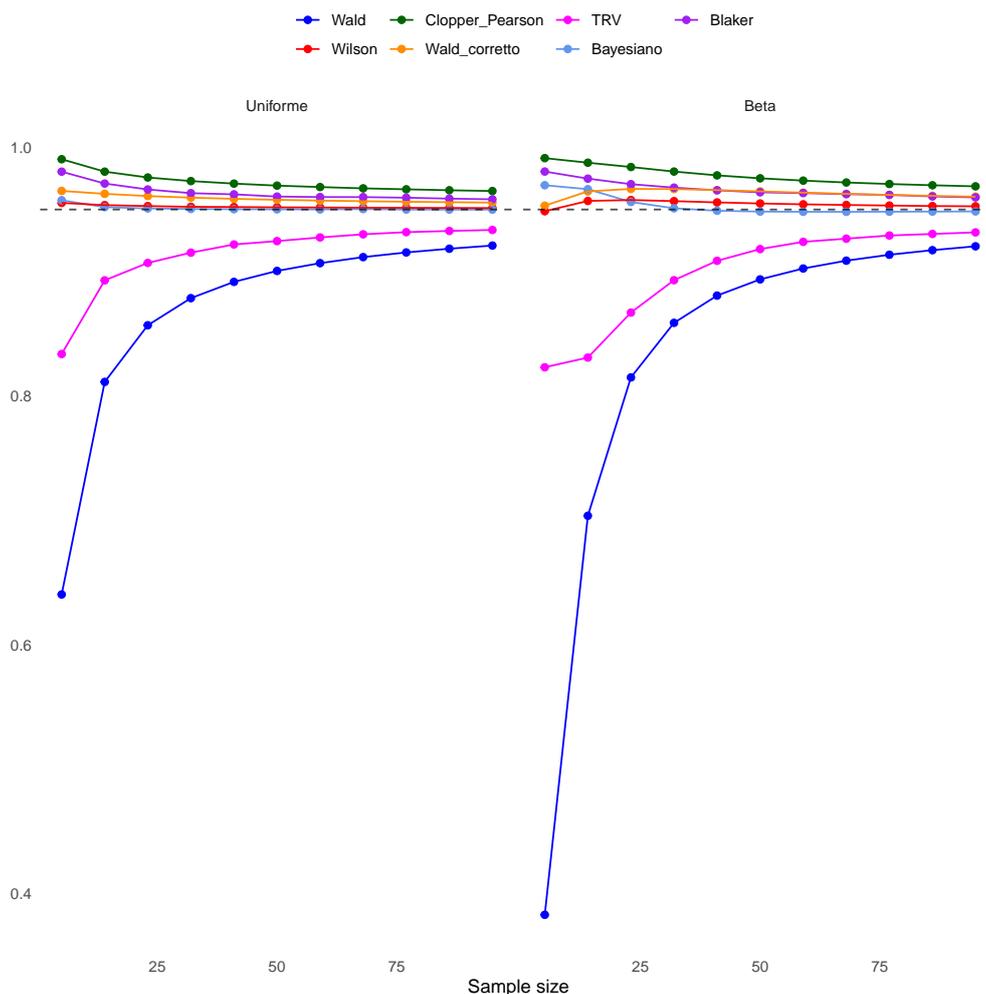
L'obiettivo di questo capitolo è la presa in esame delle probabilità di copertura per un'analisi quantitativa dei metodi di stima intervallare per una v.c. binomiale. Vi è però un ostacolo al confronto diretto tra gli intervalli, ovvero la dipendenza di  $C_n$  dal valore di  $p$ . Al fine di liberarla da questa dipendenza è possibile utilizzare una quantità usata in Agresti & Coull (1998); essa sfrutta un approccio bayesiano, selezionando una *distribuzione a priori* per il parametro e valutando la probabilità di copertura per ogni possibile valore di  $p$ . La quantità ottenuta è chiamata *probabilità media di copertura*, la cui formula è:

$$\bar{C}_n = \int_0^1 C_n(p)g(p)dp. \quad (3.1)$$

Per quanto riguarda la scelta della distribuzione a priori  $g(p)$ , si opta per le due distribuzioni selezionate in Agresti & Coull (1998): una Beta(0.35, 3.15) e una Uniforme(0, 1) per via del supporto del parametro.

Da 3.4 si nota che l'IC di Clopper-Pearson è decisamente conservativo, con coperture eccessivamente elevate quando lavora con campioni esigui, mentre il Blaker si conferma un'alternativa migliore a primo.

Il log-rapporto di verosimiglianza non dà origine ad una stima intervallare



**Figura 3.4:** Probabilità media di copertura in funzione della numerosità campionaria per gli intervalli di confidenza di livello 95% quando  $p$  ha un distribuzione Uniforme  $(0, 1)$  e Beta  $(0.35, 3.15)$

con coperture adatte che, considerata la complessità della sua costruzione, lo rende un metodo poco ideale. L'intervallo di Wald tende ad avere probabilità di copertura troppo inferiore al livello nominale, soprattutto per  $n$  piccoli, ma è sorprendente come questa rimanga sempre insufficiente anche per campioni di dimensioni notevoli. L'aggiustamento all'intervallo di Wald è estremamente efficace, rendendolo un intervallo valido anche per campioni di dimensioni ridotte. L'intervallo basato sulla funzione di punteggio sembra quello più equilibrato assieme al Wald corretto e al bayesiano, con probabilità di copertura nei pressi del livello nominale per qualsiasi  $n$  valutato.

La scelta tra le distribuzioni a priori diverse per il parametro  $p$  non influenza in modo significativo la copertura degli intervalli, l'unico cambiamento degno di

nota è quello sull'intervallo di Wald il quale ha una copertura minore passando da una Uniforme ad una Beta.

La decisione riguardo il un metodo di stima intervallare con copertura "ottimale" non è oggettiva e dipende dalle necessità di ricerca. Gli intervalli esatti riescono a garantire una probabilità sempre superiore al livello nominale, ma in media avranno ovviamente una copertura superiore di quella richiesta. Se si desidera un intervallo con copertura che sia in media pari ad  $1 - \alpha$ , la scelta dovrebbe ricadere sui metodi approssimati, facendo attenzione a sceglierne uno efficace. L'approssimazione offerta da Wilson, dall'intervallo bayesiano o dal Wald corretto sembrano in generale le più adatte, mentre l'intervallo di Blaker supera chiaramente il Clopper-Pearson nell'ambito di metodi esatti.

### 3.2 Ampiezza dell'intervallo

Come menzionato precedentemente, la probabilità di copertura è bene che venga accompagnata da una misura della lunghezza dell'intervallo, per dare un'indicazione riguardo quanto l'intervallo sia "efficiente" nell'individuazione di un range per il parametro. Un indice atto a misurare la lunghezza è la *lunghezza attesa* dell'intervallo, ovvero il valore atteso della lunghezza dell'intervallo valutata su tutti i possibili valori di  $X$  dati  $p$  ed  $n$ . Nel modello in esame, prendendo il generico intervallo di confidenza  $IC_i = (t_1(x), t_2(x))$ , la lunghezza attesa sarà:

$$\mathbb{E}[t_2(X) - t_1(X) | p] = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} [t_1(k) - t_2(k)].$$

La figura 4.10 mostra evidenti problemi nell'intervallo di Wald, il quale degenera agli estremi per  $p$ . L'analisi della copertura in 3.1 permette di scoprire che anch'essa cade vertiginosamente per valori del parametro che tendono a 0 e 1. Anche da questo punto di vista l'intervallo di Wilson ha un comportamento desiderabile dimostrandosi il più stretto per valori centrali di  $p$ . Tra i metodi approssimati che in 3.1 mostravano buone coperture, il bayesiano e il Wilson sono i più stretti, col primo che performa meglio agli estremi e il secondo che al centro risulta il migliore. L'intervallo di Wald corretto è il più ampio per valori limite di  $p$ , ma da 3.1 è anche quello con copertura maggiore tra i metodi approssimati. È proprio in questo grafico che l'intervallo di Blaker mostra la ragione per cui

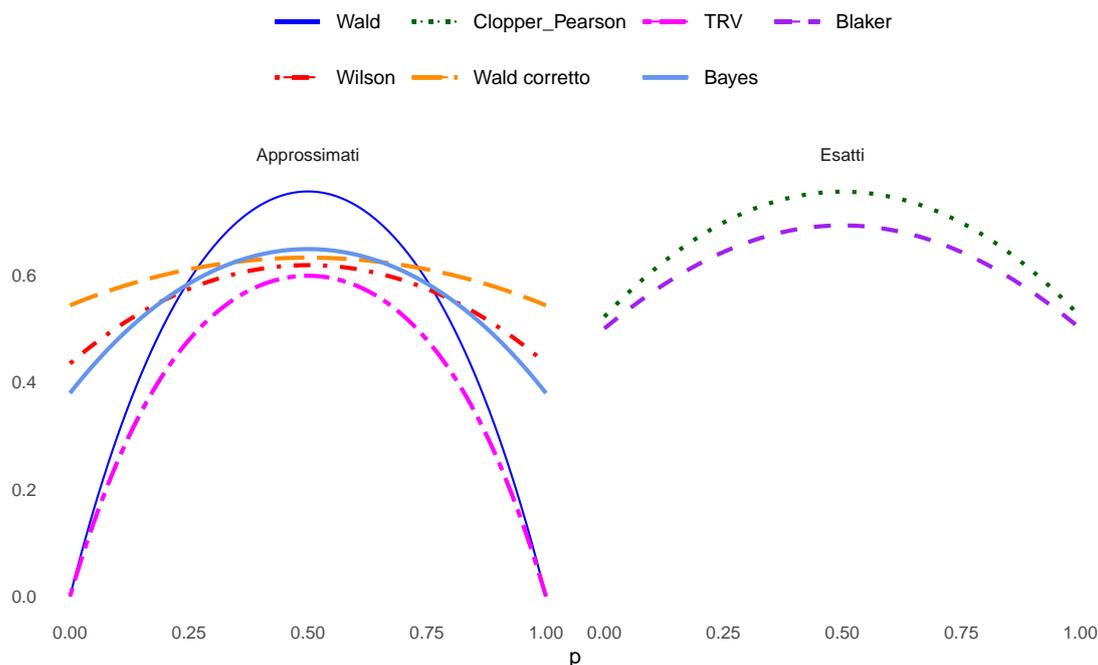


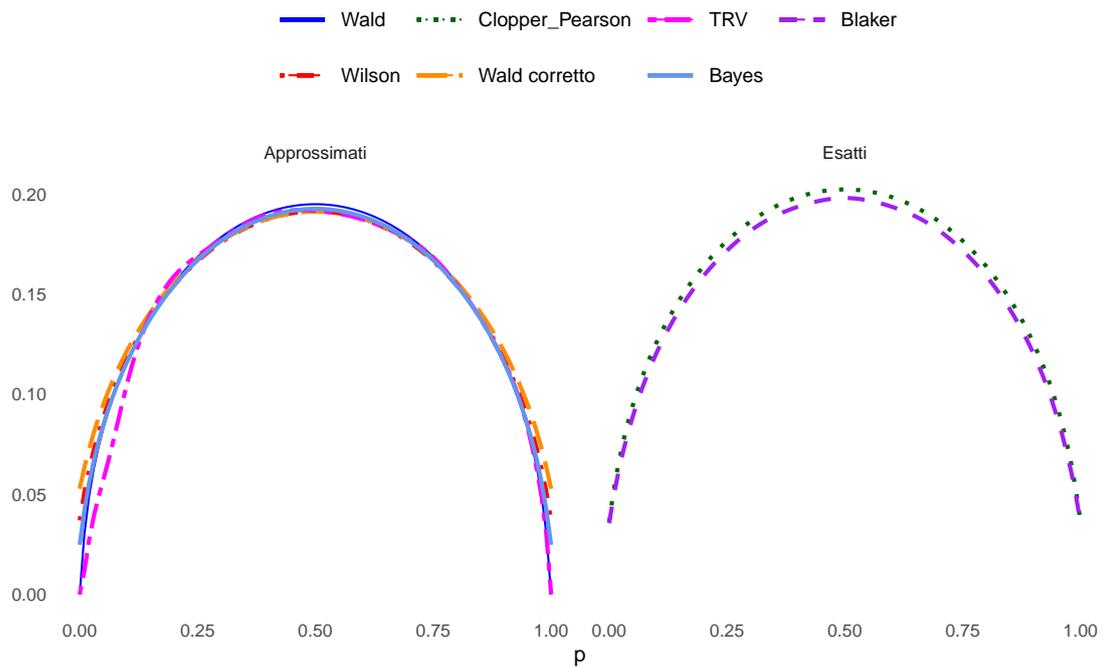
Figura 3.5: Lunghezza attesa al variare di p quando  $n = 5$

dovrebbe essere il nuovo gold-standard, esso infatti si trova sempre al di sotto del Clopper-Pearson in termini di ampiezza. L'intervallo basato sul log-rapporto di verosimiglianza risulta poco utile da analizzare, dato l'andamento della sua probabilità di copertura, poco confrontabile con gli altri metodi, sicuramente però il fatto che la lunghezza si annulli agli estremi non è apprezzabile. È interessante notare come alcuni metodi riescano ad ottenere coperture più elevate con intervalli spesso più corti. Questa coppia di caratteristiche è proprio ciò che si desidera da un processo di stima intervallare, un intervallo con buona copertura ma non troppo generico per quanto riguarda il range assunto per il parametro.

Si esamina adesso la lunghezza attesa quando la numerosità campionaria aumenta, per vedere se la gerarchia degli intervalli cambia.

Le lunghezze attese sono pressoché identiche, perciò per determinare la gerarchia tra gli intervalli si rimanda alle probabilità di copertura in 3.3.

**Conclusioni** In ambito di stima intervallare della probabilità di successo in un modello binomiale, tra i metodi approssimati i migliori risultano il Wilson ed il bayesiano. Il secondo risulta più semplice in termini di estremi e leggermente più



**Figura 3.6:** Lunghezza attesa al variare di  $p$  quando  $n = 100$

stabile nella copertura. Anche l'intervallo corretto di Wald ha buone performance, ma risulta sempre più ampio del Wilson, motivo per cui si preferisce quest'ultimo. Ovviamente queste scelte sono impraticabili se si desidera un intervallo che segua rigorosamente la definizione di intervallo di confidenza, ovvero con copertura costantemente maggiore del livello nominale. Se si desidera un metodo di questo tipo, l'intervallo di Blaker è la scelta consigliata.

## Capitolo 4

# Riparametrizzazioni

### 4.1 Quote

Si consideri una riparametrizzazione  $\psi$  di  $p$

$$\psi(p) = \frac{p}{1-p},$$

questa grandezza è chiamata *quota* dell'evento definito come successo. La forma delle quote comunica intrinsecamente il loro significato: esse rappresentano infatti la frazione tra successi ed insuccessi. Si noti come le quote siano direttamente legate al parametro naturale della v.c. binomiale, il quale è  $\theta(p) = \log(p/(1-p)) = \log(\psi(p))$ , ovvero il logaritmo di queste.

A seguito della suddetta riparametrizzazione, si definisce il nuovo modello, la sua verosimiglianza e le altre grandezze direttamente legate ad essa. Il modello è definito da  $\Psi = (\psi : \psi = \psi(p), p \in \{0, 1, \dots, n\})$ , per la verosimiglianza si considera che

$$\psi(p) = \frac{p}{1-p} \implies p(\psi) = \frac{\psi}{1+\psi}.$$

Anche il supporto cambia, difatti  $\psi \in (0, +\infty)$ .

Da (2.3), sostituendo  $p$  con  $\psi(p)$ ,

$$\ell(\psi) = \log \binom{n}{x} + x \log(\psi) + n \log \left( 1 - \frac{\psi}{1+\psi} \right),$$

da cui

$$\ell(\psi) = \log \binom{n}{x} + x \log \psi + n \log \frac{1}{1+\psi}. \quad (4.1)$$

Il calcolo delle derivate si ottiene come segue:

$$u(\psi) = \frac{\partial \ell(\psi)}{\partial \psi} = \frac{x}{\psi} - \frac{n}{1 + \psi}, \quad (4.2)$$

$$\frac{\partial^2 \ell(\psi)}{\partial \psi^2} = -\frac{x}{\psi^2} + \frac{n}{(1 + \psi)^2}.$$

Per il calcolo dell'informazione attesa di Fisher, è possibile applicare la formula appositata per riparametrizzazioni

$$I(\psi(p)) = \left[ \frac{\partial \psi(p)}{\partial p} \right]^{-2} I(p) \Big|_{p=p(\psi)},$$

oppure, utilizzando la derivata seconda di  $\ell(\psi)$ ,

$$I(\psi) = -\mathbb{E} \left[ \frac{\partial^2 \ell(\psi)}{\partial \psi^2} \right] = \mathbb{E} \left[ \frac{x}{\psi^2} - \frac{n}{(1 + \psi)^2} \right].$$

L'unico elemento di natura stocastica è  $x$  per il quale vale

$\mathbb{E}[x] = np = n\psi/(1 + \psi)$ , perciò

$$I(\psi) = \frac{n\psi}{(1 + \psi)\psi^2} - \frac{n}{(1 + \psi)^2} = \frac{n(1 + \psi) - n\psi}{(1 + \psi)^2\psi},$$

concludendo che

$$I(\psi) = \frac{n}{(1 + \psi)^2\psi}. \quad (4.3)$$

A fini applicativi si calcola anche  $i(\psi) = I(\psi)/n = 1/((1 + \psi)^2\psi)$ .

#### 4.1.1 Tre test di verosimiglianza per le quote

I tre test connessi alla verosimiglianza sono esprimibili direttamente per le quote, utilizzando le formulazioni generali del Capitolo 1. Essi saranno calcolati in un contesto di ipotesi nulla  $H_0 : \psi = \psi_0$  con alternativa bilaterale. Lo stimatore MV di  $\psi$  è fornito dal teorema di Zehna, per cui  $\hat{\psi}(p) = \psi(\hat{p}) = \hat{p}/(1 - \hat{p})$ , con  $\hat{p} = x/n$ .

### Log-rapporto di verosimiglianza

Da (1.5),  $W(x) = 2[\ell(\hat{\psi}) - \ell(\psi_0)]$ , per il parametro in questione:

$$W(x) = 2 \left[ \log \binom{n}{x} + x \log \hat{\psi} - n \log(1 + \hat{\psi}) - \log \binom{n}{x} - x \log \psi_0 + n \log(1 + \psi_0) \right],$$

$$W(x) = 2 \left[ x \log \frac{\hat{\psi}}{\psi_0} - n \log \frac{1 + \hat{\psi}}{1 + \psi_0} \right]. \quad (4.4)$$

### Test di Wald

Sfruttando (1.6),  $W_e(x) = n(\hat{\psi} - \psi_0)^2 i(\hat{\psi})$ , grazie a (4.3), si esprime come:

$$W_e(x) = \frac{n(\hat{\psi} - \psi_0)^2}{(1 + \hat{\psi})^2 \hat{\psi}}. \quad (4.5)$$

### Test di Wilson

Si ricorre ancora alla formula generica (1.7)  $W_u(x) = \ell'(\psi_0)^2 / (ni(\psi_0))$ . Sostituendo le quantità definite in (4.2) e (4.3)

$$W_u(x) = \frac{\left( \frac{x}{\psi_0} - \frac{n}{1 + \psi_0} \right)^2}{\frac{n}{(1 + \psi_0)^2 \psi_0}},$$

si sommano i termini al numeratore e si semplifica  $(1 + \psi_0)\psi_0$  giungendo a

$$W_u(x) = \frac{[x(1 + \psi_0) - n\psi_0]^2}{n\psi_0},$$

infine si raccoglie  $n^2$  al numeratore per arrivare alla forma finale

$$W_u(x) = \frac{n \left[ \frac{x}{n}(1 + \psi_0) - \psi_0 \right]^2}{\psi_0}. \quad (4.6)$$

## 4.2 Intervalli di confidenza per le quote

Dopo aver definito i test collegati alla verosimiglianza in un modello binomiale con riparametrizzazione, si passa alla costruzione degli intervalli di confidenza derivanti proprio da questi.

Prima di passare ad essi, si derivano gli intervalli esatti visti nel Capitolo 2. Per questo problema di stima intervallare si ipotizza che  $X \sim \text{Bin}(n, p)$ .

### 4.2.1 Clopper-Pearson

Il procedimento che portava alla creazione dell'intervallo di confidenza per il parametro  $p$  rimane totalmente invariato per  $\psi$ .

La monotonia della funzione applicata al parametro ci permette di sostituire  $p_0 = \psi_0/(1 + \psi_0)$  all'interno delle due equazioni (2.11), (2.12):

$$\mathbb{P}(X \leq r_1 \mid \psi = \psi_0) = \sum_{k=0}^{r_1} \binom{n}{k} \frac{\psi_0^k}{1 + \psi_0} \left(1 - \frac{\psi_0}{1 + \psi_0}\right)^{n-k} = \alpha/2$$

$$\mathbb{P}(X \geq r_2 \mid \psi = \psi_0) = \sum_{k=r_2}^n \binom{n}{k} \frac{\psi_0^k}{1 + \psi_0} \left(1 - \frac{\psi_0}{1 + \psi_0}\right)^{n-k} = \alpha/2.$$

Per i risultati numerici riguardanti le quote, gli estremi si possono ricavare sostituendo quelli trovati per  $p_0$  in  $\psi_0 = p_0/(1 - p_0)$ , sfruttando quindi la biunivocità della funzione  $\psi(p)$ .

Le soluzioni per  $p$  sono  $p_1 = q_{\alpha/2, x, n-x+1}^B$  e  $p_2 = q_{1-\alpha/2, x+1, n-x}^B$ , allora

$$\psi_1 = \frac{q_{\alpha/2, x, n-x+1}^B}{1 - q_{\alpha/2, x, n-x+1}^B},$$

$$\psi_2 = \frac{q_{1-\alpha/2, x+1, n-x}^B}{1 - q_{1-\alpha/2, x+1, n-x}^B}.$$

### 4.2.2 Blaker

L'intervallo esatto di Blaker non subisce cambiamenti dovuti alla riparametrizzazione, è facile mostrare che il procedimento mostrato per  $p$  rimane invariato. Si noti che l'ordine dei  $\gamma(\psi, x)$  non cambia grazie alla monotonia della trasformazione, di conseguenza anche le  $x_i : \gamma(\psi, x_i) \leq \gamma(\psi, x)$  rimangono le stesse. Da ciò ne consegue che  $\alpha(\psi, x) = \alpha(p, x)$  essendo esso calcolato come la somma delle  $\mathbb{P}(X = x_i \mid \psi) = \mathbb{P}(X = x_i \mid p(\psi))$ . L'unica differenza consiste chiaramente negli estremi, che saranno uguali agli estremi per  $p$  a cui sarà applicata la funzione  $\psi(\cdot)$ .

### 4.2.3 Log-rapporto di verosimiglianza

La costruzione dell'intervallo per il test TRV è di fatto uguale a quella vista per la stima intervallare di  $p$ . L'intervallo in sé è equivalente, ma espresso in termini di  $\psi$ :

$$C(x) = \{p_0 : W(x) \leq \lambda_\alpha\} = \{\psi_0 : W(x) - \lambda_\alpha \leq 0\}$$

la regione è ancora un intervallo per le stesse motivazioni presentate in precedenza. Si può mostrare ciò in relazione all'esempio visto in 2.3.3. A livello pratico, è ancora richiesto il passaggio tramite via numerica. La motivazione dell'equivalenza tra gli intervalli si ritrova nel fatto che  $\psi(p)$  è una funzione monotona, di conseguenza l'ordine delle  $W(x)$  al variare di  $x$  rimarrà tale, come anche le probabilità associate a ciascuno dei valori che può assumere. Con la formula in (4.4) è possibile calcolare i  $W(x)$  direttamente con le quote, ma i valori non cambieranno. Queste due affermazioni implicano che la tabella in 2.3.3 e quindi la soglia  $\lambda_\alpha$  rimarranno inalterate, motivo per cui i due intervalli sono equivalenti.

L'unico cambiamento si riscontra negli estremi, che a livello teorico dovrebbero essere estratti risolvendo l'equazione  $W(x) - \lambda_\alpha = 0$  in  $\psi$  invece che in  $p$ . A livello pratico la differenza è inesistente, in quanto basterà sostituire le soluzioni trovate per  $p$  con  $\psi(p) = p/(1-p)$ , sfruttando la monotonia della parametrizzazione applicata.

### 4.2.4 Wald

Dalla quantità pivotale approssimata in (4.5)

$$W_e(x) = \frac{n(\hat{\psi} - \psi)^2}{(1 + \hat{\psi})^2 \hat{\psi}} \xrightarrow{d} Z^2 \sim \chi_1^2,$$

ne consegue che

$$\frac{\sqrt{n}(\hat{\psi} - \psi)}{(1 + \hat{\psi})\sqrt{\hat{\psi}}} \xrightarrow{d} Z \sim N(0, 1).$$

Perciò, al divergere di  $n$ ,

$$\mathbb{P} \left( z_{\alpha/2} < \frac{\sqrt{n}(\hat{\psi} - \psi)}{(1 + \hat{\psi})\sqrt{\hat{\psi}}} < z_{1-\alpha/2} \right) \rightarrow \alpha,$$

isolando  $\psi$  si ricava l'intervallo

$$z_{\alpha/2}(1 + \hat{\psi})\sqrt{\hat{\psi}} < \sqrt{n}(\hat{\psi} - \psi) < z_{1-\alpha/2}(1 + \hat{\psi})\sqrt{\hat{\psi}}$$

con estremi

$$\hat{\psi} \pm z_{\alpha/2}(1 + \hat{\psi})\sqrt{\frac{\hat{\psi}}{n}}.$$

### Wald corretto

Il calcolo dell'intervallo di confidenza di Wald corretto si effettua semplicemente sostituendo  $p$  con  $\tilde{p} = (X + 2)/(n + 4)$ . Per la riparametrizzazione in quote si sostituisce  $\tilde{\psi} = \tilde{p}/(1 - \tilde{p})$ , a  $\hat{\psi}$  nell'intervallo di Wald. I risultanti estremi saranno

$$\tilde{\psi} \pm z_{\alpha/2}(1 + \tilde{\psi})\sqrt{\frac{\tilde{\psi}}{n}}.$$

### 4.2.5 Wilson

Si ricorre a (4.6) per la q.p. approssimata

$$W_u(x) = \frac{n \left[ \frac{x}{n}(1 + \psi) - \psi \right]^2}{\psi} \xrightarrow{d} Z^2 \sim \chi_1^2,$$

ragion per cui si può scrivere

$$W_u(x) = \frac{\frac{x}{n}(1 + \psi) - \psi}{\sqrt{\frac{\psi}{n}}} \xrightarrow{d} Z \sim N(0, 1).$$

All'intervallo di confidenza per  $\psi$  si giunge isolando  $\psi$  in

$$z_{\alpha/2} < \frac{\frac{x}{n}(1 + \psi) - \psi}{\sqrt{\frac{\psi}{n}}} < z_{1-\alpha/2}.$$

Si cercano gli estremi da

$$\frac{\frac{x}{n}(1 + \psi) - \psi}{\sqrt{\frac{\psi}{n}}} = \pm z_{\alpha/2},$$

si moltiplica per  $\sqrt{\psi/n}$  e si eleva al quadrato

$$\left[\frac{x}{n}(1 + \psi) - \psi\right]^2 = z_{\alpha/2}^2 \frac{\psi}{n}.$$

Si riscrive il termine sinistro

$$\left[\frac{x}{n} + \psi \left(\frac{x}{n} - 1\right)\right]^2 = z_{\alpha/2}^2 \frac{\psi}{n},$$

svolgendo il quadrato

$$\psi^2 \left(\frac{x}{n} - 1\right)^2 + 2\frac{x}{n}\psi \left(\frac{x}{n} - 1\right) + \left(\frac{x}{n}\right)^2 = z_{\alpha/2}^2 \frac{\psi}{n},$$

raccogliendo i termini dello stesso grado in  $\psi$

$$\psi^2 \left(\frac{x}{n} - 1\right)^2 + 2\psi \left[\frac{x}{n} \left(\frac{x}{n} - 1\right) - \frac{z_{\alpha/2}^2}{2n}\right] + \left(\frac{x}{n}\right)^2 = 0.$$

Adesso è possibile applicare la formula per equazioni di secondo grado quando  $X \neq n$  altrimenti il polinomio risulta di primo grado. Il risultato di questa applicazione è:

$$\psi = \frac{-\left[\frac{x}{n} \left(\frac{x}{n} - 1\right) - \frac{z_{\alpha/2}^2}{2n}\right] \pm \sqrt{\left[\frac{x}{n} \left(\frac{x}{n} - 1\right) - \frac{z_{\alpha/2}^2}{2n}\right]^2 - \left(\frac{x}{n}\right)^2 \left(\frac{x}{n} - 1\right)^2}}{\left(\frac{x}{n} - 1\right)^2},$$

all'interno della radice, il primo termine al quadrato si annulla con l'addendo esterno:

$$\psi = \frac{-\frac{x^2}{n^2} + \frac{x}{n} + \frac{z_{\alpha/2}^2}{2n} \pm \sqrt{\frac{z_{\alpha/2}^4}{4n^2} - \frac{z_{\alpha/2}^2}{n} \frac{x}{n} \left(\frac{x}{n} - 1\right)}}{\left(\frac{x}{n} - 1\right)^2},$$

si raccoglie  $z_{\alpha/2}^2$  all'interno della radice

$$\psi = \frac{-\frac{x^2}{n^2} + \frac{x}{n} + \frac{z_{\alpha/2}^2}{2n} \pm z_{1-\alpha/2} \sqrt{\frac{z_{\alpha/2}^2}{4n^2} - \frac{x}{n^2} \left(\frac{x}{n} - 1\right)}}{\left(\frac{x-n}{n}\right)^2}.$$

Semplificando  $n$  si conclude che gli estremi saranno

$$\psi = \frac{-\frac{x^2}{n} + x + \frac{z_{\alpha/2}^2}{2} \pm z_{1-\alpha/2} \sqrt{\frac{z_{\alpha/2}^2}{4} - x \left(\frac{x}{n} - 1\right)}}{(x - n)^2}.$$

Il caso  $X = n$  andrà analizzato a parte, riprendendo (4.6) la statistica test risulta

$$W_u(x) = \frac{n \left[ \frac{n}{n}(1 + \psi) - \psi \right]^2}{\psi} = \frac{n}{\psi},$$

da cui si passa a

$$z_{\alpha/2} < \sqrt{\frac{n}{\psi}} < z_{1-\alpha/2},$$

riscritto come

$$\begin{cases} \sqrt{\frac{n}{\psi}} > z_{\alpha/2}, \\ \sqrt{\frac{n}{\psi}} < z_{1-\alpha/2}. \end{cases}$$

La prima disequazione risulta sempre valida sapendo che  $\psi \in [0, +\infty)$  e  $z_{\alpha/2} < 0$  se  $\alpha \in (0, 1)$ ; la seconda diventa

$$\frac{n}{\psi} < z_{1-\alpha/2}^2,$$

isolando  $\psi$

$$\psi > \frac{n}{z_{1-\alpha/2}^2}.$$

L'IC nel caso di  $X = n$  sarà quindi  $(n/z_{1-\alpha/2}^2, +\infty)$ .

#### 4.2.6 Bayesiano

Nel problema di stima intervallare del parametro  $p$  in un contesto bayesiano si era fatto ricorso ad una priori di Jeffreys che aveva dato luogo ad una posteriori:

$$\pi(p | x) \propto \text{Beta} \left( x + \frac{1}{2}, n - x + \frac{1}{2} \right).$$

Per quanto riguarda le quote, essendo esse pari a  $\psi = \frac{p}{1-p}$ , la loro distribuzione a posteriori sarà

$$\pi(\psi | x) = \text{Beta}' \left( x + \frac{1}{2}, n - x + \frac{1}{2} \right),$$

ovvero una distribuzione *Beta prime*. Sfruttando i quantili di questa distribuzione si giungerà all'intervallo di credibilità

$$\left( q_{\alpha/2, x+1/2, n-x+1/2}^{B'}, q_{1-\alpha/2, x+1/2, n-x+1/2}^{B'} \right).$$

#### 4.2.7 Applicazione ai dati

Si mostrano adesso gli estremi ottenuti dai metodi di stima intervallare presentati, utilizzando i dati in (2.9) e (2.10).

Metodo $\alpha = 95\%$	n=5 $\hat{\psi} = 4$		n=100 $\hat{\psi} = 1.44$	
	Inf	Sup	Inf	Sup
Clopper-Pearson	0.396	197	0.95	2.199
Blaker	0.521	97.02	0.960	2.192
TRV	0.602	76.90	0.974	2.154
Wald	-4.77	12.77	0.866	2.012
Wald corretto	-1.72	5.719	0.854	1.983
Wilson	0.601	26.61	0.969	2.138
Bayesiano	0.591	43.42	0.969	2.152

**Tabella 4.1:** Estremi dei vari intervalli di confidenza per le quote al 95% per i dati (2.9) e (2.10).

Nell'analisi degli estremi si nota immediatamente come i due intervalli di Wald sfiorino il supporto, raggiungendo valori negativi per l'estremo inferiore. Essi sono anche i due con estremi superiori più bassi, dimostrandosi forse troppo liberali nella stima. Nel caso di  $n = 5$  risultano eccessivamente più stretti degli altri, considerando che tutti i metodi si mantengono molto generici data lo scarso numero di informazioni disponibili. Esaminando il secondo insieme di dati, sembra che gli approcci derivanti dal test di Wald siano proprio traslati rispetto agli altri. Per quanto riguarda il Wald corretto è intuibile che l'aggiunta di due successi e due insuccessi sposti la stima verso  $p = 0.5$  e quindi  $\psi = 1$ , ma anche senza correzione si nota questa tendenza. Il Clopper-Pearson è esageratamente ampio nel primo insieme di dati, con un estremo superiore di gran lunga più elevato anche di quello dell'altro intervallo esatto. Il Wilson e il bayesiano sono i due metodi maggiormente simili, ma comunque rimangono piuttosto distanti per campioni esigui. Nella stima per il campione con numerosità campionaria pari a 100 non vi sono differenze notevoli tra gli approcci.

### 4.2.8 Valutazione degli intervalli

In questa sezione saranno valutati i vari metodi di stima intervallare per le quote. L'analisi si svolgerà sotto due punti di vista: la probabilità di copertura e l'ampiezza dell'intervallo.

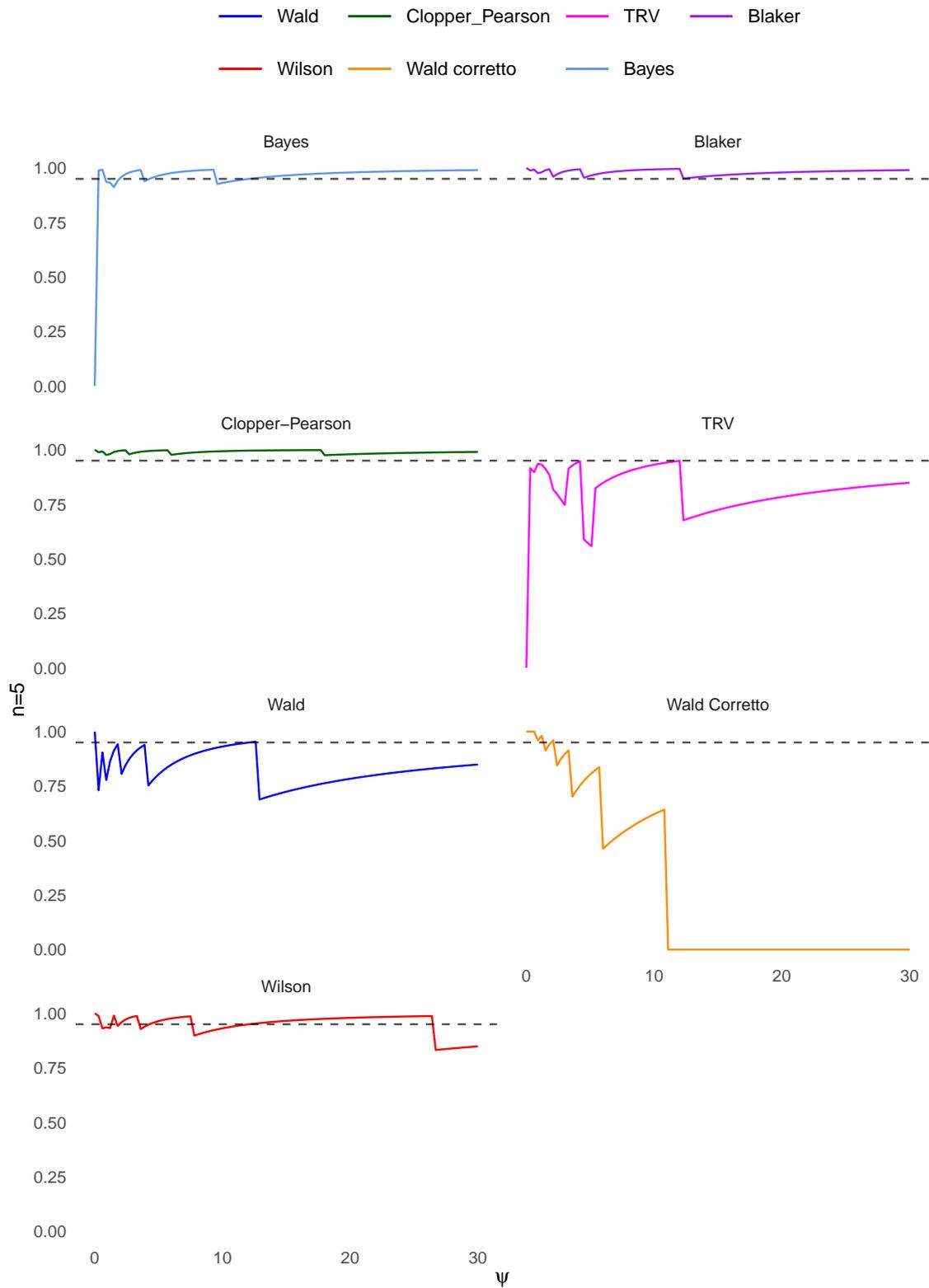
#### Probabilità di copertura

Il primo indice considerato è la probabilità di copertura, valutata per  $n = 5$  e  $n = 100$ . Dal momento che le quote sono comprese in  $[0, +\infty)$  non è possibile rappresentare il valore della probabilità per ogni quota, si opta quindi per un intervallo di  $[0, 30]$ .

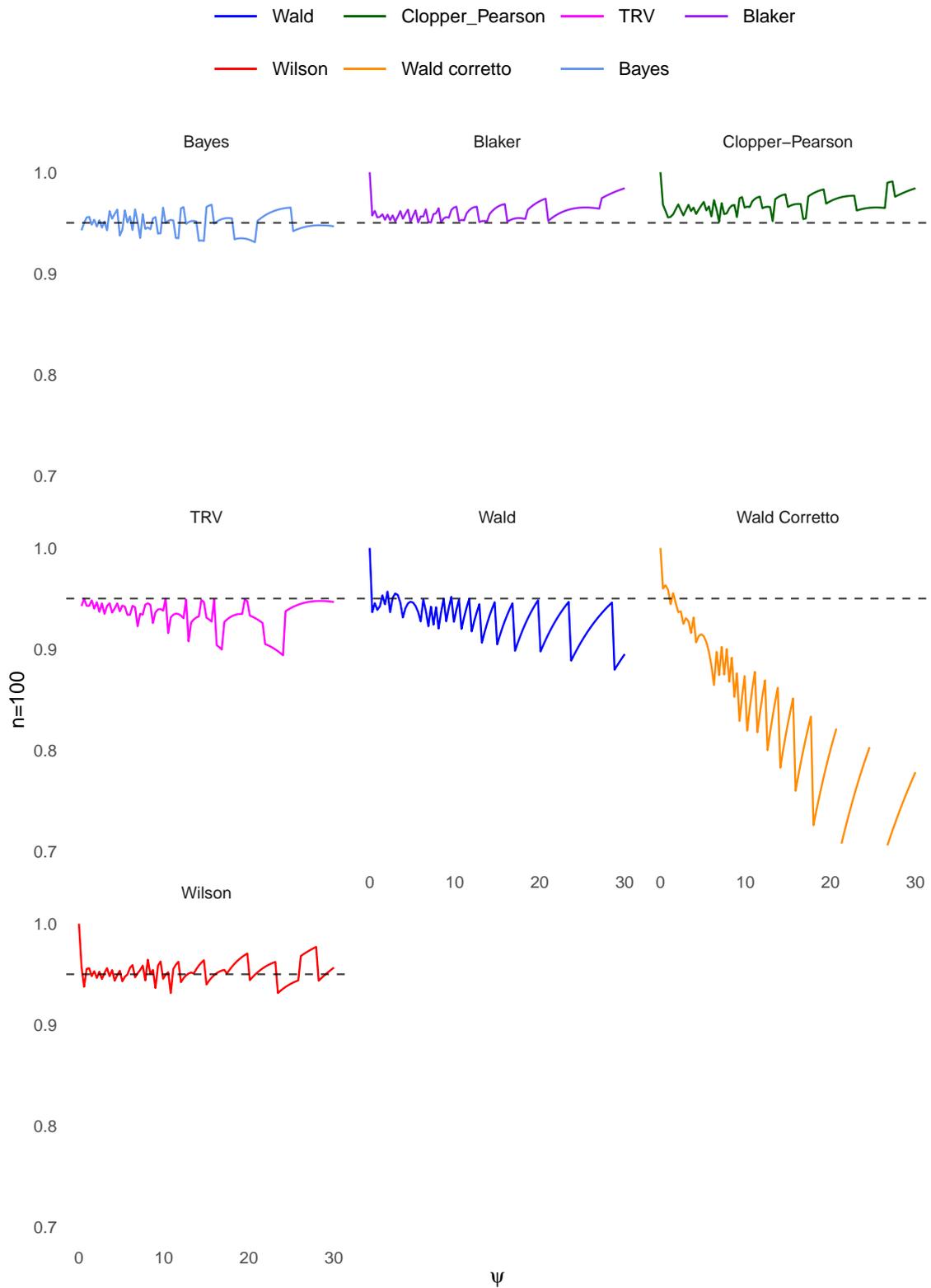
È opportuno notare che alcuni metodi risultano invarianti alle riparametrizzazioni dal punto di vista della copertura. Gli unici intervalli che non possiedono questa proprietà sono il Wald e la sua versione corretta. Tuttavia il parametro in considerazione ha un supporto differente rispetto a  $p$ , perciò i grafici con la probabilità di copertura risulteranno diversi.

I metodi di stima intervallare in 4.1 più stabili dal punto di vista della copertura sono quello bayesiano, il Clopper-Pearson e il Blaker. Sia il bayesiano sia il TRV hanno copertura nulla in corrispondenza di  $\psi = 0$ . Questo grafico evidenzia il comportamento estremamente inadatto dell'intervallo di Wald corretto. La correzione fatta tramite l'aggiunta di due successi e due insuccessi trasla la stima verso il valore di  $p = 0.5$  e quindi  $\psi = 1$ . A  $n$  esigui, questa traslazione è troppo influente e sposta l'intervallo lontano dal vero valore del parametro se questo è elevato. Si spiega quindi perché la copertura è nulla a quote alte. L'intervallo basato sul TRV ha un andamento irregolare, con coperture inferiori al livello nominale, anche Wald si trova in una situazione simile. L'intervallo di Wilson persiste nei pressi del valore 0.95, ma è caratterizzato da fluttuazioni leggere. Adesso si esamina il caso in cui  $n = 100$ .

Il comportamento dei metodi approssimati in 4.1, come ci si può aspettare, si affina all'aumentare di  $n$ . L'intervallo di Wald si riconferma però un'approssimazione scadente, con copertura costantemente inferiore al livello nominale superato un certo valore di  $\psi$ . La differenza tra i due intervalli esatti non è particolarmente marcata, la probabilità di copertura cresce per entrambi a quote alte superando di un margine troppo ampio il livello nominale. Nonostante



**Figura 4.1:** Probabilità di copertura quando  $n = 5$  degli intervalli di confidenza di livello 95% per le quote



**Figura 4.2:** Probabilità di copertura quando  $n = 100$  degli intervalli di confidenza di livello 95% per le quote

la numerosità campionaria decisamente elevata, l'intervallo corretto di Wald incorre negli stessi problemi di prima, perdendo affidabilità quando le quote sono elevate. L'intervallo bayesiano si mantiene sempre molto vicino a 0.95 ma è doveroso puntualizzare che, come per  $n = 5$ , la sua copertura si annulla per  $\psi = 0$ , comportamento condiviso dal metodo basato sul TRV.

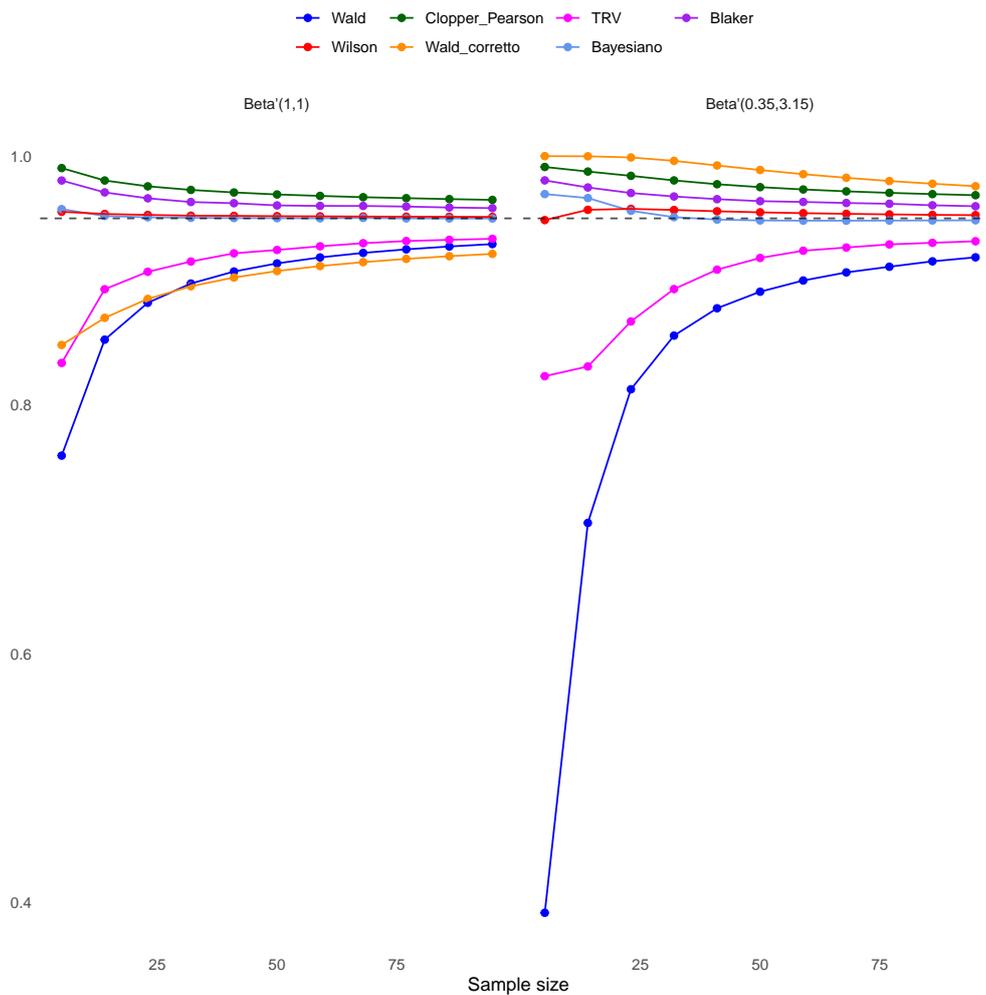
### Probabilità media di copertura

Si valuta la copertura degli intervalli con la quantità  $\bar{C}_n$  vista in (3.1). Per le distribuzioni a priori si considerano quelle usate per  $p$  e si deducono le conseguenti distribuzioni di  $\psi$ . Nel Capitolo 3 erano state considerate una distribuzione  $g(p) = \text{Uniforme}(0, 1) = \text{Beta}(1, 1)$ , di conseguenza come prima priori si impiegherà una  $h(\psi) = \text{Beta}'(1, 1)$ . La seconda:  $g'(p) = \text{Beta}(0.35, 3.15)$  che risulta in una  $h'(\psi) = \text{Beta}'(0.35, 3.15)$ .

Entrambe le priori mostrano come l'intervallo di Wald e quello basato sul TRV siano ancora quelli con performance peggiori. Per il secondo non è sorprendente, in quanto invariante a riparametrizzazioni, mentre per il Wald è una riconferma della sua inadeguatezza. È molto interessante il comportamento dell'intervallo di Wald corretto, esso sembra eccessivamente conservativo quando  $h(\psi) = \text{Beta}'(0.35, 3.15)$  e possiede una copertura totalmente inadatta con l'altra priori. Ciò accade perché la seconda priori ha funzione di densità concentrata in valori "bassi" di  $\psi$  ma, come si è visto nelle analisi delle probabilità di copertura, questo intervallo subisce gravi cali di performance per quote elevate. La  $\text{Beta}'(1, 1)$  assegna densità uguali per tutti i valori di  $\psi$ , risultando in coperture insufficienti da parte dell'intervallo di Wald corretto, che tuttavia rimane superiore alla sua versione originale con campioni di dimensioni contenute.

### Ampiezza attesa

È interessante notare che, per  $X = n$ , alcuni intervalli di confidenza "degenerano" in quanto il loro estremo superiore diventa  $+\infty$  dal momento che anche  $\hat{\psi} = +\infty$ . Tecnicamente questo comportamento è corretto in quanto, come gli intervalli per la probabilità di successo includevano il valore  $p = 1$  quando  $X = n$ , così le stime intervallari per le quote includono  $\psi(1) = 1/(1 - 1) = +\infty$ . Vi sono due metodi che non possiedono questa caratteristica, essi sono il bayesiano e il Wald corretto. Il secondo non degenera in quanto utilizza  $\tilde{\psi}$  che non è mai pari a  $+\infty$ .

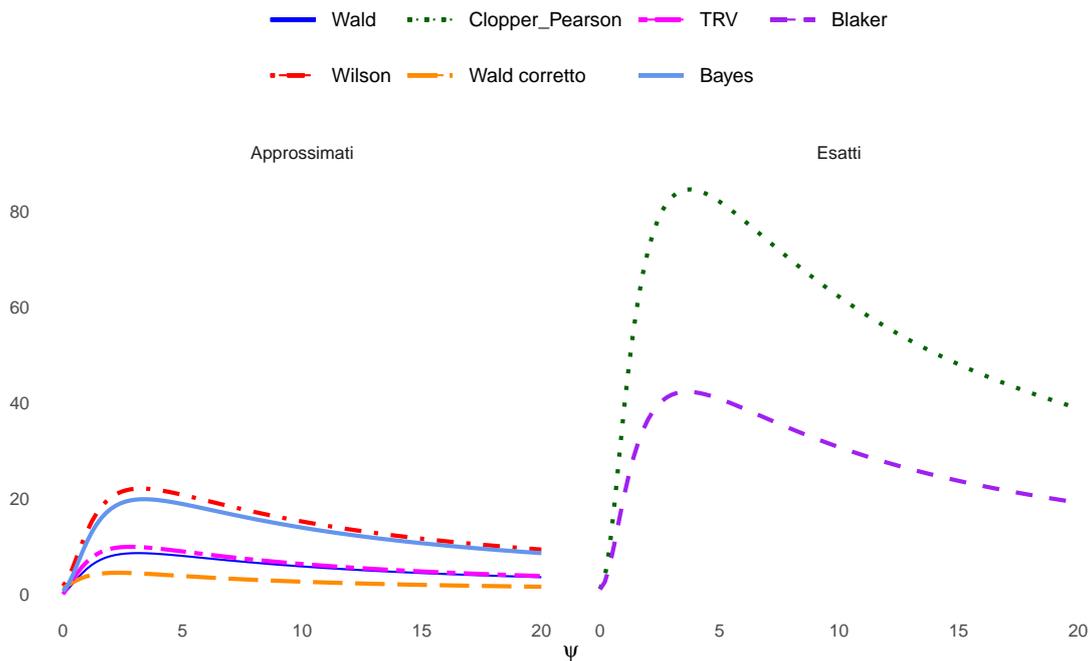


**Figura 4.3:** Probabilità media di copertura in funzione della numerosità campionaria per gli intervalli di confidenza di livello 95% per le quote quando  $\psi$  ha un distribuzione Beta' (1, 1) e Beta' (0.35, 3.15)

Il primo semplicemente assume sempre valori finiti, poiché utilizza quantili di una Beta' con parametri sempre diversi da 0. Nel calcolo dell'ampiezza attesa per gli intervalli, il valore atteso presente nella formula implica che verrà anche considerato l'IC di lunghezza infinita portando questa grandezza a degenerare. Per ovviare a questo problema il calcolo della lunghezza attesa per questo tipo di intervalli farà uso di un'espressione leggermente modificata:

$$\sum_{k=0}^{n-1} \binom{n}{k} p^k (1-p)^{n-k} [t_1(k) - t_2(k)],$$

dove  $t_1(k)$ ,  $t_2(k)$  sono gli estremi di un IC quando  $X = k$ .



**Figura 4.4:** Lunghezza attesa al variare di  $\psi$  quando  $n = 5$

La lunghezza attesa dell'intervallo di Wald corretto è da prendere con cautela, in quanto la copertura è del tutto insufficiente. Anche il Wald e il TRV sono molto corti ma la loro copertura non è molto alta. Il bayesiano e il Wilson sembrano quelli che si comportano meglio, rappresentando un buon compromesso tra la lunghezza degli intervalli esatti con coperture elevate e la scarsità di copertura degli altri intervalli.

La 4.5 mostra una sorprendente caratteristica dell'IC di Wilson, ovvero il fatto che la sua ampiezza risulta molto maggiore di tutti gli altri i metodi a  $n = 100$ , per valori elevati di  $\psi$ .

**Conclusioni** La riparametrizzazione in quote conferma essenzialmente ciò che è stato detto nel capitolo precedente. La grossa differenza sta nell'intervallo corretto di Wald, assolutamente inutilizzabile per valori alti di  $\psi$ . Il bayesiano e il Wilson si riconfermano ottimi metodi, con coperture accettabili e lunghezze ideali. Tuttavia il secondo risulta stranamente impreciso per valori alti di  $\psi$  a numerosità campionarie alte.

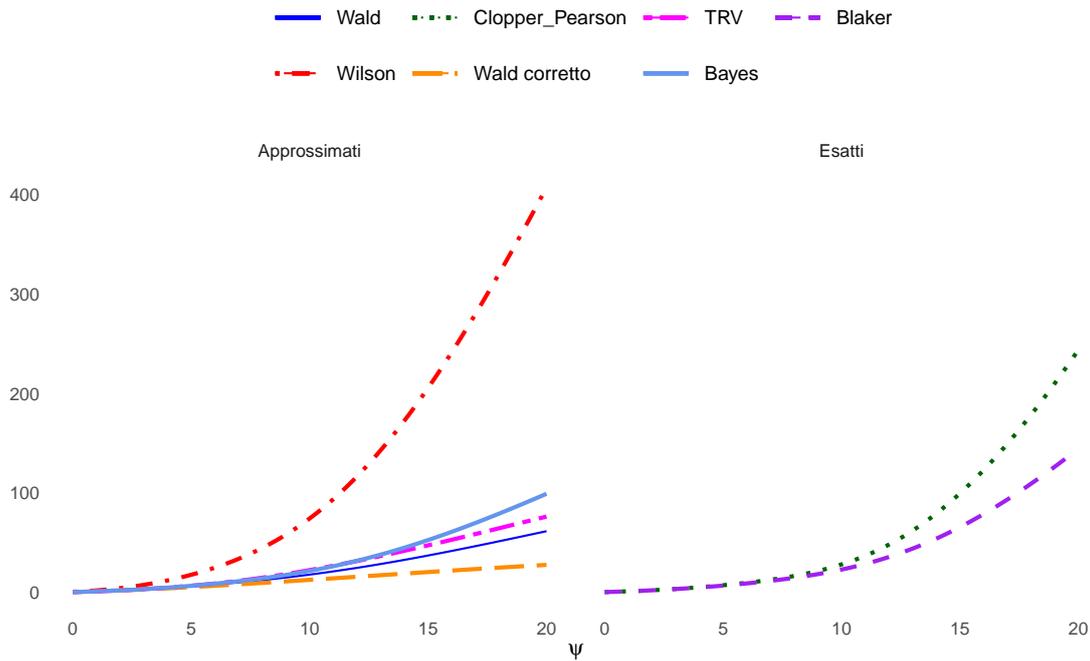


Figura 4.5: Lunghezza attesa al variare di  $\psi$  quando  $n = 100$

### 4.3 Log-quote

Si consideri adesso la parametrizzazione  $\theta$  con

$$\theta(p) = \log(\psi) = \log\left(\frac{p}{1-p}\right),$$

questa è la *parametrizzazione naturale* di una v.c. binomiale. La quantità  $\theta$  è detta *log-quota* poiché appunto è il logaritmo della quota, essa ha supporto  $(-\infty, +\infty)$ . Il modello così definito è  $\Theta = (\theta : \theta = \theta(p), p \in \{0, 1, \dots, n\})$  con verosimiglianza:

$$\ell(\theta) = \log\binom{n}{x} + x\theta - n \log(1 + e^\theta), \quad (4.7)$$

ottenuta da (2.3), sapendo che  $p(\theta) = e^\theta / (1 + e^\theta)$ .

Le funzioni legate ad essa sono:

$$u(\theta) = \frac{\partial \ell(\theta)}{\partial \theta} = x - n \frac{e^\theta}{1 + e^\theta}, \quad (4.8)$$

$$\frac{\partial^2 \ell(\theta)}{\partial \theta^2} = -n \frac{e^\theta}{(1 + e^\theta)^2},$$

di conseguenza

$$I(\theta) = -\mathbb{E} \left[ \frac{\partial^2 \ell(\theta)}{\partial \theta^2} \right] = \mathbb{E} \left[ \frac{ne^\theta}{(1+e^\theta)^2} \right] = \frac{ne^\theta}{(1+e^\theta)^2}, \quad (4.9)$$

ovviamente  $i(\theta) = I(\theta)/n = e^\theta/(1+e^\theta)$ .

### 4.3.1 Tre test di verosimiglianza per le log-quote

Si costruiscono i tre test connessi alla verosimiglianza nel sistema d'ipotesi con  $H_0 : \theta = \theta_0$  con alternativa bilaterale, la stima SMV sarà  $\hat{\theta} = \theta(\hat{p}) = \log[\hat{p}/(1 - \hat{p})]$ .

#### Log-rapporto di verosimiglianza

La derivazione di questo test è immediata, partendo da (1.5):

$$W(x) = 2 \left[ \log \binom{n}{x} + x\hat{\theta} - n \log(1 + e^{\hat{\theta}}) - \log \binom{n}{x} - x\theta_0 + n \log(1 + e^{\theta_0}) \right],$$

concludendo che

$$W(x) = 2 \left[ x(\hat{\theta} - \theta_0) + n \log \frac{1 + e^{\theta_0}}{1 + e^{\hat{\theta}}} \right]. \quad (4.10)$$

#### Test di Wald

Da (1.6)  $W_e(x) = (\hat{\theta} - \theta_0)^2 I(\theta)$ , quindi con (4.9):

$$W_e(x) = \frac{ne^{\hat{\theta}}(\hat{\theta} - \theta_0)^2}{(1 + e^{\hat{\theta}})^2}. \quad (4.11)$$

#### Test di Wilson

La statistica test ha espressione  $W_u(x) = \ell'(\theta_0)^2 / (I(\theta_0))$ , con queste quantità definite in (4.8) e (4.9):

$$W_u(x) = \frac{\left[ x - n \frac{e^{\theta_0}}{1+e^{\theta_0}} \right]^2}{n \frac{e^{\theta_0}}{(1+e^{\theta_0})^2}} \quad (4.12)$$

## 4.4 Intervalli di confidenza per le log-quote

Si deriva l'equivalente per le log-quote degli intervalli di confidenza con tutti i metodi utilizzati.

### 4.4.1 Clopper-Pearson

Come per le quote, anche la costruzione di questo IC per le log-quote non cambia, sempre per la monotonia della trasformazione applicata. È possibile riscrivere le equazioni per gli estremi:

$$\mathbb{P}(X \leq r_1 \mid \theta = \theta_0) = \sum_{k=0}^{r_1} \binom{n}{k} \frac{e_0^{\theta_0}}{1 + e_0^{\theta_0}}^k \left(1 - \frac{e_0^{\theta_0}}{1 + e_0^{\theta_0}}\right)^{n-k} = \alpha/2,$$

$$\mathbb{P}(X \geq r_2 \mid \theta = \theta_0) = \sum_{k=r_2}^n \binom{n}{k} \frac{e_0^{\theta_0}}{1 + e_0^{\theta_0}}^k \left(1 - \frac{e_0^{\theta_0}}{1 + e_0^{\theta_0}}\right)^{n-k} = \alpha/2.$$

Gli estremi si ricavano dalle soluzioni in  $p$ , giungendo a

$$\theta_1 = \log \frac{q_{\alpha/2, x, n-x+1}^B}{1 - q_{\alpha/2, x, n-x+1}^B},$$

$$\theta_2 = \log \frac{q_{1-\alpha/2, x+1, n-x}^B}{1 - q_{1-\alpha/2, x+1, n-x}^B}.$$

### 4.4.2 Blaker

Il discorso fatto per le quote vale anche per le log-quote, essendo anch'essa una riparametrizzazione tramite funzione monotona. Gli estremi saranno perciò gli stessi di quelli per  $p$  ma a cui sarà applicata la funzione  $\theta(\cdot)$ .

### 4.4.3 Log-rapporto di verosimiglianza

Anche in questo caso la riparametrizzazione implica soltanto l'applicazione della funzione  $\theta(\cdot)$  agli estremi in  $p$  per trovare l'IC. La motivazione di ciò si ritrova in [4.2.3](#), dove essa viene data nel caso delle quote.

#### 4.4.4 Wald

Essendo la distribuzione asintotica di (4.11) una  $\chi_1^2$ ,

$$\frac{(\hat{\theta} - \theta_0)\sqrt{e^{\hat{\theta}}n}}{1 + e^{\hat{\theta}}} \xrightarrow{d} Z \sim N(0, 1).$$

Per giungere all'intervallo di confidenza si svolgono i seguenti passaggi:

$$z_{\alpha/2} \leq \frac{(\hat{\theta} - \theta_0)\sqrt{e^{\hat{\theta}}n}}{1 + e^{\hat{\theta}}} \leq z_{1-\alpha/2}$$

$$z_{\alpha/2} \frac{1 + e^{\hat{\theta}}}{\sqrt{ne^{\hat{\theta}}}} \leq \hat{\theta} - \theta_0 \leq z_{1-\alpha/2} \frac{1 + e^{\hat{\theta}}}{\sqrt{ne^{\hat{\theta}}}}.$$

L'intervallo risultante avrà estremi:

$$\hat{\theta} \pm z_{\alpha/2} \frac{1 + e^{\hat{\theta}}}{\sqrt{ne^{\hat{\theta}}}}.$$

#### Wald corretto

Gli estremi dell'intervallo corretto di Wald saranno

$$\tilde{\theta} \pm z_{\alpha/2} \frac{1 + e^{\tilde{\theta}}}{\sqrt{ne^{\tilde{\theta}}}},$$

con  $\tilde{\theta} = \theta(\tilde{p})$  dove  $\tilde{p} = (X + 2)/(n + 4)$ .

#### 4.4.5 Wilson

La statistica test di Wilson (4.12) ha sempre una distribuzione asintotica  $\chi_1^2$ , da cui:

$$\frac{x - n \frac{e^{\theta_0}}{1 + e^{\theta_0}}}{\frac{\sqrt{ne^{\theta_0}}}{1 + e^{\theta_0}}} = \frac{x(1 + e^{\theta_0}) - ne^{\theta_0}}{\sqrt{ne^{\theta_0}}} \xrightarrow{d} Z \sim N(0, 1),$$

quindi per il calcolo degli estremi dell'intervallo si parte da

$$\frac{x(1 + e^{\theta_0}) - ne^{\theta_0}}{\sqrt{ne^{\theta_0}}} = \pm z_{\alpha/2}.$$

Si moltiplica per il denominatore e si eleva al quadrato, risultando in

$$[x(1 + e^{\theta_0}) - ne^{\theta_0}]^2 = z_{\alpha/2}^2 ne^{\theta_0},$$

svolgendo il quadrato nel termine sinistro

$$x^2(1 + e^{\theta_0})^2 + n^2 e^{2\theta_0} - 2nx e^{\theta_0}(1 + e^{\theta_0}) = z_{\alpha/2}^2 n e^{\theta_0},$$

adesso si raccolgono i termini allo stesso grado di  $e^{\theta_0}$

$$e^{2\theta_0}(x^2 + n^2 - 2nx) + 2e^{\theta_0} \left( x^2 - nx - n \frac{z_{\alpha/2}^2}{2} \right) + x^2 = 0.$$

Ora è possibile applicare la formula per equazioni di secondo grado quando il coefficiente del termine  $e^{2\theta_0}$  non si annulla, ovvero quando  $X \neq n$ . Si otterrà:

$$e^{\theta_0} = \frac{- \left( x^2 - nx - n \frac{z_{\alpha/2}^2}{2} \right) \pm \sqrt{\left( x^2 - nx - n \frac{z_{\alpha/2}^2}{2} \right)^2 - x^2(x^2 + n^2 - 2nx)}}{x^2 + n^2 - 2nx},$$

si prende in esame la radice quadrata

$$\sqrt{x^4 + n^2 x^2 + n^2 \frac{z_{\alpha/2}^4}{4} - 2nx^3 - 2x^2 n \frac{z_{\alpha/2}^2}{2} + 2n^2 x \frac{z_{\alpha/2}^2}{2} - x^4 - n^2 x^2 + 2nx^3}$$

alcuni termini si annullano

$$\sqrt{n^2 \frac{z_{\alpha/2}^4}{4} - x^2 n z_{\alpha/2}^2 + n^2 x z_{\alpha/2}^2} = z_{1-\alpha/2} \sqrt{n^2 \frac{z_{\alpha/2}^2}{4} - nx^2 + n^2 x}.$$

Ritornando all'espressione completa

$$e^{\theta_0} = \frac{-x^2 + nx + n \frac{z_{\alpha/2}^2}{2} \pm z_{1-\alpha/2} \sqrt{n^2 \frac{z_{\alpha/2}^2}{4} - nx^2 + n^2 x}}{x^2 + n^2 - 2nx}$$

si raccoglie  $n$  in entrambi i membri della frazione, giungendo agli estremi:

$$\theta = \frac{-\frac{x^2}{n} + x + \frac{z_{\alpha/2}^2}{2} \pm z_{1-\alpha/2} \sqrt{\frac{z_{\alpha/2}^2}{4} - \frac{x^2}{n} + x}}{\frac{x^2}{n} + n - 2x}.$$

Come per le quote il caso  $X = n$  si esamina a parte, in questo caso la radice della statistica test 4.12 diventerà

$$\frac{n - n \frac{e^{\theta_0}}{1+e^{\theta_0}}}{\frac{\sqrt{ne^{\theta_0}}}{1+e^{\theta_0}}} = \frac{n(1 + e^{\theta_0}) - ne^{\theta_0}}{\sqrt{ne^{\theta_0}}} = \frac{n}{\sqrt{ne^{\theta_0}}} = \sqrt{\frac{n}{e^{\theta_0}}}.$$

Si considera l'intervallo

$$z_{\alpha/2} < \sqrt{\frac{n}{e^\theta}} < z_{1-\alpha/2},$$

scritto come sistema

$$\begin{cases} \sqrt{\frac{n}{e^\theta}} > z_{\alpha/2}, \\ \sqrt{\frac{n}{e^\theta}} < z_{1-\alpha/2}. \end{cases}$$

La prima disequazione risulta sempre vera sapendo che  $e^\theta \in (0, +\infty)$  e  $z_{\alpha/2} < 0$  se  $\alpha \in (0, 1)$ . Nella seconda

$$\frac{n}{e^\theta} < z_{1-\alpha/2}^2,$$

si isola  $\theta$

$$\theta > \log \left( \frac{n}{z_{\alpha/2}^2} \right).$$

Gli estremi dell'intervallo quando  $X = n$  saranno  $(\log[n/z_{\alpha/2}^2], +\infty)$ .

#### 4.4.6 Bayesiano

Per l'intervallo di credibilità si mantiene la scelta della priori di Jeffreys, nel caso delle quote la posteriori risultava  $\pi(\psi | x) = \text{Beta}'(x + 1/2, n - x + 1/2)$ . Le log-quote avranno di conseguenza una distribuzione con funzione di densità:

$$f(x, \alpha, \beta) = \frac{1}{B(\alpha, \beta)} \frac{e^{-\beta x}}{(1 + e^{-x})^{\alpha+\beta}}.$$

Perciò gli estremi dell'intervallo saranno i quantili di questa distribuzione:

$$\left( q_{\alpha/2, x+1/2, n-x+1/2}^{\pi(\theta|x)}, q_{1-\alpha/2, x+1/2, n-x+1/2}^{\pi(\theta|x)} \right).$$

#### 4.4.7 Applicazione ai dati

Si presentano i risultati numerici dei processi di stima intervallare applicati sui i dati (2.9) e (2.10).

Metodo $\alpha = 95\%$	$n=5 \hat{\theta} = 1.39$		$n=100 \hat{\theta} = 0.364$	
	Inf	Sup	Inf	Sup
Clopper-Pearson	-0.927	5.283	-0.051	0.788
Blaker	-0.652	4.575	-0.041	0.785
TRV	-0.508	4.343	-0.026	0.767
Wald	-0.805	3.578	-0.035	0.762
Wald corretto	-1.166	2.552	-0.048	0.748
Wilson	-0.509	3.281	-0.032	0.76
Bayesiano	-0.526	3.771	-0.031	0.766

**Tabella 4.2:** Estremi dei vari intervalli di confidenza per le log-quote al 95% per i dati (2.9) e (2.10).

Come ci aspetteremmo, i due intervalli esatti rimangono molto ampi per  $n = 5$ , col Clopper-Pearson che si riconferma eccessivamente conservativo. L'IC derivante dal log-rapporto di verosimiglianza è sorprendentemente simile al Blaker, nonostante il primo sia un metodo asintotico e il secondo esatto. La correzione del Wald implica una traslazione smisurata dell'intervallo verso sinistra, nel primo insieme di dati. Per  $n = 100$  nessun intervallo esclude la possibilità che  $p = 0.5$  e quindi  $\theta = 0$ , anche in questo caso il comportamento tra i vari metodi non cambia in modo rilevante.

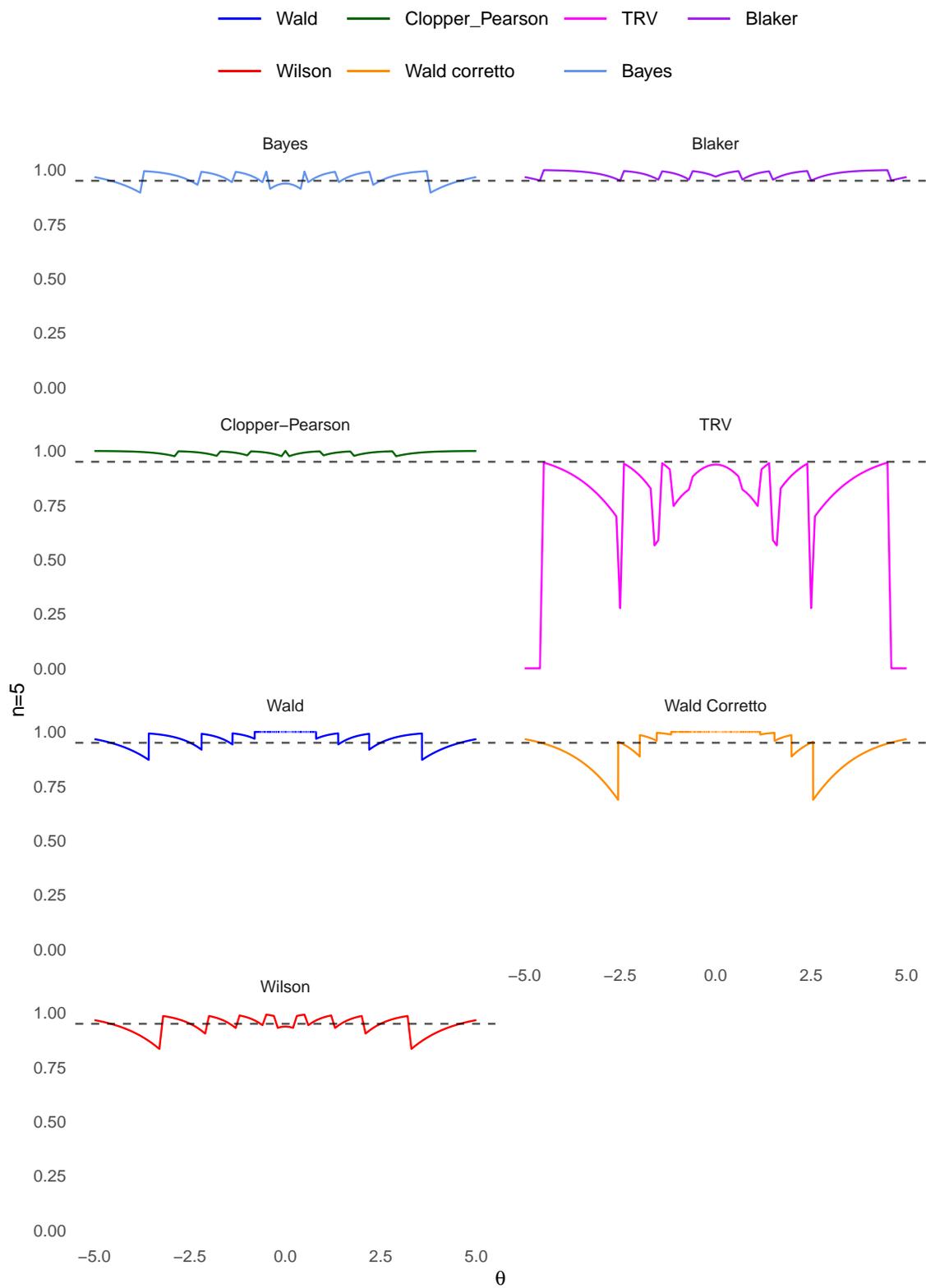
#### 4.4.8 Valutazione degli intervalli

##### Probabilità di copertura

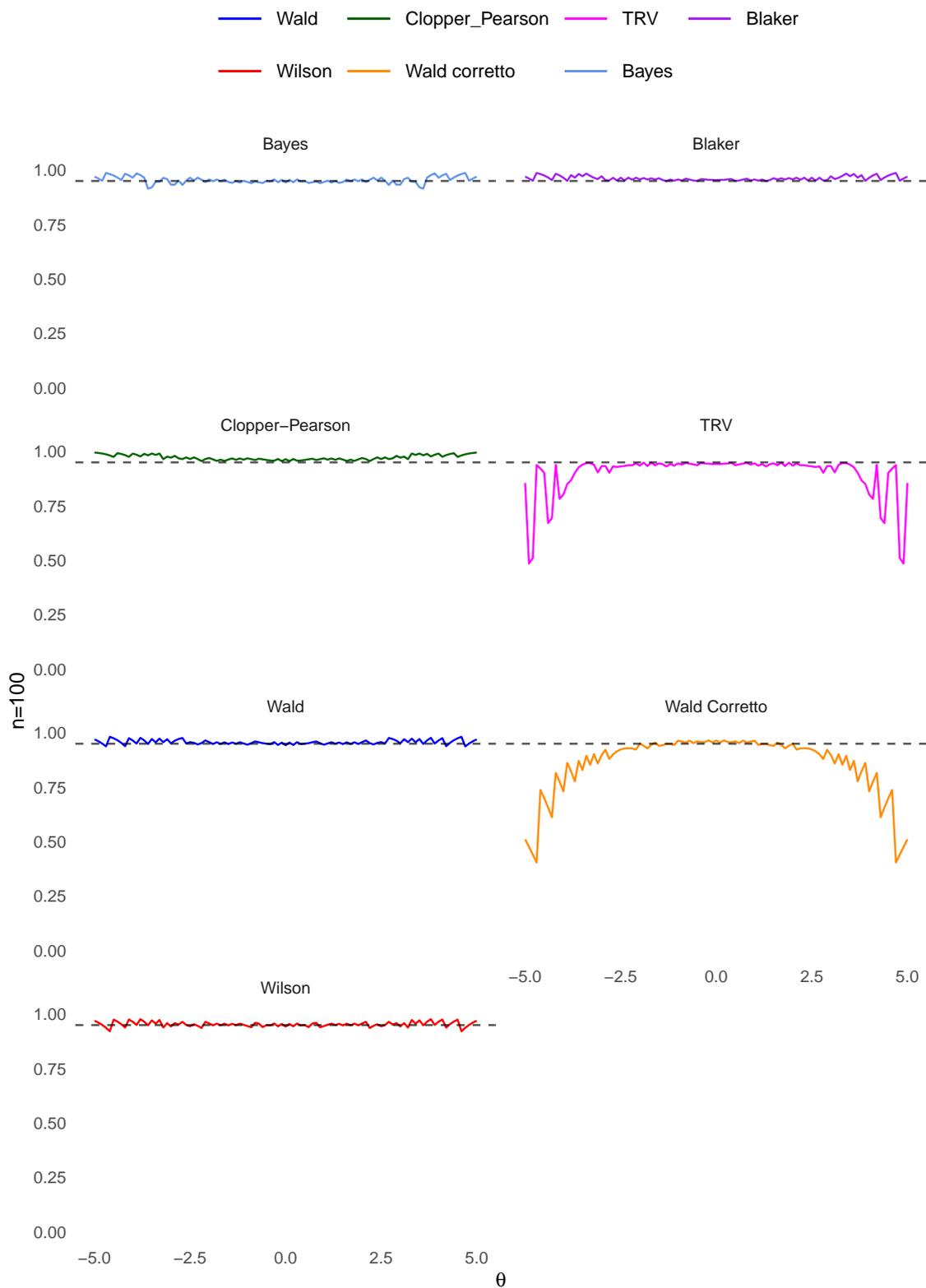
La riparametrizzazione in  $\theta(\cdot)$  altera le coperture solo degli IC di Wald e Wald corretto.

Risulta subito evidente che l'intervallo di Wald trova nelle log-quote la parametrizzazione più favorevole, in cui riesce ad avere performance decisamente migliori rispetto alle altre due. L'intervallo in questione si comporta sorprendentemente bene, con coperture sempre nei pressi del livello nominale, nonostante la numerosità campionaria esigua. In questo esempio la correzione dell'intervallo di Wald risulta inutile e anzi dannosa, abbassando la copertura decisamente sotto il 95%.

Con una numerosità campionaria così elevata in 4.7 sorprende il comportamento dell'intervallo di Wald corretto, che per i  $\theta$  molto elevati in valore assoluto subisce gravi cali nella copertura. L'intervallo di Wald migliora solamente con l'aumento della numerosità campionaria, come ci si aspetterebbe da un metodo approssimato come questo.

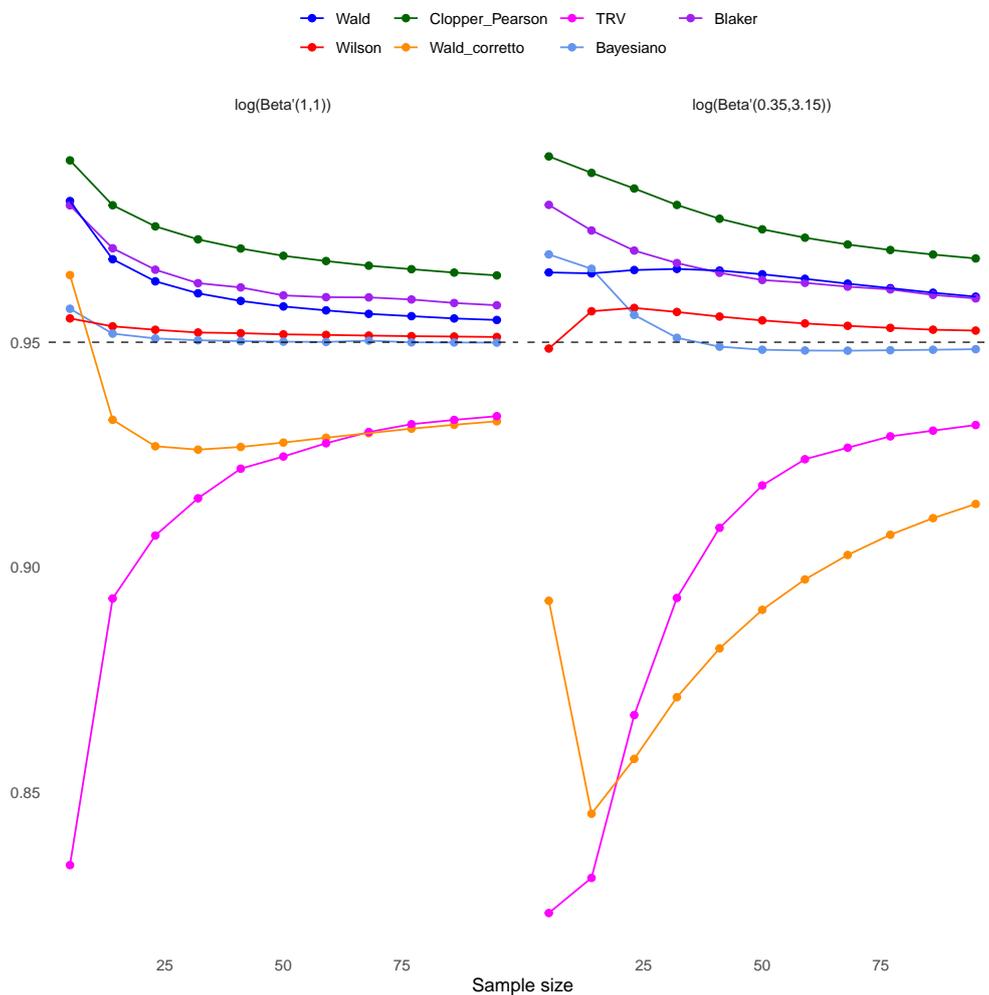


**Figura 4.6:** Probabilità di copertura quando  $n = 5$  degli intervalli di confidenza di livello 95% per le log-quote



**Figura 4.7:** Probabilità di copertura quando  $n = 100$  degli intervalli di confidenza di livello 95% per le log-quote

Probabilità media di copertura



**Figura 4.8:** Probabilità media di copertura in funzione della numerosità campionaria per gli intervalli di confidenza di livello 95% per le log-quote quando  $\theta$  ha un distribuzione  $\log \text{Beta}'(1, 1)$  e  $\log \text{Beta}'(0.35, 3.15)$

La 4.8 conferma la qualità del metodo di Wald nella stima intervallare delle log-quote. É inoltre evidente l’inadeguatezza dell’IC di Wald corretto, che per la prima volta è chiaramente peggiore della sua versione originale con entrambe le priori. Nonostante ciò, essa non è sempre molto vicina al livello nominale, a differenza per esempio del Wilson. Un’analisi della lunghezza dirà se questa copertura forse eccessiva sacrifica precisione della stima.

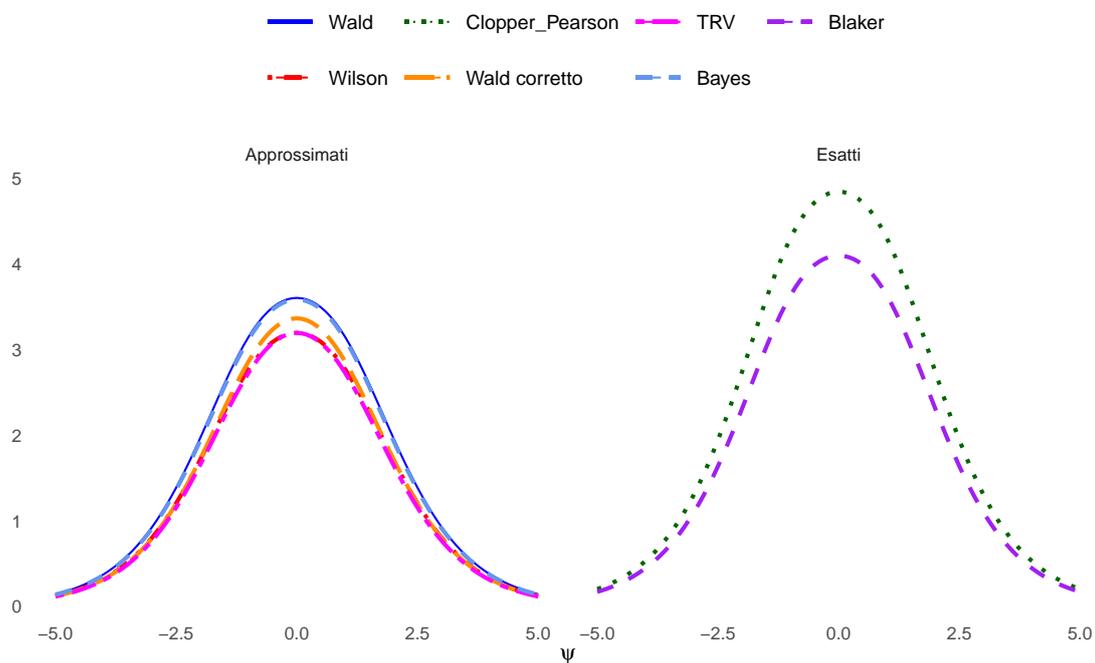
**Ampiezza attesa**

Anche con questa riparametrizzazione riscontriamo la stessa questione di prima, ovvero intervalli che hanno lunghezza infinita. Stavolta ciò accade anche per

$X = 0$ , oltre che per  $X = n$ , valori per cui si ha  $\hat{\theta} = \pm\infty$ . Vi sono sempre due metodi che non incorrono in questi "problemi", ovvero il bayesiano ed il Wald corretto. Si ricorre ancora ad una formula leggermente cambiata per il calcolo dell'ampiezza attesa:

$$\sum_{k=1}^{n-1} \binom{n}{k} p^k (1-p)^{n-k} [t_1(k) - t_2(k)],$$

dove  $t_1(k), t_2(k)$  sono gli estremi di un IC quando  $X = k$ .

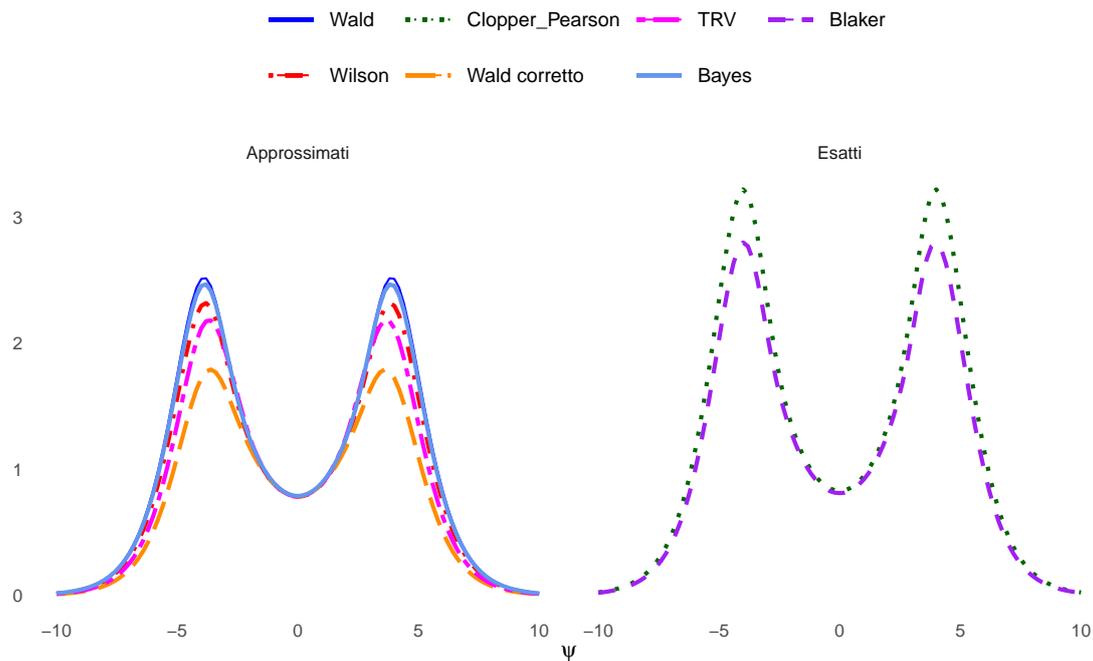


**Figura 4.9:** Lunghezza attesa al variare di  $\theta$  quando  $n = 5$

L'intervallo di Wald è di fatto più ampio del Wilson, ma non eccede in conservatività, soprattutto se comparato con i metodi esatti. L'ampiezza del Wilson è sorprendente, si mantiene simile a quella del TRV nonostante le coperture completamente diverse. Certamente in questo problema la scelta di un metodo potrebbe escludere quelli esatti, in quanto anche le approssimazioni offerte sono eccellenti.

Come per  $p$ , le lunghezze attese sono quasi identiche per  $n = 100$ , implicando che la copertura sarà la grandezza determinante per valutare questi metodi.

**Conclusioni** Nonostante la crescita prestazionale osservata con questa riparametrizzazione, l'intervallo di Wald non è ancora il metodo approssimato più ideale.



**Figura 4.10:** Lunghezza attesa al variare di  $\theta$  quando  $n = 100$

Questo titolo spetta al metodo di Wilson, il quale è comparabile al bayesiano ed al Wald in termini di copertura, ma risulta costantemente più preciso. Nell'ambito di stima intervallare tramite metodi esatti le conclusioni non cambiano, dimostrando che il Blaker ha effettivamente sostituito il Clopper-Pearson come gold standard.



## Conclusioni finali

La ricerca del metodo migliore per la stima intervallare in un modello binomiale ha, come spesso accade nella statistica, un risposta non univoca. Tra i metodi presentati non ne esiste uno che sia sempre il più preciso e che si possa definire ideale. Nei Capitoli 3 e 4 gli intervalli sono stati valutati dal punto di vista delle performance in termini di copertura e lunghezza, sotto questo metro di giudizio si riscontravano approcci totalmente inadatti: alcuni per coperture insufficienti (Wald e TRV), altri per ampiezza eccessiva dovuta a copertura esageratamente alta (Clopper-Pearson). Questa valutazione, sebbene corretta dal punto di vista statistico, non considera un aspetto molto importante: la comprensibilità. Spesso infatti la stima intervallare e i suoi metodi vengono presentati in ambiti esterni alla statistica, come quello biomedico, in cui la semplicità di un intervallo e l'immediatezza nella sua costruzione sono pregi considerevoli. L'intervallo di Wald, criticato in [Andersson \(2023\)](#) per la sua inadeguatezza nella stima, risulta facile da costruire e parte da una statistica test molto semplice da capire: essa sfrutta infatti il teorema del limite centrale e si distribuisce come una normale standard, distribuzione decisamente nota. [Andersson \(2023\)](#) critica fortemente la presentazione di un metodo così inadatto nei libri per la didattica, tuttavia la semplicità dell'intervallo di Wald rende più comprensibile il processo dietro alla costruzione di un intervallo di confidenza. La sua versione corretta è certamente migliore in termini di copertura, ma l'idea dietro alla correzione può risultare troppo vaga per uno studente o un ricercatore non specializzato in statistica. L'altro approccio criticato nel Capitolo 3, ovvero quello basato sul TRV, possiede una delle statistiche test più immediate dal punto di vista del significato della formula: il rapporto tra verosimiglianze, avendo definito la verosimiglianza come quanto un certo campione rifletta un valore del parametro, è chiaramente una quantità sensata per confrontare valori da assegnare al parametro stesso. Nonostante ciò, la costruzione dell'intervallo segue un processo troppo complesso per essere presentata come metodo principale di stima intervallare. Gli approcci

che hanno mostrato performance migliori risultano i più difficili da spiegare: il bayesiano mostra estremi relativamente semplici ma per essere compreso necessita di un'introduzione all'intera statistica bayesiana, l'intervallo di Wilson presenta estremi piuttosto complessi e una statistica test poco immediata da ricavare e l'intervallo di Blaker, che secondo [Blaker \(1998\)](#) dovrebbe rappresentare il nuovo gold standard, risulta problematico da comprendere e da costruire, per via dei nuovi concetti che introduce.

In conclusione, se si desidera affrontare un problema di stima intervallare nel modo migliore da un punto di vista di precisione nella stima, gli intervalli consigliati sono il Wilson ed il bayesiano, nell'ipotesi che non venga richiesta un metodo esatto, e l'intervallo di Blaker altrimenti. Nonostante ciò, l'intervallo di Wilson non è perfetto, con dei cali in copertura per valori estremi di  $p$ . Quanto appena detto rende evidente che un problema di stima intervallare non dovrebbe essere affrontato cercando il metodo perfetto, ma quello più adatto alla situazione in cui ci si trova. La soluzione non dovrebbe consistere solamente nell'intervallo con i suoi estremi, ma dovrebbe essere accompagnata da un'analisi della copertura del metodo utilizzato in quel contesto di campionamento.

## Bibliografia

(2017). Dataworld, ml classification: Career longevity for nba players (dataset).

(2024). Kaggle, predict survival of patients with heart failure (dataset).

AGRESTI, A. & COULL, B. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician* **52**, 119–126.

ANDERSSON, G. (2023). The wald confidence interval for a binomial  $p$  as an illuminating "bad" example. *The American Statistician* **77**, 443–448.

AZZALINI, A. (2008). *Inferenza Statistica*. Springer Verlag.

BLAKER, H. (1998). Confidence curves and improved exact confidence intervals for discrete distributions. *The Canadian Journal of Statistics* **28**, 783–798.

CASELLA, G. & BERGER, R. L. (2008). *Statistical Inference*. Brooks/Cole.

PACE, L. & SALVAN, A. (2001). *Introduzione alla statistica II*. Cedam.



## Appendice A

### Codice R

---

```
library(tidyverse)
library(cowplot)
library(gridGraphics)
library(extraDistr)
library(binom)

#####
#Capitolo 2#
#####

TRV_IC <- function(p, n = 10, x = 3, alpha = 0.05) {
  TRV <- -2 * log(dbinom(x = x, size = n, prob = p) / dbinom(x = x, size =
    n, prob = x / n))
  TRVs <- -2 * log(dbinom(c(0:n), n, p) / dbinom(c(0:n), n, c(0:n) / n))

  p_vals <- apply(as.matrix(TRVs), 1, function(x) {
    sum(dbinom(c(0:n), n, p)[which(TRVs >= x)])
  })
  lambda <- max(TRVs[p_vals > alpha])
  return(TRV - lambda)
}

TRV_IC <- Vectorize(TRV_IC, vectorize.args = "p")

#### PLOT TRV ####
(plot_lrt <- ggplot() +
  geom_function(fun = TRV_IC, args = list(n = 50, x = 20), n = 1000,
    linewidth = 0.5) +
```

```

xlab("p") +
ylim(c(-5, 70)) +
ylab(expression(paste("W(x)-", lambda[alpha]))) +
theme(
  axis.text = element_text(size = 12),
  axis.title = element_text(size = 14), axis.ticks = element_blank(),
  legend.background = element_blank(),
  legend.key = element_blank(), panel.background = element_blank(),
  panel.border = element_blank(), strip.background = element_blank(),
  plot.background = element_blank(), complete = TRUE, plot.caption =
  element_text(hjust = 0.5)
) +
geom_hline(aes(yintercept = 0, col = "darkred"), linewidth = 0.7,
  linetype = 2, show.legend = F))

#####
#Capitolo 3#
#####

IC_Blaker <- function(p, x, n, alpha = 0.05) {
  gamm <- min(pbinom(x, n, p), pbinom(x - 1, n, p, lower.tail = F))
  gammas <- apply(cbind(pbinom(c(0:n), n, p), pbinom(c(0:n) - 1, n, p,
    lower.tail = F)), 1, min)
  Xs <- which(gammas <= gamm) - 1
  sum(dbinom(Xs, n, p)) - alpha
}
IC_Blaker <- Vectorize(IC_Blaker, vectorize.args = "p")

cov_Blaker <- function(p, n, alpha = 0.05) {
  gammas <- apply(cbind(pbinom(c(0:n), n, p), pbinom(c(0:n) - 1, lower.tail
    = F, n, p)), 1, min)
  sorted_gammas <- sort(gammas, decreasing = F, index.return = T)$ix - 1
  Xs <- sorted_gammas[cumsum(dbinom(sorted_gammas, n, p)) > alpha]
  sum(dbinom(Xs, n, p))
}
cov_Blaker <- Vectorize(cov_Blaker, vectorize.args = "p")
cov_Wald <- function(p = 0.5, n = 50, alpha = 0.05) {
  p_hat <- c(0:n) / n

```

```

z <- qnorm(alpha / 2)
I_Wald <- c(rep(p, n + 1) >= p_hat + z * sqrt(p_hat * (1 - p_hat) / n) &
  rep(p, n + 1) <= p_hat - z * sqrt(p_hat * (1 - p_hat) / n))
k_Wald <- which(I_Wald == 1) - 1
sum(dbinom(k_Wald, n, p))
}

cov_Clopper_Pearson <- function(p = 0.5, n = 50, alpha = 0.05) {
  p_hat <- c(0:n) / n
  z <- qnorm(alpha / 2)
  I_Clopper_Pearson <- c(rep(p, n + 1) >= qbeta(alpha / 2, c(0:n), n -
    c(0:n) + 1) & rep(p, n + 1) <= qbeta(1 - alpha / 2, c(0:n) + 1, n -
    c(0:n)))
  k_Clopper_Pearson <- which(I_Clopper_Pearson == 1) - 1
  sum(dbinom(k_Clopper_Pearson, n, p))
}

cov_Wilson <- function(p = 0.5, n = 50, alpha = 0.05) {
  p_hat <- c(0:n) / n
  z <- qnorm(alpha / 2)
  I_Wilson <- c(rep(p, n + 1) >= (p_hat + z^2 / (2 * n) + z * sqrt((p_hat *
    (1 - p_hat) + z^2 / (4 * n)) / n)) / (1 + z^2 / n) &
  rep(p, n + 1) <= (p_hat + z^2 / (2 * n) - z * sqrt((p_hat * (1 - p_hat)
    + z^2 / (4 * n)) / n)) / (1 + z^2 / n))
  if(p==1){I_Wilson[n+1]=T}
  k_Wilson <- which(I_Wilson == 1) - 1
  sum(dbinom(k_Wilson, n, p))
}

cov_Bayesiano <- function(p = 0.5, n = 50, alpha = 0.05) {
  p_hat <- c(0:n) / n
  z <- qnorm(alpha / 2)
  I_Bayesiano <- c(rep(p, n + 1) >= qbeta(alpha / 2, c(0:n) + 0.5, n -
    c(0:n) + 0.5) & rep(p, n + 1) <= qbeta(1 - alpha / 2, c(0:n) + 0.5, n
    - c(0:n) + 0.5))
  k_Bayesiano <- which(I_Bayesiano == 1) - 1
  sum(dbinom(k_Bayesiano, n, p))
}

```

```

cov_Wald_corretto <- function(p = 0.5, n = 50, alpha = 0.05) {
  z <- qnorm(alpha / 2)
  n2 <- n + 4
  p_hat2 <- c((0:n) + 2) / n2
  I_Wald_corretto <- c(rep(p, n + 1) >= p_hat2 + z * sqrt(p_hat2 * (1 -
    p_hat2) / n2) & rep(p, n + 1) <= p_hat2 - z * sqrt(p_hat2 * (1 -
    p_hat2) / n2))
  k_Wald_corretto <- which(I_Wald_corretto == 1) - 1
  sum(dbinom(k_Wald_corretto, n, p))
}

cov_Wald <- Vectorize(cov_Wald, vectorize.args = "p")
cov_Clopper_Pearson <- Vectorize(cov_Clopper_Pearson, vectorize.args = "p")
cov_Wilson <- Vectorize(cov_Wilson, vectorize.args = "p")
cov_Wald_corretto <- Vectorize(cov_Wald_corretto, vectorize.args = "p")
cov_Bayesiano <- Vectorize(cov_Bayesiano, vectorize.args = "p")

cov_prob_TRV <- function(p, n, alpha = 0.05) {
  W <- -2 * log(dbinom(c(0:n), n, p) / dbinom(c(0:n), n, c(0:n) / n))
  W_order <- sort(W, index.return = T)$ix
  x_in_IC <- W_order[cumsum(dbinom(W_order - 1, n, p)) <= 0.95] - 1
  return(sum(dbinom(x_in_IC, n, p)))
}

cov_prob_TRV <- Vectorize(cov_prob_TRV, vectorize.args = "p")

#### Probabilita media di copertura ####
nnn <- seq(5, 100, by = 9)
dati_Wald <- apply(as.matrix(nnn), 1, function(n) {
  integrate(function(p) cov_Wald(p, n = n, alpha = 0.05), subdivisions =
    2000, upper = 1, lower = 0)$value
})
dati_Clopper_Pearson <- apply(as.matrix(nnn), 1, function(n) {
  integrate(function(p) cov_Clopper_Pearson(p, n = n, alpha = 0.05),
    subdivisions = 2000, upper = 1, lower = 0)$value
})
dati_Wilson <- apply(as.matrix(nnn), 1, function(n) {
  integrate(function(p) cov_Wilson(p, n = n, alpha = 0.05), subdivisions =
    2000, upper = 1, lower = 0)$value
})

```

```
})
dati_Wald_corretto <- apply(as.matrix(nnn), 1, function(n) {
  integrate(function(p) cov_Wald_corretto(p, n = n, alpha = 0.05),
    subdivisions = 2000, upper = 1, lower = 0)$value
})
dati_Bayesiano <- apply(as.matrix(nnn), 1, function(n) {
  integrate(function(p) cov_Bayesiano(p, n = n, alpha = 0.05), subdivisions
    = 2000, upper = 1, lower = 0)$value
})
dati_TRV <- apply(as.matrix(nnn), 1, function(n) {
  integrate(function(p) cov_prob_TRV(p, n = n, alpha = 0.05), subdivisions
    = 2000, upper = 1, lower = 0)$value
})
dati_Blaker <- apply(as.matrix(nnn), 1, function(n) {
  integrate(function(p) cov_Blaker(p, n = n, alpha = 0.05), subdivisions =
    2000, upper = 1, lower = 0)$value
})
dati_unif_wide <- as_tibble(cbind(dati_Wald, dati_Clopper_Pearson,
  dati_Wilson, dati_Wald_corretto, dati_Bayesiano, dati_TRV, dati_Blaker))
dati_unif <- dati_unif_wide %>%
  pivot_longer(cols = starts_with("dati"), cols_vary = "slowest", names_to
    = "Metodo", names_prefix = "dati_", values_to = "Cov_bar") %>%
  mutate(Priori = "Uniforme")

# Beta prior####
a <- 35 / 10
b <- 315 / 10
dati_beta_Wald <- apply(as.matrix(seq(5, 100, by = 9)), 1, function(n) {
  integrate(function(p) dbeta(p, a, b) * cov_Wald(p, n = n, alpha = 0.05),
    subdivisions = 2000, upper = 1, lower = 0)$value
})
dati_beta_Clopper_Pearson <- apply(as.matrix(seq(5, 100, by = 9)), 1,
  function(n) {
  integrate(function(p) dbeta(p, a, b) * cov_Clopper_Pearson(p, n = n,
    alpha = 0.05), subdivisions = 2000, upper = 1, lower = 0)$value
})
dati_beta_Wilson <- apply(as.matrix(seq(5, 100, by = 9)), 1, function(n) {
```

```

integrate(function(p) dbeta(p, a, b) * cov_Wilson(p, n = n, alpha =
  0.05), subdivisions = 2000, upper = 1, lower = 0)$value
})
dati_beta_Wald_corretto <- apply(as.matrix(seq(5, 100, by = 9)), 1,
  function(n) {
integrate(function(p) dbeta(p, a, b) * cov_Wald_corretto(p, n = n, alpha
  = 0.05), subdivisions = 2000, upper = 1, lower = 0)$value
})
dati_beta_Bayesiano <- apply(as.matrix(seq(5, 100, by = 9)), 1, function(n)
  {
integrate(function(p) dbeta(p, a, b) * cov_Bayesiano(p, n = n, alpha =
  0.05), subdivisions = 2000, upper = 1, lower = 0)$value
})
dati_beta_TRV <- apply(as.matrix(seq(5, 100, by = 9)), 1, function(n) {
integrate(function(p) dbeta(p, a, b) * cov_prob_TRV(p, n = n, alpha =
  0.05), subdivisions = 2000, upper = 1, lower = 0)$value
})
dati_beta_Blaker <- apply(as.matrix(seq(5, 100, by = 9)), 1, function(n) {
integrate(function(p) dbeta(p, a, b) * cov_Blaker(p, n = n, alpha =
  0.05), subdivisions = 2000, upper = 1, lower = 0)$value
})
dati_beta_wide <- as_tibble(cbind(dati_beta_Wald,
  dati_beta_Clopper_Pearson, dati_beta_Wilson, dati_beta_Wald_corretto,
  dati_beta_Bayesiano, dati_beta_TRV, dati_beta_Blaker))
dati_beta <- dati_beta_wide %>%
  pivot_longer(cols = starts_with("dati"), cols_vary = "slowest", names_to
    = "Metodo", names_prefix = "dati_beta_", values_to = "Cov_bar") %>%
  mutate(Priori = "zBeta")
dati_cov_bar <- dati_unif %>%
  add_row(dati_beta) %>%
  mutate(Priori = as.factor(Priori))
nomipriori <- c(zBeta = "Beta", Uniforme = "Uniforme")

(plot_cov_bar <- ggplot(dati_cov_bar, aes(x = rep(nnn, 14), y = Cov_bar,
  col = Metodo)) +
  geom_point(size = 1.7) +
  geom_line() +
  geom_hline(yintercept = 0.95, linetype = 2, alpha = 0.7) +

```

```

theme(
  legend.position = "top", axis.ticks = element_blank(), legend.background
    = element_blank(),
  legend.key = element_blank(), panel.background = element_blank(),
  panel.border = element_blank(), strip.background = element_blank(),
  plot.background = element_blank(), complete = TRUE, plot.caption =
    element_text(hjust = 0.5)
) +
xlab("Sample size") +
ylab("") +
scale_color_manual(
  name = "", breaks = c("Wald", "Wilson", "Clopper_Pearson",
    "Wald_corretto", "TRV", "Bayesiano", "Blaker"),
  values = c(
    "Wald" = "blue", "Wilson" = "red", "Clopper_Pearson" = "darkgreen",
    "Wald_corretto" = "darkorange", "TRV" = "magenta",
    "Bayesiano" = "cornflowerblue", "Blaker" = "purple"
  )
) +
facet_grid(~Priori, labeller = labeller(Priori = nomipriori))

#### Lunghezza attesa ####
expected_length_Wald <- function(p, n, alpha = 0.05) {
  p_hat <- c(0:n) / n
  z <- qnorm(alpha / 2)
  Length_Wald <- p_hat - z * sqrt(p_hat * (1 - p_hat) / n) - (p_hat + z *
    sqrt(p_hat * (1 - p_hat) / n))
  sum(dbinom(0:n, n, p) * Length_Wald)
}
expected_length_Clopper_Pearson <- function(p, n, alpha = 0.05) {
  p_hat <- c(0:n) / n
  z <- qnorm(alpha / 2)
  Length_Clopper_Pearson <- -(qbeta(alpha / 2, c(0:n), n - c(0:n) + 1)) +
    qbeta(1 - alpha / 2, c(0:n) + 1, n - c(0:n))
  sum(dbinom(0:n, n, p) * Length_Clopper_Pearson)
}
expected_length_Wilson <- function(p, n, alpha = 0.05) {

```

```

p_hat <- c(0:n) / n
z <- qnorm(alpha / 2)
Length_Wilson <- -((p_hat + z^2 / (2 * n) + z * sqrt((p_hat * (1 - p_hat)
  + z^2 / (4 * n)) / n)) / (1 + z^2 / n)) +
  (p_hat + z^2 / (2 * n) - z * sqrt((p_hat * (1 - p_hat) + z^2 / (4 * n))
  / n)) / (1 + z^2 / n)
sum(dbinom(0:n, n, p) * Length_Wilson)
}
expected_length_Wald_2_2 <- function(p, n, alpha = 0.05) {
  p_hat <- c(0:n) / n
  z <- qnorm(alpha / 2)
  n2 <- n + 4
  p_hat2 <- c((0:n) + 2) / n2
  Length_Wald_2_2 <- p_hat2 - z * sqrt(p_hat2 * (1 - p_hat2) / n2) -
    (p_hat2 + z * sqrt(p_hat2 * (1 - p_hat2) / n2))
  sum(dbinom(0:n, n, p) * Length_Wald_2_2)
}
expected_length_TRV <- function(p, n, alpha = 0.05) {
  p_hat <- c(0:n) / n
  z <- qnorm(alpha / 2)
  Length_TRV <- apply(as.matrix(p_hat[-c(1, n + 1)]), 1, function(x) {
    uniroot(function(k) TRV_IC(k, n = n, alpha = alpha, x = x * n), interval
      = c(x, 1))$root - uniroot(function(l) TRV_IC(l, n = n, alpha =
      alpha, x = x * n), interval = c(0, x))$root
  })
  sum(dbinom(0:n, n, p) * c(0, Length_TRV, 0))
}
expected_length_Bayes <- function(p, n, alpha = 0.05) {
  p_hat <- c(0:n) / n
  z <- qnorm(alpha / 2)
  Length_Bayes <- -(qbeta(alpha / 2, c(0:n) + 0.5, n - c(0:n) + 0.5)) +
    qbeta(1 - alpha / 2, c(0:n) + 0.5, n - c(0:n) + 0.5)
  sum(dbinom(0:n, n, p) * Length_Bayes)
}
expected_length_Blaker <- function(p, n, alpha = 0.05) {
  p_hat <- c(0:n) / n
  z <- qnorm(alpha / 2)

```

```

Length_Blaker <- apply(as.matrix(p_hat[-c(1, n + 1)]), 1, function(x) {
  uniroot(function(k) IC_Blaker(k, n = n, alpha = alpha, x = x * n),
    interval = c(x, 1))$root - uniroot(function(k) IC_Blaker(k, n = n,
    alpha = alpha, x = x * n), interval = c(0, x))$root
})
Length_Blaker <- c(
  uniroot(function(k) IC_Blaker(k, n = n, alpha = alpha, x = 0), interval
    = c(0, 1))$root - 0,
  Length_Blaker,
  1 - uniroot(function(k) IC_Blaker(k, n = n, alpha = alpha, x = n),
    interval = c(0, 1))$root
)
sum(dbinom(0:n, n, p) * Length_Blaker)
}
expected_length_Wald <- Vectorize(expected_length_Wald, vectorize.args =
  "p")
expected_length_Clopper_Pearson <-
  Vectorize(expected_length_Clopper_Pearson, vectorize.args = "p")
expected_length_Wilson <- Vectorize(expected_length_Wilson, vectorize.args
  = "p")
expected_length_Wald_2_2 <- Vectorize(expected_length_Wald_2_2,
  vectorize.args = "p")
expected_length_TRV <- Vectorize(expected_length_TRV, vectorize.args = "p")
expected_length_Bayes <- Vectorize(expected_length_Bayes, vectorize.args =
  "p")
expected_length_Blaker <- Vectorize(expected_length_Blaker, vectorize.args
  = "p")

# plot n=5
(plot_exp_length_n5 <- ggplot() +
  geom_function(fun = expected_length_Wald, args = list(n = 5), data =
    data.frame(x = c(0, 1), fun_name = "Approssimati"), aes(linetype =
    "Wald", col = "Wald")) +
  geom_function(fun = expected_length_Clopper_Pearson, args = list(n = 5),
    data = data.frame(x = c(0, 1), fun_name = "Esatti"), aes(linetype =
    "Clopper_Pearson", col = "Clopper_Pearson"), linewidth = 1) +

```

```

geom_function(fun = expected_length_Wilson, args = list(n = 5), data =
  data.frame(x = c(0, 1), fun_name = "Approssimati"), aes(linetype =
    "Wilson", col = "Wilson"), linewidth = 1) +
geom_function(fun = expected_length_Wald_2_2, args = list(n = 5), data =
  data.frame(x = c(0, 1), fun_name = "Approssimati"), aes(linetype =
    "Wald corretto", col = "Wald corretto"), linewidth = 1) +
geom_function(fun = expected_length_TRV, args = list(n = 5), data =
  data.frame(x = c(0, 1), fun_name = "Approssimati"), aes(linetype =
    "TRV", col = "TRV"), linewidth = 1) +
geom_function(fun = expected_length_Bayes, args = list(n = 5), data =
  data.frame(x = c(0, 1), fun_name = "Approssimati"), aes(linetype =
    "Bayes", col = "Bayes"), linewidth = 1) +
geom_function(fun = expected_length_Blaker, args = list(n = 5), data =
  data.frame(x = c(0, 1), fun_name = "Esatti"), aes(linetype = "Blaker",
    col = "Blaker"), linewidth = 1) +
xlab("p") +
ylab("") +
facet_wrap(~fun_name) +
scale_linetype_manual(
  name = "", breaks = c("Wald", "Wilson", "Clopper_Pearson", "Wald
    corretto", "TRV", "Bayes", "Blaker"),
  values = c("Wald" = 1, "Wilson" = 4, "Clopper_Pearson" = 3, "Wald
    corretto" = 5, "TRV" = 6, "Bayes" = 7, "Blaker" = 8)
) +
theme(
  legend.position = "top", axis.ticks = element_blank(), legend.background
    = element_blank(),
  legend.key = element_blank(), panel.background = element_blank(),
  panel.border = element_blank(), strip.background = element_blank(),
  plot.background = element_blank(),
  legend.key.size = unit(1, "cm"), complete = TRUE, legend.text =
    element_text(size = 10)
) +
scale_color_manual(
  name = "", breaks = c("Wald", "Wilson", "Clopper_Pearson", "Wald
    corretto", "TRV", "Bayes", "Blaker"),
  values = c(
    "Wald" = "blue", "Wilson" = "red", "Clopper_Pearson" = "darkgreen",

```

```
"Wald corretto" = "darkorange", "TRV" = "magenta",
"Bayes" = "cornflowerblue", "Blaker" = "purple"
)
))

# n=100
(plot_exp_length_n100 <- ggplot() +
  geom_function(fun = expected_length_Wald, args = list(n = 100), data =
    data.frame(x = c(0, 1), fun_name = "Approssimati"), aes(linetype =
      "Wald", col = "Wald")) +
  geom_function(fun = expected_length_Clopper_Pearson, args = list(n =
    100), data = data.frame(x = c(0, 1), fun_name = "Esatti"),
    aes(linetype = "Clopper_Pearson", col = "Clopper_Pearson"), linewidth
    = 1) +
  geom_function(fun = expected_length_Wilson, args = list(n = 100), data =
    data.frame(x = c(0, 1), fun_name = "Approssimati"), aes(linetype =
      "Wilson", col = "Wilson"), linewidth = 1) +
  geom_function(fun = expected_length_Wald_2_2, args = list(n = 100), data
    = data.frame(x = c(0, 1), fun_name = "Approssimati"), aes(linetype =
      "Wald corretto", col = "Wald corretto"), linewidth = 1) +
  geom_function(fun = expected_length_TRV, args = list(n = 100), data =
    data.frame(x = c(0, 1), fun_name = "Approssimati"), aes(linetype =
      "TRV", col = "TRV"), linewidth = 1) +
  geom_function(fun = expected_length_Bayes, args = list(n = 100), data =
    data.frame(x = c(0, 1), fun_name = "Approssimati"), aes(linetype =
      "Bayes", col = "Bayes"), linewidth = 1) +
  geom_function(fun = expected_length_Blaker, args = list(n = 100), data =
    data.frame(x = c(0, 1), fun_name = "Esatti"), aes(linetype = "Blaker",
      col = "Blaker"), linewidth = 1) +
  xlab("p") +
  ylab("") +
  facet_wrap(~fun_name) +
  scale_linetype_manual(
    name = "", breaks = c("Wald", "Wilson", "Clopper_Pearson", "Wald
      corretto", "TRV", "Bayes", "Blaker"),
    values = c("Wald" = 1, "Wilson" = 4, "Clopper_Pearson" = 3, "Wald
      corretto" = 5, "TRV" = 6, "Bayes" = 7, "Blaker" = 8)
  ) +
```

```

theme(
  legend.position = "top", axis.ticks = element_blank(), legend.background
    = element_blank(),
  legend.key = element_blank(), panel.background = element_blank(),
  panel.border = element_blank(), strip.background = element_blank(),
  plot.background = element_blank(),
  legend.key.size = unit(1, "cm"), complete = TRUE, legend.text =
    element_text(size = 10)
) +
scale_color_manual(
  name = "", breaks = c("Wald", "Wilson", "Clopper_Pearson", "Wald
    corretto", "TRV", "Bayes", "Blaker"),
  values = c(
    "Wald" = "blue", "Wilson" = "red", "Clopper_Pearson" = "darkgreen",
    "Wald corretto" = "darkorange", "TRV" = "magenta",
    "Bayes" = "cornflowerblue", "Blaker" = "purple"
  )
))

```

#### Probabilita di copertura ####

#n=5

```

(plot_cov_n5 <- ggplot() +
  geom_function(fun = cov_Wald, n = 5000, aes(col = "Wald"), data =
    data.frame(x = c(0, 1), fun_name = "Wald"), args = list(n = 5, alpha =
    0.05)) +
  geom_function(fun = cov_Wald_corretto, n = 5000, aes(col = "Wald
    corretto"), data = data.frame(x = c(0, 1), fun_name = "Wald
    corretto"), args = list(n = 5, alpha = 0.05)) +
  geom_function(fun = cov_Wilson, n = 5000, aes(col = "Wilson"), data =
    data.frame(x = c(0, 1), fun_name = "Wilson"), args = list(n = 5, alpha
    = 0.05)) +
  geom_function(fun = cov_Clopper_Pearson, n = 5000, aes(col =
    "Clopper_Pearson"), data = data.frame(x = c(0, 1), fun_name =
    "Clopper_Pearson"), args = list(n = 5, alpha = 0.05)) +
  geom_function(fun = cov_Bayesiano, n = 5000, aes(col = "Bayes"), data =
    data.frame(x = c(0, 1), fun_name = "Bayes"), args = list(n = 5, alpha
    = 0.05)) +

```

```

geom_function(fun = cov_prob_TRV, n = 5000, aes(col = "TRV"), data =
  data.frame(x = c(0, 1), fun_name = "TRV"), args = list(n = 5, alpha =
    0.05)) +
geom_function(fun = cov_Blaker, n = 5000, aes(col = "Blaker"), data =
  data.frame(x = c(0, 1), fun_name = "Blaker"), args = list(n = 5, alpha
    = 0.05)) +
xlim(c(0, 1)) +
ylim(c(0, 1)) +
geom_hline(yintercept = 0.95, linetype = 2, alpha = 0.7) +
facet_wrap(~fun_name, ncol = 2) +
xlab("p") +
ylab("n=5") +
theme(
  legend.position = "top", axis.ticks = element_blank(), legend.background
    = element_blank(),
  legend.key = element_blank(), panel.background = element_blank(),
  panel.border = element_blank(), strip.background = element_blank(),
  plot.background = element_blank(),
  legend.key.size = unit(1, "cm"), complete = TRUE, legend.text =
    element_text(size = 10)
) +
scale_color_manual(
  name = "", breaks = c("Wald", "Wilson", "Clopper_Pearson", "Wald
    corretto", "TRV", "Bayes", "Blaker"),
  values = c(
    "Wald" = "blue", "Wilson" = "red", "Clopper_Pearson" = "darkgreen",
    "Wald corretto" = "darkorange", "TRV" = "magenta",
    "Bayes" = "cornflowerblue", "Blaker" = "purple"
  )
)
)

# confronto tra coperture intervalli esatti
(plot_cov_esatti_n5 <- ggplot() +
  geom_function(fun = cov_Clopper_Pearson, aes(col = "Clopper_Pearson"),
    data = data.frame(x = c(0, 1), fun_name = "Clopper_Pearson"), args =
    list(n = 5, alpha = 0.05)) +
  geom_function(fun = cov_Blaker, aes(col = "Blaker"), data = data.frame(x
    = c(0, 1), fun_name = "Blaker"), args = list(n = 5, alpha = 0.05)) +

```

```

xlim(c(0, 1)) +
ylim(c(0.95, 1)) +
geom_hline(yintercept = 0.95, linetype = 2, alpha = 0.7) +
geom_hline(yintercept = 0.975, col = "red", linetype = 2, alpha = 0.7) +
facet_wrap(~fun_name, ncol = 2) +
xlab("p") +
ylab("n=5") +
theme(
  legend.position = "none", axis.ticks = element_blank(),
  legend.background = element_blank(),
  legend.key = element_blank(), panel.background = element_blank(),
  panel.border = element_blank(), strip.background = element_blank(),
  plot.background = element_blank(), strip.text = element_text(size = 15),
  legend.key.size = unit(1, "cm"), complete = TRUE, legend.text =
  element_text(size = 15)
) +
scale_color_manual(
  name = "", breaks = c("Wald", "Wilson", "Clopper_Pearson", "Wald
  corretto", "TRV", "Bayes", "Blaker"),
  values = c(
    "Wald" = "blue", "Wilson" = "red", "Clopper_Pearson" = "darkgreen",
    "Wald corretto" = "darkorange", "TRV" = "magenta",
    "Bayes" = "cornflowerblue", "Blaker" = "purple"
  )
)
))

# Analisi coperture minime agli estremi IC Wilson
nlminb(objective = function(p) cov_Wilson(p, n = 5), lower = 0, upper =
  0.99, start = 0.03)
nlminb(objective = function(p) cov_Wilson(p, n = 5), lower = 0, upper =
  0.99999, start = 0.97)
nlminb(objective = function(p) cov_Wilson(p, n = 100), lower = 0, upper =
  0.99999, start = 0.0015)
nlminb(objective = function(p) cov_Wilson(p, n = 100), lower = 0, upper =
  0.99999, start = 0.999)

# n=100
(plot_cov_n100 <- ggplot() +

```

```
geom_function(fun = cov_Wald, aes(col = "Wald"), data = data.frame(x =
  c(0, 1), fun_name = "Wald"), args = list(n = 100, alpha = 0.05)) +
geom_function(fun = cov_Wald_corretto, aes(col = "Wald corretto"), data =
  data.frame(x = c(0, 1), fun_name = "Wald corretto"), args = list(n =
  100, alpha = 0.05)) +
geom_function(fun = cov_Wilson, aes(col = "Wilson"), data = data.frame(x
  = c(0, 1), fun_name = "Wilson"), args = list(n = 100, alpha = 0.05)) +
geom_function(fun = cov_Clopper_Pearson, aes(col = "Clopper_Pearson"),
  data = data.frame(x = c(0, 1), fun_name = "Clopper_Pearson"), args =
  list(n = 100, alpha = 0.05)) +
geom_function(fun = cov_Bayesiano, aes(col = "Bayes"), data =
  data.frame(x = c(0, 1), fun_name = "Bayes"), args = list(n = 100,
  alpha = 0.05)) +
geom_function(fun = cov_prob_TRV, aes(col = "TRV"), data = data.frame(x =
  c(0, 1), fun_name = "TRV"), args = list(n = 100, alpha = 0.05)) +
geom_function(fun = cov_Blaker, aes(col = "Blaker"), data = data.frame(x
  = c(0, 1), fun_name = "Blaker"), args = list(n = 100, alpha = 0.05)) +
xlim(c(0, 1)) +
ylim(c(0.7, 1)) +
geom_hline(yintercept = 0.95, linetype = 2, alpha = 0.7) +
facet_wrap(~fun_name, ncol = 2) +
xlab("p") +
ylab("n=100") +
theme(
  legend.position = "top", axis.ticks = element_blank(), legend.background
    = element_blank(),
  legend.key = element_blank(), panel.background = element_blank(),
  panel.border = element_blank(), strip.background = element_blank(),
  plot.background = element_blank(),
  legend.key.size = unit(1, "cm"), complete = TRUE, legend.text =
    element_text(size = 10)
) +
scale_color_manual(
  name = "", breaks = c("Wald", "Wilson", "Clopper_Pearson", "Wald
    corretto", "TRV", "Bayes", "Blaker"),
  values = c(
    "Wald" = "blue", "Wilson" = "red", "Clopper_Pearson" = "darkgreen",
    "Wald corretto" = "darkorange", "TRV" = "magenta",
```

```

    "Bayes" = "cornflowerblue", "Blaker" = "purple"
  )
))

#####
#Capitolo 4#
#####

#### Quote ####

odds_cov_Wald <- function(p = 0.5, n = 50, alpha = 0.05) {
  th.true <- p / (1 - p)
  p_hat <- c(0:n) / n
  th_hat <- p_hat / (1 - p_hat)
  z <- qnorm(alpha / 2)
  I_Wald_odds <- c(rep(th.true, n + 1) >= th_hat + z * (1 + th_hat) *
    sqrt(th_hat / n) & rep(th.true, n + 1) <= th_hat - z * (1 + th_hat) *
    sqrt(th_hat / n))
  I_Wald_odds[n + 1] <- T
  k_Wald_odds <- which(I_Wald_odds == 1) - 1
  sum(dbinom(k_Wald_odds, n, p))
}

odds_cov_Wald_corretto <- function(p = 0.5, n = 50, alpha = 0.05) {
  th.true <- p / (1 - p)
  z <- qnorm(alpha / 2)
  n2 <- n + 4
  p_tilde <- c((0:n) + 2) / n2
  th_tilde <- p_tilde / (1 - p_tilde)
  I_Adj_Wald_odds <- c(rep(th.true, n + 1) >= th_tilde + z * (1 + th_tilde)
    * sqrt(th_tilde / n) & rep(th.true, n + 1) <= th_tilde - z * (1 +
    th_tilde) * sqrt(th_tilde / n))
  k_Adj_Wald_odds <- which(I_Adj_Wald_odds == 1) - 1
  sum(dbinom(k_Adj_Wald_odds, n, p))
}

#### Probabilita media di copertura ####

```

```
odds_cov_Wald <- Vectorize(odds_cov_Wald, vectorize.args = "p")
odds_cov_Wald_corretto <- Vectorize(odds_cov_Wald_corretto, vectorize.args
  = "p")

dati_odds_Wald <- apply(as.matrix(nnn), 1, function(n) {
  integrate(function(p) odds_cov_Wald(p, n = n, alpha = 0.05), subdivisions
    = 2000, upper = 1, lower = 0)$value
})
dati_odds_Clopper_Pearson <- dati_Clopper_Pearson
dati_odds_Wilson <- dati_Wilson
dati_odds_Wald_corretto <- apply(as.matrix(nnn), 1, function(n) {
  integrate(function(p) odds_cov_Wald_corretto(p, n = n, alpha = 0.05),
    subdivisions = 2000, upper = 1, lower = 0)$value
})
dati_odds_Bayesiano <- dati_Bayesiano
dati_odds_Blaker <- dati_Blaker
dati_odds_TRV <- dati_TRV

dati_odds_unif_wide <- as_tibble(cbind(dati_odds_Wald,
  dati_odds_Clopper_Pearson, dati_odds_Wilson, dati_odds_Wald_corretto,
  dati_odds_Bayesiano, dati_odds_TRV, dati_odds_Blaker))
dati_odds_unif <- dati_odds_unif_wide %>%
  pivot_longer(cols = starts_with("dati"), cols_vary = "slowest", names_to
    = "Metodo", names_prefix = "dati_odds_", values_to = "odds_cov_bar")
  %>%
  mutate(Priori = "Uniforme")

# Beta prior
a <- 35 / 10
b <- 315 / 10
dati_beta_odds_Wald <- apply(as.matrix(seq(5, 100, by = 9)), 1, function(n)
  {
  integrate(function(p) dbeta(p, a, b) * odds_cov_Wald(p, n = n, alpha =
    0.05), subdivisions = 2000, upper = 1, lower = 0)$value
  })
dati_beta_odds_Clopper_Pearson <- dati_beta_Clopper_Pearson
```

```

dati_beta_odds_Wilson <- dati_beta_Wilson
dati_beta_odds_Wald_corretto <- apply(as.matrix(seq(5, 100, by = 9)), 1,
  function(n) {
    integrate(function(p) dbeta(p, a, b) * odds_cov_Wald_corretto(p, n = n,
      alpha = 0.05), subdivisions = 2000, upper = 1, lower = 0)$value
  })
dati_beta_odds_Bayesiano <- dati_beta_Bayesiano

dati_beta_odds_Blaker <- dati_beta_Blaker
dati_beta_odds_TRV <- dati_beta_TRV

dati_beta_odds_wide <- as_tibble(cbind(
  dati_beta_odds_Wald, dati_beta_odds_Clopper_Pearson,
  dati_beta_odds_Wilson,
  dati_beta_odds_Wald_corretto, dati_beta_odds_Bayesiano,
  dati_beta_odds_TRV, dati_beta_odds_Blaker
))
dati_beta_odds <- dati_beta_odds_wide %>%
  pivot_longer(cols = starts_with("dati"), cols_vary = "slowest", names_to =
    "Metodo", names_prefix = "dati_beta_odds_", values_to =
    "odds_cov_bar") %>%
  mutate(Priori = "zBeta")
dati_odds_cov_bar <- dati_odds_unif %>%
  add_row(dati_beta_odds) %>%
  mutate(Priori = as.factor(Priori))

nomiprioriodds <- c(zBeta = "Beta'(0.35,3.15)", Uniforme = "Beta'(1,1)")

(plot_cov_bar_odds <- ggplot(dati_odds_cov_bar, aes(x = rep(nnn, 14), y =
  odds_cov_bar, col = Metodo)) +
  geom_point(size = 1.7) +
  geom_line() +
  geom_hline(yintercept = 0.95, linetype = 2, alpha = 0.7) +
  theme(
    legend.position = "top", axis.ticks = element_blank(), legend.background
    = element_blank(),
    legend.key = element_blank(), panel.background = element_blank(),
    panel.border = element_blank(), strip.background = element_blank(),

```

```

    plot.background = element_blank(), complete = TRUE, plot.caption =
      element_text(hjust = 0.5)
  ) +
  xlab("Sample size") +
  ylab("") +
  scale_color_manual(
    name = "", breaks = c("Wald", "Wilson", "Clopper_Pearson",
      "Wald_corretto", "TRV", "Bayesiano", "Blaker"),
    values = c(
      "Wald" = "blue", "Wilson" = "red", "Clopper_Pearson" = "darkgreen",
      "Wald_corretto" = "darkorange", "TRV" = "magenta",
      "Bayesiano" = "cornflowerblue", "Blaker" = "purple"
    )
  ) +
  facet_grid(~Priori, labeller = labeller(Priori = nomiprioriodds))

```

#### Probabilita di copertura ####

#n=5

```

(plot_cov_odds_n5 <- ggplot() +
  geom_function(fun = function(th) odds_cov_Wald(p = th / (1 + th), n = 5,
    alpha = 0.05), data = data.frame(x = c(0, 1), fun_name = "Wald"),
    aes(col = "Wald")) +
  geom_function(fun = function(th) odds_cov_Wald_corretto(p = th / (1 +
    th), n = 5, alpha = 0.05), data = data.frame(x = c(0, 1), fun_name =
    "Wald Corretto"), aes(col = "Wald corretto")) +
  geom_function(fun = function(th) cov_Wilson(p = th / (1 + th), n = 5,
    alpha = 0.05), data = data.frame(x = c(0, 1), fun_name = "Wilson"),
    aes(col = "Wilson")) +
  geom_function(fun = function(th) cov_Clopper_Pearson(p = th / (1 + th), n
    = 5, alpha = 0.05), data = data.frame(x = c(0, 1), fun_name =
    "Clopper-Pearson"), aes(col = "Clopper_Pearson")) + # , ) +
  geom_function(fun = function(th) cov_Bayesiano(p = th / (1 + th), n = 5,
    alpha = 0.05), data = data.frame(x = c(0, 1), fun_name = "Bayes"),
    aes(col = "Bayes")) +
  geom_function(fun = function(th) cov_prob_TRV(p = th / (1 + th), n = 5,
    alpha = 0.05), data = data.frame(x = c(0, 1), fun_name = "TRV"),
    aes(col = "TRV")) +

```

```

geom_function(fun = function(th) cov_Blaker(p = th / (1 + th), n = 5,
  alpha = 0.05), data = data.frame(x = c(0, 1), fun_name = "Blaker"),
  aes(col = "Blaker")) +
xlim(c(0, 30)) +
ylim(c(0, 1)) +
xlab(expression(paste(psi))) +
ylab("n=5") +
geom_hline(yintercept = 0.95, linetype = 2, alpha = 0.7) +
facet_wrap(~fun_name, ncol = 2) +
theme(
  legend.position = "top", axis.ticks = element_blank(), legend.background
    = element_blank(),
  legend.key = element_blank(), panel.background = element_blank(),
  panel.border = element_blank(), strip.background = element_blank(),
  plot.background = element_blank(),
  legend.key.size = unit(1, "cm"), complete = TRUE, legend.text =
    element_text(size = 10)
) +
scale_color_manual(
  name = "", breaks = c("Wald", "Wilson", "Clopper_Pearson", "Wald
    corretto", "TRV", "Bayes", "Blaker"),
  values = c(
    "Wald" = "blue", "Wilson" = "red", "Clopper_Pearson" = "darkgreen",
    "Wald corretto" = "darkorange", "TRV" = "magenta", "Bayes" =
      "cornflowerblue", "Blaker" = "purple"
  )
)
))

# n=100
(plot_cov_odds_n100 <- ggplot() +
  geom_function(fun = function(th) odds_cov_Wald(p = th / (1 + th), n =
    100, alpha = 0.05), data = data.frame(x = c(0, 1), fun_name = "Wald"),
  aes(col = "Wald")) +
  geom_function(fun = function(th) odds_cov_Wald_corretto(p = th / (1 +
    th), n = 100, alpha = 0.05), data = data.frame(x = c(0, 1), fun_name =
    "Wald Corretto"), aes(col = "Wald corretto")) +

```

```

geom_function(fun = function(th) cov_Wilson(p = th / (1 + th), n = 100,
  alpha = 0.05), data = data.frame(x = c(0, 1), fun_name = "Wilson"),
  aes(col = "Wilson")) +
geom_function(fun = function(th) cov_Clopper_Pearson(p = th / (1 + th), n
  = 100, alpha = 0.05), data = data.frame(x = c(0, 1), fun_name =
  "Clopper-Pearson"), aes(col = "Clopper_Pearson")) + # , ) +
geom_function(fun = function(th) cov_Bayesiano(p = th / (1 + th), n =
  100, alpha = 0.05), data = data.frame(x = c(0, 1), fun_name =
  "Bayes"), aes(col = "Bayes")) +
geom_function(fun = function(th) cov_prob_TRV(p = th / (1 + th), n = 100,
  alpha = 0.05), data = data.frame(x = c(0, 1), fun_name = "TRV"),
  aes(col = "TRV")) +
geom_function(fun = function(th) cov_Blaker(p = th / (1 + th), n = 100,
  alpha = 0.05), data = data.frame(x = c(0, 1), fun_name = "Blaker"),
  aes(col = "Blaker")) +
xlim(c(0, 30)) +
ylim(c(0.7, 1)) +
xlab(expression(paste(psi))) +
ylab("n=100") +
geom_hline(yintercept = 0.95, linetype = 2, alpha = 0.7) +
facet_wrap(~fun_name) +
theme(
  legend.position = "top", axis.ticks = element_blank(), legend.background
    = element_blank(),
  legend.key = element_blank(), panel.background = element_blank(),
  panel.border = element_blank(), strip.background = element_blank(),
  plot.background = element_blank(),
  legend.key.size = unit(1, "cm"), complete = TRUE, legend.text =
    element_text(size = 10)
) +
scale_color_manual(
  name = "", breaks = c("Wald", "Wilson", "Clopper_Pearson", "Wald
    corretto", "TRV", "Bayes", "Blaker"),
  values = c(
    "Wald" = "blue", "Wilson" = "red", "Clopper_Pearson" = "darkgreen",
    "Wald corretto" = "darkorange", "TRV" = "magenta", "Bayes" =
    "cornflowerblue", "Blaker" = "purple"
  )
)

```

```

))

#### Lunghezza attesa ####
expected_length_odds_Wald <- function(psi, n = 15, alpha = 0.05) {
  p_hat <- c(0:(n - 1)) / n
  th_hat <- p_hat / (1 - p_hat)
  z <- qnorm(alpha / 2)
  Exp_Wald_odds <- (th_hat - z * (1 + th_hat) * sqrt(th_hat / n)) - (th_hat
    + z * (1 + th_hat) * sqrt(th_hat / n))
  sum(dbinom(0:(n - 1), n, psi / (1 + psi)) * Exp_Wald_odds)
}
expected_length_odds_Clopper_Pearson <- function(psi, n = 15, alpha = 0.05)
{
  Exp_Clopper_Pearson_odds <- (qbeta(1 - alpha / 2, c(0:n) + 1, n - c(0:n))
    / (1 - qbeta(1 - alpha / 2, c(0:n) + 1, n - c(0:n)))) - (qbeta(alpha /
    2, c(0:n), n - c(0:n) + 1) / (1 - qbeta(alpha / 2, c(0:n), n - c(0:n)
    + 1)))
  Exp_Clopper_Pearson_odds <- Exp_Clopper_Pearson_odds[-(n + 1)]
  sum(dbinom(0:(n - 1), n, psi / (1 + psi)) * Exp_Clopper_Pearson_odds)
}
expected_length_odds_Wilson <- function(psi, n = 15, alpha = 0.05) {
  p_hat <- c(0:(n - 1)) / n
  th_hat <- p_hat / (1 - p_hat)
  z <- qnorm(alpha / 2)
  Exp_Wilson_odds <- (-p_hat^2 + p_hat + z^2 / (2 * n) - z * sqrt(z^2 / (4
    * n) - p_hat / n * (p_hat - 1))) / (p_hat - 1)^2 - (-p_hat^2 + p_hat +
    z^2 / (2 * n) + z * sqrt(z^2 / (4 * n) - p_hat / n * (p_hat - 1))) /
    (p_hat - 1)^2
  sum(dbinom(0:(n - 1), n, psi / (1 + psi)) * Exp_Wilson_odds)
}
expected_length_odds_Wald_2_2 <- function(psi, n = 15, alpha = 0.05) {
  p_hat <- c(0:(n - 1)) / n
  th_hat <- p_hat / (1 - p_hat)
  z <- qnorm(alpha / 2)
  n2 <- n + 4
  p_tilde <- c((0:n) + 2) / n2
  th_tilde <- p_tilde / (1 - p_tilde)

```

```

Exp_Adj_Wald_odds <- (th_tilde - z * (1 + th_tilde) * sqrt(th_tilde / n))
  - (th_tilde + z * (1 + th_tilde) * sqrt(th_tilde / n))
Exp_Adj_Wald_odds <- Exp_Adj_Wald_odds[-(n + 1)]
sum(dbinom(0:(n - 1), n, psi / (1 + psi)) * Exp_Adj_Wald_odds)
}
expected_length_odds_TRV <- function(psi, n = 15, alpha = 0.05) {
  p_hat <- c(0:(n - 1)) / n
  th_hat <- p_hat / (1 - p_hat)
  Exp_TRV_odds <- apply(as.matrix(p_hat[-1]), 1, function(x) {
    p_u <- uniroot(function(k) TRV_IC(k, n = n, alpha = alpha, x = x * n),
      interval = c(x, 1))$root
    p_l <- uniroot(function(l) TRV_IC(l, n = n, alpha = alpha, x = x * n),
      interval = c(0, x))$root
    p_u / (1 - p_u) - p_l / (1 - p_l)
  })
  sum(dbinom(0:(n - 1), n, psi / (1 + psi)) * c(0, Exp_TRV_odds))
}
expected_length_odds_Bayes <- function(psi, n = 15, alpha = 0.05) {
  Exp_Bayes_odds <- (qbetapr(1 - alpha / 2, c(0:n) + 1 / 2, n - c(0:n) + 1
    / 2)) - (qbetapr(alpha / 2, c(0:n) + 1 / 2, n - c(0:n) + 1 / 2))
  Exp_Bayes_odds <- Exp_Bayes_odds[-(n + 1)]
  sum(dbinom(0:(n - 1), n, psi / (1 + psi)) * Exp_Bayes_odds)
}
expected_length_odds_Blaker <- function(psi, n = 15, alpha = 0.05) {
  p_hat <- c(0:(n - 1)) / n
  Exp_Blaker_odds <- apply(as.matrix(p_hat[-1]), 1, function(x) {
    p_u <- uniroot(function(k) IC_Blaker(k, n = n, alpha = alpha, x = x *
      n), interval = c(x, 1))$root
    p_l <- uniroot(function(l) IC_Blaker(l, n = n, alpha = alpha, x = x *
      n), interval = c(0, x))$root
    p_u / (1 - p_u) - p_l / (1 - p_l)
  })
  p_Blaker_u <- uniroot(function(k) IC_Blaker(k, n = n, alpha = alpha, x =
    0), interval = c(0, 1))$root

  Exp_Blaker_odds <- c(
    p_Blaker_u / (1 - p_Blaker_u) - 0,
    Exp_Blaker_odds
  )
}

```

```

)
  sum(dbinom(0:(n - 1), n, psi / (1 + psi)) * Exp_Blaker_odds)
}

expected_length_odds_Wald <- Vectorize(expected_length_odds_Wald,
  vectorize.args = "psi")
expected_length_odds_Clopper_Pearson <-
  Vectorize(expected_length_odds_Clopper_Pearson, vectorize.args = "psi")
expected_length_odds_Wilson <- Vectorize(expected_length_odds_Wilson,
  vectorize.args = "psi")
expected_length_odds_Wald_2_2 <- Vectorize(expected_length_odds_Wald_2_2,
  vectorize.args = "psi")
expected_length_odds_TRV <- Vectorize(expected_length_odds_TRV,
  vectorize.args = "psi")
expected_length_odds_Bayes <- Vectorize(expected_length_odds_Bayes,
  vectorize.args = "psi")
expected_length_odds_Blaker <- Vectorize(expected_length_odds_Blaker,
  vectorize.args = "psi")

#n=5
(plot_exp_length_odds_n5 <- ggplot() +
  geom_function(fun = expected_length_odds_Wald, args = list(n = 5), data =
    data.frame(x = c(0, 1), fun_name = "Approssimati"), aes(linetype =
    "Wald", col = "Wald")) +
  geom_function(fun = expected_length_odds_Clopper_Pearson, args = list(n =
    5), data = data.frame(x = c(0, 1), fun_name = "Esatti"), aes(linetype
    = "Clopper_Pearson", col = "Clopper_Pearson"), linewidth = 1) +
  geom_function(fun = expected_length_odds_Wilson, args = list(n = 5), data
    = data.frame(x = c(0, 1), fun_name = "Approssimati"), aes(linetype =
    "Wilson", col = "Wilson"), linewidth = 1) +
  geom_function(fun = expected_length_odds_Wald_2_2, args = list(n = 5),
    data = data.frame(x = c(0, 1), fun_name = "Approssimati"),
    aes(linetype = "Wald corretto", col = "Wald corretto"), linewidth = 1)
  +
  geom_function(fun = expected_length_odds_TRV, args = list(n = 5), data =
    data.frame(x = c(0, 1), fun_name = "Approssimati"), aes(linetype =
    "TRV", col = "TRV"), linewidth = 1) +

```

```

geom_function(fun = expected_length_odds_Bayes, args = list(n = 5), data
  = data.frame(x = c(0, 1), fun_name = "Approssimati"), aes(linetype =
    "Bayes", col = "Bayes"), linewidth = 1) +
geom_function(fun = expected_length_odds_Blaker, args = list(n = 5), data
  = data.frame(x = c(0, 1), fun_name = "Esatti"), aes(linetype =
    "Blaker", col = "Blaker"), linewidth = 1) +
xlab(expression(paste(psi))) +
xlim(c(0, 20)) +
ylab("") +
facet_wrap(~fun_name) +
scale_linetype_manual(
  name = "", breaks = c("Wald", "Wilson", "Clopper_Pearson", "Wald
    corretto", "TRV", "Bayes", "Blaker"),
  values = c("Wald" = 1, "Wilson" = 4, "Clopper_Pearson" = 3, "Wald
    corretto" = 5, "TRV" = 6, "Bayes" = 7, "Blaker" = 8)
) +
theme(
  legend.position = "top", axis.ticks = element_blank(), legend.background
    = element_blank(),
  legend.key = element_blank(), panel.background = element_blank(),
  panel.border = element_blank(), strip.background = element_blank(),
  plot.background = element_blank(),
  legend.key.size = unit(1, "cm"), complete = TRUE, legend.text =
    element_text(size = 10)
) +
scale_color_manual(
  name = "", breaks = c("Wald", "Wilson", "Clopper_Pearson", "Wald
    corretto", "TRV", "Bayes", "Blaker"),
  values = c(
    "Wald" = "blue", "Wilson" = "red", "Clopper_Pearson" = "darkgreen",
    "Wald corretto" = "darkorange", "TRV" = "magenta",
    "Bayes" = "cornflowerblue", "Blaker" = "purple"
  )
)
)
# n=100
(plot_exp_length_odds_n100 <- ggplot() +

```

```

geom_function(fun = expected_length_odds_Wald, args = list(n = 100), data
  = data.frame(x = c(0, 1), fun_name = "Approssimati"), aes(linetype =
  "Wald", col = "Wald")) +
geom_function(fun = expected_length_odds_Clopper_Pearson, args = list(n =
  100), data = data.frame(x = c(0, 1), fun_name = "Esatti"),
  aes(linetype = "Clopper_Pearson", col = "Clopper_Pearson"), linewidth
  = 1) +
geom_function(fun = expected_length_odds_Wilson, args = list(n = 100),
  data = data.frame(x = c(0, 1), fun_name = "Approssimati"),
  aes(linetype = "Wilson", col = "Wilson"), linewidth = 1) +
geom_function(fun = expected_length_odds_Wald_2_2, args = list(n = 100),
  data = data.frame(x = c(0, 1), fun_name = "Approssimati"),
  aes(linetype = "Wald corretto", col = "Wald corretto"), linewidth = 1)
+
geom_function(fun = expected_length_odds_TRV, args = list(n = 100), data
  = data.frame(x = c(0, 1), fun_name = "Approssimati"), aes(linetype =
  "TRV", col = "TRV"), linewidth = 1) +
geom_function(fun = expected_length_odds_Bayes, args = list(n = 100),
  data = data.frame(x = c(0, 1), fun_name = "Approssimati"),
  aes(linetype = "Bayes", col = "Bayes"), linewidth = 1) +
geom_function(fun = expected_length_odds_Blaker, args = list(n = 100),
  data = data.frame(x = c(0, 1), fun_name = "Esatti"), aes(linetype =
  "Blaker", col = "Blaker"), linewidth = 1) +
xlab(expression(paste(psi))) +
xlim(c(0, 20)) +
ylab("") +
facet_wrap(~fun_name) +
scale_linetype_manual(
  name = "", breaks = c("Wald", "Wilson", "Clopper_Pearson", "Wald
  corretto", "TRV", "Bayes", "Blaker"),
  values = c("Wald" = 1, "Wilson" = 4, "Clopper_Pearson" = 3, "Wald
  corretto" = 5, "TRV" = 6, "Bayes" = 7, "Blaker" = 8)
) +
theme(
  legend.position = "top", axis.ticks = element_blank(), legend.background
  = element_blank(),
  legend.key = element_blank(), panel.background = element_blank(),
  panel.border = element_blank(), strip.background = element_blank(),

```

```

    plot.background = element_blank(),
    legend.key.size = unit(1, "cm"), complete = TRUE, legend.text =
      element_text(size = 10)
  ) +
  scale_color_manual(
    name = "", breaks = c("Wald", "Wilson", "Clopper_Pearson", "Wald
      corretto", "TRV", "Bayes", "Blaker"),
    values = c(
      "Wald" = "blue", "Wilson" = "red", "Clopper_Pearson" = "darkgreen",
      "Wald corretto" = "darkorange", "TRV" = "magenta",
      "Bayes" = "cornflowerblue", "Blaker" = "purple"
    )
  )
))

#### Log-quote ####

logodds_cov_Wald <- function(p, n, alpha = 0.05) {
  th_hat <- log(c(0:n) / n / (1 - c(0:n) / n))
  th.true <- log(p / (1 - p))
  z <- qnorm(alpha / 2)
  I_Wald_logodds <- c(rep(th.true, n + 1) >= th_hat + z * (1 + exp(th_hat))
    / (sqrt(n * exp(th_hat))) &
    rep(th.true, n + 1) <= th_hat - z * (1 + exp(th_hat)) / (sqrt(n *
      exp(th_hat))))
  I_Wald_logodds[c(1, n + 1)] <- T
  k_Wald_logodds <- which(I_Wald_logodds == 1) - 1
  sum(dbinom(k_Wald_logodds, n, p))
}

logodds_cov_Wald <- Vectorize(logodds_cov_Wald, vectorize.args = "p")

logodds_cov_Wald_corretto <- function(p, n, alpha = 0.05) {
  p_tilde <- (c(0:n) + 2) / (n + 4)
  th_tilde <- log(p_tilde / (1 - p_tilde))
  th.true <- log(p / (1 - p))
  z <- qnorm(alpha / 2)
  I_Wald_corretto_logodds <- c(rep(th.true, n + 1) >= th_tilde + z * (1 +
    exp(th_tilde)) / (sqrt(n * exp(th_tilde))) &

```

```

    rep(th.true, n + 1) <- th_tilde - z * (1 + exp(th_tilde)) / (sqrt(n *
      exp(th_tilde)))
  I_Wald_corretto_logodds[c(1, n + 1)] <- T
  k_Wald_corretto_logodds <- which(I_Wald_corretto_logodds == 1) - 1
  sum(dbinom(k_Wald_corretto_logodds, n, p))
}
logodds_cov_Wald_corretto <- Vectorize(logodds_cov_Wald_corretto,
  vectorize.args = "p")

#### Probabilita media di copertura ####

dati_logodds_Wald <- apply(as.matrix(nnn), 1, function(n) {
  integrate(function(p) logodds_cov_Wald(p, n = n, alpha = 0.05),
    subdivisions = 2000, upper = 1, lower = 0)$value
})
dati_logodds_Clopper_Pearson <- dati_Clopper_Pearson
dati_logodds_Wilson <- dati_Wilson
dati_logodds_Wald_corretto <- apply(as.matrix(nnn), 1, function(n) {
  integrate(function(p) logodds_cov_Wald_corretto(p, n = n, alpha = 0.05),
    subdivisions = 2000, upper = 1, lower = 0)$value
})
dati_logodds_Bayesiano <- dati_Bayesiano
dati_logodds_Blaker <- dati_Blaker
dati_logodds_TRV <- dati_TRV

dati_logodds_unif_wide <- as_tibble(cbind(dati_logodds_Wald,
  dati_logodds_Clopper_Pearson, dati_logodds_Wilson,
  dati_logodds_Wald_corretto, dati_logodds_Bayesiano, dati_logodds_TRV,
  dati_logodds_Blaker))
dati_logodds_unif <- dati_logodds_unif_wide %>%
  pivot_longer(cols = starts_with("dati"), cols_vary = "slowest", names_to
    = "Metodo", names_prefix = "dati_logodds_", values_to =
    "logodds_cov_bar") %>%
  mutate(Priori = "Uniforme")

a <- 35 / 10
b <- 315 / 10

```

```

dati_beta_logodds_Wald <- apply(as.matrix(seq(5, 100, by = 9)), 1,
  function(n) {
    integrate(function(p) dbeta(p, a, b) * logodds_cov_Wald(p, n = n, alpha =
      0.05), subdivisions = 2000, upper = 1, lower = 0)$value
  })
dati_beta_logodds_Clopper_Pearson <- dati_beta_Clopper_Pearson

dati_beta_logodds_Wilson <- dati_beta_Wilson
dati_beta_logodds_Wald_corretto <- apply(as.matrix(seq(5, 100, by = 9)), 1,
  function(n) {
    integrate(function(p) dbeta(p, a, b) * logodds_cov_Wald_corretto(p, n =
      n, alpha = 0.05), subdivisions = 2000, upper = 1, lower = 0)$value
  })
dati_beta_logodds_Bayesiano <- dati_beta_Bayesiano
dati_beta_logodds_Blaker <- dati_beta_Blaker
dati_beta_logodds_TRV <- dati_beta_TRV

dati_beta_logodds_wide <- as_tibble(cbind(
  dati_beta_logodds_Wald, dati_beta_logodds_Clopper_Pearson,
  dati_beta_logodds_Wilson,
  dati_beta_logodds_Wald_corretto, dati_beta_logodds_Bayesiano,
  dati_beta_logodds_TRV, dati_beta_logodds_Blaker
))
dati_beta_logodds <- dati_beta_logodds_wide %>%
  pivot_longer(cols = starts_with("dati"), cols_vary = "slowest", names_to
    = "Metodo", names_prefix = "dati_beta_logodds_", values_to =
    "logodds_cov_bar") %>%
  mutate(Priori = "zBeta")
dati_logodds_cov_bar <- dati_logodds_unif %>%
  add_row(dati_beta_logodds) %>%
  mutate(Priori = as.factor(Priori))

nomipriorilogodds <- c(zBeta = "log(Beta'(0.35,3.15))", Uniforme =
  "log(Beta'(1,1))")

(plot_cov_bar_logodds <- ggplot(dati_logodds_cov_bar, aes(x = rep(nnn, 14),
  y = logodds_cov_bar, col = Metodo)) +

```

```

geom_point(size = 1.7) +
geom_line() +
geom_hline(yintercept = 0.95, linetype = 2, alpha = 0.7) +
theme(
  legend.position = "top", axis.ticks = element_blank(), legend.background
    = element_blank(),
  legend.key = element_blank(), panel.background = element_blank(),
  panel.border = element_blank(), strip.background = element_blank(),
  plot.background = element_blank(), complete = TRUE, plot.caption =
    element_text(hjust = 0.5)
) +
xlab("Sample size") +
ylab("") +
scale_color_manual(
  name = "", breaks = c("Wald", "Wilson", "Clopper_Pearson",
    "Wald_corretto", "TRV", "Bayesiano", "Blaker"),
  values = c(
    "Wald" = "blue", "Wilson" = "red", "Clopper_Pearson" = "darkgreen",
    "Wald_corretto" = "darkorange", "TRV" = "magenta",
    "Bayesiano" = "cornflowerblue", "Blaker" = "purple"
  )
) +
facet_grid(~Priori, labeller = labeller(Priori = nomipriorilogodds))

```

#### Probabilita di copertura ####

```

(plot_cov_logodds_n5 <- ggplot() +
  geom_function(n = 5000, fun = function(th) logodds_cov_Wald(p = exp(th) /
    (1 + exp(th)), n = 5, alpha = 0.05), data = data.frame(x = c(0, 1),
    fun_name = "Wald"), aes(col = "Wald")) +
  geom_function(n = 5000, fun = function(th) logodds_cov_Wald_corretto(p =
    exp(th) / (1 + exp(th)), n = 5, alpha = 0.05), data = data.frame(x =
    c(0, 1), fun_name = "Wald Corretto"), aes(col = "Wald corretto")) +
  geom_function(fun = function(th) cov_Wilson(p = exp(th) / (1 + exp(th)),
    n = 5, alpha = 0.05), data = data.frame(x = c(0, 1), fun_name =
    "Wilson"), aes(col = "Wilson")) +

```

```

geom_function(fun = function(th) cov_Clopper_Pearson(p = exp(th) / (1 +
  exp(th)), n = 5, alpha = 0.05), data = data.frame(x = c(0, 1),
  fun_name = "Clopper-Pearson"), aes(col = "Clopper_Pearson")) + # , ) +
geom_function(fun = function(th) cov_Bayesiano(p = exp(th) / (1 +
  exp(th)), n = 5, alpha = 0.05), data = data.frame(x = c(0, 1),
  fun_name = "Bayes"), aes(col = "Bayes")) +
geom_function(fun = function(th) cov_prob_TRV(p = exp(th) / (1 +
  exp(th)), n = 5, alpha = 0.05), data = data.frame(x = c(0, 1),
  fun_name = "TRV"), aes(col = "TRV")) +
geom_function(fun = function(th) cov_Blaker(p = exp(th) / (1 + exp(th)),
  n = 5, alpha = 0.05), data = data.frame(x = c(0, 1), fun_name =
  "Blaker"), aes(col = "Blaker")) +
xlim(c(-5, 5)) +
ylim(c(0, 1)) +
xlab(expression(paste(theta))) +
ylab("n=5") +
geom_hline(yintercept = 0.95, linetype = 2, alpha = 0.7) +
facet_wrap(~fun_name, ncol = 2) +
theme(
  legend.position = "top", axis.ticks = element_blank(), legend.background
    = element_blank(),
  legend.key = element_blank(), panel.background = element_blank(),
  panel.border = element_blank(), strip.background = element_blank(),
  plot.background = element_blank(),
  legend.key.size = unit(1, "cm"), complete = TRUE, legend.text =
    element_text(size = 10)
) +
scale_color_manual(
  name = "", breaks = c("Wald", "Wilson", "Clopper_Pearson", "Wald
  corretto", "TRV", "Bayes", "Blaker"),
  values = c(
    "Wald" = "blue", "Wilson" = "red", "Clopper_Pearson" = "darkgreen",
    "Wald corretto" = "darkorange", "TRV" = "magenta", "Bayes" =
    "cornflowerblue", "Blaker" = "purple"
  )
)
)
# n=100

```

```

(plot_cov_logodds_n100 <- ggplot() +
  geom_function(fun = function(th) logodds_cov_Wald(p = exp(th) / (1 +
    exp(th)), n = 100, alpha = 0.05), data = data.frame(x = c(0, 1),
    fun_name = "Wald"), aes(col = "Wald")) +
  geom_function(fun = function(th) logodds_cov_Wald_corretto(p = exp(th) /
    (1 + exp(th)), n = 100, alpha = 0.05), data = data.frame(x = c(0, 1),
    fun_name = "Wald Corretto"), aes(col = "Wald corretto")) +
  geom_function(fun = function(th) cov_Wilson(p = exp(th) / (1 + exp(th)),
    n = 100, alpha = 0.05), data = data.frame(x = c(0, 1), fun_name =
    "Wilson"), aes(col = "Wilson")) +
  geom_function(fun = function(th) odds_cov_Clopper_Pearson(p = exp(th) /
    (1 + exp(th)), n = 100, alpha = 0.05), data = data.frame(x = c(0, 1),
    fun_name = "Clopper-Pearson"), aes(col = "Clopper_Pearson")) + # , ) +
  geom_function(fun = function(th) odds_cov_Bayesiano(p = exp(th) / (1 +
    exp(th)), n = 100, alpha = 0.05), data = data.frame(x = c(0, 1),
    fun_name = "Bayes"), aes(col = "Bayes")) +
  geom_function(fun = function(th) cov_prob_TRV(p = exp(th) / (1 +
    exp(th)), n = 100, alpha = 0.05), data = data.frame(x = c(0, 1),
    fun_name = "TRV"), aes(col = "TRV")) +
  geom_function(fun = function(th) cov_Blaker(p = exp(th) / (1 + exp(th)),
    n = 100, alpha = 0.05), data = data.frame(x = c(0, 1), fun_name =
    "Blaker"), aes(col = "Blaker")) +
  xlim(c(-5, 5)) +
  ylim(c(0, 1)) +
  xlab(expression(paste(theta))) +
  ylab("n=100") +
  geom_hline(yintercept = 0.95, linetype = 2, alpha = 0.7) +
  facet_wrap(~fun_name, ncol = 2) +
  theme(
    legend.position = "top", axis.ticks = element_blank(), legend.background
    = element_blank(),
    legend.key = element_blank(), panel.background = element_blank(),
    panel.border = element_blank(), strip.background = element_blank(),
    plot.background = element_blank(),
    legend.key.size = unit(1, "cm"), complete = TRUE, legend.text =
    element_text(size = 10)
  ) +
  scale_color_manual(

```

```

name = "", breaks = c("Wald", "Wilson", "Clopper_Pearson", "Wald
  corretto", "TRV", "Bayes", "Blaker"),
values = c(
  "Wald" = "blue", "Wilson" = "red", "Clopper_Pearson" = "darkgreen",
  "Wald corretto" = "darkorange", "TRV" = "magenta", "Bayes" =
    "cornflowerblue", "Blaker" = "purple"
)
))

```

#### Lunghezza attesa ####

```

expected_length_logodds_Clopper_Pearson <- function(p, n, alpha = 0.05) {
  p_hat <- c(0:(n - 1)) / n
  z <- qnorm(alpha / 2)
  length_Clopper_Pearson <- log(qbeta(1 - alpha / 2, c(1:(n - 1)) + 1, n -
    c(1:(n - 1)))) / (1 - qbeta(1 - alpha / 2, c(1:(n - 1)) + 1, n - c(1:(n
    - 1)))) - log(qbeta(alpha / 2, c(1:(n - 1)), n - c(1:(n - 1)) + 1) /
    (1 - qbeta(alpha / 2, c(1:(n - 1)), n - c(1:(n - 1)) + 1)))
  sum(dbinom(c(1:(n - 1)), n, p) * length_Clopper_Pearson)
}

expected_length_logodds_Clopper_Pearson <-
  Vectorize(expected_length_logodds_Clopper_Pearson, vectorize.args = "p")

expected_length_logodds_Blaker <- function(p, n = 15, alpha = 0.05) {
  p_hat <- c(0:(n - 1)) / n
  th_hat <- log(p_hat / (1 - p_hat))
  th.true <- log(p / (1 - p))
  z <- qnorm(alpha / 2)
  length_Blaker <- apply(as.matrix(c(1:(n - 1))), 1, function(x) {
    log((uniroot(function(p) IC_Blaker(p, x, n, .05), interval = c(x / n,
      1))$root) / (1 - (uniroot(function(p) IC_Blaker(p, x, n, .05),
      interval = c(x / n, 1))$root))) - log((uniroot(function(p)
      IC_Blaker(p, x, n, .05), interval = c(0.0001, x / n))$root) / (1 -
      (uniroot(function(p) IC_Blaker(p, x, n, .05), interval = c(0.00001,
      x / n))$root)))
  })

  sum(dbinom(c(1:(n - 1)), n, p) * length_Blaker)
}

```

```

expected_length_logodds_Blaker <- Vectorize(expected_length_logodds_Blaker,
  vectorize.args = "p")

expected_length_logodds_TRV <- function(p, n = 15, alpha = 0.05) {
  p_hat <- c(0:(n - 1)) / n
  th_hat <- log(p_hat / (1 - p_hat))
  th.true <- log(p / (1 - p))
  z <- qnorm(alpha / 2)
  length_TRV <- apply(as.matrix(p_hat[-1]), 1, function(x) {
    p_u <- uniroot(function(k) TRV_IC(k, n = n, alpha = alpha, x = x * n),
      interval = c(x, 1))$root
    p_l <- uniroot(function(l) TRV_IC(l, n = n, alpha = alpha, x = x * n),
      interval = c(0, x))$root
    log(p_u / (1 - p_u)) - log(p_l / (1 - p_l))
  })

  sum(dbinom(c(1:(n - 1)), n, p) * length_TRV)
}

expected_length_logodds_TRV <- Vectorize(expected_length_logodds_TRV,
  vectorize.args = "p")

expected_length_logodds_Wald <- function(p, n = 15, alpha = 0.05) {
  p_hat <- c(1:(n - 1)) / n
  th_hat <- log(p_hat / (1 - p_hat))
  z <- qnorm(alpha / 2)
  length_Wald <- (th_hat - z * (1 + exp(th_hat)) / (sqrt(n * exp(th_hat))))
  - (th_hat + z * (1 + exp(th_hat)) / (sqrt(n * exp(th_hat))))
  sum(dbinom(c(1:(n - 1)), n, p) * length_Wald)
}

expected_length_logodds_Wald <- Vectorize(expected_length_logodds_Wald,
  vectorize.args = "p")

expected_length_logodds_Wald_corretto <- function(p, n = 15, alpha = 0.05) {
  p_tilde <- (c(1:(n - 1)) + 2) / (n + 4)
  th_tilde <- log(p_tilde / (1 - p_tilde))
  z <- qnorm(alpha / 2)
  length_Wald_corretto <- (th_tilde - z * (1 + exp(th_tilde)) / (sqrt(n *
    exp(th_tilde)))) - (th_tilde + z * (1 + exp(th_tilde)) / (sqrt(n *
    exp(th_tilde))))
  sum(dbinom(c(1:(n - 1)), n, p) * length_Wald_corretto)
}

```

```

}
expected_length_logodds_Wald_corretto <-
  Vectorize(expected_length_logodds_Wald_corretto, vectorize.args = "p")
expected_length_logodds_Wilson <- function(p, n = 15, alpha = 0.05) {
  p_hat <- c(1:(n - 1)) / n
  th_hat <- log(p_hat / (1 - p_hat))
  z <- qnorm(alpha / 2)
  length_Wilson <- (log(exp(th_hat) + (z^2 / 2 - z * sqrt(z^2 / 4 + (n *
    exp(th_hat)) / (1 + exp(th_hat))^2)) / (n / (1 + exp(th_hat))^2))) -
    (log(exp(th_hat) + (z^2 / 2 + z * sqrt(z^2 / 4 + (n * exp(th_hat)) /
    (1 + exp(th_hat))^2)) / (n / (1 + exp(th_hat))^2)))
  sum(dbinom(c(1:(n - 1)), n, p) * length_Wilson)
}
expected_length_logodds_Wilson <- Vectorize(expected_length_logodds_Wilson,
  vectorize.args = "p")
expected_length_logodds_Bayesiano <- function(p, n = 15, alpha = 0.05) {
  length_Bayesiano <- (-log((qbeta(0.975, c(1:(n - 1)) + 0.5, n - c(1:(n -
    1)) + 0.5)^(-1)) - 1)) - (-log((qbeta(0.025, c(1:(n - 1)) + 0.5, n -
    c(1:(n - 1)) + 0.5)^(-1)) - 1))
  sum(dbinom(c(1:(n - 1)), n, p) * length_Bayesiano)
}
expected_length_logodds_Bayesiano <-
  Vectorize(expected_length_logodds_Bayesiano, vectorize.args = "p")

(plot_exp_length_logodds_n5 <- ggplot() +
  geom_function(fun = function(theta)
    expected_length_logodds_Wald(exp(theta) / (1 + exp(theta)), n = 5,
    .05), data = data.frame(x = c(0, 1), fun_name = "Approssimati"),
    aes(linetype = "Wald", col = "Wald")) +
  geom_function(fun = function(theta)
    expected_length_logodds_Clopper_Pearson(exp(theta) / (1 + exp(theta)),
    n = 5, .05), data = data.frame(x = c(0, 1), fun_name = "Esatti"),
    aes(linetype = "Clopper_Pearson", col = "Clopper_Pearson"), linewidth
    = 1) +
  geom_function(fun = function(theta)
    expected_length_logodds_Wilson(exp(theta) / (1 + exp(theta)), n = 5,
    .05), data = data.frame(x = c(0, 1), fun_name = "Approssimati"),
    aes(linetype = "Wilson", col = "Wilson"), linewidth = 1) +

```

```

geom_function(fun = function(theta)
  expected_length_logodds_Wald_corretto(exp(theta) / (1 + exp(theta)), n
= 5, .05), data = data.frame(x = c(0, 1), fun_name = "Approssimati"),
  aes(linetype = "Wald corretto", col = "Wald corretto"), linewidth = 1)
+
geom_function(fun = function(theta)
  expected_length_logodds_TRV(exp(theta) / (1 + exp(theta)), n = 5,
  .05), data = data.frame(x = c(0, 1), fun_name = "Approssimati"),
  aes(linetype = "TRV", col = "TRV"), linewidth = 1) +
geom_function(fun = function(theta)
  expected_length_logodds_Bayesiano(exp(theta) / (1 + exp(theta)), n =
  5, .05), data = data.frame(x = c(0, 1), fun_name = "Approssimati"),
  aes(linetype = "Bayes", col = "Bayes"), linewidth = 1) +
geom_function(fun = function(theta)
  expected_length_logodds_Blaker(exp(theta) / (1 + exp(theta)), n = 5,
  .05), data = data.frame(x = c(0, 1), fun_name = "Esatti"),
  aes(linetype = "Blaker", col = "Blaker"), linewidth = 1) +
theme_minimal() +
facet_wrap(~fun_name) +
xlim(c(-5, 5)) +
xlab(expression(paste(psi))) +
ylab("") +
scale_linetype_manual(
  name = "", breaks = c("Wald", "Wilson", "Clopper_Pearson", "Wald
  corretto", "TRV", "Bayes", "Blaker"),
  values = c("Wald" = 1, "Wilson" = 4, "Clopper_Pearson" = 3, "Wald
  corretto" = 5, "TRV" = 6, "Bayes" = 8, "Blaker" = 8)
) +
scale_color_manual(
  name = "", breaks = c("Wald", "Wilson", "Clopper_Pearson", "Wald
  corretto", "TRV", "Bayes", "Blaker"),
  values = c(
    "Wald" = "blue", "Wilson" = "red", "Clopper_Pearson" = "darkgreen",
    "Wald corretto" = "darkorange", "TRV" = "magenta",
    "Bayes" = "cornflowerblue", "Blaker" = "purple"
  )
) +
theme(

```

```

legend.position = "top", axis.ticks = element_blank(), legend.background
  = element_blank(),
legend.key = element_blank(), panel.background = element_blank(),
panel.border = element_blank(), strip.background = element_blank(),
plot.background = element_blank(),
legend.key.size = unit(1, "cm"), complete = TRUE, legend.text =
  element_text(size = 10)
) )

```

```
#n=100
```

```

(plot_exp_length_logodds_n100 <- ggplot() +
  geom_function(fun = function(theta)
    expected_length_logodds_Wald(exp(theta) / (1 + exp(theta)), n = 100,
    .05), data = data.frame(x = c(0, 1), fun_name = "Approssimati"),
    aes(linetype = "Wald", col = "Wald")) +
  geom_function(fun = function(theta)
    expected_length_logodds_Clopper_Pearson(exp(theta) / (1 + exp(theta)),
    n = 100, .05), data = data.frame(x = c(0, 1), fun_name = "Esatti"),
    aes(linetype = "Clopper_Pearson", col = "Clopper_Pearson"), linewidth
    = 1) +
  geom_function(fun = function(theta)
    expected_length_logodds_Wilson(exp(theta) / (1 + exp(theta)), n = 100,
    .05), data = data.frame(x = c(0, 1), fun_name = "Approssimati"),
    aes(linetype = "Wilson", col = "Wilson"), linewidth = 1) +
  geom_function(fun = function(theta)
    expected_length_logodds_Wald_corretto(exp(theta) / (1 + exp(theta)), n
    = 100, .05), data = data.frame(x = c(0, 1), fun_name =
    "Approssimati"), aes(linetype = "Wald corretto", col = "Wald
    corretto"), linewidth = 1) +
  geom_function(fun = function(theta)
    expected_length_logodds_TRV(exp(theta) / (1 + exp(theta)), n = 100,
    .05), data = data.frame(x = c(0, 1), fun_name = "Approssimati"),
    aes(linetype = "TRV", col = "TRV"), linewidth = 1) +
  geom_function(fun = function(theta)
    expected_length_logodds_Bayesiano(exp(theta) / (1 + exp(theta)), n =
    100, .05), data = data.frame(x = c(0, 1), fun_name = "Approssimati"),
    aes(linetype = "Bayes", col = "Bayes"), linewidth = 1) +

```

```

geom_function(fun = function(theta)
  expected_length_logodds_Blaker(exp(theta) / (1 + exp(theta)), n = 100,
    .05), data = data.frame(x = c(0, 1), fun_name = "Esatti"),
  aes(linetype = "Blaker", col = "Blaker"), linewidth = 1) +
theme_minimal() +
facet_wrap(~fun_name) +
xlim(c(-10, 10)) +
xlab(expression(paste(psi))) +
ylab("") +
scale_linetype_manual(
  name = "", breaks = c("Wald", "Wilson", "Clopper_Pearson", "Wald
    corretto", "TRV", "Bayes", "Blaker"),
  values = c("Wald" = 1, "Wilson" = 4, "Clopper_Pearson" = 3, "Wald
    corretto" = 5, "TRV" = 6, "Bayes" = 7, "Blaker" = 8)
) +
scale_color_manual(
  name = "", breaks = c("Wald", "Wilson", "Clopper_Pearson", "Wald
    corretto", "TRV", "Bayes", "Blaker"),
  values = c(
    "Wald" = "blue", "Wilson" = "red", "Clopper_Pearson" = "darkgreen",
    "Wald corretto" = "darkorange", "TRV" = "magenta",
    "Bayes" = "cornflowerblue", "Blaker" = "purple"
  )
) +
theme(
  legend.position = "top", axis.ticks = element_blank(),
  legend.background = element_blank(),
  legend.key = element_blank(), panel.background = element_blank(),
  panel.border = element_blank(), strip.background = element_blank(),
  plot.background = element_blank(),
  legend.key.size = unit(1, "cm"), complete = TRUE, legend.text =
    element_text(size = 10)
) )

```

---