Stima della significativita' di un segnale

Problema On/Off

Andrea Belfiore
DFNT Universita' di Pavia

Problema On/Off

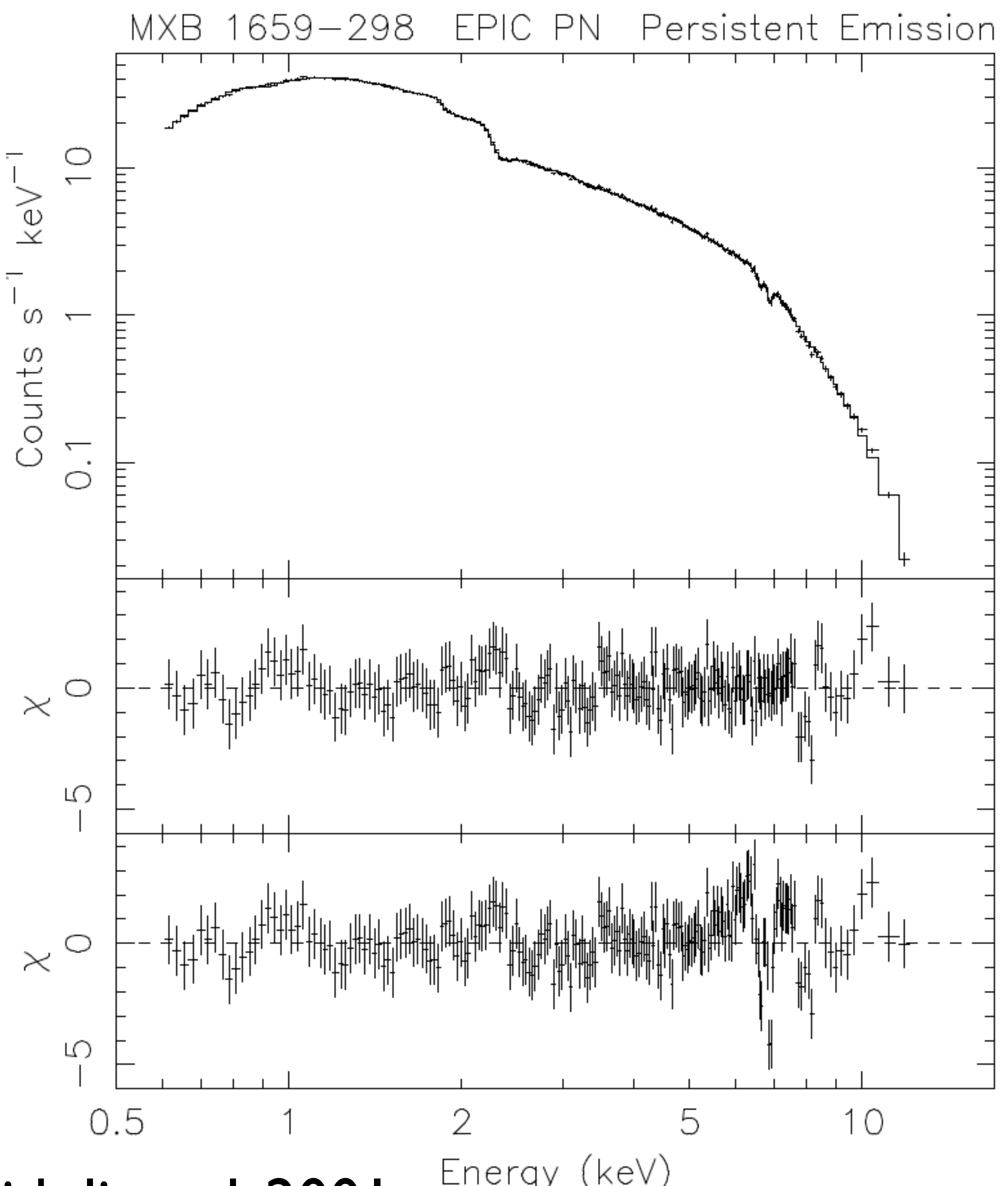
- Descrizione del problema
- Descrizione dei test MonteCarlo
- Soluzione analitica frequentista
- Soluzioni adottate in letteratura
- Confronto tra la significativita' dichiarata e quella ottenuta da simulazione

Descrizione del problema

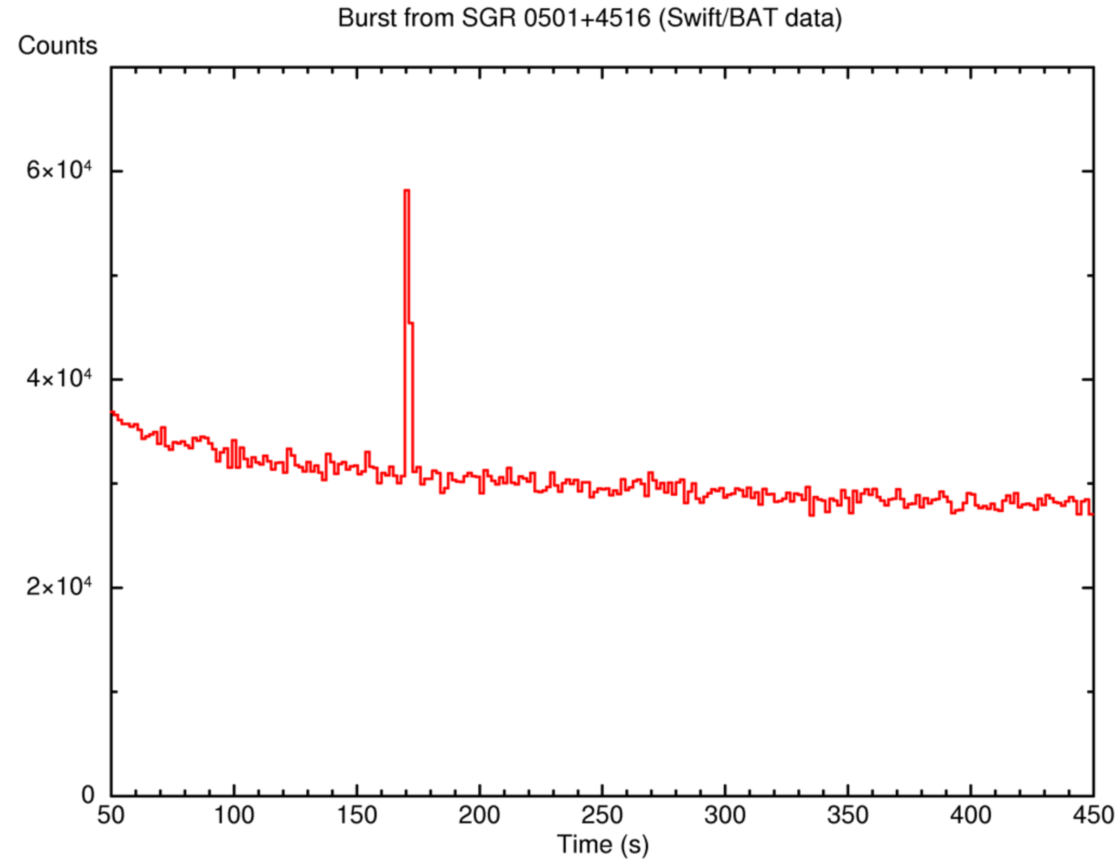
- Misura "on source": $\mu_{on}=\mu_s+\mu_b$
- Misura "off source": $\mu_{off}=\tau \cdot \mu_b$
- Presuppone che il rapporto tra i valori veri dei conteggi di fondo sia noto esattamente
- H_0: il segnale e' una fluttuazione del fondo
- Confronto il tasso di errori di I tipo e il livello di confidenza dichiarato dai test

Descrizione del problema

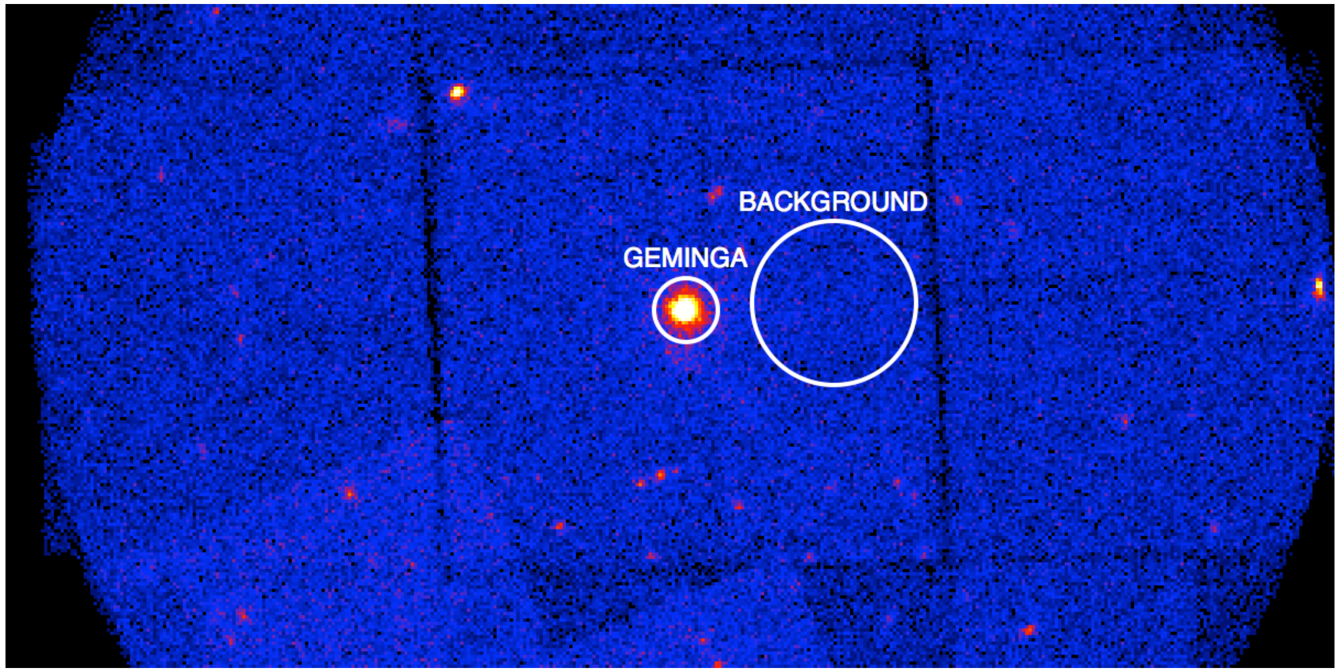
Varie applicazioni, tutte piu' complesse del modello



Sidoli et al. 2001



Grazie Gio!



Inferenza statistica

Test di ipotesi

1. Definire l'ipotesi puntuale H_0
2. Definire l'ipotesi alternativa H_1
3. Scegliere il livello di confidenza α
4. Confrontare con i risultati sperimentali
5. Stabilire se e' possibile rigettare H_0

Errore di I tipo: H_0 rigettata ma vera ($p_I=\alpha$)
Errore di II tipo: H_0 non rigettata ma falsa ($p_{II}=\beta$)

Inferenza statistica

livello di significativita' α :

probabilita' che, vera H_0, si ottengano valori tanto estremi da non verificare il test

potenza $1-\beta$:

probabilita' che, vera H_1, si ottengano valori tanto estremi da non verificare il test

p-value :

probabilita' che, vera H_0, si ottengano valori tanto o piu' estremi del valore misurato

parametro Z :

numero di deviazioni standard nel test gaussiano a una coda, equivalente in α

Descrizione dei test

- Genero i processi poissoniani N_{on} e N_{off}, sotto l'ipotesi H_0: $\mu_{on}=\mu_b$ e $\mu_{off}=\tau \cdot \mu_b$
- MonteCarlo con parametri μ_b e τ
- Stimo la significativita' del segnale per ogni realizzazione con vari metodi
- Calcolo la frequenza attesa e misurata di errori del I tipo per diversi valori di Z

Soluzione analitica frequentista

Przyborowski & Wilenski (1940)

- Scomposizione della probabilita' complessiva in una Poissoniana e una Binomiale
- Riformulazione di H_0 in termini di ρ e τ
- Stima della probabilita' di errore del I tipo per una distribuzione binomiale
- Estrazione dello Z-value dal p-value

Soluzione analitica frequentista

$$Pr(n_{on}, n_{off}|\mu_{on}, \mu_{off}) = \frac{\mu_{on}^{n_{on}}}{n_{on}!} e^{-\mu_{on}} \cdot \frac{\mu_{off}^{n_{off}}}{n_{off}!} e^{-\mu_{off}} =$$

$$= \frac{(\mu_{on} + \mu_{off})^{n_{on}+n_{off}}}{(n_{on} + n_{off})!} e^{-(\mu_{on}+\mu_{off})}.$$

$$\cdot \frac{(n_{on} + n_{off})!}{n_{on}! \cdot n_{off}!} \left(\frac{\mu_{on}}{\mu_{on} + \mu_{off}}\right)^{n_{on}} \left(\frac{\mu_{off}}{\mu_{on} + \mu_{off}}\right)^{n_{off}} =$$

$$= Pr(n_{on} + n_{off}|\mu_{on} + \mu_{off}) \cdot Pr(n_{on}|n_{on} + n_{off}, \varrho = \frac{\mu_{on}}{\mu_{on} + \mu_{off}})$$

Soluzione analitica frequentista

H_0: $\mu_{on}=\mu_{off}/\tau \Rightarrow \rho=1/(1+\tau)$

H_1: $\mu_{on}>\mu_{off}/\tau \Rightarrow \rho>1/(1+\tau)$

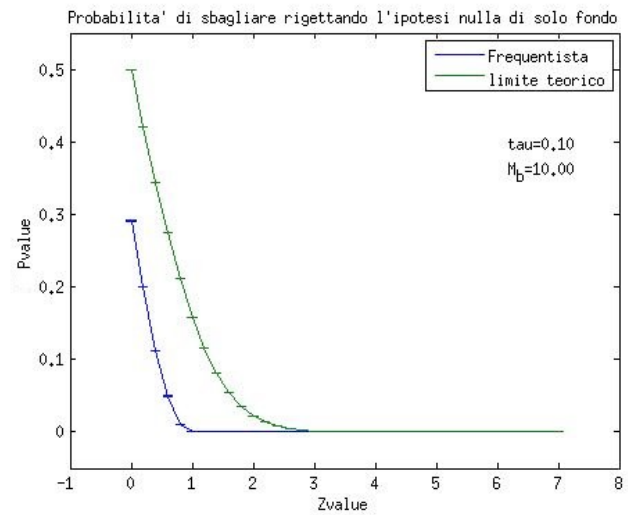
$$p = \sum_{n_{on}\leq i\leq n_{on}+n_{off}} Pr(i|n_{on}+n_{off}, \varrho = \frac{\mu_{on}}{\mu_{on}+\mu_{off}}) =$$

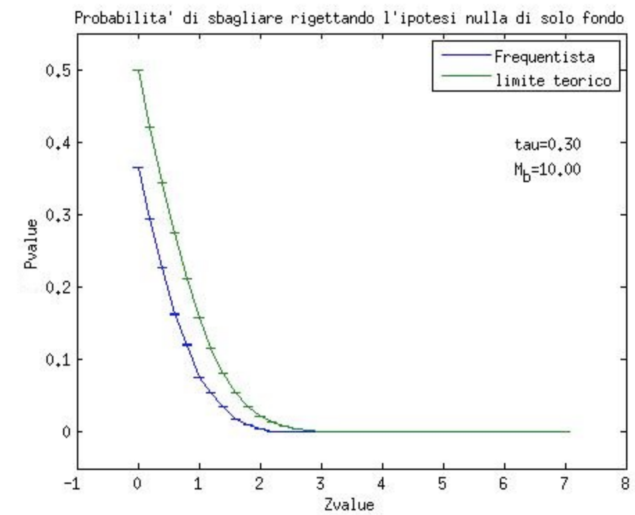
$$= B(\varrho, n_{on}, 1+n_{off})$$

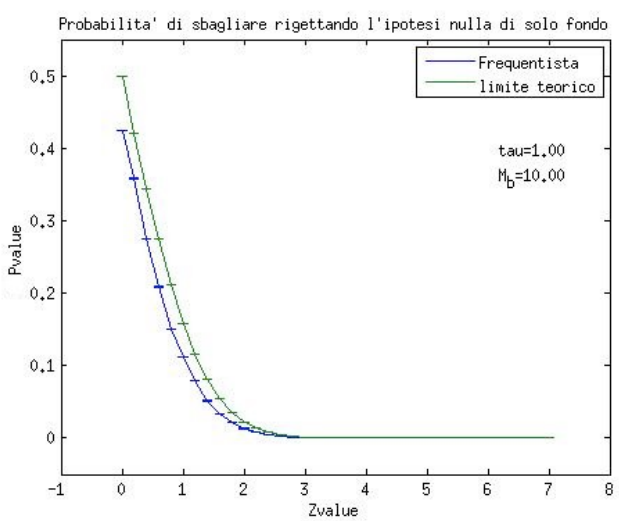
$$Z = \sqrt{2}erf^{-1}(1-2p)$$

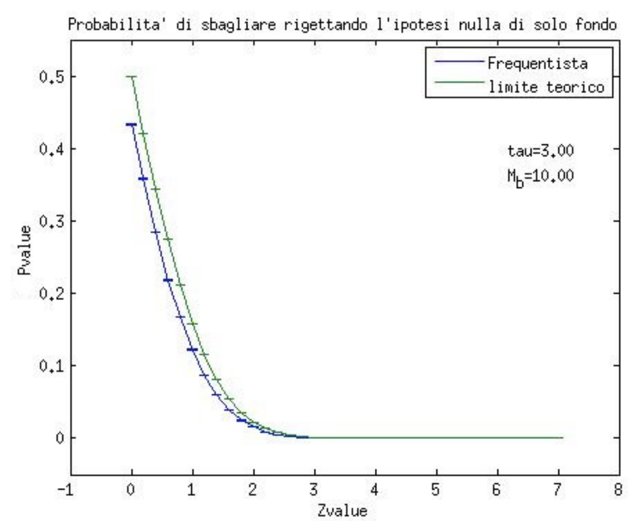
Soluzione analitica frequentista

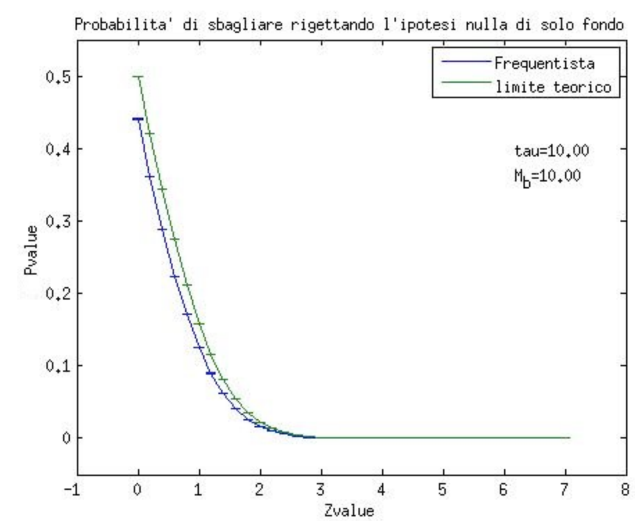
Fondo: μ_b=10
$\tau=\mu_{off}/\mu_{on}$: 0.1 0.3 1 3 10
Conservativo (sottostima Z)
Effetto piu' accentuato
quando τ e' basso





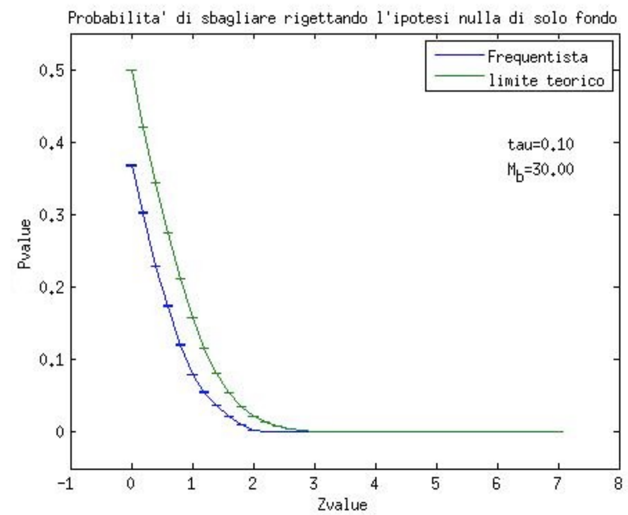


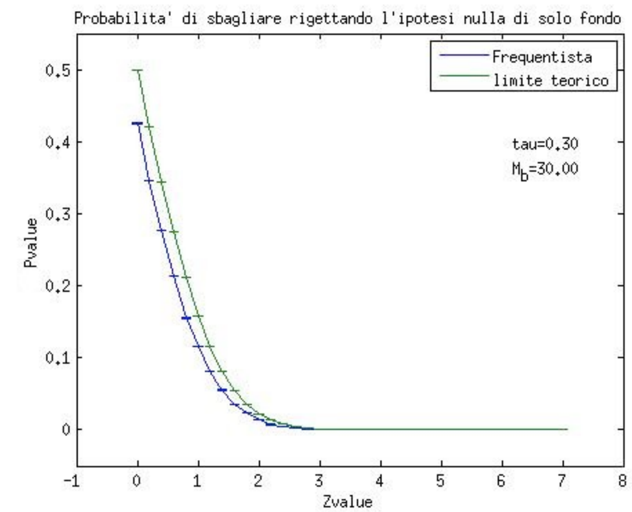


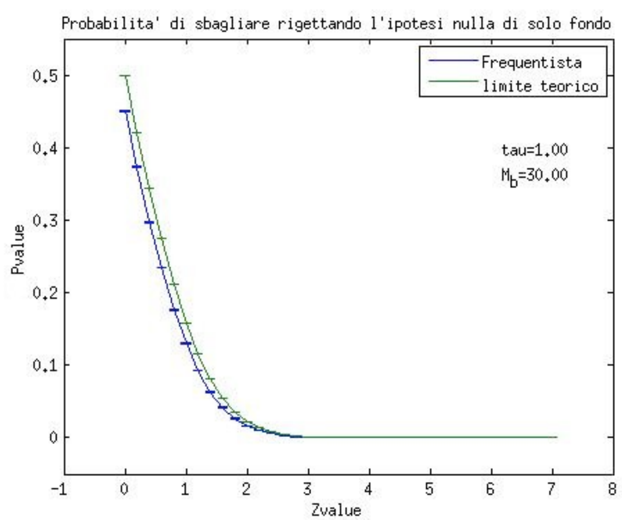


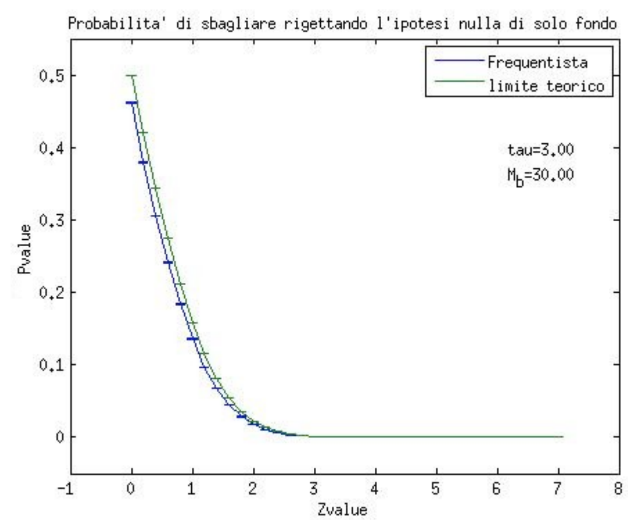
Soluzione analitica frequentista

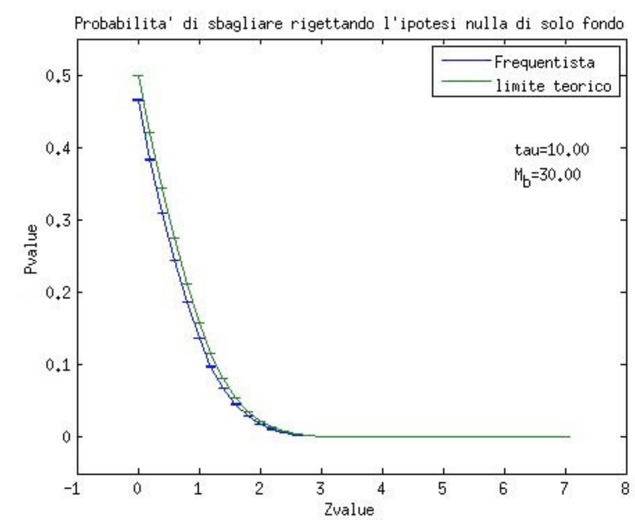
Fondo: μ_b=30
$\tau=\mu_{off}/\mu_{on}$: 0.1 0.3 1 3 10
Conservativo (sottostima Z)
Effetto piu' accentuato
quando τ e' basso





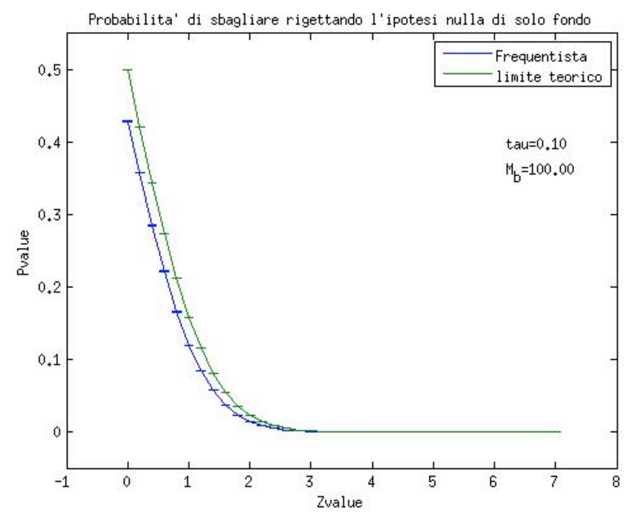


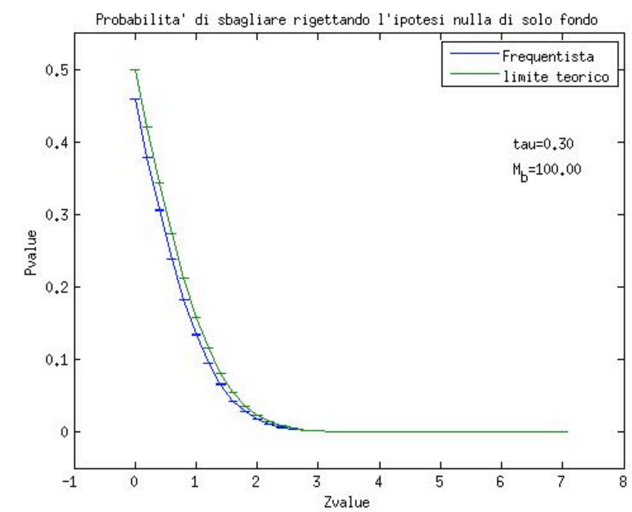


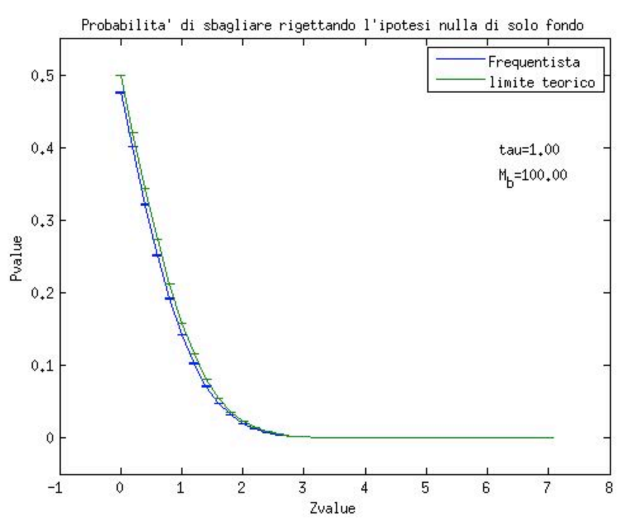


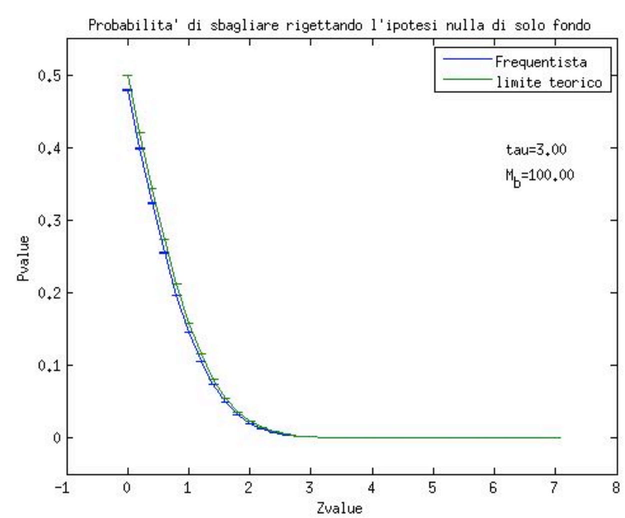
Soluzione analitica frequentista

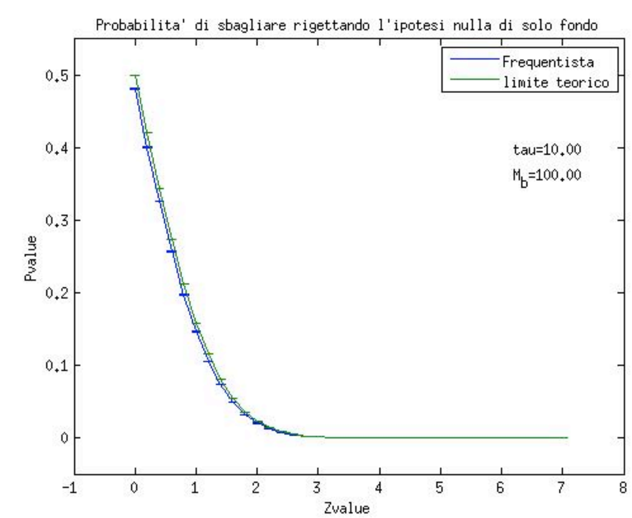
Fondo: μ_b=100
$\tau=\mu_{off}/\mu_{on}$: 0.1 0.3 1 3 10
Conservativo (sottostima Z)
Al crescere di μ_{off} e di μ_{on}
converge al limite teorico











Soluzioni adottate in letteratura

- Rapporto Segnale/Rumore gaussiano
- Rapporto Segnale/Rumore poissoniano
- Rapporto Segnale/Rumore binomiale
- Rapporto di verosimiglianza
- Trasformata a stabilizzazione di varianza

Rapporto S/R gaussiano

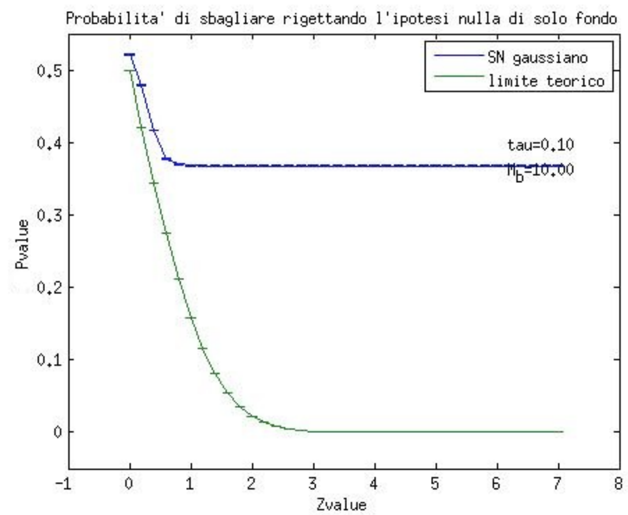
$$Z_{SNG} = \frac{\tau \cdot n_{on} - n_{off}}{\sqrt{n_{off}}}$$

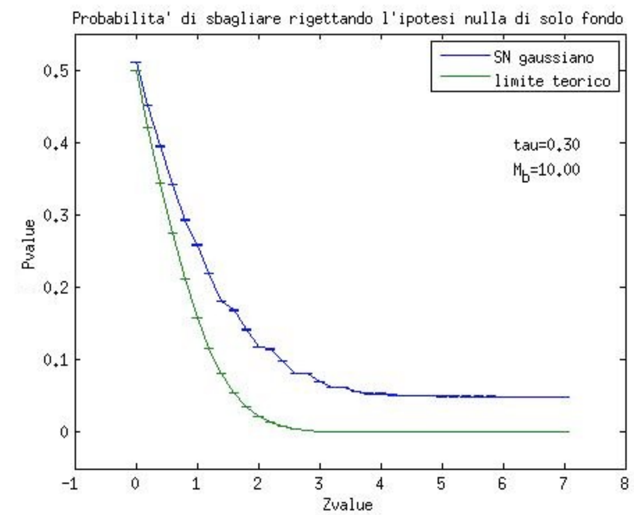
- Suppone esatta la stima del fondo: $\mu_b \cong E[\mu_b]$
- Approssima la statistica poissoniana con quella gaussiana per stimare Z
- E' semplice e maneggevole
- Sovrastima sistematicamente il valore di Z

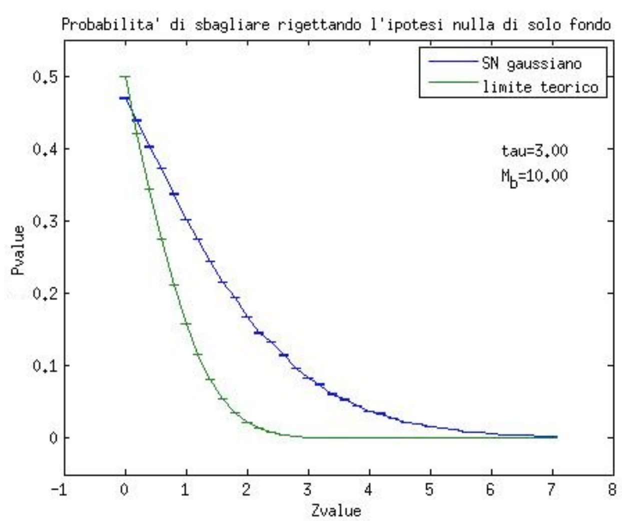
Rapporto S/R gaussiano

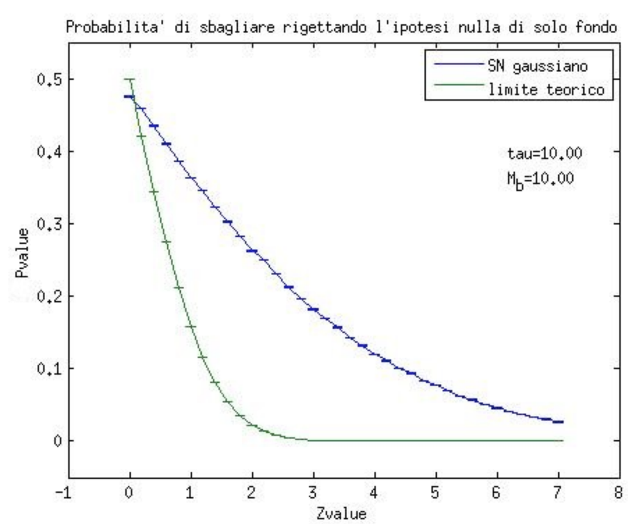
$$Z_{SNG} = \frac{\tau \cdot n_{on} - n_{off}}{\sqrt{n_{off}}}$$

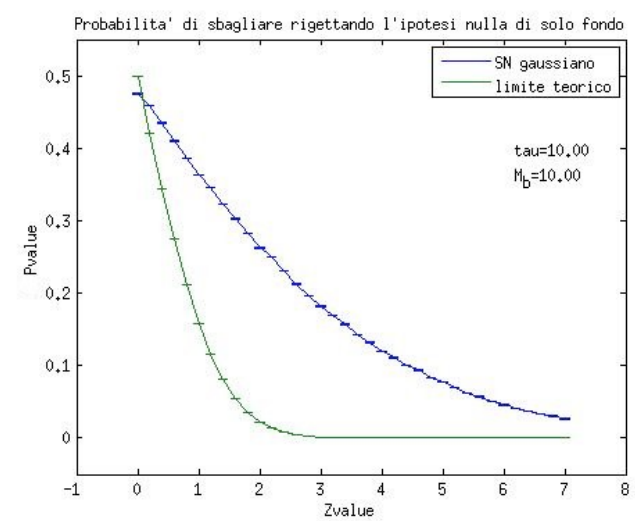
Fondo: μ_b=10
$\tau=\mu_{off}/\mu_{on}$: 0.1 0.3 1 3 10
Sovrastima sistematica
del livello di significativita'





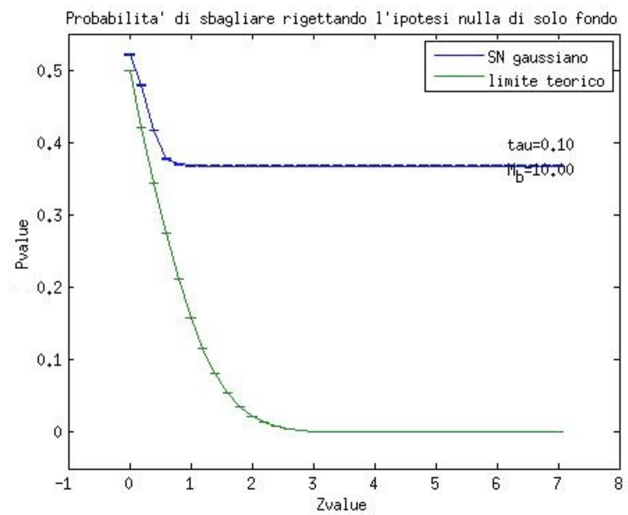


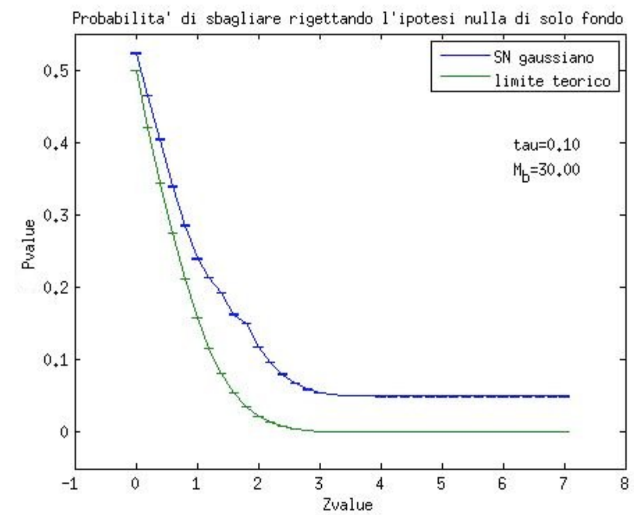


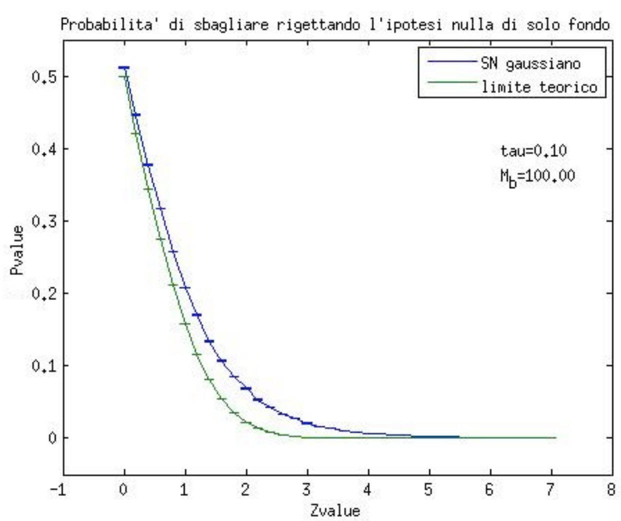


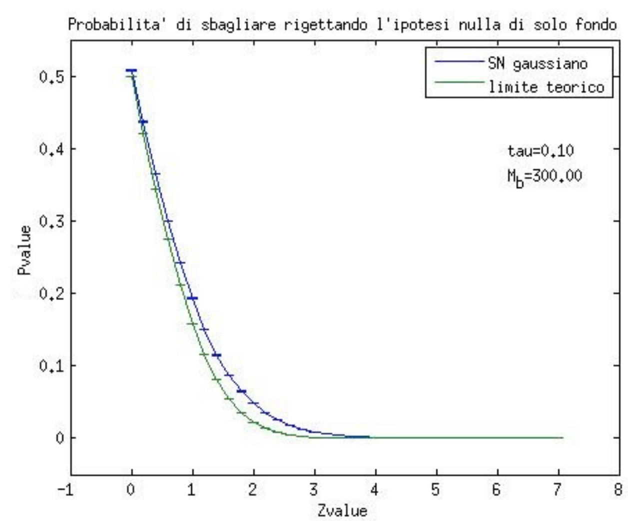
Rapporto S/R gaussiano

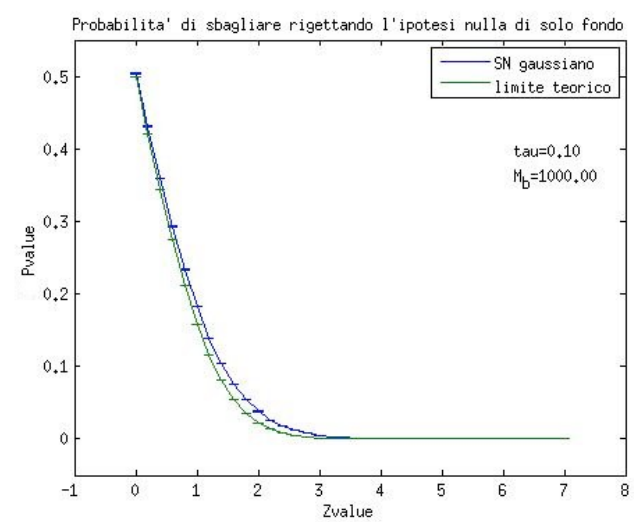
Fondo μ_b:
10, 30, 100, 300, 1000
$\tau = \mu_{off}/\mu_{on} = 0.1$
Stima di Z converge
per grandi valori di μ_b





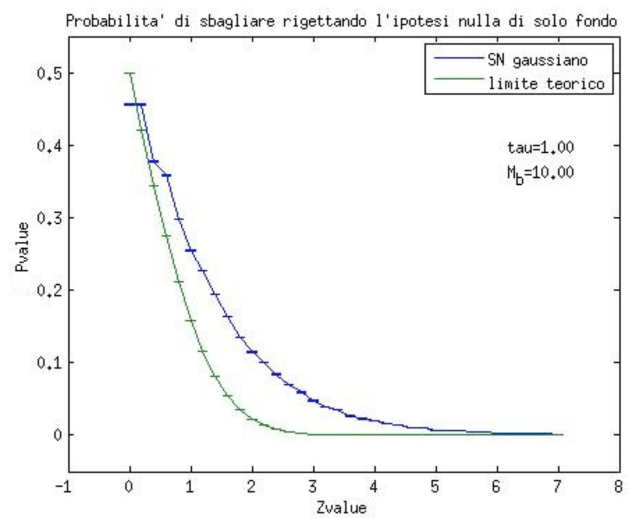


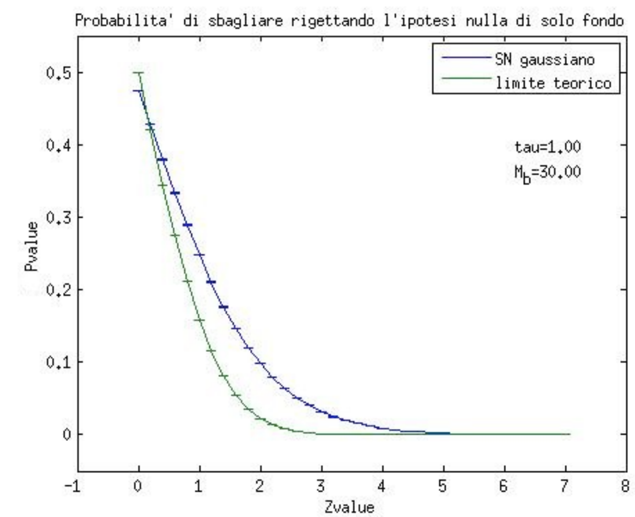


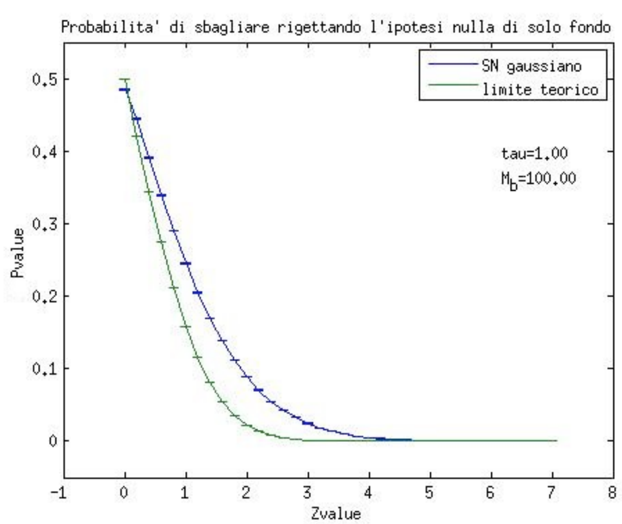


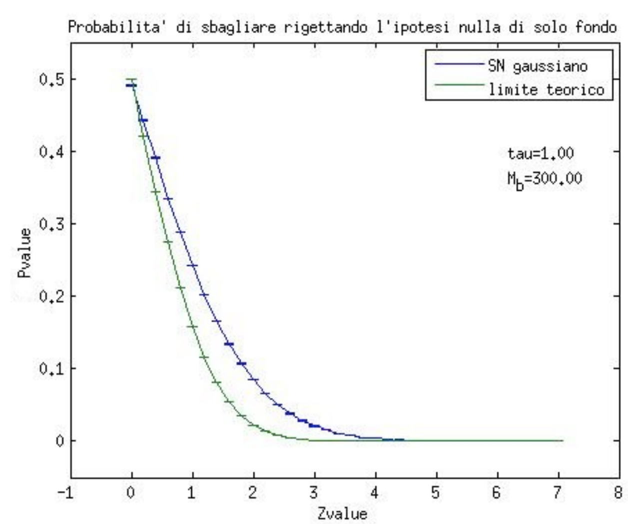
Rapporto S/R gaussiano

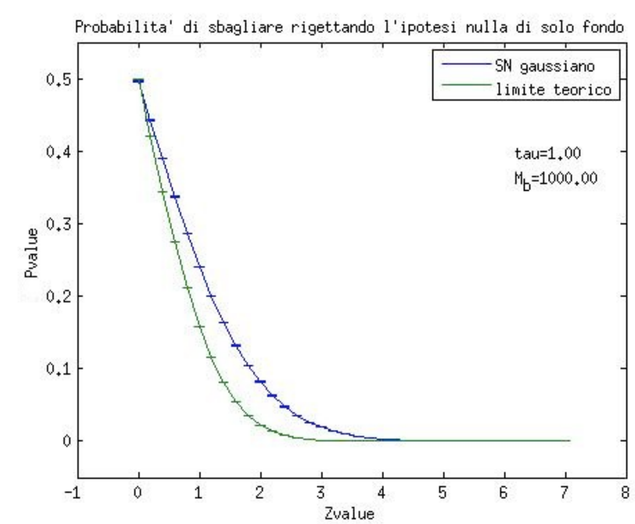
Fondo μ_b:
10, 30, 100, 300, 1000
$\tau=\mu_{off}/\mu_{on}=1$
Stima di Z converge
poco al crescere di μ_b





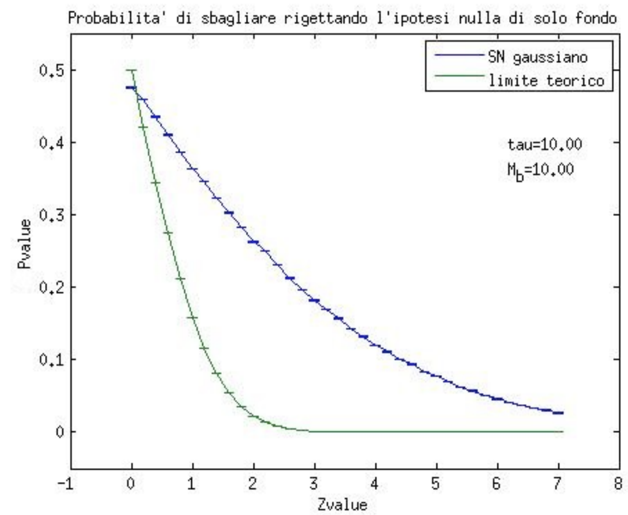


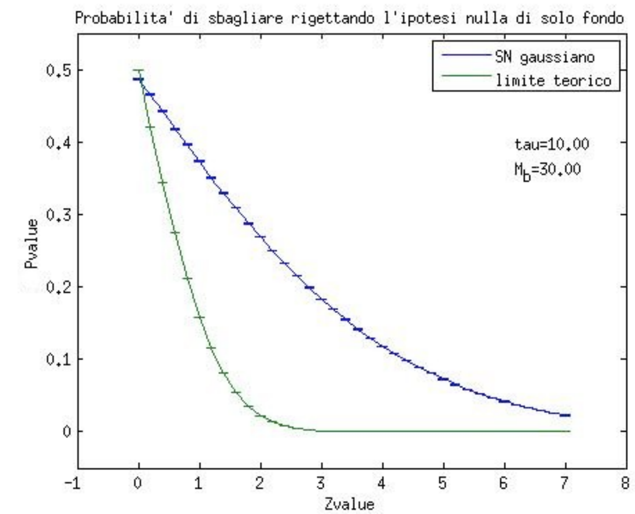


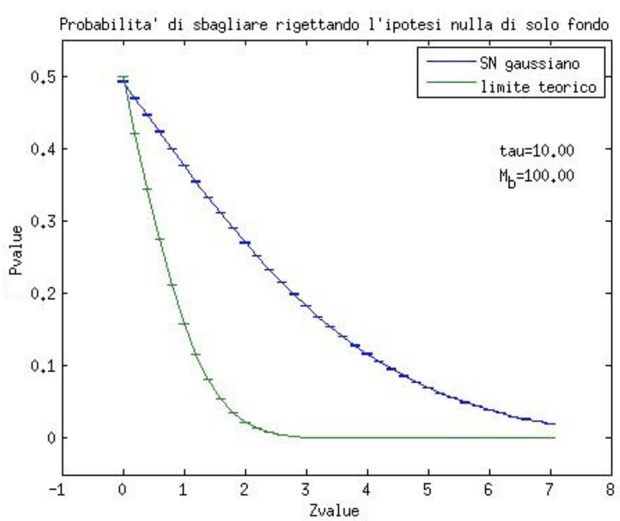


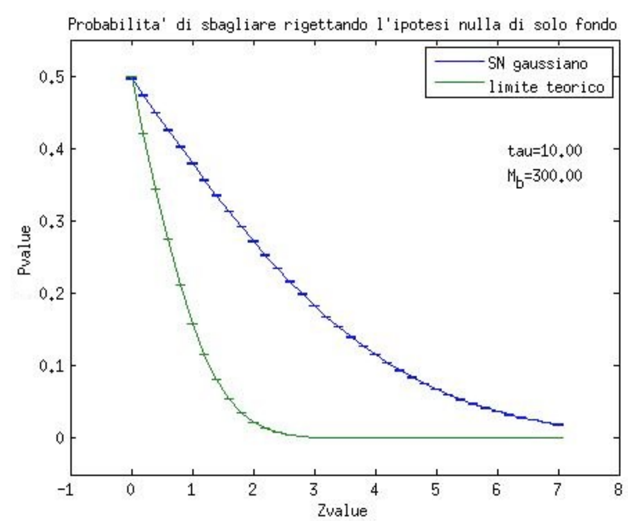
Rapporto S/R gaussiano

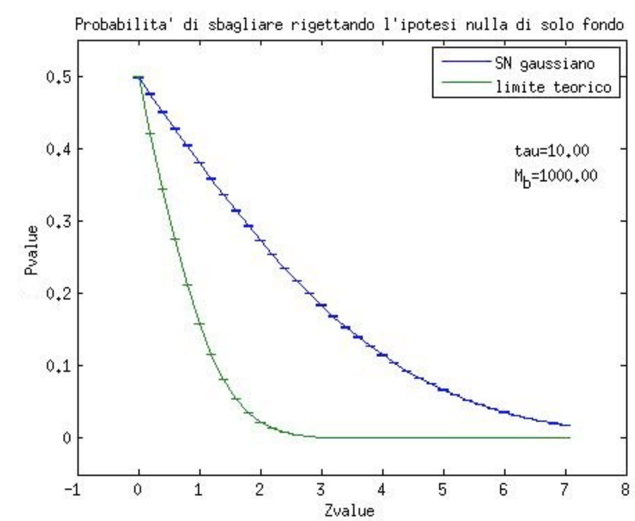
Fondo μ_b:
10, 30, 100, 300, 1000
$\tau=\mu_{off}/\mu_{on}=10$
Stima di Z non converge
neppure a grandi valori di μ_b





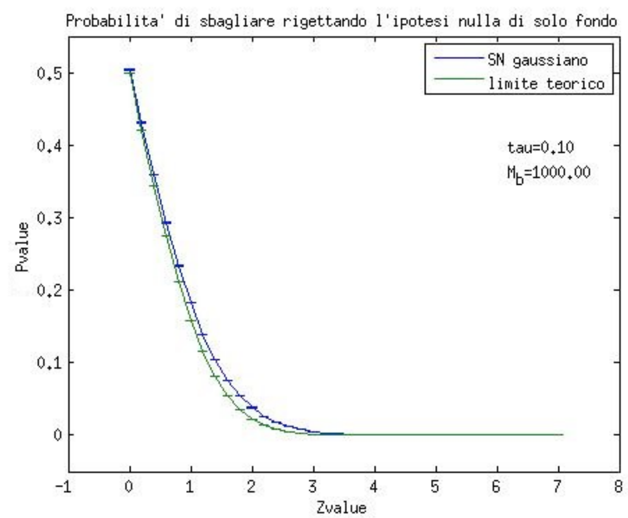


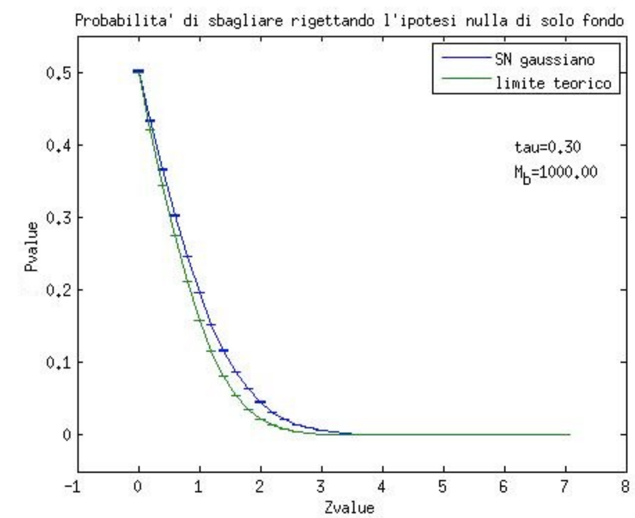


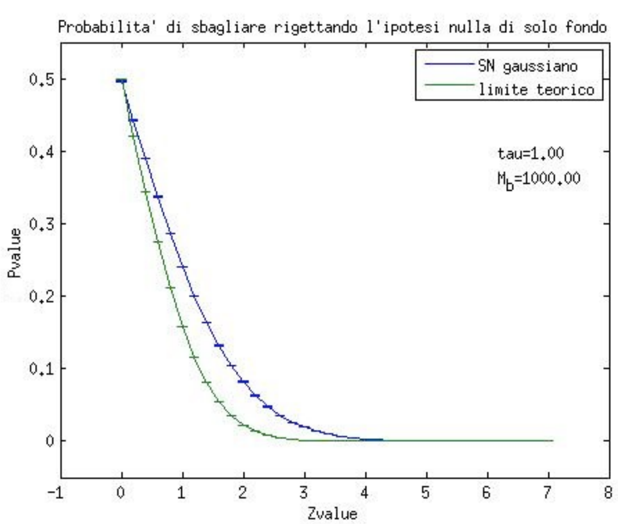


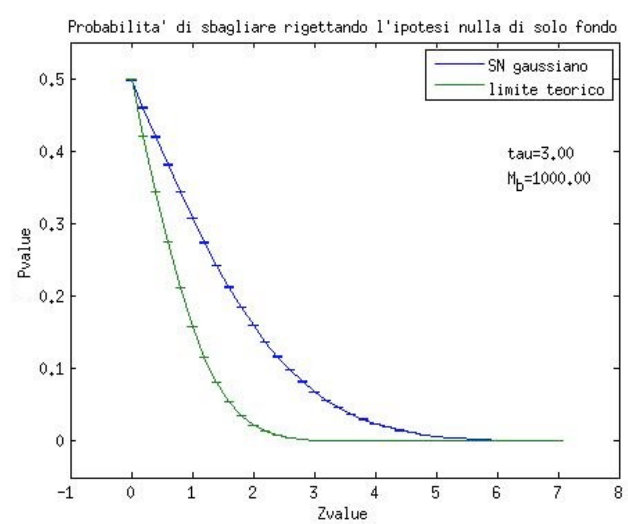
Rapporto S/R gaussiano

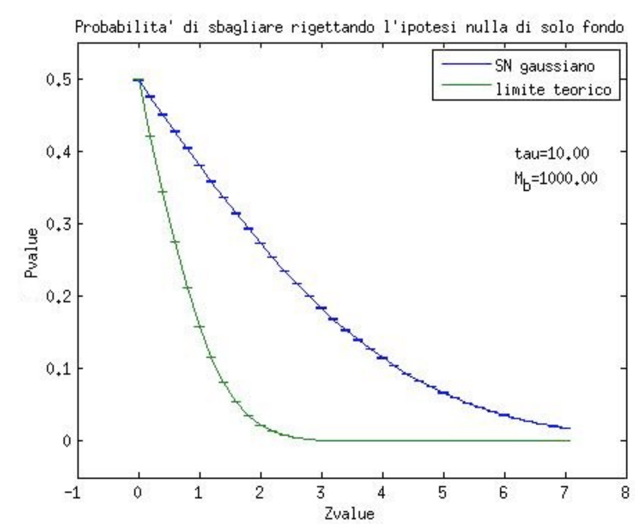
Fondo μ_b=1000
$\tau=\mu_{off}/\mu_{on}$:
0.1, 0.3, 1, 3, 10
Al crescere di τ la stima di Z
non converge piu‘ a grandi μ_b











Rapporto S/R poissoniano

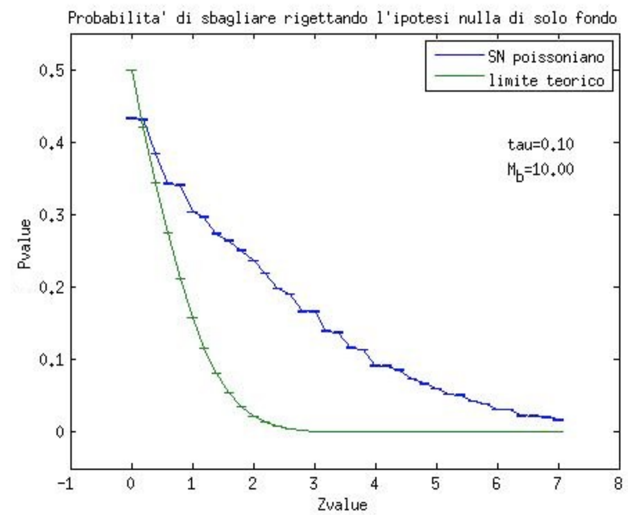
$$p = \sum_{i \geq n_{on}} \frac{\mu_b^i}{i!} e^{-\mu_b} = \Gamma(\mu_b, 0, n_{on})$$

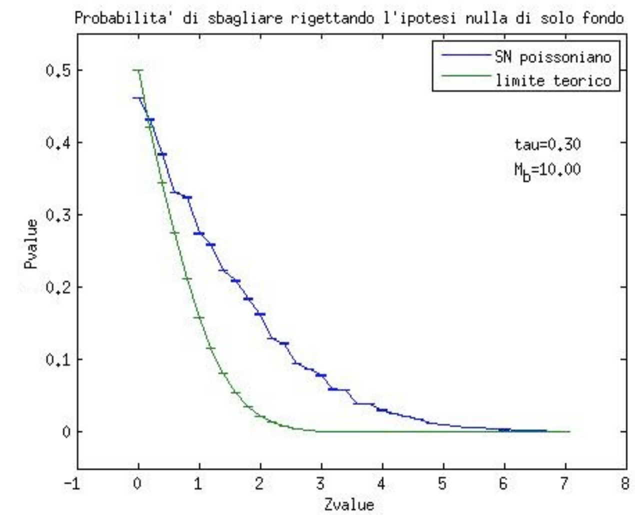
$$Z_{SNP} = \sqrt{2} erf^{-1}(1 - 2p)$$

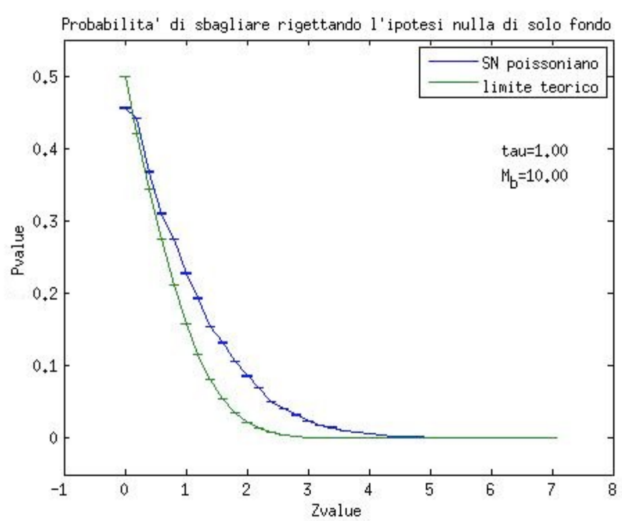
- Suppone che la stima del fondo sia esatta
- Usa la statistica poissoniana per calcolare il p-value e ne deriva lo Z equivalente

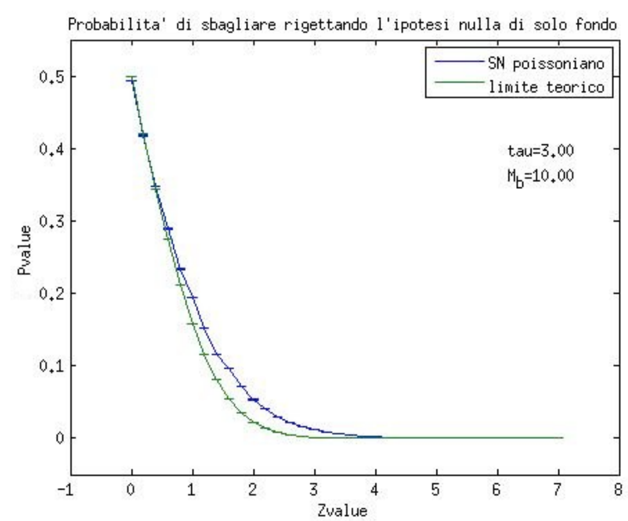
Rapporto S/R poissoniano

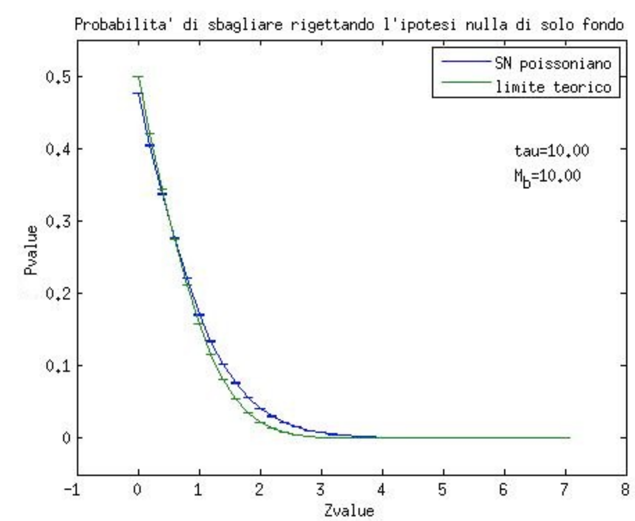
Fondo: μ_b=10
$\tau=\mu_{off}/\mu_{on}$: 0.1 0.3 1 3 10
Sovrastima sistematica
del livello di significativita'
Comportamento migliore
di S/R gaussiano ad alti Z





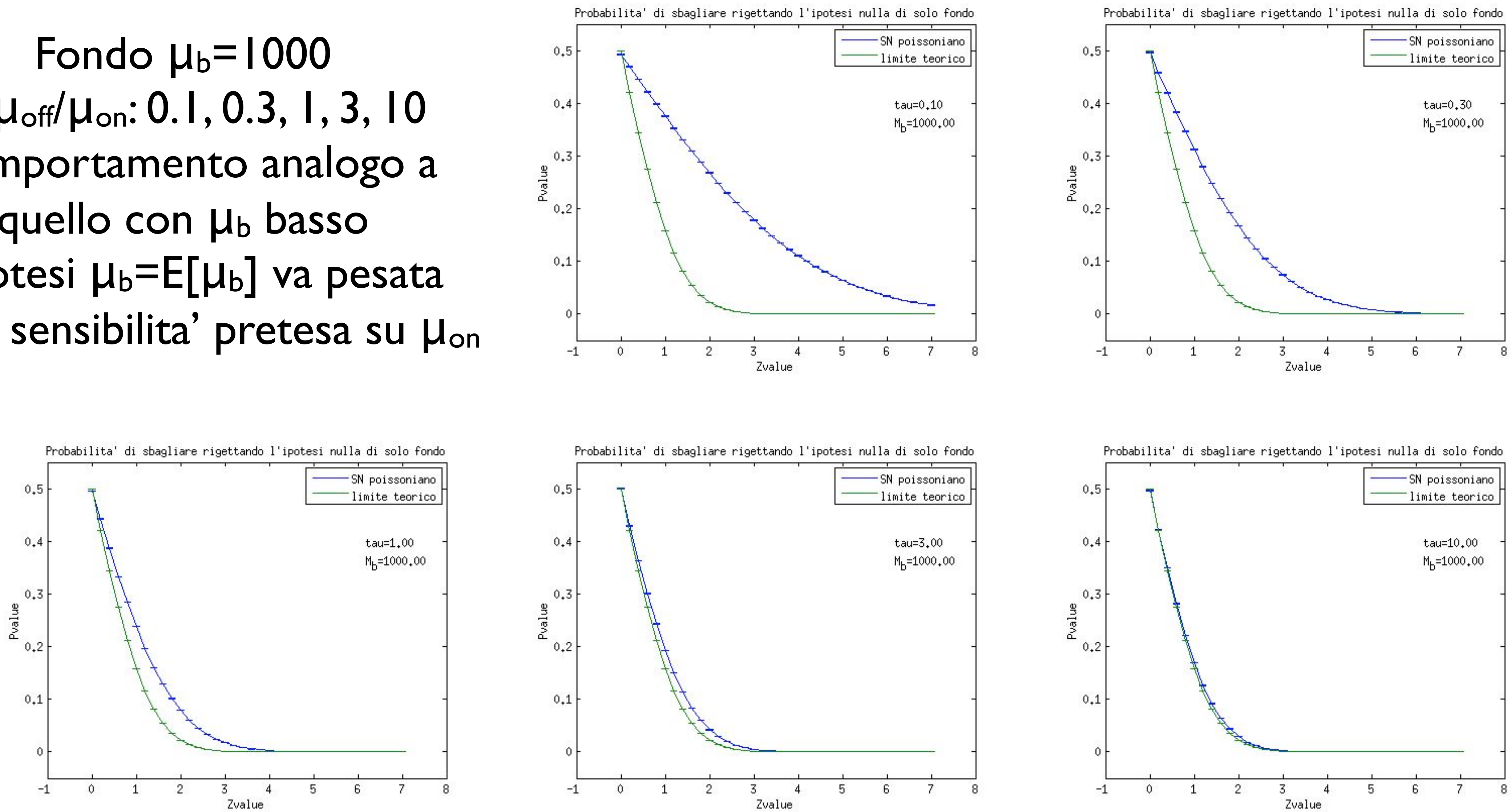






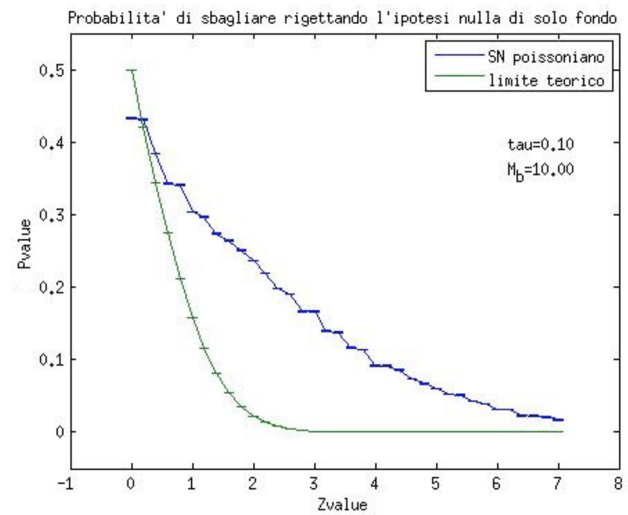
Rapporto S/R poissoniano

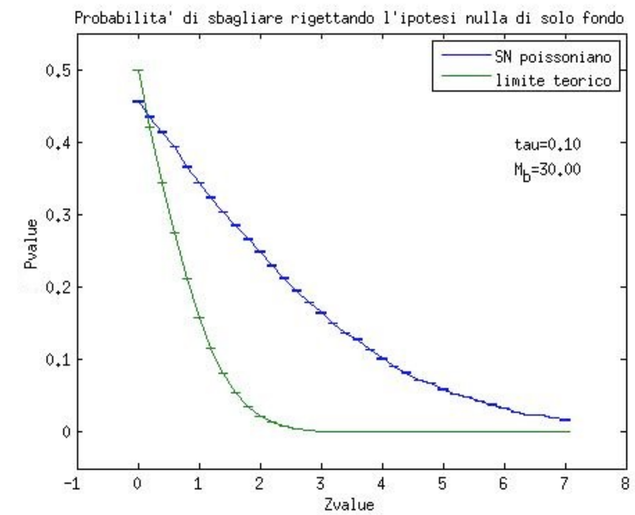
Fondo μ_b=1000
$\tau=\mu_{off}/\mu_{on}$: 0.1, 0.3, 1, 3, 10
Comportamento analogo a quello con μ_b basso
L'ipotesi μ_b=E[μ_b] va pesata con la sensibilita' pretesa su μ_{on}

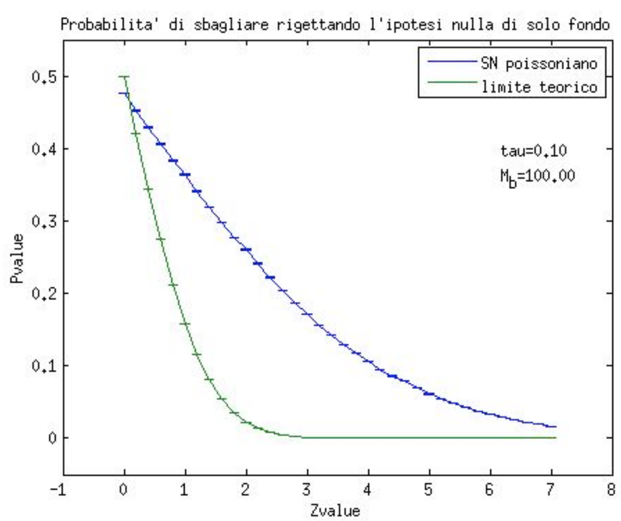


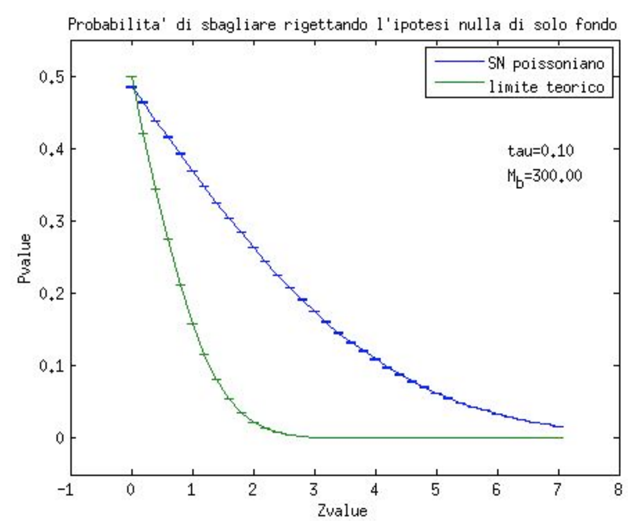
Rapporto S/R poissoniano

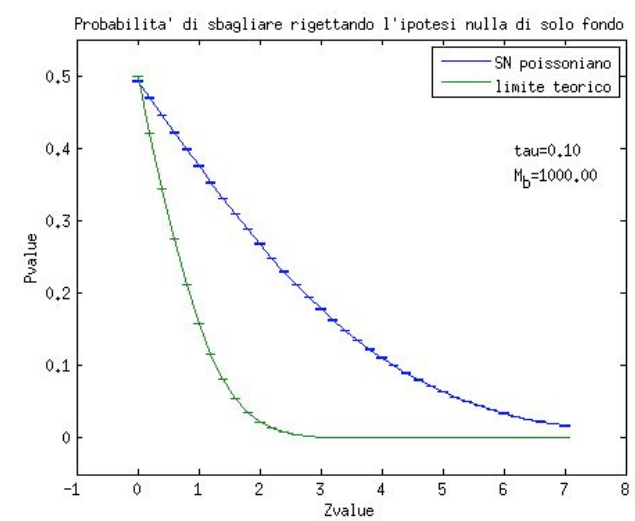
Fondo μ_b:
10, 30, 100, 300, 1000
$\tau=\mu_{off}/\mu_{on}$: 0.1
Sovrastima marcata di Z per qualsiasi livello di fondo





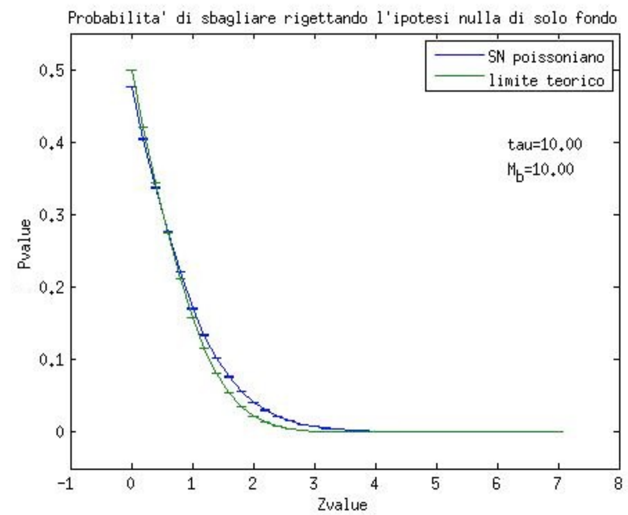


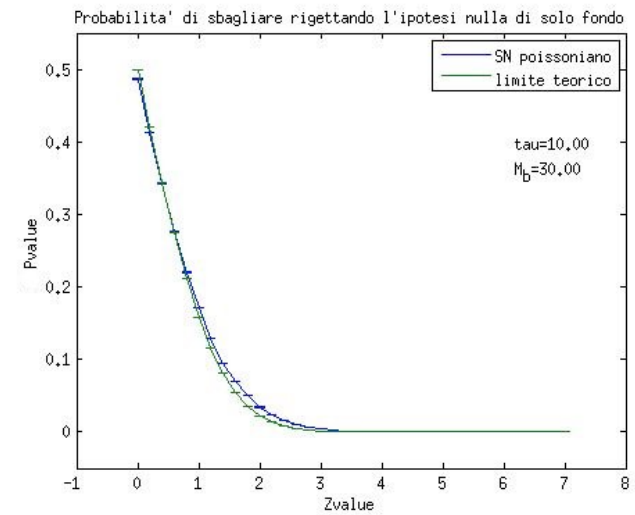


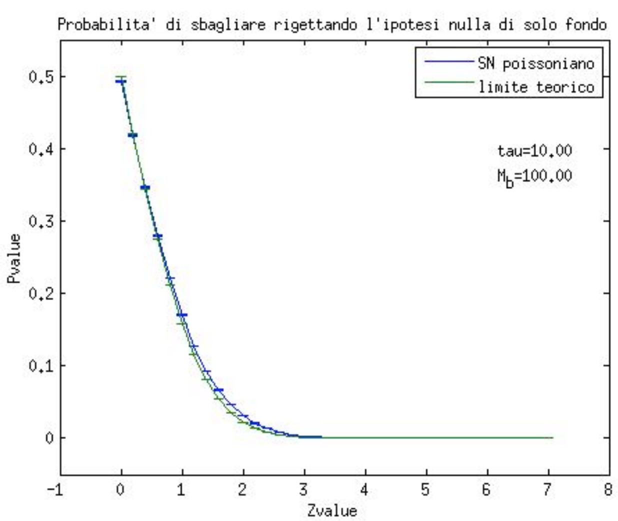


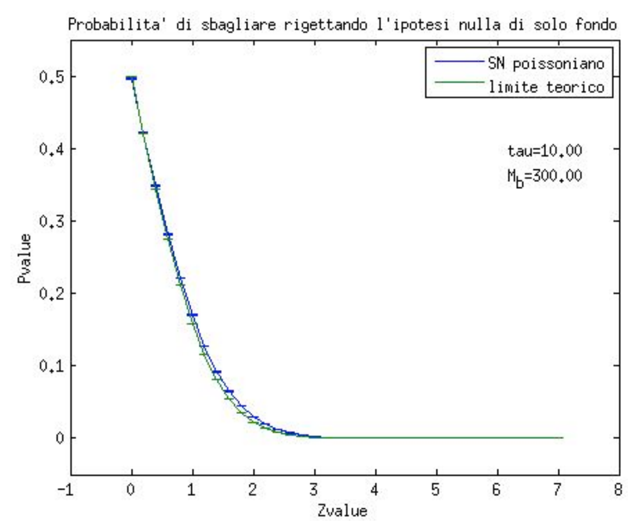
Rapporto S/R poissoniano

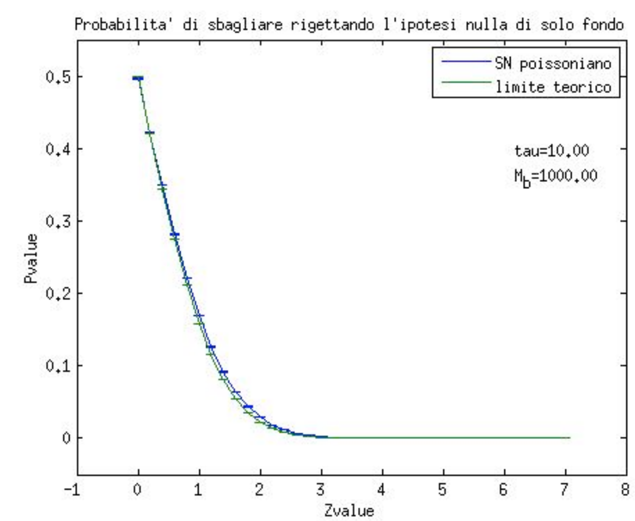
Fondo μ_b:
10, 30, 100, 300, 1000
$\tau=\mu_{off}/\mu_{on}$: 10
Sovrastima lieve di Z per qualsiasi livello di fondo











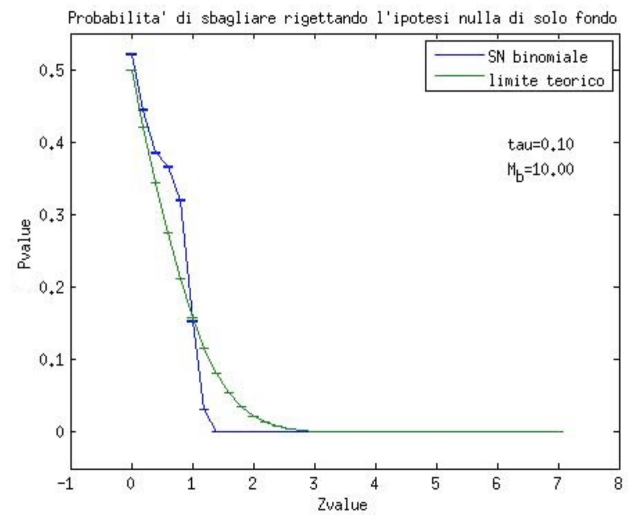
Rapporto S/R binomiale

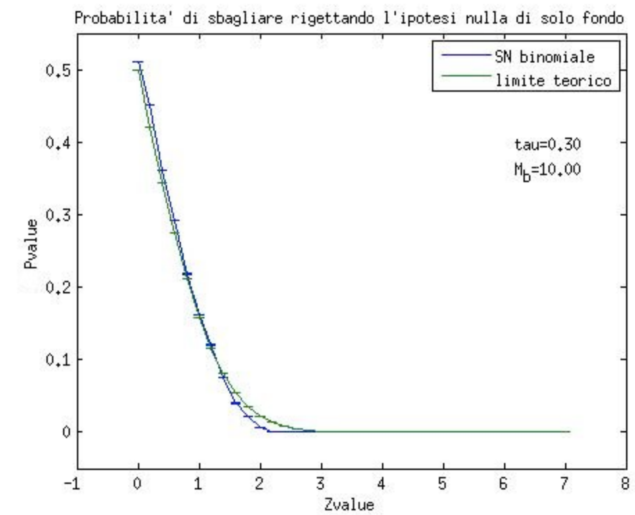
$$Z_{SNB} = \frac{\tau \cdot n_{on} - n_{off}}{\sqrt{\tau \cdot (n_{on} + n_{off})}}$$

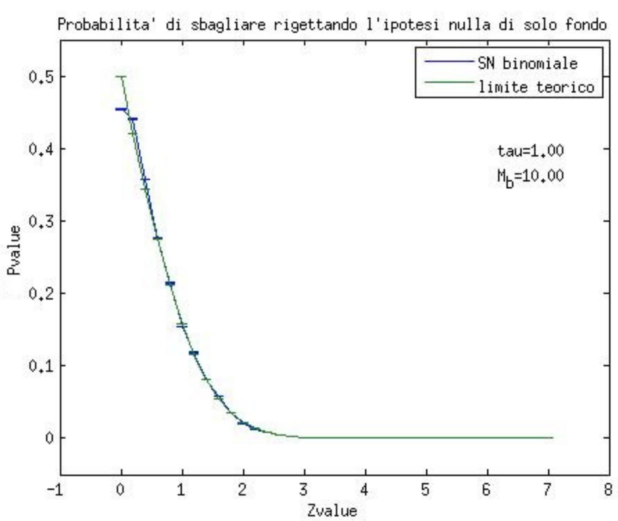
- Assunta H_0 si analizza come il fondo si ripartisce tra le misure On e Off
- Approssima la statistica binomiale con quella gaussiana per stimare Z

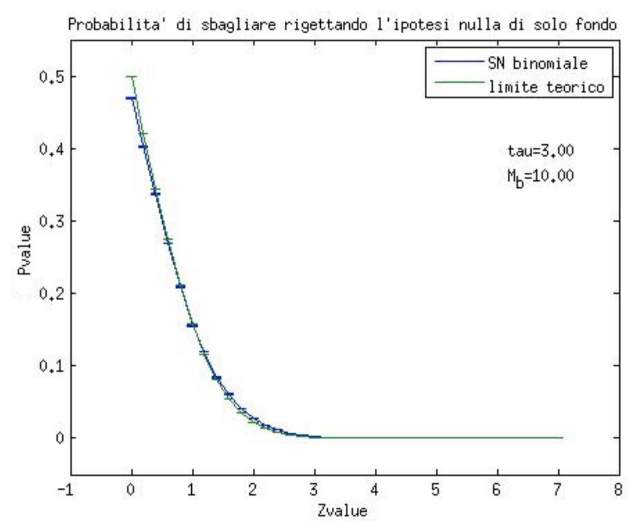
Rapporto S/R binomiale

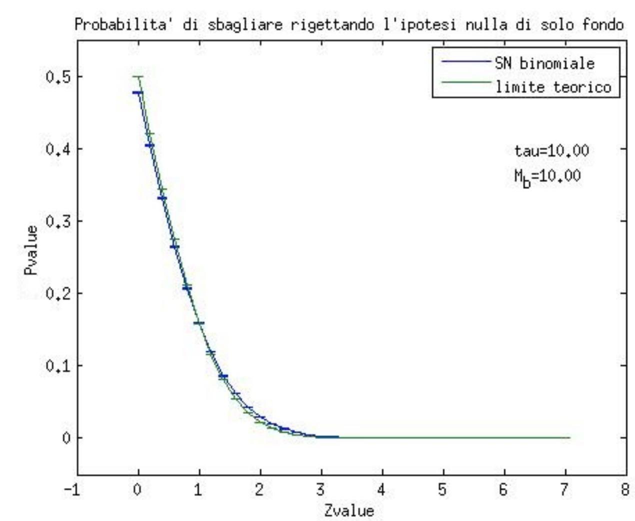
Fondo: μ_b=10
$\tau=\mu_{off}/\mu_{on}$: 0.1 0.3 1 3 10
Sovrastima Z per Z<1
Sottostima Z per Z>1
Converge al crescere di τ





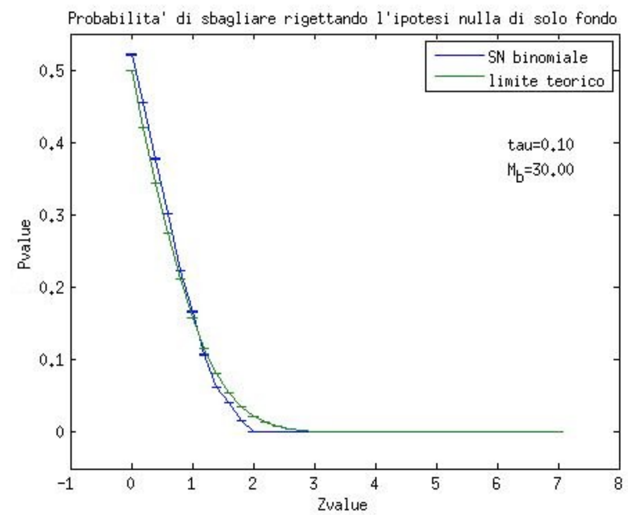


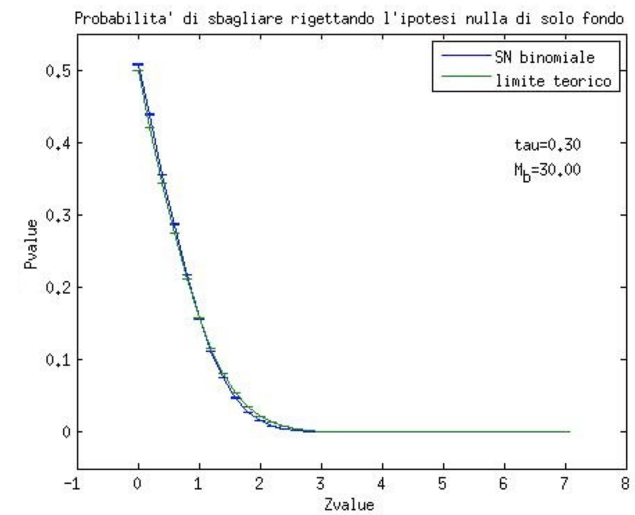


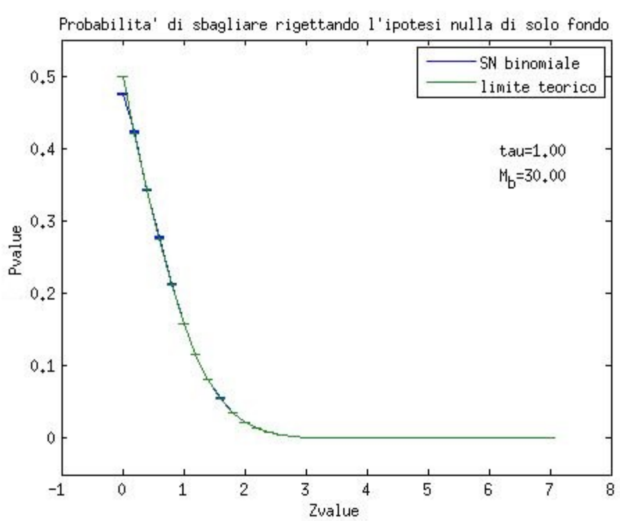


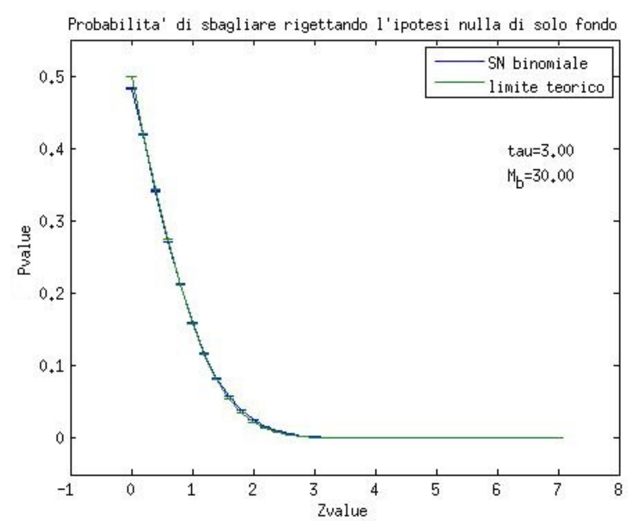
Rapporto S/R binomiale

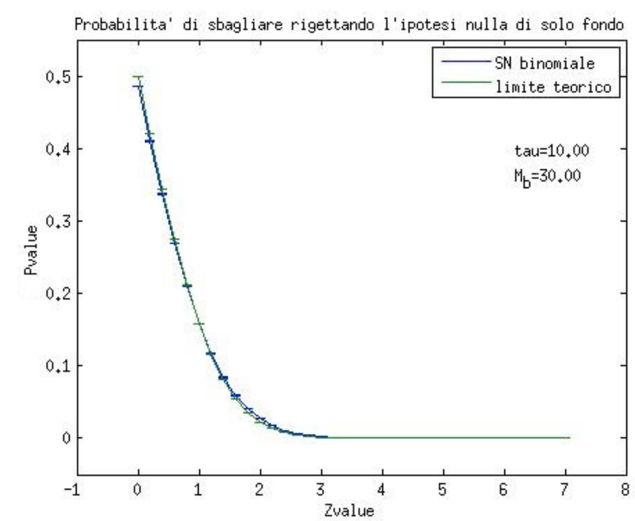
Fondo: μ_b=30
$\tau=\mu_{off}/\mu_{on}$: 0.1 0.3 1 3 10
Stesso comportamento
al crescere di μ_b
Converge al valore atteso
al crescere di $\mu_{off}+\mu_{on}$





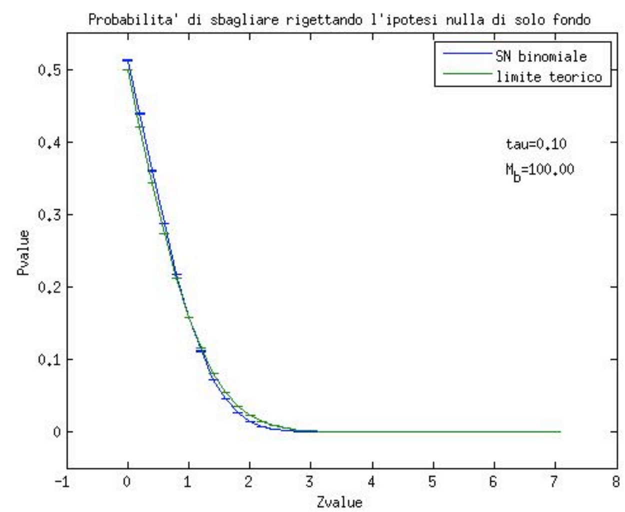


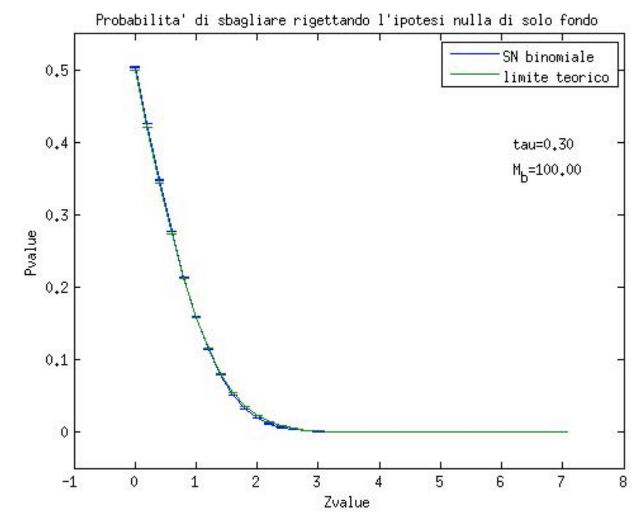


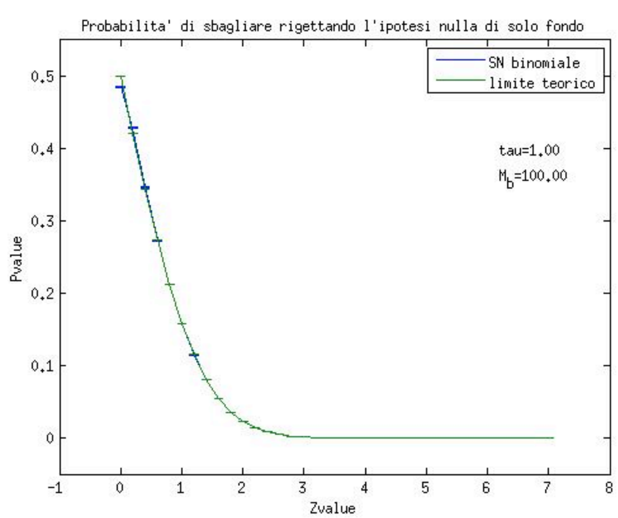


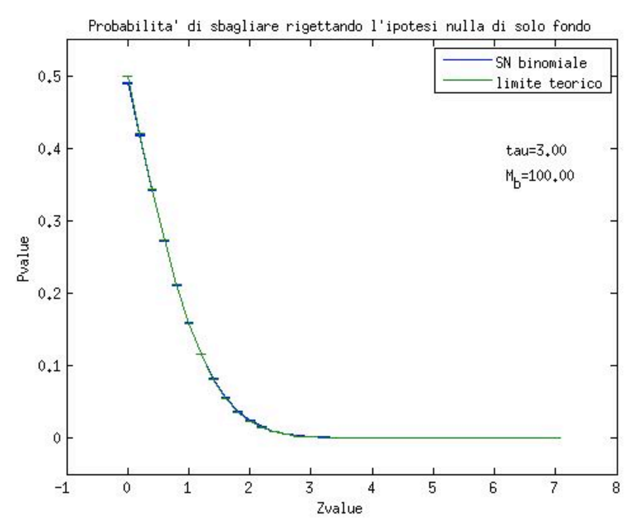
Rapporto S/R binomiale

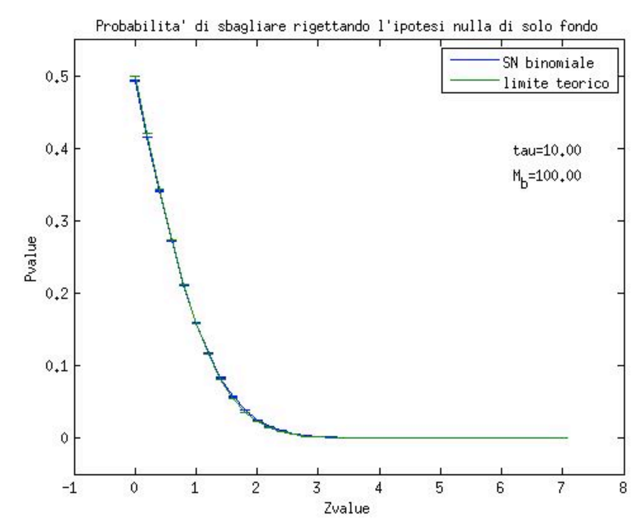
Fondo: μ_b=100
$\tau=\mu_{off}/\mu_{on}$: 0.1 0.3 1 3 10
Ottima convergenza
con $\mu_{off}+\mu_{on}$>100











Rapporto di verosimiglianza

Teorema di Wilks (1938)

- Ipotesi H_1 dipendente da p parametri
- Ipotesi H_0 ottenuta da H_1 costringendo v di questi parametri al loro valore vero
- Funzioni di verosimiglianza abbastanza regolari
- Il rapporto TS tra i logaritmi di verosimiglianza converge al crescere del numero di punti n ad una distribuzione di tipo χ^2 con p-v gradi di liberta', a meno di un fattore $\propto n^{-1/2}$
- La distribuzione asintotica di TS non dipende dalla forma della funzione di verosimiglianza

Rapporto di verosimiglianza

- Il teorema di Wilks per il problema On/Off porta a un TS con 1 solo grado di liberta'
- Per H_1 la massima verosimiglianza si ha per $E[\mu_b] = n_{off}/\tau \quad E[\mu_s] = n_{on} - n_{off}/\tau$
- Per H_0 si ha $E[\mu_b|\mu_s=0] = (n_{on}+n_{off})/(1+\tau)$
- La variabile stocastica $Y=[TS(X)]^{1/2}$ tende a una distribuzione gaussiana per grandi n

Rapporto di verosimiglianza

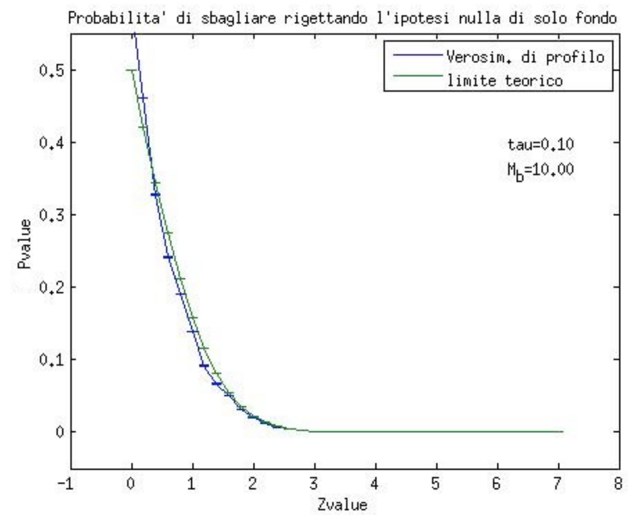
Li & Ma (1983)

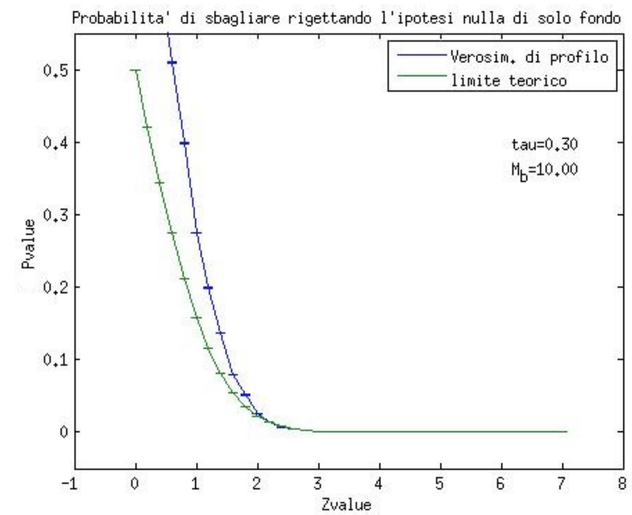
$$Z_{MV} = \sqrt{-2 \ln \frac{\mathcal{L}(0, \frac{n_{on}+n_{off}}{1+\tau})}{\mathcal{L}(n_{on} - \frac{n_{off}}{\tau}, \frac{n_{off}}{\tau})}} =$$

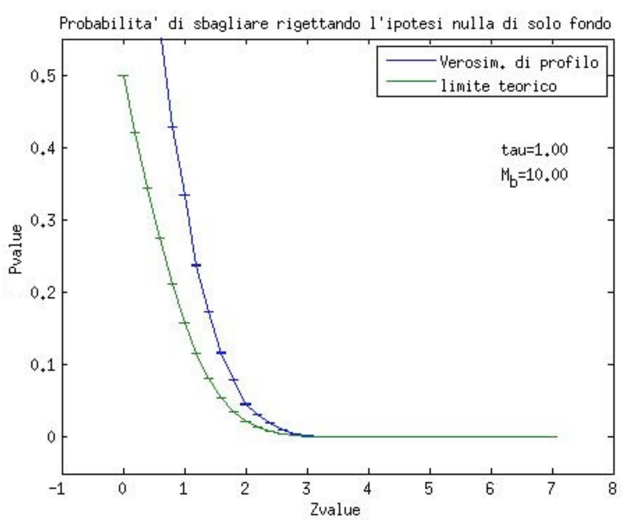
$$= \sqrt{2} \left(n_{on} \cdot \ln \frac{n_{on}(1+\tau)}{n_{on}+n_{off}} + n_{off} \cdot \ln \frac{n_{off}(1+1/\tau)}{n_{on}+n_{off}} \right)^{1/2}$$

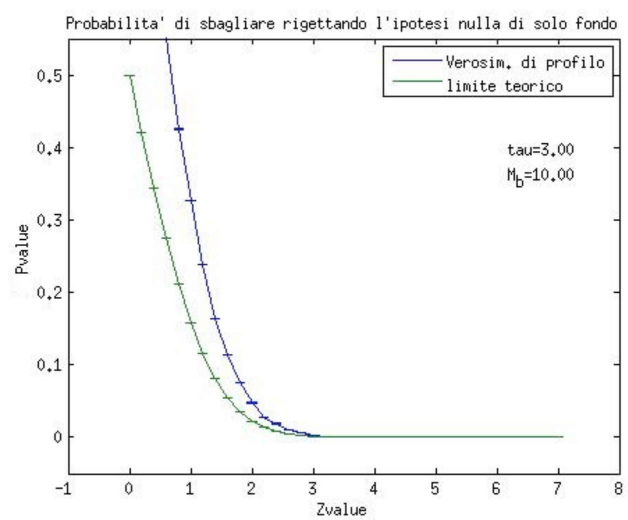
Rapporto di verosimiglianza

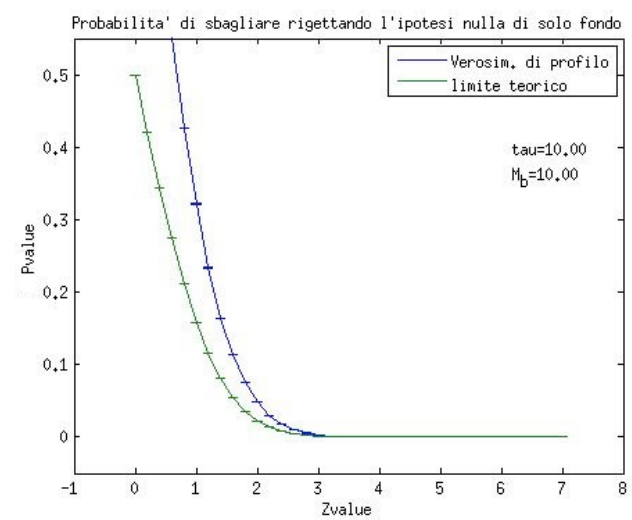
Fondo: μ_b=10
$\tau=\mu_{off}/\mu_{on}$: 0.1 0.3 1 3 10
Sottostima Z al crescere di τ
Si comporta bene quando
ci sono pochi conteggi





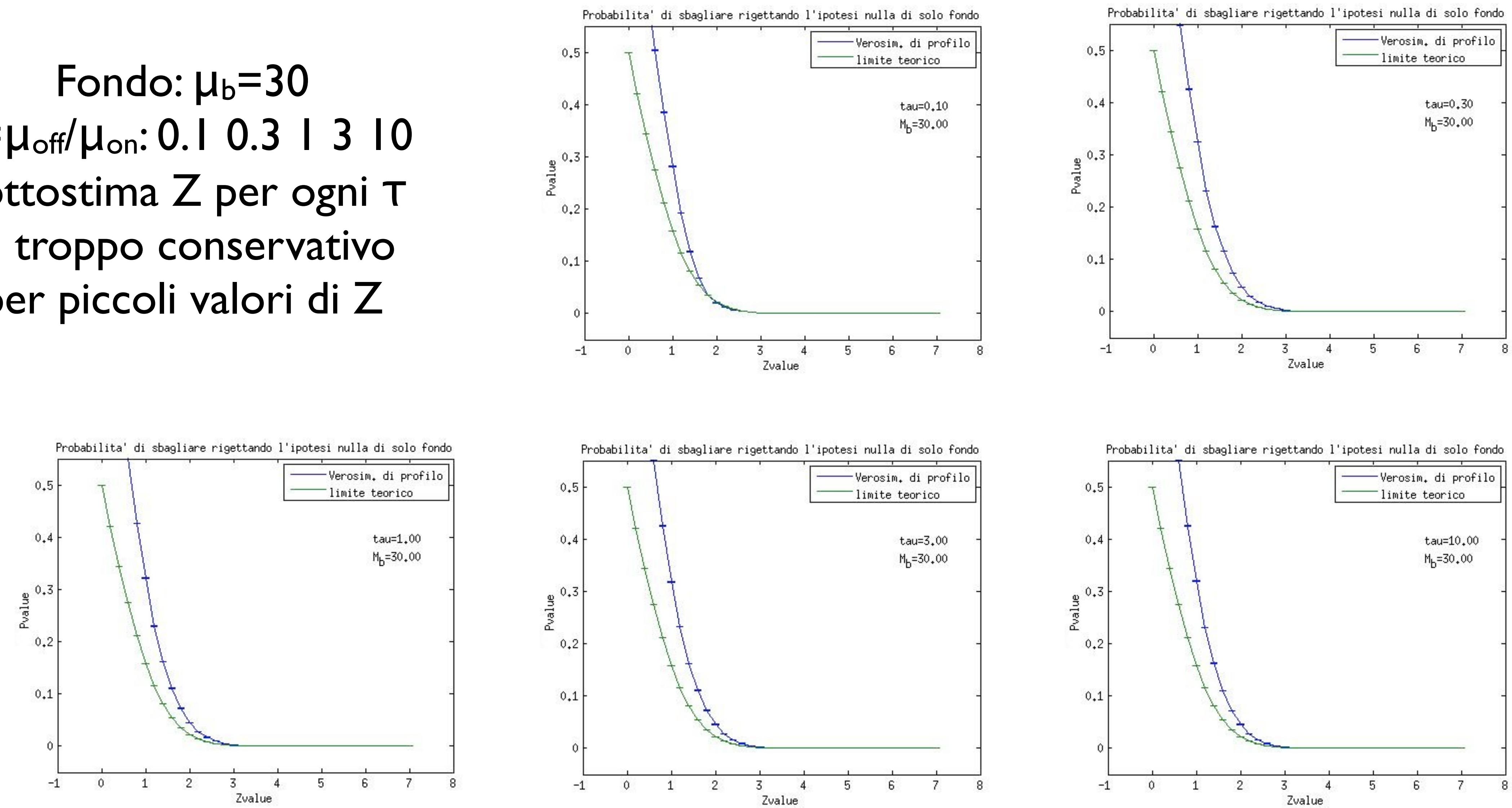






Rapporto di verosimiglianza

Fondo: μ_b=30
$\tau=\mu_{off}/\mu_{on}$: 0.1 0.3 1 3 10
Sottostima Z per ogni τ
E' troppo conservativo
per piccoli valori di Z



Stabilizzazione di varianza

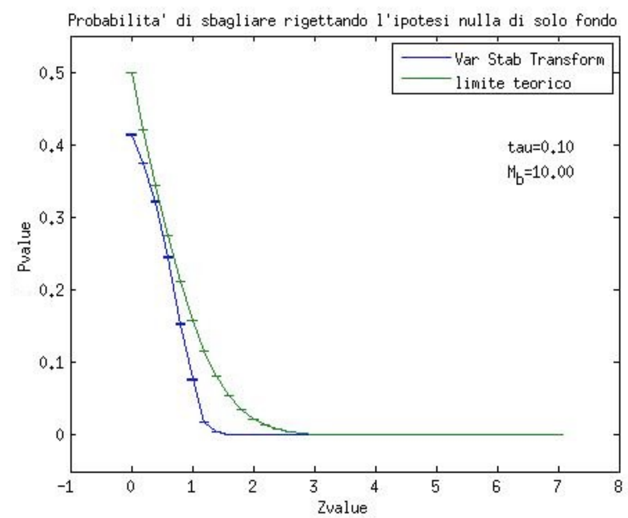
Zhang & Ramsden (1989)

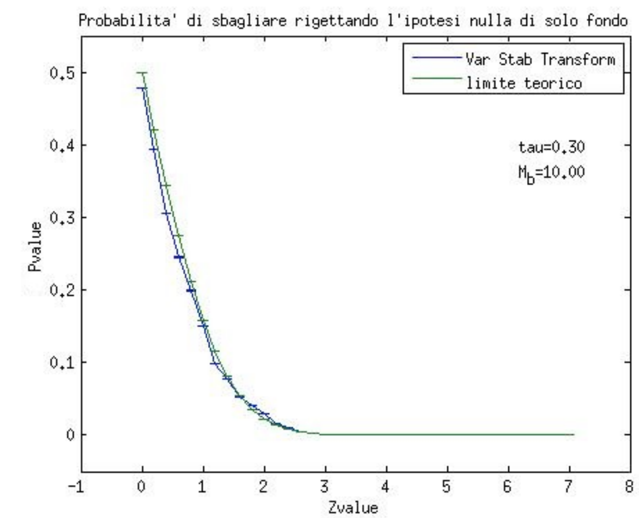
- Viene applicata una trasformazione alla variabile N_{on}-N_{off}/τ in modo che assuma profilo di distribuzione gaussiano con σ=1
- Accorgimento per tenere conto della distribuzione discreta (Anscombe 1948)

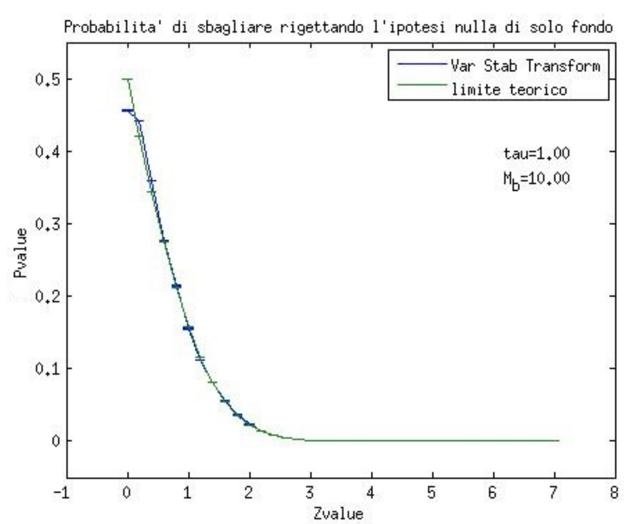
$$Z_{TSV} = \frac{2}{\sqrt{1+1/\tau}}(\sqrt{n_{on}+3/8} - \sqrt{(n_{off}+3/8)/\tau})$$

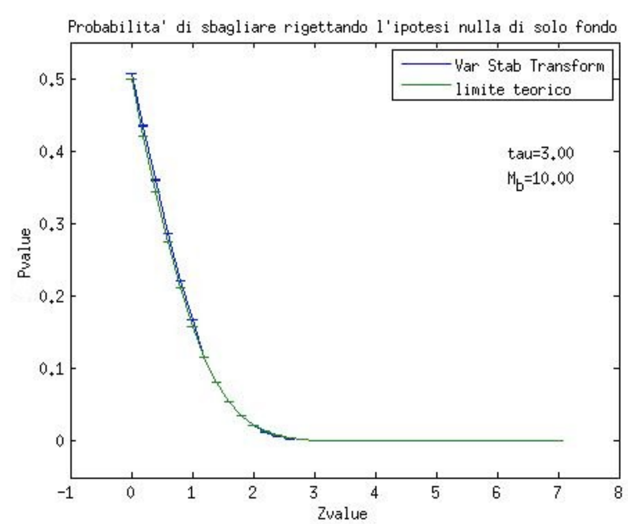
Stabilizzazione di varianza

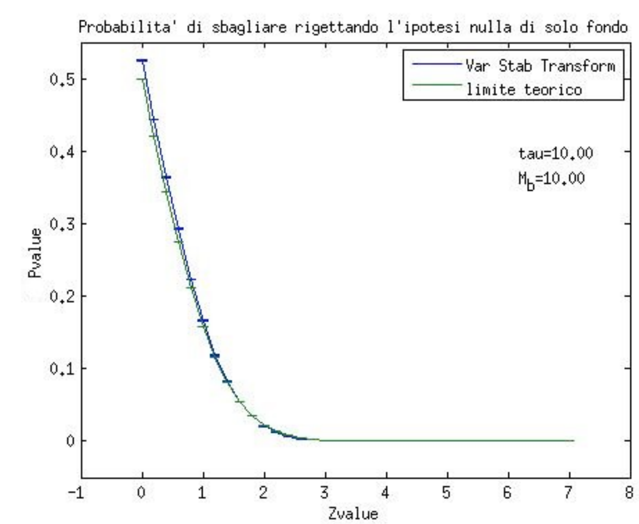
Fondo: $\mu_b=10$
$\tau=\mu_{off}/\mu_{on}$: 0.1 0.3 1 3 10
Sovrastima Z per Z<1
Sottostima Z per Z>1
Converge al crescere di τ





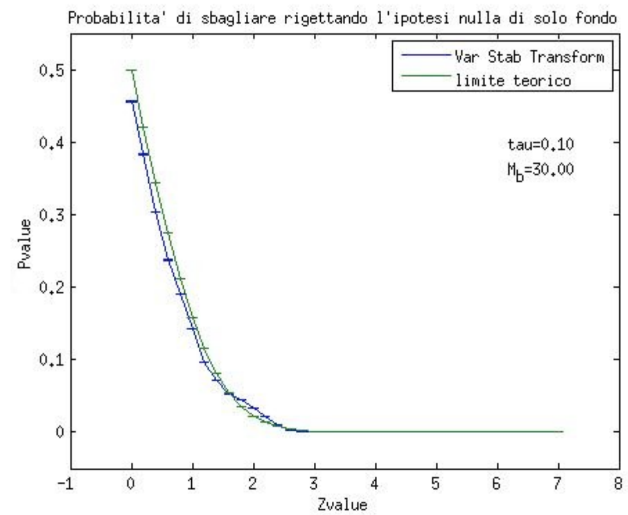


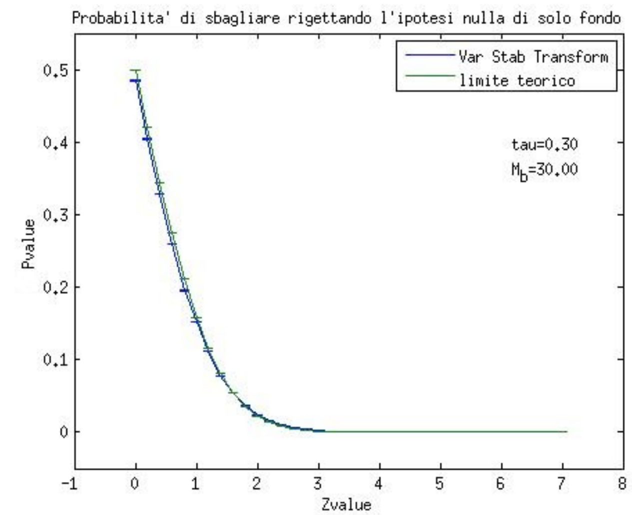


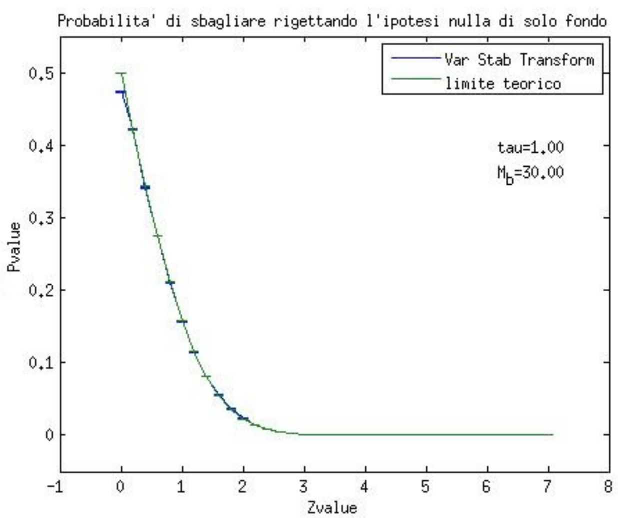


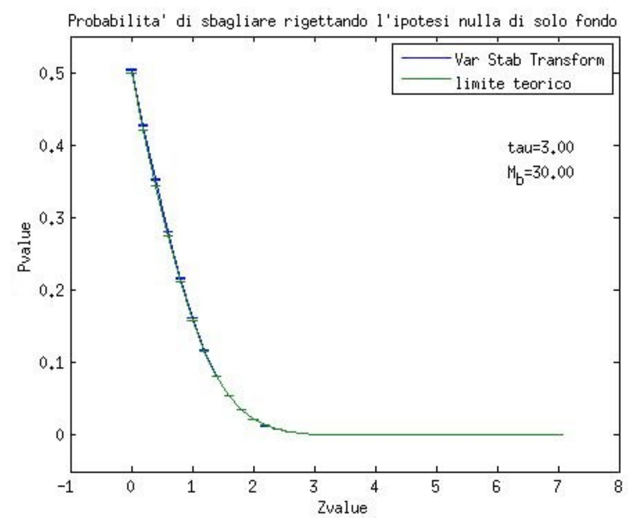
Stabilizzazione di varianza

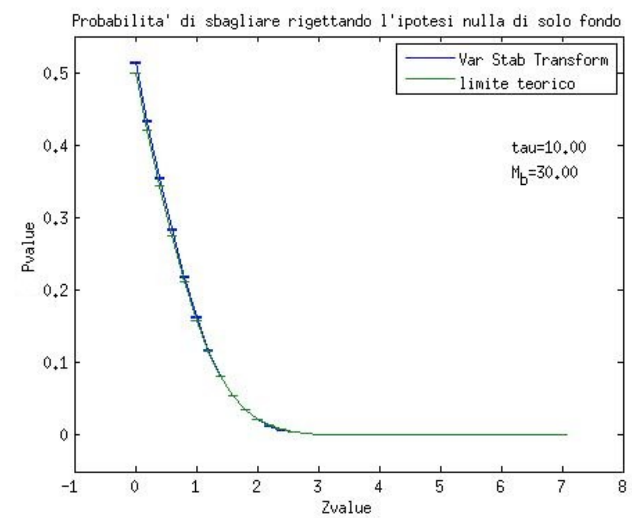
Fondo: μ_b=30
$\tau=\mu_{off}/\mu_{on}$: 0.1 0.3 1 3 10
Ottima aderenza
al valore di Z dichiarato





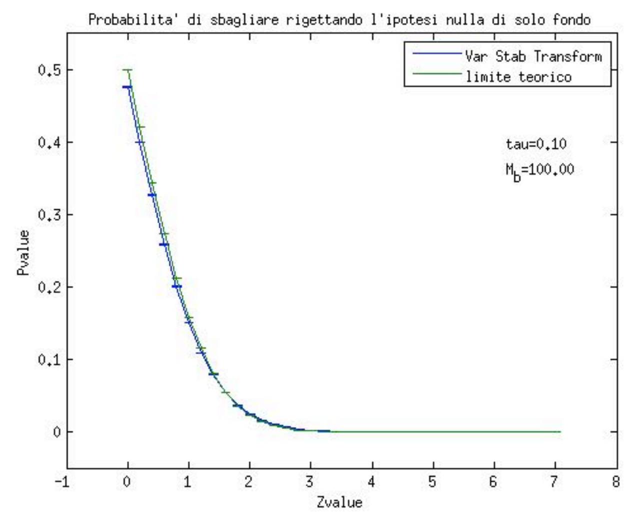


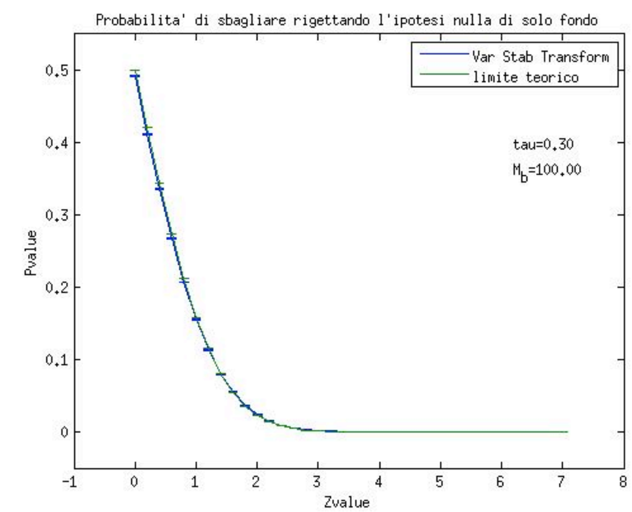


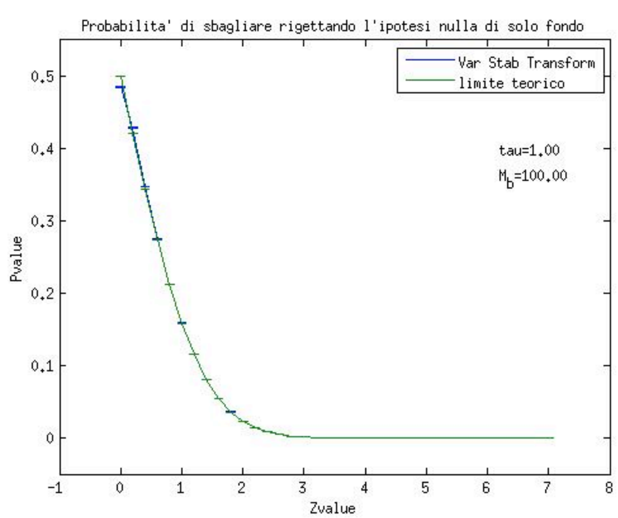


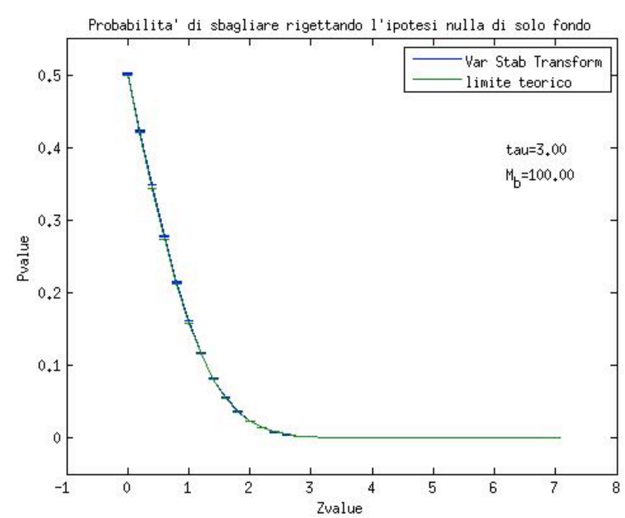
Stabilizzazione di varianza

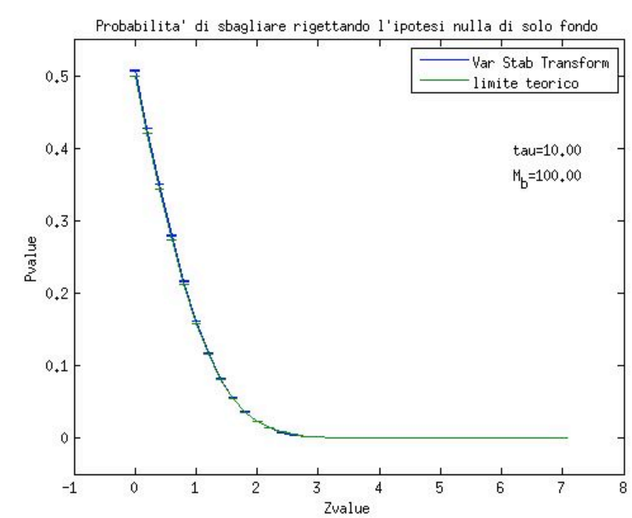
Fondo: μ_b=100
$\tau=\mu_{off}/\mu_{on}$: 0.1 0.3 1 3 10
Ottima aderenza
al valore di Z dichiarato





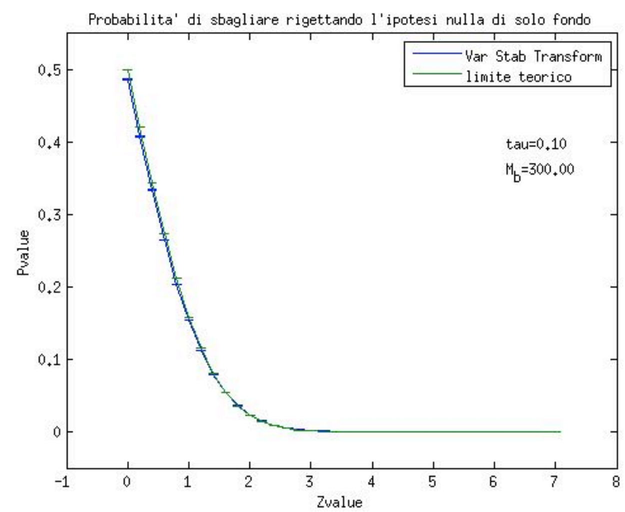


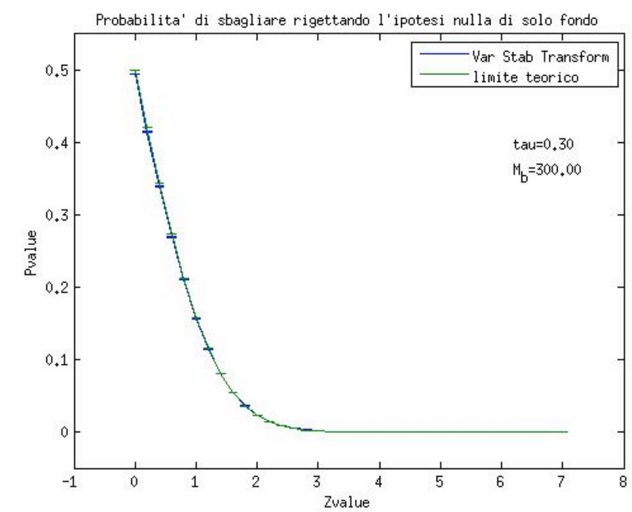


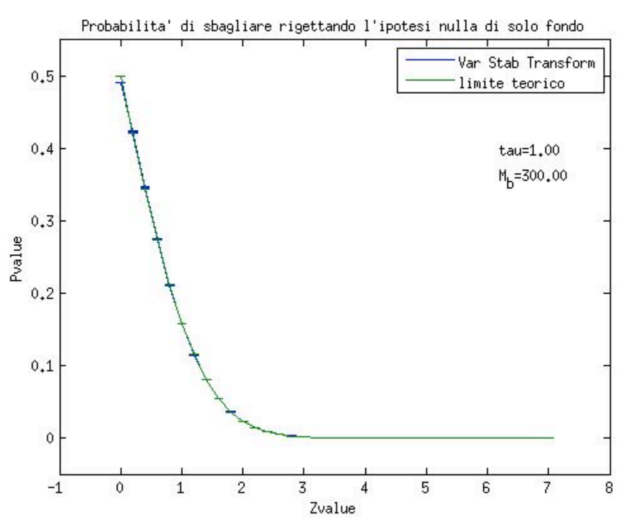


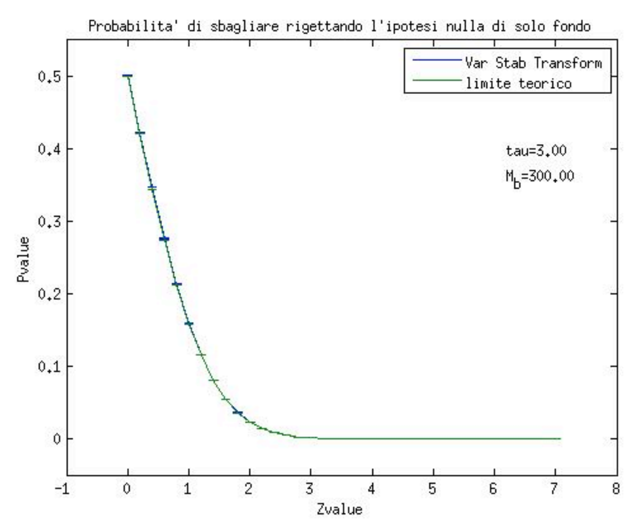
Stabilizzazione di varianza

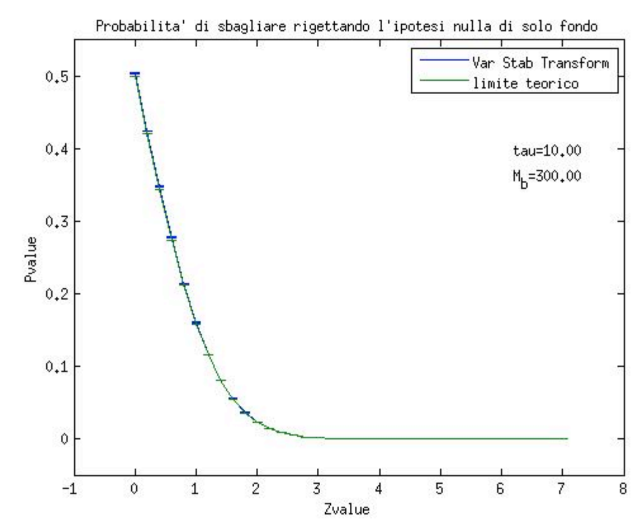
Fondo: μ_b=300
$\tau=\mu_{off}/\mu_{on}$: 0.1 0.3 1 3 10
Ottima aderenza
al valore di Z dichiarato











Stabilizzazione di varianza

Fondo: μ_b=1000
$\tau=\mu_{off}/\mu_{on}$: 0.1 0.3 1 3 10
Ottima aderenza
al valore di Z dichiarato