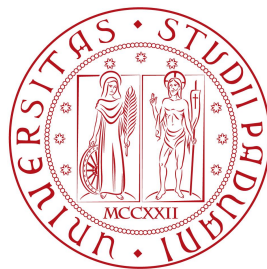


Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea in

Statistica per l'Economia e l'Impresa



**Analisi di alcuni test statistici per l'identificazione di
segnali nella fisica delle alte energie**

Relatore: Prof.ssa Giovanna Menardi
Dipartimento di Scienze Statistiche

Laureando: Davide Benussi
Matricola n. 1237102

Anno Accademico 2021/2022

Indice

Introduzione	3
1 Il contesto di riferimento	6
1.1 La Fisica delle Particelle e il Modello Standard . . .	6
1.2 Gli esperimenti in HEP	8
1.2.1 Collisione di particelle negli acceleratori . .	8
1.2.2 Dall'esperimento fisico al dato statistico . . .	10
1.3 Oltre il Modello Standard	11
1.4 Formalizzazione del problema	14
1.5 Test per il confronto tra distribuzioni	15
2 La ricerca di segnali tramite l'analisi di istogrammi	17
2.1 Formalizzazione	17
2.2 I test di bontà di adattamento per il confronto tra istogrammi	20
2.3 Approcci <i>naive</i> per la misura della significatività . .	23
2.4 Approcci avanzati	25
2.4.1 Incertezza sul <i>background</i>	25
2.4.2 Parametri di disturbo e verosimiglianza profilo	26
2.4.3 Modelli gerarchici e approcci bayesiani . . .	30
2.5 Estensione all'analisi su più <i>bin</i>	36
3 Analisi empirica	40
3.1 Obiettivi	40
3.2 Presentazione degli scenari esplorati	41
3.3 Discussione dei risultati	43
Conclusioni	59

Introduzione

The tests themselves give no final verdict, but as tools help the worker who is using them to form his final decision.

— Neyman e Pearson, *Biometrika* (1928)

L'universo in cui viviamo è regolato da una moltitudine di processi complessi, la maggior parte dei quali è ancora ignoto all'uomo. La realtà che ci circonda è governata dal caso e dal caos: in un contesto di incertezza e imprevedibilità è difficile disporre di ricette o verità oggettive. Nonostante ciò, da sempre, l'uomo sente il bisogno di mettere ordine alla complessità della vita. Egli sente il bisogno di descrivere in modo preciso e oggettivo l'insieme dei fenomeni naturali, in particolare quelli che, proprio per la loro intrinseca complessità, lo spaventano ma allo stesso tempo affasciano. Le scienze naturali sono nate proprio con lo scopo di spiegare all'uomo ciò che egli non capiva e non era in grado di spiegare senza fare riferimento ad una qualche entità trascendente. Tra le scienze naturali, la fisica ha sempre rivestito un ruolo cruciale nel processo di rappresentazione critica della realtà. Nel corso dei secoli tale disciplina ha permesso di indagare fenomeni sempre più complessi e di elaborare teorie e modelli in grado di spiegare con enorme accuratezza anche le più piccole particelle elementari alla base della materia dell'intero universo. Fino ad alcuni decenni fa l'evoluzione delle discipline fisiche si giovava soprattutto di avanzamenti di natura teorica. Successivamente, grazie allo sviluppo delle tecnologie e al conseguente potenziamento di strumenti impiegati in esperimenti sempre più complessi, si è arrivati ad un cambiamento di paradigma in cui la fisica sperimentale è diventata principale protagonista degli avanzamenti della ricerca.

In questo contesto, è stato, ed è tuttora fondamentale il ruolo della statistica. La statistica permette di introdurre nelle descrizioni deterministiche della realtà esistente quell'incertezza e imprevedibilità che l'uomo ha cercato di sottrarre per cercare di ottenere un sistema logico che rispondesse a delle corrispondenze esatte e predeterminate. Come spiegato dallo statistico inglese Sir David Cox in un commento nel *Journal of the Royal Statistical Society A* (1995): *“The very word model implies simplification and idealization. The idea that complex physical, biological or sociological systems can be exactly described by a few formulae is patently absurd. The construction of idealized representations that capture important stable aspects of such systems is, however, a vital part of general scientific analysis and statistical models, especially substantive ones, do not seem essentially different from other kinds of model.”*

Proprio grazie alla collaborazione tra la comunità fisica e la comunità statistica, alla ricerca di un equilibrio da determinismo e casualità, tra realtà ed astrazione, è stato possibile procedere lungo la strada della conoscenza della complessità della configurazione della materia dell'universo. Nonostante ciò, è ancora lungo il percorso, molti gli aspetti che rimangono inesplorati e molte le domande a cui dare una risposta. Tra queste, una su tutte: esiste Nuova Fisica oltre a quella nota? E se sì, l'uomo è in grado di poterla conoscere e comprendere?

Come riportato nella pagina web¹ del CERN dedicata all'importanza del ruolo della statistica nella fisica delle alte energie: *“All of our physics results depend on statistical procedures that we perform on the data we obtain”*, commentano i membri della CMS Statistics Committee, aggiungendo: *“You rarely observe such a huge new signal that you do not need to quantify its size or whether it may be due to a fluctuation. You have to have a procedure with which you can give a mathematical interpretation of your data in terms of how significant your result is.”*

In questo contesto, la presente relazione si pone l'obiettivo di approfondire lo studio di alcune tecniche statistiche utilizzate nella fisica delle alte energie per l'identificazione di Nuova Fisica. Pur facendo principale riferimento alla fisica delle particelle, i temi trattati si applicano anche ad altri rami della fisica, quali, ad esempio, l'astrofisica

¹<https://cms.cern/news/importance-statistics-high-energy-physics>

o più in generale a tutte le discipline in cui è d'interesse identificare un possibile segnale raro in un processo solo potenzialmente noto.

Coerentemente con la letteratura, l'analisi si concentra sull'utilizzo di test statistici per dati di conteggio, costruiti a partire da istogrammi; di questi, si discutono i vantaggi e gli svantaggi, da un punto di vista teorico ed empirico.

Nel Capitolo 1 viene fornita una breve introduzione al contesto della fisica delle particelle e viene formalizzato il problema relativo all'identificazione di un segnale; in aggiunta, vengono brevemente discussi alcuni approcci standard per il confronto tra distribuzioni. In seguito, nel Capitolo 2, vengono illustrati alcuni approcci per l'identificazione di segnali basati sull'analisi di istogrammi, passando in rassegna numerosi test con crescente livello di complessità. Infine, nel Capitolo 3, viene proposto uno studio di simulazione per confrontare le tecniche discusse a livello teorico nei capitoli precedenti.

Capitolo 1

Il contesto di riferimento

1.1 La Fisica delle Particelle e il Modello Standard

La fisica delle particelle elementari, o fisica delle alte energie (*High Energy Physics*, HEP) è un branca della fisica che si occupa di studiare i costituenti fondamentali della materia, ossia le particelle elementari, e le loro interazioni. La ricerca nella fisica delle particelle elementari si occupa della formulazione delle leggi fondamentali della natura, che descrivano il più similmente possibile tutti i fenomeni esistenti e tutto ciò che costituisce l'universo.

Le questioni che affronta la fisica delle particelle sono le stesse che hanno guidato lo sviluppo del pensiero filosofico nel corso dei millenni. Quali sono i componenti ultimi della materia? Come si muovono? Che cosa li muove? L'origine storica della fisica delle particelle elementari può essere ricondotta ai filosofi greci, e in particolare a Democrito e ai pochi frammenti pervenuti dei suoi scritti sulla teoria degli atomi. Per Democrito l'atomo costituiva l'elemento originario e fondamentale dell'universo, nonché il fondamento metafisico della realtà fisica. Per il filosofo gli atomi erano quindi dei corpuscoli non ulteriormente divisibili e impercettibili che costituivano l'essenza della realtà esistente, a cui si poteva contrapporre solo il vuoto, ovvero l'assenza di atomi.

Il viaggio che ha portato dalla teoria degli atomi di Democrito alla moderna concezione della fisica della realtà che ci circonda è stato piuttosto lungo e ricco di scoperte rivoluzionarie, rese possibili grazie all'affermarsi del metodo scientifico. Ciò che è cambiato nel corso dei secoli è la percezione della vera complessità dei fenomeni, la capacità di eseguire misure quantitative, ed i criteri epistemologici di adeguatezza della loro descrizione. Nel XIX secolo viene introdotta una prima rudimentale teoria circa la struttura dell'atomo, relativa agli studi di Rutherford, che ha poi portato alla scoperta dell'elettrone (Thomson, 1897) e dell'esistenza del nucleo atomico (Rutherford, 1911). Negli anni successivi l'interesse verso questo campo scientifico cresce notevolmente e grazie all'evoluzione delle strumentazioni scientifiche si riesce a guardare dentro la materia con un dettaglio via via maggiore. Questo porta ad eccezionali scoperte. Dopo quella del protone (Thomson, 1919), l'individuazione dell'esistenza del neutrone (Chadwick, 1932) dà origine ad una nuova struttura atomica, in cui protoni e neutroni costituivano i nuclei. Le particelle elementari sembravano essere non più atomi, ma protoni, neutroni ed elettroni. Negli anni '30 vengono individuate nuove particelle nei raggi cosmici e nei primi acceleratori ad alta energia (principalmente adroni, cioè particelle soggette alla forza nucleare, come i nucleoni). Nel corso degli anni vengono portate alla luce particolari strutture di regolarità negli adroni e viene scoperta la struttura interna degli stessi. Le particelle elementari diventarono quindi i quark, costituenti i nuclei e altri adroni, e i leptoni, non soggetti all'interazione nucleare, come gli elettroni.

Ad oggi la formulazione teorica di riferimento è il Modello Standard (*Standard Model*, SM, Beringer, 2012), che a partire dagli anni '70, si è imposto come principale punto di riferimento nel contesto della fisica delle particelle. Si tratta di una teoria dei campi che, unita alla teoria di Einstein della gravità, fornisce una descrizione sorprendentemente accurata della maggior parte dei fenomeni osservati. Questo complesso *framework* teorico descrive le principali particelle elementari e tre delle quattro forze fondamentali presenti in natura: l'interazione elettromagnetica, quella debole (unificate nella cosiddetta interazione elettrodebole) e l'interazione forte.

Al tempo della prima formulazione del Modello Standard, molte delle particelle postulate da questa teoria generale non erano ancora

state scoperte empiricamente.

Solo recentemente, è stata data prova dell'esistenza di alcuni elementi mancanti alla teoria Modello Standard, tra cui il *top quark*, (CDF Collaboration, 1995), il *neutrino tau*, (DONUT Collaboration, 2001), e il *bosone di Higgs*, (ATLAS Collaboration, 2012; CMS Collaboration, 2012).

Queste scoperte rivoluzionare hanno portato il Modello Standard, con gli adattamenti apportati nel corso degli anni, ad essere la teoria alla base della fisica delle particelle più largamente accettata. Il Modello Standard rappresenta la perfetta sintesi tra teoria quantistica e relativistica: esso coniuga la meccanica quantistica e la teoria della relatività che costituiscono le basi della fisica moderna.

1.2 Gli esperimenti in HEP

1.2.1 Collisione di particelle negli acceleratori

Il Modello Standard è una complessa teoria che descrive con enorme accuratezza un piccolo insieme di particelle elementari e come esse interagiscono tra loro. Tuttavia, è noto che tale teoria è incompleta, in quanto non è in grado di spiegare alcuni fenomeni, quali l'esistenza della materia oscura e dell'energia oscura. Di conseguenza, la comunità scientifica cerca di elaborare nuove teorie che giustifichino la presenza di tali fenomeni, prevedendo ad esempio, l'esistenza di nuove particelle, o di perfezionare la conoscenza della fisica già nota. Per questi scopi, la comunità fisica progetta ed esegue esperimenti, che richiedono la produzione di un'enorme energia da concentrare in uno spazio infinitesimale. In un collisore particellare tale energia viene raggiunta facendo collidere particelle fra loro. Poiché la produzione di particelle di massa maggiore richiede collisioni più energetiche, i ricercatori sono alla costante ricerca di tecniche per costruire acceleratori sempre più potenti.

La produzione di nuove particelle rappresenta solo un aspetto della questione, in quanto tali particelle devono essere anche rilevate tramite un rivelatore. Per comprendere in modo preciso come le nuove particelle si potrebbero manifestare e come i rivelatori dovrebbero entrare in azione, gli scienziati simulano un grande numero di collisioni. Queste simulazioni si basano sia sulle leggi fisiche note, ossia sul

Modello Standard, sia sulle previsioni dei nuovi modelli che vengono esplorati. Sofisticati software simulano le collisioni tra particelle, la successiva produzione di particelle e il loro eventuale decadimento in particelle di massa inferiore. Essi simulano anche le traiettorie delle particelle, le interazioni con i materiali del rivelatore e le risposte elettroniche dei sistemi. Infine, un programma per la ricostruzione analizza come un eventuale segnale sarebbe stato osservato da questo ipotetico rivelatore, nello stesso modo in cui si sarebbe analizzata l'osservazione di un evento reale.

La progettazione, la prova e la costruzione di acceleratori e rivelatori possono richiedere molti anni di lavoro. Per esempio, la progettazione del Large Hadron Collider (LHC) è iniziata nel 1984 e la sua costruzione è stata portata a termine solamente nel 2008. L'acceleratore circolare lungo 27 km permette di osservare due fasci contro-rotanti di protoni che sono accelerati ad un'energia pari a 6.5 TeV e fatti collidere in 4 diversi punti. Allo stesso tempo, centinaia di fisici compongono dei team di lavoro per progettare e costruire i rivelatori che misurano le particelle prodotte nelle collisioni alla ricerca di nuovi fenomeni. Nel caso dell'LHC, ATLAS e CMS rappresentano due esempi di rivelatori multi-scopo, basati su un'architettura differente e analizzati da team diversi, ciascuno costituito da più di 3000 fisici. All'LHC sono presenti anche due rivelatori più specifici: LHCb, che studia la simmetria tra materia e anti-materia, e ALICE, che analizza le proprietà degli stati primordiali di quark e gluoni.

La produzione di particelle dotate di una grande quantità di massa è un evento piuttosto raro che può essere osservato meno di una volta su un miliardo di collisioni. Per questo, l'LHC è stato progettato per poter produrre più di un miliardo di collisioni al secondo. Il numero totale di collisioni viene usualmente espresso in *inverse femtobarn*, dove 1fb^{-1} corrisponde a circa 80000 miliardi di collisioni.

Ogni collisione genera migliaia di particelle che vengono analizzate dai rivelatori opportunamente posizionati. I rivelatori possono essere immaginati come enormi telecamere 3D, costituiti da circa 100 milioni di sensori, organizzati in diversi strati e in grado di fornire informazioni sulla posizione ed energia di ogni particella. Le particelle cariche che attraversano il sensore in un *tracker* producono un piccolo segnale elettrico, che viene amplificato e registrato assieme alla posizione del sensore. Un intenso campo magnetico piega le traietto-

rie delle particelle cariche, permettendo che il loro momento (ossia il prodotto di massa per velocità) sia misurato. L'energia delle particelle neutre e delle particelle cariche è misurata da specifici calorimetri, che sono disposti su diversi strati attorno al *tracker*.

I protoni sono costituiti da quark e gluoni. Quando due protoni collidono, la probabilità che due di questi costituenti si incontrino è molto bassa. Dall'altro lato, solamente queste interazioni ravvicinate sono in grado di produrre particelle massive. Tali eventi sono caratterizzati dallo spargimento di un gran numero di particelle in tutte le direzioni, le quali lasciano delle tracce nei diversi strati del rivelatore. Un sistema piuttosto rapido di filtraggio riesce a scartare molti degli eventi di scarso rilievo. Quindi, un insieme efficiente di circuiti elettrici misura l'energia rilasciata negli strati finali del rivelatore, in seguito sistemi automatizzati assimilano e sincronizzano l'informazione proveniente dalle diverse parti del rivelatore con lo scopo di individuare particelle dotate di molta energia, come muoni, elettroni e fotoni.

1.2.2 Dall'esperimento fisico al dato statistico

Attraverso i passaggi descritti, ogni evento grezzo, che contiene tutte le informazioni registrate dagli strati differenti del rivelatore, viene trasferito al data centre del CERN e diventa dato statistico.

La fase successiva, a partire dai dati grezzi, è quella relativa alla ricostruzione dell'evento, in cui si identificano le traiettorie delle particelle e si determinano i loro parametri, quali la direzione e la curvatura.

Se si immagina la collisione come un punto nello spazio, allora è possibile definire un sistema cartesiano in cui l'asse x è l'asse che passa per il punto di collisione e per il centro del rivelatore, l'asse y è l'asse che esce dal punto di collisione ed è ortogonale all'asse x , mentre l'asse z è tangente al fascio di collisione. Ogni collisione è quindi identificata originariamente da quattro variabili: l'energia e le tre componenti assiali del momento, rispettivamente rilevate dai calorimetri e dal rivelatore. I depositi di energia nei calorimetri vengono sommati e l'informazione proveniente dagli altri strati del rivelatore viene sfruttata per ricostruire oggetti fisici, quali elettroni, muoni e getti di particelle. Sommando l'energia di tutte le particelle misurate e confrontandola con l'energia della collisione, si procede stimando l'e-

nergia “mancante” (*missing energy*) delle particelle che sono sfuggite alla rilevazione, ad esempio i neutrini. Uno dei passaggi decisivi è relativo alla scelta di quali variabili utilizzare per le analisi successive: utilizzare solamente le variabili grezze (*low level features*) o sfruttare anche alcune quantità ottenute sulla base di esse e sulla conoscenza dei processi fisici (*high level features*). Le prime identificano le variabili direttamente misurate dai rilevatori o che possono essere derivate immediatamente a partire da queste (come, ad esempio, la posizione delle particelle o il momento), mentre le seconde sono basate su una combinazione delle *low level features* e permettono di evidenziare il diverso comportamento delle particelle. Nella maggior parte di casi, la scelta ricade nell'utilizzare contemporaneamente i due tipi di variabili.

La ricostruzione dell'evento viene usualmente ripetuta più volte, in modo da riuscire ad ottenere misurazioni più accurate delle traiettorie delle particelle. L'evento viene classificato secondo determinate caratteristiche fisiche e registrato come *event summary data* per le successive fasi dell'analisi. L'intero processo di ricostruzione degli eventi richiede uno sforzo computazionale non indifferente in cui potrebbero celarsi possibili errori nell'elaborazione dei dati e che, pertanto, deve essere eseguito con estrema attenzione, in quanto i dati prodotti sono quelli che verranno effettivamente analizzati per la convalida dei modelli teorici e per la ricerca di Nuova Fisica.

1.3 Oltre il Modello Standard

Come anticipato in precedenza, nonostante il Modello Standard abbia introdotto elementi di significativa novità e abbia ricevuto diverse conferme empiriche, la comunità scientifica ritiene che si tratti di una teoria incompleta, incapace, almeno per ora, di giustificare molti fenomeni, tra i quali la forza di gravità, la natura della materia oscura e dell'energia oscura.

Per queste ragioni, a partire già dagli anni Novanta, ci sono stati svariati tentativi di dimostrare l'inconsistenza della teoria del Modello Standard e diverse le nuove teorie proposte. Tutti i tentavi che mirano a superare le carenze del Modello Standard vanno sotto il nome di *Physics Beyond the Standard Model*, ovverosia di “fisica oltre il Modello Standard”.

Molti degli esperimenti come quelli descritti nella sezione precedente hanno come loro principale scopo proprio la ricerca di Nuova Fisica. Diversi enti di ricerca, tra cui le Collaborazioni ATLAS e CMS, svolgono un'intensa attività sperimentale per studiare i fenomeni che riguardano le particelle elementari, per scoprire nuove particelle e per far luce su teorie fisiche oltre il Modello Standard. Le collaborazioni che partecipano alla costruzione degli apparati sperimentali sono composte da centinaia o anche migliaia di fisici provenienti da tutto il mondo. Questo mostra come il campo della fisica delle particelle abbia bisogno della collaborazione e dello sforzo congiunto di ricercatori di diversi ambiti scientifici per poter sostenere il passo dell'evoluzione. In un contesto simile, la statistica trova terreno fertile, grazie alla forte multidisciplinarietà e al rigore nel metodo che la contraddistinguono.

Finora le ricerche svolte sul campo non hanno portato alla luce nessuna prova dell'esistenza di Nuova Fisica oltre a quella prevista dal Modello Standard. Tuttavia non è da escludersi l'ipotesi che eventuali particelle o interazioni sconosciute si nascondano in regimi cinematici o stati finali non ancora esplorati. Poiché la mole di fenomeni fisici da analizzare è enorme, la necessità è quella di sviluppare delle tecniche che permettano di indagare l'ingente quantità di dati prodotti dagli esperimenti in modo strutturato e automatizzato. Per questo motivo, la comunità scientifica è alla costante ricerca di qualche miglioramento alle procedure algoritmiche escogitate per poter analizzare tutti i possibili stati finali prodotti dalla collisioni tra particelle.

I due mondi a confronto sono quindi, da un lato, il *background* del Modello Standard, che rappresenta la fisica nota fino ad ora, e dall'altro lato, la fisica dell'ignoto, che suppone l'esistenza di segnali sconosciuti.

Inizialmente le prime ricerche sul campo sono state condotte principalmente con un approccio "modello-dipendente", basandosi sulla teoria descritta dal Modello Standard e cercando possibili conferme a questa teoria. Questo modo di procedere ha, tuttavia, prodotto scarsi risultati. Pertanto, negli ultimi anni si è sviluppato un approccio più generale, detto "modello-indipendente". Tale approccio cerca di limitare al più possibile ogni forma di assunzione e mira a ricercare nuovi segnali piuttosto che ad appoggiare qualche congettura fisica. Ad oggi le ricerche "modello-dipendenti" vengono quindi utilizzate per vali-

dare empiricamente nuovi modelli fisici di riferimento che integrano o sostituiscono la teoria del Modello Standard.

Ciò detto, la comunità scientifica continua tuttora ad interrogarsi se sia possibile migliorare la nostra conoscenza rimanendo all'interno del *framework* del Modello Standard o se invece sia necessario andare oltre al Modello Standard, costruendo un'intera nuova teoria.

Se si dovesse identificare un segnale sconosciuto, diverse sarebbero le domande da porsi. Quali sono le proprietà di questa nuova particella? Tale particella è stata prodotta da sola o in associazione ad altre particelle? La particella osservata è decaduta in altre particelle (ad esempio come nel caso del bosone di Higgs)? Ci sono altri esperimenti indipendenti in cui si è osservata la stessa particella con parametri comparabili? Quale modello teorico è coerente con la presenza di tale particella? La presenza di tale particella esclude la validità dei modelli fisici noti? E' necessario svolgere ulteriori analisi con l'ausilio di acceleratori più potenti? Tutte queste domande aprono la strada ad una maggiore conoscenza, comprensione, a nuovi modelli e nuovi esperimenti.

1.4 Formalizzazione del problema

Lo scopo principale di molti degli esperimenti relativi alla fisica delle particelle è quello di dimostrare l'eventuale esistenza di particelle sconosciute oltre a quelle previste dalle teorie fisiche finora note. L'idea di fondo è che oltre al processo di *background* (Y_b) descritto dal Modello Standard ci possa essere un processo sconosciuto (Y_s) che si nasconde in regioni non esplorate dello spazio parametrico delle variabili comunemente indagate negli esperimenti condotti. Si può dunque immaginare che il processo effettivamente osservato (Y_{bs}) sia potenzialmente "contaminato" dalla possibile presenza di questo segnale (Y_s). Più formalmente,

$$p_{bs}(y) = (1 - \lambda)p_b(y) + \lambda p_s(y) \quad (1.1)$$

ossia il modello che descrive il processo Y_{bs} è un modello mistura in cui compare certamente la componente relativa al processo di *background*, che ci si attende essere prevalente, e potenzialmente anche la componente relativa al segnale che, se presente, ci si attende si manifesti in piccola frazione $\lambda \geq 0$.

Se fosse vero quanto previsto dal Modello Standard, allora la componente di segnale dovrebbe essere assente. In termini del paradigma della verifica di ipotesi, l'ipotesi nulla di riferimento è quindi

$$H_0 : p_{bs}(\cdot) = p_b(\cdot) \iff \lambda = 0 \quad (1.2)$$

Se è vera H_0 allora nella distribuzione del processo osservato non è presente alcun segnale, se invece è falsa, allora significa che il processo osservato non è relativo al solo processo di *background*.

Si osservi che tutte le quantità che compaiono nel modello (1.1) non sono note a priori. Pertanto, nella pratica, la distribuzione del processo osservato $p_{bs}(\cdot)$ viene ricostruita a partire da dati sperimentali, mentre quella del *background* $p_b(\cdot)$ viene ottenuta a partire da simulazioni Monte Carlo, o mediante dati sperimentali selezionati in regioni in cui è nota l'assenza del segnale. Nel seguito, dunque, si assumerà di avere a disposizione un campione di numerosità n per i dati sperimentali $(y_1^{bs}, \dots, y_n^{bs})$ e un campione, per semplicità della stessa numerosità n , dalla distribuzione del *background* (y_1^b, \dots, y_n^b) .

1.5 Test per il confronto tra distribuzioni

Il problema del confronto tra le distribuzioni $p_{bs}(\cdot)$ e $p_b(\cdot)$ può essere affrontato con un'impostazione classica, facendo uso di un test per il confronto tra distribuzioni.

La distribuzione del processo di *background* $p_b(\cdot)$ viene usualmente assunta nota ma, nella pratica, essa viene ricostruita mediante simulazioni Monte Carlo di campioni da Y_b . Pertanto, per confrontare $p_{bs}(\cdot)$ e $p_b(\cdot)$ è possibile fare riferimento sia a test a due campioni che test di bontà di adattamento, privilegiando l'approccio non parametrico.

I test di bontà di adattamento verificano se un campione è compatibile con una certa distribuzione. I test di bontà di adattamento possono essere utilizzati anche per valutare il grado di comparabilità tra due campioni, e permettono di concludere se sia plausibile ritenere che tali campioni siano stati generati dalla stessa distribuzione.

In letteratura tra i test di bontà di adattamento più comuni, vi sono il test X^2 , il test di Kolmogorov-Smirnov, di Cramér-von-Mises e di Anderson-Darling.

Se si assume nota la funzione di ripartizione $F_b(\cdot)$ del processo di *background*, allora l'ipotesi nulla da saggiare è l'ipotesi

$$H_0 : F_{bs}(y) = F_b(y) \quad (1.3)$$

dove $F_{bs}(\cdot)$ è la vera, e ignota, funzione di ripartizione dei dati Y_{bs} . Risulta naturale quindi confrontare la funzione di ripartizione $F_b(\cdot)$ con l'equivalente quantità empirica di $F_{bs}(\cdot)$, ossia $\hat{F}_n(\cdot)$, ottenuta come:

$$\hat{F}_n(y) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(-\infty, y)}(y_i)$$

per $i = 1, \dots, n$.

Per testare H_0 si può considerare il processo empirico

$$W_b(y) = \sqrt{n}[\hat{F}_n(y) - F_b(y)] = \frac{1}{\sqrt{n}} \sum_{i=1}^n [\mathbb{1}_{(-\infty, y)}(y_i) - F_b(y)]$$

A partire da W_b si possono ottenere le usuali statistiche test per la bontà di adattamento

$$\begin{aligned} T_{KS} &= \sup_y |W_b(y)| \\ T_{CvM} &= \int |W_b(y)|^2 dF_b(y) \\ T_{AD} &= \int \left| \frac{W_b(y)}{F_b(y)(1 - F_b(y))} \right|^2 dF_b(y) \end{aligned}$$

rispettivamente la statistica test di Kolmogorov-Smirnov, di Cramér-von-Mises e di Anderson-Darling. Per ulteriori approfondimenti si veda Govindarajulu (2007).

I test di bontà di adattamento basati sulla funzioni di ripartizione empiriche presentano numerosi vantaggi rispetto agli usuali test per il confronto tra due campioni, come il t test o il test di Mann-Whitney per il confronto rispettivamente tra medie e mediane, in quanto sono in grado di individuare un insieme più ampio di possibili differenze tra le due distribuzioni confrontate.

Nonostante il test di Kolmogorov-Smirnov sia maggiormente usato, l'evidenza empirica mostra come in molti casi il test di Cramér-von-Mises e di Anderson-Darling risultino più potenti per un'ampia classe di ipotesi alternative.

Capitolo 2

La ricerca di segnali tramite l'analisi di istogrammi

2.1 Formalizzazione

I test per il confronto tra distribuzioni descritti nel Capitolo 1 possono essere utilizzati per valutare il livello di concordanza tra due campioni ma, in caso di rifiuto dell'ipotesi nulla, non danno indicazioni precise su i motivi che portano a tale rifiuto. Inoltre, questi metodi potrebbero rivelarsi meno efficienti nell'individuazione di segnali che si presentano in piccole porzioni dei dati, in quanto sono volti a determinare il livello complessivo di somiglianza tra le due distribuzioni confrontate.

Pertanto, l'approccio che viene seguito nella maggior parte dei casi per la ricerca dei segnali è leggermente diverso. In pratica, molti dei test vengono eseguiti costruendo l'istogramma relativo alla distribuzione sperimentale $p_{bs}(\cdot)$ e andando a valutare cella per cella (*bin by bin*) se il numero di eventi contenuti al loro interno è coerente con le aspettative della distribuzione $p_b(\cdot)$ secondo le ipotesi del Modello Standard. Questo significa che per ogni *bin* si vuole testare se l'osservazione sperimentale è coerente con la distribuzione attesa per quel *bin* o se, al contrario, c'è sufficiente evidenza empirica per sostenere che questa osservazione sia stata generata da una distribuzione differente, che potrebbe contenere una traccia del segnale.

Questo approccio dovrebbe essere più sensibile alla presenza di un eventuale segnale quando λ è prossimo a zero, e inoltre ha il vantaggio di permettere di capire dove potrebbe essere collocato l'eventuale fenomeno sconosciuto.

Il problema può essere contestualizzato nel modo seguente. Per semplicità nel seguito si limita il problema all'analisi univariata sebbene, in principio, il problema sia multivariato.

Dall'osservazione della variabile Y_{b_s} si costruisce l'istogramma \mathcal{H}_{b_s} con N bin,

$$\mathcal{H}_{b_s} = (n_{b_s,1}, \dots, n_{b_s,N}) \quad (2.1)$$

dove ciascuno *bin* contiene un certo numero di eventi $n_{b_s,i}$. Si può assumere che tale numero di eventi $n_{b_s,i}$ sia realizzazione di una variabile casuale $N_{b_s,i}$. Nel seguito, con un leggero abuso di notazione, quando si farà riferimento al singolo *bin* si ometterà il pedice i .

La più semplice modellazione del problema prevede di assumere che

$$N_{b_s} \sim \mathcal{P}(\mu_{b_s}) \quad (2.2)$$

Per tenere conto del fatto che N_{b_s} descrive, oltre al numero di eventi del *background*, anche possibili eventi del segnale, si può assumere che

$$\mu_{b_s} = \mu_b + \lambda\mu_s$$

dove

$$\mu_s = n_{s\bullet} \int_{bin_i} p_s(y) dy = \text{numero medio di eventi del segnale,}$$

$$\mu_b = n_{b\bullet} \int_{bin_i} p_b(y) dy = \text{numero medio di eventi del } background$$

con $\sum_i n_{s,i} = n_{s\bullet}$ e $\sum_i n_{b,i} = n_{b\bullet}$, mentre $0 \leq \lambda \leq 1$ rappresenta la forza del segnale. Dunque $\lambda = 0$ in assenza di segnale, mentre $\lambda = 1$ se l'intero segnale si concentra sul *bin* considerato. Per semplicità, si assume che λ sia binario, ossia $\lambda \in \{0, 1\}$.

Dunque il problema della verifica della presenza di un segnale può essere formalizzato come un test di ipotesi, in cui, per ogni *bin*

l'ipotesi nulla coincide con

$$H_0 : \mu_{bs} = \mu_b \quad (2.3)$$

che è equivalente alle ipotesi

$$H_0 : \lambda = 0 \quad \text{o} \quad H_0 : \mu_s = 0$$

Nel seguito, talvolta, ci si riferirà all'ipotesi nulla (2.3) come ipotesi *background-only* in quanto prevede la presenza del solo processo di *background*.

L'ipotesi alternativa, invece, prevede che

$$H_1 : \mu_{bs} = \mu_b + \mu_s$$

che è equivalente all'ipotesi

$$H_1 : \lambda = 1$$

Sotto l'ipotesi alternativa si avrà quindi

$$n_{bs} = n_b + n_s$$

o, qualora si volesse esplicitare la dipendenza dal *bin*

$$n_{bs,i} = n_{b,i} + n_{s,i}$$

In fisica delle alte energie ci si riferisce al problema appena formalizzato come “*counting experiment*”: si vuole valutare l'evidenza della presenza di un segnale in presenza di altre particelle provenienti da un processo di *background*.

Molti degli esperimenti condotti in fisica non portano ad ottenere nessuna nuova scoperta. Le motivazioni alla base di questo fatto sono essenzialmente due:

- la fonte del fenomeno ricercato non esiste;
- il fenomeno esiste ma l'esperimento condotto non è stato sufficientemente sensibile per riuscire a catturarlo.

Se fosse vera la prima situazione allora si potrebbe escludere il fenomeno dalle ricerche che verranno condotte successivamente. Se, invece, fosse vera la seconda, l'unica conclusione che si può trarre è che bisogna costruire un esperimento migliore.

Per queste ragioni, in fisica, il concetto di “significatività di un segnale” è molto importante: esso quantifica la forza dell'evidenza empirica a sostegno della presenza di Nuova Fisica. Nella fase di progettazione di un nuovo esperimento è fondamentale poter conoscere quale valore della significatività aspettarsi se nei dati fosse presente effettivamente un segnale, o almeno poterne determinare un valore limite. I valori scelti per la significatività risultano infatti determinanti per poter affermare o negare l'effettiva significatività del fenomeno osservato.

In fisica, la significatività viene usualmente quantificata mediante una statistica test tipicamente indicata con la lettera Z (o S), valore- Z o Z -value, in quanto vengono applicate opportune trasformazioni in modo che Z abbia distribuzione Gaussiana.

Tipicamente, come sottolineato da Cowan et al. (2011), per valutare la sensibilità all'eventuale presenza di un segnale, si fa riferimento al valore medio o mediano di Z sotto determinate ipotesi circa un possibile valore di μ_s , ovvero assumendo che oltre al processo di *background* sia presente anche un segnale nei dati. Utilizzare il valore mediano del valore- Z è preferibile in quanto è indipendente dalla metrica utilizzata ed è meno sensibile ad eventuali valori anomali ed errori di stima. Usualmente ci si riferisce a questa quantità come “significatività attesa” (*expected significance*). Si osservi che Z è funzione dei dati, e dunque è essa stessa una variabile casuale caratterizzata da una certa distribuzione campionaria. Dunque, sotto H_1 , la distribuzione di Z dipenderà dall'ipotesi fatta su μ_s .

2.2 I test di bontà di adattamento per il confronto tra istogrammi

I test di bontà di adattamento, brevemente discussi nel Capitolo 1, possono essere facilmente adattati per confrontare due istogrammi, in aggiunta può essere introdotta la statistica X^2 .

L'approccio statistico al confronto tra istogrammi si basa sul fatto che un istogramma sia interpretabile come punto di partenza per una stima empirica della funzione di densità di probabilità sottostante ai dati considerati. Più in generale, si definisce una misura di distanza tra i due istogrammi e si valuta la distribuzione di tale distanza sotto l'ipotesi nulla di uguaglianza dei due istogrammi. Esistono altre tecniche che non verranno discusse nel seguito, usualmente applicate in contesi relativi all'*image indexing* e all'*information retrieval*, come quelle proposte da Cha e Srihari (2002), che permettono di valutare la "somiglianza" tra due istogrammi basandosi su ragionamenti di natura esclusivamente geometrica.

Si considerino i due istogrammi, uno relativo al flusso di eventi del processo di *background*, \mathcal{H}_b , e l'altro relativo al flusso di eventi sperimentali introdotto precedentemente, \mathcal{H}_{bs} , con lo stesso numero N di *bin*. Si denoti con $n_{b,i}$ il numero di eventi per il *bin* i -esimo per l'istogramma \mathcal{H}_b e con $n_{bs,i}$ il numero di eventi per il *bin* i -esimo per l'istogramma \mathcal{H}_{bs} . Siano inoltre $n_{b\bullet} = \sum_{i=1}^N n_{b,i}$ il numero totale di eventi per l'istogramma \mathcal{H}_b e $n_{bs\bullet} = \sum_{i=1}^N n_{bs,i}$ il numero di totale di eventi per l'istogramma \mathcal{H}_{bs} .

Per valutare la compatibilità tra \mathcal{H}_b e \mathcal{H}_{bs} si potrebbe considerare la statistica X^2 introdotta da Pearson (1900). Essa viene impiegata nelle situazioni in cui è d'interesse valutare un confronto in termini di bontà di adattamento, omogeneità e indipendenza.

In particolare, per la sua semplicità interpretativa, la statistica X^2 è molto utilizzata in fisica in tutte le circostanze in cui si debbano confrontare due insiemi di dati, per valutarne il grado di compatibilità, o per verificare l'adattamento di un campione ad una certa distribuzione, nonostante, come evidenziato da Bohm e Zech (2017), in molte delle applicazioni siano disponibili test più accurati e potenti.

L'idea alla base dell'utilizzo della statistica X^2 di Pearson per la bontà di adattamento è di valutare una sorta di *distanza* fra frequenze attese e frequenze osservate. Se si assume nota la distribuzione $p_b(\cdot) = p_b^0(\cdot)$, e dato un campione da Y_{bs} , un modo generale per testare l'ipotesi nulla

$$H_0 : p_{bs}(y) = p_b^0(y)$$

consiste nel considerare la statistica

$$X^2 = \sum_{i=1}^N \frac{(n_{bs,i} - n_{b,i})^2}{n_{b,i}} \quad (2.4)$$

Se il numero di osservazioni in ciascun *bin* è sufficientemente grande, allora la statistica X^2 ha distribuzione approssimata χ_{N-1}^2 , e pertanto si rifiuta l'ipotesi nulla per valori grandi della (2.4).

I limiti nell'utilizzo della statistica X^2 riguardano il fatto che essa è strettamente legata all'arbitrarietà della scelta del numero di *bin*. Inoltre, se i conteggi negli intervalli non sono sufficientemente elevati, il ricorso alla distribuzione approssimata χ^2 non è corretto.

Se invece non si assume nota la distribuzione $p_b(\cdot)$, un modo per tenere conto della incertezza dei dati di *background* è considerare una variante della statistica X^2 per il confronto tra due campioni

$$X^2 = \sum_{i=1}^N \frac{\left(\frac{n_{bs,i}}{n_{bs\bullet}} - \frac{n_{b,i}}{n_{b\bullet}}\right)^2}{\frac{n_{bs,i}}{n_{bs\bullet}^2} + \frac{n_{b,i}}{n_{b\bullet}^2}} \sim \chi_{N-1}^2 \quad (2.5)$$

Come già sottolineato, la statistica X^2 mantiene la proprietà di avere distribuzione indipendente dal modello solo se il numero di conteggi in ogni *bin* è sufficientemente elevato. Negli istogrammi prodotti dagli esperimenti fisici, in particolare quando si considera il problema multivariato, questa ipotesi non è spesso soddisfatta, pertanto, come suggerito da Lyons (2008), la statistica X^2 non dovrebbe essere utilizzata.

In alternativa, come proposto da Porter (2008), è possibile modificare opportunamente le statistiche test di Kolmogorov-Smirnov, Cramér-von-Mises e Anderson-Darling. Indicando con

$$\hat{F}_b(y) = \sum_{j=1}^i n_j^b / n_{b\bullet}$$

$$\hat{F}_{bs}(y) = \sum_{j=1}^i n_j^{bs} / n_{bs\bullet}$$

per $i = 1, \dots, N$, le distribuzioni empiriche cumulate degli istogrammi per un punto y nel *bin* j -esimo, allora si ottengono le quantità

$$T_{KS} = \max_i |\hat{F}_{bs}(y_j) - \hat{F}_b(y_j)|$$

$$T_{CvM} = \frac{n_{b\bullet} n_{bs\bullet}}{(n_{b\bullet} + n_{bs\bullet})^2} \sum_{j=1}^N (n_j^b + n_j^{bs}) [\hat{F}_{bs}(y_j) - \hat{F}_b(y_j)]^2$$

$$T_{AD} = \frac{1}{n_{b\bullet} + n_{bs\bullet}} \sum_{j=1}^N \frac{n_j^b + n_j^{bs}}{(\sum_{i=1}^j (n_i^b + n_i^{bs}))(n_{b\bullet} + n_{bs\bullet} - \sum_{i=1}^j (n_i^b + n_i^{bs}))} \cdot \left\{ \frac{B^2}{n_{b\bullet}} + \frac{BS^2}{n_{bs\bullet}} \right\}$$

con

$$B = [(n_{b\bullet} + n_{bs\bullet}) \sum_{i=1}^j n_i^b - n_{b\bullet} \sum_{i=1}^j (n_i^b + n_i^{bs})]$$

$$BS = [(n_{b\bullet} + n_{bs\bullet}) \sum_{i=1}^j n_i^{bs} - n_{bs\bullet} \sum_{i=1}^j (n_i^b + n_i^{bs})]$$

dove, se per qualche *bin* il numero di eventi del *background* o del segnale è pari a 0, allora il loro contributo alla somma è nullo.

2.3 Approcci *naive* per la misura della significatività

Si consideri la situazione ideale rappresentata dal caso in cui il numero atteso di eventi per il *background* sia noto. Sotto l'ipotesi nulla (2.3), la (2.2) diventa

$$N_{bs} \sim \mathcal{P}(\mu_b)$$

Dunque si può direttamente calcolare la probabilità $P(N_{bs} \geq n_{bs})$ di ottenere n_{bs} eventi o più quando si assume nota l'intensità μ_b sommando le probabilità Poisson

$$P_{H_0}(N_{bs} \geq n_{bs}) = \sum_{i=n_{bs}}^{\infty} \frac{\mu_b^i e^{-\mu_b}}{i!} \quad (2.6)$$

Poiché valori grandi di n_{bs} rappresentano evidenza empirica a sfavore di H_0 la (2.6) viene considerata a tutti gli effetti un p -value.

Se il numero di eventi attesi per il *background* è sufficientemente elevato, un'alternativa possibile è la seguente. Sotto l'ipotesi (2.3)

$$N_{bs} \sim \mathcal{P}(\mu_b) \Rightarrow N_{bs} \simeq \mathcal{N}(\mu_b, \mu_b)$$

quindi

$$Z = \frac{N_{bs} - \mu_b}{\sqrt{\mu_b}} \simeq \mathcal{N}(0, 1)$$

Avendo osservato n_{bs} eventi, il p -value per l'ipotesi (2.3) risulta

$$p = 1 - \Phi\left(\frac{n_{bs} - \mu_b}{\sqrt{\mu_b}}\right) \quad (2.7)$$

Invece, se H_0 non è vera, si ha che $\mathbb{E}(Z) \neq 0$, in quanto

$$\mathbb{E}(Z) = \frac{\mu_s}{\sqrt{\mu_b}} \quad (2.8)$$

La spiegazione intuitiva di questo risultato è che la deviazione standard di n_{bs} sotto l'ipotesi *background-only* è pari a $\sqrt{\mu_b}$, pertanto la quantità $\mu_s / \sqrt{\mu_b}$ rappresenta il rapporto tra la dimensione del segnale e l'errore atteso su n_{bs} nell'ipotesi di assenza di segnale. Il valore atteso $\mathbb{E}(Z)$ nella pratica viene stimato in modo robusto attraverso la mediana di Z .

Questo modo di procedere, tuttavia, non è esente da limiti. Da una parte l'approssimazione alla distribuzione Gaussiana è basata sull'assunzione che la media $\mu_b + \mu_s$ sia sufficientemente elevata, dunque potrebbe non essere soddisfatta sulle code della distribuzione del processo osservato. Inoltre, massimizzare $\mu_s / \sqrt{\mu_b}$ potrebbe portare ad ottenere un numero di eventi attesi μ_b molto piccolo, ad esempio minore di 10^{-4} e un numero di eventi attesi μ_s ad esempio pari a 0.1. Ciò significherebbe che $\mu_s / \sqrt{\mu_b} = 10$, che sembrerebbe essere un buon risultato, ma nella pratica più di 9 volte su 10 non si osserverebbe nemmeno un evento.

Un modo alternativo di procedere, dato il modello (2.2), è ragionare in termini del log-rapporto di verosimiglianza. La funzione di verosimiglianza per μ_{bs} è data da

$$L(\mu_{bs}) = \frac{\mu_{bs}^{n_{bs}} e^{-\mu_{bs}}}{n_{bs}!} \quad (2.9)$$

da cui

$$\begin{aligned}
\Lambda &= -2\{l(\hat{\mu}_{bs}) - l(\mu_{bs}|H_0)\} = \\
&= -2\{l(\hat{\mu}_{bs}) - l(\mu_b)\} = \\
&= -2\{n_{bs} \log \hat{\mu}_{bs} - \hat{\mu}_{bs} - n_{bs} \log \mu_b - \mu_b\} = \\
&= -2\left\{n_{bs} \log \left(\frac{\hat{\mu}_{bs}}{\mu_b}\right) + \mu_b - \hat{\mu}_{bs}\right\}
\end{aligned}$$

Sostituendo a $\hat{\mu}_{bs}$ il numero di eventi osservato n_{bs} si ha che, per costruzione, $\Lambda \sim \chi_1^2$, da cui

$$Z = \sqrt{\Lambda} = \sqrt{2\left(n_{bs} \ln \frac{n_{bs}}{\mu_b} + \mu_b - n_{bs}\right)} \sim \mathcal{N}(0, 1) \quad (2.10)$$

dove Z è semplicemente la statistica test radice con segno associata a Λ .

Cowan et al. (2011) sottolinea che, essendo n_{bs} stima di massima verosimiglianza di $\mu_{bs} = \mu_b + \mu_s$, la quantità (2.10) può essere approssimata da

$$\sqrt{2\left((\mu_b + \mu_s) \ln \left(1 + \frac{\mu_s}{\mu_b}\right) - \mu_s\right)} = \frac{\mu_s}{\sqrt{\mu_b}}(1 + O(\mu_s/\mu_b))$$

dove l'uguaglianza deriva da uno sviluppo in serie di Taylor del logaritmo di μ_s/μ_b , e la precedente è nota come approssimazione di Asimov.

Dunque al limite, per $\mu_s \ll \mu_b$ la quantità (2.10) si riduce all'usuale espressione *naive* $\mu_s/\sqrt{\mu_b}$ trovata nella (2.8). Nonostante l'espressione $\mu_s/\sqrt{\mu_b}$ sia largamente utilizzata per una stima del valore- Z per valori elevati di $\mu_b + \mu_s$, il suo utilizzo è valido solamente sotto l'ipotesi che $\mu_s \ll \mu_b$. Dunque l'espressione completa (2.10) fornisce una migliore approssimazione quando non è soddisfatta tale ipotesi.

2.4 Approcci avanzati

2.4.1 Incertezza sul *background*

Come già accennato in precedenza, per quanto riguarda il processo di *background*, nonostante il fenomeno fisico Y_b sia noto, la distribuzione del *background* $p_b(\cdot)$ non è esattamente nota e viene ricostruita

empiricamente. Di conseguenza, come discusso in Linnemann et al. (2007), le previsioni per ciascun *bin* per il processo *background* soffrono dei limiti delle simulazioni Monte Carlo. In aggiunta, le misurazioni del segnale risentono delle incertezze strumentali, che introducono ulteriori fonti di incertezza e parametri di disturbo. Nel seguito, per semplicità, si prenderanno in considerazione solamente le incertezze sul *background*.

Dunque, nella maggior parte dei casi, il valore atteso del numero di eventi per il *background* non è noto con esattezza. Da un punto di vista statistico, ciò equivale a dire che la distribuzione del processo di *background* è caratterizzata dalla presenza di parametri di disturbo a cui, in HEP, ci si riferisce come “incertezze sistematiche”.

I modi in cui si può procedere sono sostanzialmente due: o si simulano effettivamente i dati per il *background* oppure si utilizzano informazioni ausiliarie provenienti da *bin* vicini al *bin* di interesse ma in cui si è certi dell’assenza del segnale (regioni di controllo). A partire da questi dati si ottengono delle stime per μ_b e le si utilizza sotto H_0 per i test descritti nel paragrafo precedente. Ad esempio, indicata con σ_b l’incertezza sistematica relativa al processo di *background*, un modo *naive* per tenere conto sia di questa che dell’errore statistico relativi a μ_b è modificare l’espressione (2.8) in

$$\mathbb{E}(Z) = \frac{\mu_s}{\sqrt{\mu_b + \sigma_b^2}} \quad (2.11)$$

dove la quantità σ_b^2 deve essere stimata in qualche modo. Questa quantità viene usualmente utilizzata come misura di riferimento della significatività Z nei casi in cui l’incertezza nel *background* non può essere trascurata.

Questo modo di procedere porta con sé diversi problemi in quanto non si tiene conto della variabilità delle stime. Inoltre, ci potrebbero essere ulteriori ragioni dell’incertezza sul *background* tra cui possibili errori di misurazione e altri errori sistematici.

2.4.2 Parametri di disturbo e verosimiglianza profilo

Uno degli approcci possibili che permettono di ottenere una stima di μ_b prevede l’utilizzo di una “misurazione ausiliaria”, ossia valutare il

numero di eventi del *background* n_b in una regione in cui non si attende poterci essere alcun segnale e in cui il numero medio di eventi possa essere messo in relazione con il numero di eventi del *background* per la misurazione n_{b_s} . In HEP la regione ausiliaria è tipicamente rappresentata da una regione laterale (*sideband*), vicina alla regione del segnale ma in cui possono verificarsi solo eventi dal *background*.

Per esempio, si può assumere che n_b sia una realizzazione di una v.c. di Poisson N_b di media $\tau\mu_b$

$$N_b \sim \mathcal{P}(\tau\mu_b)$$

dove τ rappresenta un fattore di scala che può essere supposto noto, ed è spesso legato all'efficienza dei rilevatori.

La funzione di verosimiglianza basata sulla misurazione n_{b_s} e la misurazione di controllo n_b può essere espressa, assumendo indipendenza, come il prodotto delle rispettive distribuzioni Poisson

$$L(\mu_s, \mu_b) = \frac{(\mu_s + \mu_b)^{n_{b_s}} e^{-(\mu_s + \mu_b)}}{n_{b_s}!} \frac{(\tau\mu_b)^{n_b} e^{-(\tau\mu_b)}}{n_b!} \quad (2.12)$$

Per l'inferenza su μ_s si può costruire un test basato sul log-rapporto di verosimiglianza profilo, considerando τ e μ_b come parametri di disturbo e μ_s come parametro di interesse. Passando alla log-verosimiglianza, si ottiene

$$l(\mu_s, \mu_b) = n_{b_s} \log(\mu_s + \mu_b) - \mu_s - \mu_b + n_b \log(\tau\mu_b) - \tau\mu_b$$

Uguagliando a zero la derivata parziale di $l(\mu_s, \mu_b)$ rispetto a μ_s si ottiene

$$\frac{\partial l(\mu_s, \mu_b)}{\partial \mu_s} = \frac{n_{b_s}}{\mu_s + \mu_b} - 1 = 0 \iff \hat{\mu}_s = n_{b_s} - \mu_b$$

La log-verosimiglianza profilo per μ_b dato μ_s risulta essere

$$l_{\mu_s}(\mu_b) = n_b \log(\tau\mu_b) - \tau\mu_b$$

Uguagliando a zero la derivata di $l_{\mu_s}(\mu_b)$ rispetto a μ_b si ottiene

$$\frac{\partial l_{\mu_s}(\mu_b)}{\partial \mu_b} = \frac{n_b}{\mu_b} - \tau = 0 \iff \hat{\mu}_b = \frac{n_b}{\tau}$$

infine uguagliando a zero la derivata parziale di $l(\mu_s, \mu_b)$ rispetto a μ_b

$$\frac{\partial l(\mu_s, \mu_b)}{\partial \mu_b} = -\frac{(1 + \tau)\mu_b^2 - (n_{bs} + n_b - (1 + \tau)\mu_s)\mu_b - n_b\mu_s}{(\mu_s + \mu_b)\mu_b} = 0 \iff$$

$$\hat{\mu}_b(\mu_s) = \frac{n_{bs} + n_b - (1 + \tau)\mu_s + \sqrt{(n_{bs} + n_b - (1 + \tau)\mu_s)^2 + 4(1 + \tau)\mu_s n_b}}{2(1 + \tau)}$$

Sotto l'ipotesi nulla (2.3) si ha $\mu_s = 0$, dunque

$$\hat{\mu}_b(0) = \frac{n_{bs} + n_b}{1 + \tau}$$

Per $n_{bs} > \mu_b$, la statistica test del log-rapporto di verosimiglianza profilo risulta essere pari a

$$\Lambda = 2\left\{\left(-n_{bs} \log\left(\frac{n_{bs} + n_b}{1 + \tau}\right) + (1 + \tau)\frac{n_{bs} - n_b}{1 + \tau} - n_b \log\left(\frac{\tau(n_{bs} + n_b)}{1 + \tau}\right)\right) + \left(n_{bs} \log n_{bs} - n_{bs} + n_b \log n_b - n_b\right)\right\}$$

da cui

$$\Lambda = 2\left\{n_{bs} \ln\left(\frac{(1 + \tau)n_{bs}}{n_{bs} + n_b}\right) + n_b \ln\left(\frac{(1 + \tau)n_b}{\tau(n_{bs} + n_b)}\right)\right\} \quad (2.13)$$

L'espressione (2.13) coincide con il noto risultato ottenuto da Li e Ma (1983).

Come proposto da Cowan et al. (2011), passando alla radice della (2.13) e sostituendo a n_{bs} e n_b i rispettivi valori attesi $\mu_s + \mu_b$ e $\tau\mu_b$ si ottiene l'approssimazione "Asimov"

$$\sqrt{2}\left\{(\mu_s + \mu_b) \ln\left(\frac{\mu_s + (1 + \tau)\mu_b}{(1 + \tau)(\mu_s + \mu_b)}\right) + \tau\mu_b \ln\left(1 + \frac{\mu_s}{(1 + \tau)\mu_b}\right)\right\}^{1/2} \quad (2.14)$$

Nel caso in cui la misurazione di controllo n_b sia caratterizzata da un'incertezza sistematica relativamente piccola, cioè per valori elevati di τ , allora l'espressione (2.13) si riduce all'espressione di Z per il caso di *background* noto.

L'espressione (2.14) può essere riscritta in termini dell'incertezza che si attribuirebbe al *background* sulla base della misurazione ausiliaria n_b . Osservando che la varianza di n_b è per costruzione uguale alla sua media, $\tau\mu_b$, allora la varianza di $\hat{\mu}_b$ risulta pari a

$$\sigma_b^2 = \text{Var}(\hat{\mu}_b) = \mu_b/\tau$$

Utilizzando tale espressione per eliminare τ dall'espressione (2.14) si ottiene

$$\sqrt{2} \left\{ (\mu_s + \mu_b) \ln \left(\frac{(\mu_s + \mu_b)(\mu_b + \sigma_b^2)}{\mu_b^2 + (\mu_s + \mu_b)\sigma_b^2} \right) - \frac{\mu_b^2}{\sigma_b^2} \ln \left(1 + \frac{\sigma_b^2 \mu_s}{\mu_b(\mu_b + \sigma_b^2)} \right) \right\}^{1/2}$$

che sfruttando le espansioni in serie di Taylor risulta asintoticamente equivalente a

$$\frac{\mu_s}{\sqrt{\mu_b + \sigma_b^2}} (1 + O(\mu_s/\mu_b) + O(\sigma_b^2/\mu_b)) \quad (2.15)$$

L'espressione semplificata (2.15) dunque si riduce all'espressione *naive* (2.11). Tale espressione ci si attende essere valida nei casi in cui $\mu_s \ll \mu_b$ e $\sigma_b^2 \ll \mu_b$. Il requisito che $\sigma_b^2 \ll \mu_b$ è equivalente a richiedere che $\tau \gg 1$. Ciò significa che l'espressione (2.15) è valida solo se il numero di eventi attesi nella regione di controllo è sufficientemente grande rispetto al numero di eventi del *background* che contribuiscono alla misurazione n_{b_s} .

Per valori elevati di μ_b le espressioni di Z ottenute nel caso di *background* noto e nel caso di presenza di incertezza danno risultati essenzialmente equivalenti. Tuttavia, nel caso di valori piccoli di μ_b , l'espressione (2.13) è l'unica che fornisce dei valori sufficientemente accurati. Questo perché al decrescere di μ_b , le quantità μ_s/μ_b e σ_b^2/μ_b aumentano, pertanto viene meno l'ipotesi $\mu_s \ll \mu_b$ e la validità dell'approssimazione alla quantità (2.11).

Come mostrato, ad esempio, da Cousins (2008) il problema può essere affrontato in maniera leggermente diversa riscrivendo la verosimiglianza (2.12) come prodotto della distribuzione Poisson per il numero totale di eventi $n_{b_s} + n_b$ per la probabilità binomiale che il numero totale di eventi sia ripartito così come osservato, ottenendo un'espressione analoga al risultato (2.13) fornito da Li e Ma (1983).

2.4.3 Modelli gerarchici e approcci bayesiani

Come discusso da Cousins (2008) un altro modo possibile per tenere conto delle incertezze sul *background* prevede di introdurre degli approcci bayesiani nelle usuali procedure frequentiste, dando luogo a delle “ricette” ibride. La combinazione dei due paradigmi tuttavia richiede cautela, come riportato dall’autore, in quanto, talvolta, questi approcci non hanno una solida giustificazione metodologica.

L’idea di fondo è quella di non vedere μ_b come un parametro da stimare ma considerare μ_b come una variabile casuale con una propria distribuzione, che può dipendere da un insieme di parametri θ non di diretto interesse. In sostanza

$$\mu_b \sim p_{\mu_b}(\cdot) = p_{\mu_b}(\mu_b^0, \theta)$$

A partire dalla distribuzione di N_{bs} e di μ_b si costruisce la distribuzione congiunta

$$p_{N_{bs}, \mu_b}(n_{bs}; \mu_{bs}, \mu_b^0, \theta) = p_{N_{bs} | \mu_b}(n_{bs}; \mu_{bs}) \cdot p_{\mu_b}(\mu_b^0, \theta) \quad (2.16)$$

A questo punto, a partire dalla distribuzione congiunta $p_{N_{bs}, \mu_b}(\cdot)$ è possibile costruire la funzione di verosimiglianza per μ_{bs} e μ_b^0

$$\begin{aligned} L(\mu_{bs}, \mu_b^0) &= p_{N_{bs} | \mu_b}(n_{bs}; \mu_{bs}) \cdot p_{\mu_b}(\mu_b^0, \theta) \\ &= \frac{(\mu_b + \lambda \mu_s)^{n_{bs}} e^{-(\mu_b + \lambda \mu_s)}}{n_{bs}!} \cdot p_{\mu_b}(\mu_b^0, \theta) \end{aligned}$$

Questo rappresenta il punto di partenza anche per la procedura proposta da ATLAS Collaboration (2011) per la scoperta del bosone di Higgs.

Un caso particolare di questo approccio è quello introdotto da Cousins (2008) e discusso da Vianello (2018) in cui si assume che μ_b abbia distribuzione Gaussiana, ossia

$$\mu_b \sim \mathcal{N}(\mu_b^0, \sigma_b^2)$$

Questo è comune nella pratica, in particolare quando la stima del numero di eventi per il *background* non è dovuta a una misurazione (ad esempio in una regione laterale) ma da un modello per il *background* o da una procedura che restituisce come valori di *default* il numero medio di eventi μ_b assieme alla deviazione standard σ_b .

La funzione di verosimiglianza che descrive la misurazione n_{bs} e la “misurazione” Gaussiana per il *background* è ottenuta come prodotto delle rispettivi distribuzioni

$$L(\mu_s, \mu_b^0) = \frac{(\mu_s + \mu_b^0)^{n_{bs}} e^{-(\mu_s + \mu_b^0)}}{n_{bs}!} \frac{1}{\sqrt{2\pi}\sigma_b} \exp\left\{-\frac{(\mu_b - \mu_b^0)^2}{2\sigma_b^2}\right\}$$

Passando alla log-verosimiglianza, si ottiene

$$l(\mu_s, \mu_b^0) = n_{bs} \log(\mu_s + \mu_b^0) - \mu_s - \mu_b^0 - \frac{(\mu_b - \mu_b^0)^2}{2\sigma_b^2}$$

Uguagliando a zero la derivata parziale di $l(\mu_s, \mu_b^0)$ rispetto a μ_s si ottiene

$$\frac{\partial l(\mu_s, \mu_b^0)}{\partial \mu_s} = \frac{n_{bs}}{\mu_s + \mu_b^0} - 1 = 0 \iff \hat{\mu}_s = n_{bs} - \mu_b^0$$

La log-verosimiglianza profilo per μ_b^0 dato μ_s risulta essere

$$l_{\mu_s}(\mu_b^0) = n_{bs} \log(n_{bs}) - n_{bs} - \frac{(\mu_b - \mu_b^0)^2}{2\sigma_b^2}$$

Uguagliando a zero la derivata di $l_{\mu_s}(\mu_b^0)$ rispetto a μ_b^0 si ottiene

$$\frac{\partial l_{\mu_s}(\mu_b^0)}{\partial \mu_b^0} = -\frac{(\mu_b - \mu_b^0)^2}{2\sigma_b^2} = 0 \iff \hat{\mu}_b^0 = \mu_b$$

infine uguagliando a zero la derivata parziale di $l(\mu_s, \mu_b^0)$ rispetto a μ_b^0

$$\frac{\partial l(\mu_s, \mu_b)}{\partial \mu_b} = -\frac{(\mu_b^0)^2 - (\mu_s + \mu_b - \sigma_b^2)\mu_b^0 + \sigma_b^2\mu_s - n_{bs}\sigma_b^2}{(\mu_s + \mu_b^0)\sigma_b^2} = 0 \iff$$

$$\hat{\mu}_b^0(\mu_s) = \frac{1}{2}\left(\mu_s + \mu_b - \sigma_b^2 + \sqrt{(\mu_s + \mu_b - \sigma_b^2)^2 - 4(\sigma_b^2\mu_s - n_{bs}\sigma_b^2)}\right)$$

Sotto l'ipotesi nulla (2.3) si ha $\mu_s = 0$, dunque

$$\hat{\mu}_b^0(0) = \frac{1}{2}\left(\mu_b - \sigma_b^2 + \sqrt{\mu_b^2 - 2\mu_b\sigma_b^2 + 4n_{bs}\sigma_b^2 + \sigma_b^4}\right)$$

Sfruttando ancora una volta il test basato sul log-rapporto di verosimiglianza profilo si ottiene che

$$\Lambda = 2 \left\{ n_{bs} \ln \left(\frac{n_{bs}}{\hat{\mu}_b^0} \right) + \frac{(\mu_b - \hat{\mu}_b^0)^2}{2\sigma_b^2} + \hat{\mu}_b^0 - n_{bs} \right\}$$

Come in precedenza, $\Lambda \sim \chi_1^2$, dunque $Z = \sqrt{\Lambda} \sim \mathcal{N}(0, 1)$, con

$$Z = \sqrt{2} \left\{ n_{bs} \ln \left(\frac{n_{bs}}{\hat{\mu}_b^0} \right) + \frac{(\mu_b - \hat{\mu}_b^0)^2}{2\sigma_b^2} + \hat{\mu}_b^0 - n_{bs} \right\}^{1/2} \quad (2.17)$$

L'espressione (2.17) risulta essere valida solo se $n_{bs} \gg \mu_b$.

Un approccio alternativo a quello appena discusso in cui a partire dalla distribuzione congiunta di N_{bs} e di μ_b si costruisce la rispettiva funzione di verosimiglianza, prevede di ottenere la distribuzione marginale di N_{bs} integrando rispetto a μ_b . La distribuzione condizionata di N_{bs} è tale che

$$p_{N_{bs}|\mu_b}(n_{bs}; \mu_{bs}) \sim \mathcal{P}(\mu_{bs}) \quad (2.18)$$

inoltre

$$\mu_b \sim p_{\mu_b}(\mu_b^0, \theta)$$

e sotto l'ipotesi nulla (2.3) si ha che $\mu_{bs} = \mu_b$.

Di conseguenza da

$$p_{N_{bs}, \mu_b}(n_{bs}; \mu_{bs}, \mu_b^0, \theta) = p_{N_{bs}|\mu_b}(n_{bs}; \mu_{bs}) \cdot p_{\mu_b}(\mu_b^0, \theta)$$

marginalizzando rispetto a μ_b

$$p_{N_{bs}}(n_{bs}) = \int p_{N_{bs}, \mu_b}(\mu_{bs}, \mu_b^0, \theta) d\mu_b = p_{N_{bs}}(n_{bs}; \mu_b^0, \theta) \quad (2.19)$$

Questa formulazione permette di ottenere direttamente un p -value per l'ipotesi (2.3) in quanto la quantità (2.19) può essere considerata a tutti gli effetti una statistica test, come nel caso (2.6).

Come discusso da Cousins (2008) ciò è equivalente a “pesare” il p -value relativo all'ipotesi *background-only* con la distribuzioni a

priori assunta per μ_b . Con questo approccio si ottiene la seguente espressione per il p -value

$$p = \int P(N_{bs} \geq n_{bs}) \cdot p_{\mu_b}(\mu_b^0, \theta) d\mu_b \quad (2.20)$$

Un caso particolare di questo approccio è quello in cui si assume che μ_b abbia distribuzione Gaussiana, ossia

$$\mu_b \sim \mathcal{N}(\mu_b^0, \sigma_b^2)$$

Questa scelta è alla base del test proposto da ATLAS Collaboration (2019) per l'identificazione di un segnale nel processo di *background*. Nel caso del test ATLAS ci si pone fin da subito sotto l'ipotesi nulla (2.3), dove $\mu_{bs} = \mu_b$.

In questo caso la distribuzione marginale di N_{bs} risulta essere

$$\begin{aligned} p_{N_{bs}}(n_{bs}; \mu_b^0, \sigma_b^2) &= \int_0^{+\infty} \frac{e^{-\mu_b} \mu_b^{n_{bs}}}{n_{bs}!} \cdot \mathcal{N}(\mu_b^0, \sigma_b^2) d\mu_b \\ &= \int_0^{+\infty} \frac{e^{-\mu_b} \mu_b^{n_{bs}}}{n_{bs}!} \frac{1}{\sqrt{2\pi}\sigma_b} \exp\left\{-\frac{(\mu_b - \mu_b^0)^2}{2\sigma_b^2}\right\} d\mu_b \end{aligned} \quad (2.21)$$

dove la distribuzione di μ_b è stata troncata ai soli valori positivi per assicurare la non negatività della media della distribuzione Poisson per il numero di eventi del *background*. Dunque, seppur coerente con le ipotesi fisiche, il modello gerarchico (2.21) non rappresenta un modello naturale per N_{bs} , in quanto è richiesto che μ_b assuma solo valori positivi, ma $\mu_b \sim \mathcal{N}(\mu_b^0, \sigma_b^2)$. In altre parole, la distribuzione $\mathcal{N}(\mu_b^0, \sigma_b^2)$ non coniuga in modo naturale con la media di una distribuzione Poisson, e ciò comporta che non è disponibile una forma chiusa maneggevole per la (2.21), diversamente da quanto accade per altri modelli misti (modello *Poisson-Lognormale*, *Poisson-Gaussiana Inversa*, etc.).

Si osservi che come caso estremo, se il valore atteso di N_{bs} fosse noto senza incertezza, cioè se $\sigma_b^2 = 0$, allora la componente Gaussiana potrebbe essere sostituita dalla Delta di Dirac.

Conseguentemente N_{bs} seguirebbe una pura distribuzione di Poisson di media $\mu_b = \mu_b^0$

$$p_{N_{bs}}(n_{bs}; \mu_b^0) = \int_0^{+\infty} p_{N_{bs}|\mu_b}(n_{bs}; \mu_b) \cdot \delta(\mu_b - \mu_b^0) d\mu_b = \mathcal{P}(\mu_b^0)$$

ritornando al caso precedente (2.6) di *background* noto.

Il test ATLAS Collaboration (2019) è volto a saggiare un'ipotesi puntuale sul valore di μ_b . Se si denota con μ_0 il numero medio di eventi attesi sotto le teorie fisiche note, è possibile costruire il seguente sistema d'ipotesi per la vera media della componente Gaussiana:

$$\begin{cases} H_0 : & \mu_b^0 = \mu_0 \\ H_1 : & \mu_b^0 > \mu_0 \end{cases} \quad (2.22)$$

Se fosse vera l'ipotesi nulla H_0 , significherebbe che la variabile casuale N_{bs} seguirebbe esattamente la distribuzione di cui sopra, con componente gaussiana di media $\mu_b^0 = \mu_0$. Se l'ipotesi alternativa è unilaterale destra, come quella considerata dal sistema di ipotesi (2.22), allora significa che la variabile casuale N_{bs} sotto H_1 è relativa ad un processo che genera mediamente un numero di eventi maggiore rispetto a quello specificato sotto H_0 . Se n_{bs} è il valore osservato per un test con regione di rifiuto unilaterale destra, e assumendo di conoscere σ_b^2 , poiché valori grandi di n_{bs} sono sintomo di disaccordo tra i dati e H_0 , una misura del grado di conformità dei dati all'ipotesi nulla è data da:

$$p = P(N_{bs} \geq n_{bs})$$

Di conseguenza:

$$\begin{aligned} p &= P(N_{bs} \geq n_{bs} | \mu_0, \sigma_b^2) = \\ &= P(N_{bs} = n_{bs} | \mu_0, \sigma_b^2) + P(N_{bs} = n_{bs} + 1 | \mu_0, \sigma_b^2) + \dots = \\ &= A \int_0^{+\infty} \sum_{k=n_{bs}}^{+\infty} \frac{e^{-\mu_b} \mu_b^k}{k!} \mathcal{N}(\mu_0, \sigma_b^2) d\mu_b \end{aligned} \quad (2.23)$$

dove, a causa delle approssimazioni numeriche per il calcolo esplicito della (2.23), si introduce la costante di normalizzazione

$$A = \frac{1}{\int_0^{+\infty} \sum_0^{+\infty} \frac{e^{-\mu_b} \mu_b^k}{k!} \mathcal{N}(\mu_0, \sigma_b^2) d\mu_b}$$

La costante di normalizzazione A assicura che $0 \leq p \leq 1$. Un valore di p prossimo a zero è indice di scarsa conformità dei dati all'ipotesi nulla H_0 , nella direzione prevista da H_1 .

In realtà l'interesse è quello di testare l'ipotesi nulla contro l'alternativa bidirezionale, cioè se il numero di eventi n_{bs} si discosta dalla distribuzione nulla di N_{bs} sotto le ipotesi della teoria fisica nota, in una qualsiasi direzione

$$H_1 : \mu_b^0 \neq \mu_0$$

Se fosse vera H_1 , per un test con regione critica bilaterale, significa che N_{bs} è relativa ad un processo che genera mediamente un numero di eventi significativamente maggiore o minore rispetto a quello specificato sotto H_0 . Il livello di significatività osservato per un test bilaterale è dato da:

$$p = 2 \cdot \min P\{(N_{bs} \leq n_{bs}), P(N_{bs} \geq n_{bs})\} \quad (2.24)$$

In Aktas et al., H1 Collaboration (2004), le due probabilità che compaiono nella definizione (2.24) del p -value sono espresse coerentemente con il caso unilaterale (2.23)

$$P(N_{bs} \geq n_{bs}) = A \int_0^{+\infty} \sum_{k=n_{bs}}^{+\infty} \frac{e^{-\mu_b} \mu_b^k}{k!} \mathcal{N}(\mu_0, \sigma_b^2) d\mu_b$$

e

$$P(N_{bs} \leq n_{bs}) = A \int_0^{+\infty} \sum_{k=0}^{n_{bs}} \frac{e^{-\mu_b} \mu_b^k}{k!} \mathcal{N}(\mu_0, \sigma_b^2) d\mu_b$$

Al contrario in ATLAS Collaboration (2017) e ATLAS Collaboration (2019) compare anche un termine additivo nel caso del p -value sulla coda sinistra, pertanto, tale probabilità risulta essere espressa come

$$P(N_{bs} \leq n_{bs}) = A \int_0^{+\infty} \sum_{k=0}^{n_{bs}} \frac{e^{-\mu_b} \mu_b^k}{k!} \mathcal{N}(\mu_0, \sigma_b^2) d\mu_b + \int_{-\infty}^0 \mathcal{N}(\mu_0, \sigma_b^2) d\mu_b$$

In ATLAS Collaboration (2019) viene presentata una giustificazione della presenza del termine additivo Gaussiano. Secondo gli autori, questo termine fornisce la “probabilità di non osservare alcun evento dato un valore atteso negativo per le fluttuazioni al ribasso

delle incertezze sistematiche”. Più formalmente

$$\begin{aligned} \int_{-\infty}^0 \sum_{k=0}^{n_{bs}} \lim_{\mu_b \rightarrow 0} \left(\frac{e^{-\mu_b} \mu_b^k}{k!} \right) \mathcal{N}(\mu_0, \sigma_b^2) d\mu_b &= \int_{-\infty}^0 \left(\sum_{k=0}^{n_{bs}} \delta_{k0} \right) \mathcal{N}(\mu_0, \sigma_b^2) d\mu_b \\ &= \int_{-\infty}^0 \mathcal{N}(\mu_0, \sigma_b^2) d\mu_b \end{aligned}$$

dove con $\delta_{k0}(\cdot)$ si è indicata la funzione Delta di Kronecker. Dunque l'introduzione del termine additivo Gaussiano sembrerebbe essere giustificata dalla necessità di tutelarsi dal caso in cui in numero di eventi del *background* risultasse essere troppo piccolo. Nonostante ciò, tale modifica non sembra avere una solida giustificazione statistica e porta il p -value (2.24), in alcuni casi limite, ad assumere valori maggiori di 1.

2.5 Estensione all'analisi su più *bin*

La maggior parte dei test descritti nei paragrafi precedenti si applicano sullo specifico *bin* dell'istogramma \mathcal{H}_{bs} in cui si ritiene si possa manifestare il segnale. Nella pratica, tuttavia, raramente la posizione del segnale è nota, e quindi la valutazione deve estendersi alla totalità dei *bin*.

Il problema di svolgere procedure di test *multi-bin* per la presenza o l'assenza di Nuova Fisica non è banale.

Uno degli stratagemmi utilizzati per evitare di incorrere nel rischio di concedere con troppa facilità l'acclarazione di Nuova Fisica, è quello di fissare una soglia per il p -value associato al rifiuto dell'ipotesi *background-only* (2.3) particolarmente stringente.

Nelle procedure di test di ipotesi risulta cruciale la scelta della probabilità con la quale si è disposti a rifiutare l'ipotesi nulla quando questa è vera, ossia quello che viene comunemente chiamato livello di significatività nominale α . Poiché ci si attende che l'eventuale segnale, se presente, compaia in una piccola proporzione, la potenza dei test è già di per sé non così elevata. Pertanto, fissare un α troppo piccolo potrebbe essere rischioso, in quanto ciò potrebbe portare a ridurre ancora di più la potenza del test. D'altra parte, fissare α troppo grande potrebbe portare a rifiutare H_0 con troppa facilità.

Nello scenario degli esperimenti fisici condotti, alcuni studi, Linemann et al. (2007), hanno evidenziato come fissare α ad un livello comunemente utilizzato in statistica, come 0.05 o 0.001 comporta un eccesso di fiducia nell'affermare la presenza di Nuova Fisica quando in realtà non è presente alcun fenomeno sconosciuto. Per questo motivo, nella fisica delle alte energie, si è soliti fissare α pari a $1 - \Phi(5) \simeq 2.86 \cdot 10^{-7}$, con $\Phi(\cdot)$ la funzione di ripartizione della normale standard. Spesso ci si riferisce a questo valore come soglia 5σ , e in fisica si parla di “significatività della scoperta” (*discovery significance*).

Questo rappresenta un tentativo di arginare il problema dei confronti multipli o, come viene chiamato in fisica, *look-elsewhere effect*.

Infatti, come spiegato in Lyons (2008), mentre la possibilità che si verifichi un'osservazione corrispondente alla soglia 5σ per uno specifico *bin* di \mathcal{H}_{b_s} è remota, bisogna ricordare che il numero N di *bin* di \mathcal{H}_{b_s} è probabilmente elevato. Dunque, bisogna prendere in considerazione la possibilità che una fluttuazione statistica così significativa si possa manifestare ovunque in \mathcal{H}_{b_s} .

Spesso, inoltre, gli istogrammi \mathcal{H}_{b_s} vengono analizzati utilizzando differenti criteri di selezione o cambiando la scelta del numero di *bin*. Pertanto, la possibilità di osservare una fluttuazione pari a 5σ non diventa poi così remota.

La conseguenza di fissare un livello di significatività nominale così basso è che si corre il rischio di trascurare la possibile presenza di segnali, inoltre, la rigidità con cui, in fisica, questa soglia viene considerata nella maggior parte degli esperimenti, a prescindere da quali siano le ipotesi sottostanti, potrebbe avere un impatto non indifferente nell'interpretazione dei risultati.

Come già sottolineato, la presenza di Nuova Fisica potrebbe manifestarsi in più regioni dell'istogramma \mathcal{H}_{b_s} . La cosa migliore sarebbe valutare la loro significatività in modo congiunto, ma ciò spesso risulta complicato. Una possibile soluzione è quella di combinare tra loro i p -value relativi all'ipotesi *background-only* nei *bin* di interesse. In letteratura sono stati proposti diversi approcci per la combinazione dei p -value associati a procedure di test multiple. Tra questi, la funzione combinante di Fisher, Elston (1991), è una delle più utilizzate.

Tale funzione permette di combinare N p -value in unica statistica test

$$p^F = -2 \sum_{i=1}^N \log(p_i)$$

la quale, sotto l'ipotesi nulla per ciascun test e di indipendenza tra le statistiche test, ha distribuzione χ_{2N}^2 . Se le statistiche test dovessero essere fra loro positivamente correlate, il rischio è che il p -value associato alla statistica p^F risulti troppo piccolo, e pertanto anticonservativo.

In generale, i problemi relativi alle procedura di combinazione di p -value riguardano:

- la presenza di effetti diversi: il p -value di una serie di *bin* valutati singolarmente potrebbe essere minore di quello richiesto per affermare la presenza di Nuova Fisica se tutti questi sono in disaccordo con quanto previsto dal Modello Standard;
- la non-unicità della combinazione: non esiste un modo solo con cui i diversi p -value possono essere combinati;
- distorsione da selezione: combinare i p -value non equivale a valutare congiuntamente la possibile presenza di Nuova Fisica, in particolare se vengono considerati solo quelli più “interessanti”.

La strategia che si basa su combinare i p -value è, ad esempio, stata adottata in ATLAS Collaboration (2019).

Un approccio alternativo a quello *bin per bin* appena esposto prevede di costruire una funzione di verosimiglianza complessiva per l'istogramma \mathcal{H}_{bs} . In pratica, questo significa assumere indipendenza tra i diversi *bin* di \mathcal{H}_{bs} ed ottenere una verosimiglianza complessiva moltiplicando le verosimiglianze associate agli N *bin* di \mathcal{H}_{bs} . In assenza di parametri di disturbo la verosimiglianza complessiva assumerà una forma del tipo

$$\begin{aligned} L_{\mathcal{H}_{bs}}(\mu_{bs}) &= \prod_{i=1}^N \frac{(\mu_{bs,i})^{n_{bs,i}}}{n_{bs,i}!} e^{-(\mu_{bs,i})} \\ &= \prod_{i=1}^N \frac{(\mu_{b,i} + \lambda \mu_{s,i})^{n_{bs,i}}}{n_{bs,i}!} e^{-(\mu_{b,i} + \lambda \mu_{s,i})} \end{aligned}$$

In presenza di eventuali parametri di disturbo per il *background*, invece, è possibile, ad esempio, utilizzare lo stesso approccio visto nel paragrafo (2.4.2) se si hanno a disposizione delle misurazioni ausiliarie. Come mostrato da Cowan et al. (2011) le regioni di controllo possono essere considerate come K *bin* addizionali di un istogramma \mathcal{H}_b con un numero rispettivo di eventi $(n_{b,1}, \dots, n_{b,K})$ tali per cui, per $i = 1, \dots, K$

$$\mathbb{E}[n_{b,i}] = g_i(\mu_b)$$

dipendenti solo da μ_b tramite una qualsiasi funzione $g_i(\cdot)$, estendendo il caso (2.12) visto in precedenza in cui $\mathbb{E}[n_{b,i}] = \tau\mu_{b,i}$.

Dunque la funzione di verosimiglianza per μ_{bs} basata sui dati di \mathcal{H}_{bs} e \mathcal{H}_b può essere scritta come

$$\begin{aligned} L_{\mathcal{H}_{bs}, \mathcal{H}_b}(\mu_{bs}) &= L_{\mathcal{H}_{bs}}(\mu_{bs}) \cdot L_{\mathcal{H}_b}(\mu_b) \\ &= \prod_{i=1}^N \frac{(\mu_{b,i} + \lambda\mu_{s,i})^{n_{bs,i}}}{n_{bs,i}!} e^{-(\mu_{b,i} + \lambda\mu_{s,i})} \cdot \prod_{j=1}^K \frac{(g_j(\mu_b))^{n_{b,j}}}{n_{b,j}!} e^{-g_j(\mu_b)} \end{aligned} \quad (2.25)$$

In alternativa, si può procedere come nel caso (2.16), e costruire la verosimiglianza complessiva come

$$L_{\mathcal{H}_{bs}}(\mu_{bs}) = \prod_{i=1}^N \frac{(\mu_{b,i} + \lambda\mu_{s,i})^{n_{bs,i}}}{n_{bs,i}!} e^{-(\mu_{b,i} + \lambda\mu_{s,i})} \cdot p_{\mu_{b,i}}(\mu_{b,i}^0, \theta) \quad (2.26)$$

Questa secondo approccio è quello che è stato adottato, ad esempio, in ATLAS Collaboration (2011). Seguendo tale approccio, si può considerare il test basato sul log-rapporto di verosimiglianza profilo Λ . Indicando con $p_{\Lambda}(\mu_{bs})$ la distribuzione della statistica test Λ , e con Λ^{oss} il valore osservato della statistica test su i dati a disposizione, allora il p -value dell'ipotesi *background-only* può essere scritto come

$$p = \int_{\Lambda^{oss}}^{\infty} p_{\Lambda}(\mu_{bs}) d\Lambda$$

dove la distribuzione $p_{\Lambda}(\mu_{bs})$ viene tipicamente ricostruita tramite simulazioni Monte Carlo. Alcuni risultati asintotici sulla distribuzione della statistica test del log-rapporto di verosimiglianza profilo Λ sono mostrati, ad esempio, in Cowan et al. (2011).

Capitolo 3

Analisi empirica

3.1 Obiettivi

La presente analisi empirica ha lo scopo di esplorare numericamente alcune delle procedure di test discusse nei capitoli precedenti. A partire dagli stessi dati e da un'impostazione comune, l'obiettivo è quello di valutare e confrontare le principali proprietà dei test, al variare di alcune condizioni. Il fatto di valutare le proprietà dei test sulla base di un'impostazione comune permette di confrontare i risultati in modo critico ed imparziale, non agevolando una procedura rispetto ad un'altra. Allo stesso tempo, studiare il comportamento dei test al variare delle medesime condizioni, permette di capire sotto quali specifiche ipotesi le procedure esplorate forniscono risultati soddisfacenti e, allo stesso modo, riuscire a coglierne i limiti. Se le specifiche assunzioni sui cui poggiano i test discussi in precedenza sono effettivamente verificate, l'attesa è che i test producano i risultati coerenti a quelli teorici, mentre si vuole cercare di determinare quali sono i risultati che si otterrebbero laddove, più realisticamente parlando, alcune di queste ipotesi venissero a mancare.

Con riferimento ai test esplorati, i principali obiettivi perseguiti sono:

- valutare l'accuratezza del test in termini di probabilità di commettere un errore I tipo;

- valutare la capacità del test di discriminare tra ipotesi nulla e alternativa, ovvero la sua potenza.

Queste due proprietà sono valutate per ogni test anche con riferimento a:

- l'intensità dell'eventuale segnale;
- numerosità campionaria;
- robustezza rispetto alla conoscenza del modello di *background*;
- conoscenza o meno dell'eventuale posizione del segnale.

3.2 Presentazione degli scenari esplorati

Nell'ottica di considerare una situazione più realistica possibile, è stato svolto uno studio di simulazione, fissando le condizioni generali di seguito illustrate. Coerentemente con quanto descritto nei capitoli precedenti, il processo osservato è stato generato da un modello mistura

$$p_{bs}(\cdot) = (1 - \lambda)p_b(\cdot) + \lambda p_s(\cdot)$$

dove si è scelto di descrivere la distribuzione del *background* mediante $p_b(\cdot) = \chi_3^2$, mentre la distribuzione del segnale è stata supposta essere $p_s(\cdot) = \mathcal{N}(\mu_s = 7.5, \sigma_s^2 = 1/9)$.

A partire da questo macroscenario, sono stati generati 500 campioni Monte Carlo, sotto le condizioni di seguito descritte

- (1) $\lambda \in \{0.001, 0.005, 0.01, 0.02, 0.03, 0.04\}$;
- (2) numerosità campionaria $n \in \{5000, 50000\}$;
- (3) per quanto riguarda la proprietà di robustezza dei test rispetto alla conoscenza del modello di *background* sono state considerate le seguenti condizioni:
 - (A) si è supposto di conoscere perfettamente $p_b(\cdot)$;
 - (B) si è supposto di conoscere solo parzialmente $p_b(\cdot)$ mediante l'osservazione di un campione generato da essa;

- (C) si è supposto di conoscere parzialmente e con un maggiore grado di incertezza $p_b(\cdot)$ mediante l'osservazione di un campione generato da una distribuzione con una variabilità maggiore di quella descritta dal modello. In particolare, si è scelto di generare i conteggi per ciascun *bin* da una distribuzione binomiale negativa di media pari al valore teorico dei conteggi sotto $p_b(\cdot)$ e fissando il parametro di forma pari a 2 volte il valore teorico dei *bin*.
- (4) per quanto riguarda, infine, la conoscenza o meno dell'eventuale posizione del segnale, ogni test è stato condotto sia sotto l'ipotesi di conoscere con esattezza il *bin* dove si sarebbe dovuto verificare il segnale, sia sotto l'ipotesi di non conoscere la posizione del segnale, e dunque considerando l'intero istogramma.

A partire dai campioni generati, sono stati costruiti gli istogrammi corrispondenti, fissando l'ampiezza dei *bin* costante a pari ad 1. Per quanto riguarda l'analisi sul singolo *bin*, i test considerati nello studio di simulazione sono stati i seguenti:

- test (2.6): il test basato sulla funzione di ripartizione Poisson, sotto l'assunzione di *background* noto;
- test (2.7): il test basato sull'approssimazione normale del test (2.6);
- test (2.10): il test basato sul log-rapporto di verosimiglianza corrispondente al test (2.6), sempre sotto l'assunzione di *background* noto;
- test (2.13): il test basato sul log-rapporto di verosimiglianza profilo che permette di stimare eventuali parametri di disturbo nella distribuzione del *background* supponendo di avere delle misurazioni ausiliarie per il *background*;
- test (2.17): il test basato sul log-rapporto di verosimiglianza profilo derivato a partire dalla distribuzione congiunta del processo osservato e dalla distribuzione del *background* supposta essere Gaussiana;

- test (2.23): il test proposto da ATLAS Collaboration (2019), derivato tramite marginalizzazione della distribuzione congiunta alla base del test (2.17).

Per quanto riguarda l'analisi sull'intero istogramma, i test valutati sono stati i seguenti:

- test (2.4): il test X^2 ;
- test (2.5): il test X^2 a 2 campioni;
- test (2.10), (2.13) e (2.17) estesi a tutti i *bin*;
- test (2.6), (2.7) e (2.23) estesi a tutti i *bin*.

I test (2.4) e (2.5) si applicano per costruzione all'intero istogramma mentre gli altri test sono stati estesi all'analisi su tutti i *bin* secondo i metodi descritti nel paragrafo 2.5: i test basati sulla verosimiglianza sono stati estesi, sotto l'ipotesi di indipendenza tra *bin*, costruendo la verosimiglianza complessiva degli istogrammi e ottenendo la statistica test come somma delle statistiche test relative all'analisi sul singolo *bin*. Al contrario, i test non basati sulla verosimiglianza, ossia i test (2.6), (2.7) e (2.23) sono stati estesi all'analisi su più *bin* utilizzando gli approcci di combinazione dei *p*-value. Nello specifico, tra le diverse tecniche presenti in letteratura, si è scelto di utilizzare la funzione combinante di Fisher, brevemente discussa nel paragrafo 2.5.

L'intera procedura di simulazione è schematizzata nella Figura 3.1. Il software scelto per svolgere operativamente le simulazioni è stato R Core Team (2022).

3.3 Discussione dei risultati

I grafici riportati nelle Fig. 3.1-3.12, riassumono le analisi svolte e mostrano il comportamento di ciascuno dei test in termini di probabilità di errore di I tipo e di potenza, al variare delle condizioni descritte nel paragrafo precedente.

Con riferimento al controllo della probabilità di commettere un errore di I tipo, le Fig. 3.2-3.8 riportano la funzione di ripartizione empirica dei *p*-value ottenuti al variare delle simulazioni. Sotto H_0 , un test accurato dovrebbe produrre, per ogni $0 < \alpha < 1$, una frazione

Simulazione

```
nsam ← numero di simulazioni;
n ← numerosità campionaria;
pb; ps ← distribuzione del background e del segnale;
λ ← proporzione del segnale;
Hb ← genera Hb teorico da pb (A);
for (ogni i fino a nsam) do
  yb ← genera un campione di numerosità n da pb;
  Hbemp ← ottieni Hb empirico (B);
  ybs ← genera un campione di numerosità n dalla mistura pbs;
  Hbs ← ottieni Hbs;
  Hbpiù var ← ottieni Hb introducendo più variabilità nel campione yb (C);
  ...
  esegui i test su Hbs sotto le condizioni (A), (B) e (C);
  ...
  results ← salva i risultati dei test;
end for
```

Fig. 3.1. Schema della procedura di simulazione.

di p -value pari ad α stesso, ovvero la distribuzione dei p -value dovrebbe essere uniforme. Al contrario, valori della funzione di ripartizione empirica dei p -value inferiori ad α indicano che il test è conservativo, poiché, sotto H_0 , l'ipotesi viene rifiutata in una proporzione di campioni Monte Carlo inferiore al valore di α nominale. Similmente, valori della funzione di ripartizione empirica superiori ad α , indicano che il test è anticonservativo, ossia che tende a rifiutare l'ipotesi nulla troppo spesso.

Sotto H_0 , quando i test sono effettuati sul singolo *bin* del segnale, il livello di accuratezza generale è tendenzialmente buono in tutti i test e non molto sensibile alla dimensione del campione (Fig. 3.2 e 3.6).

Tuttavia, guardando da più vicino il comportamento dei test (Fig. 3.3), si nota una tendenza generale anticonservativa dei test per valori piccoli (< 0.05) di α . Solo per i test (2.13), (2.17) e (2.23) e quando il *background* è completamente noto (scenario (A)), si osserva una tendenza più conservativa.

In generale, al crescere dell'incertezza sulla distribuzione del *background*, in tutti i test si osserva una tendenza a rifiutare l'ipotesi nulla con maggiore frequenza. Questo porta ad evidenziare il comportamento anticonservativo dei test (2.6), (2.7) e (2.13), e migliora invece

i risultati dei test (2.13), (2.17) e (2.23) che sotto l'irrealistica ipotesi di conoscenza perfetta di $p_b(\cdot)$ risultavano troppo conservativi.

Come è lecito attendersi, i risultati tendenzialmente peggiorano quando i test, progettati per la valutazione di un singolo *bin*, vengono estesi alla valutazione dell'intero supporto di Y (Fig. 3.4) mediante procedure basate su ipotesi di indipendenza tipicamente non soddisfatte. Anche in questo frangente, il comportamento migliore è quello dei test (2.13), (2.17) e (2.23), seppur mostrino una leggera tendenza ad essere anticonservativi quando il *background* presenta una variabilità maggiore rispetto a quanto atteso (scenario (C)).

Per quanto riguarda invece i test (2.4) e (2.5), si conferma il fatto che, nonostante i test riescano a soddisfare con maggiore facilità le ipotesi alla base, essi hanno una potenza che cresce più lentamente se confrontata con i test basati sulla verosimiglianza.

Con riferimento al comportamento dei test sotto H_1 , la tendenza dei test ad essere anticonservativi, osservata in precedenza, si traduce con dei livelli di potenza già elevati anche quando l'intensità del segnale è molto contenuta ($\lambda = 2\%$). Tutti i test riportano, quindi, buoni livelli di potenza (ottimi per n grande), con una crescita solo leggermente più lenta in quei test capaci di controllare meglio la probabilità di errore di I tipo.

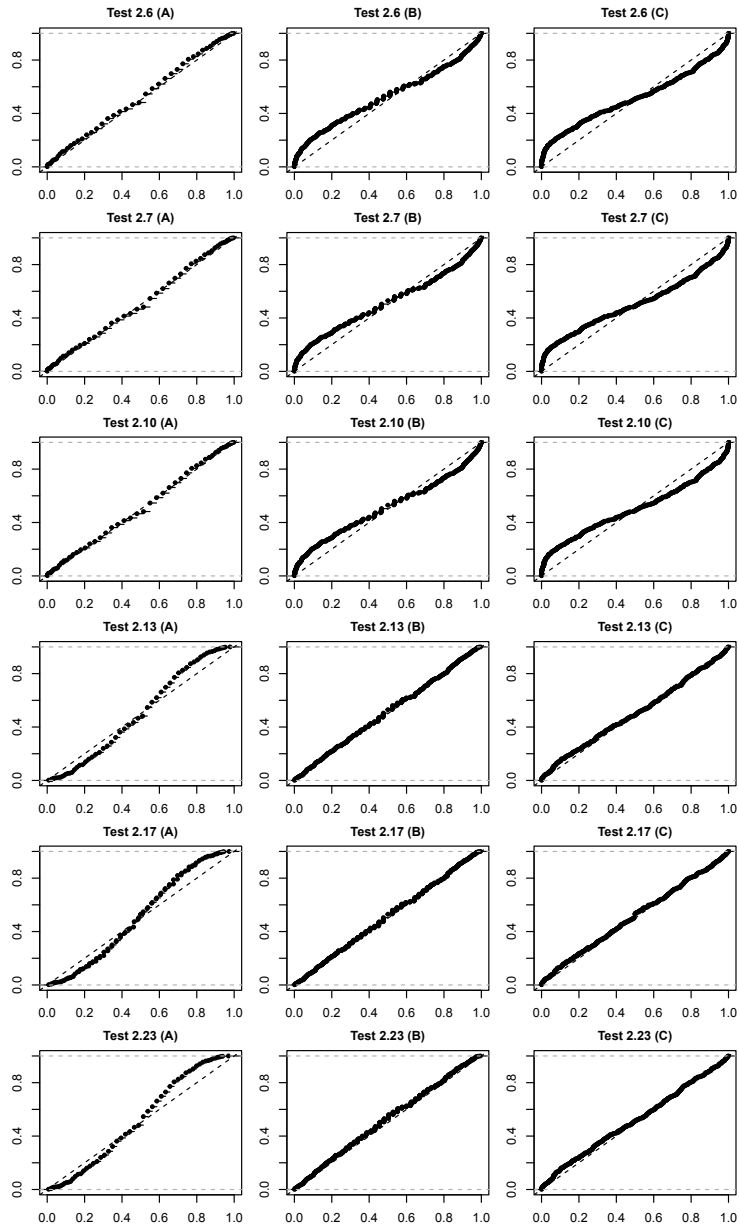


Fig. 3.2. Controllo dell'errore di I tipo: funzione di ripartizione empirica dei p -value ottenuti simulando sotto H_0 e sottoponendo a verifica l'esistenza di un segnale su un singolo bin ; $n = 5000$.

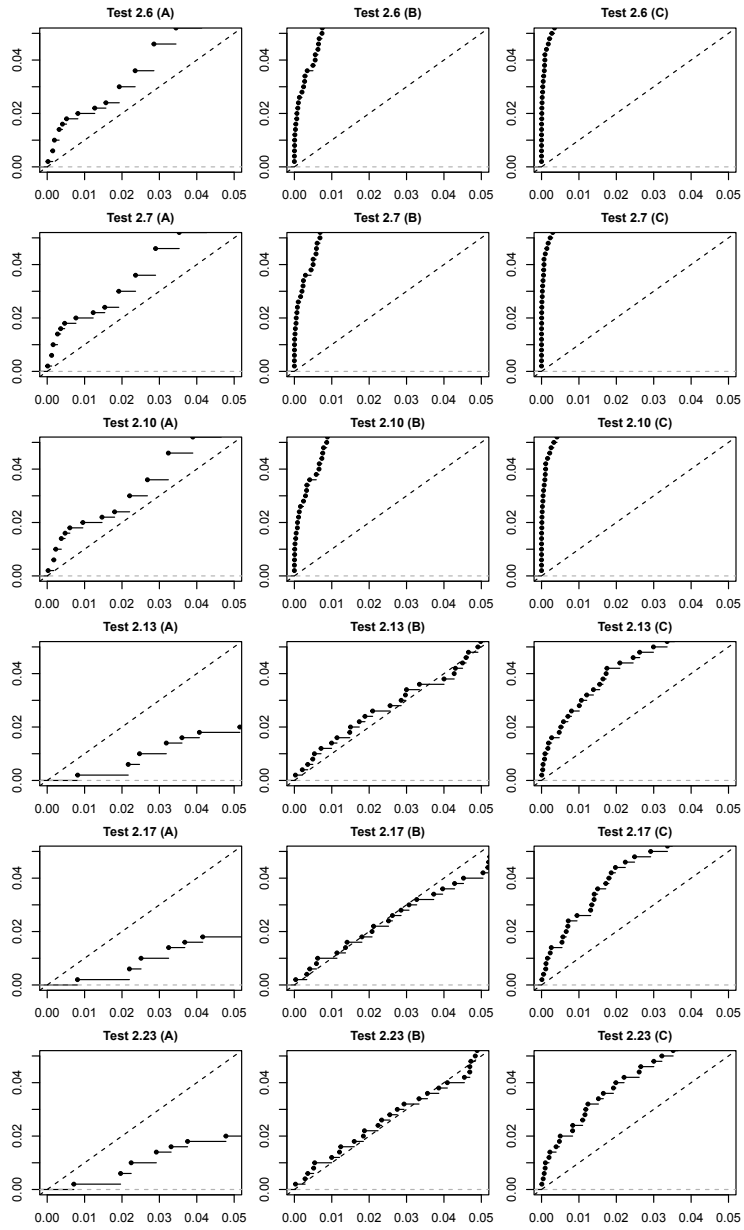


Fig. 3.3. Controllo dell'errore di I tipo: i grafici rappresentano un primo piano dei corrispondenti grafici riportati in Fig. 3.2 per valori di $0 \leq \alpha \leq 0.05$.

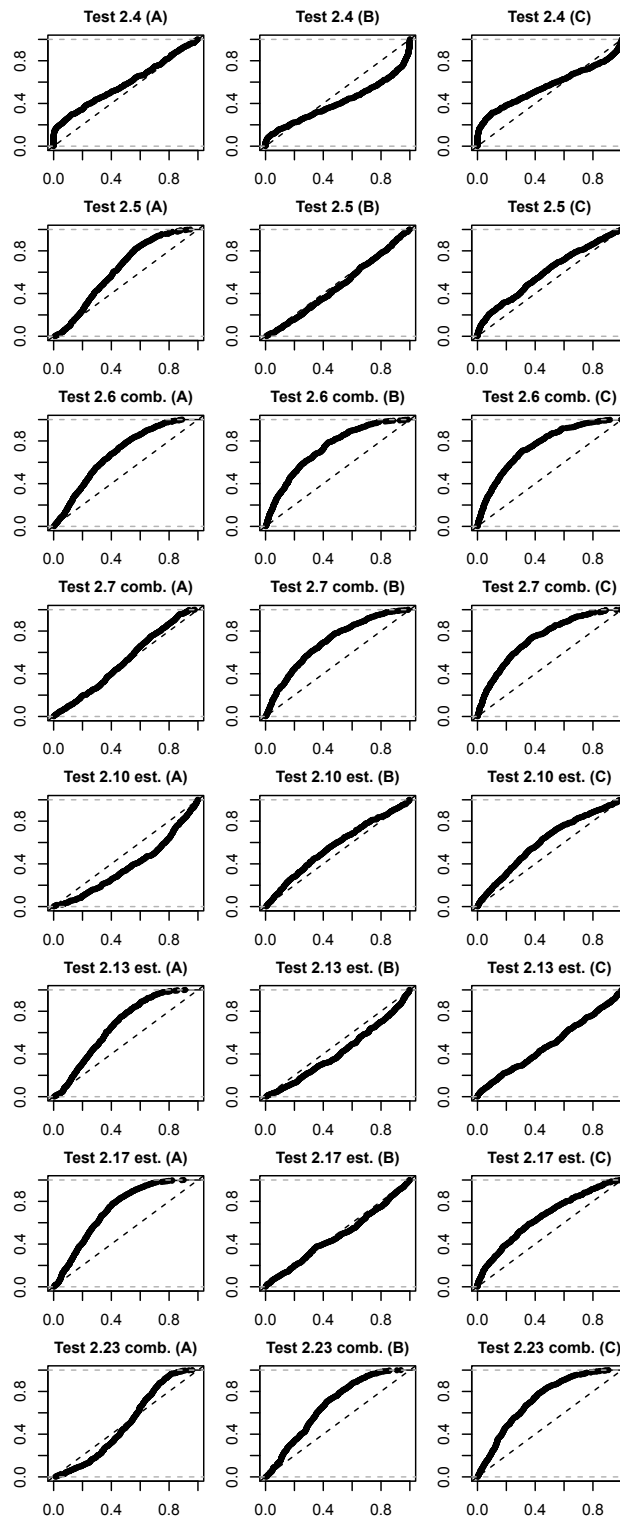


Fig. 3.4. Controllo dell'errore di I tipo: cfr. Fig. 3.2. Test estesi a tutti i *bin*; $n = 5000$.

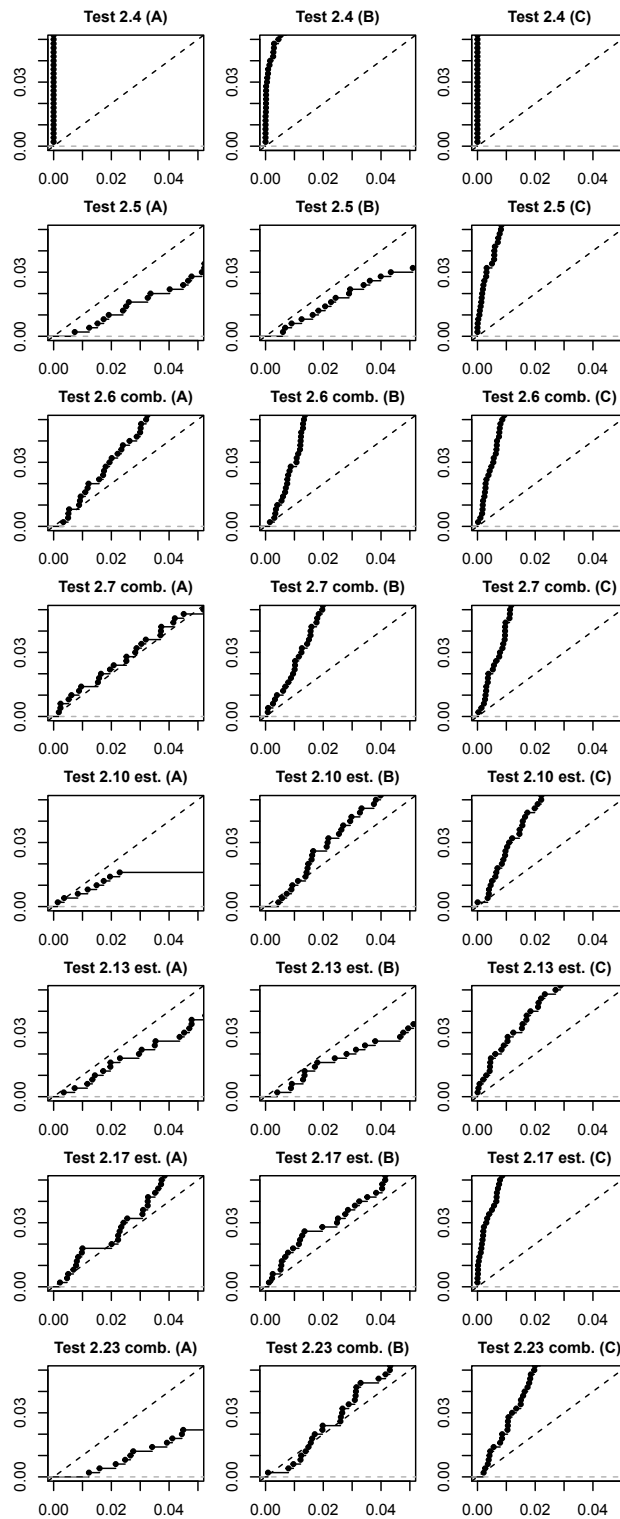


Fig. 3.5. Controllo dell'errore di I tipo: cfr. Fig. 3.3. Test estesi a tutti i *bin*; $n = 50000$.

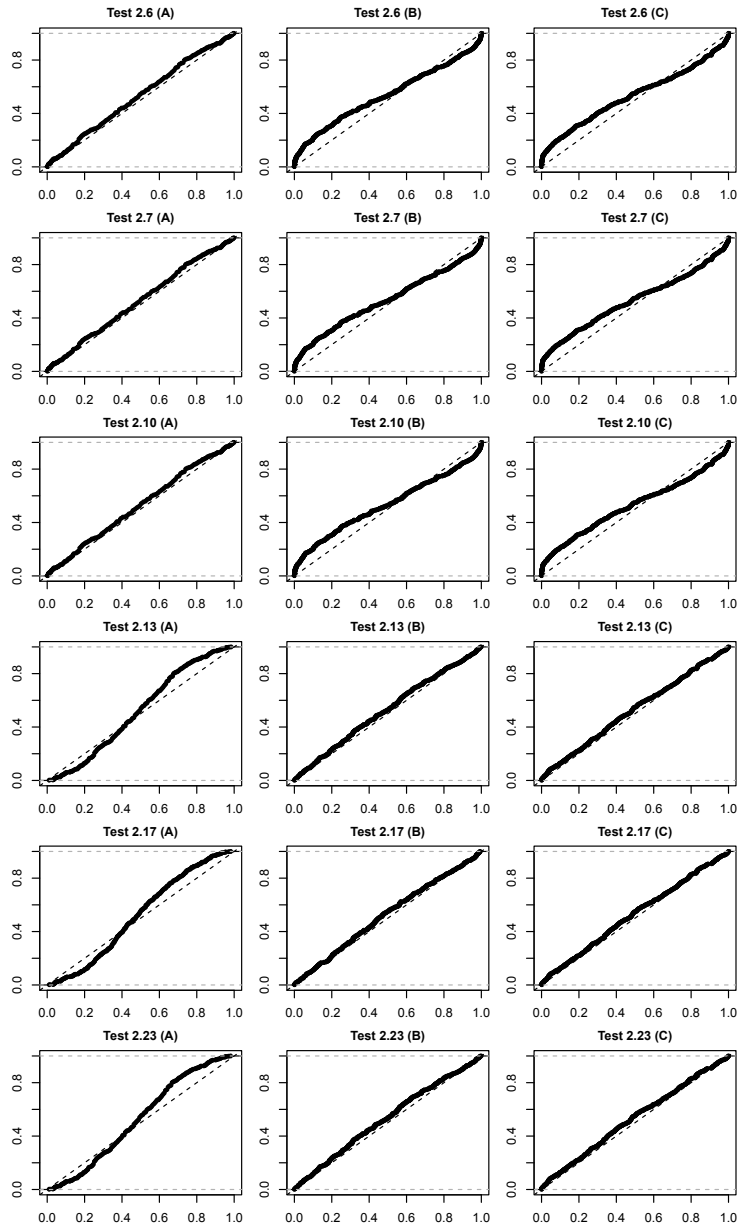


Fig. 3.6. Controllo dell'errore di I tipo: cfr. Fig. 3.2; $n = 50000$.

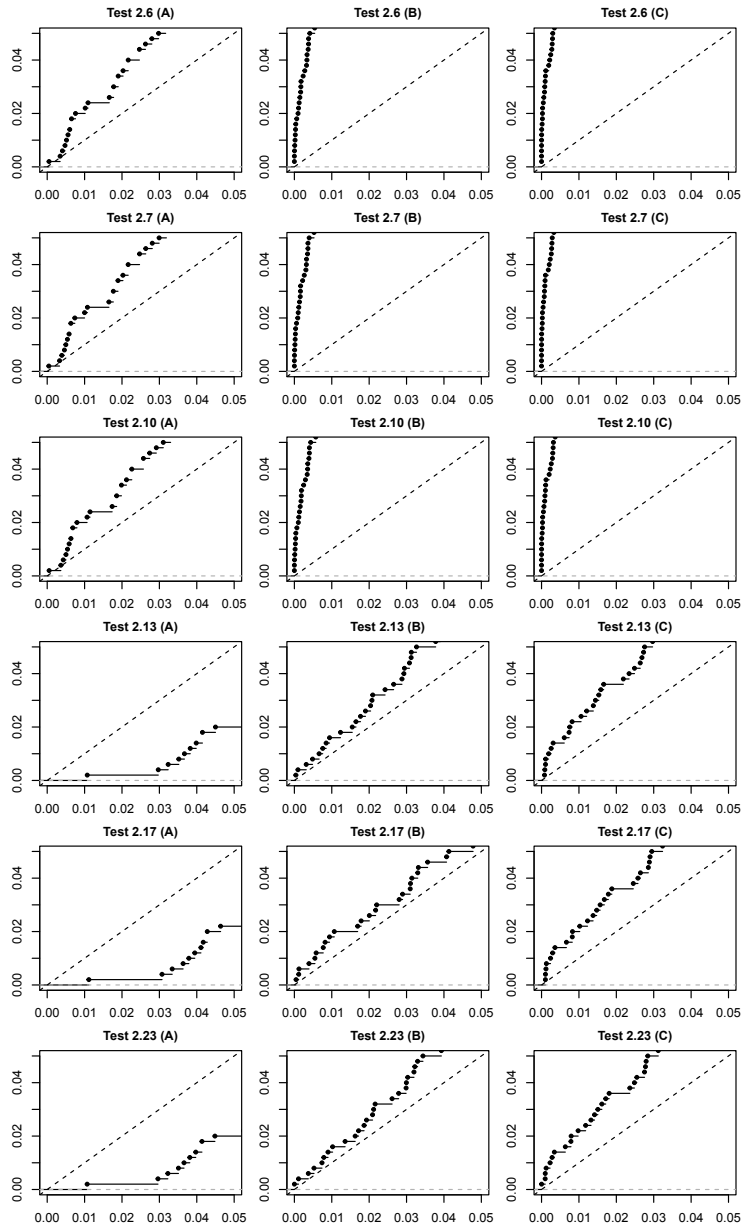


Fig. 3.7. Controllo dell'errore di I tipo: cfr. Fig. 3.3; $n = 50000$.

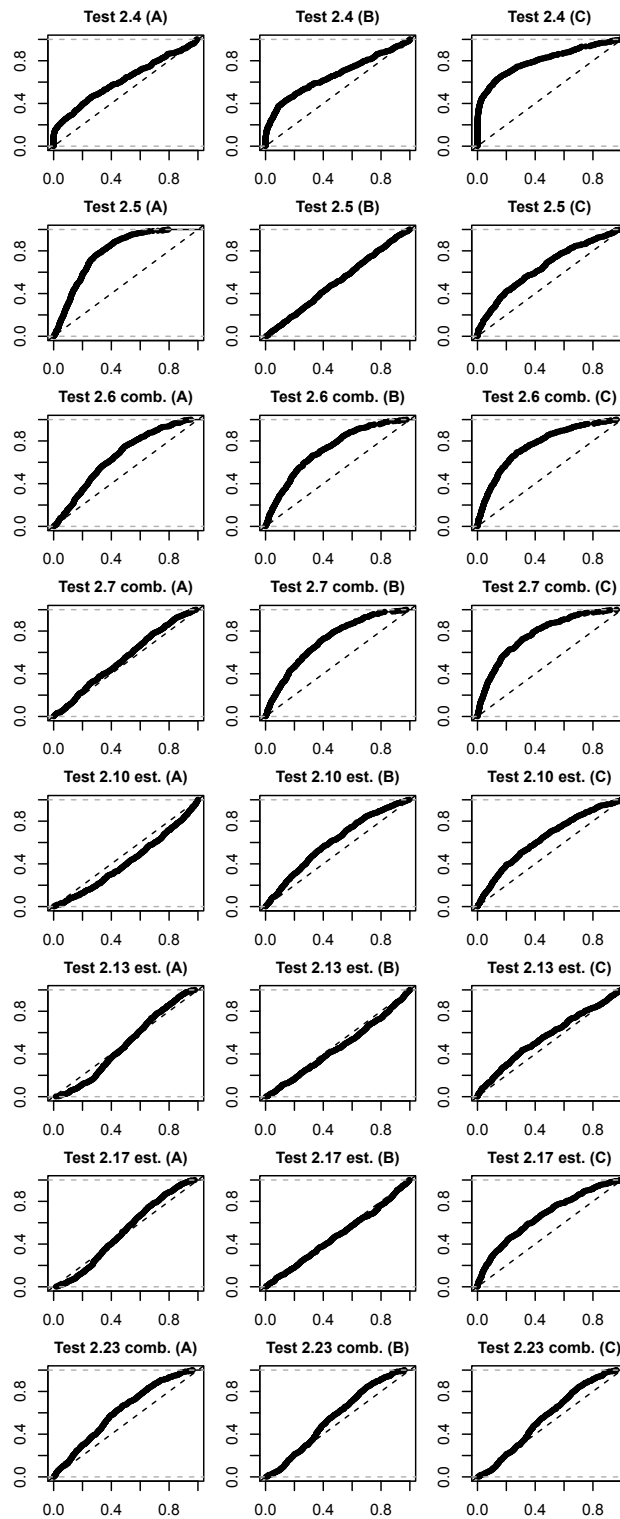


Fig. 3.8. Controllo dell'errore di I tipo: cfr. Fig. 3.4; $n = 50000$.

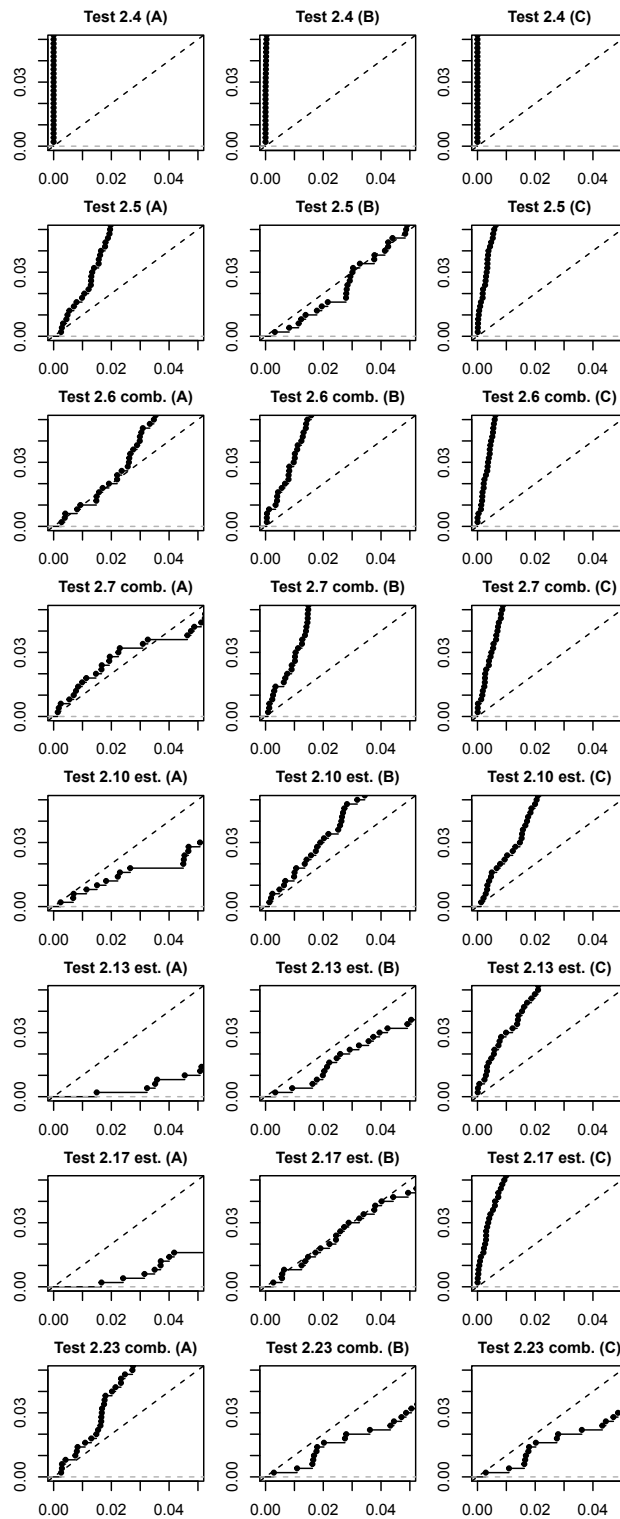


Fig. 3.9. Controllo dell'errore di I tipo: cfr. Fig. 3.5; $n = 50000$.

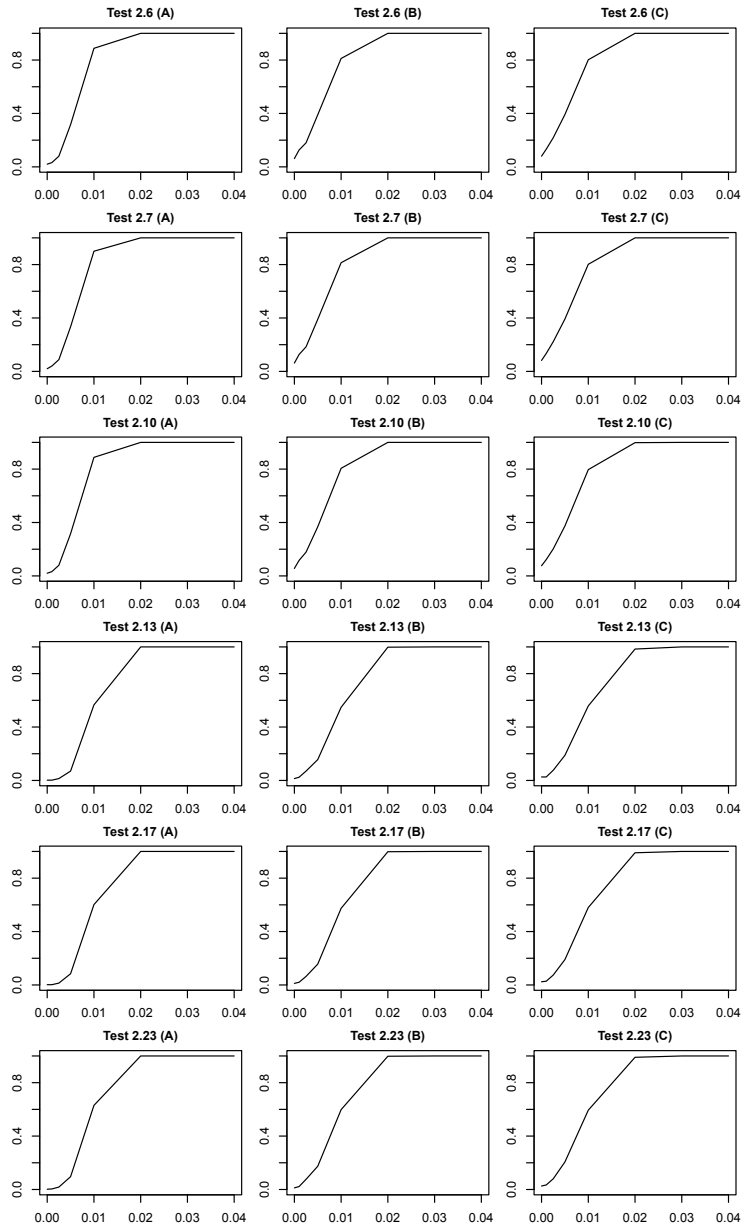


Fig. 3.10. Potenza dei test al variare dell'intensità del segnale. Risultati ottenuti sottoponendo a verifica l'esistenza di un segnale su un singolo *bin*; $n = 5000$.

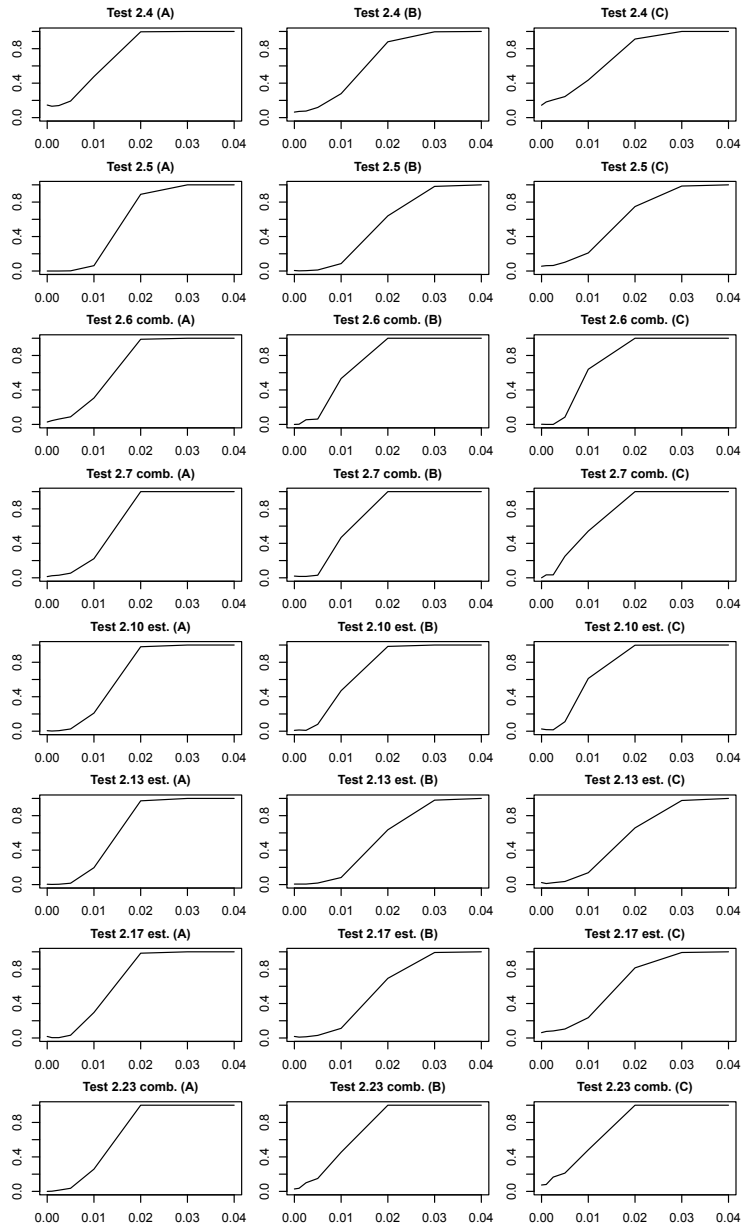


Fig. 3.11. Potenza dei test al variare dell'intensità del segnale. Test estesi a tutti i *bin*; $n = 5000$.

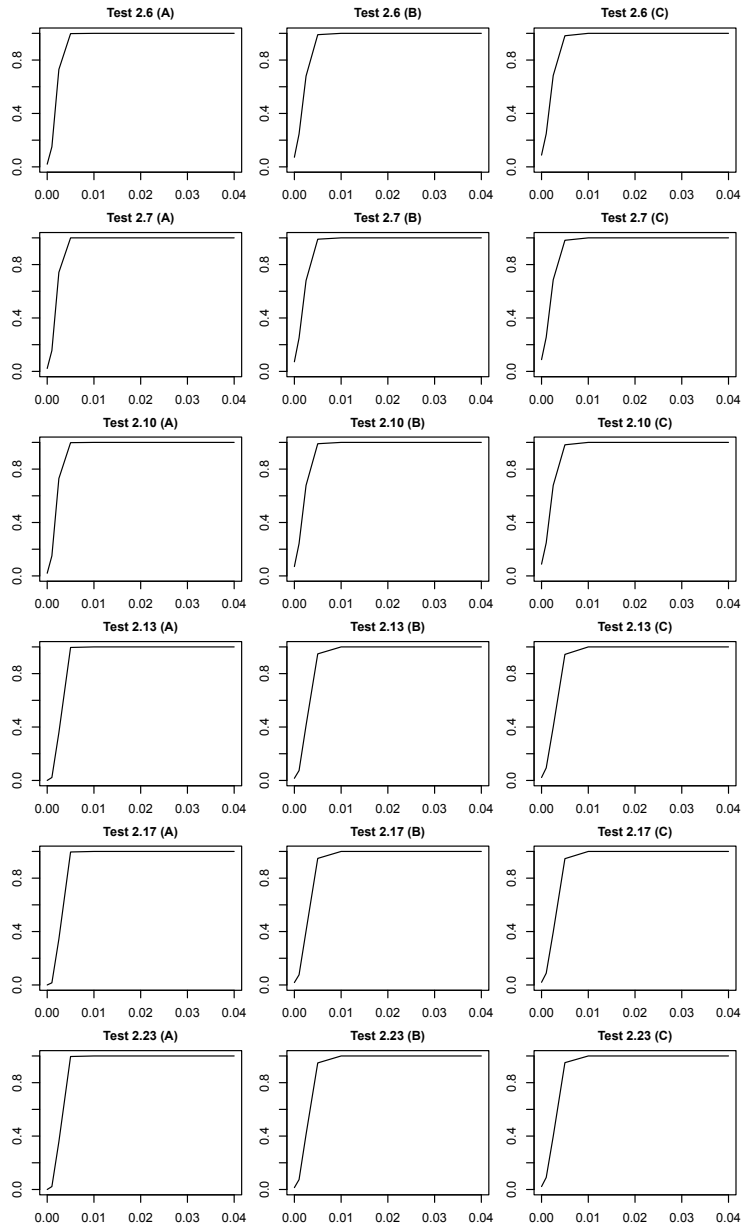


Fig. 3.12. Potenza dei test: cfr. Fig. 3.10; $n = 50000$.

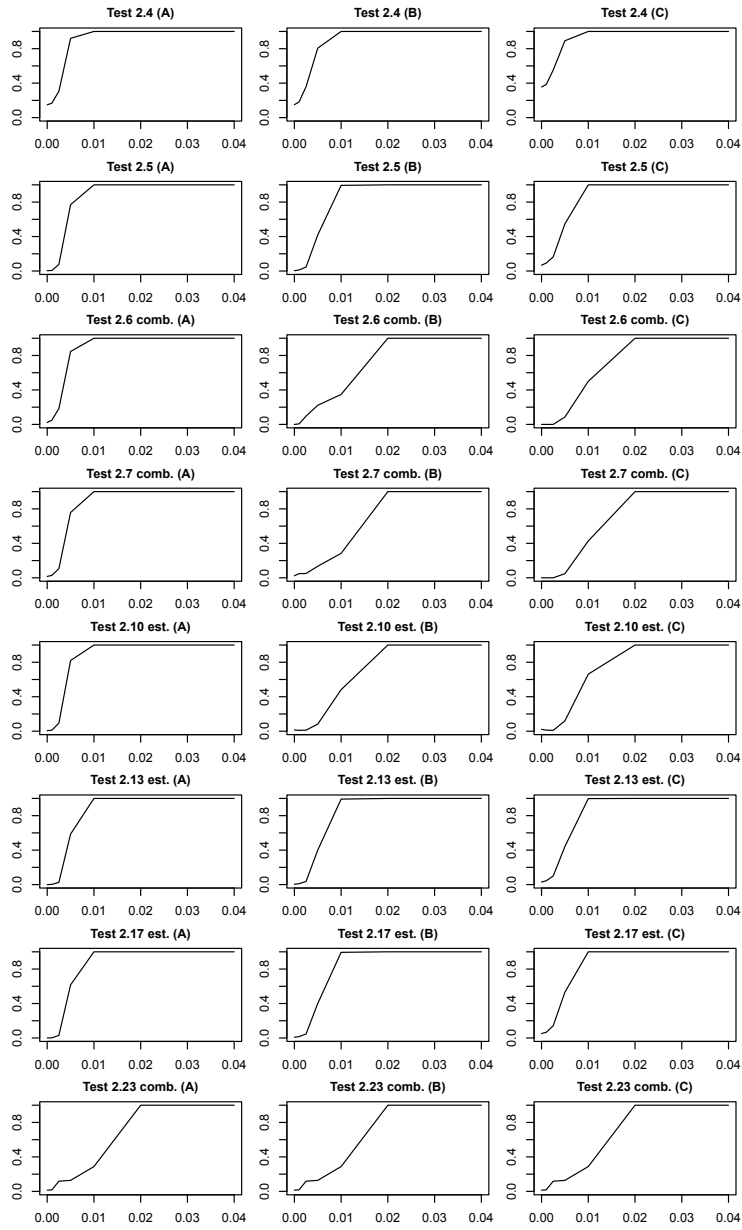


Fig. 3.13. Potenza dei test: cfr. Fig. 3.11; $n = 50000$.

Conclusioni

In questa tesi si sono esplorate alcune tecniche statistiche utilizzate nella fisica delle alte energie per l'identificazione di Nuova Fisica, in particolare, test per dati di conteggio costruiti a partire da istogrammi, con lo scopo di evidenziare l'eventuale presenza di un segnale nella distribuzione del processo di *background*. Queste tecniche sono state analizzate sia da un punto di vista teorico, descrivendone le ipotesi e il metodo, sia da un punto di vista empirico, mediante uno studio di simulazione.

Molti degli approcci esplorati si sono dimostrati validi per la ricerca di Nuova Fisica, seppur in alcuni casi si è potuto notare come questi fossero strettamente legati alle assunzioni sottostanti, tipicamente non verificabili. In aggiunta, i risultati empirici hanno mostrato come i test più complessi non portano sempre ad ottenere risultati di gran lunga migliori se confrontati con test statistici più semplici, come i test X^2 .

I test presentati, seppur non coprono l'intera varietà dei test presenti nella recente letteratura relativa all'analisi di dati di conteggio per la ricerca di segnali, mostrano come l'intrinseca difficoltà del problema sia stata affrontata arricchendo via via con elementi di maggiore complessità alcune procedure standard, con lo scopo di superarne i limiti ma con l'effetto collaterale di dare vita a degli approcci spesso tortuosi e di cui risulta talvolta arduo riconoscere e verificare le assunzioni sottostanti.

Miglioramenti alle tecniche finora elaborate potrebbero derivare da una più stretta collaborazione tra la comunità fisica e la comunità statistica. Una compenetrazione più profonda tra le due discipline potrebbe, infatti, da un lato, permettere alla comunità fisica di entrare in contatto anche con approcci e metodologie differenti da quelle esplorate fino a questo momento. Dall'altro lato, ciò permetterebbe

alla comunità statistica di conoscere più a fondo i dati fisici, per loro natura complessi, e di riuscire a plasmare in modo più consapevole le procedure statistiche di analisi, suggerendo possibili modifiche a quelle attuali o integrandole con nuovi metodi.

Bibliografia

- Aktas et al., H1 Collaboration (2004). A general search for new phenomena in ep scattering at hermes. *Phys.Lett. B* **602**, **14**.
- ATLAS Collaboration (2011). Procedure for the LHC Higgs boson search combination in summer 2011. *ATL-PHYS-PUB-2011-11*, *CMS NOTE-2011*, **005**.
- ATLAS Collaboration (2012). Observation of a new particle in the search for the standard model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, **716**.
- ATLAS Collaboration (2017). A model-independent search for new phenomena with the ATLAS detector $\sqrt{s} = 13$ TeV. *ATLAS CONF Note*, **001**.
- ATLAS Collaboration (2019). A strategy for a general search for new phenomena using data-derived signal regions and its application within the ATLAS experiment. *Eur. Phys. J. C* **79**, **120**.
- Beringer J. (2012). Review of particle physics. *Physical Review D*, **86**.
- Bohm G.; Zech G. (2017). *Introduction to Statistics and Data Analysis for Physicists*. Verlag Deutsches Elektronen-Synchrotron, Hamburg.
- CDF Collaboration (1995). Observation of top quark production in pp collisions with the collider detector at Fermilab. *Physical Review Letters*, **74**.
- Cha S. H.; Srihari S. N. (2002). Comparison of histograms in physical research. *Pattern Recognition*, **35**.

- CMS Collaboration (2012). Observation of a new boson at a mass of 125 gev with the cms experiment at the LHC. *Physics Letters B*, **716**.
- Cousins R. D. (2008). Evaluation of three methods for calculating statistical significance when incorporating a systematic uncertainty into a test of the background-only hypothesis for a poisson process. *Nuclear Instruments and Methods in Physics Research A* 595, **480**.
- Cowan et al. (2011). Asymptotic formulae for likelihood-based tests of new physics. *The European Physical Journal C*, **71**.
- DONUT Collaboration (2001). Observation of tau neutrino interactions. *Physics Letters B*, **504**.
- Elston R. C. (1991). On fisher's method of combining p-values. *Biometrical Journal*, **33**.
- Govindarajulu Z. (2007). *Nonparametric Inference*. World Scientific Publishing Company, Singapore.
- Li T.; Ma Y. (1983). Analysis methods for results in gamma-ray astronomy. *Astrophysical Journal*, **272**.
- Linnemann et al. (2007). Statistical inference problems in high energy physics and astronomy. *Report for Workshop 06w5054*.
- Lyons L. (2008). Open statistical issues in particle physics. *The Annals of Applied Statistics. Oxford University*, **2**.
- Pearson K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Phil. Mag. (5)*, **50**.
- Porter F. (2008). Testing consistency of two histograms. *California Institute of Technology*.
- R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Vianello G. (2018). The significance of an excess in a counting experiment: Assessing the impact of systematic uncertainties and the case with a gaussian background. *The Astrophysical Journal Supplement Series*, **236**.