

La ricerca si colloca nell'ambito dello sviluppo di metodologie statistiche per la valutazione della separazione e della sovrapposizione tra popolazioni, con particolare riferimento a contesti biomedici e diagnostici nei quali le popolazioni considerate corrispondano a differenti stati di malattia. In tali applicazioni, la disponibilità di misure affidabili dell'overlap tra distribuzioni riveste un ruolo rilevante, poiché consente di quantificare il grado di distinzione tra gruppi e, conseguentemente, la capacità di una variabile o di un test diagnostico di discriminare tra differenti condizioni cliniche. Il progetto si concentra sull'indice SIK, recentemente introdotto come misura di sovrapposizione tra più popolazioni in presenza di vincoli di ordinamento stocastico ad albero o ad ombrello, con l'obiettivo di estenderne l'inferenza a situazioni in cui il vero stato di appartenenza dei soggetti non sia osservato per l'intero campione.

In molti studi diagnostici, infatti, la classificazione certa dei soggetti richiede l'impiego di un test di riferimento o gold standard, che può essere costoso, invasivo o difficilmente applicabile a tutti gli individui inclusi nello studio. Di conseguenza, soltanto una parte dei soggetti viene sottoposta a verifica, mentre per i rimanenti il vero stato di malattia rimane non osservato. Se la probabilità di essere sottoposti a verifica dipende dalle caratteristiche del soggetto, dal risultato del test diagnostico o dallo stesso stato di malattia, le stime ottenute utilizzando esclusivamente i soggetti verificati possono risultare sistematicamente distorte. Tale fenomeno, noto come verification bias, è ampiamente studiato nell'ambito dell'analisi ROC, ma richiede ulteriori sviluppi metodologici quando l'obiettivo è l'inferenza su misure di overlap tra più popolazioni soggette a vincoli di ordinamento.

L'obiettivo principale della ricerca sarà pertanto lo sviluppo di procedure di stima corrette per il verification bias dell'indice SIK. Saranno considerati differenti meccanismi di mancata osservazione dello stato reale dei soggetti, con particolare attenzione ai casi Missing At Random (MAR), nei quali la probabilità di verifica può essere descritta in funzione delle variabili osservate, e Missing Not At Random (MNAR), nei quali essa può dipendere anche dallo stato di malattia non osservato. Nell'ambito MAR saranno studiate procedure basate su tecniche di imputazione, volte a ricostruire probabilisticamente l'appartenenza dei soggetti non verificati alle diverse popolazioni, e procedure di riponderazione, quali l'Inverse Probability Weighting (IPW), nelle quali il contributo di ciascun soggetto verificato viene pesato in funzione della probabilità stimata di essere sottoposto a verifica. Saranno inoltre considerate strategie che combinino imputazione e riponderazione, con l'obiettivo di ottenere stimatori più robusti rispetto alla specificazione dei modelli utilizzati.

Un ulteriore aspetto della ricerca riguarderà lo studio delle proprietà statistiche degli stimatori proposti. Saranno analizzati, ove possibile, consistenza, distribuzione asintotica, bias e varianza, nonché sviluppate procedure per la costruzione di intervalli di confidenza e per la quantificazione dell'incertezza inferenziale. Particolare attenzione sarà dedicata all'effetto del meccanismo di verifica sulla stima dell'overlap e alla capacità delle diverse procedure di recuperare correttamente il valore dell'indice SIK in presenza di differenti livelli di selettività del processo di verifica.

Le proprietà in campioni finiti delle metodologie proposte saranno infine valutate mediante un ampio studio di simulazione, considerando differenti configurazioni delle distribuzioni delle popolazioni, diversi gradi di overlap, differenti strutture di ordinamento stocastico e differenti meccanismi e percentuali di verifica. Qualora disponibili, le procedure saranno inoltre applicate a dati reali di ambito biomedico o diagnostico, al fine di valutarne l'utilità pratica e di fornire strumenti metodologici utilizzabili in contesti applicativi caratterizzati da informazione incompleta sul vero stato dei soggetti.