

Multiple tests for high-throughput data assuming a factor modelling of dependence

ABSTRACT

Dependence between variables is known to generate instability in multiple testing procedures. Therefore, an increasing interest, particularly in microarray data analysis, is to control actual error rates rather than expected ones. We propose to assume a factor-analysis model for the covariance matrix that allows identifying common and specific variability of data. Close form expressions for the variance of error rates are derived. A new test statistic is also defined, taking into account the factor structure. It is shown to reduce the variance of error rates, improving power in simultaneous tests.

(90)

Modèle en Facteurs pour les Procédures de Tests Multiples en Grande Dimension

RESUME

La corrélation entre variables entraîne des problèmes d'instabilité dans les procédures de tests multiples. Ainsi, en particulier dans l'analyse de données de biopuces, il est important de contrôler les taux d'erreurs réels plutôt que leurs espérances. Nous proposons de considérer un modèle d'analyse en facteur qui suppose la décomposition de la variance en une partie commune et une partie spécifique. Une formulation exacte de la variance des taux d'erreurs en découle. Une nouvelle statistique de test est définie également, tenant compte de la structure en facteurs proposée. Cela permet de réduire la variance des taux d'erreur, améliorant la puissance des tests simultanés.

(102)