

# Grandes déviations et comptage de mots

Mireille Régnier

## **Abstract**

La signification statistique de la surrepresentation de mots exceptionnels est un sujet important en bioinformatique. Nous utilisons la théorie des grandes déviations sur deux problèmes. En premier, nous considérons un motif  $H$  surreprésenté. Les grandes déviations permettent un calcul très efficace de la queue de la distribution, la  $p$ -valeur des biologistes. En second, nous considérons les changements dans la distribution des mots induite par la surrepresentation d'un mot donné. La suppression de ce bruit permet alors la detection de signaux plus faibles. Les conséquences algorithmiques associés et la complexité sont comparés aux résultats existants.