Outlook
Washington University School of Medicine

$$Q^n_{j,d,s}(E) = 2^{-n_{j,d,s}D(P_{j,d,s}\|Q_{j,d,s})}$$

Adapting Sanov's theorem resulted in a more efficient variation-finder.

Doing the math

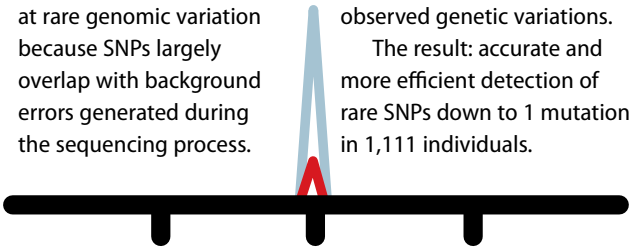
The question was how to tell an unusual gene from a technical glitch.

Researchers sought to determine variations that altered a single letter in the genetic code (A, C, G or T), called single nucleotide polymorphisms (SNPs). This problem becomes really challenging when looking at rare genomic variation because SNPs largely overlap with background errors generated during the sequencing process.

In order to distinguish a real SNP from sequencing noise, the research team developed a novel algorithm based on an application of Sanov's theorem, which was borrowed from the field of information theory.

An added mathematical model of known error rates aided comparisons with the observed genetic variations.

The result: accurate and more efficient detection of rare SNPs down to 1 mutation in 1,111 individuals.



SNPs (red) align with expected errors (blue) in this detail of a larger graph.

Washington University in St. Louis
School of Medicine