4 (Ms)CAVIAR: Likelihood-Driven Modeling of Cross-Study Causal Consistency

4.1 CAVIAR: an LD-driven model of causal configurations

Within the unified Bayesian framework of fine-mapping, CAVIAR (CAusal Variants Identification in Associated Regions) represents a class of models in which inference is driven primarily by the likelihood structure. In contrast to fastPAINTOR, which incorporates functional annotations through an explicit prior, CAVIAR relies solely on GWAS summary statistics and LD structure to infer the posterior distribution over causal configurations, without invoking external biological information (Lapierre et al., 2020).

Formally, CAVIAR models the joint distribution of GWAS Z-scores. For a genomic region containing m SNPs, let $Z=(Z_1,\ldots,Z_m)$ denote the vector of marginal association statistics. Conditional on a causal configuration γ, the model assumes:

$$Z\sim N(0,\Sigma+\sigma^2 I)$$

where $\Sigma$is the covariance matrix determined by LD structure, and σ^2captures residual variance. The key feature of this model lies in its explicit incorporation of LD structure into the likelihood function, thereby allowing correlated signals and causal signals to be distinguished within the same statistical framework.

Under this framework, CAVIAR evaluates (either exactly or approximately) the space of possible causal configurations and computes posterior inclusion probabilities (PIPs) for each SNP, from which credible sets are constructed under a specified coverage constraint (e.g., 95%). Importantly, the model allows for multiple causal variants ($K\geq1$) within a locus, thereby addressing the “signal ambiguity” that arises in LD-rich regions.

From a unified perspective, CAVIAR characterizes a causal probability distribution that is primarily determined jointly by LD structure and association statistics, namely an LD-driven causal estimand. Compared with models that rely on annotation information, it is not directly affected by the quality of external functional data. As a result, it often shows greater robustness when functional annotations are unavailable, when the annotations themselves may be biased, or when the study is conducted in a cross-species context.

4.2 MsCAVIAR: extending the likelihood to cross-study causal consistency

Building on the single-study formulation, MsCAVIAR (Multiple Study CAVIAR) extends the inferential target from identifying causal variants within a single study to assessing causal sharing across studies or populations.

This extension fundamentally alters the causal estimand: rather than estimating the posterior probability that a variant is causal within a single dataset, MsCAVIAR estimates the posterior probability that a variant is jointly causal across multiple studies.

To achieve this, MsCAVIAR introduces a random-effects model to capture heterogeneity in effect sizes across studies. Let β_{js}denote the effect of SNP $j$in study s, then:

$$\beta_{js}\sim N(\mu_j,\tau_j^2)$$

Where μ_j represents the shared (mean) effect across studies and τ_j^2captures cross-study heterogeneity. By jointly modeling summary statistics from multiple studies together with their respective LD structures, MsCAVIAR can estimate not only study-specific PIPs, that is, within-study causal probabilities, but also the shared causal probability across studies.

The importance of this modeling strategy lies in the fact that differences in LD structure across populations are no longer treated merely as background conditions of the analysis, but rather as a source of information that can be actively exploited. If the correlation structure of a given SNP differs across populations, then the association strength generated by non-causal variants will usually fluctuate with changes in LD patterns, whereas true causal variants are more likely to exhibit relatively consistent signals across studies. MsCAVIAR takes advantage of this contrast to eliminate spurious signals and further reduce the size of the credible set (Lapierre et al., 2020).