revisit my blog post "GWAS simulation setting - constant value vs random variable" at 11/Jul/2019

Huanwei Wang

2024-01-19

After reading Section 4.4 in the book Casella and Berger, Statistical Inference (2001), I would like to revisit this post.

The GWAS simulation setting, showed below, can be seen as a hierarchical model:

$$y = wu + e$$

, where

- $f \sim U(0.1, 0.5)$

- $x|f \sim Bin(2, f)$ (x is a binomial distribution condition on f)

- $w = \frac{x - 2f}{\sqrt{(2f(1-f))}}$, so

  - $E(w|f) = 0$
  - $Var(w|f) = 1$

- $u \sim N(0, 1)$, so

  - $E(wu|u, f) = uE(w|f) = u \times 0 = 0$
  - $Var(wu|u, f) = u^2 Var(w|f) = u^2$
  - $E(wu) = E(E(wu|u, f)) = E(0) = 0$
  - $Var(wu) = E(Var(wu|f, u)) + Var(E(wu|f, u)) = E(u^2) + Var(0) = 1 + 0 = 1$

- $e \sim N(0, var(wu)[\frac{1}{h2} - 1])$ (h2 is a parameter)

This provides a more detailed derivation about the mean and variance of wu based on the conditional probability and the hierarchical model.