

一对群体的模拟

假定一对群体都从一个祖先群体衍生而来。

当观测到这对群体 s_1, s_2 时，它们的 F_{ST} 是 f ，则群体可以模拟如下。

- ① 先模拟采样祖先群体基因频率 (假定服从均匀分布) $p = \text{runif}(p_1, p_2)$ 但可以在任意值
- ② 从祖先群体频率通过遗传漂变 (genetic drift) 产生两个群体。此步骤可以用 Beta (α, β) 产生。

$$\alpha = p \frac{1-f}{f}, \beta = (1-p) \frac{1-f}{f}$$

根据 Beta 分布，有

$$E(x) = \frac{\alpha}{\alpha + \beta} = p$$

$$V(x) = \frac{\alpha \beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)} = p(1-p)$$

Step 1. 产生 m 个位点 (祖先位点)。

$$p = \text{runif}(m, p_1, p_2). \quad (p_1, p_2) \text{ 是频率范围}$$

Step 2. 产生子群体频率。

$$p_{s1} = r\text{Beta}(m, p \frac{1-f}{f}, (1-p) \frac{1-f}{f})$$

$$p_{s2} = r\text{Beta}(m, p \frac{1-f}{f}, (1-p) \frac{1-f}{f})$$

Step 3 根据 p_{s1}, p_{s2} 产生 n_1, n_2 个样本，分别作为 s_1, s_2 的抽样。

$$G_{s1} = r\text{Binomial}(n_1, 2, p_{s1}) \quad ; \quad G_{s1} = r\text{Binomial}(n_1, 1, p_{s1}) \cdot 2$$

$$G_{s2} = r\text{Binomial}(n_2, 2, p_{s2}) \quad ; \quad G_{s2} = r\text{Binomial}(n_2, 1, p_{s2}) \cdot 2$$

杂交群体

自交群体。

Step 4 以上产生的是 genetic drift 自然分化位点，基于 selection 产生的位点需要更大的 F_{ST} (f_a) ($f_a > f$)。其余步骤如上 step 1~3。

根据以上步骤，可以生成其它类型群体。



统计功效

根据 Eigen GWAS 的线性模型, 可知回归系数可用卡方测验.

$$E = a + b_i \cdot \text{SNP}_i + e.$$

经过 λ_{GC} 校正.

第 i 个 SNP 标记对应的卡方测验为

$$E(\chi^2_{1\lambda_{GC}}) = \frac{f_{a_i}}{f} \quad (\text{R Heredity 2016, 117: 51~61, Eq 5}).$$

其中 f 是群体水平 F_{ST} (所有位点的平均 F_{ST} 决定).

f_{a_i} 是这个特定位点的 F_{ST} .

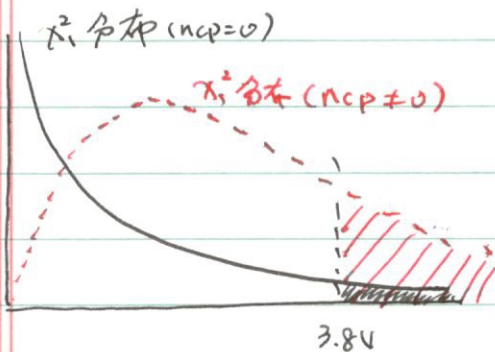
只有当 $f_{a_i} > f$ 时, 才有统计功效

卡方测验的统计功效与 n_{cp} (non-centrality parameter) 有关.

针对一个检验位点.

$$\text{对于自交群体 } n_{cp} = (n_1 + n_2) \cdot \frac{n_1}{n_1 + n_2} \cdot \frac{n_2}{n_1 + n_2} \cdot \frac{(\bar{p}_{s_1} - \bar{p}_{s_2})^2}{p(1-p)} = \frac{n_1 \cdot n_2}{n_1 + n_2} f_{a_i}$$

$$\text{对于异交群体 } n_{cp} = 4(n_1 + n_2) \cdot \frac{n_1}{n_1 + n_2} \cdot \frac{n_2}{n_1 + n_2} \cdot \frac{(\bar{p}_{s_1} - \bar{p}_{s_2})^2}{p(1-p)} = \frac{4n_1 \cdot n_2}{n_1 + n_2} f_{a_i}$$

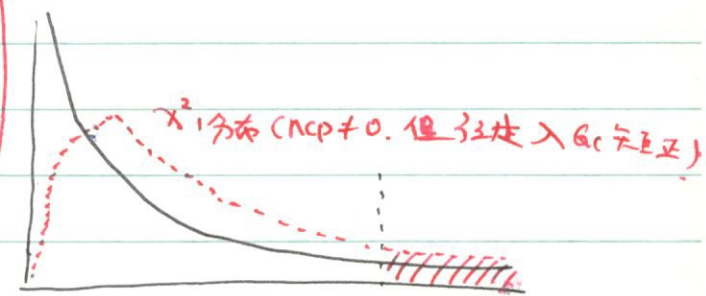


对于给定的阈值, 比如 3.84 ($p=0.05$)

$n_{cp}=0$ 情况下, $\text{power} = 0.05$ (黑色区).

$n_{cp} \neq 0$, $\text{power} > 0.05$ (红色区).

若对统计量进行 λ_{GC} 校正.



power 为红色阴影部分.

除非 $f_{a_i} = f$ (也就是这个位点不是受选择), 校正后阴影区小于左侧示例.