

0.1 A naive sampling model

Assume we have a set of n clonotypes having population abundance ρ in a sample containing N clonotypes in total. Upon re-sampling we detect m clonotypes in sample of size M . We can model sampling rate r and clonotype incidence as

$$r \sim \text{Beta}(m, n - m) \quad (1)$$

$$n \sim \text{Binom}(\rho, N) \quad (2)$$

$$m \sim \text{Binom}(\rho, M) \quad (3)$$

We can link clonotype size s and population frequency as

$$s \sim \text{Poiiss}(\rho N) \quad (4)$$

The probability to re-sample a clonotype from the set is

$$r = 1 - e^{-\rho M} \quad (5)$$

Averaging according to our uncertainty in ρ we have

$$\hat{r} = 1 - \int_0^1 e^{-\rho M} P(\rho) d\rho \quad (6)$$

If there is no uncertainty

$$\log(1 - \hat{r}) \approx -s \frac{M}{N} \quad (7)$$