

Given a (sub)population of size n , and a count $X = (X_A, X_B, X_{AB}, X_O)$ of the number of individuals with blood type A , B , AB or O ; under the mild assumption that each blood type is equally likely to be encountered, we may consider $X \sim \text{Multinom}(n, p_A, p_B, p_{AB}, p_O)$.

Furthermore, we may decompose the event of an individual being of type A into the disjoint union of the events that he/she has genotype AA or AO . Hence, $p_A = \pi_{AA} + \pi_{AO}$. Similarly, we may write $p_B = \pi_{BB} + \pi_{BO}$. Thus, we may write the likelihood function of X as

$$L_n(X) \propto (\pi_{AA} + \pi_{AO})^{X_A} (\pi_{BB} + \pi_{BO})^{X_B} p_{AB}^{X_{AB}} p_O^{X_O}$$

and assuming one can perform an MLE-style optimization for parameters $(\pi_{AA}, \pi_{AO}, \pi_{BB}, \pi_{BO}, p_{AB}, p_O) \in [0, 1]^6$ we may estimate the count of (say) the A alleles in the following way:

$$\begin{aligned} \# \text{ of } A\text{'s} &= 2 \cdot X_A \cdot \text{frequency}_A(AA) + X_A \cdot \text{frequency}_A(AO) + X_{AB} \\ &\approx 2X_A \frac{\hat{\pi}_{AA}}{\hat{\pi}_{AA} + \hat{\pi}_{AO}} + X_A \frac{\hat{\pi}_{AO}}{\hat{\pi}_{AA} + \hat{\pi}_{AO}} + X_{AB} \end{aligned}$$

since $\hat{\pi}_{AA} + \hat{\pi}_{AO} = \hat{p}_A$ and thus $\hat{\pi}_{AA}/(\hat{\pi}_{AA} + \hat{\pi}_{AO})$ should estimate π_{AA}/p_A which is the proportion of people with genotype AA inside populations of people with blood type A .