# Population Structure Exercises

# Balding Sampling Formula

For a case, suppose $n$ alleles have been seen among the known and typed contributors, and $n_A$ of these are of type $A$.

If allele $A$ in the evidence profile must be contributed by an unknown contributor under some hypothesis, the probability of that allele is

$$\Pr(A|n_A \text{ of } n) = \frac{n_A\theta + (1-\theta)p_A}{1+(n-1)\theta}$$

The $A$ allele is then added to the $n$ known alleles, and the probability of the next required allele, say $B$, is then calculated (if there were $n_B$ among the original $n$ alleles):

$$\Pr(B|n_B \text{ of } n+1) = \frac{n_B\theta + (1-\theta)p_B}{1+(n+1-1)\theta}$$

etc.

# Effect of $\theta$

Use the Balding sampling formula to find a formula for:

| | | $\theta \neq 0$ | $\theta = 0$ |
|---|---|---|---|
| $n = 0, n_A = 0$ | $\Pr(A)$ | $p_A$ | $p_A$ |
| $n = 1, n_A = 1$ | $\Pr(A\|A)$ | $\theta + (1-\theta)p_A$ | $p_A$ |
| | $\Pr(B\|A)$ | | |
| $n = 2, n_A = 2$ | $\Pr(A\|AA)$ | | |
| | $\Pr(B\|AA)$ | | |
| $n = 2, n_A = n_B = 1$ | $\Pr(A\|AB)$ | | |
| | $\Pr(B\|AB)$ | | |
| | $\Pr(C\|AB)$ | | |

# Effect of $\theta$

Use the Balding sampling formula to evaluate:

| | | $p = 0.10$ | | $p = 0.01$ | |
|---|---|---|---|---|---|
| | | $\theta = 0$ | $\theta = 0.01$ | $\theta = 0$ | $\theta = 0.01$ |
| $n = 0, n_A = 0$ | $\Pr(A)$ | 0.10 | 0.10 | 0.01 | 0.01 |
| $n = 1, n_A = 1$ | $\Pr(A\|A)$ | 0.10 | 0.109 | 0.01 | 0.0199 |
| | $\Pr(B\|A)$ | | | | |
| $n = 2, n_A = 2$ | $\Pr(A\|AA)$ | | | | |
| | $\Pr(B\|AA)$ | | | | |
| $n = 2, n_A = n_B = 1$ | $\Pr(A\|AB)$ | | | | |
| | $\Pr(B\|AB)$ | | | | |
| | $\Pr(C\|AB)$ | | | | |