**Supplementary Material S2 – Derivation of imputation accuracy when imputation is based on allele frequencies**

An alternative derivation to demonstrate that the imputation accuracy is zero when imputation is solely based on the allele frequencies, considering genotypes instead of alleles, is as follows. Consider the 3×3 possibilities of true ($g$) and predicted genotypes ($\hat{g}$) and using 2, 1 and 0 for the genotypes:

|  |  |
| --- | --- |
|  | Imputed genotype |
| Observed genotype | 2 | 1 | 0 |
| 2 | $$p\_{i}^{4}$$ | $$2p\_{i}^{3}(1-p\_{i})$$ | $$p\_{i}^{2}(1-p\_{i})^{2}$$ |
| 1 | $$2p\_{i}^{3}(1-p\_{i})$$ | $$4p\_{i}^{2}(1-p\_{i})^{2}$$ | $$2p\_{i}(1-p\_{i})^{3}$$ |
| 0 | $$p\_{i}^{2}(1-p\_{i})^{2}$$ | $$2p\_{i}(1-p\_{i})^{3}$$ | $$(1-p\_{i})^{4}$$ |

The covariance between the observed and predicted genotype is then

$$cov\left(\hat{g},g\right)=E\left(xy\right)-E\left(x\right)E\left(y\right)=\left(2×2\right)p^{4}+\left(2×1\right)2p^{3}\left(1-p\right)+\left(2×0\right)p^{2}\left(1-p\right)^{2}+\left(1×2\right)2p^{3}\left(1-p\right)+\left(1×1\right)4p^{2}\left(1-p\right)^{2}+\left(1×0\right)2p\left(1-p\right)^{3}+\left(0×2\right)p^{2}\left(1-p\right)^{2}+\left(0×1\right)2p\left(1-p\right)^{3}+\left(0×0\right)\*\left(1-p\right)^{4}-4p^{2}=4p^{2}\left(p^{2}+2p\left(1-p\right)+\left(1-p\right)^{2}\right)-4p^{2}=0$$