Figure 1: Effects of including full siblings in a sample on the distribution and variance of allele frequency estimates. The population is at HWE for a locus with two codominant alleles of the same frequency p=0.5. A sample contains n unrelated individuals from the population, and m_i full siblings from family i (i=1~5) in the same population. The sample structure, $\{n, m_1, m_2, m_3, m_4, m_5\}$, is {10, 10, 0, 0, 0, 0}, {10, 8, 2, 0, 0, 0}, {10, 6, 2, 2, 0, 0}, {10, 4, 2, 2, 2, 0}, and {10, 2, 2, 2, 2, 2}, denoted by E1~E5 respectively. Removing all but one sibling from each family yields samples {10, 1, 0, 0, 0, 0}, {10, 1, 1, 0, 0, 0}, {10, 1, 1, 1, 0, 0}, {10, 1, 1, 1, 1, 0}, and {10, 1, 1, 1, 1, 1}, denoted by e1~e5 respectively. The distributions and the sampling variance of $\hat{p}$ obtained from the 10 samples are shown in the upper and lower panels. The black and grey lines show the results for sample E1~E5 and e1~e5, respectively, calculated analytically. Simulation results of sampling variance are shown for the naïve estimator (i.e. E1~E5, stars), bold estimator (i.e. e1~e5, triangles), weighted estimator (applied to E1~E5, filled boxes), and likelihood estimator (applied to E1~E5, unfilled boxes). Note the $\hat{p}$ distributions in the upper panel are discrete (say, $\hat{p} = 0, 1/22, 2/22,…$), but for clarity are shown in solid lines as if the distributions were continuous. This, together with unequal sample sizes, causes an illusion that distributions e1~e5 become more dispersed in that order, but the opposite is true as shown in the lower panel.
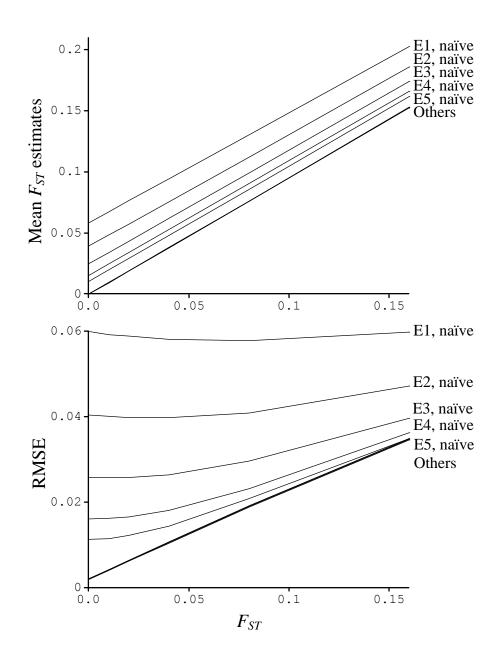
Figure 2: Effects of including full siblings in a sample on F_{ST} estimates. A population in Wright's island model with 20 subpopulations, and a locus with 4 codominant alleles with frequencies in uniform distribution, are assumed. A sample from the population consists of n unrelated individuals, and m_i full siblings from family i (i=1~5). The sample structure, {n, m_1, m_2, m_3, m_4, m_5}, is {50, 50, 0, 0, 0, 0}, {50, 40, 10, 0, 0, 0}, {50, 30, 10, 10, 0, 0}, {50, 20, 10, 10, 10, 0}, and {50, 10, 10, 10, 10, 10}, denoted by E1~E5 respectively. For each simulated F_{ST} value (x axis), estimated F_{ST} was calculated using allele frequencies estimated by the naïve, bold and weighted estimators. The means (upper panel) and RMSEs (lower panel) of the estimators are calculated. Note, the bold and weighted estimators have indistinguishable means and RMSEs for samples E1~E5, and are indicated by *Others* in the graph.
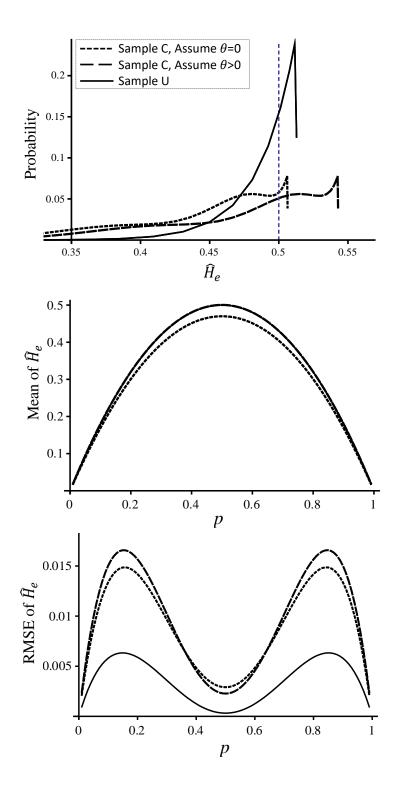
Figure 3: Effect of relatedness on estimated H_e. A combined sample C contains n=20 unrelated individuals and m=20 individuals drawn at random from a single full-sib family. Removing all but one of the m siblings yields the unrelated sample U, which has n=20 and m=1. The population is assumed at HWE for a locus with two codominant alleles of frequency p and 1-p. The upper panel shows the distributions of H_e estimates when p=0.5, the middle and lower panels show the means and RMSEs, respectively, of H_e estimates as a function of p. The broken and dotted lines show the results obtained from samples C by assuming related ($\theta > 0$) and unrelated ($\theta = 0$) individuals respectively, and solid lines show the results obtained from sample U. In the upper panel, the vertical broken line shows the population value of $H_e = 0.5$ being estimated. Note the broken and solid lines overlap in the middle panel.
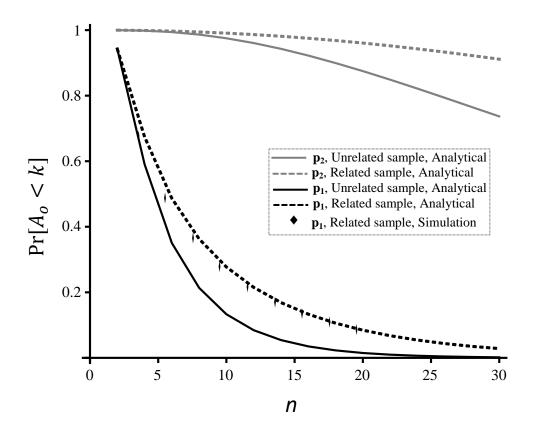
Figure 4: The probability $\Pr[A_o < k]$ of less than $k=4$ alleles being observed in a sample of n individuals. The population allele frequency distribution is $\mathbf{p_1}$=(0.1, 0.2, 0.3, 0.4) or $\mathbf{p_2}$=(0.01, 0.02, 0.03, 0.94). An unrelated sample contains n individuals drawn at random from a large population at HWE, and a related sample contains $n/2$ individuals drawn at random from the population and $n/2$ individuals drawn at random from a single full-sib family in the same population. The simulated values are shown in diamonds for the case of $\mathbf{p_1}$ and related samples.
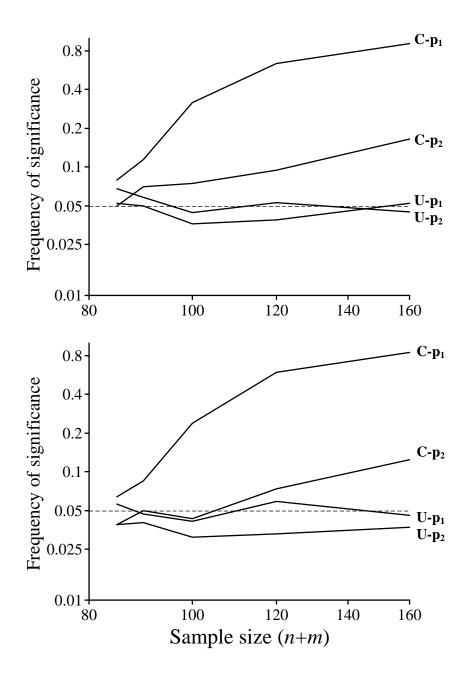
Figure 5: Frequency of significant (at 5% level) deviation from HWE (upper panel) and frequency of significant (at 5% level) LD (lower panel) as a function of the number of full siblings included in a sample. A unrelated sample, U, of a fixed number of n=80 diploid individuals were drawn at random from a large population at HWE and LE. A number of m=5, 10, 20, 40, or 80 full siblings were taken from a single family and were added to sample U to constitute the combined sample, C. Both U and C samples were tested for deviation from HWE (upper panel) or for LD (lower panel) using the exact test with the permutation approach. The allele frequency distribution is assumed to be $\mathbf{p_1}$ or $\mathbf{p_2}$, where $\mathbf{p_1}$=(0.1, 0.2, 0.3, 0.4) and $\mathbf{p_2}$=(0.01, 0.02, 0.03, 0.94). For LD tests (lower panel), the two loci have an identical allele frequency distribution. The broken line shows the 5% expected frequency. Note, the x axis shows the sizes of the C samples only, and the U samples have a fixed size of 80.