

# Supplementary Material for “Quantifying the Uncertainty in Heritability”

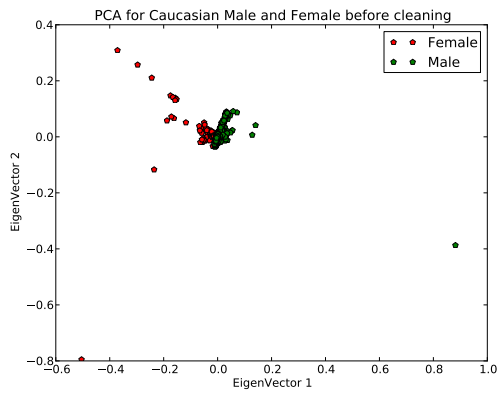
Nicholas A. Furlotte<sup>1,2,\*</sup>, Christoph Lippert<sup>1,\*</sup> and David Heckerman<sup>1,\*</sup>

<sup>1</sup> Microsoft Research, Los Angeles, California

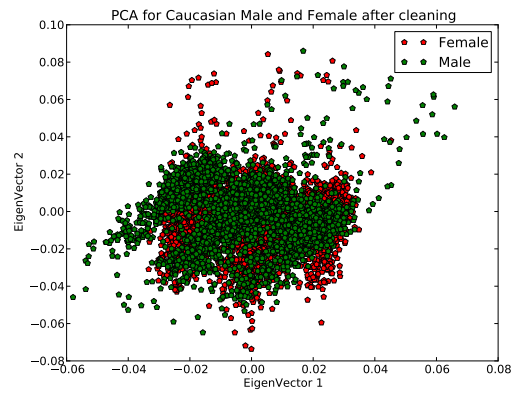
<sup>2</sup> University of California Los Angeles, Los Angeles, California

\* These authors contributed equally to this work

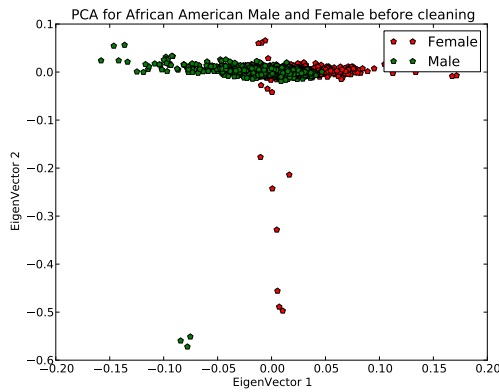
# 1 PCA plots for ARIC cohorts before and after filtering



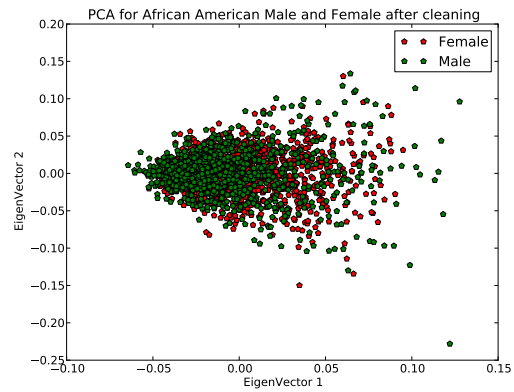
(a) PCA for Caucasians Before Filtering



(b) PCA for Caucasians After Filtering



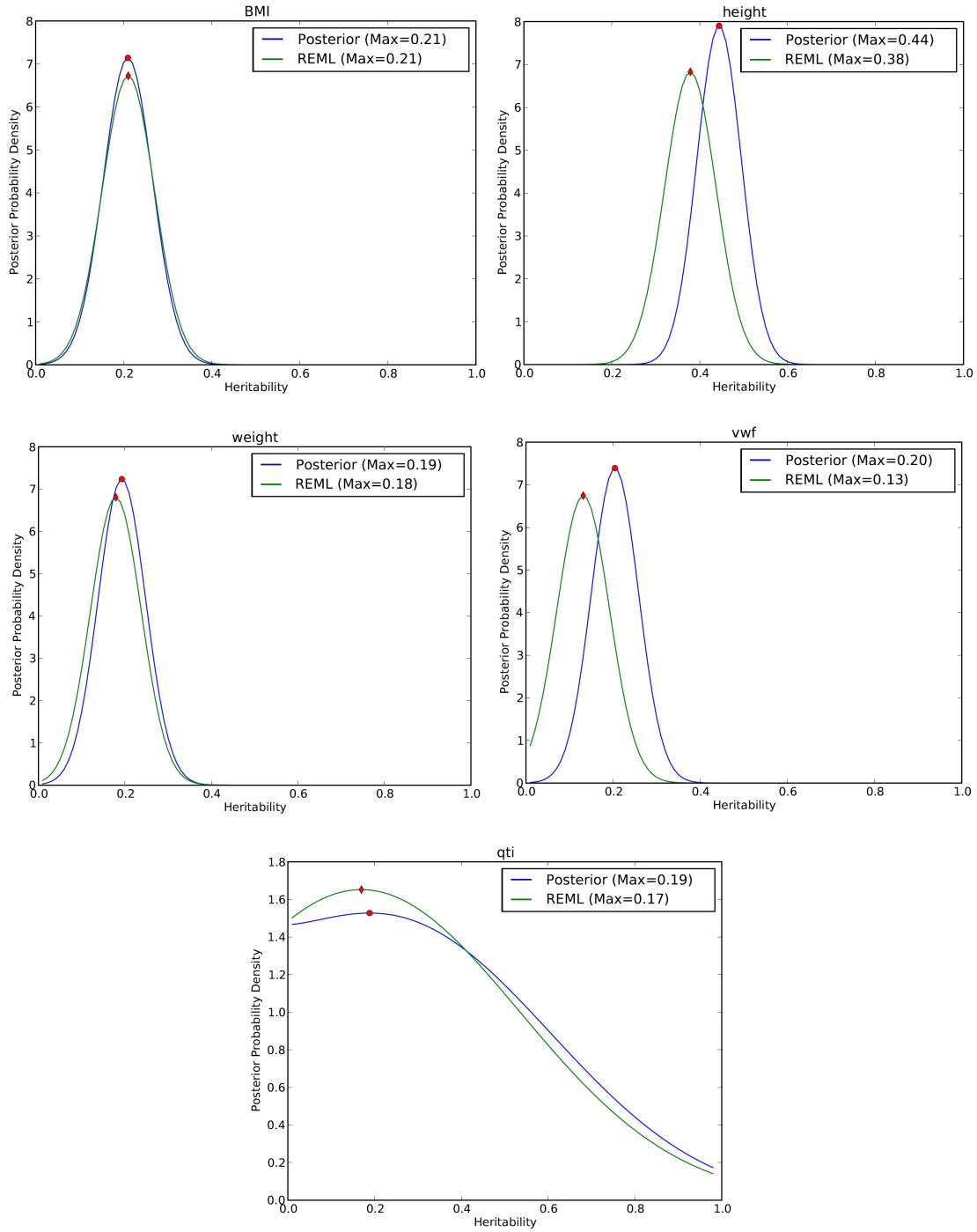
(c) PCA for African Americans Before Filtering



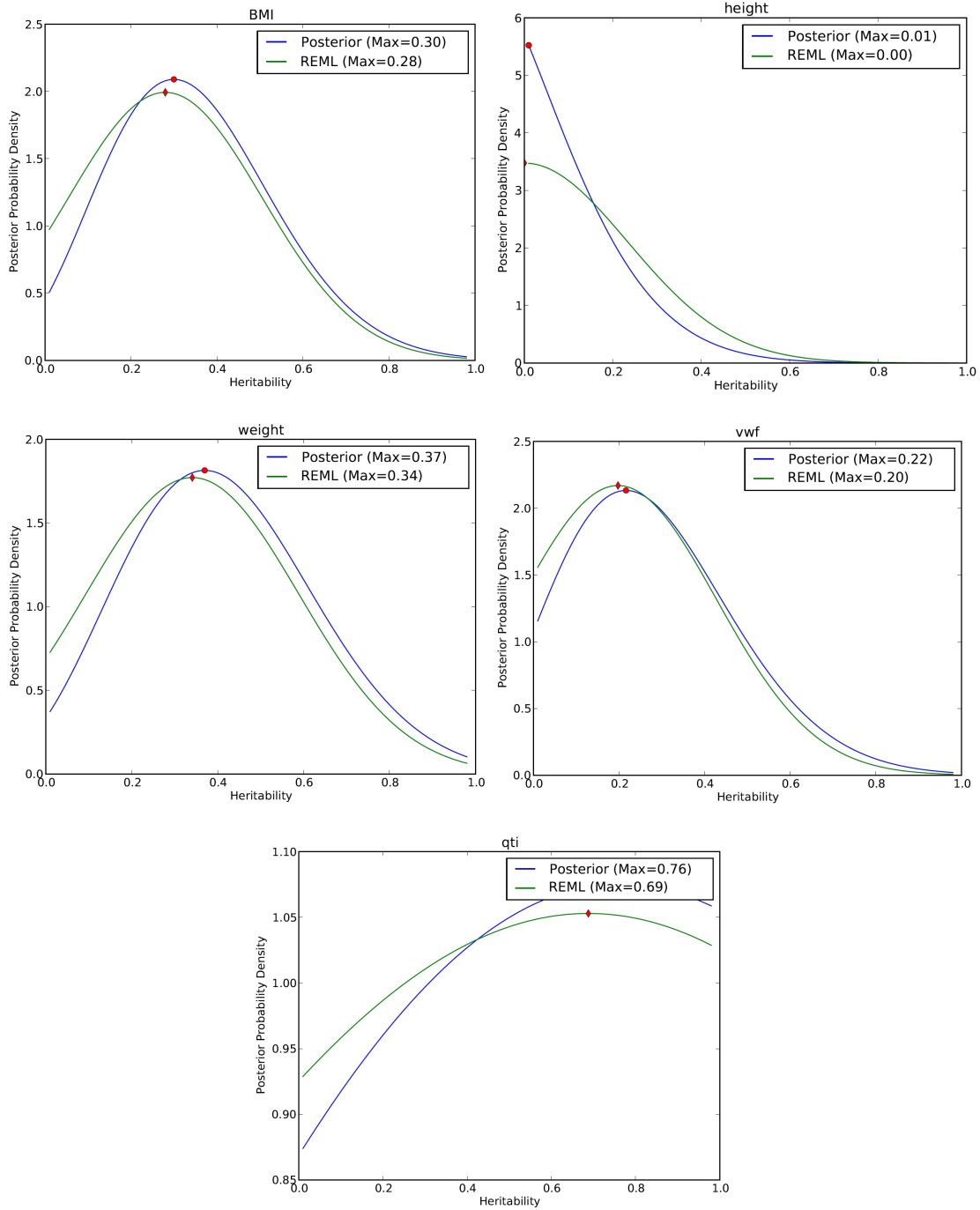
(d) PCA for African Americans After Filtering

Figure 1: **PCA plots for Caucasians and African Americans.** Prior to filtering, some individuals deviate significantly from the remainder of the group. After filtering individuals based on genetic relatedness (see Methods), individuals are relatively homogeneous with respect to ancestry.

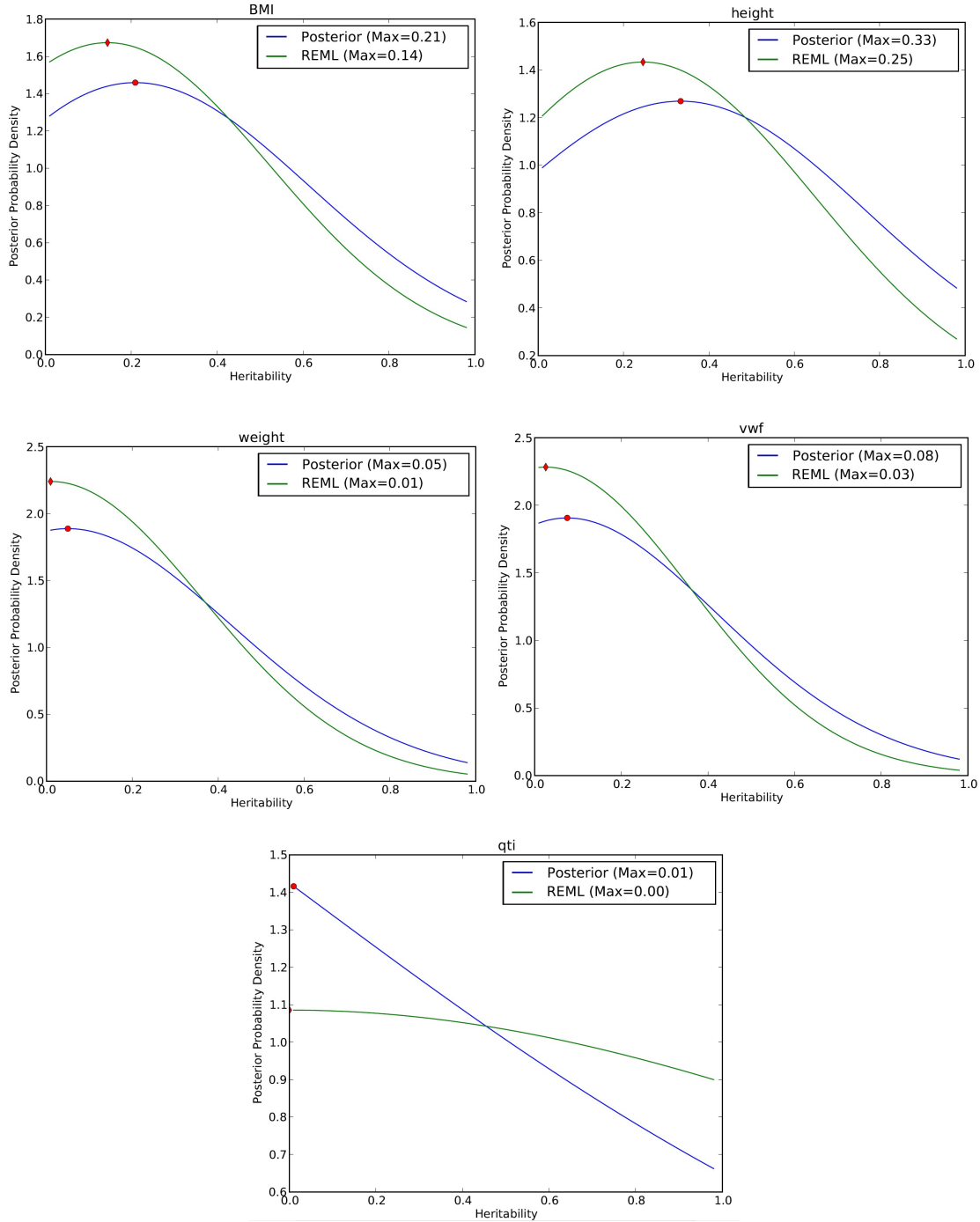
## 2 Comparison of Bayesian posterior with frequentist distribution for the full ARIC cohort



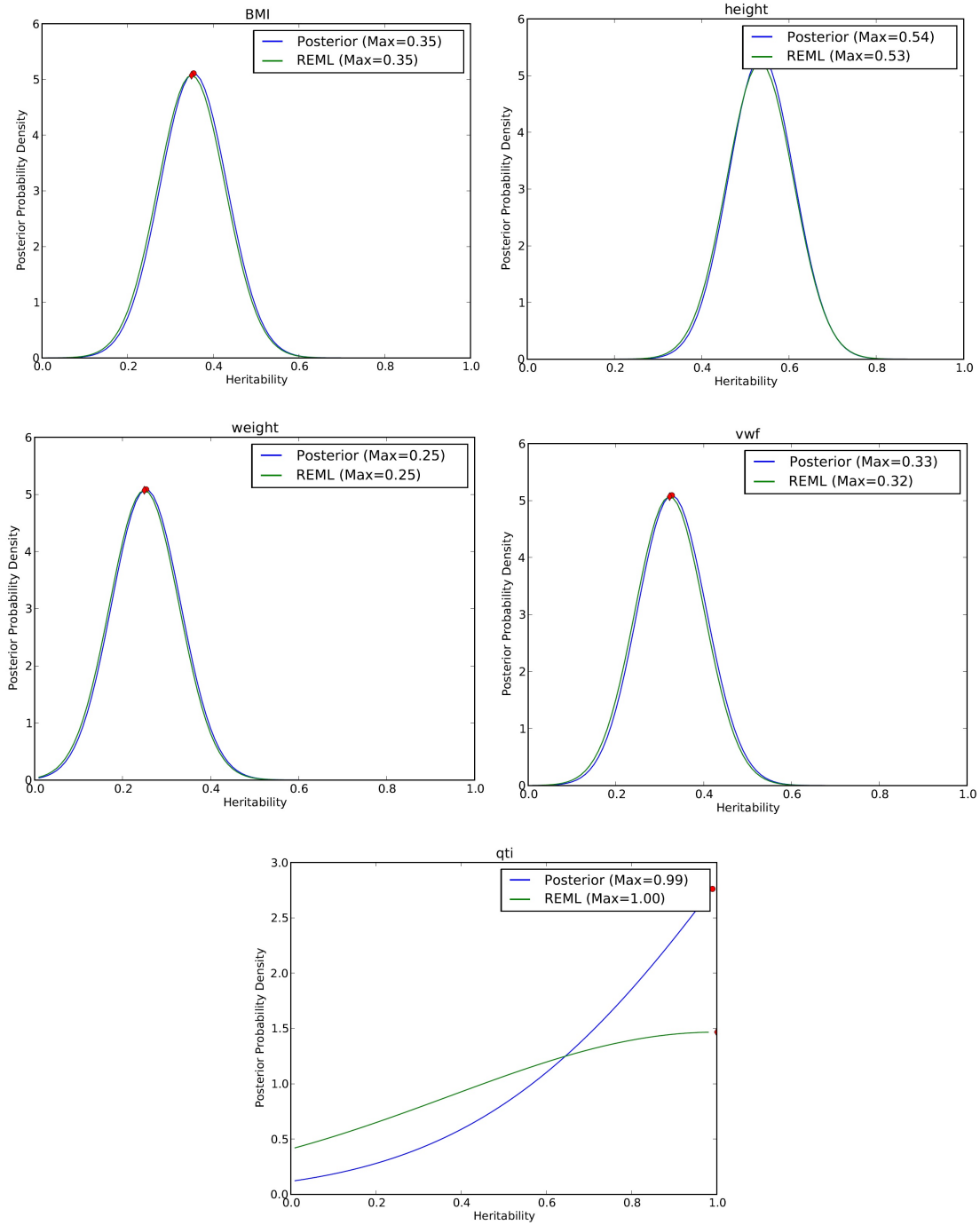
### 3 Comparison of Bayesian posterior with frequentist distribution for African American females



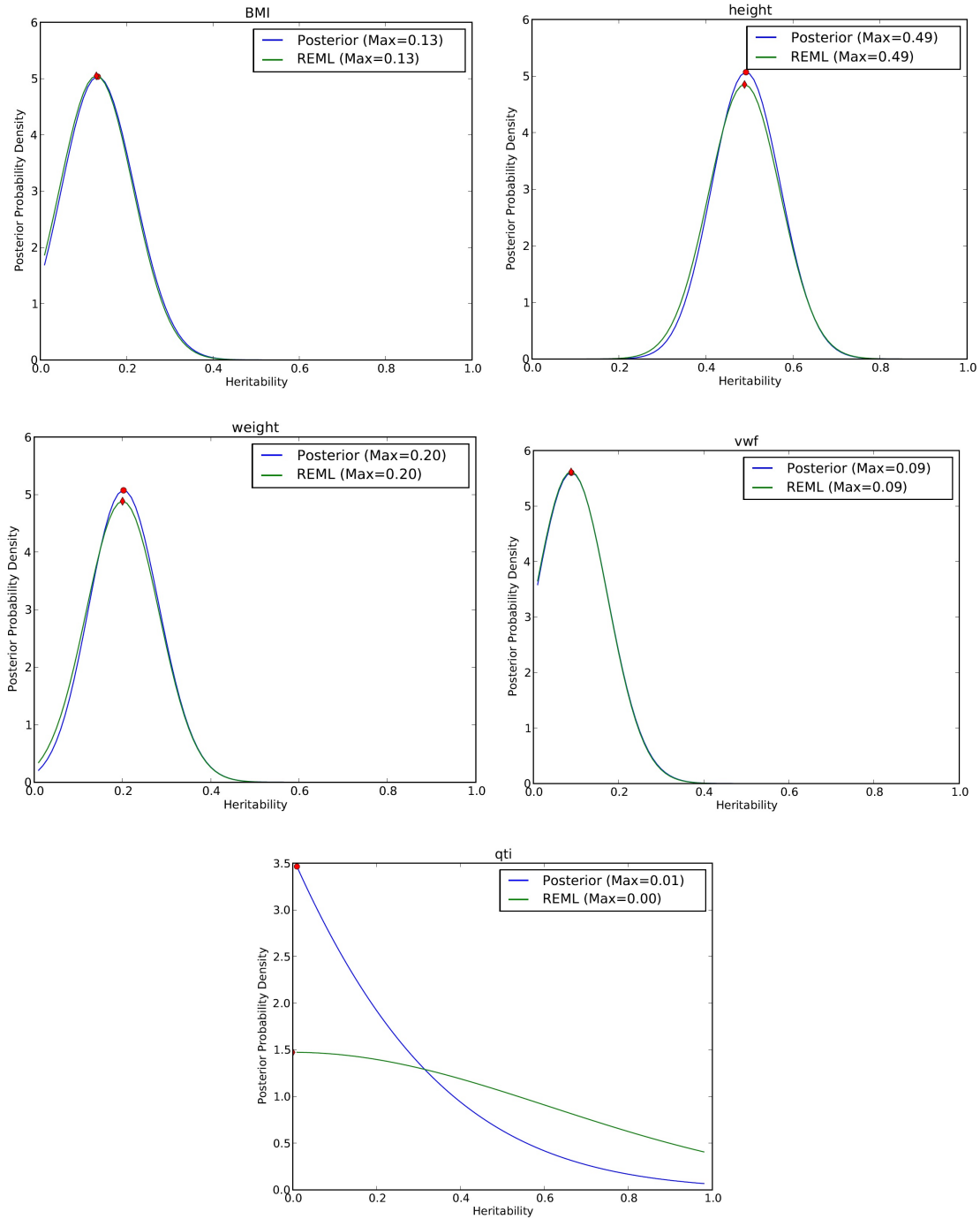
# 4 Comparison of Bayesian posterior with frequentist distribution for African American males



## 5 Comparison of Bayesian posterior with frequentist distribution for Caucasian females



## 6 Comparison of Bayesian posterior with frequentist distribution for Caucasian males



## 7 A theoretical comparison of the posterior distribution of heritability and the sampling distribution of the maximum-likelihood estimator for heritability

Consider the log-likelihood of a set of data  $\mathbf{y}$   $L(\mathbf{y} | \Theta)$ , where  $\Theta$  is a set of parameters. A maximum likelihood estimator for  $\Theta$ , denoted by  $\hat{\Theta}$ , is any parameter configuration that satisfies the following equation, where  $\dot{L}$  represents the gradient of the likelihood function with respect to  $\Theta$

$$\dot{L}(\mathbf{y} | \hat{\Theta}) = 0. \quad (1)$$

As the size of the data,  $N$ , increases to infinity, the sampling distribution of the maximum likelihood estimator is known to follow a normal distribution with mean centered at the true  $\Theta$  and with variance lower-bounded by the inverse of the expected information matrix<sup>1</sup>. The Fisher's information matrix is defined as the expected value of the negative Hessian of the log-likelihood (*i.e.*, of the observed information matrix). We denote the Hessian of the log-likelihood as  $\ddot{L}$  and define the expected information matrix  $F(\Theta)$  as follows<sup>1</sup>:

$$F(\Theta) = -E[\ddot{L}(\mathbf{y} | \Theta)]. \quad (2)$$

With this we specify the distribution of the maximum-likelihood estimator as

$$N(\hat{\Theta} | \Theta, F^{-1}(\Theta)). \quad (3)$$

Therefore, the sampling variance of the estimate of  $\Theta$  can be quantified using the distributional properties of this normal distribution.

We now show that the posterior distribution of a parameter set  $\Theta$  given the data  $\mathbf{D}$  asymptotically follows a normal distribution similar to the one shown in Equation 3. A thorough treatment of this concept may be found in ref<sup>2</sup> among others, although we follow closely the



proof given by ref<sup>3</sup>. Let the posterior distribution of  $\Theta$  be denoted by the following<sup>3</sup>:

$$p(\Theta | \mathbf{y}) \propto p(\Theta)p(\mathbf{y} | \Theta), \quad (4)$$

$$= \exp \{ \log(p(\Theta)) + \log(p(\mathbf{D} | \Theta)) \}. \quad (5)$$

Let  $\theta_0$  and  $\theta_1$  denote parameter settings that maximize  $\log(p(\Theta))$  and  $\log(p(\mathbf{y} | \Theta))$ ,

respectively. We can write the second order Taylor expansion around these maxima as follows<sup>2</sup>:

$$\log(p(\Theta)) = \log(p(\theta_0)) - \frac{1}{2}(\Theta - \theta_0)^T \mathbf{H}_0(\Theta - \theta_0) + R_0, \quad (6)$$

$$\log(p(\mathbf{y} | \Theta)) = \log(p(\mathbf{y} | \theta_1)) - \frac{1}{2}(\Theta - \theta_1)^T \mathbf{H}_1(\Theta - \theta_1) + R_1. \quad (7)$$

Here we have let  $R_0$  and  $R_1$  represent the remainders and we have let  $\mathbf{H}_0$  and  $\mathbf{H}_1$  represent the negative of the Hessian matrix for  $p(\Theta)$  and  $p(\mathbf{y} | \Theta)$ , evaluated at their respective maxima.

Assuming that, with large sample size,  $R_0$  and  $R_1$  will be small, we ignore them. Additionally, we may disregard the terms  $\log(p(\theta_0))$  and  $\log(p(\mathbf{y} | \theta_1))$  as they are constants. We then have the following result:

$$p(\Theta | \mathbf{y}) \propto \exp \left\{ -\frac{1}{2}(\Theta - \theta_0)^T \mathbf{H}_0(\Theta - \theta_0) - \frac{1}{2}(\Theta - \theta_1)^T \mathbf{H}_1(\Theta - \theta_1) \right\}. \quad (8)$$

Combining terms, we can show the following:

$$p(\Theta | \mathbf{y}) \propto \exp \left\{ -\frac{1}{2}(\Theta - \hat{\theta})^T \hat{\mathbf{H}}(\Theta - \hat{\theta}) \right\}, \quad (9)$$

$$\hat{\theta} = \mathbf{H}_1^{-1}(\mathbf{H}_0\theta_0 + \mathbf{H}_1\theta_1), \quad (10)$$

$$\hat{\mathbf{H}} = \mathbf{H}_0 + \mathbf{H}_1. \quad (11)$$

Therefore, up to a constant of proportionality, the posterior distribution of  $\Theta$  given the data follows a normal distribution  $p(\Theta | \mathbf{y}) = N(\Theta | \hat{\theta}, \hat{\mathbf{H}}^{-1})$ , where the mean and variance both have a prior-dependent term and a data-dependent term. The variance term is a combination of the prior precision (inverse variance) and the observed precision. The latter is the observed information matrix. When using a uniform prior, the prior precision will be zero and thus the posterior mean will be the maximum-likelihood solution; and the posterior variance will be determined by the observed information matrix,  $N(\Theta | \theta_1, \mathbf{H}_1^{-1})$ . By definition, the observed information is an unbiased estimate of the expected information matrix defined in Equation 2. Note that the expected information matrix is defined over the joint multivariate log-likelihood of a dataset of size  $N$ .

## References

1. Searle, S. R., Casella, G. & McCulloch, C. *Variance Components, Volume 631*. (Wiley-Interscience, 2006).
2. Bernardo, J. & Smith, A. *Bayesian Analysis*. (John Wiley, 1994).
3. Houlston, R. S. *et al.* Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nature genetics* **40**, 1426–35 (2008).