(1) *Poisson regression and iteratively reweighted least squares (IRLS) via Newton Raphson.*

(a) Derive the iteratively reweighted least squares algorithm for Poisson regression.

(b) Implement IRLS for Poisson regression.

(c) Generate 500 Gaussian data points (see: rnorm) $x$ and, for three different $\theta$ values (one negative, one between 0 and 1, and one greater than one) generate 500 poisson-distributed points $y_i, i \in \{1, \ldots, 500\}$ with mean $\mu_i = \exp \theta x_i$ (see: rpois). Plot these three data sets (try $x$ versus $y$, $x$ versus $\log y$).

(d) Estimate, using your IRLS code, the $\theta$ parameters for $x$ each of the three $\theta$ and $y$. How many iterations does it take to converge to something reasonable? How far away is this estimate from the truth? Show (three on the same plot – see: lines) the value of each $\theta$ estimate (y-axis) at each iteration (x-axis).

(2) *Bayesian regression.* In class, we discussed the expected value of the effect size $\beta$ in a linear model in Bayesian linear regression, given a Gaussian prior on $\beta$.

We might let this expected value of the effect size be our estimate of $\beta$ in the Bayesian setting, call this $\tilde{\beta}$. Call the (ML) estimate of $\beta$ in the frequentist setting $\hat{\beta}$.

(a) Write code to compute $\tilde{\beta}$ for a given set $D = \{(x_1, y_1), \ldots, (x_n, y_n)\}$, where $x_i$ are scalar, and $\mu_0$ (the mean of the distribution on $\beta$) and $\lambda_0$ (the variance of $\beta$) (although there is a scalar predictor $x$, remember to include an intercept term $\beta_0$).

(b) Download genotypes and gene expression data from our earlier homework. Compute $\tilde{\beta}$ and $\hat{\beta}$ for each pair of SNPs and gene expression values for $\mu_0 = 0$, $\lambda_0 = 1$. Plot the values (one on the x-axis, one on the y-axis) and draw a line representing $x = y$. What do you notice about the Bayesian estimate as compared to the frequentist estimates?

(c) Create the same plot for i) a very large value of $\lambda_0$ and ii) a very small value of $\lambda_0$ (non-negative). What effect does this value have on the $\tilde{\beta}$?

(d) Try varying $\mu_0$. What impact do different values of $\mu_0$ have on $\tilde{\beta}$?