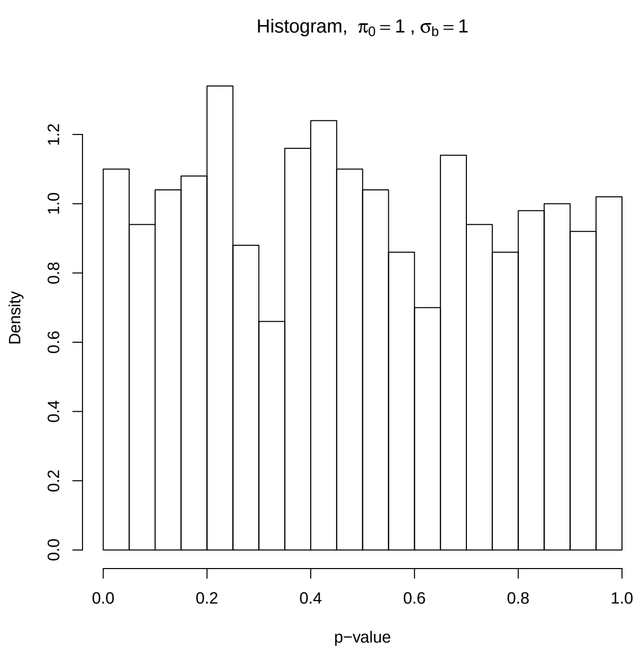
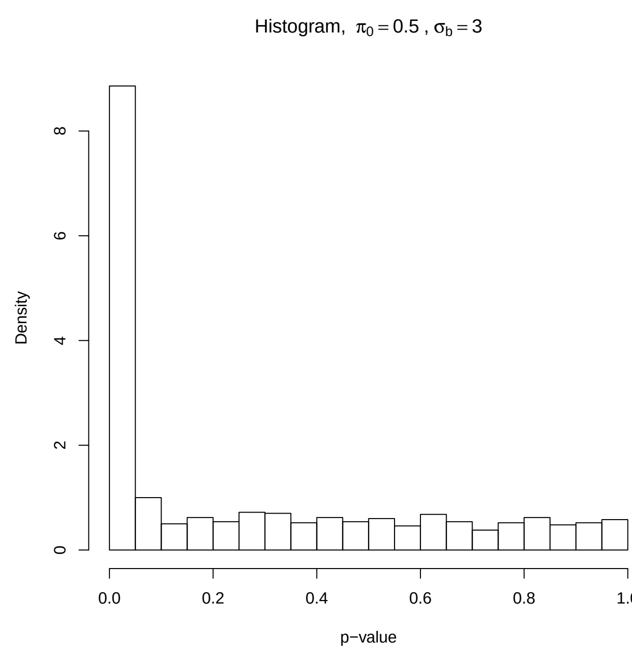
1. i) Together, (1) and (2) define a hierarchical model since they relate observed values $D_{ij}$ drawn independently based on parameters $\beta_i$ to a set of hyperparameters $\pi_0$ and $\sigma_b$ in such a way that the $D_{ij}$ can provide inference on each other through the hyperparameters. In other words, even though the observations are conditionally independent, they are still informative on each other through the hyperparameters.Since the $\beta_j$ are iid given the hyperparameters, they are necessarily exchangeable. Not conditioned on $\pi_0, \sigma_b$, the $\beta_j$ should still be exchangeable assuming the same value of $\pi_0$ and $\sigma_b$ are used to draw each $\beta_j$; there is just additional variability in the initial draw of the hyperparameters. $D_{ij} \mid \beta, \sigma$ will be exchangeable for all $i$ since they are iid, but $D_{ij} \mid \beta_j$ won't be exchangeable across $j$ since $P(D_{ij} \mid \beta_j = a, D_{ij'} \mid \beta_{j'} = b) = P(D_{ij'} \mid \beta_{j'} = a, D_{ij} \mid \beta_j = b)$ only if $\beta_{j'} = \beta_j$. Finally, the $D_{ij}$ should be exchangable accross $i$ and $j$ if unconditioned, or conditioned on $\pi, \sigma_b$, since then they will be iid.

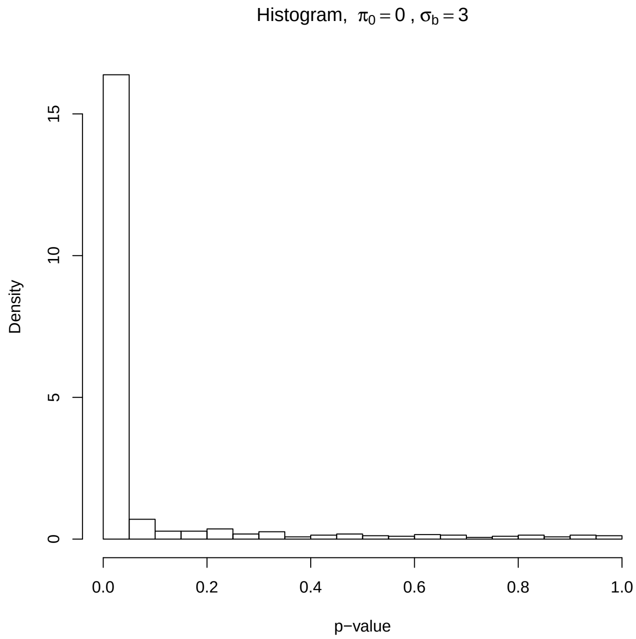
   ii) This command will calculate $\beta$, $D$, and also the p-values.

```
### Problem 1
## ii - simulate simple hierarchical model
p.values <- function(x){
    return(t.test(x)$p.value)
}

simulateD <- function(pi0,sigb,n=10,m=1000,sigj=1){
    output <- list()
    output$beta     <- ifelse(rbinom(m,1,pi0),0,rnorm(m,0,sigb))
    output$D        <- replicate(n,rnorm(m,output$beta,sigj))
    output$p.values <- apply(output$D,1,p.values)
```

   iii) I calculated the usual two sided t-test p-value using the R command. They were distributed as such:






As one would expect, for the first one where the $\beta$ are all 0, the p-values are uniformly distributed. In the second, with $\pi_0 = .5$, we see a big lump of low p-values near 0 corresponding to most of the $\beta_j$ which are non-zero, and a approximately uniform distribution over the rest of the p-values corresponding to the $\beta_j$ which are 0. Finally, in the last histogram, where none of the $\beta_j$ are 0, we see overwhelmingly low p-values corresponding to the strong evidence against $\beta_j = 0$.

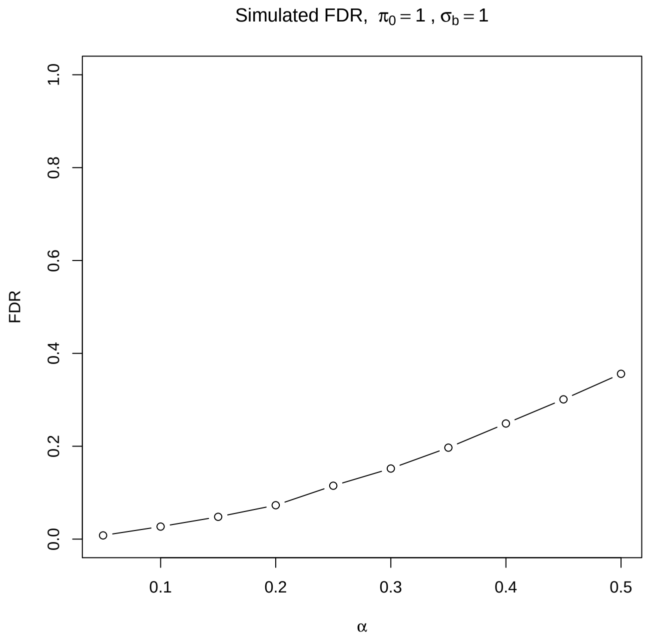
iv) This is my function to apply Benjamini-Hochberg:

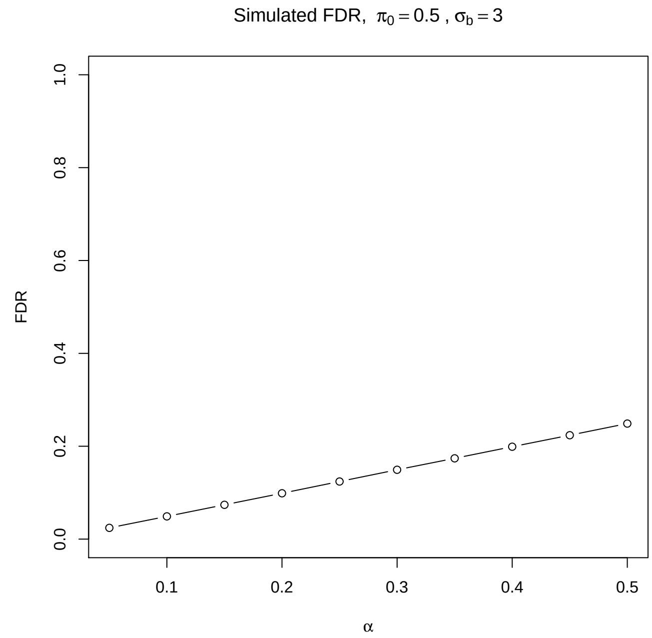
```
    }
}

## iv - implement Benjamini-Hochberg
BH <- function(pvals,alpha){
    m    <- length(pvals)
    ord  <- order(pvals)
```

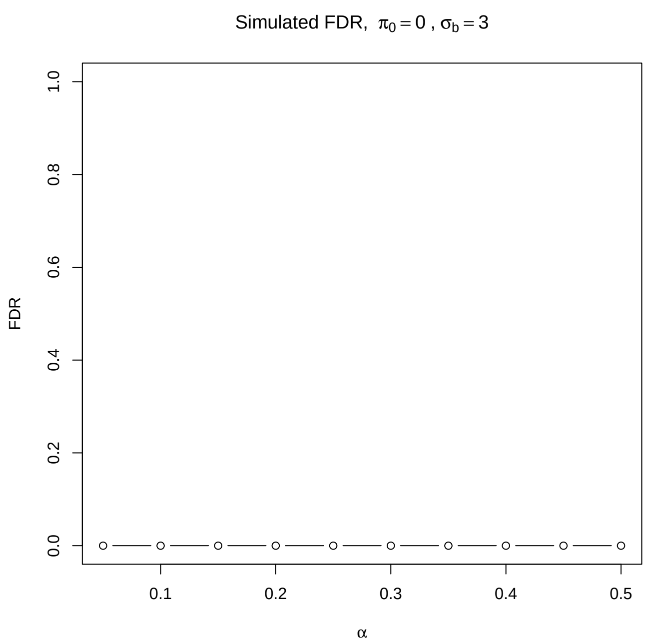
v) This is my function to computer the empirical FDR (and pFDR):

```
    return(test)
}

## v - implement FDR
FDR <- function(beta,gamma,pFDR=FALSE){
    R <- sum(gamma)
```

vi) To simulate $E(V/R)$, I generated the dataset D 1000 times for each scenario, and averaged the emperical FDR for each *alpha* over the data sets. The results are ploted below:

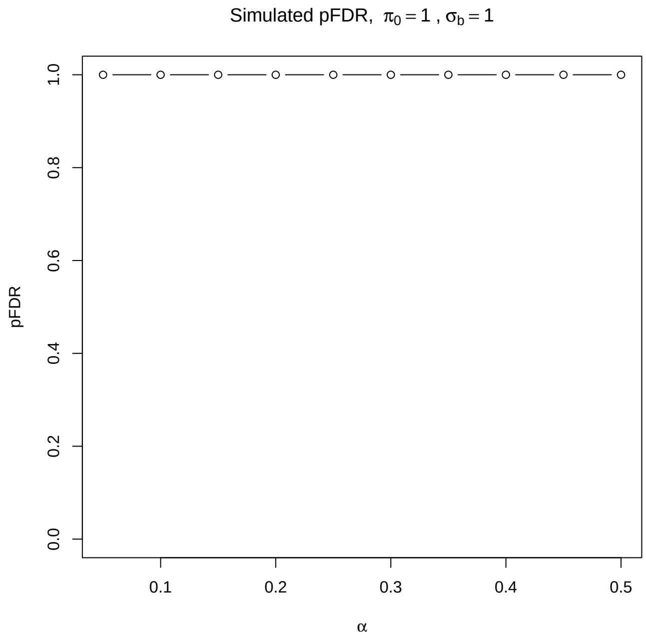


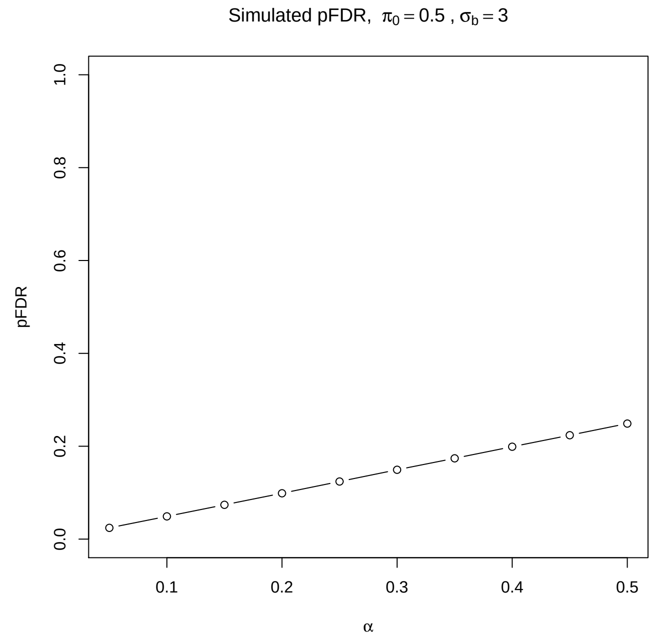



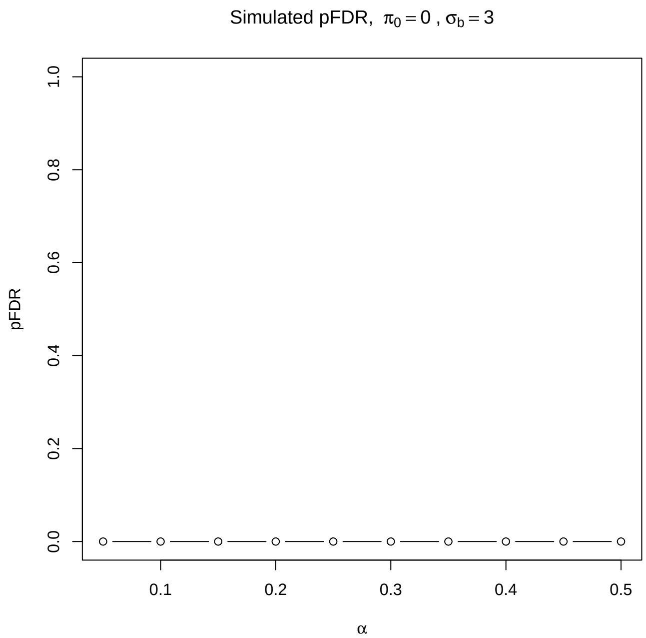




In the first scenario, despite the fact that all the discoveries are false, our FDR statistic is less than 1, which is due to the definiton. In any event, The FDR rate is properly controlled, coming in under $\alpha$. For the second scenario, the FDR rate seems to come in at almost exactly half of $\alpha$ for all levels, providing conservative control. For the final scenario, since all $\beta \neq 0$, the FDR is necessarily 0, so FDR is inherently controlled.

vii) I repeated the same procedure as above, but this time caclulated the pFDR. The results are below:

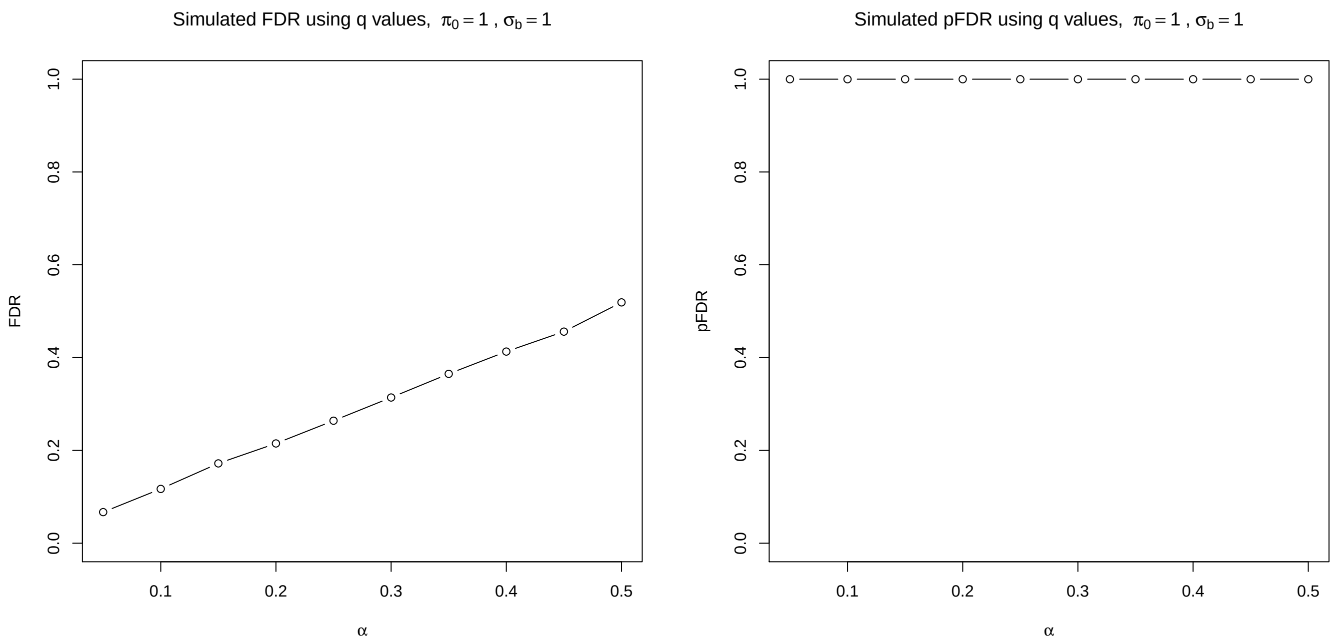


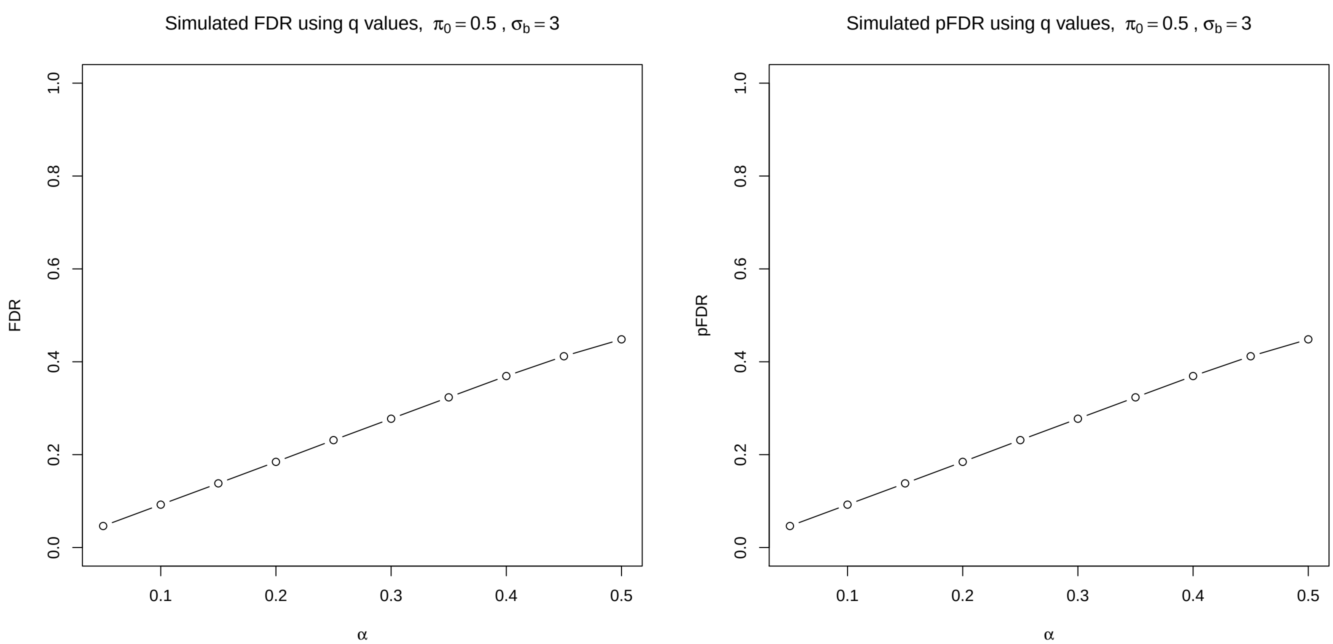




In the first case, with all discoveries false, the pFDR is naturally at 1 for all $\alpha$ and not controlled. The second case is nearly identical as with FDR, controlled at about half of $\alpha$. Again, for the final case, since no discoveries are false, the pFDR is controlled perfectly at 0.

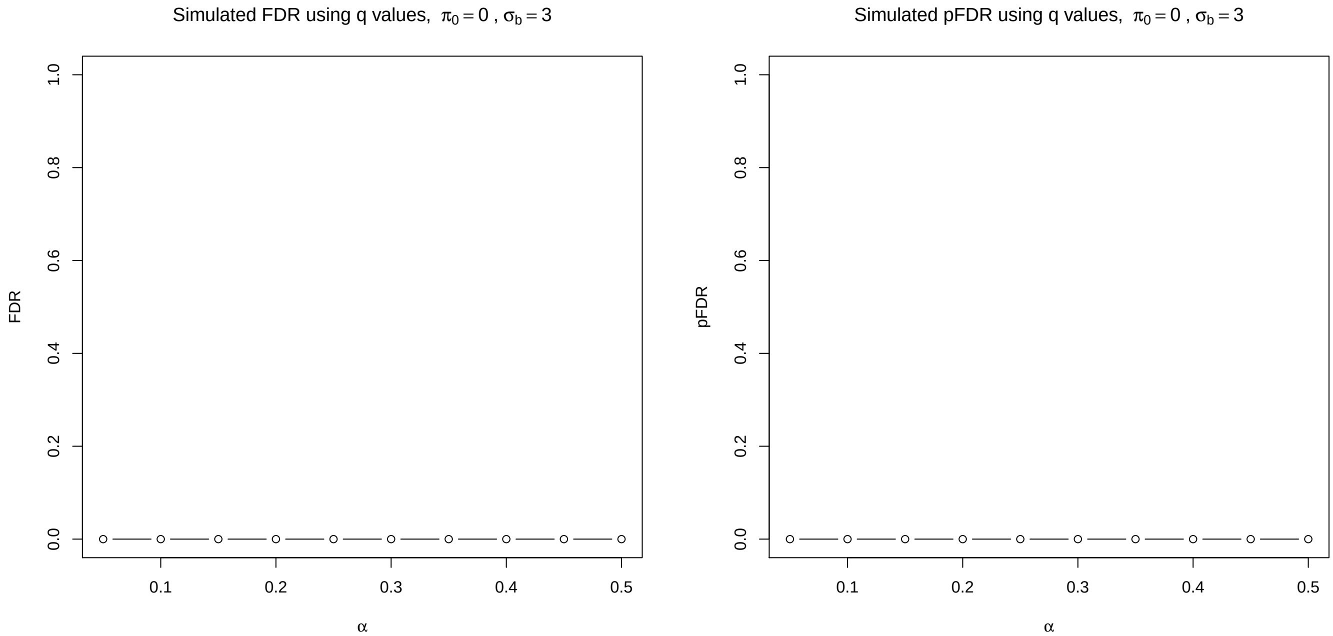
2. i) Repeating the same procedure as above, except with reject decisions based on q values, I get these results:




In this first scenario, the FDR seems to be controlled at pretty much exactly the desired rate, while the pFDR is still necessarily 1.

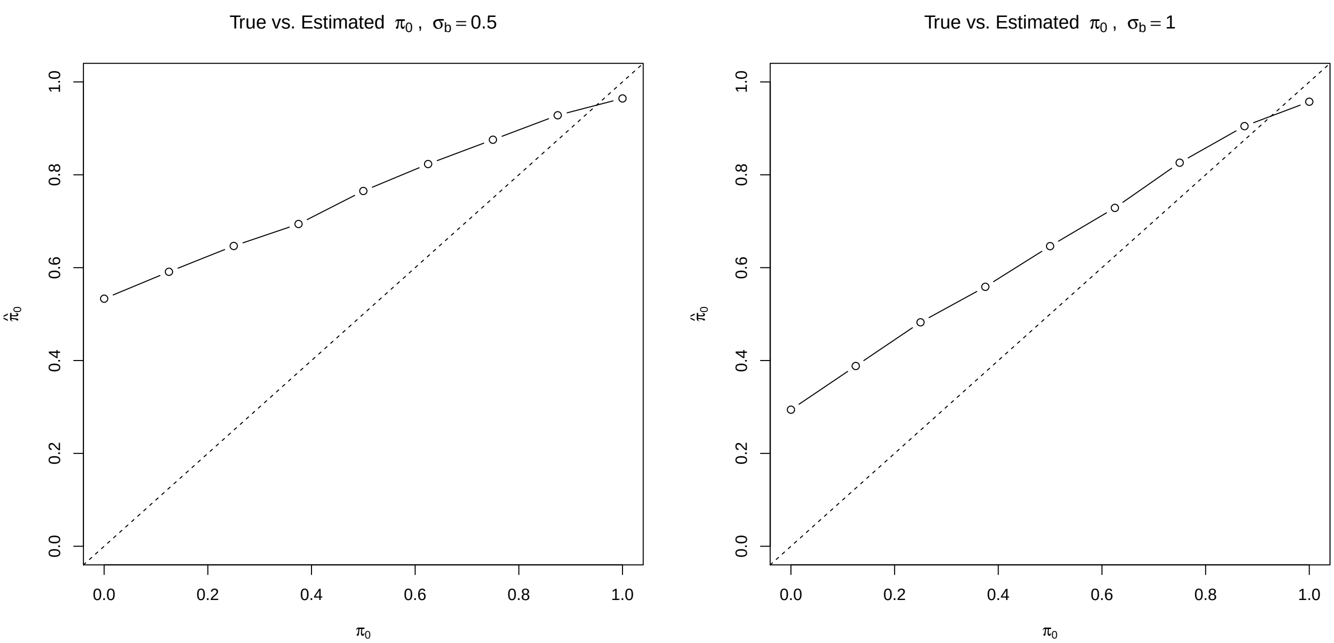



In this second scenario, we have control of both FDR and pFDR at the desired rate.

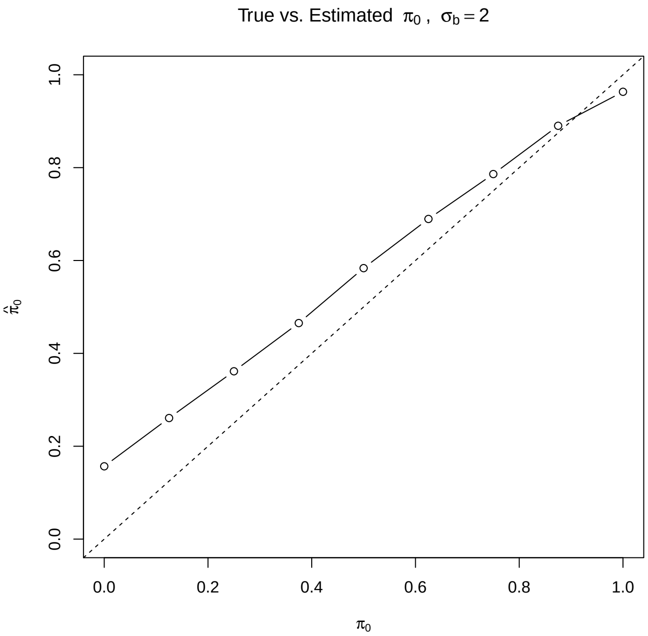


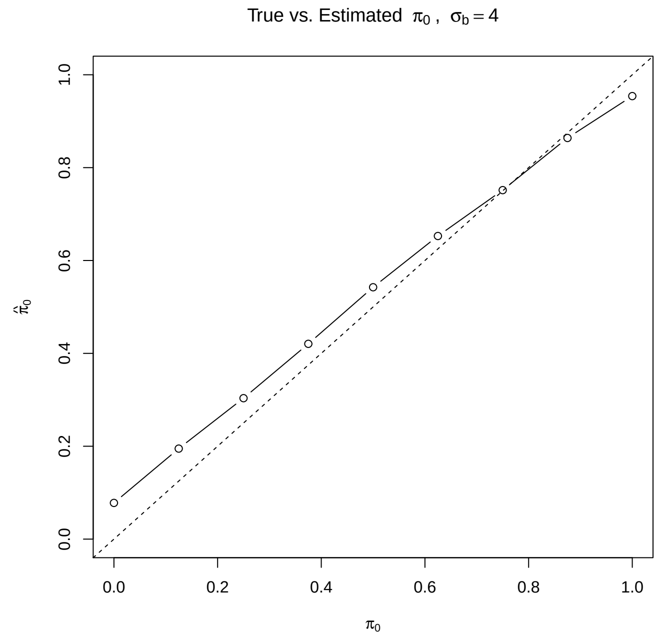

With no false discoveries, FDR and pFDR are still 0.

2) I have plotted the estimated $\pi_0$ vs actual $\pi_0$ over a range of values for $\sigma_b = .5, 1, 2, 4$. Each estimate is the average over 100 simulations.




As we can see above and below, the estimates get increasingly good as the variance goes up and the $\beta$ tend to fall farther from 0. The predictions are uniformly higher than the actual $\pi_0$, except for values near 1, and the worst predictions are for low values of $\pi_0$. In general, this method seem to provide a pretty solid estimate.




3. i) We can prove this by using the conditional independence of the $D_ij$ and decomposing the likelihood with $\sigma$ fixed:

$$
\begin{aligned}
\mathrm{P}(D \mid \beta) &= \prod_{j=1}^{m}\prod_{i=1}^{n} \mathrm{P}(D_{ij} \mid \beta_j) \\
\mathrm{P}(D \mid \beta) &\propto \prod_{j=1}^{m}\prod_{i=1}^{n} \exp\left\{-\frac{1}{2\sigma^2}(\beta_j^2 - 2\beta_j D_{ij})\right\} \\
\mathrm{P}(D \mid \beta) &\propto \prod_{j=1}^{m} \exp\left\{-\frac{n}{2\sigma^2}\left(\beta_j^2 - 2\beta_j \sum_{j=1}^{n}\frac{D_{ij}}{n}\right)\right\} \\
\mathrm{P}(D \mid \beta) &\propto \prod_{j=1}^{m} \exp\left\{-\frac{n}{2\sigma^2}\left(\beta_j^2 - 2\beta_j \bar{D}_j\right)\right\} \\
\mathrm{P}(D \mid \beta) &\propto \prod_{j=1}^{m} \exp\left\{-\frac{1}{2\frac{\sigma^2}{n}}\left(\beta_j^2 - 2\beta_j \bar{D}_j\right)\right\} \\
\mathrm{P}(D \mid \beta) &\propto \prod_{j=1}^{m} \mathrm{P}(\bar{D}_j \mid \beta_j) \\
\mathrm{P}(D \mid \beta) &\propto \mathrm{P}(\bar{D} \mid \beta)
\end{aligned}
$$

Proving that $\bar{D}$ is a sufficient statistic for the full matrix $D$.

ii) We can derive the log-likelihood by first integrating out $\beta$ to get the pdf of $\bar{D}$, and then taking the log.

Let $\sigma_n^2 = \frac{\sigma^2}{n}$.

$$\mathrm{P}(\bar{D} \mid \pi_0, \sigma_b) = \prod_{j=1}^{m} \mathrm{P}(\bar{D}_j \mid \pi_0, \sigma_b)$$

$$\mathrm{P}(\bar{D} \mid \pi_0, \sigma_b) = \prod_{j=1}^{m} \int_{\mathbb{R}} \mathrm{P}(\bar{D}_j \mid \beta_j) d\mathrm{P}_{\beta_j \mid \pi_0, \sigma_b}$$

$$\mathrm{P}(\bar{D} \mid \pi_0, \sigma_b) = \prod_{j=1}^{m} \left( \pi_0 \mathrm{P}(\bar{D}_j \mid \beta_j = 0) + \frac{1-\pi_0}{\sigma_b\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{P}(\bar{D}_j \mid \beta_j) \exp\left\{ -\frac{\beta_j^2}{2\sigma_b^2} \right\} d\beta_j \right)$$

$$\mathrm{P}(\bar{D} \mid \pi_0, \sigma_b) \propto \prod_{j=1}^{m} \left( \pi_0 + \frac{1-\pi_0}{\sigma_b\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left\{ -\frac{1}{2\sigma_n^2} \left(\beta_j^2 - 2\beta_j \bar{D}_j\right) \right\} \exp\left\{ -\frac{\beta_j^2}{2\sigma_b^2} \right\} d\beta_j \right)$$

$$\mathrm{P}(\bar{D} \mid \pi_0, \sigma_b) \propto \prod_{j=1}^{m} \left( \pi_0 + \frac{1-\pi_0}{\sigma_b\sqrt{2\pi}} \int_{\mathbb{R}} \exp\left\{ -\frac{\sigma_n^2+\sigma_b^2}{2\sigma_n^2\sigma_b^2} \left( \beta_j^2 - 2\frac{\sigma_b^2}{\sigma_n^2+\sigma_b^2}\beta_j \bar{D}_j \right) \right\} d\beta_j \right)$$

$$\mathrm{P}(\bar{D} \mid \pi_0, \sigma_b) \propto \prod_{j=1}^{m} \left( \pi_0 + \frac{1-\pi_0}{\sigma_b\sqrt{2\pi}} \exp\left\{ \frac{\sigma_b^2}{2\sigma_n^2(\sigma_n^2+\sigma_b^2)} \bar{D}_j^2 \right\} \int_{\mathbb{R}} \exp\left\{ -\frac{\sigma_n^2+\sigma_b^2}{2\sigma_n^2\sigma_b^2} \left( \beta_j - \frac{\sigma_b^2}{\sigma_n^2+\sigma_b^2} \bar{D}_j \right)^2 \right\} d\beta_j \right)$$

$$\mathrm{P}(\bar{D} \mid \pi_0, \sigma_b) \propto \prod_{j=1}^{m} \left( \pi_0 + \frac{1-\pi_0}{\sigma_b\sqrt{2\pi}} \exp\left\{ \frac{\sigma_b^2}{2\sigma_n^2(\sigma_n^2+\sigma_b^2)} \bar{D}_j^2 \right\} \sqrt{\frac{2\pi\sigma_n^2\sigma_b^2}{\sigma_n^2+\sigma_b^2}} \right)$$

$$\mathrm{P}(\bar{D} \mid \pi_0, \sigma_b) \propto \prod_{j=1}^{m} \left( \pi_0 + \frac{(1-\pi_0)\sigma_n}{\sqrt{\sigma_n^2+\sigma_b^2}} \exp\left\{ \frac{\sigma_b^2}{2\sigma_n^2(\sigma_n^2+\sigma_b^2)} \bar{D}_j^2 \right\} \right)$$

$$l(\pi_0, \sigma_b) = c + \sum_{j=1}^{m} \log\left( \pi_0 + \frac{(1-\pi_0)\sigma_n}{\sqrt{\sigma_n^2+\sigma_b^2}} \exp\left\{ \frac{\sigma_b^2}{2\sigma_n^2(\sigma_n^2+\sigma_b^2)} \bar{D}_j^2 \right\} \right)$$

iii) I've implemented a method that maximizes the likelihood given above. In testing, it converged fine, but there must be some mistake in my derivation or my code I can't figure out, since it yields consistantly the wrong answer. None the less, this is my code:

```
# compute log-likelihood for single observation in the hierarchical model
hloglik.sin <- function(Dbar, pi, sig.b, sig.n=1/sqrt(10)) {

    # fraction part second term in log
    frac  <- (1-pi)*sig.n/sqrt(sig.b^2+sig.n^2)
    # exponential part second term in log
    expon <- exp(sig.b^2*Dbar^2/(2*sig.n^4 + 2*sig.b^2))

    # actual log-likelihood
    loglik <- log(pi + frac*expon)
    return(loglik)
}

# Vectorize the log-likelihood for a single term
v.hloglik.sin <- Vectorize(hloglik.sin,c('pi','sig.b'))

# function to return the entire log-likelihood
hloglik <- function(Dbar,pi,sig.b) {
   return(sum(v.hloglik.sin(Dbar,pi=pi,sig.b=sig.b)))
```

```
}

# wrapper of above for optim
hloglik.optim <- function(param,Dbar){
    # convert to form in equation
    pi     <- inv.logit(param[1])
    sig.b <- exp(param[1])
    return(hloglik(Dbar,pi,sig.b))
}

# converts output optim back into normal form
conv.param <- function(param) {
    return(c(inv.logit(param[1]),exp(param[2])))
}
```

iv) Using the expression for likelihood we derived at the second to last step of ii, we can calculate $\mathrm{P}(\beta_j = 0 \mid \bar{D}_j, \pi_0, \sigma_b)$ as

$$\mathrm{P}(\beta_j = 0 \mid \bar{D}_j, \pi_0, \sigma_b) = \frac{\mathrm{P}(\beta_j = 0 \mid \pi_0, \sigma_b)\mathrm{P}(\bar{D}_j \mid \beta_j = 0)}{\mathrm{P}(\bar{D}_j \mid \pi_0, \sigma_b)}$$
$$\mathrm{P}(\beta_j = 0 \mid \bar{D}_j, \pi_0, \sigma_b) = \frac{\pi_0}{\pi_0 + \frac{(1-\pi_0)\sigma_n}{\sqrt{\sigma_n^2+\sigma_b^2}} \exp\left\{\frac{\sigma_b^2}{2\sigma_n^2(\sigma_n^2+\sigma_b^2)}\bar{D}_j^2\right\}}$$
$$\mathrm{P}(\beta_j = 0 \mid \bar{D}_j, \pi_0, \sigma_b) := \pi_1$$

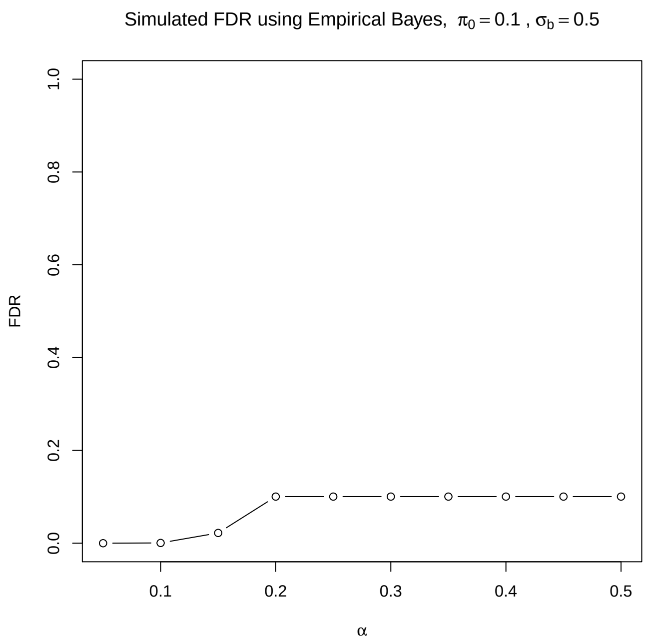
We can then derive $\mathrm{P}(\beta_j \mid \bar{D}_j, \pi_0, \sigma_b, \beta_j \neq 0)$ as

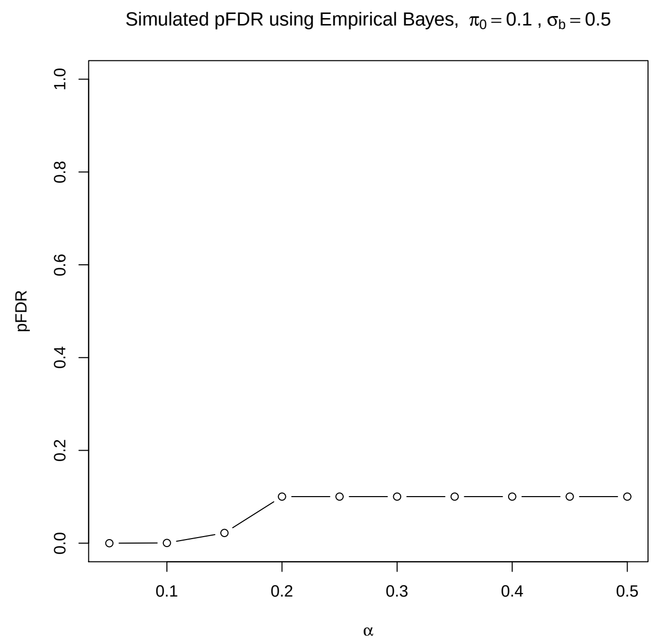
$$\mathrm{P}(\beta_j \mid \bar{D}_j, \pi_0, \sigma_b, \beta_j \neq 0) \propto \mathrm{P}(\bar{D}_j \mid \beta_j)\mathrm{P}(\beta_j \mid \pi_0, \sigma_b, \beta_j \neq 0)$$
$$\mathrm{P}(\beta_j \mid \bar{D}_j, \pi_0, \sigma_b, \beta_j \neq 0) \propto \exp\left\{-\frac{1}{2\sigma_n^2}\left(\beta_j^2 - 2\beta_j\bar{D}_j\right) - \frac{1}{2\sigma_b^2}\beta_j^2\right\}$$
$$\mathrm{P}(\beta_j \mid \bar{D}_j, \pi_0, \sigma_b, \beta_j \neq 0) \propto \exp\left\{-\frac{\sigma_n^2+\sigma_b^2}{2\sigma_n^2\sigma_b^2}\left(\beta_j^2 - 2\frac{\sigma_b^2}{\sigma_n^2+\sigma_b^2}\bar{\beta}_j D_j\right)\right\}$$
$$\mathrm{P}(\beta_j \mid \bar{D}, \pi_0, \sigma_b, \beta_j \neq 0) = \frac{1}{\sqrt{\frac{\sigma_n^2\sigma_b^2}{\sigma_n^2+\sigma_b^2}2\pi}} \exp\left\{-\frac{\sigma_n^2+\sigma_b^2}{2\sigma_n^2\sigma_b^2}\left(\beta_j - \frac{\sigma_b^2}{\sigma_n^2+\sigma_b^2}\bar{D}_j\right)^2\right\}$$
$$\mathrm{P}(\beta_j \mid \bar{D}, \pi_0, \sigma_b, \beta_j \neq 0) = \sqrt{\frac{1}{\sigma_b^2} + \frac{1}{\sigma_n^2}\frac{1}{2\pi}} \exp\left\{-\left(\frac{1}{\sigma_b^2} + \frac{1}{\sigma_n^2}\right)\left(\beta_j - \frac{\sigma_b^2}{\sigma_n^2+\sigma_b^2}\bar{D}_j\right)^2\right\}$$

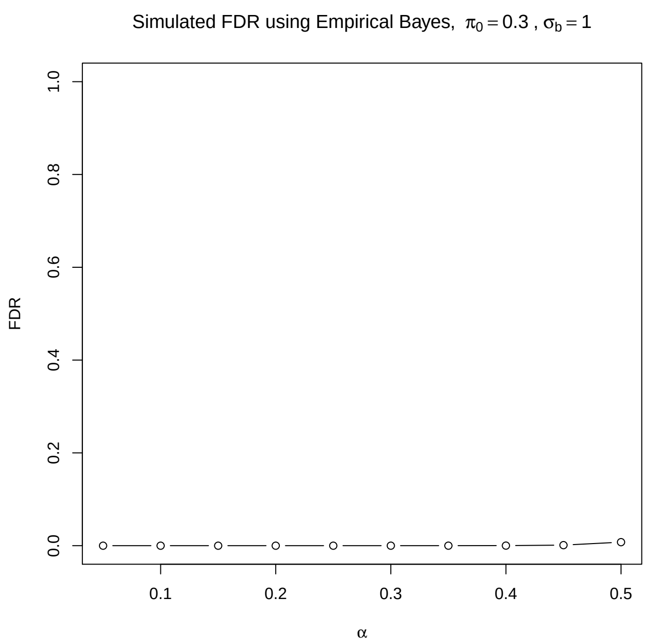
Giving us a total estimate of

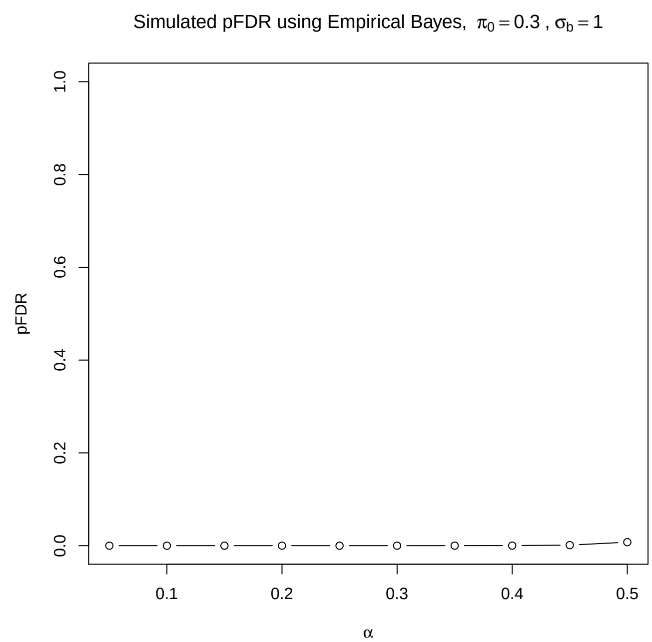
$$\mathrm{P}(\beta_j \mid \bar{D}, \pi_0, \sigma_b) = \begin{cases} (1-\pi_1)\sqrt{\frac{1}{\sigma_b^2} + \frac{1}{\sigma_n^2}\frac{1}{2\pi}} \exp\left\{-\left(\frac{1}{\sigma_b^2} + \frac{1}{\sigma_n^2}\right)\left(\beta_j - \frac{\sigma_b^2}{\sigma_n^2+\sigma_b^2}\bar{D}_j\right)^2\right\} & \text{if } \beta_j \neq 0 \\ \pi_1 & \text{if } \beta_j = 0 \end{cases}$$

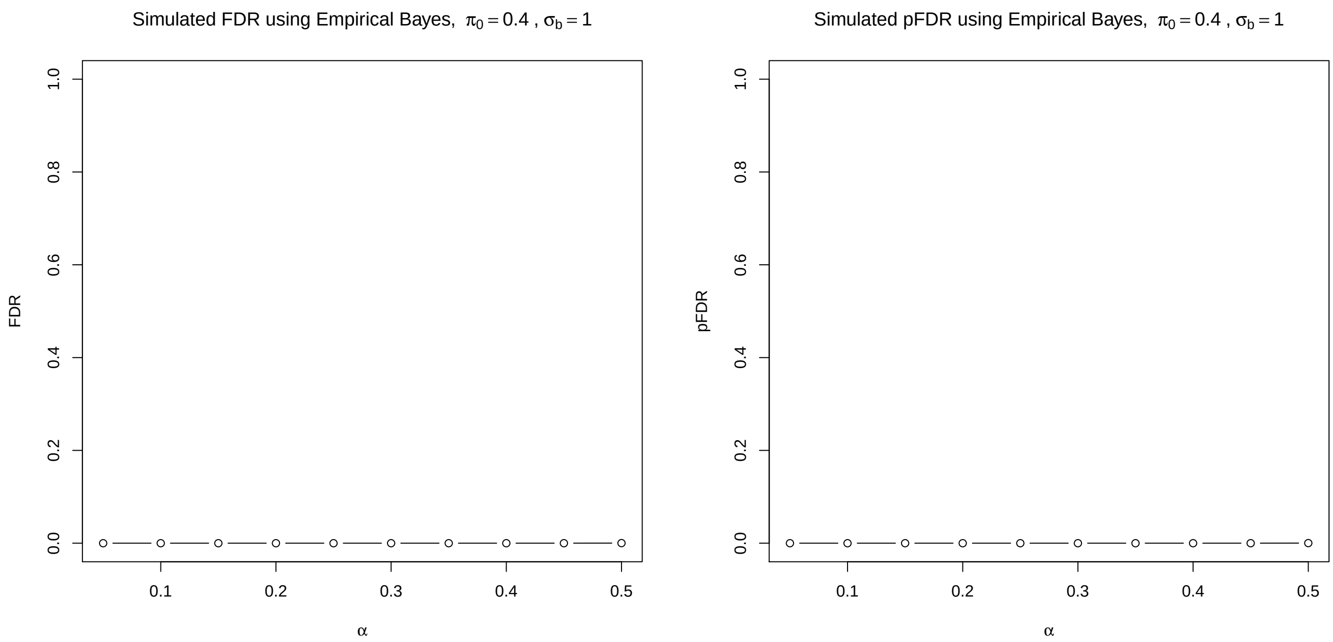
v) Since my likelihood is screwed up, I didn't actually use the optimized parameters and cheated. However, I still use the likelihood in part of my calculation of p-values as stated above, so the end result seems to still be that something is off. Non withstanding, here are my results:


Simulated FDR using Empirical Bayes, $\pi_0 = 0.1$ , $\sigma_b = 0.5$
FDR
α


Simulated pFDR using Empirical Bayes, $\pi_0 = 0.1$ , $\sigma_b = 0.5$
pFDR
α


Simulated FDR using Empirical Bayes, $\pi_0 = 0.3$ , $\sigma_b = 1$
FDR
α


Simulated pFDR using Empirical Bayes, $\pi_0 = 0.3$ , $\sigma_b = 1$
pFDR
α




I think my mistake made my test way overly conservative in a way I don't quite understand.