Progress Report

Victor Huang

November 29, 2020

Overview:

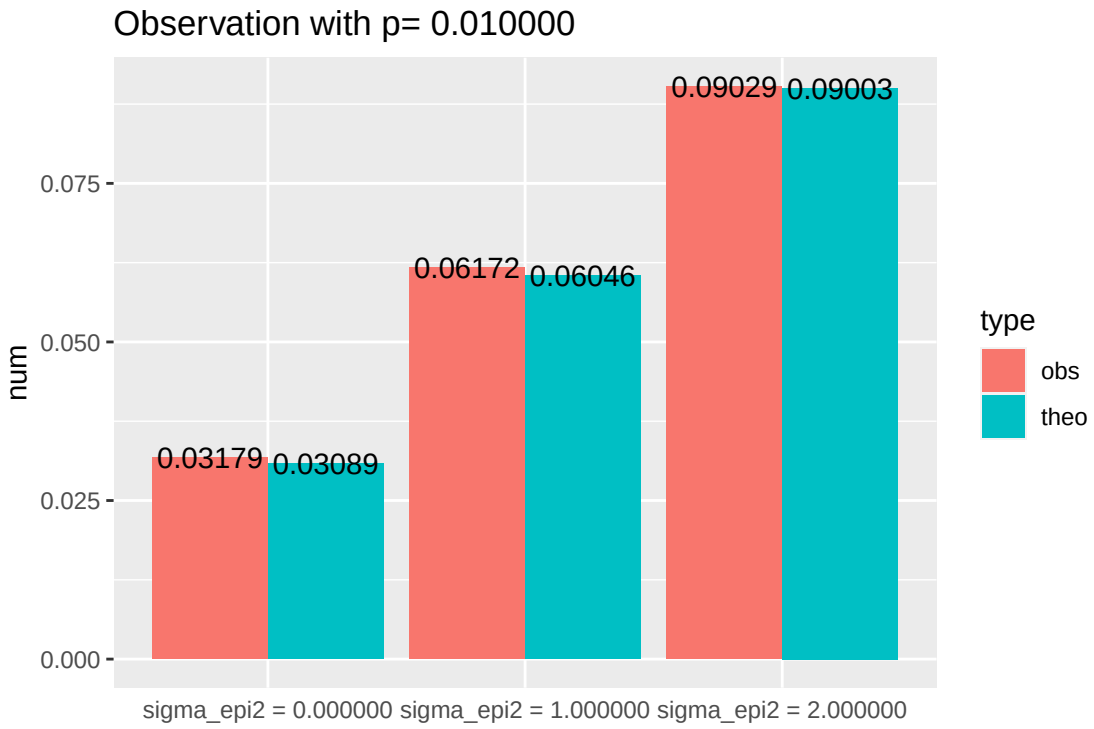
Concerned variables and underlying Monte-Carlo model:

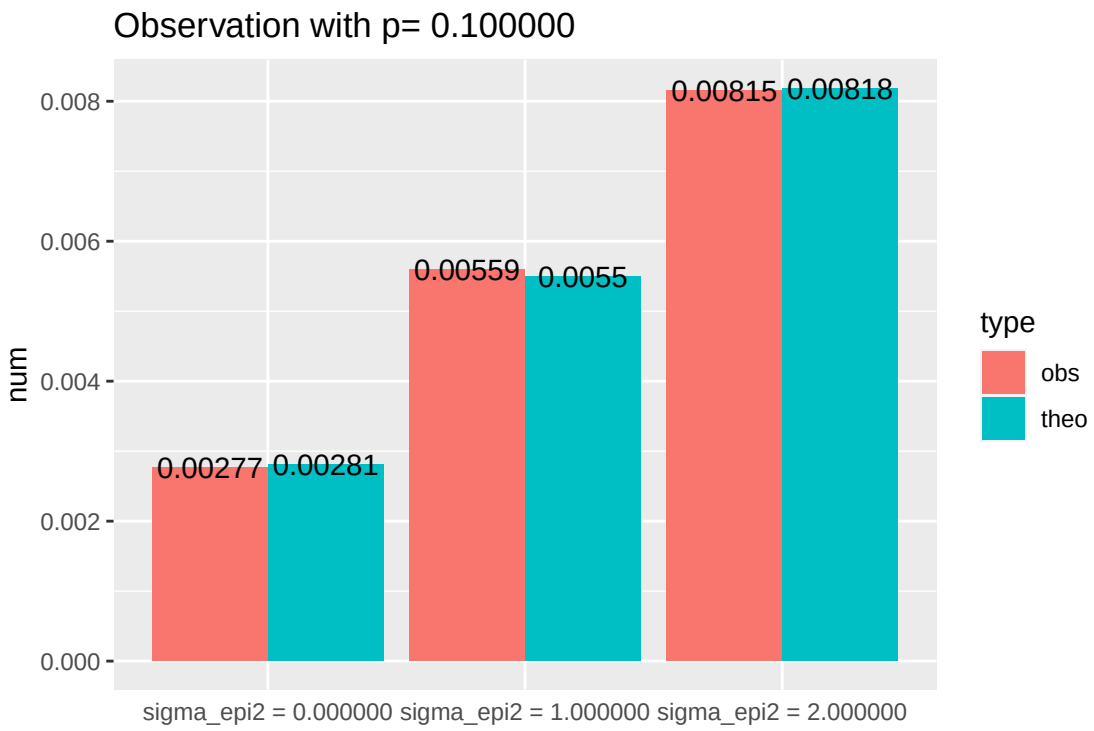
```
nSims: Number of simulations to run for each groups of parameter.
N: Size of population.
p: Probablity an observation is sampled.
sigma_e2: Variance of Model error e.
sigma_epi2: variance of unpredicted error epislon.
yi: Dependent variable without unpredicted error.
xi: Independent variable.
cap_y: Dependent variable with unpredicted error.

Simulation2: A function that randomly draws samples from population given. Each observation \\
has probability p of being samapled. It repeats this operation nSims times and will output a \\
vector of beta, the empirical value for an OLS fitting model based on cap_y and xi.
```

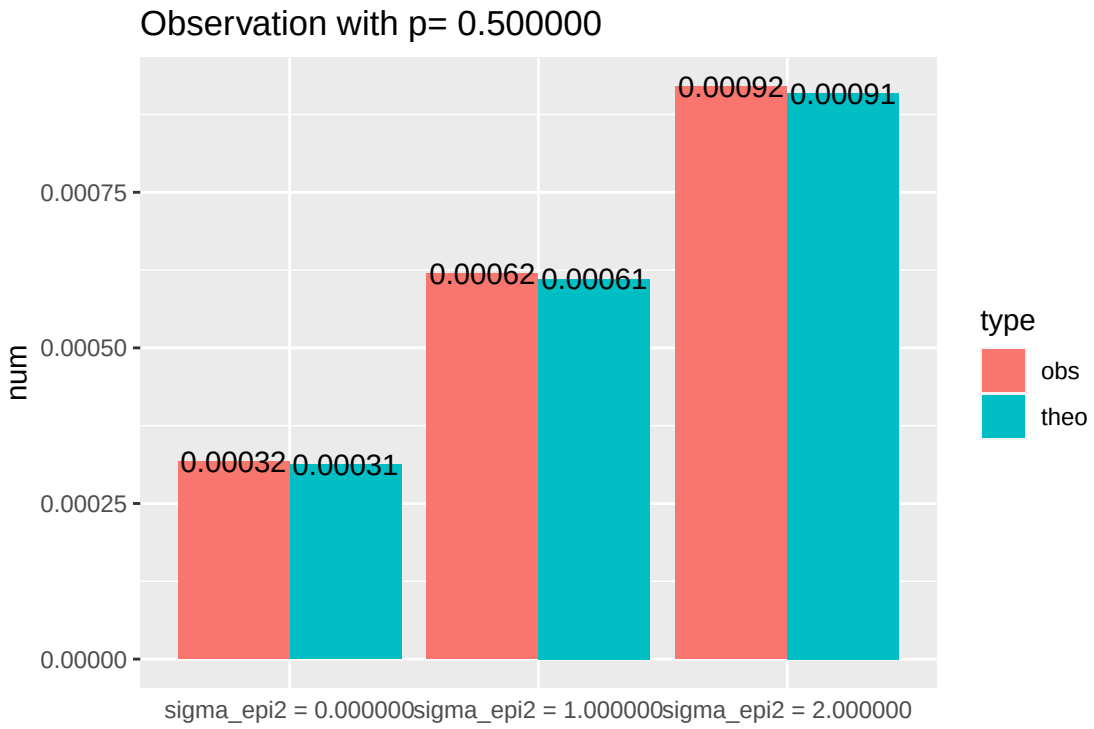
```r
# setting parameters
sigma_e2<-1
N<-10000
beta<-2
nSims<-10000

p<-c(0.01,0.1,0.5,0.9,0.99)
sigma_epi2<-c(0,1,2)

x_para<-set_para(0,1)
e_para<-set_para(0,sigma_e2)

# generating population
xi<-pop_gen(x_para,N,'uniform')
ei<-pop_gen(e_para,N,'normal')

yi<-xi*beta+ei



result<-NULL

# for each group of simulations, generate one population with
```

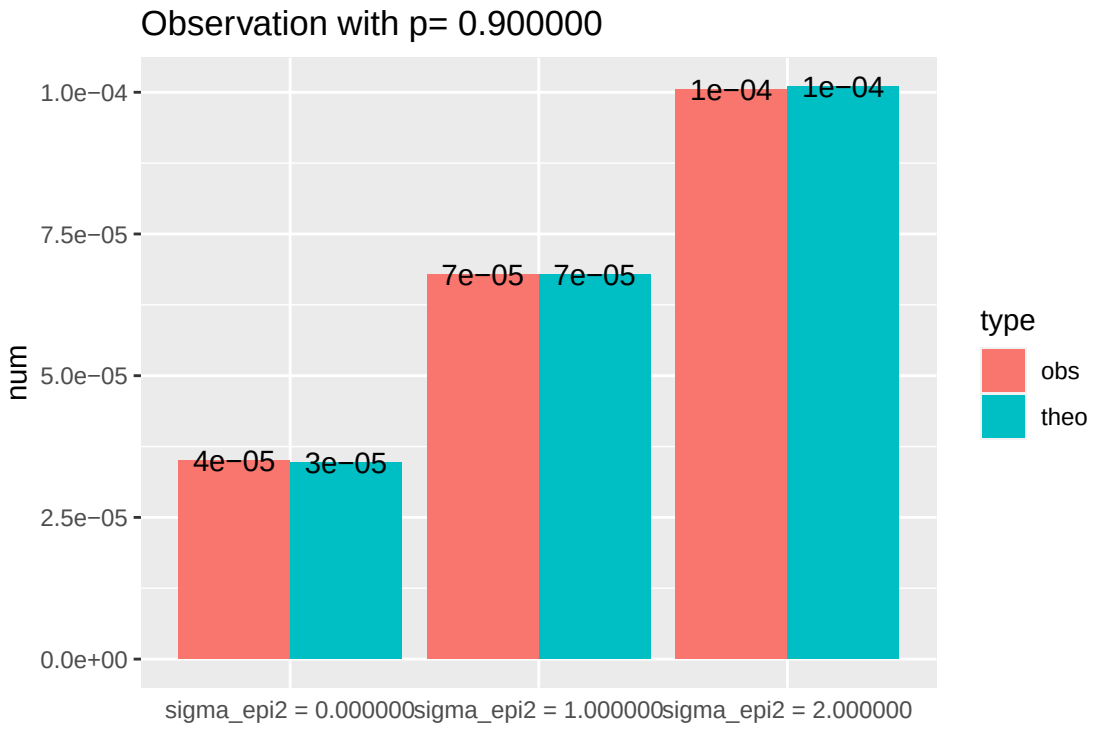
```
# independent error epislon.

for (j in c(1:length(sigma_epi2))){
    epi_para<- set_para(0,sigma_epi2[j])
    epi<-pop_gen(epi_para,N,'normal')

    cap_y<-yi+epi

# sample from the population with probability p and return the
# variance of the beta.
  for (i in c(1:length(p))){
    start.time.small<-Sys.time()
    var_beta<-var(simulation2(cap_y,xi,p[i],nSims))
    theo_var<-((1-p[i])/p[i])*
      ((sum(xi**2*ei**2))/((sum(xi**2))**2)+sigma_epi2[j]/(sum(xi**2)))

    end.time.small<-Sys.time()

# rearranging the data for presentation and making graph
    out<-cbind(obs=var_beta,theo=theo_var,prob=sprintf('p = %f ',p[i]),
               sigma=sprintf('sigma_epi2 = %f',sigma_epi2[j]),
               time=end.time.small-start.time.small)
    result<-rbind(result,out)
  }
}
```

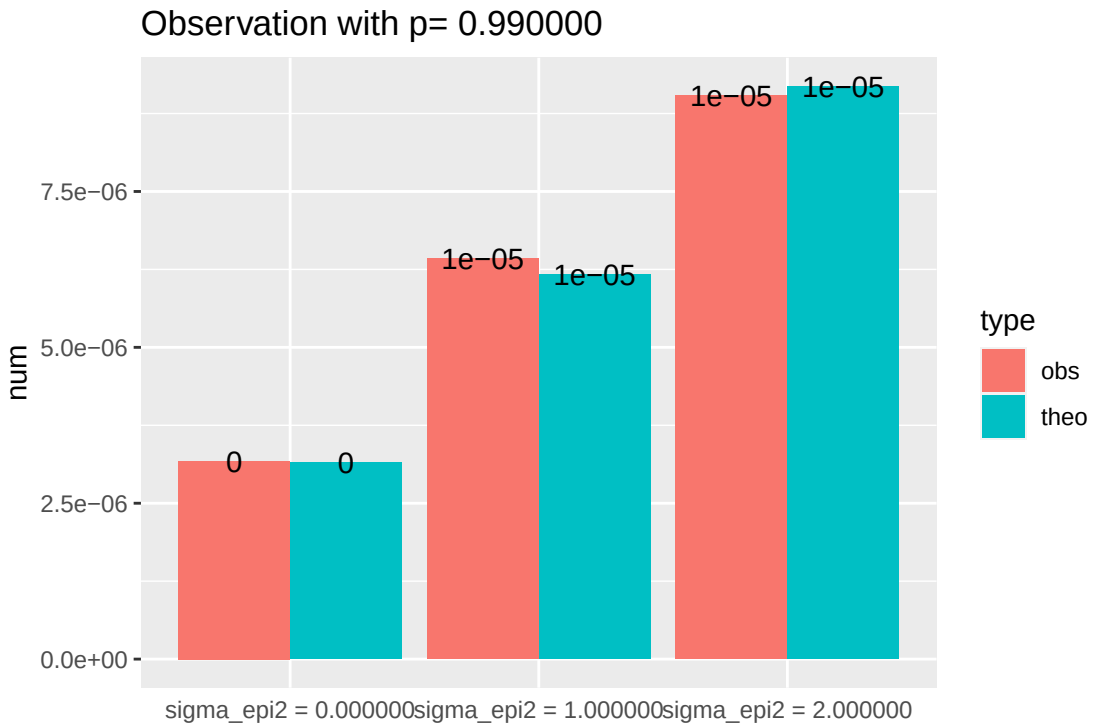
```
result1<-as.data.frame(result) %>% gather(type,num,obs:theo) %>% type_convert()

for (i in (c(1:length(p)))){
a<-ggplot(result1 %>% filter(prob==sprintf('p = %f',p[i])))+
    geom_col(aes(sigma,num,fill=type),position='dodge',width = 0.9)+
  geom_text(aes(sigma,num,label=round(as.numeric(num),5),group=type),
            position = position_dodge(width = 0.9))+
  theme(axis.title.x=element_blank(),
        axis.ticks.x=element_blank(),
        plot.margin = unit(c(1,1,1,1), "cm"))+
  ggtitle(sprintf('Observation with p= %f',p[i]))
plot(a)
}
```


Observation with p= 0.010000
num
0.075
0.050
0.025
0.000
0.03179
0.03089
0.06172
0.06046
0.09029
0.09003
sigma_epi2 = 0.000000
sigma_epi2 = 1.000000
sigma_epi2 = 2.000000
type
obs
theo

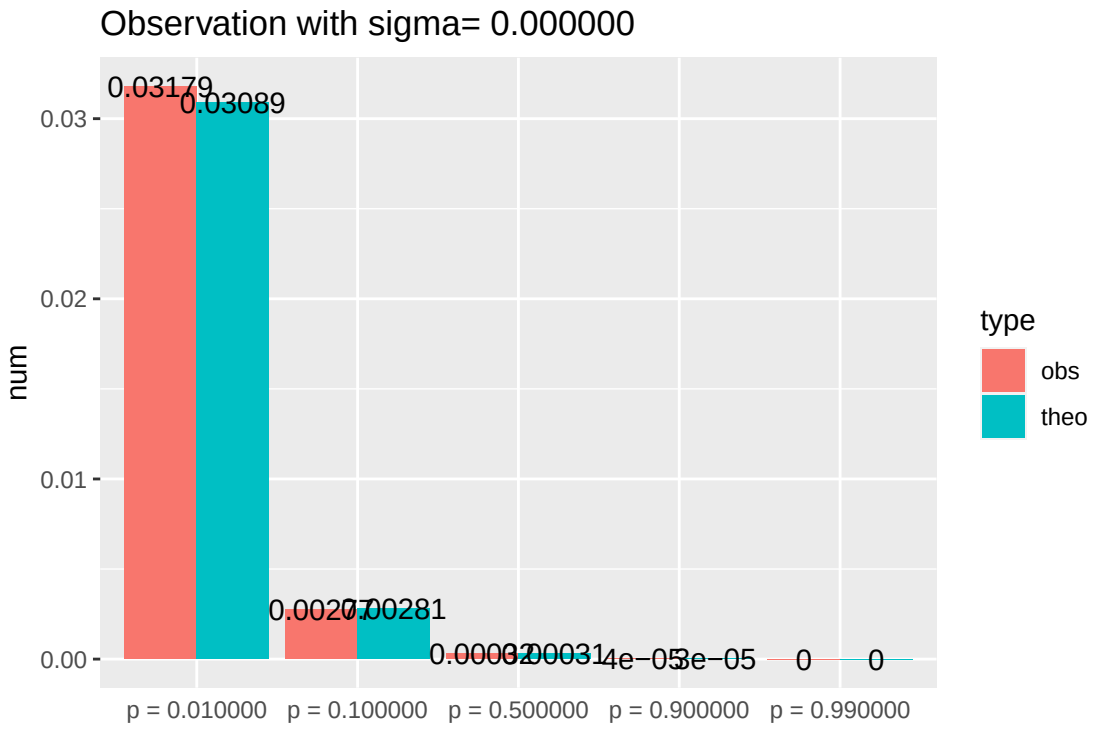

Observation with p= 0.100000
num
0.008
0.006
0.004
0.002
0.000
0.00277
0.00281
0.00559
0.0055
0.00815
0.00818
sigma_epi2 = 0.000000
sigma_epi2 = 1.000000
sigma_epi2 = 2.000000
type
obs
theo

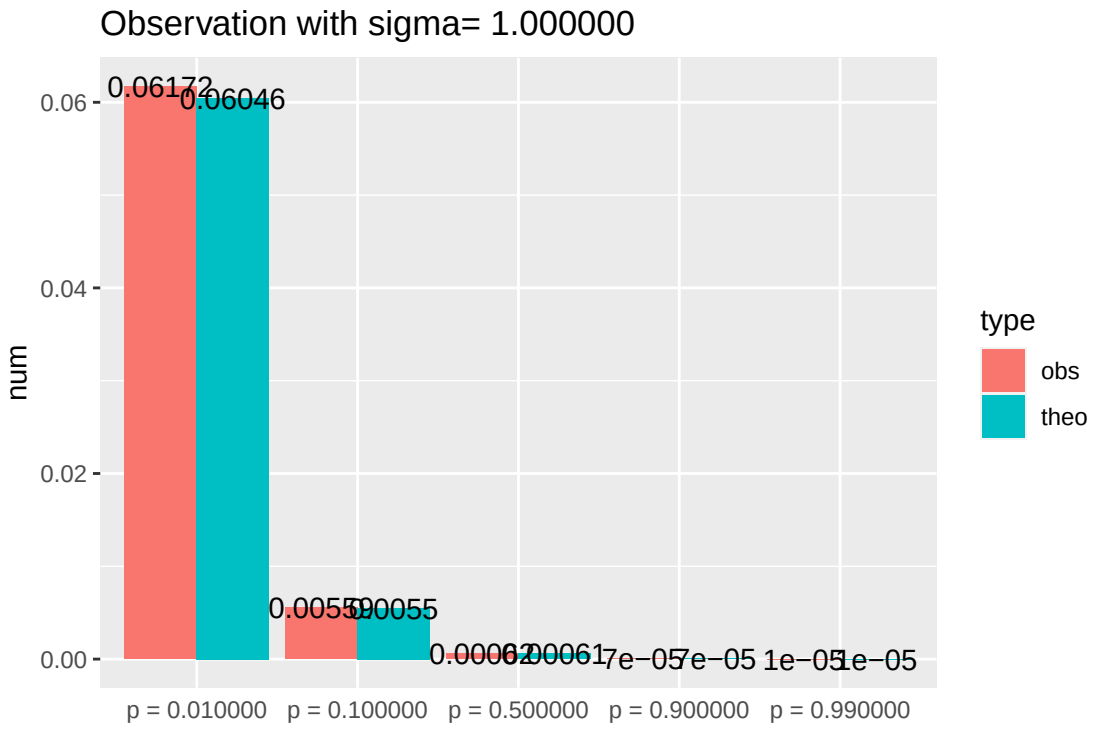

Observation with p= 0.500000
num
0.00075
0.00050
0.00025
0.00000
0.00032
0.00031
0.00062
0.00061
0.00092
0.00091
sigma_epi2 = 0.000000
sigma_epi2 = 1.000000
sigma_epi2 = 2.000000
type
obs
theo

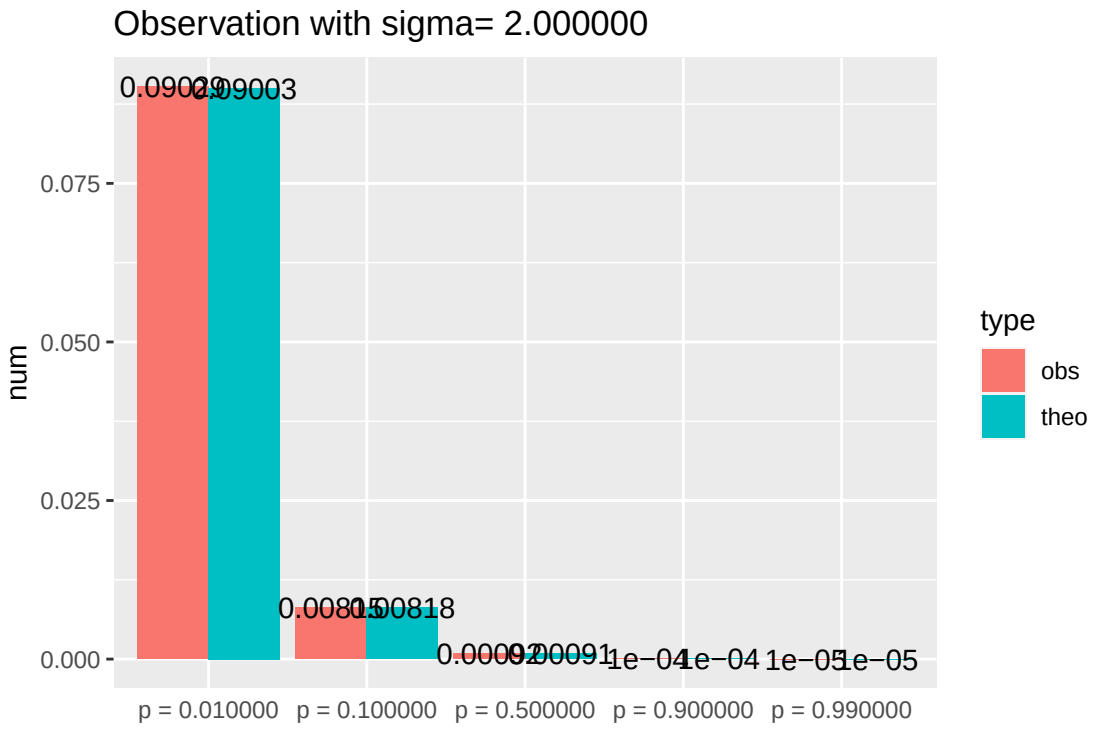

Observation with p= 0.900000
num
1.0e−04
7.5e−05
5.0e−05
2.5e−05
0.0e+00
4e−05
3e−05
7e−05
7e−05
1e−04
1e−04
sigma_epi2 = 0.000000
sigma_epi2 = 1.000000
sigma_epi2 = 2.000000
type
obs
theo




```
for (i in (c(1:length(sigma_epi2)))){
a<-ggplot(result1 %>% filter(sigma==sprintf('sigma_epi2 = %f',sigma_epi2[i])))+
    geom_col(aes(prob,num,fill=type),position='dodge',width = 0.9)+
  geom_text(aes(prob,num,label=round(as.numeric(num),5),group=type),
            position = position_dodge(width = 0.9))+
  theme(axis.title.x=element_blank(),
        axis.ticks.x=element_blank(),
        plot.margin = unit(c(1,1,1,1), "cm"))+
  ggtitle(sprintf('Observation with sigma= %f',sigma_epi2[i]))
plot(a)
}
```


Observation with sigma= 0.000000
num
0.03
0.02
0.01
0.00
0.03179
0.03089
0.00277
0.00281
0.00062
0.00031
4e−05
3e−05
0
0
p = 0.010000
p = 0.100000
p = 0.500000
p = 0.900000
p = 0.990000
type
obs
theo


Observation with sigma= 1.000000
num
0.06
0.04
0.02
0.00
0.06172
0.06046
0.00559
0.00055
0.00062
0.00061
7e−05
7e−05
1e−05
1e−05
p = 0.010000
p = 0.100000
p = 0.500000
p = 0.900000
p = 0.990000
type
obs
theo



prob	sigma	obs	theo
p = 0.010000	sigma_epi2 = 0.000000	0.0317938235664232	0.0308928726969307
p = 0.100000	sigma_epi2 = 0.000000	0.00277142149239537	0.00280844297244824
p = 0.500000	sigma_epi2 = 0.000000	0.000317255457571464	0.000312049219160916
p = 0.900000	sigma_epi2 = 0.000000	3.50728525193868e-05	3.4672135462324e-05
p = 0.990000	sigma_epi2 = 0.000000	3.17584783624905e-06	3.15201231475673e-06
p = 0.010000	sigma_epi2 = 1.000000	0.0617195849737955	0.0604618278421749
p = 0.100000	sigma_epi2 = 1.000000	0.00559174654208386	0.00549652980383408
p = 0.500000	sigma_epi2 = 1.000000	0.000619398770939344	0.000610725533759343
p = 0.900000	sigma_epi2 = 1.000000	6.77835878564246e-05	6.78583926399269e-05
p = 0.990000	sigma_epi2 = 1.000000	6.41970563267083e-06	6.16894478544791e-06
p = 0.010000	sigma_epi2 = 2.000000	0.0902945155309989	0.0900307829874192
p = 0.100000	sigma_epi2 = 2.000000	0.0081490937821573	0.00818461663521992
p = 0.500000	sigma_epi2 = 2.000000	0.000920149770052507	0.000909401848357769
p = 0.900000	sigma_epi2 = 2.000000	0.000100420738031362	0.00010104464981753
p = 0.990000	sigma_epi2 = 2.000000	9.0356336407792e-06	9.18587725613909e-06

prob	sigma	time
p = 0.010000	sigma_epi2 = 0.000000	3.89063000679016
p = 0.100000	sigma_epi2 = 0.000000	4.2536289691925
p = 0.500000	sigma_epi2 = 0.000000	5.25095796585083
p = 0.900000	sigma_epi2 = 0.000000	5.10132598876953
p = 0.990000	sigma_epi2 = 0.000000	5.04953122138977
p = 0.010000	sigma_epi2 = 1.000000	4.07208013534546
p = 0.100000	sigma_epi2 = 1.000000	4.62167382240295
p = 0.500000	sigma_epi2 = 1.000000	5.33071804046631
p = 0.900000	sigma_epi2 = 1.000000	5.14327812194824
p = 0.990000	sigma_epi2 = 1.000000	5.1153199672699
p = 0.010000	sigma_epi2 = 2.000000	3.99129390716553
p = 0.100000	sigma_epi2 = 2.000000	4.09105920791626
p = 0.500000	sigma_epi2 = 2.000000	5.29387998580933
p = 0.900000	sigma_epi2 = 2.000000	5.80045413970947
p = 0.990000	sigma_epi2 = 2.000000	5.24896788597107
time in total	time in total	1.25047199726105