Progress Report

Victor Huang

November 29, 2020

Overview:

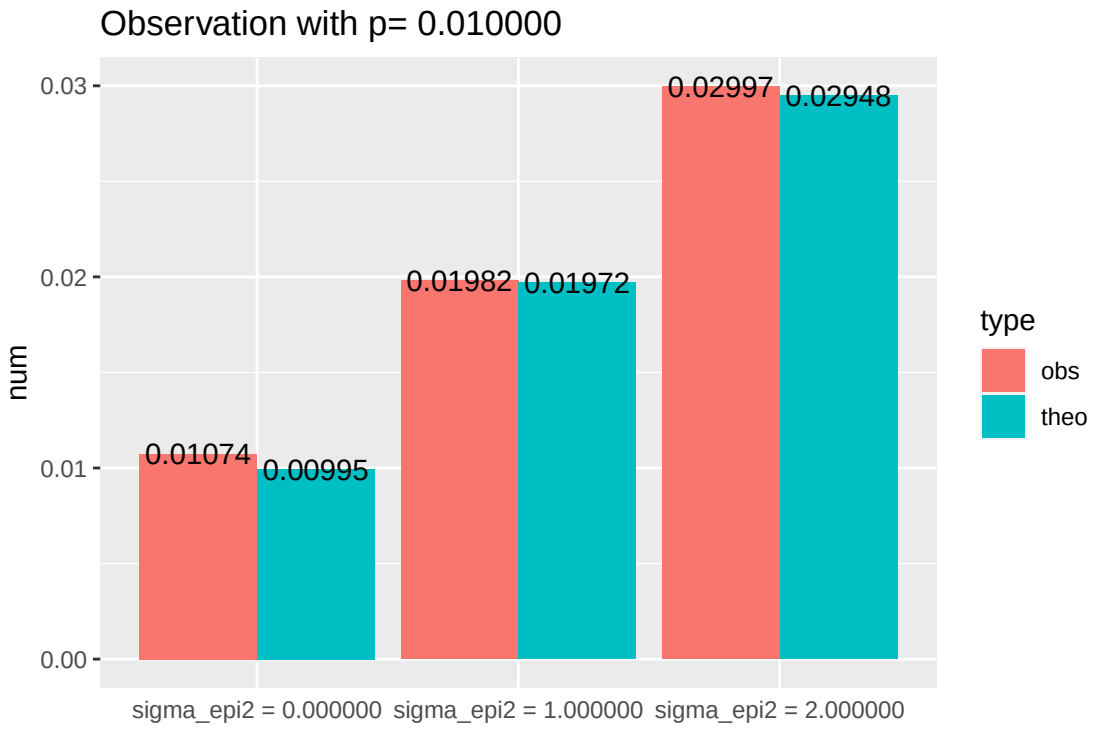
Concerned variables and underlying Monte-Carlo model:

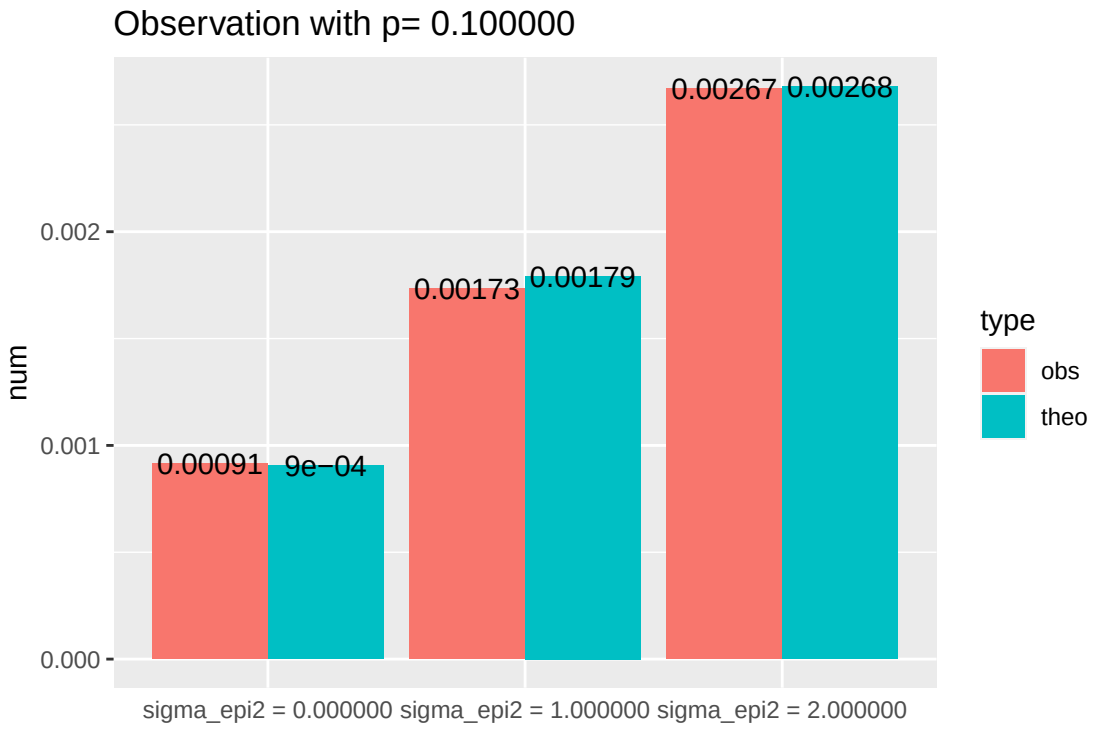
```
nSims: Number of simulations to run for each groups of parameter.
N: Size of population.
p: Probablity an observation is sampled.
sigma_e2: Variance of Model error e.
sigma_epi2: variance of unpredicted error epislon.
yi: Dependent variable without unpredicted error.
xi: Independent variable.
cap_y: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

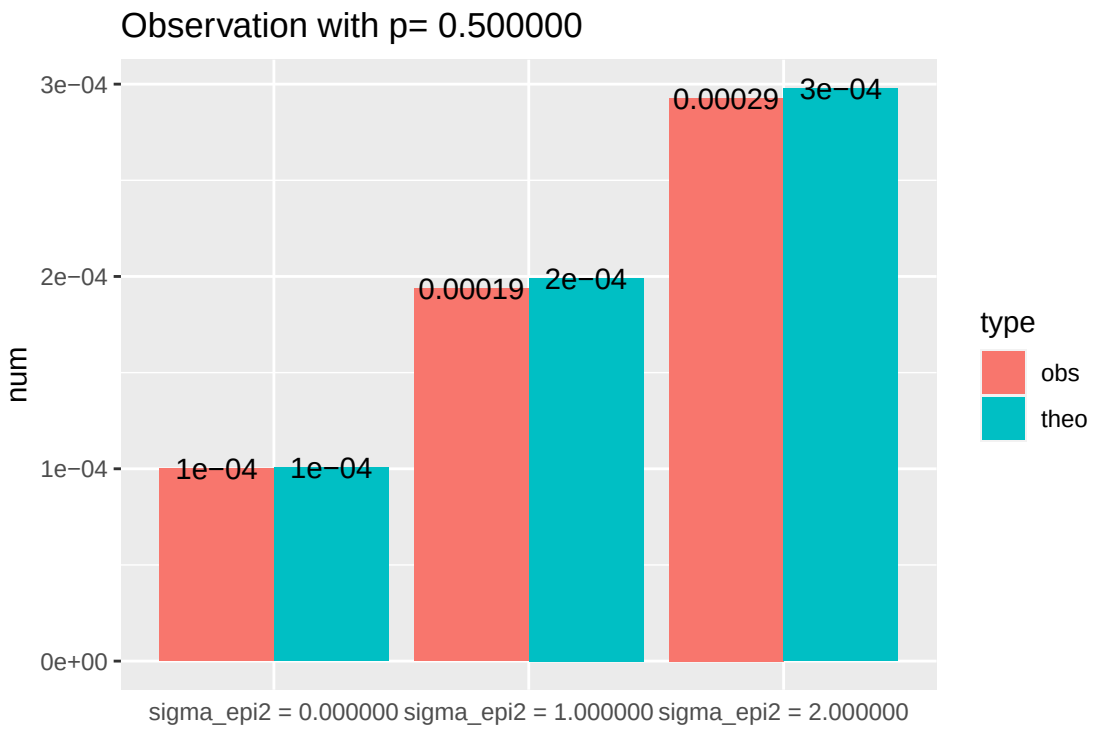
```r
# setting parameters
sigma_e2<-1
N<-10000
beta<-2
nSims<-10000

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,'normal')
ei<-pop_gen(e_para,N,'normal')

yi<-xi*beta+ei



result<-NULL

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

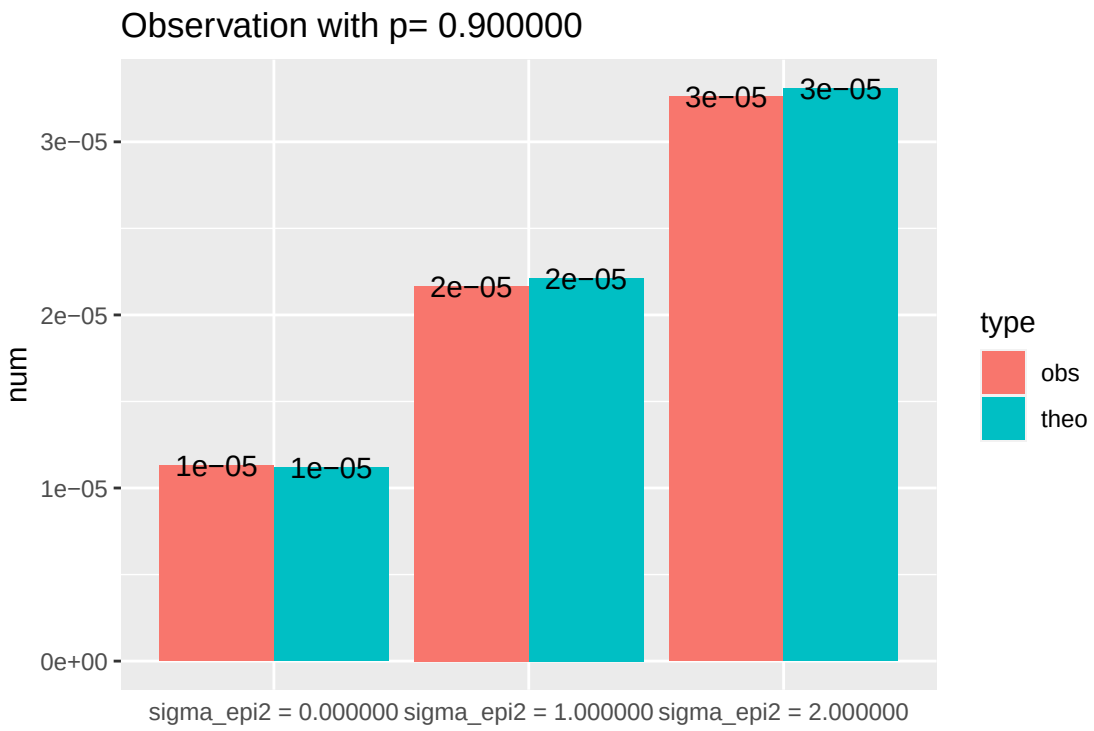
```
# independent error epislon.

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)
  }
}
```

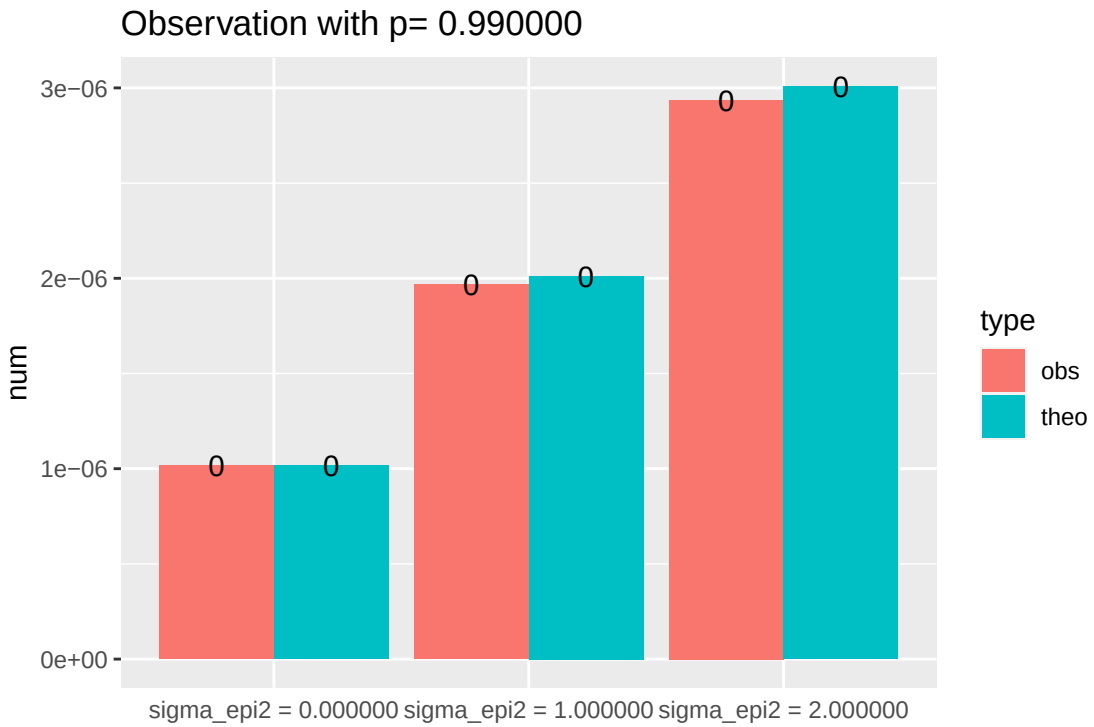
```
result1<-as.data.frame(result) %>% gather(type,num,obs:theo) %>% type_convert()

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
0.03
0.02
0.01
0.00
0.01074
0.00995
0.01982
0.01972
0.02997
0.02948
sigma_epi2 = 0.000000
sigma_epi2 = 1.000000
sigma_epi2 = 2.000000
type
obs
theo

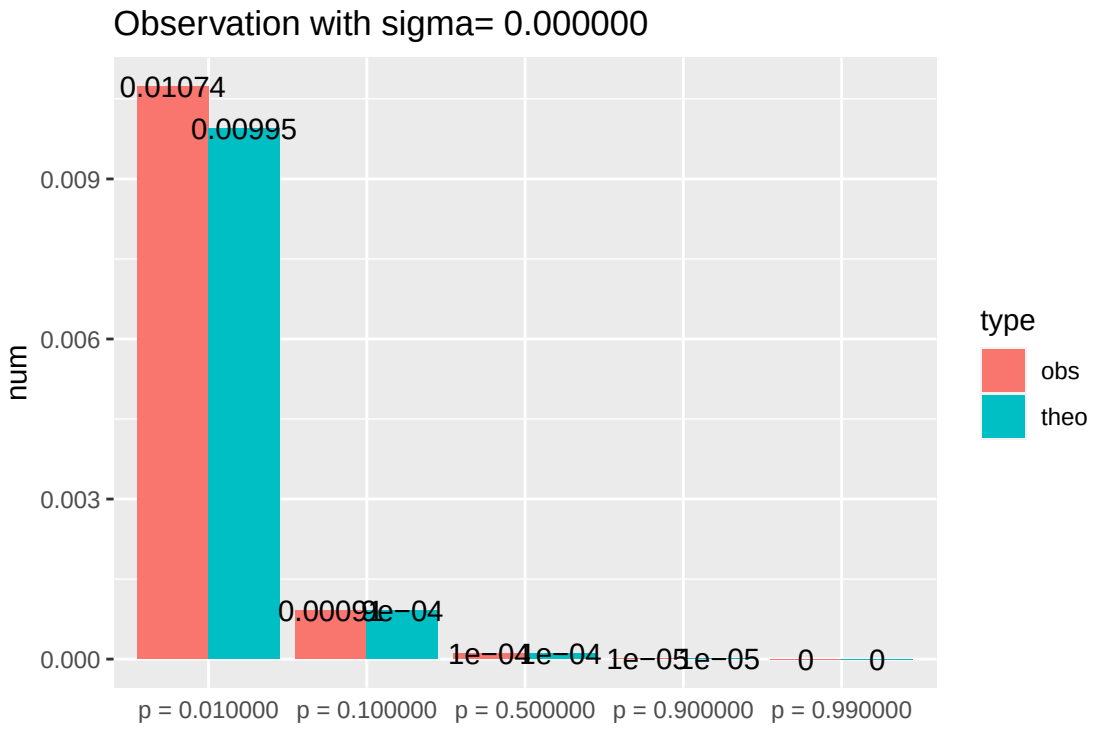

Observation with p= 0.100000
num
0.002
0.001
0.000
0.00091
9e−04
0.00173
0.00179
0.00267
0.00268
sigma_epi2 = 0.000000
sigma_epi2 = 1.000000
sigma_epi2 = 2.000000
type
obs
theo

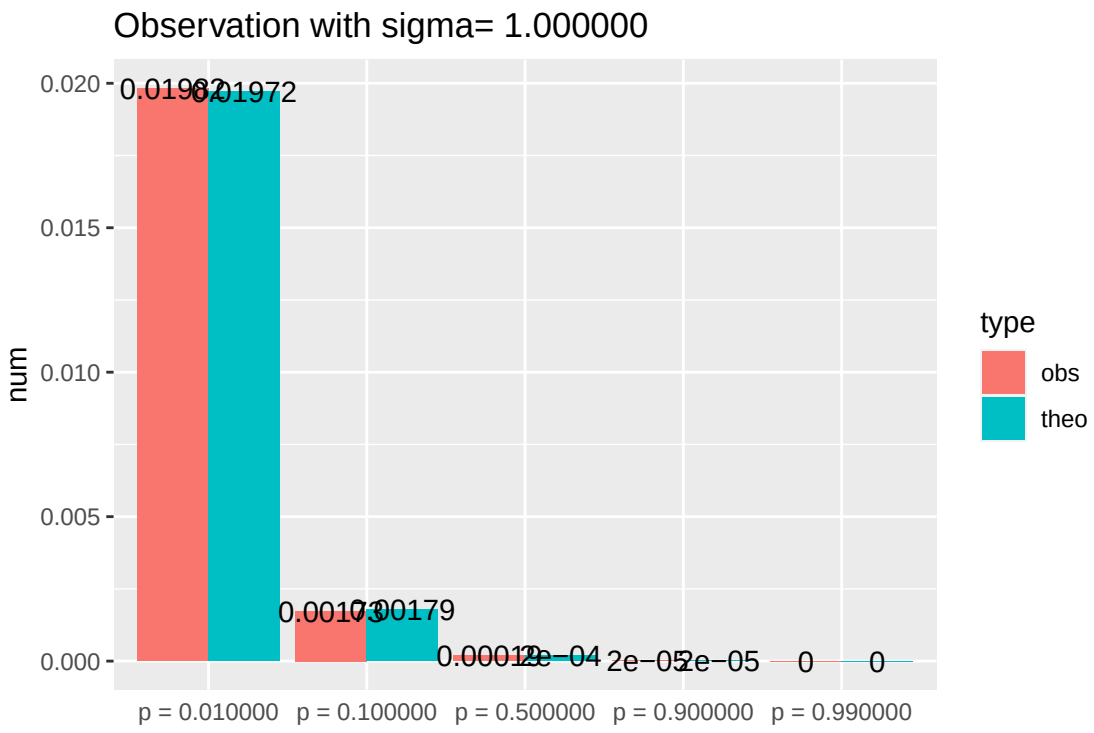

Observation with p= 0.500000
3e−04
2e−04
1e−04
0e+00
num
1e−04
1e−04
0.00019
2e−04
0.00029
3e−04
sigma_epi2 = 0.000000
sigma_epi2 = 1.000000
sigma_epi2 = 2.000000
type
obs
theo

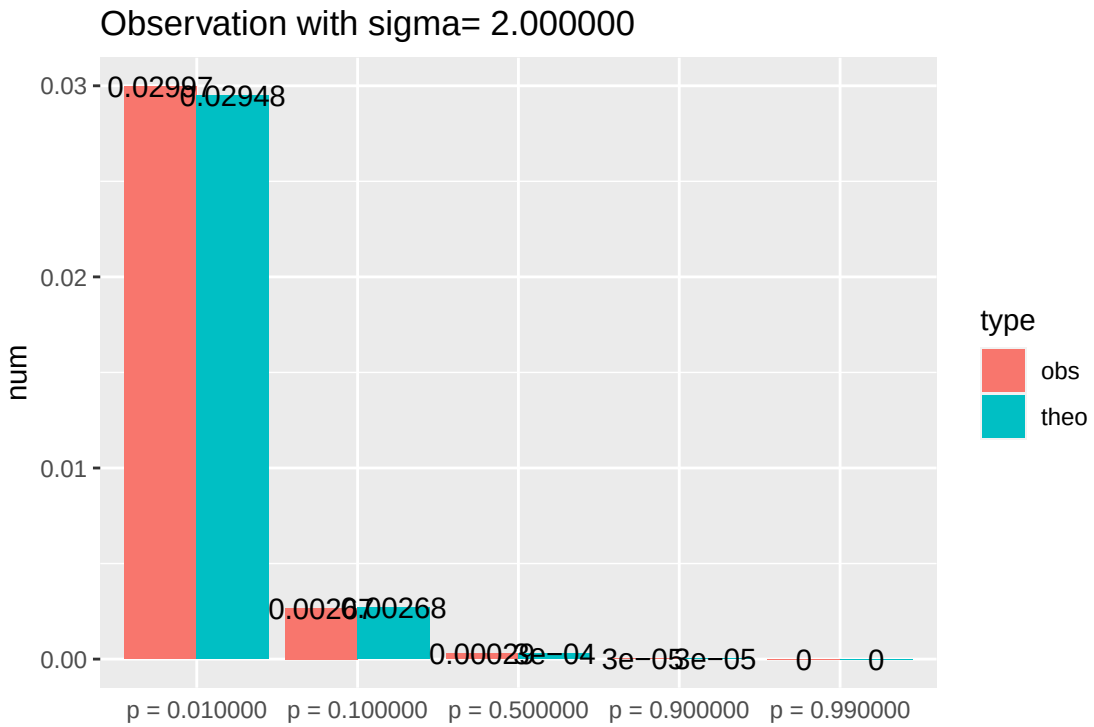

Observation with p= 0.900000
num
3e−05
2e−05
1e−05
0e+00
1e−05
1e−05
2e−05
2e−05
3e−05
3e−05
sigma_epi2 = 0.000000
sigma_epi2 = 1.000000
sigma_epi2 = 2.000000
type
obs
theo




```
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
0.009
0.006
0.003
0.000
0.01074
0.00995
0.00099
1e−04
1e−04
1e−04
1e−05
1e−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
0.020
0.015
0.010
0.005
0.000
0.019882
0.019720
0.001730
0.001790
0.000120
2e−04
2e−05
2e−05
0
0
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	0.0107351716856739	0.00994779387593741
p = 0.100000	sigma_epi2 = 0.000000	0.000914028736408776	0.000904344897812491
p = 0.500000	sigma_epi2 = 0.000000	9.99490743653437e-05	0.00010048276642361
p = 0.900000	sigma_epi2 = 0.000000	1.12802009108874e-05	1.11647518248456e-05
p = 0.990000	sigma_epi2 = 0.000000	1.01796988414198e-06	1.01497743862233e-06
p = 0.010000	sigma_epi2 = 1.000000	0.019824360698349	0.0197155975086595
p = 0.100000	sigma_epi2 = 1.000000	0.00173379381176622	0.00179232704624177
p = 0.500000	sigma_epi2 = 1.000000	0.000193587397300893	0.000199147449582419
p = 0.900000	sigma_epi2 = 1.000000	2.16657610590346e-05	2.21274943980466e-05
p = 0.990000	sigma_epi2 = 1.000000	1.96603828135991e-06	2.01159039982242e-06
p = 0.010000	sigma_epi2 = 2.000000	0.02997040729814	0.0294834011413816
p = 0.100000	sigma_epi2 = 2.000000	0.00266955792009593	0.00268030919467106
p = 0.500000	sigma_epi2 = 2.000000	0.000292805710898252	0.000297812132741229
p = 0.900000	sigma_epi2 = 2.000000	3.26242695421498e-05	3.30902369712476e-05
p = 0.990000	sigma_epi2 = 2.000000	2.93629270284254e-06	3.00820336102251e-06

prob	sigma	time
p = 0.010000	sigma_epi2 = 0.000000	3.95243191719055
p = 0.100000	sigma_epi2 = 0.000000	4.24368596076965
p = 0.500000	sigma_epi2 = 0.000000	5.08839416503906
p = 0.900000	sigma_epi2 = 0.000000	4.93181204795837
p = 0.990000	sigma_epi2 = 0.000000	4.9098379611969
p = 0.010000	sigma_epi2 = 1.000000	4.02327108383179
p = 0.100000	sigma_epi2 = 1.000000	4.44408512115479
p = 0.500000	sigma_epi2 = 1.000000	5.2460470199585
p = 0.900000	sigma_epi2 = 1.000000	5.0504629611969
p = 0.990000	sigma_epi2 = 1.000000	5.05451583862305
p = 0.010000	sigma_epi2 = 2.000000	4.03324294090271
p = 0.100000	sigma_epi2 = 2.000000	4.21171092987061
p = 0.500000	sigma_epi2 = 2.000000	5.201092004776
p = 0.900000	sigma_epi2 = 2.000000	5.03955817222595
p = 0.990000	sigma_epi2 = 2.000000	5.07842087745667
time in total	time in total	1.21575103203456