

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: Probability an observation is sampled.

sigma_e2: Variance of Model error e.

sigma_epi2: variance of unpredicted error epsilon.

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 sampled. 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.

```
# 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(2,3,4)

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 epsilon.

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<-sigma_epi2[j]/(sum(xi**2))+((1-p[i])/p[i])*
      ((sum(xi**2*ei**2))/(sum(xi**2)**2)+sigma_epi2[j]/(sum(xi**2)))

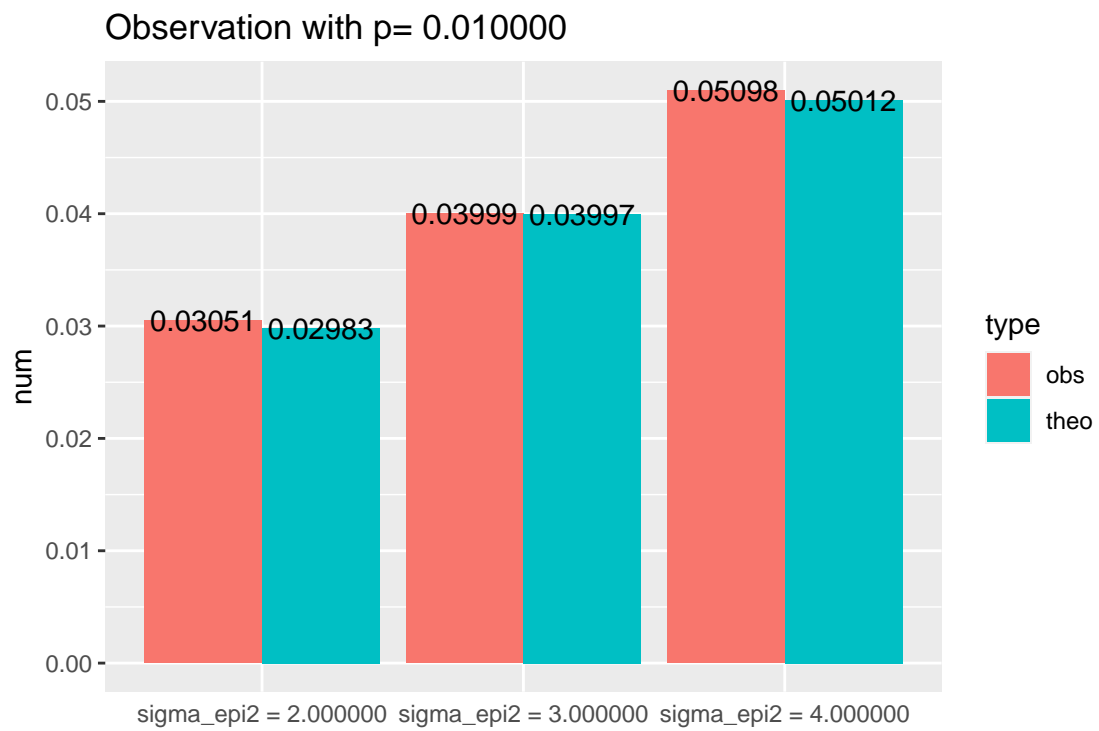
    end.time.small<-Sys.time()

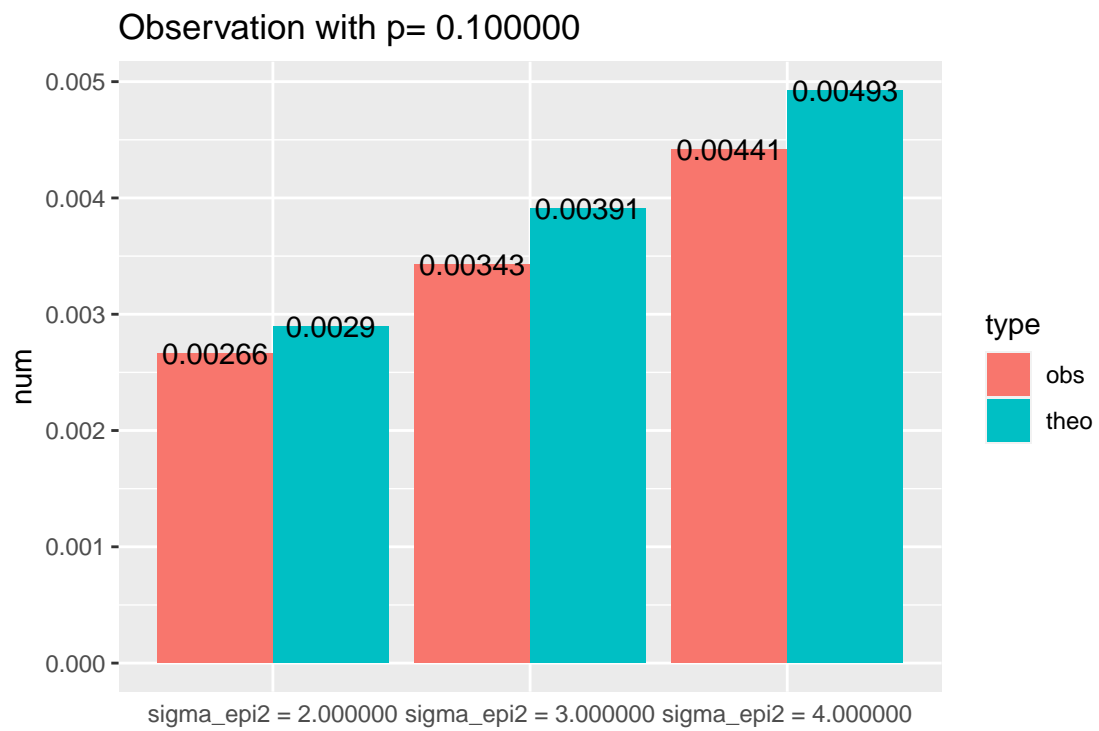
# rearranging the data for presentation and making graph
    out<-cbind(obs=var_beta,theo=theo_var,prob=sprintf('p = %f ',p[i]),
              sigma=sprintf('sigma_epi2 = %f',sigma_epi2[j]),
              time=end.time.small-start.time.small)
    result<-rbind(result,out)
  }
}

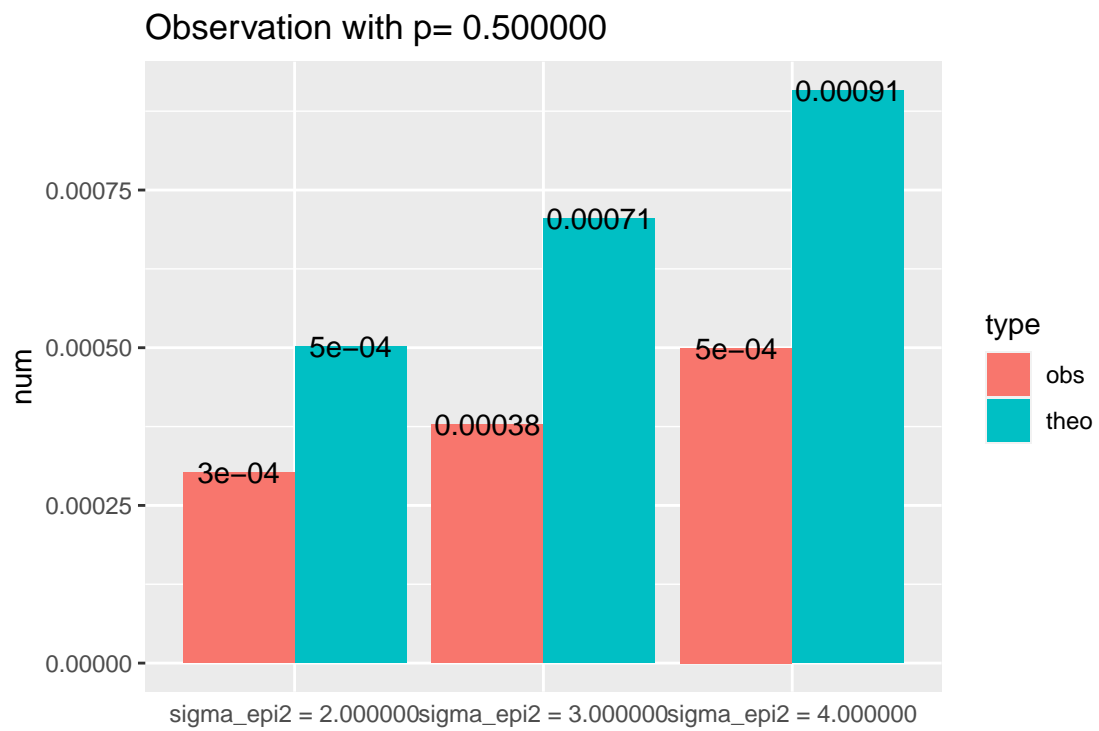
result1<-as.data.frame(result) %>% gather(type,num,obs:theo) %>% type_convert()

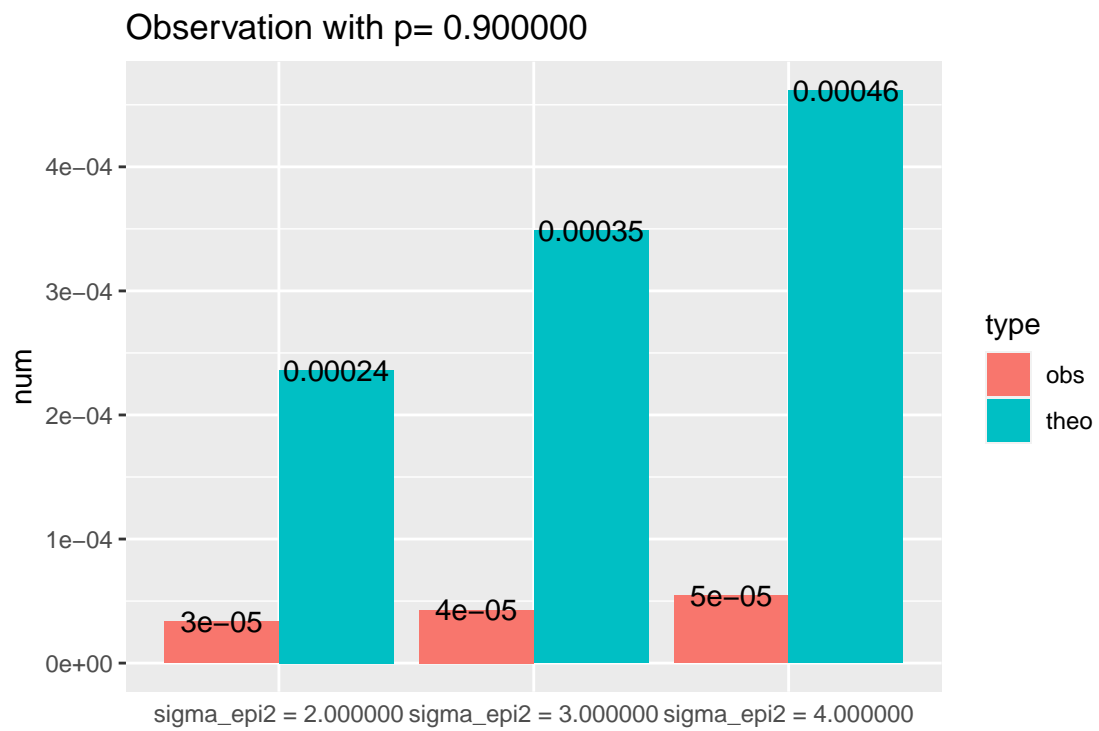
for (i in (c(1:length(p)))){
  a<-ggplot(result1 %>% filter(prob==sprintf('p = %f',p[i]))) +
    geom_col(aes(sigma,num,fill=type),position='dodge',width = 0.9)+
    geom_text(aes(sigma,num,label=round(as.numeric(num),5),group=type),
              position = position_dodge(width = 0.9))+
    theme(axis.title.x=element_blank(),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

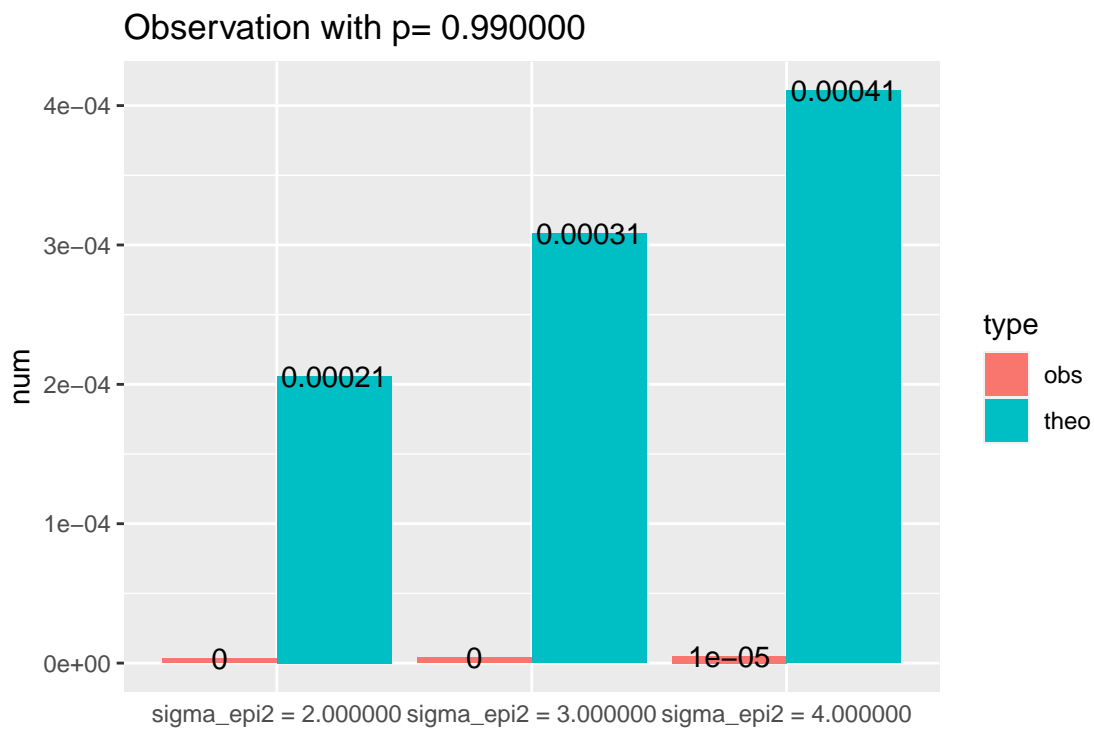
```



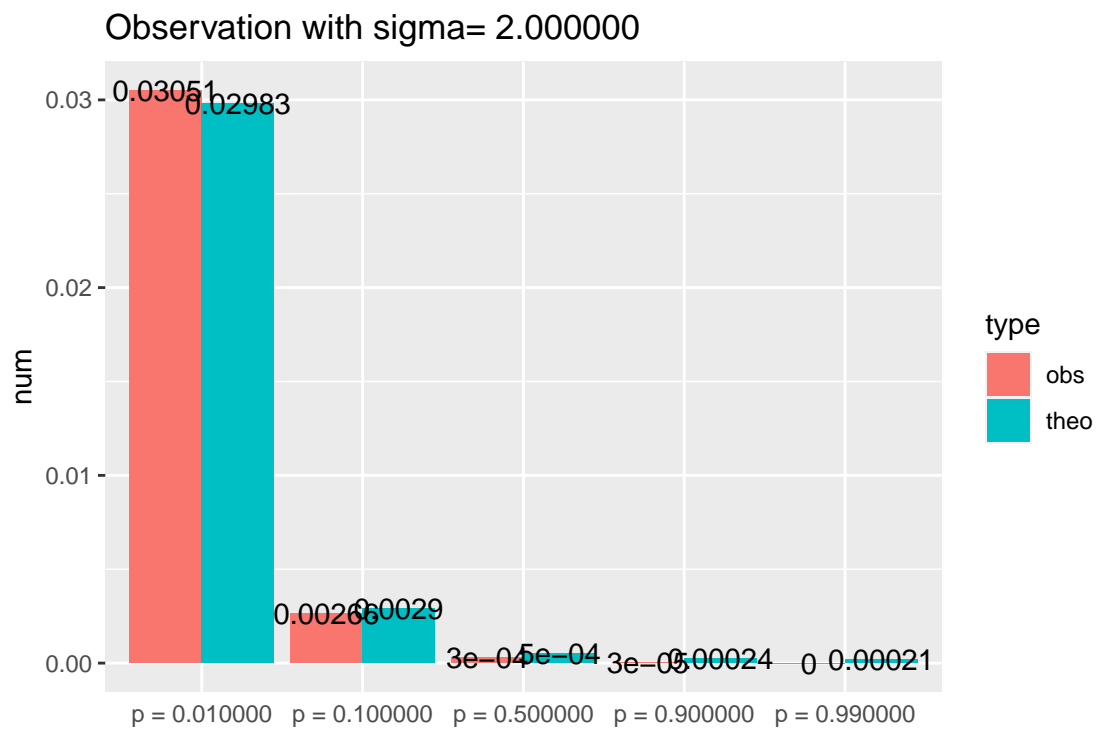


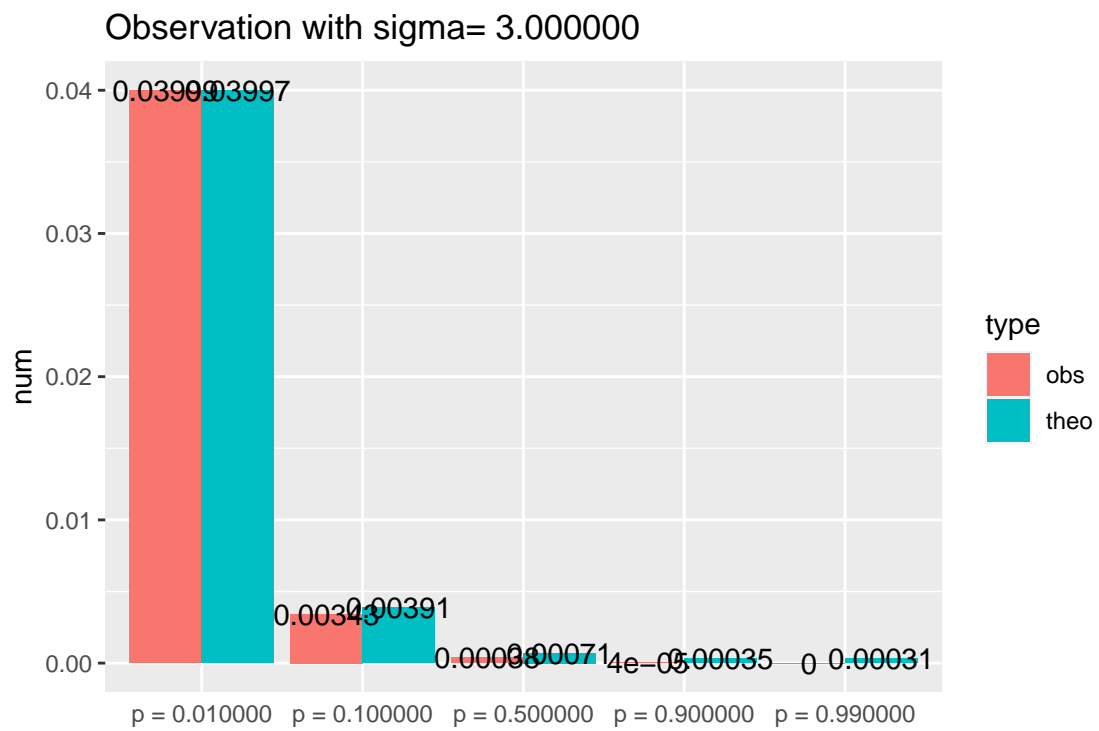


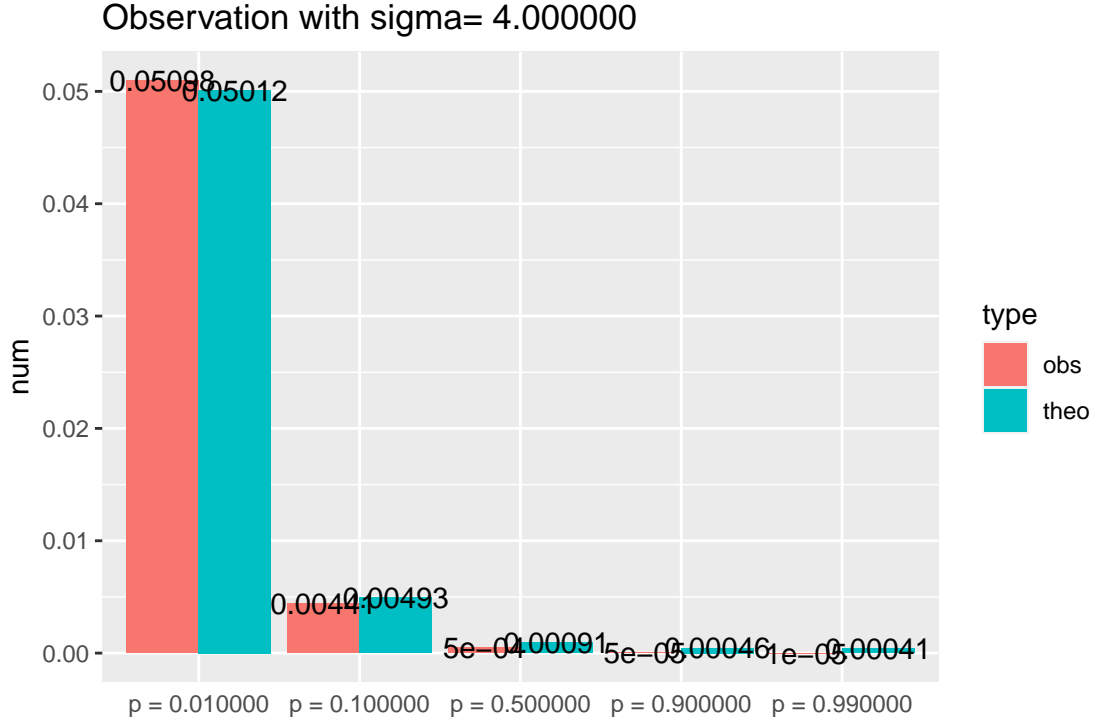




```
for (i in (c(1:length(sigma_epi2)))){
a<-ggplot(result1 %>% filter(sigma==sprintf('sigma_epi2 = %f',sigma_epi2[i]))) +
  geom_col(aes(prob,num,fill=type),position='dodge',width = 0.9) +
  geom_text(aes(prob,num,label=round(as.numeric(num),5),group=type),
    position = position_dodge(width = 0.9)) +
  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)
}
```







prob	sigma	obs	theo
p = 0.010000	sigma_epi2 = 2.000000	0.0305073662495741	0.0298262940607052
p = 0.100000	sigma_epi2 = 2.000000	0.00266047440781049	0.00289596233569465
p = 0.500000	sigma_epi2 = 2.000000	0.000302066077135416	0.000502155071249268
p = 0.900000	sigma_epi2 = 2.000000	3.35910493656514e-05	0.000236176486310892
p = 0.990000	sigma_epi2 = 2.000000	3.01970112516108e-06	0.000205951647113349
p = 0.010000	sigma_epi2 = 3.000000	0.0399934103645111	0.039972752220385
p = 0.100000	sigma_epi2 = 3.000000	0.00343041961027471	0.00391060815166263
p = 0.500000	sigma_epi2 = 3.000000	0.000377748439982266	0.000705084234442863
p = 0.900000	sigma_epi2 = 3.000000	4.28745516655709e-05	0.000348914910307334
p = 0.990000	sigma_epi2 = 3.000000	3.89051746607916e-06	0.000308441123473751
p = 0.010000	sigma_epi2 = 4.000000	0.0509808980236246	0.0501192103800647
p = 0.100000	sigma_epi2 = 4.000000	0.00441487085368125	0.0049252539676306
p = 0.500000	sigma_epi2 = 4.000000	0.000499555990829677	0.000908013397636458
p = 0.900000	sigma_epi2 = 4.000000	5.46486149051725e-05	0.000461653334303775
p = 0.990000	sigma_epi2 = 4.000000	5.07304011797874e-06	0.000410930599834152

prob	sigma	time
p = 0.010000	sigma_epi2 = 2.000000	4.47998809814453
p = 0.100000	sigma_epi2 = 2.000000	4.49601197242737
p = 0.500000	sigma_epi2 = 2.000000	5.50228095054626
p = 0.900000	sigma_epi2 = 2.000000	5.18513894081116
p = 0.990000	sigma_epi2 = 2.000000	5.54715013504028
p = 0.010000	sigma_epi2 = 3.000000	4.3114709854126
p = 0.100000	sigma_epi2 = 3.000000	4.59870386123657
p = 0.500000	sigma_epi2 = 3.000000	5.55118417739868
p = 0.900000	sigma_epi2 = 3.000000	5.66086196899414
p = 0.990000	sigma_epi2 = 3.000000	5.23697304725647
p = 0.010000	sigma_epi2 = 4.000000	4.25063395500183
p = 0.100000	sigma_epi2 = 4.000000	4.42321419715881
p = 0.500000	sigma_epi2 = 4.000000	5.49526500701904
p = 0.900000	sigma_epi2 = 4.000000	5.61498594284058
p = 0.990000	sigma_epi2 = 4.000000	5.00422596931458
time in total	time in total	1.28586364984512