```
#install.packages("cumstats")
setwd("G:\\My Drive\\Research\\Contextual Bandits\\code\\bandits\\variance_convergence_demo")

library(cumstats)
library(ggplot2)
```

Hypothesis

Three scenarios

1. Same mean, different variance

2. Same variance, different mean (small variance)

3. Same variance, different mean (large variance)

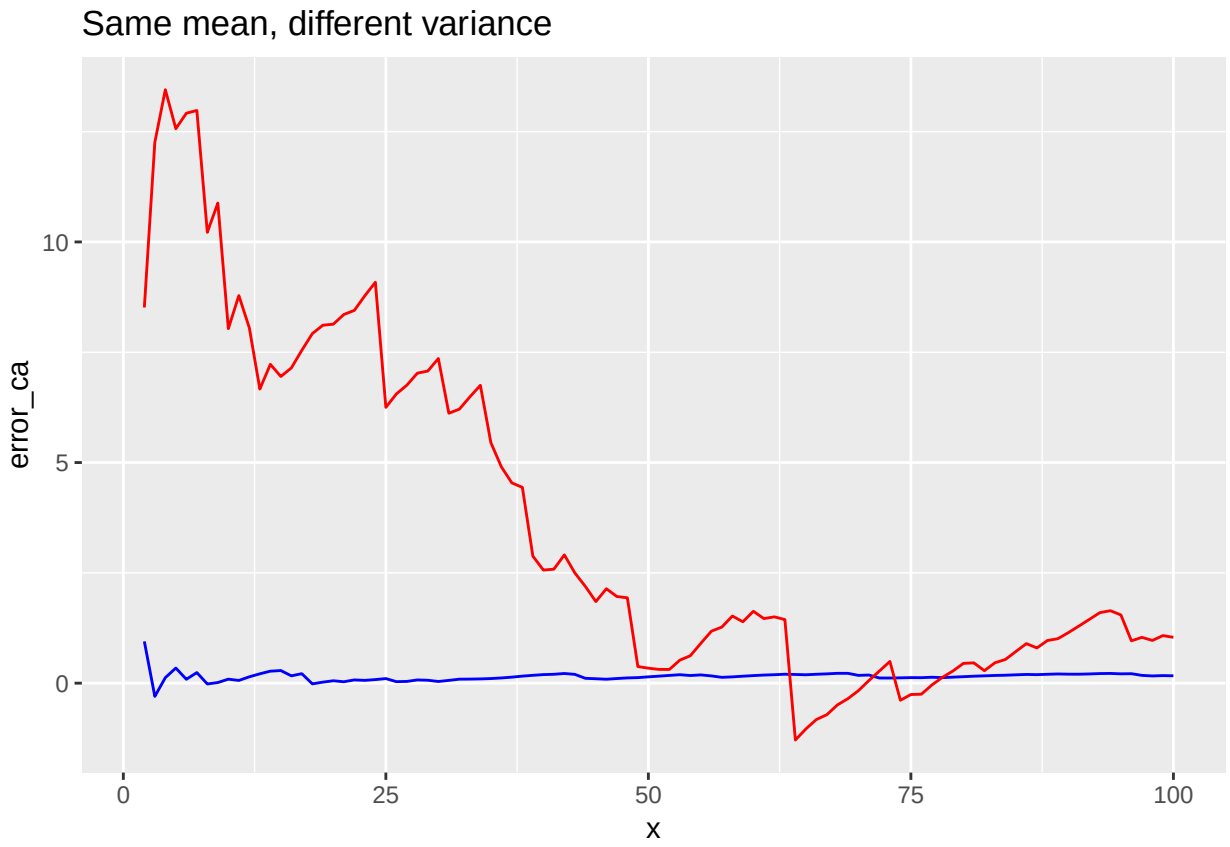
For scenario 1, we expect the error of variance estimate i.e (sigma - sugma_hat)

to converge faster if variance value is less. For scenarios 2,3 it should not matter

1. Same mean, different variance

```
x = c(1:100)
set.seed(123)
a = rnorm(100,0,1)
b = rnorm(100,0,4)
ca = cumvar(a)
cb = cumvar(b)
error_ca = 1-ca
error_cb = 16-cb
df = data.frame(x=x, error_ca = error_ca, error_cb=error_cb)
ggplot(df) + geom_line(aes(x,error_ca), color='blue') + geom_line(aes(x,error_cb), color='red') +
  ggtitle("Same mean, different variance") + ggsave("Same mean different variance.png")
```

```
## Saving 6.5 x 4.5 in image
```



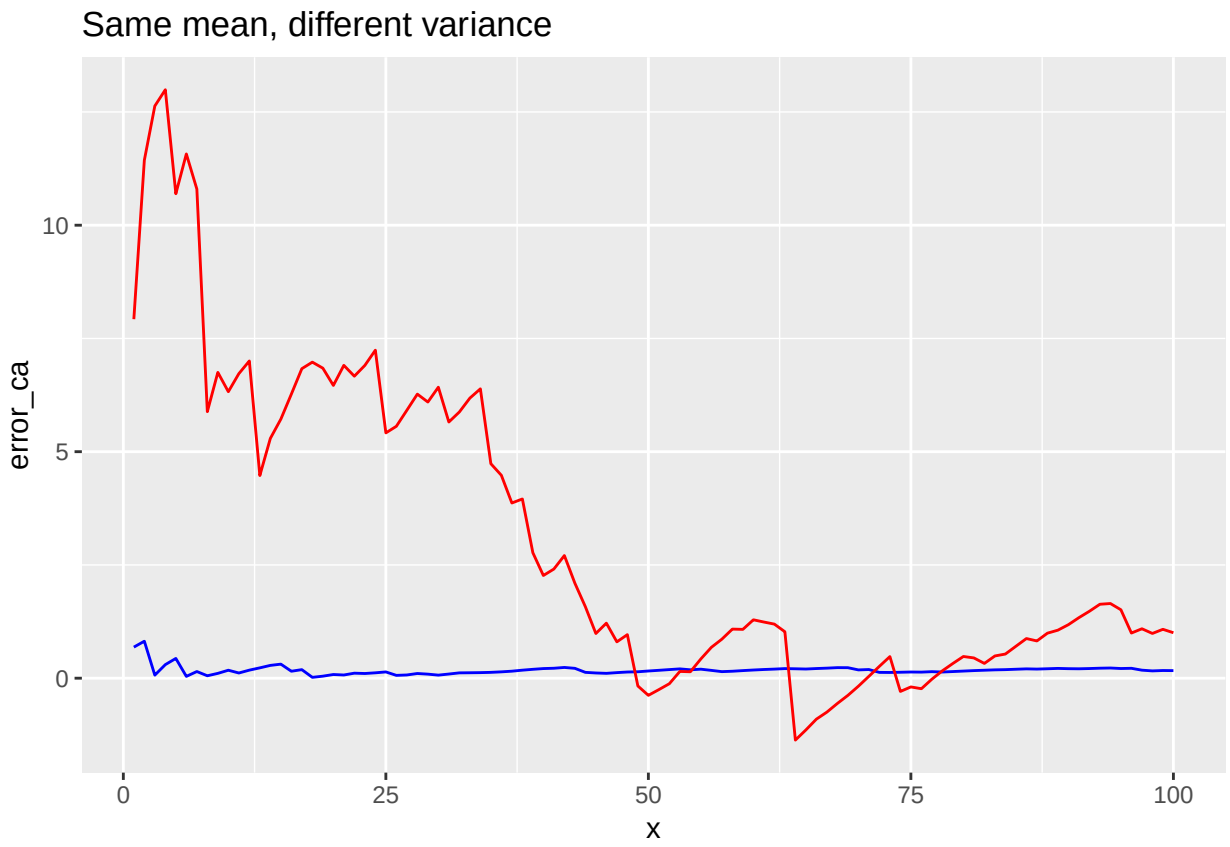
1a. Same mean, different variance

```
var_func <- function(vec, m){
  out = vector()
  for(i in 1:length(vec)){
    new_vec = vec[1:i]
    v = 0
    for(j in new_vec){
    v = v+(m-j)**2}
  out = c(out, v/length(new_vec))
    }
  return(out)
}


ca = var_func(a, 0)
cb = var_func(b, 0)


error_ca = 1-ca
error_cb = 16-cb
df = data.frame(x=x, error_ca = error_ca, error_cb=error_cb)
ggplot(df) + geom_line(aes(x,error_ca), color='blue') + geom_line(aes(x,error_cb), color='red') +
  ggtitle("Same mean, different variance") + ggsave("Same mean different variance1.png")
```

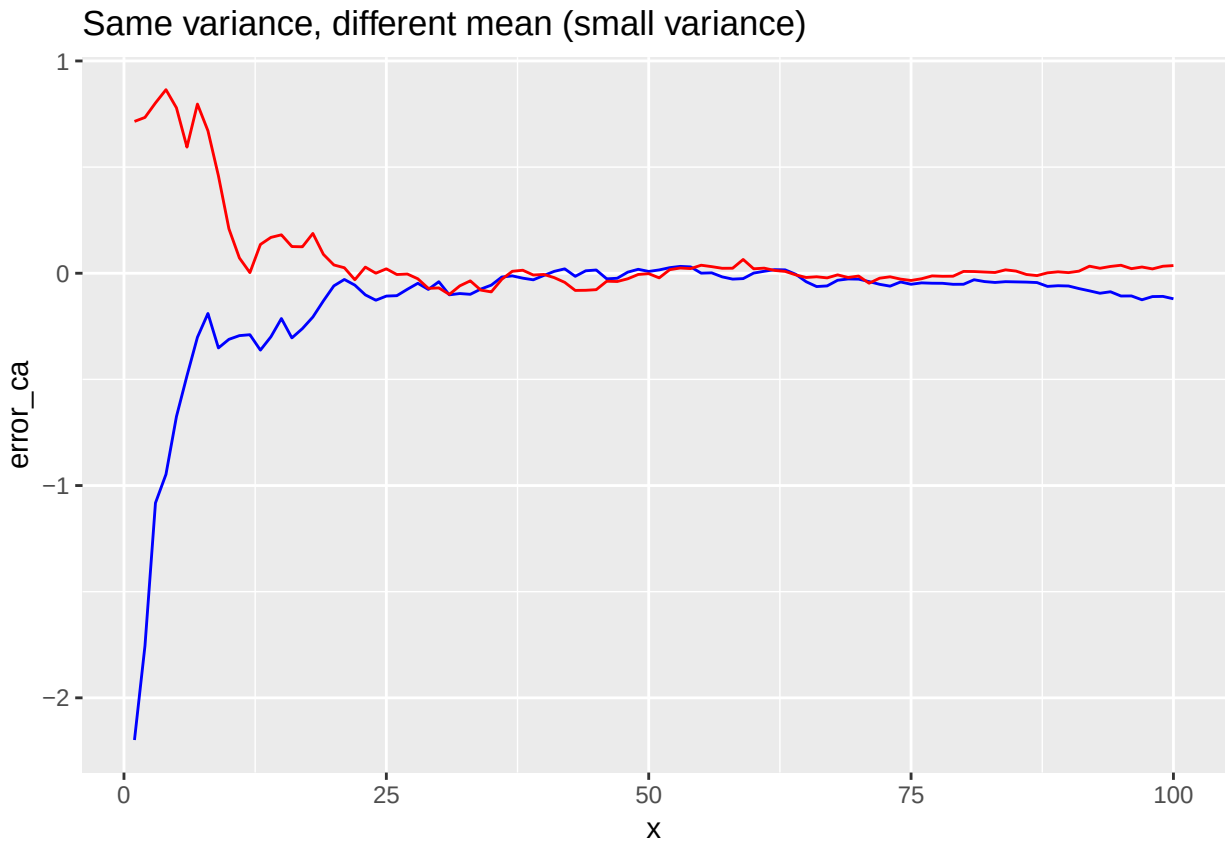
```
## Saving 6.5 x 4.5 in image
```



2. Same variance, different mean (small variance)

```
x = c(1:100)
a = rnorm(100,1,1)
b = rnorm(100,16,1)
ca = cummean(a)
cb = cummean(b)
error_ca = 1-ca
error_cb = 16-cb
df = data.frame(x=x, error_ca = error_ca, error_cb=error_cb)
ggplot(df) + geom_line(aes(x,error_ca), color='blue') + geom_line(aes(x,error_cb), color='red')+
  ggtitle("Same variance, different mean (small variance)") + ggsave("Same variance different mean (sma
```

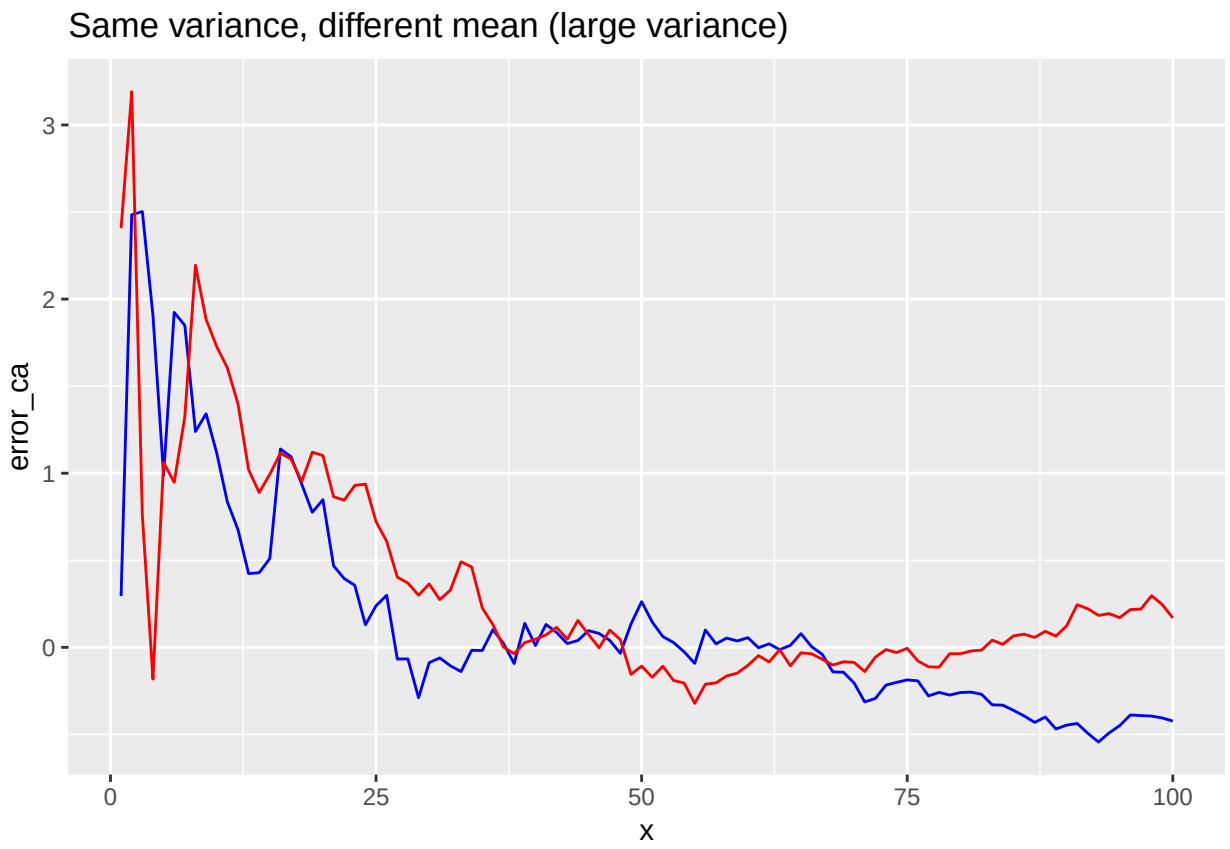
```
## Saving 6.5 x 4.5 in image
```



3. Same variance, different mean (large variance)

```
x = c(1:100)
a = rnorm(100,1,4)
b = rnorm(100,16,4)
ca = cummean(a)
cb = cummean(b)
error_ca = 1-ca
error_cb = 16-cb
df = data.frame(x=x, error_ca = error_ca, error_cb=error_cb)
ggplot(df) + geom_line(aes(x,error_ca), color='blue') + geom_line(aes(x,error_cb), color='red')+
  ggtitle("Same variance, different mean (large variance)") + ggsave("Same variance different mean (lar
```

```
## Saving 6.5 x 4.5 in image
```



All hypothesis are proven

In Graph 1 red and blue lines are clearly not following similar path even on different seeds.

Graph 2 and 3 are following similar path for different seeds