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

library(cumstats)
library(ggplot2)
```

Hypothesis

There are two outcome distributions. Control distribution is N(m_0, s_0^2), Treatment distribution is N(m_1, s_1^2). The treatment effect distribution should be N(m_1-m_0, $s\_0^{2+s}\_1 2$)

In order to estimate treatment effect distribution accurately, There are two ways 1. Would sampling from both distributions equally be able to estimate more accurately 2. Would sampling from both distributions unequally be able to estimate more accurately

Is there a difference?

```r
x = c(1:1000)
set.seed(123)
con = rnorm(100,0,1)
trt = rnorm(100,1,4)
func_equal <- function(len){
  out = vector()
  for(i in 1:len){
    if(sample(c(1,2), 1, replace = TRUE)==1){
      out = c(out, rnorm(1,0,1))
    } else {out = c(out, rnorm(1,1, 4))}
    }
  return(out)
}

a <- func_equal(1000)


func_unequal <- function(len){
  out = vector()
  for(i in 1:len){
    if(sample(c(1,2), 1, replace = TRUE, prob=c(0.8,0.2))==1){
      out = c(out, rnorm(1,0,1))
    } else {out = c(out, rnorm(1,1, 4))}
    }
  return(out)
}

b <- func_unequal(1000)


ca = cumvar(a)
cb = cumvar(b)
error_ca = 17-ca
error_cb = 17-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

Same mean, different variance