Optimization with two parameters in Gaussian distribution

Anderson Borba

25 October 2021

Example of Edge Detection with Stochastic Distances

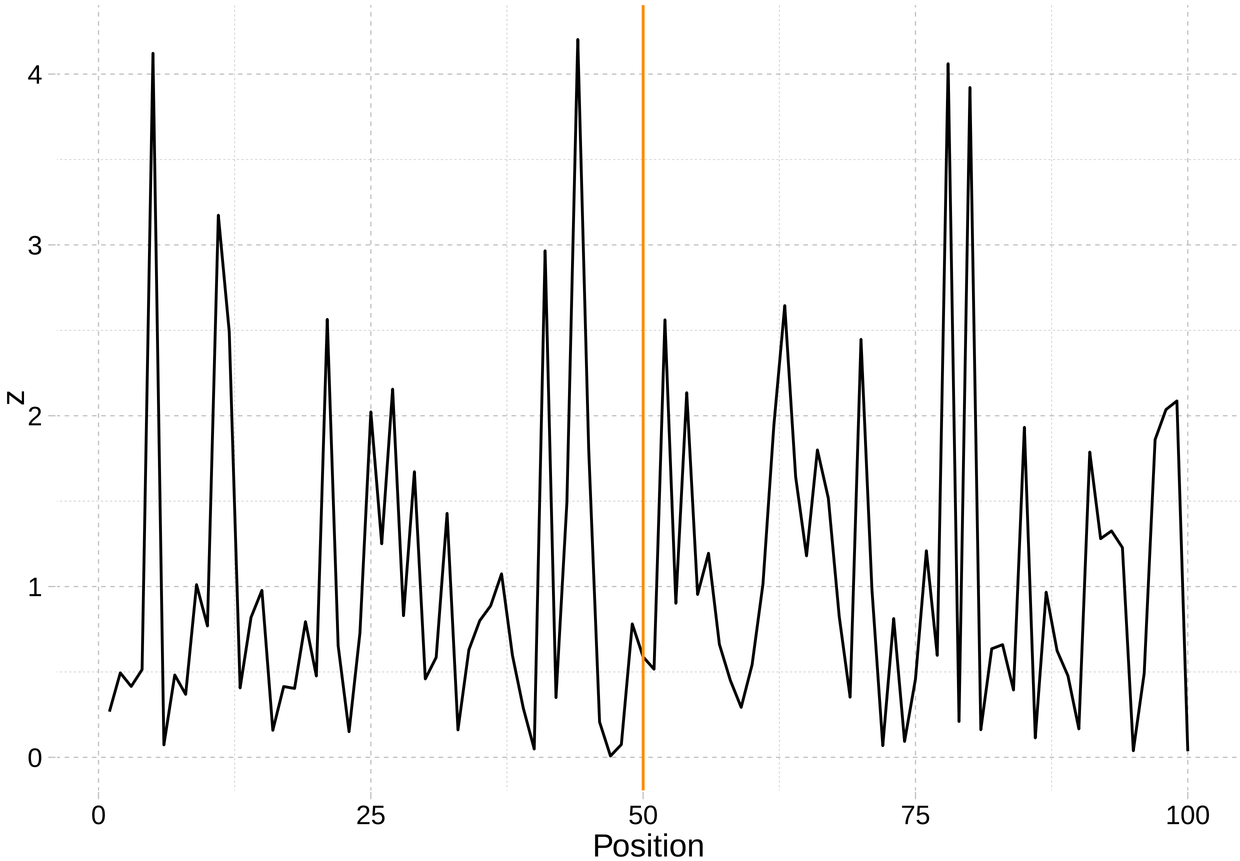
I will illustrate in the following the idea of detecting an edge using tests statistics based on stochastic distances. I will use exponentially-distributed data.

First, we simulate a strip of data of height $m = 1$ and length $n = 100$. The first half will have mean $\lambda_1 = 1$ and the second half $\lambda_2 = 10$ (these values can be altered).

```
lambda1 <- 1
lambda2 <- 1.2

z <- c(rexp(50, rate=1/lambda1), rexp(50, 1/lambda2))

ggplot(data.frame(z), aes(x=1:100, y=z)) +
  geom_line() +
  geom_vline(xintercept = 50, col="darkorange") +
  xlab("Position")
```



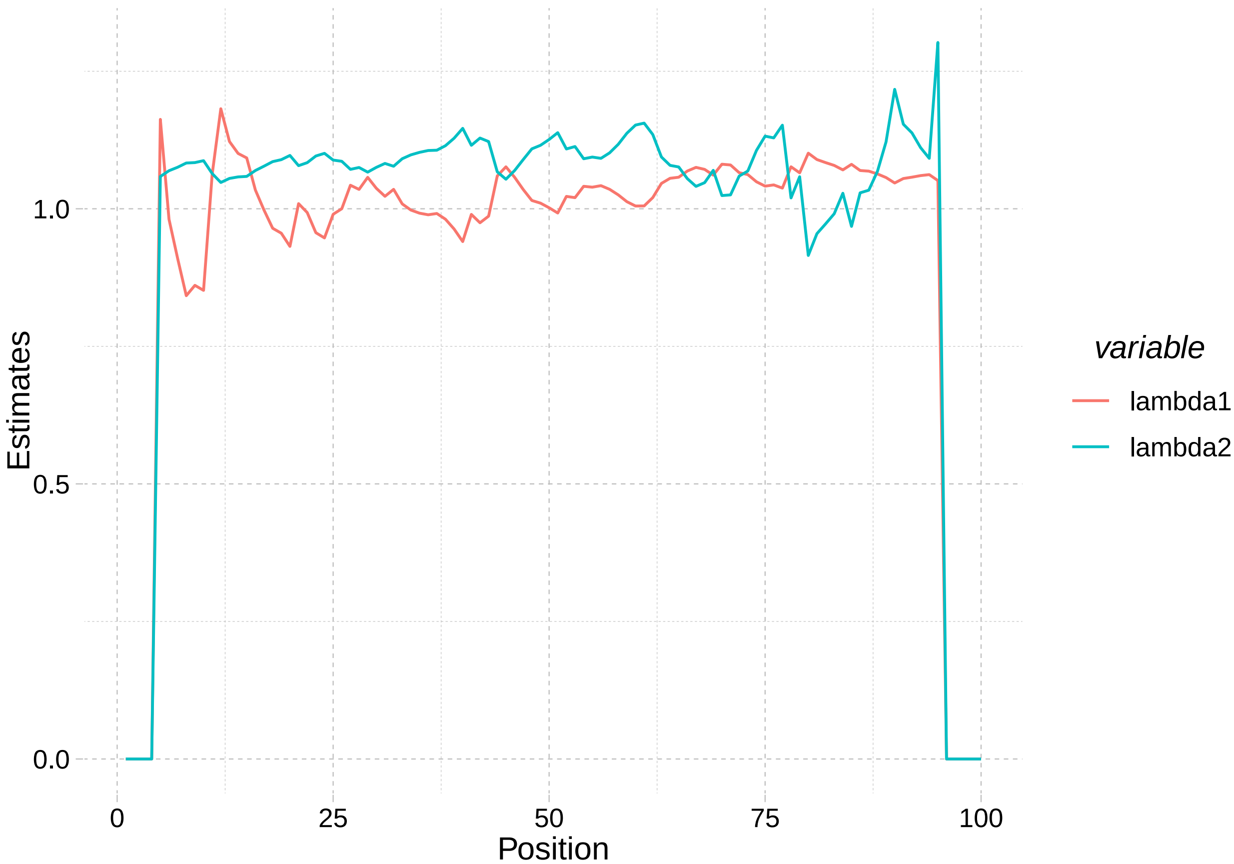
Next, we estimate the parameters λ_1 and λ_2 with $\widehat{\lambda}_1$ and $\widehat{\lambda}_2$, and store the results. We leave margins of five observations on each side.

```
Estim <- matrix(data=rep(0, 300), nrow=3, ncol=100)
Estim[1,] <- 1:100

for(el in 5:95) {
  Estim[2, el] <- mean(z[1:el])
  Estim[3, el] <- mean(z[(el+1):100])
}

Estim.df <- melt(data.frame(Position=Estim[1,], lambda1=Estim[2,], lambda2=Estim[3,]),
                 measure.vars = 2:3,
                 value.name = "Estimates")

ggplot(Estim.df, aes(x=Position, y=Estimates, group=variable, col=variable)) +
  geom_line()
```



Now we compute the distance at each point.

We assumed Exponential distributions, so the densities are:

$$f_{Z_{\mathrm{L}}}(z;\lambda_1) = \frac{1}{\lambda_1}\exp\{-x/\lambda_1\}$$

and

$$f_{Z_{\mathrm{R}}}(z;\lambda_2) = \frac{1}{\lambda_2}\exp\{-x/\lambda_2\},$$

both with positive support.

The Hellinger distance between these distributions is

$$d_{\mathrm{H}}(\lambda_1,\lambda_2) = 1 - \int_0^\infty \sqrt{f_{Z_{\mathrm{L}}}(z;\lambda_1) f_{Z_{\mathrm{L}}}(z;\lambda_2) dz} = 1 - \frac{1}{\sqrt{\lambda_1\lambda_2}}\int_0^\infty \sqrt{\exp\{-z/\lambda_1\}\exp\{-z/\lambda_2\}dz} = 1 - \frac{1}{\sqrt{\lambda_1\lambda_2}}\int_0^\infty \sqrt{\exp\Big\{-z\Big($$

We obtain the Hellinger distance as a member of the h-ϕ family of distances by setting $h(y) = y/2$ and $\phi(x) = (\sqrt{x}-1)^2$. With this we have that $h'(0) = 1/2$ and that $\phi''(1) = 1/2$. The test statistic is, thus,

$$S_{\mathrm{H}}(\widehat{\lambda}_1,\widehat{\lambda}_2) = \frac{8n_1n_2}{n_1+n_2} d_{\mathrm{H}}(\widehat{\lambda}_1,\widehat{\lambda}_2).$$

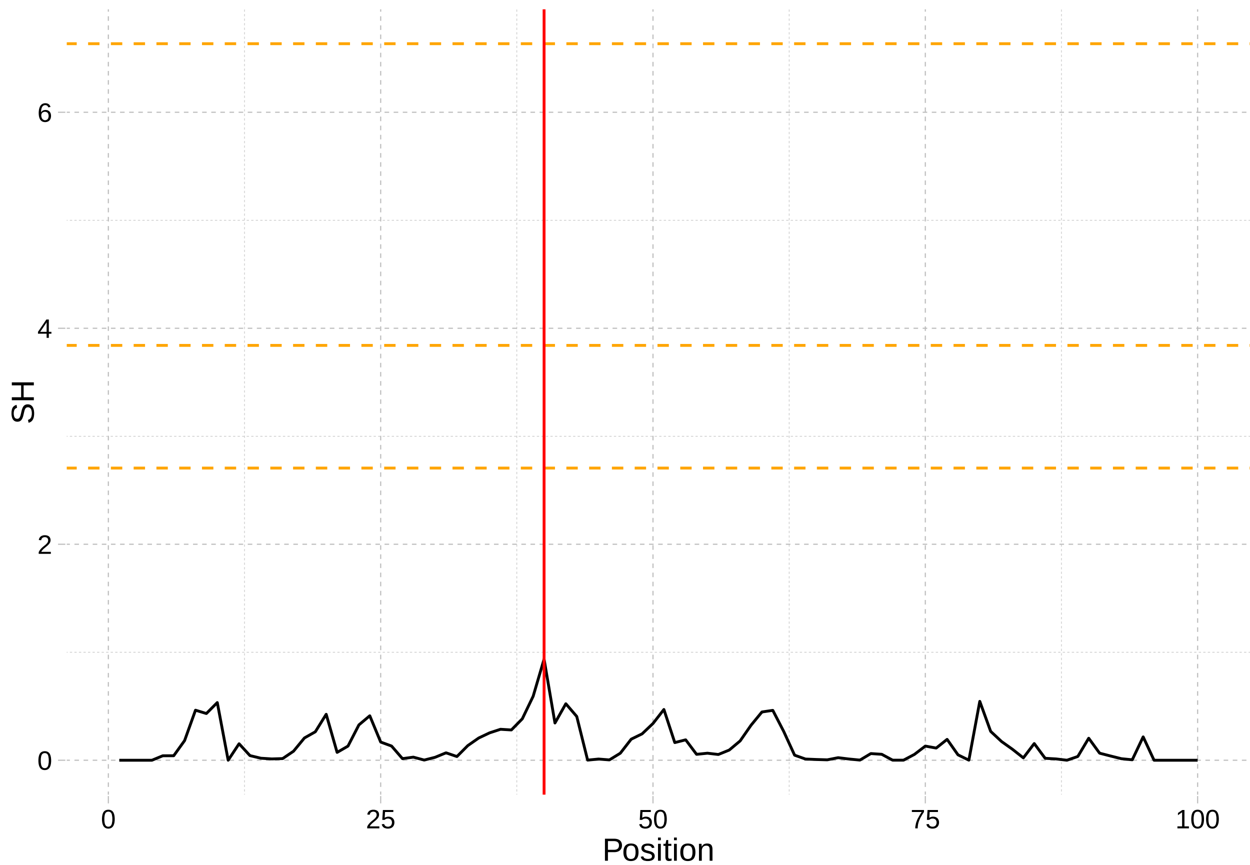
Now we compute and plot the test statistic at each point, along with the 90%, 95%, and 99% levels at which we reject the hypothesis that there is no edge (values below the orange lines do not provide enough evidence that there is an edge).

```
SH <- rep(0, 100)

for(el in 5:95) {
  lambda1 <- Estim[2,el]
  lambda2 <- Estim[3,el]
  SH[el] <- 0.08 * el * (100-el) * (1- 2*sqrt(lambda1*lambda2)/(lambda1+lambda2) )
}

edge <- which.max(SH)

ggplot(data.frame(SH), aes(x=1:100, y=SH)) +
  geom_line() +
  xlab("Position") +
  geom_hline(yintercept = qchisq(c(.9, .95, .99), df=1),
             col="orange", linetype=2) +
  geom_vline(xintercept = edge, col="red")
```



Not only we found evidence of edge at almost every point, but the technique found the edge at position 40.