

histoSplineExample.R

nychka

2022-12-05

```
#example of a log histo spline
suppressMessages(library( fields))
# gamma sample
set.seed( 223)
# random sample gamma distribution
Y<- rgamma( 20000, shape=2, rate=1.0)
# 50 bins in Histogram
N<-50
# histogram breakpoints
brk <- seq( min(Y), max(Y),length.out= (N+1) )
delta<- brk[2]- brk[1]

# in this example s are histogram bin midpoints
# and the response are the counts in each bin (OK if some are zero)
s<- (brk[1:N] + brk[2:(N+1)])/2
# counts in bins
y <- hist(Y, breaks = brk, plot = FALSE)$counts

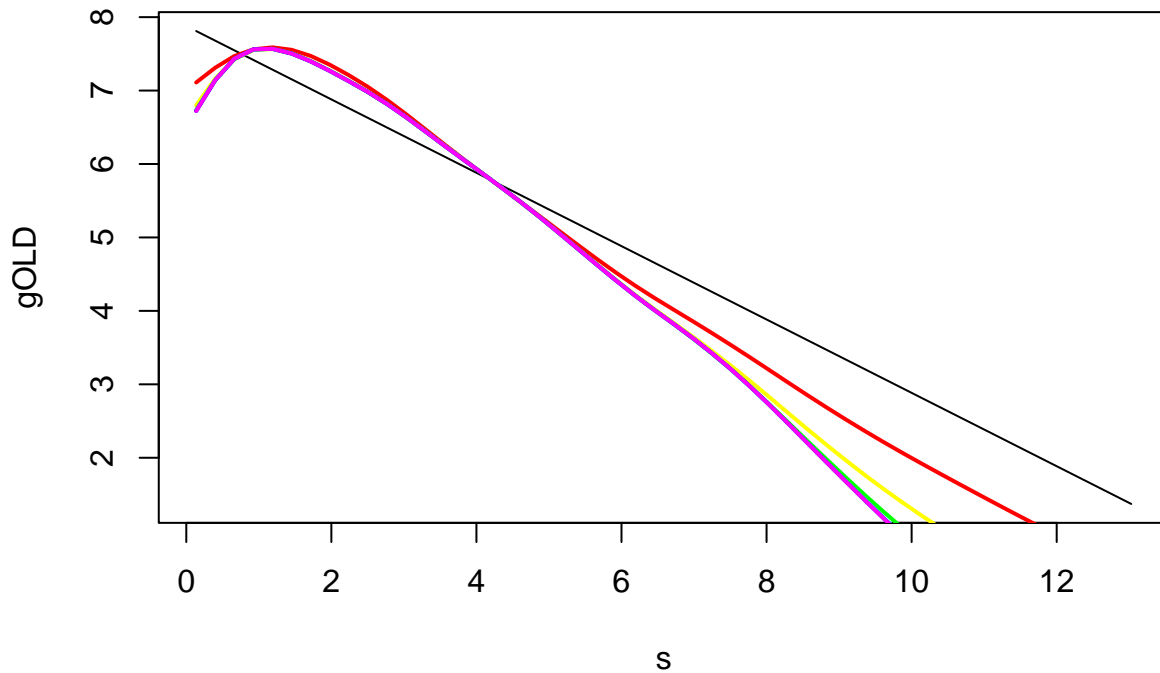
#####
## method will be valid for any case with locations s and Poisson counts, y
## note that the Tps method here will automatically handle fitting a 2D, 3D surface.
## This example fixes the degrees of freedom but in practice this will need to be chosen
## e.g. by CV or something else.
#####

# fit a GLM model to use for starting values
glmFit<- glm(y ~ s, family = poisson())
lambda<- .01
# starting value (this does not need to be a GLM estimate but need to be positive)
# poorer estimates may not give convergence.
gOLD<- predict( glmFit)
# plot iterates
plot( s, gOLD, type="l", col=1)
coltab<-rainbow(6)
for( I in 1:6){
  fHat <- exp(gOLD)
  z <- c(y - fHat)/fHat + gOLD
  weights<- c(fHat)
  TpsObj <- suppressWarnings(
    Tps( s,z, weights=weights,lambda=lambda, give.warnings=FALSE,
        df = 10)
  )
}
```

```

gNEW<- c(predict( TpsObj))
testTol <- sqrt(mean((gNEW - gOLD)^2)/mean(gOLD^2))
cat( I, testTol, fill=TRUE)
gOLD<- gNEW
# add new estimate
lines( s, gNEW, col=coltab[I], lwd=2)
}

```



```

## 1 0.1267753
## 2 0.09866334
## 3 0.06750762
## 4 0.02708224
## 5 0.003053868
## 6 3.223677e-05

```

approximate GCV at convergence

```
GCV<- mean( TpsObj$residuals^2)/ (1-TpsObj$eff.df/N )^2
```

plot the estimated density

```
fHat<- exp( gNEW)
```

normalize from counts to probability

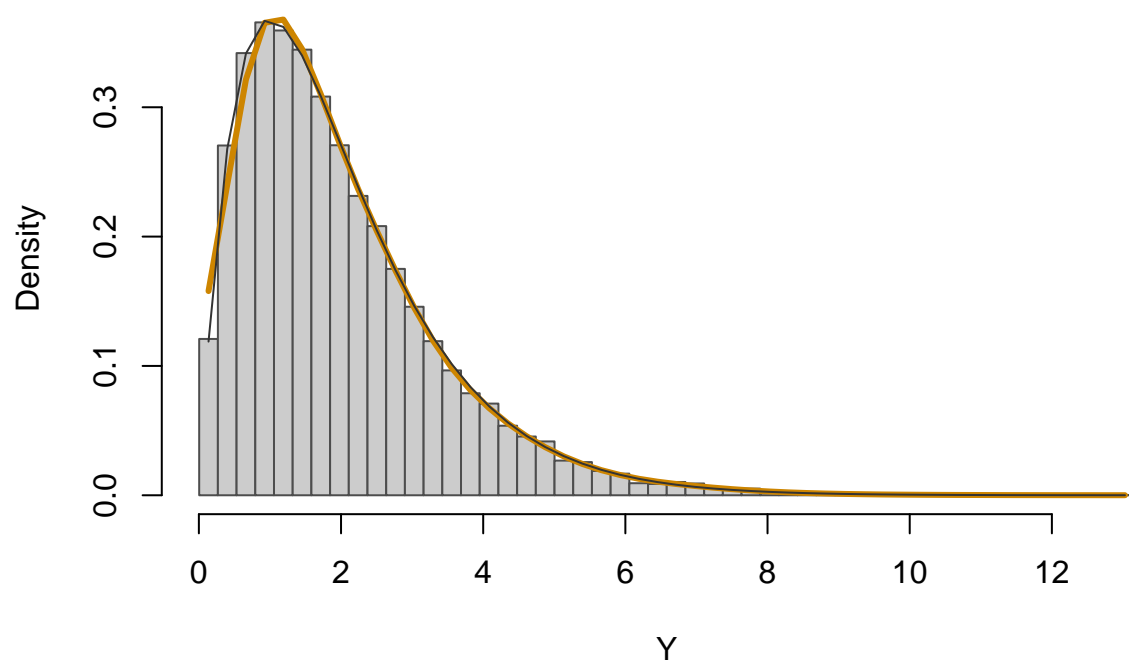
```
fHat<- fHat/ (sum(fHat)*delta)
```

```
hist( Y, breaks=brk, prob=TRUE, col="grey80", border="grey30")
```

```
lines( s, fHat, col="orange3", lwd=3)
```

```
lines( s, dgamma(s,shape=2, rate=1.0), col="grey20", lwd=1 )
```

Histogram of Y



```
# look at the tails  
plot( s, log( dgamma( s, shape=2.0)), type="l",  
      lwd=2,  
      ylab="log pdf")  
lines( s, log(fHat), col="orange3", lwd=2)
```

