

AjusteFinal

Daniel Ulate Montero

2022-11-17

Proyecto

Con base en las últimas recomendaciones, y para tener un medio de trabajo más sintetizado se crea un nuevo markdown.

```
base <- read.csv('~/.GitHub/II-2022/Distribución de pérdidas/Fire Incidents Data.csv')

perd <- base$Estimated_Dollar_Loss
#Los elementos de la base donde no se cuantificaron perdidas no interesan
base <- base[!is.na(perd),]
perd <- perd[!is.na(perd)]

base$logEstimado <- log(base$Estimated_Dollar_Loss)
base$logEstimado <- case_when(base$Estimated_Dollar_Loss == 0 ~ 0,
                              TRUE ~ log(base$Estimated_Dollar_Loss))

base$TFS_Alarm_Time <- as.POSIXct(base$TFS_Alarm_Time, format = "%Y-%m-%dT%H:%M:%S")
base$TFS_Arrival_Time <- as.POSIXct(base$TFS_Arrival_Time, format = "%Y-%m-%dT%H:%M:%S")

base$tiempoAtencion <- as.numeric(difftime(base$TFS_Arrival_Time, base$TFS_Alarm_Time))
```

Estamos desarrollando un modelaje POT (Peak Over Threshold), entonces requerimos encontrar el umbral más adecuado para los datos.

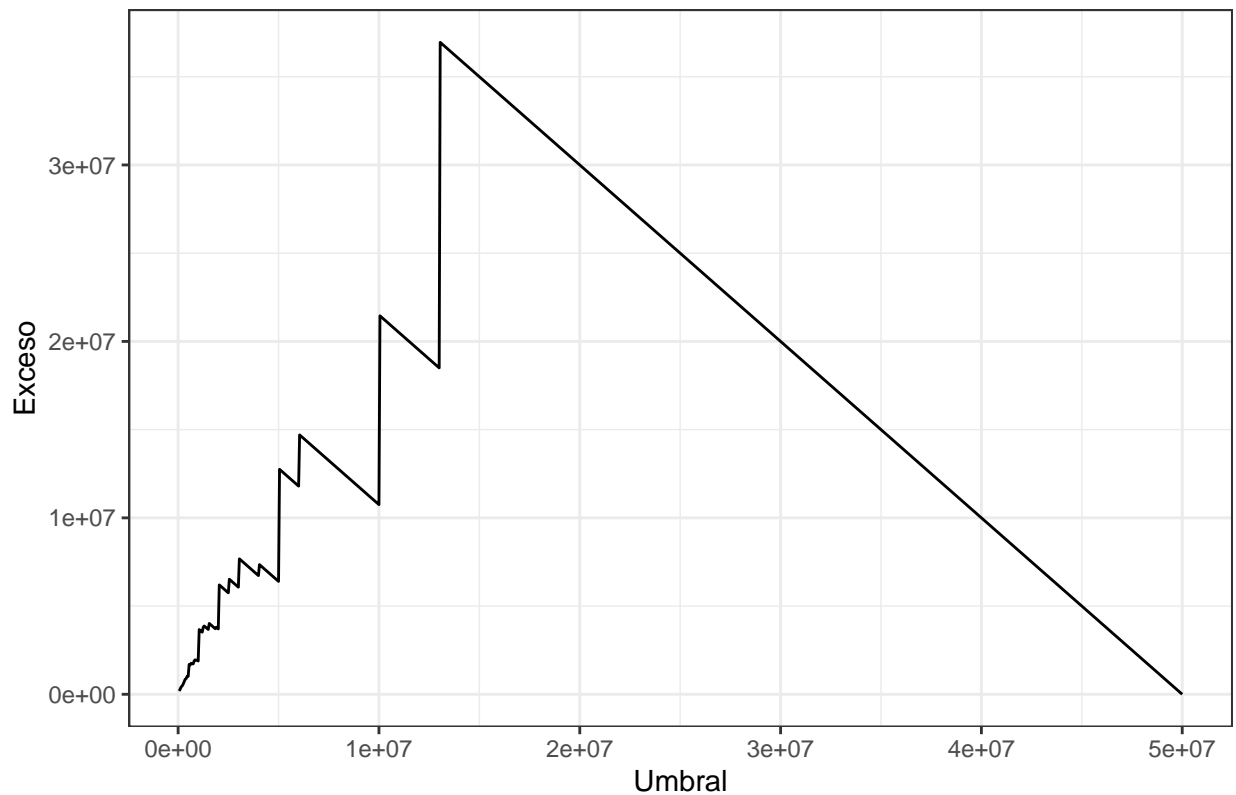
```
perd <- base$Estimated_Dollar_Loss
n <- 1000 #La cantidad de puntos
d1 <- quantile(perd, 0.9)
d<-seq(d1,max(perd),length.out=n)

e <- function(d, X = perd){ # Funcion media de exceso para cualquier vector
  index <- lapply(d, function(e){X[X>=e]-e} )
  sapply(index, mean)
}

p <- tibble(Umbral = d, Exceso = e(d))

ggplot(p, aes(x = Umbral, y = Exceso)) + geom_line() + theme_bw() +
  labs(title = "Función media de exceso")
```

Función media de exceso



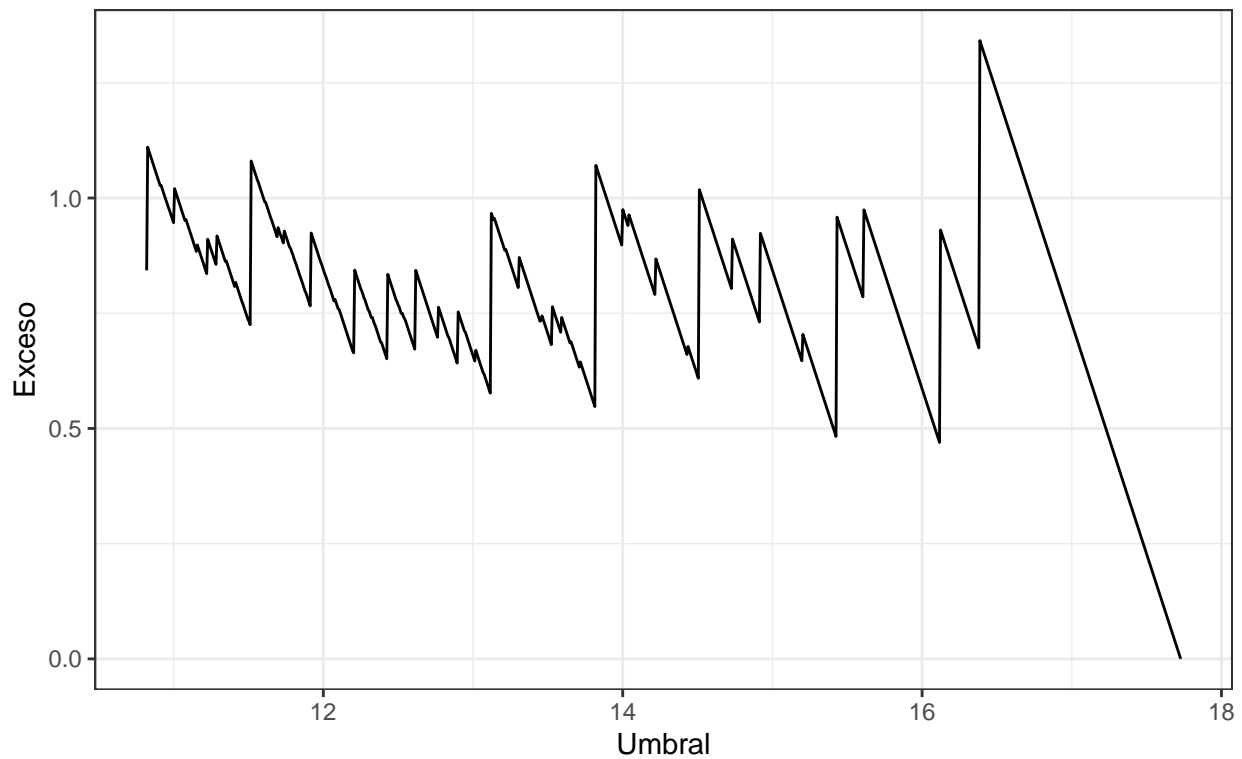
```
regular = perd[perd>d1]-d1
```

Un problema evidente es que el gráfico muestra un comportamiento distinto al aplicar la escala logarítmica de las pérdidas.

```
perd <- base$logEstimado
d1 <- quantile(perd, 0.9)
d<-seq(d1,max(perd),length.out=n)
p <- tibble(Umbral = d, Exceso = e(d))
# d1 <- nrow(p) - which.max(diff(rev(p$Exceso))<0)
# d1 <- p$Umbral[d1]

ggplot(p, aes(x = Umbral, y = Exceso)) + geom_line() + theme_bw() +
  labs(title = "Función media de exceso log-pérdida",
       caption = "Fuente: Elaboración propia")
```

Función media de exceso log-pérdida



Fuente: Elaboración propia

```
logaritmica <- perd[perd>d1]-d1
```

Ajustemos el logaritmico con cada opción a ver qué pez

```
df <- data.frame(matrix(NA, 2,0))
rownames(df) <- c('neg-verosim', 'AIC')
mod.gp <- fit.gpd(logaritmica)
df$gp <-c(mod.gp$nlh, AIC(mod.gp))

#mod.gp$param

mod.exp <- fitdist(logaritmica, 'exp')
df$exp <- c(-mod.exp$loglik, mod.exp$aic)

#mod.exp$estimate

mod.wei <- fitdist(logaritmica, 'weibull')
df$wei <- c(-mod.wei$loglik, mod.wei$aic)

df
```

```
##           gp      exp      wei
## neg-verosim 1607.9 1646.164 1512.183
## AIC         3219.8 3294.328 3028.366
```

El regular da lo siguiente:

```
df <- df[, -(1:3)]
mod.gp <- fit.gpd(regular)

df$gp <- c(mod.gp$nlh, AIC(mod.gp))
#mod.gp$param
dexp1 <- function(x, rate){
  rate * exp(-rate * x) * (x>=0)
}

#library(extRemes)
#mod.exp <- fevd(x = regular, data = data.frame(regular), type = 'Exponential')
# -mod.exp$loglik
# mod.exp$aic
#mod.exp$estimate

mod.wei <- fitdist(regular, 'weibull')

## Warning in sqrt(diag(varcovar)): Se han producido NaNs

## Warning in sqrt(1/diag(V)): Se han producido NaNs

## Warning in cov2cor(varcovar): diag(.) had 0 or NA entries; non-finite result is
## doubtful

df$wei <- c(-mod.wei$loglik, mod.wei$aic)
df

##              gp      wei
## neg-verosim 19250.58 19463.84
## AIC         38505.16 38931.68
```

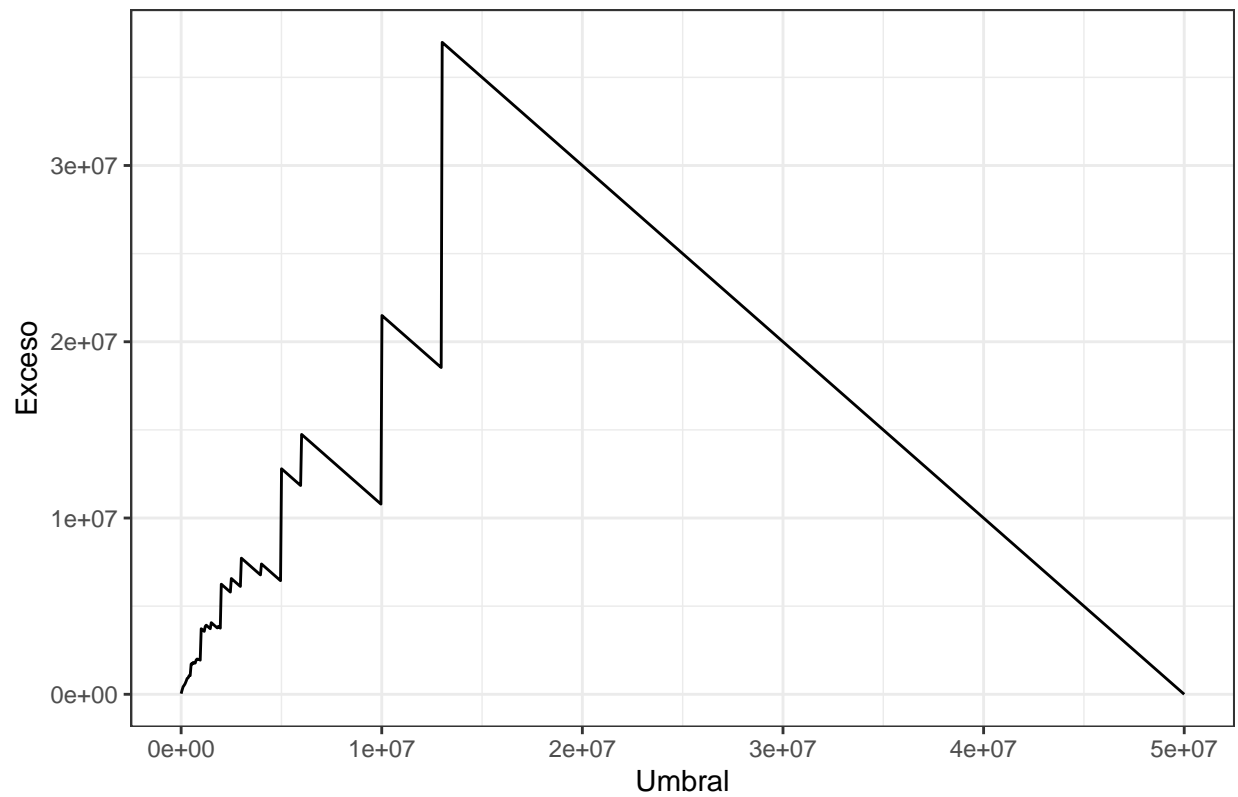
Los graficos completos por si acaso:

```
perd <- base$Estimated_Dollar_Loss
d1 <- min(perd)
d<-seq(d1,max(perd),length.out=n)

p <- tibble(Umbral = d, Exceso = e(d))

ggplot(p, aes(x = Umbral, y = Exceso)) + geom_line() + theme_bw() +
  labs(title = "Función media de exceso")
```

Función media de exceso



```
perd <- base$logEstimado
d1 <- min(perd)
d<-seq(d1,max(perd),length.out=n)

p <- tibble(Umbral = d, Exceso = e(d))

ggplot(p, aes(x = Umbral, y = Exceso)) + geom_line() + theme_bw() +
  labs(title = "Función media de exceso")
```

Función media de exceso

