2.5 Exercises

1. **The film90 example:** Familiarize yourself with the **gamlss** functions and packages by repeating the commands given in this chapter analyzing the **film90** data.

2. **The abdom data:** Information on the abdominal data is given on page 90. Fit different response distributions and choose the ‘best’ model according to the GAIC criterion:
   
   (a) Load the **abdom** data, print the variable names and plot the data.
   
   (b) Fit the normal distribution model, using `pb()` to fit P-spline smoothers for the predictors for \( \mu \) and \( \sigma \) with automatic selection of smoothing parameters:
   
   ```r
   mNO <- gamlss(y~pb(x), sigma.fo=~pb(x), data=abdom, family=NO)
   ```
   
   (c) Try fitting alternative distributions:
      
      i. two-parameter distributions: GA, IG, GU, RG, LO,
      
      ii. three-parameter distributions: PE, TF, BCCG,
      
      iii. four-parameter distributions: BCT, BCPE.
      
      Apply `pb()` to all parameters of each distribution. Make sure to use different model names.
   
   (d) Compare the fitted models using GAIC with each of the penalties \( k=2 \), \( k=3 \) and \( k=\log(\text{length(abdom$y)}) \), e.g.
   
   ```r
   GAIC(mNO,mGA,mIG,mGU,mRG,mLO,mPE,mTF,mBCCG,mBCT,mBCPE,k=2)
   ```
   
   (e) Check the residuals for your chosen model, say \( m \), by `plot(m)` and `wp(m)`.
   
   (f) For a chosen model, say \( m \), look at the total effective degrees of freedom used in the fitted model `edfAll(m)`, plot the fitted parameters, `fittedPlot(m,x=abdom$x)`, and plot the data by `plot(y~x,data=abdom)`, and the fitted \( \mu \) against \( x \) by `lines(fitted(m)~x, data=abdom)`.
   
   (g) For a chosen model, examine the centile curves using `centiles(m,abdom$x)`.

3. **The fabric data:** The data are 32 observations on faults in rolls of fabric.

   | R data file: fabric in package **gamlss.data** of dimensions 32 × 3 variables |
   | leng : the length of the fabric roll |
y : the number of faults in the roll of the fabric
x  the log of the length of the roll

**purpose:** to fit count data distributions and choose the `best` model (according to GAIC criterion).

(a) Load the **fabric** data, print the variable names and plot the data.

(b) Fit the Poisson distribution model using first a linear term in x, and then a smooth function in x, i.e. pb(x):

```r
mPO1<-gamlss(y~x, data=fabric, family=PO)
mPO2<-gamlss(y~pb(x), data=fabric, family=PO)
```

(c) Compare the models using the AIC.

(d) Check the residuals of your chosen model.

(e) Try fitting alternative distributions instead of the Poisson,

   i. two-parameter distributions: NBI, NBII, PIG, ZIP, ZAP,

   ii. three-parameter distributions: SICHEL, ZINBI, ZANBI.

   Apply `pb()` to all parameters of each distribution. Make sure to use different model names.

(f) Compare the fitted models using GAIC with each of the penalties \( k=2, k=3 \) and \( k = \log(\text{length}(\text{fabric}$y$)) \).

(g) For a chosen model look at the total effective degrees of freedom, plot the fitted parameters and plot the data and fitted \( \mu \) against x.

(h) Check the model using diagnostic plots.