Fitting distributions to dietary exposure data

Workshop on Dietary Assessment and Measurement Error

Brecht.Devleesschauwer@Sciensano.be

Department of Epidemiology and Public Health, Sciensano, Belgium

October 09, 2018

https://github.com/brechtdv/fitdist

Agenda

- Motivating example
- Individual observations
- Quantiles

Motivating Example

Motivating Example

What is the contribution of current red meat consumption levels to the colorectal cancer disease burden?

Motivating Example

What is the contribution of current red meat consumption levels to the colorectal cancer disease burden?

Comparative Risk Assessment

Systematic evaluation of the changes in burden of disease which would result from modifying the population distribution of exposure to a **theoretical minimum risk exposure distribution (TMRED)** that would imply minimum health loss, keeping all other risk factors unchanged

Calculation of **Population Attributable Fraction**

$$PAF = \frac{\int P(x)RR(x)dx - \int P'(x)RR(x)dx}{\int P(x)RR(x)dx}$$

Motivating Example

What is the contribution of current red meat consumption levels to the colorectal cancer disease burden?

Comparative Risk Assessment

Systematic evaluation of the changes in burden of disease which would result from modifying the population distribution of exposure to a **theoretical minimum risk exposure distribution (TMRED)** that would imply minimum health loss, keeping all other risk factors unchanged

Calculation of **Attributable Burden**

$$AB = B * PAF$$

Motivating Example

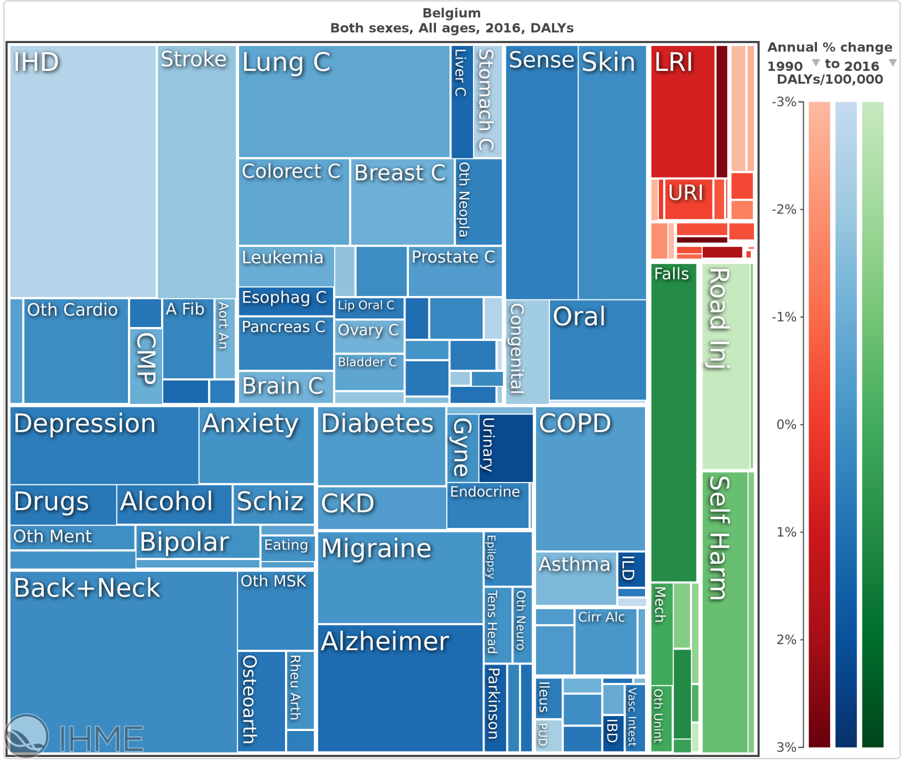
What is the contribution of current red meat consumption levels to the colorectal cancer disease burden?

Data requirements

- Current disease burden of colorectal cancer
- Relative risk function
- Current red meat consumption levels
- Ideal consumption level (TMREL)

Motivating Example

Colorectal cancer disease burden



Institute for Health Metrics and Evaluation

Global Burden of Disease, 2016

- 8,863 new cases
- 3,725 deaths
- 57,283 DALYs

Motivating Example

Relative risk function

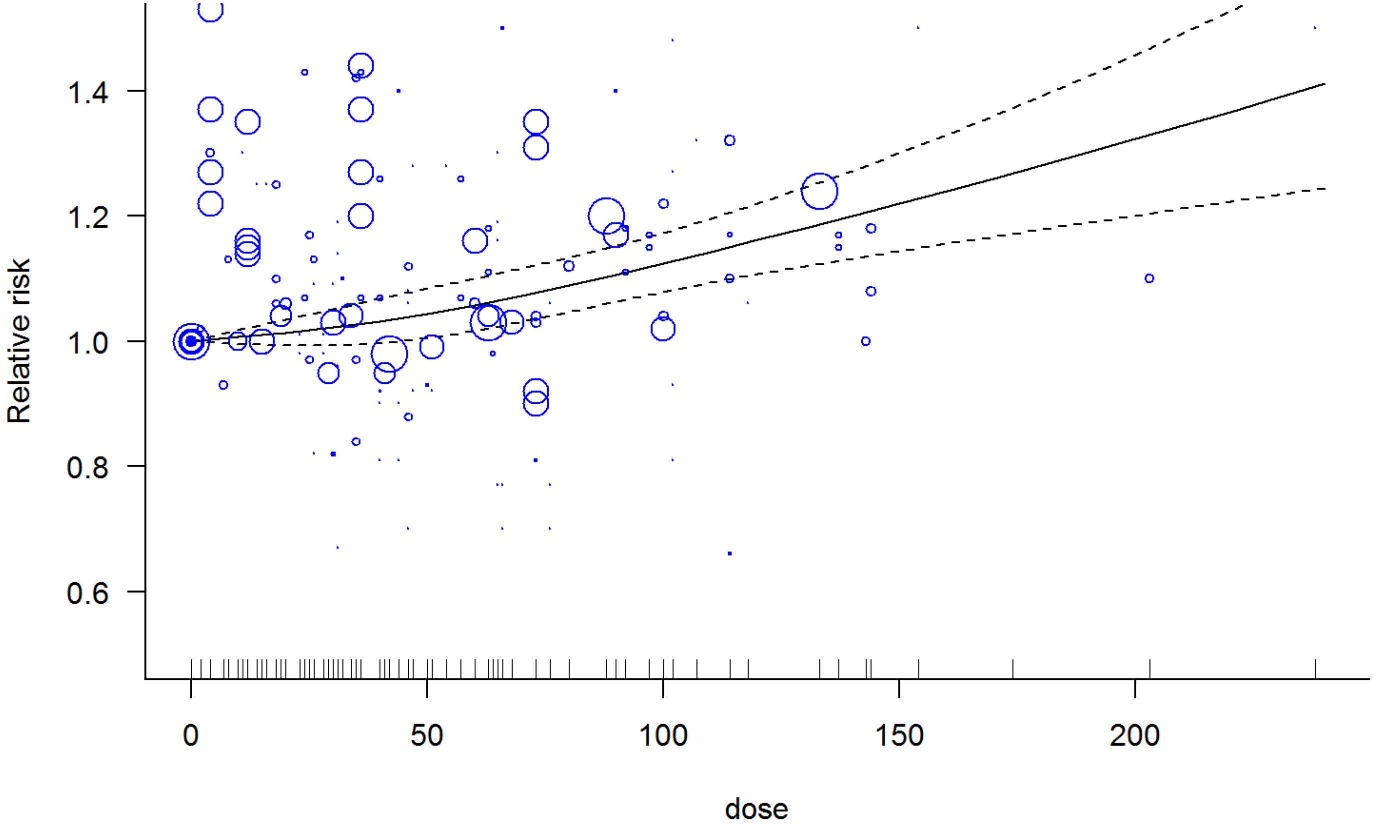
Food groups and risk of colorectal cancer. Schwingshackl L, Schwedhelm C, Hoffmann G, Knüppel S, Laure Preterre A, Iqbal K, Bechthold A, De Henauw S, Michels N, Devleesschauwer B, Boeing H, Schlesinger S. Int J Cancer. 2018 May 1;142(9):1748-1758. doi: 10.1002/ijc.31198

Non-linear dose-response function

dosresmeta package

Motivating Example

Relative risk function



Motivating Example

Red meat consumption levels

FFQ and 24h recalls

- Statistical Program to Assess Dietary Exposure (SPADE)

Individual observations

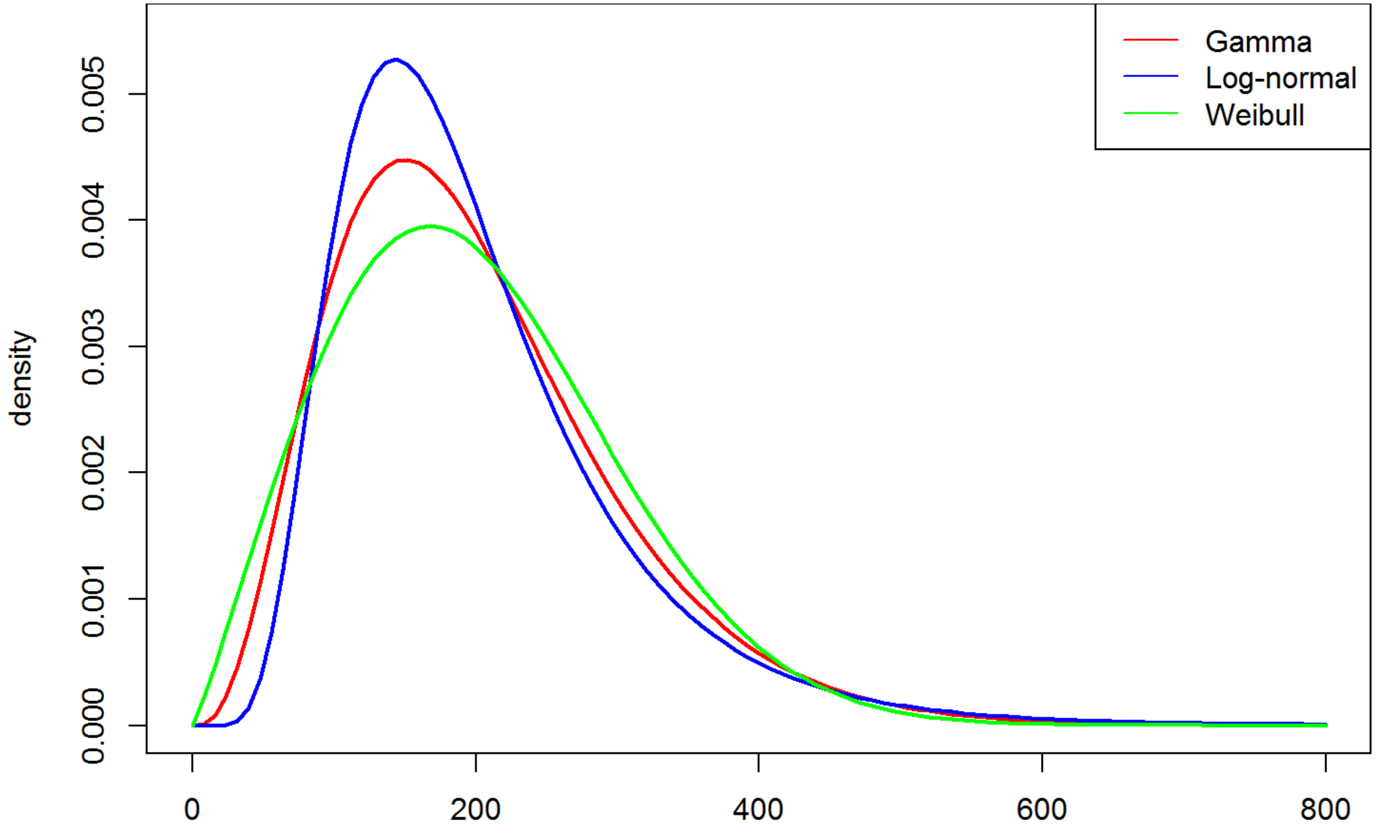
- Bootstrap
- Method of moments
- Maximum likelihood

Quantiles

- Optimization

Which distribution?

Main characteristic: non-negative, continuous



Which distribution?

Excessive zeros

Mixture of Bernoulli process (0/1) and exposure distribution

Zero-inflated models

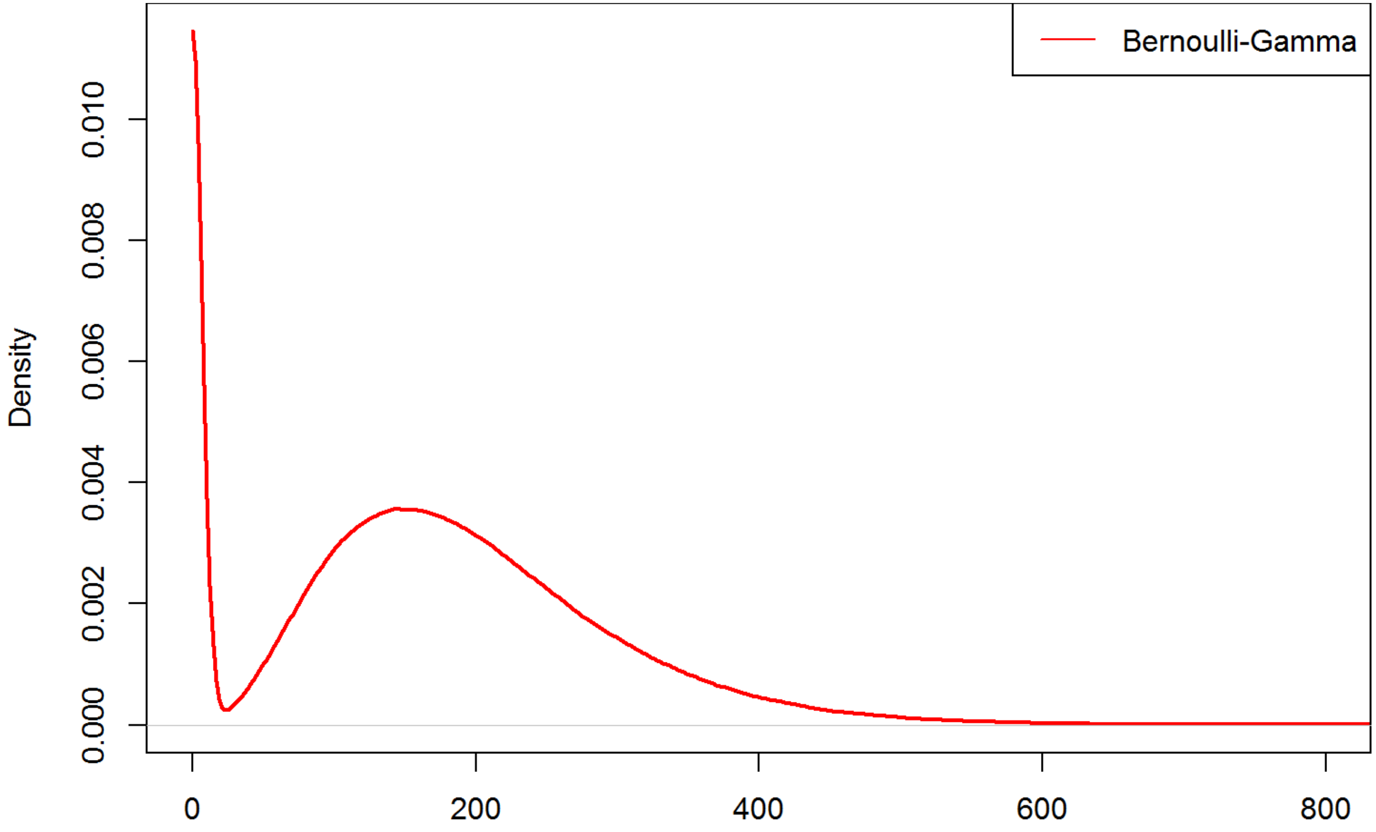
- Zeros may arise from Bernoulli process or exposure distribution
- Zeros modelled as “true” and “apparent” zeros

Hurdle models

- Zeros only arise from Bernoulli process
- Zeros and non-zeros modelled as two separate processes

Which distribution?

Excessive zeros

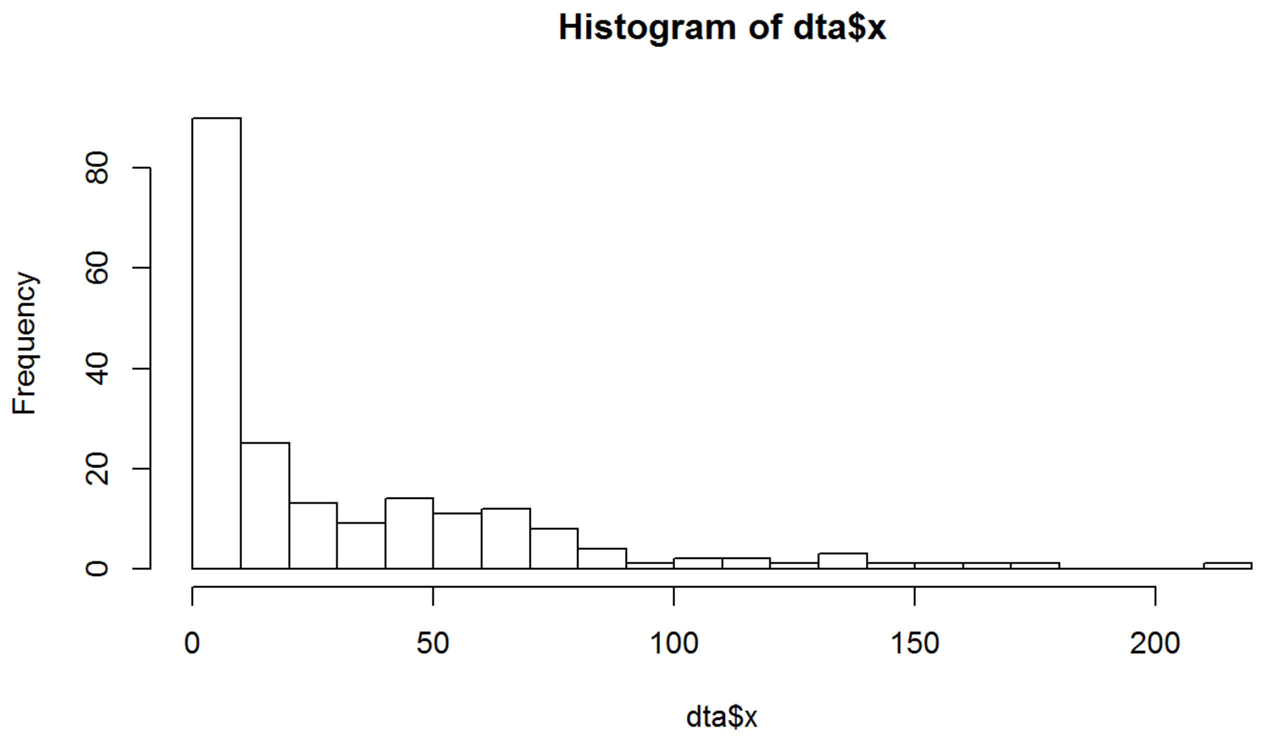


Individual observations

Individual observations

Import data

```
dta <- read.csv("consumption.csv")
hist(dta$x, breaks = 20)
```



Individual observations

Bootstrap > PAF

$$PAF = \frac{\sum P(x)RR(x) - \sum P'(x)RR(x)}{\sum P(x)RR(x)}$$

$$P(x) = 1/n$$

```
rr_fun <- rcsplineFunction(attr(fit_crc$model[[2]], "parms"), coef(fit_crc))
PRR <- mean(exp(rr_fun(dta$x)))
(PRR - 1) / PRR
```

```
## [1] 0.03224897
```

Individual observations

Fitting distributions

Motivation

- Smoothing
- Generalisability
- Mathematical ease
- Computational ease

Methods

- Method of moments
- Maximum likelihood

Individual observations

Method of moments

```
args(dgamma)
```

```
## function (x, shape, rate = 1, scale = 1/rate, log = FALSE)
## NULL
```

$$E[X] = \alpha/\beta$$

$$Var(X) = \alpha/\beta^2$$

$$\iff$$

$$\beta = E[X]/Var(X)$$

$$\alpha = E[X] * \beta$$

Individual observations

Method of moments

```
m <- mean(dta$x)
v <- var(dta$x)
b <- m / v
a <- m * b
c(a, b)
```

```
## [1] 0.63778702 0.02040918
```

```
c(m, a / b)
```

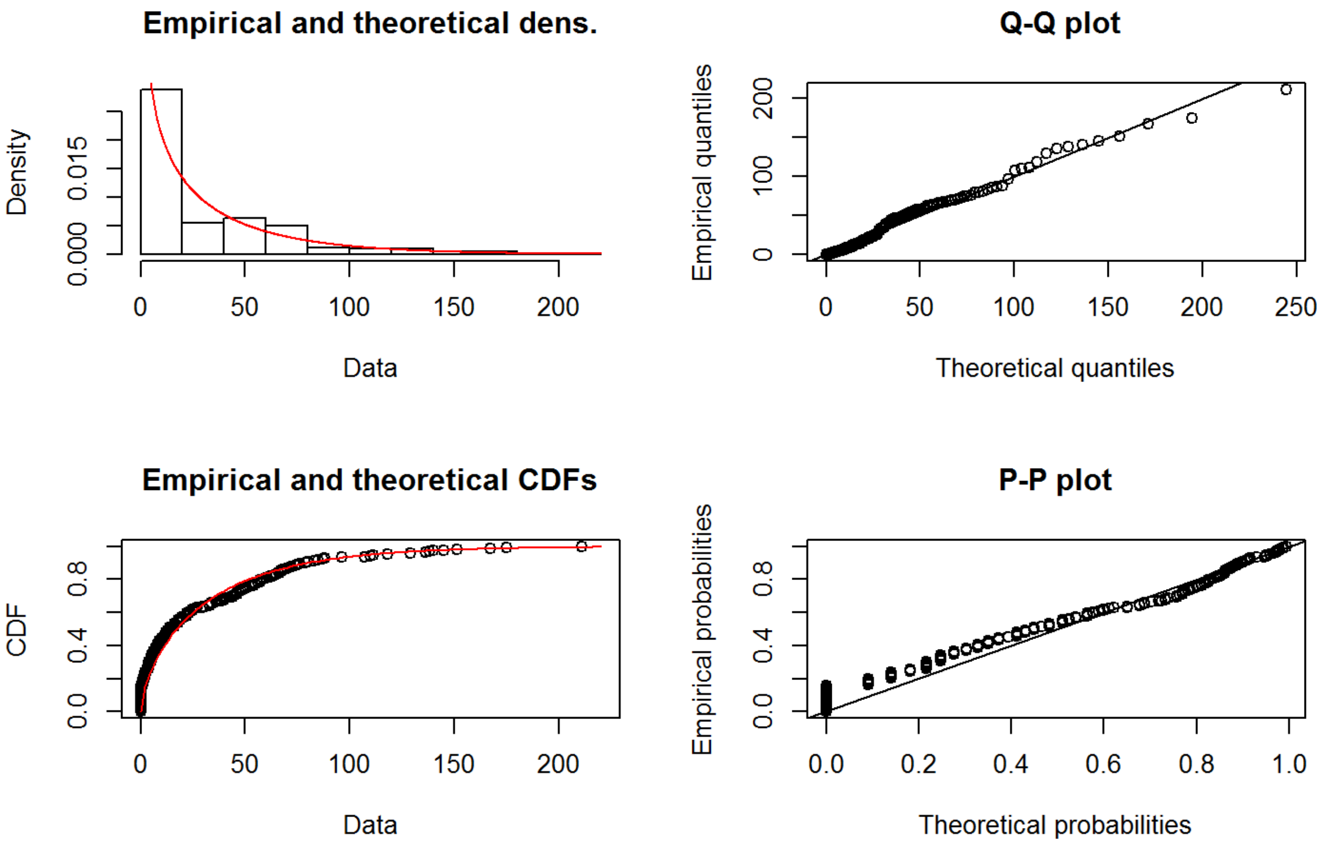
```
## [1] 31.25 31.25
```

```
c(v, a / b^2)
```

```
## [1] 1531.173 1531.173
```

Individual observations

Method of moments



Individual observations

Method of moments > PAF

```
int <-
  integrate(
    function(x)
      dgamma(x, fit_mme$`estimate`[1], fit_mme$`estimate`[2]) *
      exp(rr_fun(x)),
    lower = 0,
    upper = Inf)
PRR <- int$value
(PRR - 1) / PRR
```

```
## [1] 0.03175991
```

Individual observations

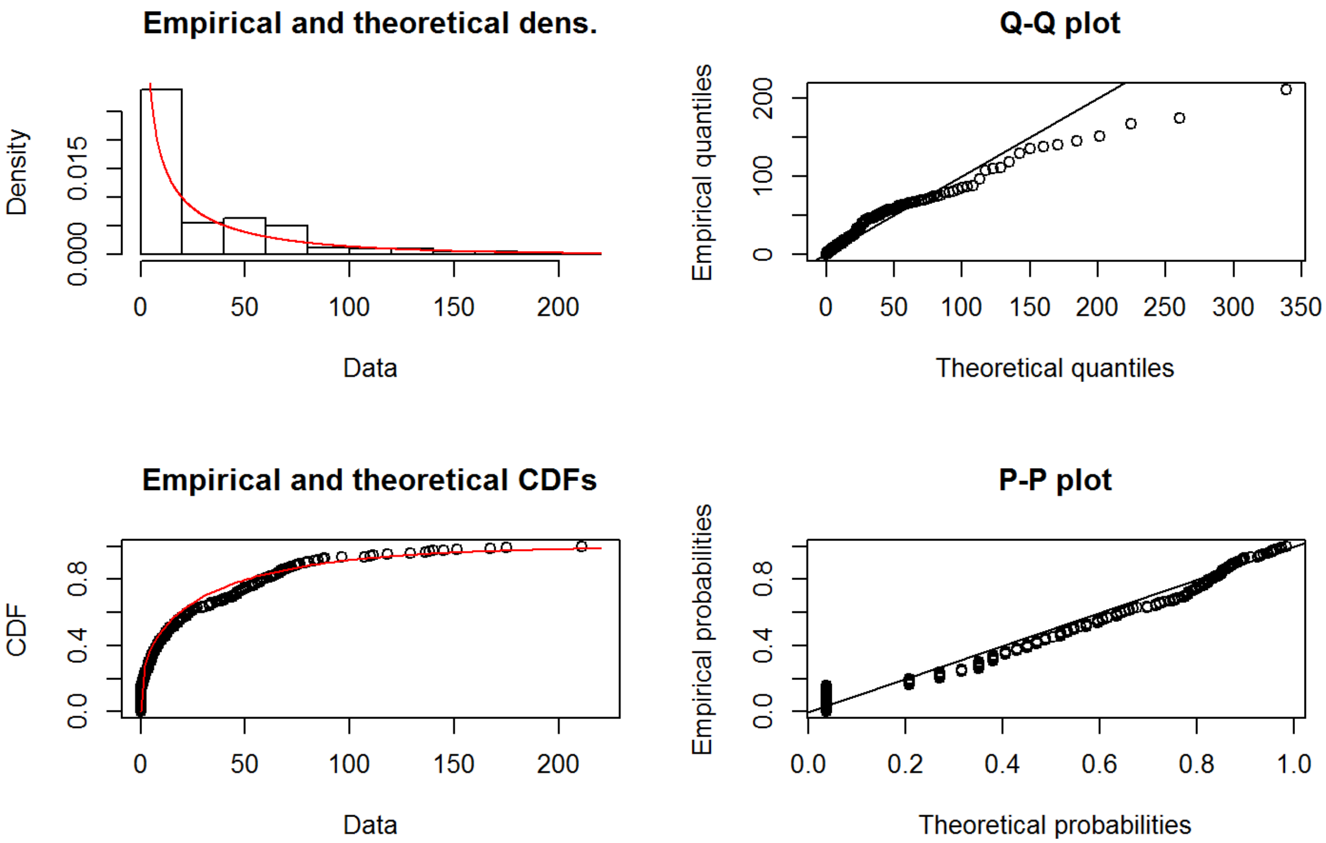
Maximum likelihood

```
dta2 <- dta$x
dta2[dta2 == 0] <- 1e-2
fit_mle <- fitdistrplus::fitdist(dta2, dgamma, "mle")
fit_mle
```

```
## Fitting of the distribution ' gamma ' by maximum likelihood
## Parameters:
##         estimate  Std. Error
## shape 0.38273119 0.030772349
## rate  0.01225496 0.001698539
```

Individual observations

Maximum likelihood



Individual observations

Maximum likelihood > PAF

```
int <-
  integrate(
    function(x)
      dgamma(x, fit_mle$`estimate`[1], fit_mle$`estimate`[2]) *
      exp(rr_fun(x)),
    lower = 0,
    upper = Inf)
PRR <- int$value
(PRR - 1) / PRR
```

```
## [1] 0.03574357
```

Individual observations

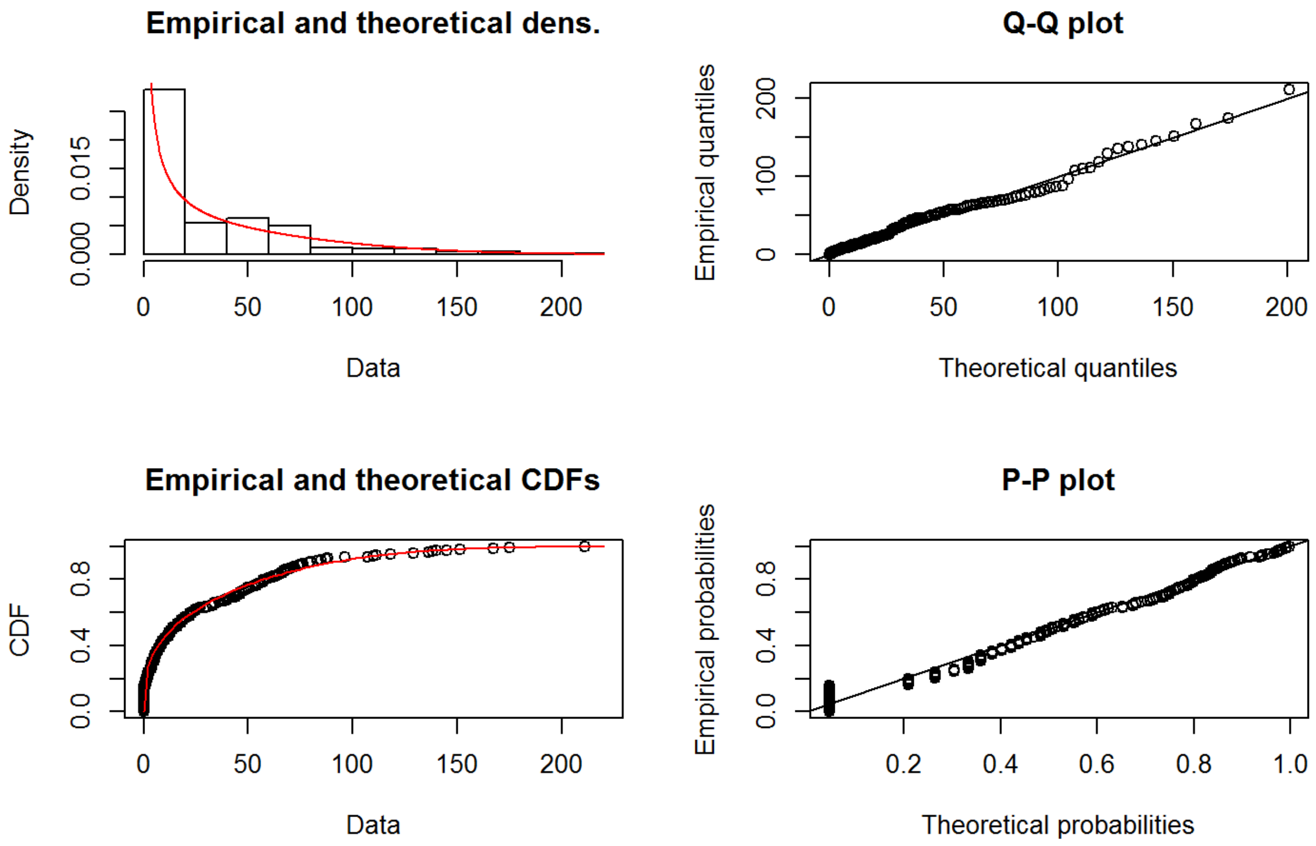
Maximum likelihood (bis)

```
library(flexsurv)
fit_mle2 <-
  fitdistrplus::fitdist(
    dta2,
    dgengamma,
    "mle",
    start = function(d)
      list(
        mu = mean(d),
        sigma = sd(d),
        Q = 0))
fit_mle2
```

```
## Fitting of the distribution ' gengamma ' by maximum likelihood
## Parameters:
##       estimate Std. Error
## mu    4.042400  0.1820763
## sigma 1.110498  0.1273631
## Q     2.697479  0.3893348
```

Individual observations

Maximum likelihood (bis)



Individual observations

Maximum likelihood (bis) > PAF

```r
int <-
  integrate(
    function(x)
      dgengamma(
        x,
        fit_mle2$`estimate`[1],
        fit_mle2$`estimate`[2],
        fit_mle2$`estimate`[3]) *
      exp(rr_fun(x)),
    lower = 0,
    upper = Inf)
PRR <- int$value
(PRR - 1) / PRR
```

```
## [1] 0.03199674
```

Individual observations

Summary

	PAF	Cases	Deaths	DALY
Bootstrap	0.0322	286	120	1847
MOM	0.0318	281	118	1819
MLE1	0.0357	317	133	2047
MLE2	0.0320	284	119	1833

Quantiles

Quantiles

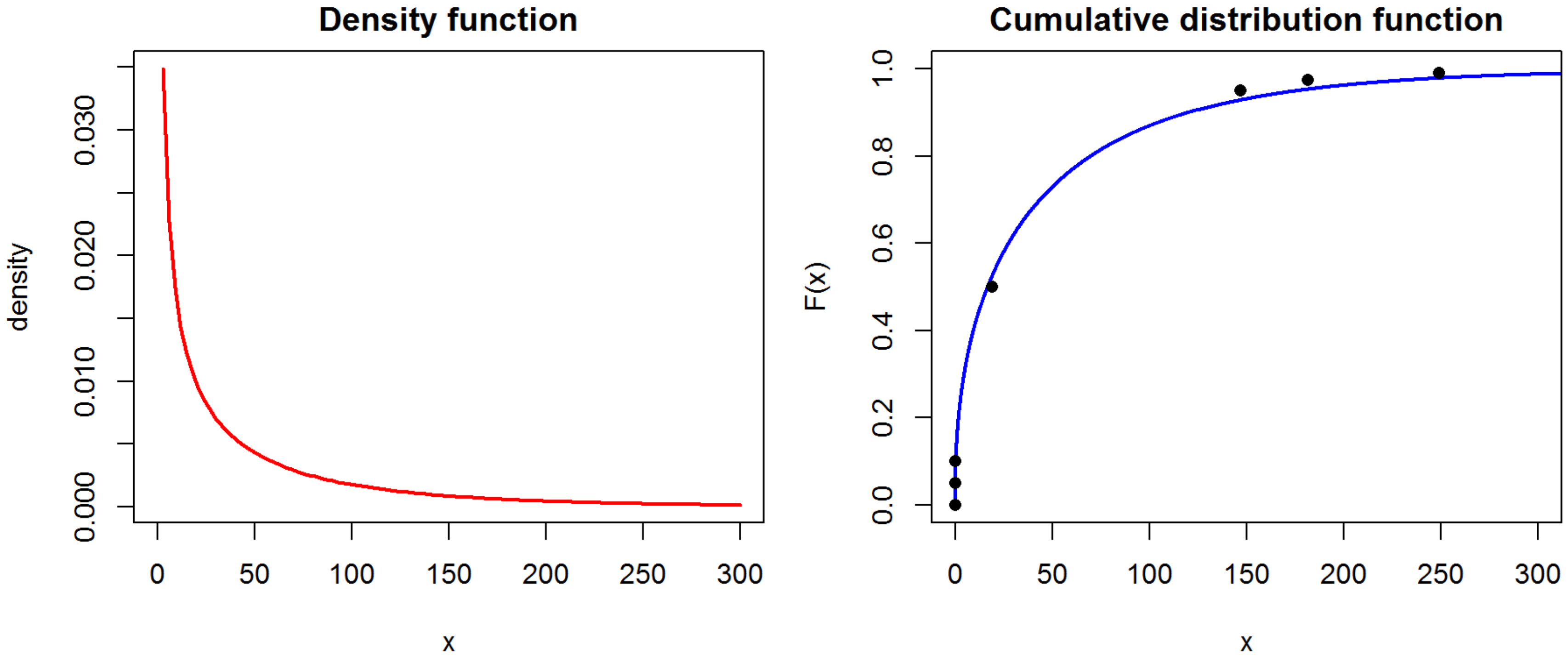
Motivating example

EFSA data on red meat consumption in Belgium

mean	P5	P10	P50	P95	P975	P99	SD
42.79	0	0	19.05	146.8	181.4	248.75	65.87

Quantiles

Method of moments



Quantiles

Optimization

Find a distribution that minimizes squared distance between observed and fitted quantiles

```r
## calculate sum of squared differences
f_gamma <-
function(par, p, q) {
  qfit <- qgamma(p = p, shape = par[1], rate = par[2])
  return(sum((qfit - q)^2))
}

## optimize
optim_gamma <-
function(p, q) {
  optim(par = c(1, 1), fn = f_gamma, p = p, q = q)
}
```

Quantiles

Optimization

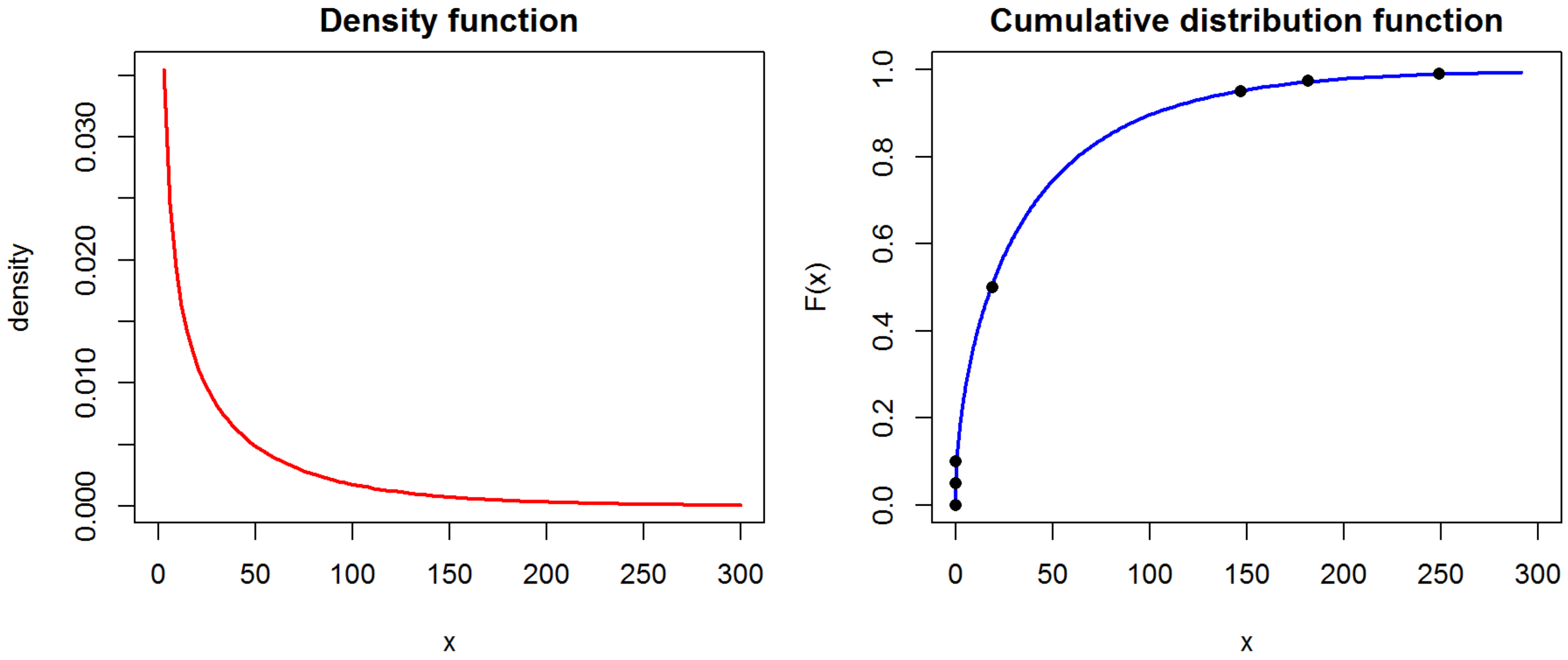
```
## find best fit
p <- c(0.05, 0.10, 0.50, 0.95, 0.975, 0.99)
fit <- optim_gamma(p = p, q = unlist(quant[1, 2:7]))
fit
```

```
## $par
## [1] 0.52763523 0.01381648
##
## $value
## [1] 50.69343
##
## $counts
## function gradient
##       97       NA
##
## $convergence
## [1] 0
##
## $message
## NULL
```

Quantiles

Optimization

Best fitting Gamma distribution



Quantiles

Optimization

	mean	P5	P10	P50	P95	P975	P99	SD
observed	42.79000	0.0000000	0.0000000	19.05000	146.8000	181.4000	248.7500	65.87000
MOM	42.79000	0.0629442	0.3259071	16.47898	174.5139	232.2550	311.5919	65.87000
optim	38.18883	0.1978381	0.7395386	18.19534	143.9096	187.2145	246.0642	52.57381

Quantiles

Optimization > PAF

	PAF	Cases	Deaths	DALY
MOM	0.0529	469	197	3032
OPTIM	0.0433	383	161	2479

Thank you