Kernel based estimation of quantiles and Risk Measures for heavy tailed distributions

Arthur Charpentier

charpentier.arthur@uqam.ca

http://freakonometrics.hypotheses.org/

based on joint work with **E. Flachaire**

initiated by some joint work with
A. Oulidi, J.D. Fermanian, O. Scaillet, G. Geenens and D. Paindaveine

(KU Leuven, AFI Seminar, 2014)

Stochastic Dominance and Related Indices

- **First Order Stochastic Dominance** (cf standard stochastic order, $\preceq_{st}$)

$$X \preceq_1 Y \Longleftrightarrow F_X(t) \geq F_Y(t), \forall t \Longleftrightarrow \text{VaR}_X(u) \leq \text{VaR}_Y(u), \forall u$$

- **Convex Stochastic Dominance** (cf martingale property)

$$X \preceq_{cx} Y \Longleftrightarrow \mathbb{E}[\tilde{Y}|\tilde{X}] = \tilde{X} \Longleftrightarrow \text{ES}_X(u) \leq \text{ES}_Y(u), \forall u \text{ and } \mathbb{E}(X) = \mathbb{E}(Y)$$

- **Second Order Stochastic Dominance** (cf submartingale property, stop-loss order, $\preceq_{icx}$)

$$X \preceq_2 Y \Longleftrightarrow \mathbb{E}[\tilde{Y}|\tilde{X}] \geq \tilde{X} \Longleftrightarrow \text{ES}_X(u) \leq \text{ES}_Y(u), \forall u$$

- **Lorenz Stochastic Dominance** (cf dilatation order)

$$X \preceq_L Y \Longleftrightarrow \frac{X}{\mathbb{E}[X]} \preceq_{cx} \frac{X}{\mathbb{E}[Y]} \Longleftrightarrow \text{L}_X(u) \leq \text{L}_Y(u), \forall u$$

Stochastic Dominance and Related Indices

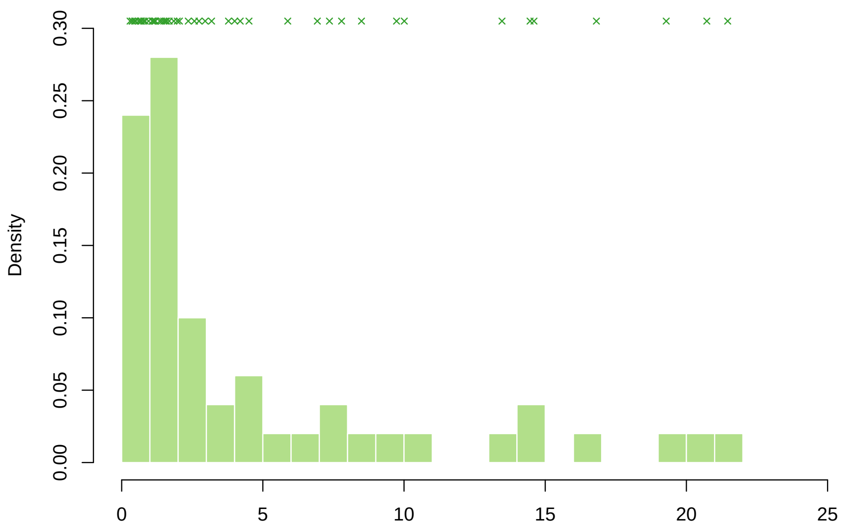
- **Parametric Model(s)**

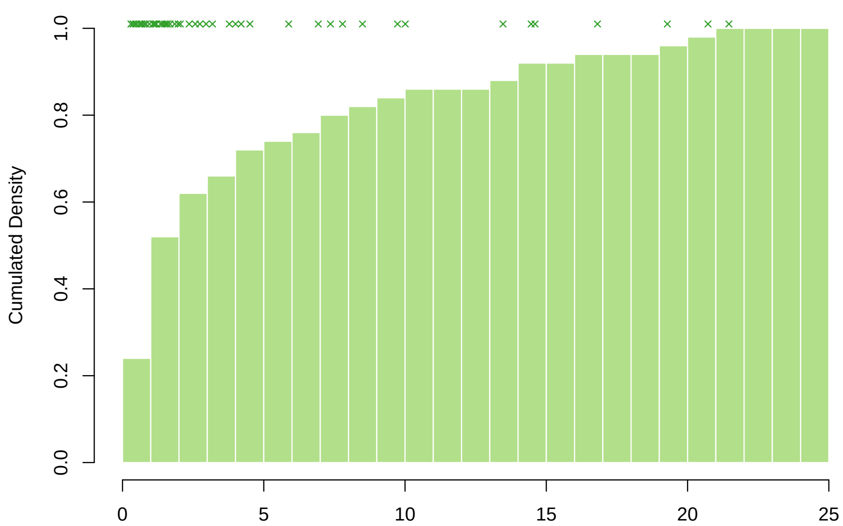
E.g. $\mathcal{N}(\mu_X,\sigma_X^2)\preceq_1\mathcal{N}(\mu_Y,\sigma_Y^2)\Longleftrightarrow\mu_X\leq\mu_Y$ and $\sigma_X^2=\sigma_Y^2$

$$\mathcal{N}(\mu_X,\sigma_X^2)\preceq_{cx}\mathcal{N}(\mu_Y,\sigma_Y^2)\Longleftrightarrow\mu_X=\mu_Y \text{ and } \sigma_X^2\leq\sigma_Y^2$$

$$\mathcal{N}(\mu_X,\sigma_X^2)\preceq_2\mathcal{N}(\mu_Y,\sigma_Y^2)\Longleftrightarrow\mu_X\leq\mu_Y \text{ and } \sigma_X^2\leq\sigma_Y^2$$

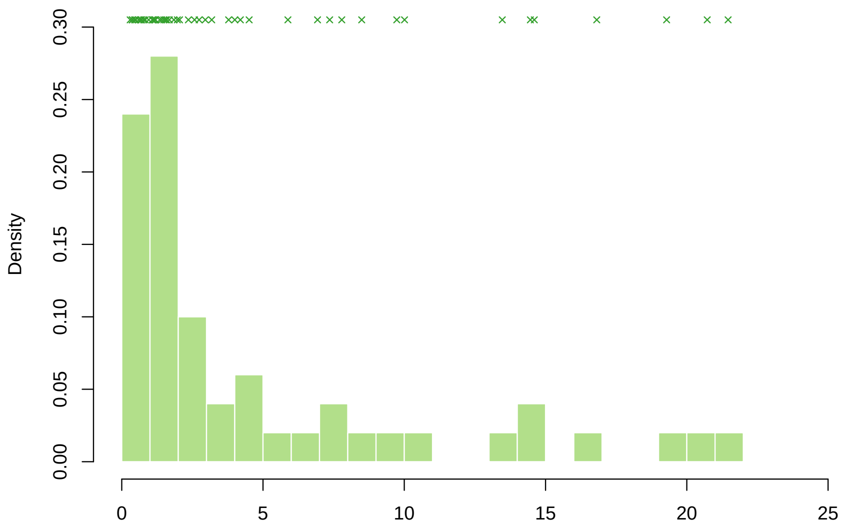
Or other parametric distribution. E.g. a lognormal distribution for losses

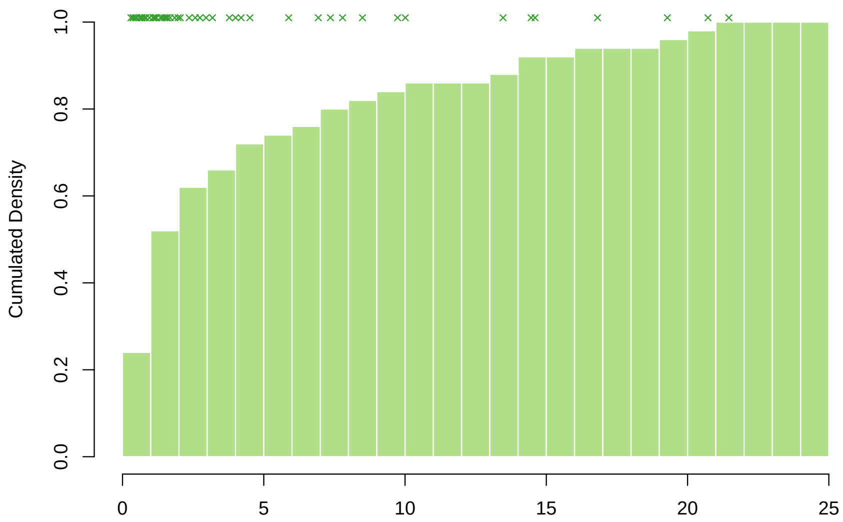




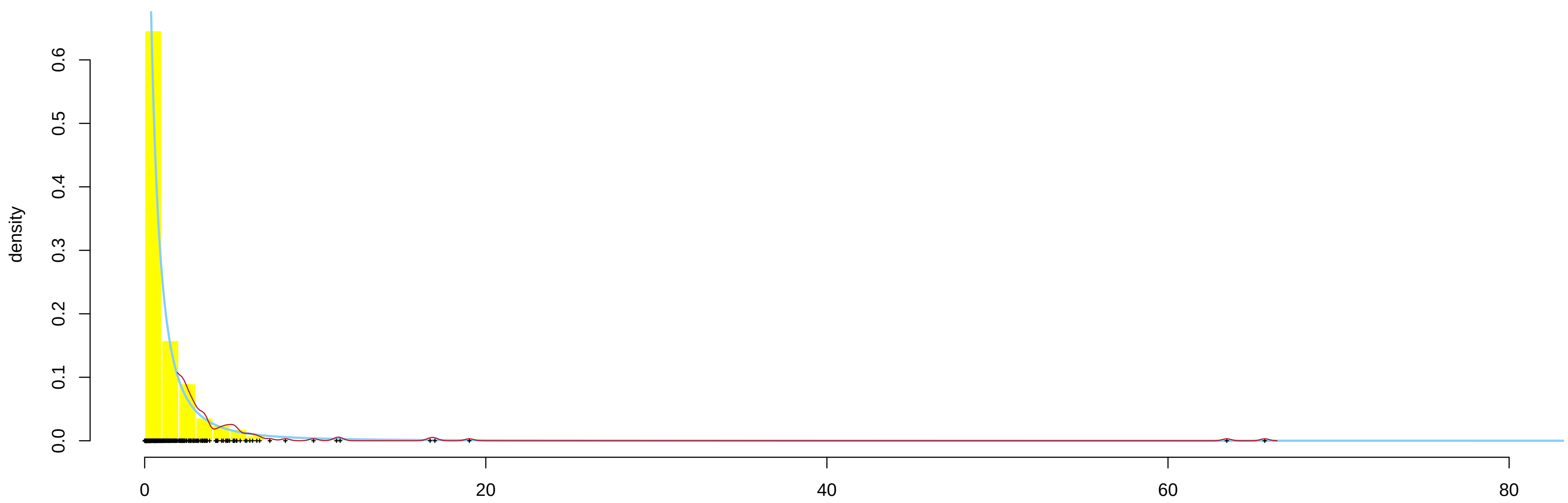
Stochastic Dominance and Related Indices

- **Non-parametric Model(s)**





Nonparametric estimation of the density



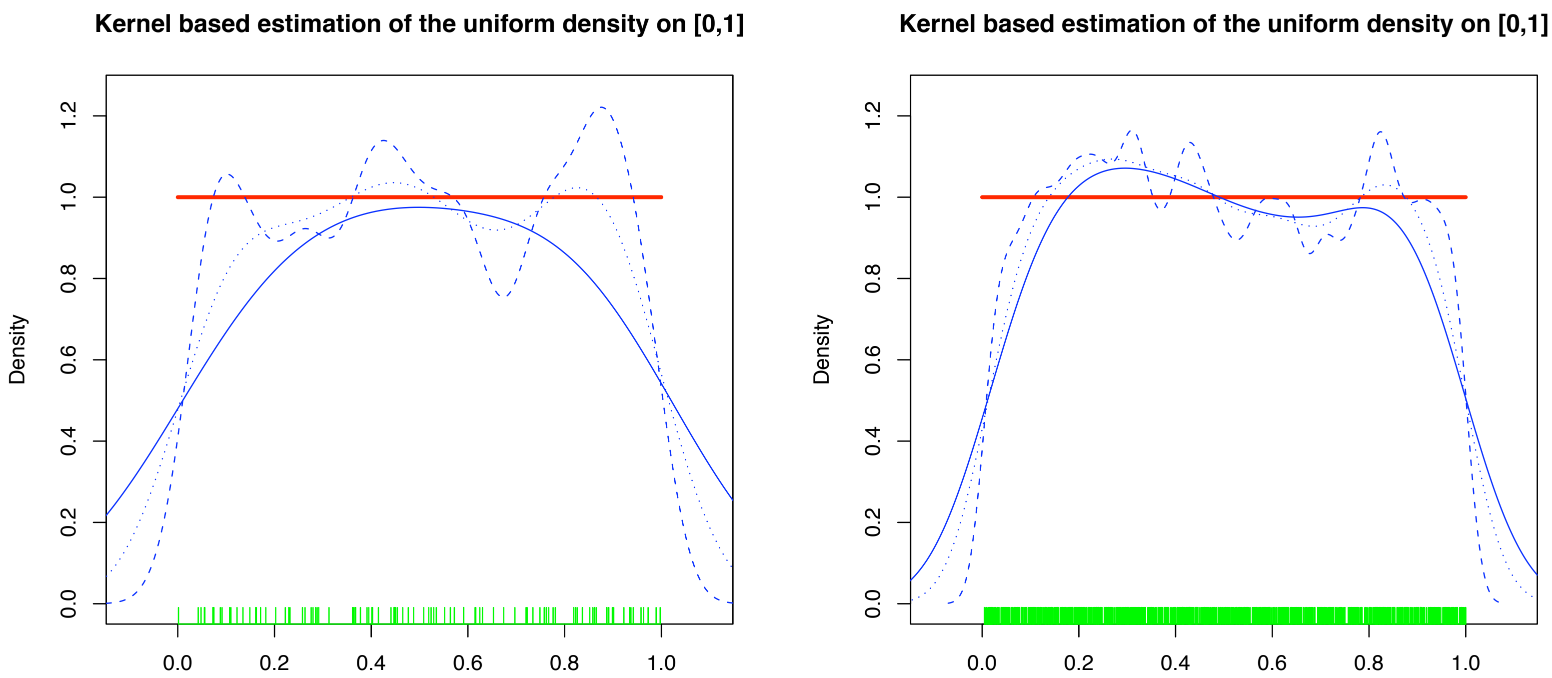
Agenda

$$
\text{sample } \{y_1,\cdots,y_n\} \longrightarrow \widehat{f}_n(\cdot) \begin{array}{l} \nearrow \quad \widehat{F}_n(\cdot) \text{ or } \widehat{F}_n^{-1}(\cdot) \\ \searrow \quad \mathcal{R}(\widehat{f}_n) \end{array}
$$

- **Estimating densities of copulas**
  - Beta kernels
  - Transformed kernels
- **Combining transformed and Beta kernels**
- **Moving around the Beta distribution**
  - Mixtures of Beta distributions
  - Bernstein Polynomials
- **Some probit type transformations**

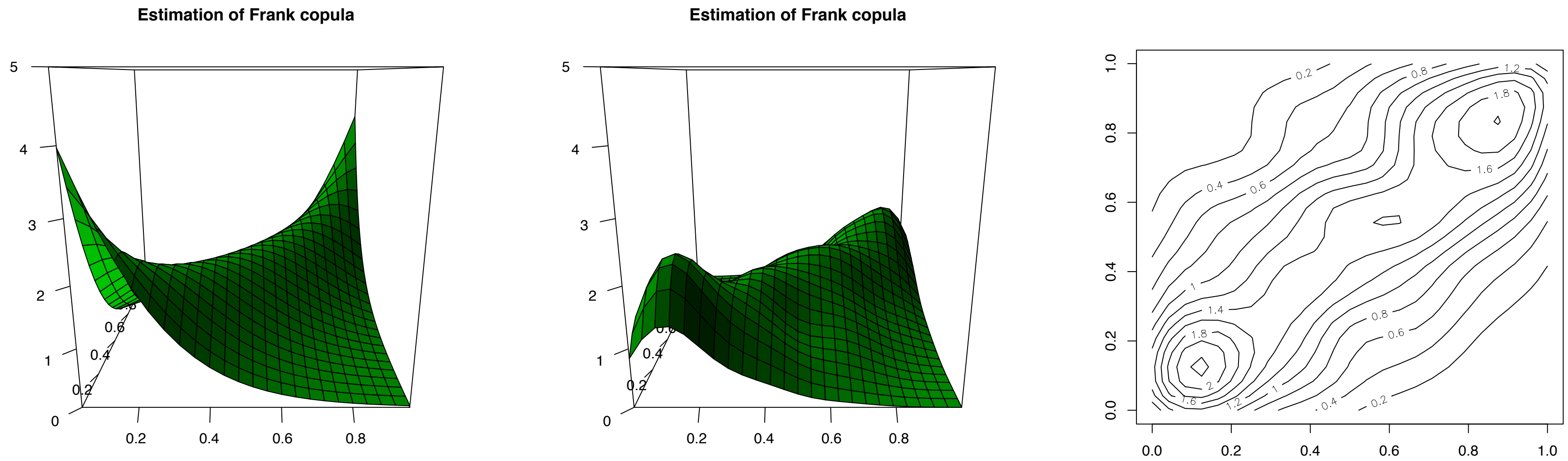
Non parametric estimation of copula density

see C., Fermanian & Scaillet (2005), bias of kernel estimators at endpoints



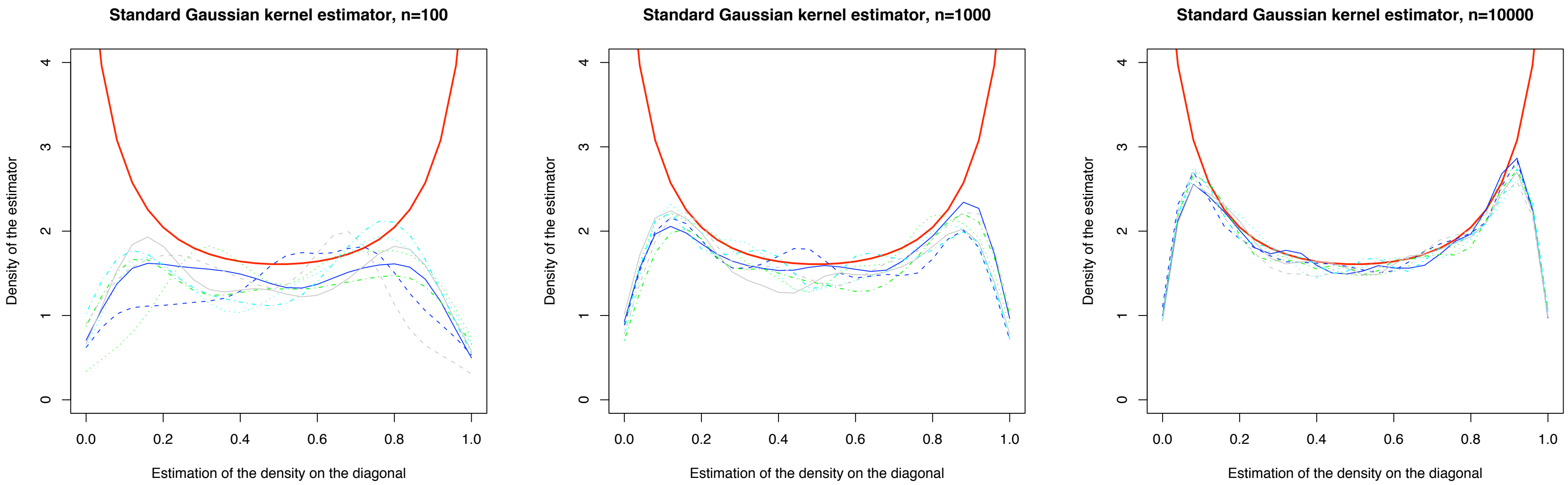
Non parametric estimation of copula density

e.g. $\mathbb{E}(\widehat{c}(0,0,h)) = \frac{1}{4} \cdot c(u,v) - \frac{1}{2}[c_1(0,0) + c_2(0,0)] \int_0^1 \omega K(\omega) d\omega \cdot h + o(h)$



with a symmetric kernel (here a Gaussian kernel).

Non parametric estimation of copula density

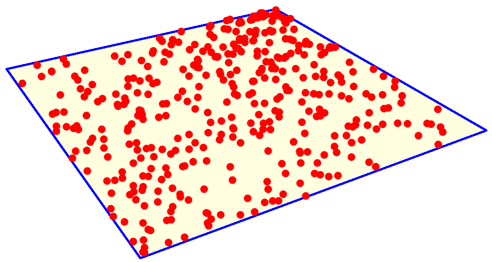


Nice asymptotic properties, see Fermanian *et al.* (2005)... but still: on finite sample, bad behavior on borders.

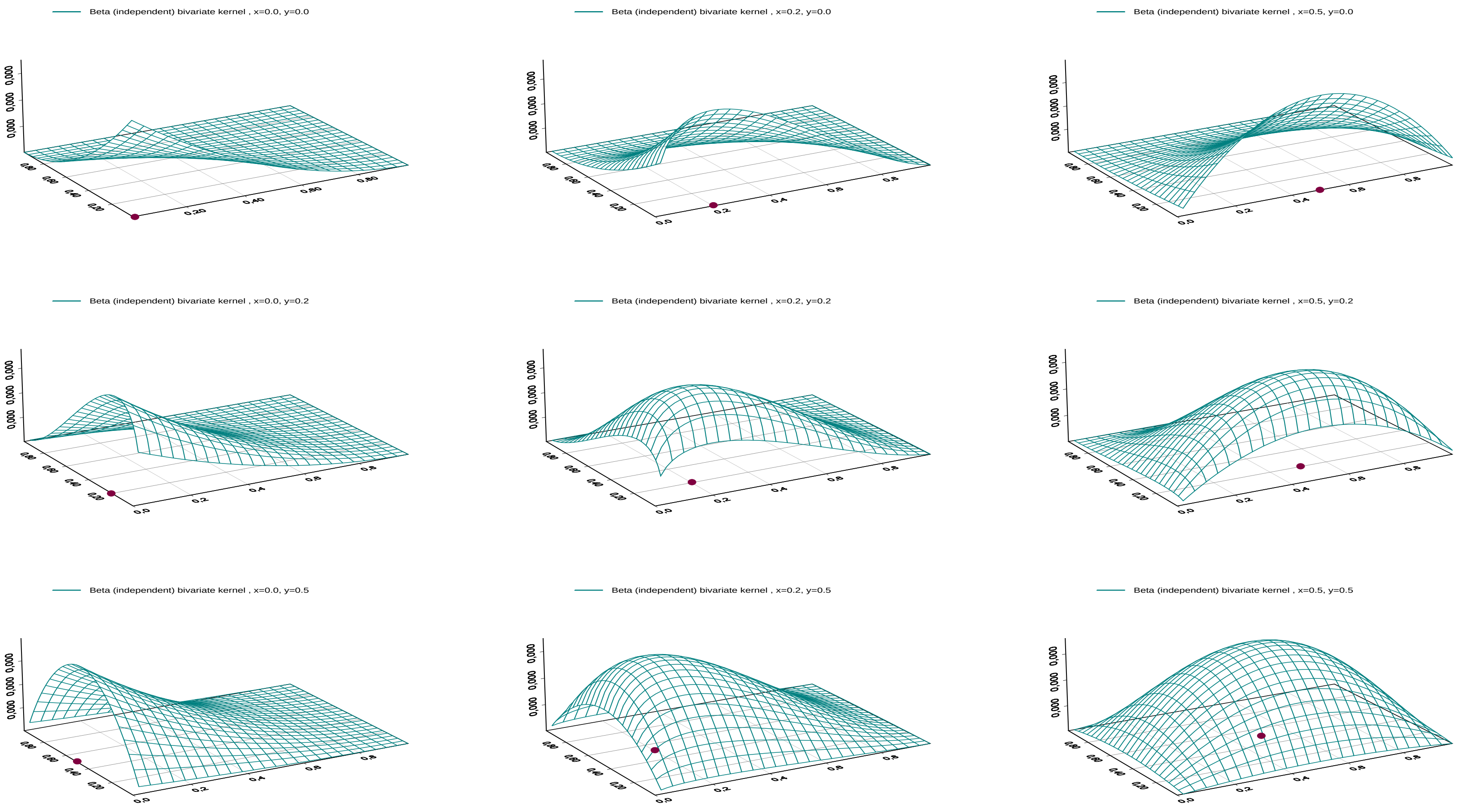
Beta kernel idea (for copulas)

see Chen (1999, 2000), Bouezmarni & Rolin (2003),

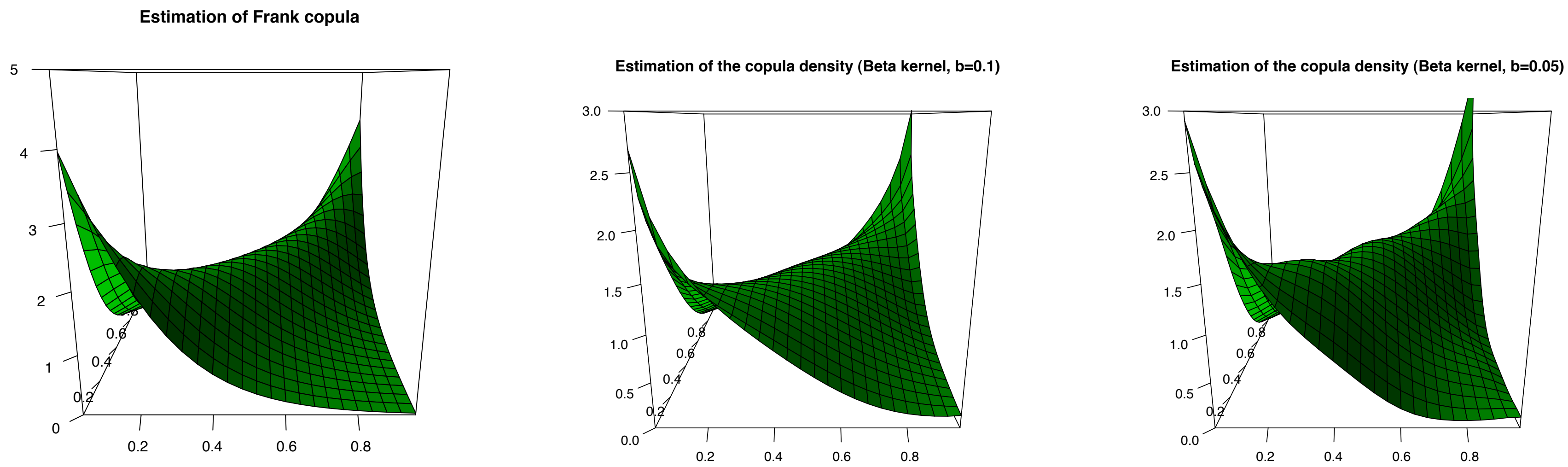
$$K_{\boldsymbol{x}_i}(\boldsymbol{u}) \propto \exp\left(-\frac{(\boldsymbol{u}-\boldsymbol{x}_i)^2}{h^2}\right) \text{ vs. } K_{\boldsymbol{x}_i}(\boldsymbol{u}) \propto \left(u_1^{\frac{x_{1,i}}{b}}[1-u_1]^{\frac{x_{1,i}}{b}}\right)\cdot\left(u_2^{\frac{x_{2,i}}{b}}[1-u_2]^{\frac{x_{2,i}}{b}}\right)$$



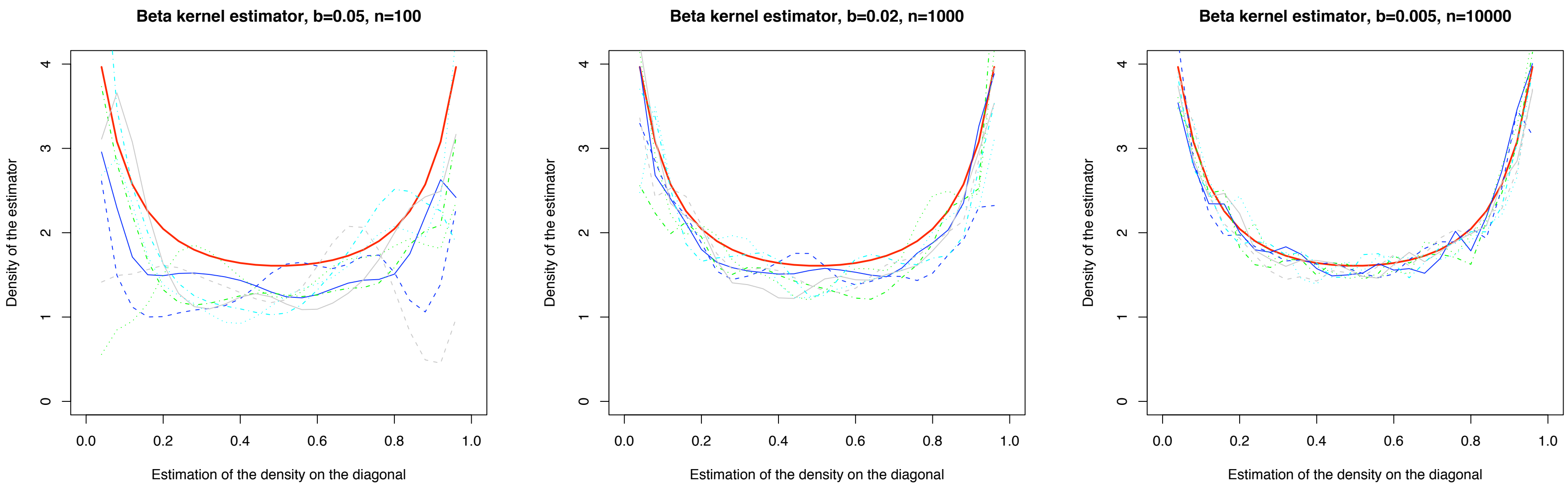
Beta kernel idea (for copulas)



Beta kernel idea (for copulas)

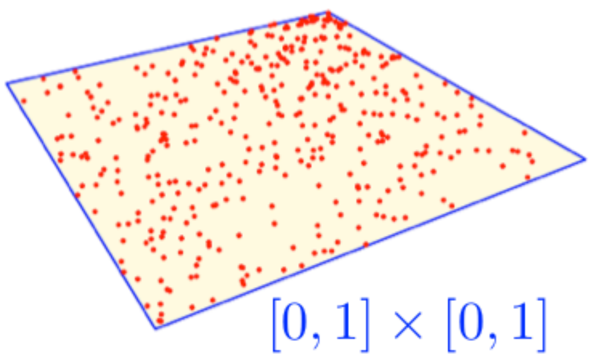


Beta kernel idea (for copulas)

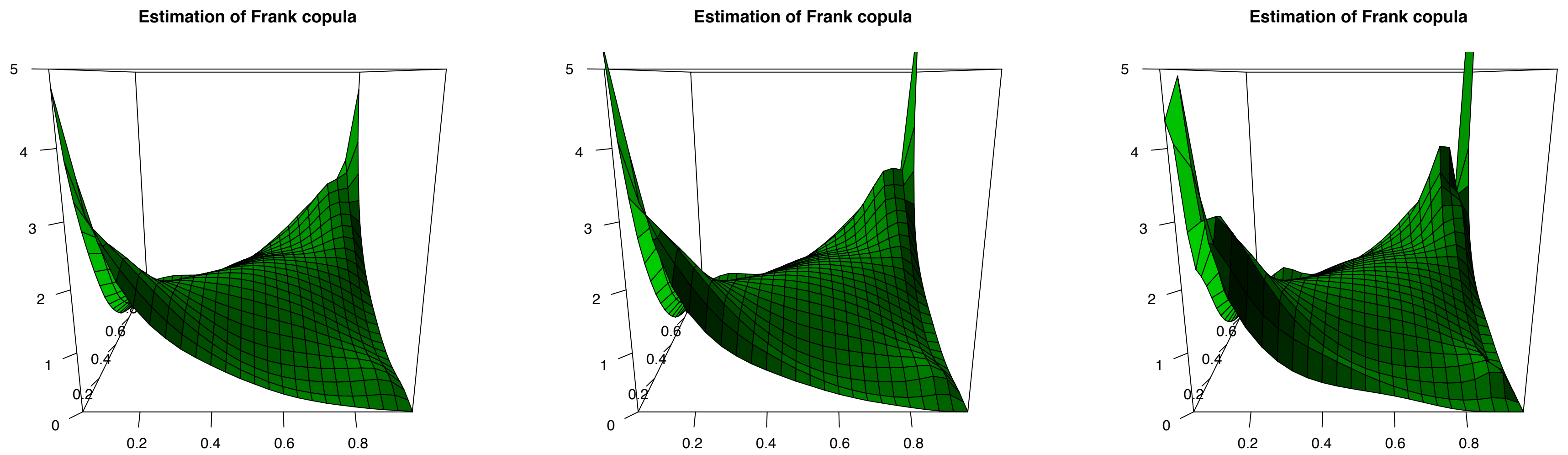


Transformed kernel idea (for copulas)

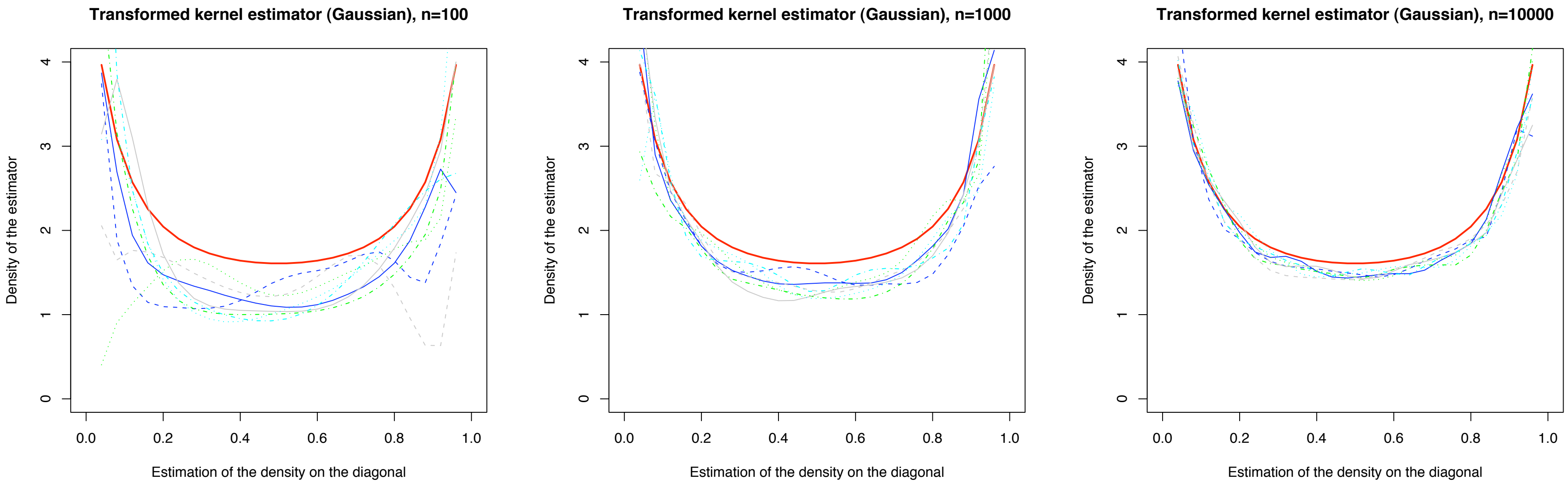
$$[0,1]\times[0,1]\to\mathbb{R}\times\mathbb{R}\to[0,1]\times[0,1]$$



Transformed kernel idea (for copulas)



Transformed kernel idea (for copulas)



see Geenens, C. & Paindaveine (2014) for more details on probit transformation for copulas.

Combining the two approaches

See Devroye & Györfi (1985), and Devroye & Lugosi (2001)

CHAPTER 9

The Transformed Kernel Estimate

... use the transformed kernel the other way, $\mathbb{R} \rightarrow [0,1] \rightarrow \mathbb{R}$

The *transformed kernel estimate* (Devroye et al., 1983) is based upon a transformation T: $R^1 \rightarrow [0,1]$ which is strictly monotonically increasing, continuously differentiable, one-to-one and onto, and which has a continuously differentiable inverse. The transformed data sequence is $Y_1, \ldots, Y_n$, where $Y_i = T(X_i)$. Note that Y_1 has density

$$g(x) = f(T^{-1}(x))T^{-1\prime}(x).$$

Now, g is estimated by g_n from $Y_1, \ldots, Y_n$, and f is estimated by

$$f_n(x) = g_n(T(x))T'(x). \tag{2}$$

Devroye & Györfi (1985) - Devroye & Lugosi (2001)

Interesting point, the optimal T should be F,

The only unknown in the design at this moment is our transformation T. We point out that for a transformed histogram estimate, the optimal T gives a uniform $[0, 1]$ density and should therefore be equal to $T(x) = F(x)$, all x. The h to be used in the histogram estimate is $(2\pi n)^{-1/3}$ (Table 5.1).

thus, T can be $F_{\widehat{\theta}}$

The key observation is that if g_n is a density on $[0, 1]$, the f_n is a density on R^1, and furthermore,

$$\int |f_n - f| = \int |g_n - g|.$$

Consistency 251

For variable transformations T, we must worry about the consistency of the resulting estimate.

The transformation $Y_i = T(X_i)$ is usually of the form

$$Y_i = T_n(X_i; X_1, \ldots, X_n),$$

Heavy Tailed distribution

Let X denote a (heavy-tailed) random variable with tail index $\alpha \in (0, \infty)$, i.e.

$$\mathbb{P}(X > x) = x^{-\alpha}\mathcal{L}_1(x)$$

where $\mathcal{L}_1$ is some regularly varying function.

Let T denote a $\mathbb{R} \to [0, 1]$ function, such that $1 - T$ is regularly varying at infinity, with tail index $\beta \in (0, \infty)$.

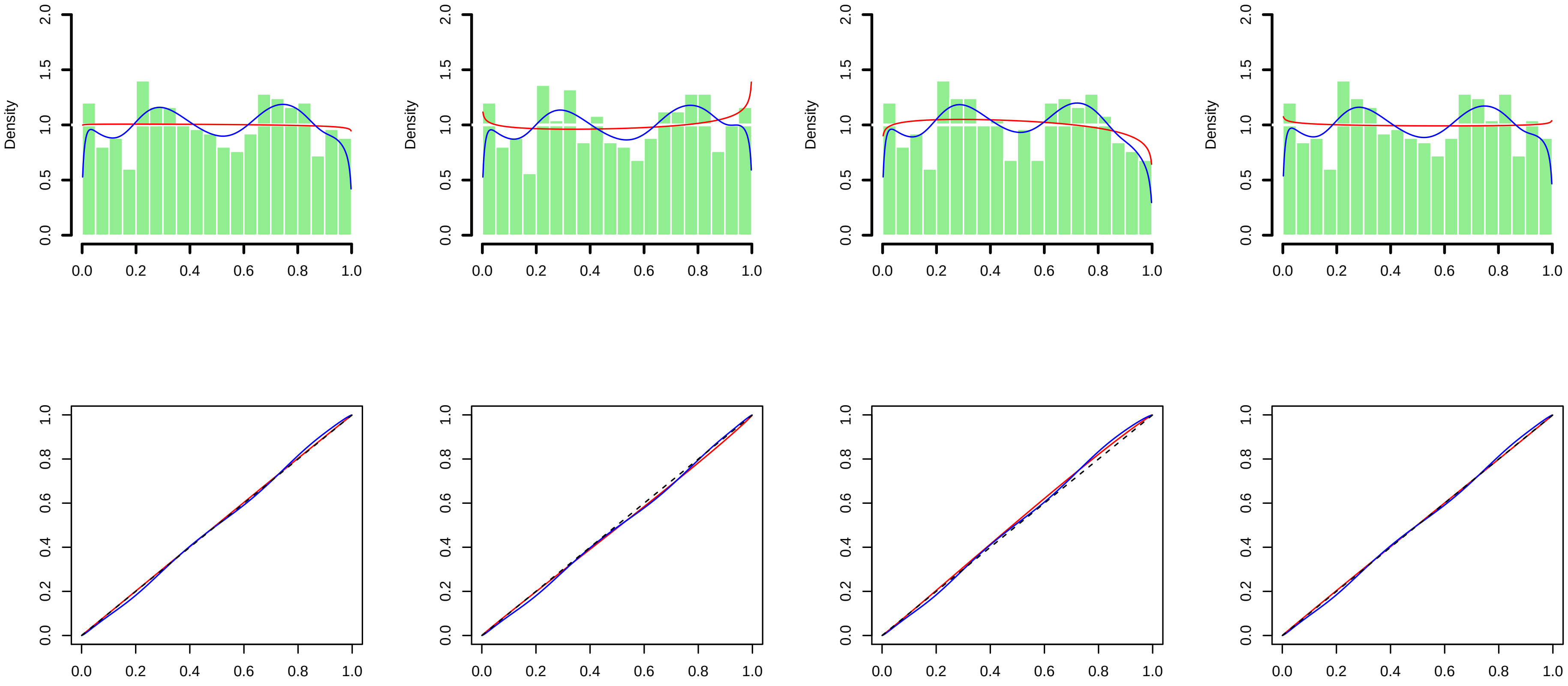
Define $Q(x) = T^{-1}(1 - x^{-1})$ the associated tail quantile function, then $Q(x) = x^{1/\beta}\mathcal{L}_2^\star(1/x)$, where $\mathcal{L}_2^\star$ is some regularly varying function (the de Bruyn conjugate of the regular variation function associated with T). Assume here that $Q(x) = bx^{1/\beta}$

Let $U = T(X)$. Then, as $u \to 1$

$$\mathbb{P}(U > u) \sim (1 - u)^{\alpha/\beta}.$$

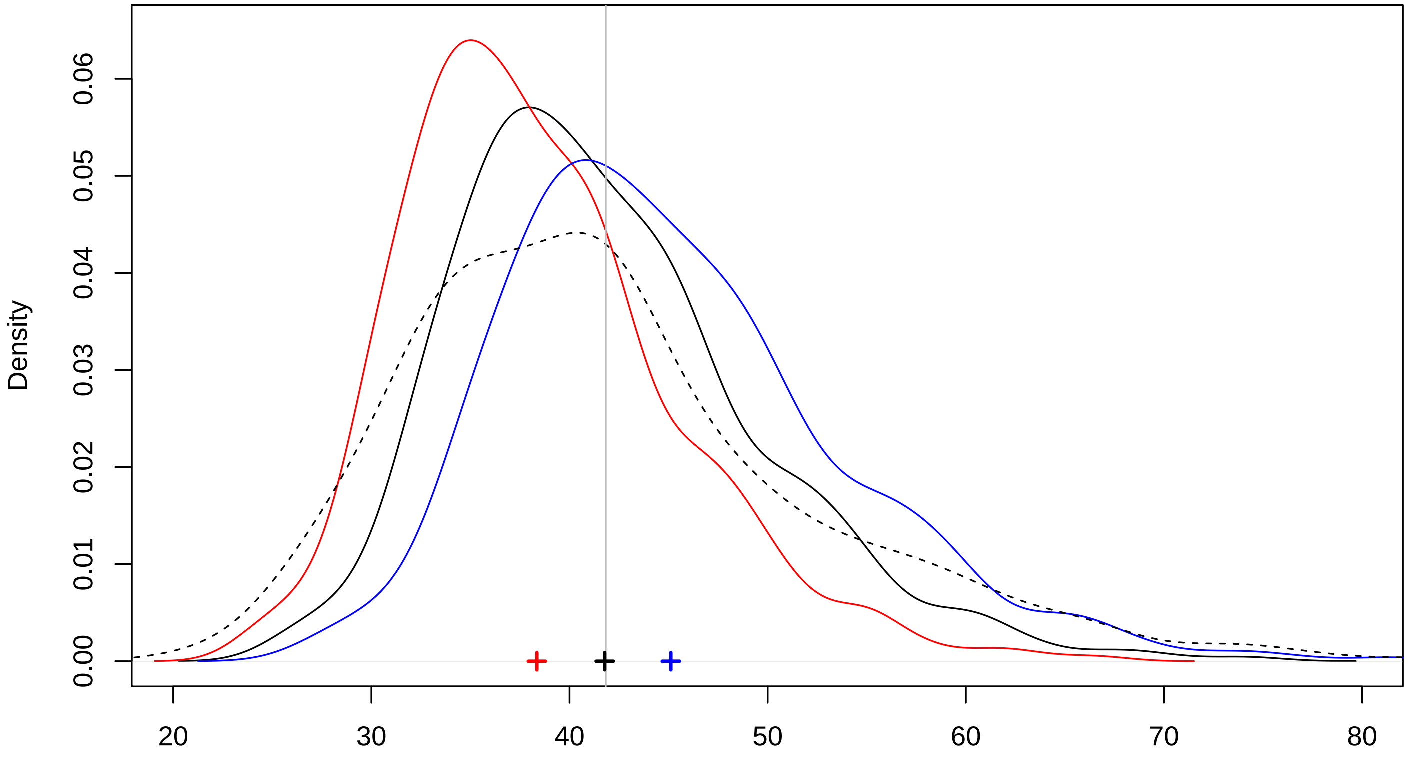
Heavy Tailed distribution

see C. & Oulidi (2007), $\alpha = 0.75^{-1}$, $T_{0.75^{-1}}$, $\underbrace{T_{0.65^{-1}}}_{\text{lighter}}$, $\underbrace{T_{0.85^{-1}}}_{\text{heavier}}$ and $T_{\hat{\alpha}}$



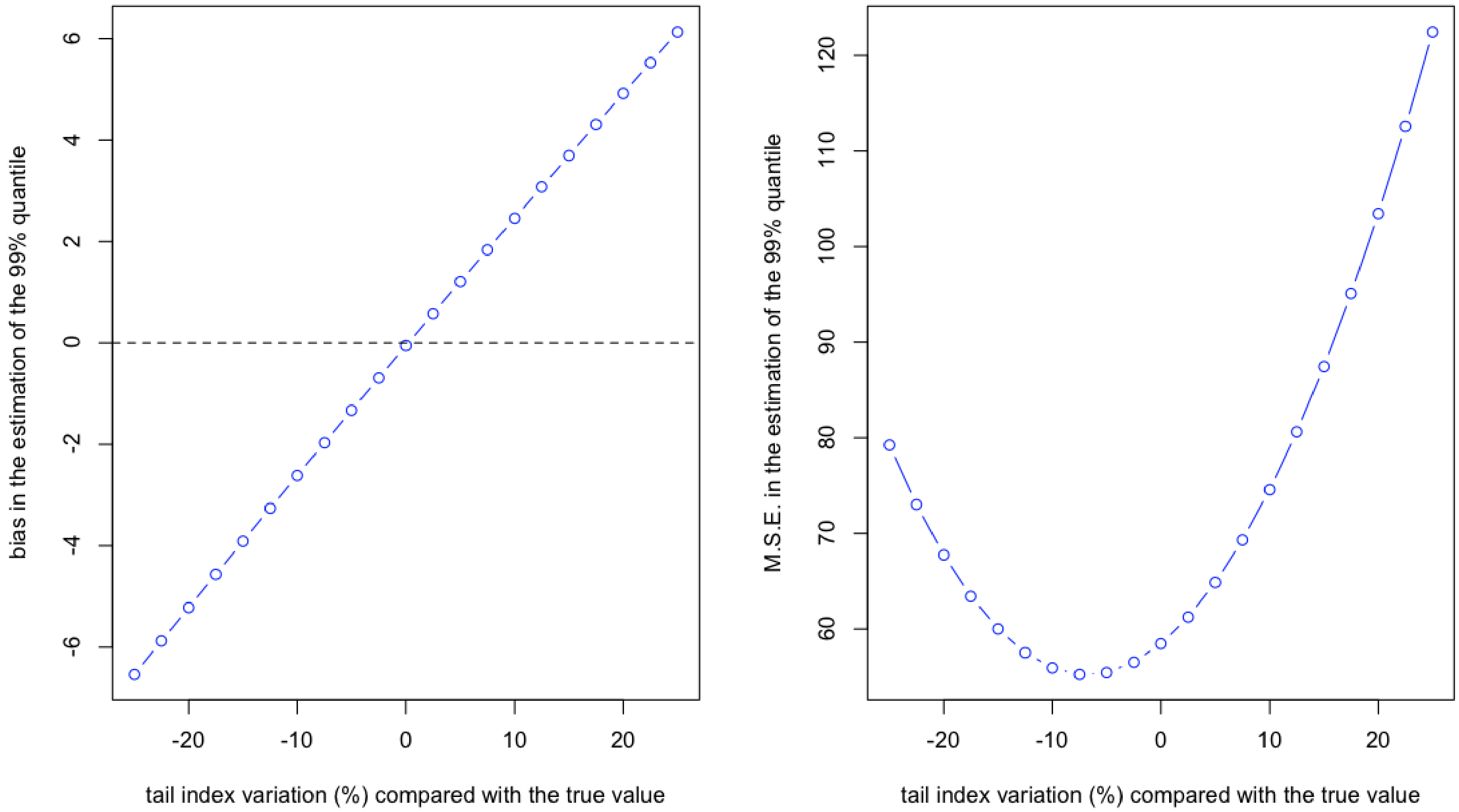
Heavy Tailed distribution

see C. & Oulidi (2007), impact on quantile estimation ?



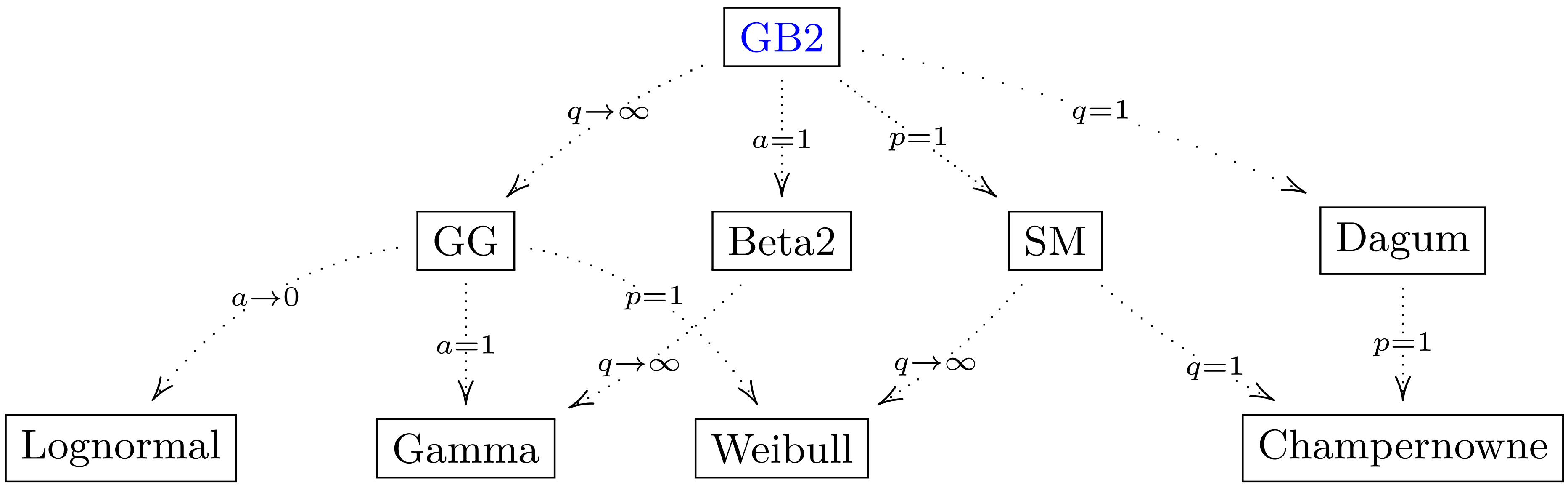
Heavy Tailed distribution

see C. & Oulidi (2007), impact on quantile estimation ? bias ? m.s.e. ?



Which transformation ?

$$\text{GB2}: \quad t(y;a,b,p,q)=\frac{|a|y^{ap-1}}{b^{ap}\mathrm{B}(p,q)[1+(y/b)^a]^{p+q}}, \qquad \text{for } y>0,$$



Estimating a density on $\mathbb{R}^+$

- **Stay on $\mathbb{R}+$: x_i's**
- **Get on $[0,1]$: $u_i = T_{\widehat{\theta}}(x_i)$ (distribution as uniform as possible)**
  - Use Beta Kernels on u_i's
  - Mixtures of Beta distributions on u_i's
  - Bernstein Polynomials on u_i's
- **Get on $\mathbb{R}$: use standard kernels (e.g. Gaussian)**
  - On $x_i^\star = \log(x_i)$
  - On $x_i^\star = \text{BoxCox}_{\widehat{\lambda}}(x_i)$
  - On $x_i^\star = \Phi^{-1}[T_{\widehat{\theta}}(x_i)]$

Beta kernel

$$\widehat{g}(u) = \sum_{i=1}^{n} \frac{1}{n} \cdot b\left(u; \frac{U_i}{h}, \frac{1-U_i}{h}\right) \quad u \in [0,1].$$

with some possible boundary correction, as suggested in Chen (1999),

$$\frac{u}{h} \to \rho(u,h) = 2h^2 + 2.5 - (4h^4 + 6h^2 + 2.25 - u^2 - u/h)^{1/2}$$

Problem : choice of the bandwidth $h^\star$? Standard loss function

$$L(h) = \int [\widehat{g}_n(u) - g(u)]^2 du = \underbrace{\int [\widehat{g}_n(u)]^2 du - 2\int \widehat{g}_n(u) \cdot g(u) du}_{CV(h)} + \int [g(u)]^2 du$$

where

$$\widehat{CV}(h) = \left(\int \widehat{g}_n(u) du\right)^2 - \frac{2}{n}\sum_{i=1}^{n} \widehat{g}_{(-i)}(U_i)$$

Mixture of Beta distributions

$$\widehat{g}(u)=\sum_{j=1}^{k}\widehat{\pi}_j\cdot b\left(u;\widehat{\alpha}_j,\widehat{\beta}_j\right)\ \ u\in[0,1].$$

Problem : choice the number of components k (and estimation...). Use of stochastic EM algorithm (or sort of) see Celeux & Diebolt (1985).

Bernstein approximation

$$\widehat{g}(u)=\sum_{k=1}^{m}[m\omega_k]\cdot b\left(u;k,m-k\right)\ \ u\in[0,1].$$

where $\omega_k=\widehat{G}\left(\frac{k}{m}\right)-\widehat{G}\left(\frac{k-1}{m}\right)$.

On the log-transform

With a standard Gaussian kernel

$$\widehat{f}_X(x) = \frac{1}{n}\sum_{i=1}^{n} \phi(x; x_i, h)$$

A Gaussian kernel on a log transform,

$$\widehat{f}_X(x) = \frac{1}{x}\widehat{f}_{X^\star}(\log x) = \frac{1}{n}\sum_{i=1}^{n} \lambda(x; \log x_i, h)$$

where $\lambda(\cdot; \mu, \sigma)$ is the density of the log-normal distribution. Here, in 0,

$$\text{bias}[\widehat{f}_X(x)] \sim \frac{h^2}{2}[f_X(x) + 3x \cdot f'_X(x) + x^2 \cdot f''_X(x)]$$

and

$$\text{Var}[\widehat{f}_X(x)] \sim \frac{f_X(x)}{xnh}$$

On the Box-Cox-transform

More generally, instead of transformed sample $Y_i = \log[X_i]$, consider

$$Y_i = \frac{X_i^\lambda - 1}{\lambda} \text{ when } \lambda \neq 0.$$

Find the optimal transformation using standard regression techniques (least squares)

$$X_i^\star = \frac{X_i^{\lambda^\star} - 1}{\lambda^\star} \text{ when } \lambda^\star \neq 0$$

and $X_i^\star = \log[X_i]$ if $\lambda^\star = 0$. The density estimation is here

$$\widehat{f}_X(x) = x^{\lambda^\star - 1}\widehat{f}_{X^\star}\left(\frac{x^{\lambda^\star} - 1}{\lambda^\star}\right)$$

Illustration with Log-normal samples

Standard kernel (— Silvermans's rule $h^\star$)

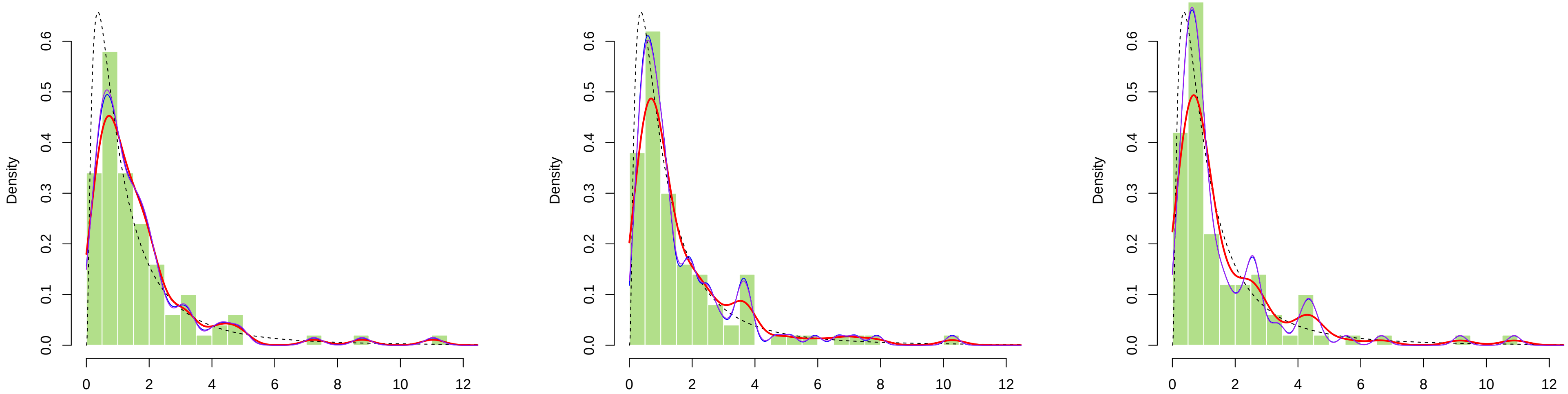


Illustration with Log-normal samples

Log transform, $x_i^\star = \log x_i$ + Gaussian kernel

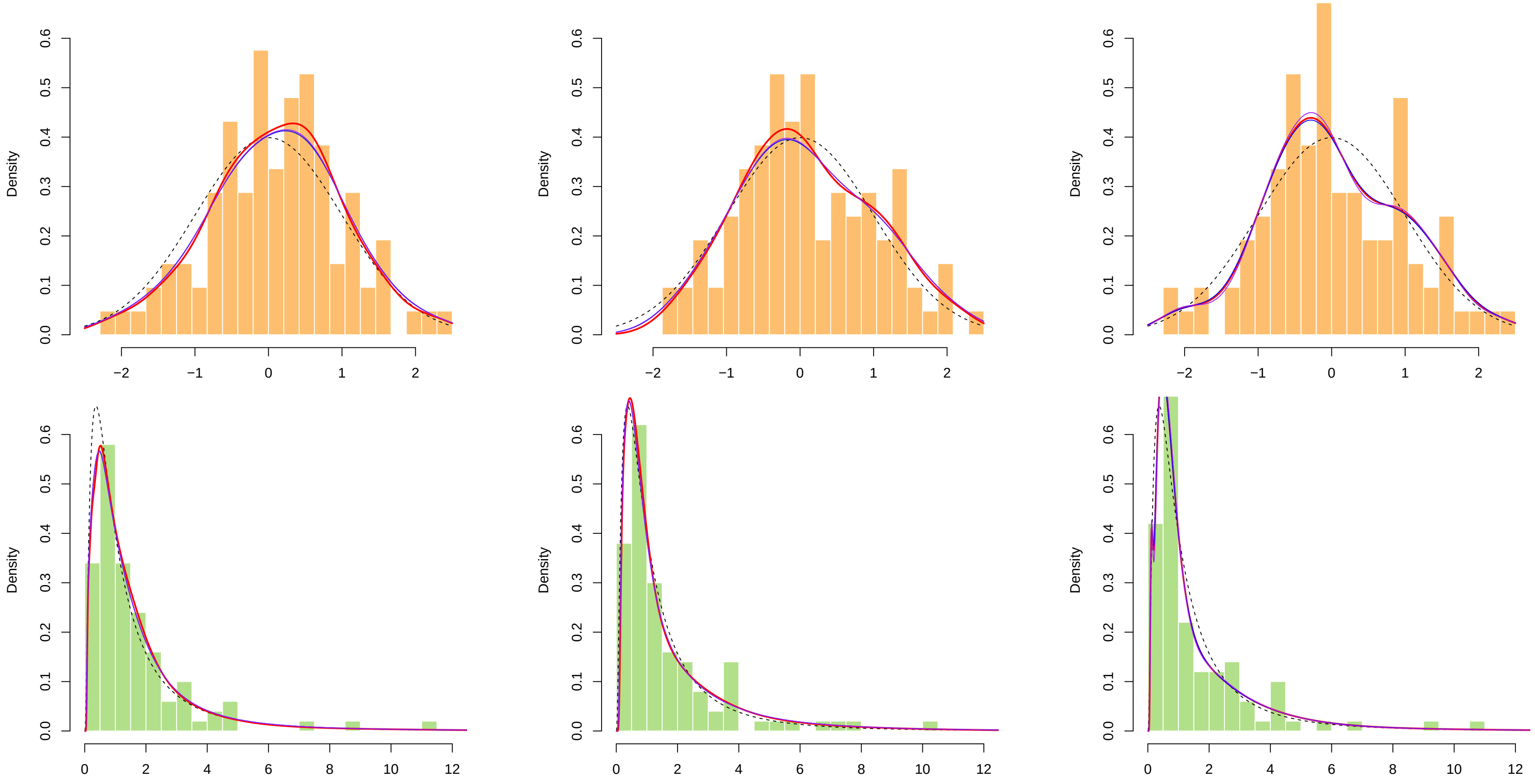


Illustration with Log-normal samples

Probit-type transform, $x_i^\star = \Phi^{-1}[T_{\widehat{\theta}}(x_i)]$ + Gaussian kernel

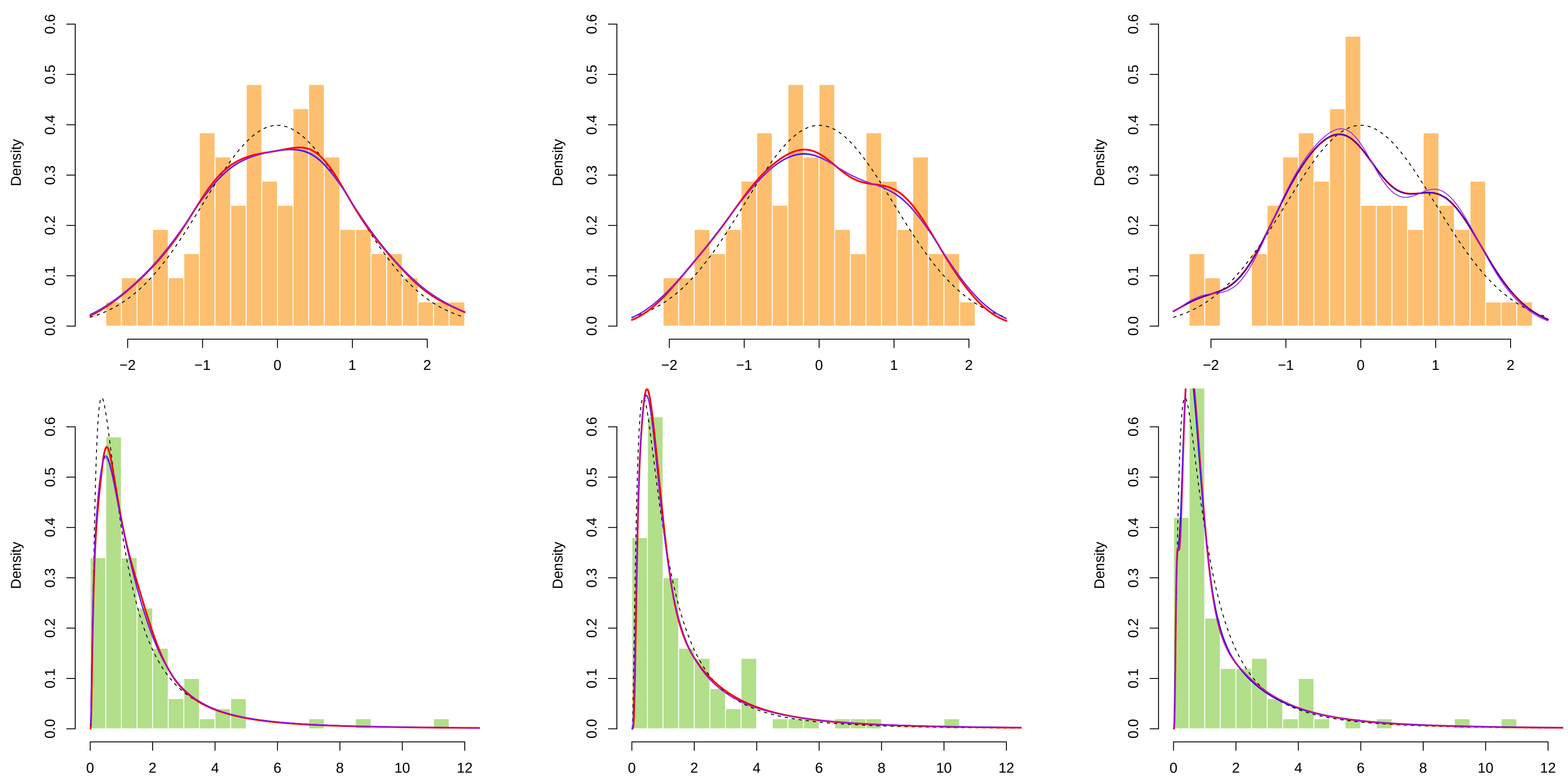


Illustration with Log-normal samples

Box-Cox transform, $x_i^\star = \text{BoxCox}_{\widehat{\lambda}}(x_i)$ + Gaussian kernel

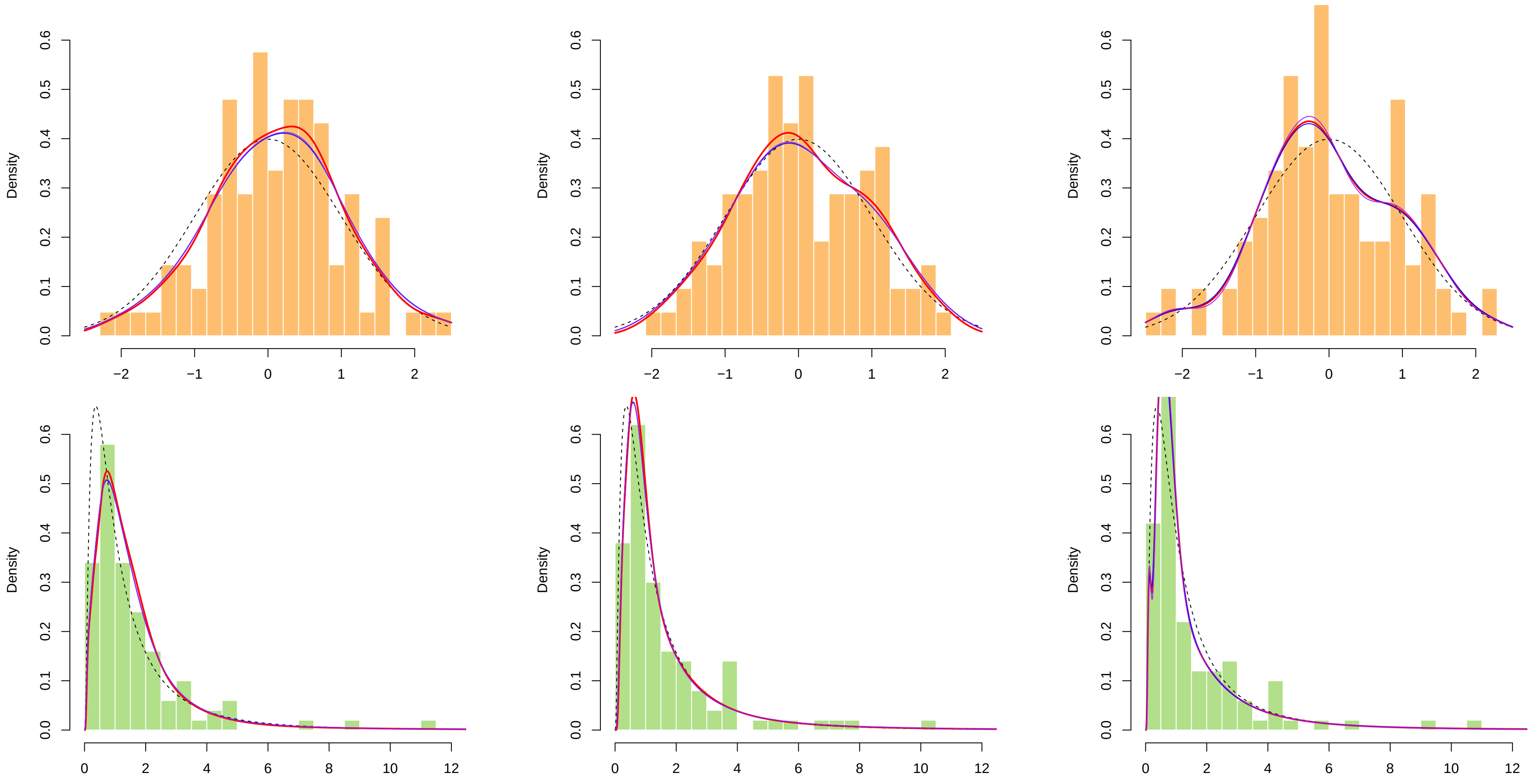


Illustration with Log-normal samples

$u_i = T_{\widehat{\theta}}(x_i)$ + Mixture of Beta distributionsl

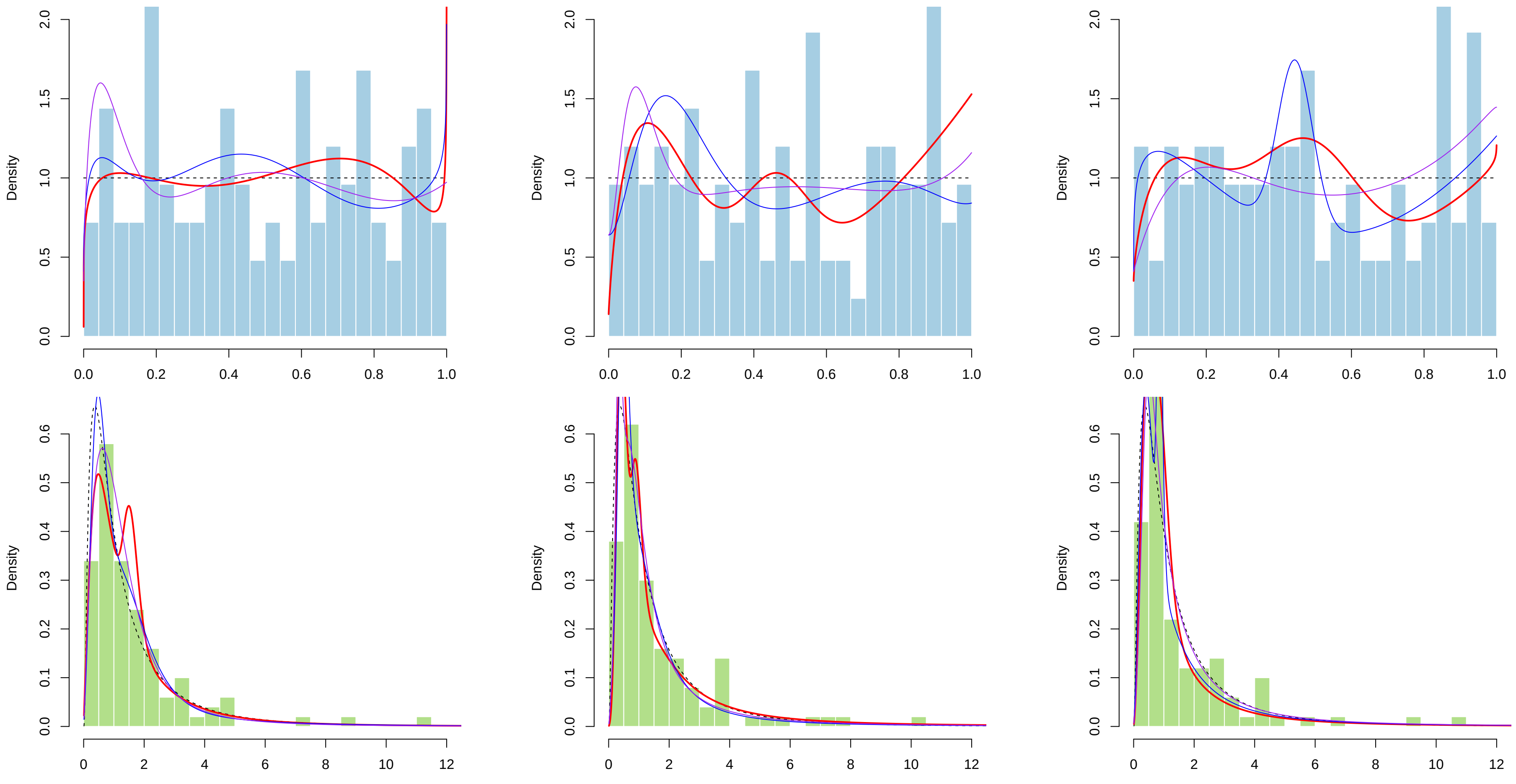
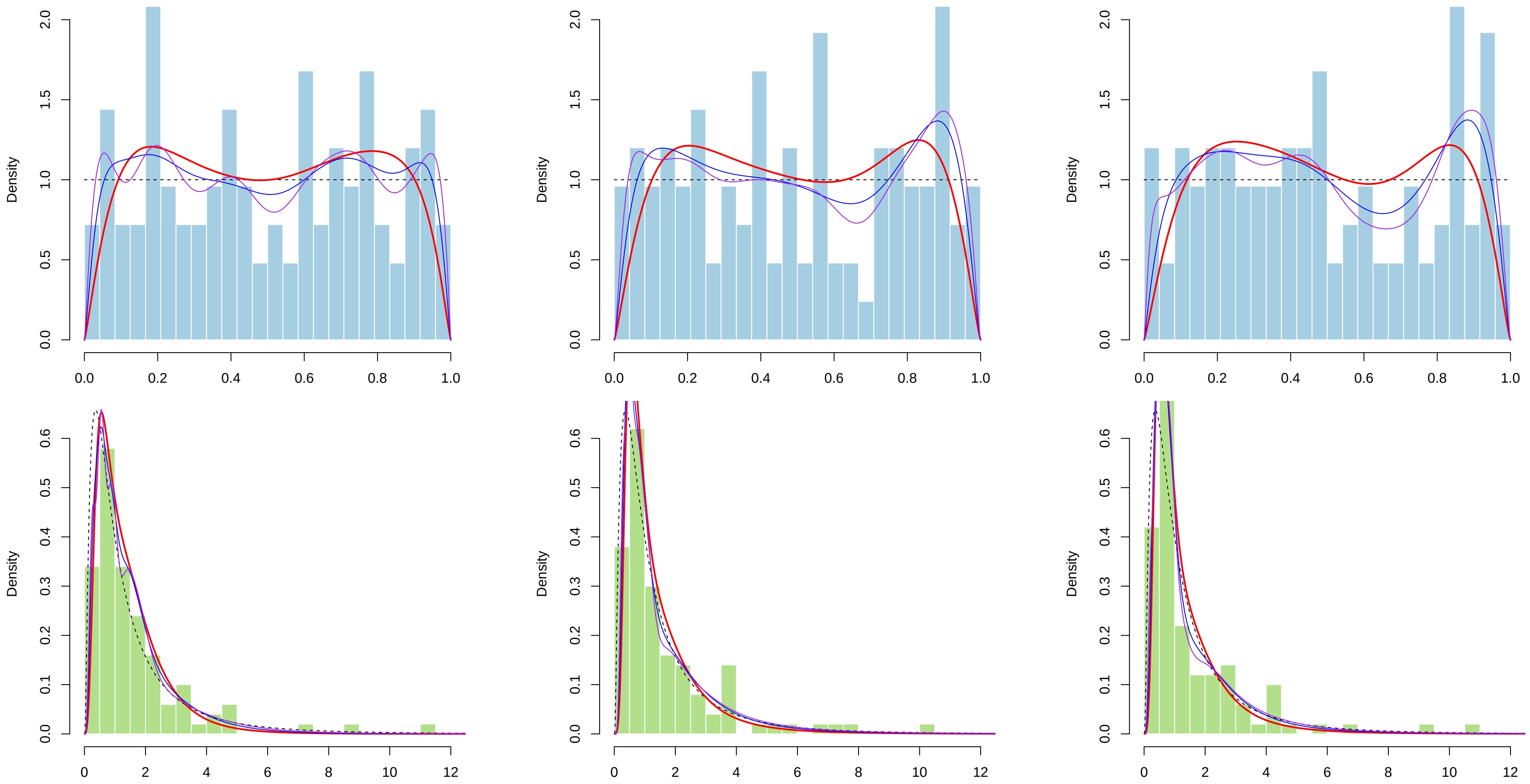


Illustration with Log-normal samples

$u_i = T_{\widehat{\theta}}(x_i)$ + Beta kernel estimation



Quantities of interest

Standard statistical quantities

- miae, $\left(\int_0^\infty \left|\widehat{f}_n(x) - f(x)\right| \,\mathrm{d}x\right)$
- mise, $\left(\int_0^\infty \left[\widehat{f}_n(x) - f(x)\right]^2 \,\mathrm{d}x\right)$
- miae$_w$, $\left(\int_0^\infty \left|\widehat{f}_n(x) - f(x)\right| \, |x| \,\mathrm{d}x\right)$
- mise$_w$, $\left(\int_0^\infty \left[\widehat{f}_n(x) - f(x)\right]^2 x^2 \,\mathrm{d}x\right)$

Quantities of interest

Inequality indices and risk measures, based on $F(x)=\int_0^x f(t)dt$,

- Gini, $\dfrac{1}{\mu}\displaystyle\int_0^\infty F(t)[1-F(t)]dt$
- Theil, $\displaystyle\int_0^\infty \frac{t}{\mu}\log\left(\frac{t}{\mu}\right)f(t)dt$
- Entropy $-\displaystyle\int_0^\infty f(t)\log[f(t)]dt$
- VaR-quantile, x such that $F(x)=\mathbb{P}(X\leq x)=\alpha$, i.e. $F^{-1}(\alpha)$
- TVaR-expected shorfall, $\mathbb{E}[X|X>F^{-1}(\alpha)]$

where $\mu=\displaystyle\int_0^\infty[1-F(x)]dx$.

Computations Aspects

Here, for each method, we return two functions,

- function $\widehat{f}_n(\cdot)$
- a random generator for distribution $\widehat{f}_n(\cdot)$

∘ H-transform and Gaussian kernel

$$\text{draw } i \in \{1, \cdots, n\} \text{ and } X = H^{-1}(Z) \text{ where } Z \sim \mathcal{N}(H(x_i), b^2)$$

∘ H-transform and Beta kernel

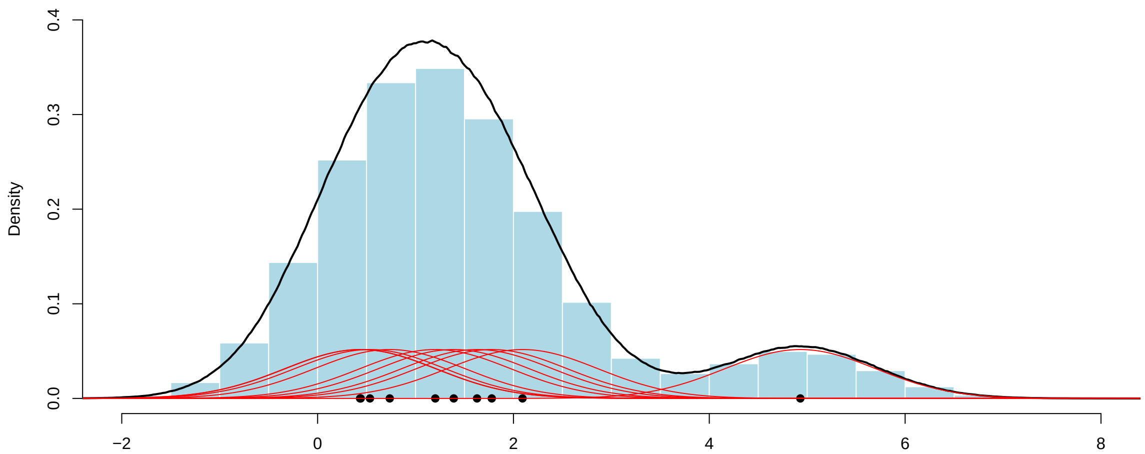
$$\text{draw } i \in \{1, \cdots, n\} \text{ and } X = H^{-1}(U) \text{ where } U \sim \mathcal{B}(H(x_i)/h, [1 - H(x_i)]/h)$$

∘ H-transform and Beta mixture

$$\text{draw } k \in \{1, \cdots, K\} \text{ and } X = H^{-1}(U) \text{ where } U \sim \mathcal{B}(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k)$$

◦ 'standard' Gaussian kernel (benchmark)

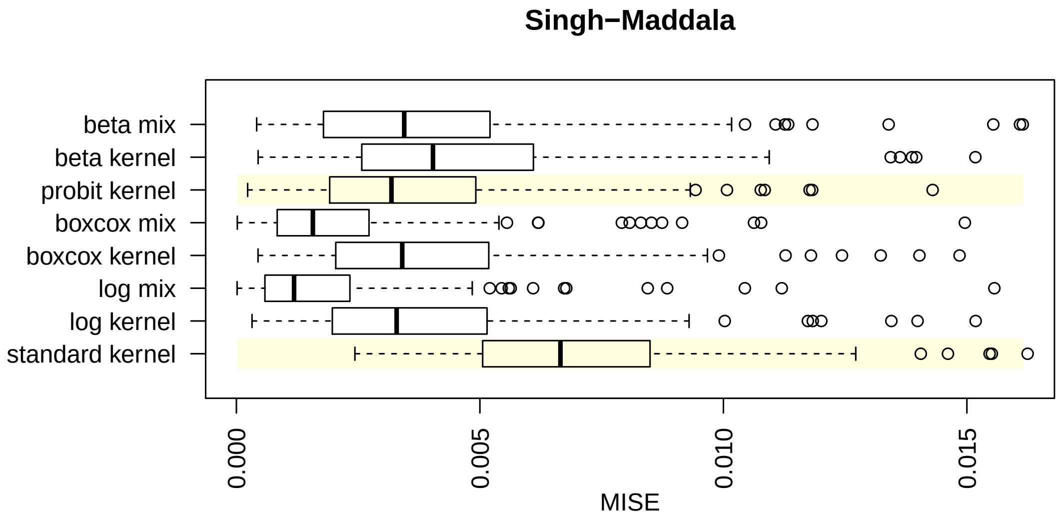
$$\text{draw } i \in \{1, \cdots, n\}, \text{ and } X \sim \mathcal{N}(x_i, b^2) \quad (almost)$$

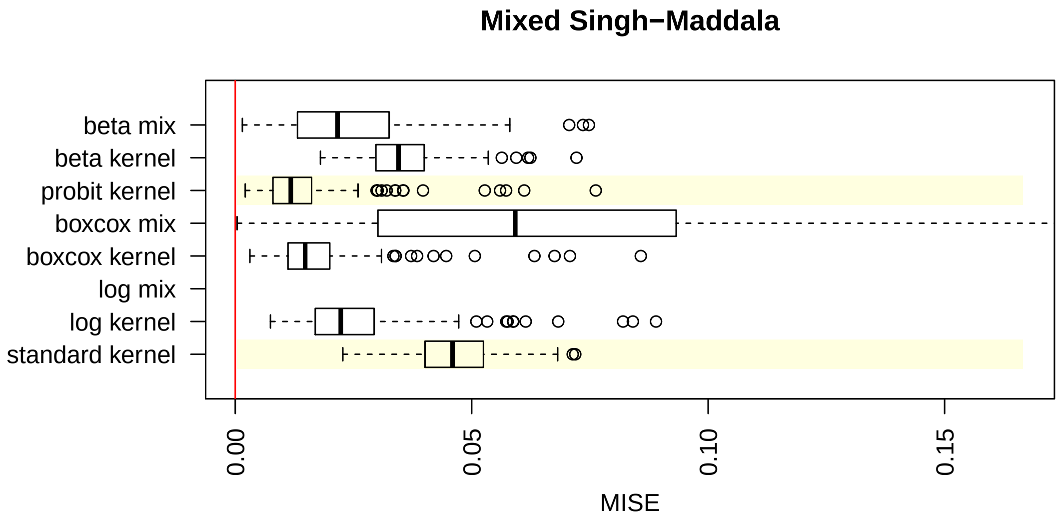


up to some normalization, $\cdot \mapsto \dfrac{\widehat{f}_n(\cdot)}{\int_0^\infty \widehat{f}_n(x)dx}$

MISE

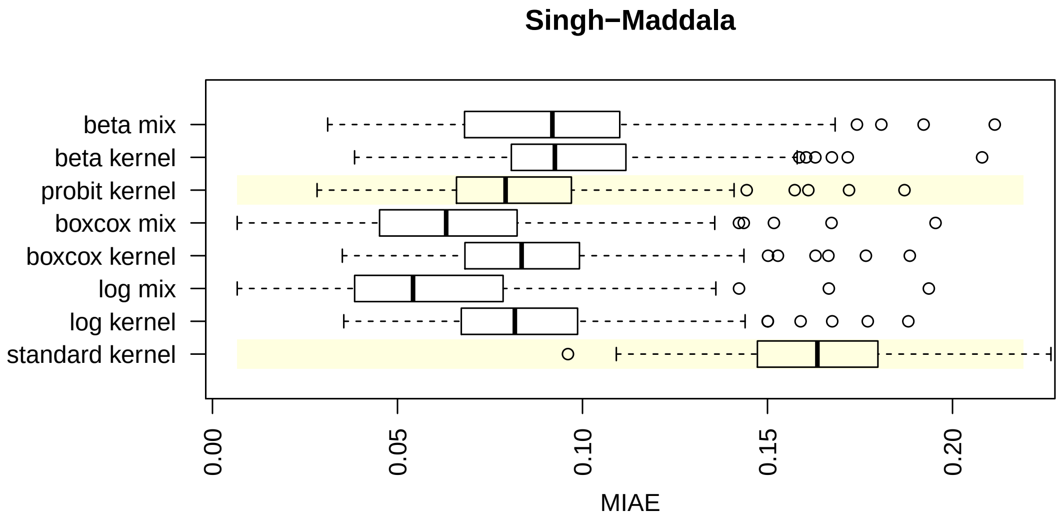
$$\int_0^\infty \left[\widehat{f}_n^{(s)}(x) - f(x)\right]^2 \, \mathrm{d}x$$

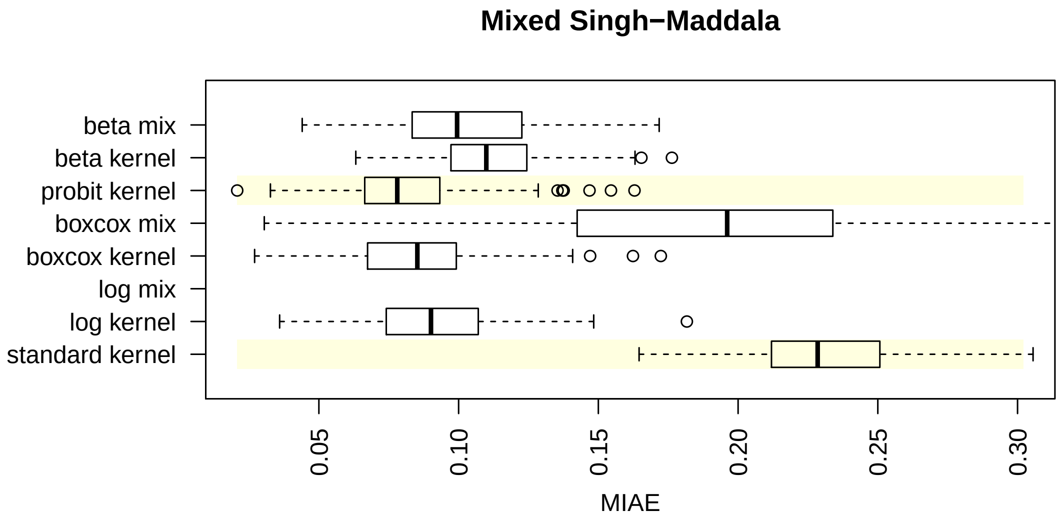




MIAE

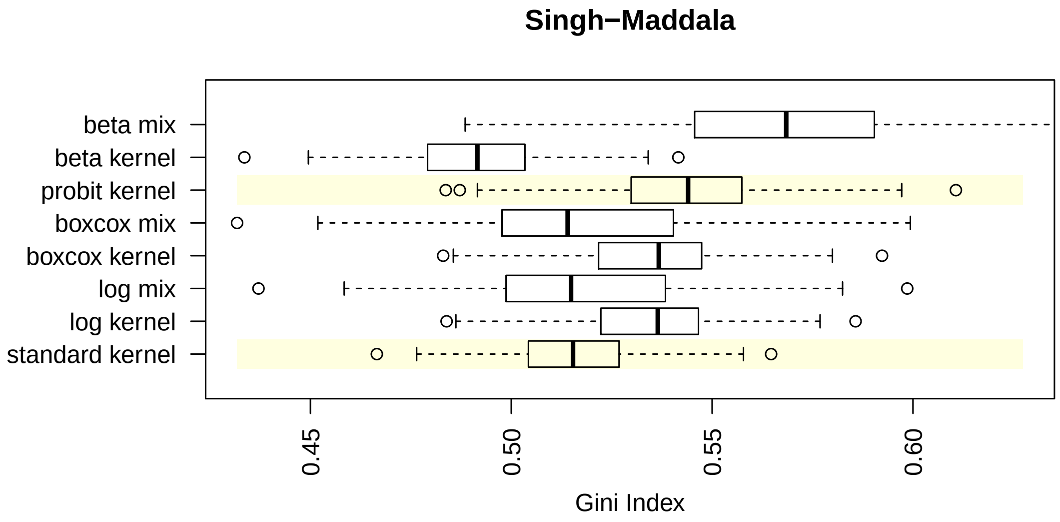
$$\int_0^\infty |\widehat{f}_n^{(s)}(x) - f(x)|\, \mathrm{d}x$$

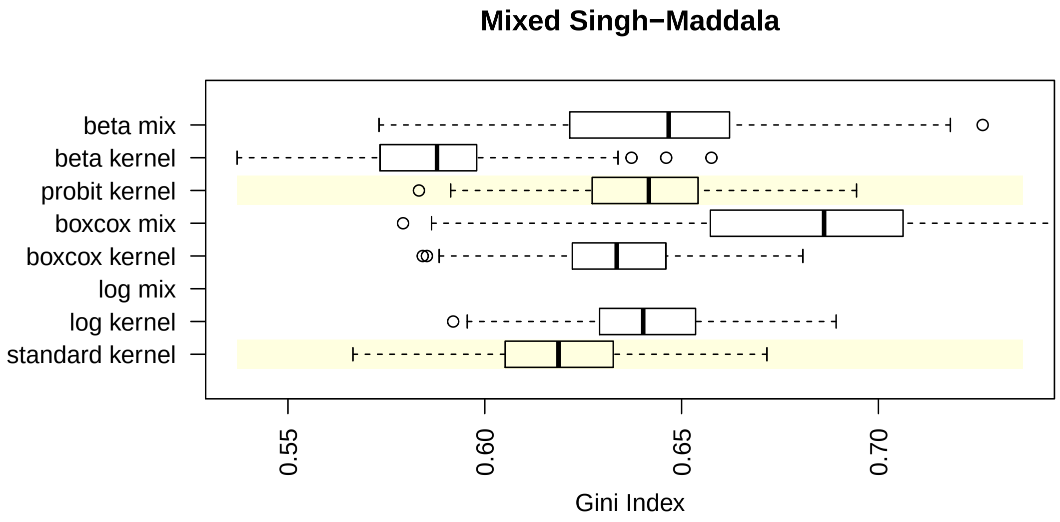




Gini Index

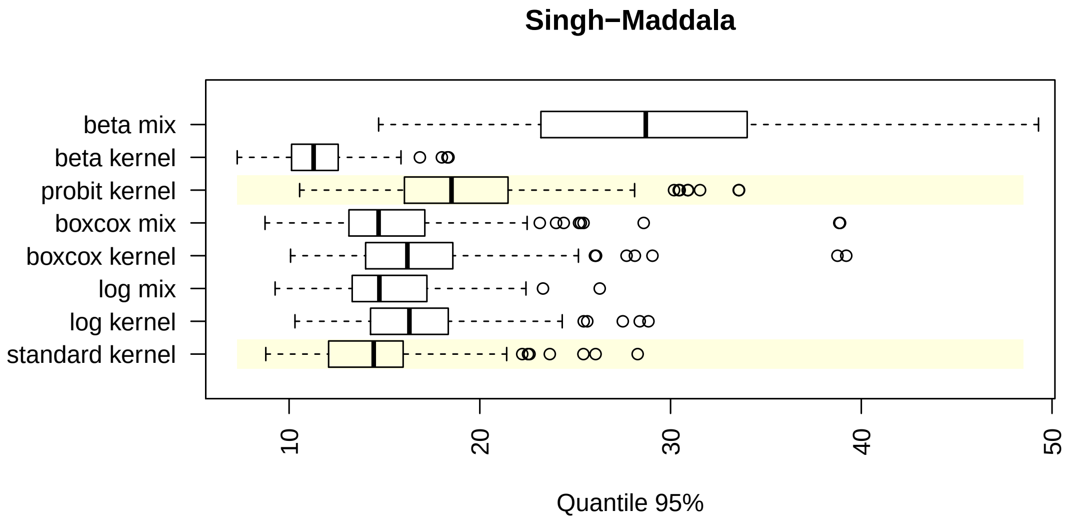
$$\frac{1}{\widehat{\mu}_n^{(s)}}\int_0^\infty \widehat{F}_n^{(s)}(t)[1-\widehat{F}_n^{(s)}(t)]dt$$

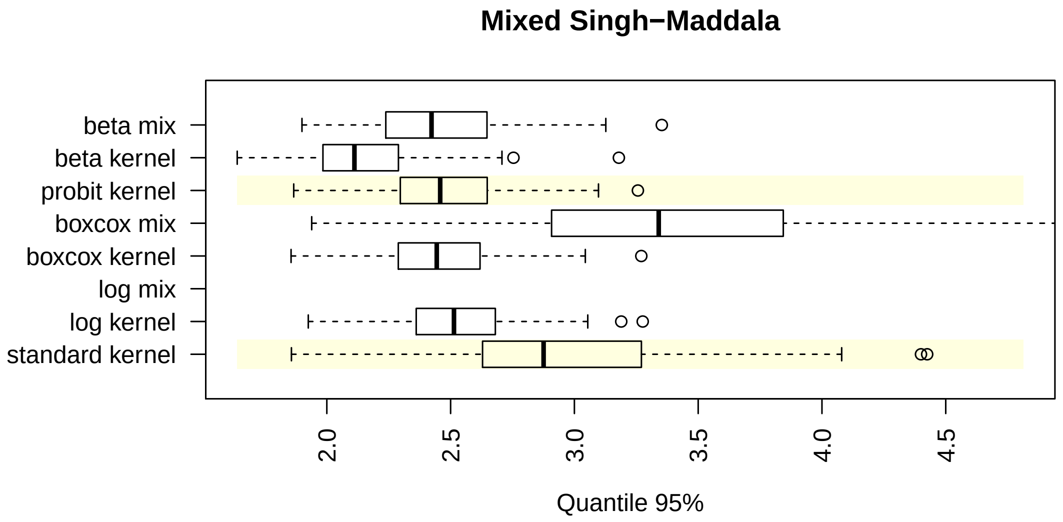




Value-at-Risk, 95%

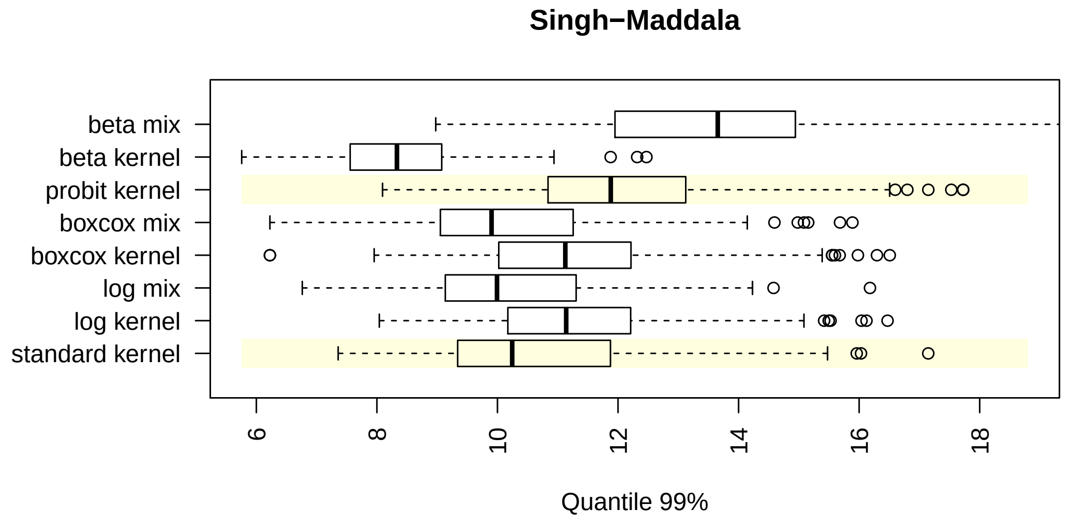
$$\widehat{Q}_n^{(s)}(\alpha) = \inf\{x, \alpha \leq \widehat{F}_n^{(s)}(x)\}$$

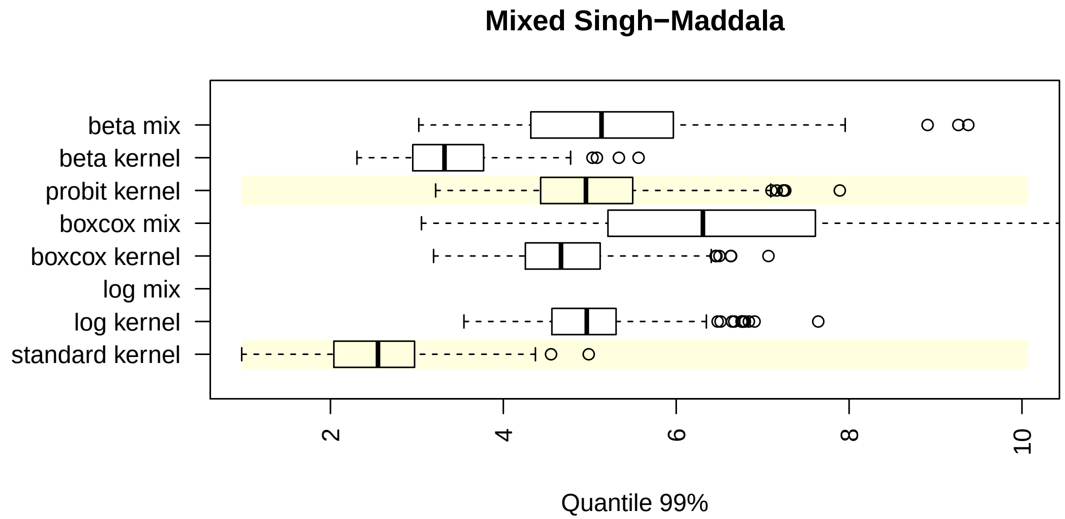




Value-at-Risk, 99%

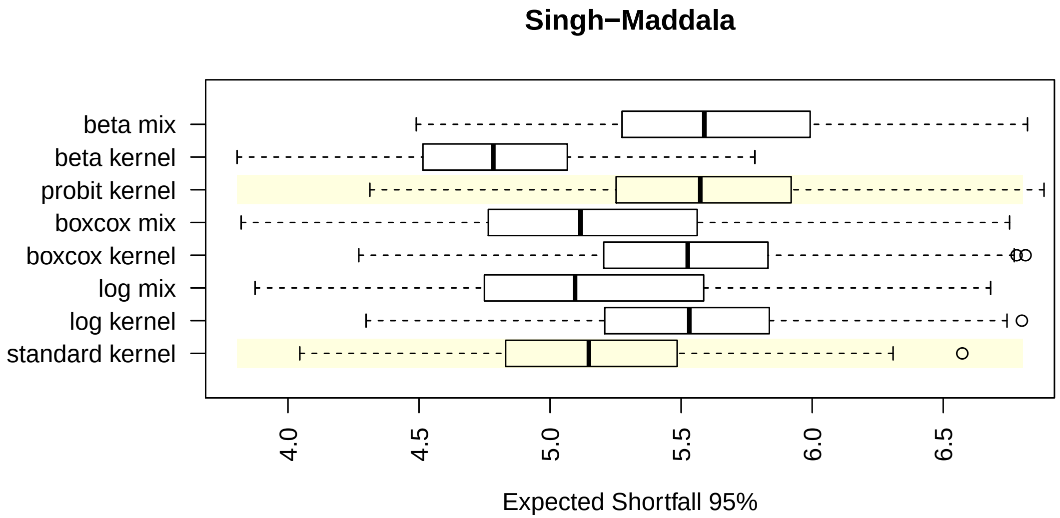
$$\widehat{Q}_n^{(s)}(\alpha) = \inf\{x, \alpha \leq \widehat{F}_n^{(s)}(x)\}$$

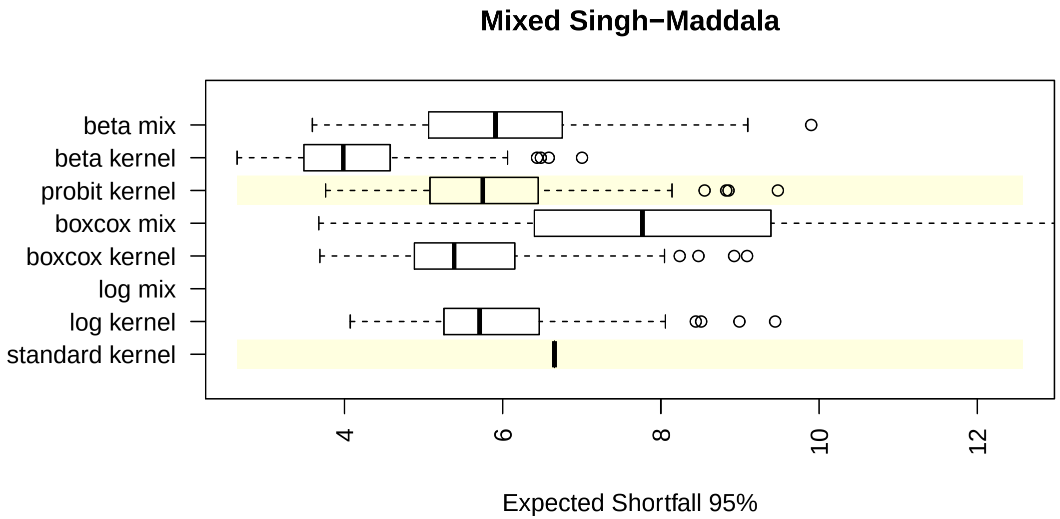




Tail Value-at-Risk, 95%

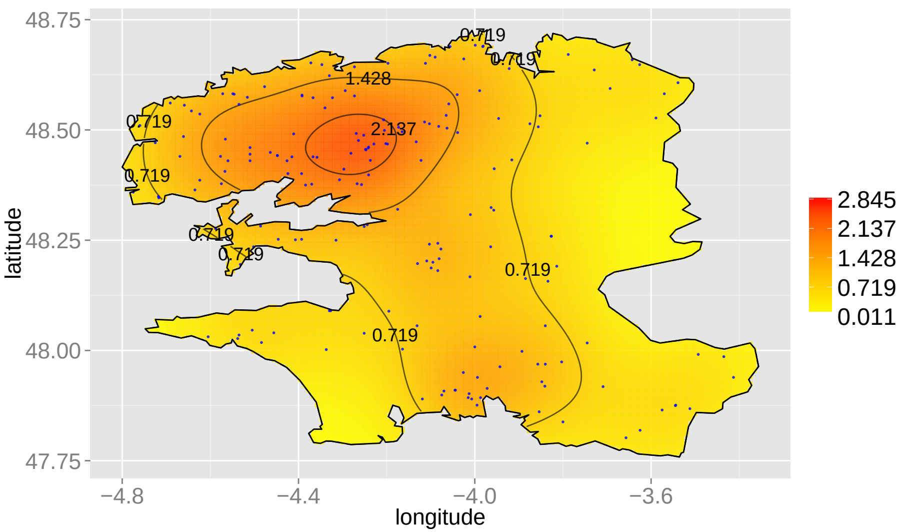
$$\mathbb{E}[X|X > \widehat{Q}_n^{(s)}(\alpha)]$$

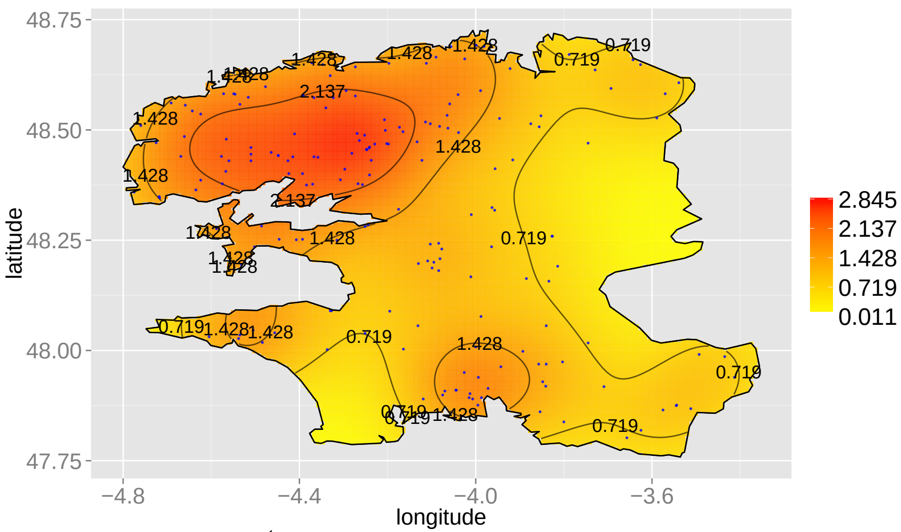




Possible conclusion ?

- estimating densities on transformated data is definitively a good idea
- but we need to find a good transformation
  - ✓ parametric + beta
  - ✓ parametric + probit
  - ✓ log-transform
  - ✓ Box-Cox





(joint work with **E. Gallic**)