Return Level Bounds for Extreme Values

J. Diebolt[1], A. Guillou[2], P. Naveau[3] and P. Ribereau[2]

(1) CNRS & Université Marne-la-Vallée

(2) Université Paris VI

(3) Laboratoire des Sciences du Climat et de l’Environnement

Work in progress

General framework

- $\{X_i\}_{i=1,\ldots,t}$ a sequence of non-negative iid random variables with continuous df F

General framework

- $\{X_i\}_{i=1,\dots,t}$ a sequence of non-negative iid random variables with continuous df F

  Question: What is the level z_t such that one may expect a single event larger that z occuring in t years ?

General framework

- $\{X_i\}_{i=1,\ldots,t}$ a sequence of non-negative iid random variables with continuous df F

  Question: What is the level z_t such that one may expect a single event larger that z occuring in t years ?

- This is equivalent to solve

$$\mathbb{E}\left(\sum_{i=1}^{t} \mathbb{1}(X_i > z_t)\right) = 1$$

with t given and z_t unknown

General framework

- $\{X_i\}_{i=1,\dots,t}$ a sequence of non-negative iid random variables with continuous df F

  Question: What is the level z_t such that one may expect a single event larger that z occuring in t years ?

- This is equivalent to solve

$$\mathbb{E}\left(\sum_{i=1}^{t} \mathbb{1}(X_i > z_t)\right) = 1$$

with t given and z_t unknown

$$\overline{F}(z_t) = 1/t$$

Objective

- To propose and study bounds for z_t

  - that can be easily estimated

  - whose properties are good and simply to derive

Objective

- To propose and study bounds for z_t

  - that can be easily estimated

  - whose properties are good and simply to derive

- We do not want to make any strong assumptions about the tail behavior of F

  $\Longrightarrow$ our bound may not always be necessarily close to z_t

Objective

- To propose and study bounds for z_t

  - that can be easily estimated

  - whose properties are good and simply to derive

- We do not want to make any strong assumptions about the tail behavior of F

  $\Longrightarrow$ our bound may not always be necessarily close to z_t

  The practitioner is often ready to overestimate the height of a dike if he knows that the upper bound is well estimated and can provide a reasonable “worst case scenario"

Extreme Value Theory

$$\hat{z}_t = \begin{cases} z^* + \frac{\hat{\sigma}}{\hat{\xi}} \left[\left(t\widehat{\overline{F}}(z^*) \right)^{\hat{\xi}} - 1 \right] & \text{if } \hat{\xi} \neq 0 \\ \\ z^* + \hat{\sigma} \log \left(t\widehat{\overline{F}}(z^*) \right) & \text{if } \hat{\xi} = 0 \end{cases}$$

(σ, ξ): scale and shape parameters of the GPD that is assume to fit $\mathbb{P}(X > z | X > z^*)$

Markov's inequality

$$\overline{F}(z) \leq \frac{\mathbb{E}[h(X)]}{h(z)} \qquad \text{for any } z > 0$$

Markov's inequality

$$\overline{F}(z) \leq \frac{\mathbb{E}[h(X)]}{h(z)} \qquad \text{for any } z > 0$$

Examples:

Markov's inequality

$$\overline{F}(z) \leq \frac{\mathbb{E}[h(X)]}{h(z)} \qquad \text{for any } z > 0$$

Examples:

- $h(z) = \exp(\lambda z)$ with $\lambda \geq 0$ leads to the famous *Chernoff bound*

$$\inf_{\lambda \geq 0} \left\{ e^{-\lambda z} \int e^{\lambda x} dF(x) \right\}$$

Markov's inequality

$$\overline{F}(z) \leq \frac{\mathbb{E}[h(X)]}{h(z)} \qquad \text{for any } z > 0$$

Examples:

- $h(z) = \exp(\lambda z)$ with $\lambda \geq 0$ leads to the famous *Chernoff bound*

$$\inf_{\lambda \geq 0} \left\{ e^{-\lambda z} \int e^{\lambda x} dF(x) \right\}$$

- $h(z) = z^n$ for every integer n yields to the *moment bound*

$$\inf_{n \geq 0} \frac{\mathbb{E}(X^n)}{z^n}$$

Markov's inequality (cont'd)

- Within the context of return levels:

$$h(z) = u(z)v(F(z))$$

$$(u, v) : \mathbb{R}^+ \times [0, 1] \to \mathbb{R}^+ \times \mathbb{R}^+$$

Bound

$$\Theta(u, v) := \mathbb{E}[h(X)] = \mathbb{E}[u(X)v(F(X))]$$

$$u(x) = x^r, v(x) = x^s \longrightarrow \text{PWM}$$

Generalized PWM: see e.g. Diebolt et al. (2003, 2004)

Bound

$$\Theta(u,v) := \mathbb{E}[h(X)] = \mathbb{E}[u(X)v(F(X))]$$

$$u(x) = x^r, v(x) = x^s \longrightarrow \text{PWM}$$

- z_t defined as $\overline{F}(z_t) = 1/t$:

$$z_t \le b_t(u,v) = u^{\leftarrow}\left[\frac{t\Theta(u,v)}{v(1-1/t)}\right] \qquad \text{where } u \uparrow \text{ and } v \uparrow$$

$$\Longrightarrow z_t \le \inf\left\{b_t(u,v) : u(.) \text{ and } v(.) \text{ non-decreasing}\right\}$$

Generalized PWM: see e.g. Diebolt et al. (2003, 2004)

Other bounds

- $v(.)$ is still non-decreasing but $u(.)$ is now non-increasing

$$
\begin{aligned}
\Theta^*(u,v) &= \mathbb{E}\left[u(X)v(\overline{F}(X))\mathbb{1}(X > z_t)\right] \\
&\quad + \mathbb{E}\left[u(X)v(\overline{F}(X))\mathbb{1}(X \leq z_t)\right] \\
&= T_1 + T_2
\end{aligned}
$$

Other bounds

- $v(.)$ is still non-decreasing but $u(.)$ is now non-increasing

$$
\begin{aligned}
\Theta^*(u,v) &= \mathbb{E}\left[u(X)v(\overline{F}(X))\mathbb{1}(X > z_t)\right] \\
&\quad + \mathbb{E}\left[u(X)v(\overline{F}(X))\mathbb{1}(X \leq z_t)\right] \\
&= T_1 + T_2
\end{aligned}
$$

$$
T_1 \leq \frac{u(z_t)v(1/t)}{t} \quad \text{and} \quad T_2 \leq \Big[\Theta^*(u^p, v^p)\Big]^{1/p} (1 - 1/t)^{1/q}
$$

Other bounds

- $v(.)$ is still non-decreasing but $u(.)$ is now non-increasing

$$\begin{aligned}\Theta^*(u,v) &= \mathbb{E}\left[u(X)v(\overline{F}(X))\mathbb{1}(X > z_t)\right] \\ &\quad + \mathbb{E}\left[u(X)v(\overline{F}(X))\mathbb{1}(X \leq z_t)\right] \\ &= T_1 + T_2\end{aligned}$$

$$T_1 \leq \frac{u(z_t)v(1/t)}{t} \quad \text{and} \quad T_2 \leq \left[\Theta^*(u^p, v^p)\right]^{1/p} (1-1/t)^{1/q}$$

$$z_t \leq b_t^*(u,v) = u^{\leftarrow}\left[t\frac{\Theta^*(u,v) - (1-1/t)^{1-1/p}\left[\Theta^*(u^p,v^p)\right]^{1/p}}{v(1/t)}\right]$$

Lower bounds

- $v(.)$ is now non-increasing and $u(.)$ non-decreasing

$$T_2 \leq (1-1/t)u(z_t)v(1/t) \quad \text{and} \quad T_1 \leq t^{-1/q}\left[\Theta^*(u^p, v^p)\right]^{1/p}$$

Lower bounds

- $v(.)$ is now non-increasing and $u(.)$ non-decreasing

$$T_2 \le (1 - 1/t)u(z_t)v(1/t) \quad \text{and} \quad T_1 \le t^{-1/q}\left[\Theta^*(u^p, v^p)\right]^{1/p}$$

$$z_t \ge l_t^*(u,v) = u^{\leftarrow}\left[\frac{\Theta^*(u,v) - t^{1/p-1}\left[\Theta^*(u^p,v^p)\right]^{1/p}}{v(1/t)(1-1/t)}\right]$$

$$z_t \ge \sup\left\{l_t^*(u,v) : u(.) \ge 0,\ v(.) \ge 0,\ u(.)\uparrow,\ v(.)\downarrow \text{ and } p > 1\right\}$$

Estimators

- $\Theta(u, v)$ and $\Theta^*(u, v)$ have to be estimated efficiently

Estimators

- $\Theta(u, v)$ and $\Theta^*(u, v)$ have to be estimated efficiently
- the choice of the functions $u(.)$ and $v(.)$ is fundamental

Estimators

- $\Theta(u, v)$ and $\Theta^*(u, v)$ have to be estimated efficiently
- the choice of the functions $u(.)$ and $v(.)$ is fundamental

$$\Theta(u, v) = \mathbb{E}[u(X)v(F(X))] = \int_0^\infty u'(x)V(F(x))dx$$

Estimators

- $\Theta(u, v)$ and $\Theta^*(u, v)$ have to be estimated efficiently
- the choice of the functions $u(.)$ and $v(.)$ is fundamental

$$\Theta(u, v) = \mathbb{E}[u(X)v(F(X))] = \int_0^\infty u'(x)V(F(x))dx$$

$$\widehat{\Theta}_n(u, v) = \int_0^\infty u'(x)V(\mathbb{F}_n(x))dx$$

Estimators

- $\Theta(u,v)$ and $\Theta^*(u,v)$ have to be estimated efficiently
- the choice of the functions $u(.)$ and $v(.)$ is fundamental

$$\Theta(u,v) = \mathbb{E}[u(X)v(F(X))] = \int_0^\infty u'(x)V(F(x))dx$$

$$\widehat{\Theta}_n(u,v) = \int_0^\infty u'(x)V(\mathbb{F}_n(x))dx$$

$$\widehat{\Theta}^*_n(u,v) = \int_0^\infty u'(x)V(\overline{\mathbb{F}}_n(x))dx$$

Theorem 1

$$\widehat{\Theta}_n(u,v) - \Theta(u,v) = O\left(\frac{\sqrt{2\log\log n}}{\sqrt{n}}\right) \quad \text{a.s.}$$

$$\widehat{\Theta}^*_n(u,v) - \Theta(u,v) = O\left(\frac{\sqrt{2\log\log n}}{\sqrt{n}}\right) \quad \text{a.s.}$$

Theorem 2

Assumptions on $u(.)$ and $v(.)$:

- $u > 0$, $v > 0$ and both $\uparrow$, $u \in \mathcal{C}^1[0, \infty)$ and $v \in \mathcal{C}^1[0, 1]$
- $u(0) = 0$, $u'(x) = O(x^{a-1})$, $a > 0$

Assumptions on $F(.)$:

- F admits a density $f > 0$ on $[0, \infty)$, $F(0) = 0$
- $\dfrac{x^{a-1}\,[1 - F(x)]^{\frac{3}{2}-\zeta}}{f(x)} \leq C$, $0 < \zeta < 1$, $x \sim \infty$

we have

$$\sqrt{n}\Big(\widehat{\Theta}_n(u, v) - \Theta(u, v)\Big) \xrightarrow{d} -\int_0^\infty u'(x)\mathbb{B}(F(x))v(F(x))dx$$

Theorem 3

Assumptions on $u(.)$ and $v(.)$:

- $u > 0$, $v > 0$, $u \downarrow$ and $v \uparrow$, $u \in \mathcal{C}^1[0, \infty)$ and $v \in \mathcal{C}^1[0, 1]$
- $u(\infty) = 0$, $u(0) < \infty$, $u'(t) = O\left(t^{-\frac{3}{2}+\xi}\right), 0 < \xi < 1, t \sim 0$

Assumptions on $F(.)$:

- F admits a density $f > 0$ on $[0, \infty)$, $F(0) = 0$
- $\dfrac{[1 - F(t)]^{\frac{3}{2}-\zeta}}{f(t)} = O\left(\dfrac{1}{|u'(t)|}\right), 0 < \zeta < 1, t \sim \infty$

we have

$$\sqrt{n}\Big(\widehat{\Theta}_n^*(u, v) - \Theta^*(u, v)\Big) \xrightarrow{d} - \int_0^\infty u'(x)\mathbb{B}(\overline{F}(x))v(\overline{F}(x))dx$$

Theorem 4

Assumptions on $u(.)$ and $v(.)$:

- $u > 0$, $v > 0$ and both $\uparrow$, $u \in \mathcal{C}^1[0, \infty)$ and $v \in \mathcal{C}^1[0, 1]$
- $u(0) = 0$, $u'(x) = O(x^{a-1})$, $a > 0$

Assumptions on $F(.)$:

- F admits a density $f > 0$ on $[0, \infty)$, $F(0) = 0$
- $\dfrac{x^{a-1}\,[1 - F(x)]^{\frac{3}{2} - \zeta - \nu}}{f(x)} \leq C$, $0 < \zeta < 1, 0 < \nu < 1/4$, $x \sim \infty$

we have for $\beta_n = o\left(n^{\min(\nu,\zeta)}\right)$

$$\beta_n \left[\sqrt{n}\Big(\widetilde{\widehat{\Theta}}_n(u, v) - \Theta(u, v)\Big) + \int_0^\infty u'(x)\mathbb{B}(F(x))v(F(x))dx \right] = o_{\mathbb{P}}(1)$$

Theorem 5

Assumptions on $u(.)$ and $v(.)$:

- $u > 0$, $v > 0$, $u \downarrow$ and $v \uparrow$, $u \in \mathcal{C}^1[0,\infty)$ and $v \in \mathcal{C}^1[0,1]$
- $u(\infty) = 0$, $u(0) < \infty$, $u'(t) = O\left(t^{-\frac{3}{2}+\xi+\nu}\right), 0 < \xi < 1, t \sim 0$

Assumptions on $F(.)$:

- F admits a density $f > 0$ on $[0,\infty)$, $F(0) = 0$
- $\dfrac{[1-F(t)]^{\frac{3}{2}-\zeta-\nu}}{f(t)} = O\left(\dfrac{1}{|u'(t)|}\right), 0 < \zeta < 1, t \sim \infty$

we have for $\beta_n^* = o\left(n^{\min(\nu,\zeta,\xi)}\right)$

$$\beta_n^*\left[\sqrt{n}\left(\widetilde{\widehat{\Theta}}_n^*(u,v) - \Theta^*(u,v)\right) + \int_0^\infty u'(x)\mathbb{B}(\overline{F}(x))v(\overline{F}(x))dx\right] = o_{\mathbb{P}}(1)$$

Simulation study

$$\Theta(u,v) = \mathbb{E}[u(X)v(F(X))] \quad \text{or} \quad \Theta^*(u,v) = \mathbb{E}[u(X)v(\overline{F}(X))]$$

$$z_t \leq \inf\left\{u^{\leftarrow}\left[\frac{t\Theta(u,v)}{v(1-1/t)}\right] : u(.)\uparrow \text{ and } v(.)\uparrow\right\}$$

$$z_t \geq \sup\left\{u^{\leftarrow}\left[\frac{\Theta^*(u,v) - t^{1/p-1}\left[\Theta^*(u^p,v^p)\right]^{1/p}}{v(1/t)(1-1/t)}\right] : u(.)\uparrow \text{ and } v(.)\downarrow\right\}$$

We tried different functions $u(.)$ and $v(.)$. The results below are given in case where $u(x) = x^a$ and $v(x) = x^b$ for the upper bound and $u(x) = x^a$ and $v(x) = x^{-b}$ for the lower bound

Simulation study (cont'd)

- Different distributions from the 3 domains of attraction
- We simulate 100 samples of size n from each distribution
- We compute

  - the median,

  - the root mean squared error (RMSE)

  - the median absolute deviation (MAD)

  - the true z_t

  - the theoretical upper and lower bounds

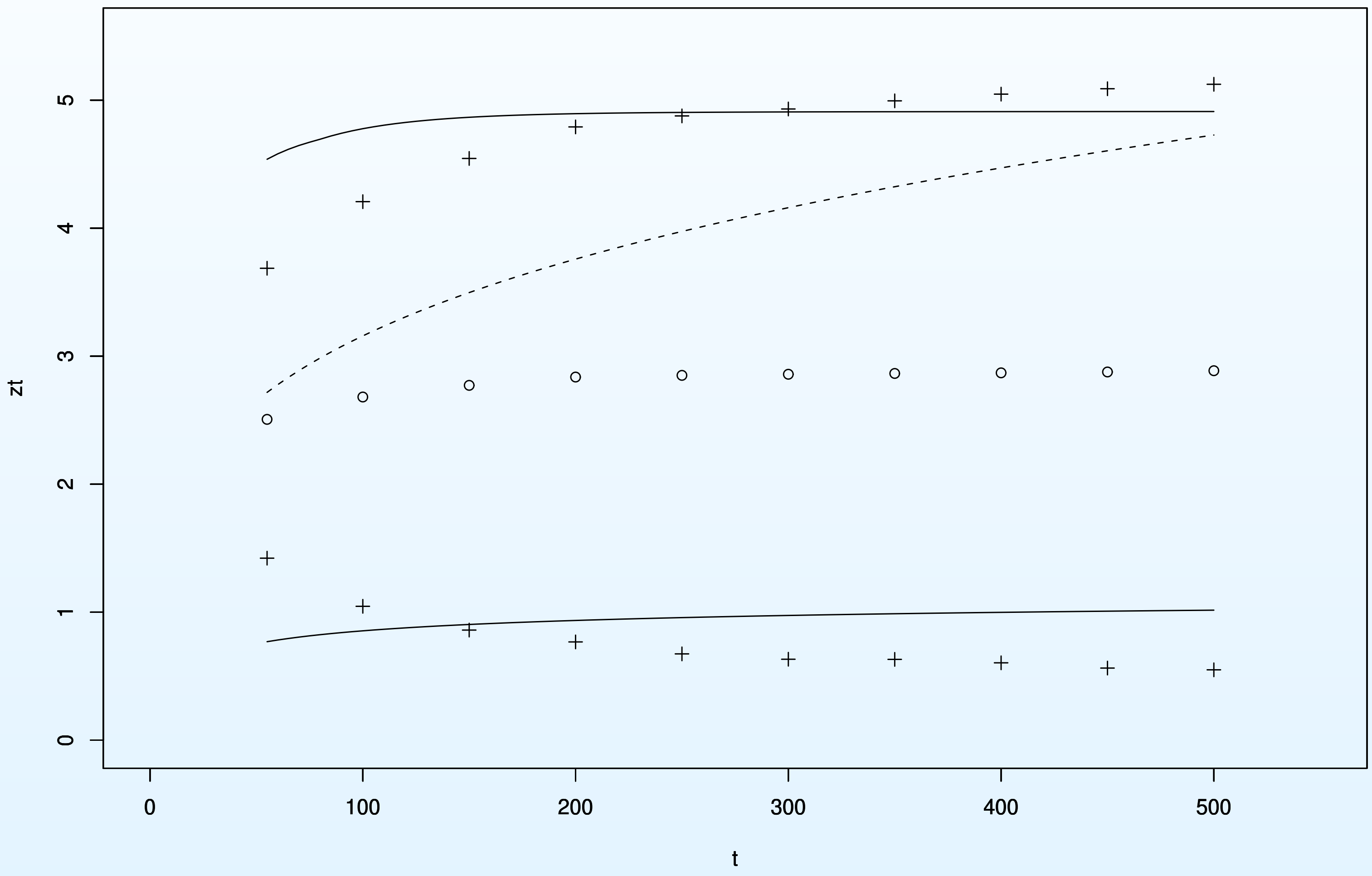
Simulation study (cont'd)

	RMSE	MAD	Median U.B.	Th. U.B.	True z_t
Fréchet	1.29	0.81	4.38	4.946	2.65
|Normal|	0.42	0.46	2.67	2.92	2.33
Beta	0.05	0.04	0.867	1.03	0.79

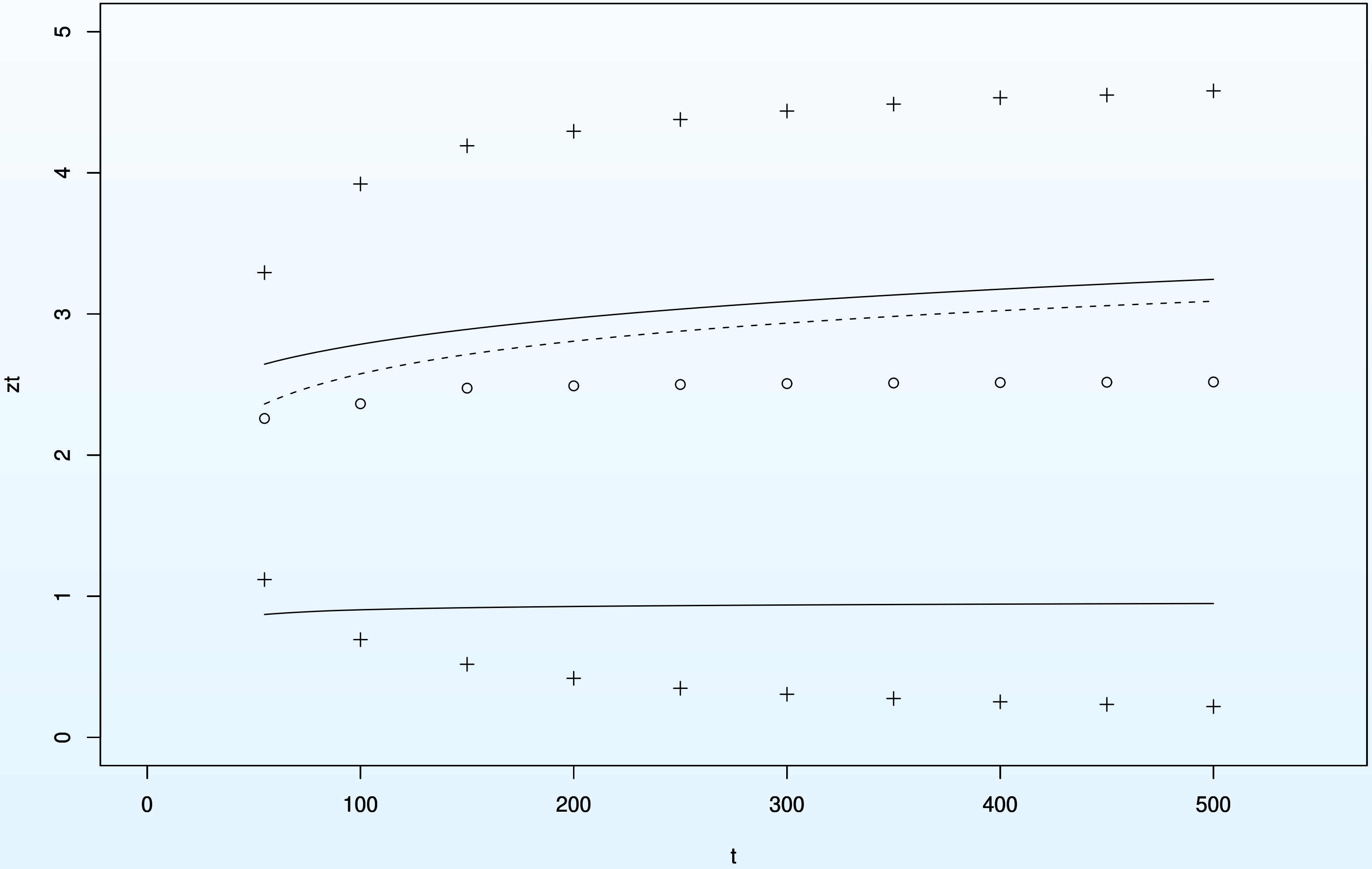
	RMSE	MAD	Median L.B.	Th. L.B.	True z_t
Fréchet	0.013	0.013	0.509	0.018	2.65
|Normal|	0.093	0.079	0.88	0.58	2.33
Beta	0.028	0.032	0.577	0.678	0.79

100 samples of size 50 with $t = 50$

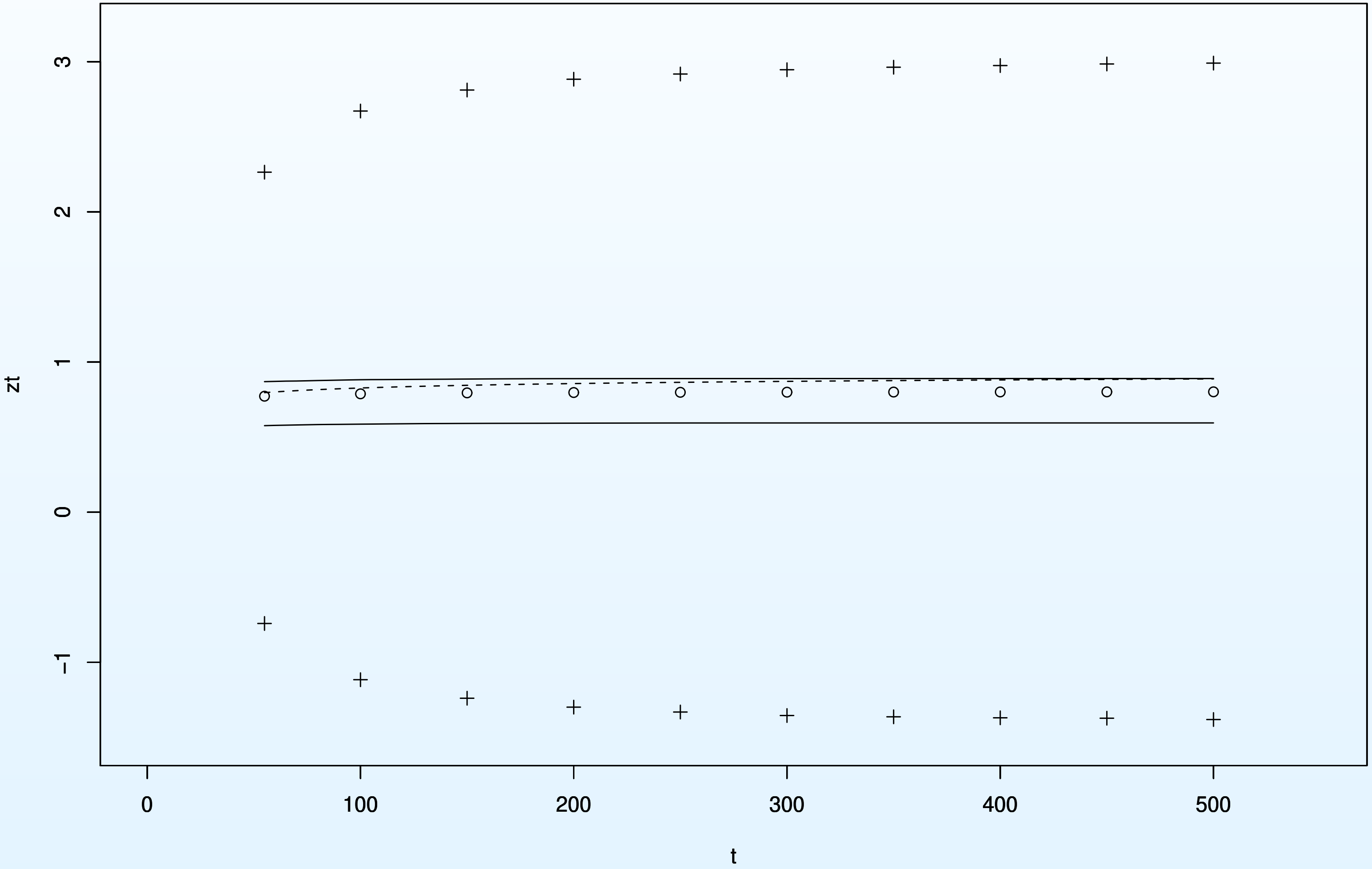
Fréchet distribution



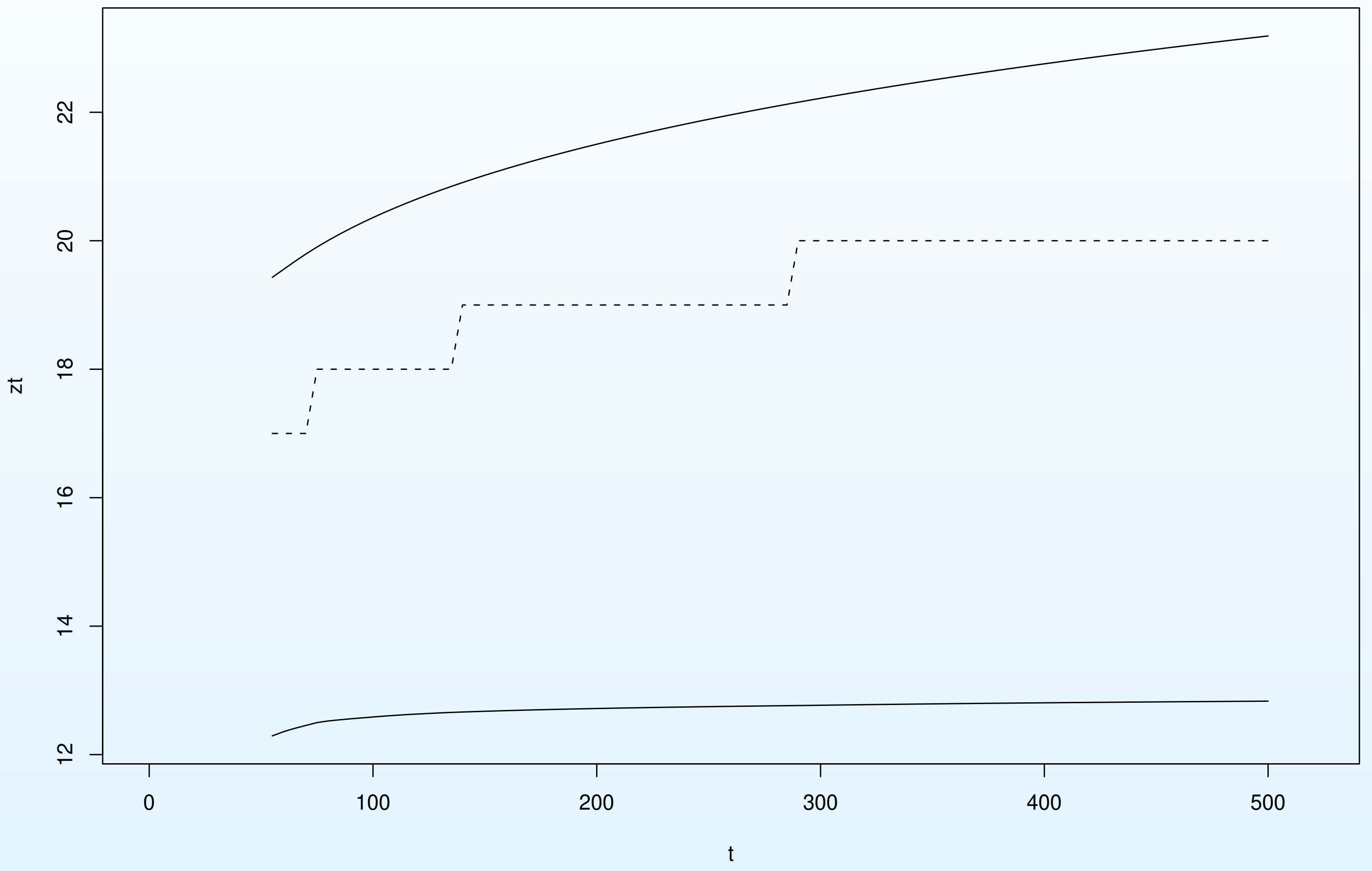
Normal distribution



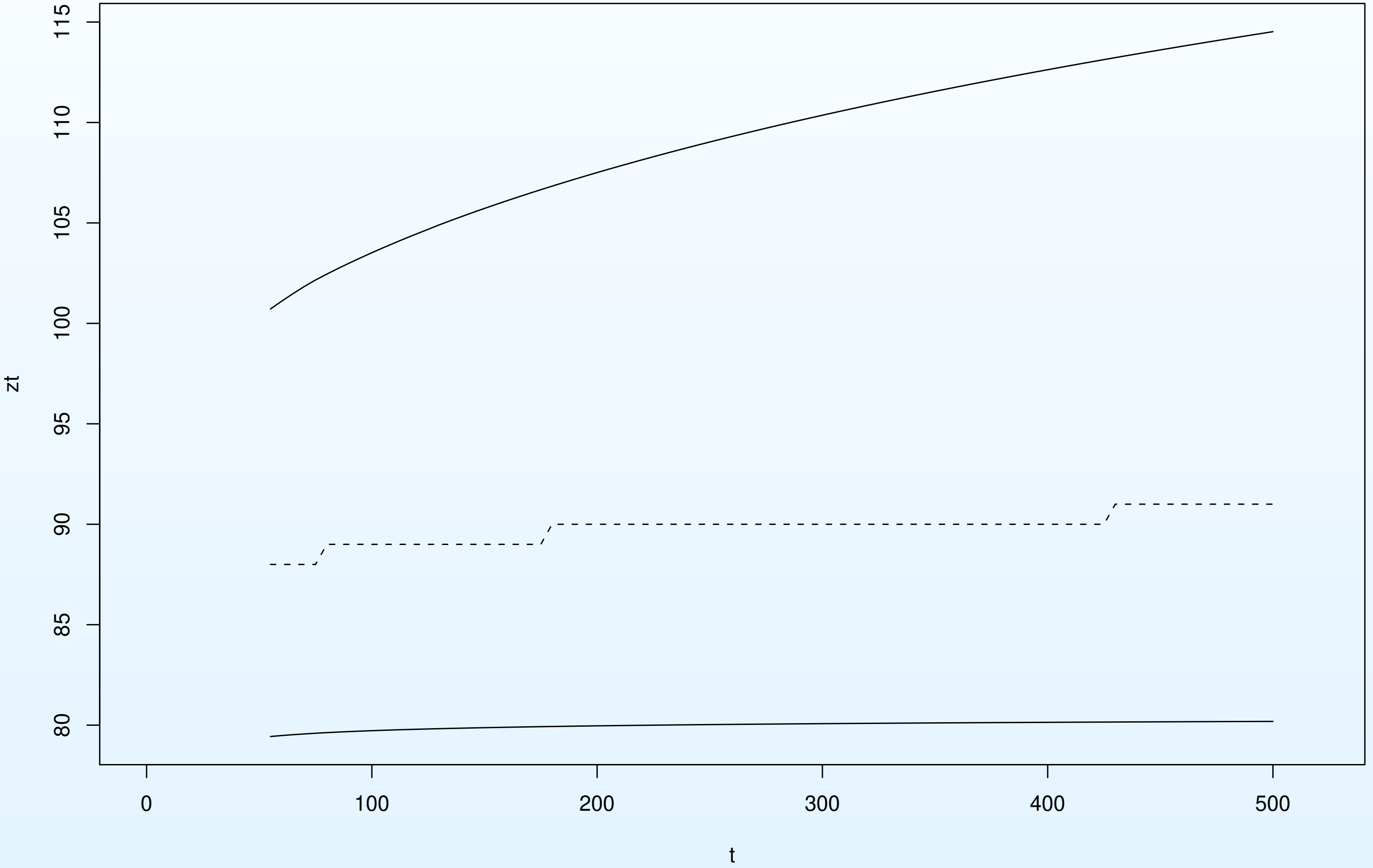
Beta distribution



Poisson distribution



Binomial distribution



Proof of Theorem 1

$$
\begin{aligned}
&\widehat{\Theta}_n(u,v) - \Theta(u,v) \\
&= -\int_0^\infty u'(x)\frac{\alpha_n(F(x))}{\sqrt{n}} v(F(x))dx \\
&\quad -\int_0^\infty u'(x)\int_0^1 (1-t)\frac{\alpha_n^2(F(x))}{n} v'\left(F(x) + t\frac{\alpha_n(F(x))}{\sqrt{n}}\right) dtdx \\
&= T_{1,n} + T_{2,n}
\end{aligned}
$$

Proof of Theorem 1 (cont'd)

- Let $\mathbb{H}$ be the Finkelstein set and q a continuous positive function on $[0,1]$, symmetric about $\frac{1}{2}$ and such that

$$q \nearrow \quad \text{and} \quad q(t)/\sqrt{t} \searrow \quad \text{on} \quad \left[0, \frac{1}{2}\right]$$

then if

$$\int_0^1 \frac{1}{q^2(t) \log\log[1/t(1-t)]} dt < \infty$$

$\dfrac{\alpha_n}{q\sqrt{2\log\log n}}$ is almost surely relatively compact with limit set

$$\mathbb{H}_q := \{h/q : h \in \mathbb{H}\}$$

Proof of Theorem 1 (cont'd)

$$q(t) = [t(1-t)]^{\frac{1}{2}-\nu}$$

$$|T_{1,n}| \leq \left\| \frac{\alpha_n}{q\sqrt{2\log\log n}} \right\| \frac{\sqrt{2\log\log n}}{\sqrt{n}} \int_0^\infty u'(x)q(F(x))v(F(x))dx$$

$$|T_{2,n}| \quad \leq \quad \left\| \frac{\alpha_n^2}{2q^2\log\log n} \right\| \frac{2\log\log n}{n} \int_0^\infty u'(x)q^2(F(x))dx$$

- If $\{f_n\}$ is almost surely relatively compact with limit set $\mathcal{F}$ then $\forall\,\Phi$ continuous, we have

$$\limsup_{n\to\infty} \Phi(f_n) = \sup_{f\in\mathcal{F}} \Phi(f) \quad \text{a.s.}$$

Proof of Theorem 2

$$
\begin{aligned}
&\widehat{\Theta}_n(u,v) - \Theta(u,v) \\
&= -\int_0^\infty u'(x)\frac{\alpha_n(F(x))}{\sqrt{n}}v(F(x))dx \\
&\quad -\int_0^\infty u'(x)\int_0^1 (1-t)\frac{\alpha_n^2(F(x))}{n}v'\left(F(x)+t\frac{\alpha_n(F(x))}{\sqrt{n}}\right)dtdx \\
&= T_{1,n} + T_{2,n}
\end{aligned}
$$

Proof of Theorem 2

$$
\begin{aligned}
&\widehat{\Theta}_n(u,v) - \Theta(u,v) \\
&= -\int_0^\infty u'(x)\frac{\alpha_n(F(x))}{\sqrt{n}} v(F(x))dx \\
&\quad - \int_0^\infty u'(x)\int_0^1 (1-t)\frac{\alpha_n^2(F(x))}{n} v'\left(F(x) + t\frac{\alpha_n(F(x))}{\sqrt{n}}\right) dt dx \\
&= T_{1,n} + T_{2,n}
\end{aligned}
$$

$$
\int_0^{1/n} \qquad \int_{1/n}^{1-1/n} \qquad \int_{1-1/n}^{1}
$$

Proof of Theorem 2

$$\widehat{\Theta}_n(u,v) - \Theta(u,v)$$

$$\begin{aligned} &= -\int_0^\infty u'(x)\frac{\alpha_n(F(x))}{\sqrt{n}}v(F(x))dx \\ &\quad -\int_0^\infty u'(x)\int_0^1 (1-t)\frac{\alpha_n^2(F(x))}{n}v'\left(F(x)+t\frac{\alpha_n(F(x))}{\sqrt{n}}\right)dtdx \\ &= T_{1,n} + T_{2,n} \end{aligned}$$

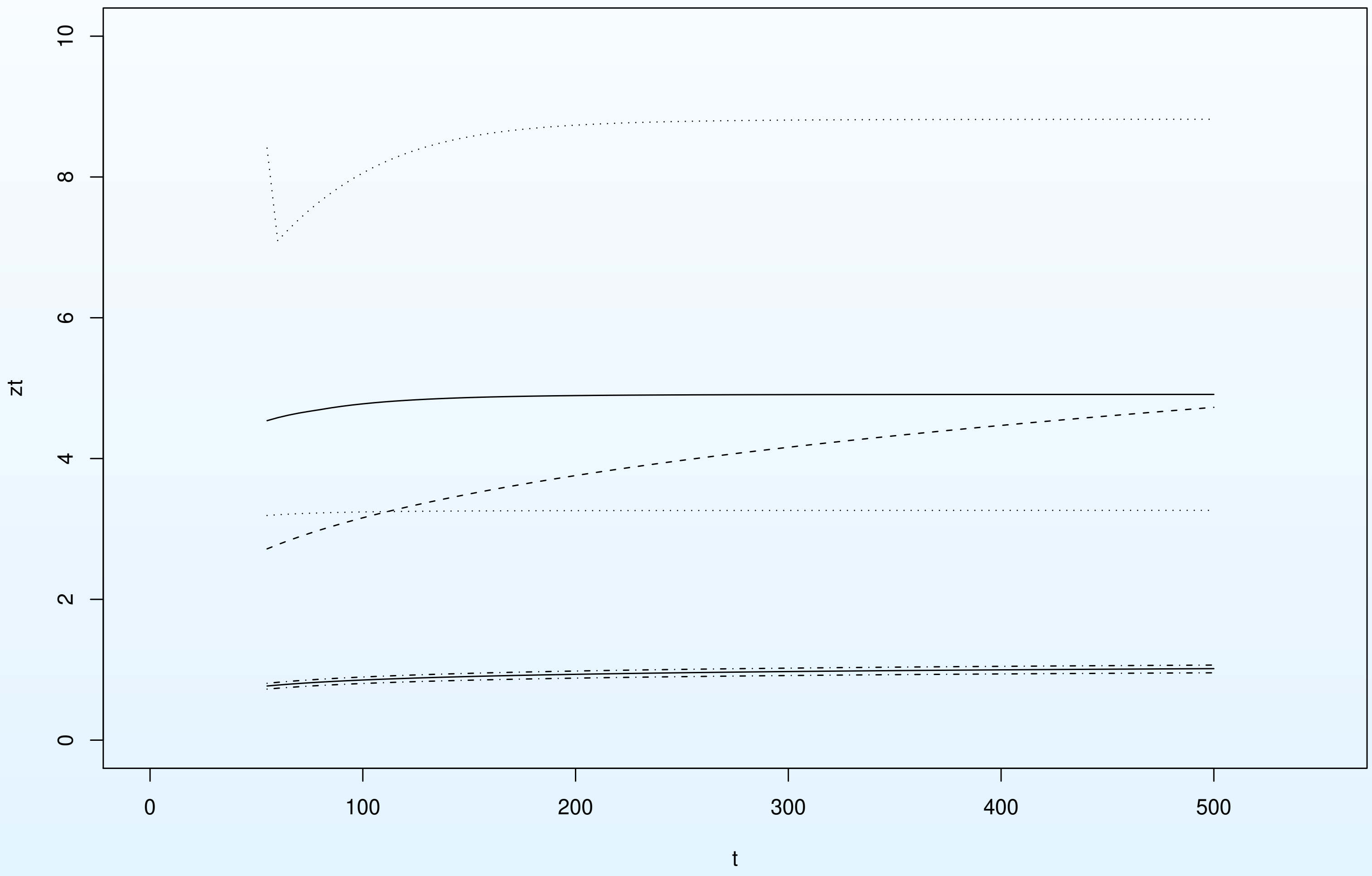
$$\int_0^{1/n} \qquad \int_{1/n}^{1-1/n} \qquad \int_{1-1/n}^{1}$$

$$T_{1,n} \xrightarrow{\mathbb{P}} -\int_0^1 u'\Big(F^{-1}(z)\Big)\mathbb{B}(z)v(z)\frac{dz}{f(F^{-1}(z))} \quad \text{and} \quad T_{2,n} \xrightarrow{\mathbb{P}} 0$$

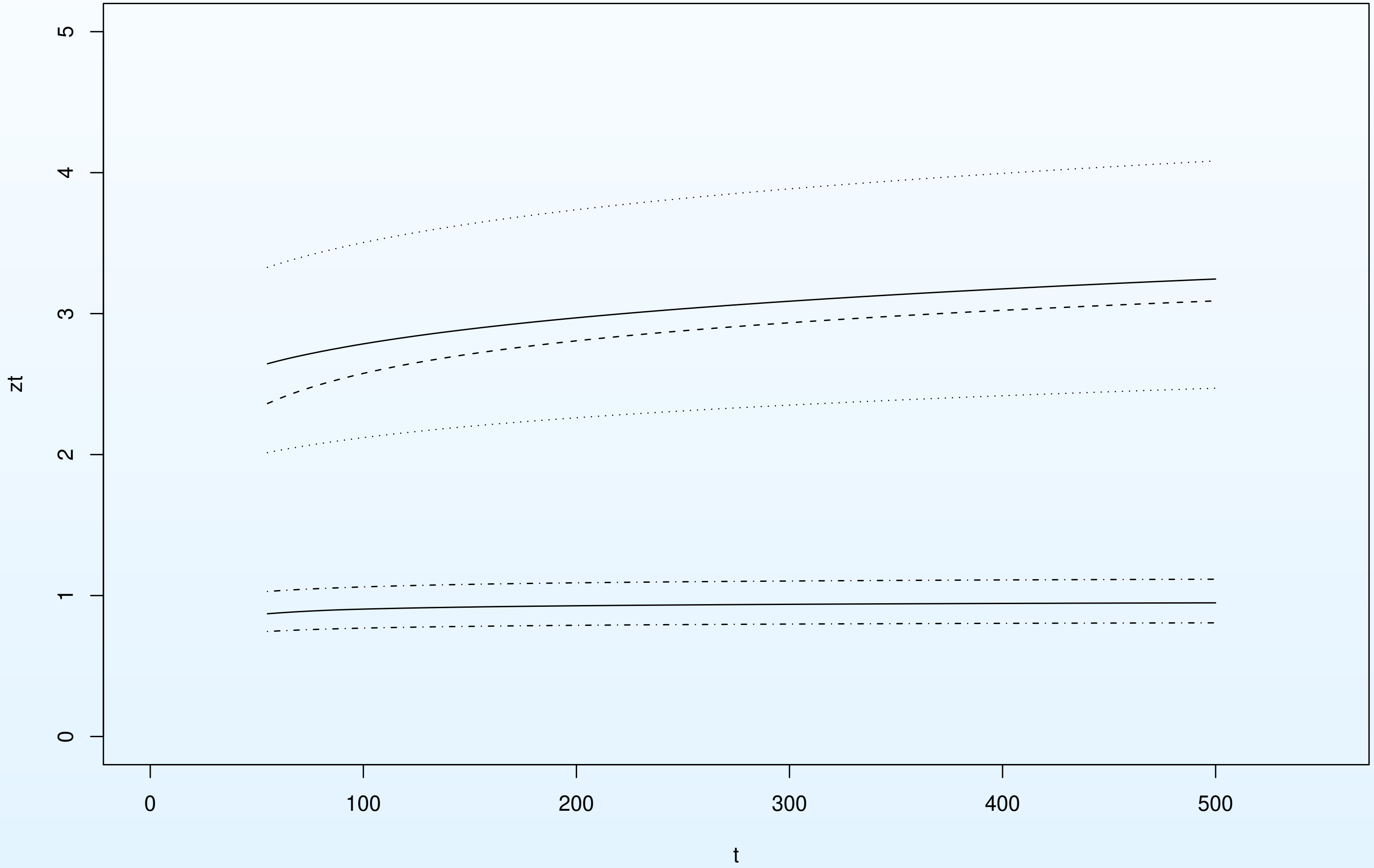
Concluding remarks

- Our bounds are easy to compute and lead to good results
- Our method is very general and can also be applied to mixture distributions, discrete random variables, ...
- Our bounds are valid for all values of n or all values of t
- These bounds are not "optimal" in the sense that we focus on a subfamily of functions u and v, but how to optimize ?
- Further research: stationarity ? bivariate case ? ...

Fréchet distribution



Normal distribution



Beta distribution