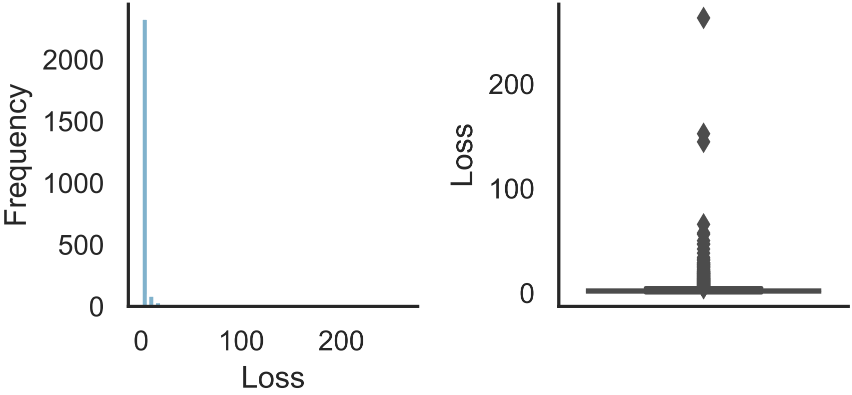
Sequential Monte Carlo samplers to fit and compare insurance loss models

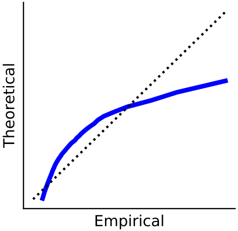
Pierre-O. Goffard

Institut de Science Financières et d'Assurances
`pierre-olivier.goffard@univ-lyon1.fr`

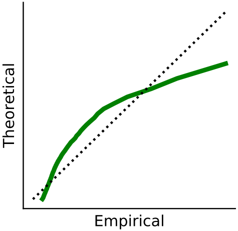
9 mars 2022



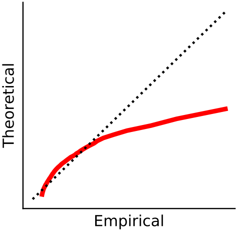
Empirical distribution of the danish fire insurance losses.



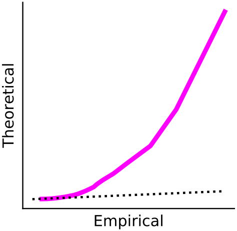
Gamma$(\widehat{r} = 1.30, \widehat{m} = 2.61)$



Weib$(\widehat{k} = 0.96, \widehat{\beta} = 3.29)$



LogNorm$(\widehat{\mu} = 0.79, \widehat{\sigma} = 0.72)$



Par$(\widehat{\alpha} = 0.54, \widehat{\gamma} = 0.31)$

Composite models

The **pdf** of a composite model is defined as

$$f(x) = \begin{cases} p\frac{f_1(x)}{F_1(\gamma)}, \text{ si } x \leq \gamma, \\ (1-p)\frac{f_2(x)}{1-F_2(\gamma)}, \text{ si } x > \gamma, \end{cases}$$

- f_1 and F_1 are the **pdf** and **cdf** of the bulk distribution
- f_2 and F_2 are the **pdf** and **cdf** of the tail distribution
- p is the mixing parameter
- γ is the threshold parameter

estimation of γ

- Extreme value theory
    - Graphical aids
    - Automatic threshold selection method
- Simultaneous estimation
    - MLE
    - MCMC

Y. Wang, I. H. Haff, and A. Huseby, "Modelling extreme claims via composite models and threshold selection methods," *Insurance : Mathematics and Economics*, vol. 91, pp. 257–268, mar 2020.

Components of the composite models

- f_1 parametric or non-parametric
- f_2 heavy tailed, often generalized Pareto

C. Scarrott and A. MacDonald, “A review of extreme value threshold es-timation and uncertainty quantification,” *REVSTAT–Statistical Journal*, vol. 10, no. 1, pp. 33–60, 2012.

Assumptions on f

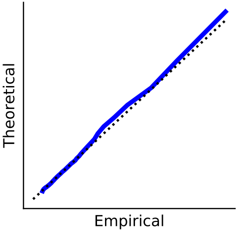
- Discontinuity at $x = \gamma$
- Discontinuity at $x = \gamma$ with fixed $p = F_1(\gamma)$
- Continuity at $x = \gamma$ with

$$p = \frac{f_1(\gamma)}{F_1(\gamma)} \Big/ \frac{f_2(\gamma)}{1 - F_2(\gamma)}. \tag{1}$$

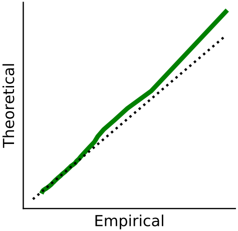
- Continuity and differentiability at $x = \gamma$ with (1) and γ solution of

$$\frac{f_1'(\gamma)}{f_1(\gamma)} - \frac{f_2'(\gamma)}{f_2(\gamma)} = 0.$$

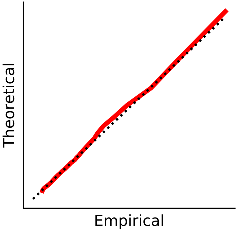
B. Grün and T. Miljkovic, "Extending composite loss models using a general framework of advanced computational tools," *Scandinavian Actuarial Journal*, vol. 2019, pp. 642–660, apr 2019.



Gamma$(\widehat{r}=35.68)-$**Par**$(\widehat{\alpha}=1.31,\widehat{\gamma}=1.16)$



Weib$(\widehat{k}=14.03)-$**Par**$(\widehat{\alpha}=1.26,\widehat{\gamma}=1.00)$



LogNorm$(\widehat{\sigma}=0.19)-$**Par**$(\widehat{\alpha}=1.32,\widehat{\gamma}=1.21)$

Application to insurance

Over a given time period, the aggregate cost of claim is

$$S = \sum_{i=1}^{N} X_i,$$

where

- N is a counting random variable ($\sim \text{Pois}(\lambda)$)
- The X_i's are non-negative **iid** random variables and independent from N

⚠ Distribution of S almost never available

Risk management

Actuaries are interested in the high order quantiles of S that they called the Value-at-Risk

$\hookrightarrow$ Very much sensitive to the occurence of extreme X_i's

… And reinsurance

To mitigate the risk, the insurer can transfer part of it to a reinsurer. We have

$$S = R + D$$

where

- R is taken on by the reinsurer
- D stays with the first line insurer

The premiums Π are also subdivided into

$$\Pi = \Pi_R + \Pi_D$$

where

- Π is paid by the policyholder
- Π_R stays with the first line insurer

XL reinsurance

Excess-of-Loss reinsurance

The reinsurance takes on the part of each loss in excess of the priority P up to a limit L

We have

$$R = \sum_{i=1}^{N} \min\{(X_i - P)_+, L\},$$

and

$$D = \sum_{i=1}^{N} \left[\min\{X_i, P\}\mathbb{I}_{\{X_i \leq P+L\}} + (X_i - L)\mathbb{I}_{\{X_i > P+L\}}\right]$$

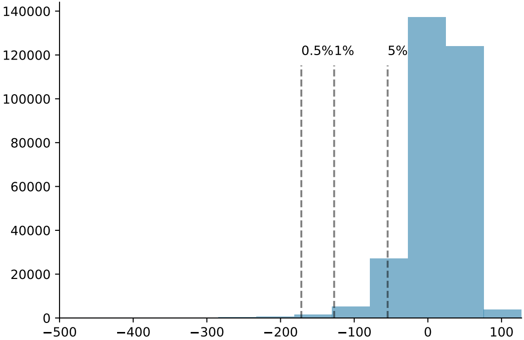
Practical risk capital allocation

Once λ and θ (of the composite model) has been estimated

We are after a reserve $u > 0$ such that

$$\mathbb{P}(u + \Pi_D - D \leq 0) = \alpha \text{ (small)}$$

- $\Pi_D = \mathbb{E}(D) \Rightarrow$ Numerical integration
- Distribution of $\Pi_D - D \Rightarrow$ CMC



Accurate premiums calculation and quantile estimation requires a model to encapsulate both large and morderate claim amounts

Bayesian statistics

Let $\mathscr{M}$ be a model with parameters θ, and x some observed data.

- Bayesian statistics targets the posterior distribution of the parameters

$$\pi(\theta|\mathrm{x}) = \frac{p(\mathrm{x}|\theta)\pi(\theta)}{\int_\Theta p(\mathrm{x}|\theta)\pi(\theta)\mathrm{d}\theta} = \frac{p(\mathrm{x}|\theta)\pi(\theta)}{Z(\mathrm{x})},$$

by updating the prior $\pi(\theta)$ via the likelihood $p(\mathrm{x}|\theta)$.

↪ Model calibration

If many models $\mathscr{M}_1,\ldots,\mathscr{M}_K$ are competing

- The posterior model evidence of each model follow from

$$\pi(M_i|\mathrm{x}) = \frac{p(\mathrm{x}|M_i)\pi(\mathscr{M}_i)}{\sum_{j=1}^K p(\mathrm{x}|M_j)\pi(\mathscr{M}_j)},\ i = 1,\ldots,K.$$

↪ Select or combine models

Importance sampling

Bayesian inference reduces to compute

$$\mathbb{E}_{\pi(\theta|\mathrm{x})}(\varphi) = \int_{\Theta} \varphi(\theta)\pi(\theta|\mathrm{x})\mathrm{d}\theta \approx \frac{1}{N}\sum_{i=1}^{N}\varphi(\theta_i),$$

where $\theta_i \overset{iid}{\sim} \pi(\theta|\mathrm{x})$. Let g be a proposal distribution on Θ then the following identity holds

$$\begin{aligned}
\mathbb{E}_{\pi(\theta|\mathrm{x})}(\varphi) &= \int_{\Theta} \varphi(\theta)\pi(\theta|x)\mathrm{d}\theta \\
&= \int_{\Theta} \varphi(\theta)\frac{L(\mathrm{x}|\theta)\pi(\theta)}{Z(\mathrm{x})}\mathrm{d}\theta \\
&= Z(\mathrm{x})^{-1}\int_{\Theta} \varphi(\theta)\frac{L(\mathrm{x}|\theta)\pi(\theta)}{g(\theta)}g(\theta)\mathrm{d}\theta \\
&= Z(\mathrm{x})^{-1}\int_{\Theta} \varphi(\theta)w(\theta)g(\theta)\mathrm{d}\theta \\
&= Z(\mathrm{x})^{-1}\mathbb{E}_{g}(\varphi\cdot w),
\end{aligned}$$

Importance sampling

Taking $\varphi = 1$ yields

$$Z(\mathrm{x}) = \mathbb{E}_g(w) \approx \frac{1}{N}\sum_{i=1}^{N} w(\theta_i^*),$$

where $\theta_i^* \overset{iid}{\sim} g$. We also have

$$\mathbb{E}_{\pi(\theta|\mathrm{x})}(\varphi) \approx \sum_{i=1}^{N} W_i \varphi(\theta_i^*),$$

with

$$W_i = \frac{w(\theta_i^*)}{\sum_{j=1}^{N} w(\theta_j^*)},\ \ i = 1, \dots, N.$$

Sequential Monte Carlo Sampler

Let $\pi_s, \; s = 0, \ldots, t$ be a sequence of intermediary distributions such that $\pi = \pi_0$ and $\pi_t = \pi(\cdot|\mathrm{x})$ represented by particle clouds $(W_i^s, \theta_i^s), \; i = 1, \ldots, N$.

1. **Reweight** :

$$W_i^{s+1} \propto w_i^{s+1} = \frac{\pi_{s+1}(\theta_i^s)}{\pi_s(\theta_i^s)} \text{ such that } ESS > \rho \cdot N$$

2. **Select** :

$$(\widetilde{\theta}_1^{s+1} \ldots, \widetilde{\theta}_N^{s+1}) \sim \left\{(W_1^{s+1}, \theta_1^s) \ldots, (W_N^{s+1}, \theta_1^s)\right\}$$

3. **Move** :

$$\theta_i^{s+1} \sim K_H(\widetilde{\theta}_i^{s+1}, \cdot) \text{ and } W_i^{s+1} \leftarrow 1/N \text{ for } i = 1, \ldots, N$$

We have

$$\theta_1^t, \ldots, \theta_N^t \sim \pi(\cdot|\mathrm{x}), \text{ and } Z(\mathbf{x}) \approx \prod_{s=1}^{t} \left(\frac{1}{N} \sum_{i=1}^{N} w_i{}^s\right)$$

. P. D. Moral, A. Doucet, and A. Jasra, "Sequential monte carlo samplers," *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, vol. 68, pp. 411–436, jun 2006.

How to define these intermediary distributions ?

- Simulated annealing

$$\pi_s(\theta) \propto L(\mathrm{x}|\theta)^{\tau_s}\pi(\theta),$$

  where $0 = \tau_0 < \ldots < \tau_t = 1$.

- Data by batches

$$\pi_s(\theta) \propto L(x_1, \ldots, x_{n_s}|\theta)\pi(\theta)$$

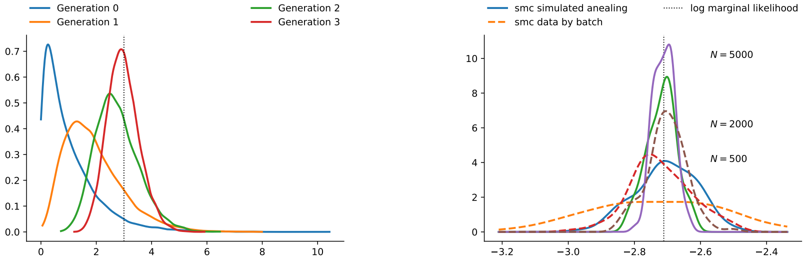
  where $0 = n_0 < \ldots < n_t = n$.

. R. M. Neal, "Annealed importance sampling," *Statistics and Computing*, vol. 11, no. 2, pp. 125–139, 2001.

. N. Chopin, "A sequential particle filter method for static models," *Biometrika*, vol. 89, pp. 539–552, aug 2002.

SMC illustrations

$$x_1, \dots, x_{25} \overset{iid}{\sim} \text{Exp}(\delta = 3), \text{ and } \pi(\theta) \sim \text{Gamma}(a = 1, b = 1)$$



Posterior distribution convergence

Normalizing constant convergence

Bayesian Model Averaging

Let Δ be a quantity of interest, e.g. the 95% or 99% quantile.

- For each model m_j, $j = 1, \ldots, J$, we hold an estimation $\widehat{\Delta}_j$ of Δ
- We can combine them using the posterior model probabilities as

$$\widehat{\Delta} := \mathbb{E}(\Delta|\mathrm{x}) \approx \sum_{j=1}^{J} \widehat{\Delta}_j \pi(m_j|\mathrm{x}),\ j = 1, \ldots, J.$$

This is called Bayesian model averaging.

J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky, "Bayesian model averaging : A tutorial," *Statistical Science*, vol. 14, no. 4, pp. 382–401, 1999.

`bayes-splicing` package

Name	Parameters		pdf
Exponential	$\text{Exp}(\lambda)$	$\lambda>0$	$\lambda e^{-\lambda x},\ x>0$
Gamma	$\text{Gamma}(r,m)$	$r,m>0$	$\frac{x^{r-1}e^{-x/m}}{\Gamma(r)m^r},\ x>0$
Weibull	$\text{Weibull}(k,\beta)$	$k,\beta>0$	$\frac{k}{\beta}\left(\frac{x}{\beta}\right)^{k-1}e^{-(x/\beta)^k},\ x>0$
Lognormal	$\text{Lognormal}(\mu,\sigma)$	$\mu\in\mathbb{R},\ \sigma>0$	$\frac{1}{x\sigma\sqrt{2\pi}}\exp\left[-\frac{(\log(x)-\mu)^2}{2\sigma^2}\right],\ x>0$
Inverse-Gaussian	$\text{Inverse-Gaussian}(\mu,\lambda)$	$\mu,\lambda>0$	$\sqrt{\frac{\lambda}{2\pi x^3}}\exp\left(\frac{\lambda(x-\mu)^2)}{2\mu^2 x}\right),\ x>0$
Inverse-Gamma	$\text{Inverse-Gamma}(r,m)$	$r,m>0$	$\frac{e^{-m/x}m^r}{x^{r+1}\gamma(r)},\ x>0$
Inverse-Weibull	$\text{Inverse-Weibull}(k,\beta)$	$r,m>0$	$k\beta^k x^{-k-1}e^{-(\beta/x)^k},\ x>0$
Lomax	$\text{Lomax}(\alpha,\sigma)$	$\alpha,\sigma>0$	$\frac{\alpha\sigma^\alpha}{(\sigma+x)^{\alpha+1}},\ x>0$
Log-Logistic	$\text{Log-Logistic}(\beta,\sigma)$	$\beta,\sigma>0$	$\frac{\beta\sigma^\beta x^{\beta-1}}{(\sigma^\beta+x^\beta)^2},\ x>0$
Burr	$\text{Burr}(\alpha,\beta,\sigma)$	$\alpha,\beta,\sigma>0$	$\frac{\alpha\beta\sigma^{\alpha\beta}x^{\beta-1}}{(\sigma^\beta+x^\beta)^{\alpha+1}},\ x>0$
Pareto	$\text{Pareto}(\alpha,\gamma)$	$\alpha,\gamma>0$	$\frac{\gamma^\alpha}{x^{\alpha+1}},\ x>\gamma$
Generalized Pareto	$\text{GPD}(\xi,\sigma,\gamma)$	$\xi,\sigma,\gamma>0$	$\sigma^{-1}\left[1+\frac{\xi(x-\gamma)}{\sigma}\right]^{-(\xi+1)/\xi},\ x\geq 0$

`https://youtu.be/MVXUm9AWVVs`

Simulation : well-specified case

We sample

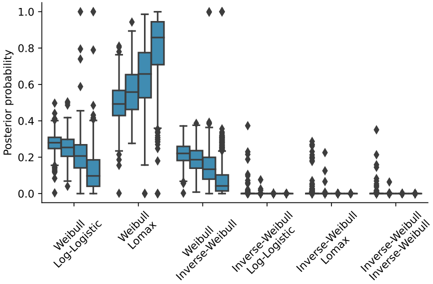
$$x_1, \ldots, x_n \sim \text{Weibull}(k = 1/2, \beta = 1) - \text{Lomax}(\alpha = 2.5, \sigma = 3) - (\gamma = 1.5)$$

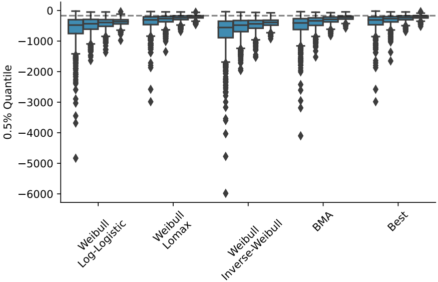
- $n \in \{500, 1,000, 2,000\}$
- $1,000$ times

We fit composite models as combination of

- $f_1 \in \left\{\text{Weibull}, \text{Inverse-Weibull}\right\}$
- $f_2 \in \left\{\text{Lomax}, \text{Log-Logistic}, \text{Inverse-Weibull}\right\}$

That makes 6 models in total.


Posterior probability
1.0
0.8
0.6
0.4
0.2
0.0
Weibull
Log-Logistic
Weibull
Lomax
Weibull
Inverse-Weibull
Inverse-Weibull
Log-Logistic
Inverse-Weibull
Lomax
Inverse-Weibull
Inverse-Weibull


0.5% Quantile
0
−1000
−2000
−3000
−4000
−5000
−6000
Weibull
Log-Logistic
Weibull
Lomax
Weibull
Inverse-Weibull
BMA
Best

Simulation : miss-specified case

We sample

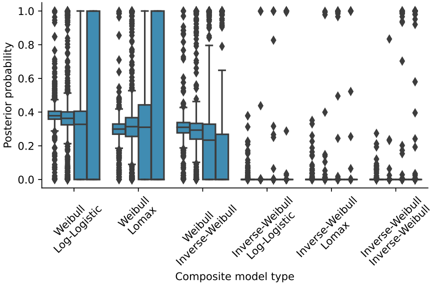
$$x_1, \ldots, x_n \sim \text{Exp}(\lambda = 1/2) - \text{Burr}(\alpha = 1.8, \beta = 2, \sigma = 3) - (\gamma = 2.5)$$

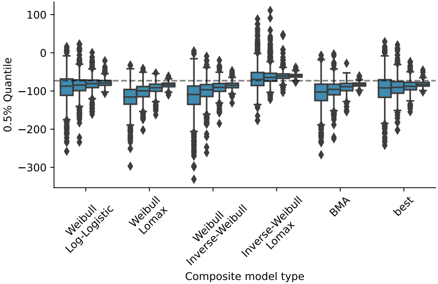
- $n \in \{500, 1,000, 2,000\}$
- 1,000 times

We fit composite models as combination of

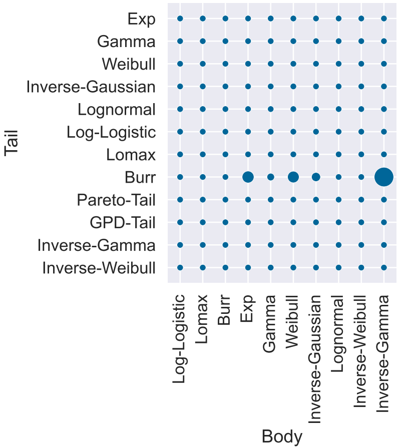
- $f_1 \in \left\{\text{Weibull}, \text{Inverse-Weibull}\right\}$
- $f_2 \in \left\{\text{Lomax}, \text{Log-Logistic}, \text{Inverse-Weibull}\right\}$

That makes 6 models in total.

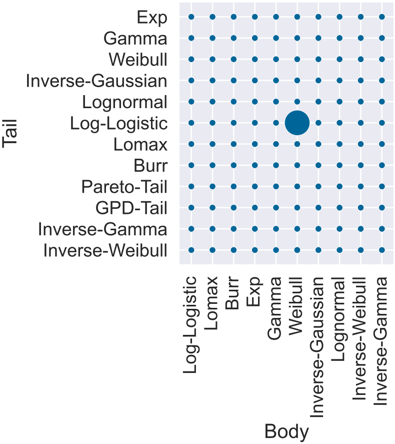

Posterior probability
1.0
0.8
0.6
0.4
0.2
0.0
Weibull Log-Logistic
Weibull Lomax
Weibull Inverse-Weibull
Inverse-Weibull Log-Logistic
Inverse-Weibull Lomax
Inverse-Weibull Inverse-Weibull
Composite model type


0.5% Quantile
100
0
−100
−200
−300
Weibull Log-Logistic
Weibull Lomax
Weibull Inverse-Weibull
Inverse-Weibull Lomax
BMA
best
Composite model type

Danish fire insurance data : Which model ?

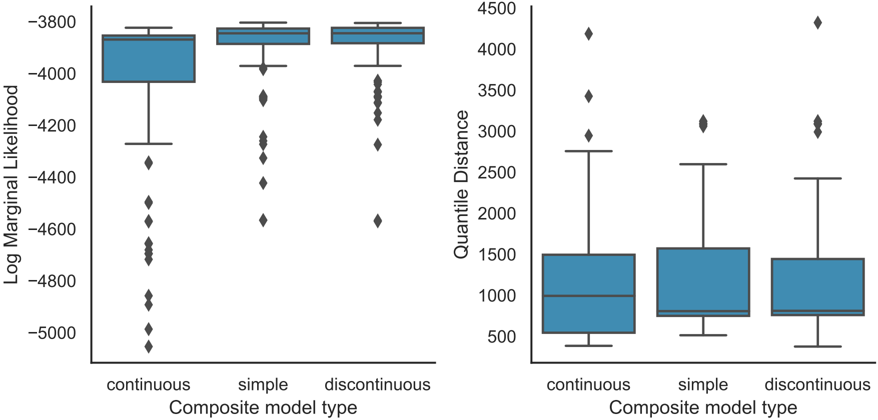


Discontinuous



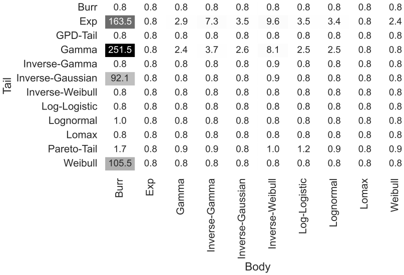
Continuous

Danish fire insurance data : Continuous or not ?



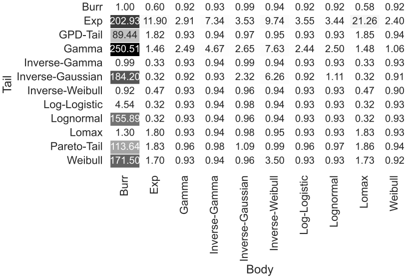
. **Quantile Distance** $= \sum_{i=1}^{n} \left| x_{i:n} - F^{-1}(i/n) \right|$

Danish fire insurance data : threshold ?



	Burr	Exp	Gamma	Inverse-Gamma	Inverse-Gaussian	Inverse-Weibull	Log-Logistic	Lognormal	Lomax	Weibull
Burr	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
Exp	163.5	0.8	2.9	7.3	3.5	9.6	3.5	3.4	0.8	2.4
GPD-Tail	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
Gamma	251.5	0.8	2.4	3.7	2.6	8.1	2.5	2.5	0.8	0.8
Inverse-Gamma	0.8	0.8	0.8	0.8	0.8	0.9	0.8	0.8	0.8	0.8
Inverse-Gaussian	92.1	0.8	0.8	0.8	0.8	0.9	0.8	0.8	0.8	0.8
Inverse-Weibull	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
Log-Logistic	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
Lognormal	1.0	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
Lomax	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
Pareto-Tail	1.7	0.8	0.9	0.9	0.8	1.0	1.2	0.9	0.8	0.9
Weibull	105.5	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8

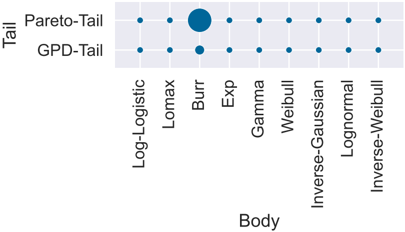
Discontinuous



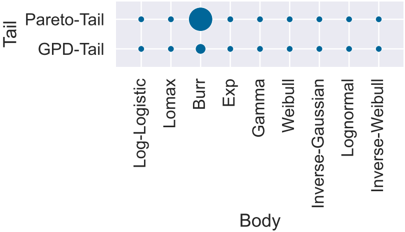
	Burr	Exp	Gamma	Inverse-Gamma	Inverse-Gaussian	Inverse-Weibull	Log-Logistic	Lognormal	Lomax	Weibull
Burr	1.00	0.60	0.92	0.93	0.99	0.94	0.92	0.92	0.58	0.92
Exp	202.93	11.90	2.91	7.34	3.53	9.74	3.55	3.44	21.26	2.40
GPD-Tail	89.44	1.82	0.93	0.94	0.97	0.95	0.93	0.93	1.85	0.94
Gamma	250.51	1.46	2.49	4.67	2.65	7.63	2.44	2.50	1.48	1.06
Inverse-Gamma	0.99	0.33	0.93	0.94	0.99	0.94	0.93	0.93	0.33	0.93
Inverse-Gaussian	184.20	0.32	0.92	0.93	2.32	6.26	0.92	1.11	0.32	0.91
Inverse-Weibull	0.92	0.47	0.93	0.94	0.96	0.94	0.93	0.93	0.47	0.90
Log-Logistic	4.54	0.32	0.93	0.94	0.98	0.94	0.93	0.93	0.32	0.93
Lognormal	155.89	0.32	0.93	0.94	0.96	0.94	0.93	0.93	0.32	0.93
Lomax	1.30	1.80	0.93	0.94	0.98	0.95	0.93	0.93	1.83	0.93
Pareto-Tail	113.64	1.83	0.96	0.98	1.09	0.99	0.96	0.97	1.86	0.94
Weibull	171.50	1.70	0.93	0.94	0.96	3.50	0.93	0.93	1.73	0.92

Continuous

Danish fire insurance data : fixed threshold



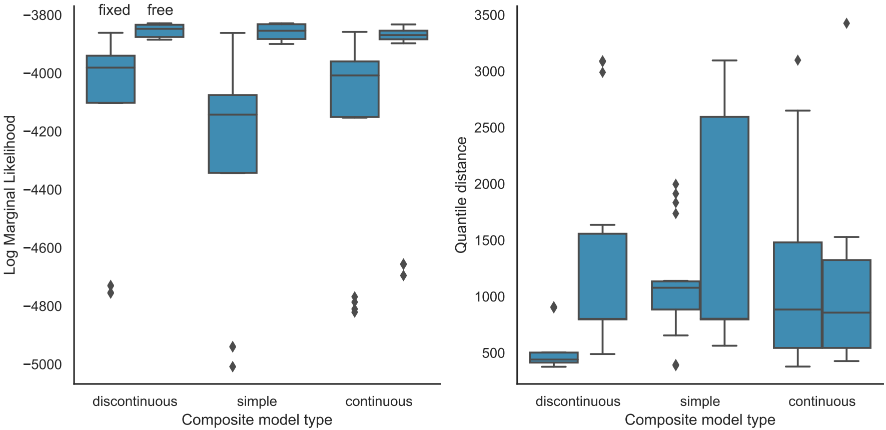
Discontinuous



Continuous

M. Bladt, H. Albrecher, and J. Beirlant, “Threshold selection and trimming in extremes,” *Extremes*, vol. 23, pp. 629–665, jul 2020.

Danish fire insurance data : fixed VS free



Conclusions

Implementation of the SMC sampler to fit and compare composite models .

P.-O. Goffard, “Sequential Monte Carlo samplers to fit and compare insurance loss models.” working paper or preprint, June 2021.

. Code available at `https://github.com/LaGauffre/bayes-splicing`

Perspectives

⚠ Improve the methodology

- Use particle recycling and better perturbation kernel
  - L. F. South, A. N. Pettitt, and C. C. Drovandi, "Sequential monte carlo samplers with independent markov chain monte carlo proposals," *Bayesian Analysis*, vol. 14, pp. 753–776, sep 2019.
- Use cross validation pocedure to combine the models
  - Y. Yao, A. Vehtari, D. Simpson, and A. Gelman, "Using stacking to average bayesian predictive distributions (with discussion)," *Bayesian Analysis*, vol. 13, sep 2018.

Perspectives

⚠ Better estimate of the higher order quantiles ⇒ minimum distance estimator

- Use the Wasserstein distance
  - E. Bernton, P. E. Jacob, M. Gerber, and C. P. Robert, "On parameter estimation with the wasserstein distance," *Information and Inference : A Journal of the IMA*, vol. 8, pp. 657–676, oct 2019.
- Approximate Bayesian Computation to get posterior distributions
  - E. Bernton, P. E. Jacob, M. Gerber, and C. P. Robert, "Approximate bayesian computation with the wasserstein distance," *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, vol. 81, pp. 235–269, feb 2019.
  - P.-O. Goffard and P. J. Laub, "Approximate bayesian computations to fit and compare insurance loss models," *Insurance : Mathematics and Economics*, vol. 100, pp. 350–371, sep 2021.

Perspective linked to EcoDEP

- ⚠ Model meteorological events, i.e. wind speed and precipitation
- ⚠ Multivariate composite models and inspect the joint posterior distribution of the extreme value thresholds