Presentation

“A nonparametric approach to calculating value-at-risk”

Oscar Johansson
oscar.johansson@student.kuleuven.be

Heavy tail distributions

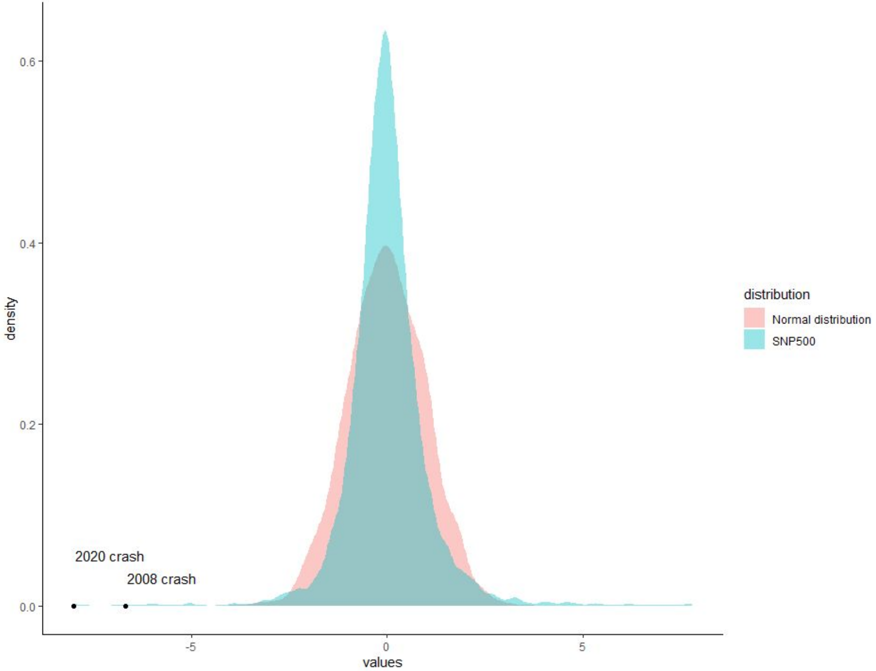


Figure 1: Gaussian kernel density of S&P500 daily logreturns compared to normal approximation.

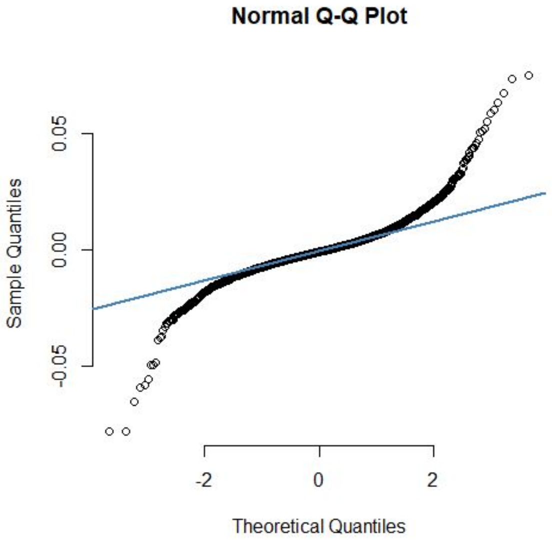


Figure 2: Q-QPlot of daily lognormal returns of S&P500 compared to normal estimate(line)

$$VaR_\alpha(X) = inf\{x : F_X(x) \geq \alpha\} = F_X^{-1}(\alpha)$$

	Empirical S&P500	Normal
$\alpha = 0.1$	-0.96	-1.28
$\alpha = 0.05$	-1.35	-1.64
$\alpha = 0.01$	-2.54	-2.33
$\alpha = 0.001$	-5.99	-3.09

Problem premise

Accurate VaR ⟷ Accurate cdf

Non-parametric solution?

ELSEVIER

Insurance: Mathematics and Economics

Volume 52, Issue 2, March 2013, Pages 255-262

A nonparametric approach to calculating value-at-risk

Ramon Alemany, Catalina Bolancé, Montserrat Guillén

Recall:

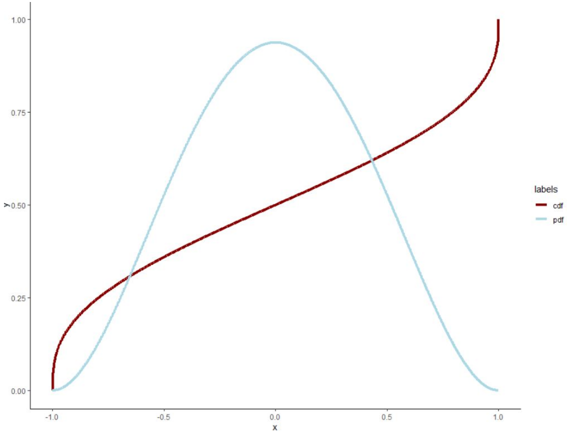
$$
\begin{aligned}
AMISE = & \frac{1}{n}\int F_X(x)[1 - F_X] \\
& - \frac{1}{n}b(t)[1 - K(t)]dt \\
& + \frac{1}{4}b^4 \int [f'_X(x)]^2 dx \left(\int t^2 k(t) dt \right)^2
\end{aligned}
$$

- Minimize with respect to b
- Minimize with respect to $[f'_X(x)]^2$

George R. Terrell showed in 1990 that the Beta(3,3) scaled to [-1,1] minimize the AMISE in the set of all densities with known variance[1].

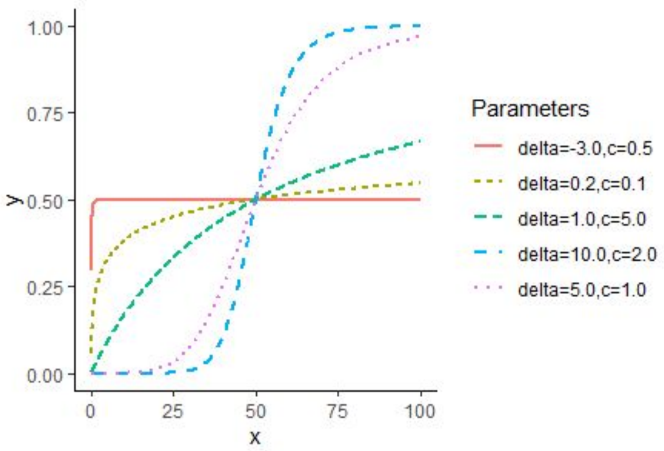
How do we get our data to be Beta(3,3) distributed?

$$m(x) = \frac{15}{16}(1 - x^2)^2$$
$$M(x) = \frac{1}{16}(15x - 5x^3 + x^5 + 11)$$

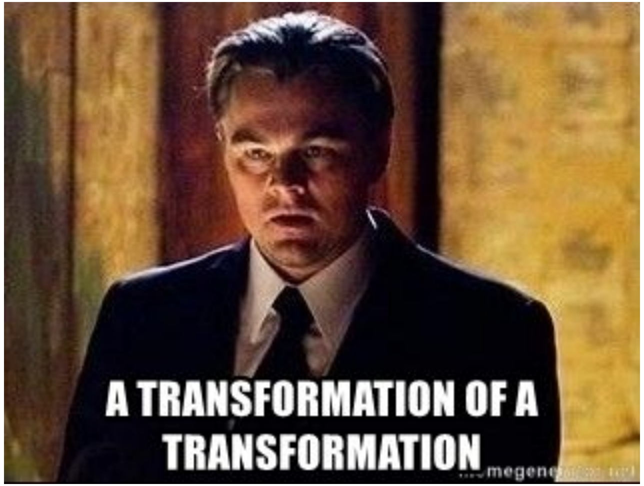


Champernowne distribution[2]

$$T(x) = \frac{(x+c)^\delta - c^\delta}{(x+c)^\delta + (M+c)^\delta - 2c^\delta}.$$



$$Y = M(T^{-1}(X))$$



Algorithm 1 The DTKE algorithm

Require: $\mathbf{X}, n = len(X), \alpha$

$Z \leftarrow T(X)$

$Y \leftarrow M^{-1}(Z)$

$x_\alpha \leftarrow M^{-1}(\alpha)$

$b_\alpha^{Clas} \leftarrow (\frac{9}{35}\frac{m(x_\alpha)}{\frac{1}{25}m'(x_\alpha)})^{-\frac{1}{3}}n^{-\frac{1}{3}}$

$\bar{F}_Y(x) = \frac{1}{n}\sum_{i=1}^{n} K(\frac{x-Y_i}{b_\alpha^{Clas}})$

$\hat{F}_x = \bar{F}_Y(x) / \int_0^1 \bar{F}_Y(x)dx, x \in [0,1]$

q=$\hat{F}_X^{-1}(\alpha)$

Return: $VaR_\alpha = T^{-1}(M(q))$

Using the Epanechnikov kernel we have different choices:

- Optimal AMISE bandwidth:

$$\mathrm{MISE}\left\{\widehat{F}_X(x)\right\} = E\left\{\int \left[F_X(x) - \widehat{F}_X(x)\right]^2 dx\right\} \longrightarrow \hat{b}^* = 3^{\frac{1}{3}} n^{-\frac{1}{3}}.$$

- Optimal AWISE bandwidth:

$$WISE\left\{\widehat{F}_X(x)\right\} = E\left\{\int \left[F_X(x) - \widehat{F}_X(x)\right]^2 x^2 dx\right\} \longrightarrow \hat{b}^{**} = (9/7)^{\frac{1}{3}} n^{-\frac{1}{3}}.$$

- Optimal bandwith around α:

$$\mathrm{MISE}\left\{\widehat{F}_X(x)\right\} = E\left\{\int \left[F_X(x) - \widehat{F}_X(x)\right]^2 dx\right\} \longrightarrow b^{Clas}_{y=0.78872} = 0.88321 n^{-\frac{1}{3}}.$$

Results

$$W = 0.3LN(0, 0.5)+0.7P(1, 1)$$

LN is a lognormal distribution and P the pareto distribution.

n=5000, 100 runs, ± std

VaR_α	$\alpha = 0.99$	$\alpha = 0.999$
$Analytical$	139.0	699.0
CKE_w	159 ± 129	244 ± 205
CKE_x	162 ± 199	487 ± 244
$DTKE_w$	243 ± 100	359 ± 48.0
$DTKE_x$	200 ± 84	561 ± 49.1
TKE_w	195 ± 7.4	762 ± 159

Discussion

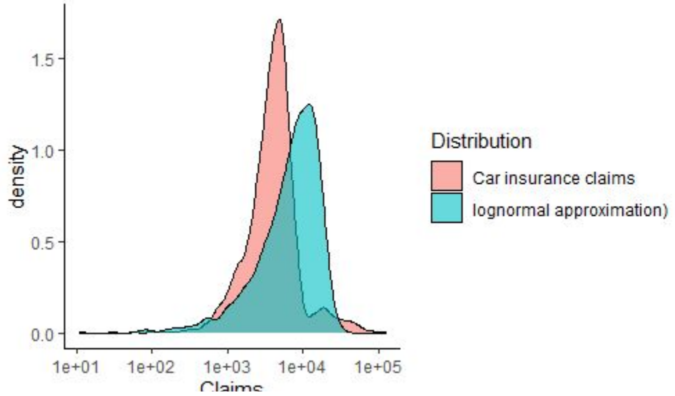
- Does not state assumption and restrictions.
- Unclear methodology.
- Over optimistic results.

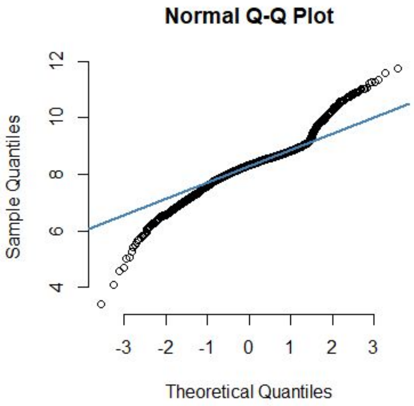
Example:

- Car insurance data from Kaggle[3].
- Heavy tailed distribution.

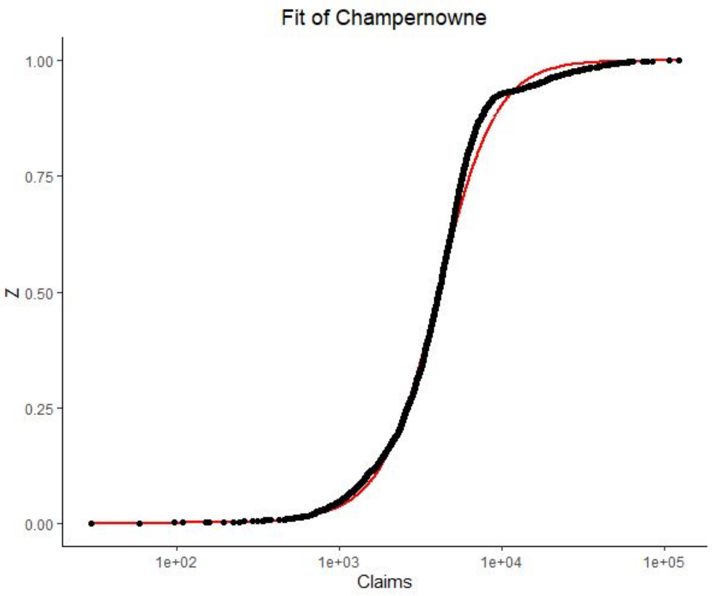
Question:

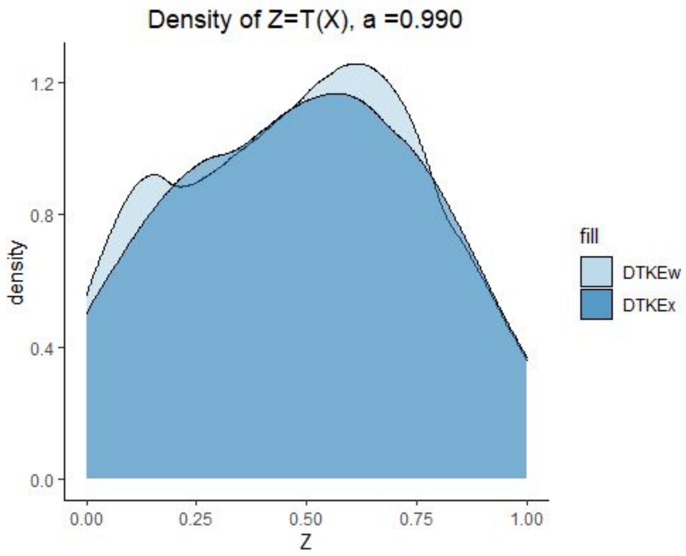
- Set new budget based on VaR.
- Cost worst case ≈ $n_{claims} * VaR_{\alpha}$.
- Let α=0.99.



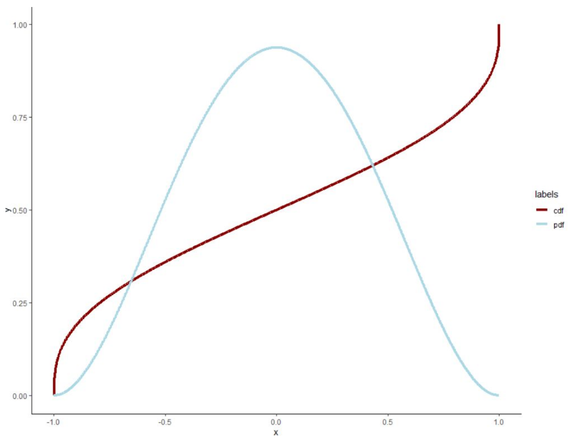


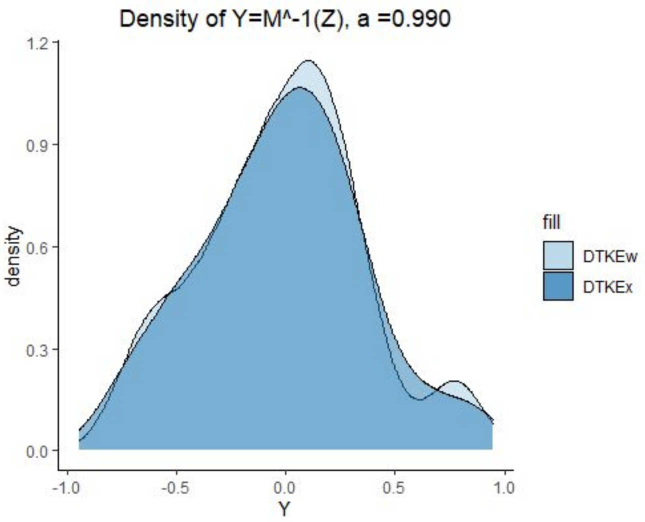
Example: Car insurance data





Example: Car insurance data





Example: Car insurance data

5 data splits,
Mean ± std

Method	Esstimate
CKE_w	37129 ± 24399
CKE_x	47338 ± 11850
TKE	18147 ± 1407
$DTKE_w$	25280 ± 6673
$DTKE_x$	24748 ± 5237

Implications

5 data splits, ± std

Method	Esstimate
CKE_w	37129 ± 24399
CKE_x	47338 ± 11850
TKE	18147 ± 1407
$DTKE_w$	25280 ± 6673
$DTKE_x$	24748 ± 5237

Analysis:

- Varying results.
- CKE_x would imply double budget of DTKE.
- Transforms likely underestimate
- Under or over estimate?

Conclusion

- Varying implications.
- Accuracy not most important but margin.
- In real life other solution:
  - Claim limit.
  - VaR on claim sum instead.
  - Overestimate.

References:

[1] George R Terrell. 1990. “The maximal smoothing principle in density estimation” Journal of the American statistical association, 85(410):470–477.

[2] Tine Buch-Larsen, Jens Perch Nielsen, Montserrat Guillen, and Catalina Bolanc ´ e. 2005. Kernel density estimation for heavy-tailed distributions using the champernowne transformation. Statistics,39(6):503–516

[3] Bunty Shah. 2017. Auto insurance claims data.https://www.kaggle.com/buntyshah/auto-insurance-claims-data.

Algorithm 1 The DTKE algorithm

Require: $\mathbf{X}, n = len(X), \alpha$

$Z \leftarrow T(X)$

$Y \leftarrow M^{-1}(Z)$

$x_\alpha \leftarrow M^{-1}(\alpha)$

$b_\alpha^{Clas} \leftarrow (\frac{9}{35}\frac{m(x_\alpha)}{\frac{1}{25}m'(x_\alpha)})^{-\frac{1}{3}}n^{-\frac{1}{3}}$

$\bar{F}_Y(x) = \frac{1}{n}\sum_{i=1}^{n} K(\frac{x-Y_i}{b_\alpha^{Clas}})$

$\hat{F}_x = \bar{F}_Y(x)/\int_0^1 \bar{F}_Y(x)dx, x \in [0,1]$

q=$\hat{F}_X^{-1}(\alpha)$

Return: $VaR_\alpha = T^{-1}(M(q))$

Questions

VaR_α	$\alpha = 0.99$	$\alpha = 0.999$
$Analytical$	139.0	699.0
CKE_w	159 ± 129	244 ± 205
CKE_x	162 ± 199	487 ± 244
$DTKE_w$	243 ± 100	359 ± 48.0
$DTKE_x$	200 ± 84	561 ± 49.1
TKE_w	195 ± 7.4	762 ± 159

Questions

$$T(x) = \frac{(x+c)^\delta - c^\delta}{(x+c)^\delta + (M+c)^\delta - 2c^\delta}.$$