Vybrané metódy a aplikácie štatististickej inferencie založené na numerickej inverzii charakteristickej funkcie[1]

Selected Methods and Applications of Statistical Inference Based on Numerical Inversion of the Characteristic Function

Viktor WITKOVSKÝ

Ústav merania SAV, Bratislava
witkovsky@savba.sk

ROBUST 2016

19. letní škola JČMF
Sporthotel Kurzovní, Rejhotice/Loučná nad Desnou, 11.-16. september 2016

[1] Táto práca bola podporovaná Agentúrou na podporu výskumu a vývoja (APVV) na základe Zmluvy č. APVV-15-0295 a Vedeckou grantovou agentúrou MŠVVaŠ SR a SAV (VEGA), projekt VEGA 2/0047/15 a VEGA 2/0011/16.

Outline

1 Introduction

2 Methods for numerical inversion of the characteristic function
- Characteristic functions and their combinations
- Numerical methods for the inverse Fourier transform

3 Applications
- Parametric inference
  - 1. Exact distribution of the LRT statistic in normal linear regression
- Non-parametric inference
  - 2. Gaussian processes and their distribution
  - 3. Distribution of the Cramér-von Mises and the Anderson-Darling statistics
- Inference based on the empirical characteristic function
  - 4. PDF/CDF estimation by numerical inversion of the smoothed ECF
  - 5. Testing symmetry based on the ECF
  - 6. Compound distribution derived by numerical inversion of the compound ECF
- Other stuff based on ECF
  - 7. Bootstrap distribution by numerical inversion of the convolved ECF
  - 8. Re-weighted and smoothed bootstrap distribution of the kth order statistic (quantile estimator)

MOTIVATION:

- The exact, asymptotic, and/or approximate methods of the statistical inference are frequently based on distributions (PDF, CDF and/or QF) derived by inverting the appropriate characteristic functions (CF).
  - **Metrology** – combining information from independent sources, combining expert knowledge with experimental evidence, expressing uncertainty in measurement, the common mean problem in inter-laboratory comparisons (leading e.g. to a linear combination of independent Student's t or Fisher-Snenedor F distributed random variables).
  - **Operational risk and insurance** – aggregate loss, compound distributions of frequency and severity of claims, combining historical evidence with expert knowledge (i.e. mixing empirical and parametric distributions), empirical distribution and heavy tails.
- Working with CFs provides an alternative and frequently more simple route than working directly with PDFs and/or CDFs.
- In particular, combining CFs is a simple and trivial task for
  - **convolutions** or **linear combinations** of independent random variables (RVs),
  - specific **compound distributions**,
  - weighted **mixtures of distributions**,
  - **transformed (empirical/approximate) distributions**, e.g., by transforming the (approximate/empirical) CFs.

MOTIVATION FOR NUMERICAL METHODS:

A standard numerical approach for estimating distribution of combined random variables is based on using Monte Carlo methods.

There is strong need for efficient methods and algorithms for arithmetic computations with random variables and their distributions
(see e.g. Pacal: A Python package for arithmetic computations with random variables by Korzen and Jaroszewicz (2014)).

- Derivation of the exact distribution functions by using the analytical inverse of the Fourier transform is frequently too complicated – available only in special cases.
- In most practical situations, it is possible and sufficient to rely on numerical methods for derivation of the PDF/CDF from the CF.
- The numerical inversion of the appropriate CF is applicable in parametric as well as in nonparametric settings (based on using the empirical CDF or CF).
    - In particular, numerical inversion of the appropriate CF can be very useful for the applications related to the stochastic Gaussian processes.

Introduction

- Efficient numerical evaluation of the (inverse) Fourier transform is a well-known and intensively studied problem for a long time.
- Frequently, it is connected with the problem of computing integrals of highly oscillatory (complex) functions.
    - The methods for computing **integrals of oscillatory functions** include, e.g., Sidi (1982), Sidi (1988), Levin (1996), Milovanović (1998), Sidi (2012), Asheim & Huybrechs (2013).
    - The methods for **inverting CF for obtaining the probability distribution functions** include, e.g., Gil-Pelaez (1951), Imhof (1961), Bohman (1970, 1972), Davies (1980), Abate & Whitt (1992), Shephard (1991), Waller, Turnbull & Hardin (1995), Zieliński (2001), Strawderman (2004).
- Surprisingly, such methods are still not so much widespread among statisticians. One possible reason might be that the characteristic functions and the algorithms for numerical inversions are not directly available in standard statistical packages, e.g., R, SAS, MATLAB.

Introduction

- In general, methods for numerical inversion seems to be considered difficult.
- However, here we shall present brief overview of the very simple methods for numerical inversion of the CFs, which could serve as a useful basic tool for approximate statistical methods.
- In particular, here we focus on approximate numerical methods for computing PDF/CDF of univariate continuous random variables from their CF, based on
  - the **Gil-Pelaez inversion** formulae,
  - and the **Fast Fourier Transform (FFT)** algorithm.

Characteristic functions and their combinations

Let Y denotes the continuous univariate RV with its PDF $\mathrm{pdf}_Y(y)$. Recall that,

- CF of the distribution of Y is given by the Fourier transform,

$$\mathrm{cf}_Y(t) = \mathsf{E}\left[e^{itY}\right] = \int_{-\infty}^{\infty} e^{ity}\,\mathrm{pdf}_Y(y)\,dy.$$

- Analytical expressions of the characteristic functions are known for many standard probability distributions, see e.g. Lukacs (1970), or other available sources.

- Otherwise CFs can be computed by available software – analytically or numerically.

Characteristic functions and their combinations

Table: Characteristic functions of continuous univariate distributions used in metrological applications (selected symmetric zero-mean distributions and non-negative distributions). Here, $K_\nu(z)$ denotes the modified Bessel function of the second kind and $J_\nu(z)$ is the Bessel function of the first kind.

Probability distribution	Characteristic function
Gaussian $N(0,1)$	$\mathrm{cf}(t) = \exp\left(-\frac{1}{2}t^2\right)$
Student's t_ν	$\mathrm{cf}(t) = \frac{1}{2^{\frac{\nu}{2}-1}\Gamma(\frac{\nu}{2})}\left(\nu^{\frac{1}{2}}\lvert t\rvert\right)^{\frac{\nu}{2}} K_{\frac{\nu}{2}}\left(\nu^{\frac{1}{2}}\lvert t\rvert\right)$
Rectangular $R(-1,1)$	$\mathrm{cf}(t) = \frac{\sin(t)}{t}$
Triangular $T(-1,1)$	$\mathrm{cf}(t) = \frac{2-2\cos(t)}{t^2}$
Arcsine $U(-1,1)$	$\mathrm{cf}(t) = J_0(t)$

Characteristic functions and their combinations

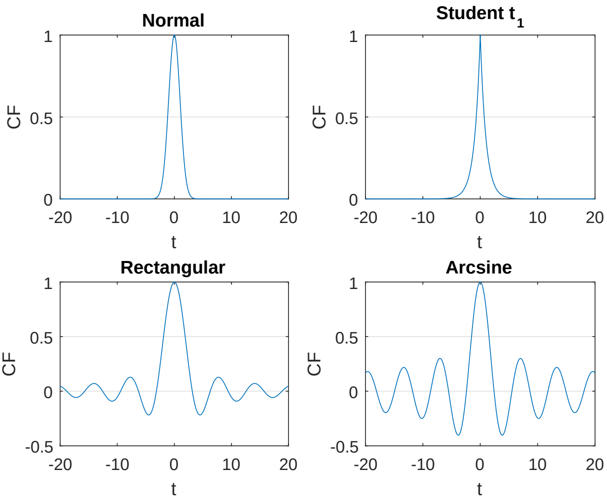


Figure: Characteristic functions of selected symmetric distributions.

Characteristic functions and their combinations

If the analytical form of the CF is unknown or it is too complicated, as it depends on nonstandard special functions and/or complicated series expansions (as is the case, e.g., for the Pareto, Weibul, log-normal and/or log-logistic distributions), such CFs can be still evaluated numerically, either directly from its definition, and/or any other suitable representation.

- For example, by using the half-space Fourier integral transformation for a positive continuous random variable X (defined for $X \geq 0$) with its PDF given by an analytical function $\mathrm{pdf}_X(z)$, which is well defined for complex $z \in \mathbb{C}$ and decays at infinity, (as e.g. for the Weibul, Pareto, log-normal and log-logistic distributions), we get

$$\mathrm{cf}_X(t) = \int_0^\infty \frac{\mathrm{i}}{t}\,\mathrm{pdf}_X\left(\frac{\mathrm{i}x}{t}\right) e^{-x}\,dx, \quad t \in \mathbb{R},$$

  e.g. Asheim & Huybrechs (2013).

- Moreover, by using a suitable stabilizing transformation from $(0,\infty)$ to $(0,1)$, the CF can be numerically evaluated by using a simple Gaussian quadrature rule of a well behaved integral,

$$\mathrm{cf}_X(t) = \int_0^1 \frac{\mathrm{i}}{t}\,\mathrm{pdf}_X\left(\frac{\mathrm{i}}{t}\left(\frac{x}{1-x}\right)^2\right) \frac{2x\,e^{-\left(\frac{x}{1-x}\right)^2}}{(1-x)^3}\,dx.$$

Characteristic functions and their combinations

Table: Selected characteristic functions used as the components of the compound CFs of the aggregate loss distributions. Here, $U(a,b,z)$ denotes the confluent hypergeometric function of the second kind.

Probability distribution	Characteristic function
Poisson	$\mathrm{cf}_N(t) = \exp\left(\lambda\left(e^{\mathrm{i}t} - 1\right)\right)$
Binomial	$\mathrm{cf}_N(t) = \left(1 - p + pe^{\mathrm{i}t}\right)^n$
Negative Binomial	$\mathrm{cf}_N(t) = p^r\left(1 - (1-p)e^{\mathrm{i}t}\right)^{-r}$
Exponential	$\mathrm{cf}_X(t) = \frac{\lambda}{\lambda - it}$
Gamma	$\mathrm{cf}_X(t) = \left(1 - \frac{it}{\beta}\right)^{-\alpha}$
Fisher-Snedecor F_{ν_1,ν_2}	$\mathrm{cf}(t) = \frac{\Gamma(\frac{\nu_1}{2}+\frac{\nu_2}{2})}{\Gamma(\frac{\nu_2}{2})} U\left(\frac{\nu_1}{2}, 1 - \frac{\nu_2}{2}, -\frac{\nu_2}{\nu_1}it\right)$
Pareto (Type I)	$\mathrm{cf}_X(t) = \alpha e^{\mathrm{i}t\sigma}\, U(1, 1-\alpha, -\mathrm{i}t\sigma)$, or alternatively
	$\mathrm{cf}_X(t) = e^{\mathrm{i}t\sigma} \times \mathrm{cf}^0_X(t)$
	$\mathrm{cf}^0_X(t)$ by numerical integration from $\mathrm{pdf}_X(z) = \alpha\sigma^\alpha(\sigma+z)^{-(\alpha+1)}$
Weibull	$\mathrm{cf}_X(t)$ by numerical integration from $\mathrm{pdf}_X(z) = \frac{\alpha}{\sigma}\left(\frac{z}{\sigma}\right)^{\alpha-1} e^{-\left(\frac{z}{\sigma}\right)^\alpha}$
Log-normal	$\mathrm{cf}_X(t)$ by numerical integration from $\mathrm{pdf}_X(z) = \frac{1}{z\sigma\sqrt{2\pi}} e^{-\frac{(\log(z)-\mu)^2}{2\sigma^2}}$
Log-logistic	$\mathrm{cf}_X(t)$ by numerical integration from $\mathrm{pdf}_X(z) = \frac{\frac{\beta}{\alpha}\left(\frac{z}{\alpha}\right)^{\beta-1}}{\left(1+\left(\frac{z}{\alpha}\right)^\beta\right)^2}$
Generalized Pareto	$\mathrm{cf}_X(t) = e^{\mathrm{i}t\theta} \times \mathrm{cf}^0_X(t)$, where θ is the threshold parameter, and
	$\mathrm{cf}^0_X(t)$ by numerical integration from $\mathrm{pdf}_X(z) = \frac{1}{\sigma}\left(1 + \xi\frac{z}{\sigma}\right)^{-(\frac{1}{\xi}+1)}$

Characteristic functions and their combinations

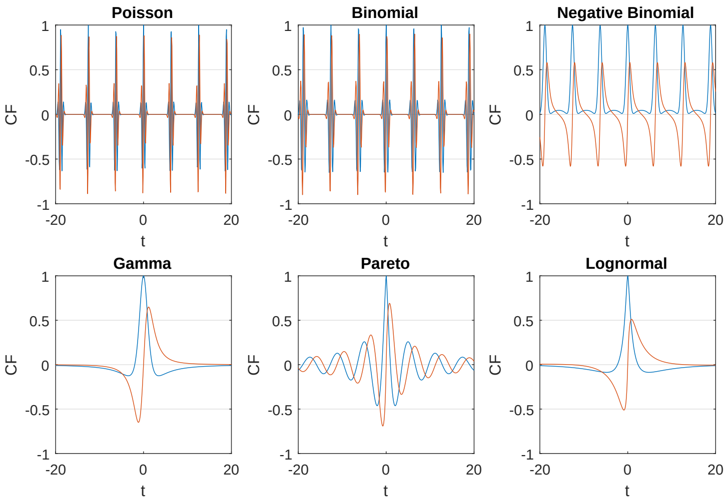


Figure: Selected characteristic functions used as the components of the compound CFs of the aggregate loss distributions.

Working with CFs provides an alternative and frequently more simple route than working directly with PDFs and/or CDFs. In particular, derivation of the CF of a weighted sum of independent random variable is a very simple trivial task.

- CF of a linear combination, $Y = c_1X_1 + \cdots + c_nX_n$, where X_j are independent RVs with known $\mathrm{cf}_{X_j}(t)$ and coefficients c_j, is given by

$$\mathrm{cf}_Y(t) = \prod_{j=1}^{n} \mathrm{cf}_{X_j}(c_j t).$$

- CF of a stochastic convolution, $Y = X_1 + \cdots + X_N$, (compound distribution), where X_j are i.i.d. RVs with common $\mathrm{cf}_X(t)$ and N is a discrete RV defined on non-negative integers with $\mathrm{cf}_N(t)$, is given by

$$\mathrm{cf}_Y(t) = \mathrm{cf}_N\Big(-\mathrm{i}\log\left(\mathrm{cf}_X(t)\right)\Big).$$

Characteristic functions and their combinations

- CF of a weighted mixture of distribution $F_w = \sum_{j=1}^{n} w_j F_{X_j}$, with $\sum_{j=1}^{n} w_j = 1$, is

$$\mathrm{cf}_{F_w}(t) = \sum_{j=1}^{n} w_j \, \mathrm{cf}_{X_j}(t).$$

- ECF — the empirical characteristic function, based on the observed data $x_1, \ldots, x_n$ (realization of the random sample $X_1, \ldots, X_n$, where $X_j \sim F$) is defined as a (equally weighted) mixture of the characteristic functions of the Dirac random variables (concentrated at the values x_j, i.e. $\mathrm{cf}_{x_j}(t) = e^{itx_j}$):

$$\mathrm{cf}_{\hat{F}_n}(t) = \sum_{j=1}^{n} w_j \, \mathrm{cf}_{x_j}(t) = \frac{1}{n}\sum_{j=1}^{n} e^{itx_j} \overset{\text{smoothing}}{\Longrightarrow} \mathrm{cf}_{\tilde{F}_n}(t) = \mathrm{cf}_{\hat{F}_n}(t) \times \mathrm{cf}_Z(\sigma_b t),$$

where $\mathrm{cf}_Z(t) = e^{-\frac{t^2}{2}}$ and σ_b is the pre-selected bandwidth parameter.

- The approximate (discretized/empirical) characteristic function of a RV Y with known distribution F and quantile function $F^{-1}(p)$, is defined as a mixture of the Dirac random variables at $q_j = F^{-1}(p_j)$, for equidistant $p_j \in (0, 1)$:

$$\hat{\mathrm{cf}}_Y(t) = \frac{1}{n}\sum_{j=1}^{n} e^{itq_j} \overset{\text{transforming}}{\Longrightarrow} \hat{\mathrm{cf}}_{g(Y)}(t) = \frac{1}{n}\sum_{j=1}^{n} e^{itg(q_j)} \overset{\text{smoothing}}{\Longrightarrow} \ldots .$$

Characteristic functions and their combinations

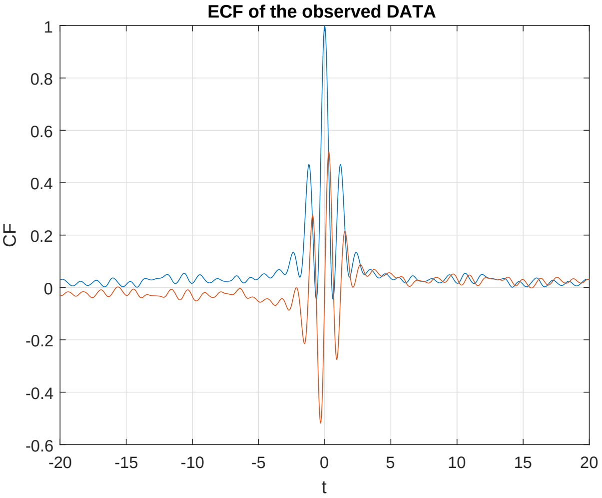


Figure: Empirical characteristic function based on observed data.

The Gil-Pelaez inversion formulae

PDF is given by the inverse Fourier transform,

$$\mathrm{pdf}_Y(y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ity}\,\mathrm{cf}_Y(t)\,dt$$

Gil-Pelaez (1951) derived the inversion formulae of the absolutely integrable CFs over $(-\infty,\infty)$, suitable for numerical evaluation of the PDF and/or the CDF, which require integration of a real-valued functions, only. In particular,

- $$\begin{aligned}\mathrm{pdf}_Y(y) &= \frac{1}{\pi}\int_{0}^{\infty} \Re\left(e^{-ity}\,\mathrm{cf}_Y(t)\right)\,dt \\ &\approx \frac{\delta_t}{\pi}\sum_{j=0}^{N} w_j \Re\left(e^{-it_jy}\,\mathrm{cf}_Y(t_j)\right)\end{aligned}$$

  $$\begin{aligned}\mathrm{cdf}_Y(y) &= \frac{1}{2} - \frac{1}{\pi}\int_{0}^{\infty} \Im\left(\frac{e^{-ity}\,\mathrm{cf}_Y(t)}{t}\right)\,dt \\ &\approx \frac{1}{2} - \frac{\delta_t}{\pi}\sum_{j=0}^{N} w_j \Im\left(\frac{e^{-it_jy}\,\mathrm{cf}_Y(t_j)}{t_j}\right),\end{aligned}$$

- N is sufficiently large integer, w_j are the quadrature weights (for trapezoidal rule use $w_0 = w_N = \frac{1}{2}$, otherwise $w_j = 1$), t_j denote the appropriate nodes (here equidistant) from $(0, T)$, for sufficiently large T, e.g. $T = \frac{N2\pi}{B-A}$, where $(A, B) = \mathrm{mean}(Y) \mp 6\,\mathrm{std}(Y)$.

TECHNICAL DETAILS:

- Particular selection of the values N and T influences the total approximation error, i.e. combination of the truncation error and the integration error, of the used integral approximation based on the trapezoidal quadrature rule. The trade-off between N and T strongly depends on the particular distribution of Y and its CF.
- If the optimum values of N and T are unknown, we suggest, as a simple rule of thumb, to start with the application of the following six-sigma-rule.
- For that, set $\delta_t = 2\pi/(B-A)$, where the interval $(A,B) = \mathrm{E}(Y) \mp k\sqrt{var(Y)}$ with $k = 6$ (or other more suitable value of the multiplication coefficient k) specifies the substantial part of the distribution support of the random variable Y, and then set N and $T = N\delta_t$ such that the value of the integrand function is sufficiently small for all $t > T$, say $\left|\Im\left(e^{-it_j y}\,\mathrm{cf}_Y(t)/t\right)\right| \leq |\mathrm{cf}_Y(t)/t| < \varepsilon$, with $\varepsilon = 10^{-12}$.

FURTHER TECHNICAL DETAILS:

- For computing the first term of the Gil-Pelaez formula for CDF, we can use the following result:

$$\lim_{t \to 0} \Im \left(\frac{e^{-ity}\, \mathrm{cf}_Y(t)}{t} \right) = E(Y) - y.$$

- The required location and scale (dispersion) parameters, i.e. $E(Y)$ and $var(Y)$, can be evaluated either analytically, from the moments of the distribution (i.e. the expectation and the variance of Y, if they exist and are known), or approximately, by using numerical differentiation of the (known) characteristic function of Y, $\mathrm{cf}_Y(t)$. In particular,

$$E(Y) \approx \frac{1}{12ih} \begin{pmatrix} \mathrm{cf}_Y(-2h) - 8\,\mathrm{cf}_Y(-h) \\ +8\,\mathrm{cf}_Y(h) - \mathrm{cf}_Y(2h) \end{pmatrix},$$

$$var(Y) \approx E\left(Y^2\right) - [E(Y)]^2,$$

where

$$E\left(Y^2\right) \approx -\frac{1}{144h^2} \begin{pmatrix} \mathrm{cf}_Y(-4h) - 16\,\mathrm{cf}_Y(-3h) \\ +64\,\mathrm{cf}_Y(-2h) + 16\,\mathrm{cf}_Y(-h) \\ -130 \\ +16\,\mathrm{cf}_Y(h) + 64\,\mathrm{cf}_Y(2h) \\ -16\,\mathrm{cf}_Y(3h) + \mathrm{cf}_Y(4h) \end{pmatrix},$$

for any small $h > 0$, e.g., $h = 10^{-4}$.

The Gil-Pelaez inversion formulae

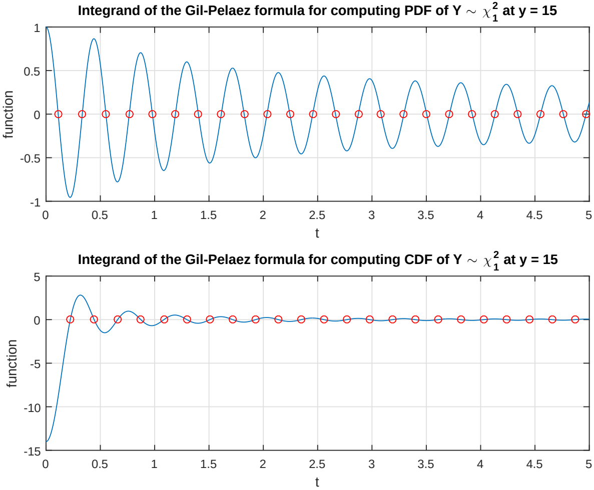


Figure: Example: Integrands of the Gil-Pelaez formulae for computing PDF/CDF. Here $f_{pdf}(t) = \Re\left(e^{-ity}\,\mathrm{cf}_Y(t)\right) = \Re\left(e^{-it15}(1-2it)^{-1/2}\right)$, $f_{cdf}(t) = \Im\left(e^{-it15}(1-2it)^{-1/2}/t\right)$.

Numerical inversion by using the FFT algorithm

- Recall that the continuous Fourier transform (CFT) of the function $f(u)$ is given by

$$F(y) = \int_{-\infty}^{\infty} f(u)\, e^{-i2\pi u y}\, du$$

- CFT can be approximated by the discrete Fourier transform (DFT), which is defined, for the complex numbers $f_0, \ldots, f_{N-1}$, by the following relation

$$F_k = \sum_{j=0}^{N-1} f_k e^{-i2\pi k \frac{j}{N}} \qquad k = 0, \ldots, N-1$$

- Numerically, DFT can be efficiently evaluated by the Fast Fourier Transform (FFT) algorithm:

$$\mathbf{F}_N = \mathrm{FFT}(\mathbf{f}_N)$$

where $\mathbf{f}_N = (f_0, \ldots, f_{N-1})$ and $\mathbf{F}_N = (F_0, \ldots, F_{N-1})$.

Numerical inversion by using the FFT algorithm

- We can represent the inverse Fourier transform

$$\mathrm{pdf}_Y(y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-ity}\,\mathrm{cf}_Y(t)\,dt$$

as CFT and approximate it by DFT/FFT, see e.g., Hürlimann (2013):

- Let N is a sufficiently large integer and (A, B) is a sufficiently large interval (approximate support), where the distribution of Y is concentrated. A reasonable starting rule for (A, B) can be, for example, the six-sigma-rule: $(A, B) = \mathrm{mean}(Y) \mp 6\,\mathrm{std}(Y)$.
- Denote
  - $y_k = A + k\delta_y, \quad \delta_y = \frac{B-A}{N}, \quad k = 0, \dots, N-1,$
  - $t_k = -T + k\delta_t, \quad \delta_t = \frac{2\pi}{B-A}, \quad T = \frac{\pi(N-1)}{B-A}, \quad k = 0, \dots, N-1,$
  - $\mathbf{C} = (C_0, \dots, C_{N-1}), \quad C_k = \frac{1}{B-A}(-1)^{[(1-\frac{1}{N})(\frac{NA}{B-A}+k)]}, \quad k = 0, \dots, N-1,$
  - $\mathbf{D} = (D_0, \dots, D_{N-1}), \quad D_k = (-1)^{\frac{-2A}{B-A}k}, \quad k = 0, \dots, N-1,$
  - $\mathbf{cf}_N = (\mathrm{cf}_Y(t_0), \dots, \mathrm{cf}_Y(t_{N-1})),$
  - $\mathbf{pdf}_N = (\mathrm{pdf}_Y(y_0), \dots, \mathrm{pdf}_Y(y_{N-1})).$
- Then,

$$\mathbf{pdf}_N \approx \mathbf{C} \odot \mathrm{FFT}(\mathbf{D} \odot \mathbf{cf}_N).$$

- CDF can be approximated by interpolation from the simple cumulative sum from the evaluated PDF values. QF can be evaluated by interpolation from the CDF.

Applications

1. Exact distribution of the LRT statistic in normal linear regression

- Let $Y \sim N(X\beta, \sigma^2 I)$ be an n-dimensional normally distributed random vector with full-ranked $(n \times k)$-design matrix X and parameters $\beta \in \mathbb{R}^k$ and $\sigma^2 > 0$.

- Consider testing $H_0 : (\beta, \sigma^2) = (\beta_0, \sigma_0^2)$, vs. $H_A : (\beta, \sigma^2) \neq (\beta_0, \sigma_0^2)$, based on the log-LRT statistic

$$
\begin{aligned}
lrt &= -2\left(\operatorname{loglik}\left(\beta_0, \sigma_0^2 \mid Y\right) - \operatorname{loglik}\left(\hat\beta, \hat\sigma^2 \mid Y\right)\right) \\
&= \frac{1}{\sigma_0^2}(Y - X\beta_0)^T(Y - X\beta_0) - n\log\left(\frac{\hat\sigma^2}{\sigma_0^2}\right) - n \\
&\overset{H_0}{\sim} Q_k + \left\{(Q_\nu - n) - n\log\left(\frac{Q_\nu}{n}\right)\right\} \equiv Q_k + W_\nu,
\end{aligned}
$$

where $Q_k \sim \chi_k^2$ and $Q_\nu \sim \chi_\nu^2$, with $\nu = n - k$, are independent RVs.

- Here, $W_\nu = (Q_\nu - n) - n\log\left(\frac{Q_\nu}{n}\right)$ denotes the log-Lambert $W \times \chi^2$ random variable, with known distribution and its characteristic function.

- That is, under H_0, the LRT statistic is distributed as a linear combination of two independent RVs, with χ_k^2 and $LW(\chi_\nu^2, \theta)$ distributions, where $\nu = n - k$ and $\theta = (n(\log(n) - 1), n, 1)$.

V. Witkovský, G. Wimmer and T. Duby, Logarithmic Lambert $W \times \mathcal{F}$ random variables for the family of chi-squared distributions and their applications, *Statistics & Probability Letters* **96**, 223 (2015).

2. Gaussian processes and their distribution

Let $\{X(\mathbf{t})\}$ denote a centered Gaussian processes defined on $[0,1]^d$ with $d \geq 1$. Denote the covariance function of X by

$$K(\mathbf{t},\mathbf{s}) = E\,(X(\mathbf{t})X(\mathbf{s}))\,,\ \text{for } \mathbf{t},\mathbf{s} \in [0,1]^d,$$

then by Mercer's theorem,

$$K(\mathbf{t},\mathbf{s}) = \sum_{i=1}^{\infty} \lambda_i e_i(\mathbf{t}) e_i(\mathbf{s}),$$

where λ_i and $e_i(\mathbf{t}$ are the set of eigenvalues and normalized eigenvectors of the integral operator corresponding to the covariance function in the sense of

$$\lambda f(\mathbf{t}) = \int_{[0,1]^d} K(\mathbf{t},\mathbf{s}) f(\mathbf{s})\,d\mathbf{s},\ \ \mathbf{t},\mathbf{s} \in [0,1]^d,$$

The well-known Karhunen–Loéve (KL) expansion for Gaussian process $X(\mathbf{t})$ on $[0,1]^d$ is

$$X(\mathbf{t}) = \sum_{i=1}^{\infty} \lambda_i e_i(\mathbf{t}) Z_i,$$

where Z_i are i.i.d. standard normal random variables. Note that $e_i(\mathbf{t})$ forms an orthogonal base in $L_2([0,1]^d)$ and thus a natural consequence of the KL expansion is the distributional identity

$$\int_{[0,1]^d} X^2(\mathbf{t})\,d\mathbf{t} \stackrel{\mathcal{L}}{=} \sum_{i=1}^{\infty} \lambda_i Z_i^2.$$

3. Distribution of the Cramér-von Mises and the Anderson-Darling statistics

The CvM and the A-D statistics belong to the class of quadratic goodness-of-fit test statistics based on the empirical distribution function. By using the theory of stochastic processes, the asymptotic distributions are derived from the KL representation of functionals of the Brownian motion (resp. Brownian bridge) processes. Such approach can be used to decompose also other stochastic processes.

Let $\hat{F}_n(x)$ denotes the empirical CDF based on n i.i.d. observations $X_1, \ldots, X_n$ from continuous distribution F, i.e. $X_j \sim F$.

- The Cramér-von Mises statistic is defined by

$$W_n = n \int_{-\infty}^{\infty} \left(\hat{F}_n(x) - F(x)\right)^2 dF(x) \quad \xrightarrow{D} \quad W_\infty \equiv \int_0^1 B^2(t)\, dt \sim \sum_{j=1}^{\infty} \frac{Z_j^2}{(j\pi)^2},$$

$$\mathrm{cf}_{W_\infty}(t) = \prod_{j=1}^{\infty} \mathrm{cf}_{\chi_1^2}\left(\tfrac{t}{(j\pi)^2}\right) = \prod_{j=1}^{\infty} \left(1 - \tfrac{2it}{(j\pi)^2}\right)^{-\frac{1}{2}} = \sqrt{\frac{\sqrt{2it}}{\sin(\sqrt{2it})}}.$$

- The Anderson-Darling statistic is a (weighted) generalization of W_n, defined by

$$A_n = n \int_{-\infty}^{\infty} \frac{\left(\hat{F}_n(x) - F(x)\right)^2}{F(x)(1-F(x))}\, dF(x) \quad \xrightarrow{D} \quad A_\infty \equiv \int_0^1 \frac{B^2(t)}{t(1-t)}\, dt \sim \sum_{j=1}^{\infty} \frac{Z_j^2}{j(j+1)},$$

$$\mathrm{cf}_{A_\infty}(t) = \prod_{j=1}^{\infty} \mathrm{cf}_{\chi_1^2}\left(\tfrac{t}{j(j+1)}\right) = \prod_{j=1}^{\infty} \left(1 - \tfrac{2it}{j(j+1)}\right)^{-\frac{1}{2}} = \sqrt{\frac{-2\pi it}{\cos\left(\frac{\pi}{2}\sqrt{1+8it}\right)}}.$$

- Analytical inversion of the characteristic function leads to a complicated and computationally rather strange expressions, see Anderson and Darling (1952):

$$Pr(A_\infty \leq z) = \frac{\sqrt{2\pi}}{z} \sum_{j=0}^{\infty} \binom{-\frac{1}{2}}{j} e^{-\frac{(4j+1)^2\pi^2}{8z}} \int_0^\infty \exp\left\{\frac{z}{8(1+w^2)} - \frac{w^2(4j+1)^2\pi^2}{8z}\right\} dw$$

- REMARK: Recently, an efficient method for numerical evaluation of the asymptotic Anderson-Darling distribution, based on a sophisticated recurrence relation, was proposed by Marsaglia and Marsaglia (2004). The algorithm is currently available also in the R-package `goftest`.

2. Distribution of the Cramér-von Mises and the Anderson-Darling statistics

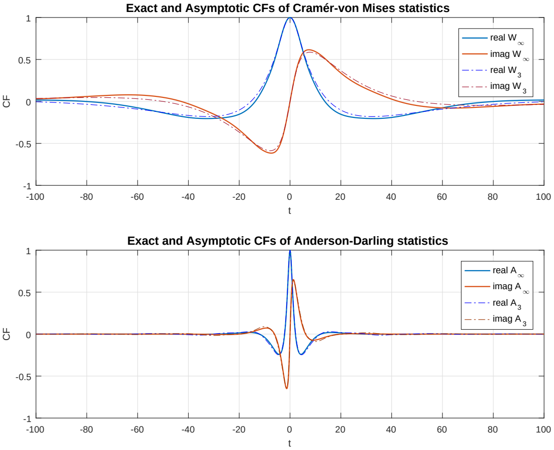


Figure: The exact and the asymptotic CFs of the Cramér-von Mises and the Anderson-Darling statistics. The ’exact’ CFs were approximated by using the Chebyshev polynomials, and based on similar iterative procedure as suggested by Knott (1974) and further improved by Csörgő and Faraway (1996).

3. Distribution of the Cramér-von Mises and the Anderson-Darling statistics

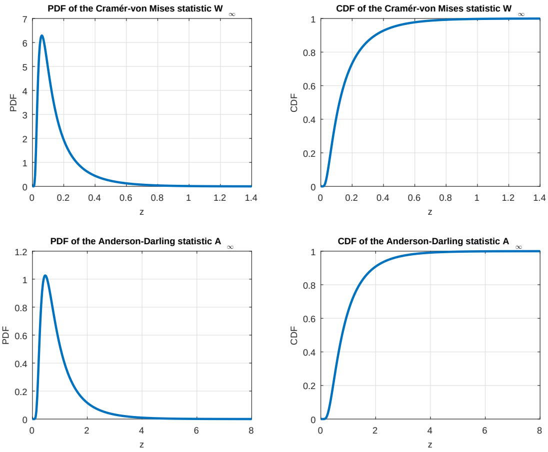


Figure: Distribution functions (PDF/CDF) of the asymptotic distribution of the Cramér-von Mises statistic and the Anderson-Darling statistic calculated by the FFT numerical inversion of their CFs.

- Numerical inversion of the smoothed empirical characteristic function is an alternative method for the kernel density estimation.
- Convolution of the empirical distribution with the kernel (e.g., Gaussian distribution) is an efficient tool for getting smooth PDF estimator.

  The shape of the estimated PDF depends critically on the value of the bandwidth parameter.
- The smoothed ECF is defined as

$$\mathrm{cf}_{\tilde{F}_n}(t) = \mathrm{cf}_{\hat{F}_n}(t) \times \mathrm{cf}_Z(\sigma_b t) = \left(\frac{1}{n}\sum_{j=1}^{n} e^{\mathrm{i}tx_j}\right) \times e^{-\frac{1}{2}\sigma_b^2 t^2} = \frac{1}{n}\sum_{j=1}^{n} e^{\mathrm{i}x_j t - \frac{1}{2}\sigma_b^2 t^2},$$

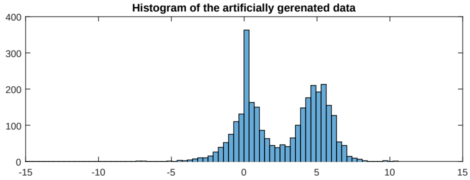
  where $\mathrm{cf}_Z(\sigma_b t)$ is the kernel CF. Here, $\mathrm{cf}_Z(\sigma_b t) = e^{-\frac{1}{2}\sigma_b^2 t^2}$, is CF of a normally distributed RV $Z \sim N(0, \sigma_b^2)$, with a given bandwidth parameter σ_b.

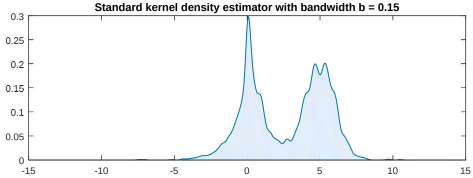
  The resulted distribution defined by the smoothed ECF with Gaussian kernel is an equally weighted mixture of normal distributions.
- The kernel estimate of PDF/CDF can be evaluated by numerical inversion of the smoothed ECF.

4. PDF/CDF estimation by numerical inversion of the smoothed ECF

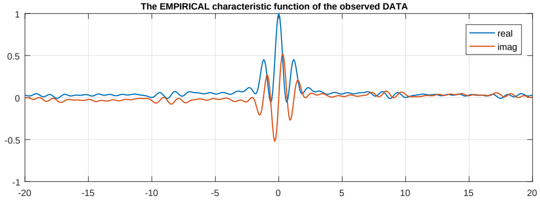
- For illustration, we have generated the data $x_1, \ldots, x_n$:
  A random sample of size $n = 3000$ from a mixture distribution defined by

$$\mathrm{pdf}_F(x) = \frac{3}{6}\,\mathrm{pdf}_{N(5,1)}(x) + \frac{2}{6}\,\mathrm{pdf}_{t_3}(x) + \frac{1}{6}\,\mathrm{pdf}_{\chi^2_1}(x)$$





4. PDF/CDF estimation by numerical inversion of the smoothed ECF



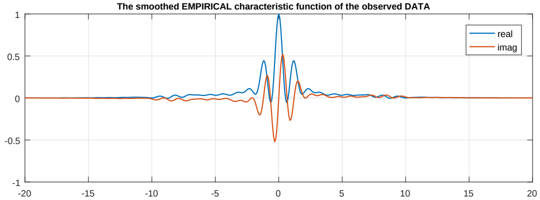


Figure: Plot of the empirical characteristic function and its smoothed version: A convolution of the ECF and the Gaussian kernel distribution with $\sigma^2 = 0.15$.

4. PDF/CDF estimation by numerical inversion of the smoothed ECF

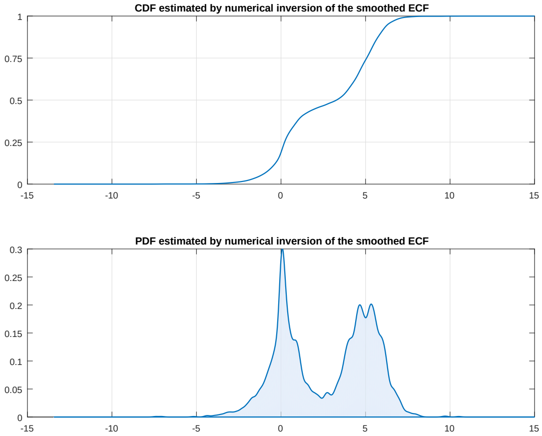


Figure: Plot of the estimated CDF and PDF computed by numerical FFT inversion of the smoothed empirical characteristic function.

5. Distribution of the test statistic for testing symmetry based on ECF

- Feuerverger and Mureika (1977) suggested the following test statistic based on the ECF $\mathrm{cf}_{\hat{F}_n}(t)$ for testing symmetry of the distribution, see also Meintanis *etal* (2016),

$$nT_n = \int_{-\infty}^{\infty} \Im(\mathrm{cf}_{\hat{F}_n}(t))^2\, dG(t) = \frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n}\Big(\mathrm{cf}_G(X_j - X_k) - \mathrm{cf}_G(X_j + X_k)\Big),$$

$X_1, \ldots, X_n$ is a random sample from F, $\hat{F}_n$ is the ECDF, and $G(t)$ is the chosen weighing zero-mean symmetric CDF (e.g. normal, triangular on (-1,1), or U-shaped arcsine on (-1,1)) with its (known) characteristic function $\mathrm{cf}_G(t)$.

- The suggested approximate distribution:

$$nT_n = \int_{-\infty}^{\infty} W_n(t)^2\, dG(t) \xrightarrow{D} \int_{-\infty}^{\infty} W(t)^2\, dG(t) \sim \sum_{j=1}^{\infty}\lambda_j Z_j^2 \approx \sum_{j=1}^{n}\hat{\lambda}_j Z_j^2,$$

$W(t)$ is a zero-mean Gaussian process (with the same covariance function as $W_n(t)$), $Z_j^2 \overset{iid}{\sim} \chi_1^2$, $\hat{\lambda}_j$ are the eigenvalues of the $(n \times n)$-matrix D, defined by its elements

$$D_{jk} = \mathrm{cf}_G(x_j - x_k) - \mathrm{cf}_G(x_j + x_k).$$

- Thus, the approximate FM distribution can be evaluated by numerical inversion of the characteristic function of a linear combination of independent chi-squared random variables: $\hat{\mathrm{cf}}_{FM}(t) = \prod_{j=1}^{n}\left(1 - 2\mathrm{i}\hat{\lambda}_j t\right)^{-1/2}$.

5. Compound distribution derived by inversion of the compound ECF

- In financial risk management, estimation of the operational risk capital requires evaluation of aggregate (compound) loss distributions of the operational loss $L = \sum_{m=1}^{M} S_m$, where S_m is the sum of losses of the mth cell of the portfolio, see e.g., Kaas *etal* (2008).

- As a special case, consider the collective risk model in insurance which requires distribution of the aggregate loss $S = \sum_{j=1}^{N} X_j$ of an insurance portfolio in a certain period of time, e.g. 1 year, defined as a compound distribution of the severity and frequency distributions.

- The sizes of these claims are taken to be iid RVs $X_1, \ldots, X_N$, with $X_j \sim F_X$, where F_X is the continuous severity distribution (e.g., gamma, Pareto, Weibull, log-normal), independent of the random number of insurance claims generated in the given time period N, with $N \sim F_N$, where F_N is the discrete frequency distribution (e.g., Poisson, binomial, negative-binomial).

5. Compound distribution derived by inversion of the compound ECF

- CDF of the collective risk $S = \sum_{j=1}^{N} X_j$ is given as a mixture $F_S = \sum_{n=0}^{\infty} \Pr(N = n) F_X^n$, where F_X^n denotes the n-times convolved distribution, and its CF is given by

$$\mathrm{cf}_S(t) = \mathrm{cf}_N\left(-\mathrm{i}\log\left(\mathrm{cf}_X(t)\right)\right).$$

- Let $\hat{F}_N$ denotes the ECDF of the observed historic figures of the realized claims $n_1, \ldots, n_J$, in each of J historic years, with its ECF $\mathrm{cf}_{\hat{F}_N}(t) = \frac{1}{J}\sum_{j=1}^{J} e^{\mathrm{i}tn_j}$.

  Let $\hat{F}_X$ denotes the ECDF based on K observed historic values of claims $x_1, \ldots, x_K$, with its ECF $\mathrm{cf}_{\hat{F}_X}(t) = \frac{1}{K}\sum_{k=1}^{K} e^{\mathrm{i}tx_k}$. Then, the compound empirical CF is

$$\hat{\mathrm{cf}}_S(t) = \mathrm{cf}_{\hat{F}_N}\left(-\mathrm{i}\log\left(\mathrm{cf}_{\hat{F}_X}(t)\right)\right) = \frac{1}{J}\sum_{j=1}^{J}\left(\frac{1}{K}\sum_{k=1}^{K} e^{\mathrm{i}tx_k}\right)^{n_j}.$$

5. Compound distribution derived by inversion of the compound ECF

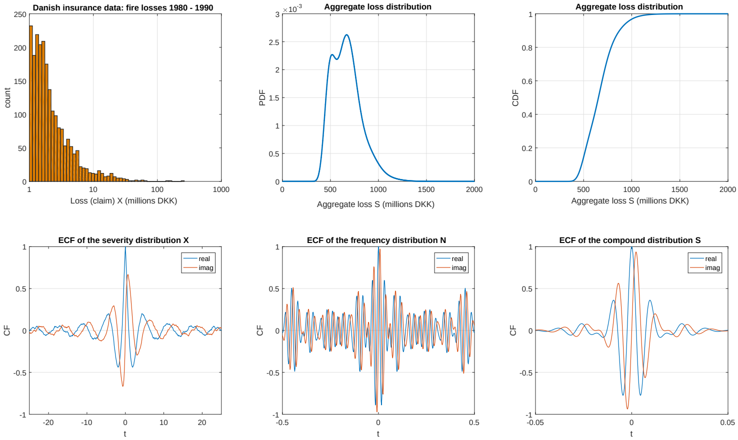


Figure: Compound aggregate loss (collective risk) distribution for 1-year period, based on historic Danish insurance data (2167 fire losses $X \geq 1$ mil. DKK, and 11 frequencies N, with $\bar{n} = 197$, observed in 1980-1990), derived by numerical inversion from ECF $\hat{\mathrm{cf}}_S(t) = \frac{1}{J}\sum_{j=1}^{J}\left(\frac{1}{K}\sum_{k=1}^{K} e^{itx_k}\right)^{n_j}$

5. Compound distribution derived by inversion of the compound ECF

- Combination of the empirical CDF and the fitted generalized Pareto CDF is frequently used for modeling the heavy tailed (severity) distributions, based on the observed data.

- The characteristic function of such distribution, say $\mathrm{cf}_{\widehat{F_X}}(t)$, can be expressed as a weighted mixture of the empirical CF and the generalized Pareto CF,

$$\mathrm{cf}_{\widehat{F_X}}(t) = p \times \mathrm{cf}_{\hat{F}_{X_L}}(t) + (1-p) \times \mathrm{cf}_{GP}(t),$$

  where $p \in (0,1)$ is chosen probability level specifying the tail part of the distribution, typically $p = 0.8$ or $p = 0.9$

- $\mathrm{cf}_{\hat{F}_{X_L}}(t)$ denotes the empirical CF based on the lower p-part of the observed values ($x_k \leq \theta$, where θ is the threshold selected as the p-quantile of the distribution),

- $\mathrm{cf}_{GP}(t)$ denotes the CF of the fitted generalized Pareto distribution, $GP(\xi, \sigma, \theta)$, with the parameters ξ and σ estimated (e.g., by the maximum likelihood estimation method) from the observed values $x_k - \theta \geq 0$, $k = 1, \ldots, K$.

6. Compound distribution derived by inversion of the compound ECF

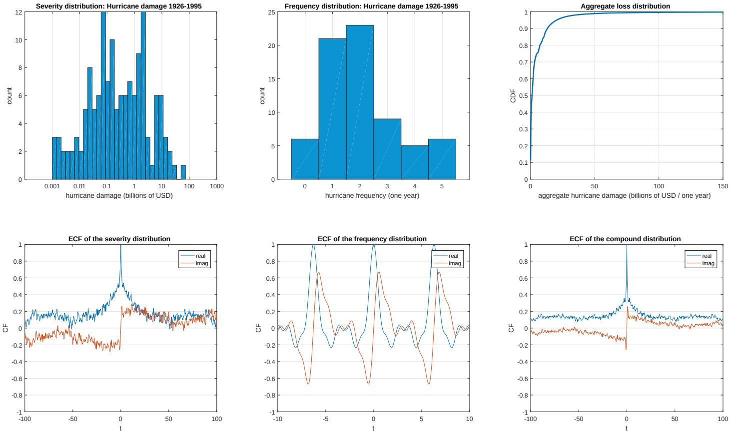


Figure: Hurricane damage data (1926-1995). ICAT developed website to provide easy access to historical hurricane damage information. All information is open source and based upon publicly available data. The data has been normalized to reflect current inflation, wealth, and population from what existed at the time of the actual storm activity, see Pielke, R., Jr., Gratz, J., Landsea, C., Collins, D., Saunders, M., and Musulin, R. (2008). Normalized Hurricane Damage in the United States: 1900–2005. Nat. Hazards Rev. (2008), 29-42.

7. Bootstrap distribution by numerical inversion of the convolved ECF

- In simple cases (e.g., for estimators that are linear functions of the sample), the bootstrap distribution can be derived directly, without any simulations, by the numerical inversion of the n-times convolved empirical CF.
- For example, assume that we want to derive the bootstrap distribution of the sample mean $\bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$, based on the observed data $x_1, \ldots, x_n$.
- Recall that ECF is defined by

$$\mathrm{cf}_{\hat{F}_n}(t) = \frac{1}{n}\sum_{j=1}^{n} e^{\mathrm{i}tx_j},$$

and the bootstrap version of the sample mean is $\bar{X}^* = \frac{1}{n}\sum_{j=1}^{n} X_j^*$, where $X_j^* \sim \hat{F}_n$.

- The distribution of the bootstrap sample mean $\bar{X}^*$ is defined by its bootstrap characteristic function

$$\mathrm{cf}_{\bar{X}^*}(t) = \mathrm{cf}_{\hat{F}_n}\left(\tfrac{t}{n}\right) \times \cdots \times \mathrm{cf}_{\hat{F}_n}\left(\tfrac{t}{n}\right) = \left(\mathrm{cf}_{\hat{F}_n}\left(\tfrac{t}{n}\right)\right)^n = \left(\frac{1}{n}\sum_{j=1}^{n} e^{\mathrm{i}\frac{t}{n}x_j}\right)^n.$$

7. Bootstrap distribution by numerical inversion of the convolved ECF

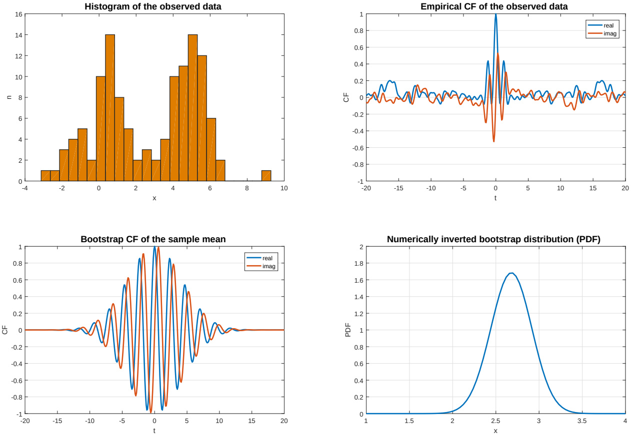


Figure: Numerically inverted distribution of the bootstrap sample mean $\bar{X}^*$, based on $n = 100$ observed data, $x_1, \ldots, x_{100}$, generated from the mixture distribution $\mathrm{pdf}_F(x) = \frac{3}{6}\,\mathrm{pdf}_{N(5,1)}(x) + \frac{2}{6}\,\mathrm{pdf}_{t_3}(x) + \frac{1}{6}\,\mathrm{pdf}_{\chi^2_1}(x)$ with the true mean value $\mu = 2.67$.

8. Re-weighted and smoothed bootstrap distribution of the kth order statistic

- The qth population quantile, for a fixed $q \in (0,1)$, can be estimated by the kth order statistic, where k is suitable chosen order (e.g. $k = \lfloor q \times (n-1) + 1 \rfloor$).
- Let $\hat{F}_n$ denotes the empirical CDF, based on the observed values $x_1, \ldots, x_n$. Then the bootstrap distribution of the kth order statistic, say $\hat{F}_k^*$, is a discrete CDF function with possible non-zero steps (jumps) concentrated at distinct values of the observed data $x_1, \ldots, x_n$.
- The bootstrap CDF of the kth order statistic $\hat{F}_k^*$ can be evaluated at $x_j \in \{x_1, \ldots, x_n\}$ directly (without simulations) as

$$\hat{F}_k^*(x_j) = B\left(\hat{F}_n(x_j), k, n+1-k\right),$$

  where $B(z, a, b)$ is the incomplete beta function.
- It is known that the bootstrap confidence intervals for the qth population quantile have notably poor coverages, especially for the large population quantiles. Different approaches have been suggested to improve the coverage probability of the bootstrap confidence intervals, see e.g. Ho and Lee (2005).
- Here we illustrate alternative methods for smoothing and re-weighting the bootstrap distribution of the kth order statistic, which is used as an estimator of the qth population quantile.

8. Re-weighted and smoothed bootstrap distribution of the kth order statistic

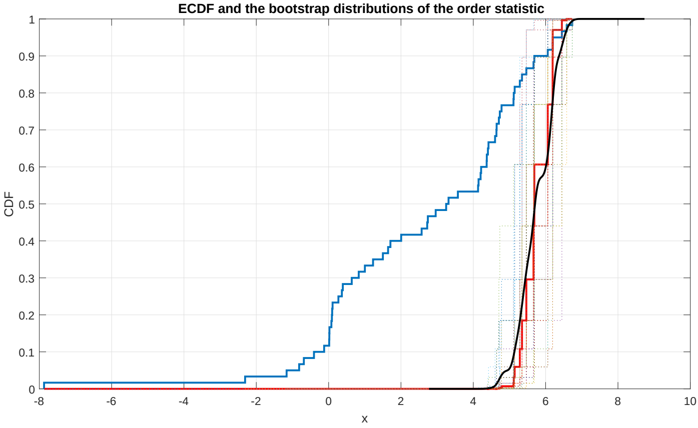


Figure: The empirical CDF (blue) and the bootstrap distribution (red) of the kth order statistic ($k = \lfloor 0.9 \times (n-1) + 1 \rfloor = 54$), based on $n = 60$ observed data, $x_1, \ldots, x_{60}$, generated from the mixture distribution $\mathrm{pdf}_F(x) = \frac{3}{6}\,\mathrm{pdf}_{N(5,1)}(x) + \frac{2}{6}\,\mathrm{pdf}_{t_3}(x) + \frac{1}{6}\,\mathrm{pdf}_{\chi^2_1}(x)$ plotted together with the re-weighted and smoothed bootstrap distribution (black), re-weighted from other (iterated) bootstrap distributions of the kth order statistic, based on the bootstrapped samples $x_1^*, \ldots, x_{60}^*$.

Conclusions

- The characteristic functions represent a complete characterization of the distribution of the random variables. However, analytical inversion of the characteristic functions (if it is possible and available) frequently lead to a complicated and computationally rather strange expressions of the corresponding PDF/CDF.

- As an alternative, here we advocate to **use the efficient methods for numerical inversion of the characteristic functions** - based e.g. on the Gil-Pelaez inversion formulae with trapezoidal rule, used for the required integration, or based on the computationally efficient fast FFT algorithm.

- **At present time, the standard statistical software packages (as e.g. R, SAS, MATLAB) to not offer efficient tools and algorithms for computing, combining and inverting the characteristic functions.**

THANK YOU!

REFERENCES

- Abate, J., Whitt, W. (1992). The Fourier-series method for inverting transforms of probability distributions. *Queueing systems*, 10(1-2), 5-87.
- Anderson, T. W., Darling, D. A. (1952). Asymptotic theory of certain goodness of fit criteria based on stochastic processes. *The Annals of Mathematical Statistics*, 193-212.
- Asheim, A., Huybrechs, D. (2013). Complex Gaussian quadrature for oscillatory integral transforms. *IMA Journal of Numerical Analysis*, drs060.
- Bohman, H. (1970). A method to calculate the distribution function when the characteristic function is known. *BIT Numerical Mathematics*, 10(3), 237-242.
- Bohman, H. (1972). From characteristic function to distribution function via Fourier analysis. *BIT Numerical Mathematics*, 12(3), 279-283.
- Csörgő, S., Faraway, J. J. (1996). The exact and asymptotic distributions of Cramér-von Mises statistics. *Journal of the Royal Statistical Society*. Series B (Methodological), 221-234.
- R. Davies, Algorithm AS 155: The distribution of a linear combinations of χ^2 random variables, *Applied Statistics* **29**, 232 (1980).
- Feuerverger, A., Mureika, R. A. (1977). The empirical characteristic function and its applications. *The Annals of Statistics*, 88-97.
- Gil-Pelaez, J. (1951). Note on the inversion theorem. *Biometrika*, 38(3-4), 481-482.
- Ho, Y. H., Lee, S. M. (2005). Iterated smoothed bootstrap confidence intervals for population quantiles. *The Annals of Statistics*, 33(1), 437-462.
- Hürlimann, W. (2013). Improved FFT approximations of probability functions based on modified quadrature rules. In *International Mathematical Forum* (Vol. 8, No. 17, pp. 829-840).

REFERENCES

- Imhof, J. P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48(3/4), 419-426.
- Kaas, R., Goovaerts, M., Dhaene, J., Denuit, M. (2008). *Modern Actuarial Risk Theory: Using R* (Vol. 128). Springer Science & Business Media.
- Knott, M. (1974). The distribution of the Cramér-von Mises statistic for small sample sizes. *Journal of the Royal Statistical Society*. Series B (Methodological), 36(3), 430-438.
- Levin, D. (1996). Fast integration of rapidly oscillatory functions. *Journal of Computational and Applied Mathematics*, 67(1), 95-101.
- Marsaglia, G., Marsaglia, J. (2004). Evaluating the Anderson-Darling distribution. *Journal of Statistical Software*, 9(2), 1-5.
- Meintanis, S. G., Allison, J., Santana, L. (2016). Goodness-of-fit tests for semiparametric and parametric hypotheses based on the probability weighted empirical characteristic function. *Statistical Papers*. Accepted for publication 2016.
- Milovanović, G. V. (1998). Numerical calculation of integrals involving oscillatory and singular kernels and some applications of quadratures. *Computers & Mathematics with Applications*, 36(8), 19-39.
- Shephard, N. G. (1991). From characteristic function to distribution function: a simple framework for the theory. *Econometric theory*, 7(04), 519-529.
- Sidi, A. (1982). The numerical evaluation of very oscillatory infinite integrals by extrapolation. *Mathematics of Computation*, 38(158), 517-529.

REFERENCES

- Sidi, A. (1988). A user-friendly extrapolation method for oscillatory infinite integrals. *Mathematics of Computation*, 51(183), 249-266.
- Sidi, A. (2012). A user-friendly extrapolation method for computing infinite range integrals of products of oscillatory functions. *IMA Journal of Numerical Analysis*, 32(2), 602-631.
- Strawderman, R. L. (2004). Computing tail probabilities by numerical Fourier inversion: the absolutely continuous case. *Statistica Sinica*, 175-201.
- Waller, L. A., Turnbull, B. W., Hardin, J. M. (1995). Obtaining distribution functions by numerical inversion of characteristic functions with applications. *The American Statistician*, 49(4), 346-350.
- V. Witkovský, On the exact computation of the density and of the quantiles of linear combinations of t and F random variables, *Journal of Statistical Planning and Inference* **94**, 1 (2001).
- V. Witkovský, Matlab algorithm TDIST: The distribution of a linear combination of Student's t random variables, in *COMPSTAT 2004 Symposium*, ed. J. Antoch (Physica-Verlag/Springer, Heidelberg, Germany, 2004) pp. 1995–2002.
- V. Witkovský, Computing the distribution of a linear combination of inverted gamma variables, *Kybernetika* **37**, 79 (2001).
- V. Witkovský, G. Wimmer and T. Duby, Logarithmic Lambert $W \times \mathcal{F}$ random variables for the family of chi-squared distributions and their applications, *Statistics & Probability Letters* **96**, 223 (2015).
- Zieliński, R. (2001). High-Accuracy Evaluation of the Cumulative Distribution Function of α-Stable Symmetric Distributions. *Journal of Mathematical Sciences*, 105(6), 2630-2632.