An expectile computation cookbook

Abdelaati Daouia[1†], Gilles Stupfler[2†], Antoine Usseglio-Carleve[3*†]

[1]Toulouse School of Economics, University of Toulouse Capitole, France
ORCID: 0000-0003-2621-8860.

[2*]Univ Angers, CNRS, LAREMA, SFR MATHSTIC, F-49000 Angers, France
ORCID: 0000-0003-2497-9412.

[3]Avignon Université, LMA UPR 2151, 84000 Avignon, France
ORCID: 0000-0002-8148-3758.

*Corresponding author(s). E-mail(s): antoine.usseglio-carleve@univ-avignon.fr;
Contributing authors: abdelaati.daouia@tse-fr.eu; gilles.stupfler@univ-angers.fr;
†These authors contributed equally to this work.

Abstract

A substantial body of work in the last 15 years has shown that expectiles constitute an excellent candidate for becoming a standard tool in probabilistic and statistical modeling. Surprisingly, the question of how expectiles may be efficiently calculated has been left largely untouched. We fill this gap by, first, providing a general outlook on the computation of expectiles that relies on the knowledge of analytic expressions of the underlying distribution function and mean residual life function. We distinguish between discrete distributions, for which an exact calculation is always feasible, and continuous distributions, where a Newton-Raphson approximation algorithm can be implemented and a list of exceptional distributions whose expectiles are in analytic form can be given. When the distribution function and/or the mean residual life is difficult to compute, Monte-Carlo algorithms are introduced, based on an exact calculation of sample expectiles and on the use of control variates to improve computational efficiency. We discuss the relevance of our findings to statistical practice and provide numerical evidence of the performance of the considered methods.

Keywords: Control variates, Exact computation, Expectiles, Monte-Carlo sampling, Newton-Raphson method, Quadratic convergence.

1 Introduction

Expectiles are least squares analogs of quantiles and define an important probabilistic concept that characterizes a probability distribution just as the quantile function does. The expectiles of a given probability distribution μ on $\mathbb{R}$, endowed with its Borel $\sigma-$algebra, are obtained by minimizing the following asymmetric squared loss problem:

$$\xi_\tau = \underset{\theta\in\mathbb{R}}{\arg\min} \int_{\mathbb{R}} (\eta_\tau(x-\theta) - \eta_\tau(x))\mu(\mathrm{d}x),$$
$$\text{with } \eta_\tau(x) = |\tau - \mathbb{1}_{\{x\le 0\}}|\, x^2 \text{ and } \tau \in (0,1). \quad (1)$$

The expectile ξ_τ is well-defined, finite and unique for each $\tau \in (0,1)$ if and only if μ has a finite first moment, *i.e.* $\int_{\mathbb{R}} |x|\mu(\mathrm{d}x) < \infty$. In this case, $\xi_{1/2} = \int_{\mathbb{R}} x\,\mu(\mathrm{d}x)$ is the expectation of μ, and two probability distributions with finite first moment are equal if and only if they have the same expectiles: the latter property was first noted by Newey and Powell (1987), for sufficiently regular distributions. The original motivation for the use of expectiles was to test for homoskedasticity and conditional symmetry in linear regression problems.

The concept of expectiles has recently gathered substantial momentum for a number of reasons, including the fact that they induce the only law-invariant, coherent (Artzner et al., 1999) and elicitable (Gneiting, 2011) risk measure, and they also define the only $M-$quantiles (Breckling and Chambers, 1988) that are coherent risk measures, see Bellini et al. (2014) and Ziegel (2016). Expectiles can thus be viewed as canonical risk measures for their ability to simultaneously abide by the diversification principle in finance and insurance and to be backtested in a straightforward manner. For these reasons and other probabilistic merits, including the fact that they account for both the frequency of tail observations and their values, unlike quantiles which only rely on frequencies, as well as various angles of interpretation they benefit from (see nine of them in Philipps, 2022), considerable effort has gone into expectile estimation and inference over the past 15 years. Prominent among many statistical works are the contributions of Taylor (2008), Kuan et al. (2009) and Sobotka and Kneib (2012) from a methodological perspective, Holzmann and Klar (2016) and Krätschmer and Zähle (2017) for deep asymptotic results about the estimation of expectiles of fixed order τ, or Daouia et al. (2018, 2020, 2021) and Girard et al. (2021) for the estimation of tail expectiles, obtained as $\tau \uparrow 1$, in heavy-tailed models that describe well the tail structure of many actuarial, financial and environmental data.

One key difficulty in working with expectiles is that they are rarely available in closed or analytic form. Indeed, it is a simple exercise to show that the loss function giving rise to the expectile ξ_τ in (1) is differentiable, so that ξ_τ must cancel the first derivative of this loss function. As a consequence, the τth expectile is the unique x satisfying the equation

$$\frac{\varphi(x)}{2\varphi(x)+x-m} = 1-\tau, \text{ with} \tag{2}$$
$$\varphi(x) = \int_{\mathbb{R}} (t-x)\mathbb{1}_{\{t>x\}}\,\mu(\mathrm{d}t),\ m = \int_{\mathbb{R}} t\,\mu(\mathrm{d}t) = \xi_{1/2}.$$

It follows that, as observed by Jones (1994), expectiles are themselves quantiles of the transformed distribution function E defined by $E(x) = 1 - \varphi(x)/(2\varphi(x)+x-m)$, built on the function φ which is very closely related to the so-called mean residual life function. The issue at play here is that while E is explicit in a wide range of commonly used probabilistic models, it is very often impossible to invert in closed or analytic form, even if the distribution function of μ can be inverted to produce explicit quantiles. An instructive example is the Pareto distribution with extreme value index $\gamma > 0$, having distribution function $x \mapsto 1-x^{-1/\gamma}$ for $x > 1$, for which characterization (2) leads to the equation

$$(1-\gamma)(1-\tau)\xi_\tau^{1/\gamma} - (1-\tau)\xi_\tau^{1/\gamma-1} - \gamma(2\tau-1) = 0.$$

This equation cannot, in general, be solved in closed form for every $\tau \in (0,1)$: for example, when $1/\gamma$ is an integer greater than or equal to 5, it is well-known (as a consequence of Galois theory) that it cannot be solved in radicals. By contrast, it is immediate that the τth quantile of this Pareto distribution is $q_\tau = (1-\tau)^{-\gamma}$. It is even more complex to work with expectiles in other examples such as the Weibull distribution, whose quantiles are known in simple closed form but whose mean residual life function can in general only be expressed in terms of the upper incomplete gamma function. Despite these difficulties, the question of how to compute expectiles for a given distribution has been left largely untouched, even though it is crucial when it comes to assessing the quality of expectile estimation methods in practice. The state of the art in expectile computation appears to be mainly based on solving Equation (2) through either a bisection search, implemented in the `R` package `expectreg` (Otto-Sobotka et al., 2022), or, in the `R` package `ExtremeRisks` (Padoan and Stupfler, 2020), using quasi-Newton techniques that are only valid for a very small set of distributions. When the function φ is intractable, the approach in `ExtremeRisks` resorts to naive Monte-Carlo sampling together with a quasi-Newton method.

Our contribution is to provide more efficient strategies for the calculation of expectiles. We first attack this problem by reformulating Equation (2) as a fixed point equation (called identification equation in Z-estimation) involving the function φ. We then show that, for discrete distributions, the function φ is piecewise linear with explicit coefficients, so that the identification equation consists of a collection of linear equations, one and one only of which has a solution. As such, expectiles of a discrete distribution can always be exactly calculated. Then, we note that for distributions having a continuous density function with respect to the Lebesgue measure, the identification equation amounts to finding the unique root of a convex function. This motivates a Newton-Raphson algorithm which is readily implemented when φ is explicitly determined or at least can be accurately calculated using standard software: this encompasses, among many others, the logistic, Weibull and Gaussian distributions, as well as their mixtures. We show that this Newton-Raphson algorithm has quadratic convergence and will therefore converge faster than existing methods, whose convergence is linear. We also show that, in the challenging scenario of extreme expectile calculation for heavy-tailed distributions, the relative error of the Newton-Raphson approximation converges quadratically, which provides the right scale on which to measure the accuracy of the Newton-Raphson algorithm for tail expectiles. We moreover discuss a list of exceptional continuous distributions whose expectiles can be expressed in closed or analytic form, by solving low-degree polynomial equations or transcendental equations derived from (2). In doing so, we expand upon preliminary work undertaken by Koenker (1993), Zou (2014) and Bellini and Di Bernardino (2017), where only a handful of such distributions appeared. These expressions are useful in order to check that expectile computation algorithms are correct, and in certain statistical simulation contexts where the exact value of the expectile must be known.

For continuous distributions whose distribution function and/or mean residual life function is difficult to compute, we provide another angle of attack by revisiting Monte-Carlo computation. More precisely, if one can simulate independent data points $X_1, \ldots, X_n$ from the distribution μ, then sample expectiles derived by solving (1) for the empirical distribution $\widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$ (where δ_{X_i} is the Dirac probability mass at X_i) are nothing but least asymmetrically weighted squares (LAWS) estimators of the true expectiles, known to be consistent and asymptotically normal under reasonable conditions. We first leverage the fact that these LAWS estimators are expectiles of a discrete uniform distribution, in order to show that they can be exactly calculated, therefore bypassing the use of iteratively reweighted least squares or quasi-Newton methods for their calculation, which is prevalent in the state of the art. We then show that if the expectation m of the distribution is known, the performance of this Monte-Carlo approach can be substantially improved by using the expectation as a control variate. When the target expectile is extreme, lying in the right tail of μ, we construct an analogous algorithm based on the use of a suitably chosen extreme quantile as control variate. We quantify the improvement in the variance of Monte-Carlo sample expectiles in both settings. We illustrate the performance of these algorithms in examples spanning compound Poisson processes, time series and stochastic differential equations, all chosen for their relevance to financial and actuarial practice.

The paper is organized as follows. We start by investigating the calculation of expectiles using their characteristic equation in terms of the mean residual life in Section 2. We then introduce and study Monte-Carlo algorithms based on the LAWS estimator of the target expectile in Section 3. Both sections are illustrated by theoretical and finite-sample examples of application of our results and algorithms. We conclude with a discussion of our work and potential extensions in Section 4. Appendix A contains all mathematical proofs, Appendix B provides the details of the calculations for some of our examples, and Appendix C gives a catalog of expectile functions of continuous distributions, including a list of exceptional cases where expectiles can be found in closed or analytic form. Our methods have been incorporated into the R package `Expectrem`[1].

2 Expectile computation with the mean residual life

Let μ be a probability measure on $\mathbb{R}$ endowed with its Borel $\sigma-$algebra, and assume that μ has a finite first moment, *i.e.* $\int_{\mathbb{R}} |x|\mu(\mathrm{d}x) < \infty$. The characterization of ξ_τ (where $\tau \in (0, 1)$) as the critical

[1] Available on GitHub at https://github.com/AntoineUC/Expectrem

point of the loss function in (1) leads to (2), or equivalently to

$$\frac{2\tau-1}{1-\tau}\varphi(\xi_\tau)+m=\xi_\tau. \tag{3}$$

This means that ξ_τ is the unique solution of a fixed point equation or, equivalently, the unique root of

$$g_\tau : x \mapsto \frac{2\tau-1}{1-\tau}\varphi(x)+m-x.$$

We note that, in this expression, the function φ is equivalently rewritten as

$$\begin{aligned}\varphi(x) &= \int_{\mathbb{R}} t\mathbb{1}_{\{t>x\}}\,\mu(\mathrm{d}t) - x\mu((x,+\infty)) \\ &= \mu((x,+\infty))\left(\frac{\int_{\mathbb{R}} t\mathbb{1}_{\{t>x\}}\,\mu(\mathrm{d}t)}{\mu((x,+\infty))}-x\right).\end{aligned}$$

The quantity $\overline{F}(x)=\mu((x,+\infty))$ is nothing but the probability $\mathbb{P}(X>x)$ if X is a random variable having distribution μ, *i.e.* the survival function associated to μ, and

$$\begin{aligned}\frac{\varphi(x)}{\mu((x,+\infty))} &= \frac{\int_{\mathbb{R}} t\mathbb{1}_{\{t>x\}}\,\mu(\mathrm{d}t)}{\mu((x,+\infty))}-x \\ &= \mathbb{E}(X-x\,|\,X>x)=e(x)\end{aligned}$$

is the so-called mean residual life $e(x)$ above level x, obtained by subtracting x to the expected shortfall $\mathrm{ES}(x)=\mathbb{E}(X\,|\,X>x)$ (also called conditional Value-at-Risk in actuarial mathematics). In other words, the functions φ and g_τ have a closed or analytic form as soon as the survival function and mean residual life/expected shortfall related to μ do.

Solving the nonlinear equation (3) in closed or analytic form is in general impossible, even if φ and g_τ can be written in closed or analytic form. For discrete distributions, however, the functions φ and g_τ are in fact piecewise linear, which makes it possible to compute expectiles exactly, and also to give explicit formulae for the expectiles of a number of classical distributions. This is the focus of the next section.

2.1 Discrete distributions: Exact computation

Here is a general result for the computation of expectiles of discrete distributions, where we allow the set of indices I to be of the form $\{1,\ldots,n\}$ (distribution on a finite set), $\mathbb{N}$ (for one-tailed discrete distributions with countable support) or $\mathbb{Z}$ (for two-tailed discrete distributions).

Theorem 2.1. *Assume that μ is a discrete distribution with finite first moment, supported on the set $\{a_i,\, i\in I\}$, where the a_i are arranged in increasing order, with probability mass function $p_i=\mu(\{a_i\})>0$ for any $i\in I$. Fix $\tau\in(0,1)$. Then there is a unique index $i=i(\tau)$ such that the inequalities*

$$\begin{aligned}&\frac{\sum_{k<i}p_k(a_i-a_k)}{\sum_{k<i}p_k(a_i-a_k)+\sum_{k>i}p_k(a_k-a_i)}\leq\tau \\ &<\frac{\sum_{k<i+1}p_k(a_{i+1}-a_k)}{\sum_{k<i+1}p_k(a_{i+1}-a_k)+\sum_{k>i+1}p_k(a_k-a_{i+1})}\end{aligned}$$

hold. With this index i,

$$\begin{aligned}\xi_\tau &= \frac{\tau\sum_{k>i}p_ka_k+(1-\tau)\sum_{k\leq i}p_ka_k}{\tau\sum_{k>i}p_k+(1-\tau)\sum_{k\leq i}p_k} \\ &= \frac{(2\tau-1)\sum_{k>i}p_ka_k+(1-\tau)\sum_{k\in I}p_ka_k}{(2\tau-1)\sum_{k>i}p_k+1-\tau}.\end{aligned}$$

It should be noted that since the inequalities in Theorem 2.1 involve the tail probability $\mu((a_i,+\infty))=\sum_{k>i}p_k$ and the tail mean $\sum_{k>i}p_ka_k$ of the distribution μ, they can be used to search for the relevant index $i=i(\tau)$, and therefore to give an explicit value of the expectile ξ_τ when the survival function and expected shortfall of the distribution μ are explicit. Even when this is not the case, one can always find the index $i=i(\tau)$ numerically and therefore compute the exact value of ξ_τ.

We draw two corollaries of Theorem 2.1. The first one deals with the case when the support of μ is finite with n elements, in which case there are at most $n-1$ inequalities to check in order to calculate the exact value of the expectile.

Corollary 2.1. *Assume that μ is a distribution supported on the finite set $\{a_i,\, 1\leq i\leq n\}$, where $a_1<a_2<\cdots<a_n$, with probability mass function $p_i=\mu(\{a_i\})>0$. Fix $\tau\in(0,1)$. Then there is a unique index $i=i(\tau)\in\{1,\ldots,n-1\}$ such that the inequalities*

$$\frac{\sum_{k=1}^{i-1}p_k(a_i-a_k)}{\sum_{k=1}^{i-1}p_k(a_i-a_k)+\sum_{k=i+1}^{n}p_k(a_k-a_i)}\leq\tau$$

$$< \frac{\sum_{k=1}^{i} p_k(a_{i+1}-a_k)}{\sum_{k=1}^{i} p_k(a_{i+1}-a_k) + \sum_{k=i+2}^{n} p_k(a_k - a_{i+1})}$$

hold. With this index i,

$$\begin{aligned}\xi_\tau &= \frac{\tau \sum_{k=i+1}^{n} p_k a_k + (1-\tau)\sum_{k=1}^{i} p_k a_k}{\tau \sum_{k=i+1}^{n} p_k + (1-\tau)\sum_{k=1}^{i} p_k} \\ &= \frac{(2\tau-1)\sum_{k=i+1}^{n} p_k a_k + (1-\tau)\sum_{k=1}^{n} p_k a_k}{(2\tau-1)\sum_{k=i+1}^{n} p_k + 1 - \tau}.\end{aligned}$$

The second corollary focuses on the setting where the distribution is not only supported on a finite set but is also uniform. This is relevant to expectile estimation in statistical applications, where only a finite sample of observations from a given distribution is available; if the underlying distribution is continuous, then the (realized) empirical distribution of the observations is discrete and uniform on the set of data points (see Example 2.6 below).

Corollary 2.2. *Assume that μ is the uniform distribution on the set $\{a_i, 1 \le i \le n\}$, where $a_1 < a_2 < \cdots < a_n$. Fix $\tau \in (0,1)$. Then there is a unique index $i = i(\tau) \in \{1, \ldots, n-1\}$ such that the inequalities*

$$\begin{aligned}&\frac{\sum_{k=1}^{i-1}(a_i - a_k)}{\sum_{k=1}^{i-1}(a_i - a_k) + \sum_{k=i+1}^{n}(a_k - a_i)} \le \tau \\ &< \frac{\sum_{k=1}^{i}(a_{i+1} - a_k)}{\sum_{k=1}^{i}(a_{i+1}-a_k) + \sum_{k=i+2}^{n}(a_k - a_{i+1})}\end{aligned}$$

hold. With this index i,

$$\begin{aligned}\xi_\tau &= \frac{\tau \sum_{k=i+1}^{n} a_k + (1-\tau)\sum_{k=1}^{i} a_k}{\tau(n-i) + (1-\tau)i} \\ &= \frac{\tau \sum_{k=i+1}^{n} a_k + (1-\tau)\sum_{k=1}^{i} a_k}{\tau n - (2\tau-1)i}.\end{aligned}$$

We give below a few examples as applications of the above results.

Example 2.1 (Bernoulli and Rademacher distributions)**.** For the Bernoulli distribution with parameter $p \in (0,1)$, it immediately follows from Corollary 2.1 that

$$\xi_\tau = \frac{\tau p}{(2\tau-1)p + 1 - \tau} \quad \text{for any } \tau \in (0,1).$$

In particular, the expectile function of the Bernoulli distribution with parameter $1/2$ is $\xi_\tau = \tau$, and for the Rademacher distribution putting equal probability $1/2$ on the values 1 and -1, the expectile function is $\xi_\tau = (1+\tau)/2$.

Example 2.2 (Distribution supported on a set with three elements)**.** Let μ be the probability distribution on $\{0,1,2\}$ with $\mu(\{1\}) = p$ and $\mu(\{2\}) = q$, with $p, q > 0$ and $p + q < 1$. Then, from Corollary 2.1,

$$\xi_\tau = \begin{cases} \dfrac{\tau(p+2q)}{(2\tau-1)(p+q)+1-\tau} & \text{for } \tau \le 1 - \dfrac{q}{1-p}, \\ \dfrac{2\tau q + (1-\tau)p}{(2\tau-1)q + 1 - \tau} & \text{otherwise.} \end{cases}$$

Appendix B contains more general expressions when μ is the distribution on a set of three points $\{a, b, c\}$ as well as an application to the distribution of the sum of two independent Bernoulli random variables.

Example 2.3 (Uniform distribution on $\{1, \ldots, n\}$)**.** Fix $n \ge 2$. For the uniform distribution on $\{1, \ldots, n\}$, solving the inequalities of Corollary 2.2 is equivalent to finding the unique index $i \in \{1, \ldots, n-1\}$ such that

$$\begin{aligned}&\frac{i(i-1)}{i(i-1)+(n-i)(n-i+1)} \le \tau \\ &\qquad < \frac{i(i+1)}{i(i+1)+(n-i)(n-i-1)}.\end{aligned}$$

This is equivalent to finding the unique solution (known to exist, by Corollary 2.2) to the inequalities $P_\tau(i+1) < 0 \le P_\tau(i)$ for $i \in \{1, \ldots, n-1\}$, where P_τ is the polynomial

$$P_\tau(x) = (2\tau-1)x^2 - \{2\tau(n+1)-1\}x + \tau n(n+1).$$

Straightforward calculations, found in Appendix B, then entail

$$\xi_\tau = \begin{cases} \dfrac{n(n+1)}{2} & \text{when } \tau = 1/2, \\ \dfrac{\tau n(n+1) - (2\tau-1)\lfloor x_\tau \rfloor(\lfloor x_\tau \rfloor + 1)}{2\tau n - 2(2\tau-1)\lfloor x_\tau \rfloor} & \text{otherwise,} \end{cases}$$

with

$$x_\tau = \frac{2\tau(n+1) - 1 - \sqrt{4\tau(1-\tau)(n+1)(n-1)+1}}{2(2\tau-1)}.$$

Example 2.4 (Geometric distribution)**.** For the geometric distribution with success probability $p \in (0,1)$, namely, $\mu(\{k\}) = p(1-p)^{k-1}$ for any positive integer k, the inequalities of Theorem 2.1 read as

$$\frac{(1-p)^i - (1-pi)}{2(1-p)^i - (1-pi)} \le \tau < \frac{(1-p)^{i+1} - (1-p(i+1))}{2(1-p)^{i+1} - (1-p(i+1))}.$$

Solving these inequalities is equivalent to finding the index $i \ge 1$ such that $h_\tau(i+1) < 0 \le h_\tau(i)$, where

$$h_\tau(x) = (2\tau-1)(1-p)^x - (1-\tau)(px-1).$$

Straightforward calculations, found in Appendix B and involving the transcendental equation defining the main branch of the Lambert W function, reveal that

$$\xi_\tau = \frac{(2\tau-1)(1-p)^{\lfloor x_\tau \rfloor}(1+p\lfloor x_\tau \rfloor) + 1 - \tau}{p\{(2\tau-1)(1-p)^{\lfloor x_\tau \rfloor} + 1 - \tau\}}$$
$$\text{with } x_\tau = \frac{1}{p} - \frac{W\left(-\frac{(1-p)^{1/p}\log(1-p)}{p}\frac{2\tau-1}{1-\tau}\right)}{\log(1-p)}.$$

The main branch of the Lambert function is available numerically in `R` using (for instance) the `gsl` package (Hankin et al., 2023), acting as a wrapper for the GNU Scientific Library.

Example 2.5 (Poisson distribution)**.** For the Poisson distribution with parameter $\lambda > 0$, namely, $\mu(\{k\}) = e^{-\lambda}\lambda^k/k!$ for any nonnegative integer k, the inequalities of Theorem 2.1 are

$$\frac{iF_\lambda(i) - \lambda F_\lambda(i-1)}{2(iF_\lambda(i) - \lambda F_\lambda(i-1)) - (i-\lambda)} \le \tau < \frac{(i+1)F_\lambda(i+1) - \lambda F_\lambda(i)}{2((i+1)F_\lambda(i+1) - \lambda F_\lambda(i)) - (i+1-\lambda)},$$

where $F_\lambda(i) = \sum_{k=0}^{i} \frac{\lambda^k}{k!} e^{-\lambda}$ denotes the distribution function of the Poisson distribution, which is readily calculated using the function `ppois` in base `R`. It follows that

$$\begin{aligned}\xi_\tau &= \lambda \frac{\tau - (2\tau-1)F_\lambda(i_\tau - 1)}{\tau - (2\tau-1)F_\lambda(i_\tau)} \\ &= \lambda\left(1 + \frac{(2\tau-1)\lambda^{i_\tau}/i_\tau!}{\tau e^\lambda - (2\tau-1)\sum_{k=0}^{i_\tau}\lambda^k/k!}\right),\end{aligned}$$

where i_τ is the unique nonnegative integer i such that

$$(2\tau-1)(iF_\lambda(i) - \lambda F_\lambda(i-1)) - \tau(i-\lambda) \ge 0 \text{ and}$$
$$(2\tau-1)((i+1)F_\lambda(i+1) - \lambda F_\lambda(i)) - \tau(i+1-\lambda) < 0.$$

Example 2.6 (Sample expectiles)**.** Suppose that $X_1, \ldots, X_n$ are independent random draws from a distribution μ having a continuous distribution function, and consider the empirical probability measure $\widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ where δ_a is the Dirac probability mass at a. Then the τth sample expectile is the τth expectile of the empirical measure $\widehat{\mu}_n$, that is,

$$\begin{aligned}\widehat{\xi}_{\tau,n} &= \arg\min_{\theta\in\mathbb{R}} \int_{\mathbb{R}} (\eta_\tau(x-\theta) - \eta_\tau(x))\widehat{\mu}_n(\mathrm{d}x) \\ &= \arg\min_{\theta\in\mathbb{R}} \sum_{i=1}^n \left|\tau - \mathbb{1}_{\{X_i \le \theta\}}\right| (X_i - \theta)^2.\end{aligned}$$

This is the so-called Least Asymmetrically Weighted Squares (LAWS) estimator of ξ_τ. Since F is continuous, there are no ties within the X_i with probability 1, so that, according to Corollary 2.2,

$$\widehat{\xi}_{\tau,n} = \frac{\tau\sum_{k=i+1}^n X_{k:n} + (1-\tau)\sum_{k=1}^i X_{k:n}}{\tau n - (2\tau-1)i}$$

where $X_{1:n} < X_{2:n} < \cdots < X_{n:n}$ are the order statistics of the sample $(X_1, \ldots, X_n)$, and $i = \widehat{i}_n(\tau) \in \{1, \ldots, n-1\}$ is the unique index such that the inequalities

$$\frac{\sum_{k=1}^{i-1}(X_{i:n} - X_{k:n})}{\sum_{k=1}^{i-1}(X_{i:n} - X_{k:n}) + \sum_{k=i+1}^n (X_{k:n} - X_{i:n})} \le \tau < \frac{\sum_{k=1}^{i}(X_{i+1:n} - X_{k:n})}{\sum_{k=1}^{i}(X_{i+1:n} - X_{k:n}) + \sum_{k=i+2}^n (X_{k:n} - X_{i+1:n})}$$

hold. This means that, at the price of a numerical search for this index i, the point estimate

$\widehat{\xi}_{\tau,n}$ can be exactly computed. To the best of our knowledge, this had not been noticed before in the statistical literature, with the preferred methods in `R` so far apparently being either an iteratively reweighted least squares algorithm in the `expectreg` package (Otto-Sobotka et al., 2022) or a quasi-Newton method via the `optim` routine using the `method="BFGS"` argument in the `ExtremeRisks` package (Padoan and Stupfler, 2020).

Example 2.7 (Kernel expectile regression)**.** Let $(\boldsymbol{X}_1, Y_1), \ldots, (\boldsymbol{X}_n, Y_n)$ be independent random copies of a continuous random pair $(\boldsymbol{X}, Y) \in \mathbb{R}^p \times \mathbb{R}$. Let $\mu_{\boldsymbol{x}}$ denote the conditional probability distribution of Y given $\boldsymbol{X} = \boldsymbol{x}$ (this is well-defined by disintegration of probability measures on $\mathbb{R}^{p+1}$) and let $F(\cdot|\boldsymbol{x})$ be its distribution function. Let g be the probability density function of $\boldsymbol{X}$ on $\mathbb{R}^p$ and fix $\boldsymbol{x} \in \mathbb{R}^p$ with $g(\boldsymbol{x}) > 0$. Consider the standard kernel estimator of $F(\cdot|\boldsymbol{x})$:

$$\widehat{F}_n(y|\boldsymbol{x}) = \frac{1}{nh_n^p\,\widehat{g}_n(\boldsymbol{x})} \sum_{i=1}^n \mathbb{1}_{\{Y_i \le y\}} K\left(\frac{\boldsymbol{x} - \boldsymbol{X}_i}{h_n}\right)$$
$$\text{with } \widehat{g}_n(\boldsymbol{x}) = \frac{1}{nh_n^p} \sum_{i=1}^n K\left(\frac{\boldsymbol{x} - \boldsymbol{X}_i}{h_n}\right).$$

Here K is a kernel probability distribution function on $\mathbb{R}^p$ and h_n is a (positive) bandwidth, with $\widehat{g}_n$ being the Parzen-Rosenblatt estimator of g. Then the smoothed empirical distribution $\widehat{\mu}_{\boldsymbol{x},n}$ associated to $\widehat{F}_n(\cdot|\boldsymbol{x})$ is discrete and its expectiles

$$\widehat{\xi}_{\tau,n}(\boldsymbol{x}) = \arg\min_{\theta\in\mathbb{R}} \int_{\mathbb{R}} (\eta_\tau(y-\theta) - \eta_\tau(y))\widehat{\mu}_{\boldsymbol{x},n}(\mathrm{d}y),$$

called kernel regression expectiles (Girard et al., 2022; Daouia et al., 2023), can in fact be exactly calculated, while the standard up to now in the literature seems to have been restricted to Brent's method via the `uniroot` function in `R`. Indeed, by Corollary 2.1, one has

$$\widehat{\xi}_{\tau,n}(\boldsymbol{x}) = \frac{\tau \sum_{k=i+1}^n \omega_{k,n}(\boldsymbol{x}) Y_{k:n}}{\tau \sum_{k=i+1}^n \omega_{k,n}(\boldsymbol{x}) + (1-\tau)\sum_{k=1}^i \omega_{k,n}(\boldsymbol{x})} + \frac{(1-\tau)\sum_{k=1}^i \omega_{k,n}(\boldsymbol{x}) Y_{k:n}}{\tau \sum_{k=i+1}^n \omega_{k,n}(\boldsymbol{x}) + (1-\tau)\sum_{k=1}^i \omega_{k,n}(\boldsymbol{x})}$$

where $\omega_{k,n}(\boldsymbol{x}) = K((\boldsymbol{x} - \boldsymbol{X}_{[k:n]})/h_n)$ and $\boldsymbol{X}_{[k:n]}$ is the covariate value concomitant to the order statistic $Y_{k:n}$ (*i.e.* $\boldsymbol{X}_{[k:n]} = \boldsymbol{X}_j$ if and only if $Y_{k:n} = Y_j$), and $i = \widehat{\imath}_{\boldsymbol{x},n}(\tau) \in \{1, \ldots, n-1\}$ is the unique index such that the following inequalities hold:

$$\frac{\sum_{k=1}^{i-1} \omega_{k,n}(\boldsymbol{x}) Z_{k,i:n}}{\sum_{k=1}^{i-1} \omega_{k,n}(\boldsymbol{x}) Z_{k,i:n} - \sum_{k=i+1}^n \omega_{k,n}(\boldsymbol{x}) Z_{k,i:n}} \le \tau < \frac{\sum_{k=1}^{i} \omega_{k,n}(\boldsymbol{x}) Z_{k,i+1:n}}{\sum_{k=1}^{i} \omega_{k,n}(\boldsymbol{x}) Z_{k,i+1:n} - \sum_{k=i+2}^n \omega_{k,n}(\boldsymbol{x}) Z_{k,i+1:n}},$$

where $Z_{k,i:n} = Y_{i:n} - Y_{k:n}$.

2.2 Continuous distributions: A Newton-Raphson algorithm

The computation of expectiles for a continuous distribution is in general more difficult, and apart from a few exceptions, has to be done numerically. It is easy to show that, when $\tau < 1/2$, the τth expectile $\xi_\tau = \xi_\tau(\mu)$ is linked to the $(1-\tau)$th expectile of the probability measure ν, uniquely determined by its values on half-lines as $\nu((-\infty, t]) = \mu([-t, +\infty))$, as $\xi_\tau(\mu) = -\xi_{1-\tau}(\nu)$. Therefore, we focus in this section on the case where $\tau > 1/2$ and the distribution function $F : x \mapsto \mu((-\infty, x])$ related to μ is continuous.

Since $\varphi(x) = \int_x^\infty \overline{F}(t)\,\mathrm{d}t$ with $\overline{F} = 1 - F$, the function φ is absolutely continuous and non-increasing, and so is the function g_τ whose unique root is ξ_τ. A simple idea to calculate ξ_τ is then to use a bisection search. This is the current standard in the `R` package `expectreg` (Otto-Sobotka et al., 2022), whose routines rely on a bisection search and on a closed form of the function g_τ, or at least on the latter being written in terms of standard special functions available in `R`. However, the reliance on a bisection search makes these routines not only relatively slow, but also inaccurate for large τ, because interval bounds for the bisection search are built *a priori* in the routines without possible change by the user. The bisection search has also been recently paired with quantum computing in Laudagé and Turkalj (2022). An alternative option, used in the `ExtremeRisks` package for a very small set of distributions, is Brent's method which, roughly speaking, pairs a bisection search with a quasi-Newton method (Padoan and Stupfler, 2020), and converges at a linear rate as the bisection search does.

A faster alternative is found by exploiting the fact that φ is actually convex, from which it follows that g_τ is convex as well. A convergent approximation is then given by the Newton-Raphson iterative method for the root of the equation $g_\tau(x) = 0$:

$$\begin{aligned} x_{n+1} &= x_n - \frac{g_\tau(x_n)}{g'_\tau(x_n)} \\ &= \frac{(2\tau-1)(\varphi(x_n) + x_n\overline{F}(x_n)) + (1-\tau)m}{(2\tau-1)\overline{F}(x_n) + 1 - \tau}. \end{aligned}$$

If μ is the probability distribution of a random variable X, it is interesting to note that the iterative algorithm can be rewritten as

$$x_{n+1} = \frac{(2\tau-1)\mathbb{E}(X\mathbb{1}_{\{X>x_n\}}) + (1-\tau)\mathbb{E}(X)}{(2\tau-1)\mathbb{P}(X > x_n) + 1 - \tau}.$$

The approximation x_n thus depends, like the target expectile ξ_τ, on the values and frequencies of tail observations from X.

The starting point x_0 of the algorithm can be any value smaller than ξ_τ. Since the expectile function is monotonic, one should in practice at least choose $x_0 > \xi_{1/2} = m$, the mean of the distribution under consideration. This Newton-Raphson algorithm is readily implemented when $\overline{F}$ and φ have a simple closed form or are efficiently calculated numerically; a list of classical examples of continuous distributions with their respective values of $\overline{F}$ and φ is provided in Table 2. Our next main result makes the rate of convergence of the proposed Newton-Raphson algorithm explicit.

Theorem 2.2. *Let μ be a distribution with finite first moment and having a density function f with respect to the Lebesgue measure. Fix $\tau > 1/2$ and a starting point $x_0 < \xi_\tau$, and assume that f is continuous on $[x_0, \xi_\tau]$. Then the Newton-Raphson sequence of iterates (x_n) is nondecreasing, converges to ξ_τ and satisfies*

$$|x_{n+1} - \xi_\tau| \le \frac{1}{2}\frac{(2\tau-1)\max_{[x_n,\xi_\tau]} f}{1-\tau+(2\tau-1)\overline{F}(x_n)}|x_n - \xi_\tau|^2.$$

for all $n \in \mathbb{N}$. In particular, for all $n, p \in \mathbb{N}$,

$$\begin{aligned} |x_{n+p} - \xi_\tau| \le \left(\frac{1}{2}\frac{(2\tau-1)\max_{[x_p,\xi_\tau]} f}{1-\tau+(2\tau-1)\overline{F}(\xi_\tau)}\right)^{2^n-1} \\ \times |x_p - \xi_\tau|^{2^n}. \end{aligned}$$

One should of course expect the Newton-Raphson method to be much faster than the bisection search and Brent's method, since the former has a quadratic rate of convergence, by Theorem 2.2, while the latter only have linear rates of convergence.

Among interesting cases, the following corollary concentrates on the situation where the probability density function of μ is in fact decreasing on a suitable interval containing the target value, in which case one obtains a cruder, but often useful, control on the error involving the hazard function of μ.

Corollary 2.3. *Under the conditions of Theorem 2.2, if f is moreover nonincreasing on $[x_0, \xi_\tau]$, then (x_n) satisfies*

$$\forall n \in \mathbb{N},\ |x_{n+1} - \xi_\tau| \le \frac{1}{2}h(x_n)|x_n - \xi_\tau|^2$$

where $h(x) = f(x)/\overline{F}(x)$. In particular, if $h(x)$ is bounded by h_0 on $[x_0, \xi_\tau]$, then

$$\forall n, p \in \mathbb{N},\ |x_{n+p} - \xi_\tau| \le \left(\frac{h_0}{2}\right)^{2^n-1}|x_p - \xi_\tau|^{2^n}.$$

The next corollary examines the important situation when f is bounded. In this case, one can find a very simple control on the error, which will actually be sharper than the above control using the hazard rate for τ not too close to 1.

Corollary 2.4. *Under the conditions of Theorem 2.2, if f is moreover bounded by M on $[x_0, \xi_\tau]$, then (x_n) satisfies*

$$\forall n \in \mathbb{N},\ |x_{n+1} - \xi_\tau| \le \frac{(2\tau-1)M}{2(1-\tau)}|x_n - \xi_\tau|^2.$$

In particular, for all $n, p \in \mathbb{N}$,

$$|x_{n+p} - \xi_\tau| \le \left(\frac{(2\tau-1)M}{2(1-\tau)}\right)^{2^n-1}|x_p - \xi_\tau|^{2^n}.$$

An insightful consequence of Corollary 2.4 is on the computation of expectiles ξ_τ with $(2\tau - 1)/(1-\tau) \in \mathbb{N}\setminus\{0\}$, *i.e.* $\tau = \tau_m = (m+1)/(m+2) = 2/3, 3/4, 4/5, \ldots$ [These levels τ_m make the factor in front of $\varphi(\xi_\tau)$ in Equation (3) an integer.] Corollary 2.4 yields

$$\forall n \in \mathbb{N},\ |x_{n+1} - \xi_{\tau_m}| \le \frac{m}{2}\sup_{\mathbb{R}} f \times |x_n - \xi_{\tau_m}|^2.$$

Such simple bounds are useful in the definition of stopping criteria for the iterative algorithm. For $m = 2$ and $\tau = 3/4 = 0.75$, for example, we obtain

$$|x_n - \xi_{3/4}| \leq \frac{1}{\sup_{\mathbb{R}} f} \left(\sup_{\mathbb{R}} f \times |x_0 - \xi_{3/4}| \right)^{2^n}.$$

for all $n \in \mathbb{N}$. If the starting point x_0 can be chosen such that $|x_0 - \xi_{3/4}| < 1/(2 \sup_{\mathbb{R}} f)$ (for instance following a preliminary evaluation of g_τ on a grid of step size $1/(2 \sup_{\mathbb{R}} f)$), then

$$\forall n \in \mathbb{N}, \ |x_n - \xi_{3/4}| \leq \frac{2^{-2^n}}{\sup_{\mathbb{R}} f}.$$

In this situation the approximation x_n is guaranteed to be accurate within 10^{-k} of the target value $\xi_{3/4}$ as soon as

$$n > \frac{1}{\log 2} \log \left(\frac{k \log 10 - \log(\sup_{\mathbb{R}} f)}{\log 2} \right).$$

This number of iterations grows logarithmically fast in the number of significant digits k.

We now highlight a few examples where this construction of the Newton-Raphson algorithm for the computation of expectiles applies without difficulty. It should be noted that if μ is a mixture of distributions, say $\mu = \sum_{j=1}^d p_j \mu_j$, where the probabilities p_j add up to 1, then, with obvious notation, $\overline{F} = \sum_{j=1}^d p_j \overline{F}_j$ and $\varphi = \sum_{j=1}^d p_j \varphi_j$. As a consequence, if one can write a Newton-Raphson algorithm for each of the μ_j, then writing a Newton-Raphson algorithm for any of their mixtures is straightforward. This principle is used in Example 2.9 below and can also be applied to construct an expectile computation algorithm for a mixture of Gaussian distributions, in conjunction with Example 2.11.

Example 2.8 (Logistic distribution)**.** Consider the centered logistic distribution with unit scale, having survival function $\overline{F}(x) = 1/(1 + e^x)$ and probability density function $f(x) = e^x/(1 + e^x)^2$. Straightforward calculus leads to $\varphi(x) = \log(1 + e^{-x})$, so the iterations of the Newton-Raphson algorithm are

$$x_{n+1} = (2\tau - 1) \times \frac{(1 + \exp(x_n)) \log(1 + \exp(x_n)) - x_n \exp(x_n)}{\tau + (1 - \tau) \exp(x_n)}.$$

It is readily shown that $f'(x) = e^x(1 - e^x)/(1 + e^x)^3 < 0$ on $(0, \infty)$, so f is continuous and decreasing on the positive half-line. The hazard function of the logistic distribution is given by $h(x) = e^x/(1 + e^x)$, which is bounded above by 1. Then, according to Corollary 2.3,

$$|x_{n+1} - \xi_\tau| \leq \frac{1}{2} |x_n - \xi_\tau|^2$$

when $\tau > 1/2$, for any starting point $x_0 \in (0, \xi_\tau)$.

Example 2.9 (Hall-Weiss distribution)**.** The Hall-Weiss distribution has the survival function $\overline{F}(x) = (1/2)x^{-\alpha} + (1/2)x^{-\alpha-\beta}$ for $x > 1$, where $\alpha > 0$ and $\beta \geq 0$; the case $\beta = 0$ produces the Pareto distribution. We consider the case $\alpha > 1$, which is necessary and sufficient for this distribution to have a finite first moment. Obviously, for all $x > 1$,

$$\varphi(x) = \frac{1}{2} \left(\frac{1}{\alpha - 1} x^{1-\alpha} + \frac{1}{\alpha + \beta - 1} x^{1-\alpha-\beta} \right)$$

so the iterations of the Newton-Raphson algorithm are

$$x_{n+1} = \frac{(2\tau - 1)\left(\frac{\alpha}{\alpha-1} x_n^{1-\alpha} + \frac{\alpha+\beta}{\alpha+\beta-1} x_n^{1-\alpha-\beta} \right)}{(2\tau - 1)(x_n^{-\alpha} + x_n^{-\alpha-\beta}) + 2(1 - \tau)} + \frac{\frac{2(\alpha-1)(\alpha+\beta)+\beta}{(\alpha-1)(\alpha+\beta-1)}(1 - \tau)}{(2\tau - 1)(x_n^{-\alpha} + x_n^{-\alpha-\beta}) + 2(1 - \tau)}$$

Since, for $x > 1$, $2f(x) = \alpha x^{-\alpha-1} + (\alpha + \beta)x^{-\alpha-\beta-1}$, f is clearly decreasing on $(1, \infty)$. Straightforward calculations yield the hazard function as

$$\forall x > 1, \ h(x) = \frac{1}{x} \left(\alpha + \frac{\beta}{x^\beta + 1} \right).$$

Conclude, from Corollary 2.3, that when $\tau > 1/2$, whatever the starting point $x_0 \in [1, \xi_\tau)$ of the algorithm,

$$\begin{aligned} |x_{n+1} - \xi_\tau| &\leq \frac{1}{2x_n} \left(\alpha + \frac{\beta}{x_n^\beta + 1} \right) |x_n - \xi_\tau|^2 \\ &\leq \frac{1}{2} \left(\alpha + \frac{\beta}{2} \right) |x_n - \xi_\tau|^2. \end{aligned}$$

Example 2.10 (Weibull distribution)**.** Consider the Weibull distribution with unit scale and shape parameter $\beta > 0$, having survival function $\overline{F}(x) = \exp(-x^\beta)$, $x > 0$. Here

$$\varphi(x) = \left\{\Gamma\left(1+\frac{1}{\beta}\right) - \Gamma_{x^\beta}\left(1+\frac{1}{\beta}\right)\right\} - x\exp(-x^\beta), \quad \text{for } x > 0,$$

where $\Gamma_x(a) = \int_0^x t^{a-1}e^{-t}\,dt$ is the lower incomplete Gamma function and $\Gamma(a) = \Gamma_\infty(a)$ is the Gamma function at a. The mean of this Weibull distribution is $m = \Gamma(1+1/\beta)$, and thus the Newton-Raphson iterative sequence is given by

$$x_{n+1} = \frac{(2\tau-1)\{\Gamma(1+1/\beta) - \Gamma_{x_n^\beta}(1+1/\beta)\}}{(2\tau-1)\exp(-x_n^\beta)+1-\tau} + \frac{(1-\tau)\Gamma(1+1/\beta)}{(2\tau-1)\exp(-x_n^\beta)+1-\tau}.$$

The incomplete Gamma function can be computed using `pgamma` and multiplying its output by the (complete) Gamma function found through the `gamma` routine in base `R`. The probability density function of the Weibull distribution is $f(x) = \beta x^{\beta-1}\exp(-x^\beta)$, $x > 0$. For $\beta \le 1$ the function f is decreasing on $(0,\infty)$, so Corollary 2.3 applies and yields, for any starting point $x_0 \in [m, \xi_\tau) = [\Gamma(1+1/\beta), \xi_\tau)$,

$$|x_{n+1}-\xi_\tau| \le \frac{\beta}{2}x_n^{\beta-1}|x_n-\xi_\tau|^2 \le \frac{\beta}{2}x_0^{\beta-1}|x_n-\xi_\tau|^2 \le \frac{\beta}{2}\{\Gamma(1+1/\beta)\}^{\beta-1}|x_n-\xi_\tau|^2.$$

In particular, for the unit exponential distribution (with $\beta = 1$), $|x_{n+1}-\xi_\tau| \le (1/2)|x_n-\xi_\tau|^2$. For $\beta > 1$, Corollary 2.3 cannot be used, but Corollary 2.4 applies because

$$\max_{x>0} f(x) = \beta^{1/\beta}(\beta-1)^{1-1/\beta}e^{-(1-1/\beta)}.$$

Using Corollary 2.4 then yields

$$|x_{n+1}-\xi_\tau| \le \frac{2\tau-1}{2(1-\tau)}\beta^{1/\beta}(\beta-1)^{1-1/\beta}e^{-(1-1/\beta)}|x_n-\xi_\tau|^2.$$

It is worth noting that Corollary 2.4 does not apply when $\beta < 1$ because f is then not bounded. It does however apply in the special case of the exponential distribution, thus yielding

$$|x_{n+1}-\xi_\tau| \le \frac{2\tau-1}{2(1-\tau)}|x_n-\xi_\tau|^2.$$

By combining this with the inequality $|x_{n+1} - \xi_\tau| \le (1/2)|x_n-\xi_\tau|^2$, which is more precise if and only if $\tau \in (2/3,1)$, we obtain, for the exponential distribution specifically, that for any starting point $x_0 \in [1,\xi_\tau)$,

$$|x_{n+1}-\xi_\tau| \le \frac{|x_n-\xi_\tau|^2}{2}\begin{cases}\dfrac{2\tau-1}{1-\tau} & \text{if } \tau \in [1/2, 2/3], \\ 1 & \text{if } \tau \in (2/3,1).\end{cases}$$

Example 2.11 (Gaussian distribution)**.** Consider the standard Gaussian distribution with density function $\phi : x \mapsto (2\pi)^{-1/2}\exp(-x^2/2)$ and distribution function Φ. A simple calculation yields $\varphi(x) = \phi(x) - x(1-\Phi(x))$, so the Newton-Raphson iterations are

$$x_{n+1} = \frac{(2\tau-1)\phi(x_n)}{\tau-(2\tau-1)\Phi(x_n)}.$$

The functions ϕ and Φ can be computed using, for example, `dnorm` and `pnorm` in base `R`. Since ϕ is decreasing on the positive half-line, Corollary 2.3 applies and provides, when $\tau > 1/2$ and for any starting point $x_0 \in (0,\xi_\tau)$,

$$|x_{n+1}-\xi_\tau| \le \frac{1}{2r(x_n)}|x_n-\xi_\tau|^2 \quad \text{with } r(x) = \frac{1-\Phi(x)}{\phi(x)} = \text{Mills' ratio}.$$

Using the inequality $1/r(x) \le (x+\sqrt{4+x^2})/2$ due to Birnbaum (1942), we find

$$|x_{n+1}-\xi_\tau| \le \frac{x_n+\sqrt{4+x_n^2}}{4}|x_n-\xi_\tau|^2 \le \frac{x_n+1}{2}|x_n-\xi_\tau|^2.$$

We illustrate how Corollary 2.4 leads to a better control for moderately large values of τ. Since

$\phi(x) \leq 1/\sqrt{2\pi}$ for any x, we obtain from Corollary 2.4 that

$$|x_{n+1} - \xi_\tau| \leq \frac{1}{2} \times \frac{2\tau - 1}{(1-\tau)\sqrt{2\pi}} |x_n - \xi_\tau|^2.$$

Since, for $\tau \geq 1/2$, $(2\tau - 1)/((1-\tau)\sqrt{2\pi}) < 1$ if and only if $\tau \in [1/2, (1+\sqrt{2\pi})/(2+\sqrt{2\pi})] \approx [1/2, 0.778]$, the latter bound is more precise than the former at least when the Gaussian expectile to be computed is smaller than the "third expectile-quartile" $\xi_{3/4}$.

When the target level τ tends to 1, the quantity

$$\frac{1}{2} \frac{(2\tau - 1) \max_{[x_0, \xi_\tau]} f}{1 - \tau + (2\tau - 1)\overline{F}(\xi_\tau)},$$

appearing in Theorem 2.2, diverges since its denominator tends to 0 as $\tau \uparrow 1$. One can then expect that the convergence of the Newton-Raphson method for the calculation of extreme expectiles is typically slower than for central expectiles, and the bounds in Theorem 2.2 are less useful for extreme expectile calculation. The intuition here is that, at least for unbounded distributions, measuring the quality of the approximate solution x_n of the extreme expectile computation problem on the standard scale is too difficult, because the target value ξ_τ will be very large. We conclude this section with a result showing that, when μ is a heavy-tailed probability distribution, the right scale on which to measure the accuracy of the Newton-Raphson approximation of extreme expectiles is the relative scale.

Theorem 2.3. *Under the conditions of Theorem 2.2, if f fulfills the von Mises condition $xf(x)/\overline{F}(x) \to 1/\gamma$ as $x \to \infty$, where $\gamma \in (0,1)$, and if $x_0 > (1-\varepsilon)\xi_\tau$ for some $\varepsilon > 0$, then (x_n) satisfies*

$$\left|\frac{x_{n+1}}{\xi_\tau} - 1\right| \leq \left((1-\varepsilon)^{-1/\gamma - 1} \frac{1/\gamma - 1}{2} + r(\tau)\right) \times \left|\frac{x_n}{\xi_\tau} - 1\right|^2, \quad \text{for all } n \in \mathbb{N},$$

where $r(\tau) = r(\tau, \mu, \varepsilon) \to 0$ as $\tau \uparrow 1$. In particular

$$\forall n, p \in \mathbb{N}, \ \left|\frac{x_{n+p}}{\xi_\tau} - 1\right| \leq \left((1-\varepsilon)^{-1/\gamma - 1} \frac{1/\gamma - 1}{2} + r(\tau)\right)^{2^n - 1} \left|\frac{x_p}{\xi_\tau} - 1\right|^{2^n}.$$

One can give an asymptotic equivalent of the function r using so-called second-order properties of $\overline{F}$ in a neighborhood of infinity and an asymptotic expansion of ξ_τ as $\tau \uparrow 1$, provided in Daouia et al. (2020). We omit this discussion for the sake of brevity.

An important benefit of having fast and accurate expectile computation algorithms is that it allows one to construct expectile tables and draw the expectile function on $(0,1)$, just as one would construct quantile tables and draw the quantile function for reference distributions. We give a few examples of expectile tables in Tables C1–C4 for the standard Gaussian, log-normal, Student and chi-squared distributions, as well as graphical comparisons between quantile and expectile curves for these same distributions in Figures C1–C4. In Table 1 below, we also illustrate the difference in computational time when applying the Newton-Raphson algorithm to the calculation of standard Gaussian, chi-squared, log-normal and Student expectiles compared to pre-implemented routines part of the `expectreg` package in `R` (Otto-Sobotka et al., 2022). It is readily seen that the Newton-Raphson algorithm is typically four times faster than the methods from the `expectreg` package. The gap in performance decreases as one approaches the right tail of the distribution: this is due to the fact that the starting point from the Newton-Raphson is set to be the mean of the distribution, which in this case is far from the target value. We also note that in two instances (log-normal and Student with 2 degrees of freedom), the default current implementation of the `expectreg` routines (version 0.52) failed to converge to the expectile at level $\tau = 0.9995$, the reason being that the target expectile lies outside the pre-specified bounds for the bisection search employed in these routines.

2.3 Continuous distributions: Exceptional cases

2.3.1 Closed-form expressions through low-degree polynomial equations

Equation (3) is a nonlinear equation whose solution generally does not exist in closed form. An

Distribution	Value of τ Method used	Average/median computational time
Gaussian	$\tau = 0.75$ Newton-Raphson `expectreg` v0.52	88.3 μs / 35.3 μs 308 μs / 199 μs
Gaussian	$\tau = 0.95$ Newton-Raphson `expectreg` v0.52	138 μs / 58.1 μs 356 μs / 255 μs
Gaussian	$\tau = 0.9995$ Newton-Raphson `expectreg` v0.52	154 μs / 85.2 μs 288 μs / 204 μs
Chi-squared $\nu = 5$	$\tau = 0.75$ Newton-Raphson `expectreg` v0.52	119 μs / 65.4 μs 507 μs / 355 μs
Chi-squared $\nu = 5$	$\tau = 0.95$ Newton-Raphson `expectreg` v0.52	146 μs / 83.1 μs 535 μs / 376 μs
Chi-squared $\nu = 5$	$\tau = 0.9995$ Newton-Raphson `expectreg` v0.52	195 μs / 145 μs 397 μs / 321 μs
Log-normal	$\tau = 0.75$ Newton-Raphson `expectreg` v0.52	171 μs / 61.3 μs 408 μs / 287 μs
Log-normal	$\tau = 0.95$ Newton-Raphson `expectreg` v0.52	175 μs / 84.1 μs 376 μs / 329 μs
Log-normal	$\tau = 0.9995$ Newton-Raphson `expectreg` v0.52	236 μs / 190 μs Convergence failed
Student $\nu = 2$	$\tau = 0.75$ Newton-Raphson `expectreg` v0.52	139 μs / 117 μs 822 μs / 769 μs
Student $\nu = 2$	$\tau = 0.95$ Newton-Raphson `expectreg` v0.52	150 μs / 122 μs 609 μs / 446 μs
Student $\nu = 2$	$\tau = 0.9995$ Newton-Raphson `expectreg` v0.52	212 μs / 169 μs Convergence failed
Student $\nu = 4$	$\tau = 0.75$ Newton-Raphson `expectreg` v0.52	116 μs / 98.4 μs 751 μs / 639 μs
Student $\nu = 4$	$\tau = 0.95$ Newton-Raphson `expectreg` v0.52	144 μs / 111 μs 642 μs / 486 μs
Student $\nu = 4$	$\tau = 0.9995$ Newton-Raphson `expectreg` v0.52	297 μs / 253 μs 604 μs / 520 μs

Table 1: Comparison between the Newton-Raphson algorithm and the `enorm`, `echisq`, `elnorm` and `et` routines of the `expectreg` package run with their default settings. Computational times reported in the third column are based on 1,000 consecutive calls to each function. The stopping criterion for the Newton-Raphson algorithm is the same as in the `expectreg` routines, *i.e.* the algorithm stops when the evaluation of the distribution function $E : x \mapsto 1-\varphi(x)/(2\varphi(x)+x-m)$ at the approximated expectile is within 10^{-10} of the level τ.

interesting subproblem is to consider this equation when it is in fact polynomial, which typically (but not exclusively) happens when the function φ is a rational function. Then, if the resulting equation is polynomial with degree ≤ 4, a solution can always be found in closed form. When the degree is 3 or 4, this can be done using the Cardano or Ferrari formulae, which we recall in Appendix B. We explain below in more detail how this idea can be used in a few examples; the distributions we consider are all heavy-tailed with extreme value index 1/4, representing the borderline situation where a fourth moment exists. The existence of a fourth moment is important in a number of problems, including for showing the asymptotic normality of maximum likelihood estimators in time series (Brockwell and Davis, 1991; Francq and Zakoïan, 2004), and the index 1/4 is sometimes considered a reference level for exploratory extreme value analysis (del Castillo et al., 2019). An expanded list of continuous distributions for which expectiles can be written in closed form is found in Tables C5, C6, C7 and C8.

Example 2.12 (Student distribution with $\nu = 4$ degrees of freedom)**.** Consider the Student distribution with ν degrees of freedom, having probability density function

$$f(x) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\Gamma(\nu/2)}\left(1+\frac{x^2}{\nu}\right)^{-(\nu+1)/2}, \ x \in \mathbb{R}.$$

When $\nu = 4$, one finds (see Appendix B for further details)

$$\varphi(x) = \frac{1}{2}\left(\frac{x^2+2}{\sqrt{x^2+4}} - x\right).$$

Equation (3) then becomes

$$\xi_\tau^4 + 4\xi_\tau^2 - \frac{(2\tau-1)^2}{\tau(1-\tau)} = 0.$$

This is a biquadratic equation, leading to $\xi_\tau^2 = -2 + 1/\sqrt{\tau(1-\tau)}$ because $\xi_\tau^2 \geq 0$, and then

$$\xi_\tau = \mathrm{sign}(2\tau-1)\sqrt{\frac{1}{\sqrt{\tau(1-\tau)}} - 2}.$$

In general, the distribution function and mean residual life function of the Student distribution involve the hypergeometric function. It is not hard to see that, while the distribution function and mean residual life function can in fact be written in closed form when ν is an even integer, resulting in a polynomial equation characterizing ξ_τ, only the cases $\nu \in \{2, 4, 6\}$ result in an equation of degree 4 or lower.

The expectiles of the Student distribution with $\nu = 2$ degrees of freedom (Koenker, 1993), the uniform distribution (see Example 3.1 in Bellini and Di Bernardino, 2017), the Pareto distribution with $\alpha = 2$ (see Example 3.6 in Bellini and Di Bernardino, 2017) and the Dagum distribution (with survival function $\overline{F}(x) = 1 - (1 + x^{-\alpha})^{-\beta}$, $y > 0$) with $\alpha = 2$ and $\beta = 1/2$ can also be found by solving a quadratic equation. For the Student distribution with 4 degrees of freedom the equation is actually biquadratic.

Example 2.13 (Fisher distribution with $(4, 4)$ degrees of freedom)**.** The Fisher distribution with degrees of freedom $\nu_1 > 0$ and $\nu_2 > 0$ has density function

$$f(x) = \frac{(\nu_1/\nu_2)^{\nu_1/2}}{B(\nu_1/2, \nu_2/2)} x^{\nu_1/2-1}(1+\nu_1 x/\nu_2)^{-(\nu_1+\nu_2)/2},$$

for $x > 0$, where B is the Beta function. In the specific case $\nu_1 = \nu_2 = 4$, one finds $\varphi(x) = (3x + 2)/(x + 1)^2$ for $x > 0$, and $m = 2$. Equation (3) is thus equivalent to the cubic equation

$$\xi_\tau^3 - \frac{3\tau}{1-\tau}\xi_\tau - \frac{2\tau}{1-\tau} = 0.$$

Straightforward calculations involving Cardano's method (see Appendix B for details) then yield

$$\xi_\tau = \begin{cases} \sqrt[3]{\dfrac{\tau}{1-\tau}} \left(\sqrt[3]{1 + \sqrt{\dfrac{1-2\tau}{1-\tau}}} + \sqrt[3]{1 - \sqrt{\dfrac{1-2\tau}{1-\tau}}} \right) & \text{if } \tau \le 1/2, \\ 2\sqrt{\dfrac{\tau}{1-\tau}} \cos\left(\dfrac{1}{3}\arccos\left(\sqrt{\dfrac{1-\tau}{\tau}}\right)\right) & \text{if } \tau > 1/2. \end{cases}$$

Other examples obtained through solving cubic equations include the Beta distribution with $(\alpha, \beta) = (1, 2)$ or $(2, 1)$, the triangular distribution (obtained as the convolution of two standard uniform distributions), the Hall-Weiss distribution with $\alpha = 2$ and $\beta = 1$, and the Pareto distribution with extreme value index $\gamma = 1/3$ and $\gamma = 2/3$ (the latter using the change of variables $z_\tau = \sqrt{\xi_\tau}$).

Example 2.14 (Pareto distribution with extreme value index $1/4$)**.** The Pareto distribution with extreme value index $\gamma > 0$ has survival function $\overline{F}(x) = x^{-1/\gamma}$ for $x > 1$. This distribution has a finite first moment when $\gamma < 1$, and since $\varphi(x) = \gamma x^{1-1/\gamma}/(1-\gamma)$ for $x > 1$ and $m = 1/(1-\gamma)$, Equation (3) leads to

$$(1-\gamma)(1-\tau)\xi_\tau^{1/\gamma} - (1-\tau)\xi_\tau^{1/\gamma-1} - \gamma(2\tau - 1) = 0.$$

When $\gamma = 1/4$, this is the quartic equation $\xi_\tau^4 + b\xi_\tau^3 + c\xi_\tau^2 + d\xi_\tau + e = 0$, where $b = -4/3$, $c = 0$, $d = 0$ and $e = (1-2\tau)/(3(1-\tau))$. Ferrari's method (see Appendix B) leads to

$$\xi_\tau = \frac{1}{2}\sqrt{-2\lambda_\tau - \frac{8}{3}\frac{\lambda_\tau}{\sqrt{2\lambda_\tau + \frac{4}{9}}} + \frac{4}{3}\sqrt{2\lambda_\tau + \frac{4}{9}} + \frac{8}{9}} + \frac{1}{2}\sqrt{2\lambda_\tau + \frac{4}{9}} + \frac{1}{3},$$

where

$$\lambda_\tau = \sqrt[3]{\frac{1 - 2\tau + |1-2\tau|\sqrt{\frac{\tau}{1-\tau}}}{27(1-\tau)}} + \sqrt[3]{\frac{1 - 2\tau - |1-2\tau|\sqrt{\frac{\tau}{1-\tau}}}{27(1-\tau)}}.$$

Other distributions leading to closed-form expectiles through solving a quartic equation are the Beta distribution with $(\alpha, \beta) = (2, 2)$, $(3, 1)$ or $(1, 3)$, the Fisher distribution with $(\nu_1, \nu_2) = (4, 6)$ or $(6, 4)$, and the Hall-Weiss distribution with $(\alpha, \beta) = (3/2, 1/2)$, $(3, 1)$ or $(2, 2)$. A suitable change of variables ($z_\tau = \sqrt[3]{\xi_\tau}$) also leads to a quartic equation for the expectile of the Pareto distribution with extreme value index $\gamma = 3/4$. For the Student distribution with $\nu = 6$ degrees of freedom, the equation is actually biquartic.

2.3.2 Analytic expressions through transcendental equations

The above distributions, whose expectiles are obtained by solving a polynomial equation, are closely related to the Pareto distribution, either because they are heavy-tailed, or, in the case of the Beta distribution, because it can be transformed into a heavy-tailed distribution in a simple manner: if X has a Beta distribution, then $1/(1-X)$ has a heavy-tailed distribution. The purpose of this section is to highlight a couple of distributions closely related to the exponential distribution whose expectiles are in analytic form. This builds upon earlier work of Bellini and Di Bernardino (2017), who showed in their Example 3.2 that expectiles of the exponential distribution may be expressed using the Lambert W function. It is then not hard to show that this is also the case for the Laplace distribution (also called "double exponential distribution"). We discuss below the cases of the Inverse Gamma distribution and the chi-squared distribution in detail.

Example 2.15 (Inverse-Gamma distribution)**.** The Inverse-Gamma distribution with scale parameter $\lambda > 0$ and shape parameter $\alpha > 0$ has density function

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp(-\lambda/x), \ x > 0.$$

It has a finite first moment if and only if $\alpha > 1$, in which case $m = \lambda/(\alpha-1)$. In general, the function φ depends on the incomplete Gamma function, but it has a remarkably simple form when $\alpha = 2$:

$$\begin{aligned}\varphi(x) &= \int_x^\infty (y-x) f(y)\, \mathrm{d}y = \int_0^{1/x} \lambda^2(1-xv)e^{-\lambda v}\, \mathrm{d}v \\ &= \lambda - x(1-\exp(-\lambda/x)), \ x > 0.\end{aligned}$$

Equation (3) is thus exactly the transcendental equation

$$\begin{aligned}&(2\tau-1)\exp\left(-\frac{\lambda}{\xi_\tau}\right) + \tau\left(\frac{\lambda}{\xi_\tau} - 1\right) = 0 \\ \Leftrightarrow &\left(\frac{\lambda}{\xi_\tau} - 1\right)\exp\left(\frac{\lambda}{\xi_\tau} - 1\right) = -\frac{2\tau-1}{\tau}e^{-1}.\end{aligned}$$

Since $\lambda/\xi_\tau - 1$ is by construction greater than -1, and since the right-hand side is, for any $\tau \in (0,1)$, greater than $-e^{-1}$, one obtains the solution in terms of the principal branch of the Lambert W function:

$$\xi_\tau = \frac{\lambda}{1 + W(-\frac{2\tau-1}{\tau}e^{-1})}.$$

Example 2.16 (Chi-squared distribution with $\nu = 2$ or 4 degrees of freedom)**.** The chi-squared distribution with $\nu > 0$ degrees of freedom has density function

$$f(x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} x^{\nu/2-1} \exp(-x/2), \ x > 0.$$

When $\nu = 2$, this is nothing but an exponential distribution with $\lambda = 1/2$, whose expectiles involve the Lambert W function:

$$\xi_\tau = \frac{1}{2}\left(1 + W\left(\frac{2\tau-1}{1-\tau}e^{-1}\right)\right).$$

When $\nu = 4$, straightforward calculations yield $\varphi(x) = (x+4)\exp(-x/2)$ for $x > 0$ and $m = 4$, leading to the equation

$$(2\tau-1)(\xi_\tau+4)\exp(-\xi_\tau/2) - (1-\tau)(\xi_\tau-4) = 0.$$

This is equivalently rewritten as

$$\exp\left(\frac{\xi_\tau}{2}\right)\frac{\frac{\xi_\tau}{2} - t}{\frac{\xi_\tau}{2} - s} = a,$$

where $t = 2$, $s = -2$ and $a = (2\tau-1)/(1-\tau)$. According to Mező and Baricz (2017), the solution of this equation is $\xi_\tau = 2W\left(\begin{smallmatrix}2\\-2\end{smallmatrix}; \frac{2\tau-1}{1-\tau}\right)$, where $W\left(\begin{smallmatrix}\cdot\\\cdot\end{smallmatrix}; \cdot\right)$ is the generalized Lambert W function. Theorem 1 in Mező and Baricz (2017) provides a power series expansion for this special function which may be used for a numerical implementation.

3 Monte-Carlo computation

When the distribution function related to μ is continuous but the functions $\overline{F}$ and φ are not explicit and their numerical calculation is difficult or unstable, a calculation of expectiles in analytic form is not possible and writing a Newton-Raphson algorithm that performs well in practice is much harder. In this setting, if one can at least efficiently simulate realizations from the distribution μ, one may instead use Monte-Carlo computation. Suppose that $X_1, \ldots, X_n$ are independent random draws from the distribution μ having continuous distribution function F, and consider the empirical probability measure $\widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$. The results of Holzmann and Klar (2016) entail that the τth expectile $\widehat{\xi}_{\tau,n}$ of the distribution $\widehat{\mu}_n$, *i.e.* the LAWS estimator of ξ_τ, is $\sqrt{n}-$consistent when μ has a finite second moment, and can be exactly computed (see Example 2.6). This provides a first reasonable approximation for the target expectile ξ_τ when n is sufficiently large.

This Monte-Carlo computational approach can be further improved using very simple devices such as control variates. Since expectiles extend the mean of a distribution, and the latter is very often known exactly, it makes sense to set the mean m as a control variate and seek the joint asymptotic behavior of the LAWS estimator $\widehat{\xi}_{\tau,n}$ and the sample mean $\overline{X}_n$ in order to find the linear combination $\widehat{\xi}_{\tau,n}+c(\overline{X}_n-m)$ having the lowest possible (asymptotic) variance. Define

$$\varphi^{(k)}(x) = \int_{\mathbb{R}} (t-x)^k \mathbb{1}_{\{t>x\}}\,\mu(\mathrm{d}t),$$
$$\text{with then } \varphi(x) = \varphi^{(1)}(x).$$

It turns out that (see Corollary 4 in Holzmann and Klar, 2016)

$$\sqrt{n}(\widehat{\xi}_{\tau,n}-\xi_\tau,\, \overline{X}_n-m) \xrightarrow{d} \mathcal{N}((0,0),\mathbf{\Sigma}) \text{ as } n\to\infty,$$

as soon as μ has a finite variance σ^2 and puts no mass at ξ_τ, where the 2×2 symmetric matrix $\mathbf{\Sigma}$ is defined as

$$\Sigma_{11} = \frac{(1-\tau)^2\mathbb{E}((X-\xi_\tau)^2)+(2\tau-1)\varphi^{(2)}(\xi_\tau)}{(1-\tau+(2\tau-1)\overline{F}(\xi_\tau))^2},$$
$$\Sigma_{12} = \frac{(1-\tau)\mathbb{E}((X-\xi_\tau)^2)+(2\tau-1)\varphi^{(2)}(\xi_\tau)}{1-\tau+(2\tau-1)\overline{F}(\xi_\tau)}$$
$$\text{and } \Sigma_{22} = \sigma^2.$$

A straightforward calculation shows that

$$\widehat{\xi}_{\tau,n} - \frac{\Sigma_{12}}{\Sigma_{22}}(\overline{X}_n-m) = \widehat{\xi}_{\tau,n} - \frac{\Sigma_{12}}{\sigma^2}(\overline{X}_n-m)$$

is the asymptotically unbiased linear combination of $\widehat{\xi}_{\tau,n}$ and $\overline{X}_n-m$ with lowest asymptotic variance, equal to $\Sigma_{11}(1-\Sigma_{12}^2/(\Sigma_{11}\Sigma_{22}))$. Of course, Σ_{12} and Σ_{22} are not known, but are readily estimated by

$$\widehat{\Sigma}_{12,n} = \frac{\frac{1-\tau}{n}\sum_{i=1}^n (X_i-\widehat{\xi}_{\tau,n})^2+(2\tau-1)\widehat{\varphi}_n^{(2)}(\widehat{\xi}_{\tau,n})}{1-\tau+(2\tau-1)\widehat{\overline{F}}_n(\widehat{\xi}_{\tau,n})}$$
and $\widehat{\Sigma}_{22,n}=\widehat{\sigma}_n^2$, where

$$\widehat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i-\overline{X}_n)^2, \widehat{\overline{F}}_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{X_i>x\}},$$
$$\widehat{\varphi}_n^{(k)}(x) = \frac{1}{n}\sum_{i=1}^n (X_i-x)^k \mathbb{1}_{\{X_i>x\}}.$$

This leads to a further, practically feasible Monte-Carlo approximation of ξ_τ given the control variate m as

$$\widetilde{\xi}_{\tau,n} = \widehat{\xi}_{\tau,n} - \frac{\widehat{\Sigma}_{12,n}}{\widehat{\sigma}_n^2}(\overline{X}_n-m),$$

whose asymptotic behavior is established in the following result.

Theorem 3.1. *Assume that μ has a finite and positive variance σ^2 and let the X_i be independent random copies of a random variable X having distribution μ. If μ puts no mass at ξ_τ then $\sqrt{n}(\widetilde{\xi}_{\tau,n}-\xi_\tau)$ is asymptotically a centered Gaussian random variable with variance*

$$\frac{(1-\tau)^2\mathbb{E}((X-\xi_\tau)^2)+(2\tau-1)\varphi^{(2)}(\xi_\tau)}{(1-\tau+(2\tau-1)\overline{F}(\xi_\tau))^2}(1-R(\tau,\mu))$$

where $R(\tau,\mu)=$

$$\frac{1}{\sigma^2}\frac{((1-\tau)\mathbb{E}((X-\xi_\tau)^2)+(2\tau-1)\varphi^{(2)}(\xi_\tau))^2}{(1-\tau)^2\mathbb{E}((X-\xi_\tau)^2)+(2\tau-1)\varphi^{(2)}(\xi_\tau)}.$$

The quantity $\widetilde{\xi}_{\tau,n}$ has the lowest asymptotic variance among all asymptotically unbiased linear combinations of $\widehat{\xi}_{\tau,n}$ and $\overline{X}_n-m$. In addition, one has $R(1/2,\mu)=1$, and if the distribution function pertaining to μ is continuous on $[\xi_{\tau_1},\xi_{\tau_2}]$, for $0\le\tau_1<\tau_2\le 1$, then the function $\tau\mapsto 1-R(\tau,\mu)$ is continuously differentiable on $I=(\tau_1,\tau_2)$, decreasing on $(0,1/2]\cap I$ and increasing on $[1/2,1)\cap I$.

For continuous distributions, the variance reduction factor $1-R(\tau,\mu)$ is therefore monotonic and smooth as the target expectile deviates from the mean. It is not, however, possible to give a simple expression of this variance reduction factor, unlike for Monte-Carlo simulation of a quantile q_τ with the median as control variate. Indeed, as regards the latter, when μ has a continuous and positive density function f with respect to the Lebesgue measure, a straightforward application of the joint asymptotic normality result for several sample quantiles (see p.72 in Koenker,

2005) results in the asymptotic variance-optimal and unbiased linear combination $\widetilde{q}_{\tau,n}$ of $\widehat{q}_{\tau,n}$ and $\widehat{q}_{1/2,n} - q_{1/2}$ satisfying

$$\sqrt{n}(\widetilde{q}_{\tau,n} - q_\tau) \xrightarrow{d} \mathcal{N}\left(0, \frac{\tau(1-\tau)}{(f(q_\tau))^2}\left(1 - \frac{\min(\tau, 1-\tau)}{\max(\tau, 1-\tau)}\right)\right).$$

For quantiles, then, the variance reduction factor is the explicit and universal number $1 - \min(\tau, 1-\tau)/\max(\tau, 1-\tau)$. For expectile computation, even though no simple expression of the variance reduction factor $1 - R(\tau, \mu)$ is feasible, we shall illustrate that in typical interesting examples, it is very small for $\tau \in [1/4, 3/4]$. This can already be visualized in the toy example of the Fréchet distribution (where expectiles are known to a high degree of accuracy using the Newton-Raphson algorithm), see Figure 1. As a result, Monte Carlo computations with the mean as control variate will generally drastically improve upon vanilla Monte Carlo when the target expectile lies within what could be called the "interexpectile range", by analogy to the interquartile range.

Example 3.1 (Finite sums and products of random variables)**.** Conceptually simple yet interesting examples where simulation is straightforward but analytical computations may be impossible are sums and products of random variables. As a toy example, let us first consider the Irwin-Hall model with parameter $d \geq 2$, whereby μ is the probability distribution of the sum of d independent standard uniform random variables. The case $d = 2$ is the triangular distribution, whose expectiles are known in closed form (see Table C5), and for large d, the central limit theorem entails that the Irwin-Hall distribution is essentially the Gaussian distribution with mean $d/2$ and variance $d/12$. However, no simple analytic expression is available for the function φ when $d > 2$, and as such, the calculation of expectiles using the Newton-Raphson algorithm is well-nigh impossible. We display in Figure 2 a comparison between estimated expectiles $\widehat{\xi}^{(d)}_{\tau,n}$ using vanilla Monte-Carlo computation and the estimates $\widetilde{\xi}^{(d)}_{\tau,n}$ produced using the Monte-Carlo algorithm with the mean $m = d/2$ of the Irwin-Hall distribution as a control variate, both on $N = 1{,}000$ replicated independent samples of size $n = 10{,}000$. The Monte-Carlo method using the mean as a control variate appears to be much more accurate than vanilla Monte-Carlo, with variance reduced by a factor of about 10 at level $\tau = 0.75$. For $d = 2$, the Monte-Carlo estimates are consistent with the closed-form expression in Table C5. For large d, the Monte-Carlo approach reassuringly recovers the expectiles of the Gaussian distribution with $d/2$ and variance $d/12$.

A more complex example that is relevant in insurance and finance is the computation of expectiles for the stationary distribution of an ARMA-GARCH model. For the sake of simplicity, we focus here on the ARMA(1,1)-GARCH(1,1) model: recall that the time series (X_t) follows an ARMA(1,1)-GARCH(1,1) model with mean 0 if

$$X_t = \phi X_{t-1} + \theta\varepsilon_{t-1} + \varepsilon_t, \text{ with } \varepsilon_t = \sigma_t\eta_t \\ \text{and } \sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2,\ t \in \mathbb{Z}.$$

Here $\phi, \theta \in \mathbb{R}$, $\omega > 0$, $\alpha, \beta \geq 0$ and the η_t are independent, identically distributed, nondegenerate, centered random variables with variance 1. It is well-known that a unique nonanticipative strictly stationary solution to these ARMA-GARCH equations exists if, for instance, $\phi, \theta \in (-1, 1)$ and $\alpha + \beta < 1$ (see, for example, Francq and Zakoïan, 2004, and particularly Equation (2.5), Assumptions (A2) and (A8) and p.612 therein). In this model, one may ask how expectiles of the stationary distribution of (X_t) can be computed; this is relevant for applications to long-term risk management, for instance when trying to evaluate measures of extreme risks over dozens or hundreds of years. In doing so we do not, of course, tackle the different problem of dynamic expectile computation, *i.e.* the calculation of the expectile of X_t given its past: this is essentially a trivial problem, since it only requires the calculation of expectiles of the innovation distribution. Figure 3 compares the performance of the vanilla Monte-Carlo algorithm for the computation of expectiles of the stationary distribution of the (X_t) with that of the Monte-Carlo algorithm using the mean (which is zero) as a control variate, again on $N = 1{,}000$ replicated independent samples of size $n = 10{,}000$, generated via the `ugarchsim` routine from the `R` package `rugarch` (Galanos and Kley, 2022), from two ARMA(1,1)-GARCH(1,1) models having standard Gaussian innovations. The improvement in terms of variability brought

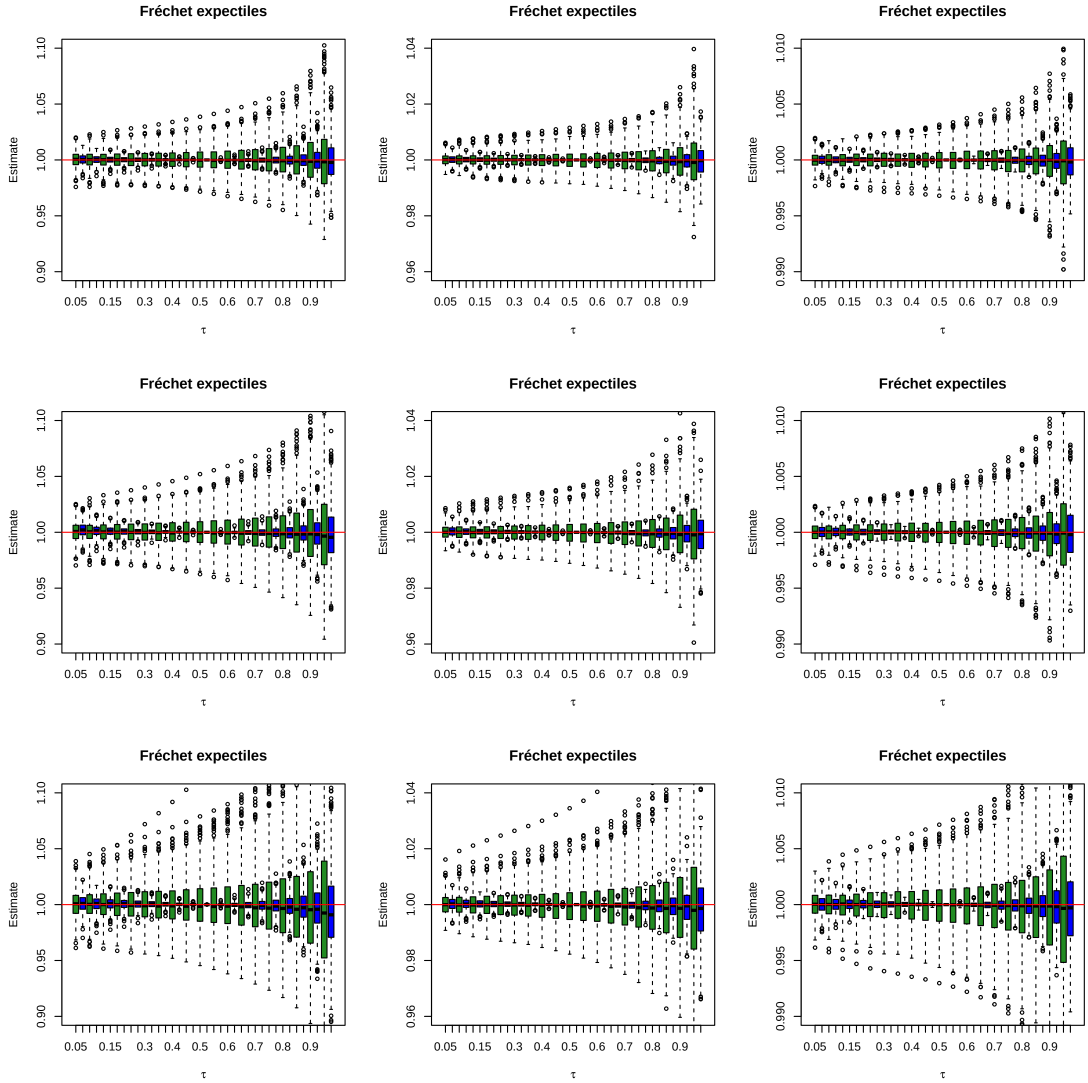

Fig. 1: Example of the Fréchet distribution. Boxplots of $\widehat{\xi}_{\tau,n}$ (green) and $\widetilde{\xi}_{\tau,n}$ (blue), normalized by the true value ξ_τ calculated through the Newton-Raphson algorithm, for $\tau \in \{0.05, 0.10, \ldots, 0.95\}$. We take 1,000 Monte-Carlo replications of an independent sample of size $n =$ 1,000 (left), 10,000 (middle) and 100,000 (right), where $\gamma = 1/5$ (top panels), 1/4 (middle panels) and 1/3 (bottom panels).

by the control variate device still appears to be very substantial in this difficult setting where the stationary distribution is heavy-tailed.

Example 3.2 (Insurance premium calculation)**.** For insurance companies, the calculation of the premium paid by a policyholder is an important task: a premium that is too low will threaten the company's survival in the long run, while a premium that is too high will be detrimental to the company's competitivity on the open market. Adopting a model for the sum of claim amounts of a policyholder at time t is a necessary step before premium calculations. A standard such model is the compound Poisson process

$$C_t = \sum_{i=1}^{N_t} X_i,$$

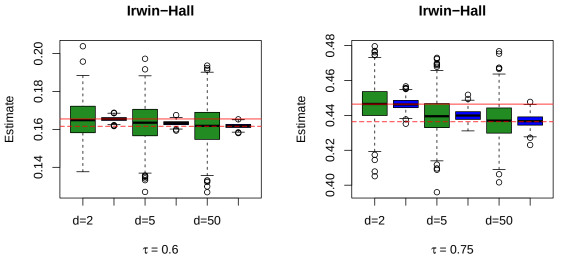

Fig. 2: Example of the Irwin-Hall distribution. Boxplots of $(\widehat{\xi}_{\tau,n}^{(d)}-d/2)/\sqrt{d/12}$ (green) and $(\widetilde{\xi}_{\tau,n}^{(d)}-d/2)/\sqrt{d/12}$ (blue) for $\tau=0.6$ (left) and $\tau=0.75$ (right). In both panels, we take, from left to right, $d\in\{2,5,50\}$, and use 1,000 Monte-Carlo replications of an independent sample of size $n=10{,}000$. The full red line is $(\xi_\tau(\mathrm{T})-1)/\sqrt{1/6}$, where $\xi_\tau(\mathrm{T})$ is the expectile of the triangular distribution (known in closed form), and the dashed red line is the expectile $\xi_\tau(\mathrm{G})$ of the standard Gaussian distribution (approximated via the Newton-Raphson algorithm).

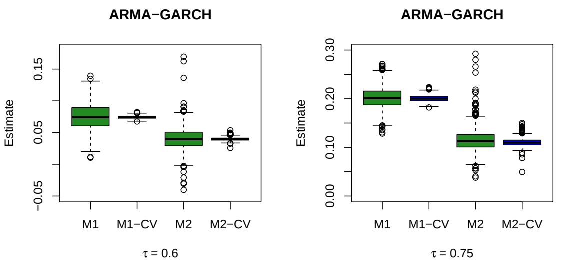

Fig. 3: Example of the ARMA(1,1)-GARCH(1,1) process. Boxplots of $\widehat{\xi}_{\tau,n}$ (green) and $\widetilde{\xi}_{\tau,n}$ (blue) for $\tau=0.6$ (left) and $\tau=0.75$ (right). In model 1 we let $(\alpha,\beta)=(0.1,0.85)$ and in model 2 we let $(\alpha,\beta)=(0.85,0.1)$. Both models have $\phi=0.9$, $\theta=0.5$, $\omega=0.001$, and the ε_t are standard Gaussian innovations. In each setting, 1,000 Monte-Carlo replications of an independent sample of size $n=10{,}000$ were used.

where the X_i are independent copies of a positive random variable X having a finite first moment and (N_t) is a homogeneous Poisson process with intensity $\lambda>0$ independent of the X_i (see for instance Asmussen and Albrecher, 2010). Given this model for the sum of a policyholder's claim amounts, representing its individual cost to the insurer, there are several possibilities in order to compute a fair premium relative to the contract up to time $t=T$. A reasonable and widely used approach is the "principle of zero utility" (see Chapter 3 in Dickson, 2016), defined as the solution π to the equation

$$\mathbb{E}[u(\pi-C_T)]=u(0)$$

where u is a suitably chosen utility function. As pointed out in Bellini et al. (2014), the choice of $u(x)=(2\tau-1)x\mathbb{1}_{\{x<0\}}+(1-\tau)x$ leads to $\pi=\xi_\tau=\xi_\tau(C_T)$, the τth expectile of C_T. For $\tau>1/2$, the choice of premium $\pi=\xi_\tau=\xi_{0.5(1+\{2\tau-1\})}$ is thus a particular case of the principle of zero utility, with $2\tau-1>0$ being seen as the analog of the loading factor appearing in the simple "expected value principle". It is, moreover, immediate that the mean of C_T is $m=m(C_T)=\lambda T\mathbb{E}(X)$, which is known as soon as the expectation of X is known. However, expressing the distribution of C_T in a tractable form is unfeasible in general, even in very simple settings such as when the X_i have a common exponential distribution: in this case, the probability density function of C_T given that $N_T>0$ involves modified Bessel functions.

As an illustration, we compute an approximation $\widetilde{\xi}_{\tau,n}$ of ξ_τ using our Monte-Carlo approach with m as control variate, and we compare it with the vanilla Monte-Carlo estimator $\widehat{\xi}_{\tau,n}$. Results, for $N=1{,}000$ replicated independent samples of size $n=10{,}000$ from C_T, are reported in Figure 4 on two models involving exponential and Pareto random variables X_i. The decrease in variability of the Monte-Carlo method when incorporating a control variate is obvious: in particular, our calculations indicate, on the two proposed examples, that the variance of $\widetilde{\xi}_{\tau,n}$ is reduced by a factor of more than 10 compared to the variance of $\widehat{\xi}_{\tau,n}$ at $\tau=0.75$.

Example 3.3 (SDEs and option pricing)**.** Let S_t, $0\le t\le T$ be the price of a financial asset. Financial asset prices are often modeled by continuous-time stochastic processes, of which a popular subclass are the solutions of stochastic differential equations (SDEs) of the form

$$\mathrm{d}S_t=rS_t\,\mathrm{d}t+\sqrt{\nu_t}S_t\,\mathrm{d}B_t,\ S_0\in\mathbb{R}$$

where (B_t) is a standard Brownian motion, r is the risk-free interest rate, and ν_t is a stochastic

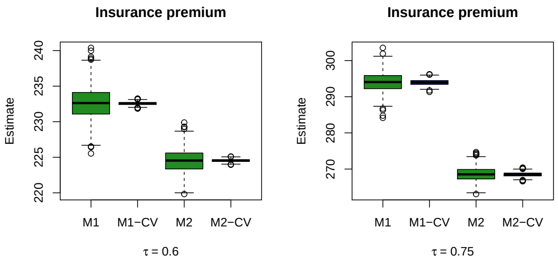

Fig. 4: Insurance premium example. Boxplots of $\widehat{\xi}_{\tau,n}$ (green) and $\widetilde{\xi}_{\tau,n}$ (blue) for $\tau = 0.6$ (left) and $\tau = 0.75$ (right). In model 1 we let the X_i have an exponential distribution with mean 100, and in model 2 we let the X_i have a Pareto distribution with extreme value index $\gamma = 1/4$, rescaled to have mean 100. In both models, $T = 20$ and $\lambda = 0.1$, and 1,000 Monte-Carlo replications of an independent sample of size $n = 10{,}000$ were used.

process representing volatility. These models are in turn used to determine the price of financial derivatives, such as call options, that allow to buy the asset at time T for a predefined strike price K. The payoff at time T of such a call option is given by the random variable

$$C_T(K) = (S_T - K)\mathbb{1}_{\{S_T > K\}}.$$

Under the risk-neutral assumption, using the equality $\mathbb{E}(S_T) = S_0 \exp(rT)$, the price at time $t = 0$ of such an option is $C_0(K) = \mathbb{E}(C_T(K))\exp(-rT)$. This means that determining the price to be paid for a call option requires computing the expected payoff $\mathbb{E}(C_T(K))$. This is in general impossible, unless one assumes, for instance, the unrealistic Black-Scholes model (Black and Scholes, 1973) where ν_t is constant, that is,

$$\mathrm{d}S_t = rS_t \,\mathrm{d}t + \sigma S_t \,\mathrm{d}B_t$$

where $\sigma > 0$. In this model (S_t) is a geometric Brownian motion and the price $C_0(K) = C_0^{(BS)}(K)$ satisfies

$$C_0^{(BS)}(K) = S_0 \Phi\left(\frac{\log\left(\frac{S_0}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}\right)$$

$$- K\exp(-rT)\Phi\left(\frac{\log\left(\frac{S_0}{K}\right) + \left(r - \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}\right)$$

where Φ is the cumulative distribution function of the standard Gaussian distribution. Outside of this simple model, the Monte-Carlo approach for the computation of $C_0(K)$ simulates N paths $(S_t^{(i)})$, for $1 \le i \le n$, of (S_t) over the period $[0, T]$ and computes

$$\widehat{C}_0(K) = \left(\frac{1}{n}\sum_{i=1}^{n}(S_T^{(i)} - K)\mathbb{1}_{\{S_T^{(i)} > K\}}\right)\exp(-rT).$$

The solution S_t can be simulated using an Euler-Maruyama-type scheme. By viewing the strike price K as an expectile ξ_τ of S_T to be estimated (with τ around $1/2$), a vanilla Monte-Carlo approach to the approximation of $C_0(K) = C_0(\xi_\tau)$ is

$$\widehat{C}_0(\xi_\tau) = \left(\frac{1}{n}\sum_{i=1}^{n}(S_T^{(i)} - \widehat{\xi}_{\tau,n})\mathbb{1}_{\{S_T^{(i)} > \widehat{\xi}_{\tau,n}\}}\right)\exp(-rT),$$

where $\widehat{\xi}_{\tau,n}$ is the Monte-Carlo estimate of ξ_τ based on the $S_T^{(i)}$. Differently from that technique, and since $\mathbb{E}(C_T(K)) = \varphi(K)$ for μ being the probability distribution of S_T, Equation (3) yields the remarkable identity

$$\mathbb{E}(C_T(\xi_\tau)) = \frac{1-\tau}{2\tau-1}(\xi_\tau - \mathbb{E}(S_T))$$

$$\text{and then } C_0(\xi_\tau) = \frac{1-\tau}{2\tau-1}\left(\xi_\tau \exp(-rT) - S_0\right).$$

An alternative option to the use of $\widehat{C}_0(\xi_\tau)$ is therefore

$$\widetilde{C}_0(\xi_\tau) = \frac{1-\tau}{2\tau-1}\left(\widetilde{\xi}_{\tau,n}\exp(-rT) - S_0\right)$$

where $\widetilde{\xi}_{\tau,n}$ is the Monte-Carlo estimate of ξ_τ using the known mean $m = \mathbb{E}(S_T) = S_0\exp(rT)$ as control variate.

We compare the computation methods $\widehat{C}_0(\xi_\tau)$ and $\widetilde{C}_0(\xi_\tau)$ on the Black-Scholes model, where the true value $C_0^{(BS)}(\xi_\tau)$ can be approximated to a high degree of accuracy using the Newton-Raphson algorithm because S_T has a log-normal distribution, and on the Heston model (Heston, 1993)

$$\mathrm{d}S_t = rS_t\,\mathrm{d}t + \sqrt{\nu_t}S_t\,\mathrm{d}B_t,$$
$$\mathrm{d}\nu_t = \kappa(\theta - \nu_t)\,\mathrm{d}t + \sigma\sqrt{\nu_t}\,\mathrm{d}W_t, \quad S_0 \in \mathbb{R},\ \nu_0 \in \mathbb{R}$$

where $\kappa, \theta, \sigma > 0$ and (B_t), (W_t) are two standard Brownian motions with correlation $\rho \in (-1, 1)$. Results are reported in Figure 5 for a number $n =$ 10,000 of independent simulations of (S_T). As in Example 3.2, the variability of the Monte-Carlo method with control variate is much lower than that of vanilla Monte-Carlo.

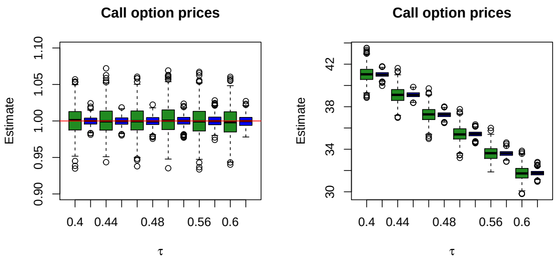

Fig. 5: Option pricing example. Left panel: Boxplots of $\widehat{C}_0^{(BS)}(\xi_\tau)/C_0^{(BS)}(\xi_\tau)$ (green) and $\widetilde{C}_0^{(BS)}(\xi_\tau)/C_0^{(BS)}(\xi_\tau)$ (blue), in the Black-Scholes model with volatility $\sigma^2 = 1$. Right panel: Boxplots of $\widehat{C}_0(\xi_\tau)$ (green) and $\widetilde{C}_0(\xi_\tau)$ (blue), in the Heston model with $\kappa = 2$, $\theta = 0.8$, $\nu_0 = 1$, $\sigma^2 =$ 0.01 and $\rho = 0.5$. In both examples, the time horizon is $T = 1$, the risk-free interest rate is $r = 5\% =$ 0.05 and the initial condition is $S_0 = 100$. We consider the values $\tau = 0.4, 0.44, 0.48, 0.52, 0.56, 0.6$, and use 1,000 Monte-Carlo replications of an independent sample of size $n =$ 10,000 from S_T.

In extremal cases, *i.e.* when τ is close to 1, the asymptotic variance of the Monte-Carlo estimator diverges to infinity. It is, however, well-known (see for example Theorem 2 in Daouia et al., 2018) that when μ is a (right) heavy-tailed distribution with finite second moment and extreme value index $\gamma \in (0, 1/2)$, and if $\tau_n \uparrow 1$ with $n(1-\tau_n) \to \infty$, then

$$\sqrt{n(1-\tau_n)}\left(\frac{\widehat{\xi}_{\tau_n,n}}{\xi_{\tau_n}} - 1\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{2\gamma^3}{1-2\gamma}\right)$$

as $n \to \infty$. For such expectiles, taking the mean as control variate makes little sense: being a parameter relevant to the bulk of the distribution, rather than to its extremes, the sample mean will have little correlation with extreme sample expectiles. However, if the inverse distribution function $q = F^{\leftarrow}$ is easy to compute, then a reasonable control variate is an extreme quantile q_{α_n} whose empirical estimator (the order statistic $\widehat{q}_{\alpha_n,n} = X_{\lceil n\alpha_n \rceil, n}$) is highly correlated with $\widehat{\xi}_{\tau_n,n}$. This leads us to consider Monte-Carlo estimates of the form $\check{\xi}_{\tau_n,n} = \widehat{\xi}_{\tau_n,n} + c(\widehat{q}_{\alpha_n,n} - q_{\alpha_n})$ for a certain constant c and a level $\alpha_n \uparrow 1$ to be determined. To this end, we first state a joint asymptotic normality result between $\widehat{\xi}_{\tau_n,n}$ and $\widehat{q}_{\alpha_n,n}$ under the following classical second-order condition on the heavy right tail behavior of μ.

$\mathcal{C}_2(\gamma, \rho, A)$ There exist $\gamma > 0$, $\rho \le 0$ and a measurable auxiliary function A having constant sign and converging to 0 at infinity such that the function $\overline{F} : x \mapsto \mu((x, +\infty))$ satisfies

$$\forall x > 0,\ \lim_{t\to\infty} \frac{1}{A(1/\overline{F}(x))}\left(\frac{\overline{F}(tx)}{\overline{F}(t)} - x^{-1/\gamma}\right) = \frac{x^{-1/\gamma}}{\gamma^2}\int_1^x s^{\rho/\gamma - 1}\,\mathrm{d}s.$$

Proposition 3.1. *Assume that μ has a finite variance and satisfies condition $\mathcal{C}_2(\gamma, \rho, A)$ with $\gamma \in (0, 1/2)$. Suppose that $\tau_n, \alpha_n \uparrow 1$, with $n(1-\tau_n) \to \infty$, $(1-\alpha_n)/(1-\tau_n) \to \lambda \in (0, \infty)$ and $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) = \mathrm{O}(1)$. Let the X_i be independent random variables with distribution μ. Then*

$$\sqrt{n(1-\tau_n)}\left(\frac{\widehat{\xi}_{\tau_n,n}}{\xi_{\tau_n}} - 1, \frac{\widehat{q}_{\alpha_n,n}}{q_{\alpha_n}} - 1\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}_2, \gamma^2\boldsymbol{V})$$

where $\boldsymbol{V}$ is the 2×2 symmetric matrix having elements $V_{11} = 2\gamma/(1-2\gamma)$, $V_{22} = 1/\lambda$ and

$$V_{12} = \frac{1}{\lambda}\left\{\min\left(\frac{\lambda}{1/\gamma - 1}, 1\right)\right\}^{1-\gamma} + \left\{\min\left(\frac{\lambda}{1/\gamma - 1}, 1\right)\right\}^{-\gamma} - 1.$$

Proposition 3.1 is Proposition 1 in Stupfler and Usseglio-Carleve (2023) with $p = 2$, except that, with the notation therein, the latter contained a typo in the off-diagonal covariance term Λ_{12}: in the first term of this covariance Λ_{12}, the quantity γ between $(p-1)$ and $(g_p(\gamma)/\theta)$ should not appear.

We may now prove the following result on the optimal choice of linear combination of $\widehat{\xi}_{\tau_n,n}$ and $\widehat{q}_{\alpha_n,n} - q_{\alpha_n}$ in terms of relative asymptotic variance. We note that there are two degrees of freedom for the optimization of the control variate algorithm: the weight of $\widehat{q}_{\alpha_n,n} - q_{\alpha_n}$ as well as the value of the quantile level α_n itself.

Theorem 3.2. *Work under the conditions of Proposition 3.1. Then*

$$\begin{aligned}\breve{\xi}_{\tau_n,n} &= \widehat{\xi}_{\tau_n,n} \\ &- \left(\frac{\lambda}{1/\gamma-1}\right)^{\gamma}\left(\left\{\min\left(\frac{\lambda}{1/\gamma-1},1\right)\right\}^{1-\gamma}\right. \\ &\left.+\lambda\left\{\min\left(\frac{\lambda}{1/\gamma-1},1\right)\right\}^{-\gamma}-\lambda\right)\times(\widehat{q}_{\alpha_n,n}-q_{\alpha_n})\end{aligned}$$

has lowest relative asymptotic variance among all asymptotically unbiased linear combinations of $\widehat{\xi}_{\tau_n,n}$ and $\widehat{q}_{\alpha_n,n} - q_{\alpha_n}$, and

$$\sqrt{n(1-\tau_n)}\left(\frac{\breve{\xi}_{\tau_n,n}}{\xi_{\tau_n}}-1\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{2\gamma^3}{1-2\gamma}(1-C(\gamma,\lambda))\right)$$

as $n \to \infty$, where

$$\begin{aligned}C(\gamma,\lambda) = \Bigg(&\left\{\min\left(\frac{\lambda}{1/\gamma-1},1\right)\right\}^{1-\gamma} \\ &+\lambda\left\{\min\left(\frac{\lambda}{1/\gamma-1},1\right)\right\}^{-\gamma}-\lambda\Bigg)^2\frac{1-2\gamma}{2\lambda\gamma}.\end{aligned}$$

The function $\lambda \mapsto C(\gamma,\lambda)$ is maximal at

$$\lambda^\star = \lambda^\star(\gamma) = \left(\frac{1}{\gamma}-1\right)\left(\frac{1-\gamma}{1-2\gamma}\right)^{-1/\gamma} < \frac{1}{\gamma}-1,$$

yielding an optimal level $\alpha_n^\star = 1-\lambda^\star(1-\tau_n)$ of the quantile control variate for which

$$\breve{\xi}_{\tau_n,n} = \widehat{\xi}_{\tau_n,n} - 2\left(\frac{1-\gamma}{1-2\gamma}\right)^{-1/\gamma}(\widehat{q}_{\alpha_n^\star,n} - q_{\alpha_n^\star}).$$

As $n \to \infty$, this random quantity satisfies

$$\sqrt{n(1-\tau_n)}\left(\frac{\breve{\xi}_{\tau_n,n}}{\xi_{\tau_n}}-1\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{2\gamma^3}{1-2\gamma}\left[1-2\left(\frac{1-\gamma}{1-2\gamma}\right)^{-1/\gamma+1}\right]\right).$$

It is readily shown that the variance reduction factor $1-2((1-\gamma)/(1-2\gamma))^{-1/\gamma+1}$ in Theorem 3.2 is a monotonic function of γ, converges to 1 as $\gamma \uparrow 1/2$ and has limit $1-2e^{-1} \approx 0.264$ as $\gamma \downarrow 0$. This is illustrated again through the toy example of the Fréchet distribution in Figure 6, where it can be seen that Monte Carlo computations with the extreme quantile $q_{\alpha_n^\star}$ as control variate considerably improve upon vanilla Monte Carlo when the target expectile approaches the upper tail of the underlying distribution.

4 Discussion

This article discusses the calculation of expectiles from several angles. We showed that an exact computation of the expectiles of any discrete distribution can always be carried out. For continuous distributions whose distribution function and mean residual life function can be expressed in analytic form, a Newton-Raphson algorithm is shown to be an efficient way of calculating expectiles to a high degree of accuracy. When the distribution function and/or the mean residual life function is hard to compute, a Monte-Carlo algorithm resting on sample expectiles with the mean (resp. an extreme quantile) as a control variate is a reasonably accurate way to approximate central (resp. tail) expectiles, as we show for difficult but interesting examples including compound Poisson processes and stochastic differential equations.

In many statistical applications, the estimation of a location parameter of interest, such as a quantile or an expectile, is sought. The choice of a parametric family of distributions to describe the observations, which is a reasonable step in statistical modeling, naturally induces a function $\boldsymbol{\theta} \mapsto \xi_\tau(\boldsymbol{\theta})$ mapping parameter values to the expectile function. In such a situation, it is intuitively more efficient to use the plug-in estimator $\xi_\tau(\widehat{\boldsymbol{\theta}}_n)$ of $\xi_\tau(\boldsymbol{\theta})$ based on a model (for example maximum likelihood) estimator $\widehat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}$, than the sample

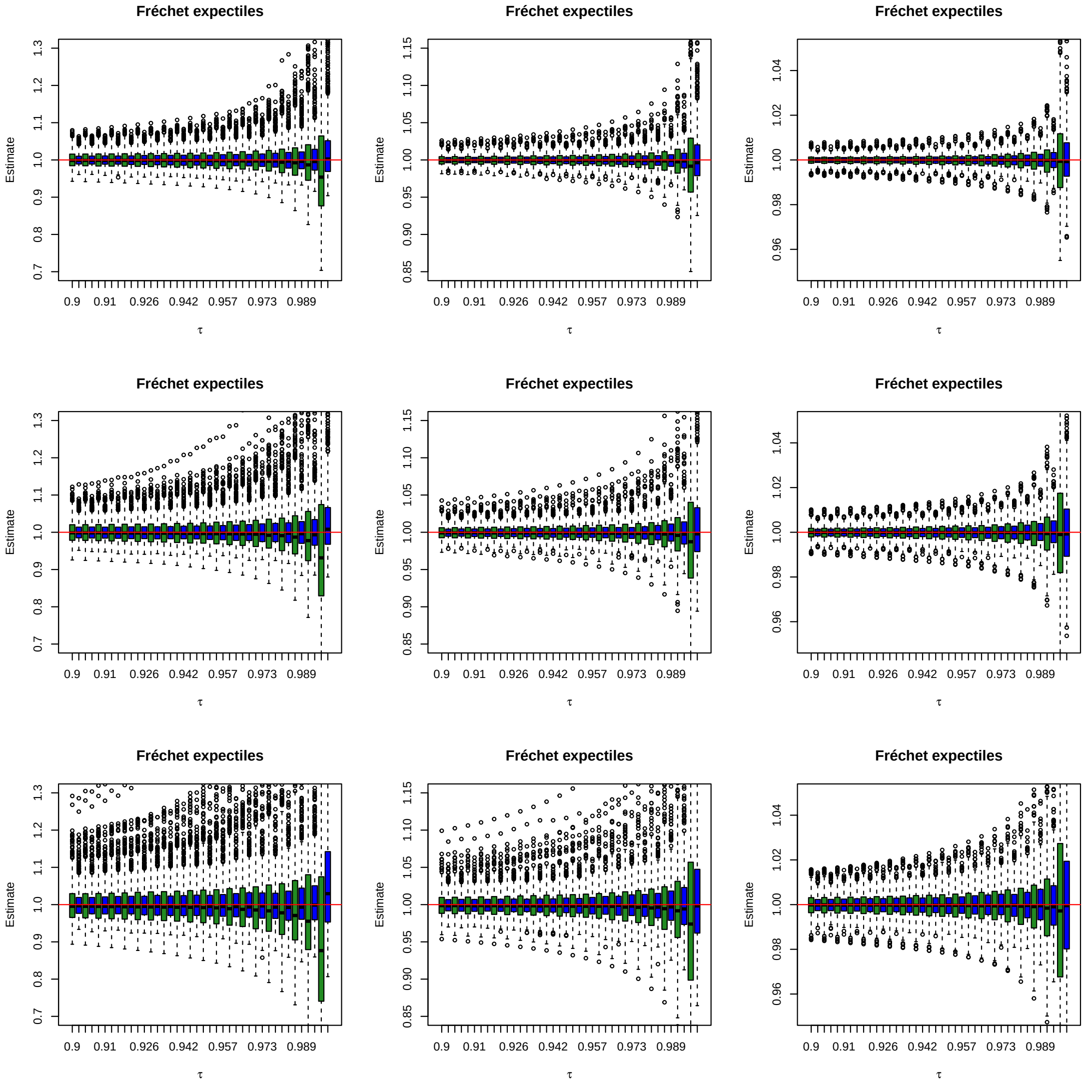

Fig. 6: Example of the Fréchet distribution. Boxplots of $\widehat{\xi}_{\tau,n}$ (green) and $\check{\xi}_{\tau,n}$ (blue), normalized by the true value ξ_τ calculated through the Newton-Raphson algorithm, for a regular grid of 20 values of $\tau \in [0.9, 0.995]$. We take 1,000 Monte-Carlo replications of an independent sample of size $n = 1{,}000$ (left), 10,000 (middle) and 100,000 (right), where $\gamma = 1/5$ (top panels), 1/4 (middle panels) and 1/3 (bottom panels).

LAWS expectile. In practice, this procedure is made possible by the computation techniques we proposed, since they allow the computation of the map $\boldsymbol{\theta} \mapsto \xi_\tau(\boldsymbol{\theta})$; quantifying the degree of improvement this brings over the LAWS estimator in statistical terms is an interesting question which is beyond the scope of this paper.

Despite its reasonable behavior, our Monte-Carlo approach with control variates is still computationally costly in the special case of time series: in our ARMA-GARCH example (see the second part of Example 3.1), for instance, ensuring that realizations have the correct, stationary distribution required a lengthy period of burn-in, and ensuring independence made us keep only the last data point in a given simulation after burn-in.

Instead, one may keep several data points that are sufficiently far apart in time (a particular, conservative case of thinning) in a given simulation after burn-in so as to preserve independence. An alternative option would be to keep whole blocks of data points, at the price of developing an appropriate asymptotic theory that allows to handle the autocorrelation structure within each block of the time series. Theoretical results along these lines are left for future research. Finally, we did not enter into the important question of extending our Monte-Carlo approach to the case when simulating from the target distribution is itself difficult, as is often the case in modern applications of Bayesian statistics. This is of genuine interest if expectiles are to be used in complex parametric models having a large number of parameters.

Acknowledgments

This research was supported by the French National Research Agency under the grants ANR-19-CE40-0013 (ExtremReg project), ANR-17-EURE-0010 (EUR CHESS) and ANR-11-LABX-0020-01 (Centre Henri Lebesgue). A. Daouia and G. Stupfler acknowledge financial support from the TSE-HEC ACPR Chair.

Statements and Declarations

The authors declare no competing interests. All authors contributed equally to the work, and read and approved the final manuscript.

References

Newey, W.K., Powell, J.L.: Asymmetric least squares estimation and testing. Econometrica **55**(4), 819–847 (1987)

Artzner, P., Delbaen, F., Eber, J.-M., Heath, D.: Coherent measures of risk. Mathematical Finance **9**(3), 203–228 (1999)

Gneiting, T.: Making and evaluating point forecasts. Journal of the American Statistical Association **106**(494), 746–762 (2011)

Breckling, J., Chambers, R.: M-quantiles. Biometrika **75**(4), 761–772 (1988)

Bellini, F., Klar, B., Müller, A., Gianin, E.R.: Generalized quantiles as risk measures. Insurance: Mathematics and Economics **54**, 41–48 (2014)

Ziegel, J.F.: Coherence and elicitability. Mathematical Finance **26**(4), 901–918 (2016)

Philipps, C.: Interpreting expectiles. Available at SSRN: `http://dx.doi.org/10.2139/ssrn.3881402` (2022)

Taylor, J.W.: Estimating Value at Risk and Expected Shortfall using expectiles. Journal of Financial Econometrics **6**(2), 231–252 (2008)

Kuan, C.-M., Yeh, J.-H., Hsu, Y.-C.: Assessing value at risk with care, the conditional autoregressive expectile models. Journal of Econometrics **150**(2), 261–270 (2009)

Sobotka, F., Kneib, T.: Geoadditive expectile regression. Computational Statistics & Data Analysis **56**(4), 755–767 (2012)

Holzmann, H., Klar, B.: Expectile asymptotics. Electronic Journal of Statistics **10**(2), 2355–2371 (2016)

Krätschmer, V., Zähle, H.: Statistical inference for expectile-based risk measures. Scandinavian Journal of Statistics **44**(2), 425–454 (2017)

Daouia, A., Girard, S., Stupfler, G.: Estimation of tail risk based on extreme expectiles. Journal of the Royal Statistical Society: Series B **80**(2), 263–292 (2018)

Daouia, A., Girard, S., Stupfler, G.: Tail expectile process and risk assessment. Bernoulli **26**(1), 531–556 (2020)

Daouia, A., Girard, S., Stupfler, G.: ExpectHill estimation, extreme risk and heavy tails. Journal of Econometrics **221**(1), 97–117 (2021)

Girard, S., Stupfler, G., Usseglio-Carleve, A.: Extreme conditional expectile estimation in heavy-tailed heteroscedastic regression models. Annals of Statistics **49**(6), 3358–3382 (2021)

Jones, M.C.: Expectiles and M-quantiles are quantiles. Statistics & Probability Letters **20**(2), 149–153 (1994)

Otto-Sobotka, F., Spiegel, E., Schnabel, S., Schulze Waltrup, L., Eilers, P., Kauermann, G., Kneib, T.: `Expectreg`: Expectile and Quantile Regression. (2022). R package version 0.52. https://CRAN.R-project.org/package=expectreg

Padoan, S., Stupfler, G.: `ExtremeRisks`: Extreme Risk Measures. (2020). R package version 0.0.4. https://cran.r-project.org/package=ExtremeRisks

Koenker, R.: When are expectiles percentiles? Econometric Theory **9**(3), 526–527 (1993)

Zou, H.: Generalizing Koenker's distribution. Journal of Statistical Planning and Inference **148**, 123–127 (2014)

Bellini, F., Di Bernardino, E.: Risk management with expectiles. The European Journal of Finance **23**(6), 487–506 (2017)

Hankin, R.K.S., Clausen, A., Murdoch, D.: `Gsl`: Wrapper for the Gnu Scientific Library. (2023). R package version 2.1-8. https://cran.r-project.org/package=gsl

Girard, S., Stupfler, G., Usseglio-Carleve, A.: Nonparametric extreme conditional expectile estimation. Scandinavian Journal of Statistics **49**(1), 78–115 (2022)

Daouia, A., Stupfler, G., Usseglio-Carleve, A.: Inference for extremal regression with dependent heavy-tailed data. https://hal.science/hal-03612202v2 (2023)

Laudagé, C., Turkalj, I.: Calculating Expectiles and Range Value-at-Risk using Quantum Computers. arXiv:2211.04456 (2022)

Birnbaum, Z.W.: An Inequality for Mill's Ratio. Annals of Mathematical Statistics **13**(2), 245–246 (1942)

Brockwell, P.J., Davis, R.A.: Time Series: Theory and Methods (Second Edition), Springer (1991)

Francq, C., Zakoïan, J.-M.: Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. Bernoulli **10**(4), 605–637 (2004)

Castillo, J., Soler, D.M., Serra, I.: `Ercv`: Fitting Tails by the Empirical Residual Coefficient of Variation. (2019). R package version 1.0.1. https://CRAN.R-project.org/package=ercv

Mező, I., Baricz, Á.: On the generalization of the Lambert W function. Transactions of the American Mathematical Society **369**(11), 7917–7934 (2017)

Koenker, R.: Quantile Regression, Cambridge University Press (2005)

Galanos, A., Kley, T.: `Rugarch`: Univariate GARCH Models. (2022). R package version 1.4-9. https://CRAN.R-project.org/package=rugarch

Asmussen, S., Albrecher, H.: Ruin Probabilities. Advanced Series On Statistical Science And Applied Probability, vol. 14. World Scientific (2010)

Dickson, D.C.: Insurance Risk and Ruin, Cambridge University Press (2016)

Black, F., Scholes, M.: The pricing of options and corporate liabilities. Journal of Political Economy **81**(3), 637–654 (1973)

Heston, S.L.: A closed-form solution for options with stochastic volatility with applications to bond and currency options. The Review of Financial Studies **6**(2), 327–343 (1993)

Stupfler, G., Usseglio-Carleve, A.: Composite bias-reduced L^p-quantile-based estimators of extreme quantiles and expectiles. Canadian Journal of Statistics **51**(2), 704–742 (2023)

Haan, L., Ferreira, A.: Extreme Value Theory: An Introduction. Springer, New York (2006)

Distribution (parameters)	Survival function $\overline{F}(x)$	Function $\varphi(x)+x\overline{F}(x)$	Expectation m	Support
Beta (α, β)	$1-\frac{B_x(\alpha,\beta)}{B(\alpha,\beta)}$	$\frac{\alpha}{\alpha+\beta}\left(1-\frac{B_x(\alpha+1,\beta)}{B(\alpha+1,\beta)}\right)$	$\frac{\alpha}{\alpha+\beta}$	$[0,1]$
Triangular	$1-\frac{x^2}{2}$ for $x\in[0,1]$, $\frac{(2-x)^2}{2}$ for $x\in[1,2]$	$-\frac{x^3}{3}+1$ for $x\in[0,1]$, $\frac{(2-x)^2(x+1)}{3}$ for $x\in[1,2]$	1	$[0,2]$
Pareto (γ)	$x^{-1/\gamma}$	$\frac{1}{1-\gamma}x^{1-1/\gamma}$	$\frac{1}{1-\gamma}$	$[1,\infty)$
Hall-Weiss (α, β)	$\frac{1}{2}\left(x^{-\alpha}+x^{-\alpha-\beta}\right)$	$\frac{1}{2}\left(\frac{\alpha}{\alpha-1}x^{1-\alpha}+\frac{\alpha+\beta}{\alpha+\beta-1}x^{1-\alpha-\beta}\right)$	$\frac{2(\alpha-1)(\alpha+\beta)+\beta}{2(\alpha-1)(\alpha+\beta-1)}$	$[1,\infty)$
Fréchet (α)	$1-\exp(-x^{-\alpha})$	$\Gamma_{x^{-\alpha}}\left(1-\frac{1}{\alpha}\right)$	$\Gamma\left(1-\frac{1}{\alpha}\right)$	$[0,\infty)$
Fisher (ν_1, ν_2)	$1-\frac{B_{\frac{\nu_1 x}{\nu_1 x+\nu_2}}\left(\frac{\nu_1}{2},\frac{\nu_2}{2}\right)}{B\left(\frac{\nu_1}{2},\frac{\nu_2}{2}\right)}$	$\frac{\nu_2}{\nu_2-2}\left(1-\frac{B_{\frac{\nu_1 x}{\nu_1 x+\nu_2}}\left(\frac{\nu_1}{2}+1,\frac{\nu_2}{2}-1\right)}{B\left(\frac{\nu_1}{2}+1,\frac{\nu_2}{2}-1\right)}\right)$	$\frac{\nu_2}{\nu_2-2}$	$[0,\infty)$
Burr (α, β)	$(1+x^\alpha)^{-\beta}$	$\beta B_{\frac{1}{1+x^\alpha}}\left(\beta-\frac{1}{\alpha},1+\frac{1}{\alpha}\right)$	$\beta B\left(\beta-\frac{1}{\alpha},1+\frac{1}{\alpha}\right)$	$[0,\infty)$
Dagum (α, β)	$1-\left(1+x^{-\alpha}\right)^{-\beta}$	$\beta\left\{B\left(\beta+\frac{1}{\alpha},1-\frac{1}{\alpha}\right)-B_{\frac{1}{1+x^{-\alpha}}}\left(\beta+\frac{1}{\alpha},1-\frac{1}{\alpha}\right)\right\}$	$\beta B\left(\beta+\frac{1}{\alpha},1-\frac{1}{\alpha}\right)$	$[0,\infty)$
Inverse-Gamma (α, λ)	$\frac{\Gamma_{\lambda/x}(\alpha)}{\Gamma(\alpha)}$	$\lambda\frac{\Gamma_{\lambda/x}(\alpha-1)}{\Gamma(\alpha)}$	$\frac{\lambda}{\alpha-1}$	$[0,\infty)$
Log-Normal (μ, s)	$\frac{1}{2}\left(1-\operatorname{erf}\left(\frac{\log(x)-\mu}{s\sqrt{2}}\right)\right)$	$\frac{1}{2}\left(1-\operatorname{erf}\left(\frac{\log(x)-\mu-s^2}{s\sqrt{2}}\right)\right)\exp\left(\mu+\frac{s^2}{2}\right)$	$\exp\left(\mu+\frac{s^2}{2}\right)$	$[0,\infty)$
Chi-squared (ν)	$1-\frac{\Gamma_{x/2}(\nu/2)}{\Gamma(\nu/2)}$	$\nu\left(1-\frac{\Gamma_{x/2}(\frac{\nu}{2}+1)}{\Gamma(\frac{\nu}{2}+1)}\right)$	ν	$[0,\infty)$
Weibull (β)	$\exp(-x^\beta)$	$\Gamma\left(1+\frac{1}{\beta}\right)-\Gamma_{x^\beta}\left(1+\frac{1}{\beta}\right)$	$\Gamma\left(1+\frac{1}{\beta}\right)$	$[0,\infty)$
Normal	$\overline{\Phi}(x)=1-\Phi(x)$	$\phi(x)$	0	$\mathbb{R}$
Student (ν)	$\overline{\Phi}_\nu(x)=1-\Phi_\nu(x)$	$\frac{\nu}{\nu-1}\left(1+\frac{x^2}{\nu}\right)\phi_\nu(x)$	0	$\mathbb{R}$
Logistic	$\frac{1}{1+e^x}$	$\log(1+e^{-x})+\frac{x}{1+e^x}$	0	$\mathbb{R}$

Table 2: The functions $\overline{F}$, φ and expectations for a catalog of continuous probability distributions. The third column reports the function $x\mapsto\varphi(x)+x\overline{F}(x)=\int_{\mathbb{R}}t\mathbb{1}_{\{t>x\}}\,\mu(\mathrm{d}t)$ appearing in the numerator of the Newton-Raphson iterative formula, and which in most situations has a simpler expression than the function φ itself. Here and throughout $\Gamma_x(a)$ (resp. $\Gamma(a)$) denotes the value of the lower incomplete Gamma function at x (resp. of the Gamma function) with parameter $a>0$, $B_x(\alpha,\beta)$ (resp. $B(\alpha,\beta)$) denotes the value of the lower incomplete Beta function at x (resp. of the Beta function) with parameters $\alpha,\beta>0$, erf denotes the Gauss error function, Φ (resp. ϕ) denotes the distribution function (resp. probability density function) of the standard Gaussian distribution, Φ_ν (resp. ϕ_ν) denotes the distribution function (resp. probability density function) of the Student distribution with ν degrees of freedom.

Appendix A Proofs of the theoretical results

Proof of Theorem 2.1. Clearly

$$\forall i \in I,\ \forall x \in [a_i, a_{i+1}),\ \varphi(x) = \int_{\mathbb{R}} (t-x)\mathbb{1}_{\{t>x\}}\,\mu(\mathrm{d}t) = \psi(a_i) - x\overline{F}(a_i)$$

where $\psi(x) = \int_{\mathbb{R}} t\mathbb{1}_{\{t>x\}}\,\mu(\mathrm{d}t)$ and $\overline{F}(x) = 1-F(x)$ with $F(x) = \mu((-\infty, x])$. Consequently, for any $i \in I$ and $x \in [a_i, a_{i+1})$,

$$(1-\tau)g_\tau(x) = -x\{(2\tau-1)\overline{F}(a_i) + 1-\tau\} + (2\tau-1)\psi(a_i) + (1-\tau)m$$

$$= -x(\tau\overline{F}(a_i) + (1-\tau)F(a_i)) + \tau\psi(a_i) + (1-\tau)(m-\psi(a_i))$$

where $m = \int_{\mathbb{R}} x\,\mu(\mathrm{d}x) = \sum_{k\in I} p_k a_k$. In other words, the function g_τ is continuous, piecewise linear and decreasing, and tends to $+\infty$ (resp. $-\infty$) as $x \to -\infty$ (resp. $x \to +\infty$). It follows that there is a unique index $i = i(\tau)$ such that the two inequalities

$$\tau\psi(a_i) + (1-\tau)(m-\psi(a_i)) \geq a_i(\tau\overline{F}(a_i) + (1-\tau)F(a_i))$$
$$\text{and } \tau\psi(a_{i+1}) + (1-\tau)(m-\psi(a_{i+1})) < a_{i+1}(\tau\overline{F}(a_{i+1}) + (1-\tau)F(a_{i+1}))$$

hold. With this index i, the expectile ξ_τ is the unique root of the linear function g_τ on the interval $[a_i, a_{i+1})$, namely:

$$\xi_\tau = \frac{\tau\psi(a_i) + (1-\tau)(m-\psi(a_i))}{\tau\overline{F}(a_i) + (1-\tau)F(a_i)}. \tag{A1}$$

Now $\psi(a_i) = \sum_{k>i} p_k a_k$ and $\overline{F}(a_i) = \sum_{k>i} p_k$, so that, for any i,

$$\tau\psi(a_i) + (1-\tau)(m-\psi(a_i)) - a_i(\tau\overline{F}(a_i) + (1-\tau)F(a_i))$$
$$= \tau\left(\sum_{k<i} p_k(a_i - a_k) + \sum_{k>i} p_k(a_k - a_i)\right) - \sum_{k<i} p_k(a_i - a_k).$$

The above pair of inequalities is therefore equivalent to

$$\frac{\sum_{k<i} p_k(a_i - a_k)}{\sum_{k<i} p_k(a_i - a_k) + \sum_{k>i} p_k(a_k - a_i)} \leq \tau < \frac{\sum_{k<i+1} p_k(a_{i+1} - a_k)}{\sum_{k<i+1} p_k(a_{i+1} - a_k) + \sum_{k>i+1} p_k(a_k - a_{i+1})}$$

as announced. The two identities involving ξ_τ follow immediately from (A1). □

Proof of Theorem 2.2. Define a continuous function $h = h_\tau$ by setting

$$h(x) = x - \frac{g_\tau(x)}{g'_\tau(x)}, \text{ so that } x_{n+1} = h(x_n).$$

Recall that $g_\tau : x \mapsto \frac{2\tau-1}{1-\tau}\varphi(x) + m - x$ is convex. It is also clear that the derivative g'_τ is negative. In particular, g_τ is decreasing, and since ξ_τ is the unique solution of the equation $g_\tau(x) = 0$, one has $g_\tau(x_0) > 0$ for any $x_0 < \xi_\tau$. Then

$$\forall x_0 < \xi_\tau,\ h(x_0) - x_0 = -\frac{g_\tau(x_0)}{g'_\tau(x_0)} > 0 \text{ and}$$

$$h(x_0)-\xi_\tau=\left\{\frac{\xi_\tau-x_0}{g_\tau(\xi_\tau)-g_\tau(x_0)}-\frac{1}{g_\tau'(x_0)}\right\}g_\tau(x_0)\leq 0$$

using the convexity property of g_τ. It follows that for any $x_0<\xi_\tau$, $x_0<h(x_0)\leq\xi_\tau$, and therefore, for any starting point $x_0<\xi_\tau$, the Newton-Raphson sequence of iterates (x_n) is nondecreasing and bounded and hence convergent. The limit must be a root of g_τ by taking limits in the equation $x_{n+1}=h(x_n)$, meaning that (x_n) converges to ξ_τ.

A Taylor expansion of g_τ on the interval $[x_n,\xi_\tau]\subset[x_0,\xi_\tau]$ (on which g_τ is twice continuously differentiable) with remainder in integral form entails

$$0=g_\tau(\xi_\tau)=g_\tau(x_n)+(\xi_\tau-x_n)g_\tau'(x_n)+\int_{x_n}^{\xi_\tau}(\xi_\tau-u)g_\tau''(u)\,\mathrm{d}u.$$

This is readily rewritten as

$$x_{n+1}-\xi_\tau=\frac{1}{g_\tau'(x_n)}\int_{x_n}^{\xi_\tau}(\xi_\tau-u)g_\tau''(u)\,\mathrm{d}u.$$

Then

$$\begin{aligned}|x_{n+1}-\xi_\tau|&\leq\frac{\max_{[x_n,\xi_\tau]}g_\tau''}{2|g_\tau'(x_n)|}(\xi_\tau-x_n)^2=\frac{1}{2}\frac{(2\tau-1)\max_{[x_n,\xi_\tau]}f}{1-\tau+(2\tau-1)\overline{F}(x_n)}|x_n-\xi_\tau|^2\\&\leq\frac{1}{2}\frac{(2\tau-1)\max_{[x_n,\xi_\tau]}f}{1-\tau+(2\tau-1)\overline{F}(\xi_\tau)}|x_n-\xi_\tau|^2\leq\frac{1}{2}\frac{(2\tau-1)\max_{[x_k,\xi_\tau]}f}{1-\tau+(2\tau-1)\overline{F}(\xi_\tau)}|x_n-\xi_\tau|^2\end{aligned}$$

for any $k<n$. The proof is complete. □

Proof of Theorem 2.3. From the last chain of inequalities in the proof of Theorem 2.2, we get, for any n,

$$|x_{n+1}-\xi_\tau|\leq\frac{1}{2}\frac{(2\tau-1)\max_{[x_0,\xi_\tau]}f}{1-\tau+(2\tau-1)\overline{F}(\xi_\tau)}|x_n-\xi_\tau|^2\leq\frac{1}{2}\frac{(2\tau-1)\max_{[(1-\varepsilon)\xi_\tau,\xi_\tau]}f}{1-\tau+(2\tau-1)\overline{F}(\xi_\tau)}|x_n-\xi_\tau|^2.$$

Take then $c>0$ sufficiently large and write

$$\overline{F}(x)=x^{-1/\gamma}\left\{c^{1/\gamma}\overline{F}(c)\exp\left(\int_c^x\eta(t)\frac{\mathrm{d}t}{t}\right)\right\}\text{ with }\eta(t)=\frac{1}{\gamma}-\frac{tf(t)}{\overline{F}(t)}.$$

The function η is continuous on $[c,\infty)$ and converges to 0 at infinity. By the representation theorem for regularly varying functions (see Theorem B.1.6 p.365 in de Haan and Ferreira, 2006), $\overline{F}$ is regularly varying with index $-1/\gamma$, and then f is regularly varying with index $-1/\gamma-1$. By the uniform convergence theorem for regularly varying functions (see Theorem B.1.4 p.363 in de Haan and Ferreira, 2006),

$$\frac{\xi_\tau}{\overline{F}(\xi_\tau)}\max_{[(1-\varepsilon)\xi_\tau,\xi_\tau]}f=\frac{\xi_\tau f(\xi_\tau)}{\overline{F}(\xi_\tau)}\max_{x\in[1-\varepsilon,1]}\frac{f(\xi_\tau x)}{f(\xi_\tau)}\to\frac{(1-\varepsilon)^{-1/\gamma-1}}{\gamma}\text{ as }\tau\uparrow 1.$$

Moreover (see Bellini et al., 2014; Daouia et al., 2018)

$$\frac{\overline{F}(\xi_\tau)}{1-\tau}\to\frac{1}{\gamma}-1\text{ as }\tau\uparrow 1.$$

Conclude from these two convergences that

$$\frac{1}{2}\frac{(2\tau-1)\max_{[(1-\varepsilon)\xi_\tau,\xi_\tau]}f}{1-\tau+(2\tau-1)\overline{F}(\xi_\tau)}=\frac{1}{\xi_\tau}\left((1-\varepsilon)^{-1/\gamma-1}\frac{1/\gamma-1}{2}+\mathrm{o}(1)\right)\text{ as }\tau\uparrow 1.$$

The proof is complete. $\square$

Proof of Theorem 3.1. We first prove the asymptotic normality statement, and for this, it is enough to show that $\widehat{\Sigma}_{12,n}/\widehat{\sigma}_n^2 \to \Sigma_{12}/\Sigma_{22}$ in probability. By the law of large numbers, $\widehat{\sigma}_n^2 \to \sigma^2$ in probability, so it suffices to show that $\widehat{\Sigma}_{12,n} \to \Sigma_{12}$ in probability. Finally, since $\widehat{\xi}_{\tau,n} \to \xi_\tau$ in probability (for example by Theorem 2 in Holzmann and Klar (2016)), it is enough to prove that $\widehat{\overline{F}}_n(\widehat{\xi}_{\tau,n}) \to \overline{F}(\xi_\tau)$ and $\widehat{\varphi}_n^{(2)}(\widehat{\xi}_{\tau,n}) \to \varphi^{(2)}(\xi_\tau)$ in probability.

Fix $\varepsilon > 0$. Then $\widehat{\xi}_{\tau,n} \in [\xi_\tau - \varepsilon, \xi_\tau + \varepsilon]$ with arbitrarily high probability as $n \to \infty$. Then clearly

$$\left|\widehat{\overline{F}}_n(\widehat{\xi}_{\tau,n}) - \widehat{\overline{F}}_n(\xi_\tau)\right| \le \frac{1}{n}\sum_{i=1}^n |\mathbb{1}_{\{X_i > \widehat{\xi}_{\tau,n}\}} - \mathbb{1}_{\{X_i > \xi_\tau\}}| \le \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{X_i \in [\xi_\tau - \varepsilon, \xi_\tau + \varepsilon]\}}$$

with arbitrarily high probability as $n \to \infty$. The upper bound converges to $\mu([\xi_\tau - \varepsilon, \xi_\tau + \varepsilon])$, by the law of large numbers, which is arbitrarily small as $\varepsilon \downarrow 0$. Conclude that $\widehat{\overline{F}}_n(\widehat{\xi}_{\tau,n}) = (\widehat{\overline{F}}_n(\widehat{\xi}_{\tau,n}) - \widehat{\overline{F}}_n(\xi_\tau)) + \widehat{\overline{F}}_n(\xi_\tau) \to \overline{F}(\xi_\tau)$ in probability, by the law of large numbers. We now prove that $\widehat{\varphi}_n(\widehat{\xi}_{\tau,n}) = \widehat{\varphi}_n^{(1)}(\widehat{\xi}_{\tau,n}) \to \varphi^{(1)}(\xi_\tau) = \varphi(\xi_\tau)$ before turning to the convergence of $\widehat{\varphi}_n^{(2)}(\widehat{\xi}_{\tau,n})$. Write

$$(X - x)\mathbb{1}_{\{X > x\}} - (X - x')\mathbb{1}_{\{X > x'\}} = (x' - x)\mathbb{1}_{\{X > x\}} + (X - x')(\mathbb{1}_{\{X > x\}} - \mathbb{1}_{\{X > x'\}})$$

to obtain

$$\begin{aligned}\left|\widehat{\varphi}_n^{(1)}(\widehat{\xi}_{\tau,n}) - \widehat{\varphi}_n^{(1)}(\xi_\tau)\right| &\le |\widehat{\xi}_{\tau,n} - \xi_\tau|\widehat{\overline{F}}_n(\widehat{\xi}_{\tau,n}) + \frac{1}{n}\sum_{i=1}^n |X_i - \xi_\tau||\mathbb{1}_{\{X_i > \widehat{\xi}_{\tau,n}\}} - \mathbb{1}_{\{X_i > \xi_\tau\}}| \\ &\le \varepsilon\left(\widehat{\overline{F}}_n(\widehat{\xi}_{\tau,n}) + \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{X_i \in [\xi_\tau - \varepsilon, \xi_\tau + \varepsilon]\}}\right) \le 2\varepsilon\end{aligned}$$

with arbitrarily high probability as $n \to \infty$. Using the law of large numbers, this again shows that $\widehat{\varphi}_n^{(1)}(\widehat{\xi}_{\tau,n}) \to \varphi^{(1)}(\xi_\tau)$ in probability. Finally

$$(X - x)^2 - (X - x')^2 = 2(x' - x)(X - x) - (x' - x)^2$$

so that, with arbitrarily high probability as $n \to \infty$,

$$\begin{aligned}\left|\widehat{\varphi}_n^{(2)}(\widehat{\xi}_{\tau,n}) - \widehat{\varphi}_n^{(2)}(\xi_\tau)\right| &\le |\widehat{\xi}_{\tau,n} - \xi_\tau|^2\widehat{\overline{F}}_n(\widehat{\xi}_{\tau,n}) + 2|\widehat{\xi}_{\tau,n} - \xi_\tau|\widehat{\varphi}_n^{(1)}(\widehat{\xi}_{\tau,n}) + \frac{1}{n}\sum_{i=1}^n (X_i - \xi_\tau)^2|\mathbb{1}_{\{X_i > \widehat{\xi}_{\tau,n}\}} - \mathbb{1}_{\{X_i > \xi_\tau\}}| \\ &\le \varepsilon\left(\varepsilon\widehat{\overline{F}}_n(\widehat{\xi}_{\tau,n}) + 2\widehat{\varphi}_n^{(1)}(\widehat{\xi}_{\tau,n}) + \varepsilon \times \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{X_i \in [\xi_\tau - \varepsilon, \xi_\tau + \varepsilon]\}}\right) \le 2\varepsilon\left(\varepsilon + \widehat{\varphi}_n^{(1)}(\widehat{\xi}_{\tau,n})\right).\end{aligned}$$

Conclude that $\widehat{\varphi}_n^{(2)}(\widehat{\xi}_{\tau,n}) \to \varphi^{(2)}(\xi_\tau)$ in probability, as required. The fact that $\widetilde{\xi}_{\tau,n}$ has the lowest asymptotic variance among all asymptotically unbiased linear combinations of $\widehat{\xi}_{\tau,n}$ and $\overline{X}_n - m$ is then obvious.

It remains to prove the assertions about the variance reduction factor $1 - R(\tau, \mu)$. This function is clearly zero at $\tau = 1/2$. Note also that

$$\varphi^{(2)}(x) = \mathbb{E}((X - x)^2\mathbb{1}_{\{X > x\}}) = 2\int_x^\infty (t - x)\overline{F}(t)\,\mathrm{d}t.$$

As a result, and since $\tau \mapsto \xi_\tau$ is continuously differentiable on $I = (\tau_1, \tau_2)$ (see Proposition 1(iii) in Holzmann and Klar, 2016), the function $\tau \mapsto 1 - R(\tau, \mu)$ is continuously differentiable on this interval. It remains to prove the statements about monotonicity. Set

$$\begin{aligned} u(\tau) = u(\tau,\mu) &= (1-\tau)\mathbb{E}((X-\xi_\tau)^2) + (2\tau-1)\varphi^{(2)}(\xi_\tau) \\ \text{and } v(\tau) = v(\tau,\mu) &= (1-\tau)^2\mathbb{E}((X-\xi_\tau)^2) + (2\tau-1)\varphi^{(2)}(\xi_\tau) \end{aligned}$$

so that $R(\tau, \mu) = (u(\tau))^2/(\sigma^2 v(\tau))$ and therefore

$$\frac{\partial R}{\partial \tau}(\tau,\mu) = \frac{u(\tau)}{\sigma^2 (v(\tau))^2}(2u'(\tau)v(\tau) - v'(\tau)u(\tau)).$$

Writing $u(\tau) = (1-\tau)\mathbb{E}((X-\xi_\tau)^2 \mathbb{1}_{\{X<\xi_\tau\}}) + \tau\mathbb{E}((X-\xi_\tau)^2 \mathbb{1}_{\{X>\xi_\tau\}})$ yields in particular that $u(\tau) > 0$ for any τ, meaning that the partial derivative $\frac{\partial R}{\partial \tau}(\tau, \mu)$ has the same sign as $2u'(\tau)v(\tau) - v'(\tau)u(\tau)$. Now

$$\begin{aligned} u'(\tau) &= 2\varphi^{(2)}(\xi_\tau) - \mathbb{E}((X-\xi_\tau)^2) - 2\frac{\mathrm{d}\xi_\tau}{\mathrm{d}\tau}((2\tau-1)\varphi(\xi_\tau) + (1-\tau)(m-\xi_\tau)) \\ &= 2\varphi^{(2)}(\xi_\tau) - \mathbb{E}((X-\xi_\tau)^2) \text{ (using (3))} \\ \text{and } v'(\tau) &= 2\varphi^{(2)}(\xi_\tau) - 2(1-\tau)\mathbb{E}((X-\xi_\tau)^2) - 2\frac{\mathrm{d}\xi_\tau}{\mathrm{d}\tau}((2\tau-1)\varphi(\xi_\tau) + (1-\tau)^2(m-\xi_\tau)) \\ &= 2\varphi^{(2)}(\xi_\tau) - 2(1-\tau)\mathbb{E}((X-\xi_\tau)^2) + 2\tau(1-\tau)\frac{\mathrm{d}\xi_\tau}{\mathrm{d}\tau}(m-\xi_\tau) \text{ (from (3) again).} \end{aligned}$$

Straightforward calculations yield

$$2u'(\tau)v(\tau) - v'(\tau)u(\tau) = 2\left((1-2\tau)\mathbb{E}((X-\xi_\tau)^2\mathbb{1}_{\{X<\xi_\tau\}})\mathbb{E}((X-\xi_\tau)^2\mathbb{1}_{\{X>\xi_\tau\}}) + \tau(1-\tau)u(\tau)\frac{\mathrm{d}\xi_\tau}{\mathrm{d}\tau}(m-\xi_\tau)\right).$$

This quantity is positive on $(0, 1/2)$ and negative on $(1/2, 1)$ because $\tau \mapsto \xi_\tau$ is strictly increasing (see Proposition 1(ii) in Holzmann and Klar, 2016) and $\xi_{1/2} = m$. The proof is complete. □

Proof of Theorem 3.2. Let $\check{\xi}_{\tau_n,n} = \widehat{\xi}_{\tau_n,n} + c(\widehat{q}_{\alpha_n,n} - q_{\alpha_n})$, where c is to be found so as to minimize the relative asymptotic variance of $\check{\xi}_{\tau_n,n}$. Write

$$\sqrt{n(1-\tau_n)}\left(\frac{\check{\xi}_{\tau_n,n}}{\xi_{\tau_n}} - 1\right) = \sqrt{n(1-\tau_n)}\left(\frac{\widehat{\xi}_{\tau_n,n}}{\xi_{\tau_n}} - 1\right) + c\frac{q_{\alpha_n}}{\xi_{\tau_n}} \times \sqrt{n(1-\tau_n)}\left(\frac{\widehat{q}_{\alpha_n,n}}{q_{\alpha_n}} - 1\right).$$

Applying Proposition 3.1 entails that the desired value of c satisfies

$$c\frac{q_{\alpha_n}}{\xi_{\tau_n}} = -\frac{V_{12}}{V_{22}} = -\left(\left\{\min\left(\frac{\lambda}{1/\gamma-1}, 1\right)\right\}^{1-\gamma} + \lambda\left\{\min\left(\frac{\lambda}{1/\gamma-1}, 1\right)\right\}^{-\gamma} - \lambda\right).$$

Apply Proposition 1(i) in Daouia et al. (2020) to obtain

$$\frac{q_{\alpha_n}}{\xi_{\tau_n}} = \frac{q_{\alpha_n}}{q_{\tau_n}}\frac{q_{\tau_n}}{\xi_{\tau_n}} \to \lambda^{-\gamma}(1/\gamma-1)^\gamma \text{ as } n \to \infty,$$

leading to a choice of c minimizing the relative asymptotic variance of $\breve{\xi}_{\tau_n,n}$ as

$$c = -\left(\frac{\lambda}{1/\gamma - 1}\right)^{\gamma} \left(\left\{\min\left(\frac{\lambda}{1/\gamma - 1}, 1\right)\right\}^{1-\gamma} + \lambda \left\{\min\left(\frac{\lambda}{1/\gamma - 1}, 1\right)\right\}^{-\gamma} - \lambda\right).$$

The statement on the asymptotic normality of $\breve{\xi}_{\tau_n,n}/\xi_{\tau_n} - 1$ with this choice of c is then immediate. Finding the maximum of $\lambda \mapsto C(\gamma, \lambda)$ is done by noting that this function is decreasing past $1/\gamma - 1$, and

$$\forall \lambda \in (0, 1/\gamma - 1), C(\gamma, \lambda) = \frac{1 - 2\gamma}{2\gamma} \left(\frac{\lambda^{1/2-\gamma}}{\gamma(1/\gamma - 1)^{1-\gamma}} - \sqrt{\lambda}\right)^2.$$

Maximizing this function over $(0, 1/\gamma - 1)$ is straightforward and leads to the value $\lambda^\star$ specified in the statement of Theorem 3.2. The last asymptotic normality result follows by plugging $\lambda^\star$ into $C(\gamma, \lambda)$. □

Appendix B Detailed calculations related to the examples

Distribution supported on a set with three elements (Example 2.2)

Let μ be the probability distribution on a set $\{a, b, c\}$ with $a < b < c$ characterized by $\mu(\{b\}) = p$ and $\mu(\{c\}) = q$, with $p, q > 0$ and $p + q < 1$. Then, from Corollary 2.1,

$$\xi_\tau = \frac{\tau(pb + qc) + (1 - \tau)(1 - p - q)a}{(2\tau - 1)(p + q) + 1 - \tau}, \text{ for } \tau \le 1 - \frac{q(c - b)}{(1 - p)(b - a) + q(a + c - 2b)},$$

and

$$\xi_\tau = \frac{\tau qc + (1 - \tau)\{(1 - p - q)a + pb\}}{(2\tau - 1)q + 1 - \tau} \text{ otherwise.}$$

In particular, for the distribution μ on $\{0, 1, 2\}$ with $\mu(\{1\}) = p$ and $\mu(\{2\}) = q$,

$$\xi_\tau = \begin{cases} \dfrac{\tau(p + 2q)}{(2\tau - 1)(p + q) + 1 - \tau} & \text{for } \tau \le 1 - \dfrac{q}{1 - p}, \\ \dfrac{2\tau q + (1 - \tau)p}{(2\tau - 1)q + 1 - \tau} & \text{otherwise.} \end{cases}$$

Taking $p = p_1(1 - p_2) + p_2(1 - p_1)$ and $q = p_1 p_2$, for $p_1, p_2 \in (0, 1)$, yields the expectile of the sum of two independent random variables having Bernoulli distributions with parameters p_1 and p_2:

$$\xi_\tau = \begin{cases} \dfrac{\tau(p_1 + p_2)}{(2\tau - 1)(p_1 + p_2 - p_1 p_2) + 1 - \tau} & \text{for } \tau \le \dfrac{1 - p_1 - p_2 + p_1 p_2}{1 - p_1 - p_2 + 2p_1 p_2}, \\ \dfrac{(2\tau - 1)2p_1 p_2 + (1 - \tau)(p_1 + p_2)}{(2\tau - 1)p_1 p_2 + 1 - \tau} & \text{otherwise.} \end{cases}$$

Uniform distribution on $\{1, \ldots, n\}$ (Example 2.3)

Fix $n \ge 2$. For the uniform distribution on $\{1, \ldots, n\}$, solving the inequalities of Corollary 2.2 is equivalent to finding the unique index $i \in \{1, \ldots, n - 1\}$ such that

$$\frac{i(i - 1)}{i(i - 1) + (n - i)(n - i + 1)} \le \tau < \frac{i(i + 1)}{i(i + 1) + (n - i)(n - i - 1)}.$$

This is equivalent to finding the unique solution (which we already know to exist, by Corollary 2.2) to the inequalities $P_\tau(i+1) < 0 \le P_\tau(i)$ for $i \in \{1, \dots, n-1\}$, where P_τ is the polynomial

$$P_\tau(x) = (2\tau - 1)x^2 - \{2\tau(n+1) - 1\}x + \tau n(n+1).$$

This polynomial has discriminant $4\tau(1-\tau)(n+1)(n-1) + 1 > 0$ and then (for $\tau \neq 1/2$) two real roots $x_{\tau,-}$ and $x_{\tau,+}$ defined as

$$x_{\tau,\pm} = \frac{2\tau(n+1) - 1 \pm \sqrt{4\tau(1-\tau)(n+1)(n-1)+1}}{2(2\tau-1)}.$$

A straightforward calculation yields $P_\tau(1) = \tau n(n-1) > 0$ and $P_\tau(n) = -(1-\tau)n(n-1) < 0$. It follows that when $\tau > 1/2$ (resp. $\tau < 1/2$), only the lowest (resp. largest) of the two roots $x_{\tau,-}$ and $x_{\tau,+}$ belongs to the interval $[1, n]$. Conclude that, in both cases, $P_\tau(i+1) < 0 \le P_\tau(i) \Leftrightarrow i \le x_{\tau,-} < i+1 \Leftrightarrow i = \lfloor x_{\tau,-} \rfloor$. With this index i,

$$\xi_\tau = \frac{\tau n(n+1) - (2\tau-1)i(i+1)}{2\tau n - 2(2\tau-1)i}.$$

Consequently

$$\xi_\tau = \begin{cases} \dfrac{n(n+1)}{2} \text{ when } \tau = 1/2, \\ \dfrac{\tau n(n+1) - (2\tau-1)\lfloor x_\tau \rfloor(\lfloor x_\tau \rfloor + 1)}{2\tau n - 2(2\tau-1)\lfloor x_\tau \rfloor} \text{ otherwise,} \end{cases}$$

with

$$x_\tau = \frac{2\tau(n+1) - 1 - \sqrt{4\tau(1-\tau)(n+1)(n-1)+1}}{2(2\tau-1)}.$$

Geometric distribution (Example 2.4)

For the geometric distribution with success probability $p \in (0,1)$, namely, $\mu(\{k\}) = p(1-p)^{k-1}$ for any positive integer k, the inequalities of Theorem 2.1 read as

$$\frac{(1-p)^i - (1-pi)}{2(1-p)^i - (1-pi)} \le \tau < \frac{(1-p)^{i+1} - (1-p(i+1))}{2(1-p)^{i+1} - (1-p(i+1))}.$$

Solving these inequalities is equivalent to finding the index $i \ge 1$ such that $h_\tau(i+1) < 0 \le h_\tau(i)$, where

$$h_\tau(x) = (2\tau - 1)(1-p)^x - (1-\tau)(px - 1).$$

Straightforward calculations reveal that the unique root x_τ of h_τ over $[1, +\infty)$ satisfies the equation

$$-\log(1-p)\left(x_\tau - \frac{1}{p}\right)\exp\left(-\log(1-p)\left(x_\tau - \frac{1}{p}\right)\right) = \frac{2\tau-1}{1-\tau}\left(-\frac{\log(1-p)}{p}(1-p)^{1/p}\right).$$

This is a transcendental equation (unless $\tau \neq 1/2$, for which $x_{1/2} = 1/p$). Nevertheless, since by construction

$$-\log(1-p)\left(x_\tau - \frac{1}{p}\right) \ge -\log(1-p)\left(1 - \frac{1}{p}\right) > -1$$

for any $p \in (0,1)$, and

$$\frac{2\tau-1}{1-\tau}\left(-\frac{\log(1-p)}{p}(1-p)^{1/p}\right) = \frac{2\tau-1}{1-\tau}\{-(1-p)^{1/p}\log((1-p)^{1/p})\} > -e^{-1}$$

for any $\tau, p \in (0,1)$, one may express x_τ using the main branch of Lambert's W function, that is

$$x_\tau = \frac{1}{p} - \frac{1}{\log(1-p)} W\left(-\frac{(1-p)^{1/p}\log(1-p)}{p}\frac{2\tau-1}{1-\tau}\right)$$

where, for $x > 0$, $W(x)$ is the unique (positive) solution to the equation $we^w = x$. The main branch of the Lambert function is available numerically in `R` using (for instance) the `gsl` package (Hankin et al., 2023), acting as a wrapper for the GNU Scientific Library.

Note further that when $\tau > 1/2$, the function h_τ is obviously decreasing, so the inequalities $h_\tau(i+1) < 0 \leq h_\tau(i)$ are equivalent to $i \leq x_\tau < i+1$, *i.e.* $i = \lfloor x_\tau \rfloor$. When $\tau < 1/2$, it is readily shown that h'_τ is decreasing and $h'_\tau(1) = -(1-2\tau)(1-p)\log(1-p) - (1-\tau)p < 0$ for any $p \in (0,1)$, so again h_τ is decreasing and $h_\tau(i+1) < 0 \leq h_\tau(i) \Leftrightarrow i = \lfloor x_\tau \rfloor$. Conclude that

$$\xi_\tau = \frac{(2\tau-1)(1-p)^{\lfloor x_\tau \rfloor}(1+p\lfloor x_\tau \rfloor)+1-\tau}{p\{(2\tau-1)(1-p)^{\lfloor x_\tau \rfloor}+1-\tau\}} \text{ with}$$
$$x_\tau = \frac{1}{p} - \frac{1}{\log(1-p)} W\left(-\frac{(1-p)^{1/p}\log(1-p)}{p}\frac{2\tau-1}{1-\tau}\right).$$

Cardano and Ferrari formulae (relevant to Section 2.3.1)

To solve a real cubic polynomial equation of the form $x^3+bx^2+cx+d=0$, Cardano's method consists in letting $p = c - b^2/3$ and $q = d + b(2b^2-9c)/27$, and in computing the discriminant $\Delta = -4p^3 - 27q^2$ of the so-called depressed cubic $X^3 + pX + q = 0$. If $\Delta \leq 0$, then the unique real root of the equation is given by

$$x = \sqrt[3]{\frac{-q+\sqrt{\frac{-\Delta}{27}}}{2}} + \sqrt[3]{\frac{-q-\sqrt{\frac{-\Delta}{27}}}{2}} - \frac{b}{3}.$$

If on the contrary $\Delta > 0$, then necessarily $p < 0$ and there are 3 real solutions, given by Viète's formula:

$$x = 2\sqrt{\frac{-p}{3}}\cos\left(\frac{1}{3}\arccos\left(\frac{3q}{2p}\sqrt{\frac{3}{-p}}\right) + \frac{2k\pi}{3}\right) - \frac{b}{3},$$

$k \in \{0,1,2\}$. To solve a real quartic polynomial equation $x^4+bx^3+cx^2+dx+e=0$, Ferrari's method first finds a root λ to the cubic equation $8\lambda^3 - 4c\lambda^2 + (2bd-8e)\lambda - b^2e + 4ce - d^2 = 0$. The four (possibly complex) solutions to the quartic equation are then

$$\frac{\varepsilon_1\sqrt{2\lambda - c + \frac{b^2}{4}}}{2} - \frac{b}{4} + \frac{\varepsilon_2\sqrt{-2\lambda - c - \varepsilon_1\left(\frac{2(d-b)\lambda}{\sqrt{2\lambda-c+\frac{b^2}{4}}} + b\sqrt{2\lambda - c + \frac{b^2}{4}}\right) + \frac{b^2}{2}}}{2}$$

where $\varepsilon_1, \varepsilon_2 \in \{-1, 1\}$.

Student distribution with $\nu = 4$ degrees of freedom (Example 2.12)

Consider the Student distribution with ν degrees of freedom, having probability density function

$$f(x) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu\pi}\Gamma(\nu/2)}\left(1+\frac{x^2}{\nu}\right)^{-(\nu+1)/2}, \ x \in \mathbb{R}.$$

Here Γ is Euler's Gamma function. When $\nu = 4$, the probability density function simplifies to

$$f(x) = \frac{3}{8}\left(1 + \frac{x^2}{4}\right)^{-5/2}, \ x \in \mathbb{R}.$$

The change of variables $t = 2\tan(\theta)$ combined with the trigonometric identities $\cos(3\phi) = 4\cos^3(\phi) - 3\cos(\phi)$, $\sin(3\phi) = 3\sin(\phi) - 4\sin^3(\phi)$ and $\sin(\arctan\theta) = \theta/\sqrt{1+\theta^2}$ then yield, after straightforward calculations, the following closed form for the survival function:

$$\overline{F}(x) = \int_x^\infty f(t)\,\mathrm{d}t = \frac{1}{2} - \frac{x}{8}\frac{3 + x^2/2}{(1 + x^2/4)^{3/2}}.$$

Further straightforward calculations based on the change of variables $u = t^2$ then provide

$$\varphi(x) = \int_x^\infty \overline{F}(t)\,\mathrm{d}t = \frac{1}{2}\left(\frac{x^2+2}{\sqrt{x^2+4}} - x\right).$$

Since the Student distribution is centered, $m = 0$ and Equation (3) is

$$\xi_\tau^4 + 4\xi_\tau^2 - \frac{(2\tau-1)^2}{\tau(1-\tau)} = 0.$$

This is a biquadratic equation, leading to $\xi_\tau^2 = -2 + 1/\sqrt{\tau(1-\tau)}$ because $\xi_\tau^2 \geq 0$, and then

$$\xi_\tau = \operatorname{sign}(2\tau - 1)\sqrt{\frac{1}{\sqrt{\tau(1-\tau)}} - 2}.$$

In general, the distribution function and mean residual life function of the Student distribution involve the hypergeometric function. It is not hard to see that, while the distribution function and mean residual life function can in fact be written in closed form when ν is an even integer, resulting in a polynomial equation characterizing ξ_τ, only the cases $\nu \in \{2, 4, 6\}$ result in an equation of degree 4 or lower.

Fisher distribution with $(4, 4)$ degrees of freedom (Example 2.13)

The Fisher distribution with degrees of freedom $\nu_1 > 0$ and $\nu_2 > 0$ has density function

$$f(x) = \frac{(\nu_1/\nu_2)^{\nu_1/2}}{B(\nu_1/2, \nu_2/2)} x^{\nu_1/2-1}(1 + \nu_1 x/\nu_2)^{-(\nu_1+\nu_2)/2},$$

$x > 0$, where B is the Beta function. In the specific case $\nu_1 = \nu_2 = 4$, one finds $\varphi(x) = (3x+2)/(x+1)^2$ for $x > 0$, and $m = 2$. Equation (3) is thus equivalent to the cubic equation

$$\xi_\tau^3 - \frac{3\tau}{1-\tau}\xi_\tau - \frac{2\tau}{1-\tau} = 0.$$

The discriminant of this equation is $\Delta = 108\tau^2(2\tau-1)/(1-\tau)^3$. If $\tau \leq 1/2$, then $\Delta \leq 0$ and the unique solution is

$$\xi_\tau = \sqrt[3]{\frac{\tau}{1-\tau}}\left(\sqrt[3]{1 + \sqrt{\frac{1-2\tau}{1-\tau}}} + \sqrt[3]{1 - \sqrt{\frac{1-2\tau}{1-\tau}}}\right).$$

If now $\tau > 1/2$, then $\Delta > 0$ and the 3 possible solutions are

$$\xi_\tau = 2\sqrt{\frac{\tau}{1-\tau}}\cos\left(\frac{1}{3}\arccos\left(\sqrt{\frac{1-\tau}{\tau}}\right) + \frac{2k\pi}{3}\right),$$

$k \in \{0,1,2\}$. Since $\tau > 1/2$, $\arccos(\sqrt{(1-\tau)/\tau}) \in [0, \pi/2]$, and therefore (taking the constraint $\xi_\tau \geq 0$ into account) $k = 0$ is the only admissible solution, namely

$$\xi_\tau = 2\sqrt{\frac{\tau}{1-\tau}}\cos\left(\frac{1}{3}\arccos\left(\sqrt{\frac{1-\tau}{\tau}}\right)\right).$$

Pareto distribution with extreme value index $1/4$ (Example 2.14)

The Pareto distribution with extreme value index $\gamma > 0$ has survival function $\overline{F}(x) = x^{-1/\gamma}$ for $x > 1$. This distribution has a finite first moment when $\gamma < 1$, and since $\varphi(x) = \gamma x^{1-1/\gamma}/(1-\gamma)$ for $x > 1$ and $m = 1/(1-\gamma)$, Equation (3) leads to

$$(1-\gamma)(1-\tau)\xi_\tau^{1/\gamma} - (1-\tau)\xi_\tau^{1/\gamma-1} - \gamma(2\tau-1) = 0.$$

When $\gamma = 1/4$, this is the quartic equation $\xi_\tau^4 + b\xi_\tau^3 + c\xi_\tau^2 + d\xi_\tau + e = 0$, where $b = -4/3$, $c = 0$, $d = 0$ and $e = (1-2\tau)/(3(1-\tau))$. Ferrari's method leads to finding $\lambda = \lambda_\tau$ which is a root of the cubic equation

$$\lambda_\tau^3 - \frac{1-2\tau}{3(1-\tau)}\lambda_\tau - \frac{2}{9}\frac{1-2\tau}{3(1-\tau)} = 0.$$

The discriminant of this equation is

$$\Delta = -\frac{4}{27}\frac{(1-2\tau)^2}{(1-\tau)^2}\frac{\tau}{1-\tau} \leq 0$$

for all $\tau \in (0,1)$, from which the unique solution of the equation involving λ is

$$\lambda_\tau = \sqrt[3]{\frac{1-2\tau+|1-2\tau|\sqrt{\frac{\tau}{1-\tau}}}{27(1-\tau)}} + \sqrt[3]{\frac{1-2\tau-|1-2\tau|\sqrt{\frac{\tau}{1-\tau}}}{27(1-\tau)}}.$$

Ferrari's method yields four possible solutions. The only real-valued solution greater than 1 is obtained with $\varepsilon_1 = \varepsilon_2 = 1$, leading to the solution

$$\xi_\tau = \frac{1}{2}\sqrt{-2\lambda_\tau - \frac{8}{3}\frac{\lambda_\tau}{\sqrt{2\lambda_\tau + \frac{4}{9}}} + \frac{4}{3}\sqrt{2\lambda_\tau + \frac{4}{9}} + \frac{8}{9}} + \frac{1}{2}\sqrt{2\lambda_\tau + \frac{4}{9}} + \frac{1}{3}.$$

Appendix C Catalog of expectile functions of continuous distributions

This section provides a catalog of expectile functions of continuous distributions, obtained either via numerical means or, in exceptional cases, in closed or analytic form. Tables C1, C2, C3 and C4 give reference values for the expectiles of the standard Gaussian, log-normal, Student and chi-squared distributions, respectively. Figures C1, C2, C3, C4 provide graphical representations of the corresponding expectile functions over $(0,1)$. Table C5 gathers closed-form expressions for expectiles of certain bounded continuous distributions. Table C6 gives analytic-form expressions for expectiles of some unbounded continuous distributions. Tables C7 and C8 list closed-form expressions for expectiles of the Hall-Weiss distribution and the Pareto distribution, respectively, with particular parameters.

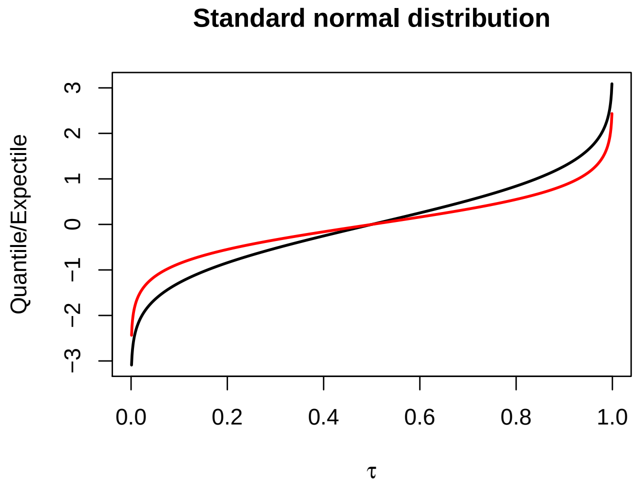

Fig. C1: Quantiles (black) and expectiles (red) of the standard Gaussian distribution, as functions of $\tau \in (0,1)$.

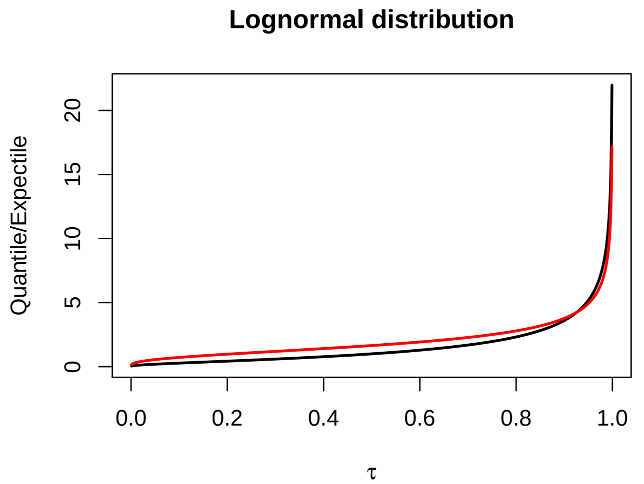

Fig. C2: Quantiles (black) and expectiles (red) of the log-normal distribution, as functions of $\tau \in (0,1)$.

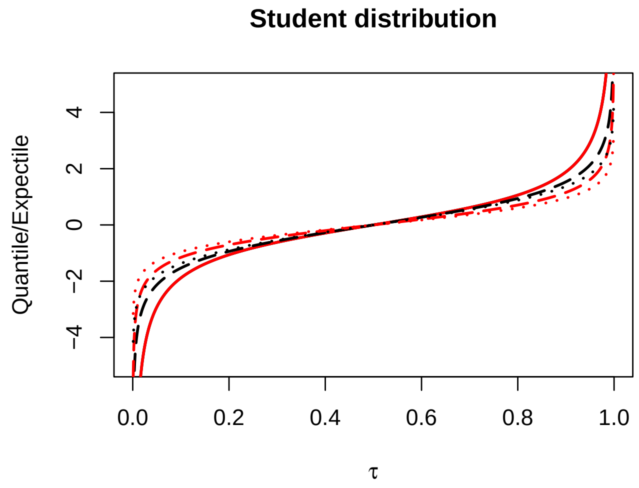

Fig. C3: Quantiles (black) and expectiles (red) of the Student distribution with 2 (solid curves), 4 (dashed curves) and 10 (dotted curves) degrees of freedom, as functions of $\tau \in (0, 1)$. The Student distribution with 2 degrees of freedom is Koenker's distribution (Koenker, 1993), for which quantiles and expectiles are identical.

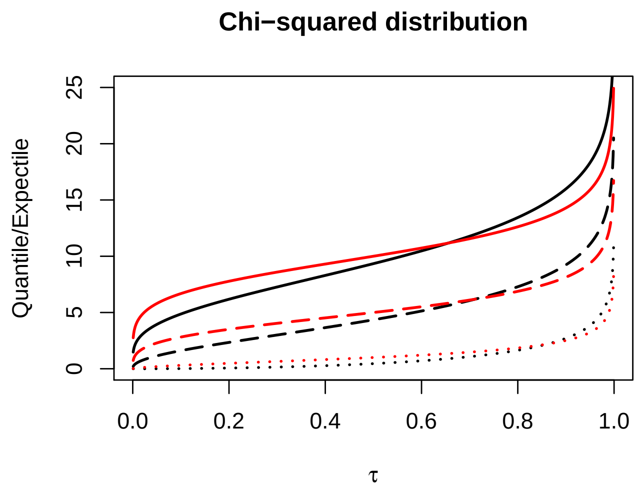

Fig. C4: Quantiles (black) and expectiles (red) of the chi-squared distribution with 1 (dotted curves), 5 (dashed curves) and 10 (solid curves) degrees of freedom, as functions of $\tau \in (0, 1)$.

τ	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	$-\infty$	-1.717	-1.478	-1.332	-1.225	-1.140	-1.069	-1.008	-0.954	-0.906
0.1	-0.862	-0.821	-0.784	-0.749	-0.716	-0.684	-0.655	-0.627	-0.600	-0.574
0.2	-0.549	-0.525	-0.502	-0.479	-0.458	-0.436	-0.416	-0.395	-0.376	-0.356
0.3	-0.337	-0.318	-0.300	-0.282	-0.264	-0.247	-0.229	-0.212	-0.195	-0.178
0.4	-0.162	-0.145	-0.129	-0.112	-0.096	-0.080	-0.064	-0.048	-0.032	-0.016
0.5	0	0.016	0.032	0.048	0.064	0.080	0.096	0.112	0.129	0.145
0.6	0.162	0.178	0.195	0.212	0.229	0.247	0.264	0.282	0.300	0.318
0.7	0.337	0.356	0.376	0.395	0.416	0.436	0.458	0.479	0.502	0.525
0.8	0.549	0.574	0.600	0.627	0.655	0.684	0.716	0.749	0.784	0.821
0.9	0.862	0.906	0.954	1.008	1.069	1.140	1.225	1.332	1.478	1.717

Table C1: Table of expectiles of the standard Gaussian distribution, computed via the Newton-Raphson algorithm (see Example 2.11).

τ	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0	0.317	0.399	0.459	0.508	0.552	0.591	0.626	0.660	0.691
0.1	0.721	0.750	0.777	0.804	0.829	0.855	0.879	0.903	0.927	0.950
0.2	0.973	0.996	1.018	1.041	1.063	1.085	1.106	1.128	1.150	1.171
0.3	1.193	1.215	1.236	1.258	1.279	1.301	1.323	1.345	1.367	1.389
0.4	1.412	1.434	1.457	1.480	1.503	1.527	1.551	1.575	1.599	1.624
0.5	1.649	1.674	1.700	1.726	1.753	1.780	1.808	1.837	1.866	1.895
0.6	1.926	1.957	1.988	2.021	2.055	2.089	2.125	2.161	2.199	2.238
0.7	2.279	2.321	2.364	2.410	2.457	2.506	2.558	2.612	2.669	2.730
0.8	2.793	2.861	2.932	3.009	3.092	3.181	3.277	3.382	3.498	3.627
0.9	3.770	3.933	4.121	4.340	4.603	4.927	5.347	5.925	6.819	8.584

Table C2: Table of expectiles of the standard log-normal distribution, computed via the Newton-Raphson algorithm.

	τ							
ν	0.5	0.9	0.95	0.975	0.99	0.995	0.999	0.9995
2	0	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0	1.320	1.890	2.549	3.626	4.656	8.121	10.270
4	0	1.155	1.609	2.099	2.837	3.490	5.444	6.537
5	0	1.077	1.480	1.899	2.503	3.011	4.430	5.173
6	0	1.032	1.407	1.788	2.321	2.756	3.914	4.494
7	0	1.002	1.359	1.717	2.206	2.598	3.606	4.095
8	0	0.981	1.326	1.667	2.128	2.491	3.403	3.834
9	0	0.966	1.302	1.631	2.072	2.414	3.259	3.651
10	0	0.954	1.283	1.604	2.029	2.356	3.152	3.516

Table C3: Table of expectiles of the Student distribution with ν degrees of freedom, computed via the Newton-Raphson algorithm. The number of degrees of freedom varies along rows, while the level τ varies along columns.

	τ								
ν	0.01	0.025	0.05	0.1	0.5	0.9	0.95	0.975	0.99
1	0.069	0.124	0.194	0.305	1	2.513	3.231	4.007	5.122
2	0.272	0.422	0.588	0.820	2	4.080	4.982	5.926	7.243
3	0.585	0.835	1.093	1.435	3	5.496	6.531	7.597	9.060
4	0.979	1.325	1.668	2.107	4	6.839	7.983	9.150	10.736
5	1.433	1.869	2.290	2.817	5	8.137	9.378	10.632	12.325
6	1.933	2.454	2.947	3.555	6	9.405	10.732	12.065	13.854
7	2.469	3.071	3.631	4.314	7	10.651	12.056	13.461	15.338
8	3.035	3.712	4.336	5.090	8	11.878	13.356	14.829	16.788
9	3.626	4.375	5.059	5.879	9	13.091	14.638	16.174	18.210
10	4.237	5.056	5.797	6.680	10	14.292	15.904	17.499	19.608

Table C4: Table of expectiles of the chi-squared distribution with ν degrees of freedom, computed via the Newton-Raphson algorithm. The number of degrees of freedom varies along rows, while the level τ varies along columns.

Distribution (parameters)	ξ_τ
Uniform ($a < b$), $\tau \neq 1/2$	$\frac{\tau b-(1-\tau)a-(b-a)\sqrt{\tau(1-\tau)}}{2\tau-1}$
Triangular, $\tau < 1/2$	$\sqrt[3]{\frac{\tau}{1-2\tau}}\left(\sqrt[3]{\sqrt{\frac{9-10\tau}{1-2\tau}}+3}-\sqrt[3]{\sqrt{\frac{9-10\tau}{1-2\tau}}-3}\right)$
Triangular, $\tau > 1/2$	$2-\sqrt[3]{\frac{1-\tau}{2\tau-1}}\left(\sqrt[3]{\sqrt{\frac{10\tau-1}{2\tau-1}}+3}-\sqrt[3]{\sqrt{\frac{10\tau-1}{2\tau-1}}-3}\right)$
Beta ($\alpha = 2$, $\beta = 1$), $\tau < 1/2$	$\sqrt[3]{\tau\frac{1+\sqrt{\frac{1-\tau}{1-2\tau}}}{1-2\tau}}+\sqrt[3]{\tau\frac{1-\sqrt{\frac{1-\tau}{1-2\tau}}}{1-2\tau}}$
Beta ($\alpha = 2$, $\beta = 1$), $\tau > 1/2$	$2\sqrt{\frac{\tau}{2\tau-1}}\cos\left(\frac{1}{3}\arccos\left(-\sqrt{\frac{2\tau-1}{\tau}}\right)+\frac{4\pi}{3}\right)$
Beta ($\alpha = 2$, $\beta = 2$), $\tau \neq 1/2$	$\operatorname{sign}(1-2\tau)\frac{\sqrt{2\lambda_\tau+1}-\sqrt{2+2\operatorname{sign}(1-2\tau)\left(\sqrt{2\lambda_\tau+1}-2\frac{\lambda_\tau-\frac{\tau}{1-2\tau}}{\sqrt{2\lambda_\tau+1}}\right)-2\lambda_\tau}}{2}+\frac{1}{2}$ where $\lambda_\tau = \sqrt[3]{\frac{\tau(1-\tau)}{2(1-2\tau)^2}}$
Beta ($\alpha = 3$, $\beta = 1$), $\tau \neq 1/2$	$\operatorname{sign}(2\tau-1)\frac{\sqrt{2\lambda_\tau}-\sqrt{\frac{8\tau}{\lvert 2\tau-1\rvert\sqrt{2\lambda_\tau}}-2\lambda_\tau}}{2}$ where $\lambda_\tau = \sqrt[3]{\frac{\tau^2+\tau\sqrt{\tau(1-\tau)}}{(1-2\tau)^2}}+\sqrt[3]{\frac{\tau^2-\tau\sqrt{\tau(1-\tau)}}{(1-2\tau)^2}}$

Table C5: Closed-form expressions for expectiles in exceptional cases of certain bounded continuous distributions.

Distribution (parameters)	ξ_τ
Dagum ($\alpha = 2$, $\beta = 1/2$), $\tau \neq 1/3$	$\frac{\operatorname{sign}(1-2\tau)\sqrt{\tau^2(1-\tau)^2-\tau(1-4\tau+3\tau^2)(3\tau-2)}-\tau(1-\tau)}{1-4\tau+3\tau^2}$
Fisher ($\nu_1 = \nu_2 = 4$), $\tau \geq 1/2$	$2\sqrt{\frac{\tau}{1-\tau}}\cos\left(\frac{1}{3}\arccos\left(\sqrt{\frac{1-\tau}{\tau}}\right)\right)$
Fisher ($\nu_1 = \nu_2 = 4$), $\tau \leq 1/2$	$\sqrt[3]{\frac{\tau}{1-\tau}}\left(\sqrt[3]{1+\sqrt{\frac{1-2\tau}{1-\tau}}}+\sqrt[3]{1-\sqrt{\frac{1-2\tau}{1-\tau}}}\right)$
Fisher ($\nu_1 = 6$, $\nu_2 = 4$)	$\frac{\sqrt{2\lambda_\tau+\frac{8}{3}\frac{\tau}{1-\tau}}+\sqrt{-2\lambda_\tau+\frac{8}{3}\frac{\tau}{1-\tau}+\frac{128\tau}{27(1-\tau)\sqrt{2\lambda_\tau+\frac{8}{3}\frac{\tau}{1-\tau}}}}}{2}$, where $\lambda_\tau = \sqrt[3]{\frac{64}{729}\frac{\tau^2}{(1-\tau)^3}\left(1-2\tau+\lvert 2\tau-1\rvert\sqrt{\frac{1-\tau}{\tau}}\right)}$ $+\sqrt[3]{\frac{64}{729}\frac{\tau^2}{(1-\tau)^3}\left(1-2\tau-\lvert 2\tau-1\rvert\sqrt{\frac{1-\tau}{\tau}}\right)}-\frac{4}{9}\frac{\tau}{1-\tau}$
Fisher ($\nu_1 = 4$, $\nu_2 = 6$), $\tau \neq 1/3$	$\frac{\operatorname{sign}(3\tau-1)\sqrt{2\lambda_\tau+\frac{9}{4}}+\sqrt{\frac{9}{2}-2\lambda_\tau-\operatorname{sign}(3\tau-1)\left(3\sqrt{2\lambda_\tau+\frac{9}{4}}-\frac{27\frac{\tau}{1-\tau}+12\lambda_\tau}{2\sqrt{2\lambda_\tau+\frac{9}{4}}}\right)}}{2}-\frac{3}{4}$, where $\lambda_\tau = \frac{9}{4}\sqrt[3]{\frac{\tau(2\tau-1)}{2(1-\tau)^2}}$
Inverse-Γ ($\alpha = 2$, $\lambda > 0$)	$\frac{\lambda}{1+W\left(\frac{1-2\tau}{\tau}\exp(-1)\right)}$
Student ($\nu = 2$)	$\frac{2\tau-1}{\sqrt{2\tau(1-\tau)}}$
Student ($\nu = 4$)	$\operatorname{sign}(2\tau-1)\sqrt{\frac{1}{\sqrt{\tau(1-\tau)}}-2}$
Student ($\nu = 6$)	$\operatorname{sign}(2\tau-1)\sqrt{\frac{\sqrt{2\lambda_\tau-27}+\sqrt{54-2\lambda_\tau-\frac{(72\lambda_\tau-972)\tau^2+(972-72\lambda_\tau)\tau-27}{2\tau(1-\tau)\sqrt{2\lambda_\tau-27}}-18\sqrt{2\lambda_\tau-27}}}{2}-\frac{9}{2}}$, where $\lambda_\tau = \sqrt[3]{\frac{729}{128}\frac{(2\tau-1)^4}{\tau^2(1-\tau)^2}}-\sqrt[3]{\frac{729}{32}\frac{(2\tau-1)^2}{\tau(1-\tau)}}+18$
Chi-squared ($k = 4$)	$2W\left(\begin{smallmatrix}2\\-2\end{smallmatrix}, \frac{2\tau-1}{1-\tau}\right)$
Exponential ($\lambda > 0$)	$\frac{1}{\lambda}\left[1+W\left(\frac{2\tau-1}{1-\tau}\exp(-1)\right)\right]$
Laplace	$\operatorname{sign}(2\tau-1)W\left(\frac{\lvert 2\tau-1\rvert}{2\min\{\tau,1-\tau\}}\right)$

Table C6: Analytic-form expressions for expectiles in exceptional cases of certain unbounded continuous distributions. Here and throughout W denotes the principal branch of the Lambert function and $W\left(\begin{smallmatrix}\cdot\\\cdot\end{smallmatrix}, \cdot\right)$ is the generalized Lambert W function of Mező and Baricz (2017).

Distribution (parameters)	ξ_τ
Hall-Weiss ($\alpha = 2$, $\beta = 1$) $\tau \in \big(0, \frac{1}{2}\big] \cup \big[\frac{375}{407}, 1\big)$	$\sqrt{\frac{25-\tau}{36(1-\tau)}} \cos\left(\frac{1}{3} \arccos\left(\frac{125-593\tau}{\tau-25}\sqrt{\frac{1-\tau}{25-\tau}}\right)\right) + \frac{7}{12}$
Hall-Weiss ($\alpha = 2$, $\beta = 1$) $\tau \in \big[\frac{1}{2}, \frac{375}{407}\big)$	$\sqrt[3]{\frac{593\tau-125}{1728(1-\tau)} + \sqrt{\frac{\tau(1-2\tau)(407\tau-375)}{6912(1-\tau)^3}}} + \sqrt[3]{\frac{593\tau-125}{1728(1-\tau)} - \sqrt{\frac{\tau(1-2\tau)(407\tau-375)}{6912(1-\tau)^3}}} + \frac{7}{12}$
Hall-Weiss ($\alpha = 3/2$, $\beta = 1/2$) $\tau \leq 1/2$	$\frac{\left(\sqrt{2\lambda_\tau+\frac{5}{2}}+\sqrt{\frac{5}{2}-2\lambda_\tau+\frac{2}{\sqrt{2\lambda_\tau+\frac{5}{2}}}\frac{2\tau-1}{1-\tau}}\right)^2}{4}$ where $\lambda_\tau = \frac{1}{6}\sqrt{\frac{49-73\tau}{1-\tau}} \cos\left(\frac{1}{3} \arccos\left(\frac{-1027\tau^2+1262\tau-343}{(73\tau-49)(1-\tau)}\sqrt{\frac{1-\tau}{49-73\tau}}\right)\right) - \frac{5}{12}$
Hall-Weiss ($\alpha = 3/2$, $\beta = 1/2$) $\tau \geq 1/2$	$\frac{\left(\sqrt{2\lambda_\tau+\frac{5}{2}}+\sqrt{\frac{5}{2}-2\lambda_\tau+\frac{2}{\sqrt{2\lambda_\tau+\frac{5}{2}}}\frac{2\tau-1}{1-\tau}}\right)^2}{4}$ where $\lambda_\tau = \sqrt[3]{\frac{\frac{1027\tau^2-1262\tau+343}{864}+\sqrt{\tau\frac{3082\tau^3-6573\tau^2+4574\tau-1029}{3456}}}{2(1-\tau)^2}}$ $+\sqrt[3]{\frac{\frac{1027\tau^2-1262\tau+343}{864}-\sqrt{\tau\frac{3082\tau^3-6573\tau^2+4574\tau-1029}{3456}}}{2(1-\tau)^2}} - \frac{5}{12}$
Hall-Weiss ($\alpha = 3$, $\beta = 1$)	$\frac{\sqrt{2\lambda_\tau+\frac{289}{576}}+\sqrt{-2\lambda_\tau-\frac{\frac{1-2\tau}{2(1-\tau)}+\frac{17}{6}\lambda_\tau}{\sqrt{2\lambda_\tau+\frac{289}{576}}}+\frac{17}{12}\sqrt{2\lambda_\tau+\frac{289}{576}}+\frac{289}{288}}}{2} + \frac{17}{48}$ where $\lambda_\tau = \frac{1}{24}\sqrt[3]{\frac{(2\tau-1)(397\tau-343)+\lvert 2\tau-1\rvert\sqrt{\tau(80605-77689\tau)}}{(1-\tau)^2}}$ $+\frac{1}{24}\sqrt[3]{\frac{(2\tau-1)(397\tau-343)-\lvert 2\tau-1\rvert\sqrt{\tau(80605-77689\tau)}}{(1-\tau)^2}}$
Hall-Weiss ($\alpha = 2$, $\beta = 2$)	$\frac{\sqrt{2\lambda_\tau+\frac{2\tau-1}{2(1-\tau)}+\frac{25}{36}}+\sqrt{-2\lambda_\tau+\frac{2\tau-1}{2(1-\tau)}-\frac{10}{3}\frac{\lambda_\tau}{\sqrt{2\lambda_\tau+\frac{2\tau-1}{2(1-\tau)}+\frac{25}{36}}}+\frac{5}{3}\sqrt{2\lambda_\tau+\frac{2\tau-1}{2(1-\tau)}+\frac{25}{36}}+\frac{25}{18}}}{2} + \frac{5}{12}$ where $\lambda_\tau = \frac{\sqrt[3]{(1-2\tau)(6\tau^2-32\tau+27)+2\lvert 2\tau-1\rvert\sqrt{\tau(540-1553\tau+1504\tau^2-491\tau^3)}}}{12(1-\tau)}$ $+\frac{\sqrt[3]{(1-2\tau)(6\tau^2-32\tau+27)-2\lvert 2\tau-1\rvert\sqrt{\tau(540-1553\tau+1504\tau^2-491\tau^3)}}}{12(1-\tau)} + \frac{1-2\tau}{12(1-\tau)}$

Table C7: Closed-form expressions for expectiles in exceptional cases of the Hall-Weiss distribution with parameters $\alpha > 0$ and $\beta \geq 0$, having distribution function $F(x) = 1 - (x^{-\alpha} + x^{-\alpha-\beta})/2$, $x > 1$.

Distribution (parameters)	ξ_τ
Pareto ($\gamma = 4/5$), $\tau = 2/3$	$\frac{1}{81}\left[2 - \frac{5^{2/3}}{\sqrt[3]{7+3\sqrt{6}}} + \sqrt[3]{5(7+3\sqrt{6})}\right]^4$
Pareto ($\gamma = 3/4$)	$\frac{\left(\sqrt{2\lambda_\tau}+\sqrt{\frac{8}{\sqrt{2\lambda_\tau}}-2\lambda_\tau}\right)^3}{8}$ where $\lambda_\tau = \sqrt[3]{1+\sqrt{\frac{\tau(7\tau^2-9\tau+3)}{(1-\tau)^3}}} + \sqrt[3]{1-\sqrt{\frac{\tau(7\tau^2-9\tau+3)}{(1-\tau)^3}}}$
Pareto ($\gamma = 2/3$), $\tau \geq 2/3$	$\left(\sqrt[3]{\frac{2\tau-1+\sqrt{\tau(3\tau-2)}}{1-\tau}} + \sqrt[3]{\frac{2\tau-1-\sqrt{\tau(3\tau-2)}}{1-\tau}}\right)^2$
Pareto ($\gamma = 2/3$), $\tau \leq 2/3$	$4\cos\left(\frac{1}{3}\arccos\left(\frac{2\tau-1}{1-\tau}\right)\right)^2$
Pareto ($\gamma = 1/2$)	$1 + \sqrt{\frac{\tau}{1-\tau}}$
Pareto ($\gamma = 1/3$), $\tau \geq 1/2$	$\sqrt[3]{\frac{3\tau-1+2\sqrt{\tau(2\tau-1)}}{8(1-\tau)}} + \sqrt[3]{\frac{3\tau-1-2\sqrt{\tau(2\tau-1)}}{8(1-\tau)}} + \frac{1}{2}$
Pareto ($\gamma = 1/3$), $\tau \leq 1/2$	$\cos\left(\frac{1}{3}\arccos\left(\frac{3\tau-1}{1-\tau}\right)\right) + \frac{1}{2}$
Pareto ($\gamma = 1/4$)	$\frac{1}{3} + \frac{\sqrt{2\lambda_\tau+\frac{4}{9}}+\sqrt{-2\lambda_\tau-\frac{8}{3}\frac{\lambda_\tau}{\sqrt{2\lambda_\tau+\frac{4}{9}}}+\frac{4}{3}\sqrt{2\lambda_\tau+\frac{4}{9}}+\frac{8}{9}}}{2}$ where $\lambda_\tau = \sqrt[3]{\frac{1-2\tau+\lvert 1-2\tau\rvert\sqrt{\frac{\tau}{1-\tau}}}{27(1-\tau)}} + \sqrt[3]{\frac{1-2\tau-\lvert 1-2\tau\rvert\sqrt{\frac{\tau}{1-\tau}}}{27(1-\tau)}}$

Table C8: Closed-form expressions for expectiles in exceptional cases of the Pareto distribution with extreme value index $\gamma > 0$, having distribution function $F(x) = 1 - x^{-1/\gamma}$, $x > 1$. Other closed-form expressions directly deduced from these cases are those of expectiles for the generalized Pareto distribution (which is a location-scale variant of the Pareto distribution) with the same extreme value indices.