On the Asymptotic Normality between Sample Quantiles and Dispersion Estimators

Marcel Bräutigam [1] & Marie Kratz [2]

[1] *ESSEC Business School Paris (CREAR) & Sorbonne University (LPSM) & LabEx MME-DII, brautigam@essec.edu*

[2] *ESSEC Business School Paris (CREAR), kratz@essec.edu*

Abstract. In this study, we derive the joint asymptotic distributions of functionals of sample quantiles and functionals of measure of dispersion estimators (the sample variance and the sample mean absolute deviation).

Keywords. Asymptotic distribution; sample quantile; measure of dispersion; non-linear dependence; correlation

1 Introduction

The joint asymptotic distribution between a measure of location estimator and a sample quantile, for an identically and independently distributed (iid) sample, has been considered in the literature in two cases, when the location measure is chosen as the sample mean and as the sample median, respectively. While this latter case is directly deduced from the well-known asymptotics of a vector of sample quantiles (the sample median being a sample quantile itself), the joint asymptotics between the sample mean and sample quantile were treated by Lin et al. (1980) and later, using another approach, by Ferguson (1999). These results have then been used by Bera et al. (2016) to introduce a new characterization and hence also test for the normal distribution.

Here we move from measures of location to measures of dispersion and present in the main result (Theorem 1) joint asymptotics for functionals of sample dispersion measures with functionals of the sample quantile. By measures of dispersion we mean well-known quantities as the variance or standard deviation, but also less frequently used ones as, for example, the mean absolute deviation (denoted MAD). The latter can relax the asymptotic constraints that come with the use of the sample variance (such as the existence of the fourth moment of the underlying distribution).

Such joint asymptotics have not been yet considered in generality in the literature. Only a few specific examples exist, see DasGupta and Haff (2006), Bos and Janus (2013). These results can be seen as special cases of Theorem 1.

The motivation for this study comes from previous work in financial risk management, see Bräutigam et al. (2019) and Zumbach (2012, 2018). A further application of the results presented concerns the risk measure estimation developed in Bräutigam and Kratz (2018): The sample quantile can be seen as a Value-at-Risk (VaR) estimator and the functional framework allows us to extend the results to Expected Shortfall (ES).

This note is part of a more general treatment and analysis of functionals of quantile estimators (sample quantile and location-scale quantile estimator) and functionals of measure of dispersion estimators (including apart from the sample variance, sample MAD also the sample median absolute deviation around the sample median) that can be found in the working paper by Bräutigam and Kratz (2018). The structure is to first present the main result in Theorem 1, namely the joint bivariate asymptotic normality of functionals of the sample quantile with either functionals of the sample variance or the sample MAD. This result is then illustrated with an example in the next section.

2 Main Result

2.1 Notation

Let $(X_1, \cdots, X_n)$ be a sample of size n, with parent random variable (rv) X, parent cumulative distribution function (cdf) F_X, (and, given they exist,) probability density function (pdf) f_X, mean μ, variance σ^2, and quantile of order p defined as $q_X(p) := \inf\{x \in \mathbb{R} : F_X(x) \geq p\}$. We denote its ordered sample by $X_{(1)} \leq ... \leq X_{(n)}$. In the special case of the standard normal distribution $\mathcal{N}(0,1)$, we use the standard notation $\Phi, \phi, \Phi^{-1}(p)$ for the cdf, pdf and quantile of order p, respectively.

In this paper, we focus on the following three estimators. First, we consider two estimators of the dispersion: the sample variance $\hat{\sigma}_n^2$, and the sample mean absolute deviation around the sample mean (MAD) $\hat{\theta}_n$, respectively. We introduce a unified notation:

$$D_i = \begin{cases} \sigma^2 & \text{for } i = 1, \\ \theta & \text{for } i = 2, \end{cases} \quad \text{and estimators} \quad \hat{D}_{i,n} = \begin{cases} \hat{\sigma}_n^2 := \dfrac{1}{n-1} \displaystyle\sum_{j=1}^{n} (X_j - \bar{X}_n)^2, & \text{for } i = 1, \\ \hat{\theta}_n := \dfrac{1}{n} \displaystyle\sum_{j=1}^{n} |X_j - \bar{X}_n|, & \text{for } i = 2, \end{cases}$$

where $\bar{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$. As third estimator we consider the sample quantile q_n, being defined as

$$q_n(p) = X_{(\lceil np \rceil)},$$

denoting by $\lceil x \rceil = \min\{m \in \mathbb{Z} : m \geq x\}$ the rounded-up integer-parts of a real number $x \in \mathbb{R}$.

In addition, to be consistent in the notation with related results in the literature, we generalise a notation used in Ferguson (1999), then in Bera (2016): Assuming that the underlying rv X has finite moments up to order l, and that η is a continuous real-valued function, we set, for $1 \leq k \leq l$ and $p \in (0,1)$,

$$\tau_k(\eta(X), p) = (1-p)\left(\mathbb{E}[\eta^k(X)|X > q_X(p)] - \mathbb{E}[\eta^k(X)]\right). \tag{1}$$

When η is the identity function, we abbreviate $\tau_k(X,p)$ as $\tau_k(p)$. Finally, the signum function is denoted by sgn and defined, as usual, by $\operatorname{sgn}(x) := \begin{cases} -1 & \text{if } x < 0, \\ 0 & \text{if } x = 0, \\ 1 & \text{if } x > 0. \end{cases}$

The standard notations $\overset{d}{\to}$ and $\overset{P}{\to}$ correspond to the convergence in distribution and in probability, respectively. Further, for a sequence of random variables X_n and constants a_n, we denote by $X_n = o_P(a_n)$ the convergence to zero in probability of X_n/a_n.

2.2 Theorem

Our main result gives the joint bivariate asymptotics of functionals of the sample quantile with functionals of either the sample variance or sample MAD. We consider functionals h_1, h_2 of the estimators that we assume to be continuous real-valued functions with existing derivatives denoted by h_1' and h_2' respectively.

Theorem 1 *Consider an iid sample with parent rv X having mean μ and variance σ^2. Assume that F_X is differentiable at $q_X(p)$ and $f_X(q_X(p)) > 0$, that $\mathbb{E}[X^{2r}] < \infty$ for $r = 1, 2$ respectively as well as $f_X(\mu) > 0$ for $r = 1$. Then the joint behaviour of the functionals h_1 of the sample quantile $q_n(p)$, for $p \in (0,1)$, and h_2 of the sample measure of dispersion $\hat{D}_{r,n}$, is asymptotically normal:*

$$\sqrt{n}\begin{pmatrix} h_1(q_n(p)) - h_1(q_X(p)) \\ h_2(\hat{D}_{r,n}) - h_2(D_r) \end{pmatrix} \xrightarrow[n\to\infty]{d} \mathcal{N}(0, \Sigma^{(r)}),$$

where the asymptotic covariance matrix $\Sigma^{(r)} = (\Sigma_{ij}^{(r)}, 1 \leq i, j \leq 2)$ satisfies

$$\Sigma_{11}^{(r)} = \frac{p(1-p)}{f_X^2(q_X(p))}\,(h_1'(q_X(p)))^2; \quad \Sigma_{22}^{(r)} = (h_2'(D_r))^2 \operatorname{Var}\left(|X-\mu|^r + (2-r)(2F_X(\mu)-1)X\right);$$

$$\Sigma_{12}^{(r)} = \Sigma_{21}^{(r)} = h_1'(q_X(p))\,h_2'(D_r) \times \frac{\tau_r(|X-\mu|, p) + (2-r)(2F_X(\mu)-1)\tau_1(p)}{f_X(q_X(p))},$$

τ_r being defined in (1).

The asymptotic correlation between the functional h_1 of the sample quantile and the functional h_2 of the measure of dispersion is - up to its sign $a_\pm = \operatorname{sgn}(h_1'(q_X(p)) \times h_2'(D_r))$ -

the same whatever the choice of h_1, h_2:

$$\lim_{n\to\infty} \mathrm{Cor}\left(h_1(q_n(p)), h_2(\hat{D}_{r,n})\right) = a_\pm \times \frac{\tau_r(|X-\mu|,p) + (2-r)(2F_X(\mu)-1)\tau_1(p)}{\sqrt{p(1-p)\,\mathrm{Var}\left(|X-\mu|^r + (2-r)(2F_X(\mu)-1)X\right)}}.$$

Note that, if F_X belongs to the class of location-scale distributions, then the asymptotic correlation becomes independent of the mean μ and variance σ^2 (see Bräutigam and Kratz (2018)). Further, corresponding results with the sample median absolute deviation around the sample median (MedianAD), and the location-scale quantile as alternative quantile estimator can also be found in Bräutigam and Kratz (2018).

2.3 Outline of the proof

The main approach in proving Theorem 1 relies on the Bahadur representation of the sample quantile (first proved in Bahadur (1966)). Here we use the version of Ghosh (1971) where, assuming F_X is differentiable at $q_X(p))$ and $f_X(q_X(p)) > 0$, it holds

$$q_n(p) = q_X(p) + \frac{1 - F_n(q_X(p)) - (1-p)}{f_X(q_X(p))} + R_{n,p},$$

with $R_{n,p} = o_P(n^{-1/2})$. Using this representation and the bivariate central limit theorem (CLT) with the sample variance gives the result in the case $r = 2$. For the case $r = 1$, i.e. the sample MAD, we have that under $f_X(\mu) > 0$ (see e.g. Babu and Rao (1992), Segers (2014))

$$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n}|X_i - \mu| + (2F_X(\mu)-1)(\bar{X}_n - \mu) + S_{n,p},$$

with $S_{n,p} = o_P(n^{-1/2})$. Again, the asymptotics in the case of the sample MAD and sample quantile follows from the use of this representation and the bivariate CLT.

3 An Example

To illustrate Theorem 1 in a specific case, we consider a normal sample (with mean μ and variance σ^2) and provide the asymptotic correlation of the sample quantile with either the sample variance $\hat{\sigma}_n^2$ or sample MAD $\hat{\theta}_n$. We obtain:

$$\lim_{n\to\infty} \mathrm{Cor}(q_n(p), \hat{\sigma}_n^2) = \frac{\phi(\Phi^{-1}(p))\Phi^{-1}(p)}{\sqrt{2p(1-p)}}$$

$$\text{and} \quad \lim_{n\to\infty} \mathrm{Cor}(q_n(p), \hat{\theta}_n) = \frac{\phi(\Phi^{-1}(p)) - (1-p)\sqrt{2/\pi}}{\sqrt{p(1-p)}\,\sqrt{1-2/\pi}}.$$

These two asymptotic correlations, as a function of the order of the quantile, $p \in (0,1)$, are plotted in Figure 1. We see that these correlations are point-symmetric around $p = 0.5$, taking the value 0 at $p = 0.5$ and tending to 0 for $p \to 0$ and 1. Both correlations, with the sample variance and sample MAD respectively, have a similar range. The maximum correlation value for both is further in the tails. At the same time we observe that the correlation with the sample MAD is higher for intermediate values, than with the sample variance, whereas it is lower for values in the tail. This is a behaviour one could expect since the MAD is a robust measure of dispersion with tail values that do not have such an influence on the measure of dispersion as for the variance.

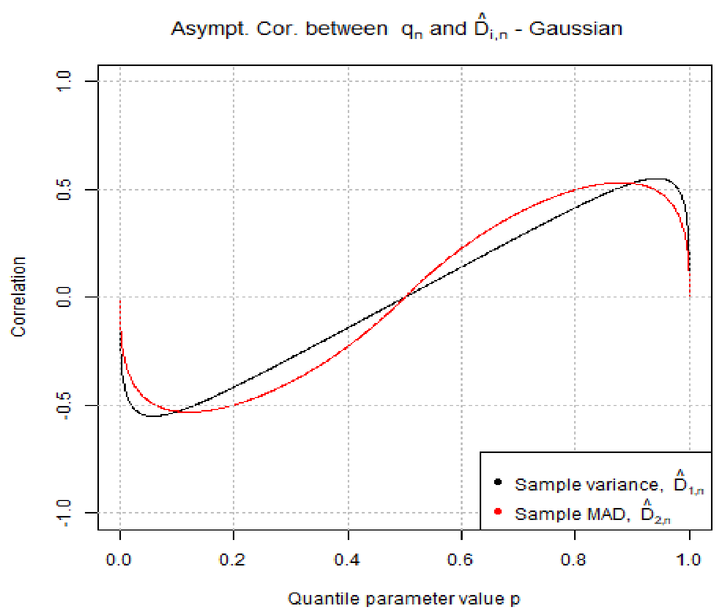

Figure 1: Asymptotic correlation between the sample quantile and either the sample variance (in black) $\hat{D}_{1,n} = \hat{\sigma}_n^2$ or the sample MAD (in red) $\hat{D}_{2,n} = \hat{\theta}_n$.

A simulation study showing the good finite sample approximation of these asymptotics has been performed in Bräutigam and Kratz (2018). Further examples and covariances results (given also in the case of using the location-scale quantile estimator or the sample MedianAD as measure of dispersion) can be found in the same reference.

Bibliography

Babu, G. and Rao, C. (1992), Expansions for statistics involving the mean absolute deviations, *Annals of the Institute of Statistical Mathematics* 44(2), 387-403

Bahadur, R. (1966), A note on quantiles in large samples, *The Annals of Mathematical Statistics* 37(3), 577-580

Bera, A. and Galvao, A. and Wang, L. and Xiao, Z. (2016), A new characterization of the normal distribution and test for normality., *Econometric Theory* 32(5), 1216–1252

Bos, C. and Janus, P. (2013), A Quantile-based Realized Measure of Variation: New Tests for Outlying Observations in Financial Data, *Tinbergen Institute Discussion Paper 13-155/III*

Bräutigam, M. and Dacorogna, M. and Kratz, M. (2019), Pro-Cyclicality of Traditional Risk Measurements: Quantifying and Highlighting Factors at its Source, *arXiv-1903.03969*

Bräutigam, M. and Kratz, M. (2018), On the Dependence between Quantiles and Dispersion Estimators, *ESSEC Working Paper*, WP 1807

DasGupta, A. and Haff, L. (2006), Asymptotic values and expansions for the correlation between different measures of spread, *Journal of Statistical Planning and Inference* 136(7), 2197-2212

Ferguson, T. (1999), Asymptotic joint distribution of sample mean and a sample quantile, *preprint* *http://www.math.ucla.edu/~tom/papers/unpublished/meanmed.pdf*

Ghosh, J. (1971), A new proof of the bahadur representation of quantiles and an application, *The Annals of Mathematical Statistics*, 1957-1961

Lin, P.-E. and Wu, K.-T. and Ahmad, I. (1980), Asymptotic joint distribution of sample quantiles and sample mean with applications, *Communications in Statistics-Theory and Methods* 9(1), 51–60

Segers, J. (2014), On the asymptotic distribution of the mean absolute deviation about the mean, *arXiv:1406.4151*

Zumbach, G. (2012), Discrete Time Series, Processes, and Applications in Finance, Springer Science & Business Media

Zumbach, G. (2018), Correlations of the realized volatilities with the centered volatility increment, http://www.finanscopics.com/figuresPage.php?figCode=corr_vol_r_VsDV0, accessed 1-October-2018