LINKING CHANGES IN MOMENTS OF A DISTRIBUTION WITH CHANGES IN ITS PERCENTILES

Fabrizio Falasca
Courant Institute of Mathematical Sciences
New York University, New York, NY, USA

Andrew Brettin
Courant Institute of Mathematical Sciences
New York University, New York, NY, USA

Laure Zanna
Courant Institute of Mathematical Sciences
New York University, New York, NY, USA

March 15, 2022

1 Abstract

We present an analytical relation between changes in moments of a distribution and changes in its percentiles. The result holds for distributions that show no large deviation from normality. The set of bases derived is shown to form an orthogonal set and found to be useful as a dimensionality reduction method in quantile regression estimations.

2 Introduction

A probability distribution function (PDF) can be characterized by its statistical moments. While theoretically all moments are necessary to describe a PDF, it is often enough to focus on the first four. These statistics are the mean, the variance, the skewness and kurtosis. The first two represent the expected value and the variance around it. The skewness quantifies the asymmetry of the distribution and the kurtosis the "fatness" of its tails.
On the other hand, a PDF can also be characterized by its quantiles. A quantile $q \in [0, 1]$ is defined as the value of the PDF under which $p = q \cdot 100\%$ of the values lie.

Given a PDF $f(x)$, we define its n^{th} moment as $\int_{-\infty}^{\infty} x^n f(x)dx$. The quantile of a PDF can be instead identified through its quantile function. Given a cumulative distribution function $F_X(x)$, a quantile function returns the value of x such that $F_X(x) = P(X \leq x) = p$. A way to write the quantile function is simply $Q(p) = \inf \{x \in \mathbb{R} : p \leq F_X(x)\}$.

Is there a relationship between moments and quantiles of a distribution? Yes. And as it turns out, people in the insurance sector or portfolio optimization know about it [1, 2]. The main study has been carried out by Cornish and Fisher in two papers dating back to 1937 and 1960 [3, 4]. In those works the authors showed that it is possible to express any percentile point of a distribution as an infinite expansion of its cumulants. Cumulants are linked to the moments of a distributions and details can be found in several sources, one of them: `http://www.scholarpedia.org/article/Cumulants`. A byproduct of their 1937 paper [3] is a truncated formula providing an approximation of values of quantiles y_q of a distribution given its first four moments. In these notes we will refer to such approximation as the Cornish-fisher Expansion.

Here, we are interested in defining suitable basis functions capturing how values at each quantile of a distribution change when changing its moments one at a time. The motivation comes from studying changes in distribution in time series data (sea level time series in our case). For a given time series we first quantify linear trends in its quantiles [5, 6]; this allows to study changes in the *full* distribution. Then we use the basis functions derived here to condense the N

quantile trends (slopes) onto 4 bases, quantifying changes in quantiles implied by changes in the first four statistical moments.

3 Cornish-Fisher (CF) Expansion (1937)

Given a random variable X with mean μ, variance σ^2, skewness γ and excess kurtosis κ, the following approximate relation holds:

$$y_q \sim \mu + \sigma w; \quad w = z_q + (z_q^2 - 1)\frac{\gamma}{6} + (z_q^3 - 3z_q)\frac{\kappa}{24} - (2z_q^3 - 5z_q)\frac{\gamma^2}{36}; \tag{1}$$

where $z_q = \mathrm{Q}(\mathrm{N}(0,1), q)$ is the value at quantile $q \in [0,1]$ assumed by a Standard Normal $\mathrm{N}(0,1)$.

The expansion is a simple polynomial function of the quantiles of a standard normal. It allows to approximate the quantiles of a general distribution by knowing its first four moments. By considering the third and fourth moments (i.e.,skewness and kurtosis), it allows to quantify deviations from normality.

Note: keep in mind that generally, γ and κ are not really the the skewness and kurtosis ($\hat{\gamma}$ and $\hat{\kappa}$) of the distribution but parameters dependent on those real values (see [1] and [2]). However, for small skewness and kurtosis (small deviations from gaussianity) it holds $\gamma \sim \hat{\gamma}$ and $\kappa \sim \hat{\kappa}$. For analytical results, it is then useful to leave γ and κ in the formula. This is especially true in our case as we will define infinitesimal changes in the neighborhood of a standard normal distribution.

3.1 Testing the CF Expansion

We test the formula using a Beta distribution with known first 4 moments: $\mu = 0.143$, $\sigma^2 = 0.0082$, $\gamma = 0.988$ and $\kappa = 4.026$. Such distribution is shown in 1. The quantile function of a Beta distribution is known analytically. Therefore we can compute the correspondent value at each quantile.

- The value of quantile $q = 0.95$ (95^{th} percentile) is 0.31634.
- The CF expansion provides a value of any quantile as a function of the first four moments. In this case: $y_{q=0.95}(\mu = 0.143, \sigma^2 = 0.0082, \gamma = 0.988, \kappa = 4.026) \sim 0.313324$.

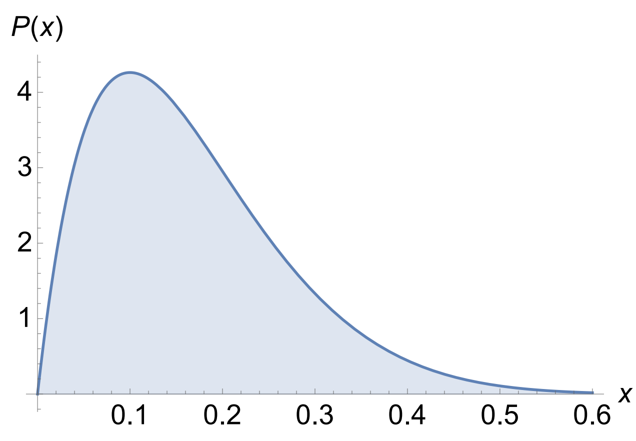

Figure 1: Beta distribution with $\mu = 0.143$, $\sigma^2 = 0.0082$, $\gamma = 0.988$ and $\kappa = 4.026$

4 Changes in quantiles driven by changes in moments

Goal. We aim in finding a set of basis functions capturing how each quantile of a distribution change when changing its moments one at a time. On the operational side, this is useful when estimating changes in distributions from time series.

Specifically, we are going to quantify (linear) trends in percentiles through quantile regression. Quantile slopes can then be summarized through the basis functions derived here.

We restrict ourselves to the simple case of focusing on linear changes in quantiles. We first estimate infinitesimal changes in y_p when changing the moments of the distribution. Then, we evaluate such changes as small deviation from normality; in other words, we estimate results in the neighborhood of ($\mu = 0$, $\sigma^2 = 1$, $\gamma = 0$, $\kappa = 0$).

Changes in y_q when changing the moments are defined as follows:

$$\begin{cases} \frac{\partial y_q}{\partial \mu} & = 1 \\ \frac{\partial y_q}{\partial (\sigma^2)} & = \frac{1}{2\sigma}(z_q + \frac{1}{6}(z_q^2 - 1)\gamma - \frac{1}{36}(2z_q^3 - 5z_q)\gamma^2 + \frac{1}{24}(z_q^3 - 3z_q)\kappa) \\ \frac{\partial y_q}{\partial \gamma} & = \sigma(\frac{1}{6}(z_q^2 - 1) - \frac{1}{18}(2z_q^3 - 5z_q)\gamma) \\ \frac{\partial y_q}{\partial \kappa} & = \frac{\sigma}{24}(z_q^3 - 3z_q) \end{cases} \tag{2}$$

Derivatives in the system of Eq. 2 define changes in distributional quantiles driven by infinitesimal changes in the first four moments. These values are defined in the neighborhood of a normal distribution and can therefore be evaluated in the point $x_* = (\mu = 0, \sigma^2 = 1, \gamma = 0, \kappa = 0)$.

$$\begin{cases} \frac{\partial y_q}{\partial \mu}|_{x_*} & = 1 \\ \frac{\partial y_q}{\partial (\sigma^2)}|_{x_*} & = \frac{z_q}{2} \\ \frac{\partial y_q}{\partial \gamma}|_{x_*} & = \frac{1}{6}(z_q^2 - 1) \\ \frac{\partial y_q}{\partial \kappa}|_{x_*} & = \frac{1}{24}(z_q^3 - 3z_q) \end{cases} \tag{3}$$

With $z_p = Q(N(0,1), p)$. Therefore, the set of basis functions is defined as $f_i(q) = \{f_1(q) = 1, f_2(q) = \frac{z_q}{2}, f_3(q) = \frac{1}{6}(z_q^2 - 1), f_4(q) = \frac{1}{24}(z_q^3 - 3z_q)\}$. These functions are defined in the interval $q \in [0, 1]$. And as it turns out...

$$\int_0^1 f_i(q) f_j(q) dp = 0; \text{ if } i \neq j. \tag{4}$$

So, this set of basis functions are actually orthogonal. This is not very surprising, as they are Hermite polynomials of the quantile function z_p; if we transform this to the real space (i.e., $x \in [-\infty, \infty]$) their orthogonality relationship becomes clear.

A plot of such basis is shown in Figure 2.

Now, let's consider a time series $x(t)$. We quantify linear changes in its statistics by computing linear trends in quantiles. The result will be N quantile slopes $s(q)$. We use the derived basis functions to condense these results onto 4 numbers, quantifying the changes in quantiles implied by the changes in the first disributional moments. In other words we find a functional form for the slopes $s(q)$ as:

$$s(q) = \sum_{i=1}^{4} a_i b_i(q); \tag{5}$$

and solve equation 5 by computing its least-squares solution.

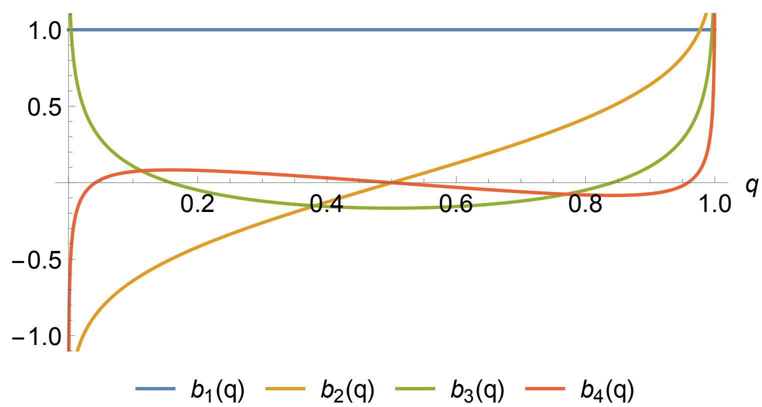

Figure 2: Basis functions capturing changes in quantiles of a distribution when changing its moments.

5 What are these bases quantifying?

What they are. The bases give us a way to condense N linear quantiles slopes onto four. Bases serve as a dimensionality reduction tool and quantify how linear changes in percentiles are implied by changes in moments. In other words, it is a way to quantify drivers of quantiles trends, i.e. from changes in one single moment or a combination of them.

What they are not. Apart from changes in the mean, regression coefficients of each basis should not be considered as the real coefficient quantifying drifts in the moments. This is not what we are after with this methodology, what we want is simply a way to identify which moment drives changes in quantile trends.

6 Synthetic datasets

We consider the following cases:

- Case (a): time-dependent Gaussian distribution;
- Case (b): time-dependent Beta distribution.

(a) Time-dependent Gaussian distribution. We define a time-dependent Gaussian process with mean $\mu = at$ and standard deviation $\sigma = (bt)^{1/2}$. With $a, b \in \mathbb{R}$ and t representing the time vector. The distribution then reads as $P(x, t; \mu = at, \sigma^2 = bt) = \frac{1}{\sqrt{2\pi bt}} e^{\frac{(x-at)^2}{2bt}}$.
Specifically we choose $a = b = 5$ and $t \in [1, 3]$ (to remove a small initial transient). At each time step $dt = 0.0001$, we randomly sample from the distribution, therefore defining a time series with drift in mean and variance. We apply the proposed framework to (a) estimate trends in quantiles and (b) projecting the slopes onto the 4 basis functions. Quantile regression is computed starting from $q = 0.01$ to $q = 0.99$ every $dq = 0.01$. Number of bootstrap samples used for the statistical significance is $n = 500$. Results are shown in Figure 3, panels (a) and (c).

(b) Time-dependent Beta distribution. We define a time-dependent process sampled from a drifting Beta distribution. A beta distribution writes as $P(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$ and is defined over the interval $[0, 1)$. It depends on two parameters, α and β, controlling the "fatness" of the tails of the distribution. A larger α (β) implies negative (positive) skewness; the distribution is symmetric if $\alpha = \beta$.

The mean, variance, skewness and kurtosis of the distribution are defined as $(\mu, \sigma^2, \gamma, \kappa) = (\frac{\alpha}{\alpha+\beta}, \frac{\alpha\beta}{(\alpha+\beta)^2(1+\alpha+\beta)}), \frac{2(\beta-\alpha)\sqrt{1+\alpha+\beta}}{\sqrt{\alpha}\sqrt{\beta}(2+\alpha+\beta)}, \frac{3(1+\alpha+\beta)(\alpha\beta(\alpha+\beta-6)+2(\alpha+\beta)^2}{\alpha\beta(2+\alpha+\beta)(3+\alpha+\beta)})$.
We define a "drifting" Beta distribution with $\alpha = at$ and $\beta = bt$. We choose $a = b = 10$, therefore defining a process with constant mean, zero skewness but time dependent variance and kurtosis. Therefore, in this case the moments are defined as: $(\mu, \sigma^2, \gamma, \kappa) = (\frac{1}{2}, \frac{1}{1+80t}, 0, 3 - \frac{6}{3+20t})$.

As in the Gaussian example, we evaluate for the time period $t \in [1, 3]$, with $dt = 0.0001$. Quantile regression is computed starting from $q = 0.01$ to $q = 0.99$ every $dq = 0.01$. Number of bootstrap samples used for the statistical significance is $n = 500$. Results are shown in Figure 3, panels (b) and (d).

In Figure 3 we show the results for the time-dependent process defined in the case of Gaussian and Beta distribution. The method allows for a correct reconstruction of quantile slopes in terms of basis functions and it allows to find the modes driving such changes.

References

[1] CO. Amédée-Manesme, F. Barthélémy, and D. Maillard. Computation of the corrected Cornish–Fisher expansion using the response surface methodology: application to VaR and CVaR. *Ann Oper Res*, 281:423–453, 2019.

[2] YDidier Maillard. A User's Guide to the Cornish Fisher Expansion. *hal-02987694*, 2020.

[3] E. A. Cornish and R. A. Fisher. Moments and Cumulants in the Specification of Distributions. *Revue De L'Institut International De Statistique / Review of the International Statistical Institute*, 5(4):307–320, 1937.

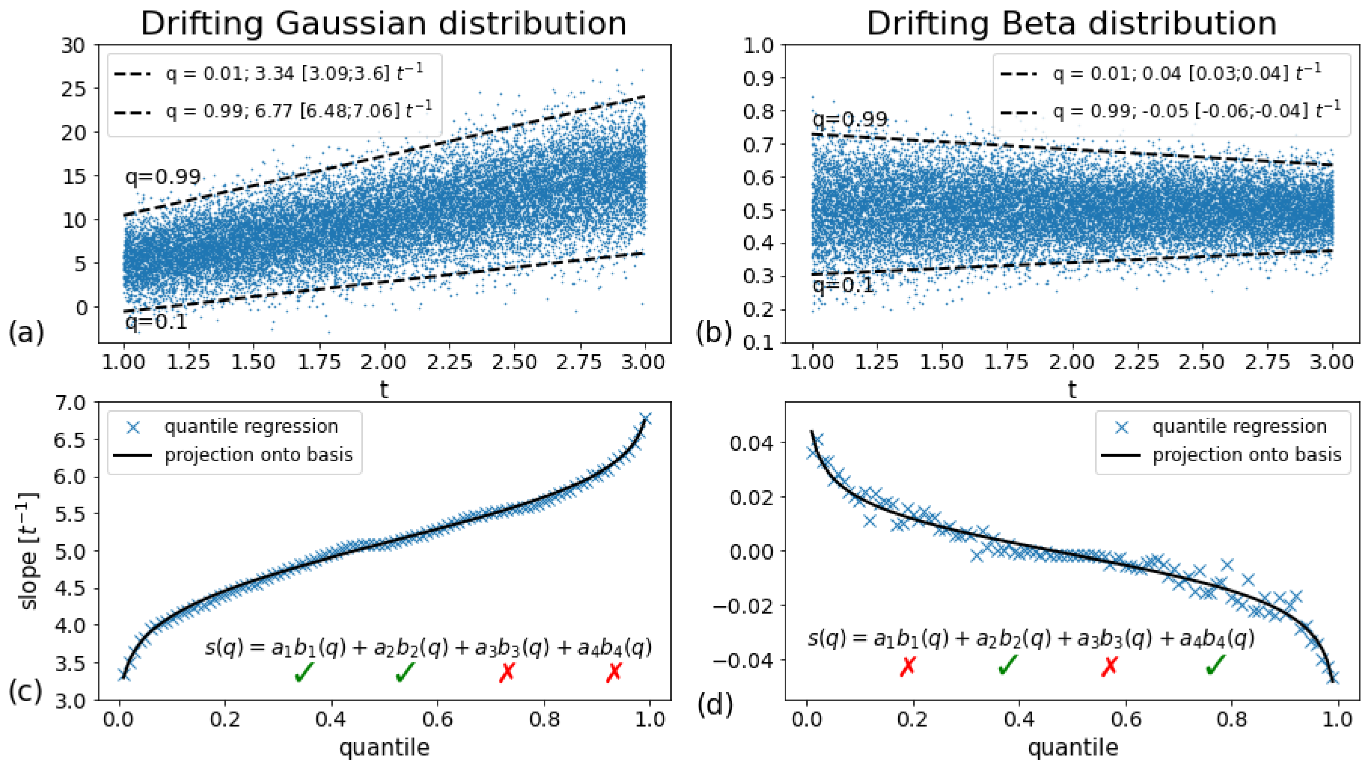

Figure 3: Panel (a,c): test for time series sampled from a time dependent Gaussian. (a) Significance trends in quantiles $q = 0.01$ and $q = 0.99$ are reported along with their 95% confidence bounds. (c) All quantile slopes along with their projection onto bases are reported; green "check" marks indicate when changes in moments are considered as statistically significant. Panel (b,d): same as panels (a) and (c) but for a time dependent Beta distribution.

[4] R. A. Fisher and E. A. Cornish. The Percentile Points of Distributions Having Known Cumulants. *Technometrics*, 2(2):209–225, 1960.

[5] R. Koenker and G. Jr. Bassett. Regression quantiles. *Econometrica*, 46:33–50, 1978.

[6] Roger Koenker and Kevin F Hallock. Quantile regression. *Journal of economic perspectives*, 15(4):143–156, 2001.