



HAL
open science

Bahadur representation of sample quantiles for functional of Gaussian dependent sequences under a minimal assumption

Jean-François Coeurjolly

► **To cite this version:**

Jean-François Coeurjolly. Bahadur representation of sample quantiles for functional of Gaussian dependent sequences under a minimal assumption. 2007. hal-00131071v1

HAL Id: hal-00131071

<https://hal.science/hal-00131071v1>

Preprint submitted on 15 Feb 2007 (v1), last revised 29 Sep 2008 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

BAHADUR REPRESENTATION OF SAMPLE QUANTILES FOR
FUNCTIONAL OF GAUSSIAN DEPENDENT SEQUENCES UNDER A
MINIMAL ASSUMPTION

BY JEAN-FRANÇOIS COEURJOLLY¹
SAGAG, Department of Statistics, Grenoble, FRANCE

February 15, 2007

We obtain a Bahadur representation for sample quantiles of nonlinear functional of Gaussian sequences with correlation function decreasing as $k^{-\alpha}$ for some $\alpha > 0$. This representation is derived under a minimal assumption.

1 Introduction

We consider the problem of obtaining a Bahadur representation of sample quantiles in a certain dependence context. Before stating in what a Bahadur representation consists, let us specify some general notation. Given some random variable Y , $F(\cdot) = F_Y(\cdot)$ is referred as the cumulative distribution function of Y , $\xi(p) = \xi_Y(p)$ for some $0 < p < 1$ as the quantile of order p . If $F(\cdot)$ is absolutely continuous with respect to Lebesgue measure, the probability density function is denoted by $f(\cdot) = f_Y(\cdot)$. Based on the observation of a vector $\mathbf{Y} = (Y(1), \dots, Y(n))$ of n random variables distributed as Y , the sample cumulative distribution function and the sample quantile of order p are respectively denoted by $\widehat{F}_Y(\cdot; \mathbf{Y})$ and $\widehat{\xi}_Y(p; \mathbf{Y})$ or simply by $\widehat{F}(\cdot; \mathbf{Y})$ and $\widehat{\xi}(p; \mathbf{Y})$.

Let $\mathbf{Y} = (Y(1), \dots, Y(n))$ a vector of n i.i.d. random variables such that $F''(\xi(p))$ exists and is bounded in a neighborhood of $\xi(p)$ and such that $F'(\xi(p)) > 0$, Bahadur proved that as $n \rightarrow +\infty$,

$$\widehat{\xi}(p) - \xi(p) = \frac{p - \widehat{F}(p)}{f(\xi(p))} + r_n,$$

¹*Key words and phrases.* Gaussian processes, Bahadur representation of sample quantiles.

with $r_n = \mathcal{O}_{a.s.}(n^{-3/4} \log(n)^{3/4})$ where a sequence of random variables U_n is said to be $\mathcal{O}_{a.s.}(v_n)$ if U_n/v_n is almost surely bounded. Kiefer obtained the exact rate $n^{-3/4} \log \log(n)^{3/4}$. Under an Assumption on $F(\cdot)$ which is quite similar to the one done by Bahadur, extensions of above results to dependent random variables have been pursued in Sen and Ghosh (1972) for ϕ -mixing variables, in Yoshihara (1995) for strongly mixing variables, and recently in Wu (2005) for short-range and long-range dependent linear processes, following works of Hesse (1990) and Ho and Hsing (1996). Finally, such a representation has been obtained by Coeurjolly (2007) for nonlinear functional of Gaussian sequences with correlation function decreasing as $k^{-\alpha}$ for some $\alpha > 0$.

Ghosh (1971) proposed in the i.i.d. case a much simpler proof of Bahadur's result which suffices for many statistical applications. He established under a weaker assumption on $F(\cdot)$ ($F'(\cdot)$ exists and is bounded in a neighborhood of $\xi(p)$ and $f(\xi(p)) > 0$) that the remainder term satisfies $r_n = o_p(n^{-1/2})$, which means that $n^{1/2}r_n$ tends to 0 in probability. This result is sufficient for example to establish a central limit theorem for the sample quantile. Our goal is to extend Ghosh's result to nonlinear functional of Gaussian sequences with correlation function decreasing as $k^{-\alpha}$. The Bahadur representation is presented in Section 2 and is applied to a central limit theorem for the sample quantile. Proofs are deferred in Section 3.

2 Main result

Let $\{Y(i)\}_{i=1}^{+\infty}$ be a stationary (centered) gaussian process with variance 1, and correlation function $\rho(\cdot)$ such that, as $i \rightarrow +\infty$

$$|\rho(i)| \sim i^{-\alpha} \tag{1}$$

for some $\alpha > 0$.

Let us recall some background on Hermite polynomials: the Hermite polynomials form an orthogonal system for the Gaussian measure and are in particular such that $\mathbf{E}(H_j(Y)H_k(Y)) = j! \delta_{j,k}$, where Y is referred to a standard Gaussian variable. For some measurable function $g(\cdot)$ defined on \mathbb{R} such that $\mathbf{E}(g(Y)^2) < +\infty$, the following expansion holds

$$g(t) = \sum_{j \geq \tau} \frac{c_j}{j!} H_j(t) \quad \text{with} \quad c_j = \mathbf{E}(g(Y)H_j(Y)),$$

where the integer τ defined by $\tau = \inf \{j \geq 0, c_j \neq 0\}$, is called the Hermite rank of the function g . Note that this integer plays an important role. For example, it is related to the correlation of $g(Y_1)$ and $g(Y_2)$, for Y_1 and Y_2 two standard gaussian variables with correlation ρ , since $\mathbf{E}(g(Y_1)g(Y_2)) = \sum_{k \geq \tau} \frac{(c_k)^2}{k!} \rho^k = \mathcal{O}(\rho^\tau)$.

Our result is based on the assumption that $F'_{g(Y)}(\cdot)$ exists and is bounded in a neighborhood of $\xi(p)$. This is achieved if the function $g(\cdot)$ satisfies the following assumption (see *e.g.* Dacunha-Castelle and Dufflo (1982), p.33).

Assumption A($\xi(p)$) : there exist U_i , $i = 1, \dots, L$, disjoint open sets such that U_i contains a unique solution to the equation $g(t) = \xi_{g(Y)}(p)$, such that $F'_{g(Y)}(\xi(p)) > 0$ and such that g is a \mathcal{C}^1 -diffeomorphism on $\cup_{i=1}^L U_i$.

Note that this assumption allows us to obtain

$$F'_{g(Y)}(\xi_{g(Y)}(p)) = f_{g(Y)}(\xi_{g(Y)}(p)) = \sum_{i=1}^L \frac{\phi(g_i^{-1}(t))}{g'(g_i^{-1}(t))},$$

where $g_i(\cdot)$ is the restriction of $g(\cdot)$ on U_i and where $\phi(\cdot)$ is referred to the probability density function of a standard Gaussian variable.

Now, define, for some real u , the function $h_u(\cdot)$ by:

$$h_u(t) = \mathbf{1}_{\{g(t) \leq u\}}(t) - F_{g(Y)}(u). \quad (2)$$

We denote by $\tau(u)$ the Hermite rank of $h_u(\cdot)$. For the sake of simplicity, we set $\tau_p = \tau(\xi_{g(Y)}(p))$. For some function $g(\cdot)$ satisfying Assumption **A($\xi(p)$)**, we denote by

$$\bar{\tau}_p = \inf_{\gamma \in \cup_{i=1}^L g(U_i)} \tau(\gamma), \quad (3)$$

that is the minimal Hermite rank of $h_u(\cdot)$ for u in a neighborhood of $\xi_{g(Y)}(p)$.

Theorem 1 *Under Assumption **A($\xi(p)$)**, the following result holds as $n \rightarrow +\infty$*

$$\hat{\xi}(p; \mathbf{g}(\mathbf{Y})) - \xi_{g(Y)}(p) = \frac{p - \hat{F}(\xi_{g(Y)}(p); \mathbf{g}(\mathbf{Y}))}{f_{g(Y)}(\xi_{g(Y)}(p))} + o_p(r_n(\alpha, \bar{\tau}_p)), \quad (4)$$

where $\mathbf{g}(\mathbf{Y}) = (g(Y(1)), \dots, g(Y(n)))$, for $i = 1, \dots, n$ and where the sequence $(r_n(\alpha, \bar{\tau}_p))_{n \geq 1}$ is defined by

$$r_n(\alpha, \bar{\tau}_p) = \begin{cases} n^{-1/2} & \text{if } \alpha \bar{\tau}_p > 1, \\ n^{-1/2} \log(n)^{1/2} & \text{if } \alpha \bar{\tau}_p = 1, \\ n^{(1-\alpha \bar{\tau}_p)/2} & \text{if } \alpha \bar{\tau}_p < 1. \end{cases} \quad (5)$$

Remark 1 The sequence $r_n(\alpha, \bar{\tau}_p)$ is related to the behaviour short-range or long-range dependent behaviour of the sequence $h_u(Y(1)), \dots, h_u(Y(n))$ for u in a neighborhood of $\xi(p)$. More precisely, it corresponds to the asymptotic behaviour of the sequence

$$\left(\frac{1}{n} \sum_{|i| < n} \rho(i)^{\bar{\tau}_p} \right)^{1/2}.$$

Corollary 2 Under Assumption $\mathbf{A}(\xi(\mathbf{p}))$, then if $\alpha \bar{\tau}_p > 1$, the following convergence in distribution holds as $n \rightarrow +\infty$

$$\sqrt{n} \left(\widehat{\xi}(p; \mathbf{g}(\mathbf{Y})) - \xi_{g(Y)}(p) \right) \xrightarrow{d} \mathcal{N}(0, \sigma_p^2), \quad (6)$$

where $\sigma_p^2 = \sum_{i \in \mathbb{Z}} \sum_{j \geq \tau_p} \frac{(c_j)^2}{j!} \rho(i)^j$

The proof of this result is omitted since it is a direct application of Theorem 1 and a central limit theorem adapted to nonlinear functional of Gaussian sequences, *e.g.* Breuer and Major (1983).

3 Proofs

3.1 Auxiliary Lemma

Lemma 3 For every $j \geq 1$ and for all positive sequence $(u_n)_{n \geq 1}$ such that $u_n \rightarrow 0$, as $n \rightarrow +\infty$, we have, under Assumption $\mathbf{A}(\xi(\mathbf{p}))$

$$I = \int_{\mathbb{R}} H_j(t) \phi(t) \mathbf{1}_{\{|g(t) - \xi_{g(Y)}(p)| \leq u_n\}} dt \sim u_n \kappa_j, \quad (7)$$

where κ_j is defined, for every $j \geq 1$, by

$$\kappa_j = \begin{cases} -2 \sum_{i=1}^L \frac{\phi'(g_i^{-1}(\xi(p)))}{g'(g_i^{-1}(\xi(p)))} & \text{if } j = 1, \\ 2(-1)^j \sum_{i=1}^L \frac{\phi^{(j-2)}(g_i^{-1}(\xi(p)))}{g'(g_i^{-1}(\xi(p)))} & \text{if } j > 1. \end{cases} \quad (8)$$

Proof. Under Assumption $\mathbf{A}(\xi(\mathbf{p}))$, there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$,

$$I = \sum_{i=1}^L I_i \quad \text{with} \quad I_i = \int_{U_i} H_j(t) \phi(t) \mathbf{1}_{\{\xi(p) - u_n \leq g(t) \leq \xi(p) + u_n\}} dt. \quad (9)$$

Assume without loss of generality that the restriction of $g(\cdot)$ on U_i (denoted by $g_i(\cdot)$) is an increasing function, we have

$$\begin{aligned} I_i &= \int_{U_i} H_j(t) \phi(t) \mathbf{1}_{\{\xi(p)-u_n \leq g(t) \leq \xi(p)+u_n\}} dt \\ &= \int_{g_i^{-1}(\xi(p)-u_n)}^{g_i^{-1}(\xi(p)+u_n)} H_j(t) \phi(t) dt \\ &= \begin{cases} \phi(m_{i,n}) - \phi(M_{i,n}) = (m_{i,n} - M_{i,n}) & \text{if } j = 1 \\ (-1)^j (\phi^{(j-1)}(M_{i,n}) - \phi^{(j-1)}(m_{i,n})) & \text{if } j > 1, \end{cases} \end{aligned}$$

where $M_{i,n} = g_i^{-1}(\xi(p) + u_n)$ and $m_{i,n} = g_i^{-1}(\xi(p) - u_n)$. Then, there exists $\omega_{n,i,j} \in [m_{i,n}, M_{i,n}]$ for every $j \geq 1$ such that

$$I_i = \begin{cases} (m_{i,n} - M_{i,n}) \phi^{(1)}(\omega_{n,i,1}) & \text{if } j = 1 \\ (-1)^j (M_{i,n} - m_{i,n}) \phi^{(j-2)}(\omega_{n,i,j}) & \text{if } j > 1. \end{cases},$$

Under Assumption **A**($\xi(\mathbf{p})$), we have, as $n \rightarrow +\infty$

$$\omega_{n,i,j} \sim g_i^{-1}(\xi(p)) \quad \text{and} \quad M_{i,n} - m_{i,n} \sim 2u_n \frac{1}{g'(g_i^{-1}(\xi(p)))},$$

which ends the proof. \blacksquare

3.2 Proof of Theorem 1

For the sake of simplicity, we set $\widehat{\xi}(p) = \widehat{\xi}(p; \mathbf{g}(\mathbf{Y}))$, $\xi(p) = \xi_{g(Y)}(p)$, $\widehat{F}(\cdot) = \widehat{F}(\cdot; \mathbf{g}(\mathbf{Y}))$, $F(\cdot) = F_{g(Y)}(\cdot)$ et $f(\cdot) = f_{g(Y)}(\cdot)$ and $r_n = r_n(\alpha, \bar{\tau}_p)$. Define,

$$V_n = r_n^{-1} \left(\widehat{\xi}(p) - \xi(p) \right) \quad \text{and} \quad W_n = r_n^{-1} \left(\frac{p - F(p)}{f(p)} \right).$$

The result is established if $V_n - W_n \xrightarrow{\mathbb{P}} 0$ as $n \rightarrow +\infty$. It suffices to prove that V_n and W_n satisfy the conditions of Lemma 1 of Ghosh (1971):

- **condition (a)** : for all $\delta > 0$, there exists $\varepsilon = \varepsilon(\delta)$ such that $\mathbb{P}(|W_n| > \varepsilon) < \delta$.
- **condition (b)** : for all $y \in \mathbb{R}$ and for all $\varepsilon > 0$

$$\lim_{n \rightarrow +\infty} \mathbb{P}(V_n \leq y, W_n \geq k + \varepsilon) \quad \text{and} \quad \lim_{n \rightarrow +\infty} \mathbb{P}(V_n \geq y + \varepsilon, W_n \geq k)$$

condition (a) : from Bienaymé-Tchebyshev's inequality it is sufficient to prove that $\mathbf{E}W_n^2 = \mathcal{O}(1)$. Rewrite $W_n = \frac{r_n^{-1}}{n} \sum_{i=\ell+1}^n h_{\xi(p)}(Y(i))$. Let c_j (for some $j \geq 0$) denote the j -th Hermite coefficient of $h_{\xi(p)}(\cdot)$. Since $h_{\xi(p)}(\cdot)$ has at least Hermite rank $\bar{\tau}_p$, then

$$\begin{aligned} \mathbf{E}W_n^2 &= \frac{r_n^{-2}}{n^2} \sum_{i_1, i_2=1}^n \mathbf{E} (h_{\xi(p)}(Y(i_1)) h_{\xi(p)}(Y(i_2))) \\ &= \frac{r_n^{-2}}{n^2} \sum_{i_1, i_2=1}^n \sum_{j_1, j_2 \geq \bar{\tau}_p} c_{j_1} c_{j_2} \mathbf{E} (H_{j_1}(Y(i_1)) H_{j_2}(Y(i_2))) \\ &= \frac{r_n^{-2}}{n^2} \sum_{i_1, i_2=1}^n \sum_{j \geq \bar{\tau}_p} \frac{(c_j)^2}{(j)!} \rho(i_2 - i_1)^j \\ &= \mathcal{O} \left(r_n^{-2} \times \frac{1}{n} \sum_{|i| < n} \rho(i)^{\bar{\tau}_p} \right) = \mathcal{O}(1), \end{aligned}$$

from Remark 1.

condition (b) : let $y \in \mathbb{R}$, we have

$$\begin{aligned} \{V_n \leq y\} &= \left\{ \widehat{\xi}(p) \leq y \times r_n + \xi(p) \right\} \\ &= \left\{ p \leq \widehat{F}(y \times r_n + \xi(p)) \right\} = \{Z_n \leq y_n\}, \end{aligned} \quad (10)$$

with

$$Z_n = \frac{r_n^{-1}}{f(\xi(p))} \left(F(y \times r_n + \xi(p)) - \widehat{F} \left(\frac{y}{\sqrt{r_n}} + \xi(p) \right) \right)$$

and

$$y_n = \frac{r_n^{-1}}{f(\xi(p))} \left(F(y \times r_n + \xi(p)) - p \right)$$

Under Assumption $\mathbf{A}(\xi(p))$, we have $y_n \rightarrow y$, as $n \rightarrow +\infty$. Now, prove that $Z_n - W_n \xrightarrow{\mathbb{P}} 0$.

Without loss of generality, assume $y > 0$. Then, we have

$$\begin{aligned} W_n - Z_n &= \frac{r_n^{-1}}{f(p)} \left(\widehat{F}(y \times r_n + \xi(p)) - F(y \times r_n + \xi(p)) - \widehat{F}(\xi(p)) + F(\xi(p)) \right) \\ &= \frac{r_n^{-1}}{n} \frac{1}{f(\xi(p))} \sum_{i=1}^n h_{\xi(p),n}(Y(i)) \end{aligned}$$

where $h_{\xi(p),n}(\cdot)$ is the function defined for $t \in \mathbb{R}$ by :

$$h_{\xi(p),n}(t) = \mathbf{1}_{\left\{ \xi(p) \leq g(t) \leq \xi(p) + y \times r_n \right\}}(t) - \mathbb{P} \left(\xi(p) \leq g(Y) \leq \xi(p) + y \times r_n \right).$$

For n sufficiently large, the function $h_{\xi(p),n}(\cdot)$ has Hermite rank $\bar{\tau}_p$. Denote by $c_{j,n}$ the j -th Hermite coefficient of $h_{\xi(p),n}(\cdot)$. From Lemma 3, there exists a sequence $(\kappa_j)_{j \geq \bar{\tau}_p}$ such that, as $n \rightarrow +\infty$

$$c_{j,n} \sim \kappa_j \times r_n.$$

Since, for all $n \geq 1$ $\mathbf{E}(h_n(Y)^2) = \sum_{j \geq \bar{\tau}_p} (c_{j,n})^2/j! < +\infty$, it is clear that the sequence $(\kappa_j)_{j \geq \bar{\tau}_p}$ is such that $\sum_{j \geq \bar{\tau}_p} (\kappa_j)^2/j! < +\infty$. By denoting λ a positive constant, we get, as $n \rightarrow +\infty$

$$\begin{aligned} \mathbf{E}(W_n - Z_n)^2 &= \frac{r_n^{-2}}{n^2} \frac{1}{f(\xi(p))^2} \sum_{i_1, i_2=1}^n \mathbf{E}(h_{\xi(p),n}(Y(i_1)) h_{\xi(p),n}(Y(i_2))) \\ &= \frac{r_n^{-2}}{n^2} \frac{1}{f(\xi(p))^2} \sum_{i_1, i_2=1}^n \sum_{j_1, j_2 \geq \bar{\tau}_p} c_{j_1, n} c_{j_2, n} \mathbf{E}(H_{j_1}(Y(i_1)) H_{j_2}(Y(i_2))) \\ &= \frac{r_n^{-2}}{n^2} \frac{1}{f(\xi(p))^2} \sum_{i_1, i_2=1}^n \sum_{j \geq \bar{\tau}_p} \frac{c_{j,n}^2}{j!} \rho(i_2 - i_1)^j \\ &\leq \lambda \frac{r_n^{-2}}{n} \sum_{j \geq \bar{\tau}_p} \frac{(\kappa_j)^2}{j!} r_n^2 \sum_{|i| < n} \rho(i)^j = \mathcal{O}\left(\frac{1}{n} \sum_{|i| < n} \rho(i)^{\bar{\tau}_p}\right) = \mathcal{O}(r_n^2), \end{aligned}$$

from Remark 1. Therefore, $W_n - Z_n$ converges to 0 in probability, as $n \rightarrow +\infty$. Thus, for all $\varepsilon > 0$, we have, as $n \rightarrow +\infty$,

$$\mathbb{P}(V_n \leq y, W_n \geq y + \varepsilon) = \mathbb{P}(Z_n \leq y_n, W_n \geq y + \varepsilon) \rightarrow 0.$$

Following the sketch of this proof, we also have $\mathbb{P}(V_n \geq y + \varepsilon, W_n \leq y) \rightarrow 0$, ensuring condition (b). Therefore, $W_n - Z_n$ converges to 0 in probability, as $n \rightarrow +\infty$. Thus, for all $\varepsilon > 0$, we have, as $n \rightarrow +\infty$,

$$\mathbb{P}(V_n \leq y, W_n \geq y + \varepsilon) = \mathbb{P}(Z_n \leq y_n, W_n \geq y + \varepsilon) \rightarrow 0.$$

Following the sketch of this proof, we also have $\mathbb{P}(V_n \geq y + \varepsilon, W_n \leq y) \rightarrow 0$, ensuring condition (b).

References

BAHADUR R.R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.* **37** 577-580.

- BREUER P. AND MAJOR P. (1983). Central limit theorems for non-linear functionals of Gaussian fields. *J. Mult. Ann.* **13** 425-441.
- COEURJOLLY J.-F. (2007). Hurst exponent estimation of locally self-similar Gaussian processes using sample quantiles. *E-print arxiv: <http://fr.arxiv.org/abs/math.ST/0506290>*, 1-44.
- DACUNHA-CASTELLE D. AND DUFLO M. (1982) *Exercices de probabilités et statistiques. Tome 1 Problèmes à temps fixe*. Collection Mathématiques Appliquées pour la maîtrise, Masson, Paris.
- GHOSH J.K. (1971). A new proof of the Bahadur Representation of Quantiles and an application. *Ann. Math. Statist.* **42** 1957-1961.
- HESSE C.H. (1990). A Bahadur-type representation for empirical quantiles of a large class of stationary, possibly infinite-variance, linear processes. *Ann. Statist.* **18** 1188-1202.
- HO H.C. AND HSING T. (1996) On the asymptotic expansion of the empirical process of long-memory moving averages. *Ann. Statist.* **24** 992-1024.
- KIEFER J. (1967). On Bahadur's representation of sample quantiles. *Ann. Math. Statist.* **38** 1323-1342.
- SEN P.K. (1972). On the Bahadur representation of sample quantiles for sequences of ϕ -mixing random variables. *J. Multivariate Anal.* **2** 77-95.
- YOSHIHARA K.-I. (1995). The Bahadur representation of sample quantiles for sequences of strongly mixing random variables. *Stat. Probab. Lett.* **24** 299-304.
- WU W.-B. (2005). On the Bahadur representation of sample quantiles for dependent sequences. *Ann. Statist.* **33** 1934-1963.

J.-F. COEURJOLLY

SAGAG, DEPARTMENT OF STATISTICS, GRENOBLE

1251 AV. CENTRALE BP 47

38040 GRENOBLE CEDEX 09

FRANCE

E-MAIL: Jean-Francois.Coeurjolly@upmf-grenoble.fr