Journal of Nonparametric Statistics
Vol. 00, No. 00, Month 2010, 1–10

RESEARCH ARTICLE

A bootstrap method to test for the existence of finite moments

Igor Fedotenkov[a*]

[a] *University of Verona, Via dell'Artigliere 19, 37129 Verona, Italy*;
(*Received 00 Month 200x; in final form 00 Month 200x*)

This paper presents a simple bootstrap test to verify the existence of finite moments. The efficacy of the test relies on the fact that in the absence of a first moment and under certain general conditions, the arithmetic average of a sample grows at a rate greater than the growth rates of the arithmetic averages of the sub-samples. Firstly, we show test consistency analytically. Then, Monte-Carlo simulations are performed to compare our test with the Hill estimator.

Keywords: Moment existence; bootstrap test; sub-sampling; consistency; Monte-Carlo simulations

AMS Subject Classification: 62F03; 62F05; 62F40; 62G32

*Corresponding author. Email: i.fedotenkov@gmail.com

ISSN: 1048-5252 print/ISSN 1029-0311 online

DOI: 10.1080/1048525YYxxxxxxxx
http://www.informaworld.com

1. Introduction

Many theorems in statistics and econometrics assume that a distribution of a random variable has a finite first, second or other moment. For example, Kolmogorov's law of large numbers requires the existence of a first finite moment. Lyapunov's central limit theorem assumes that a sample is drawn from a distribution with a finite second moment. There is, therefore, the need for a test to verify these assumptions.

In this paper a bootstrap method is applied to test the hypothesis of the existence of finite moments. Our test is primarily designed to verify the hypothesis that a first finite moment exists; however, it can also be used to verify if other moments are finite by raising the sample data to the corresponding power.

The Hill estimator (Hill (1975))[1] is the method most commonly used to make an inference regarding the heaviness of the tails. The method is as follows: Pareto-type tails are assumed, then parameters are estimated. Next, the hypothesis for the corresponding parameter is tested. However, it is not always clear how many observations should be treated as tails, since too many observations produce a large bias and too few observations a large variance. Methods devised to resolve this problem have been presented in a vast literature. Hall (1981), Beirlant et al. (1996), Danielson et al. (2001) and many others have tried to determine an optimal number of observations to be used for the estimation of the tail index. Resnick and Stărică (1997), and Martins et al. (2004) calculated the tail index averaging the Hill estimators. However, the fact that all these tests rely on a specific assumption about tail distribution remains a major unresolved problem. This may produce misleading results when the underlying distribution has a different form.

Another relevant paper is that of Bertail et al. (2001), who not only studied the estimation of distributions of diverging statistics using sub-samples, but also provided an alternative method to estimate the tail index. Although their paper applied sub-sampling to obtain an inference about the heaviness of the tails, the principal problems remained unsolved, as a rather restrictive regularity of the tails is assumed, and testing for the existence of moments is indirect: the conclusion was derived from having checked their hypothesis about the parameter of a specific functional form.

Mandelbrot (1963) proposed another frequently used testing method (see, for example, R. Cont (2001) to verify the existence of finite moments. The idea of the test is that the graph of the moment of interest as a function of time (or number of observations) tends to converge to the theoretical moment - if the moment exists. If the moment does not exist, its behavior is unstable. While this method is quite intuitive, it is informal, as the acceptance or rejection of the hypothesis depends on the intuition of the researcher, who may be relying on personal experience or other external factors.

Our test aims to resolve several of these problems. The hypothesis that a finite moment exists is checked directly and simply. The researcher need not estimate any parameters, nor take any additional steps to calculate test statistics. Furthermore, consistency does not depend on a specific functional form of the tail. Although some tail regularity is assumed in order to show the consistency of the test, these assumptions are quite inconsequential. However, the choice of sub-sample size and one additional parameter poses a similar question to that of the Hill estimator, i.e., choosing the number of observations treated as a tail.

Monte-Carlo simulations show that the test performs rather well for a large number of distributions. In borderline cases, such that a finite EX does not exist, but $EX^{1-\epsilon} < \infty$,

[1] Alternative estimators are discussed in detail by L. de Haan and L. Peng (1998).

for a small positive ϵ, the test is not very powerful, but still outperforms the Hill estimator.

In the following section the basic framework is introduced and the test is formulated. In the third section test consistency is shown. The forth section presents Monte-Carlo simulations and compares the test with the Hill estimator.

2. The test

Our testing method is designed to verify the existence of the first moment. A researcher wishing to check for the existence of other moments may apply the test for X^p, where p is the moment of interest.

Let sample $(X_1, X_2, ..., X_n)$ be a random sample independently drawn from distribution $\mathcal{F}$ with a support $\mathcal{D}$. For simplicity it will be assumed that X_i can take positive values only, hence, $\mathcal{D} \subset \mathcal{R}_+$. $\hat{\mathcal{F}}_n(x)$ is an estimate of $\mathcal{F}$, $\hat{\mathcal{F}}_n(x) = \sum_{i=1}^n \mathbb{1}(X_i \leq x)/n$. The assumption that X_i is positive is not very restrictive because it is always possible to use an absolute value. Alternatively the hypothesis of finite mean can be checked for positive and negative values separately. Let $(X^*_{k,1}, X^*_{k,2}, ..., X^*_{k,m})$ be a sub-sample of X randomly drawn from $\hat{\mathcal{F}}_n$. There are M sub-samples and the subscript k, $k = 1...M$, refers to an index of a sub-sample.

Define

$$\mu_n = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

$$\mu^*_{m,k} = \frac{1}{m}\sum_{i=1}^{m} X^*_{k,i}.$$

Hypothesis:

- H_0: distribution $\mathcal{F}$ has a finite first moment.
- H_1: $\mathcal{F}$ does not have a finite first moment.

Test:

(1) Choose a significance level $\alpha > 0$.
(2) Choose a bootstrap sample size m, $m \to \infty$ as $n \to \infty$, $m = o(n)$.[1]
(3) Draw M sub-samples from $\hat{\mathcal{F}}_n$.
(4) Calculate the p-value

$$p_n = \frac{1}{M}\sum_{k=1}^{M} \mathbb{1}(\mu^*_{m,k} > \xi\mu_n).$$

We assume that ξ is a constant $0 < \xi < 1$. (The test is more powerful when ξ is close to unity.) $\mathbb{1}(\cdot)$ is a unit indicator function.

(5) If $p_n > \alpha$ accept H_0. Otherwise reject H_0 in favor of H_1.

[1]For simplicity, $m \sim \log(n)$ will be assumed further in the analysis. A more general outcome is shown in appendix A.

The intuition of the method is as follows: if the first moment exists, both the arithmetic average of the sample and sub-samples uniformly converge to the theoretical mean. So, μ_m^* will not be far from μ_n. Reducing the critical value μ_n multiplying it by ξ, $0 < \xi < 1$, increases the share of μ_m^*, exceeding the critical value, and rules out possible counterexamples with highly skewed probability functions. On the other hand, in the absence of the first moment, under some additional conditions, μ_n expands to infinity with probability one, as reported by Derman and Robins (1955). Thus, in setting $m = o(n)$, we may expect that in general μ_m^* converges to infinity more slowly than μ_n. It may be argued that μ_m^* will "spin" around μ_n in any case, since μ_n is finite for each $n < \infty$, and $E^*\mu_m^* = \mu_n$. But distributions with no finite mean usually exhibit a small number of very large outliers, driving up the arithmetic average, so that it exceeds most observations. Thus, with an increasing number of observations, the portion of observations above the arithmetic average of the sample decreases and the probability of drawing them into a sub-sample decreases. Hence, as a rule, if the distribution does not have a finite first moment, the share of $\mu_m^* > \mu_n$ decreases as n grows. This finding is expressed formally in the following section.

First, observe what occurs when the first moment exists:

$$Pr^*\left(\frac{1}{m}\sum_{i=1}^{m} X_i^* > \xi\mu_n\right) =$$

$$Pr^*\left(X_m^* > m\xi\mu_n - \sum_{i=1}^{m-1} X_i^*\right) \xrightarrow{a.s.}$$

$$Pr\Big(X > m\xi\mu - (m-1)\mu_{m-1}^*\Big), \quad as \quad n \to \infty.$$

Where μ denotes a theoretical mean, the almost sure convergence in the bottom line follows from Kolmogorov's law of large numbers: $\mu_n \xrightarrow{a.s.} \mu$, and the Glivenko-Cantelli lemma: $Pr^* \xrightarrow{a.s.} Pr$, when $n \to \infty$.

$$Pr\Big(X > m\mu\xi - (m-1)\mu_{m-1}^*\Big) \xrightarrow{a.s.} Pr\left(X > -\infty\right) = 1, \quad as \quad m \to \infty. \tag{1}$$

$\mu_{m-1}^* \xrightarrow{a.s} \mu$ when $m \to \infty$ following Kolmogorov's law of large numbers. $\xi < 1$ insures that increasing m drives the expression $m\mu\xi - (m-1)\mu_{m-1}^*$ to $-\infty$. Hence, probability converges to unity, because of the properties of a cumulative probability function.

3. Consistency

Lemma 3.1: *Suppose the existence of constants C, $C > 0$ and β, $0 < \beta < 1$, such that $\mathcal{F}(x) \leq 1 - \frac{C}{x^\beta}$ for large x. Then $\forall\ \gamma > \beta$, $Pr(\mu_n > n^{\frac{1}{\gamma}-1}) = 1$ for sufficiently large n.*

Proof: The proof of this lemma follows directly from the proof of the theorem of Derman and Robbins (1955). They found that $\mu_n \xrightarrow{a.s.} \infty$ if the right tail of a distribution is heavier than the left. This lemma is an intermediate finding, shown to prove their theorem. □

Assumption $0 < \xi < 1$ is needed in equation (1). The following lemma, corollary and theorem are also valid for a larger range of ξ. However, to avoid misunderstandings we suppose ξ to be a constant $0 < \xi < 1$.

Lemma 3.2: *Supposing the existence of constants $\tilde{C}$, $\tilde{C} > 0$ and δ, $\delta > 2$, such that $\mathcal{F}(x) > 1 - \tilde{C}/\log^{\delta}(x)$ for large x. Then $Pr\big(\mu_n > \xi \exp(n)\big) = 0$ for sufficiently large n.*

Proof:

$$Pr\big(\mu_n > \xi \exp(n)\big) \le Pr\big(\max_{i=1\dots n} X_i > \xi \exp(n)\big) =$$

$$1 - \big(\mathcal{F}(\xi \exp(n))\big)^n \le 1 - \Big(1 - \frac{\tilde{C}}{(n + \log \xi)^{\delta}}\Big)^n.$$

As $n > max(2, -\log \xi)$, Bernoulli's inequality may be applied:

$$1 - \Big(1 - \frac{\tilde{C}}{(n + \log \xi)^{\delta}}\Big)^n \le \frac{\tilde{C} n}{(n + \log \xi)^{\delta}}.$$

With $\delta > 2$ the sum $\sum_{n > -\log \xi} \frac{n}{(n+\log \xi)^{\delta}}$ is finite (easily verifiable with the integral test for convergence). Therefore, the Borel-Cantelli lemma implies that $Pr(\mu_n > \xi \exp(n)) = 0$ for sufficiently large n.

□

Corollary 3.3: *Suppose that for large x, $\mathcal{F}(x) > 1 - \tilde{C}/\log^{\delta}(x)$ for some $\delta > 2$ and $\tilde{C} > 0$. Then $Pr^*(\mu^*_m > \xi \exp(m)) = 0$ for sufficiently large m.*

The corollary result derives from the fact that $X^*_1, ..., X^*_m$ is a sub-sample taken from $X_1, ..., X_n$. We note that according to the Glivenko-Cantelli lemma $Pr^* \xrightarrow{a.s} Pr$ as $n \to \infty$, the statement of the corollary follows from the proof of lemma 3.2.

Theorem 3.4: *Suppose that the conditions of lemma 3.1 and lemma 3.2 are satisfied, $m \le \left(\frac{1}{\gamma} - 1\right) \log n$ for some γ, $\beta < \gamma < 1$, then $\lim_{n,m\to\infty} Pr^*(\mu^*_m > \xi \mu_n) = 0$.*

Proof:

$$\begin{aligned} Pr^*\big(\mu^*_m > \xi \mu_n\big) &\le Pr^*\big(\mu^*_m > \xi n^{\frac{1}{\gamma}-1}\big) \\ &\le Pr^*\Big(\mu^*_m > \xi \exp m\Big). \end{aligned}$$

The first inequality follows from lemma 3.1. The second inequality is a result of construction $m \le \left(\frac{1}{\gamma} - 1\right) \log n$. The corollary of lemma 3.2 implies that $Pr^*\big(\mu^*_m > \xi \exp m\big) = 0$ as $m \to \infty$.

□

Theorem 3.4 shows that the test is consistent when the conditions of lemma 3.1 and lemma 3.2 are satisfied. Consistency is found, therefore, for a very large set of probability functions. However, these conditions do not apply to all cases of distributions without a finite first moment. If the right tail behaves as $1 - 1/x$ for large x the first moment does not exist, though the tail does not fulfil the assumption of lemma 3.1. Indeed, the test is not consistent for such distributions. For example, if the sample is drawn from a standard Cauchy distribution, the arithmetic averages of sample and sub-samples all have a standard Cauchy distribution as well. Hence, $Pr\big(\mu_m^* > \xi\mu_n\big)$, (where μ_n and μ_m^* are calculated from positive values only), should not converge to zero. However, it may still be greater than the level of significance.

The condition of lemma 3.2 also restricts the probability functions. Fortunately, this condition is not necessary. If a distribution does not satisfy the condition of lemma 3.2, then the condition of lemma 3.1 can be rewritten as $\mathcal{F}(x) \leq 1 - 1/\log x^\beta$ for large x and some β, $\beta \geq 2$. Correspondingly, the condition of lemma 3.2 can be extended to more heavily-tailed distributions. Moreover, this expedient may be repeated telescopically. This is a difficult procedure, but to illustrate its feasibility we show consistency for tails heavier than $\mathcal{F}(x) = 1 - 1/\log x^\beta$ with $\beta > 2$.

Theorem 3.5: *Suppose that there $\exists$ such a constant C, and a positive integer r, that $F(x) = 1 - C/\underbrace{\log ... \log x}_{r}$ for large x, $0 < \xi < 1$, moreover, $m \approx \log n$, then $\lim_{n,m\to\infty} Pr^*(\mu_m^* > \xi\mu_n) = 0$.*

Proof: The proof of theorem 3.5 is presented in appendix B. □

Theorem 3.5 extends the set of functions restricted by the assumptions of statement 3.4. It shows consistency for a different set of probability functions, which do not have a finite first moment. However, it is still possible to construct a distribution with tails heavier than $F(x) > 1 - 1/(\underbrace{\log ... \log x}_{r})^\delta$. But from the proof of lemma 3.2 it is intuitively clear that consistency can be shown for an even larger set of probability functions when we allow m to grow at a rate much slower than $m \approx \left(\frac{1}{\gamma} - 1\right)\log n$. This is taken into account in appendix A, where we present a more general result. It is less intuitive, yet still simple to apply, and we use it to prove theorem 3.5.

4. Monte-Carlo simulations, comparison with the Hill estimator

In this section Monte-Carlo simulations are performed to evaluate the efficacy of the test. First, our test is compared with the Hill estimator for independent data. Then, test performance is checked in a numerical experiment with dependent data. While test consistency is evaluated for independent data only, the performance of the test using dependent data is also interesting, as this is the case in most real applications. In the numerical experiment with independent data, observations are drawn from log-logistic distribution with the scale parameter normalised to unity and shape parameter values equal to 0.5, 0.9, 1, 1.1 and 1.5 and for the absolute values of standard Cauchy distribution. In cases of a Cauchy distribution and a log-logistic distribution with parameters 0.5, 0.9 or 1 the finite first moment does not exist. However, the Cauchy distribution and the log-logistic distribution with parameter 1 do not fulfill the requirements of lemma 3.1 and lemma A.1 (in appendix A), so test behavior is of particular interest in this case.

Table 1 shows the share of H_0 rejected by the bootstrap test. For comparison table 2 shows test performance based on the Hill estimator. The constant ξ is equalised to 0.999. The size of a bootstrap sample is approximately equal to $0.4 * \log n$, which corresponds to the functional form for m used in lemma 3.2. The constant 0.4 is used to ensure that sub-samples have had at least two observations for the smallest n used in simulations.

Table 1. Bootstrap. Share H_0 rejected

n	$ll(0.5)$	$ll(0.9)$	$ll(1)$	$\|Cauchy\|$	$ll(1.1)$	$ll(1.5)$
100	0.292	0.071	0.054	0.051	0.034	0.006
500	0.686	0.121	0.070	0.066	0.038	0.004
10^3	0.764	0.130	0.072	0.071	0.039	0.003
10^4	0.999	0.193	0.086	0.082	0.033	0.001
10^5	1	0.317	0.096	0.100	0.032	0

Table 2. Hill estimator. Share H_0 rejected

n	$ll(0.5)$	$ll(0.9)$	$ll(1)$	$\|Cauchy\|$	$ll(1.1)$	$ll(1.5)$
100	0.549	0	0	0	0	0
500	0.944	0.004	0	0	0	0
10^3	0.989	0.020	0.001	0	0	0
10^4	1	0.128	0.007	0.006	0	0
10^5	1	0.466	0.013	0.012	0	0

The number of observations used for the Hill parameter estimation is approximately equal to $n^{1/2}$.

10000 Monte-Carlo simulations were performed. As regards the bootstrap, the null hypothesis is rejected when the estimated p_n is smaller than 0.05. In the case of the Hill estimator the confidence interval can be constructed for $1/\alpha$ value, where α is a tail index, using asymptotic normality associated with some additional assumptions (see Haeusler and Teugels (1985)). However, the confidence interval for α would be biased in this case. To improve test performance we constructed a confidence interval $[\hat{\alpha} - 2\hat{\sigma}(\hat{\alpha}), \hat{\alpha} + 2\hat{\sigma}(\hat{\alpha})]$, where the standard deviation $\hat{\sigma}(\hat{\alpha})$ is estimated from the simulated $\hat{\alpha}$. We rejected the null hypothesis when the confidence interval did not intersect with the region $(1, \infty)$, where the null hypothesis is valid.

Simulations were performed for $n = 100$, 500, 10^3, 10^4, 10^5. The number of bootstrap sub-samples was 10000.

Tables 1 and 2 show that the test based on the Hill estimator performed better for very heavy tails i.e. a log-logistic distribution with the 0.5 shape parameter, but in the borderline cases (Cauchy and log-logistic with the shape parameter 1), it rarely rejected the null hypothesis because the estimate of the index tail in these cases is usually close to 1. Thus, the confidence interval very often intersects with the region in which the distributions have a finite first moment. In these cases the bootstrap test performed better: its rejection rate being greater than the significance level in all cases.

As expected, the test based on the Hill estimator classified all distributions correctly when observations were drawn from those distributions with finite first moments. The bootstrap test made few mistakes - the error rate was very low - well below the level of significance. We should note that the rate of first order mistakes does not approach the significance level even asymptotically, indicating that the test proposed in this paper is not exact.

Tables 3 and 4 compare our test with the Hill estimator in cases of serially dependent data. The data are constructed as follows: The first observation was made from log-

Table 3. Bootstrap. Share H_0 rejected. Dependent data

n	$ll(0.5)$	$ll(0.9)$	$ll(1)$	$\|Cauchy\|$	$ll(1.1)$	$ll(1.5)$
100	0.300	0.070	0.052	0.049	0.034	0.007
500	0.692	0.121	0.072	0.069	0.043	0.005
10^3	0.751	0.127	0.074	0.070	0.039	0.004
10^4	0.999	0.198	0.075	0.087	0.036	0.001
10^5	1	0.319	0.099	0.096	0.032	0

Table 4. Hill estimator. Share H_0 rejected. Dependent data

n	$ll(0.5)$	$ll(0.9)$	$ll(1)$	$\|Cauchy\|$	$ll(1.1)$	$ll(1.5)$
100	0.578	0	0	0	0	0
500	0.947	0.005	0	0	0	0
10^3	0.993	0.020	0	0	0	0
10^4	1	0.143	0.008	0.006	0	0
10^5	1	0.465	0.010	0.009	0	0

logistic or Cauchy distributions with scale parameter equal to unity - as was done for independent data. If the observation was smaller than unity, the next observation was taken from the distribution, again with the same parameters. If not smaller, the next observation was made from the distribution with scale parameter equal to 10, with other parameters remaining unchanged. Next, new observations were compared to unity and it was then determined from which distribution the next observation should be taken in the same way. The procedure continued until an adequate sample was compiled. The tests performed similarly, both with serially dependent data and with independent data, which indicates a degree of test robustness in accordance with our assumption.

Acknowledgments

I would like to thank the participants at the 23rd Nordic Conference on Mathematical Statistics and the 10th International Vilnius Conference on Probability Theory and Mathematical Statistics (2010), especially Alfredas Račkauskas, Remigijus Leipus, and Vygantas Paulauskas. Furthermore, I wish to thank Marius Radavičius and Irena Mikolajun for their help and consulting. Finally, I would like to express my appreciation to the anonymous referee for his/her useful comments.

References

Beirlant, J., Vynckier, P., and Teugels, J.L. (1996), "Tail index estimation, Pareto quantile plots, and regression diagnostics," *Journal of the American Statistical Association*, 91, 1659–1667.

Bertail, P., Haefke, C., Politis, D., and White, H., "A subsampling approach to estimating the distribution of diverging statistics with application to assessing financial market risks," *Economics Working Papers* Nr. 599 (2001).

Cont, R. (2001), "Empirical properties of asset returns: stylized facts and statistical issues," *Quantitative Finance*, 1, 223–236.
Danielson, J., de Haan, L., Peng, L., and de Vries, C.G. (2001), "Using a bootstrap method to choose the sample fraction in tail index estimation," *Journal of Multivariate Analysis*, 76, 226–248.
Derman, C., and Robbins, H. (1955), "The strong law of large numbers when the first moment does not exist," *Proceedings of the National Academy of Sciences*, 41, 586–587.
Haan, L.D., and Peng, L. (1998), "Comparison of tail index estimators," *Statistica Neerlandica*, 52, 60–70.
Haeusler, E., and Teugels, J.L. (1985), "On Asymptotic Normality of Hill's Estimator for the Exponent of Regular Variation," *The Annals of Statistics*, 13, 743–756.
Hall, P. (1981), "On Some Simple Estimates of an Exponent of Regular Variation," *Journal of the Royal Statistical Society*, 44, 37–42.
Hill, B. (1975), "A simple general approach to inference about the tail of a distribution," *The Annals of Statistics*, 3, 1163–1174.
Mandelbrot, B. (1963), "The variation of certain speculative prices," *The Journal of Business*, 36, 394–419.
Martins, M.J., Gomes, M.I., and Neves, M.M. (2004), "Averages of Hill Estimators," *Test*, 13, 113–128.
Resnick, S., and Stărică, C. (1997), "Smoothing the Hill estimator," *Advances in Applied Probability*, 29, 271–293.

Appendix A. Consistency in a general case

Suppose a sample is drawn from a distribution $\mathcal{F}(x) = 1 - 1/f(x)$, where $f(x)$ is a nondecreasing function, such that $\mathcal{F}()$ is well defined. $m = \tilde{m}(n)$ is an increasing function such that $\tilde{m}(n) \to \infty$, $\tilde{m}(n) = o(n)$ as $n \to \infty$. Furthermore, an inverse function $\tilde{m}^{-1}()$ exists.

Lemma A.1: *If such increasing functions exist $g(n)$ and $\tilde{m}(n)$, that $f(ng(n)) \leq n^\delta$ and $f(\xi g(\tilde{m}^{-1}(n))) \geq n^\gamma$ for large n, and $0 < \delta < 1$, $0 < \xi < 1$, $\gamma > 2$, then $Pr^*(\mu^*_m > \xi\mu_n) = 0$ for sufficiently large n.*

Proof:

$$Pr\Big(\mu_n < g(n)\Big) = Pr\Big(\sum_{i=1}^{n} X_i < ng(n)\Big) \leq$$

$$Pr(max_{i=1..n} X_i < ng(n)) =$$

$$\mathcal{F}(ng(n))^n = \left(1 - \frac{1}{f(g(n))}\right)^n \leq$$

$$\left(1 - \frac{1}{n^\delta}\right)^n, \quad 0 < \delta < 1.$$

The last inequality holds for large n only, say $n > \tilde{n}$. For δ in $0 < \delta < 1$, $\sum_{n>\tilde{n}}^{\infty}(1 - 1/n^\delta)^n < \infty$. The sum $\sum_{n=1}^{\tilde{n}}(1 - 1/f(g(n)))^n$ is also finite, hence, Borel - Cantelli lemma implies that for sufficiently large n $Pr(\mu_n < g(n)) = 0$. Therefore, $Pr(\mu_n \geq g(n)) = 1$ if

n is sufficiently large.

$$Pr\Big(\mu_n > \xi g(\tilde{m}^{-1}(n))\Big) \leq Pr\Big(max_{i=1..n} X_i > \xi g(\tilde{m}^{-1}(n))\Big) \leq$$
$$1 - F(\xi g(\tilde{m}^{-1}(n)))^n = 1 - \left(1 - \frac{1}{f\big(\xi g(\tilde{m}^{-1}(n))\big)}\right)^n \leq$$
$$\frac{n}{f(\xi g(\tilde{m}^{-1}(n)))} \leq n^{1-\gamma}.$$

The last inequality is valid for sufficiently large n, so if $\gamma > 2$, the series is summable. Thus the Borel-Cantelli lemma implies that $Pr\big(\mu_n > \xi g(\tilde{m}^{-1}(n))\big) = 0$ as n is sufficiently large. Applying Glivenko-Cantelli lemma we obtain $Pr^*\big(\mu^*_m > \xi g(\tilde{m}^{-1}(m))\big) = 0$ as m is sufficiently large. Thus

$$Pr^*(\mu^*_m > \xi \mu_n) \leq Pr^*(\mu^*_m > \xi g(n))$$
$$= Pr^*\big(\mu^*_m > \xi g(\tilde{m}^{-1}(m))\big) = 0$$

□

for sufficiently large n.

Appendix B. Proof of theorem 3.5

Proof: Apply lemma A.1. In this case, function $f(n) = \underbrace{\log ... \log n}_{r} / C$ with a positive integer r. If we take $g(n) = \underbrace{(\exp ... \exp n^\delta)}_{r} / n$ the condition $f(ng(n)) \leq n^\delta$ for $0 < \delta < 1$ is satisfied with equality by construction. $\tilde{m}(n) \approx \log(n)$, hence, $\tilde{m}^{-1}(m) \approx \exp(m)$.

$$\lim_{n\to\infty} \frac{f(\xi g(\tilde{m}^{-1}(n)))}{n^\gamma} \approx$$
$$\lim_{n\to\infty} \frac{\overbrace{\log ... \log}^{r-1}(log\xi - \log n + \overbrace{\exp ... \exp}^{r} n^\delta)}{Cn^\gamma} = \infty.$$

The result is due to one extra exponent in the nominator, arising from $\tilde{m}^{-1} \approx \exp n$. Thus, the condition that $f\big(\xi g(\tilde{m}^{-1}(n))\big) \geq n^\gamma$, $\gamma > 2$ is satisfied for large n. Therefore, the theorem is proved. □