Entropic CLT for Order Statistics

Martina Cardone*, Alex Dytso*, Cynthia Rush†
* University of Minnesota, Minneapolis, MN 55404, USA, Email: mcardone@umn.edu
* New Jersey Institute of Technology, Newark, NJ 07102, USA Email: alex.dytso@njit.edu
† Columbia University, New York, NY 10025, USA, Email: cynthia.rush@columbia.edu

Abstract—It is well known that central order statistics exhibit a central limit behavior and converge to a Gaussian distribution as the sample size grows. This paper strengthens this known result by establishing an entropic version of the CLT that ensures a stronger mode of convergence using the relative entropy. In particular, an order $O(1/\sqrt{n})$ rate of convergence is established under mild conditions on the parent distribution of the sample generating the order statistics. To prove this result, ancillary results on order statistics are derived, which might be of independent interest.

I. INTRODUCTION

Consider a random sample $X_1, X_2, \ldots, X_n$ drawn independently from a parent distribution having cumulative distribution function (cdf) $F$ and probability density function (pdf) $f$. Let the random variables $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ denote the order statistics of the sample. With this notation, $X_{(1)}$ corresponds to the minimum value of the sample, $X_{(n)}$ corresponds to the maximum value of the sample, and $X_{\left(\frac{n+1}{2}\right)}$ (provided that $n$ is odd) corresponds to the median of the sample. Order statistics play an important role in statistical sciences; for example, order statistics are instrumental to constructing a class of robust estimators known as $L$-estimators. The interested reader is referred to [1], [2] for comprehensive summaries on the application and theory of order statistics.

In this paper, we study the asymptotic behavior of the distribution of the central order statistics, that is we are interested in the distribution of $X_{(p)}$ for a fixed $p \in (0, 1)$ as $n \to \infty$. Our main contribution is the proof of a strong form of the central limit theorem (CLT) showing that the relative entropy between the Gaussian distribution and the distribution of the order statistics converges to 0 as $n$ grows.

A. Prior Work

The study of the asymptotic distribution of order statistics has a long history. For example, in [3] Laplace (already in 1818!) established asymptotic normality of the sample median. The asymptotic normality of central order statistics was shown by Smirnov in [4] and, in [5]. Smirnov proved a general convergence theorem establishing that a non-degenerate asymptotic distribution of $X_{(n)}$ can be of only four different types, under the condition that $\sqrt{n}(X_{(n)} - p) \to 0$ for a fixed $p \in (0, 1)$. Further, in [6], [7], Balkema and Haan established that the distribution of order statistics is dense in the space of all distributions in the following sense: there exists a parent cdf such that the limiting cdf of $X_{(n)}$ is any desired cdf for an appropriately chosen sequence $r_n$. A succinct summary of these results can be found in [8, p.145].

The asymptotic distribution of the joint order statistics has been studied in a series of papers [9], [10], [11], [12]. Berry-Esseen type results [13] for the order statistics have been shown by Reiss in [14]. The convergence in the total variational distance has been studied in [12], [15], [16] and the law of the iterated logarithm for the order statistics has been shown in [17]. Asymptotic results for random order statistics (i.e., $X_{(\nu)}$ where $\nu$ is an integer valued random variable) have been considered in [18]. All of the aforementioned results assume that the parent distribution has an absolutely continuous cdf, while the asymptotic distribution of discrete random variables has recently been considered in [19].

The entropic CLT for the sample mean has been first studied by Linnik in [20] and later considerably generalized by Barron in [21]. There has also been some recent activity around finding the rates of convergence and the interested reader is referred to [22], [23], [24], [25], [26] and references therein.

Information measures on the distribution of order statistics have also received some attention [27], [28], [29], [30], [31], [29], [32], [33]. For example, in [34] we showed that the $f$-divergence between the joint distribution of order statistics and the product distribution of order statistics does not depend on the parent distribution.

B. Contributions and Outline

We establish an entropic version of the CLT for order statistics that ensures a strong mode of convergence in terms of relative entropy. In particular, in Section II we provide mild conditions to ensure that $D(X_{(np)}||G_{n,p}) = O(1/\sqrt{n})$, where $G_{n,p}$ is a Gaussian random variable and $D(\cdot||\cdot)$ denotes the relative entropy. In order to prove the CLT, in Section III we derive some ancillary results on order statistics, which might be of independent interest. For instance, we show a rather general bound on the moments of the order statistics in terms of properties of the parent distribution. Finally, in Section IV we conclude the paper with a discussion on the necessity of the derived conditions for convergence. In particular, we show that (although mild) some of the conditions might not be necessary, whereas others (or a variation of them) do appear to be needed to ensure convergence of the relative entropy. Some of the proofs can be found in the appendices.

C. Notation

Deterministic quantities are denoted by lower case letters and random variables are denoted by upper case letters (e.g.,
The inverse cdf, also known as the quantile function, of the random variable \( X \) characterized by noting that

\[
F^{-1}(p) = \inf \{ x \in \mathbb{R} : p \leq F(x) \}, \quad p \in (0, 1).
\]

The inverse cdf will play an important role in our analysis. In particular, we will often exploit the following well-known fact [35]:

\[
X_{(k)} = F^{-1}(U_{(k)}), \quad k \in \{1, \ldots, n\},
\]

where \( \overset{d}{=} \) denotes equality in distribution and \( U_{(1)}, \ldots, U_{(n)} \) are the order statistics of a sample drawn independently from a parent distribution that is uniform on \((0, 1)\). We will also frequently use the mean and variance of \( U_{(k)} \), which can be characterized by noting that \( U_{(k)} \sim \text{Beta}(k, n + 1 - k) \) [35]: hence, for \( k \in \{1, \ldots, n\},

\[
\mathbb{E}[U_{(k)}] = \frac{k}{n + 1}, \quad \text{and} \quad \text{Var}(U_{(k)}) = \frac{k(n + 1 - k)}{(n + 1)(n + 2)^2}.
\]

The differential entropy of a random variable \( V \) having pdf \( f_V \) is defined and denoted by

\[
h(V) = -\int_{\mathbb{R}} f_V(v) \log f_V(v) \, dv.
\]

The relative entropy between \( W \sim f_W \) and \( V \sim f_V \), respectively, is defined and denoted by

\[
D(W \| V) = \int_{\mathbb{R}} f_W(x) \log \frac{f_W(x)}{f_V(x)} \, dx.
\]

The inverse cdf, also known as the quantile function, of the random variable \( V \) is denoted as

\[
F^{-1}(p) = \inf \{ x \in \mathbb{R} : p \leq F(x) \}, \quad p \in (0, 1).
\]

Our main result is that the following theorem, which provides a stronger mode of convergence than the classical CLT for order statistics at the expense of extra sufficient conditions.

**Theorem 1.** Let \( X_1, \ldots, X_n \) be a sequence of order statistics generated i.i.d. according to parent cdf \( F \) with the pdf \( f \), and let \( g_{n,p} \sim N(\mu_p, \sigma_p^2) \) where \( \mu_p, \sigma_p \) are defined in \([7]\). Fix some \( p \in (0, 1) \) and assume that

1. \( \|f\|_m < \infty \) for some \( m \in [2, \infty) \);
2. \( f(F^{-1}(p)) \neq 0 \), and \( f(F^{-1}(p)) \) is continuous at \( t = p \);
3. \( \mathbb{E}[|X|^r] < \infty \) for some \( r > 0 \).

Then,

\[
D(X_{(np)} \| G_{n,p}) = O \left( \frac{1}{\sqrt{n}} \right).
\]

**Proof:** We start by recalling the following well known property of the differential entropy (see for example [36]): given a differentiable and bijection function \( g \), we have that

\[
h \left( g(V) \right) = h(V) + \mathbb{E} \log |g'(V)|.
\]

Next, let \( \Phi_{n,p} \) and \( \phi_{n,p} \) denote the cdf and the pdf of \( G_{n,p} \) (with mean and variance in \([7]\)), and define a function

\[
g(u) = \Phi_{n,p}(F^{-1}(u)), \quad u \in (0, 1).
\]

The derivative of \( g(u) \) is given by

\[
g'(u) = \frac{d}{du} \Phi_{n,p}(F^{-1}(u)) = \frac{\phi_{n,p}(F^{-1}(u))}{f(F^{-1}(u))}, \quad u \in (0, 1),
\]

which is strictly positive. Therefore, \( g \) is bijective.

Let \( U \) be the uniform random variable on \((0, 1)\). Then,

\[
D(X_{(np)} \| G_{n,p} \| U) \overset{(a)}{=} D \left( \Phi_{n,p}(X_{(np)}) \| U \right) \overset{(b)}{=} D \left( \Phi_{n,p} \left( U_{(np)} \right) \| U \right) \overset{(c)}{=} -h \left( \Phi_{n,p} \left( X_{(np)} \right) \right) \overset{(d)}{=} -h \left( \Phi_{n,p}(F^{-1}(U_{(np)})) \right) \overset{(e)}{=} -h(U_{(np)}) - \mathbb{E} \log |g'(U_{(np)})|,
\]

where the labeled equalities follow because: (a) \( D(W \| U) = D(f(W) \| f(U)) \) for any invertible function \( f \); (b) \( F_X(X) \overset{d}{=} U \) for any random variable \( X \) with cdf \( F_X \); (c) \( D(X \| U) = -h(X) \) for any random variable \( X \in (0, 1) \); (d) \( \overset{(d)}{=} \), and (e) \( \overset{(e)}{=} \) and the definition of \( g \) in \([10]\), which is bijective.

Now we study the terms on the right-hand side of \((12)\). In particular, we start focusing on the term \( \mathbb{E} \log |g'(U_{(np)})| \). First, notice that since \( \phi_{n,p}(\cdot) \) is the Gaussian pdf having mean and variance given in \([7]\), we have

\[
\phi_{n,p}(F^{-1}(u)) = \exp \left( -\frac{1}{2\sigma_{n,p}} \left( F^{-1}(u) - F^{-1}(p) \right)^2 \right) \sqrt{\frac{1}{2\pi\sigma_{n,p}}},
\]

Therefore, using \((11)\) and the above, we obtain

\[
\mathbb{E} \log |g'(U_{(np)})|.
\]
\[ E \left[ \log \left( \phi_{n,p}(F^{-1}(U_{(np)})) \right) \right] = \frac{1}{2} \log (2\pi V_{n,p}) - \frac{1}{2V_{n,p}} E \left[ (F^{-1}(U_{(np)}) - F^{-1}(p))^2 \right] \]

\[ = -\frac{1}{2} \log (2\pi V_{n,p}) - \frac{1}{2V_{n,p}} E \left[ (F^{-1}(U_{(np)}) - F^{-1}(p))^2 \right] \]

\[ - \mathbb{E} \left[ \log \left( \frac{f(F^{-1}(U_{(np)}))}{f(F^{-1}(U_{(np)}))} \right) \right] \]

\[ + \frac{1}{2} - \frac{1}{2V_{n,p}} \mathbb{E} \left[ (F^{-1}(U_{(np)}) - F^{-1}(p))^2 \right], \] (13)

where the last equality follows by using the expression of \( V_{n,p} \) in (7). Now, combining (12) and (13), we find that

\[ K \]

the parent distribution. In Section III, we will present ancillary

\[ D(X_{(np)}\|G_{n,p}) = K_1 + K_2 + K_3, \] (14a)

where

\[ K_1 = \frac{1}{2} \log \left( \frac{2\pi e(p(1-p))}{n} \right) - h(U_{(np)}), \] (14b)

\[ K_2 = \frac{1}{2V_{n,p}} E \left[ (F^{-1}(U_{(np)}) - F^{-1}(p))^2 \right] - \frac{1}{2}, \] (14c)

\[ K_3 = \mathbb{E} \left[ \log \left( \frac{f(F^{-1}(U_{(np)}))}{f(F^{-1}(p))} \right) \right]. \] (14d)

It is worth noting that \( K_1 \) is independent of the parent distribution, whereas the terms \( K_2 \) and \( K_3 \) both depend on the parent distribution. In Section III, we will present ancillary results that show the following facts: (i) \( K_1 = O(1/n) \) by Lemma 2; (ii) \( K_2 = O(1/\sqrt{n}) \) by Corollary 1 (this requires Assumptions (2) and (3)); and (iii) \( K_3 = O(1/\sqrt{n}) \) by Lemma 3 (this requires Assumptions (1) and (2)). Combining these facts together with (14) and the non-negativity property of the relative entropy, we have that

\[ D(X_{(np)}\|G_{n,p}) = O(1/\sqrt{n}). \]

This concludes the proof of Theorem 1.

We conclude this section by highlighting that a variety of different distributions (e.g., uniform, Gaussian, exponential) satisfy the conditions of Theorem 1. Moreover, it is also interesting to note that the Cauchy distribution satisfies the conditions despite the fact that in this case \( \mathbb{E}[|X|] = \infty \); hence, the CLT for the sample mean does not hold. This can be seen since the Cauchy pdf clearly satisfies the first two conditions and \( \mathbb{E}[|X|^r] < \infty \) for \( 0 < r < 1 \).

III. ANCILLARY RESULTS

In this section we present auxiliary results needed for the proof of Theorem 1; some of which may be of interest on their own.

A. Entropy of Uniform Order Statistics

We provide the exact expression for the differential entropy of \( U_{(k)} \) for \( k \in \{1, \ldots, n\} \), and an asymptotic expression for the entropy of \( U_{(np)} \) for \( p \in (0,1) \) as \( n \to \infty \). The proof of Lemma 2 can be found in Appendix A.

Lemma 2. For any \( k \in \{1,2,\ldots,n\} \),

\[ h(U_{(k)}) = T_{k-1} + T_{n-k} - T_n - H_n, \] (15)

where for \( r \in \mathbb{N} \), with \( \mathbb{N} \) denoting the natural numbers,

\[ H_r = \sum_{k=1}^{r} \frac{1}{k}, \quad \text{and} \quad T_r = \log(r!) - rH_r. \] (16)

Moreover, if \( k = pn \) for \( p \in (0,1) \), then

\[ h(U_{(np)}) = \frac{1}{2} \log \left( \frac{2\pi e(p(1-p))}{n} \right) \]

\[ = \frac{1}{p} + \frac{1-p}{2} - 4 \frac{1}{12n^2} + O \left( \frac{1}{n^3} \right). \] (17)

It is interesting to note that in Lemma 2, the rate of convergence is \( O(1/n^2) \) instead of \( O(1/n) \) for \( p = 1/2 \).

B. Bound on the Estimation Error of the \( p \)-th Quantile

In practice, one might desire to estimate the \( p \)-th quantile \( F^{-1}(p) \) of an unknown cdf \( F \). The order statistic \( X_{(np)} \) based on an i.i.d. sample from the parent cdf \( F \) is a natural estimate for \( F^{-1}(p) \). It is well known that this estimator is consistent as \( n \to \infty \). The next result (see Appendix A for the proof) provides an upper bound on the mean squared error of estimating \( F^{-1}(p) \) with \( X_{(np)} \).

Lemma 3. Fix some \( p \in (0,1) \), and assume a pdf \( f \) such that \( f(F^{-1}(p)) \neq 0 \), and \( f'(F^{-1}(t)) \) is continuous at \( t = p \). Then, for any \( \epsilon \in (\frac{1}{n+1},p) \), we have that

\[ \mathbb{E} \left[ (F^{-1}(U_{(np)}) - F^{-1}(p))^2 \right] \]

\[ \leq 4 \sqrt{\mathbb{E} \left[ (X_{(np)})^4 \right] e^{-n(1-p)(\epsilon\pi r)^2}} + \frac{\text{Var}(U_{(np)})}{(f(F^{-1}(p)))^2} \]

\[ + C_{p,\epsilon} O \left( \frac{1}{n^2} \right), \] (18a)

where

\[ C_{p,\epsilon} = \max_{t \in [p - \epsilon, p + \epsilon]} \left| \frac{f'(F^{-1}(t))}{(f(F^{-1}(t)))^2} \right|. \] (18b)

Remark 1. Note that if the parent distribution is the uniform on \( (0,1) \), then \( F^{-1}(x) = x \) for \( x \in (0,1) \) and

\[ \mathbb{E}[(F^{-1}(U_{(np)}) - F^{-1}(p))^2] = \mathbb{E} \left[ (U_{(np)} - p)^2 \right] \]

\[ = \text{Var}(U_{(np)}) + (\mathbb{E}[U_{(np)}] - p)^2 \]

\[ = \text{Var}(U_{(np)}) + \frac{p^2}{(n+1)^2}, \]

where we have used that \( \mathbb{E}[U_{(np)}] = \frac{np}{n+1} \) (see 5) so that \( (\mathbb{E}[U_{(np)}] - p)^2 = (\frac{np}{n+1} - p)^2 \). In other words, there exists a distribution for which the bound in Lemma 2 is asymptotically tight as \( n \to \infty \).
C. New Bound on the Moments of Order Statistics

In order to control the error in Lemma 3, we need to control the fourth moment of the order statistics. Ideally, we would like to control this in terms of the properties of the parent distribution (e.g., moments). The next result, the proof of which is in Appendix D, establishes a rather general bound on the moments of the order statistics in terms of properties of the parent distribution.

Lemma 4. Let $X_1, X_2, \ldots, X_n$ be an i.i.d. random sample. For any $q, r > 0$ and $k \in \{1, 2, \ldots, n\}$, we have that

$$\mathbb{E}[|X_{(k)}|^r] \leq C_{n,k,q,r} (\mathbb{E}[|X|^r])^{\frac{k}{r}},$$

where

$$C_{n,k,q,r} = \begin{cases} \frac{\Gamma(n+1)\Gamma(k-\frac{r}{2})\Gamma(n-k+\frac{r}{2}+1)}{\Gamma(n-k+1)\Gamma(k)\Gamma(n-k+1)} & \text{if } k > \frac{r}{2}, \ n-k > \frac{r}{2}-1, \\
\infty & \text{otherwise.} \end{cases}$$

In addition,

$$\lim_{n \to \infty} C_{n,p,n,q,r} = (p-1)(p) \frac{\hat{p}}{r}. \quad (19)$$

Notice that Lemma 4 shows that all of the moments of the order statistics exist, provided that a single moment of the parent distribution, for any order $r > 0$, exists.

We now conclude this subsection with the following corollary, which is an immediate consequence of Lemma 3 and Lemma 4 and is used in the proof of Theorem 1 to show that $K_2 = O(1/\sqrt{n})$.

Corollary 5. Fix some $p \in (0, 1)$. Suppose that $f(F^{-1}(p)) \neq 0$, $f'(F^{-1}(t))$ is continuous at $t = p$, and $\mathbb{E}[|X|^r] < \infty$ for some $r > 0$. Then, for $V_{n,p}$ defined in (7),

$$\frac{1}{2V_{n,p}} \mathbb{E} \left[ (F^{-1}(U_{(np)}) - F^{-1}(p))^2 \right] - \frac{1}{2} = O \left( \frac{1}{\sqrt{n}} \right).$$

Proof: We have that

$$\frac{1}{2V_{n,p}} \mathbb{E} \left[ (F^{-1}(U_{(np)}) - F^{-1}(p))^2 \right] - \frac{1}{2} \leq \frac{\text{Var}(U_{(np)})}{2V_{n,p} (f(F^{-1}(p)))^2} - \frac{1}{2} + O \left( \frac{1}{\sqrt{n}} \right)$$

$$\leq \frac{n^2 (n+1-np)}{2(1-p)(n+1)(n+2)^2} - \frac{1}{2} + O \left( \frac{1}{\sqrt{n}} \right)$$

$$\leq \frac{n+1-np}{2(1-p)n} - \frac{1}{2} + O \left( \frac{1}{\sqrt{n}} \right)$$

$$= \frac{1}{2(1-p)n} + O \left( \frac{1}{\sqrt{n}} \right),$$

where (a) follows by using Lemma 3 and Lemma 4 since

$$2\sqrt{C_{n,p,n,q,r}} V_{n,p} (\mathbb{E}[|X|^r])^{\frac{k}{r}} e^{-n+2} \left(1 - \frac{1}{n^r} \right)^2 + \frac{C_{p,x}}{2V_{n,p}} O \left( \frac{1}{n^2} \right)$$

is $O(1/\sqrt{n})$ and (b) follows from (5) and (7).

D. Bound on $K_3$ in (14)

We now conclude this section with the following lemma, the proof of which is given in Appendix E. In particular, Lemma 6 demonstrates that $K_3 = O(1/\sqrt{n})$.

Lemma 6. Fix some $p \in (0, 1)$. Choose some $q, r \in [1, \infty]$ such that $\frac{1}{q} + \frac{1}{r} = 1$ and $\epsilon > 0$ such that

$$p > \epsilon > \max \left\{ \frac{p}{n+1}, \frac{|q-2|p-q+1}{q(n-1)+2} \right\}.$$  

Then,

$$\mathbb{E} \left[ \log \left( \frac{f(F^{-1}(U_{(np)}))}{f(F^{-1}(p))} \right) \right]$$

$$\leq 2|\log(f(F^{-1}(p)))| |e^{-2(n+2)}(\epsilon^{\frac{1}{2}})^2 + C_{p,x}^2 O \left( \frac{1}{\sqrt{n}} \right)$$

$$+ 2C_q (\|f\|_r + 1) \pi^{-\frac{1}{2}} q^{-\frac{1}{2} - 1} q^{-\frac{1}{r}}$$

$$\leq C_{p,x}^2 \max_{p-e^\epsilon \leq u \leq p+e^\epsilon} \left| \frac{f'(F^{-1}(u))}{f(F^{-1}(u))} \right|.$$  

IV. Discussion and Conclusion

In this paper, we have derived an entropic version of the CLT for the order statistics that ensures a stronger mode of convergence (in terms of relative entropy) than the convergence in distribution provided by the classical CLT for continuous central order statistics.

The three sufficient conditions in Theorem 1 are fairly mild and we discuss them now in more detail. We suspect that condition 3) (or a variation of it) in Theorem 1 might be needed for convergence. To support this claim, consider the following density

$$f_1(x) = \frac{2}{x \log^4(x)}, \quad x \in (e, \infty), \quad (20)$$

which is a canonical example of a pdf such that $\mathbb{E}[|X|^r] = \infty$ for all $r > 0$ (i.e., $f_1$ does not satisfy the third condition in Theorem 1). In Appendix H we indeed show that for this density $K_2 = O(1)$; hence, $D(X_{np})/|G_{np}| = \infty$.

Unlike condition 3), condition 1) in Theorem 1 might not be needed, as we argue next. Consider the following density

$$f_2(x) = \frac{1}{x \log^3(x)}, \quad x \in (0, e^{-1}), \quad (21)$$

which is a canonical example of a pdf such that $\|f_2\|_m = \infty$ for all $m > 1$ (i.e., $f_2$ does not satisfy the first condition in Theorem 1). Although this density does not satisfy the first condition in Theorem 1 in Appendix H we show that

$$D(X_{np})/|G_{np}| = \Theta \left( \frac{1}{n} \right),$$

that is, the claim of Theorem 1 still holds. This example shows that, even if quite mild, condition 1) in Theorem 1 might not be needed for convergence. Therefore, an interesting research direction would consist of further relaxing this condition.
APPENDIX A

PROOF OF LEMMA 2

We first recall that by (15), the random variable $U(k)$ is distributed as Beta($k, n+1-k$); hence, it has pdf

$$f_{U(k)}(x) = c_k x^{k-1}(1-x)^{n-k}, \text{ for } x \in [0, 1],$$

where $c_k = \frac{n!}{(k-1)!(n-k)!}$. Then, using the definition of differential entropy in (1), we obtain

$$h(U(k)) = \mathbb{E} \left[ \log \left( c_k U^{k-1}(1-U)^{n-k} \right) \right]$$

$$= \log(c_k) + (k-1)\mathbb{E}[\log(U_k)] + (n-k)\mathbb{E}[\log(1-U_k)].$$

(23)

Now, consider the two expectations on the right side of (23), namely $\mathbb{E}[\log(U(k))]$ and $\mathbb{E}[\log(1-U(k))]$. Notice that $U(k) \sim \text{Beta}(k, n+1-k)$, we also have $1-U(k) \sim \text{Beta}(n+1-k, k)$. Therefore, (see, e.g., (35)),

$$\mathbb{E}[\log(U(k))] = \psi(k) - \psi(n+1),$$

$$\mathbb{E}[\log(1-U(k))] = \psi(n+1-k) - \psi(n+1),$$

where $\psi(\cdot)$ is the digamma function. Next, we leverage the relationship between the digamma function and the harmonic number, $H_r$ defined in (16), namely $\psi(k) = H_{k-1} - \gamma$, where $\gamma$ is the Euler-Mascheroni constant. Putting this all together in (23), we have that

$$h(U(k)) = \log(c_k) + (k-1)(\psi(k) - \psi(n+1)) + (n-k)(\psi(n+1-k) - \psi(n+1))$$

$$= \log(c_k) + (k-1)(H_{k-1} - H_n) + (n-k)(H_{n-k} - H_n)$$

Finally, using the fact that $c_k = \frac{n!}{(k-1)!(n-k)!}$, and the definition of $T_j = \log(j!) - jH_j$ in (16), we have that

$$h(U(k)) = \log(n!) - \log((k-1)!) - \log((n-k)!)$$

$$+ (k-1)H_{k-1} + (n-k)H_{n-k} - (n+1)H_n$$

$$= T_n - T_{k-1} - T_{n-k} + H_n,$$

(24)

This shows (15). To show (17), let $k < n$ and we approximate the expression in (24) by leveraging the following expressions for the series expansion of the harmonic number and the log factorial (37),

$$H_k = \log(k) + \gamma + \frac{1}{2k} - \frac{1}{12k^2} + \frac{1}{120k^4} - O \left( \frac{1}{k^6} \right),$$

$$\log(k!) = k \log(k) - k + \frac{1}{2} \log(2\pi k) + \frac{1}{12k} - \frac{1}{360k^3} + O \left( \frac{1}{k^5} \right).$$

(25)

Combining (25) and (26), we arrive at

$$T_k = \frac{1}{2} \log \left( \frac{2\pi k}{e} \right) - (1+\gamma)k + \frac{1}{6k} - \frac{1}{90k^3} + O \left( \frac{1}{k^5} \right).$$

(27)

Now, using the expression of $T_k$ in (27) inside (24), and collecting similar terms we find

$$h(U(k)) = T_{k-1} + T_{n-k} - T_n - H_n$$

$$= \frac{1}{2} \log \left( \frac{2\pi(k-1)(n-k)}{n^3} \right) + \frac{1}{6} \left( \frac{1}{k-1} + \frac{1}{n-k} - \frac{4}{n} \right)$$

$$- \frac{1}{90} \left( \frac{1}{(k-1)^3} + \frac{1}{(n-k)^3} - \frac{1}{n^3} \right) + \frac{1}{12n^2} \left( 1 - \frac{10n^2}{40} \right)$$

$$+ O \left( \frac{1}{1-k} \right) + O \left( \frac{1}{n-k} \right) - O \left( \frac{1}{n^3} \right) + O \left( \frac{1}{n^5} \right).$$

Now, letting $k = np$ and keeping only the dominant terms in the expression above, we arrive at

$$h(U(np)) = \frac{1}{2} \log \left( \frac{2\pi(p-\frac{1}{2})(1-p)}{n} \right) + \frac{1}{12n^2}$$

$$+ \frac{1}{6n} \left( \frac{1}{p} + \frac{1}{1-p} - 4 \right) + O \left( \frac{1}{n^3} \right).$$

This concludes the proof of (17) and also of Lemma 2.

APPENDIX B

PROOF OF LEMMA 3

Recall that by assumption, $\frac{p}{n+1} < \epsilon < p$. Now, define the event

$$A = \{ p - \epsilon \leq U(np) \leq p + \epsilon \},$$

and let $1_A$ be the indicator function of the event $A$ and $1_{\overline{A}}$ be the indicator function of the complementary event. Now, using the law of total expectation we can write

$$\mathbb{E} \left[ \left( F^{-1}(U(np)) - F^{-1}(p) \right)^2 \right]$$

$$= \mathbb{E} \left[ \left( F^{-1}(U(np)) - F^{-1}(p) \right)^2 1_A \right]$$

$$+ \mathbb{E} \left[ \left( F^{-1}(U(np)) - F^{-1}(p) \right)^2 1_{\overline{A}} \right].$$

(29)

We will now analyze the two expectations on the right side of (29) separately.

**First expectation in (29).** We obtain

$$\mathbb{E} \left[ \left( F^{-1}(U(np)) - F^{-1}(p) \right)^2 1_A \right]$$

$$\leq 2 \mathbb{E} \left[ \left( F^{-1}(U(np)) \right)^2 \right] 1_A$$

$$+ 2 \mathbb{E} \left[ \left( F^{-1}(p) \right)^2 \right] 1_{\overline{A}}$$

$$\leq 2 \sqrt{\mathbb{E} \left[ \left( X(np) \right)^4 \right]} \mathbb{P}(A) + 2 \left( F^{-1}(p) \right)^2 \mathbb{P}(A^c)$$

$$\leq 2 \sqrt{\mathbb{E} \left[ \left( X(np) \right)^4 \right]} \left( F^{-1}(p) \right)^2 \sqrt{\mathbb{P}(A)}$$

$$\leq 4 \left( \sqrt{\mathbb{E} \left[ \left( X(np) \right)^4 \right]} + \left( F^{-1}(p) \right)^2 \right) e^{-(n+2)(\epsilon - \frac{p}{n+1})^2},$$

(30)

where the labeled inequalities follow from: (a) the fact that $(n-b)^2 \leq 2(a^2 + b^2)$; (b) the Cauchy-Schwarz inequality and the fact $X(np) \leq F^{-1}(U(np))$; (c) the fact that $x \leq \sqrt{x}$ for $x \in (0, 1)$; and (d) Lemma 7 below, which is proved in Appendix C.

**Lemma 7.** Let $A = \{ p - \epsilon \leq U(np) \leq p + \epsilon \}$ for $p \in (0, 1)$ and $\frac{p}{n+1} < \epsilon < p$. Then,

$$\mathbb{P}(A^c) \leq 2 \exp \left( -2(n+2) \left( \epsilon - \frac{p}{n+1} \right)^2 \right).$$

(31)
Second expectation in (29). From the definition of $A$ in (28), we know that $U_{(np)} \in [p - \epsilon, p + \epsilon]$; hence, using the Taylor remainder theorem, for any $t \in [p - \epsilon, p + \epsilon]$, where $\tilde{u}$ is some value between $p$ and $t$, we have

$$F^{-1}(t) = F^{-1}(p) + \frac{d}{du}F^{-1}(u)|_{u=p}(t-p)$$

$$+ \frac{d^2}{du^2} F^{-1}(u)|_{u=\tilde{u}} \frac{(t-p)^2}{2}$$

$$= F^{-1}(p) + g^{(1)}(p)(t-p) + \frac{1}{2} g^{(2)}(\tilde{u})(t-p)^2, \tag{31}$$

where in the last equality we let $g^{(1)}(p) := \frac{d}{du} F^{-1}(u)|_{u=p}$ and $g^{(2)}(\tilde{u}) := \frac{d^2}{du^2} F^{-1}(u)|_{u=\tilde{u}}$.

We next provide an upper bound for $g^{(2)}(\tilde{u})$. Define,

$$C_{p,\epsilon} = \max_{t \in [p-\epsilon, p+\epsilon]} |g^{(2)}(t)| < \infty, \tag{32}$$

where the last inequality follows since

$$g^{(2)}(t) = \frac{d^2}{du^2} F^{-1}(u)|_{u=t} = \frac{d}{du} \left( \frac{1}{f(F^{-1}(u))} \right)|_{u=t} = \frac{f'(F^{-1}(t))}{(f(F^{-1}(t)))^3}. \tag{33}$$

Therefore, for sufficiently small $\epsilon > 0$, we have that $C_{p,\epsilon} < \infty$ if $f(F^{-1}(p)) \neq 0$, and $f'(F^{-1}(t))$ is continuous at $t = p$; these two conditions are guaranteed by the assumptions of Lemma 3. Thus, plugging this into (31),

$$\left( F^{-1}(t) - F^{-1}(p) \right)^2 \leq \left( g^{(1)}(p) \right)^2 (t-p)^2 + \frac{C_{p,\epsilon}^2}{4} (t-p)^4$$

$$+ C_{p,\epsilon} \left| g^{(1)}(p) (t-p)^3 \right|. \tag{34}$$

Now using (33),

$$\begin{align*}
\mathbb{E} \left[ (F^{-1}(U_{(np)}) - F^{-1}(p))^2 \right] & \leq \left( g^{(1)}(p) \right)^2 \mathbb{E} \left[ (U_{(np)} - p)^2 \right] + \frac{C_{p,\epsilon}^2}{4} \mathbb{E} \left[ (U_{(np)} - p)^4 \right] \\
& \leq C_{p,\epsilon} \left| g^{(1)}(p) \mathbb{E} \left[ (U_{(np)} - p)^2 \right] \right|. \tag{35}
\end{align*}$$

Next, we individually upper bound each term on the right side of (34).

First term in (34). We have

$$\begin{align*}
\mathbb{E} \left[ (U_{(np)} - p)^2 \right] & \leq \mathbb{E} \left[ (U_{(np)} - p)^2 \right] \\
& = \text{Var}(U_{(np)}) + \frac{p^2}{(n+1)^2}, \tag{36}
\end{align*}$$

where the equality follows since

$$\begin{align*}
\mathbb{E}[(U_{(np)} - p)^2] &= \mathbb{E}[(U_{(np)} - \mathbb{E}[U_{(np)}] + \mathbb{E}[U_{(np)}] - p)^2] \\
& = \text{Var}(U_{(np)}) + \left( \mathbb{E}[U_{(np)}] - p \right)^2,
\end{align*}$$

where $\mathbb{E}[U_{(np)}] = \frac{np}{n+1}$, so $\mathbb{E}[U_{(np)} - p]^2 = \left( \frac{p}{n+1} \right)^2$.

From (34), the first term on the right side of (34) can be upper bounded as

$$\left( g^{(1)}(p) \right)^2 \mathbb{E} \left[ (U_{(np)} - p)^2 \right] \leq \left( g^{(1)}(p) \right)^2 \left( \text{Var}(U_{(np)}) + \frac{p^2}{(n+1)^2} \right). \tag{37}$$

Second term in (34). We obtain

$$\begin{align*}
\mathbb{E} \left[ (U_{(np)} - p)^4 \right] & \leq \mathbb{E} \left[ (U_{(np)} - \frac{np}{n+1} + \frac{np}{n+1} - p)^4 \right] \\
& \leq 2C^2_{p,\epsilon} \left( \frac{p^2}{(n+1)^2} + \mathbb{E} \left[ (U_{(np)} - \frac{np}{n+1} - p)^4 \right] \right) \\
& \leq 2C^2_{p,\epsilon} \left( \frac{p^2}{(n+1)^2} + \mathbb{E} \left[ (U_{(np)} - \frac{np}{n+1} - p)^4 \right] \right). \tag{38}
\end{align*}$$

where the labeled inequalities follow from: (a) the fact that $(a+b)^4 \leq 8(a^4+b^4)$, and (b) the fact that if $W \sim \text{Beta}(\alpha, \beta)$, then (see, for example [38])

$$\mathbb{E}[(W - \mathbb{E}[W])^4] = \frac{3(\alpha^2\beta^2 + 2\alpha^2\beta + \alpha\beta^3 - 2\alpha\beta^2 + 2\beta^3)}{\alpha^3(\alpha + \beta + 1)(\alpha + \beta + 2)(\alpha + \beta + 3)}; \tag{39}$$

and, in particular, for $U_{(np)} \sim \text{Beta}(np, n+1-np)$, by setting $\alpha = np$ and $\beta = n - np + 1$ in (38)

$$\mathbb{E} \left[ (U_{(np)} - \frac{np}{n+1} - p)^4 \right] = \Theta \left( \frac{1}{n^2} \right). \tag{40}$$

Third term in (34). This term can be upper bounded using Jensen’s inequality (as $\mathbb{E}[\| Y \|^3] = \mathbb{E}[(\| Y \|^4)^{3/4}] \leq (\mathbb{E}[\| Y \|^4])^{3/4}$ and the same steps as we used to get the bound in (37). Indeed,

$$\begin{align*}
\mathbb{E} \left[ g^{(1)}(p) (U_{(np)} - p)^3 \right] & \leq C_{p,\epsilon} \left| g^{(1)}(p) \right| \mathbb{E} \left[ (U_{(np)} - p)^3 \right] \\
& \leq C_{p,\epsilon} \left| g^{(1)}(p) \right| \mathbb{E} \left[ (U_{(np)} - p)^3 \right] \tag{41}
\end{align*}$$

Collecting (34) together with the bounds in (36), (37), and (39).

$$\mathbb{E} \left[ (F^{-1}(U_{(np)}) - F^{-1}(p))^2 \right] \leq \left( g^{(1)}(p) \right)^2 \text{Var}(U_{(np)}) + C_{p,\epsilon} \left( \frac{1}{n^2} \right). \tag{42}$$

Therefore, plugging (30) and (40) in (29),

$$\begin{align*}
\mathbb{E} \left[ (F^{-1}(U_{(np)}) - F^{-1}(p))^2 \right] & \leq \left( g^{(1)}(p) \right)^2 \text{Var}(U_{(np)}) + C_{p,\epsilon} \left( \frac{1}{n^2} \right) \\
& \leq \left( g^{(1)}(p) \right)^2 \left( \text{Var}(U_{(np)}) + \frac{p^2}{(n+1)^2} \right).
\end{align*}$$
where in the last equality we have used the fact that \( g^{(1)}(p) = \frac{1}{f(F^{-1}(p))} \). This concludes the proof of Lemma 5.

### APPENDIX C

#### PROOF OF LEMMA 7

Using the definition of \( A \) in (28),

\[
\mathbb{P}(A^c) = \mathbb{P}[U_{(np)} \leq p - \epsilon] + \mathbb{P}[U_{(np)} \geq p + \epsilon] = \mathbb{P}[U_{(n(1-p)+1)} \geq 1-p+\epsilon] + \mathbb{P}[U_{(np)} \geq p+\epsilon],
\]

where in the last equality we have used the fact that \( U_{(np)} \equiv 1 - U_{(n(1-p)+1)} \) where \( U_{(n(1-p)+1)} \sim \text{Beta}(n+1-np, np) \) since \( U_{(np)} \sim \text{Beta}(np, n+1-np) \).

To upper bound \( \mathbb{P}(A^c) \) in (41), we leverage the fact that Beta-distributed random variables \( U_{(np)} \) and \( U_{(n(1-p)+1)} \) are \( \sigma_0 \)-sub-Gaussian with \( \sigma_0 < \frac{1}{4(n+1)} \) (see [39] Thm. 1). Recall (see, for example, [40]), that a random variable \( X \) is sub-Gaussian with parameter \( \sigma_0 \) if for every \( \lambda \in \mathbb{R} \),

\[
\mathbb{E}[e^{\lambda(X-\mathbb{E}[X])}] \leq e^{\lambda^2 \sigma_0/2}.
\]

Moreover, if \( X \) is \( \sigma_0 \)-sub-Gaussian, then for all \( t \geq 0 \),

\[
\mathbb{P}(X - \mathbb{E}[X] \geq t) \leq e^{-\frac{t^2}{2\sigma_0}.
\]

Now, let us consider the first term on the right side of (41).

Using that \( \mathbb{E}[U_{(n(1-p)+1)}] = \frac{n+1-np}{n+1} \),

\[
\mathbb{P}[U_{(n(1-p)+1)} \geq 1 - p + \epsilon] = \mathbb{P}
\left[
U_{(n(1-p)+1)} - \mathbb{E}[U_{(n(1-p)+1)}] \geq 1 - p + \epsilon - \frac{p}{n+1}
\right]
\]

\[
\leq \exp\left(-\frac{1}{2\sigma_0} \left( \epsilon - \frac{p}{n+1} \right)^2 \right)
\]

\[
\leq \exp\left(-2(n+2) \left( \epsilon - \frac{p}{n+1} \right)^2 \right),
\]

where (a) follows from the bound in (43), and (b) is due to the fact that \( \sigma_0 < \frac{1}{4(n+1)} \).

Similarly, for the second term on the right side of (41), using that \( \mathbb{E}[U_{(np)}] = \frac{np}{n+1} \), we obtain

\[
\mathbb{P}[U_{(np)} \geq p + \epsilon] = \mathbb{P}
\left[
U_{(np)} - \mathbb{E}[U_{(np)}] \geq \frac{p}{n+1} + \epsilon
\right]
\]

\[
\leq \exp\left(-2(n+2) \left( \frac{p}{n+1} + \epsilon \right)^2 \right)
\]

\[
\leq \exp\left(-2(n+2) \epsilon^2 \right),
\]

where, again, (a) follows from the bound in (43), and the fact that \( \sigma_0 < \frac{1}{4(n+2)} \).

Now, by summing (44) and (45), we wind that \( \mathbb{P}(A^c) \) in (41) can be upper bounded as

\[
\mathbb{P}(A^c) \leq \exp\left(-2(n+2) \left( \epsilon - \frac{p}{n+1} \right)^2 \right) + \exp\left(-2(n+2) \epsilon^2 \right)
\]

\[
\leq 2 \exp\left(-2(n+2) \left( \epsilon - \frac{p}{n+1} \right)^2 \right),
\]

concluding the proof of Lemma 7.

### APPENDIX D

#### PROOF OF LEMMA 4

Recall from (41) that \( X_{(k)} \equiv F^{-1}(U_{(k)}) \), where \( U_{(k)} \) is the order statistic of a sample of uniform random variables. Thus,

\[
\mathbb{E}[|X_{(k)}|^q] = \mathbb{E}[|F^{-1}(U_{(k)})|^q] \leq (\mathbb{E}[|X'|^q])^{\frac{q}{2}} \mathbb{E}\left[\frac{1}{\min(U_{(k)}, 1-U_{(k)})^q}\right], \tag{46}
\]

where the inequality follows from Lemma 8 given below and proved in Appendix E.

#### Lemma 8

Let \( F \) be the cdf of the random variable \( X \). Then, for any \( r > 0 \),

\[
|F^{-1}(u)| \leq \sqrt{\frac{\mathbb{E}[|X'|^r]}{\min(u, 1-u)^r}}, \quad u \in (0, 1).
\]

Next, we upper bound the right side of (46) using the following:

\[
\mathbb{E}\left[\left(\frac{1}{\min(U_{(k)}, 1-U_{(k)})^q}\right)^{\frac{q}{2}}\right] \leq \mathbb{E}\left[\left(\frac{1}{U_{(k)}(1-U_{(k)})}\right)^{\frac{q}{2}}\right] \leq C_{n,k,q,r}. \tag{48}
\]

In the above, the inequalities follow from the facts that (a) \( \min(a, b) \geq \frac{ab}{a+b} \) and (b) \( U_{(k)} \sim \text{Beta}(k, n+1-k) \); hence,

\[
\mathbb{E}\left[\left(\frac{1}{U_{(k)}(1-U_{(k)})}\right)^{\frac{q}{2}}\right] = \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n+1-k)} \int_0^1 t^{k-1} (1-t)^{n-k-\frac{q}{2}} \text{dt} = \frac{\Gamma(k+\frac{q}{2})\Gamma(n-k+\frac{q}{2}+1)}{\Gamma(n+\frac{q}{2}+1)} \begin{cases} k > \frac{q}{r} & \text{and} \ n+1-k > \frac{q}{r} \\ \infty & \text{else} \end{cases} = C_{n,k,q,r}.
\]

Plugging (48) into (46), we have shown

\[
\mathbb{E}[|X_{(k)}|^q] \leq C_{n,k,q,r} \left(\mathbb{E}[|X'|]^r\right)^{\frac{q}{2}},
\]

as desired.

Finally, we notice that if \( q \) and \( r \) are fixed with \( k = np \) for fixed \( p \), then as \( n \to \infty \), we have \( k = np > \frac{q}{r} \) and
\[ n + 1 - k = n(1 - p) + 1 > \frac{2}{\beta}, \text{ eventually}. \] Then, by [57 eq. 6.1.46], for \( a, b \in \mathbb{R} \),
\[
\lim_{x \to \infty} \frac{\Gamma(x + a)}{\Gamma(x + b) x^{b-a}} = 1,
\]
which leads to the conclusion that
\[
\lim_{n \to \infty} C_{n,pn,q,r}
= \lim_{n \to \infty} \frac{\Gamma(n + 1)}{\Gamma(n - \frac{2}{\beta} + 1)} \frac{(pn - \frac{2}{\beta})}{\Gamma(pn)} \frac{\Gamma(n(1-p) - \frac{2}{\beta} + 1)}{\Gamma(n(1-p) + 1)}
= \lim_{n \to \infty} n^{-\frac{2}{\beta}} (pn - \frac{2}{\beta}) \Gamma(n(1-p) - \frac{2}{\beta} + 1) \Gamma(n(1-p) + 1)
= (p(1-p))^{\frac{2}{\beta}}.
\]
This concludes the proof of Lemma 4.

APPENDIX E
PROOF OF LEMMA 8
First, recall that for a uniform random variable over \([0, 1]\), denoted as \( U \), we have \( X \overset{d}{=} F^{-1}(U) \). Therefore,
\[
\mathbb{E}[|X|] = \mathbb{E}[|F^{-1}(U)|] = \int_0^1 |F^{-1}(t)|^r \, dt.
\]
First, assume that \( 0 \leq F(0) \leq u \leq 1 \). Then,
\[
\int_0^1 |F^{-1}(t)|^r \, dt \geq \int_u^1 |F^{-1}(t)|^r \, dt \geq (1 - u)|F^{-1}(u)|^r, \tag{49}
\]
where in the last inequality we have used that \( F^{-1}(t) \) is positive for \( t \geq u \geq F(0) \), implying \( |F^{-1}(u)|^r \leq |F^{-1}(t)|^r \) for \( t \in [u, 1] \), and that \( F^{-1}(x) \) is non-decreasing in \( x \); hence, \( F^{-1}(u) \leq F^{-1}(t) \) for \( t \in [u, 1] \).

Now assume, on the other hand, \( 0 \leq u \leq F(0) \leq 1 \). Then,
\[
\int_0^1 |F^{-1}(t)|^r \, dt \geq \int_0^u |F^{-1}(t)|^r \, dt \geq u|F^{-1}(u)|^r. \tag{50}
\]
In the last inequality, we have used that \( F^{-1}(t) \) is negative for \( t \leq u \leq F(0) \), implying \( |F^{-1}(u)|^r \leq |F^{-1}(t)|^r \) and that \( F^{-1}(x) \) is non-decreasing in \( x \); hence, \( F^{-1}(t) \leq F^{-1}(u) \) for \( t \in [0, u] \). Thus, \( |F^{-1}(u)|^r \leq |F^{-1}(t)|^r \).

The bounds in (49) and (50) imply that
\[
|F^{-1}(u)| \leq \left( \frac{\mathbb{E}[|X|^r]}{u} \right)^{\frac{1}{r}}, \quad \text{for } u \in [0, F(0)],
\]
\[
|F^{-1}(u)| \leq \left( \frac{\mathbb{E}[|X|^r]}{1 - u} \right)^{\frac{1}{r}}, \quad \text{for } u \in [F(0), 1].
\]
Taking the largest of the two bounds concludes the proof of Lemma 8.

APPENDIX F
PROOF OF LEMMA 9
Choose some \( q \in [1, \infty] \) and define an event \( A \subset \{ p - \epsilon \leq U_{(np)} \leq p + \epsilon \} \), where we assume that
\[
p > \epsilon > \max \left\{ \frac{p}{n + 1}, \frac{\lfloor (q-2)p - q + 1 \rfloor}{q(n-1)+2} \right\}.
\]
Next, notice that we can write
\[
\mathbb{E}[\log(f(F^{-1}(U_{(np)})))] = \mathbb{E}[\log(f(F^{-1}(U_{(np)})))) 1_A]
+ \mathbb{E}[\log(f(F^{-1}(U_{(np)})))) 1_{A^c}]. \tag{51}
\]
We now analyze each expectation on the right side of (51) separately.

First expectation in (51). Using Taylor’s remainder theorem we have that
\[
\log(f(F^{-1}(u))) = \log(f(F^{-1}(p))) + \frac{f'(F^{-1}(\tilde{u}))}{f(F^{-1}(\tilde{u}))^2}(u - p),
\]
where \( \tilde{u} \) is some number between \( p \) and \( u \). Therefore, using the definition
\[
C^{(2)}_\epsilon = \max_{|u - p| \leq \epsilon} \left| \frac{f'(F^{-1}(u))}{f(F^{-1}(u))^2} \right|,
\]
we find that
\[
\mathbb{E}[\log(f(F^{-1}(U_{(np)})))) 1_A]
\leq \log(f(F^{-1}(p))) \mathbb{P}[A] + C^{(2)}_\epsilon \mathbb{E}[|U_{(np)} - p| 1_A]. \tag{52}
\]
Next, we notice that
\[
\mathbb{E}[|U_{(np)} - p| 1_A \leq \sqrt{\mathbb{E}[|U_{(np)} - p|^2]} \overset{(a)}{=} O(1/\sqrt{n}), \tag{53}
\]
where in the above, (a) follows using Cauchy-Schwarz inequality and the fact that \( \mathbb{P}[A] \leq 1 \) and (b) using Remark 11. Combining (52) and (53) we have the bound
\[
\mathbb{E}[\log(f(F^{-1}(U_{(np)})))) 1_A]
\leq \log(f(F^{-1}(p))) \mathbb{P}[A] + C^{(2)}_\epsilon O \left( \frac{1}{\sqrt{n}} \right). \tag{54}
\]
Second expectation in (51). First recall that \( U_{(np)} \sim \text{Beta}(np, n+1-np) \) and let \( f_{\alpha,\beta} \) denote the beta distribution with parameters \( \alpha := np \) and \( \beta := n+1-np \). Then, using the bound \( \log(x) \leq x \), we have
\[
\mathbb{E}[\log(f(F^{-1}(U_{(np)})))) 1_{A^c}]
\leq \mathbb{E}[f(F^{-1}(U_{(np)})) 1_{A^c}]
\leq \int_0^1 1_{A^c} f_{\alpha,\beta}(u)f(F^{-1}(u)) \, du. \tag{55}
\]
Next, by Hölder’s inequality, for any \( q, r \in [1, \infty] \) such that \( \frac{1}{q} + \frac{1}{r} = 1 \), we find
\[
\int_0^1 1_{A^c} f_{\alpha,\beta}(u)f(F^{-1}(u)) \, du
\leq \left( \int_0^1 1_{A^c} f_{\alpha,\beta}^q(u) \, du \right)^{\frac{1}{q}} \left( \int_0^1 (f(F^{-1}(u)))^r \, du \right)^{\frac{1}{r}}. \tag{56}
\]
Next we simplify the two terms on the right side of (56). First notice that \( f_{\alpha,\beta}^q(u) = \frac{\Gamma(i)\Gamma(j)\Gamma(i+j)}{\Gamma(i+j)^2} \) where
\[
\alpha^* = q(\alpha - 1) + 1, \quad \beta^* = q(\beta - 1) + 1, \tag{57}
\]
and
\[
c_{i,j} = \frac{\Gamma(i)\Gamma(j)}{\Gamma(i+j)}, \quad i \in \{\alpha, \alpha^*\}, \quad j \in \{\beta, \beta^*\}.
\]
It follows that,
\[
\left( \int_0^1 1_{A^c} f_{\alpha,\beta}^q(u) \, du \right)^{\frac{1}{q}} = \frac{c_{\alpha,\beta}}{c_{\alpha^*,\beta^*}} \left( \int_0^1 1_{A^c} f_{\alpha^*,\beta^*}(u) \, du \right)^{\frac{1}{r}}.
\]
where in the final equality \( U_{\alpha^*, \beta^*} \) denotes a beta random variable with parameters \( \alpha^* \) and \( \beta^* \). For the second term on the right side of (56), we use a change of variables with \( x = F^{-1}(u) \) with \( dx = (f(F^{-1}(u)))^{-1}du \) to get
\[
\left( \int_0^1 (f(F^{-1}(u)))^r du \right) = \left( \int_{-\infty}^\infty (f(x))^{r+1} dx \right) = (\|f\|_r+1)^{r+1}.
\] (59)
Therefore, putting together (55) – (59), we have shown
\[
\mathbb{E} \left[ \log(f(F^{-1}(U_{(np)}))) \right] \leq \frac{c_{\alpha^*, \beta^*}}{c_{\alpha, \beta}} \left( \|f\|_{r+1} \right)^{\frac{r+1}{2}}. \tag{60}
\]
We now focus on further upper bounding the right-hand side of (60). Towards this end, we start by noting that the definitions in (57) and the fact that \( \alpha := np \) and \( \beta := n+1-np \), we find \( \alpha^* + \beta^* = q(\alpha + \beta - 2) + 2 = q(n-1) + 2 \), therefore
\[
\frac{\alpha^*}{\alpha^* + \beta^*} = \frac{q(\alpha + \beta - 2) + 2}{q(n-1) + 2}, \quad \frac{\beta^*}{\alpha^* + \beta^*} = \frac{q(n-1) + 1}{q(n-1) + 2}.
\]
Moreover, in Appendix G it is shown that
\[
\frac{c_{\alpha^*, \beta^*}}{c_{\alpha, \beta}} \leq C_q n^{\frac{1}{2}} \tag{61}
\]
where the universal constant \( C_q := e^{1+\frac{1}{2}} (\sqrt{2\pi})^{-\frac{1}{2}} q^{-\frac{1}{2}} \).

Finally, we note that from properties of the beta distribution, namely that \( U_{\alpha^*, \beta^*} \equiv 1 - U_{\beta^*, \alpha^*} \), we find
\[
\mathbb{P}(U_{\alpha^*, \beta^*} \in \mathcal{A}^c) = \mathbb{P} [ U_{\alpha^*, \beta^*} \leq p - \epsilon ] + \mathbb{P} [ U_{\alpha^*, \beta^*} \geq p + \epsilon ] = \mathbb{P} [ U_{\beta^*, \alpha^*} \geq 1 - p + \epsilon ] + \mathbb{P} [ U_{\alpha^*, \beta^*} \geq p + \epsilon ]. \tag{62}
\]
Now focus on the first term on the right side of (62), we have
\[
\mathbb{P} [ U_{\beta^*, \alpha^*} \geq 1 - p + \epsilon ] \leq \mathbb{P} \left[ U_{\beta^*, \alpha^*} - \frac{\beta^*}{\alpha^* + \beta^*} \geq 1 - p + \epsilon - \frac{\beta^*}{\alpha^* + \beta^*} \right] \tag{63a}
\]
\[
\leq e^{-2(q(n-1) - (\epsilon - \sqrt{\frac{q(n-1) + 1}{q(n-1) + 2}}))^2} \tag{63b}
\]
where the labeled steps follow from: (a) noting that \( 1 - p + \epsilon - \frac{\beta^*}{\alpha^* + \beta^*} = \epsilon + \frac{(q-2)p - q + 1}{q(n-1) + 2} \geq \epsilon - \frac{|(q-2)p - q + 1|}{q(n-1) + 2} \). and (b) applying (43) since \( U_{\beta^*, \alpha^*} \) is \( \sigma_0 \)-sub-Gaussian random variables with \( \sigma_0 < \frac{1}{4(\alpha^* + \beta^* + 1)} = \frac{1}{4(q(n-1) + \beta)} \) Thm. 1.

Similarly, since
\[
p + \epsilon - \frac{\alpha^*}{\alpha^* + \beta^*} = \epsilon - \frac{(q-2)p - q + 1}{q(n-1) + 2} \geq \epsilon - \frac{|(q-2)p - q + 1|}{q(n-1) + 2};
\]
along with the fact that \( U_{\alpha^*, \beta^*} \) is \( \sigma_0 \)-sub-Gaussian again with \( \sigma_0 < \frac{1}{4(\alpha^* + \beta^* + 1)} = \frac{1}{4(q(n-1) + \beta)} \), we have the same bound for the second term on the right side of (62):
\[
\mathbb{P} [ U_{\alpha^*, \beta^*} \geq p + \epsilon ] \leq \mathbb{P} \left[ U_{\alpha^*, \beta^*} - \frac{\alpha^*}{\alpha^* + \beta^*} \geq \epsilon - \frac{|(q-2)p - q + 1|}{q(n-1) + 2} \right] \leq e^{-2(q(n-1) - (\epsilon - \sqrt{\frac{q(n-1) + 1}{q(n-1) + 2}}))^2}.
\] (64)
Thus, we have shown in (62)–(64)
\[
\mathbb{P}(U_{\alpha^*, \beta^*} \in \mathcal{A}^c) \leq 2e^{-2(q(n-1) - (\epsilon - \sqrt{\frac{q(n-1) + 1}{q(n-1) + 2}}))^2}. \tag{65}
\]
Now, combining (60) with (67) and (65) we arrive at
\[
\mathbb{E} \left[ \log \left( f(F^{-1}(U_{(np)}))) \right] \right] - \log \left( f(F^{-1}(p)) \right) \leq \log(f(F^{-1}(p)))/\mathbb{P}[\mathcal{A}^c] + C_q^2 O \left( \frac{1}{\sqrt{n}} \right) + 2C_q (\|f\|_{r+1}) e^{-2(q(n-1) - (\epsilon - \sqrt{\frac{q(n-1) + 1}{q(n-1) + 2}}))^2} \leq 2\log(f(F^{-1}(p)))e^{-2(q(n-1) - (\epsilon - \sqrt{\frac{q(n-1) + 1}{q(n-1) + 2}}))^2} + C_q^2 O \left( \frac{1}{\sqrt{n}} \right) + 2C_q (\|f\|_{r+1}) e^{-2(q(n-1) - (\epsilon - \sqrt{\frac{q(n-1) + 1}{q(n-1) + 2}}))^2}, \tag{66}
\]
where the last inequality follows by bounding \( \mathbb{P}[\mathcal{A}^c] \) using Lemma 7. This concludes the proof of Lemma 6.

**APPENDIX G**

**PROOF OF THE BOUND IN (61)**

Recall from the proof of Lemma 6 that \( \alpha := np \) and \( \beta := n+1-np \) while \( \alpha^* = q(\alpha - 1) + 1 \), and \( \beta^* = q(\beta - 1) + 1 \) for some \( q > 1 \). Thus, with reference to (61), we want to show that
\[
\frac{\Gamma(\alpha)^{\frac{1}{2}} \Gamma(\beta)^{\frac{1}{2}} \Gamma(\alpha + \beta)}{\Gamma(\alpha^* + \beta^*)^{\frac{1}{2}} \Gamma(\alpha + \beta)} \leq C_q n^{\frac{1}{2}(1 - \frac{1}{q})}, \tag{67}
\]
where \( C_q := e^{1+\frac{1}{2}} (\sqrt{2\pi})^{-\frac{1}{2}} q^{-\frac{1}{2}} \).

First, using that \( \Gamma(n+1) = n! \), recall Stirling’s approximation [37], which tells us that for \( n \geq 1 \),
\[
\Gamma(n+1) = \kappa_n \sqrt{2\pi n} n^{n+\frac{1}{2}} e^{-n}, \tag{68}
\]
where \( 1 \leq \frac{\Gamma(n+1)}{\kappa_n} \leq e^{\frac{1}{12n+1}} \leq e \).

Using the bound in (68), we find
\[
\Gamma(\alpha^*) - \Gamma(\alpha - 1) \leq e^{\sqrt{2\pi}(\alpha - 1)^{1/2}} e^{-\frac{1}{2}} q^{(\alpha - 1)/2} e^{-q(\alpha - 1)/2}.
\]
and 

\[ \Gamma(\alpha) \geq \sqrt{2\pi(\alpha - 1)^{\alpha - \frac{1}{2}}} e^{-(\alpha - 1)}; \]

hence, by simplifying, we obtain

\[ \frac{\Gamma(\alpha + \beta)^*}{\Gamma(\alpha)} \leq \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)} \frac{\Gamma(\alpha + \beta)^*}{\Gamma(\alpha + \beta)} \]

Finally, we again use the bound in (68) to find

\[ \frac{\Gamma(n + \beta)}{\Gamma(n)} \leq \frac{\Gamma(n + \beta)}{\Gamma(n)} \frac{\Gamma(n + \beta)}{\Gamma(n + \beta)} \]

Combining (69), (70), and (71) we arrive at

\[ \Gamma(\alpha + \beta) \Gamma(\alpha + \beta)^* \] 

hence, \( D(X_{(np)} \| G_{n,p}) = \infty \). This suggests that condition 3) (or a variation of it) in Theorem \( \Pi \) might indeed be necessary for convergence. First, note that

\[ \mathbb{E} \left[ \frac{1}{(1 - U_{(k)})^2} \right] = c_{k,n} \int_0^1 u^{k-1}(1-u)^{n-k} du, \]

where \( c_{k,n} = \frac{\Gamma(n+1)}{\Gamma(k)(n+k)} \); hence,

\[ \mathbb{E} \left[ \frac{1}{(1 - U_{(k)})^2} \right] = \infty, \quad \frac{m}{n} \geq n - k + 1. \] (72)

Now, by using the Taylor expansion of \( e^x \), we have that

\[ \mathbb{E}[X_{(k)}] = \mathbb{E}[F^{-1}(U_{(k)})] = \frac{e^{(1-U_{(k)})^{1/2}}}{m!} \]

where in the last step we have used Tonelli’s theorem (which holds even if the series diverges). Combining (72) and (73), we conclude that for every \( k \) there exists an \( m \) such that \( \frac{m}{n} \geq n - k + 1 \); hence, \( \mathbb{E}[X_{(k)}] = \infty \). This implies that

\[ \mathbb{E} \left[ (F^{-1}(U_{(np)}) - F^{-1}(p))^2 \right] = \mathbb{E} \left[ \left( X_{(np)} - e^{-x} \right)^2 \right] = \infty, \]

which leads to \( K_2 = \infty \).

**B. Density in (21)**

For the pdf

\[ f_2(x) = \frac{1}{x \log^2(x)}, \quad x \in (0, e^{-1}), \]

the cdf and the quantile function are given by

\[ F(x) = \frac{1}{x \log^2(x)}, \quad x \in (0, e^{-1}) \]

\[ F^{-1}(p) = e^{-x}, \quad p \in (0, 1) \]

We note that \( f_2 \) satisfies conditions 2) and 3) in Theorem \( \Pi \) hence, \( K_1 = O(1/n) \) by Lemma \( 2 \) and \( K_2 = O(1/\sqrt{n}) \) by Corollary \( 5 \). However, \( f_2 \) does not satisfy condition 1) in Theorem \( \Pi \). Nevertheless, as we next show, we can still prove the convergence of \( K_3 \). We start by noting that

\[ f(F^{-1}(p)) = p^2 e^{\frac{1}{x}}, \]

hence, from (14d) we obtain

\[ K_3 = \mathbb{E} \left[ \log \left( \frac{f(F^{-1}(U_{(np)})]}{f(F^{-1}(p))} \right) \right] \]

\[ = 2 \mathbb{E}[\log U_{(np)}] + \mathbb{E} \left[ \frac{1}{U_{(np)}} - 2 \log(p) - \frac{1}{p} \right] \]

\[ = 2 \left( \psi(np) - \psi(n + 1) \right) + \frac{n}{pm - 1} - 2 \log(p) - \frac{1}{p}, \]
where \( \psi \) is the digamma function. By noting that the digamma function can be approximated as
\[
\psi(x) = \log(x) - \frac{1}{2x} + O\left(\frac{1}{x^2}\right)
\]
and also using the fact that \( \psi(x+1) = \psi(x) + \frac{1}{x} \), we arrive at
\[
K_3 = \frac{n + 1 - pn - p^2 n + p}{pn (pn - 1)} + O\left(\frac{1}{n^2}\right),
\]
which, together with \( K_1 = O(1/n) \) and \( K_2 = O(1/\sqrt{n}) \), implies
\[
D(X_{(np)}(p)|G_{n,p}) = \Theta\left(\frac{1}{n}\right).
\]

REFERENCES

[1] C. R. Rao, C. Rao, and V. Govindaraju, *Handbook of Statistics*. Elsevier, 2006, vol. 17.

[2] H. A. David and H. N. Nagaraja, *Order Statistics*, Third edition. John Wiley & Sons, 2003.

[3] P. Laplace, “Théorie analytique des probabilités, deuxième supplément,” *Oeuvres*, vol. 7, no. 2, pp. 531–580, 1818.

[4] N. V. Smirnov, “Über die Verteilung des allgemeinen Gliedes in der Variationsreihe,” *Metron*, vol. 12, pp. 59–81, 1935.

[5] ——, “Limit distributions for the terms of a variational series,” *Trudy Mathematics Research Center*, United States Army, University of Wisconsin, 1963.

[6] A. A. Balkema and L. de Haan, “Limit distributions for order statistics. I,” *Theory of Probability & Its Applications*, vol. 23, no. 1, pp. 77–92, 1978.

[7] A. A. Balkema and L. de Haan, “Limit distributions for order statistics. II,” *Theory of Probability & Its Applications*, vol. 23, no. 2, pp. 341–358, 1979.

[8] R.-D. Reiss, *Approximate distributions of order statistics: With applications to nonparametric statistics*. Springer science & business media, 2012.

[9] M. M. Siddiqui, “Distribution of quantiles in samples from a bivariate population,” *J. Res. Nat. Bur. Standards B*, vol. 64, pp. 145–150, 1960.

[10] L. Weiss, *On the asymptotic joint normality of quantiles from a multivariate distribution*. Mathematics Research Center, United States Army, University of Wisconsin, 1963.

[11] ——, “The asymptotic distribution of order statistics,” *Naval Research Logistics Quarterly*, vol. 26, no. 3, pp. 437–445, 1979.

[12] ——, “The asymptotic joint distribution of an increasing number of sample quantiles,” *Annals of the Institute of Statistical Mathematics*, vol. 21, no. 1, pp. 257–263, 1969.

[13] W. Feller, *An Introduction to Probability Theory and its Applications*. John Wiley & Sons, 2008, vol. 2.

[14] R.-D. Reiss, “On the accuracy of the normal approximation for quantiles,” *The Annals of Probability*, pp. 741–744, 1974.

[15] S. Ikeda and T. Matsunawa, “On the uniform asymptotic joint normality of sample quantiles,” *Annals of the Institute of Statistical Mathematics*, vol. 24, no. 1, pp. 33–52, 1972.

[16] M. Falk, “A note on uniform asymptotic normality of intermediate order statistics,” *Annals of the Institute of Statistical Mathematics*, vol. 41, no. 1, pp. 19–29, 1989.

[17] R. R. Bahadur, “A note on quantiles in large samples,” *The Annals of Mathematical Statistics*, vol. 37, no. 3, pp. 577–580, 1966.

[18] M. L. Puri and S. S. Ralescu, “Limit theorems for random central order statistics,” in *Probability Theory and Extreme Value Theory*. De Gruyter Mouton, 2011, pp. 154–182.

[19] Y. Ma, M. G. Genton, and E. Parzen, “Asymptotic properties of sample quantiles of discrete distributions,” *Annals of the Institute of Statistical Mathematics*, vol. 63, no. 2, pp. 227–243, 2011.

[20] J. V. Linnik, “An information-theoretic proof of the central limit theorem with Lindeberg conditions,” *Theory of Probability & Its Applications*, vol. 4, no. 3, pp. 288–299, 1959.

[21] A. R. Barron, “Entropy and the central limit theorem,” *The Annals of probability*, pp. 336–342, 1986.

[22] S. Aitkin, K. M. Ball, F. Barthe, and A. Naor, “On the rate of convergence in the entropic central limit theorem,” *Probability theory and related fields*, vol. 129, no. 3, pp. 381–390, 2004.

[23] O. Johnson and A. Barron, “Fisher information inequalities and the central limit theorem,” *Probability Theory and Related Fields*, vol. 129, no. 3, pp. 391–409, 2004.

[24] M. Madiman and A. Barron, “Generalized entropy power inequalities and monotonicity properties of information,” *IEEE Transactions on Information Theory*, vol. 53, no. 7, pp. 2317–2329, 2007.