OSCILLATIONS FOR ORDER STATISTICS OF SOME DISCRETE PROCESSES

ANDREA OTTOLINI

Abstract. Suppose \(k\) balls are dropped into \(n\) boxes independently with uniform probability, where \(n, k\) are large with ratio approximately equal to some positive real \(\lambda\). The maximum box count has a counterintuitive behavior: first of all, with high probability it takes at most two values \(m_n\) or \(m_n + 1\), where \(m_n\) is roughly \(\frac{\ln n \ln \ln n}{\ln n}\). Moreover, it oscillates between these two values with an unusual periodicity. In order to prove this statement and various generalizations, it is first shown that for \(X_1, \ldots, X_n\) independent and identically distributed discrete random variables with common distribution \(F\), under mild conditions, the limiting distribution of their maximum oscillates in three possible families, depending on the tail of the distribution. The result stated at the beginning follows from the equivalence of ensemble for the order statistics in various allocations problems, obtained via conditioning limit theory. Results about the number of ties for the maximum, as well as applications, are also provided.

1. Introduction

1.1. Extreme value theory. Even though outliers are often disregarded in statistical models, an understanding of rare and extreme events plays a central role in a variety of situations. An important example, which is analyzed in more detail later, is the occurrence of coincidences for big earthquakes.

It is well known (see [16]) that for independent identically distributed (i.i.d.) random variables \(X_1, \ldots, X_n\) with common distribution function \(F(x)\), in order for a law of large numbers for \(\max_{1 \leq i \leq n} X_i\) to hold, it is necessary and sufficient for the \(X_i\)'s to have a slowly varying tail. More precisely,

\[
\exists m_n \text{ s.t. } X_{(n)} - m_n \to 0 \text{ in probability } \iff \lim_{x \to +\infty} \frac{1 - F(x + y)}{1 - F(x)} = 0, \forall y > 0. 
\]

In case the condition above holds, necessary and sufficient conditions for the existence of a limiting distribution for \(X_{(n)}\), after rescaling, are also standard in the literature, and the limits have been widely studied. For a survey on the subject, as well as generalizations and applications, the reader is referred to [15], [11].

It is worth mentioning that (1.1) fails to capture the maxima of a variety of distributions: in particular, if the \(X_i\)'s only take integer values, the above condition cannot be satisfied, since for \(0 < y < 1\) and \(x \in \mathbb{N}\) the above limit is one. The goal of this paper is to investigate the limiting behaviour of \(X_{(n)}\) when condition (1.1) is not satisfied.

In this paper, ‘clustering’ refers to the extent to which \(X_{(n)}\) fails to satisfy a law of large numbers. Roughly speaking, it will be shown that the decay of the mass function determines the size of the cluster. As an instance, for Poisson random variables (whose mass function decays faster than geometrically) \(X_{(n)}\) clusters at two
values with high probability, while for negative binomial random variables (whose mass function has a geometric decay) $X_{(n)}$ is spread onto all integers with high probability.

The first rigorous result in this direction is due to Anderson [3]. He classifies the cluster size for maxima of discrete random variables in terms of their tails. A further analysis is carried out in [21], where lower order statistics are taken into account, as well as in [6], where the authors study the number of ties.

The first result of this paper completes the discussion given in Anderson [3].

**Theorem 1.1.** Let $X_1, \ldots, X_n$ be i.i.d. discrete random variables with common distribution $F$. Suppose the support of $F$ is not bounded from above, and that for some $\gamma \in [0, 1]$

\[
\lim_{n \to +\infty} \frac{1 - F(n + 1)}{1 - F(n)} = \gamma.
\]

Then there exist two sequences $\{m_n\} \subset \mathbb{N}, \{p_n\} \subset [0, 1]$ such that

- $\gamma = 0 \implies \mathbb{P}(X_{(n)} = m_n) \sim p_n, \mathbb{P}(X_{(n)} = m_n + 1) \sim 1 - p_n$,
- $\gamma \in (0, 1) \implies \text{for all } x \in \mathbb{Z}, \mathbb{P}(X_{(n)} \leq m_n + x) \sim p_n^x$,
- $\gamma = 1 \implies \text{for all } x \in \mathbb{Z}, \mathbb{P}(X_{(n)} \leq m_n + x) \sim p_n$.

In the first case, there exists an increasing sequence $\{N_i\}_{i \in \mathbb{N}} \subset \mathbb{N}$, with $N_i - N_{i-1} \to \infty$, such that for $n \to \infty, n \notin \{N_i\}$, one has $p_{n+1} \leq p_n$ and $p_{n+1} - p_n \to 0$.

**Remark 1.2.** The last part of Theorem 1.1 is from where the expression “oscillations” originates. Indeed, a more informal way of interpreting the result is the following: if the endpoints of $[0, 1]$ are identified to obtain a circle (and the orientation on $[0, 1]$ induces a counterclockwise orientation on the circle), then $p_n = e^{-2\pi i q_n}$, where the sequence $\{q_n\}$ is increasing and satisfies $q_n - q_{n-1} \to 0$, $\limsup q_n = +\infty$.

In words, under this identification the sequence of $p_n$’s will move clockwise on the circle with smaller and smaller steps, winding around the origin infinitely many times.

From a probabilistic point of view, in the case $\gamma = 0$ the sequences $\{p_n\}, \{1 - p_n\}$ represent, for $n$ large, the relative frequency of $X_{(n)} = m_n, X_{(n)} = m_{n+1}$ respectively. Therefore, a histogram of many samples from $X_{(n)} - m_n$ will not converge to a given shape: it will instead consist of two adjacent columns whose heights oscillate between 0 and 1. Moreover, for every fixed $p \in [0, 1]$, it is possible to find a subsequence along which the height of the left column will converge to $p$ (and correspondingly, the height of the right one will converge to $1 - p$). This again justifies the term “oscillations.”

Another natural question concerns the number of times the maximum is expected to occur in a sequence of independent and identically distributed samples from a discrete distribution. This question in only addressed here in the case $\gamma = 0$, where the result is the following.

**Theorem 1.3.** Let $X_1, \ldots, X_n$ be i.i.d. with common distribution $F$ such that

\[
\lim_{n \to +\infty} \frac{1 - F(n + 1)}{1 - F(n)} = 0.
\]

Then, for $p_n, m_n$ as in Theorem 1.1

\[
P(\text{at least } k \text{ ties for the maximum}) \sim p_n + 1 - p_n \sum_{j=0}^{k} \frac{\ln^j \left( \frac{1}{p_n} \right)}{j!}.
\]
As a byproduct, the probability of having no ties is given by \(-p_n \ln p_n\), which thus oscillates between 0 and \(\frac{1}{e}\).

In the proof of the theorem, it will be clear that in order to determine the number of ties, one first flips a \(p_n\)-biased coin to determine whether the maximum \(X_{(n)}\) will assume the value \(m_n\) or \(m_n + 1\): if the former happens, then one expects, for all fixed \(k\), to have at least \(k\) ties with high probability as \(n\) gets larger; if the latter happens, the number of ties is Poisson distributed with parameter \(\ln\left(\frac{1}{p_n}\right)\). In the case where \(p_n = 0\) or \(p_n = 1\) for some \(n\), the right side in (1.3) equals one (according to the convention \(0 \ln 0 := 0\)). Therefore, along subsequences of \(\{p_n\}\) converging to 0 (resp. 1), the \(k\)-th order statistic \(X_{(n-k)}\) for any fixed \(k\) will be \(m_n\) (resp. \(m_n + 1\)) with high probability, so that an arbitrarily large number of ties is expected.

Now, suppose that the \(p_n\)-biased coin gives \(X_{(n)} = m_n\). It is interesting to determine how many ties are expected to occur, letting \(k\) grow with \(n\). This is answered by the following.

**Theorem 1.4.** Let \(X_1, \ldots, X_n\) be as before, and let \(c > 0\) be fixed. Then, there exists a sequence \(z_n\) such that

- if \(c < 1\), \(P(X_{n-cz_n} > m_n - 1) \to 1\),
- if \(c > 1\), \(P(X_{n-cz_n} > m_n - 1) \to 0\).

In other words, there is a phase transition for the number of ties in the top position, the critical point being \(z_n\). It is worth mentioning that \(z_n\) oscillates as well.

### 1.2. Multinomial allocations and their Bayesian counterparts.

In order to adapt all previous results to the case of dependent random variables, one approach is to use Poissonization. This standard technique has been widely exploited in a variety of situations: various combinatorial problems in probability [9], cycle structure [22] and longest increasing subsequence [2] of random permutations, equivalence of ensemble in statistical mechanics [24], and many others.

When the randomization leads to an exponential family, from which the original distribution can be obtained by means of conditioning with respect to a sufficient statistic, the results belong to conditioning limit theory. The case where the sufficient statistic is given by \(\sum iX_i\) is treated in [5], while here the focus will be on the sum, already considered in [13].

In the following, two different allocation problems are considered. Suppose \(k\) balls are dropped into \(n\) boxes, and we denote by \(Y = (Y_1, \ldots, Y_n)\) the box counts. If different balls are dropped independently, \(Y\) has a multinomial distribution; if the probabilities of falling into a certain box are unknown, it is natural to consider a Dirichlet mixture of multinomial distributions.

The starting point is the conditional representation

\[(1.4) \quad P(\max(Y_1, \ldots, Y_n) \in A) = P\left(\max(X_1, \ldots, X_n) \in A \mid \sum X_i = k\right).\]

In the case \(Y\) is multinomial with parameters \((k, p_1, \ldots, p_n)\), the \(X_i\)'s can be chosen in such a way that \(X_1 \sim \text{Poi}(\lambda p_1), \ldots, X_n \sim \text{Poi}(\lambda p_n)\) for every \(\lambda\). If the \(Y\) are a mixture of multinomial distributions with symmetric Dirichlet kernel with hyperparameter \(r\), then the \(X_i\)'s can be chosen to be in the negative binomial family.
with parameter NB(r, p). Bayes’ theorem then implies

\[ P(\max(Y_1, ..., Y_n) \in A) = \frac{\sum X_i = k}{\sum X_i = k} \]

where \( \bar{X}_i \) is the law of \( X_i \) conditioned on \( \max(X_1, ..., X_n) \in A \). Notice that all the three terms on the right side only involve independent random variables. The idea is that, under the appropriate choice of the \( X_i \)’s on the right hand side, the ratio appearing there is approximately one, so that

\[ P(\max(Y_1, ..., Y_n) \in A) \sim P(\max(X_1, ..., X_n) \in A). \]

Notice that in general this does not require either side to have a limit: the only important aspect is that the distribution of \( \max(X_1, ..., X_n) \) and \( \max(Y_1, ..., Y_n) \) merge together, and then everything about the former (e.g., convergence, limit points etc.) carries over to the latter. For different notions of merging, corresponding to different notions of distance between \( f_n(X_1, ..., X_n) \) and \( f_n(Y_1, ..., Y_n) \), the reader is referred to [1].

The main results of this paper for the allocation models are the following.

**Theorem 1.5.** Let \( (Y_1, ..., Y_n) \) be multinomial with parameters \( (k, \frac{1}{n}, ..., \frac{1}{n}) \). Suppose that \( \lambda \frac{k}{n} \) is fixed. If \( m_n, p_n \) are defined as in Theorem 1.1 for \( F \) the distribution function of a Poisson with parameter \( \lambda \), then one has

\[ P(Y_n = m_n) \sim p_n, \quad P(Y_n = m_n + 1) \sim 1 - p_n. \]

**Theorem 1.6.** Let \( (Y_1, ..., Y_n) \) be multinomial with parameters \( (k, \frac{1}{n}, ..., \frac{1}{n}) \). Let \( \lambda \frac{k}{n} \) be fixed. If \( m_n, p_n, z_n \) are defined as in Theorems 1.3, 1.4, for \( F \) the distribution function of a Poisson with parameter \( \lambda \), then one has:

- As \( n \to +\infty \),

\[ P(\text{at least } t \text{ ties at } Y(n)) \sim p_n + 1 - p_n \sum_{j=0}^{t} \frac{\ln^j(\frac{1}{p_n})}{j!}. \]

- As \( n \to +\infty \),

\[ \begin{align*}
  &\quad \text{if } c < 1, \quad P(X_n - cz_n > m_n - 1) \to 1, \\
  &\quad \text{if } c > 1, \quad P(X_n - cz_n > m_n - 1) \to 0.
\end{align*} \]

**Theorem 1.7.** Let \( (Y_1, ..., Y_n) \) be a (symmetric) Dirichlet mixture of multinomials with parameters \( (k, r) \). Let \( n, k \) be chosen in such a way that for some fixed \( p \), \( \frac{k}{n} - p = \frac{1}{n} \) is fixed. If \( m_n, p_n \) are defined as in Theorem 1.1 for \( F \) the distribution function of a negative binomial random variable with parameters \( (r, p) \), then for every \( x \in \mathbb{Z} \)

\[ P(Y_n \leq m_n + x) \sim p_n^x. \]

Borrowing language from statistical mechanics, in the multinomial allocations problem (as well as in its Bayesian version), several features involving the top order statistics can be equivalently studied in the microcanonical and canonical picture [3].
2. The independent case

Let $X_1, ..., X_n$ be i.i.d. random variables with common distribution $F$, satisfying assumption (1.2). Denote by $F := 1 - F$ the tail of the distribution. A real extension $\mathcal{G}$ of $F$ will be considered, with the properties

- $\mathcal{G}(n) = F(n)$ for $n \in \mathbb{Z}$.
- $\mathcal{G}$ is continuous.
- $\mathcal{G}$ is decreasing.
- $\mathcal{G}$ is log-convex.

Such an extension always exists: for example, the log-linear one provided by Anderson in [3] works (yet, sometimes, this is not the most natural one, as in the Poisson case). Notice that the assumption of log-convexity ensures, for example, the existence of

$$\lim_{x \to \pm \infty} \frac{\mathcal{G}(x + \frac{1}{2})}{\mathcal{G}(x)},$$

since the function $x \to \frac{\mathcal{G}(x + \frac{1}{2})}{\mathcal{G}(x)}$ is decreasing. Combined with (1.2), one obtains

$$\lim_{x \to \pm \infty} \frac{\mathcal{G}(x + \frac{1}{2})}{\mathcal{G}(x)} = \sqrt{\gamma}.$$  

In a similar way, because $\mathcal{G}$ is continuous and log-convex, for every $\epsilon \in (0, 1)$ one has

$$\lim_{x \to \pm \infty} \frac{\mathcal{G}(x + \epsilon)}{\mathcal{G}(x)} = \gamma^\epsilon.$$  

Let $x_n$ be a solution to $\mathcal{G}(x_n) = \frac{1}{n}$. Owing to the continuity and monotonicity of $\mathcal{G}$, such a solution exists and it is unique. Set $m_n$ to be the floor of $x_n + \frac{1}{2}$. By definition, $m_n \in [x_n - \frac{1}{2}, x_n + \frac{1}{2}]$. Define also,

$$\theta_n := \frac{\mathcal{G}(m_n)}{\mathcal{G}(x_n)}, \quad p_n := e^{-\theta_n}.$$  

Finally, let

$$z_n := -\ln \left( \frac{\mathcal{G}(m_n - 1)}{\mathcal{G}(x_n)} \right).$$

**Remark 2.1.** While the above definitions of $m_n, x_n, p_n, z_n$ depend on the particular choice of $\mathcal{G}$, all the results stated in the above theorems are equivalent for all such choices.

Consider the case $\gamma \in (0, 1)$ in Theorem 1.1, the other results being analogous. Let $\mathcal{G}$ and \( \tilde{\mathcal{G}} \) be two different extensions of $F$, and let $\tilde{x}_n, \tilde{m}_n, \tilde{p}_n$ be the quantities corresponding to $x_n, m_n, p_n$, obtained by using the extension $\tilde{\mathcal{G}}$ instead. The first claim is that

$$x_n - \tilde{x}_n \to 0.$$  

Indeed, first notice that the floor of both $x_n$ and $\tilde{x}_n$ is the largest integer $t_n$ such that $\mathcal{F}(t_n) \geq \frac{1}{n}$. Therefore, $x_n = t_n + \epsilon_n$ and $\tilde{x}_n = \tilde{t}_n + \tilde{\epsilon}_n$, with $\epsilon_n, \tilde{\epsilon}_n \in [0, 1)$. Suppose, toward contradiction, that along some subsequence $|\epsilon_n - \tilde{\epsilon}_n| \sim \epsilon > 0$ for some fixed $\epsilon$. In this case, applying (2.2) leads to

$$\mathcal{F}(t_n) \gamma^\epsilon \sim \mathcal{G}(x_n) = \frac{1}{n} = \tilde{\mathcal{G}}(\tilde{x}_n) \sim \mathcal{F}(t_n) \gamma^\epsilon,$$

which in turn implies $\gamma^\epsilon = 1$ in the limit, a contradiction.

Therefore, along subsequences of $\{x_n\}$ (and thus $\{\tilde{x}_n\}$, since $x_n - \tilde{x}_n \to 0$) that
are bounded away from half-integers, the definition of \( m_n \) and \( \tilde{m}_n \) will eventually be the same, and consequently \( p_n \) will be eventually the same (since \( p_n \) depends on \( G \) only via \( m_n \)) regardless of the choice of \( G \).

Along subsequences of \( \{x_n\} \) (and \( \{\tilde{x}_n\} \)) for which \( x_n - [x_n] \to \frac{1}{2} \), it may be the case that \( m_n \) and \( \tilde{m}_n \), defined from \( x_n \) and \( \tilde{x}_n \) respectively, differ by one.

Without loss of generality, assume that along some subsequence \( m_n - x_n \to -\frac{1}{2} \) and \( \tilde{m}_n - \tilde{x}_n \to \frac{1}{2} \). In this case, combining (2.3), (2.1), one obtains

\[
p_n \sim e^{-\gamma}, \quad \tilde{p}_n \sim e^{-\gamma}.
\]

In particular, the use of the extension \( \tilde{G} \) leads to the same asymptotic obtained by using \( G \), since

\[
P(X_{(n)} \leq m_n + x) = P(X_{(n)} \leq \tilde{m}_n - 1 + x) \sim (\tilde{p}_n)^{\gamma^{-1+x}} \sim e^{-\gamma^{-\frac{1}{2}+x}} \sim p_n^x.
\]

**Remark 2.2.** The freedom in the choice of \( G \) is not merely an abstract curiosity. As previously mentioned, there are cases (e.g., Poisson distribution) where a certain \( G \) can be obtained by appropriately replacing a sum with an integral (in the Poisson distribution case, an incomplete gamma function), and such \( G \) can be easily checked to be log-convex (for classical discrete distribution, this often boils down to the log-convexity of the gamma function). In those cases, it is much easier to use such extensions in order to obtain numerical approximations for \( m_n \) and \( p_n \), rather than using the artificial log-linear extension introduced by Anderson [3].

### 2.1. Proof of Theorem 1.1

**Proof of Theorem 1.1.** Owing to the i.i.d. assumption, for all \( x \in \mathbb{Z} \)

\[
P(X_{(n)} \leq x) = \left(1 - F(x)\right)^n.
\]

In the following, note that for \( z > -1 \),

\[
\frac{z}{1+z} \leq \ln(1+z) \leq z.
\]

Thus, for every \( x \in \mathbb{Z} \),

\[
P(X_{(n)} \leq x) = \left(1 - F(x)\right)^n = \left(1 - \frac{F(x)}{\mathcal{G}(x_n)^{n}}\right) \in \left[e^{-\frac{F(x)}{\mathcal{G}(x_n)^{n}}}, e^{-\frac{F(x)}{\mathcal{G}(x_n)^{n}}}\right],
\]

where the definition of \( \mathcal{G}(x_n) = \frac{1}{n} \) is used in the second equality, while the bounds are derived from (2.4) applied to \( z = -\frac{F(x)}{n\mathcal{G}(x_n)} \) (so that \( 1+z = F(x) \)).

It is convenient to split the proof in the two cases \( \gamma = 0, \gamma > 0 \):

- Case \( \gamma = 0 \): the choice of \( x = m_n - 1 \) leads to

\[
P(X_{(n)} \leq m_n - 1) \leq e^{-\frac{\mathcal{G}(m_n-1)}{\mathcal{G}(x_n)}} \leq e^{-\frac{\mathcal{G}(x_n+\frac{1}{2})}{\mathcal{G}(x_n)}} \to 0
\]

where the first inequality follows from the upper bound in (2.4) and the second follows from the both the monotonicity of \( \mathcal{G} \) and that \( m_n \in [x_n - \frac{1}{2}, x_n + \frac{1}{2}] \). The last step is a consequence of the assumption \( \gamma = 0 \) and equation (2.2). If \( x = m_n + 1 \), then the lower bound in (2.5) is attained and the very same argument leads to

\[
P(X_{(n)} \leq m_n + 1) \geq \exp\left\{-\frac{\mathcal{G}(x_n+\frac{1}{2})}{\mathcal{G}(x_n)\mathcal{F}(m_n+1)}\right\}
\]

\[
\to 1.
\]
Oscillations for Order Statistics of Some Discrete Processes

This result proves the clustering effect on the two values \( m_n, m_n + 1 \). In
general, the same computations lead to
\[
\mathbb{P}(X(n) \leq m_n) \sim \exp \left\{ -\frac{\mathcal{G}(m_n)}{\mathcal{G}(x_n)} \right\} = p_n,
\]
from which the result
\[
\mathbb{P}(X(n) \leq m_n) \sim p_n, \quad \mathbb{P}(X(n) = m_n + 1) \sim 1 - p_n,
\]
follows. As for the last statement in the Theorem, notice that by the
definition of \( x_n \)
\[
\frac{\mathcal{G}(x_{n+1})}{\mathcal{G}(x_n)} = \frac{n+1}{n} \to 1.
\]
Suppose, toward contradiction, that \( x_{n+1} - x_n \to \epsilon > 0 \) along some sub-
sequence. Owing to (2.2), the limit \( \ell := \lim_{n \to +\infty} \frac{\mathcal{G}(x_n + \epsilon)}{\mathcal{G}(x_n)} \) exists and it
is equal to zero (thanks to the assumption \( \gamma = 0 \)). Thus, necessarily
\( x_{n+1} - x_n \to 0 \). By continuity of \( \mathcal{G} \),
\[
\mathcal{G}(x_n) - \mathcal{G}(x_{n+1}) \to 0.
\]
Define \( N_i \) as the increasing sequence of natural numbers for which \( x_{N_i} \leq
i + \frac{1}{2}, x_{N_i+1} > i + \frac{1}{2} \) for all integers \( i \). Consider \( x_n \) for \( n \notin \{N_i\}_{i \in \mathbb{N}} \), and
recall that \( m_n \) is the floor of \( x_n + \frac{1}{2} \). For such \( n \)'s, one has \( m_n = m_{n+1} \),
and consequently \( p_{n+1} \leq p_n \) (owing to the monotonicity of \( \mathcal{G} \) and (2.3)),
and \( p_{n+1} - p_n \to 0 \) (owing to the continuity of \( \mathcal{G} \)). Finally, notice that
\( N_{i+1} - N_i \to \infty \) since \( x_n \to +\infty, x_{n+1} - x_n \to 0 \), which concludes the
proof.

- Case \( \gamma > 0 \): for all fixed \( x \in \mathbb{Z} \) one has, owing to (2.3) and (1.2),
\[
\frac{\mathcal{F}(m_n + x)}{\mathcal{G}(x_n)} \sim \frac{\mathcal{F}(m_n + x)}{\mathcal{G}(x_n)\mathcal{F}(m_n + x)} \sim \frac{\mathcal{G}(m_n)\gamma^n}{\mathcal{G}(x_n)}.\]
Therefore, in this case
\[
\mathbb{P}(X(n) \leq m_n + x) = \left( 1 - \frac{\mathcal{G}(m_n)}{\mathcal{G}(x_n)} \right)^n \sim \exp \left\{ -\frac{\mathcal{G}(m_n)}{\mathcal{G}(x_n)} \gamma^n \right\} = p_n^{\gamma^n},
\]
which concludes the proof in the case \( \gamma > 0 \) (notice that, if \( \gamma = 1 \), one has
\( \gamma^n \equiv 1 \)).

\[
\square
\]

2.2. Proof of Theorems 1.3 and 1.4

Proof of Theorem 1.3. First of all, notice that the assumption \( \gamma = 0 \) guarantees
that \( z_n \to +\infty \). Moreover, since by definition
\[
n\mathcal{F}(m_n - 1) = z_n,
\]
and \( m_n \to +\infty \), it follows that \( z_n = o(n) \). Now, consider the binomial formula for
the order statistics
\[
\mathbb{P}(X(n-k) \leq x) = \sum_{j=0}^{k} \binom{n}{j} [\mathcal{F}(x)]^n-j[1 - \mathcal{F}(x)]^j.
\]
If \( x = m_n - 1 \), \( j \) is fixed and \( n \to +\infty \), one has
\[
\binom{n}{j} \left( 1 - \frac{z_n}{n} \right)^{n-j} \frac{z_n}{n}^j \sim \frac{z_n^j}{j!} e^{-z_n} \to 0.
\]

When \( k \) is fixed, since there are only finitely many terms in \( \sum_{j=0}^{k} \), and each of these converges to 0, one concludes
\[
(2.7) \quad P(X_{(n-k)} < m_n - 1) \to 0.
\]

Now fix \( p \in (0,1) \), and look at a subsequence \( p_n \to p \) (where for simplicity the \( p_{n_k} \) subsequence was renamed \( p_n \)). Then, since \( P(X_{(n)}) > m_n + 1 \) \( \to 0 \), along this subsequence
\[
P(X_{(n-k)} = m_n + 1) \sim P(X_{(n-k)} > m_n)
\]
\[
= 1 - \sum_{j=0}^{k} \binom{n}{j} \left( 1 - \frac{\theta_n}{n} \right)^{n-j} \left( \frac{\theta_n}{n} \right)^j
\]
\[
= p \left( \frac{1}{p} - \sum_{j=0}^{k} \frac{\ln^j \left( \frac{1}{p} \right)}{j!} \right),
\]
where it was used that
\[
F(m_n) = \frac{\mathcal{F}(m_n)}{\mathcal{G}(x_n)} \mathcal{G}(x_n) = \mathcal{G}(m_n) \mathcal{G}(x_n) \left( \frac{1}{n} \right) = \frac{\theta_n}{n}.
\]
\[
\square
\]

Before moving on to the proof of Theorem 2.4, recall the incomplete gamma function
\[
\Gamma(k, z) = \int_0^z t^{k-1} e^t dt.
\]
Notice that for \( k \) an integer, integration by parts shows that
\[
e^{-z} \sum_{j=k}^{\infty} \frac{z^j}{j!} = \frac{\Gamma(k+1, z)}{\Gamma(k+1)}.
\]

To find asymptotics for \( \Gamma(k, z) \), it is useful to recall the Laplace asymptotic formula (see, e.g., Theorem 3.5.3 in [3]):

**Theorem 2.3** (Laplace asymptotic formula). Let \( S(x) \) be a smooth function on \((a, b)\). Then
\begin{itemize}
  \item If \( S'(x) < 0 \) for all \( x \in (a, b) \), then
    \[
    \int_a^b e^{-mS(x)} f(x) dx = \frac{1}{m} \frac{1}{S'(b)} f(b) e^{-mS(b)} \left( 1 + O \left( \frac{1}{m} \right) \right), \quad m \to +\infty.
    \]
  \item If \( S \) has a unique nondegenerate minimum \( x_0 \) in \((a, b)\), then
    \[
    \int_a^b e^{-mS(x)} f(x) dx = \sqrt{\frac{2\pi}{m S''(x_0)}} f(x_0) e^{-mS(x_0)} \left( 1 + O \left( \frac{1}{m} \right) \right), \quad m \to +\infty.
    \]
\end{itemize}

**Proof of Theorem 2.4** Using (2.6)
\[
P(X_{(n-cz_n+1)} > m_n - 1) = 1 - \sum_{j=0}^{cz_n-1} \binom{n}{j} \left[ F(m_n - 1) \right]^{n-j} \left[ 1 - F(m_n - 1) \right]^j
\]
\[
= \sum_{j=cz_n}^{n} \binom{n}{j} \left( 1 - \frac{z_n(1 + o(1))}{n} \right)^{n-j} \left( \frac{z_n(1 + o(1))}{n} \right)^j.
\]
Now, given $c \in (0, +\infty)$, consider $m = m(c)$ large (to be fixed later). The sum above can be split into

$$P(X_{n-cz_n+1} > m_n - 1) = \sum_{j=cz_n}^{mcz_n-1} \binom{n}{j} \left(1 - \frac{z_n(1 + o(1))}{n}\right)^{n-j} \left(\frac{z_n(1 + o(1))}{n}\right)^j$$

$$+ \sum_{j=mcz_n}^{n} \binom{n}{j} \left(1 - \frac{z_n(1 + o(1))}{n}\right)^{n-j} \left(\frac{z_n(1 + o(1))}{n}\right)^j =: A + B$$

First, consider the second summand: since $\binom{n}{j} \leq \frac{n^j}{j!}$, each term can be bounded:

$$\left(\frac{n}{j}\right) \left(1 - \frac{z_n(1 + o(1))}{n}\right)^{n-j} \left(\frac{z_n(1 + o(1))}{n}\right)^j \leq \frac{n^j}{j!} \rightarrow 0.$$  

By choosing $m$ large enough that $mc > e$, using

$$\frac{n^j}{j!} \leq \frac{mc}{mc - 1} \left(\frac{e}{mc}\right)^{mcz_n} \rightarrow 0.$$

Going back to the first summand, to finish the proof it is enough to show that

$$A = \sum_{j=cz_n}^{mcz_n-1} \binom{n}{j} \left(1 - \frac{z_n(1 + o(1))}{n}\right)^{n-j} \left(\frac{z_n(1 + o(1))}{n}\right)^j$$

converges to $0$ if $c > 1$ and converges to $1$ if $c < 1$, regardless of $m$. In this regime, $j \to +\infty$, $j = o(n)$, so

$$A \sim \sum_{j=cz_n}^{mcz_n-1} \frac{e^{-cz_n}s_j^j}{j!} \sim \sum_{cz_n}^{+\infty} \frac{e^{-cz_n}s_j^j}{j!},$$

where the last step follows from the fact that $B \to 0$. Using (2.8),

$$A \sim \frac{\Gamma(cz_n + 1, z_n)}{\Gamma(k_n(c))} = \int_{(cz_n)!}^{+\infty} \frac{e^{t+c}e^{-t}dt}{(cz_n)!}$$

Changing the variable $t = cz_n s$ and using Stirling’s approximation, we obtain

$$A \sim \frac{1}{(cz_n)!} \int_0^{\frac{cz_n}{s}} e^{-cz_n[s \ln s + cz_n \ln(cz_n)] + cz_n \ln s} ds$$

$$\sim \frac{cz_n}{e^{-cz_n \sqrt{2\pi cz_n}} \sqrt{2\pi cz_n}} \int_0^{\frac{cz_n}{\ln s}} e^{-cz_n(s-Ln s)} ds.$$  

Since the function $S(s) = s - \ln s$ has a global minimum at $s = 1$, with $S(0) = 1$, $S''(1) = 1$, if $c > 1$ the first part of Theorem 2.3 gives

$$A \sim \frac{\sqrt{cz_n}}{e^{-cz_n \sqrt{2\pi}} \frac{1}{2} e^{(-cz_n)(\frac{1}{c} + \ln c)}} \rightarrow 0,$$

since $1 < \frac{1}{c} + \ln c$. In the case $c > 1$, the second part of Theorem 2.3 leads to

$$A \sim \frac{\sqrt{cz_n}}{e^{-cz_n \sqrt{2\pi}} \sqrt{2\pi} e^{-cz_n}} = 1,$$

which concludes the proof. □
2.3. Some examples: the Poisson, the negative binomial, and the discrete Cauchy. Consider the case where $X_1 \sim \text{Poi}(\lambda)$. Anderson ([3]) already proved the result

$$P(X(n) \in \{m_n, m_n + 1\}) \to 1.$$ 

In the language of this paper, the Poisson distribution falls into the case $\gamma = 0$ of Theorem 1.1, since

$$F(x) \to e^{-\lambda} \frac{\lambda^{x+1}}{\Gamma(x+2)}.$$ 

Notice that the most natural choice for $G$ in this case is given by the incomplete gamma function, rather than the log-linear extension. Following the proof of Theorem 1.1, it is easy to see that

$$P(X(n) \notin \{m_n, m_n + 1\}) \leq \left(\frac{\lambda}{x_n + 1}\right)^{m_n - x_n}.$$ 

This bound is important, as it shows the clustering may emerge even for small values of $n$, provided that $\lambda$ is small. As for the value of $m_n$, in [20], it is shown that in first approximation

$$x_n \sim \frac{\ln n}{\ln \ln n}.$$ 

However, this estimate is extremely poor, as is shown in [7]. In particular, if $W(z)$ is the solution to $e^{W(z)}W(z) = z$ (known as the Lambert function, see [10]), then a much better approximation is given by

$$\tilde{x}_n = y_n + \frac{\ln \lambda - \frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln(y_n)}{\ln(y_n) - \ln \lambda}, \quad y_n = \frac{\ln n}{W\left(\frac{\ln n}{\lambda^2}\right)}.$$ 

The Negative binomial distribution $N(r, p)$ falls into the second category of Theorem 1.1 with $\gamma = p$, since

$$\frac{F(n+1)}{F(n)} = \frac{\Gamma(n + 2 + r) \Gamma(n + 1) \Gamma(r)}{\Gamma(n + 1) \Gamma(n + r)} \frac{\int_0^t t^{n+1}(1-t)^{r-1}dt}{\int_0^t t^n(1-t)^{r-1}}.$$ 

and, using [23] and the property of the gamma function, it is easy to obtain

$$\frac{F(n+1)}{F(n)} = \frac{n + 1 + r}{n + 1} p \left(1 + o\left(\frac{1}{n}\right)\right) \to p.$$ 

Finally, the discrete Cauchy distribution falls into the third regime, since in that case

$$\frac{F(n+1)}{F(n)} \to \frac{1}{1 + x_n^2} \to 1.$$ 

3. The dependent case

As explained in the introduction, it is possible to export the previous results to a certain class of allocation problems. The main ingredient is the local central limit theorem (see, e.g., [17]).

Lemma 3.1 (Local Central Limit theorem). Let $X_1, ..., X_n$ be discrete i.i.d. random variables, with $E(X_1) = \mu, \text{Var}(X_1) = \sigma^2$, such that the values taken on by $X_1$
are not contained in some infinite progression $a + q\mathbb{Z}$ for integers $a, q$ with $q > 1$. Then, for every integer $t$,
\[
\mathbb{P}\left(\sum_{i=1}^{n} X_i = t\right) = \frac{1}{\sqrt{2\pi n\sigma}} \exp\left(-\frac{(t - n\mu)^2}{2n\sigma^2}\right) + o\left(\frac{1}{\sqrt{n}}\right),
\]
the error being uniform in $t$.

3.1. Multinomial allocations. First, consider the case of multinomial allocations, all boxes being equally likely.

**Proof of theorem 1.5.** Let $X_1, \ldots, X_n$ be i.i.d. with $x_i \sim \text{Poi}(\lambda)$, $X_1 \sim \text{Poi}(\lambda)$. By means of (1.5),
\[
\mathbb{P}(Y_{(n)} \in A) = \mathbb{P}(X_{(n)} \in A) \frac{\mathbb{P}(\sum X_i = k, X_{(n)} \in A)}{\mathbb{P}(\sum X_i = k)}.
\]

Notice that
\[
\mathbb{P}(X_{(n)} = m_n) \sim p_n, \quad \mathbb{P}(X_{(n)} = m_{n+1}) \sim 1 - p_n
\]
owing to Theorem 1.1. Moreover, $\sum_{i=1}^{n} X_i \sim \text{Poi}(k)$, so that
\[
(3.1) \quad \mathbb{P}\left(\sum_{i=1}^{n} X_i = k\right) = e^{-k} \frac{k^k}{k!} \sim \frac{1}{\sqrt{2\pi k}}.
\]

It remains to estimate the “tilted” version of the $X_i$’s. If $A = \tilde{m}_n$, where $\tilde{m}_n = m_n$ or $\tilde{m}_n = m_n + 1$ then $\{X_i\}_{i=1}^{n-1} = \{X_i\}_{i=1}^{n}\backslash X_{(n)}$ are still independent and identically distributed according to
\[
\mathbb{P}(\tilde{X}_1 = t) = \frac{e^{-\lambda} \lambda^t}{F(\tilde{m}_n)}, \quad t = 0, \ldots, \tilde{m}_n,
\]
$F$ being the cumulative distribution of Poisson as in section 3. By symmetry, each $X_i$ is equally likely to be the maximum, so that $\mathbb{P}(X_{(n)} = X_i) = \frac{1}{n}$. Therefore,
\[
\mathbb{P}\left(\sum_{i=1}^{n} X_i = k | X_{(n)} = \tilde{m}_n\right) = \mathbb{P}\left(\sum_{i=1}^{n-1} \tilde{X}_i = n - \tilde{m}_n\right),
\]
which can be now estimated by means of Theorem 3.1 (notice that the condition that $X_1$ does not belong to a subprogression is obviously satisfied). The first moment is
\[
\mathbb{E}(\tilde{X}_1) = \frac{1}{F(\tilde{m}_n)} \sum_{j=0}^{\tilde{m}_n} e^{-\lambda} \frac{\lambda^j}{j!} = \lambda \frac{F(\tilde{m}_n - 1)}{F(\tilde{m}_n)} = \lambda(1 + o(1)).
\]

Similarly, the variance is given by
\[
\text{Var}(\tilde{X}_1^2) = \frac{\sum_{j=0}^{\tilde{m}_n} e^{-\lambda} \frac{\lambda^{j+2}}{j!}}{F(\tilde{m}_n)} - \lambda^2(1 + o(1))
\]
\[
= \frac{\lambda^2 F(\tilde{m}_n - 2) + \lambda F(\tilde{m}_n - 1)}{F(\tilde{m}_n)} - \lambda^2(1 + o(1))
\]
\[
= \lambda(1 + o(1)).
\]
Hence, the local central limit theorem leads to
\[
P\left(\sum_{i=1}^{n-1} X_i = n - \tilde{m}_n\right) = \frac{1}{\sqrt{2\pi(n - 1)\lambda}} \exp\left\{\frac{(k - \tilde{m}_n - (n - 1)\lambda(1 + o(1)))}{2(n - 1)\lambda}\right\}
\]
\[
\sim \frac{1}{\sqrt{2\pi k}} \exp\left\{\frac{\lambda - \tilde{m}_n)^2}{2k}\right\}
\]
\[
\sim \frac{1}{\sqrt{2\pi k}}
\]
where the last step follows from \(\tilde{m}_n^2 = o(k)\), a consequence of \(\tilde{m}_n \sim \frac{\ln n}{n}\) and \(k = \lambda n\). Therefore,
\[
P(Y(n) = m_n) \sim P(X(n) = m_n) \sim p_n, \quad P(Y(n) = m_n + 1) \sim 1 - p_n,
\]
as desired.

For the proofs of Theorem 1.6, the very same argument can be applied. Indeed, the only difference is that the number of copies of \(\tilde{X}\)'s is now \(n - t, n - t_n(c)\) respectively. However, this does not affect the central limit theorem, since \(t\) and \(t_n(c)\) are much smaller than \(n\) (so that the asymptotic in the central limit theorem remains the same).

### 3.2. A Bayesian version.

Consider now the Bayesian variant of the multinomial allocation problem. The idea is again the same, but a proof is sketched for the sake of completeness.

**Proof of Theorem 1.7** Fix \(x \in \mathbb{Z}\). By means of (1.5) and the conditional representation of Dirichlet mixture of multinomials as negative binomials, it suffices to show that for \(X_1, ..., X_n\) i.i.d. with \(X_i \sim NB(r, p)\) and \(\frac{e^t}{e^p} = \frac{k}{\pi}\),
\[
P\left(\sum_{i=1}^{n} X_i = k\right) \sim P\left(\sum_{i=1}^{n} X_i = k|X(n) \leq m_n + x\right),
\]
as \(n, k \to +\infty\). As before, the right hand side can be rewritten as
\[
P\left(\sum_{i=1}^{n} \tilde{X}_i = k\right),
\]
where \(\tilde{X}_i\) is the tilded version of \(X_i\), given by
\[
P(\tilde{X}_i = s) = \frac{P(X_1 = s)}{P(X_1 \leq m_n + x)}, \quad s \leq m_n + x.
\]
Since both the mean and the variance are asymptotically the same for \(x, X_i\) and \(\tilde{X}_i\) (using that \(m_n + x \to +\infty\)), the local central limit theorem can be applied to conclude the proof.

### 4. Numerical results and applications

While theoretically satisfactory, the question remains of whether these asymptotic results are of any use in simulations or real models (or whether \(n\) has to be unreasonably large for the effect to be manifest). Here are the main take-aways from some simulations for i.i.d. discrete random variables and random allocation models:

- The merging of dependent and independent cases works well for reasonable values of \(k\) and \(n\). If the theory gives good approximations in some regime for the independent random variables, it also works for the dependent ones.
• In order to detect the oscillations of the maxima (as well as the other features) in the Poisson case, the quantity \( \frac{\lambda}{x_n + 1} \) has to be small. Since \( x_n \) grows sublogarithmically, \( n \) has to be extremely large compared to \( \lambda \) (in particular, it is necessary to have \( n \gg e^\lambda \)). This explains why simulations essentially fail for \( \lambda \gg 1 \), why they work for \( \lambda = O(1) \) provided \( n \) is large (for \( \lambda = 1 \), in order to obtain \( p_n \) within an error of \( \epsilon \), it is necessary to have at least \( n \geq e^{\epsilon} \)), and why they are excellent for \( \lambda \ll 1 \), even with relatively small \( n \).

4.1. The role of the mean. Here are some numerical values for \( m_n, x_n, p_n \), depending on \( n \). For now, we take \( \lambda = 1 \), (but, as explained above, soon \( \lambda \) will be small).

| \( n \) | \( x_n \) | \( m_n \) | \( p_n \) |
|------|------|------|------|
| 10^3 | 4.63591 | 5 | 0.58694674 |
| 10^4 | 5.84299 | 6 | 0.47741767 |
| 10^5 | 6.95712 | 7 | 0.40055502 |
| 10^6 | 8.00608 | 8 | 0.36296353 |
| 10^7 | 10.89530 | 11 | 0.46225972 |
| 10^8 | 10.8993 | 11 | 0.45873497 |
| 10^9 | 40.0255 | 40 | 0.333090 |

Values of \( m_n, x_n, p_n \) from Theorem 1.1 as a function of \( n \) (with \( \lambda = 1 \)). The value \( x_n \) can be obtained e.g., via approximating the Lambert function as in [10], or by means of numerical methods.

Here are some observations from the table:

• The value \( x_n \) grows slowly. At first sight, it seems logarithmic, as the factor \( \ln \ln n \) in the asymptotic (2.9) is hard to detect for reasonable values of \( n \). Only the last entry gives an insight in this direction,

• The absence of a law of large numbers, as well as the oscillations, already emerges in this picture: the value of \( p_n \) does not exhibit any limiting behavior,

• The period of the oscillations (i.e., the difference \( N_{i+1} - N_i \) in the language of Theorem 1.1) is increasing in \( n \).

In the following, 1000 trials of the experiment “drop \( \lambda n \) balls into \( n \) boxes independently” were simulated. Because of Theorem 1.5, the maximum box count should be \( m_n \) or \( m_n + 1 \) with probabilities given by, respectively, \( p_n \) or \( 1 - p_n \). The outcomes are represented in the following table:

| \( n \) | \( \lambda \) | \( m_n \) | \( p_n \) | \( f_n \) | \( o_n \) |
|------|------|------|------|------|------|
| 10^3 | 0.1 | 3 | 0.675268 | 0.69 | 0.005 |
| 10^4 | 1 | 7 | 0.40055502 | 0.353 | 0.11 |
| 10^5 | 10 | 25 | 0.325168 | 0.162 | 0.467 |

Comparison between \( p_n \) and the relative frequency \( f_n \) for \( m_n \) out of 1000 trials of the experiment “drop \( \lambda n \) balls into \( n \) boxes”. The last column represents the fraction \( o_n \) of maxima outside the cluster values \( m_n, m_n + 1 \).

Here are some observations:

• For large \( \lambda \), the approximation is useless. This is not surprising since the quantity \( \frac{\lambda}{x_n + 1} \) is far from being negligible.

• For small \( \lambda \), the approximation works well, and the theory can be fully appreciated for reasonable \( n \), since the quantity \( \frac{\lambda}{x_n + 1} \) is small,
If \( \lambda \) increases, the value of \( m \) also increases. However, the \( \lambda \)-correction in \( x \) (and hence, in \( m \)) is rather small. This is the reason why small \( \lambda \) is preferable in order to see the results from the theory;

- Notice that since \( x \) grows sublogarithmically, for fixed \( \lambda \) we need to significantly increase \( n \) to see an improvement. On the other hand, once \( \lambda \) is small, the theory works even for \( n \) small (e.g., \( n = 1000 \)).

That being said, the focus will be now on the regime \( \lambda = 0.01 \) in order to even better capture the “oscillating behavior” of \( p_n \). This is an experiment of dropping \( \lambda n \) balls into \( n \) boxes, for various values of \( n \),

| \( n \)  | \( m_n \) | \( p_n \)  | \( f_n \) |
|-----|-----|-----|-----|
| 2000 | 1   | 0.8902 | 0.9073 |
| 4000 | 1   | 0.8039 | 0.8171 |
| 8000 | 1   | 0.6602 | 0.6646 |
| 16000| 1   | 0.4492 | 0.4548 |
| 32000| 1   | 0.2106 | 0.2047 |
| 64000| 1   | 0.0469 | 0.0357 |
| 128000| 1   | 0.0023 | 0.0017 |
| 256000| 1   | 0.0000 | 0.0000 |
| 512000| 2   | 0.9103 | 0.9181 |

Oscillation of \( p_n \) for \( n = 1000 \times 2^m, m \in \{1, \ldots, 9\} \), and \( \lambda = 0.01 \). For each \( n \), I simulated 10000 times the experiment of dropping \( \lambda n \) balls into \( n \) boxes. As before, \( f_n \) denotes the relative frequency of \( m_n \).

The oscillation is visible in the last step: \( p_n \) “refreshes” at 1 after \( x_n - m_n \) changes its sign, a phenomenon that happens on a long scale.

Moving to the number of ties, Theorem 1.3 implies that the probability of having \( t \) ties at the value of the maximal order statistic is given by

\[
\mathbb{P}(t \text{ ties for the maximal order statistic}) \sim p \frac{\ln^{t+1} \left( \frac{1}{p} \right)^{t+1}}{(t+1)!}
\]

Here is a simulation of the process:

| \( t \) | \( t_n \) | \( f_n \) |
|-----|-----|-----|
| 0   | 0.35948 | 0.3613 |
| 1   | 0.14382 | 0.1431 |
| 2   | 0.03836 | 0.0385 |
| 3   | 0.0076  | 0.008  |

The result of 10000 simulation of dropping 160 balls into 16000 boxes (\( \lambda = 0.01 \), \( p_n = 0.44924115 \)) and counting the number of ties \( t \). The relative frequencies \( f_n \) are compared to the theoretical probabilities \( t_n \).

The results are very accurate for small numbers of ties. Finally, here are simulations for the result for the cluster size on the top two spots for the same values of \( n \) and \( \lambda \): the theoretical result is that about 156.65 boxes should have a count of 1 or 2 balls. The average of 10000 experiments gives the result 159.21.

4.2. Coincidence for earthquakes. In the popular imagination, big earthquakes are one of the main instances of randomness in natural events. Heuristically, big
earthquakes are not independent of each other (as everyone who lives in a seismic area knows), and they instead tend to clump together. As such, a reasonable model is that of inter-arrival times (forgetting about any geographic information) which are distributed according a negative binomial (see [15]), which is suitable for representing positively correlated events.

In the following, we adopt this model and use our theory to study the occurrence of multiple big earthquakes in a given window of time, using data from [23]. For instance, can we explain the occurrence of multiple earthquakes in a given hour by purely statistical arguments, without any “cause-effect” arguments?

In the language of the previous section, $X_i$ is the number of earthquakes of magnitude above 6 which occurred in an hour $i$ of the day, with $i$ running from 1 to 24. We consider realizations over three periods of time: the 70s, the 80s, and the 90s, which correspond respectively to 3652, 3653, and 3652 instances of the $X_i$’s.

| Decade | $E(X_i)$ | $Var(X_i)$ | $r$   | $p$   |
|--------|---------|------------|------|------|
| 70's   | 0.01226497 | $6.089 \times 10^{-4}$ | 0.0496 | 0.0472 |
| 80's   | 0.01382425 | $7.112 \times 10^{-4}$ | 0.0514 | 0.0489 |
| 90's   | 0.01669177 | $8.6302 \times 10^{-4}$ | 0.0517 | 0.0489 |

Occurrence of earthquakes in a given hour across three different decades, with corresponding estimators with a negative binomial model.

We compare the results between our theory, a result of $10^6$ simulations of negative binomial random variables with the same parameter, and the empirical data. We expect the maximum number of earthquakes in a single hour within a day to be either 0, 1, or 2 (theoretically, numerically, and empirically it is almost impossible to observe more than 3 earthquakes in a given hour). Here are the results:

| Decade | theory | numerics | empirical data |
|--------|--------|----------|----------------|
| 70's   | 75.17 - 23.49 - 1.28 | 75.06 - 24.08 - 0.81 | 76.01 - 23.11 - 0.84 |
| 80's   | 72.52 - 25.92 - 1.48 | 72.34 - 26.65 - 0.97 | 74.24 - 24.89 - 0.95 |
| 90's   | 67.88 - 30.24 - 1.79 | 67.65 - 31.07 - 1.22 | 69.52 - 29.30 - 1.01 |

Maximum number of earthquakes in a single hour within a day. The notation $a - b - c$ denotes the percentages of days with 0, 1, or 2 as a maximum.

ACKNOWLEDGEMENT

The author wishes to thank Persi Diaconis for suggesting the problem, and for many helpful discussions on the subject. The author is also indebted to Daniel Dore and two referees for their careful revision of the first drafts.

REFERENCES

[1] A. D’Aristotile, P. Diaconis, D. Freedman On merging of probabilities Sankhya: The Indian Journal of Statistics, 1988
[2] D.J. Aldous, P. Diaconis Hammersley’s interacting particle process and longest increasing subsequences Probability Theory and Related Fields, 1995.
[3] C.W. Anderson, Extreme value theory for a class of discrete distributions with applications, J. Appl. Probab. 7, 1970.
[4] G.W. Anderson, A. Guionnet, O. Zeitouni An introduction to Random Matrices, Cambridge University Press (2006)
[5] R. Arratia, D. Barbour, S. Tavaré Limits of logarithmic combinatorial structures, Annals of Probability, 2000.
[6] J.S. Athreya, S. Sethuraman On the asymptotic of discrete order statistics, Statistics & Probability Letters 54, 2001.
[7] K. Briggs, L. Song, T. Prellberg A note on the distribution of the maximum of a set of Poisson random variables, https://arxiv.org/abs/0903.4373.
[8] J.M.V. Campenhout, T.M. Cover Maximum entropy and Conditional probability, IEEE Transaction on Information Theory, Vol. IT-27, No. 4, July 1981.
[9] S. Chatterjee, P. Diaconis, E. Meckes Exchangeable pairs and Poisson approximation, Probability Surveys, 2005
[10] R.M. Corless, G.H. Gonnet, D.E.G. Hare, D.J. Jeffrey, D.E. Knuth On the Lambert W function, Advances in Computational Mathematics (1996).
[11] L. de Haan, A.F. Ferreira Extreme Value Theory: An Introduction, Springer Series in Operations Research and Financial Engineering (2006).
[12] P. Diaconis, D. Freedman Partial exchangeability and sufficiency, Proc. IndianStat. Inst. Golden Jubilee Int’l Conf. Stat.: Applications and New Directions, J. K. Ghosh and J. Roy (eds.), Indian Statistical Institute, Calcutta, 205-236, (1984)
[13] P. Diaconis, S. Holmes A Bayesian peek into Feller volume I, The Indian Journal of Statistics, 2002
[14] P. Diaconis, F. Mosteller Methods for studying coincidences, Journal of the American Statistical Association, Vol 84, No. 408 (Dec., 1989), pp. 853-861.
[15] J. Galambos The asymptotic Theory of Extreme Order Statistics, Krieger Publishing Co., Melbourne, FL, 1987.
[16] B.V. Gnedenko, A.N. Kolmogorov Limit distributions for sums of independent random variables, Addison-Wesley (1954) (Translated from Russian).
[17] J. Greenhough, I. G. Main A Poisson model for earthquake frequency uncertainties in seismic hazard analysis
[18] K. Joav-Deg, F. Proschan Negative association of random variables, with applications, Ann. Statist., 11, 286-295.
[19] A.C. Kimber A note on Poisson maxima Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete, (1983)
[20] S. Sethuraman A clustering law for some discrete order statistics, J. Appl. Prob. 40, 2003.
[21] L.A. Sheep, S.P. Lloyd Ordered cycle lengths in a random permutation, Transactions of the American Mathematical Society, 1966
[22] USGS https://earthquake.usgs.gov/.
[23] S.L. Zabell Rates of convergence for Conditional Expectations, Annals of Probability, 1980.

(A. Ottolini) Department of Mathematics, Stanford University, 450 Jane Stanford Way, Stanford CA 94305 (USA).
Email address: ottolini@stanford.edu