# The Exact and Asymptotic Distributions of $X$ mod $k$ and Various Analytical Connections

ANIRBAN DASGUPTA*, THOMAS M. SELLKE†, SANTOSH S. VENKATESH‡

February 29, 2016

ABSTRACT For a general sequence of non-negative integer-valued random variables $X_n$ and a fixed integer $k \geq 2$, the distribution of $X_n$ mod $k$ for fixed $n$ and asymptotic distributions as $n \to \infty$ are studied by using various mathematical tools. The speed of convergence is also considered. The convergence to uniformity is linked to several things, such as Fourier analysis, Weyl's equidistribution theorem and convolutions, the Euler-MacLaurin summation formula, a Tauberian theorem, and Markov chains.

In particular, Euler-Maclaurin summation allows one to write almost a necessary and sufficient condition for convergence to uniformity, and also explains oscillation on the way to convergence.

It is shown that in some important special cases, such as binomial, negative binomial, and Poisson, the exact distributions for fixed $n$ admit closed form formulas in terms of special functions, the limit being uniform in appropriate limiting paradigms.

Convergence to uniformity can fail when polynomials of lattice variables are considered. In such cases, results on congruences tell us what the limit distribution will be, and when the limit will still be uniform.

Extensions to the multidimensional case as well as the case where the modulus itself is random are indicated, and the theorems are illustrated with many examples and plots.

*Department of Statistics, Purdue University, dasgupta@stat.purdue.edu. Research partially supported by ELS-LSIGJLXJ2K356E.

†Department of Statistics, Purdue University, tsellke@purdue.edu.

‡Department of Electrical and Systems Engineering, University of Pennsylvania, venkatesh@seas.upenn.edu. This research was funded in part by grant CNS-0915697 from the National Science Foundation.

# 1 Introduction

Let $X$ be a non-negative integer-valued random variable, and $k \geq 2$ a fixed positive integer. We study the distribution of $X$ mod $k$ and interesting functionals, for example the moments, using a number of mathematical tools. The tools used include analysis, in particular Tauberian theorems, Fourier analysis, number theory, asymptotic expansions, and Markov chains in some special cases. In a more general case of this setting we consider a sequence of non-negative integer-valued random variables, say $X_n$, and study the asymptotic distribution of $X_n \bmod k$, convergence of the moments, and the rate of convergence; e.g., is convergence exponentially fast?

The problem we treat here is in some sense a natural extension of Poincare's roulette problem. We do not claim to have broken new ground probabilistically in this article. But we have given a set of rather pretty calculations and examples, and we have shown how the same problem is connnected to many classical facets of mathematics. The appeal of our article is in the connections and the calculations.

The Poisson distribution is an obvious special case. If $X_n$ is Poisson with mean $n$, then, modulo $k$, $X_n$ admits the recursion $X_n = X_{n-1} + Y_n$, where the $Y_n$ sequence is i.i.d. The mod $k$ problem in such a convolution setting arises if one wants to study the position of a random walker on the non-negative integers, where counting starts over when the walker reaches or goes past a given $k$. The study of the asymptotic aspects of this case is amenable to classical Markov chain techniques; two specifically relevant references are Hildebrand (1990) and Diaconis (1992). Hildebrand and Diaconis both point out that second eigenvalue techniques do not generally lead to the best possible answers in this case. Fourier methods have more to offer.

The results of this article have a few different aspects: one aspect is studying the distribution of $X \bmod k$ for a fixed lattice random variable $X$ in interesting special cases, such as the Poisson, binomial, and the negative binomial; in these cases we provide explicit closed form formulas for the modulo residues in terms of hypergeometric functions.

A second aspect arises by taking a sequence $X_n$ of lattice random variables and studying the asymptotic distribution of $X_n$ mod $k$. In this story, we show that there is an underlying connection of this problem to the theme of Weyl's equidistribution theorem. If the distribution of $X_n$ has a convolutional structure, $X_n = Y_1 + \cdots + Y_n$, where the $Y_i$ are i.i.d. and the common distribution $F$ of the $Y_i$ puts positive mass at both $0$ and $1$, then the distribution of $X_n \bmod k$ will approach the uniform distribution on the set of residues $\{1, 2, \cdots, k-1\}$ when $n \to \infty$, and this condition cannot be relaxed. But we also show that $X_n$ need not have a convolution structure for convergence to uniformity for $X_n$ mod $k$ to take place; if (for large $n$) $X_n$ has a finite number of modes $m_n$ and a global maximum $M_n$, then the distribution of $X_n \bmod k$ will approach the uniform on $\{1, 2, \cdots, k-1\}$ as

long as $m_n M_n \to 0$.

A third aspect is determining the speed of convergence to uniformity (when it happens) and understanding the oscillatory nature of this convergence. Here quite a different technique becomes useful. We show by use of the Euler-MacLaurin summation formula that if the mass function $f_n$ of $X_n$ is inherited from a very smooth density function on $[0, \infty)$, then convergence to uniformity takes place, and additionally, we can say something about how fast the convergence occurs.

A fourth aspect is understanding cases of failure and salvaging the failures. Here, we find other connections. We show by use of well-known theorems in congruence that prime powers and well-behaved polynomials conform to convergence to uniformity, but otherwise in general it fails.

Here is a short description of what is in this article:

a) Section 2 starts out with a number of examples that illustrate a diversity of phenomena; for instance, oscillation, lack of oscillation, rapid convergence, slow convergence, failures of a general theorem, etc.

b) Section 4 lays out a general story of convergence and moments. The connection to convolution structures is laid out in Theorem 4.3 together with Corollary 4.3 and Corollary 4.4. Theorem 4.4 and Corollary 4.5 show the multidimensional extensions.

The connection to smooth interpolation and the Euler-MacLaurin formula are laid out in Theorem 4.5 and Corollary 4.6. Theorem 4.6 and Theorem 4.7 give Fourier bounds on rapidity of the convergence of moments. Section 4.4 shows where the oscillatory behavior comes from in certain cases.

c) The important Poisson case is picked up in Section 3. Theorem 3.1 gives the exact formula for the distribution of the residues of a Poisson variable modulo $k$ by using hypergeometric functions. Lemma 3.1 is the Tauberian connection. As is well known, the Poisson case is suited for Markov chain methods; this is shown for completeness as part of the proof of Theorem 3.2. Failures and connections to prime powers are described in Section 3.3.

d) Section 3.4 is the binomial and negative binomial analog of what is in Section 3. Again, exact formulas are given (Theorem 3.6, Theorem 3.7). Convergence to uniformity is handled at the same time.

e) Many examples, figures, and tables are given in various sections to motivate and illustrate the results.

## 2 First Illustrative Examples

We begin by a consideration of the usual lattice distributions.

**Example 2.1.** *(The binomial).* Suppose $X \sim \text{Binomial}(n, 1/2)$ represents the accumulated successes in $n$ tosses of a fair coin. Then,

$$P(5 \text{ divides } X) = P(X \bmod 5 = 0) = \sum_{j=0}^{\lfloor n/5 \rfloor} \binom{n}{5j} 2^{-n}$$
$$= \frac{1}{5} + \frac{2}{5}\left[\left(\frac{\varphi}{2}\right)^n \cos\left(\frac{n\pi}{5}\right) + \left(\frac{1}{2\varphi}\right)^n \cos\left(\frac{2n\pi}{5}\right)\right] \quad (1)$$

where $\varphi = \big(\sqrt{5}+1\big)/2 = 1.618\cdots$ is the golden ratio and $1/\varphi = \big(\sqrt{5}-1\big)/2 = 0.618\cdots$ is the golden ratio conjugate. Thus, for example, if $n$ is an even multiple of $5$, then $P(X \bmod 5 = 0)$ has a positive jump, while if $n$ is an odd multiple of $5$, then $P(X \bmod 5 = 0)$ has a negative jump. The function is oscillatory.

Since $\varphi/2$ and $1/(2\varphi)$ are both less than $1$, we see that $P(X \bmod 5 = 0)$ converges to $1/5$ as $n \to \infty$. It is interesting to note, however, that the second term in the formula (1) is oscillatory, although it decays exponentially quickly with $n$. Figure 1 shows that the oscillations are significant for $n$ up to about $30$.

Oscillatory behaviour arises in other directions as well: Figure 2 shows that the probabilities $P\big(5 \mid \text{Binomial}(n,p)\big)$ are also oscillatory in $p$ for given $n$, the oscillation gradually dying off as $n$ increases.

**Example 2.2.** *(The Poisson case).* Suppose $X \sim \text{Poisson}(\lambda)$, the Poisson distribution with mean $\lambda$. Then, by a direct calculation

$$P(3 \text{ divides } X) = P(X \bmod 3 = 0) = \sum_{j=0}^{\infty} \frac{e^{-\lambda}\lambda^{3j}}{(3j)!} = \frac{1}{3} + \frac{2}{3}e^{-3\lambda/2}\cos\left(\frac{\lambda\sqrt{3}}{2}\right). \quad (2)$$

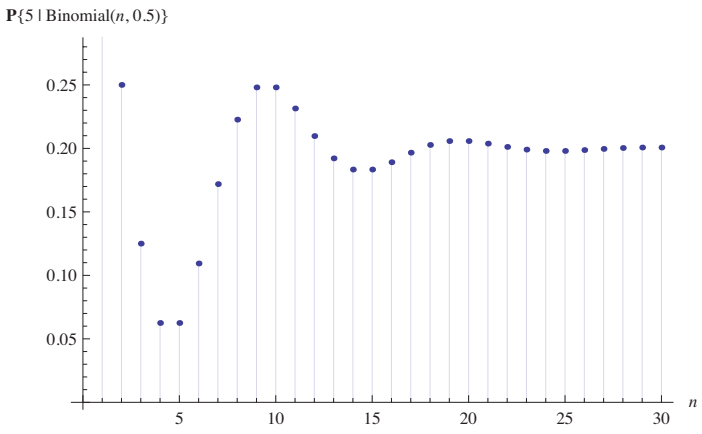


***Figure 1:*** *Oscillatory convergence of $P\big(5 \mid \text{Binomial}(n,0.5)\big)$ as $n$ increases.*

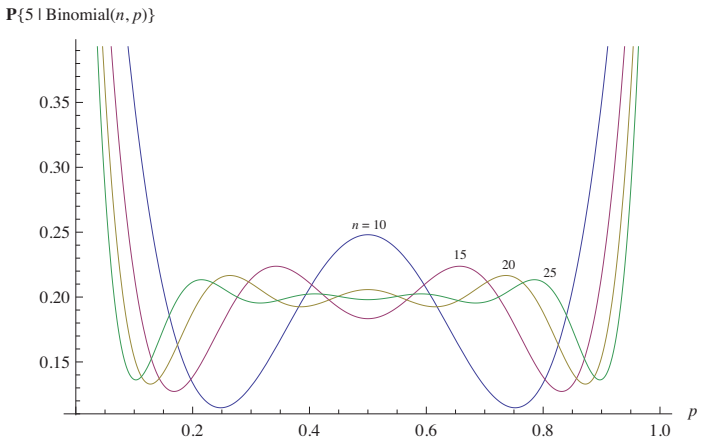


***Figure 2:*** *Oscillations in $p$ for the distribution of binomial residues modulo* $5$.

This obviously converges to $1/3$ when $\lambda \to \infty$. However, again, it is interesting that it contains a very small oscillatory term $\frac{2}{3}e^{-3\lambda/2}\cos\left(\frac{\lambda\sqrt{3}}{2}\right)$ converging to zero.

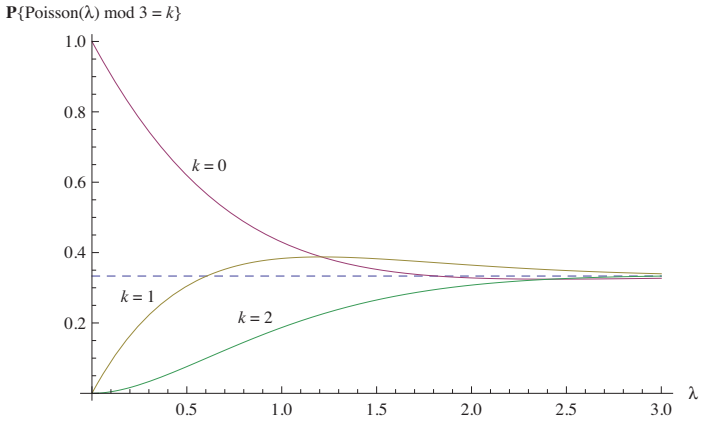


***Figure 3:*** *Oscillatory convergence of the distribution of Poisson residues modulo* $3$.

In fact, there is nothing particularly special about $P(X \bmod 3 = 0)$; similar exact formulas for $P(X \bmod 3 = 1)$ and $P(X \bmod 3 = 2)$ show that these too converge to $1/3$, while possessing small oscillatory terms. The exact formulas are:

$$P(X \bmod 3 = 1) = \frac{1}{3} + \frac{2}{3}e^{-3\lambda/2}\sin\left(\frac{3\sqrt{3}\lambda - \pi}{6}\right),$$
$$P(X \bmod 3 = 2) = \frac{1}{3} - \frac{2}{3}e^{-3\lambda/2}\sin\left(\frac{3\sqrt{3}\lambda + \pi}{6}\right).$$

Thus, in this specific illustrative example, $P(X \bmod 3 = j) \to 1/3$ for each $j = 0, 1, 2$, while $P(X \bmod 3 = j) - 1/3$ is an $O(e^{-3\lambda/2})$ oscillatory function of $\lambda$. Convergence to uniformity is very fast in this case.

There is nothing special about studying $X \bmod 3$; for any positive integer $k$, the probabilities $P(X \bmod k = j)$ depict analogous convergence as well as oscillatory phenomena. Figure 4 shows the convergence to uniformity of $X \bmod 5$ in the Poisson case as $\lambda$ grows. We shall see in Section 4.3 that these phenomena are more general and not just

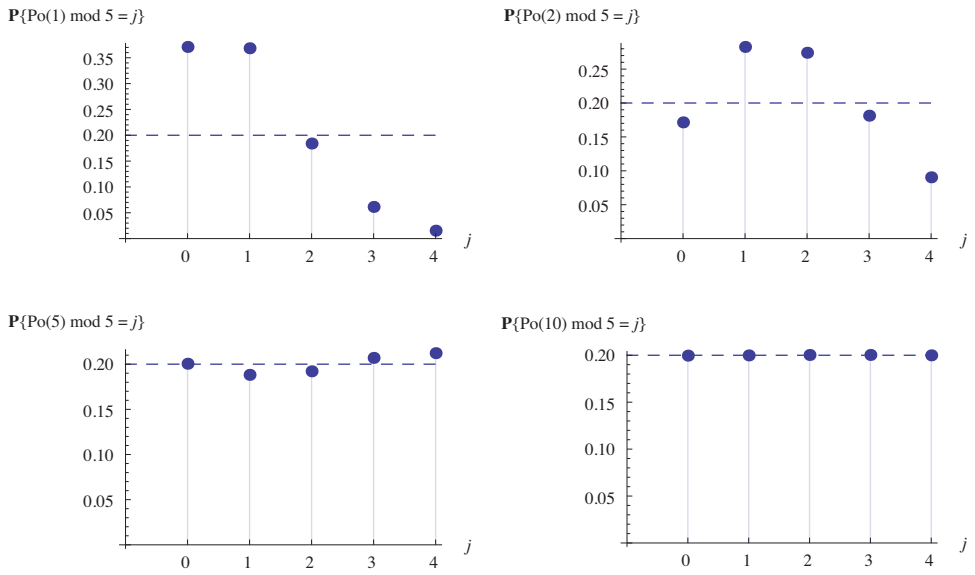


***Figure 4:*** *Distribution of Poisson residues modulo* $5$.

confined to the Poisson or the binomial case.

In the previous examples the underlying distributions had finite moments of all orders. The next example is designed to show that existence of moments is not necessary for the phenomenon of convergence to uniformity for the distribution of $X \bmod k$ to prevail.

**Example 2.3.** *(The zeta distribution and connections to Stieltjes constants).* Suppose, given $z > 1$, the arithmetic random variable $X$ has the mass function

$$P(X = n) = \frac{1}{\zeta(z)\, n^z} \qquad (n = 1, 2, 3, \dots), \tag{3}$$

where $\zeta(z)$ is the Riemann zeta function. When $z \to 1$, these distributions *shift to the right* in the sense that the probability of any compact set of integers converges to zero as $z \to 1$. The distribution has finite expectation only for $z > 2$.

Now fix any $k \geq 2$ and $0 < j \leq k - 1$. For $j = 0$, the calculation below needs a very small obvious modification. Then,

$$P(X \bmod k = j) = \frac{1}{\zeta(z)} \sum_{n=0}^{\infty} \frac{1}{(kn + j)^z} = \frac{1}{k^z\, \zeta(z)} \sum_{n=0}^{\infty} \frac{1}{(n + \frac{j}{k})^z} = \frac{\zeta(z, \frac{j}{k})}{k^z\, \zeta(z)}, \tag{4}$$

where $\zeta(z,q)$ denotes the Hurwitz zeta function $\sum_{n=0}^{\infty}\frac{1}{(n+q)^z}, q>0, z>1$. Using the known facts

$$\lim_{z\to 1}(z-1)\zeta(z)=1 \quad \text{and} \quad \lim_{z\to 1}\left[\zeta(z,q)-\frac{1}{z-1}\right]=-\psi(q),$$

where $\psi$ denotes the digamma function, we get, for any $k\geq 2$ and $j>0$,

$$P(X \bmod k=j)=\frac{\frac{1}{z-1}-\psi(q)+o(1)}{k^z\,\frac{1}{z-1}(1+o(1))}=\frac{1-(z-1)\psi(q)+o(z-1)}{k^z\,(1+o(1))}\to\frac{1}{k},$$

as $z\to 1$. The limit is $1/k$ for the case $j=0$ as well for trivial reasons as in this case $P(X \bmod k=0)=k^{-z}$. The case $X \bmod 5$ is illustrated in Figure 5.

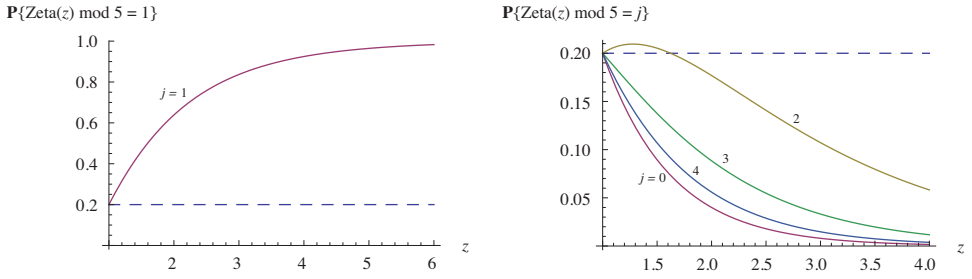


***Figure 5:*** *Convergence to the uniform of the distribution of zeta residues modulo* $5$ *as* $z\downarrow 1$.

We can obtain an asymptotic expansion for $P(X \bmod k=j)$ via classical analysis. Begin with the Laurent series expansions of the Riemann and the Hurwitz zeta functions:

$$\zeta(z)=\frac{1}{z-1}+\sum_{n=0}^{\infty}\frac{(-1)^n\gamma_n}{n!}\,(z-1)^n, \tag{5}$$

$$\zeta(z,q)=\frac{1}{z-1}+\sum_{n=0}^{\infty}\frac{(-1)^n\gamma_n(q)}{n!}\,(z-1)^n. \tag{6}$$

The terms $\gamma_n,\gamma_n(q)$ in these series expansions are the celebrated *Stieltjes constants*. For example, $\gamma_0=C$ is Euler's constant, and $\gamma_0(q)=-\psi(q)$ where, once again $\psi$ is the digamma function. The first few Stieltjes constants $\gamma_n$ are shown to four decimal places in Table 1, as also values of $\gamma_n(j/5)$ for the case $k=5$ [for more extensive tables see Finch (2003) and Musser (2011), for example].

Combining (4), (5), and (6), and writing $k^{-z}=e^{-z\log k}=\frac{1}{k}\,e^{-(z-1)\log k}$, a little algebra gives us a two term asymptotic expansion

$$P(X \bmod k=j)=k^{-1}\Big[1+\big\{\gamma_0\big(\tfrac{j}{k}\big)-\gamma_0-\log(k)\big\}(z-1)+\Big\{\tfrac{1}{2}\log(k)^2 \\ +\big\{\gamma_0-\gamma_0\big(\tfrac{j}{k}\big)\big\}\big(\log(k)+\gamma_0\big)+\gamma_1-\gamma_1\big(\tfrac{j}{k}\big)\Big\}(z-1)^2+O\big((z-1)^3\big)\Big]. \tag{7}$$

For the case $k=5$, for instance, the coefficients needed to carry out (7) may be read out from Table 1.

| $n$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $\gamma_n$ | 0.5772 | −0.0728 | −0.0097 | 0.0021 | 0.0023 |

| $j$ | $\gamma_0(j/5)$ | $\gamma_1(j/5)$ | $\gamma_2(j/5)$ | $\gamma_3(j/5)$ | $\gamma_4(j/5)$ |
|---|---|---|---|---|---|
| 1 | 5.2890 | −8.0302 | 12.9407 | −20.8487 | 33.5483 |
| 2 | 2.5614 | −2.2528 | 2.1017 | −1.9268 | 1.7603 |
| 3 | 1.5406 | −0.8308 | 0.4456 | −0.2210 | 0.1120 |
| 4 | 0.9650 | −0.2982 | 0.0691 | −0.0094 | 0.0040 |

***Table 1:*** *Some Stieltjes constants.*

**Example 2.4.** *(An example with exceptional structure: the geometric distribution).* Let $Y$ have the exponential distribution with mean $\lambda$, and let $X$ be its integer part. Then $X$ has the waiting time mass function

$$P(X = n) = w(n; 1 - e^{-1/\lambda}) = e^{-n/\lambda} - e^{-(n+1)/\lambda} = e^{-n/\lambda}(1 - e^{-1/\lambda}) \qquad (n \geq 0),$$

that is to say, $X \sim$ Geometric$(1 - e^{-1/\lambda})$ has the geometric distribution with parameter $p = 1 - e^{-1/\lambda}$. By direct summing we see now that

$$P(X \bmod k = j) = \sum_{n=0}^{\infty} w(kn + j; 1 - e^{-1/\lambda}) = \frac{(e^{1/\lambda} - 1)\, e^{(k-j-1)/\lambda}}{e^{k/\lambda} - 1}$$
$$= \frac{e^{(k-j-1)/\lambda}}{1 + e^{1/\lambda} + e^{2/\lambda} + \cdots + e^{(k-1)/\lambda}}. \quad (8)$$

It is apparent from (8) that, for any given $k$ and $\lambda$, $P(X \bmod k = j)$ is monotone decreasing in $j$ [see Figure 6]. This is not true, for example, in the Poisson case.

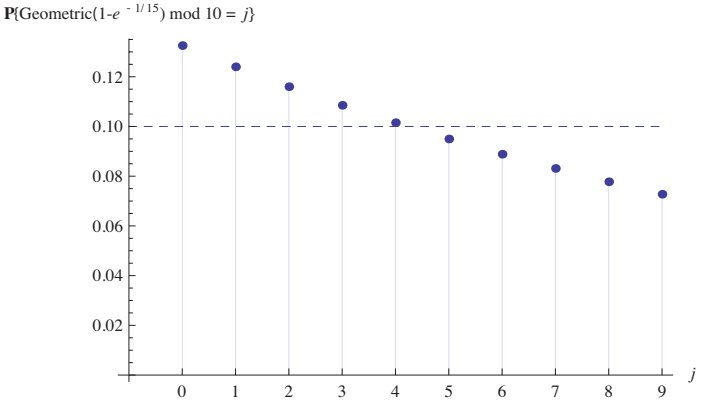


***Figure 6:*** *Monotone behaviour with $j$ for the distribution of geometric residues modulo* 10.

Certain monotonicities in $\lambda$ are also visible; for instance, for a given $k$, a simple differentiation shows that, as $\lambda$ increases, $P(X \bmod k = 0)$ is decreasing while $P(X \bmod k = k-1)$ is increasing; non-monotone behaviour is possible for intermediate values of $j$ [see Figure 7].

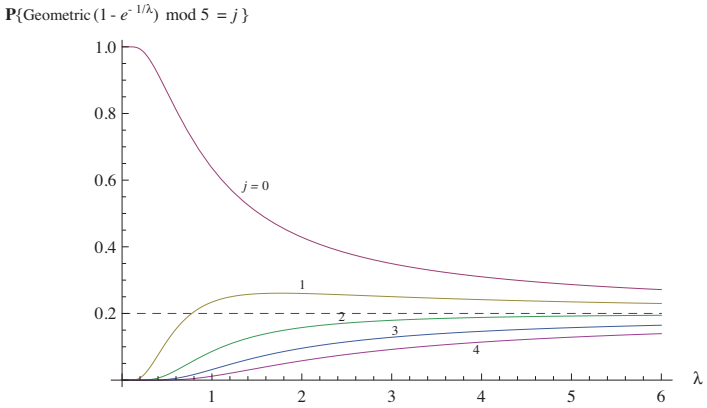


***Figure 7:*** *Convergence to the uniform of the distribution of geometric residues modulo* $5$.

We can investigate the asymptotics in $\lambda$ by a Taylor series expansion of the exponentials in (8). By a simple calculation, one gets

$$P(X \bmod k = j) = \frac{1}{k} + \frac{k-2j-1}{2k\,\lambda} + \frac{k^2-3k+2+6j(j+1-k)}{12k\,\lambda^2} + O(\lambda^{-3}).$$

Hence, except when $k$ is odd and $j = \frac{k-1}{2}$, convergence to uniformity is at the slow rate of $\frac{1}{\lambda}$. This is also in contrast to the Poisson case. In Figure 6 we also see illustrated the significant practical deviation from uniformity for the distribution of $X \bmod 10$ at the large value of $\lambda = 15$.

**Example 2.5.** *(A case of failure).* There are fairly natural integer-valued random variables for which the phenomenon of convergence to uniformity of the distribution of $X \bmod k$ fails. On introspection, this is not surprising. But as a simple concrete example, take $X$ to be the square of a Poisson random variable $Y$ with mean $\lambda$. Since $X$ is even if, and only if, $Y$ is even, it follows that

$$P(X \bmod 2 = 0) = e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^{2j}}{(2j)!} = e^{-\lambda} \left( \frac{e^{\lambda} - e^{-\lambda}}{2} \right) = \frac{1}{2} - \frac{1}{2} e^{-2\lambda} \to \frac{1}{2} \qquad (\lambda \to \infty).$$

Likewise, $P(X \bmod 2 = 1)$ of course also goes to $1/2$ asymptotically. However, convergence to uniformity fails for the distribution of $X \bmod 3$. Since $3n+2$ can never be a perfect square for an integer $n$, we do not have convergence to uniformity in this case. Indeed, in this case,

$$P(X \bmod 3 = 0) = \frac{1}{3} + \frac{2}{3} e^{-3\lambda/2} \cos \frac{\lambda\sqrt{3}}{2} \to \frac{1}{3},$$

$$P(X \bmod 3 = 1) = \frac{2}{3} - \frac{2}{3} e^{-3\lambda/2} \cos \frac{\lambda\sqrt{3}}{2} \to \frac{2}{3},$$

$$P(X \bmod 3 = 2) = 0.$$

Basically, $X \bmod k$ will converge in law to $U_k^2 \bmod k$ where $U_k$ denotes a variable uniform on $\{0, 1, \ldots, k-1\}$; however, see Theorem 4.5 for a crisper description. Table 2 lists the limiting distributions of $X \bmod k$ for some specific values of $k$.

| | $j$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $k$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 2 | $\frac{1}{2}$ | $\frac{1}{2}$ | | | | | | | | |
| 3 | $\frac{1}{3}$ | $\frac{2}{3}$ | 0 | | | | | | | |
| 4 | $\frac{1}{2}$ | $\frac{1}{2}$ | 0 | 0 | | | | | | |
| 5 | $\frac{1}{5}$ | $\frac{2}{5}$ | 0 | 0 | $\frac{2}{5}$ | | | | | |
| 10 | $\frac{1}{10}$ | $\frac{1}{5}$ | 0 | 0 | $\frac{1}{5}$ | $\frac{1}{10}$ | $\frac{1}{5}$ | 0 | 0 | $\frac{1}{5}$ |

***Table 2:*** *Limiting distributions of $Y^2 \bmod k$ when $Y$ is Poisson.*

# 3 Exact Formulas for Poisson and Other Standard Cases

We treat the important Poisson, binomial, and negative binomial cases in detail in this section. It turns out that in each of these important special cases, we can write closed form formulas for $P(X \bmod k = j)$ for every fixed parameter value of the distribution, and for every fixed $k$ and $j$. These exact formulas are in terms of hypergeometric functions. These exact formulations lend themselves to a variety of direct proofs of asymptotic uniformity and we explore some of the ramifications. These proofs bring out some special features and special connections that are not present in general. We start with the Poisson case.

## 3.1 Exact and Asymptotic Distributions of Poisson mod $k$

Let $X \sim$ Poisson$(\lambda)$. We prove, using a few different methods, that the distribution of $X \bmod k$ converges weakly to the discrete uniform distribution on the set $\{0, 1, \cdots, k-1\}$. Each proof has its own virtue, each making a connection to very different fields of mathematics: the first proof exhibits a Tauberian link via an exact formula for the probabilities; the second proof exposes a Fourier connection via an identity for characteristic functions; and the third proof exploits a Markov chain embedding. The explicit closed

form formulas that we provide for the distribution of $X$ mod $k$ are of independent interest.

**Theorem 3.1.** *Let $X \sim$ Poisson$(\lambda)$ and, in the usual notation, let ${}_0F_{k-1}$ denote the generalized hypergeometric function ${}_pF_q$ with parameters $p = 0$ and $q = k - 1$. Then, for each pair of integers $j$ and $k$ satisfying $k \geq 2$ and $0 \leq j \leq k - 1$, one has the exact formula*

$$\begin{aligned} p_j(\lambda) &:= P(X \bmod k = j) \\ &= \frac{e^{-\lambda}\lambda^j}{j!}\, {}_0F_{k-1}\left(\frac{j+1}{k}, \frac{j+2}{k}, \ldots, \frac{k-1}{k}, \frac{k+1}{k}, \ldots, \frac{j+k}{k}; \frac{\lambda^k}{k^k}\right). \quad (9) \end{aligned}$$

*Proof.* For the given choices of $j$, $k$, and $\lambda$, a little algebraic massaging shows that

$$\begin{aligned} p_j(\lambda) = e^{-\lambda}\sum_{r=0}^{\infty}\frac{\lambda^{kr+j}}{(kr+j)!} = \frac{e^{-\lambda}\lambda^j}{j!}\Bigg[1 + \frac{k}{j+1}\cdot\frac{k}{j+2}\cdots\frac{k}{j+k}\cdot\frac{\lambda^k}{k^k} \\ + 2\frac{k}{j+1}\cdot\frac{k}{j+k+1}\cdot\frac{k}{j+2}\cdot\frac{k}{j+k+2}\cdots\frac{k}{j+k}\cdot\frac{k}{j+2k}\cdot\frac{(\frac{\lambda}{k})^{2k}}{2!} + \cdots\Bigg] \end{aligned}$$

and we may simply identify the expression in square brackets on the right with the hypergeometric function ${}_0F_{k-1}$ appearing in (9). □

For special values of $k$, the exact formula (9) simplifies; here is a record of it.

**Corollary 3.1.** *Let $X \sim$ Poisson$(\lambda)$. Then:*

$$P(X \bmod 2 = j) = \begin{cases} e^{-\lambda}\cosh\lambda & \text{if } j = 0, \\ e^{-\lambda}\sinh\lambda & \text{if } j = 1; \end{cases}$$

$$P(X \bmod 3 = j) = \begin{cases} \frac{1}{3} + \frac{2}{3}\,e^{-3\lambda/2}\cos\left(\frac{\lambda\sqrt{3}}{2}\right) & \text{if } j = 0, \\ \frac{1}{3} + \frac{2}{3}\,e^{-3\lambda/2}\sin\left(\frac{\lambda\sqrt{3}}{2} - \frac{\pi}{6}\right) & \text{if } j = 1, \\ \frac{1}{3} - \frac{2}{3}\,e^{-3\lambda/2}\sin\left(\frac{\lambda\sqrt{3}}{2} + \frac{\pi}{6}\right) & \text{if } j = 2; \end{cases}$$

$$P(X \bmod 4 = j) = \begin{cases} e^{-\lambda}\frac{\cos\lambda + \cosh\lambda}{2} & \text{if } j = 0, \\ e^{-\lambda}\frac{\sin\lambda + \sinh\lambda}{2} & \text{if } j = 1, \\ e^{-\lambda}\frac{\cosh\lambda - \cos\lambda}{2} & \text{if } j = 2, \\ e^{-\lambda}\frac{\sinh\lambda - \sin\lambda}{2} & \text{if } j = 3. \end{cases}$$

**Theorem 3.2.** *For any $k$ and $0 \leq j \leq k - 1$, we have $P(X \bmod k = j) \to 1/k$ as $\lambda \to \infty$.*

Proof 1: First, the Tauberian approach. For this we show that $p_j(\lambda)$ converges in mean under a suitably chosen sequence of weight functions $g_\mu(\lambda)$, i.e., $\int_0^\infty p_j(\lambda) g_\mu(\lambda)\, d\lambda \to 1/k$ when $\mu \to \infty$. To demonstrate that convergence in mean in this setting will indeed imply that $p_j(\lambda) \to 1/k$ when $\lambda \to \infty$, we appeal to the following Tauberian theorem [Hardy (1956, Theorem 221, p. 286)]; we state it as a lemma.

**Lemma 3.1.** Suppose $f$ is bounded and slowly oscillating at $\infty$, and that the Fourier transform of $g$ has no real zeroes. Suppose $\int g(\theta - t) f(t)\, dt \to l$ as $\theta \to \infty$. Then $f(x) \to l$ as $x \to \infty$.

In order to use Lemma 3.1, we let $g(t) = e^{-t} e^{-e^{-t}}$, $f(t) = p_j(e^t)$ where $p_j(\cdot)$ is the function in (9), and $\theta = \log \mu$ with $\mu > 0$. Then, by a change of variable,

$$\int_{-\infty}^{\infty} g(\theta - t) f(t)\, dt = \int_{-\infty}^{\infty} e^{t-\theta} e^{-e^{t-\theta}} p_j(e^t)\, dt = \int_0^\infty \frac{1}{\mu} e^{-\frac{\lambda}{\mu}} p_j(\lambda)\, d\lambda$$
$$= \frac{1}{\mu\, j!} \int_0^\infty \lambda^j e^{-\lambda(\mu+1)/\mu}\, {}_0F_{k-1}\left( \frac{j+1}{k}, \frac{j+2}{k}, \ldots, \frac{k-1}{k}, \frac{k+1}{k}, \ldots, \frac{j+k}{k}; \frac{\lambda^k}{k^k} \right) d\lambda$$

The integral on the right, curiously, simplifies to an exact form and, after simplification, we obtain the formula

$$\int_{-\infty}^{\infty} g(\theta - t) f(t)\, dt = \frac{\mu^j (\mu+1)^{k-j-1}}{(\mu+1)^k - \mu^k}. \tag{10}$$

Note that for the right-hand side of (10) to tend to $1/k$ as $\mu \to \infty$ it is necessary and sufficient that $\theta = \log \mu \to \infty$.

Now observe that $g(t)$ is the density of the standard Gumbel distribution which is infinitely divisible. Hence, its Fourier transform *cannot have any real zeroes*. With this, the only thing left to verify is that the function $f(t)$ is slowly oscillating, and then Lemma 3.1 applies and we have our desired result $f(x) \to 1/k$ as $x \to \infty$, which means $p_j(\lambda) \to 1/k$ as $\lambda \to \infty$.

Now, a sufficient condition for $f$ to be slowly oscillating is that $f'(t) = O(1)$ [see Hardy (1956, p. 287)]. But, since $f(t) = p_j(e^t)$, we have $f'(t) = e^t\, p_j'(e^t)$, and so, we need to show that $\lambda p_j'(\lambda) = O(1)$. Now, directly from the definition of $p_j(\lambda)$, by term-by-term differentiation of the absolutely convergent power series, we obtain

$$p_j'(\lambda) = -e^{-\lambda} \lambda^j \sum_{r=0}^{\infty} \frac{\lambda^{kr}}{(kr)!} + j e^{-\lambda} \lambda^{j-1} \sum_{r=0}^{\infty} \frac{\lambda^{kr}}{(kr)!} + e^{-\lambda} \lambda^j \sum_{r=1}^{\infty} \frac{kr \lambda^{kr-1}}{(kr)!}$$
$$= -p_j(\lambda) + \frac{j}{\lambda} p_j(\lambda) + e^{-\lambda} \lambda^j \sum_{r=1}^{\infty} \frac{(kr + j - j) \lambda^{kr-1}}{(kr)!}$$

$$= -p_j(\lambda) + \frac{j}{\lambda} p_j(\lambda) + e^{-\lambda}\lambda^{j-1}\left[\sum_{r=0}^{\infty} \frac{\lambda^{kr}}{(kr+j-1)!} - \frac{1}{(j-1)!}\right]$$
$$- je^{-\lambda}\lambda^{j-1}\left[\sum_{r=0}^{\infty} \frac{\lambda^{kr}}{(kr+j)!} - \frac{1}{j!}\right]$$
$$= -p_j(\lambda) + \frac{j}{\lambda} p_j(\lambda) + p_{j-1}(\lambda) - \frac{e^{-\lambda}\lambda^{j-1}}{(j-1)!} - \frac{j}{\lambda} p_j(\lambda) + \frac{e^{-\lambda}\lambda^{j-1}}{(j-1)!}.$$

After cancellations on the right-hand side, we hence obtain the rather remarkable identity

$$p_j'(\lambda) = p_{j-1}(\lambda) - p_j(\lambda) \tag{11}$$

valid formally for $1 \le j \le k-1$. We may extend the validity of (11) to $j = 0$ as well by interpreting subscripts modulo $k$ so that $0 - 1 \equiv k - 1 \pmod k$: thus, for $j = 0$, the corresponding identity is

$$p_0'(\lambda) = p_{k-1}(\lambda) - p_0(\lambda). \tag{12}$$

We will see shortly that we may interpret (11,12) from a Markov process standpoint. From (11,12), for any given $j, k$, one gets that $p_j'(\lambda) = O(e^{-\epsilon\lambda})$ for some suitable $\epsilon = \epsilon_{j,k} > 0$. Hence, $\lambda p_j'(\lambda) = o(1)$ as $\lambda \to \infty$, which means that it is also $O(1)$. Thus, the function $f(t) = e^t\, p_j(e^t)$ is slowly oscillating at $\infty$. This completes *Proof 1* of Theorem 3.2. □

Proof 2: This is an efficient Fourier analytic proof of the result which has the advantage of applying more generally but which, in a sense, may lack a little in mathematical intuition. This approach should be compared to the material on moments in Section 4.5.

Let $\psi_X(t) = \mathbf{E}[e^{itX}]$ be the characteristic function of $X$. By inverting the transform, we obtain the Fourier identity

$$p_j(\lambda) = \frac{1}{k}\sum_{r=0}^{k-1} e^{-i2\pi jr/k}\psi_X\left(\frac{2\pi r}{k}\right) = \frac{1}{k} + \frac{1}{k}\sum_{r=1}^{k-1} e^{-i2\pi jr/k}\psi_X\left(\frac{2\pi r}{k}\right) \tag{13}$$

which applies in general. In our Poisson case, $\psi_X(t) = e^{\lambda(e^{it}-1)}$, and so the identity (13) specializes to

$$p_j(\lambda) = \frac{1}{k} + \frac{1}{k}\sum_{r=1}^{k-1} e^{-i2\pi jr/k}\left[e^{\lambda\{\cos(2\pi r/k)-1\}}\left\{\cos\left(\lambda\sin\frac{2\pi r}{k}\right) + i\sin\left(\lambda\sin\frac{2\pi r}{k}\right)\right\}\right]. \tag{14}$$

Now, for each $r$ between 1 and $k-1$, we have $\cos(2\pi r)/k < 1$, and hence each of the terms in the sum on the right in (14) tends to 0 as $\lambda \to \infty$. It follows that $P(X \bmod k = j) \to 1/k$ as $\lambda \to \infty$. □

Proof 3: This is also a short proof using a Markov chain convergence argument. Fix $k \ge 2$. Writing $t$ for time, let $N(t)$, $t \ge 0$, be a Poisson process with a constant rate equal

to 1, and let $\tilde{X}(t) = N(t) \bmod k$ for $t \geq 0$. Then $\tilde{X}(t)$ is a continuous time Markov chain on the state space $\{0, 1, \cdots, k-1\}$. Using the same notation as in Ross (1995, p. 251), the embedded discrete time chain has the transition probabilities

$$P_{i,j} = \begin{cases} 1 & \text{if } i = 0, 1, \ldots, k-2 \text{ and } j = i+1, \\ 1 & \text{if } i = k-1 \text{ and } j = 0. \end{cases}$$

Furthermore, the departure rates $\nu_i$ corresponding to the $k$ different states are each equal to 1. Hence a unique solution of the system of equations $\pi_j = \sum_i \pi_i P_{i,j}$ satisfying the constraint $\sum_i \pi_i = 1$ is $\pi_j \equiv 1/k$. Now, the limiting probabilities are characterized by

$$p_j := \lim_{t \to \infty} P_{i,j}(t) = \frac{\pi_j/\nu_j}{\sum_i \pi_i/\nu_i} \equiv \frac{1}{k},$$

which is the desired result. □

We remark that the identities (11,12) that we derived above from first principles happen to be exactly the Kolmogorov forward equations for the process $\tilde{X}(t)$ [see Ross (1995, pp. 246–250), or Feller (1971, p. 328)].

## 3.2 Nonuniformity when $k \to \infty$ with $\lambda$

If $X \sim$ Poisson($\lambda$), then the distribution of $X \bmod k$ converges weakly to the discrete uniform distribution on the set $\{0, 1, \ldots, k-1\}$ for every fixed $k$ when $\lambda \to \infty$. However, this is not necessarily true, when $k$ grows with $\lambda$ and also goes to $\infty$. Depending on the rate of growth of $k$ as $\lambda \to \infty$, various types of limiting distributions are possible with an appropriate centering and scaling of $X \bmod k$. For example, if $k = \lambda$, then the distribution of $X \bmod k$ is approximated by a mixture of two half-normal distributions, one centered at $k$ and the other centered at zero. On the other hand, if $k \gg \lambda$, then $X \bmod k$ will have an approximately normal distribution with a suitable centering $\mu(k)$ and scaling $\sigma(k)$. Figure 8 contains representative plots illustrating the various behaviours that arise.

## 3.3 Poisson Powers and Connections to Number Theory

We showed by an example (Example 2.5) that convergence to uniformity on the residue class $\{0, 1, \ldots, k-1\}$ fails for the square of a Poisson. In this section we provide an explicit general result for Poisson powers when $k$ is a prime; if $k$ is not a prime, the result is not explicit. We begin by introducing some notation.

Consider the equation $x^r = j \bmod p$. Here $j = 1, 2, \ldots, p-1$ and we assume $p$ to be a prime. Let $d$ be the greatest common divisor of $r$ and $p-1$. Let $g$ be any generator of

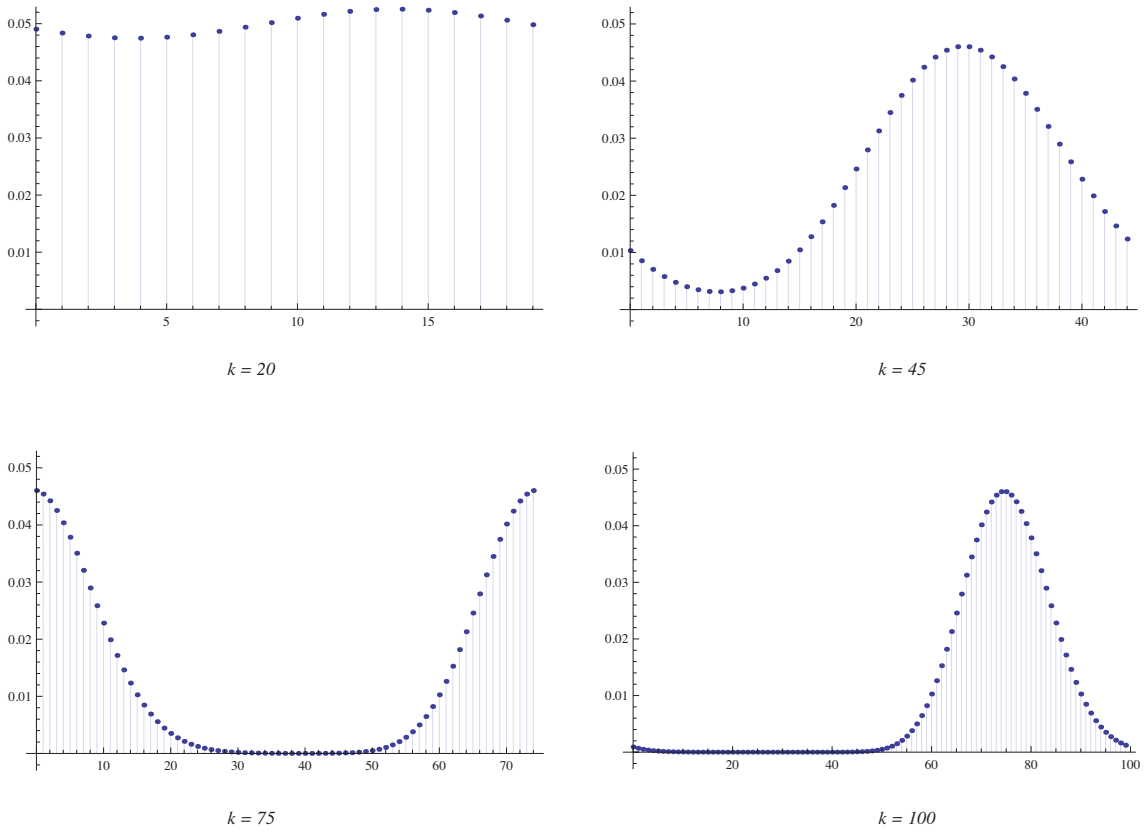


***Figure 8:*** *The change in the form of the limiting distribution in the case of the Poisson distribution of mean* $75$ *when* $k$ *runs through the values* $20$*,* $45$*,* $75$*, and* $100$*.*

the multiplicative quotient group $\mathcal{Z}/p\mathcal{Z}$. Let $n$ be the degree of $j$; thus, $g^n = j \bmod p$. Define the function

$$N(j) = \begin{cases} 1 & \text{if } n = 0, \\ d & \text{if } n \neq 0 \text{ and } n = 0 \bmod d, \\ 0 & \text{if } n \neq 0 \bmod d. \end{cases}$$

With this for preparation, here is our general result.

**Theorem 3.3.** *Suppose* $Y \sim$ *Poisson*$(\lambda)$ *and let* $X = Y^r$ *for* $r \geq 2$*. Let* $k \geq 2$ *and, modifying our earlier notation, let* $U_k$ *denote a random variable distributed uniformly on* $\{0, 1, \ldots, k-1\}$*. Then:*

*(a) For general* $k$*,* $X \bmod k \overset{\mathcal{D}}{\Rightarrow} U_k^r \bmod k$ *as* $\lambda \to \infty$*.*

*(b) If* $k = p$ *is prime then* $P(X \bmod p = j) \to N(j)/p$ *as* $\lambda \to \infty$ *for each integer* $0 \leq j \leq p-1$*.*

The explicit formulation in part (b) follows from equation (4.86), p. 100, in the proof of Theorem 4.4.8 in Miller and Takloo-Bighash (2006).

We illustrate Theorem 3.3 with an example (also refer to Example 2.5).

**Example 3.1.** Let $r = 5$ and $k = p = 11$. Then $d = 5$. A generator of $\mathcal{Z}/11\mathcal{Z}$ is $g = 2$. Table 3 lists the degrees $n$ corresponding to each $j$ for the congruence $2^n = j \bmod 11$. By inspection of the elements of the table we see that $n \neq 0 \bmod 5$ for $j \in$

| $j$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | 10 | 1 | 8 | 2 | 4 | 9 | 7 | 3 | 6 | 5 |

***Table 3:*** $2^n = j \bmod 11$.

$\{2, 3, 4, 5, 6, 7, 8, 9\}$, while $n \neq 0$ and $n = 0 \bmod 5$ for $j \in \{1, 10\}$. Thus, as $\lambda \to \infty$,

$$P(X = j \bmod 11) \to \begin{cases} 1/11 & \text{if } j = 0, \\ 5/11 & \text{if } j \in \{1, 10\}, \\ 0 & \text{if } j \in \{2, 3, 4, 5, 6, 7, 8, 9\}. \end{cases}$$

We see hence that the limit of the fifth power is drastically nonuniform.

There is one interesting special case where the power would still be asymptotically uniform on the residue class $\{0, 1, \ldots, k-1\}$. We present the result for the Poisson case but it generalizes to suitable non-Poisson cases as well.

**Theorem 3.4.** *Suppose $Y \sim$ Poisson$(\lambda)$ and let $X = Y^p$ where $p$ is a prime. Let $0 \leq j \leq p-1$. Then $P(X \bmod p = j) \to 1/p$ as $\lambda \to \infty$.*

*Proof.* By Fermat's little theorem, if $p$ is a prime, then for any integer $n$ we have $n^p \equiv n \pmod{p}$, whence $P(X \bmod p = j) = P(Y \bmod p = j) \to 1/p$ by Theorem 3.2. □

Theorem 3.4 generalizes to certain polynomials of multiple Poisson variables, a special case being convolutions of Poisson powers. Theorem 3.5 below will generalize to many non-Poisson cases.

**Theorem 3.5.** *Let $Y_1, \ldots, Y_m$ be a sequence of independent random variables where $Y_i \sim$ Poisson$(\lambda_i)$, and suppose $X_i = Y_i^p$ where $p$ is a prime. Let $f(y_1, \ldots, y_m)$ be a polynomial in $(y_1, \ldots, y_m)$ with integer coefficients. Suppose that for $0 \leq j \leq p-1$, we have*

$$P\big(f(Y_1, \ldots, Y_m) \bmod p = j\big) \to 1/p \tag{15}$$

*asymptotically as $\min_i \lambda_i \to \infty$. Then*

$$P\big(f(X_1, \ldots, X_m) \bmod p = j\big) \to 1/p$$

*for each $j = 0, 1, \ldots, p-1$.*

**Corollary 3.2.** *With notation as in Theorem 3.5, suppose only that* $\sum_{i=1}^{m} \lambda_i \to \infty$. *Then*

$$P\left(\sum_{i=1}^{m} X_i \bmod p = j\right) \to 1/p$$

*for each* $j = 0, 1, \ldots, p-1$.

*Proof.* We use the following strong generalization of Fermat's little theorem [see Pólya and Szego (1998, Vol. II, p. 148), or LeVeque (1996, p. 57)]: for any prime $p$ and any polynomial $f(y_1, \ldots, y_m)$ with integer coefficients, $[f(y_1, \cdots, y_m)]^p \equiv f(y_1^p, \cdots, y_m^p) \pmod{p}$. The proof follows easily by using the hypothesis (15) of the theorem. □

Notice that in the specialization of the theorem to convolutions, not all of the $\lambda_i$ need go to $\infty$; it is enough if one of them does.

## 3.4 The Binomial and the Negative Binomial Cases

The binomial and the negative binomial cases exhibit one significant point of difference with the Poisson case. In the Binomial$(n, p)$ case convergence to uniformity occurs in two separate asymptotic paradigms: 1) $p$ is fixed and $n \to \infty$, and 2) $p = p(n) \to 0$ and $np \to \infty$. The second paradigm resembles the Poisson case. Likewise, in the NB$(r, p)$ case convergence to uniformity occurs in the two asymptotic paradigms, 1) $p$ is fixed and $r \to \infty$, and 2) $r$ is fixed and $p \to 0$ (a necessary and sufficient condition is provided in (17) below). A special case of the latter asymptotic paradigm is seen in the case of the geometric distribution with $p \to 0$.

On the principle that it is always good to possess an exact formula, in the case of the negative binomial we derive an explicit expression for $P(X \bmod k = j)$ for each fixed $r$, $p$, $j$, $k$. We have not, however, been successful in obtaining a similar clean, exact formula in the case of the binomial except for special values of $k$.

**Theorem 3.6.** *Let* $X \sim NB(r, p)$ *and let* $q = 1 - p$. *Let* $k \geq 2$ *be any given integer. Then, for* $0 \leq j \leq k-1$,

$$P(X \bmod k = j) = \tbinom{r+j-1}{r-1} p^r (1-p)^j \\ \times {}_kF_{k-1}\left(\tfrac{r+j}{k}, \tfrac{r+j+1}{k}, \ldots, \tfrac{r+j+k-1}{k}; \tfrac{j+1}{k}, \tfrac{j+2}{k}, \ldots, \tfrac{k-1}{k}, \tfrac{k+1}{k}, \ldots, \tfrac{k+j}{k}; (1-p)^k\right) \quad (16)$$

*where the usual accommodations have to be made for the end cases* $j = 0$ *and* $j = k-1$ *where many of the terms in the argument of the hypergeometric distribution vanish.*

We omit the proof as the derivation is similar to that of (9). For specific $r$ and $k$, formula (16) simplifies. Here are a few specific cases as examples.

**Corollary 3.3.** *Let $X_r \sim NB(r,p)$ and let $q = 1-p$. Then:*

*a)* $P(X_1 \bmod k = j) = \dfrac{pq^j}{1-q^k} \qquad (0 \le j \le k-1).$

*b)* $P(X_2 \bmod k = j) = \dfrac{p^2q^j\big(j+1+q^k(k-j-1)\big)}{(1-q^k)^2} \qquad (0 \le j \le k-1).$

*c)* $P(X_3 \bmod 2 = j) = \begin{cases} \frac{(1+3q^2)}{(1+q)^3} & \text{if } j=0, \\ \frac{q(3+q^2)}{(1+q)^3} & \text{if } j=1. \end{cases}$

*d)* $P(X_3 \bmod 3 = j) = \begin{cases} \frac{(1+7q^3+q^6)}{(1+q+q^2)^3} & \text{if } j=0, \\ \frac{3q(1+2q^3)}{(1+q+q^2)^3} & \text{if } j=1, \\ \frac{3q^2(2+q^3)}{(1+q+q^2)^3} & \text{if } j=2. \end{cases}$

*e)* $P(X_3 \bmod 4 = j) = \begin{cases} \frac{(1+12q^4+3q^8)}{(1+q+q^2+q^3)^3} & \text{if } j=0, \\ \frac{q(3+12q^4+q^8)}{(1+q+q^2+q^3)^3} & \text{if } j=1, \\ \frac{2q^2(3+5q^4)}{(1+q+q^2+q^3)^3} & \text{if } j=2, \\ \frac{2q^3(5+3q^4)}{(1+q+q^2+q^3)^3} & \text{if } j=3. \end{cases}$

We now present the corresponding convergence theorem.

**Theorem 3.7.** *Let $X \sim NB(r,p)$. Then $P(X \bmod k = j) \to 1/k$ for each $k \ge 2$ and $0 \le j \le k-1$ if, and only if, for each real number $a \in [-1,1)$,*

$$\left(\frac{p}{\sqrt{(1-qa)^2+q^2(1-a^2)}}\right)^r \to 0. \tag{17}$$

*Proof.* The approach uses Fourier methods. Fix $0 < \theta < 2\pi$, and denote $a = \cos\theta$. Bear in mind that $a < 1$. Suppose $X \sim \text{NB}(r,p)$ and denote its characteristic function by $\psi_X(t)$. With $q = 1-p$ as before, note that

$$\psi_X(t) = \left(\frac{p}{1-qe^{it}}\right)^r.$$

By the Fourier identity (13), $P(X \bmod k = j) \to 1/k$ for each fixed $k$ and each $0 \le j < k$ if, and only if,

$$\left|\left(\frac{p}{1-qe^{it}}\right)^r\right| \to 0.$$

But this condition may be rewritten in the form

$$\left(\frac{p}{\sqrt{1-2q\cos t+q^2}}\right)^r \to 0$$

which, in turn, is equivalent to the statement (17) to be shown. □

In particular, (17) holds for each fixed $r \geq 1$ if $p \to 0$, or for each fixed $p$ if $r \to \infty$.

**Theorem 3.8.** *Let $X \sim$ Binomial$(n, p)$ and let $q = 1-p$. Then $P(X \bmod k = j) \to 1/k$ for each $k \geq 2$ and each $0 \leq j \leq k-1$ if, and only if, for each real number $a \in [-1, 1)$,*

$$\big((pa+q)^2 + p^2(1-a^2)\big)^n \to 0. \tag{18}$$

The proof of (18) follows the same lines as that of (17) by exploiting the fact that the characteristic function of a Binomial$(n, p)$ distribution is given by $\psi_X(t) = (1-p+pe^{it})^n$. We omit the details.

# 4 General Theory and General Convergence Theorems

Having treated the important standard lattice distributions in Section 3, we now move on to generalities. The purpose of this section is three-fold: (a) Obtain a general expression for $P(X \bmod k = j)$ for a more or less general lattice random variable $X$, by using the classic tool of the Euler-Maclaurin summation formula. (b) For a sequence of lattice random variables $X_n$ having a convolution structure, draw an analogy to the famous Weyl equidistribution theorem, and exploit this analogy to write a nearly necessary and sufficient condition for convergence of $X_n \bmod k$ to uniformity. (c) Again, for a sequence of lattice random variables $X_n$, give a widely applicable sufficient condition for convergence to uniformity by using the modal properties of the mass functions of $X_n$.

Because of the technical nature of the proofs of all these general theorems, we have given these proofs in a separate appendix.

We start with a general theorem on the convergence of the distribution of a sequence of integer-valued random variables $X_n$ mod $k$ to the uniform distribution on $\{0, 1, \ldots, k-1\}$ and provide an elementary proof in the spirit of Poincaré's method of arbitrary functions. The conditions imposed are close to being necessary and sufficient and, if convergence is all that is desired, this theorem will cover all the standard cases. For instance, it will already explain the distributional convergence of variables reduced modulo $k$ in the examples of the previous section.

Our basic theorem provides a bound on the total variation distance between the distribution of $X \bmod k$ and the uniform distribution on $\{0, 1, \ldots, k-1\}$ for a given distribution for $X$. While the total variation bound suffices to deduce conditions for the convergence of moduli residues to the uniform, it is a rather blunt instrument and leaves open the question of how such distributions may be systematically constructed. Nor does the total variation bound provide sufficiently precise information about finer questions such as the rate of convergence or the origins of oscillatory phenomena. For these we have to dig deeper.

The first of these deeper questions may be partially answered by a variation on the theme of Weyl's equidistribution theorem by a consideration of convolutional structures arising out of random walks on the additive group $\mathbb{Z}_k$. We shall see that such processes characterize a large family of distributions for which the moduli residues converge to the uniform.

We follow this limit theorem by a characterization of a rather different flavor by connecting the distribution of $X \bmod k$ to absolutely continuous distributions via the Euler-Maclaurin summation formula. When a tractable smooth interpolation of a discrete distribution is available, a finer analysis now allows one to not only inherit distributional convergence for the modulo residues from a class of absolutely continuous distributions, but, by a consideration of the residual terms in the Euler-Maclaurin sum, explicate the origins of the oscillation that was evidenced in the illustrative examples of Section 2.

## 4.1 First General Theorem and a Bound on the Total Variation Distance

Our basic result is a variation on the theme of H. Poincaré's classical analysis of roulette in 1896 [see Poincaré (1912, pp. 122–130)]. The body of work emerging from Poincaré's analysis, memorably called "the method of arbitrary functions" by Hopf, builds upon the fundamental result that if $X$ is absolutely continuous and $t$ a real number then the fractional part of $tX$ converges in distribution to the uniform distribution on the unit interval as $t \to \infty$. [For a rich tapestry of applications, see Engel (1992).] The following theorem is a discrete analogue. It asserts, roughly speaking, that if the distribution of a non-negative, integer-valued random variable $X$ is "honest" in the sense that it is suitably well spread out then $X \bmod k$ is approximately uniformly distributed for any $k$. The proof we give here exploits partitioning in the spirit of Poincaré's approach [see Venkatesh (2013, pp. 288–290), for instance].

A little notation streamlines the presentation. For any $k \geq 2$, let $U^{(k)}$ denote a random variable distributed uniformly on $\{0, 1, \ldots, k-1\}$, and, for any non-negative integer-valued random variable $X$, write $X^{(k)} = X \bmod k$. If $U$ and $V$ are random variables, write

$$d_{\text{TV}}\big(\mathcal{L}(U), \mathcal{L}(V)\big) = \sup_A \big|P(U \in A) - P(V \in A)\big|$$

for the total variation distance between the laws of $U$ and $V$. The supremum is over all Borel sets; $\mathcal{L}(\cdot)$ stands for distribution or law of the random variable in the argument.

We recall that a mode of an arithmetic random variable is any value at which its mass function achieves a (local) maximum. We identify any *maximal* collection of consecutive integers all achieving the same (locally) maximum value as an equivalence class of

modes and select as representative any member of the class. By definition, any two equivalence classes of modes must either coincide or be disjoint. *Distinct* modes correspond to representatives in different equivalence classes.

**Theorem 4.1.** *Let $X$ be an integer-valued random variable taking values in $0, 1, 2, \ldots$. Suppose the mass function $f(i) = P(X = i)$ of $X$ has $m$ distinct modes and write $M = \max_i f(i)$. Then $\frac{1}{k}\, d_{TV}\big(\mathcal{L}(X^{(k)}), \mathcal{L}(U^{(k)})\big) \le 2mM$ for every $k \ge 2$.*

Example 2.5 illustrates what can go wrong if the number of modes is not finite. On the other hand, if the number of modes is finite then with an explicit bound in hand it is easy to write down a general convergence theorem.

**Corollary 4.1** (CONVERGENCE CRITERION)**.** *Let $\{X_n\}$ be a sequence of non-negative integer-valued random variables. Suppose the mass function $f_n(i) = P(X_n = i)$ has $m_n$ distinct modes and attains a maximum value $M_n = \max_i f_n(i)$. If $m_n M_n \to 0$ then $d_{TV}\big(\mathcal{L}(X_n^{(k)}), \mathcal{L}(U^{(k)})\big) \to 0$ for each $k \ge 2$.*

The usual setting where the convergence theorem comes into play is when the sequence $\{m_n\}$ is bounded, the unimodal case being the most important in practice. In this case the convergence criterion takes a very simple form: *if $\{m_n\}$ is bounded and $M_n \to 0$ then $d_{TV}\big(\mathcal{L}(X_n^{(k)}), \mathcal{L}(U^{(k)})\big) \to 0$ for each $k \ge 2$.* Our convergence theorem yields a quick and direct verification of the cases of convergence in the examples of the previous section without the necessity of a laborious case-by-case analysis (though finer behavior will have to await a more careful analysis).

**Example 4.1.** *(The binomial, revisited.)* The Binomial$(n, p)$ distribution $b_n(k; p) = \binom{n}{k} p^k (1-p)^{n-k}$ is unimodal and attains its maximum value $M_n$ at the point $\lceil np \rceil$. An easy application of Stirling's formula shows that

$$M_n = b_n\big(\lceil np \rceil; p\big) \sim \frac{1}{\sqrt{2\pi p(1-p)n}} \to 0 \qquad (n \to \infty)$$

for every fixed $p$ in the open unit interval $0 < p < 1$. Thus, if $X_n \sim$ Binomial$(n, p)$ and $0 < p < 1$, then $d_{\text{TV}}\big(\mathcal{L}(X_n^{(k)}), \mathcal{L}(U^{(k)})\big) \to 0$ as $n \to \infty$. Notice that the conclusion that $M_n \to 0$ is also obtainable from the local limit theorem because of the convolution structure of the binomial distribution. We will return to a more careful study of the binomial in Section 3.4.

**Example 4.2.** *(The Poisson distribution, revisited.)* The Poisson$(n)$ distribution $p(k; n) = e^{-n} k^n / k!$ $(n \ge 0)$ is also unimodal and attains its maximum value $M_n$ at the point $k = n$. It follows that

$$M_n = p(n; n) \sim \frac{1}{\sqrt{2\pi n}} \to 0 \qquad (n \to \infty),$$

and so, if $X_n \sim \text{Poisson}(n)$, then $d_{\text{TV}}\big(\mathcal{L}(X_n^{(k)}), \mathcal{L}(U^{(k)})\big) \to 0$. Once again, we may also conclude that $M_n \to 0$ from thelocal limit theorem as the Poisson also has a convolution structure. We will study the Poisson case in detail in Section 3.

**Example 4.3.** *(The zeta distribution, revisited).* The zeta distribution

$$f(n; z) = \frac{1}{\zeta(z) n^z} \qquad (n \geq 1)$$

of Example 2.3 is unimodal and achieves its maximum value of $1/\zeta(z)$ at $n = 1$. As $\zeta(z) \to \infty$ as $z \to 1$, it follows that, if $X_z \sim f(\cdot; z)$, then $X_z^{(k)}$ converges in distribution to the uniform as $z \downarrow 1$.

**Example 4.4.** *(The geometric distribution, revisited.)* Suppose $X_\lambda$ is concentrated on the non-negative integers with the geometric distribution with parameter $1 - e^{-1/\lambda}$. Its mass function $w(j; 1-e^{-1/\lambda}) = e^{-j/\lambda}(1-e^{-1/\lambda})$ is unimodal and achieves its maximum value of $w(0; 1 - e^{-1/\lambda}) = 1 - e^{-1/\lambda}$ at the origin. It follows that $d_{\text{TV}}\big(\mathcal{L}(X_\lambda^{(k)}), \mathcal{L}(U^{(k)})\big) \to 0$ as $\lambda \to \infty$.

**Example 4.5.** *(The negative binomial distribution.)* Fix any positive integer $n$ and $0 < p < 1$. Write $q = 1 - p$. The negative binomial distribution

$$w_n(j; p) = \binom{-n}{j} (-q)^j p^n \qquad (j \geq 0)$$

is unimodal and achieves its maximum value $M_{n,p}$ at the point $j = \lceil (n-1)q/p \rceil$. An easy application of Stirling's formula (or the local limit theorem) shows that

$$M_{n,p} = \max_j w_n(j; p) = w_n\big(\lceil (n-1)q/p \rceil; p\big) \sim \frac{p}{\sqrt{2\pi q n}} \qquad (n \to \infty).$$

Thus, if $X_{n,p} \sim w_n(\cdot; p)$, then $d_{\text{TV}}\big(\mathcal{L}(X_{n,p}^{(k)}), \mathcal{L}(U^{(k)})\big) \to 0$ as $n \to \infty$.

Another asymptotic convergence is obtained by allowing $p \downarrow 0$. The asymptotic estimates are unchanged and so $d_{\text{TV}}\big(\mathcal{L}(X_{n,p}^{(k)}), \mathcal{L}(U^{(k)})\big) \to 0$ also when $p \to 0$.

The simplest examples of multimode distributions arise out of mixtures of unimodal (or other multimodal) distributions. The Poisson case provides a typical illustrative example.

**Example 4.6.** *Poisson mixtures.* In the notation of Example 4.2, write

$$p(i; \lambda) = e^{-\lambda} \frac{\lambda^i}{i!} \qquad (i \geq 0)$$

for the Poisson distribution with mean $\lambda$. Then the mixture distribution

$$f(i) = \sum_{j=1}^{m} \alpha_j p(i; \lambda_j) \qquad (i \geq 0)$$

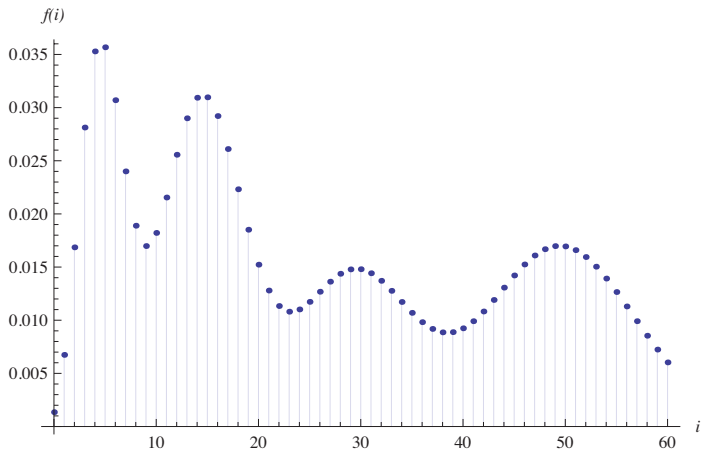


***Figure 9:*** *The Poisson mixture distribution* $f(i) = 0.2p(i;5) + 0.3p(i;15) + 0.2p(i;30) + 0.3p(i;50)$ *has four modes.*

(where the $\alpha_j$ are positive and sum to 1) has up to $m$ modes as illusrated in Figure 9. If $X$ is a random variable distributed according to $f(\cdot)$ then, by Theorem 4.1,

$$d_{\text{TV}}\big(\mathcal{L}(X^{(k)}), \mathcal{L}(U^{(k)})\big) \le 2kmM$$

where we may bound

$$M = \max_i f(i) \le \sum_{j=1}^{m} \alpha_j \max_i p(i;\lambda_j) \le \frac{C}{\sqrt{2\pi}} \sum_{j=1}^{m} \frac{\alpha_j}{\sqrt{\lambda_j}} \le \frac{C}{\sqrt{2\pi}} \cdot \frac{1}{\min_j \sqrt{\lambda_j}}.$$

The quantity $C$ on the right represents an absolute positive constant which we may take to be $e^{1/13}$. The explicit bound on the right yields a simple convergence criterion in terms of the Poisson parameters. Suppose $\{m_n, n \ge 1\}$ is a positive integer sequence and $\{\lambda_{jn}, 1 \le j \le m_n, n \ge 1\}$ is a doubly indexed sequence of positive real values such that $m_n \big/ \big(\min_{1\le j\le m_n} \sqrt{\lambda_{j,n}}\big) \to 0$. Let $\{X_n, n \ge 1\}$ be a sequence of random variables where, for each $n$, $X_n$ has a Poisson mixture distribution of the form $f_n(\cdot) = \sum_{j=1}^{m_n} \alpha_{j,n} p(\cdot;\lambda_j)$ (where $\alpha_{j,n} \ge 0$ and $\sum_{j\le m_n} \alpha_{j,n} = 1$). Then $d_{\text{TV}}\big(\mathcal{L}(X_n^{(k)}), \mathcal{L}(U^{(k)})\big) \to 0$ for each $k \ge 2$.

The convergence theorem extends effortlessly to settings where the modulus is random. Some notation first: to each distribution $\alpha = (\alpha_k, k \ge 1)$ with support in the positive integers associate the distribution $h(j;\alpha) = \sum_{k\ge j+1} \alpha_k/k$ with support in the non-negative integers $j \ge 0$.

**Theorem 4.2.** *Suppose* $K$ *is a lattice variable whose distribution* $\alpha = (\alpha_k, k \ge 1)$ *has support in the positive integers. If* $X$ *is as in Theorem 4.1 and* $K$ *is independent of* $X$ *then* $\sup_j \big|P\big(X^{(K)} = j\big) - h(j;\alpha)\big| \le 4mM$.

**Corollary 4.2.** *If the sequence $\{X_n\}$ is as in Corollary 4.1 and $K$ is independent of $\{X_n\}$ then $\sup_j \left|P\big(X_n^{(K)} = j\big) - h(j;\alpha)\right| \to 0$ as $n \to \infty$ provided $m_n M_n \to 0$.*

## 4.2 A Variation on Weyl's Equidistribution Theorem and Automatic Constructions

The convergence theorem of the previous section identifies structural features (in terms of modes) of arithmetic distributions for which there is distributional convergence for moduli residues. If a criticism were to be levied it could be that the discovery of distributions satisfying the convergence theorem has been carried out on a trial and error basis and that the examples that we have encountered so far are mostly the standard lattice distributions. It would be useful to bolster our understanding by *systematic* constructions. Fortunately, a large (but, as we shall see, not exhaustive) family of distributions is available to us through the classical theory. First, we need some notation.

As a preliminary, we recall the cyclic convolution operator on the additive group $\mathbb{Z}_k$ with group operation addition modulo $k$. If $F$ and $G$ are distributions with support in $\mathbb{Z}_k$, the cyclic convolution of $F$ and $G$, denoted $F \star G$, is the distribution with mass function

$$F \star G\{t\} = \sum_{s \in \mathbb{Z}_k} G\{t-s\}F\{s\} \qquad (t \in \mathbb{Z}_k)$$

where sums and differences of elements in the additive group $\mathbb{Z}_k$ are to be interpreted modulo $k$. The operation is commutative and associative and so there is no harm in writing $F_1 \star F_2 \star \cdots \star F_n$ for a repeated convolution.

Suppose $X$ is a random variable whose distribution $F$ has support contained in the set of non-negative integers. Write $X^{(k)} = X \bmod k$, let $F^{(k)}$ be the distribution of $X^{(k)}$, and write $V^{(k)}$ for that subset of $\{0, 1, \ldots, k-1\}$ on which $F^{(k)}$ is concentrated, i.e., $V^{(k)}$ is the support of $F^{(k)}$. In the algebraic viewpoint we are working on the additive group $bbZ_k$. Write $V_*^{(k)}$ for the subgroup of $\mathbb{Z}_k$ generated by $V^{(k)}$ and let $F_*^{(k)}$ denote the atomic distribution concentrated on $V_*^{(k)}$ which places equal mass on each of the elements of $V_*^{(k)}$.

*We shall henceforth restrict attention to distributions $F$ for which $V^{(k)}$ contains the point* $0$ as otherwise fairly trivial examples can be constructed that do not converge.

Let $\{X_n, n \geq 1\}$ be a sequence of random variables obtained by independent sampling from $F$. Form the partial sums $S_n = X_1 + \cdots + X_n$. Each $S_n$ has distribution $F_{\star n} = F \star \cdots \star F$ obtained as the $n$-fold convolution of $F$ with itself. Write $S_n^{(k)} = S_n \bmod k$ and let $F_{\star n}^{(k)}$ denote its distribution. Then $F_{\star n}^{(k)}$ is concentrated on some subset $V_n^{(k)}$ of the set $\{0, 1, \ldots, k-1\}$. As $V_1^{(k)} = V^{(k)}$ contains the origin, it is clear that the sequence $\big\{V_n^{(k)}, n \geq 1\big\}$ increases to a limit set $\bigcup_n V_n^{(k)}$ which we may, in fact, identify with the

subgroup $V_*^{(k)}$ generated by $V^{(k)}$. As we are working on the finite group $\mathbb{Z}_k$, it is clear that $V_n^{(k)}$ must actually coincide with $V_*^{(k)}$ for all sufficiently large $n$.

We are now ready for the following variation on a classical result.

**Theorem 4.3.** *Suppose $F^{(k)}$ is concentrated on $V^{(k)}$ and suppose also that $0 \in V^{(k)}$. Then $\{F_{\star n}^{(k)}, n \geq 1\}$ converges weakly to the uniform atomic distribution $F_*^{(k)}$ which places equal mass on each of the elements of the subgroup $V_*^{(k)}$ generated by $V^{(k)}$.*

If we identify the half-closed unit interval $[0, 1)$ with the circle $\mathbb{T}$ of unit length then, by wrapping the real line around $\mathbb{T}$ (in one or the other orientation) we may identify points $x$, $x \pm 1$, $x \pm 2$, ... and so on. In this viewpoint the sequence of (normalized) partial sums, $\{\frac{1}{k} S_n^{(k)}, n \geq 1\}$ represents a random walk on the vertices $0, 1/k, \ldots, (k-1)/k$ of a regular polygon embedded in $\mathbb{T}$. Theorem 4.3 is hence a statement about the weak convergence of a particular sequence of distributions concentrated on a regularly spaced set of points on the circle $\mathbb{T}$.

The conditions are reminiscent of Weyl's equidistribution theorem. In that setting we begin with an atomic distribution $F_1 = F$ placing all its mass at a single irrational point $x$ in $\mathbb{T}$; we now construct a sequence of distributions$\{F_n, n \geq 1\}$ with $F_n$ placing equal mass on the points $x$, $2x$, ..., $nx$ in $\mathbb{T}$. The analysis now proceeds in exactly the same fashion with the conclusion now that $\{F_n, n \geq 1\}$ converges weakly to the uniform distribution on $\mathbb{T}$ or, in the language of analysis, the sequence $\{nx, n \geq 1\}$ is equidistributed on $\mathbb{T}$.

**Corollary 4.3.** *For a given $k \geq 2$, if $F^{(k)}$ places positive mass on both $0$ and $1$, or if $F^{(k)}$ places positive mass on $0$, $s$, and $t$ where $s$ and $t$ are relatively prime, then $d_{TV}\big(\mathcal{L}(S_n^{(k)}), \mathcal{L}(U^{(k)})\big) \to 0$ as $n \to \infty$.*

**Corollary 4.4.** *If $F$ places positive mass on $0$ and $1$ then $\lim_n d_{TV}\big(\mathcal{L}(S_n^{(k)}), \mathcal{L}(U^{(k)})\big) = 0$ for every $k \geq 2$.*

We have stated Theorem 4.3 for lattice distributions on the line to keep the notation and presentation unencumbered. But a little introspection shows that the proof carries through essentially *in toto* in two and more dimensions. With proper notation, we can present the multidimensional result in the same form as in the one dimensional case. Suppose $d \geq 1$ is a positive integer and $k = (k_1, \ldots, k_d)$ a vector of moduli. We work in $(\mathbb{Z}^+)^d$ and identify the additive group $\mathbb{Z}_k = \mathbb{Z}_{k_1} \times \cdots \times \mathbb{Z}_{k_d} = \{0, 1, \ldots, k_1 - 1\} \times \cdots \times \{0, 1, \ldots, k_d - 1\}$ to be the lattice points in a $d$-dimensional rectangle with vector modulo $k$ operations interpreted component-wise. We suppose now that $X = (X_1, \ldots, X_d)$ is a lattice vector whose distribution $F$ has support in $(\mathbb{Z}^+)^d$. Reusing notation, let $X^{(k)} = X \bmod k$ denote the vector residue modulo $k$, and write $F^{(k)}$ for its distribution and $V^{(k)}$ for the subset of $\mathbb{Z}_k$ on which $F^{(k)}$ is concentrated. As before, we write $V_*^{(k)}$ for the

subgroup of $\mathbb{Z}_k$ generated by $V^{(k)}$ and $F_*^{(k)}$ for the vector atomic distribution placing equal mass at each of the points in $V_*^{(k)}$.

**Theorem 4.4.** *Theorem 4.3 holds in a $d$-dimensional setting where $k = (k_1, \ldots, k_d)$ represents any vector of moduli.*

**Corollary 4.5.** *For a given vector of moduli $k = (k_1, \ldots, k_d)$, suppose $F^{(k)}$ places positive mass at the origin as well as on the $d$ lattice points $(i_1, \ldots, i_d)$ neighboring the origin where precisely one of the indices $i_j$ takes value $1$ with all the others being $0$. Let $S_n^{(k)} = S_n \bmod k$ represent the modulo residues of the corresponding vector random walk and let $U^{(k)}$ represent a random vector with the uniform distribution on $\mathbb{Z}_k$. Then $d_{TV}\big(\mathcal{L}(S_n^{(k)}), \mathcal{L}(U^{(k)})\big) \to 0$ as $n \to \infty$.*

The adaptation of the proofs in one dimension requires nothing more than to strike out the references to a scalar modulus $k$ and replace them by a vector modulus $k = (k_1, \ldots, k_d)$, and replace references to $\operatorname{card} \mathbb{Z}_k = k$ in the scalar case by $\operatorname{card} \mathbb{Z}_k = k_1 \cdots k_d$ in the vector case.

Theorems 4.3 and 4.4 show how distributional convergence of modulo residues can arise from a wide range of basic distributions via convolutions. Many of the standard distributions fall under its sway: the Binomial$(n, p)$ distribution arises out of repeated convolutions of the Bernoulli$(p)$ distribution; the Poisson distribution is closed under convolutions and hence the Poisson$(n)$ distribution is obtained by repeated convolutions of the Poisson$(1)$ distribution; the Negative Binomial$(n, p)$ distribution arises out of repeated convolutions of the Geometric$(p)$ distribution. As the Bernoulli$(p)$, Poisson$(1)$, and Geometric$(p)$ distributions all put mass at both $0$ and $1$, it follows as a special case that there is distributional convergence for the modulo residues for the binomial, Poisson, and negative binomial cases [see Examples 2.1, 2.2, 4.5]. However, if we take $F$ to be the distribution of twice a Poisson random variable, then $F$ does not put any mass at $1$ and the modulo residues do not converge to the uniform. We believe that Theorem 4.3 explains a critical feature of the problem: basically long convolutions made out of a distribution putting mass at both $0$ and $1$ will exhibit the convergence to uniformity property.

While the theorem provides a concrete construction of a large family of settings for which there is distributional convergence for modulo residues, it is not exhaustive. The zeta and geometric distributions provide examples of distributions for which modulo residues converge weakly but *these distributions do not arise out of convolutions* [see Examples 2.3, 2.4]. The convergence of moduli residues with the parameter $p$ for the negative binomial distribution going to $0$ [Example 4.5] illustrates another setting where the convolutional character of the distribution is incidental.

Theorem 4.3 provides a useful *constructive* counterpart to the *structural* characterization of Theorem 4.1. Both, however, leave finer questions on the rate of convergence

open. We turn to this next.

## 4.3 Connection to the Euler-Maclaurin Formula

If the sequence of mass functions, $\{f_n, n \geq 1\}$, possesses a sufficiently smooth, tractable interpolation then Euler-Maclaurin summation provides an elegant approach to getting at convergence rates.

Suppose that, for each $n$, the mass function $f_n(i)$ with support in the non-negative integers $\{0, 1, 2, \dots\}$ has been extended via a smooth interpolation to a function $f_n(x)$ on the half-line $[0, \infty)$. By "smooth" we mean that the function $f_n$ extended in this fashion possesses sufficiently many continuous derivatives on $(0, \infty)$. Now fix integers $k \geq 2$, $0 \leq j \leq k-1$, $q > 0$, and $m > 0$. Let also, as usual, $B_s$, $s \geq 2$, denote the *Bernoulli numbers*: these are non-zero only for even values $s$, $B_2 = 1/6$, $B_4 = -1/30$, etc. If $f_n$ is $2m$ times continuously differentiable on $(0, \infty)$, then, by a change of variable of integration in the Euler-Maclaurin formula [see, e.g., Olver (1997, pp. 279–286)],

$$\begin{aligned}\sum_{l=0}^{q} f_n(kl+j) &= \int_0^q f_n(kx+j)\,dx + \frac{1}{2}\big[f_n(j) + f_n(kq+j)\big] \\ &\qquad + \sum_{s=1}^{m-1} \frac{k^{2s-1}B_{2s}}{(2s)!}\big[f_n^{(2s-1)}(kq+j) - f_n^{(2s-1)}(j)\big] + R_q(m) \\ &= \frac{1}{k}\int_0^{kq+j} f_n(x)\,dx - \frac{1}{k}\int_0^{j} f_n(x)\,dx + \frac{1}{2}\big[f_n(j) + f_n(kq+j)\big] \\ &\qquad + \sum_{s=1}^{m-1} \frac{k^{2s-1}B_{2s}}{(2s)!}\big[f_n^{(2s-1)}(kq+j) - f_n^{(2s-1)}(j)\big] + R_q(m).\end{aligned}$$

where, writing $\mathcal{V}_I$ for total variation over an interval $I$, the remainder term $R_q(m)$ may be bounded absolutely by $|R_q(m)| \leq C(k,m)\mathcal{V}_{[j,kq+j]}(f_n^{(2m-1)})$ for an absolute positive constant $C(k,m)$ determined solely by $k$ and $m$. Letting $q \to \infty$ we obtain

$$\begin{aligned}P(X_n \bmod k = j) = \sum_{l=0}^{\infty} f_n(kl+j) &= \frac{1}{k} - \frac{1}{k}\int_0^j f_n(x)\,dx + \frac{1}{2}f_n(j) \\ &\qquad - \sum_{s=1}^{m-1} \frac{k^{2s-1}B_{2s}}{(2s)!} f_n^{(2s-1)}(j) + R_m, \quad (19)\end{aligned}$$

where the remainder may be bounded absolutely by $|R_m| \leq C(k,m)\,\mathcal{V}_{[0,\infty)}(f_n^{(2m-1)})$. The expansion is valid asymptotically provided $f_n$ and its first $2m-3$ derivatives vanish as $x \to \infty$ and $f_n^{(2m-1)}$ is of bounded variation on the half-line $[0, \infty)$. When such a tractable smooth interpolation exists, (19) provides an effective practical method for numerical approximation of $P(X_n \bmod k = j)$.

**Example 4.7.** *Interpolating the geometric distribution.* Consider the failure geometric distribution with mass function $P(X = n) = pq^n$ with $q = 1 - p$. Direct summation gives the exact expression

$$P(X \bmod k = j) = \frac{pq^j}{1-q^k} = \frac{1}{k} + \left(\frac{1}{2} - \frac{j+\frac{1}{2}}{k}\right)p + \frac{k^2 - 6jk + 6j^2 - 1}{12k}p^2 + O(p^3). \tag{20}$$

This is the same as the expression obtained in Example 2.4 with $p = 1 - e^{-1/\lambda}$ and $\lambda$ serving in the role of the asymptotic parameter.

The natural interpolation of the geometric distribution is, of course, $f_n(x) = pq^x$ with support in $[0, \infty)$. Applying (19) with $m = 2$ hence gives

$$\begin{aligned} P(X \bmod k = j) &= \frac{1}{k} - \frac{p}{k}\int_0^j q^x dx + \frac{pq^j}{2} - \frac{k}{12}pq^j \log q + R_2(p) \\ &= \frac{1}{k} - \frac{p}{k}\frac{q^j - 1}{\log q} + \frac{pq^j}{2} - \frac{k}{12}pq^j \log q + R_2(p) \\ &= \frac{1}{k} + \left(\frac{1}{2} - \frac{j}{k}\right)p + \frac{k^2 - 6jk + 6j^2}{12k}p^2 + O(p^3) + R_2(p). \end{aligned} \tag{21}$$

A comparison of (20,21) shows that the coefficients of $p$ and $p^2$ are slightly different. This is because the remainder term $R_2(p)$ has not been exactly written out; when it is, the two asymptotic expansions will coincide.

It is useful to compare these findings with those of Examples 2.1 and 2.2. Unlike those cases where convergence to the limiting value is exponentially fast, in the geometric case, convergence occurs only at a polynomial rate as was seen in Example 2.4 and as (20,21) again aver.

**Example 4.8.** *Interpolating the zeta distribution.* The zeta distribution of Example 2.3 admits the smooth interpolation

$$f(x) = f(x; z) = \frac{x^{-z}}{\zeta(z)} \qquad (x \geq 1)$$

with parameter $z > 1$. Restricted to the positive integers this recovers the zeta distribution (3). Then

$$f^{(2s-1)}(j) = -z(z+1)\dots(z+2s-1)\frac{j^{-z-2s+1}}{\zeta(z)} = -\big[(2s-1)! + O(z-1)\big] \cdot \frac{e^{-(z-1)\log(j)}}{j^{2s}\zeta(z)}$$

and it is easy to verify that $P(X \bmod k = j) = \frac{1}{k} + O(z-1)$ as $z \downarrow 1$ though interestingly, the convergence is very slow.

For a given lattice distribution, the utility of the approach depends upon our ability to find a natural, smooth interpolation as in these examples. In reverse, we may also start

with a suitably smooth density on the half-line $[0, \infty)$ and, by considering its restriction to the non-negative integers, inherit a lattice distribution for which the moduli residues converge in distribution. (A normalizing constant may appear in general but it is easily handled.) Such settings are relevant in applications in vector quantization.

Where available, the Euler-Maclaurin expansion (19) also gives us useful information about when convergence to the limiting value $1/k$ should be exponentially fast. Suppose the lattice-valued random variable $X_n$ has a convolution form, $X_n = Y_1 + Y_2 + \cdots + Y_n$, where the $Y_i$ are i.i.d. have a moment generating function in a neighborhood of zero. Then, by large deviation theory for partial sums [see, e.g., Varadhan (1984)], the term $\int_0^j f_n(x)dx \approx P(X_n \leq j)$ would be exponentially small. By large deviation theory for local limit theorems [c.f. Richter (1957), Chaganty and Sethuraman (1985)], the term $f_n(j)$ would also be exponentially small. Thus, if the oscillations of the sequence of functions $f_n$ die at exponential rates, then by (19), $P(X_n \bmod k = j) - \frac{1}{k}$ would converge to zero at a suitable exponential rate. In the absence of such a convolution structure it would be difficult to assert anything general about exponentially fast convergence.

## 4.4 Writing Out the Oscillatory Term

Expression (19) also explains the reason for the oscillation in $P(X_n \bmod k = j)$ that we evidenced in some of the illustrative examples in Section 2. The remainder term in (19) has the precise form

$$R_m = \int_0^\infty \frac{B_{2m} - B_{2m}(x - \lfloor x \rfloor)}{(2m)!} f_n^{(2m)}(x)\, dx,$$

where $B_{2m}$ is the $2m$th Bernoulli number and $B_{2m}(t)$ is the Bernoulli polynomial of degree $2m$. Using the Fourier expansion

$$B_{2m}(t) = \frac{2(-1)^{m+1}(2m)!}{(2\pi)^{2m}} \sum_{j=1}^{\infty} \frac{\cos 2\pi jt}{j^{2m}}, \tag{22}$$

together with the formula for even-ordered Bernoulli numbers

$$B_{2m} = \frac{2(-1)^{m+1}(2m)!}{(2\pi)^{2m}} \zeta(2m) = \frac{2(-1)^{2m+1}(2m)!}{(2\pi)^{2m}} \sum_{j=1}^{\infty} \frac{1}{j^{2m}}, \tag{23}$$

[see, e.g., Olver (1997, pp. 283–285), for (22,23)], the remainder may be written explicitly in the form

$$R_m = \frac{2(-1)^{m+1}}{(2\pi)^{2m}} \sum_{j=1}^{\infty} \frac{1}{j^{2m}} \int_0^\infty \big[1 - \cos\big(2\pi j(x - \lfloor x \rfloor)\big)\big] f_n^{(2m)}(x)\, dx, \tag{24}$$

provided $f_n^{(2m)}$ is absolutely integrable on $(0,\infty)$ so that the integration can be performed term by term. The expression (24) for the remainder gives an explanation for the presence of an oscillatory term in $P(X_n \bmod k = j)$.

By combining the arguments leading to (19) and (24), we obtain the following general theorem.

**Theorem 4.5.** *Let $X$ be an integer-valued random variable taking values $0, 1, 2, \cdots$ with mass function $P(X = j) = f(j)$, where $f\colon [0,\infty) \to [0,\infty)$ is a density function satisfying the following properties:*

*(a) $f$ is ultimately decreasing;*

*(b) There is an even integer $2M$ such that $f$ is $2M$ times continuously differentiable on $(0,\infty)$;*

*(c) $\lim_{x\to\infty} f(x) = \lim_{x\to\infty} f^{(2s-1)}(x) = 0$ for every $s = 1, 2, \cdots, M-1$;*

*(d) $f^{(2M)}$ is absolutely integrable on $(0,\infty)$.*

*Fix any integer $k \geq 2$ and $0 \leq j \leq k-1$. Then,*

$$P(X \bmod k = j) = \frac{1}{k} - \frac{1}{k}\int_0^j f(x)\,dx + \frac{1}{2}f(j) - \sum_{s=1}^{M-1} \frac{k^{2s-1}B_{2s}}{(2s)!} f^{(2s-1)}(j) + \frac{2(-1)^{M+1}}{(2\pi)^{2M}} \sum_{j=1}^{\infty} \frac{1}{j^{2M}} \int_0^\infty \big[1 - \cos\big(2\pi j(x - \lfloor x \rfloor)\big)\big] f^{(2M)}(x)\,dx. \quad (25)$$

We may now simply read out the conditions for distributional convergence of modulo residues.

**Corollary 4.6.** *Suppose that the non-negative integer-valued random variable $X_n$ has a mass function for which there exists a smooth, $2m$-times differentiable interpolation $f_n$ satisfying the following asymptotic properties:*

*(a) For any fixed $t$, $\int_0^t f_n(x)\,dx \to 0$ and also $f_n^{(2l-1)}(t) \to 0$ for $0 \leq l \leq m-1$.*

*(b) $\mathcal{V}_{[0,\infty)}\big(f_n^{(2m-1)}\big) \to 0$.*

*Then $d_{TV}\big(\mathcal{L}(X_n^{(k)}), \mathcal{L}(U^{(k)})\big) \to 0$ for each $k \geq 2$.*

*Theorem 4.5 gives a purely analytical explanation for why lattice random variables that have a flat and well-behaved mass function are approximately uniform when measured mod $k$ for any $k$; the approximate uniformity is transparent from the representation on the right in* (25) *as every term other than $\frac{1}{k}$ would be small.*

We close this subsection with the observation that it is possible to use the mulivariate version of the Euler-Maclaurin summation formula [see Bhattacharya and Rao (2010, p. 296)] to obtain a version of Theorem 4.5 in two or more dimensions much in the same way as Theorem 4.4 expanded the purview of Theorem 4.3. The cost is in a rapidly burgeoning notational complexity and we eschew the details.

## 4.5 The Moments

Let $X$ be a general non-negative integer-valued random variable and let $k \geq 2$ be a fixed integer. The formula in (25) for the mass function of $X \bmod k$ does not lead to any worthwhile general formula for the moments of $X \bmod k$. Although neat general formulas seem quite difficult, a Fourier approach, instead of a direct attack by using equation (25), does help in writing a general formula for the moments, and this formula leads to some simplification of the low order moments. Here is the general Fourier formula. We first need to lay out some notation.

We denote

$$\begin{aligned}
X^{(k)} &= X \bmod k; \\
p_j(k) &= P(X^{(k)} = j); \\
\mu_n(k) &= \mathbf{E}[X^{(k)}]^n; \\
\psi_X(t) &= \text{The characteristic function of } X; \\
\psi_{c,X}(t) &= \text{The Fourier cosine transform of } X; \\
\psi_{s,X}(t) &= \text{The Fourier sine transform of } X; \\
z_r &= \tfrac{2\pi r}{k}; \\
B_k &= k\text{th Bernoulli number}; \\
\alpha_{n,m}(t) &= \textstyle\sum_{j=1}^{m} j^n \cos(tj); \\
\beta_{n,m}(t) &= \textstyle\sum_{j=1}^{m} j^n \sin(tj).
\end{aligned}$$

In particular, we have $\psi_X(z_r) = \psi_{c,X}(z_r) + i\psi_{s,X}(z_r)$ while we may also identify

$$\sum_{j=1}^{k-1} j^n (e^{-iz_r})^j = \alpha_{n,k-1}(z_r) - i\beta_{n,k-1}(z_r).$$

Identifying $z_r = 2\pi r/k$ in the Fourier identity (13), we obtain

$$\begin{aligned}
\mu_n(k) &= \sum_{j=1}^{k-1} j^n p_j(k) = \sum_{j=1}^{k-1} j^n \left[\frac{1}{k} + \frac{1}{k}\sum_{r=1}^{k-1} (e^{-i\,z_r})^j \psi_X(z_r)\right] \\
&= \frac{1}{k}\left[\sum_{j=1}^{k-1} j^n + \sum_{j=1}^{k-1}\sum_{r=1}^{k-1} \psi_X(z_r) j^n (e^{-iz_r})^j\right] = \frac{1}{k}\left[\sum_{j=1}^{k-1} j^n + \sum_{r=1}^{k-1} \psi_X(z_r) \sum_{j=1}^{k-1} j^n (e^{-i\,z_r})^j\right]
\end{aligned} \tag{26}$$

In view of the identity

$$\sum_{j=1}^{p} j^n = \frac{1}{n+1}\sum_{j=0}^{n}(-1)^j \binom{n+1}{j} B_j\, p^{n+1-j},$$

we may combine (13,26), to get the following general result.

**Theorem 4.6.**

$$\mu_n(k) = \frac{1}{k}\Bigg[\frac{1}{n+1}\sum_{j=0}^{n}(-1)^j\binom{n+1}{j}B_j\,(k-1)^{n+1-j} + \sum_{r=1}^{k-1}\big\{\alpha_{n,k-1}(z_r)\psi_{c,X}(z_r) + \beta_{n,k-1}(z_r)\psi_{s,X}(z_r)\big\}\Bigg].$$

This is a general moment formula valid for arbitrary $n$ and $k$ and general $X$. No further simplification seems possible, except in specific cases. For example, if $n = 1$, the functions $\alpha_{n,k-1}$ and $\beta_{n,k-1}$ simplify and we obtain the following corollary.

**Corollary 4.7.** *For general $X$ and $k$, the first moment of $X \bmod k$ equals*

$$\mu_1(k) = \frac{1}{k}\Bigg[\frac{k(k-1)}{4} - \frac{1}{4}\sum_{r=1}^{k-1}\csc^2\Big(\frac{z_r}{2}\Big)\Big\{\big(1+(k-1)\cos(kz_r) - k\cos((k-1)z_r)\big)\psi_{c,X}(z_r) + \big((k-1)\sin(kz_r) - k\sin((k-1)z_r)\big)\psi_{s,X}(z_r)\Big\}\Bigg]. \quad (27)$$

Here is an example that uses Corollary 4.7.

**Example 4.9.** Suppose $X \sim$ Poisson$(\lambda)$. Consider the case $k = 3$; thus, $X \bmod k$ takes the values $0, 1, 2$. We would expect its mean to be c;ose to $1$ for large $\lambda$; how close? For example:

How large a $\lambda$ is needed to make $|1 - \mathbf{E}(X \bmod 3)| < \epsilon$?

In the Poisson case, the Fourier cosine and the Fourier sine transforms can be written in closed form:

$$\psi_{c,X}(t) = \tfrac{1}{2}\big(e^{\lambda(e^{it}-1)} + e^{\lambda(e^{-it}-1)}\big); \quad \psi_{s,X}(t) = e^{\lambda(\cos t - 1)}\,\sin(\lambda \sin t).$$

The values of $z_r$ are $\frac{2}{3}\pi, \frac{4}{3}\pi$, and $\cos(kz_r), \cos((k-1)z_r), \sin(kz_r), \sin((k-1)z_r)$, and $\csc(\frac{z_r}{2})$ are easily found in closed form. After some intermediate calculations, formula (27) leads to

$$\mathbf{E}(X \bmod 3) = 1 - e^{-3\lambda/2}\left[\cos\left(\frac{\lambda\sqrt{3}}{2}\right) + \frac{\sin\left(\frac{\lambda\sqrt{3}}{2}\right)}{\sqrt{3}}\right]. \quad (28)$$

So, $|1 - \mathbf{E}(X \bmod 3)| = O(e^{-3\lambda/2})$. Going further, the function $\cos(\frac{\lambda\sqrt{3}}{2}) + \frac{\sin(\frac{\lambda\sqrt{3}}{2})}{\sqrt{3}}$ is bounded in absolute value by 1. Hence, using (28), we see that $|1 - \mathbf{E}(X \bmod 3)| < \epsilon$ for $\lambda > \log \epsilon^{-2/3}$.

Theorem 4.6 also results in this following general approximation bound on the moments of $X \bmod k$.

**Theorem 4.7.** *Let $U^{(k)}$ denote a uniform random variable on the set $\{0, 1, \ldots, k-1\}$ and let $\nu_n(k) = \mathbf{E}\big[\big(U^{(k)}\big)^n\big]$. Then, for any $k$ and $n$,*

$$\sup_{n \geq 1} \left| \frac{\mu_n(k)}{\nu_n(k)} - 1 \right| \leq (k-1) \sup_{\frac{2\pi}{k} \leq t \leq \frac{2\pi(k-1)}{k}} |\psi_X(t)|.$$

The main appeal of Theorem 4.7 is that it will often provide an exponential convergence result for the moments.

# References


Bhattacharya, R. and Rao, R. (2010). Normal Approximation and Asymptotic Expansions, SIAM, Philadelphia.

Chaganty, N. and Sethuraman, J. (1985). Large deviation local limit theorems for arbitrary sequences of random variables, Ann. Prob., 13, 97-114.

Diaconis, P. (1992). Finite Fourier methods: Access to tools, Proc. Symp. Appl. Math, 44, 171-194, AMS, Providence.

Engel, E. (1992). A Road to Randomness in Physical Systems. Springer-Verlag, Lecture Notes in Statistics, 71.

Feller, W. (1971). An Introduction to Probability Theory with Applications, Volume 2, Wiley, New York.

Finch, S. (2003). Mathematical Constants, Cambridge University Press, Cambridge, UK.

Hardy, G. (1956). Divergent Series, Clarendon, Oxford, UK.

Hildebrand, M. (1990). Random transvections, Preprint, Dept. of Mathematics, Univ. of Michigan.

LeVeque, W. (1996). Fundamentals of Number Theory, Dover, New York.

Miller, S. and Takloo-Bighash, R. (2006). An Invitation to Modern Number Theory, Princeton Univ. Press, Princeton.

Musser, J. (2011). Higher derivatives of the Hurwitz Zeta function, PhD Thesis, Western Kentucky Univ.

Niven, I. and Zuckerman, H. S. (1980). The Theory of Numbers, Wiley, New York.

Olver, F. (1997). Asymptotics and Special Functions, AK Peters, Wellesley, MA.

Poincaré, H. (1912). Calcul des Probabilités. Gauthier-Villars, Paris.

Pólya, G. and Szego, G. (1998). Problems and Theorems in Analysis, II, Springer-Verlag, Heidelberg.
Richter, W. (1957). Local limit theorems for large deviations, Theor. Prob. Appls., 2, 206-219.
Ross, S. (1995). Stochastic Processes, Wiley, New York.
Shapiro, H. (1983). Introduction to the Theory of Numbers, Dover, New York.
Varadhan, S. (1984). Large Deviations and Applications, SIAM, Philadelphia.
Venkatesh, S. S. (2013). The Theory of Probability: Explorations and Applications, Cambridge University Press, Cambridge, UK.

# 5 Appendix

*Proof of Theorem 4.1*: Select any $k \geq 2$. We will prove a little more than advertised and show indeed that

$$\left|P\big(X^{(k)} = j\big) - \tfrac{1}{k}\right| \leq 4mM \qquad (0 \leq j \leq k-1). \tag{29}$$

The claimed result follows from the observation that

$$d_{\text{TV}}\big(\mathfrak{L}(X^{(k)}), \mathfrak{L}(U^{(k)})\big) = \frac{1}{2}\sum_{j=0}^{k-1}\left|P\big(X^{(k)} = j\big) - \frac{1}{k}\right|.$$

For $i \geq 0$, set $\Omega_i = \{ik, ik+1, \ldots, (i+1)k-1\}$. Fix any $0 \leq j \leq k-1$. Each $\Omega_i$ then contains a unique element $t_{ij}$ of the form $\alpha k + j$. Then $t_{ij} \equiv j \pmod{k}$, and the collection $E_j = \{t_{ij}, i \geq 0\}$ forms an equivalence class of elements all congruent to $j$ modulo $k$. We then see that

$$\begin{aligned} P\big(X^{(k)} = j\big) - \frac{1}{k} = P(X \in E_j) - \frac{1}{k} &= \sum_i f(t_{ij}) - \frac{1}{k}\sum_t f(t) \\ &= \sum_i \frac{1}{k}\sum_{t\in\Omega_i} f(t_{ij}) - \frac{1}{k}\sum_i\sum_{t\in\Omega_i} f(t) \end{aligned}$$

as, on the one hand, $\operatorname{card}\Omega_i = k$ for each $i$, and, on the other, the collection $\{\Omega_i, i \geq 0\}$ partitions $\mathbb{Z}^+$. By consolidating the sums we hence obtain

$$\left|P\big(X^{(k)} = j\big) - \frac{1}{k}\right| = \left|\sum_i \frac{1}{k}\sum_{t\in\Omega_i}\big(f(t_{ij}) - f(t)\big)\right| \leq \sum_i \frac{1}{k}\sum_{t\in\Omega_i}\big|f(t_{ij}) - f(t)\big|. \tag{30}$$

We now proceed by partitioning to exploit the modal nature of the distribution. As $X$ has $m$ distinct (equivalence classes of) modes we may identify an increasing sequence $-\infty = \eta_0 < \xi_1 < \eta_1 < \xi_2 < \eta_2 < \cdots < \eta_{m-1} < \xi_m < \eta_m = +\infty$ such that, for each $j$,

$f(t)$ is non-decreasing for $\eta_{j-1} \le t \le \xi_j$ and non-increasing for $\xi_j \le t \le \eta_j$. Write $I$ for the set of indices $i$ for which $\Omega_i$ contains one or more of the extremum points $\xi_1$, $\eta_1$, $\xi_2$, …, $\eta_{m-1}$, $\xi_m$. It is clear that $1 \le \operatorname{card}(I) \le 2m-1$.

Split the sum over $i$ on the right in (30) and write $\sum_i = \sum_{i\in I} + \sum_{i\notin I}$. We may accordingly bound

$$\big|P\big(X^{(k)} = j\big) - \tfrac{1}{k}\big| \le \sum\nolimits' + \sum\nolimits'' \tag{31}$$

where $\sum'$ accumulates the terms for $i \in I$ and $\sum''$ accumulates the remaining terms for $i \notin I$.

Simple bounds suffice for the contribution from the terms $i \in I$. As $|f(s)-f(t)| \le M$ for all integers $s$ and $t$, it follows that

$$\sum\nolimits' = \sum_{i\in I} \frac{1}{k} \sum_{t\in\Omega_i} \big|f(t_{ij}) - f(t)\big| \le M \operatorname{card}(I) < 2mM.$$

We will need to be a little more careful in bounding the contribution from the terms $i \notin I$. After removing the set $I$, the remaining integers in $\mathbb{Z}^+ \setminus I$ may be grouped into maximal, pairwise disjoint collections of successive indices, say, $\{I_\lambda, \lambda \in \Lambda\}$, where $\operatorname{card}\Lambda \le 2m$, and each $I_\lambda$ is a maximal unbroken sequence of successive integers such that $\bigcup_{i\in I_\lambda} \Omega_i$ is either contained in an open interval of the form $(\eta_{j-1}, \xi_j)$ or contained in an open interval of the form $(\xi_j, \eta_j)$ for some $j$. In the former case, the values $f(t)$ are non-decreasing as $t$ varies through $\bigcup_{i\in I_\lambda} \Omega_i$ and we say that $I_\lambda$ is "increasing"; in the latter case, the values $f(t)$ are non-increasing as $t$ varies through $\bigcup_{i\in I_\lambda} \Omega_i$ and we say that $I_\lambda$ is "decreasing". By partitioning the sum further, we see that

$$\sum\nolimits'' = \sum_{i\notin I} \frac{1}{k} \sum_{t\in\Omega_i} \big|f(t_{ij}) - f(t)\big| = \sum_{\lambda\in\Lambda} \bigg( \sum_{i\in I_\lambda} \frac{1}{k} \sum_{t\in\Omega_i} \big|f(t_{ij}) - f(t)\big| \bigg). \tag{32}$$

Suppose first that $I_\lambda$ is increasing. Then there exists $j$ such that $\bigcup_{i\in I_\lambda} \Omega_i \subset (\eta_{j-1}, \xi_j)$. If $i \in I_\lambda$ then, for all $s, t \in \Omega_i$, we have $|f(s)-f(t)| \le f\big((i+1)k\big) - f(ik)$ and accordingly,

$$\begin{aligned} \sum_{i\in I_\lambda} \frac{1}{k} \sum_{t\in\Omega_i} \big|f(t_{ij}) - f(t)\big| &\le \sum_{i\in I_\lambda} \frac{1}{k} \sum_{t\in\Omega_i} \big[f\big((i+1)k\big) - f(ik)\big] \\ &\overset{(*)}{=} \sum_{i\in I_\lambda} \big[f\big((i+1)k\big) - f(ik)\big] \le f(\xi_j) - f(\eta_{j-1}) \le M \end{aligned}$$

as the intervals $\{\Omega_i, i \in I_\lambda\}$ form an unbroken succession and so the sum in $(*)$ telescopes to its endpoints. Suppose next that $I_\lambda$ is decreasing. Then there exists $j$ such that $\bigcup_{i\in I_\lambda} \Omega_i \subset (\xi_j, \eta_j)$. If $i \in I_\lambda$ then, for all $s, t \in \Omega_i$, we have $|f(s) - f(t)| \le$

$f(ik) - f\big((i+1)k\big)$ and accordingly, by a similar telescoping argument,

$$\sum_{i\in I_\lambda} \frac{1}{k} \sum_{t\in\Omega_i} \big|f(t_{ij}) - f(t)\big| \le \sum_{i\in I_\lambda} \frac{1}{k} \sum_{t\in\Omega_i} \big[f(ik) - f\big((i+1)k\big)\big]$$
$$= \sum_{i\in I_\lambda} \big[f(ik) - f\big((i+1)k\big)\big] \le f(\xi_j) - f(\eta_j) \le M.$$

Thus, in all cases, the term enclosed in round brackets in (32) is bounded above by $M$ and so $\sum'' \le M \operatorname{card}\Lambda \le 2mM$. Pooling bounds on the right in (31) gives (29). □

*Proof of Theorem 4.2*: The proof follows by conditioning on $K$ and leveraging (29). □

*Proof of Theorem 4.3*: The proof follows a standard line and we simply sketch the main points. We begin by initially supposing that the support $V^{(k)}$ of $F^{(k)}$ already coincides with $V_*^{(k)}$. Then all subsequent convolutional iterates have the same support: $V_n^{(k)} = V_*^{(k)}$.

By Helly's selection principle, there exists a subsequence $\big\{F^{(k)}_{\star n_i}, i \ge 1\big\}$ converging weakly to a distribution, say, $G$ concentrated on $V_*^{(k)}$. Starting anew with $G$, we construct a new sequence of distributions $\{G_n, n \ge 0\}$ by repeated convolution with $F^{(k)}$ by first setting $G_0 = G$ and, for $n \ge 1$, recursively forming $G_n = F^{(k)} \star G_{n-1}$. As $G$ is concentrated on $V_*^{(k)}$ and $F^{(k)}$ places non-zero mass on the point $0$ it follows that each of the convolutional distributions $G_n$ is also concentrated on $V_*^{(k)}$.

Now $F^{(k)}_{\star n_i+1} = F^{(k)} \star F^{(k)}_{\star n_i}$ and $G_1 = F^{(k)} \star G_0$. As $F^{(k)}_{\star n_i} \to G_0$ weakly, it follows that $\big\{F^{(k)}_{\star n_i+1}, i \ge 1\big\}$ converges weakly to $G_1$. By induction, it follows that $\big\{F^{(k)}_{\star n_i+j}, i \ge 1\big\}$ converges weakly to $G_j$ for each $j \ge 0$.

For $n \ge 0$, let $M_n = \max_{t\in\mathbb{Z}_k} F^{(k)}_{\star n}\{t\}$ denote the largest atomic mass of the distribution $F^{(k)}_{\star n}$. Clearly, $0 < M_n \le 1$. Then

$$F^{(k)}_{\star n+1}\{t\} = \sum_{s\in\mathbb{Z}_k} F^{(k)}_{\star n}\{t-s\}F^{(k)}\{s\} \le M_n \sum_{s\in\mathbb{Z}_k} F^{(k)}\{s\} = M_n F^{(k)}(\mathbb{Z}_k) = M_n$$

for every $t \in \mathbb{Z}_k$, and so $M_{n+1} \le M_n$. As $\{M_n, n \ge 0\}$ is a non-increasing sequence bounded below, it has a limit, say, $M$. As every subsequence of a convergent sequence has the same limit, it follows that $M_{n_i+j} \to M$ for each $j \ge 0$. We conclude that $\max_{t\in\mathbb{Z}_k} G_j\{t\} = M$ for each $j \ge 0$.

We now argue that, in fact, $G_0\{t\} = M$ for each $t \in V_*^{(k)}$. Suppose, on the contrary, that $G\{t\} < M$ for some $t \in V_*^{(k)}$. We then see that

$$G_1\{\tau\} = G_0\{t\}F^{(k)}\{\tau - t\} + \sum_{s\in\mathbb{Z}_k\setminus\{\tau-t\}} G_0\{\tau - s\}F^{(k)}\{s\}$$
$$< MF^{(k)}\{\tau - t\} + \sum_{s\in\mathbb{Z}_k\setminus\{\tau-t\}} G_0\{\tau - s\}F^{(k)}\{s\}.$$

The inequality is strict as, by assumption, $F^{(k)}$ has support in all of $V_*^{(k)}$ and so it is the case that $F^{(k)}\{\tau - t\} > 0$. Bounding the remaining terms in the sum on the right, we obtain

$$G_1\{\tau\} < MF^{(k)}\{\tau - t\} + MF^{(k)}\big(\mathbb{Z}_k \setminus \{\tau - t\}\big) = M$$

for every $\tau \in \mathbb{Z}_k$. It must hence follow under the reductio ad absurdum assumption that $M > \max_\tau G_1\{\tau\} = M$ and we have a contradiction. We conclude that $G_0\{t\} = M$ for every $t \in V_*^{(k)}$ and, by induction, that, for each $j \geq 0$, we must have $G_j\{t\} = M$ for all $t \in V_*^{(k)}$. We have hence shown that there is a constant $M$ such that $F_{\star n}^{(k)}\{t\} \to M$ for every $t \in V_*^{(k)}$. But then this constant must be equal to $1/\operatorname{card} V_*^{(k)}$ and so $F_{\star n}^{(k)}$ converges weakly to $F_*^{(k)}$, the uniform distribution concentrated on $V_*^{(k)}$.

To complete the proof we now remove our initial assumption that $V^{(k)}$ is a subgroup of $\mathbb{Z}_k$ to begin with. Suppose now that the generator $V^{(k)}$ is a proper subset of the subgroup $V_*^{(k)}$ generated by it. As $V^{(k)}$ includes the point $0$, we recall that the sequence of supports $\big\{V_n^{(k)}, n \geq 0\big\}$ is increasing with limit set $\bigcup_n V_n^{(k)} = V_*^{(k)}$. As $\operatorname{card} V_*^{(k)} \leq k$, there must in fact exist an integer $N$ such that the support $V_N^{(k)}$ of the distribution $F_{\star N}^{(k)}$ coincides with the subgroup $V_*^{(k)}$. It follows that $V_n^{(k)} = V_*^{(k)}$ for all $n \geq N$ and, in particular, the subsequence $\{F_{\star nN}^{(k)}, n \geq 1\}$ consists of distributions all concentrated on the subgroup $V_*^{(k)}$. As convolution is associative, $F_{\star 2N}^{(k)} = F_{\star N}^{(k)} \star F_{\star N}^{(k)}$, $F_{\star 3N}^{(k)} = F_{\star N}^{(k)} \star F_{\star 2N}^{(k)}$, and so on, so that the entire subsequence can be built up by repeated convolutions of the basic distribution $F_{\star N}^{(k)}$. By the just concluded proof of the simplified setting, the sequence $\big\{F_{\star nN}^{(k)}, n \geq 1\big\}$ converges weakly to $F_*^{(k)}$.

Now consider the subsequence $\big\{F_{\star nN+1}^{(k)}, n \geq 1\big\}$ obtained by convolving each member of the sequence $\big\{F_{\star nN}^{(k)}, n \geq 1\big\}$ with $F^{(k)}$. On the one hand, $F_{\star nN+1}^{(k)} = F^{(k)} \star F_{\star nN}^{(k)}$, while, on the other,

$$F^{(k)} \star F_*^{(k)}\{t\} = \sum_{s \in \mathbb{Z}_k} F_*^{(k)}\{t - s\}F^{(k)}\{s\} = \frac{1}{\operatorname{card} V_*^{(k)}} \sum_{s \in V^{(k)}} F^{(k)}\{s\} = \frac{1}{\operatorname{card} V_*^{(k)}}$$

for every $t \in V_*^{(k)}$ so that $F^{(k)} \star F_*^{(k)} = F_*^{(k)}$. It follows that $\big\{F_{\star nN+1}^{(k)}, n \geq 1\big\}$ converges weakly to $F_*^{(k)}$ and, by induction, $\big\{F_{\star nN+j}^{(k)}, n \geq 1\big\}$ converges weakly to $F_*^{(k)}$ for each $j \geq 0$. We conclude that $\big\{F_{\star n}^{(k)}, n \geq 1\big\}$ converges weakly to $F_*^{(k)}$. □

Corollaries 4.3 and 4.4 follow quickly by Bézout's identity [see, for example, Niven and Zukerman (1980, Theorem 1.3, p. 7)] as, in both cases, $V^{(k)}$ is a generator of the additive group $\mathbb{Z}_k$.