A. PROOFS

Proof of Proposition 1. For convenience, let 0 denote the last digit of s_1. If the last digit of X is distributed uniformly, the difference in density with which different numerals occur must on average be zero. Formally,

$$\sum_{a=\frac{s_1}{b}}^{\frac{s_2-b}{b}} (f(ab+d_1) - f(ab+d_2)) = 0 \quad \forall d_1, d_2 \in \{0, \ldots, b-1\}. \tag{A1}$$

If g can be approximated linearly over consecutive intervals of size b, each starting at some $a \in \{\frac{s_1}{b}, \ldots, \frac{s_2-b}{b}\}$, we have

$$g(ab+d) = g(ab) + k_a d, \text{ and so} \tag{A2}$$

$$g(ab) + k_a b = g((a+1)b) \tag{A3}$$

for any $d \in \{0, \ldots, b-1\}$, with $g(ab)$ constant over the given interval and k_a denoting the linear coefficient for that interval.

From (1) and (A2) it follows that

$$\begin{aligned} f(ab+d) &= \int_{ab+d}^{ab+d+1} g(x)\,dx \\ &= \int_{ab}^{ab+1} g(x)\,dx + \int_{ab+1}^{ab+d+1} g(x)\,dx \\ &= f(ab) + (g(ab) + k_a x)|_{ab+1}^{ab+d+1} \\ &= f(ab) + k_a d. \end{aligned} \tag{A4}$$

Using (A4), we can rewrite (A1) as

$$\sum_{a=\frac{s_1}{b}}^{\frac{s_2-b}{b}} (f(ab) + k_a d_1 - f(ab) - k_a d_2) = 0, \text{ and hence}$$

$$(d_1 - d_2) \sum_{a=\frac{s_1}{b}}^{\frac{s_2-b}{b}} k_a = 0. \tag{A5}$$

It now remains to be shown that $\sum_{a=\frac{s_1}{b}}^{\frac{s_2-b}{b}} k_a = 0$.

Recall from (A3) that we can write

$$g(s_2) = g(s_1) + b \sum_{a=\frac{s_1}{b}}^{\frac{s_2-b}{b}} k_a.$$

Since $g(s_1) = g(s_2)$ and $b > 0$, this implies $\sum_{a=\frac{s_1}{b}}^{\frac{s_2-b}{b}} k_a = 0$. □

Proof of Proposition 2. Recall that proposition 1 holds if equation (A5) is true. Given probability density $f(ab) + k_a d + f_e(ab + d)$, and recalling equation (A1), we rewrite (A5) as

$$(d_1 - d_2) \sum_{a=\frac{s_1}{b}}^{\frac{s_2-b}{b}} (k_a + f_e(ab + d_1) - f_e(ab + d_2)) = 0, \text{ which implies}$$

$$\sum_{a=\frac{s_1}{b}}^{\frac{s_2-b}{b}} f_e(ab + d_1) = \sum_{a=\frac{s_1}{b}}^{\frac{s_2-b}{b}} f_e(ab + d_2). \tag{A6}$$

□

Proof of Corollary 3. Suppose to the contrary that proposition 1 holds if d is additively separable from f_e. Then (A6) can be written as

$$\sum_{a=\frac{s_1}{b}}^{\frac{s_2-b}{b}} f_e(ab+d_1) = \sum_{a=\frac{s_1}{b}}^{\frac{s_2-b}{b}} f_e(ab+d_2), \text{ and hence}$$

$$h_e(d_1) = h_e(d_2),$$

which is not true if $h_e(d)$ is not constant over all $d \in \{0, \ldots, b-1\}$. Similarly we can show that proposition 1 does not hold if d is multiplicatively separable from f_e. □

Proof of Proposition 3. Consider any sequence $\{z, \ldots, z+2(b-1)\}$, where $z \in \{s_1, \ldots, s_2 - 2(b-1)\}$. Let this sequence be denoted q, and let Q denote the set of all such sequences of size $2b-1$ on the domain of f. We can approximate f in this sequence by arithmetic progression, which yields $f(z+d') = f(z) + k_z d' + f_e(z+d')$, where k_z is the common difference of successive elements of the sequence, f_e is some function that gives the error in approximation, and $d' \in \{0, \ldots, 2(b-1)\}$. Since we want to assess the average relative densities with which last digits appear across all sequences of size b, let's average f across all sequences of size b inside q. There are b unique sequences of size b wholly contained in $\{z, \ldots, z+2(b-1)\}$. Note that each last digit $d \in \{0, \ldots, b-1\}$ appears exactly once in each sequence of size b, each number $z+d$ appears in $d+1$ sequences, and correspondingly each number $z+b+d$ that is contained in q appears in $b-(d+1)$ sequences. We can then write the sum of weighted densities for numbers ending in d (i.e. the numbers $z+d$ and $z+b+d$) as

$$\begin{aligned}
& (d+1)f(z+d) + (b-(d+1))f(z+b+d) \\
= \quad & (d+1)(f(z) + k_z d + f_e(z+d)) \\
& \quad + (b-(d+1))(f(z) + k_z(b+d) + f_e(z+b+d)) \\
= \quad & f(z) + k_z(b^2 - b) + (d+1)f_e(z+d) - (b-(d+1))f_e(z+d+b).
\end{aligned}$$

In expectation we have $E[f_e(z+d)] = 0$, and so by taking expectations we are left with $\mathrm{E}[f(z)] + k_z(b^2 - b)$. Note that this density is not a function of d, i.e. in expectation it is identical for all $d \in \{0, \dots, b-1\}$. Thus last digits of the random variable X' are uniformly distributed in expectation, where X' has probability density $f(x)$ weighted by the probability with which x is included in an arbitrary sequence of length b in q. In other words, we have shown that the (unnormalized) density function $f(x)h(x,q)$ produces a uniform distribution of last digits, where $h(x,q)$ gives the probability that number x is included in any sequence of size b in q. It remains to be shown that $\sum_{q \in Q} h(x,q)$ is proportional to a constant (i.e. does not vary with x), or equivalently, that $\sum_{q \in Q} f(x)h(x,q)$ can be normalized to $f(x)$.

Function $h(x,q)$ is clearly not constant within the sequence $\{z, \dots, z+2(b-1)\}$, since the number of sequences of size b that include x varies with the position of x relative to z. But there are $2b-1$ sequences in Q that include x, and x is in a different position relative to z in each of these sequences. For any $x \in \{s_1 + 2(b-1), \dots, s_2 - 2(b-1)\}$, summing over Q then yields

$$
\begin{aligned}
\sum_{q \in Q} f(x)h(x,q) &= f(x) \sum_{q \in Q} h(x,q) \\
&= f(x) \left(\sum_{d=0}^{b-1} (d+1) + \sum_{d=0}^{b-1} (b - (d+1)) \right) \\
&= f(x) b^2 \sum_{d=0}^{b-1} ((d+1) - (d+1)) \\
&= f(x) b^2 \\
&\propto f(x).
\end{aligned}
$$

This leaves $x \in \{s_1, \dots, s_1 + 2b - 3; s_2 - 2b + 3, \dots , s_2\}$, that is x at the boundaries of the domain of f. For x at the lower bound, we have

$$\sum_{q\in Q} h(x,q) = \sum_{d=0}^{x-s_1}(d+1) \quad \text{for } x \in \{s_1,\ldots,s_1+b-1\}, \text{ and}$$

$$\sum_{q\in Q} h(x,q) = \sum_{d=0}^{b-1}(d+1) + \sum_{d=0}^{x-(s_1+b-1)}(b-(d+1))$$
$$\text{for } x \in \{s_1+b,\ldots,s_1+2b-3\},$$

where the sum of $h(x,q)$ over all elements of Q varies with x. This follows equivalently for x at the upper bound.

Hence we can normalize $\sum_{q\in Q} f(x)h(x,q)$ to $f(x)$ only if $f(x) = 0$ for $x \in \{s_1,\ldots,s_1+2b-3; s_2-2b+3,\ldots,s_2\}$. In other words, the density attributed to x at the upper and lower bounds of the domain determines the extent to which $f(x)$ is different from the (normalized) density $\sum_{q\in Q} f(x)h(x,q)$ and thus the extent to which last digits may follow a non-uniform distribution. For the relevant density at the lower bound of x we have

$$\begin{aligned} f(s_1)+\ldots+f(s_1+2b-3) &= \frac{f(s_1)+f(s_1+2b-3)}{2}(2b-2) \\ &= (b-1)(f(s_1)+f(s_1+2b-3)). \end{aligned}$$

Similarly we can compute the density over $x \in \{s_2-2b+3,\ldots,s_2\}$. It follows that as

$$(b-1)(f(s_1)+f(s_1+2b-3)+f(s_2-2b+3)+f(s_2)) \to 0$$

or, less generally, as $f(x)$ approaches 0 for $x \le s_1+2b-3$ and $x \ge s_2-2b+3$, the last digits of random variable X approach a uniform distribution. □