The Bernoulli-Trials Distribution and Wavelet

"..with my God, I can scale a wall" -David in Psalm 18

Lang P. Withers, Jr.

ABSTRACT: This paper begins by using the Haar wavelet to analyze the cumulative distribution function of a stream of two-valued Bernoulli trials. We find that this function maps binary numbers in [0,1] into non-uniform binary numbers in [0,1]. More generally, for r-valued Bernoulli trials, the distribution function likewise maps usual base-r numbers into non-uniform base-r numbers. We also find this distribution function obeys a two-scale dilation equation and can be used to construct a family of wavelets. This family contains the Haar wavelet and the piecewise-linear wavelet as special cases.

1. Introduction: the distribution of Bernoulli sequences

Consider a sequence $B_1B_2B_3\cdots$ of Bernoulli trials, a stream of independent, identically distributed random variables, each taking one of only two values. For $j = 1, 2, 3, \cdots$, let $B_j = 0$ with probability $p \geq 0$, and $B_j = 1$ with probability $q = 1-p$. This is a model of repeated tosses of a biased coin, for example. Let us represent a Bernoulli sequence $B_1B_2B_3\cdots$ by the random binary number $X = .B_1B_2B_3\cdots$ in [0,1]; that is, $X = \sum_j \frac{1}{2^j}B_j$. The B_j now serve as random bits. (In mapping Bernoulli sequences onto $[0,1]$, we identify those pairs of sequences that begin with the same J bits, but end in $0111\cdots$ and $1000\cdots$, respectively. The set of such pairs is countable with Lebesgue measure 0 [**3**].) Accordingly, we say that X has a meta-binomial or Bernoulli-trials distribution. (The binomial distribution of the random integer $\sum_{j=1}^{J} B_j$ is not used in this paper, but is also named for James Bernoulli.)

If we limit X to J bits, then for $k = 0, ..., 2^J - 1$, X takes the values $.k$ if we express k in base 2. For example, for $J = 2$, X takes the values $.00, .01, .10, .11$. So X takes the value $.k$ with probability

$$\begin{aligned}\pi_{J,k} &= \text{Prob}(X = .k = .b_1b_2\cdots b_J) \\ &= p_1p_2\cdots p_J,\end{aligned}$$

where $p_j = p$ whenever $b_j = 0$, and $p_j = q$ whenever $b_j = 1$. We define $\pi_{0,0} = 1$. For $J = 2$, we have the discrete probabilities $\pi_{2,00} = pp = p^2$, $\pi_{2,01} = pq$, $\pi_{2,10} = qp$, and $\pi_{2,11} = qq = q^2$. Thus X (with J bits) has a cumulative distribution function $\widetilde{F}_J(x)$ of scale J given by

$$\widetilde{F}_J(x) = \sum_{.k \leq x} \pi_{J,k}. \tag{1}$$

This function is continuous from the left. Example graphs of $\widetilde{F}_1, \widetilde{F}_2, \widetilde{F}_3$ are shown in Figure 1.

These distribution functions become easier to work with if we translate them to the right by one unit of length at scale J. We define the unit-translates F_J as

$$F_J(x) = \widetilde{F}_J\left(x - \tfrac{1}{2^J}\right). \tag{2}$$

Example graphs of F_1, F_2, F_3 are shown in Figure 2. Over each dyadic subinterval at scale

Math. Subject Classifications: 60E05 42C40 26A30
Keywords and Phrases: probability distribution, wavelet, singular function

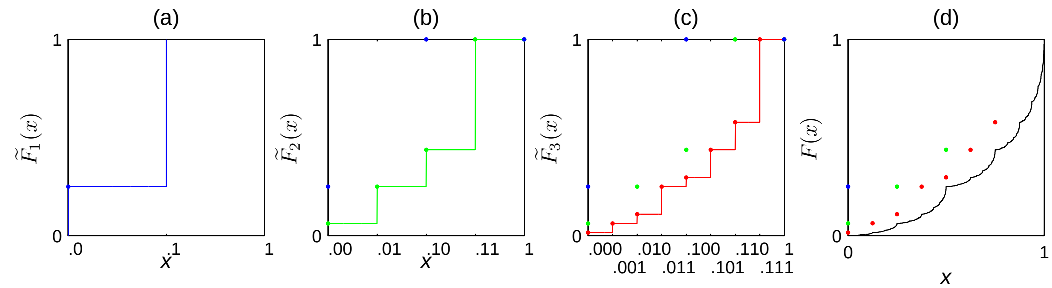

Figure 1: (a,b,c) The distribution functions $\widetilde{F}_J$ for $p = 1/4$, $J = 1, 2, 3$ and (d) their limit F as $J \to \infty$.

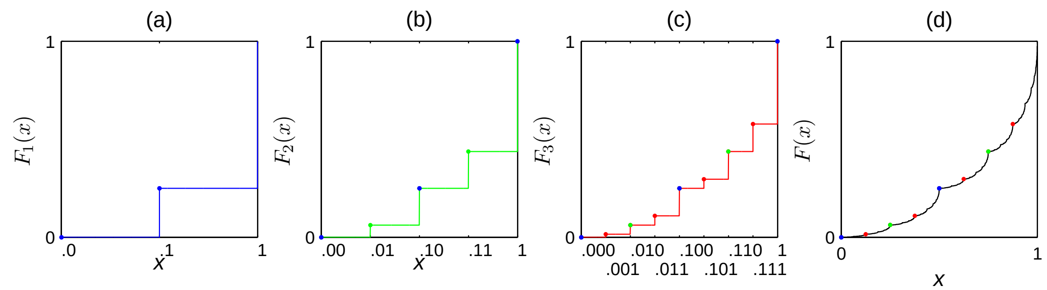

Figure 2: (a,b,c) The unit-translated distribution functions F_J for $p = 1/4$, $J = 1, 2, 3$ and (d) their limit F as $J \to \infty$.

J, the functions $F_J(x)$ increase or stay the same as J increases, while the original distribution functions $\widetilde{F}_J(x)$ decrease or stay the same. ($F_J(x)$ accretes, $\widetilde{F}_J(x)$ erodes.) However, as J increases, F_J behaves much more simply than $\widetilde{F}_J$. F_J is a refinement of F_{J-1}. The graph of F_J inherits all jump points of F_{J-1}, and between them interpolates an equal number of new ones (see Figure 2abc):

Lemma 1. For every scale-$(J-1)$ dyadic rational $x \in [0,1]$, $F_J(x) = F_{J-1}(x)$.

Proof. Using the definition of F_J and translation by a unit $u = 2^{-J}$ at scale J, we have the two partial sums $F_{J-1}(x) = \sum_{.k \le x-2u} \pi_{J-1,k} = \sum_{.k+2u \le x} \pi_{J-1,k}$ and $F_J(x) = \sum_{.k' \le x-u} \pi_{J,k'} = \sum_{.k'+u \le x} \pi_{J,k'}$. In the sum for F_{J-1}, at scale $J-1$, each dyadic rational atom $.k = .b_1 b_2 \ldots b_{J-1}$, with probability mass $\pi_{J-1,k}$, is translated to $.k + 2u$, and included in the sum if less than or equal to x. This atom splits into two atoms at scale J, $.k = .b_1 b_2 \ldots b_{J-1} 0$ and $.k + u = .b_1 b_2 \ldots b_{J-1} 1$, with probability masses $\pi_{J-1,k} \cdot p$ and $\pi_{J-1,k} \cdot q$, respectively, which give a total mass $\pi_{J-1,k}$ for the pair of atoms. These atoms are translated by u in the sum for F_J to $.k + u$ and $.k + 2u$, respectively; note that the second scale-J atom lies exactly where the scale-$(J-1)$ atom is, and the first scale-J atom lies one unit to their left. Therefore, each mass in the sum for F_{J-1} (from $.k = 0$ through $.k = x$) is exactly replaced by the pair of masses in the sum for F_J, and the two sums are equal. (For example, for $J = 2$ with unit $u = \frac{1}{4} = .01$, the scale-1 atom at $.0 + 2u = .1$ has mass p, and the scale-2 atoms at $.0 + u = .01$ and $.0 + 2u = .10$ have masses p^2 and pq, and $p^2 + pq = p$). $\square$

Though $\widetilde{F}_J$ introduces as many new jump points as F_J does when we change from scale $J-1$ to J, the old jump points of $\widetilde{F}_{J-1}$ do not remain fixed, but all shift to the right by one unit of length (2^{-J}) at scale J (see Figure 1abc). Once introduced, each jump point moves horizontally by half of the previous step, like the runner in Xeno's paradox [**4**]. In the limit as $J \to \infty$, each new jump point of $\widetilde{F}_J$ will reach the corresponding stationary jump point of F_J. As $J \to \infty$, we have $F_J(x) \uparrow F(x)$, since $F_{J+1}(x) - F_J(x) \le pq^J$ for $p \le q$, or qp^J

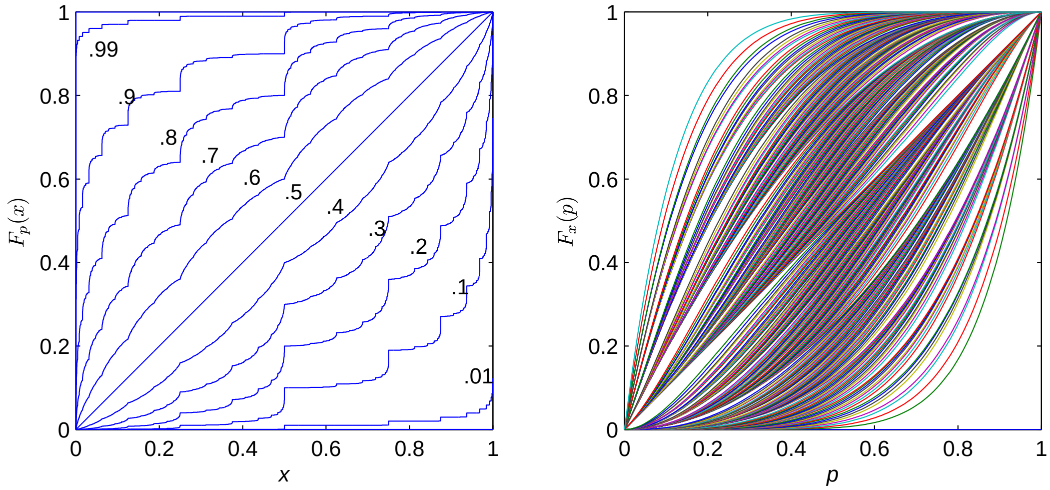

Figure 3: The Bernoulli-sequence distribution function F: as a function of x for various values of p (left side); as a function of p for $x = k/256$, $k = 0, \ldots, 255$ (right side). On the right, note that $F_0(p) \equiv 0, F_1(p) \equiv 1$, $F_{\frac{1}{2}}(p) = p$. Also note the bundling or non-uniformity of the curves on the right.

for $q \le p$. Similarly, $\widetilde{F}_J(x) \downarrow \widetilde{F}(x)$. Continuity of both limit functions is easily verified; this implies that $\widetilde{F}(x) = F(x)$ since the two limits coincide at every dyadic rational x.

Partitioning the area above $\widetilde{F}_J(x)$ and below 1 into the horizontal rectangles to the left of each jump in the graphs in Figure 1 [**2**], we easily verify that X_J has average value

$$\begin{aligned} \mathrm{E}(X_J) &= \sum .k\pi_{J,k} = \int_0^\infty (1 - \widetilde{F}_J(x))dx = 1 - \int_0^1 \widetilde{F}_J(x)dx \\ &= \frac{2^J - 1}{2^J} q. \end{aligned} \tag{3}$$

Since F_J is the cumulative distribution function of $Y_J = X_J + 2^{-J}$, $\mathrm{E}(Y_J) = q + 2^{-J}p$. We also observe that

$$F_{J+1}(x) = \begin{cases} pF_J(2x) & \text{when } x < \frac{1}{2}, \\ p + qF_J(2x - 1) & \text{when } x \ge \frac{1}{2}. \end{cases} \tag{4}$$

$\widetilde{F}_J$ satisfies this same recursion.

For $p = q = \frac{1}{2}$, $F(x) = x$ for all $x \in [0, 1]$, the uniform distribution function. Otherwise, $F(x)$ is known to be singular; that is, $F'(x) = 0$ except on a set of measure 0, those x's whose bit average is not p. Also $F(x)$ increases as x goes from 0 to 1 [**2**]. Sometimes we label F with its parameter p, as F_p. For $x = \frac{1}{2}$, $F_p(1/2) = p$ for all $p \in [0, 1]$. Plots of $F_p(x)$ as a function of x and as a function of p are given in Figure 3. From (3), we have $\mathrm{E}(X) = q$ and $\int_0^1 F(x)dx = p$. $F(x)$ is self-similar (fractal), since from (4),

$$F(x) = \begin{cases} pF(2x) & \text{when } x < \frac{1}{2}, \\ p + qF(2x - 1) & \text{when } x \ge \frac{1}{2}. \end{cases} \tag{5}$$

The distribution $F(x)$ also appears as the probability of winning a game of bold bets on Bernoulli trials, beginning with an amount of money $x \in [0, 1]$; the game stops when $x = 0$ (loss) or $x = 1$ (win) [**2**].

2. The Bernoulli-trials distribution as viewed by a Haar wavelet basis

Since $\tilde{F}_J$ is a staircase function of scale J supported on $[0,1]$, it has an expansion in the orthonormal Haar wavelet basis on $[0,1)$,

$$\tilde{F}_J(x) = \langle \tilde{F}_J, p_{0,0} \rangle p_{0,0}(x) + \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \langle \tilde{F}_J, h_{j,k} \rangle h_{j,k}(x), \tag{6}$$

where $p_{0,0} = \chi_{[0,1)}$ is the Haar scaling function, $h_{0,0} = \chi_{[0,\frac{1}{2})} - \chi_{[\frac{1}{2},1)}$ is the Haar wavelet, with its dilated and translated copies given by $h_{j,k}(x) = 2^{\frac{j}{2}} h_{0,0}(2^j x - k)$ for $k = 0, 1, \ldots 2^j - 1$ [**1**]. We always have $\tilde{F}_J(1) = 1$, so we do not need an expansion for the endpoint $x = 1$.

Theorem 2. For every $x \in [0,1)$,

$$\tilde{F}_J(x) = \alpha_{0,0} p_{0,0}(x) + \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \beta_{j,k} h_{j,k}(x), \tag{7}$$

where for $j = 0 : J-1$

$$\begin{aligned} \alpha_{0,0}^{(J)} &= p + \frac{1}{2^J} q, \\ \beta_{j,k}^{(J)} &= -\left(\frac{1}{2^{J-\frac{j}{2}}} q + \frac{2^{J-j-1} - 1}{2^{J-\frac{j}{2}-1}} pq \right) \cdot \pi_{j,k}. \end{aligned}$$

Proof. By induction on J. For $J = 1$, we have $\alpha_{0,0}^{(1)} = \langle \tilde{F}_1, p_{0,0} \rangle = \frac{1}{2} p + \frac{1}{2} = p + \frac{1}{2} q$. Likewise, $\beta_{0,0}^{(1)} = \langle \tilde{F}_1, h_{j,k} \rangle = \frac{1}{2} p - \frac{1}{2} = -\frac{1}{2} q$. Assume the coefficients are correct as given for scale J. Then using (4), we have

$$\begin{aligned} \alpha_{0,0}^{(J+1)} &= \langle \tilde{F}_{J+1}, p_{0,0} \rangle \\ &= \frac{1}{2} p \cdot \langle \tilde{F}_J, p_{0,0} \rangle + \frac{1}{2} p + \frac{1}{2} q \cdot \langle \tilde{F}_J, p_{0,0} \rangle \\ &= \frac{1}{2} p + \frac{1}{2} \left(p + \frac{1}{2^J} q \right) \pi_{0,0} \\ &= p + \frac{1}{2^{J+1}} q, \end{aligned}$$

$$\begin{aligned} \beta_{0,0}^{(J+1)} &= \langle \tilde{F}_{J+1}, h_{0,0} \rangle \\ &= \frac{1}{2} p \cdot \langle \tilde{F}_J, p_{0,0} \rangle - \left(\frac{1}{2} p + \frac{1}{2} q \cdot \langle \tilde{F}_J, p_{0,0} \rangle \right) \\ &= -\frac{1}{2} p + \frac{1}{2} (p - q) \left(p + \frac{1}{2^J} q \right) \pi_{0,0} \\ &= -\left(\frac{1}{2^{J+1} p} + \frac{2^J - 1}{2^J} \right) pq. \end{aligned}$$

For $j > 0$ and $k = 0, \ldots, 2^{j-1} - 1$, the induction step follows by using the top of (4) and the fact that $\pi_{j,k} = p \pi_{j-1,k}$:

$$\begin{aligned} \beta_{j,k}^{(J+1)} &= \langle \tilde{F}_{J+1}, h_{j,k} \rangle \\ &= \frac{1}{2} p \sqrt{2} \langle \tilde{F}_J, h_{j-1,k} \rangle = 2^{-\frac{1}{2}} p \beta_{j-1,k}^{(J)}. \end{aligned}$$

For $j > 0$ and $k' = k + 2^{j-1} = 2^{j-1}, \dots, 2^j - 1$, the induction step follows by using the bottom of (4) and the fact that $\pi_{j,k'} = q\pi_{j-1,k}$:

$$\begin{aligned} \beta_{j,k'}^{(J+1)} &= \langle \tilde{F}_{J+1}, h_{j,k'} \rangle \\ &= \frac{1}{2} q\sqrt{2} \langle \tilde{F}_J, h_{j-1,k} \rangle = 2^{-\frac{1}{2}} q \beta_{j-1,k}^{(J)}. \ \square \end{aligned}$$

Taking the right and left hand limits as $J \to \infty$ in Theorem 1, we immediately find

Corollary 3.

$$F(x) = p \cdot p_{0,0}(x) + \sum_{j=0}^{\infty} \sum_{k=0}^{2^j - 1} -pq\pi_{j,k} \cdot 2^{-\frac{j}{2}} h_{j,k}(x). \tag{8}$$

Now $2^{-\frac{j}{2}} h_{j,k}(x)$ is just the wavelet supported on the half-open interval $I_{j,k} = [2^{-j}k, 2^{-j}(k+1))$ with amplitude ± 1. Imagine the vertical line through a fixed value x. For each scale j, it slices through the Haar wavelet with support over the $I_{j,k}$ which contains x. At scale j, we keep only that k in the inside sum in (8). This will not change the sum for this x value. Indeed, if we express $x = .b_1 b_2 b_3 \cdots$ in binary form, for scale j, $k_{x,j} = b_1 b_2 \cdots b_j$ is just x truncated to j bits. The corresponding probability $\pi_{j,k} = p_1 p_2 \cdots p_j$ also follows the pattern of the bit sequence of x, since we have ve defined $p_i = p$ if $b_i = 0$ and $p_i = q$ if $b_i = 1$. As the scale increases from $j - 1$ to j, we go to the left half-interval $I_{j,k}^{\ell}$ if the last bit b_j is 0, and to the right half-interval $I_{j,k}^{r}$ if it is 1. So the sign of the Haar wavelet $h_{j,k}$ we keep at scale j will be +1 or -1, according as the next bit is 0 or 1, respectively. Thus the Haar wavelet expansion (8) reduces to a closed-form series for $F(x)$:

Corollary 4. *Given $x = .b_1 b_2 b_3 \cdots$,*

$$\begin{aligned} F(x) &= p - pq \sum_{j=0}^{\infty} (-1)^{k_{x,j+1}} \pi_{j,k_{x,j}} \\ &= p - pq \sum_{j=0}^{\infty} (-1)^{b_{j+1}} p_1 p_2 \cdots p_j. \end{aligned} \tag{9}$$

For example, if $x = .010010 \cdots$,

$$F(x) = p - pq \left(1 - p + pq + pqp - pqpp + pqppq \pm pqppqp \pm \cdots \right).$$

Since for $p = \frac{1}{2}$, $F_p(x) = x$ for $x \in [0,1]$, Corollary 3 implies $\frac{1}{2} - \frac{1}{4} \sum_{j=0}^{\infty} (-1)^{b_{j+1}} (\frac{1}{2})^j = \sum_{j=1}^{\infty} b_j (\frac{1}{2})^j$. There is a more general form of this identity:

Lemma 5. *Given $x = .b_1 b_2 b_3 \cdots$,*

$$\frac{p}{q} \sum_{j=1}^{\infty} b_j \cdot p_1 p_2 \cdots p_j \quad = \quad p - pq \sum_{j=0}^{\infty} (-1)^{b_{j+1}} p_1 p_2 \cdots p_j. \tag{10}$$

Proof. First we prove this for the case of dyadic rationals of scale J, $x = .b_1 b_2 \cdots b_J$. Here J is the finite index of the last $b_j = 1$; that is, when $b_j = 0$ (and $p_j = p$) for every $j > J$. For $J = 1$, the left side is 0 if $b_1 = 0$, and $\frac{p}{q} \cdot q = p$ if $b_1 = 1$. (If $q = 0$, we define $q/q = 1$.) The right side is $p - pq\left((-1)^{b_1} + p_1 \cdot (1 + p + p^2 + \cdots)\right) = p - pq\left((-1)^{b_1} + p_1 \cdot \frac{1}{q}\right)$. This also reduces to 0 or p, for $b_1 = 0$ or 1, respectively.

Now assume the identity true when the last 1-bit occurs at index J. When the $(J+1)$st bit b_{J+1} changes from 0 to 1, the left-hand sum increases by the new term $\ell_J = p \cdot p_1 \cdots p_J$. The sum of terms on the right-hand side changes by $-pq \cdot \left(p_1 \cdots p_J q \cdot \frac{1}{q} - p_1 \cdots p_J\right) - pq \cdot$

$\left(p_1 \cdots p_J \cdot \frac{1}{q}\right) = -pq \cdot \left(0 - p_1 \cdots p_J \cdot \frac{1}{q}\right) = p_1 \cdots p_J p$. This completes the induction step. Since the identity holds for every $J \geq 1$, it is clear that it also holds in the limit as the series converge. □

Lemma 5 lends new meaning to Corollary 4, which can now be expressed as

Theorem 6. *Given* $x = .b_1 b_2 b_3 \cdots$,

$$F(x) = \frac{p}{q} \sum_{j=1}^{\infty} b_j p_1 p_2 \cdots p_j. \tag{11}$$

We can regard the right-hand series as a p-binary representation of the number $y = F_p(x)$, and write it as $(.b_1 b_2 b_3 \cdots)_p$. The usual binary representation of x is $(.b_1 b_2 b_3 \cdots)_{\frac{1}{2}}$. The meaning of F_p now becomes clear. The map F_p is one-to-one, and takes $x = (.b_1 b_2 b_3 \cdots)_{\frac{1}{2}}$ to $y = (.b_1 b_2 b_3 \cdots)_p$. The p-binary representation (11) of a number x is an infinite polynomial or power series when considered as a function of the variables p and q. Each bit is the coefficient of a product of a sequence of p's and q's that encodes the history of the preceding bits as well. This product can be regarded as the signature or pedigree of the bit. The usual case $p = \frac{1}{2}$ is a degenerate case, in which the bit coefficient $\frac{1}{2^j}$ is fixed and carries no implicit history of the bits before it. Just as the usual binary expansion is said to be for base 2, so also we may think of this as a twin Hölder base with $\frac{1}{a} + \frac{1}{b} = 1$, $a = 1/p \geq 1$. Note that $F_q(x) = 1 - F_p(1 - x)$. To find $F_p(x)$, one must have x in binary form. Likewise, to find $F_p^{-1}(y)$, one only needs to digitize y. We can digitize a real number y in the base $\left(\frac{1}{p}, \frac{1}{q}\right)$ in the same manner that we find the bits of a number x base 2; but we will not detail this here.

Since $F(x)$ equals $F_J(x)$ for dyadic rationals x of scale J, we see that the partial sums on both sides of (10) in Lemma 4 are actually the expansions for $F_J(x)$ (not $\tilde{F}_J(x)$):

Theorem 7. *Given* $x = .b_1 b_2 b_3 \ldots b_J$,

$$F_J(x) = \sum_{j=1}^{J} p_1 p_2 \cdots p_{j-1} P_j \tag{12}$$

$$= \frac{p}{q} \sum_{j=1}^{J} b_j p_1 p_2 \cdots p_j, \tag{13}$$

where we define $P_j = 0$ *if* $b_j = 0$, and $P_j = p$ *if* $b_j = 1$.

Proof. By induction: the equality (12) for $J = 1$ is easily verified (the empty product $p_1 \cdots p_{j-1}$ for $j = 1$ is always defined as 1). Assume the equality holds for $J - 1 > 0$, and let $x = .k = .b_1 b_2 b_3 \ldots b_{J-1} b_J = .k' b_J = x' + b_J u$, where $u = 2^{-J}$ is again the scale-J unit, and $x' = .k'$ is a scale-$(J-1)$ dyadic rational number formed by the first $J - 1$ bits of x. If $b_J = 0$, $x = x'$ and the left side of (12) is $F_J(x) = F_J(x') = F_{J-1}(x')$ by Lemma 1; the right side clearly reduces to the scale-$(J-1)$ expression for x' as well. If $b_J = 1$, this bit increases each side above its scale-$(J-1)$ expression by just one term: since $F_J = \tilde{F}_J(x-u)$, $F_J(x) = F_J(x') + \pi_{J,k'0} = F_{J-1}(x') + \pi_{J,k'0}$, where the increase is $\pi_{J,k'0} = p_1 p_2 \cdots p_{j-1} \cdot p$. But evidently the right side is also the sum for scale $J - 1$ increased by this same term. In both cases, then, equality of both sides of (12) holds by the induction hypothesis. The expression (13) now follows by noting that by definition $P_j = \frac{p}{q} p_j \cdot b_j$, where we take $\frac{q}{q} = 1$ even when $q = 0$. □

Remark: The sum for $F_J(x)$ based on (1) has on average $2^J/2$ terms, while the right side of equality (12) reduces this to only J terms. The proof only depends on Lemma 1. Thus, having seen the relation (12) by means of the Haar expansion, we could begin anew with this relation and forget how we originally saw it. In Section 4, we will do this when we come to generalize this relation.

We can reexpress identity (4) as

$$\begin{aligned} F\left(\tfrac{1}{2}x\right) &= pF(x) \\ F\left(\tfrac{1}{2}x+\tfrac{1}{2}\right) &= p+qF(x), \end{aligned} \tag{14}$$

for every $x \in [0,1]$. Theorem 6 makes it easy to generalize that identity to this:

Corollary 8. *Given $x \in [0,1]$ and $0 \le k \le 2^j - 1$. Then*

$$F(2^{-j}(x+k)) \quad = \quad F(2^{-j}k) + \pi_{j,k}F(x). \tag{15}$$

Proof. In usual binary terms, let $x = .b_1b_2b_3\cdots$, and $k = k_1k_2\cdots k_j$. Then $2^{-j}k \le 1-2^{-j}$, and $2^{-j}x \le 2^{-j}$. Therefore $2^{-j}k + 2^{-j}x = .k_1k_2\cdots k_jb_1b_2b_3\cdots$ and, by Theorem 5, $F_p(2^{-j}k+2^{-j}x) = F_p(2^{-j}k) + \pi_{j,k}(p,q)F_p(x)$. $\square$

3. The Bernoulli-trials wavelet

The identity (4) looks like a two-scale dilation equation. Indeed, it can be rewritten as one with only two coefficients p and q:

$$F_{J+1}(x) = pF_J(2x) + qF_J(2x-1). \tag{16}$$

Taking the limit as the number of bits $J \to \infty$, we have a two-scale dilation equation

$$F(x) = pF(2x) + qF(2x-1). \tag{17}$$

Remark. It is remarkable that this two-scale relation arises naturally in the context of the probability distribution of a random binary number. This goes to show that such two-scale relations are not just a human contrivance.

To turn $F(x)$ into a scaling function in $L^1(\mathbb{R})$, we proceed as in [**1**] for the piecewise-linear case ($p = \frac{1}{2}$) and define, for each integer $J > 0$,

$$G_J(x) = F_J\left(x+\tfrac{1}{2}\right) - F_J\left(x-\tfrac{1}{2}\right). \tag{18}$$

This scaling function obeys a two-scale dilation equation with three coefficients,

$$G_{J+1}(x) = pG_J(2x+1) + G_J(2x) + qG_J(2x-1). \tag{19}$$

Taking the Fourier transform of both sides gives us

$$\begin{aligned} \hat{G}_{J+1}(\gamma) &= \frac{1}{2}\left(pe^{i\pi\gamma} + 1 + qe^{-i\pi\gamma}\right)\hat{G}_J(\gamma/2) \\ &= \frac{1}{2}\left(pe^{i\pi\gamma} + 1 + qe^{-i\pi\gamma}\right)\hat{G}_J(\gamma/2) \\ &= m_0(\gamma/2)\hat{G}_J(\gamma/2), \end{aligned} \tag{20}$$

where

$$\begin{aligned} m_0(\gamma) &= \frac{1}{2}\left(pe^{i2\pi\gamma} + 1 + qe^{-i2\pi\gamma}\right) \\ &= \frac{1}{2}\left(e^{i\pi\gamma} + e^{-i\pi\gamma}\right)\left(pe^{i\pi\gamma} + qe^{-i\pi\gamma}\right) \\ &= \cos\pi\gamma\left(pe^{i\pi\gamma} + qe^{-i\pi\gamma}\right). \end{aligned} \tag{21}$$

Now we apply the two-scale relation (20) over and over and take the limit as $J \to \infty$ to get

$$\begin{aligned}\hat{G}(\gamma) &= \prod_{j=0}^{\infty} m_0(\gamma/2^j) \\ &= \prod_{j=1}^{\infty} \cos(\pi\gamma/2^j) \cdot \prod_{j=1}^{\infty} \left(pe^{i\pi\gamma/2^j} + qe^{-i\pi\gamma/2^j}\right) \\ &= \operatorname{sinc}(\pi\gamma) \cdot \operatorname{inc}_p(\pi\gamma), \end{aligned} \tag{22}$$

where we use the identity [**2**] [**1**]

$$\operatorname{sinc}(\pi\gamma) = \prod_{j=1}^{\infty} \cos(\pi\gamma/2^j), \tag{23}$$

for $\operatorname{sinc}(\alpha) = \sin\alpha/\alpha$, and we also define

$$\operatorname{inc}_p(\pi\gamma) = \prod_{j=1}^{\infty} \left(pe^{i\pi\gamma/2^j} + qe^{-i\pi\gamma/2^j}\right). \tag{24}$$

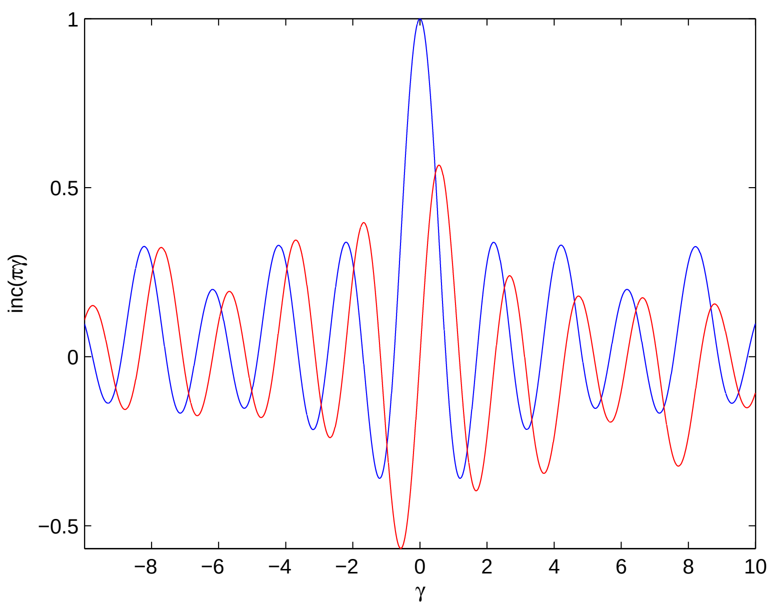

Figure 4: Graph of $\operatorname{inc}_{\frac{1}{4}}$. The real part is even (blue), the imaginary part is odd (red).

Let us pause to examine this inc_p function before going on. Figure 4 shows $\operatorname{inc}_{\frac{1}{4}}$ as an example. For $p = \frac{1}{2}$, $\operatorname{inc}_{\frac{1}{2}}(\pi\gamma) = \operatorname{sinc}(\pi\gamma)$, and for $p = 0, 1$, $\operatorname{inc}_p(\pi\gamma) = e^{\pi\gamma}, e^{-\pi\gamma}$, respectively. Other expressions for inc_p are as follows:

$$\operatorname{inc}_p(\pi\gamma) = \prod_{j=1}^{\infty} \left(\cos(\pi\gamma/2^j) + i(p-q)\sin(\pi\gamma/2^j)\right) \tag{25}$$

$$= \prod_{j=1}^{\infty} \left(e^{i\theta_j}\sqrt{1 - 4pq\sin^2(\pi\gamma/2^j)}\right), \tag{26}$$

where we put $\tan\theta_j = (p-q)\tan(\pi\gamma/2^j)$. (Note that $|p-q| = \sqrt{(1-4pq)}$.) Since these angles just reverse sign if we exchange the roles of p and q, we see that $\mathrm{inc}_q = \overline{\mathrm{inc}}_p$. From now on, we will write inc_p without its subscript p.

Taking the inverse Fourier transform of (22), we have a convolution

$$G(x) \quad = \quad \chi_{[-1,1]}(x) * \circledast_{j=1}^{\infty} \left(-p \cdot \delta\left(x + \tfrac{1}{2^j}\right) + q \cdot \delta\left(x - \tfrac{1}{2^j}\right)\right). \tag{27}$$

Here we see that the right-hand convolution is the inverse Fourier transform of inc_p and generates a dense series of terms such as $-pq \cdot \delta\left(x - \frac{1}{2^j} + \frac{1}{2^k}\right)$ for $j \neq k$. The window $\chi_{[-1,1]}$ integrates these δ-terms as it moves from left to right, to give G.

Now we resume our wavelet development. The next step is to orthonormalize the translates of $G(x)$ to make it a true scaling function. We use the method of Lemma 7.7 [**1**] to do this in the Fourier domain, and put

$$\hat{\varphi}(\gamma) \quad = \quad \Phi^{-\frac{1}{2}}(\gamma) \cdot \hat{G}(\gamma), \tag{28}$$

for the 1-periodization of $\left|\hat{G}(\gamma)\right|^2$,

$$\Phi(\gamma) \quad = \quad \sum_{n\in\mathbb{Z}} \left|\hat{G}(\gamma+n)\right|^2. \tag{29}$$

We will verify below that $\Phi(\gamma) > 0$. Then since $\Phi(\gamma)$ has period 1, $\sum_{n\in\mathbb{Z}} |\hat{\varphi}(\gamma+n)|^2 = \Phi^{-2}(\gamma) \cdot \sum_{n\in\mathbb{Z}} \left|\hat{G}(\gamma+n)\right|^2 = 1$. This condition is equivalent to orthonormality of the translates $\varphi(x-n), n \in \mathbb{Z}$ by Lemma 7.4 [**1**]. By Exercise 7.11 in [**1**],

$$\Phi(\gamma) \quad = \quad \sum_{n\in\mathbb{Z}} \langle G(x), T_n G(x)\rangle e^{-i2\pi\gamma n}, \tag{30}$$

where the translation operator T_n is defined for any function $f(x)$ by $T_n f(x) = f(x-n)$.

In this case, the support of $G(x)$ is [-1,1]. Define the translation coefficients by $t_n^{(J)} = \langle G_J(x), T_n G_J(x)\rangle$. Clearly $t_n^{(J)} = 0$ except for n=-1,0,1. To find t_{-1}, t_0, t_1, we observe that $t_{-1} = \bar{t}_1$, and note they have recursive relations:

Lemma 9. *For a=pq and a+b=1, and* $J > 0$*,*

$$t_0^{(J+1)} \quad = \quad b t_0^{(J)} + t_1^{(J)} \tag{31}$$

$$2t_1^{(J+1)} \quad = \quad a t_0^{(J)} + t_1^{(J)} \tag{32}$$

$$1 \quad = \quad t_0^{(J)} + 2t_1^{(J)}. \tag{33}$$

Proof. This follows by induction on J from the original relations (4).

By Lemma 7, taking limits as $J \to \infty$, we have

$$\begin{aligned} t_0 &= bt_0 + t_1 \\ 2t_1 &= at_0 + t_1 \\ 1 &= t_0 + 2t_1. \end{aligned}$$

The first two relations both reduce to $t_1 = at_0$. So we use the third relation to solve for t_0, t_1 and find

$$t_0 \quad = \quad \frac{1}{1+2pq} \tag{34}$$

$$t_1 \quad = \quad \frac{pq}{1+2pq}. \tag{35}$$

Therefore, we arrive at

$$\begin{aligned}\Phi(\gamma) &= t_{-1}e^{i2\pi\gamma} + t_0 + t_1 e^{-i2\pi\gamma}. \\ &= \frac{1}{1+2pq}\left(1 + 2pq\cos(2\pi\gamma)\right). \end{aligned} \tag{36}$$

It is easy to verify that this Φ meets the frame condition (7.4) of [**1**]: Since $pq = p(1-p) \leq \frac{1}{4}$, we have

$$\frac{1}{3} \leq \frac{1-2pq}{1+2pq} \leq \Phi(\gamma) \leq \frac{1+2pq}{1+2pq} = 1. \tag{37}$$

Thus, in the Fourier domain, we have the scaling function given by

$$\hat{\varphi}(\gamma) = \sqrt{\frac{1+2pq}{1+2pq\cos(2\pi\gamma)}} \cdot \hat{G}(\gamma). \tag{38}$$

The Fourier transform of the wavelet is now given by

$$\hat{\psi}(\gamma) = m_1(\gamma/2)\hat{\varphi}(\gamma/2) \tag{39}$$

$$= -e^{-i\pi\gamma} \cdot \frac{1}{2}\left(1 - pe^{i\pi\gamma} - qe^{-i\pi\gamma}\right) \cdot \sqrt{\frac{(1-2pq\cos(\pi\gamma)) \cdot (1+2pq)}{(1+2pq\cos(2\pi\gamma)) \cdot (1+2pq\cos(\pi\gamma))}} \cdot \hat{G}(\gamma/2). \tag{40}$$

To compute the scaling function φ in the time domain, we express $\Phi^{-\frac{1}{2}}(\gamma)$ as an inverse Fourier series $\Phi^{-\frac{1}{2}}(\gamma) = \sum_n c(n)e^{i2\pi\gamma n}$. Then we have

$$\varphi(x) = \sum_n c(n)G(x-n). \tag{41}$$

Similarly, to compute the wavelet ψ, we express the function that multiplies $\hat{G}(\gamma/2)$ in (40) as an inverse Fourier transform to obtain

$$\psi(x) = \sum_n d(n)G(2x-n). \tag{42}$$

Note that $m_0(0) = 1$, and by using the last expression in (26), we readily verify that

$$|m_0(\gamma)|^2 + |m_0(\gamma+\tfrac{1}{2})|^2 = \cos^2(\pi\gamma) \cdot \left(1 - 4pq\sin^2(\pi\gamma)\right) \frac{(1+2pq\cos(2\pi\gamma))}{(1+2pq\cos(4\pi\gamma))} \tag{43}$$

$$+ \sin^2(\pi\gamma) \cdot \left(1 - 4pq\cos^2(\pi\gamma)\right) \frac{(1-2pq\cos(2\pi\gamma))}{(1+2pq\cos(4\pi\gamma))}$$

$$= 1, \tag{44}$$

for all values of γ. By Theorem 8.13 in [**1**], φ and ψ comprise an orthonormal multi-resolution analysis (MRA).

Plots of φ and ψ and their associated functions are given for $p = \frac{1}{4}$ in Figure (5). We can see the spectrum $\hat{\psi}(\gamma)$ is near-zero in regular intervals, about the zeroes at $\gamma = n\pi$ due to the factor $\mathrm{sinc}(\pi\gamma)$ that $\hat{\psi}$ gets from $\hat{G}$. This trait is most pronounced for $p = \frac{1}{2}$. Plots of φ and ψ for various values of p are given in Figure (6). For $p = \frac{1}{2}$, we have the piecewise-linear MRA. For $p = 0$, we have the Haar MRA. As $p \to 0$, φ and ψ decay more rapidly outside $[-1, 1]$ until the support outside vanishes at $p = 0$. For $p' = 1 - p$, φ and ψ are the same as for p but reversed. So for $p = 1$, we have the reversed Haar functions.

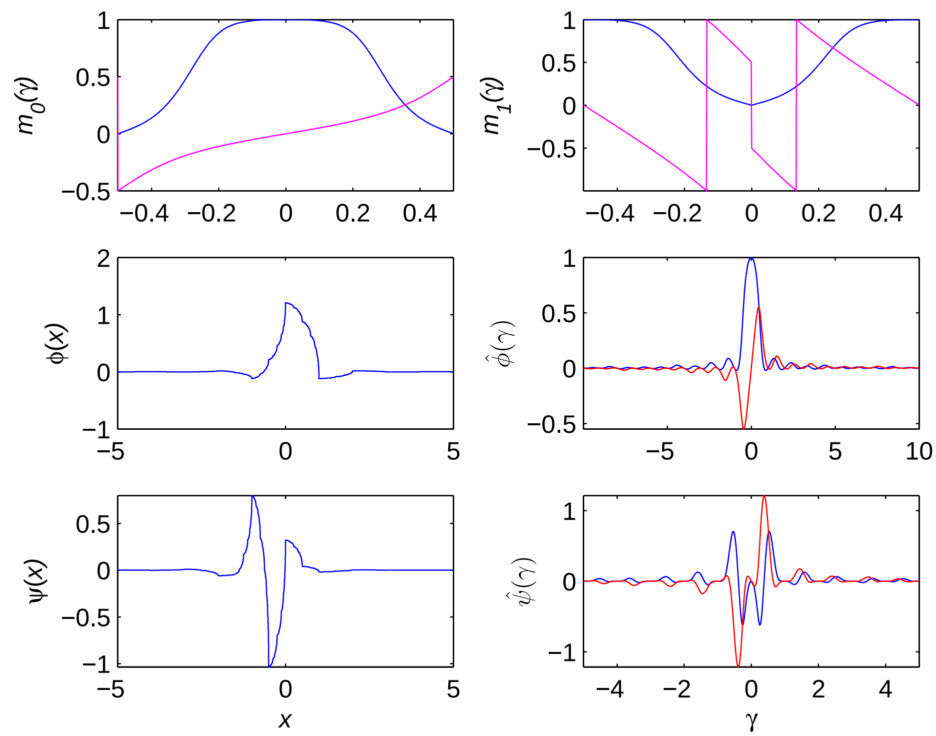

Figure 5: The Bernoulli scaling function $\varphi(x)$ and wavelet $\psi(x)$ and associated functions for $p = \frac{1}{4}$. The real and imaginary parts of the Fourier transforms $\hat{\varphi}(\gamma)$ and $\hat{\psi}(\gamma)$ are plotted; the real parts are even functions. For m_0 and m_1, amplitude and phase (argument) are plotted; the amplitudes are even.

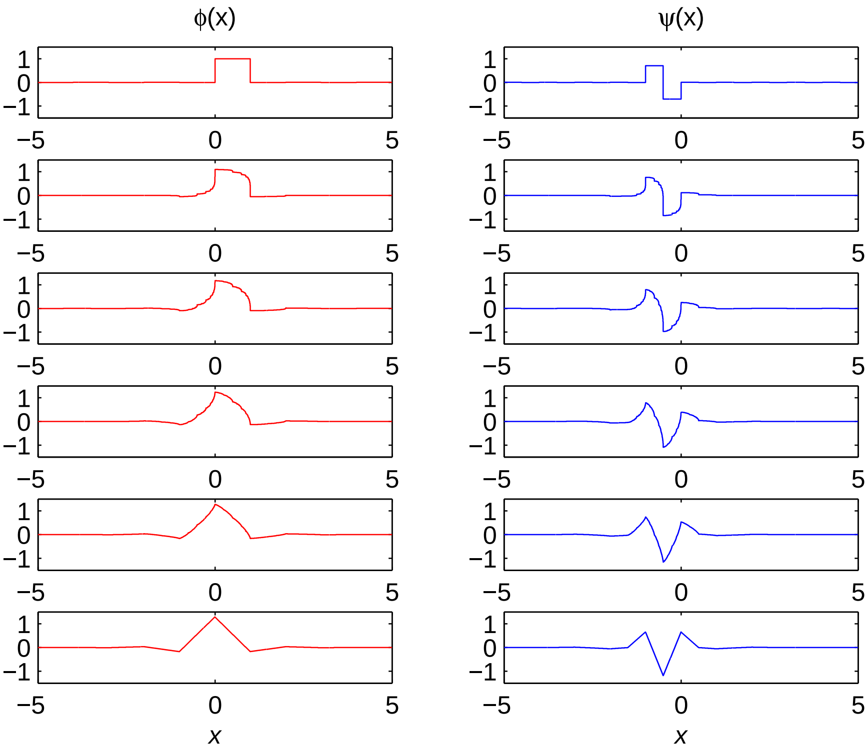

Figure 6: The Bernoulli scaling function ϕ and wavelet ψ for $p = 0, .1, .2, .3, .4, .5$ (top to bottom).

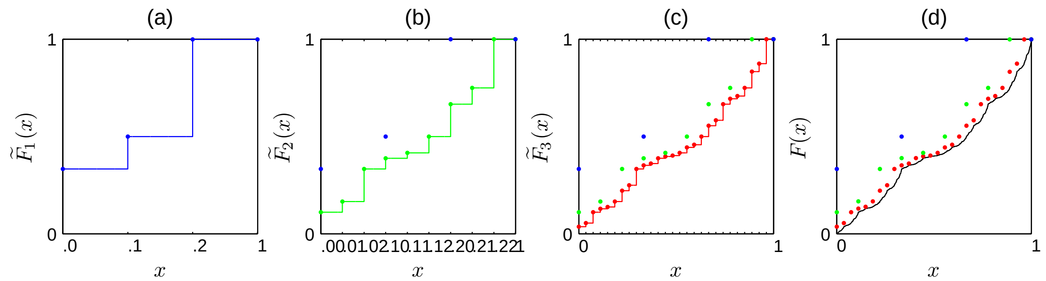

Figure 7: (a,b,c) The distribution functions $\widetilde{F}_J$ for $(p_1, p_2, p_3) = (\frac{1}{3}, \frac{1}{6}, \frac{1}{2})$, $J = 1, 2, 3$ and (d) their limit F as $J \to \infty$.

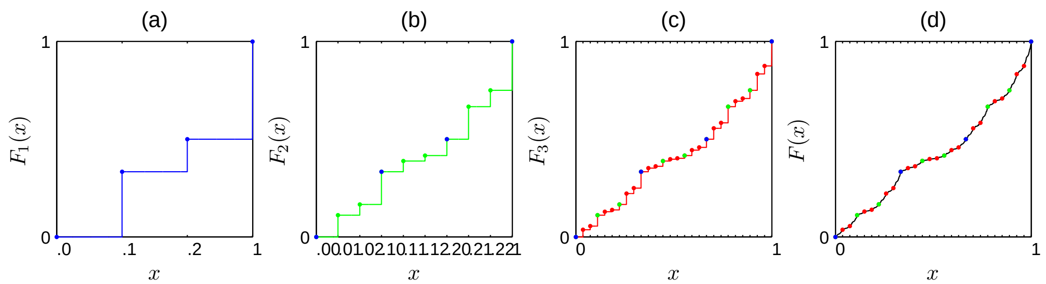

Figure 8: (a,b,c) The unit-translated distribution functions F_J for $(p_1, p_2, p_3) = (\frac{1}{3}, \frac{1}{6}, \frac{1}{2})$, $J = 1, 2, 3$; and (d) their limit F as $J \to \infty$.

4. The general Bernoulli-trials distribution

In [**2**], Patrick Billingsley treated the general case of the meta-multinomial distribution $F(x)$, for multi-valued independent Bernoulli trials. Each trial now has r possible outcomes, $k = 0, 1, 2, \ldots, r-1$, with corresponding probabilities p_k such that $\sum p_k = 1$. He showed that, as for $r = 2$, this function is continuous and singular, with one exception: when each trial has uniformly distributed outcomes ($p_k = \frac{1}{r}$), the stream of trials is uniformly distributed ($F(x) = x$).

Proceeding as for the binary case ($r = 2$), we restrict X to $J \geq 1$ digits. Then for $k = 0, ..., r^J - 1$, X takes the values $.k$ if we express k in base r. For example, for $r = 3, J = 2$, X takes the values $.00, .01, .02, .10, .11, .12, .20, .21, .22$. So X takes the value $.k$ with probability

$$\begin{aligned} \pi_{J,k} &= \text{Prob}(X = .k = .d_1 d_2 \cdots d_J) \\ &= p_{d_1} p_{d_2} \cdots p_{d_J} \end{aligned}$$

The random variable X (with J digits), representing a finite Bernoulli stream of J r-valued outcomes or symbols, has the cumulative distribution function $\widetilde{F}_J(x)$ of scale J given by

$$\widetilde{F}_J(x) = \sum_{.k \leq x} \pi_{J,k}. \tag{45}$$

These distribution functions for multi-valued Bernoulli trials again become easier to use if we translate them to the right by one unit of length at scale J. We define the unit-translates F_J as

$$F_J(x) = \widetilde{F}_J\left(x - r^{-J}\right). \tag{46}$$

Example graphs of $\widetilde{F}_J$ and F_J are shown in Figures 7 and 8 for three-valued trials ($r = 3$). For $r = 3$ and $p_0 = p_2 = \frac{1}{2}$, $p_1 = 0$, $F(x)$ is the Cantor function [**2**] [**6**], shown in Figures 9

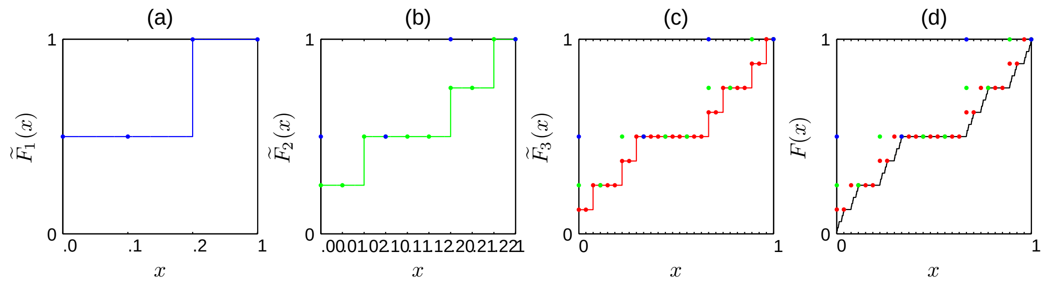

Figure 9: (a,b,c) The distribution functions $\widetilde{F}_J$ for $(p_1, p_2, p_3) = (\frac{1}{2}, 0, \frac{1}{2})$, $J = 1, 2, 3$; and (d) their limit F as $J \to \infty$, the Cantor function.

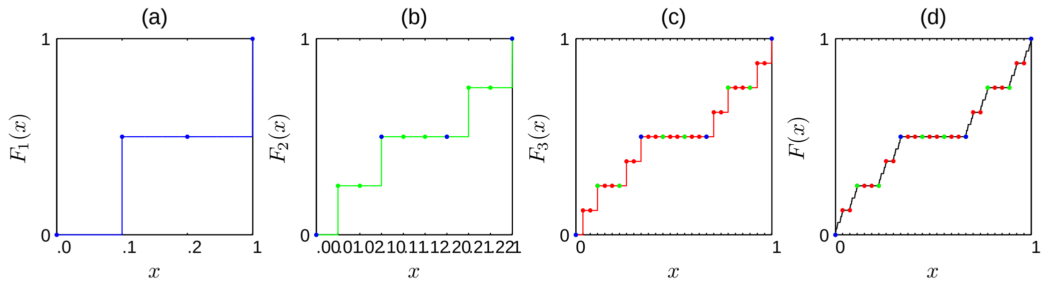

Figure 10: (a,b,c) The unit-translated distribution functions F_J for $(p_1, p_2, p_3) = (\frac{1}{2}, 0, \frac{1}{2})$, $J = 1, 2, 3$; and (d) their limit F as $J \to \infty$. This is essentially the usual way of constructing the the Cantor function.

and 10. The Cantor function is usually constructed as the limit of a sequence of functions equal to the unit-translated distribution functions F_J on the middle thirds of every scale-J unit of length r^{-J}. Over each middle third unit at scale J, $F_{J+n}, n > 0$ remains a constant step of slope 0 that is never subdivided again. But the definition of F_J on the two outside thirds of each unit is almost arbitrary [**6**], since at each change to the next finer scale, these thirds are subdivided to insert a step at their middle thirds; and the union of middle thirds (for all J) is readily seen to have measure 1.

F_J is still a refinement of F_{J-1}. The graph of F_J inherits all jump points of F_{J-1}, and between them interpolates r new ones:

Lemma 10. For every scale-$(J-1)$ r-adic rational $x \in [0, 1]$, $F_J(x) = F_{J-1}(x)$.

Proof. Using the definition of F_J and translation by a unit $u = r^{-J}$ at scale J, we have the two partial sums $F_{J-1}(x) = \sum_{.k \le x - ru} \pi_{J-1,k} = \sum_{.k+ru \le x} \pi_{J-1,k}$ and $F_J(x) = \sum_{.k' \le x - u} \pi_{J,k'} = \sum_{.k'+u \le x} \pi_{J,k'}$. In the sum for F_{J-1}, at scale $J-1$, each r-adic rational atom $.k = .d_1 d_2 \ldots d_{J-1}$, with probability mass $\pi_{J-1,k}$, is translated to $.k + ru$, and included in the sum if less than or equal to x. This atom splits into r atoms at scale J, $.k + du = .d_1 d_2 \ldots d_{J-1} d$, for $d = 0, \ldots, r-1$, with probability masses $\pi_{J-1,k} \cdot p_d$, respectively, which sum to a total mass $\pi_{J-1,k}$ for all of these atoms. These atoms are translated by u in the sum for F_J to positions $.k + (d+1)u$; note that the final scale-J atom, at $.k + ru$, lies exactly where the scale-$(J-1)$ atom is, and the preceding scale-J atoms lie to their left. Therefore, each mass in the sum for F_{J-1} (from $.k = 0$ through $.k = x$) is exactly replaced by the corresponding r masses in the sum for F_J, and the two sums are equal. □

In general, F maps base-r numbers into $(p_0, \ldots, p_{r-1})$-base numbers in a way similar to (11):

Theorem 11. *Given* $x = .d_1d_2d_3 \ldots d_J$,

$$F_J(x) \;=\; \sum_{j=1}^{J} p_{d_1}p_{d_2}\cdots p_{d_{j-1}} \cdot P_{d_j}, \tag{47}$$

where for any digit d *we define* $P_d = 0$ *if* $d = 0$, and $P_d = (p_1 + \cdots + p_{d-1})$ *if* $d > 0$.

Proof. By induction: the equality (47) for $J = 1$ is easily verified for $j = 1$. Assume the equality holds for $J - 1 > 0$, and let $x = .k = .d_1d_2d_3 \ldots d_{J-1}d_J = .k'd_J = x' + d_J u$, where $u = r^{-J}$ is again the scale-J unit, and $x' = .k'$ is a scale-$(J-1)$ r-adic rational number formed by the first $J - 1$ digits of x. If $d_J = 0$, $x = x'$ and the left side of (47) is $F_J(x) = F_J(x') = F_{J-1}(x')$ by Lemma 10; the right side clearly reduces to the scale-$(J-1)$ expression for x' as well. If $d_J > 0$, this digit increases each side above its scale-$(J-1)$ expression by just a sum of d_J terms: since $F_J = \tilde{F}_J(x - u)$, $F_J(x) = F_J(x') + \sum_{d=0}^{d_J-1} \pi_{J,k'd} = F_{J-1}(x') + \sum_{d=0}^{d_J-1} \pi_{J,k'd}$, again by Lemma 10. This increase equals $\pi_{J-1,k'} \cdot \sum_{d=0}^{d_J-1} p_d = p_{d_1}p_{d_2}\cdots p_{d_{J-1}} \cdot P_{d_J}$, with P_{d_J} as defined. But evidently the right side is also the sum for scale $J - 1$ increased by this same amount. In any case, then, equality of both sides of (47) holds by the induction hypothesis. □

Taking the limit of F_J as $J \to \infty$, we have

Theorem 12. *Given* $x = .d_1d_2d_3\cdots$,

$$F(x) \;=\; \sum_{j=1}^{\infty} p_{d_1}p_{d_2}\cdots p_{d_{j-1}} \cdot P_{d_j}. \tag{48}$$

For $r = 2$, we recognize $P_0 = 0$ and $P_1 = p = q \cdot \frac{p}{q}$ in expression (11). For uniform $p_k = \frac{1}{r}$, we have $\frac{1}{p_d}P_d = d$, so that the jth term in (48) equals $r^{-j}d_j$ as in the usual base-r expansion of $x \in [0, 1]$, and $F(x) = x$. Note that (48) is expressible as a nested (Horner) expansion (cf. the index reversal mapping in [**7**], Section 2.2.3):

$$F(x) \;=\; P_{d_1} + p_{d_1}\left(P_{d_2} + p_{d_2}\left(P_{d_3} + p_{d_3}\left(\cdots\right)\right)\right). \tag{49}$$

For uniform $p_k = \frac{1}{r}$, this becomes the usual Horner expansion to evaluate a polynomial of one variable r^{-1}:

$$F(x) \;=\; d_0 + r^{-1}\left(d_1 + r^{-1}\left(d_2 + r^{-1}\left(\cdots\right)\right)\right), \tag{50}$$

where we include the constant term $d_0 = 0$.

5.0 Conclusion and Questions

The Haar wavelet was used to analyze the Bernoulli-trials distribution function, which is known to be continuous, increasing, and singular for $p \neq \frac{1}{2}$ [**2**]. This analysis revealed that the function just maps usual binary numbers into p-binary numbers. (In a different application to probability, integrals of Haar wavelets were used to synthesize continuous Brownian motions, which are singular functions of time [**5**] [**2**].) Since it happens to obey a two-scale dilation equation, the Bernoulli trials distribution function could be used to construct a new family of wavelets. The Bernoulli family includes the Haar wavelet ($p = 0, 1$) and the piecewise-linear wavelet ($p = \frac{1}{2}$) as special cases. (These are also the B-spline wavelets of orders 0 and 1, respectively [**1**].) For intermediate values of p, we have wavelets that offer more nearly compact support as $p \to 0, 1$, or smoother approximation (with wider periodic spectral gaps) as $p \to \frac{1}{2}$.

We also found that the multi-valued Bernoulli distribution function maps usual base-r numbers into base-$(p_1, \ldots, p_r)$ numbers. The wavelet family given in Section 3 appears to belong to a much larger one. Dr. Billingsley also gave a base-r two-scale equation that is equivalent to the two-scale dilation equation

$$F(x) = \sum_{k=0}^{r-1} p_k F(rx - k).$$

From this equation, it should be possible to construct a wavelet family having r filter coefficients, along a path similar to that for $r = 2$. For radix-r, the basis consists of dilations and translations of a scaling function φ and $(r-1)$ mutually orthogonal wavelets $\psi_1, \ldots, \psi_{r-1}$ [**8**]. This family for $r = 3$ would include a Cantor wavelet basis, given by a scaling function and two wavelets.

That p-binary numbers are found in a natural setting and develop Bernoulli wavelets stirs us to ask questions about arithmetic. First, is it possible to multiply or add p-binary numbers, where each bit has a unique signature? Is there any computational benefit? Much more broadly, can a wavelet (Bernoulli or otherwise) be used to represent real numbers and to do arithmetic on the real line, possibly using the fast wavelet transform? The goal would be to perform bit-level arithmetic operations with uniform precision regardless of scale. This could serve to replace floating-point arithmetic, which loses precision as the scale (exponent) increases.

Acknowledgements

I would like to thank Dr. Ryszard Syski for introducing me to the Bernoulli-trials distribution function,and Dr. David Walnut for teaching me everything I needed to know about wavelets.

References

[1] David Walnut, *Introduction to Wavelet Analysis*, Birkhäuser, 2002.

[2] Patrick Billingsley, *Probability and Measure*, 3rd ed., John Wiley and Sons, 1995.

[3] Malcolm Adams and Victor Guillemin, *Measure Theory and Probability*, Birkhäuser, 1996.

[4] Tom M. Apostol, *Calculus*, vol.I, Blaisdell Publishing, 1967.

[5] Samuel Karlin and Howard M. Taylor, *A First Course in Stochastic Processes*, 2nd ed., Academic Press, 1975.

[6] H.S. Bear, *A Primer of Lebesgue Integration*, Academic Press, 2001.

[7] Charles Van Loan, *Computational Frameworks for the Fast Fourier Transform*, SIAM Press, 1992.

[8] Martin Vetterli and Jelena Kovačević, *Wavelets and Subband Coding*, Prentice-Hall, 1995.

School of Computational Sciences, George Mason University, Fairfax, Virginia