Beyond the law of large numbers

Uwe Saint-Mont

Nordhausen University of Applied Sciences

Hochschule Nordhausen

Abstract

In probability theory and statistics, the IID model represents a single population, and a large, potentially infinite sample from this population. Main theorems assure convergence, making asymptotic statistics possible. In particular the law of large number (LLN) states that the sample mean approaches the population mean, i.e., $\bar{x}_n \to \mu$.

To *avoid* convergence, it is thus straightforward to consider two populations H_0, H_1 with expected values μ_0 and μ_1, respectively, and a sample that continually fluctuates between them. Quite obviously, with two points of attraction, $\bar{x}_n$ cannot converge. However, given a new *exponential* kind of sampling, the mean quite straightforwardly produces

1. a geometric version of Pascal's triangle
2. a new class of distributions
3. a corresponding classic fractal

Although these results may seem quite unexpected, they are straightforward and can be obtained with a minimum of technical effort (mainly sums and series).

Exponential Sampling

Consider two populations (distributions) H_0 and H_1. If one switches between the populations at a constant rate, i.e., if j observations from H_0 are followed by j observations from H_1, and so forth, the arithmetic mean of this sequence will converge.

However, if 2^j observations from H_0 are followed by 2^{j+1} observations from H_1, etc., one then obtains the desired effect. (On a logarithmic scale, taking ld $= \log_2$, the distance $(j+1)-j=1$ is a constant. Thus, there, one switches at a constant rate.)

From $H_0 : x_1$;
From $H_1 : x_2, x_3$;
From $H_0 : x_4, x_5, x_6, x_7$;
From $H_1 : x_8, x_9, x_{10}, \ldots, x_{15}$;
...

Instead of switching deterministically, a particularly elegant way to alternate between H_0 and H_1 is to take the next batch of 2^j observations ($j = 0, 1, \ldots$) from H_0 with probability $1-p$, and from H_1 with probability p. Set $f = p/(1-p)$.

Weaving

Example: Let $n = 3$. Thus the sample size is $2^0 + 2^1 + 2^2 = 2^3 - 1 = 7$. The binary numbers $(000)_2$ to $(111)_2$ encode the result of sampling: For example, $(101)_2$ says that $1 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0 = 5$ of these observations come from H_1. The fourth choice would add $2^{4-1} = 8$ observations, and a prefix to $(101)_2$ which is 0 for H_0 and 1 for H_1.
Altogether, one gets threads that interweave:

0 **1**
00 01 *10* **11**
000 001 *010* **011** 100 101 *110* **111**
...

However, they do *not* merge: Weaving (left) vs. Binomial structure (right)

```
          .                         .
        ↙   ↘                    ↙       ↘
      0        1              0              1
    ↙    ↙↘       ↘         ↙   ↘         ↙   ↘
  00   01   10    11      00     (01, 10)      11
          ...                      ...
```

Geometric Triangle

$j = 0:$ 1
$j = 1:$ 1 | f
$j = 2:$ 1 | f || f | f^2
$j = 3:$ 1 | f || f | f^2 ||| f | f^2 || f^2 | f^3
...

Global interpretation (self-similarity): $\mathbf{f}_0 = 1$, and $\mathbf{f}_{j+1} = (\mathbf{f}_j, f \cdot \mathbf{f}_j)$
Lokal interpretation (cascade): Every entry a in row j becomes $(a, f \cdot a)$ in row $j+1$.
Applying the logarithm base f to every a yields the exponents, i.e., the numbers:

j		Sum s_j
0	0	0
1	0 | 1	1
2	0 | 1 || 1 | 2	4
3	0 | 1 || 1 | 2 ||| 1 | 2 || 2 | 3	12
	...	

In general, $s_0 = 0$, and $s_{j+1} = 2s_j + 2^j$ for $j = 0, 1, \ldots$

The geometric triangle is the combinatoric core of the Weaver's distribution:
$\mathbf{p}_0 = 1$, and $\mathbf{p}_n = p_0(\mathbf{f}_{n-1}, f\mathbf{f}_{n-1})$, where $p_0 = (1-p)^n$.

The Weaver's Distribution $W(n, p)$

Make n independent choices: $\mathbf{B}_n = (B_1, \ldots, B_n)$ where $B_j = 0$ with prob. $1-p$ and $B_j = 1$ with prob. p ($j = 1, \ldots, n$). Let $\mu_0 = 0$, $\mu_1 = 1$ and

$$Y_n = E(\bar{X}_n | \mathbf{B_n}) = E\left(\frac{\sum_{i=1}^{2^n-1} X_i}{2^n - 1} | \mathbf{B_n}\right)$$

Then:

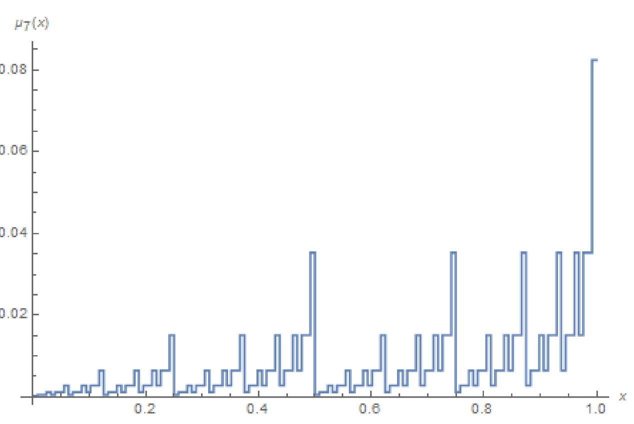
1. Y_n assumes the values $y_k = y_{k,n} = k/(2^n - 1)$ for $k = 0, 1, \ldots, 2^n - 1$.
2. Suppose $\mathbf{B}_n = \mathbf{b}_n = (b_{n-1}, \ldots, b_1, b_0)$. Note that b_{i-1} can be interpreted as the ith digit in the binary representation of a natural number $k \in \{0, \ldots, 2^n - 1\}$, i.e., $k = \sum_{i=0}^{n-1} b_i 2^i$. Then the probability p_k at the point y_k is given by
$$p_k = p^{\sum_{i=0}^{n-1} b_i}(1-p)^{n - \sum_{i=0}^{n-1} b_i} = p^{\#1}(1-p)^{\#0} \geq 0,$$
where #1 and #0 denote the number of ones and zeros in $\mathbf{b}_n$, respectively.
3. The probabilities corresponding to row n can be constructed by the following simple scheme which has the same structure as the geometric triangle.

1
$1-p$ | p
$(1-p)^2$ | $p(1-p)$ || $p(1-p)$ | p^2
$(1-p)^3$ | $p(1-p)^2$ || $p(1-p)^2$ | $p^2(1-p)$ ||| $p(1-p)^2$ | $p^2(1-p)$ || $p^2(1-p)$ | p^3
...

Properties of $W(n, p)$

Let $Y_n \sim W(n, p)$. Then

1. $EY_n = p$
2. $\sigma^2(Y_n) = \frac{\sum_{i=0}^{n-1} 2^{2i}}{(2^n-1)^2} \, p\,(1-p)$
3. $W(n, p)$ is a discrete version of Mandelbrot's "binomial measure:"
Start with the uniform distribution on the unit interval. Next, the proportion $1-p$ is uniformly distributed on the interval $(0, 1/2)$, and the proportion p is uniformly distributed on $(1/2, 1)$. Then, split the masses further (locally) according to the geometric triangle, i.e., mass $(1-p)^2$ to the interval $(0, 1/4)$, mass $(1-p)p$ to $(1/4, 1/2)$, mass $p(1-p)$ to $(1/2, 3/4)$, and mass p^2 to the interval $(3/4, 1)$, etc.



Ideas of proof.
1. and 2. Sophisticated bookkeeping upon noticing that the random variable $\#\mathbf{1}$ has a binomial distribution.
3. Every value $y_{k,n}$ may be located in an interval of length $1/2^n$ with density $g_{k,n}(x) = 2^n p^{\#1}(1-p)^{\#0}$ there.

Note: The "roughness" of the density (measured by f^j) grows at the same rate as the number of intervals. Thus $\ln(f^n)/\ln(2^n) = \ln(f)/\ln(2)$ is a constant, called the fractal dimension.

The limit distribution $W(p)$

Suppose $F_n(x)$ is the distribution function of Y_n. Then $\lim_{n\to\infty} F_n(x) = F(x)$ exists. $F(x)$ may be interpreted as the distribution function of a random variable $Y \sim W(p)$.
Properties:

1. $EY = p$
2. $\sigma^2(Y) = p\,(1-p)/3$
3. $W(p)$ is equivalent to the uniform distribution if $p = 1/2$,
4. $W(p)$ is equivalent to Mandelbrot's asymptotic "binomial measure" if $p \neq 1/2$ which has no density.

Idea of proof. Pass to the limit.

The process $\bar{X}_n$

Suppose, in addition to $\mu_0 = 0$ and $\mu_1 = 1$, that the populations H_0, H_1 have finite variances σ_0^2 and σ_1^2, respectively. Then

1. $E\bar{X}_n = p$
2. $\sigma^2(\bar{X}_n) = p\,(1-p) + \frac{\sigma_0^2 + \sigma_1^2}{2^n - 1}$
3. The limit distribution is Bernoulli $B(p)$.

Ideas of Proof.

1. Total mean of a mixture of 2^n distributions.
2. Decomposition of variance (between and within the populations) and advanced bookkeeping.
3. Asymptotically, all mass is located in the unit interval. The limit distribution is centered in μ and its variance is $p(1-p)$, i.e., only the variance between the populations is relevant. This implies the result.

Further research

1. General switching schemes (e.g., Markov)
2. "Augmented" distributions, e.g., the Geometric. That is, $P(T = 2^{i-1})$ with probability $(1-p)^{i-1}p$, where $i = 1, 2, \ldots$ and $0 < p < 1$. Thus $ET = p\sum_{i=0}^{\infty} 2^i (1-p)^i = p\sum_{i=0}^{\infty}(2-2p)^i$ which converges if $p > 1/2$, and ET^2 converges if $p > 3/4$.
3. General cascades (e.g., Cheng 2014)
4. If switching occurs very often (e.g., n observations from H_0, n observations from H_1, etc.), the populations merge into one, and we are back to the classical theory ($\bar{X}_n$ converging to a fixed number). If switching occurs seldom, in particular, if sampling is exponential, both populations remain distinct asymptotically, and $\bar{X}_n$ converges in distribution. It would be interesting to know more about the "line" separating these two situations. What are necessary and sufficient conditions for either kind of convergence of $\bar{X}_n$?
5. Several populations (connection with Benford's law)

References

[1] Cheng, Q. (2014). Generalized binomial multiplicative cascade processes and asymmetrical multifractal distributions. Nonlin. Processes Geophys. **21**, 477-487.

[2] Mandelbrot, B. (1974). Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier. Journal of Fluid Mechanics **62**, 331-358.

[3] Saint-Mont, U. (2017). Beyond the law of large numbers: Introducing progressive sampling, weaving, the geometric triangle, and corresponding distributions. arxiv.org/abs/1709.10281