An Almost-Sure Functional Central Limit Theorem

1. Introduction

If $\{z_t\}$ is a sequence of IID random variables with $\mathrm{E}(z_t) = 0$ and $\mathrm{E}(z_t^2) = 1$, then the central limit theorem tells us that the sequence with typical element

$$Z_n \equiv n^{-1/2} \sum_{t=1}^{n} z_t \tag{1}$$

is asymptotically normal. Specifically, the sequence $\{Z_n\}$ tends in distribution to the standard normal distribution:

$$Z_n \xrightarrow[D]{} \mathrm{N}(0,1).$$

We may define a stochastic process on the $[0,1]$ interval by means of the sequence $\{z_t\}$ as follows:

$$W^n(t) = n^{-1/2} \sum_{t=1}^{\lfloor nt \rfloor} z_t, \quad t \in [0,1]. \tag{2}$$

Then the functional central limit theorem tells us that, as $n \to \infty$,

$$W^n(t) \xrightarrow[D]{} W(t),$$

where $W(t)$ is a standard Wiener process, or Brownian motion, on $[0,1]$.

A difficulty with this construction is that the limit is *only* in distribution, and not in probability, still less almost sure. To see this, suppose that $Z_n \to Z$ in probability, where Z is some random variable. Then we show that Z and the summands z_t are independent. Note that, for the given t,

$$Z_{tn} \equiv n^{-1/2} \sum_{s=t+1}^{t+n} z_s$$

also tends to N(0,1) in distribution as $n \to \infty$. Under the assumption that $Z_n \to Z$ in probability, it follows also that $Z_{tn} \to Z$ in probability. Consider the joint characteristic function of z_t and Z. It is, for arbitrary real arguments s and r,

$$\begin{aligned}
\mathrm{E}\exp(isZ + irz_t) &= \lim_{n\to\infty} \mathrm{E}\exp(isZ_{tn} + irz_t) \\
&= \lim_{n\to\infty} \mathrm{E}\exp(isZ_{tn})\mathrm{E}\exp(irz_t) \qquad (Z_{tn} \text{ is independent of } z_t) \\
&= \mathrm{E}\exp(isZ)\mathrm{E}\exp(irz_t).
\end{aligned}$$

The factorisation of the joint characteristic function demonstrates the independence of Z and z_t, for any t. Intuitively, the weight of z_t in the partial sums Z_n gets smaller as $n \to \infty$, and in the limit we have independence. A straightforward extension of this proof shows that Z_n is independent of Z for any n.

Now $Z \sim \mathrm{N}(0,1)$, and so $-Z \sim \mathrm{N}(0,1)$ as well. By independence, therefore, the joint distribution of Z_n and Z is the same as that of Z_n and $-Z$. Consequently,

$$\Pr(|Z_n - Z| > \varepsilon) = \Pr(|Z_n + Z| > \varepsilon)$$

for all n and all $\varepsilon > 0$. But this means that $Z_n \to -Z$ in probability, which is incompatible with $Z_n \to Z$ unless $Z = 0$. But that too is contradicted by the fact that $Z \sim \mathrm{N}(0,1)$, and so we conclude that the probability limit Z cannot exist.

2. A Different Construction

Despite the above result, it is possible to construct a sequence $\{Z_n\}$ of variables, where each Z_n has the same distribution as the partial sum (1) with IID summands, and $Z_n \to Z$ almost surely, with $Z \sim \mathrm{N}(0,1)$. The key is to fill in the terms of the partial sum from the middle, rather than continually appending new, independent, terms at the end.

We begin with the simplest case, in which the summands z_t are NID(0,1) themselves. As we will see, the construction provides as a by-product a means for simulating a Wiener process in continuous time directly. The starting point is to generate the realisation of $W(1)$, of which the marginal distribution is just N(0,1). At step i of the construction, we have a sequence z_{ti}, $t = 0, 1, \ldots, 2^i$, of NID(0,1) variables that define the stochastic process $W^i(t)$ through the formula (2) with $n = 2^i$. The process $W^i(t)$ is such that, for all $j < i$,

$$W^j(2^{-j}k) = W^i(2^{-j}k) \text{ for } k = 0, 1, \ldots 2^j.$$

This means that the values of the processes W^i at the dyadic points $2^{-j}k$, $k = 0, 1, \ldots, 2^j$, are the same for all i with $i \geq j$.

We can simplify notation by omitting obvious powers of 2. Thus, instead of $W^i(2^{-i}k)$, we may write simply $W^i(k)$ for $k = 0, 1, \ldots, 2^i$. To go from step i to step $i+1$, we must establish the values of W^{i+1} at the points $2^{-(i+1)}(2k+1)$, $k = 0, 1, \ldots, 2^i - 1$. These are the points midway between the points at which values are permanently established at step i, and, with them, constitute the set of points at which values are permanently established at step $i+1$. For the construction to be correct, it must be the case that $W^i(k) \sim \mathrm{N}(0, 2^{-i}k)$. In addition, we require that the increments $W^i(k+1) - W^i(k) \sim \mathrm{N}(0, 2^{-i})$, and that they are independent across k. Suppose that we have achieved these requirements at step i; we show how to maintain them at step $i+1$. Note that these requirements are enough for our claim that the distribution of

$$Z_{2^i} \equiv 2^{-i/2} \sum_{k=0}^{2^i - 1} 2^{i/2} \big(W^i(k+1) - W^i(k)\big)$$

is that of a sum of 2^i NID(0,1) variables, divided by $2^{i/2}$.

What then is the distribution of $W^{i+1}(2k+1)$ conditional on the $W^i(k)$, $k = 0, 1, \ldots, 2^i$? Since $W^i(0) = 0$ by construction, the conditioning is equivalent to conditioning on the increments $W^i(k+1) - W^i(k)$, $k = 0, 1, \ldots, 2^i - 1$. The conditional distribution is established if we can find that of the increment $W^{i+1}(2k+1) - W^{i+1}(2k)$, which is just $W^{i+1}(2k+1) - W^i(k)$, since the value at $2^{-i}k$ is permanently established at step i. Since we want both this increment and the next one, namely $W^i(k+1) - W^{i+1}(2k+1)$, to be independent of all other summands at step $i+1$, we see that it is enough to condition on $W^i(k+1) - W^i(k)$, which is independent of all the other step i increments.

Let

$$X_1 = 2^{i/2}\big(W^{i+1}(2k+1) - W^i(k)\big) \text{ and}$$
$$X_2 = 2^{i/2}\big(W^i(k+1) - W^{i+1}(2k+1)\big).$$

Then we require that X_1 and X_2 should be independent, each with marginal distribution $\mathrm{N}(0, 1/2)$, and such that $X_1 + X_2 = 2^{i/2}\big(W^i(k+1) - W^i(k)\big)$, of which the marginal distribution is N(0,1). The joint distribution of X_1 and $X \equiv X_1 + X_2$ is normal, with covariance matrix

$$\begin{bmatrix} 2^{-1} & 2^{-1} \\ 2^{-1} & 1 \end{bmatrix}.$$

The correlation is therefore $1/\sqrt{2}$. It follows that the expectation of X_1 conditional on X is $X/2$, and the conditional variance is $1/4$. Thus we may set $X_1 = \frac{1}{2}(X + U)$, where $U \sim \mathrm{N}(0, 1)$, independent of any random number used up to step i. We can check that, with this definition, $\mathrm{Var}(X_1) = 1/2$, as required. In addition $X_2 = X - X_1 = (X - U)/2$ has a variance of $1/2$. Further, the covariance of X_1 and X_2 is

$$\mathrm{cov}(X_1, X_2) = \tfrac{1}{4}\mathrm{E}\big((X+U)(X-U)\big) = \tfrac{1}{4}(1-1) = 0,$$

so that X_1 and X_2 are independent. In terms of the W^i, we have, for $k = 0, 1, \ldots, 2^i - 1$,

$$W^{i+1}(2k+1) = \tfrac{1}{2}\big(W^i(k) + W^i(k+1)\big) + 2^{-(i+2)/2}U_{i+1,k},$$

where we have indexed the innovation U so as to make it clear when it is used in the whole procedure.

We may denote by $\mathcal{F}_i$ the sigma-algebra generated by all the innovations used up to and including step i. Then $\mathcal{F}_0$ is generated by $U_{0,0} \sim \mathrm{N}(0, 1)$, and the process W^i is $\mathcal{F}_i$–measurable.

For any number $t \in [0, 1]$ with a finite dyadic expansion

$$t = \sum_{j=1}^{n} b_j 2^{-j}, \quad b_j = 0 \text{ or } 1,$$

for some finite n, it is clear that the sequence $\{W^i(t)\}$ converges to $W^n(t)$ as $i \to \infty$, since $W^i(t) = W^n(t)$ for all $i \geq n$. It remains to find the best argument to show that $\{W^i(t)\}$ converges almost surely for all real $t \in [0, 1]$.

3. Non-Gaussian Innovations

If the z_t in (1) are not Gaussian, then the distribution of Z_n depends in general on n. Our goal is still to construct a sequence $\{Z_n\}$ such that, for each n, the distribution of Z_n is that of a sum of n IID random variables, each of expectation 0 and variance 1, drawn from whatever non-Gaussian distribution we wish, subject only to the requirement that these sums obey the central limit theorem, and so tend in distribution to N(0,1). However, unlike (1), we wish our sequence to converge almost surely to some well-defined N(0,1) variable.

The innovations $U_{i,k}$ are now IID drawings from the uniform U(0,1) distribution. Let F be the CDF of the desired distribution for the summands. Then let F_i be the distribution of a sum of i IID variables each of which has distribution F. Clearly the F_i can be defined recursively by convolutions:

$$F_1(x) = F(x), \quad F_{i+1}(x) = \int F_i(x-y)\,dF(y).$$

We also need the distribution of a sum of 2^i IID summands conditional on the value of this sum plus another such sum, independent of the first. We denote this conditional CDF by $F_{2^i|2^{i+1}}$. To avoid undue complexity, we assume that all the distributions we consider, the F_i and the $F_{2^i|2^{i+1}}$, are strictly increasing functions of their argument, and thus have inverses, G_i and $G_{2^i|2^{i+1}}$, say.

We begin by generating $W^0(1)$ as $G_1(U_{0,0})$. Thus $W^0(1)$ follows the distribution with CDF F. At step 1, we wish $W^1(2)$ to have a different distribution from $W^0(1)$, namely, the F_2 distribution, divided by $\sqrt{2}$. Thus we set $W^1(2) = G_2(U_{0,0})/\sqrt{2}$. Next, we wish to generate $W^1(1)$, which should follow the F_1 distribution, divided by $\sqrt{2}$. We therefore draw a variable from the conditional distribution $F_{1|2}$, with the value of the conditioning variable given by $\sqrt{2}W^1(2)$, and then have

$$W^1(1) = 2^{-1/2} G_{1|2}\big(U_{1,0} \,|\, \sqrt{2}W^1(2)\big),$$

where $U_{1,0}$ is a $\mathcal{F}_1$–measurable U(0,1) variable.

Here, note that, if U_1 and U_2 are independent U(0,1) variables, and if X and Y are random variables such that F_X is the marginal CDF of X and $F_{Y|X}(Y \,|\, X)$ is the CDF of Y conditional on X, then the couple

$$F_X^{-1}(U_1) \text{ and } F_{Y|X}^{-1}\big(U_2 \,|\, F_X^{-1}(U_1)\big)$$

follows the joint distribution of X and Y. (Proof in Appendix.)

Now consider step i. We begin by computing quantities $V^i(k)$, which have the distributions of sums of IID variables from F_1. Subsequently, we obtain all the $W^i(k)$ by dividing the $V^i(k)$ by $2^{i/2}$. We begin by evaluating $V^i(2^i)$ as $G_{2^i}(U_{0,0})$. This means that $V^i(2^i)$ is distributed like the sum of 2^i IID variables from F_1. Then we compute $V^i(2^{i-1})$, which is the value at the midpoint of the interval. It is generated as

$$V^i(2^{i-1}) = G_{2^{i-1}|2^i}\big(U_{1,0}\,|\,V^i(2^i)\big),$$

so that it has the distribution of a sum of 2^{i-1} summands from F_1, conditional on being the first 2^{i-1} out of the 2^i summands that add up to $V^i(2^i)$. Then we interpolate at the 1/4 and 3/4 points. We have

$$V^i(2^{i-2}) = G_{2^{i-2}|2^{i-1}}\big(U_{2,0}\,|\,V^i(2^{i-1})\big)$$

and

$$V^i(3.2^{i-2}) = V^i(2^{i-1}) + G_{2^{i-2}|2^{i-1}}\big(U_{2,1}\,|\,V^i(2^i) - V^i(2^{i-1})\big).$$

For the last of these, we reason again in terms of increments. The conditioning value, $V^i(2^i) - V^i(2^{i-1})$, is the sum of the last 2^{i-1} summands, and we add to the value $V^i(2^{i-1})$ an increment that is half of these.

The approach above can now be generalised with no special difficulty. The main difference relative to the Gaussian case is that, at each step i, everything must be reevaluated, although new random numbers are used only for the newly interpolated points. The first two steps are described in the preceding paragraph, where we obtain values of the process at step i at the points $0, 1/4, 1/2, 3/4$, and 1. For the endpoint, we use the random number $U_{0,0}$; for the midpoint $U_{1,0}$, and for the quarter and three-quarter points $U_{2,0}$ and $U_{2,1}$. There are 2^{k-1} points in the $[0,1]$ interval of the form $(2j+1)/2^k$, and these points use the 2^{k-1} random numbers $U_{k,j}$, $j = 0, 1, \ldots, 2^{k-1} - 1$, that belong to $\mathcal{F}_k$. At step $i \geq k$, these random numbers are used to generate the quantities $V^i(2^{i-k}(2j+1))$, which are therefore all $\mathcal{F}_k$–measurable. On division by $2^{i/2}$, we get the $W^i(2^{i-k}(2j+1))$, which are the values of the step-i process at the points $2^{-k}(2j+1)$ of the $[0,1]$ interval.

At stage k of step i, we generate an increment, conditional on a larger increment that is $\mathcal{F}_{k-1}$–measurable. The appropriate formula is

$$V^i(2^{i-k}(2j+1)) = V^i(2^{i-k+1}j) + G_{2^{i-k}|2^{i-k+1}}\big(U_{k,j}\,|\,V^i(2^{i-k+1}(j+1)) - V^i(2^{i-k+1}j)\big).$$

This formula gives us the values at step i of those V that are $\mathcal{F}_k$–measurable but not $\mathcal{F}_{k-1}$–measurable. At this point, everything that is $\mathcal{F}_{k-1}$–measurable for step i has been generated, which means that the values $V^i(2^{i-k+1}j)$ for *all* $j = 0, 1, \ldots, 2^{k-1} - 1$ are available, that is, the values that define the step-i process at the dyadic points $j2^{-k+1}$. We may check that the increment $V^i(2^{i-k}(2j+1)) - V^i(2^{i-k+1}j)$ is the sum of $2^{i-k}(2j+1) - 2^{i-k+1}j$, or 2^{i-k}, summands, and that the conditioning increment, $V^i(2^{i-k+1}(j+1)) - V^i(2^{i-k+1}j)$, is the sum of 2^{i-k+1} summands, as indicated by the inverse CDF $G_{2^{i-k}|2^{i-k+1}}$.

4. An Example

It is not usually possible to obtain analytic expressions for the CDFs F_i and $F_{2^i|2^{i+1}}$ that we need for a numerical implementation of the construction of the last section. However, it can be done with little trouble if we choose for the base distribution F a chi-squared distribution, suitably centred and standardised, since sums of chi-squared variables are also chi-squared, with more degrees of freedom.

In practice, it is easiest just the generate sums of chi-squared variables, and centre and standardise them at the end. Suppose that we use the distribution with one degree of freedom for the base distribution. Then a sum of i such IID variables has the χ^2 distribution with i degrees of freedom. We also need the distribution of the sum of n IID χ^2_1 variables conditional on their being the first n out of a total of $2n$ variables for which we give the value of the sum. The density of χ^2_n is

$$f_n(x) = \frac{x^{n/2-1}e^{-x/2}}{2^{n/2}\Gamma(n/2)},$$

where $\Gamma(\cdot)$ is the gamma function. If X and $Y-X$ are two independent χ^2_n variables, then their joint density is

$$f_n(x)f_n(y-x) = \frac{x^{n/2-1}(y-x)^{n/2-1}e^{-y/2}}{2^n(\Gamma(n/2))^2}. \tag{3}$$

The density of X conditional on Y is the density of the distribution $F_{n|2n}$ that we seek. It is the joint density (3) divided by the marginal density of Y, which is the χ^2_{2n} density. We have

$$\begin{aligned}\frac{f_n(x)f_n(y-x)}{f_{2n}(y)} &= \frac{\Gamma(n)}{(\Gamma(n/2))^2}\,\frac{x^{n/2-1}(y-x)^{n/2-1}}{y^{n-1}} \\ &= \frac{1}{y\mathrm{B}(n/2,n/2)}\left(\frac{x}{y}\right)^{n/2-1}\left(1-\frac{x}{y}\right)^{n/2-1} \qquad (4),\end{aligned}$$

where $\mathrm{B}(\cdot,\cdot)$ is the beta function, defined by the relation

$$\mathrm{B}(x,y) = \frac{\Gamma(x+y)}{\Gamma(x)\Gamma(y)}.$$

The CDF associated with the conditional density (4) is

$$\begin{aligned}F_{n|2n}(x\,|\,y) &= \frac{1}{y\mathrm{B}(n/2,n/2)}\int_0^x \left(\frac{z}{y}\right)^{n/2-1}\left(1-\frac{z}{y}\right)^{n/2-1} dz \\ &= \frac{1}{\mathrm{B}(n/2,n/2)}\int_0^{x/y} w^{n/2-1}(1-w)^{n/2-1}dw \\ &= I_{x/y}\left(\frac{n}{2},\frac{n}{2}\right), \qquad (5)\end{aligned}$$

where $I_z(a,b)$ is the incomplete beta function; see Abramowitz and Stegun (1965), section 26.5.1, defined by the equation

$$I_z(a,b) = \frac{1}{\mathrm{B}(a,b)} \int_0^z t^{a-1}(1-t)^{b-1}dt, \quad 0 \le x \le 1.$$

Now the CDF of Snedecor's F distribution with n_1 and n_2 degrees of freedom is $1 - I_z(n_2/2, n_1/2)$, where $z = n_2/(n_2 + n_1 F)$, F being the argument of the CDF. Thus, if $F_{n,n}$ denotes a random variable distributed as F with n and n degrees of freedom, we have that

$$\Pr(F_{n,n} \le f) = 1 - I_{1/(1+f)}\Big(\frac{n}{2}, \frac{n}{2}\Big) \text{ and } \Pr(F_{n,n} > f) = I_{1/(1+f)}\Big(\frac{n}{2}, \frac{n}{2}\Big).$$

If we set $z = 1/(1+f)$, so that $f = 1/z - 1$, then

$$I_z\Big(\frac{n}{2}, \frac{n}{2}\Big) = \Pr\Big(F_{n,n} > \frac{1}{z} - 1\Big) = \Pr\Big(\frac{1}{1+F_{n,n}} < z\Big). \tag{6}$$

If X is a random variable of which the distribution is given by the CDF (5), then

$$\Pr(X \le x) = \Pr\Big(\frac{X}{y} \le \frac{x}{y}\Big) = I_{x/y}\Big(\frac{n}{2}, \frac{n}{2}\Big),$$

and so

$$\Pr\Big(\frac{X}{y} \le z\Big) = I_z\Big(\frac{n}{2}, \frac{n}{2}\Big). \tag{7}$$

Comparison of (6) and (7) shows that X/y and $1/(1+F_{n,n})$ have the same distribution, and so X can be generated as $y/(1+F_{n,n})$.

In the construction of the preceding section, we start from a random number drawn from the U(0,1) distribution, and apply the inverse of the CDF (5) to it. Let $G_{n,n}$ denote the inverse of the CDF of the F distribution with n and n degrees of freedom. Thus, if $U \sim \mathrm{U}(0,1)$, $G_{n,n}(U) \sim F(n,n)$, in obvious notation. From this it follows that $X \equiv y/(1+G_{n,n}(U))$ follows the distribution with CDF (5). Thus the function $G_{2^i|2^{i+1}}$ used in the construction satisfies the relation

$$G_{2^i|2^{i+1}}(U \,|\, V) = \frac{V}{1+G_{n,n}(U)}. \tag{8}$$

Code for the Construction

The following ***Ects*** code can be used in order to perform the construction with chi-squared variables. The code goes up to step 16, but the only impediment to going further is computing time.

```
set n = 16  # Number of steps
setrng kiss
set rt2 = sqrt(2)

sample 1 2^(n-1)
gen U = random(0,1) # One-time generation of all random numbers
```

```
mat fsp = rowcat(0,1) # These are used subsequently for
mat bsp = rowcat(1,0) # interpolation of new values
set linestyle = 1     # For the plot command

silent
noecho
set i = 0    # index of step
set dw = 0   # for increments
set w = 0    # for newly interpolated values

while i < n
   set W = chicrit(U(1),2^i) # Begin with the endpoint
   set j = 0                 # index of sigma algebra
   while j < i
      sample 1 2^j
      gen dW = W-lag(1,W)    # create increments
      sample 2^j+1 2^(j+1)
      del dw
      gen dw = lag(2^j,dW)   # move increments down
      set j = j+1
      set l = i-j
      gen dw = dw/(1+fishcrit(1-U,2^l,2^l)) # This is (8)
      sample 1 2^(j-1)       # Begin juggling to interpolate
      mat dw = dw(2^(j-1)+1,2^j,1,1)
      del w
      gen w = lag(1,W)+dw
      sample 1 2^j
      mat w = kron(w,bsp)
      mat W = kron(W,fsp)
      gen W = W+w            # This completes step j
   end

   set fac = rt2^(-(i+1))    # Scaling factor
   gen W = fac*(W-time(0))   # Centring

   sample 1 2^i+1
   gen x = lag(1,W)          # Move down so as to have W(0) = 0
   set i = i+1
end

quit
```

The same procedure can be used for the simpler Gaussian construction of section 2, by replacing the `fishcrit` function by `normcrit`, the inverse of the standard normal CDF. The details are omitted.

5. The Brownian Bridge

Consider an IID sample u_i, $i = 1, \ldots, n$ drawn from the U(0, 1) distribution. The empirical distribution function, or EDF, of this sample, is

$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}(x_i \le x), \quad x \in [0, 1]. \tag{9}$$

As usual, $\mathrm{I}(\cdot)$ is an indicator function of a Boolean argument. The standard asymptotic result about EDFs is that, as $n \to \infty$, $n^{1/2}(\hat{F}(x) - x)$ tends in distribution to a Brownian bridge, $B(x)$, that is, a Gaussian stochastic process defined on $[0, 1]$, with expectation zero, and covariance structure given by

$$\mathrm{E}\big(B(t)B(s)\big) = t(1 - s), \quad t \le s.$$

A Brownian bridge can be constructed from a standard Wiener process $W(x)$, by the relation

$$B(x) = W(x) - xW(1). \tag{10}$$

If this is done, then observe that

$$\begin{aligned}\mathrm{E}\big(B(x)W(1)\big) &= \mathrm{E}\big(W(x)W(1)\big) - x\mathrm{E}\big(W^2(1)\big)\\ &= x - x = 0,\end{aligned}$$

since the covariance structure of a Wiener process is

$$\mathrm{E}\big(W(t)W(s)\big) = t, \quad t \le s.$$

This implies that the Brownian bridge is independent of $W(1)$. Consequently, the Wiener process can be constructed from a Brownian bridge $B(x)$ and an independent standard normal variable Z by the relation

$$W(x) = B(x) + xZ. \tag{11}$$

Since we know how to generate a Wiener process, we can obviously generate a Brownian bridge using (10). A possibly even simpler way is to start the generation of a Wiener process with the deterministic value of zero for the first round, which gives us $W(1)$. Conditional on $W(1) = 0$, the distribution of the Wiener process $W(x)$ is that of a Brownian bridge, as can be seen directly from (11).

If an IID sample y_i, $i = 1, \ldots, n$, is drawn from a continuous distribution function G, then the set $u_i \equiv G(y_i)$ is an IID sample drawn from the U(0,1) distribution. The EDF of the y_i is

$$\begin{aligned}\hat{G}(y) &= \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}(y_i \leq y) \\ &= \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}\big(G(y_i) \leq G(y)\big) \\ &= \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}\big(u_i \leq G(y)\big) = \hat{F}(G(y)),\end{aligned}$$

where $\hat{F}$ is the EDF of the u_i. It follows that

$$n^{1/2}\big(\hat{G}(y) - G(y)\big) = n^{1/2}\big(\hat{F}(G(y)) - G(y)\big)$$

tends in distribution to $B(G(y))$.

6. The Construction

We wish to find a sequence of stochastic processes $B^n(x)$ defined on $[0,1]$ such that $B^n(x) \to B(x)$ almost surely for all $x \in [0,1]$, where $B(x)$ is a Brownian bridge. We also require that $B^n(x)$ should have, for a set x_i^n of points in $[0,1]$ that becomes dense in $[0,1]$ as $n \to \infty$, the distribution of the EDF of an IID sample of n draws from U(0,1). Such a sequence would be the analogue of the sequence of stochastic processes $W^n(t)$ that converges almost surely to a standard Wiener process $W(t)$ where, for an asymptotically dense set t_i^n, $W^n(t)$ has the distribution of the partial sum $n^{-1/2}\sum_{i=1}^{nt} x_i$, where the x_i constitute an IID sample drawn from N(0,1), or, more generally, from a distribution with continuous CDF, expectation zero, and unit variance.

We can see from the definition (9) of an EDF that $n^{1/2}(\hat{F}(x) - x)$ has, for each x, the structure of something to which a central limit theorem can be applied. We are not interested, however, in the partial sums of this process. Rather, we wish to focus on the sum of all n terms as a function of x. For this reason, the construction is somewhat different from that used for an almost sure functional central limit theorem.

It is convenient to perform the construction, just as for the almost sure central limit theorem, so that the sample size n takes on only values that are integer powers of 2. In addition, the points at which the sequence $B^n(x)$ is defined are dyadic points of the form $k2^{-i}$, $k = 1, \ldots, 2^i - 1$, where $n = 2^i$. For $k = 0$ and $k = 2^i$, that is for $x = 0$ and $x = 1$, the values $B^n(0) = 0$ and $B^n(1) = 0$ are fixed deterministically.

Consider then an IID sample u_j, $j = 1, \ldots, n$, of $n = 2^i$ drawings from U(0,1). The EDF of this sample is such that

$$\hat{F}(k2^{-i}) = 2^{-i}\sum_{j=1}^{2^i} \mathrm{I}(u_j \leq k2^{-i}), \quad k = 1, \ldots, 2^i - 1.$$

The summands here are Bernoulli variables, with probability $k2^{-i}$ of being equal to 1. The distribution of the $\hat{F}(k2^{-i})$ can therefore be worked out from the properties of the multinomial distribution. In particular, we have that

$$\begin{aligned}\Pr\Bigl(\sum_{j=1}^{2^i} \mathrm{I}(u_j \le k2^{-i}) = m\Bigr) &= \binom{2^i}{m}(k2^{-i})^m(1-k2^{-i})^{2^i-m} \\ &= 2^{-i\,2^i}\frac{(2^i)!}{m!(2^i-m)!}k^m(2^i-k)^{2^i-m}. \end{aligned} \tag{12}$$

The joint distribution of the values $\hat{F}(k2^{-i})$ can however be characterised more simply as follows. The arguments $k2^{-i}$, $i = 1, \ldots, 2^i - 1$, define a partition of $[0, 1]$ into 2^i segments. If we specify how many of the 2^i draws fall into each of these segments, then the values of the EDF at the arguments $k2^i$ are determined. Indeed, let n_k denote the number of draws in the segment $[(k-1)2^{-i}, k2^{-i}]$, $k = 1, \ldots, 2^i$. Then

$$\hat{F}(k2^{-i}) = 2^{-i}\sum_{l=1}^{k} n_l. \tag{13}$$

Since we must have $\sum_{k=1}^{2^i} n_k = 2^i$, the set of the n_k forms a partition of the integer 2^i.

Suppose now that the partition is constructed by drawing the u_j and dropping them, still labelled by the index j, into one of the segments. For notational ease, consider first the problem of dropping m labelled draws into n equally probable slots. The number of different, equally probable, ways of doing this is n^m. If the resulting (unlabelled) partition is $j_1, \ldots, j_n$, with $\sum_{i=1}^{n} j_i = m$, then this partition could be obtained in $m!/(j_1! \ldots j_n!)$ different ways. Thus the probability of obtaining the partition $j_1, \ldots, j_n$ is

$$\Pr(j_1, \ldots, j_n) = n^{-m}\frac{m!}{j_1! \ldots j_n!}.$$

If we apply this to our problem, where $m = n = 2^i$, we see that

$$\Pr(n_1, \ldots, n_{2^i}) = 2^{-i2^i}\frac{(2^i)!}{n_1! \ldots n_{2^i}!}. \tag{14}$$

This result gives the distribution of the values (13).

To alleviate notational complexity, we denote the process given by 2^i draws by $B^i(x)$ rather than $B^{2^i}(x)$. The construction of a realisation of the process $B^i(x)$ proceeds in i steps. At the first step, the random number u_1^1 is drawn from $\mathrm{U}(0, 1)$. (It is understood henceforth that a random number is a drawing from $\mathrm{U}(0, 1)$, independent of all other random numbers.) This random number is used to determine the quantities $n^i(0, 1/2)$ and $n^i(1/2, 1)$, which are the number of draws in the segments $[0, 1/2]$ and $[1/2, 1]$ respectively.

(There is no need to worry whether intervals are open or closed.) Quite generally, we denote by $n^i(a, b)$ the number of draws out of a total of 2^i that lie in $[a, b]$.

The marginal distribution of $n^i(0, 1/2)$ determines the joint distribution of $n^i(0, 1/2)$ and $n^i(1/2, 1)$ because the sum of the two random variables is 2^i. Since each $\mathrm{I}(u_j \leq 0.5)$ is a Bernoulli variable with probability 0.5, it follows that

$$\Pr\bigl(n^i(0, \tfrac{1}{2}) = n\bigr) = 2^{-2^i} \frac{(2^i)!}{n!\,(2^i - n)!}. \tag{15}$$

We refer to the discrete distribution of the sum S of m IID Bernoulli variables each with probability 0.5 as the $\mathrm{M}(m)$ distribution, with probabilities

$$p_n^m \equiv \Pr(S = n) = 2^{-m} \frac{m!}{n!\,(m - n)!}. \tag{16}$$

Thus the probability (15) is $p_n^{2^i}$. The realisation $n^i(0, 1/2)$ is obtained from the random number u_1^1 by the relation

$$\sum_{j=0}^{n^i(0,1/2)} p_j^{2^i} \leq u_1^1 < \sum_{j=0}^{n^i(0,1/2)+1} p_j^{2^i}.$$

Intuitively, $n^i(0, 1/2)$ is the u_1^1–quantile of the $\mathrm{M}(2^i)$ distribution.

When stage k is reached, we require that the quantities $n^i((l-1)2^{-k}, l2^{-k})$, $l = 1, \ldots, 2^k$, should all be determined. Therefore, the new determinations at stage k are only of the $n^i((l-1)2^{-k}, l2^{-k})$ for l even, since, if l is odd, that is, if $l = 2m - 1$ for a positive integer m, then $(l-1)2^{-k} = (m-1)2^{-k+1}$, so that, in accordance with our requirement, $n^i((m-1)2^{-k+1}, m2^{-k+1})$ has already been determined. But

$$\begin{aligned} n^i((m-1)2^{-k+1}, m2^{-k+1}) &= n^i((l-1)2^{-k}, (l+1)2^{-k}) \\ &= n^i((l-1)2^{-k}, l2^{-k}) + n^i(l2^{-k}, (l+1)2^{-k}), \end{aligned} \tag{17}$$

and so only one of the two terms in (17) needs to be realised for the first time at stage k.

Denote the sigma-algebra generated by those variables that are realised at or before stage k by $\mathcal{F}_k$. Then the variables realised at stage k are realised conditionally on $\mathcal{F}_{k-1}$. From (17), we see that what is needed is to partition $n^i((l-1)2^{-k}, (l+1)2^{-k})$ into two segments, with equal probability for a draw in either segment. Thus we use the random numbers u_l^k, $l = 1, 3, \ldots, 2^k - 1$ in order to generate the $n^i((l-1)2^{-k}, l2^{-k})$ as the u_l^k–quantiles of the $\mathrm{M}(n^i((l-1)2^{-k}, (l+1)2^{-k}))$ distribution. We now claim that the probability of the stage k partition $\{n^i((l-1)2^{-k}, l2^{-k})\}$, $l = 1, \ldots, 2^k$ is

$$\Pr\Bigl(\{n^i((l-1)2^{-k}, l2^{-k})_{l=1}^{l=2^k}\}\Bigr) = \frac{2^{-k2^i}(2^i)!}{\prod_{l=1}^{2^k} n^i((l-1)2^{-k}, l2^{-k})!} \tag{18}$$

The proof is by induction. At stage 1, the only realisation is of $n^i(0,1/2)$, and, from (12) with $k = 2^{i-1}$, we see that

$$\begin{aligned}\Pr\big((n^i(0,1/2),n^i(1/2,1)) = (n_1,n_2)\big) &= \Pr\big(n^i(0,1/2) = n_1\big)\\ &= 2^{-i2^i}\frac{(2^i)!}{n_1!\,n_2!}(2^{i-1})^{2^i} = \frac{2^{-2^i}(2^i)!}{n_1!\,n_2!},\end{aligned} \tag{19}$$

where, of course, $n_1 + n_2 = 2^i$. It is clear that (19) is just (18) for $k = 1$.

Now suppose that (18) holds with k replaced by $k-1$. The new random numbers at stage k are independent of $\mathcal{F}_{k-1}$, and so the distribution of $n^i((l-1)2^{-k}, l2^{-k})$, $l = 1, 3, 5, \ldots, 2^k - 1$, conditional on $\mathcal{F}_{k-1}$, is $\mathrm{M}(n^i((l-1)2^{-k},(l+1)2^{-k}))$. Thus, conditional on $\mathcal{F}_{k-1}$, the probabilities of the realisations $n^i((l-1)2^{-k}, l2^{-k})$ for odd l are, from (16),

$$2^{-n^i((l-1)2^{-k},(l+1)2^{-k})}\frac{n^i((l-1)2^{-k},(l+1)2^{-k})!}{n^i((l-1)2^{-k},l2^{-k})!\,n^i(l2^{-k},(l+1)2^{-k})!}. \tag{20}$$

Since the random numbers u_l^k are mutually independent, these probabilities are of mutually independent events, and so the probability of the set $\{n^i((l-1)2^{-k}, l2^{-k})\}$, for $l = 1, 2, \ldots, 2^k$, conditional on $\mathcal{F}_{k-1}$, is their product.

The product of the powers of 2 that are the first factors in (20) is 2 raised to the power

$$-\sum_{l=1,3,\ldots,2^k-1} n^i((l-1)2^{-k},(l+1)2^{-k}) = -n^i(0,1) = -2^i,$$

the product of the denominators in (20) is

$$\prod_{l=1,3,\ldots,2^k-1} n^i((l-1)2^{-k},l2^{-k})!\,n^i(l2^{-k},(l+1)2^{-k})! = \prod_{l=1}^{2^k} n^i((l-1)2^{-k},l2^{-k})!,$$

and the product of the numerators in (20) is

$$\prod_{l=1,3,\ldots,2^k-1} n^i((l-1)2^{-k},(l+1)2^{-k})! = \prod_{m=1}^{2^{k-1}} n^i((m-1)2^{-k+1},m2^{-k+1})!.$$

Thus the product of the probabilities (20) is

$$\frac{2^{-2^i}\prod_{m=1}^{2^{k-1}} n^i((m-1)2^{-k+1},m2^{-k+1})!}{\prod_{l=1}^{2^k} n^i((l-1)2^{-k},l2^{-k})!}. \tag{21}$$

The unconditional probability of the stage $k-1$ realisations is, by (18) and the induction hypothesis,

$$\frac{2^{-(k-1)2^i}(2^i)!}{\prod_{m=1}^{2^{k-1}} n^i((m-1)2^{-k+1},m2^{-k+1})!}. \tag{22}$$

Thus the unconditional probability of the stage k realisations, which is the product of the unconditional probability (22) and the conditional probability (21), is

$$\frac{2^{-k2^i}(2^i)!}{\prod_{l=1}^{2^k} n^i((l-1)2^{-k}, l2^{-k})!},$$

which is just (18). The induction is therefore complete.

Setting $k = i$ gives the probability of the final configuration $n^i_1, n^i_2, \ldots, n^i_{2^i}$, where we note that $n^i_l = n^i((l-1)2^i, l2^i)$. We find that

$$\Pr(n^i_1, \ldots, n^i_{2^i}) = \frac{2^{-i2^i}(2^i)!}{\prod_{l=1}^{2^i} n^i_l!},$$

which is exactly the distribution (14). Thus our stepwise construction leads to a realisation from the desired distribution of the EDF of a set of 2^i IID drawings from U(0, 1) evaluated at the dyadic points $k2^{-i}$. By construction, these EDFs, interpolated as we wish between the dyadic points, converge almost surely to a limiting process.

7. The Limiting Distribution

The usual result for EDFs demonstrates that the limit of the sequence of stochastic processes defined in the previous section as $i \to \infty$ is a Brownian bridge. It is interesting to show this fact explicitly making use of the properties of the multinomial distribution.

From (13), we have, for the EDF that is element i of the sequence,

$$\hat{F}^i(k2^{-i}) = 2^{-i} \sum_{l=1}^{k} n^i_l = 2^{-i} n^i(0, k2^{-i}).$$

The expression that converges to a limit as $i \to \infty$ is $2^{i/2}\big(\hat{F}^i(x) - x\big)$, for $x \in [0, 1]$. For element i, we limit attention to values of x equal to one of the dyadic points. For $x = k2^{-i}$, $2^{i/2}\big(\hat{F}^i(x) - x\big)$ becomes $2^{-i/2}(n^i(0, k2^{-i}) - k)$, which we denote as B^i_k.

The distribution of B^i_k is given by that of $n^i(0, k2^{-i})$. From (12),

$$\Pr\big(n^i(0, k2^{-i}) = m\big) = \Pr\big(B^i_k = 2^{-i/2}(m-k)\big) = \frac{2^{-i2^i}(2^i)!\, k^m\, (2^i-k)^{2^i-m}}{m!\,(2^i-m)!}.$$

In order to obtain the CDF of B^i_k, we calculate as follows:

$$\begin{aligned} \Pr(B^i_k \le x) &= \sum_{2^{-i/2}(m-k) \le x} \frac{2^{-i2^i}(2^i)!\, k^m (2^i-k)^{2^i-n}}{m!\,(2^i-m)!} \\ &= 2^{-i2^i}(2^i)! \sum_{m=0}^{\lfloor k+2^{i/2}x \rfloor} \frac{k^m (2^i-k)^{2^i-m}}{m!\,(2^i-m)!}. \end{aligned} \tag{23}$$

Now let $k2^{-i} = r$. It follows that $r \in [0, 1]$. The probability (23) is

$$(2^i)! \sum_{m=0}^{\lfloor 2^i r + 2^{i/2}x \rfloor} \frac{r^m (1-r)^{2^i-m}}{m!\,(2^i-m)!}.$$

In order to find the limit as $i \to \infty$, let $N = 2^i$ and let $N \to \infty$. The probability (23) becomes in this new notation

$$\Pr(B_k^i \le x) = \sum_{m=0}^{\lfloor Nr + x\sqrt{N} \rfloor} \binom{N}{m} r^n (1-r)^{N-m}. \tag{24}$$

We now make use of a result found in Abramowitz and Stegun (1965), equation 26.5.24, which tells us that

$$\sum_{s=a}^{n} \binom{n}{s} p^s (1-p)^{n-s} = I_p(a, n-a+1), \quad p \in [0,1], \tag{25}$$

where $I_x(a,b)$ is the incomplete beta function, defined by

$$I_x(a,b) = \frac{1}{\mathrm{B}(a,b)} \int_0^x t^{a-1}(1-t)^{b-1} dt, \tag{26}$$

with $\mathrm{B}(a,b)$ the (complete) beta function, defined as

$$\mathrm{B}(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 t^{a-1}(1-t)^{b-1} dt,$$

where $\Gamma(a)$ is the gamma function. The result (25) can be shown without much trouble by repeated integration by parts of the definition (26) of the incomplete beta function.

Using (25) in (24) gives us that

$$\begin{aligned}
\Pr(B_k^i \le x) &= 1 - \sum_{m=\lceil Nr + x\sqrt{N} \rceil} \binom{N}{m} r^m (1-r)^{N-n} \\
&= 1 - I_r(Nr + x\sqrt{N}, N(1-r) - x\sqrt{N} + 1) \\
&= I_{1-r}(N(1-r) - x\sqrt{N} + 1, Nr + x\sqrt{N}),
\end{aligned} \tag{27}$$

since $I_x(a,b) = 1 - I_{1-x}(b,a)$. If this is to correspond to a Brownian bridge, then the limit distribution of B_k^i should be $\mathrm{N}(0, r(1-r))$, and so we wish to show that (27) tends to $\Phi(x/\sqrt{r(1-r)})$ as $N \to \infty$, where $\Phi(\cdot)$ is the standard normal CDF.

An asymptotic expansion is given for the incomplete beta function in Abramowitz and Stegun (1965), equation 26.5.19, as follows:

$$I_x(a,b) \sim \Phi(y) - \phi(y)\left(a_1 + \frac{a_2(y-a_1)}{1+a_2} + \frac{a_3(1+y^2/2)}{1+a_2} + \dots\right), \text{ where} \tag{28}$$

$$y^2 = 2\Big((a+b-1)\log\frac{a+b-1}{a+b-2} + (a-1)\log\frac{a-1}{(a+b-1)x}$$

$$
\begin{aligned}
&+(b-1)\log\frac{b-1}{(a+b-1)(1-x)}\Big),\\
a_1 &= \frac{2}{3}(b-a)\big((a+b-2)(a-1))(b-1)\big)^{-1/2},\\
a_2 &= \frac{1}{12}\left(\frac{1}{a-1}+\frac{1}{b-1}-\frac{13}{a+b-1}\right), \text{ and}\\
a_3 &= -\frac{8}{15}\left(a_1\Big(a_2+\frac{3}{a+b-2}\Big)\right).
\end{aligned}
$$

The variable y is taken negative if $x < (a-1)/(a+b-2)$. Here $\phi(\cdot)$ is the standard normal density. For the expansion of (27), we have $a+b = N+1$, and so the argument y is given by

$$
\begin{aligned}
y^2 = 2\Big(-N\log\big(1-\frac{1}{N}\big) &+ \big(N(1-r)-x\sqrt{N}\big)\log\frac{N(1-r)-x\sqrt{N}}{N(1-r)}\\
&+\big(Nr+x\sqrt{N}-1\big)\log\frac{Nr+x\sqrt{N}-1}{Nr}\Big)
\end{aligned}
$$

Taylor expansion of the logarithms gives, with error of order $N^{-1/2}$,

$$
\begin{aligned}
y^2 &= 2\Bigg[1-\big(N(1-r)-x\sqrt{N}\big)\left(\frac{x}{\sqrt{N}(1-r)}+\frac{x^2}{2N(1-r)^2}\right)\\
&\quad + (Nr+x\sqrt{N}-1)\left(\frac{x\sqrt{N}-1}{Nr}-\frac{(x\sqrt{N}-1)^2}{2n^2r^2}\right)\Bigg]\\
&= 2\Big[1-x\sqrt{N}+\frac{x^2}{1-r}-\frac{x^2}{2(1-r)}+x\sqrt{N}-1+\frac{(x\sqrt{N}-1)^2}{Nr}-\frac{(x\sqrt{N}-1)^2}{2N^2r^2}\Big]\\
&= \frac{x^2}{1-r}+\frac{x^2}{r}=\frac{x^2}{r(1-r)}.
\end{aligned}
$$

Further, since $a = O(N)$ and $b = O(N)$, we see that $a_1 = O(N^{-1/2})$, $a_2 = O(N^{-1})$, and $a_3 = O(N^{-3/2})$. Thus only the first term in (28) contributes to the limit when $N \to \infty$. The limit of this term is just $\Phi(y)$ where $y^2 = x^2/(r(1-r))$. The condition for the sign of y to be negative is

$$
1-r < \frac{N(1-r)-x\sqrt{N}}{N-1} = 1-r+\frac{1-r}{N-1}-\frac{x\sqrt{N}}{N-1},
$$

that is, $x < N^{-1/2}(1-r)$. As $N \to \infty$, this condition is just $x < 0$, and so, for all real x, positive or negative, the limit of (27) is $\Phi(x/\sqrt{r(1-r)})$, as we wished to show.

Appendix

Construction of bivariate variables from random numbers

Let the marginal CDF of X be F_X, and the CDF of Y conditional on X be $F_{Y|X}$. The joint CDF is given as follows:

$$\begin{aligned} F_{X,Y}(x,y) &= \Pr\bigl((X \le x) \cap (Y \le y)\bigr) \\ &= \mathrm{E}\bigl(\mathrm{I}(X \le x)\mathrm{I}(Y \le y)\bigr) \\ &= \mathrm{E}\bigl(\mathrm{E}\bigl(\mathrm{I}(X \le x)\mathrm{I}(Y \le y) \,|\, X\bigr)\bigr) \\ &= \mathrm{E}\bigl(\mathrm{I}(X \le x)\mathrm{E}\bigl(\mathrm{I}(Y \le y) \,|\, X\bigr)\bigr) \\ &= \mathrm{E}\bigl(\mathrm{I}(X \le x)F_{Y|X}(y \,|\, X)\bigr) \\ &= \int_{-\infty}^{x} F_{Y|X}(y \,|\, z)\, dF_X(z). \end{aligned} \tag{29}$$

Now let U_1 and U_2 be two independent U(0,1) variables. We wish to show that $X \equiv F_X^{-1}(U_1)$ and $Y \equiv F_{Y|X}^{-1}(U_2 \,|\, F_X^{-1}(U_1))$ constitute a joint drawing from the distribution with CDF (29).

We have

$$\begin{aligned} \Pr\bigl((X \le x) \cap (Y \le y)\bigr) &= \Pr\bigl((U_1 \le F_X(x)) \cap (U_2 \le F_{Y|X}(y \,|\, F_X^{-1}(U_1))\bigr) \\ &= \mathrm{E}\bigl(\mathrm{I}(U_1 \le F_X(x))\mathrm{E}\bigl(\mathrm{I}(U_2 \le F_{Y|X}(y \,|\, F_X^{-1}(U_1) \,|\, U_1)\bigr)\bigr). \end{aligned} \tag{30}$$

Now

$$\mathrm{E}\bigl(\mathrm{I}(U_2 \le F_{Y|X}(y \,|\, F_X^{-1}(U_1)) \,|\, U_1\bigr) = F_{Y|X}(y \,|\, F_X^{-1}(U_1))$$

by the independence of U_1 and U_2. Thus the unconditional expectation (30) becomes

$$\mathrm{E}\bigl(\mathrm{I}(U_1 \le F_X(x))F_{Y|X}(y \,|\, F_X^{-1}(U_1)\bigr) = \int_0^{F_X(x)} F_{Y|X}(y \,|\, F_X^{-1}(u_1))\, du_1, \tag{31}$$

since $U_1 \sim \mathrm{U}(0,1)$. Now change variables by the formula $u_1 = F_X(z)$, from which we see that $du_1 = dF_X(z)$. The right-hand side of (31) becomes

$$\int_{-\infty}^{x} F_{Y|X}(y \,|\, z)\, dF_X(z),$$

which is identical to (29). This completes the proof. ■

It is clear that the proof can be extended to show how drawings from a multivariate joint distribution can be constructed from a set of independent random numbers on the basis of a set of conditional distributions.

References

Abramowitz, M., and I. A. Stegun (1965). *Handbook of Mathematical Functions*, New York, Dover.