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,

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

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 $W(1)$, or 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 $j < i$,

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

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

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 on 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 anything 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

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

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 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 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. Subsequently, we obtain all the $W^i(k)$ by dividing the $V^i(k)$ by $2^{i/2}$. 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

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

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 $z = n_2/(n_2 + n_1F)$, 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 $U \sim \text{U}(0, 1)$, $G_{n,n}(U) \sim F(n, n)$, 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.

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:

$$\begin{aligned} F_{X,Y}(x,y) &= \Pr\bigl((X \le x) \cap (Y \le y)\bigr) \\ &= \mathrm{E}\bigl(\mathrm{I}(X \le x)\mathrm{I}(Y \le y)\bigr) \\ &= \mathrm{E}\bigl(\mathrm{E}\bigl(\mathrm{I}(X \le x)\mathrm{I}(Y \le y) \,|\, X\bigr)\bigr) \\ &= \mathrm{E}\bigl(\mathrm{I}(X \le x)\mathrm{E}\bigl(\mathrm{I}(Y \le y) \,|\, X\bigr)\bigr) \\ &= \mathrm{E}\bigl(\mathrm{I}(X \le x)F_{Y|X}(y \,|\, X)\bigr) \\ &= \int_{-\infty}^{x} F_{Y|X}(y \,|\, z)\, dF_X(z). \end{aligned} \tag{9}$$

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 (9).

We have

$$\begin{aligned} \Pr\bigl((X \le x) \cap (Y \le y)\bigr) &= \Pr\bigl((U_1 \le F_X(x)) \cap (U_2 \le F_{Y|X}(y \,|\, F_X^{-1}(U_1)))\bigr) \\ &= \mathrm{E}\bigl(\mathrm{I}(U_1 \le F_X(x))\mathrm{E}\bigl(\mathrm{I}(U_2 \le F_{Y|X}(y \,|\, F_X^{-1}(U_1) \,|\, U_1)\bigr). \end{aligned} \tag{10}$$

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 (10) becomes

$$\mathrm{E}\bigl(\mathrm{I}(U_1 \le F_X(x))F_{Y|X}(y \,|\, F_X^{-1}(U_1)\bigr) = \int_0^{F_X(x)} F_{Y|X}(y \,|\, F_X^{-1}(u_1))\, du_1, \tag{11}$$

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 (11) becomes

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

which is identical to (9). This completes the proof. ■

References

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