The ECDF is one of the most useful statistics, especially in nonparametric and robust inference. It is essentially the same as the set of order statistics, so like them, it is a sufficient statistic. Its distribution at a point is binomial, and so its pointwise properties are easy to see. Its global relationship to the most fundamental measure of a probability model, however, accounts for its usefulness. The basic facts regard the convergence of the sup distance of the ECDF from the CDF, $\rho_\infty(F_n, F)$, to zero.

The Dvoretzky/Kiefer/Wolfowitz inequality provides a bound for the probability that the sup distance of the ECDF from the CDF exceeds a given value. In one-dimension, for any positive z, there is a positive constant C that does not depend on F, z, or n, such that

$$\Pr(\rho_\infty(F_n, F) > z) \leq C\mathrm{e}^{-2nz^2}.$$

Of course, in any event, $\rho_\infty(F_n, F) < 1$. This inequality is useful in proving convergence results for the ECDF. Some important results are given in Theorems 5.1 and 5.2 in Shao. A simple and more common statement of the convergence is the so-called Glivenko-Cantelli theorem.

When we consider the convergence of metrics on functions, the arguments of the functions are sequences of random variables, yet the metric integrates out the argument. One way of handling this is just to use the notation F_n and F, as Shao does. Another way is to use the notation $F_n(x, \omega)$ to indicate that the ECDF is a random variable, yet to allow it to have an argument just as the CDF does. I will use this notation occasionally, but usually I will just write $F_n(x)$. The randomness comes in the definition of $F_n(x)$, which is based on the random sample.

Theorem 1 (Glivenko-Cantelli) *If $X_1, \ldots, X_n$ be i.i.d. with CDF F and ECDF F_n, and if $D_n(\omega) = \rho_\infty(F_n, F) = \sup_x(|F_n(x, \omega) - F(x)|)$, then $D_n(\omega) \longrightarrow 0$ wp1.*

Proof. First, note by the SSLN and the binomial distribution of F_n, $\forall$ (fixed) x, $F_n(x, \omega) \longrightarrow F(x)$ wp1; that is,

$$\lim_{n\to\infty} F_n(x, \omega) = F(x)$$

$\forall x$, except $x \in A_x$, where $\Pr(A_x) = 0$.

The problem here is that A_x depends on x and so there are uncountably many such sets. The probability of their union may possibly be positive. So we must be careful.

We will work on the CDF and ECDF from the other side of x (the discontinuous side). Again, by the SSLN, we have

$$\lim_{n\to\infty} F_n(x-, \omega) = F(x-)$$

$\forall x$, except $x \in B_x$, where $\Pr(B_x) = 0$.
Now, let

$$\phi(u) = \inf\{x\ ;\ u \leq F(x)\} \quad \text{for } 0 < u \leq 1.$$

(Notice $F(\phi(u)-) \leq u \leq F(\phi(u))$. Sketch the picture.)

Now consider $x_{m,k} = \phi(k/m)$ for positive integers m and k with $1 \leq k \leq m$. (There are countably many $x_{m,k}$, and so when we consider $F_n(x_{m,k}, \omega)$ and $F(x_{m,k})$, there are countably many null-probability sets, $A_{x_{m,k}}$ and $B_{x_{m,k}}$, where the functions differ in the limit.

We immediately have the three relations:

$$F(x_{m,k}-) - F(x_{m,k-1}) \leq m^{-1}$$

$$F(x_{m,1}-) \leq m^{-1}$$

and

$$F(x_{m,m}) \geq 1 - m^{-1},$$

and, of course, F is nondecreasing.

Now let $D_{m,n}(\omega)$ be the maximum over all $k = 1, \ldots, m$ of

$$|F_n(x_{m,k}, \omega) - F(x_{m,k})|$$

and

$$|F_n(x_{m,k}-, \omega) - F(x_{m,k}-)|.$$

(Compare $D_n(\omega)$.)
We now consider three ranges for x:

$$\begin{array}{c} (-\infty\ ,\ x_{m,1}) \\ {[x_{m,k-1}\ ,\ x_{m,k})} \quad \text{for } k = 1, \ldots, m \\ {[x_{m,m}\ ,\ \infty)} \end{array}$$

Consider $[x_{m,k-1} \leq x < x_{m,k})$. In this interval,

$$\begin{aligned} F_n(x, \omega) &\leq F_n(x_{m,k}-, \omega) \\ &\leq F(x_{m,k}-) + D_{m,n}(\omega) \\ &\leq F(x) + m^{-1} + D_{m,n}(\omega) \end{aligned}$$

and

$$\begin{aligned} F_n(x, \omega) &\geq F_n(x_{m,k-1}, \omega) \\ &\geq F(x_{m,k-1}) - D_{m,n}(\omega) \\ &\geq F(x) - m^{-1} - D_{m,n}(\omega) \end{aligned}$$

Hence, in these intervals, we have

$$\begin{aligned} D_{m,n}(\omega) + m^{-1} &\geq \sup_x |F_n(x, \omega) - F(x)| \\ &= D_n(\omega). \end{aligned}$$

We can get this same inequality in each of the other two intervals.

Now, $\forall m$, except on the unions over k of $A_{x_{m,k}}$ and $B_{x_{m,k}}$, $\lim_n D_{m,n}(\omega) = 0$, and so $\lim_n D_n(\omega) = 0$, except on a set of probability measure 0 (the countable unions of the $A_{x_{m,k}}$ and $B_{x_{m,k}}$.) Hence, we have the convergence wp1; i.e., a.s. convergence.