Clearly, $nf_n(\sigma_k, W_n) = \sum_{j=1}^n \mathbf{1}_{\{X_j=\sigma_k\}} = \text{Bin}(n, p_k)$. In particular, it has mean np_k and variance $np_k(1-p_k) \le np_k$. Thus, by the Chebyshev inequality,

$$1 - \pi_n(\epsilon) \le \sum_{k=1}^m \frac{np_k}{n^2\epsilon^2} = \frac{1}{n\epsilon^2}. \tag{6.57}$$

The theorem follows. □

6.3. The Glivenko–Cantelli Theorem. A histogram is a random discrete probability distribution; it depends on $\{X_i\}_{i=1}^n$, and assigns probability $p_n(x)$ to any point $x \in \mathbf{R}$, where $p_n(x)$ is the fraction of the data that is equal to x. Its cumulative distribution function is called the empirical distribution function, and defined more formally as follows.

Definition 6.28. If $\{X_i\}_{i=1}^\infty$ are i.i.d. random variables all with distribution function F, then their *empirical distribution function* F_n is

$$F_n(x, \omega) = \frac{1}{n}\sum_{k=1}^n \mathbf{1}_{\{X_k \le x\}}(\omega) \qquad \forall x \in \mathbf{R}, \omega \in \Omega. \tag{6.58}$$

Thus, we can view $x \mapsto F_n(x, \cdot)$ as a random (cumulative) distribution function. As we did with other sorts of random variables, we suppress the dependence on the ω variable, and write $F_n(x)$ in place of $F_n(x, \omega)$.

The following is due to Glivenko (1933) and Cantelli (1933b). In data-analytic terms, this theorem presents a uniform approximation to an unknown distribution function F based on a random sample from this distribution.

Theorem 6.29. $\lim_{n\to\infty} \sup_{x\in\mathbf{R}} |F_n(x) - F(x)| = 0$ *a.s.*

Proof. Since F_n and F are right-continuous,

$$\sup_{x\in\mathbf{R}} |F_n(x) - F(x)| = \sup_{x\in\mathbf{Q}} |F_n(x) - F(x)|. \tag{6.59}$$

Thus, $\sup_{x\in\mathbf{R}} |F_n(x) - F(x)|$ is a random variable.

Fix $x \in \mathbf{R}$ and note that $nF_n(x) = \text{Bin}(n, F(x))$. By the strong law of large numbers (p. 73), for any $\epsilon > 0$, $n \ge 1$, and $x \in \mathbf{R}$,

$$|F_n(x) - F(x)| \le \epsilon \quad \text{for all but finitely many } n\text{'s, a.s.} \tag{6.60}$$

Recall that: (i) F is non-decreasing; (ii) F is right-continuous; (iii) $F(\infty) = 1$; and (iv) $F(-\infty) = 0$. Therefore, we can find $-\infty < x_0 < \cdots < x_m < \infty$ such that: $F(x_0) \le \epsilon$; $F(x_m) \ge 1 - \epsilon$; and

$$\sup_{x_{j-1} \le x < x_j} |F(x) - F(x_{j-1})| \le \epsilon \qquad \forall j = 1, \ldots, m. \tag{6.61}$$

According to (6.60),

$$\max_{0\le j\le m} |F_n(x_j) - F(x_j)| \le \epsilon \quad \text{for all but finitely many } n\text{'s, a.s.} \tag{6.62}$$

Hence follows that if $x \in [x_{j-1}, x_j)$ for some $1 \le j \le m$, then with probability one the following holds for all sufficiently large integers n:

$$\begin{aligned} F_n(x) &\le F_n(x_j) \le F(x_j) + \epsilon; \text{ and} \\ F(x_{j-1}) &\le F_n(x_{j-1}) + \epsilon \le F_n(x) + \epsilon. \end{aligned} \tag{6.63}$$

By (6.61), $F(x_j) \le F(x_{j-1}) + \epsilon$. Since F is non-decreasing it follows that with probability one,

$$\sup_{x_0\le x\le x_m} |F_n(x) - F(x)| \le 2\epsilon \quad \text{for all but finitely many } n\text{'s.} \tag{6.64}$$

Choose and fix n large enough that the previous inequality is satisfied. If $x > x_m$, then $F(x) \ge F(x_m) \ge 1 - \epsilon$ and $F_n(x) \ge F(x_m) - \epsilon \ge 1 - 2\epsilon$. Therefore,

$$|F_n(x) - F(x)| \le |1 - F(x)| + |1 - F_n(x)| \le 3\epsilon \qquad \forall x > x_m. \tag{6.65}$$

Similarly, if $x < x_0$, then $|F(x) - F_n(x)| \le F(x_0) + F_n(x_0) \le 3\epsilon$. Consequently, with probability one,

$$\sup_{x\in\mathbf{R}} |F_n(x) - F(x)| \le 3\epsilon \quad \text{for all but finitely many } n\text{'s.} \tag{6.66}$$

This proves the theorem. □

6.4. The Erdős Bound on Ramsey Numbers. Let us begin with a definition or two from graph theory.

Definition 6.30. The *complete graph* K_m on m vertices is a collection of m distinct vertices any two of which are connected by a unique edge. The nth (diagonal) *Ramsey number* R_n is the smallest integer N such that any bi-chromatic coloring of the edges of K_N yields a $K_n \subseteq K_N$ whose edges are all of the same color.

To understand this definition suppose $R_n = N$. Then, no matter how we color the edges of K_N using only the colors red and blue, somewhere inside K_N there exists a K_n whose edges are either all blue or all red, and N is the smallest such value. It is possible to check that $R_2 = 3$ and $R_3 = 6$, for example.

Ramsey (1930) introduced these and other Ramsey numbers to discuss ways of checking the consistency of a logical formula. See also Skolem (1933) and Erdős and Szekeres (1935).

As a key step in his proofs Ramsey proved that $R_n < \infty$ for all $n \ge 1$. Evidently, $R_n \to \infty$ as $n \to \infty$; in fact, it is obvious that $R_n \ge n$. The following theorem of Erdős (1948) presents a much better lower bound.