Th4: The Glivenko–Cantelli theorem, Proof, Simulations
Glivenko–Cantelli

In the theory of probability, the Glivenko–Cantelli theorem (sometimes referred to as the Fundamental Theorem of Statistics), named after Valery Ivanovich Glivenko and Francesco Paolo Cantelli, determines the asymptotic behavior of the empirical distribution function as the number of independent and identically distributed observations grows.
The uniform convergence of more general empirical measures becomes an important property of the Glivenko–Cantelli classes of functions or sets.

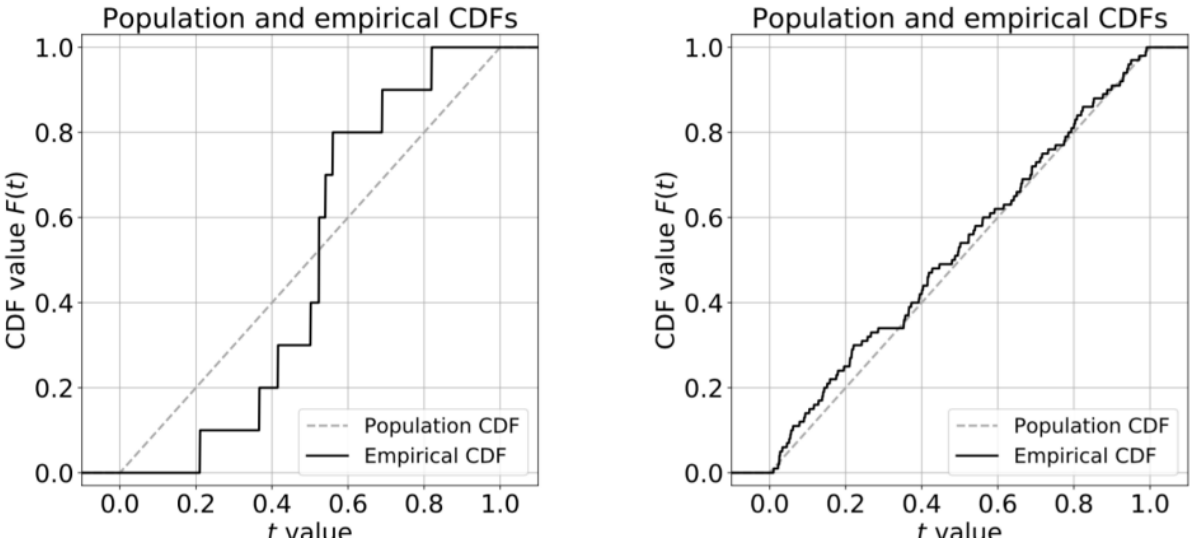
In the most general statistical setup, we have some unknown probability distribution μ, and we try to estimate properties of it by taking independent, identically distributed samples X1, X2, … , Xn with distribution μ (written Xi ~ μ). The general hope is that as we take enough samples (n → ∞), we can recover properties of the distribution μ.

The hope and general assumption of statistics is that it is possible to estimate any property of the distribution μ consistently. A distribution μ on R specifies a cumulative distribution function (CDF).

$$F_\mu(t) := \mathbb{P}_\mu(X \leq t).$$

The CDF uniquely specifies the distribution it arises from, as well.

We can see this approximation in the following graph



We have μ((a, b]) =Fμ(b) − Fμ(a), which determines the value of μ by the Caratheodory outer measure construction. So if we can consistently estimate Fμ with the data X1, . . . , Xn as n → ∞, we should be able to estimate any property of the distribution μ. The Glivenko-Cantelli theorem says that this estimation of the entire distribution is indeed possible.

So in simpler words we can say that the Glivenko-Cantelli (GC) Theorem gives the relationship between the empirical ("data-driven") CDF and the true CDF. A CDF gives the probability that a random variable takes value less than a specified number. The empirical CDF estimates this

probability with the proportion of data less than a specified number. The GC theorem says that the difference between the empirical ("data driven") CDF and the true CDF goes away as you get a larger sample size.

the Glivenko-Cantelli theorem states the following:
With probability one

$$D_n := \sup_{-\infty<t<\infty} |F_n(t) - F(t)| = \sup_{t \text{ rational}} |F_n(t) - F(t)| \longrightarrow 0$$

The proof of the Glivenko-Cantelli theorem is the following:
If $-\infty \le s \le t < u \le \infty$, then

$F_n(s) - F(s) - (F(u-) - F(s)) \le F_n(t) - F(t) \le F_n(u-) - F(s) = (F_n(u-) - F(u-)) + (F(u-) - F(s))$.

Thus if S is a finite subset of R, say, $S = \{s_1, \ldots, s_k\}$ with $s_1 < s_2 < \ldots < s_k$, with associated partition

$\pi S := \{(-\infty =: s_0, s_1), [s_1, s_2), \ldots, [s_{k-1}, s_k), [s_k, s_{k+1} := \infty)\}$,

we have

1) $Dn = \sup_t |F_n(t) - F(t)| \le \max_{s\in S} (|F_n(s) - F(s)| \vee |F_n(s-) - F(s-)|) + w\pi_S(F)$
where
2) $w\pi_S(F) := \max_{1 \le j \le k+1} |F(s_j-) - F(s_{j-1})|$.

Now observe that there exists a sequence $(S_m)_{m\ge1}$ of finite subsets $S_m \subset R$ such that $w\pi_{Sm}(F) \to 0$ as $m \to \infty$;

This is clear from inspection of a diagram I'll draw in class. Put $S_\infty := \cup_{m=1}^{\infty} S_m$. S_∞ is countable, and countably many applications of the SLLN imply that there exists a P-null set N off of which
3) $|F_n(s) - F(s)| \vee |F_n(s-) - F(s-)| \to 0$ for all $s \in S_\infty$

Now from (1), (2), (3) we get $D_n \to 0$ off of N, as desired.

A simulation of the Glivenko-Cantelli theorem can be found on this page.

Bibliography
-https://en.wikipedia.org/wiki/Glivenko–Cantelli_theorem
-https://pillowmath.github.io/Expository%20Notes/VC-Dimension-and-Glivenko-Cantelli-Notes.pdf
-https://www.ams.jhu.edu/~fill/550.621/prob_two/GCthm.pdf