6.S951 Modern Mathematical Statistics Fall 2024

Lecture 10 — October 10, 2024

Prof. Stephen Bates

Scribe: Ishan Ganguly

1 Outline:

Agenda:

1. Inference for CDFs
2. Empirical CDFs
3. DKW inequality, confidence bounds
4. Inference via simulation

For the next few lectures, we'll talk about more ways of constructing confidence intervals. In particular, we will look at nonparametric ways of inference for cumulative distribution functions (CDFs).

2 Empirical CDFs

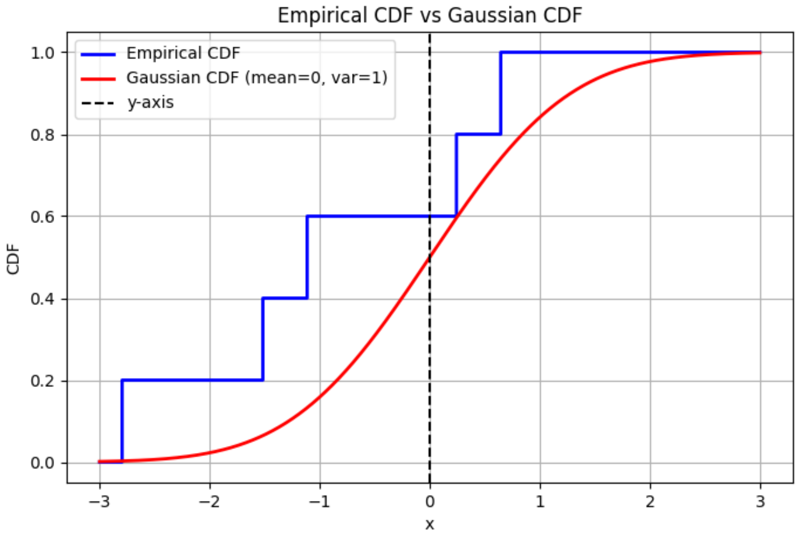
We're interested in estimating a 1-parameter distribution from a sample, along with some measure of confidence. Let our sample be $X_1, \cdots, X_n \overset{\text{i.i.d.}}{\sim} P$, with $x_i \in \mathbb{R}$. Let $F : \mathbb{R} \to [0,1]$ be the cumulative density function. We introduce the notion of an empirical CDF.

Definition 1 (Empirical CDF)**.**

$$\hat{F}_+ = \frac{1}{n} \sum_{i=1}^{n} 1_{\{t \geq x_i\}}$$

Note that this is a CDF: it is monotonically increasing and tends towards 0 at $-\infty$ and to 1 at ∞. We place a mass of $1/n$ on all observed x.

Example 2 (Gaussian data)**.** *Let $P = N(0,1)$. Then $F = \Phi$. Let the sample be* $(X_1, \cdots, X_5) = (-1.12, .64, -1.52, .24, -2.8)$.



How does this compare to the actual Gaussian CDF?

Not very close. (The data is slightly skewed to the left compared to the actual distribution).

Note that, for fixed $t \in \mathbb{R}$, $\hat{F}(t) \xrightarrow{p} F(t)$ by the law of large numbers. However, we have the stronger result that it converges uniformly for all points in $\mathbb{R}$.

Theorem 3 (Glivenko-Cantelli). *For any P, as $n \to \infty$,*

$$\sup_t |\hat{F}(t) - F(t)| \xrightarrow{p} 0.$$

Proof. The main idea is that we can prove that the functions converge on a finite grid, and using interpolation we can show that the deviations between the functions between the grid points must be bounded. The main feature that allows this is the monotonicirt of the CDFs.

Take F continuous without loss of generality. Let $t_1, \cdots, t_n$ be points such that $F(t_n) = \frac{k}{n+1}$. We can use pointwise convergence to show that, for any $\epsilon, \delta > 0$, there exists an N_k such that $P(|\hat{F}(t_k) - F(t_k)| > \epsilon) < \frac{\delta}{m}$. Since this is a finite set, we can choose $N = \max(N_k)$, which implies that for $n > N$,

$$\begin{aligned} p(\max_k |\hat{F}(t_k) - F(t_k)| \geq \epsilon) &\leq \sum_{k=1}^{m} P(|\hat{F}(t_k) - F(t_k)| \geq \epsilon) \\ &\leq \delta, \end{aligned}$$

where we used the union bound.

Now, let us bound the points in between the grid points. For any t, let k be such that $t_{k-1} \leq t \leq t_k$. Then,

$$\begin{aligned} |\hat{F}(t) - F(t) &\leq \max\{\hat{F}(t_k) - F(t), F(t) - \hat{F}(t_{k-1})\} \\ &\leq \max\{\hat{F}(t_k) - F(t_k) - (F(t) - F(t_k)), -\hat{F}(t_{k-1}) - F(t_{k-1}) - (F(t) - F(t_{k-1})\}. \end{aligned}$$

Since both of the arguments of the max deviate from $|\hat{F}(t) - F(t)|$ by no less than $\frac{1}{m+1}$, we have that, for $n > N$,

$$p\left(\sup_t |\hat{F}(t) - F(t)| > \epsilon + \frac{1}{m+1}\right) \leq \delta.$$

This completes the proof, since ϵ, δ, m are all arbitrary. □

We are now going to state, but not prove, a more quantitative version of this theorem.

3 DKW Inequality

This inequality gives a lower bound for how fast $\hat{F}$ converges to F, which is indeed very quick. Recall that, via Hoeffding's inequality,

$$P(|\hat{F}(t) - F(t)| > \epsilon) \leq 2\exp(-2n\epsilon^2).$$

It turns out, that this bound works for all points simultaneously.

Theorem 4 (Dvoretzky-Kiefer-Wolfowitz)**.** *For any distribution P,*

$$P(\sup_t |\hat{F}(t) - F(t)| > \epsilon) \leq 2\exp(-2n\epsilon^2).$$

The proof is a tour-de-force of probability theory, and was proved just 35 years ago, by Massart. We will prove a weaker version of this theorem in Homework 5.

This provides a $1-\delta$ coverage-level confidence band, $(\hat{F}(t) - \sqrt{\frac{log(2/\delta)}{2n}}, \hat{F}(t) + \sqrt{\frac{log(2/\delta)}{2n}})$. However, this is a quite loose confidence band, and too wide away from the median. We can try to improve these shortcomings by a simulation approach.

4 Inference by simulation

Proposition 5. *The distribution of $M := \sup_t |\hat{F}(t) - F(t)|$ is the same for any continuous F. We say this quantity is **pivotal**, which means its distribution doesn't depend on the true underlying distribution.*

We omit the proof, but the idea behind it is that stretching or compressing the x-axis doesn't change M. We can transform the x axis such that the true cdf is that of a uniform distribution in a reversible manner, which implies that we can transform any cdf with the same M to any other. We can now describe inference via simulation:

1. Fix $n > 0$
2. For $j = 1, \cdots, N_{sim}$, let $U_i \overset{\text{i.i.d.}}{\sim} unif[0,1]$ for $i = 1, \cdots, n$.
3. Form $\hat{G}$, which is the empirical CDF of u_i.
4. Compute $W_j = \sup_{t\in[0,1]} |\hat{G}(t) - t|$.

Let ϵ_{sim} be the $1 - \delta$ quantile of $W_1, \cdots W_{N_sim}$. Then $\hat{F}(t) \pm \epsilon_{sim}$ is a $1 - \delta$ confidence band. Note that the choice of uniform distribution was arbitrary, and any other distribution would work just as well. This is essentially the Kolmogorov-Smirnov test.