Some Basic Limit Laws in Statistics

Weak Law of Large Numbers (WLLN)

Let $X_1, X_2, \ldots \overset{IID}{\sim} X_1$ and $E(X_1)$ exists, then the sample mean $\overline{X}_n$ converges in probability to $E(X_1)$

$$\overline{X}_n \xrightarrow{P} E(X_1)$$

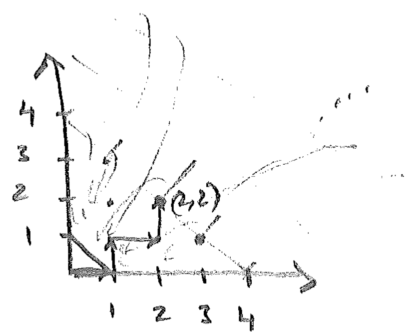
Proof of This is beyond current scope. However we can get an appreciation for a special case of the WLLN.

Ex Bernoulli WLLN & Galton's Quincunx

You should understand the meaning of :

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \theta \quad , \text{ where}$$

$$X_1, X_2, \ldots, X_n \overset{IID}{\sim} \text{Bernoulli}(\theta, 1-\theta) \ R\vec{V}s.$$



specifically, let $\theta = \frac{1}{2}$, $n = 4$.

$$P(X_1+X_2+X_3+X_4 = (2,2)) = \binom{4}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2$$
$$= \frac{{}^{2}\cancel{4}\times 3\times\cancel{2}}{\cancel{2}\times\cancel{2}}\left(\frac{1}{2}\right)^4$$
$$= \frac{3}{8}$$

Similarly find

$$P\left(\sum_{i=1}^{4} X_i = (x, 4-x)\right) = \ldots$$

Then, project it down to the unit simplex (The line connecting (1,0) and (0,1)) when dividing by 4. Now let n = 5, 6, etc & imagine...

Central Limit Theorem (CLT)

Let $X_1, X_2, \cdots \overset{IID}{\sim} X_1$ and suppose $E(X_1)$ and $V(X_1)$ exists,

then

$$\overline{X}_n \rightsquigarrow \text{Normal}\left(E(X_1), \frac{V(X_1)}{n}\right)$$

$$Z_n := \frac{\overline{X}_n - E(\overline{X}_n)}{\sqrt{V(\overline{X}_n)}} = \frac{\sqrt{n}\left(\overline{X}_n - E(X_1)\right)}{\sqrt{V(X_1)}} \rightsquigarrow Z \sim \text{Normal}(0,1)$$

CLT says that probability statements about $\overline{X}_n$ can be approximated using the Normal distribution.

Exercise:

What does CLT mean in the case of

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \rightsquigarrow \text{Normal}\left(E(X_1), \frac{V(X_1)}{n}\right), \text{ where}$$

$X_1, X_2, \ldots, X_n \overset{IID}{\sim} \text{Bernoulli}(\theta)$ RV

Try visualising with the Quincunx ...

Armed with these limit theorems we can start looking at basic estimation problems

Estimation ("big picture")

	Point Estimation	Set Estimation
parametric	MLE of $\boxed{\text{DONE}}$ finitely many parameters	confidence intervals (via CLT).
Nonparametric (∞-many parameters)	About to see..	About to see...

Nonparametric Estimation

So far we have seen parametric models.

for example,

$$X_1, X_2, \cdots, X_n \overset{IID}{\sim} \text{Bernoulli}(\theta^*), \quad \theta^* \in [0,1] \subseteq \mathbb{R}^1$$

$$X_1, X_2, \cdots, X_n \overset{IID}{\sim} \text{Exponential}(\lambda^*), \quad \lambda^* \in (0,\infty) \subseteq \mathbb{R}^1$$

$$X_1, X_2, \cdots, X_n \overset{IID}{\sim} \text{Normal}(\mu^*, \sigma^*), \quad \mu^* \in \mathbb{R}^1$$

$$\sigma^{*} \in (0,\infty)$$

$$\text{so, } (\mu^*, \sigma^{*}) \subseteq \mathbb{R}^2$$

In all these cases, the parameter space is finite dimensional, i.e. the parameter space is a subset of $\mathbb{R}^k$, with $k < \infty$.

For parametric experiments we may use the likelihood principle and estimate the parameter(s) using the MLE.

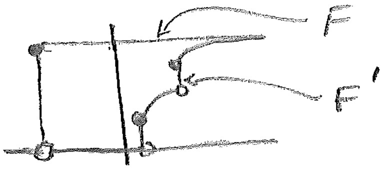
For example, for the product Bernoulli(θ^*) experiment MLE $\hat{\theta}_n = \sum_{i=1}^{n} x_i / n$ and for the product Exponential (λ^*) experiment MLE $\hat{\lambda}_n = 1 / \sum_{i=1}^{n} x_i$, etc.

Consider the following nonparametric experiment:

$$X_1, X_2, \dots, X_n \overset{\text{IID}}{\sim} F^* \in \{\text{all DFs}\}.$$

We want to produce a point estimate for F^* which is allowed to be any DF, i.e. $F^* \in \{\text{all DFs}\}$.

Note: $\{\text{all DFs}\}$ is infinite dimensional.



Two DFs in $\{\text{all DFs}\}$.

Such Estimation is possible in infinite dimensional contexts due to the following two theorems.

Glivenko-Cantelli Theorem ("Fundamental Theorem of Statistics")

Let $X_1, X_2, \dots, X_n \overset{\text{IID}}{\sim} F^* \in \{\text{all DFs}\}$

Then,

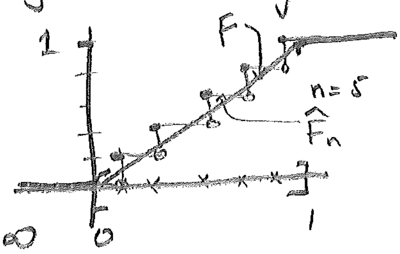
$$\sup_x \left| \hat{F}_n(x) - F^*(x) \right| \xrightarrow{P} 0$$

where, The Empirical Distribution Function (EDF) is

$$\hat{F}_n(x) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(x_i \le x), \quad \mathbb{1}(x_i \le x) = \begin{cases} 1 & \text{if } x_i \le x \\ 0 & \text{otherwise} \end{cases}$$

—x—

The proof is beyond current scope. But we can get an appreciation for the statement by considering the following simulation experiment:



Let $X_1, \dots, X_n \overset{\text{IID}}{\sim} \text{Uniform}(0,1)$

Then $\hat{F}_n(x)$ starts "wiggling" closer to the DF of Uniform(0,1) RV as $n \to \infty$

From the simulation experiment with the Uniform(0,1) product experiment it is clear that

$$\sup_x \left| \hat{F}_n(x) - F^*(x) \right| \xrightarrow{P} 0$$

This phenomenon will hold no matter what the possibly unknown F^* is.
Thus, $\hat{F}_n$ is a point estimate of F^*.

We need the following Theorem to get confidence sets or confidence band that "traps" F^* with a high probability.

<u>Dvoretzky-Kiefer-Wolfowitz (DKW) Inequality</u>

Let $X_1, X_2, \ldots, X_n \overset{IID}{\sim} F^* \in \{\text{all DFs}\}$

Then, for any $\varepsilon > 0$:

$$P\left(\sup_x \left| \hat{F}_n(x) - F^*(x) \right| > \varepsilon \right) \leq 2\exp\left(-2n\varepsilon^2\right)$$

This inequality gives us an $1-\alpha$ <u>confidence band</u>

$C_n(x) := \left[\underline{C}_n(x), \overline{C}_n(x)\right]$ about our point estimate $\hat{F}_n$ of our possibly unknown F^* such that the F^* is "trapped" by the band with probability at least $1-\alpha$. Here,

$$\underline{C}_n = \max\{\hat{F}_n(x) - \varepsilon_n, 0\}$$
$$\overline{C}_n = \min\{\hat{F}_n(x) + \varepsilon_n, 1\}, \quad \varepsilon_n = \sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)}$$

and $P\left(\underline{C}_n(x) \leq F^*(x) \leq \overline{C}_n(x)\right) \geq 1-\alpha$.

Applications

Ex 1 We can model the time between Earth-quakes or inter-earthquake times as:

$$X_1, X_2, \ldots, X_n \overset{IID}{\sim} F^* \in \{\text{all DFs}\}$$

and estimate F^* using $\hat{F}_n$ and also build a 95% confidence band. See this week's SAGE LAB.

Ex 2 We can also apply it to estimate the DF of Dosinia anus (Coarse Venus shell) diameters on one or both sides of the New Brighton Pier.

$$X_1, X_2, \ldots, X_{136} \overset{IID}{\sim} F^*$$

data from left of pier

$$Y_1, Y_2, \ldots, Y_{114} \overset{IID}{\sim} G^*$$

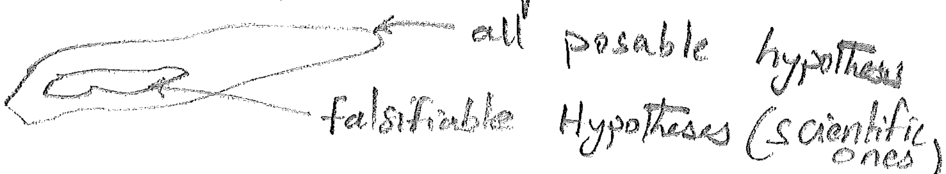
data from right of pier.

— x —

Trailer to hypothesis testing.

Suppose we are not interested in estimating F^* or G^* in the Shell Diameters problem (Ex 2 above) per se but rather in "scientifically investigating" if the distributions are the same on either side of the pier.

Then, we should attempt to reject a falsifiable hypothesis.



According to Popper's demarcation of science from nonscience via the criterion of falsifiability, we are interested in our empirical (data-based) attempt to falsify or reject the null hypothesis or H_0 in many scientific investigations against an Alternative Hypothesis H_1.

Recall Ex 2

Null Hypothesis $H_0 : F^* = G^*$, $X_1, X_2, \ldots, X_{136}, Y_1, Y_2, \ldots, Y_{114} \overset{IID}{\sim} F^* = G^*$

$H_1 : F^* \neq G^*$

In words, H_0 here means "the diameter of coarse venus shells is distributed identically on the left and right sides of the pier".

Note: That this is a falsifiable H_0 because we can generate data to attempt to falsify or reject it.

Another Example:

H_0 : The Average waiting time at The Orbiter bus stop is equal to 10 minutes.

H_1 : The Average waiting time is not 10 minutes.

Are you convinced that H_0 is falsifiable?