Asymptotics

- Sampling Model
- Limit theorems
  - (weak) Law of large numbers
  - Central limit theorem

- Pollard "User's Guide to Measure-theoretic probability"

- Probability model: a collection of probability measures $\{\mathbb{P}_\theta : \theta \in \Theta\}$ on $\{\mathcal{X}, \mathcal{X}\}$

  θ — parameter; Θ — parameter space

- Data set of size N: $\{X_i\}_{i=1}^N$ is a collection of numbers modeled by a probability model in $\mathcal{X}^N$

- Say $\{X_i\}_{i=1}^N$ is a sample from $\mathbb{P}_\theta$ if its distribution is $\mathbb{P}_\theta^N$, that is they are independent and identically distributed according to $\mathbb{P}_\theta$

$$\mathbb{E} f(X_i) g(X_j) h(X_k) = \mathbb{E} f(X_i) \mathbb{E} g(X_j) \mathbb{E} h(X_k)$$

for any i, j, k and $\mathbb{E}$ is expectation associated with $\mathbb{P}_\theta$

(actually all orders at product)

- Technicality: X_i is on $(\mathbb{X}, \mathcal{X}, \mathbb{P}_\theta)$

  $\{X_i\}_{i=1}^N$ $(\mathbb{X}^N, \mathcal{X}^N, \mathbb{P}_\theta^N)$ product space

  $1\{X_i \in A_i, X_j \in B_j\} \in \mathcal{X}^N$ for any $A_i, B_j \in \mathcal{X}$

- A statistic: a function $T_N: \mathbb{X}^N \to T$ mapping data to a space

  - if data are drawn from a probability distribution, then $T_N(\{X_i\}_{i=1}^N)$ has a distribution called the sampling distribution

- Sampling distribution often hard to calculate so we approximate it by taking limits as $N \to \infty$

  - Asymptotics embeds probability model and our data into a sequence of models with increasing N

    - Approach 1: for each N, new probability space but with the same properties

      $\{X_i\}_{i=1}^{N+1}$ lives on $(\mathbb{X}^{N+1}, \mathcal{X}^{N+1}, \mathbb{P}_\theta^{N+1})$

    - Approach 2: All on one infinite space

      $\{X_i\}_{i=1}^\infty$ $(\mathbb{X}^\infty, \mathcal{X}^\infty, \mathbb{P}_\theta^\infty)$

Weak Law of Large numbers

- Have a probability model with a probability measure with parameter $\theta_0 \in \Theta$
- Compute a statistic based on sample of size N
- Call θ_N
- Want to know whether $\theta_N \approx \theta_0$
  in particular, whether θ_N will approach θ_0 as $N \to \infty$
- Maybe also want to know how quickly
- Many definitions of convergence
- Start with convergence in probability

Def: θ_N converges in probability to θ_0

$$\theta_N \xrightarrow{p} \theta_0 \quad \text{or} \quad \operatorname{plim} \theta_N = \theta_0$$

if
$$\lim_{N \to \infty} Pr\left(|\theta_N - \theta_0| \geq \varepsilon \right) = 0$$
for all $\varepsilon > 0$

θ is an element of a normed vector space

- We will apply this to a particular statistic:
  Sample average: $T_N\left(\{X_i\}_{i=1}^N\right) = \frac{1}{N}\sum_{i=1}^N X_i$

Weak Law of Large numbers: under some conditions, $\frac{1}{N}\sum_{i=1}^N X_i \xrightarrow{p} \mathbb{E} X_i$

Preliminary result: Chebyshev's inequality

Let X with expected value μ and variance σ^2

$$\Pr(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2} \quad \text{for any } k>0$$

Markov's inequality: Let X be a nonnegative random variable and $a>0$. Then $\Pr(X \geq a) \leq \frac{E(X)}{a}$

<u>Proof</u>: $a \mathbb{1}\{X(\omega) \geq a\} \leq X(\omega)$ for all $\omega \in \Omega$

So $\quad E\, a\, \mathbb{1}\{X \geq a\} \leq E X$

$a \Pr\{X \geq a\} \leq E X$

$\Pr(X \geq a) \leq \frac{E X}{a}$

<u>Proof of Chebyshev's</u>: Let $Y = (X-\mu)^2$, $a = (k\sigma)^2$

where $E[X] = \mu$, $Var(X) = \sigma^2$

Then $\Pr(|X-\mu| \geq k\sigma) = \Pr(|X-\mu|^2 \geq (k\sigma)^2)$

by Markov $\leq \frac{E(X-\mu)^2}{(k\sigma)^2} = \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}$

Apply Chebyshev to prove weak $\underbrace{\text{law of large numbers}}_{\text{LLN}}$

- Assume we have a sequence $\{X_i\}_{i=1}^{N}$ iid with mean μ $\left(\mathbb{E}X_i = \mu\right)$ and $\text{Var}(X_i) = \sigma^2 < \infty$
- Statistic is sample average: $\bar{X}_N = \frac{1}{N}\sum_{i=1}^{N} X_i$
- We have that $\mathbb{E}[\bar{X}_N] = \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N} X_i\right]$

$$= \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}[X_i] = \frac{1}{N}\sum_{i=1}^{N} \mu = \mu$$

- Can also show $\text{Var}(\bar{X}_N) = \mathbb{E}[(\bar{X}_N - \mu)^2]$

$$= \frac{\sigma^2}{N}$$

- Apply Chebyshev to $\bar{X}_N$

$$\Pr\left(|\bar{X}_N - \mu| \geq k \underbrace{\frac{\sigma}{\sqrt{N}}}_{\text{standard deviation}}\right) \leq \frac{1}{k^2} \quad \text{for any } k > 0$$

- Let $k = \frac{\sqrt{N}\varepsilon}{\sigma}$ then

$$\Pr\left(|\bar{X}_N - \mu| \geq \varepsilon\right) \leq \frac{\sigma^2}{N\varepsilon^2}$$

so. $\Pr\left(|\bar{X}_N - \mu| < \varepsilon\right) = 1 - \Pr\left(|\bar{X}_N - \mu| \geq \varepsilon\right) \geq 1 - \frac{\sigma^2}{N\varepsilon^2}$

$\to 1$ as $N \to \infty$

therefore, since this applies for any $\varepsilon > 0$, $\bar{X}_N \xrightarrow{P} \mu$ as $N \to \infty$

Can show stronger, related results

- Finite variance: $\sigma^2 < \infty$ is sufficient but not necessary
- If X_i iid with $EX_i = M$ finite, then $\bar{X}_N \xrightarrow{p} X$
- That is the "weak law of large numbers"
- There is a strong law of large numbers
  - If X_i iid and $EX_i = M < \infty$ $(\mathcal{X}^\infty, \mathcal{X}^\infty, P^\infty)$

$$Pr(\bar{X}_N \to M) = \underline{1}$$

This is called "Almost sure convergence", denoted $\bar{X}_N \xrightarrow{as} M$

- Versions exist where $\{X_i\}_{i=1}^N$ are independent but not identically distributed by a variant of the Chebyshev proof
- Also holds for variables that are only "approximately independent". Precise statement depends on def of "approximately independent"
  - Time series: "Ergodic theorem"

- Let $g: \mathbb{R}^K \to \mathbb{R}^J$ be a function continuous at $C \in \mathbb{R}^K$
- Let $\{X_N\}_{N=1}^\infty$ a sequence of $K \times 1$ random vectors

  such that $X_N \xrightarrow{p} C$

- Then $g(X_N) \xrightarrow{p} g(C)$ as $N \to \infty$

  ie $X_N \xrightarrow{p} C \Rightarrow g(X_N) \xrightarrow{p} g(C)$ $\quad$ $\operatorname{plim} g(X_N) = g(\operatorname{plim} X_n)$

Proof: use def of continuity at x

$\forall \varepsilon > 0$, $\exists \delta > 0$ s.t. if $|x_N - x| < \delta$

then $|g(x_N) - g(x)| < \varepsilon$

- So for all $\varepsilon > 0$, $\exists \delta > 0$ s.t.

$$1 \geq \Pr(|g(x_N) - g(x)| < \varepsilon)$$

$$\geq \Pr(|x_N - x| < \delta) \to 1 \text{ as } N \to \infty$$

$$\Pr(|g(x_N) - g(x)| < \varepsilon) \to 1$$

or

$$\operatorname{plim} g(x_N) = g(x)$$

- Implications: adding, multiplying, dividing (if plim is not 0), etc., all preserve convergence in probability