2. In probability: $X_n \xrightarrow{\text{P}} X$

$$(\forall \varepsilon > 0) \lim_{n\to\infty} \mathbb{P}\left[|X_n - X| > \varepsilon\right] = 0$$

3. Almost surely (strongly): $X_n \xrightarrow{\text{as}} X$

$$\mathbb{P}\left[\lim_{n\to\infty} X_n = X\right] = \mathbb{P}\left[\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right] = 1$$

4. In quadratic mean (L_2): $X_n \xrightarrow{\text{qm}} X$

$$\lim_{n\to\infty} \mathbb{E}\left[(X_n - X)^2\right] = 0$$

Relationships

- $X_n \xrightarrow{\text{qm}} X \implies X_n \xrightarrow{\text{P}} X \implies X_n \xrightarrow{\text{D}} X$
- $X_n \xrightarrow{\text{as}} X \implies X_n \xrightarrow{\text{P}} X$
- $X_n \xrightarrow{\text{D}} X \wedge (\exists c \in \mathbb{R})\, \mathbb{P}\left[X = c\right] = 1 \implies X_n \xrightarrow{\text{P}} X$
- $X_n \xrightarrow{\text{P}} X \wedge Y_n \xrightarrow{\text{P}} Y \implies X_n + Y_n \xrightarrow{\text{P}} X + Y$
- $X_n \xrightarrow{\text{qm}} X \wedge Y_n \xrightarrow{\text{qm}} Y \implies X_n + Y_n \xrightarrow{\text{qm}} X + Y$
- $X_n \xrightarrow{\text{P}} X \wedge Y_n \xrightarrow{\text{P}} Y \implies X_n Y_n \xrightarrow{\text{P}} XY$
- $X_n \xrightarrow{\text{P}} X \implies \varphi(X_n) \xrightarrow{\text{P}} \varphi(X)$
- $X_n \xrightarrow{\text{D}} X \implies \varphi(X_n) \xrightarrow{\text{D}} \varphi(X)$
- $X_n \xrightarrow{\text{qm}} b \iff \lim_{n\to\infty} \mathbb{E}\left[X_n\right] = b \wedge \lim_{n\to\infty} \mathbb{V}\left[X_n\right] = 0$
- $X_1, \ldots, X_n$ IID $\wedge\, \mathbb{E}\left[X\right] = \mu \wedge \mathbb{V}\left[X\right] < \infty \iff \bar{X}_n \xrightarrow{\text{qm}} \mu$

Slutzky's Theorem

- $X_n \xrightarrow{\text{D}} X$ and $Y_n \xrightarrow{\text{P}} c \implies X_n + Y_n \xrightarrow{\text{D}} X + c$
- $X_n \xrightarrow{\text{D}} X$ and $Y_n \xrightarrow{\text{P}} c \implies X_n Y_n \xrightarrow{\text{D}} cX$
- In general: $X_n \xrightarrow{\text{D}} X$ and $Y_n \xrightarrow{\text{D}} Y \not\Longrightarrow X_n + Y_n \xrightarrow{\text{D}} X + Y$

10.1 Law of Large Numbers (LLN)

Let $\{X_1, \ldots, X_n\}$ be a sequence of IID RV's, $\mathbb{E}\left[X_1\right] = \mu$.
Weak (WLLN)

$$\bar{X}_n \xrightarrow{\text{P}} \mu \qquad n \to \infty$$

Strong (SLLN)

$$\bar{X}_n \xrightarrow{\text{as}} \mu \qquad n \to \infty$$

10.2 Central Limit Theorem (CLT)

Let $\{X_1, \ldots, X_n\}$ be a sequence of IID RV's, $\mathbb{E}\left[X_1\right] = \mu$, and $\mathbb{V}\left[X_1\right] = \sigma^2$.

$$Z_n := \frac{\bar{X}_n - \mu}{\sqrt{\mathbb{V}\left[\bar{X}_n\right]}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{\text{D}} Z \qquad \text{where } Z \sim \mathcal{N}(0,1)$$

$$\lim_{n\to\infty} \mathbb{P}\left[Z_n \le z\right] = \Phi(z) \qquad z \in \mathbb{R}$$

CLT notations

$$\begin{aligned} Z_n &\approx \mathcal{N}(0,1) \\ \bar{X}_n &\approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \\ \bar{X}_n - \mu &\approx \mathcal{N}\left(0, \frac{\sigma^2}{n}\right) \\ \sqrt{n}(\bar{X}_n - \mu) &\approx \mathcal{N}\left(0, \sigma^2\right) \\ \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} &\approx \mathcal{N}(0,1) \end{aligned}$$

Continuity correction

$$\mathbb{P}\left[\bar{X}_n \le x\right] \approx \Phi\left(\frac{x + \frac{1}{2} - \mu}{\sigma/\sqrt{n}}\right)$$

$$\mathbb{P}\left[\bar{X}_n \ge x\right] \approx 1 - \Phi\left(\frac{x - \frac{1}{2} - \mu}{\sigma/\sqrt{n}}\right)$$

Delta method

$$Y_n \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \implies \varphi(Y_n) \approx \mathcal{N}\left(\varphi(\mu), (\varphi'(\mu))^2 \frac{\sigma^2}{n}\right)$$

11 Statistical Inference

Let $X_1, \cdots, X_n \stackrel{iid}{\sim} F$ if not otherwise noted.

11.1 Point Estimation

- Point estimator $\widehat{\theta}_n$ of θ is a RV: $\widehat{\theta}_n = g(X_1, \ldots, X_n)$
- $\mathsf{bias}(\widehat{\theta}_n) = \mathbb{E}\left[\widehat{\theta}_n\right] - \theta$
- Consistency: $\widehat{\theta}_n \xrightarrow{\text{P}} \theta$
- Sampling distribution: $F(\widehat{\theta}_n)$
- Standard error: $\mathsf{se}(\widehat{\theta}_n) = \sqrt{\mathbb{V}\left[\widehat{\theta}_n\right]}$

- Mean squared error: $\text{MSE} = \mathbb{E}\left[(\widehat{\theta}_n - \theta)^2\right] = \mathsf{bias}(\widehat{\theta}_n)^2 + \mathbb{V}\left[\widehat{\theta}_n\right]$
- $\lim_{n\to\infty} \mathsf{bias}(\widehat{\theta}_n) = 0 \wedge \lim_{n\to\infty} \mathsf{se}(\widehat{\theta}_n) = 0 \implies \widehat{\theta}_n$ is consistent
- Asymptotic normality: $\dfrac{\widehat{\theta}_n - \theta}{\mathsf{se}} \xrightarrow{\text{D}} \mathcal{N}(0,1)$
- Slutzky's Theorem often lets us replace $\mathsf{se}(\widehat{\theta}_n)$ by some (weakly) consistent estimator $\widehat{\sigma}_n$.

11.2 Normal-Based Confidence Interval

Suppose $\widehat{\theta}_n \approx \mathcal{N}\left(\theta, \widehat{\mathsf{se}}^2\right)$. Let $z_{\alpha/2} = \Phi^{-1}(1-(\alpha/2))$, i.e., $\mathbb{P}\left[Z > z_{\alpha/2}\right] = \alpha/2$ and $\mathbb{P}\left[-z_{\alpha/2} < Z < z_{\alpha/2}\right] = 1-\alpha$ where $Z \sim \mathcal{N}(0,1)$. Then

$$C_n = \widehat{\theta}_n \pm z_{\alpha/2}\widehat{\mathsf{se}}$$

11.3 Empirical distribution

Empirical Distribution Function (ECDF)

$$\widehat{F}_n(x) = \frac{\sum_{i=1}^n I(X_i \le x)}{n}$$

$$I(X_i \le x) = \begin{cases} 1 & X_i \le x \\ 0 & X_i > x \end{cases}$$

Properties (for any fixed x)

- $\mathbb{E}\left[\widehat{F}_n\right] = F(x)$
- $\mathbb{V}\left[\widehat{F}_n\right] = \dfrac{F(x)(1-F(x))}{n}$
- $\text{MSE} = \dfrac{F(x)(1-F(x))}{n} \xrightarrow{\text{D}} 0$
- $\widehat{F}_n \xrightarrow{\text{P}} F(x)$

Dvoretzky-Kiefer-Wolfowitz (DKW) inequality ($X_1, \dots, X_n \sim F$)

$$\mathbb{P}\left[\sup_x \left|F(x) - \widehat{F}_n(x)\right| > \varepsilon\right] = 2e^{-2n\varepsilon^2}$$

Nonparametric $1-\alpha$ confidence band for F

$$L(x) = \max\{\widehat{F}_n - \epsilon_n, 0\}$$
$$U(x) = \min\{\widehat{F}_n + \epsilon_n, 1\}$$
$$\epsilon = \sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)}$$

$$\mathbb{P}\left[L(x) \le F(x) \le U(x)\ \forall x\right] \ge 1-\alpha$$

11.4 Statistical Functionals

- Statistical functional: $T(F)$
- Plug-in estimator of $\theta = (F)$: $\widehat{\theta}_n = T(\widehat{F}_n)$
- Linear functional: $T(F) = \int \varphi(x)\,dF_X(x)$
- Plug-in estimator for linear functional:

$$T(\widehat{F}_n) = \int \varphi(x)\,d\widehat{F}_n(x) = \frac{1}{n}\sum_{i=1}^n \varphi(X_i)$$

- Often: $T(\widehat{F}_n) \approx \mathcal{N}\left(T(F), \widehat{\mathsf{se}}^2\right) \implies T(\widehat{F}_n) \pm z_{\alpha/2}\widehat{\mathsf{se}}$
- p^{th} quantile: $F^{-1}(p) = \inf\{x : F(x) \ge p\}$
- $\widehat{\mu} = \bar{X}_n$
- $\widehat{\sigma}^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^n (X_i - \bar{X}_n)^2$
- $\widehat{\kappa} = \dfrac{\frac{1}{n}\sum_{i=1}^n (X_i - \widehat{\mu})^3}{\widehat{\sigma}^3}$
- $\widehat{\rho} = \dfrac{\sum_{i=1}^n (X_i - \bar{X}_n)(Y_i - \bar{Y}_n)}{\sqrt{\sum_{i=1}^n (X_i - \bar{X}_n)^2}\sqrt{\sum_{i=1}^n (Y_i - \bar{Y}_n)^2}}$

12 Parametric Inference

Let $\mathfrak{F} = \left\{f(x;\theta) : \theta \in \Theta\right\}$ be a parametric model with parameter space $\Theta \subset \mathbb{R}^k$ and parameter $\theta = (\theta_1, \dots, \theta_k)$.

12.1 Method of Moments

j^{th} moment

$$\alpha_j(\theta) = \mathbb{E}\left[X^j\right] = \int x^j\,dF_X(x)$$

j^{th} sample moment

$$\widehat{\alpha}_j = \frac{1}{n}\sum_{i=1}^n X_i^j$$

Method of Moments estimator (MoM)

$$\alpha_1(\theta) = \widehat{\alpha}_1$$
$$\alpha_2(\theta) = \widehat{\alpha}_2$$
$$\vdots = \vdots$$
$$\alpha_k(\theta) = \widehat{\alpha}_k$$