a statistical model $\mathcal{F}$ is a set of distribution functions

a parametric model is a statistical model parametrized by a finite # of parameters

$$\mathcal{F} = \{f_\theta : \theta \in \Theta\}$$

a point estimate is a "best guess" of some parameter of interest, based on observed data

Ex. Let X be output of t-step random walk with (unknown) parameter p

$$X = \sum_{i=1}^{t} Y_i \qquad Y_i = \begin{cases} +1 & \text{wp } 1-p \\ -1 & \text{wp } p \end{cases}$$

$X_1, X_2, \ldots, X_n$ iid copies of X

$$\hat{p}_n = \frac{1 - \frac{1}{tn}\sum_{i=1}^{n} X_i}{2} \quad \text{pt. est. for } p$$

$$\mathbb{E}[\hat{p}_n] = \frac{1 - \frac{1}{tn}\sum_{i=1}^{n}\mathbb{E}[X_i]}{2} = \frac{1 - \frac{1}{tn}\sum_{i=1}^{n} t(1-2p)}{2} = p$$

$\text{bias}^2 = 0$

$$se_n = \sqrt{\mathbb{V}[\hat{p}_1]} = \sqrt{\frac{p(1-p)}{nt}}$$

$$\mathbb{V}[\hat{p}_1] = \mathbb{V}\left[\frac{1}{2} - \frac{1}{2tn}\sum_{i=1}^{n} X_i\right]$$

$$= \frac{1}{4t^2n^2}\sum_{i=1}^{n}\mathbb{V}[X_i] \quad \text{HW2 \#3}$$

$$= \frac{1}{4t^2n^2}\sum_{i=1}^{n} 4tp(1-p)$$

$$= \frac{1}{nt}p(1-p)$$

$$MSE_n = \text{bias}_n^2 + se_n^2 = \sqrt{\frac{p(1-p)}{nt}}$$

Confidence Sets

Def. A $1-\alpha$ confidence interval for a parameter θ is an interval $C_n = (a_n, b_n)$, where a_n, b_n are functions of $X_1, \dots, X_n$ s.t.

$$\mathbb{P}_\theta[\theta \in C_n] \geq 1-\alpha \quad \forall \theta \in \Theta$$

C_n = random

θ = fixed

Ex. $Y = \begin{cases} \theta + 1 & \text{wp } \frac{1}{2} \\ \theta - 1 & \text{wp } \frac{1}{2} \end{cases}$ $\theta \in \mathbb{R}$ fixed, unknown parameter

Observe iid samples $Y_1, Y_2, \dots, Y_n \sim Y$

$$C_n = (\min(X_1, \dots, X_n) - \tfrac{1}{2}, \max(X_1, \dots, X_n) + \tfrac{1}{2})$$

$$\theta \in C_n \iff \max(X_1, \dots, X_n) \neq \min(X_1, \dots, X_n)$$

$\mathbb{P}[\theta \in C_n] = 1 - \frac{1}{2^{n-1}} \Rightarrow C_n$ is a $1 - \frac{1}{2^{n-1}}$ confidence interval for θ

---

if $n = 2$, $Y_1 = 15$, $Y_2 = 17$ then $C_n = (14.5, 17.5)$ but we can be 100% sure $\theta = 16$.

CI is <u>not</u> a probability statement about θ