STATISTICAL ANALYSIS OF SIMULATION DATA

STOCHASTIC MODEL M + SIMULATION ALGORITHM.

$M \longrightarrow X_{0:T}$: Trajectories $X_{0:T} : X_0, X_1, \ldots X_T$ (FOR DTMC)

sampled from $P_M(X_{0:T})$

$$\begin{matrix} X_{0:T}^{(1)}, & \ldots & X_{0:T}^{(N)} \\ \shortparallel & & \shortparallel \\ \underline{X}_1 & & \underline{X}_N \end{matrix}$$ N-independent TRAJECTORIES

$\Rightarrow \left[Y = f(X_{0:T}) \right] \Rightarrow y_1 = f(X_{0:T}^{(1)}) \ldots y_N = f(X_{0:T}^{(N)})$

GOAL $\quad \underline{\mathbb{E}[Y] = \mathbb{E}[f(X_{0:T})]}$

EXAMPLES

a) FIXED-TIME PROPERTY $\quad f(X_{0:T}) = \#$ infected at time T

b) TRAJECTORY-BASED PROPERTIES $\quad f(X_{0:T})$ = cumulative # of death individuals from 0 to T.

c) BEHAVIOURS (0-1 VALUED TRAJ. PROP.)

$f(X_{0:T}) = \begin{cases} 1 & \text{if the epidemic peak happens} \leq \bar{t} \\ 0 & \text{oth.} \end{cases}$

$X_{0:T}^{(1)}, \dots, X_{0:T}^{(N)}, \quad y_i = f(X_{0:T}^{(i)}) \in \mathbb{R} \quad Y = f(X_{0:T}) \quad \mu = E[Y]?$

$\Rightarrow y_1, \dots, y_n = \bar{y}$ i.i.d. (INDEPENDENT and IDENTICALLY DISTRIBUTED)

$$\boxed{\hat{Y}_N = \frac{1}{N}\sum_{i=1}^{N} y_i}$$

$\hat{Y}_N$ IS A RANDOM QUANTITY, seen as $\boxed{\hat{Y}_N = \frac{1}{N}\sum_{i=1}^{N} Y_i}$ $\quad Y_i \sim Y$

$$\underbrace{\mathbb{E}[\hat{Y}_N] = \mu}_{\text{UNBIASED ESTIMATOR}} \quad \left[\mathbb{E}[\hat{Y}_N] = \frac{1}{N}\mathbb{E}\left[\sum_{i=1}^{N} Y_i\right] = \frac{1}{N}\sum \mathbb{E}[Y_i] = \frac{N\mu}{N} = \mu\right]$$

X_i indep $\Rightarrow VAR[\sum X_i] = \sum VAR[X_i]$

$$VAR[\hat{Y}_N] = VAR\left[\frac{1}{N}\sum_{i} Y_i\right] = \frac{1}{N^2}\sum VAR[Y_i] = \boxed{\frac{VAR[Y]}{N}} \quad (VAR[Y] = \sigma^2)$$

$VAR[aX] = a^2 VAR[X]$

$$\underbrace{\hat{S}_N^2}_{\text{SAMPLE VARIANCE, UNBIASED}} = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i - \hat{Y}_N\right)^2 = \underbrace{\frac{1}{N-1}\sum_{i=1}^{N} Y_i^2}\ - \frac{N}{N-1}\underbrace{\hat{Y}_N^2}$$

$$\mathbb{E}[\hat{S}_N^2] = \sigma^2$$

ONLINE ESTIMATES

y: $y_1, \dots, y_N$ observed, compute

$$\hat{Y}_N = \frac{1}{N}\sum_{i=1}^{N} y_i$$

$$\tilde{Y}^2_N = \frac{1}{N-1}\sum_{i=1}^{N} y_i^2$$

$$\hat{S}^2_N = \tilde{Y}^2_N - \frac{N}{N-1}\hat{Y}_N$$

- OBSERVE y_{N+1}, then one can update previous values as:

$$\hat{Y}_{N+1} = \frac{N}{N+1}\hat{Y}_N + \frac{y_{N+1}}{N+1}$$

$$\tilde{Y}^2_{N+1} = \frac{N-1}{N}\tilde{Y}^2_N + \frac{y^2_{N+1}}{N}$$

$$\hat{Y}_N \; : \; \mathbb{E}[\hat{Y}_N] = \mu \qquad VAR[\hat{Y}_N] = \frac{VAR[Y]}{N} \approx \frac{\hat{S}_N^2}{N}$$

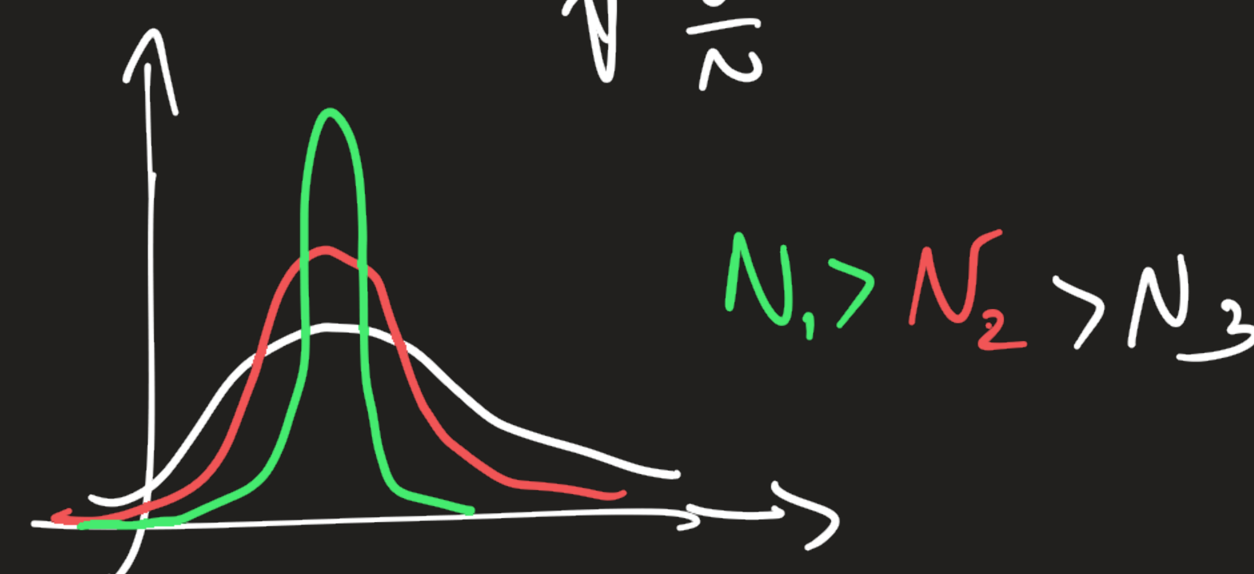
CONFIDENCE INTERVAL AT CONFIDENCE α $\quad I_N = [\underline{I}_N, \bar{I}_N]$ random variate

$$P(\mu \in I_N) \approx \alpha$$

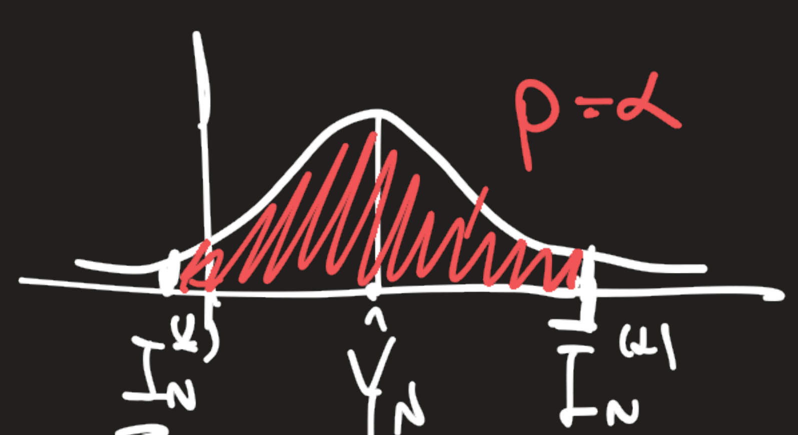
- STRONG LAW OF LARGE NUMBERS $\quad \lim_{N \to \infty} \hat{Y}_N = \mu$ a.s.
- CENTRAL LIMIT THEOREM $\quad \lim_{N \to \infty} \frac{\hat{Y}_N - \mu}{\sqrt{\frac{\sigma^2}{N}}} \to \mathcal{N}(0,1)$

$$\Downarrow$$

$$\hat{Y}_N \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{N}\right)$$



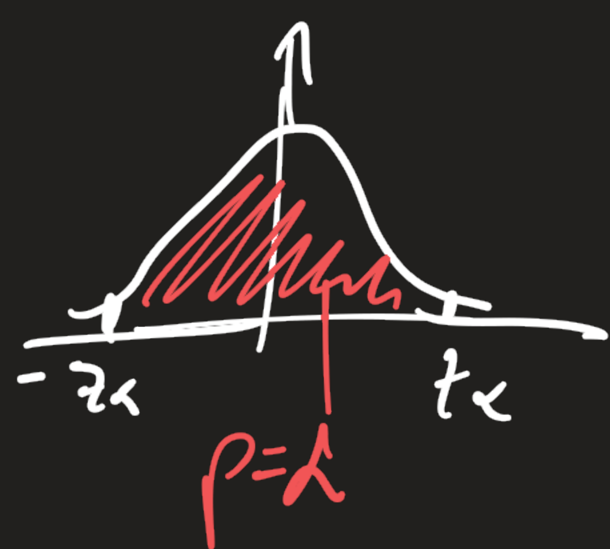
$$I_N^{(\alpha)} =$$



$$\underbrace{\mathcal{N}\left(\hat{Y}_N, \frac{\hat{S}_N^2}{N}\right) = \sqrt{\frac{\hat{S}_N^2}{N}}\,\mathcal{N}(0,1) + \hat{Y}_N}$$

FIX α $(\alpha = 0.95)$ $\quad z_\alpha = \phi^{-1}\left(\frac{\alpha+1}{2}\right)$ $\quad \phi(z) := \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-y^2/2}\,dy$

$\Updownarrow$



$$\phi(z_\alpha) - \phi(-z_\alpha) = \alpha$$

$$(\alpha = 0.95 \quad z_\alpha \approx 1.96)$$

$$I_N^{(\alpha)} = \left[\hat{Y}_N - z_\alpha\sqrt{\frac{S_N^2}{N}},\ \hat{Y}_N + z_\alpha\sqrt{\frac{S_N^2}{N}}\right]$$

$$\boxed{2z_\alpha\sqrt{\frac{S_N^2}{N}} < \varepsilon}$$ STOP CRITERION

ANALYSIS OF BEHAVIOURAL PROBABILITIES

$M: \; x_{0:T}^{(1)}, \dots, x_{0:T}^{(N)}$ $\qquad \underbrace{Z} = f(x_{0:T}) \in \{0,1\} \rightsquigarrow z_i = f(x_{0:T}^{(i)}) \in \{0,1\}$

$\underline{z} = (z_1, \dots, z_N)$

$\mathbb{E}[Z] = \underbrace{P(Z=1)} = \vartheta^*$ $\qquad Z = \text{Bernoulli}(\vartheta^*)$

$\hat{Z}_N = \frac{1}{N} \sum_{i=1}^{N} z_i = \hat{\vartheta}_N$

$\mathcal{L}(\vartheta) = \log p(\underline{z} \mid \vartheta) = \underbrace{n \log \vartheta + (N-n) \log(1-\vartheta)}$

$n = \sum_{i=1}^{N} z_i$

MAXIMUM LIKELIHOOD SETTING.

$\hat{\vartheta}_N = \arg\max \mathcal{L}(\vartheta)$

$\frac{\partial}{\partial \vartheta} \mathcal{L}(\vartheta) = 0 \implies \hat{\vartheta}_N = \frac{n}{N}$

GIVEN ACCURACY $\varepsilon > 0$ AND CONFIDENCE $\alpha \in [0,1]$ HOW MANY SAMPLES WE NEED TO HAVE

$$p\left(|\hat{\vartheta}_N - \vartheta^*| > \varepsilon\right) < 1-\alpha \; ?$$

CHERNOFF BOUND

$$p\left(|\hat{\vartheta}_N - \vartheta^*| \geq \varepsilon\right) \leq 2\exp\left(-\frac{\varepsilon^2}{2\vartheta^* + \varepsilon} N\right) \leq 1-\alpha$$

$$\boxed{N \geq \frac{2+\varepsilon}{\varepsilon^2} \log\left(\frac{2}{1-\alpha}\right)}$$

IT IS A WORST CASE BOUND

$\Downarrow$

N CAN BE MUCH LARGER THAN NEEDED.

SEQUENTIAL WALD TEST

H_0 - null hypothesis $\quad H_1$ - alternative hypothesis

	H_1 accepted	H_1 rejected
H_1 TRUE	TRUE POSITIVE TP	FALSE NEGATIVE FN
H_1 FALSE	FALSE POSITIVE FP	TRUE NEGATIVE TN

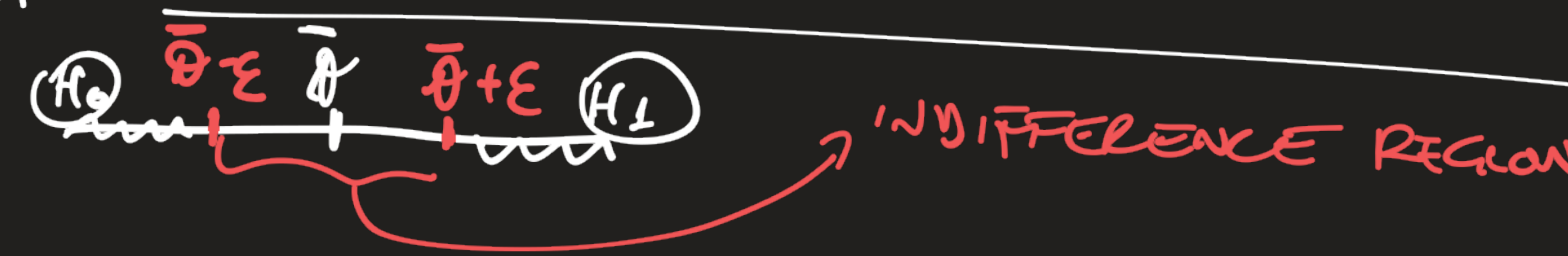
TYPE I ERROR $\leftarrow$ FP

TYPE II ERROR $\leftarrow$ FN

SIGNIFICANCE α
Prob TYPE I ERROR

POWER $1-\beta$
β Prob TYPE II ERROR

TEST $\underline{\vartheta^* \geq \bar{\vartheta} \in [0,1]}$ CONSTANT



$H_1 : \vartheta^* \geq \bar{\vartheta}$

$H_0 : \vartheta^* < \bar{\vartheta}$

FIX $\varepsilon > 0$

$H_1 : \vartheta^* \geq \bar{\vartheta}+\varepsilon$

$H_0 : \vartheta^* \leq \bar{\vartheta}-\varepsilon$

$\underline{w_n}$ TEST QUANTITY, LOG-LIKELIHOOD RATIO

$$w_n = w_{n-1} + I\{z_n=1\}\log\left(\frac{\bar{\vartheta}+\varepsilon}{\bar{\vartheta}-\varepsilon}\right) + I(z_n=0)\log\left(\frac{1-\bar{\vartheta}+\varepsilon}{1-\bar{\vartheta}-\varepsilon}\right)$$

$$w_0 = 0$$

ACCEPT
- H_0 iff $w_n \geq \log(1-\beta/\alpha)$
- H_1 iff $w_n \leq \log\left(\frac{\beta}{1-\alpha}\right)$

BAYESIAN ESTIMATION OF BEHAVIOURAL PROPERTIES

$\underline{z}: z_1, \dots, z_N \in \{0,1\}$ $\qquad$ $\vartheta^* = \text{prob}(z=1)$

$p(\vartheta)$ - PRIOR PROBABILITY ON $\vartheta \in [0,1]$

$$p(\vartheta | \underline{z}) = \frac{p(\underline{z}|\vartheta)\, p(\vartheta)}{p(\underline{z})} \propto p(\underline{z}|\vartheta)\, p(\vartheta)$$

- MAP $\quad \vartheta_N^{MAP} = \underset{\vartheta}{\operatorname{argmax}}\ p(\vartheta | \underline{z})$
- PREDICTIVE DISTRIBUTION $\quad \underbrace{\mathbb{E}[\vartheta | \underline{z}] = \int \vartheta\, p(\vartheta | \underline{z})\, d\vartheta}$

HERE

$$p(\vartheta) = \text{Beta}(\vartheta | a, b) \propto \underbrace{\vartheta^{a-1} (1-\vartheta)^{b-1}}$$

$\left[\begin{array}{l} a = b = 1 \\ \text{Beta}(\vartheta | 1,1) \simeq \text{Uniform}(0,1) \end{array} \right.$

$$p(\vartheta | \underline{z}) = \text{Beta}(\vartheta | a+n, b+N-n)$$

$n = \sum z_i$

$\Downarrow$

$$\mathbb{E}[\vartheta | \underline{z}] = \frac{a+n}{a+b+N}$$

$$p(\vartheta \in [p_1, p_2]) = \int_{p_1}^{p_2} \mathrm{Beta}(\vartheta \mid a+n, b+N-n)\, d\vartheta$$

FIX AN ERROR $\varepsilon > 0$, AND CONFIDENCE α

$$\bar{\vartheta} = \mathbb{E}[\vartheta \mid \underline{z}]$$

$$p(\vartheta \in [\bar{\vartheta} - \varepsilon, \bar{\vartheta} + \varepsilon]) > \alpha$$

use this condition in a Bayesian sequential scheme

$H_0 : \vartheta \geq p$

$H_1 \ \ \vartheta \leq p$

$p \in [0,1]$ threshold

$$B = \frac{p(\underline{z} \mid H_0)}{p(\underline{z} \mid H_1)}$$ BAYES FACTOR

$\mathrm{Beta}(\vartheta \mid a, b)$ PRIOR

$\underline{z}$ $\quad n$ successes

$\quad N-n$ failures

$$B_N = \frac{p(H_1)}{p(H_0)} \left[\left(\int_0^p \mathrm{Beta}(\vartheta \mid a+n, b+N-n)\, d\vartheta \right)^{-1} - 1 \right]$$

$$p(H_0) = \int_p^1 \mathrm{Beta}(\vartheta \mid a, b)\, d\vartheta$$

FIX $T > 1$

ACCEPT $\quad H_0$ iff $B_N > T$

$\quad H_1$ iff $B_N < \frac{1}{T}$

SIGNIF. $\frac{1}{T}$

POWER $1 - \frac{1}{T}$