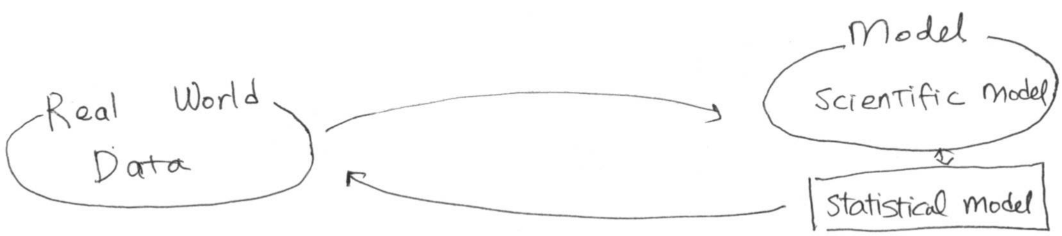
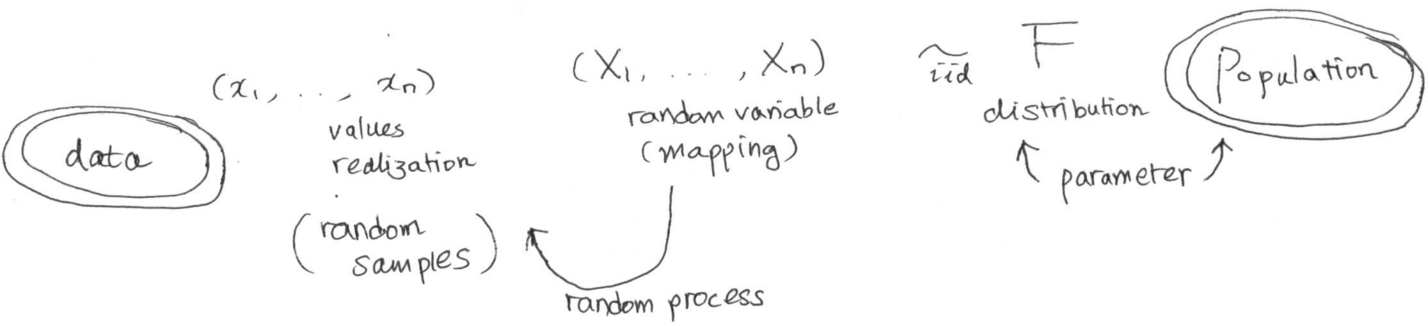
Statistical inference

Given a sample $X_1, \ldots, X_n \sim F$ how do we infer $F$?

$\mathfrak{F}$ : a statistical model, a collection of possible distribution

- parametric model $\{ f(x; \theta) : \theta \in \Theta \}$
  - $\Theta$ : parameter space
  - $\theta$ : unknown, fixed parameter
- Non-parametric model : an infinite number of parameters, distribution free.



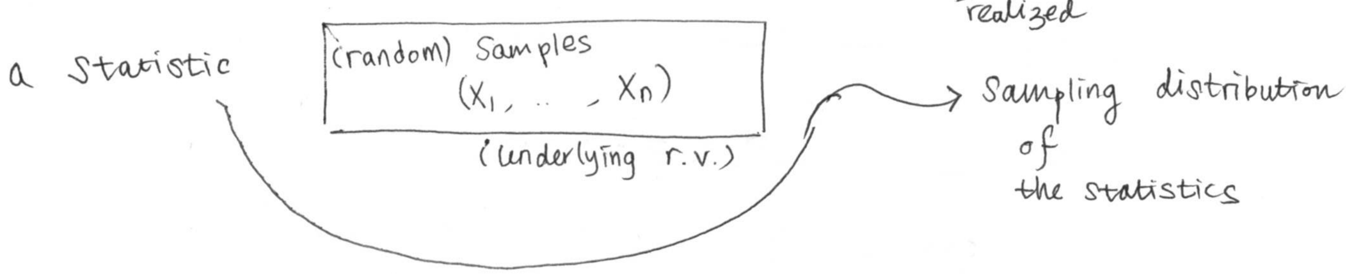

$\hat{\theta}(x_1, \ldots, x_n)$ estimate (realization)

$\hat{\theta}(X_1, \ldots, X_n)$ estimator (r.v)

$\theta$ parameter (unknown, fixed)

$\hat{\theta}(X_1, \ldots X_n)$ : estimator, random variable, a function of a sample

$\hat{\theta}(x_1, \ldots x_n)$ : estimate, a function of a sample values
realized



## properties of estimator

i) unbiased — bias of $\hat{\theta}_n = E_\theta(\hat{\theta}_n) - \theta$

ii) consistent — $\hat{\theta}_n \xrightarrow{P} \theta$

iii) asymptotic normality — $\dfrac{\hat{\theta}_n - \theta}{se(\hat{\theta}_n)} \xrightarrow{D} N(0,1)$

$$se(\hat{\theta}_n) = \sqrt{V_\theta(\hat{\theta}_n)}$$

MSE (mean squared error)

$$= E_\theta[(\hat{\theta}_n - \theta)^2]$$

$$= E_\theta[(\hat{\theta}_n - E(\hat{\theta}_n) + E(\hat{\theta}_n) - \theta)^2]$$

$$= E_\theta[(\hat{\theta}_n - E(\hat{\theta}_n))^2 + (E(\hat{\theta}_n) - \theta)^2 + 2\underbrace{(\hat{\theta}_n - E(\hat{\theta}_n))}_{0}(E(\hat{\theta}_n) - \theta)]$$

$$= E_\theta[(\hat{\theta}_n - E(\hat{\theta}_n))^2] + \text{bias}^2$$

$$= \text{variance} + \text{bias}^2$$

$$= V_\theta(\hat{\theta}_n) + \text{bias}^2(\hat{\theta}_n)$$

## Confidence interval for $\theta$

$C_n$: A $1-\alpha$ confidence interval for $\theta$

s.t $\mathbb{P}_\theta(\theta \in C_n) \geq \underbrace{1-\alpha}$ for $\forall\theta$

the coverage of the interval

## an approximate $(1-\alpha)$ confidence interval $\theta$

Suppose $\hat{\theta}_n \approx N(\theta, \hat{\sigma}_n^2)$, $z_{\alpha/2} = \Phi^{-1}(1 - \frac{\alpha}{2})$

$$C_n = (\hat{\theta}_n - z_{\alpha/2}\hat{\sigma}_n, \ \hat{\theta}_n + z_{\alpha/2}\hat{\sigma})$$

then $\mathbb{P}_\theta(\theta \in C_n) \longrightarrow 1-\alpha$

p.f) $\cdot\ Z_n = \dfrac{\hat{\theta}_n - \theta}{\hat{se}} \xrightarrow{D} Z \sim N(0,1)$

$$\mathbb{P}_\theta(\theta \in C_n) = \mathbb{P}_\theta(\hat{\theta}_n - z_{\alpha/2}\hat{se} < \theta < \hat{\theta}_n + z_{\alpha/2}\hat{se})$$

$$= \mathbb{P}_\theta\left(-z_{\alpha/2} < \frac{\hat{\theta} - \theta}{\hat{se}} < z_{\alpha/2}\right) \xrightarrow[\text{by assumption}]{} \mathbb{P}_\theta(-z_{\alpha/2} < Z < z_{\alpha}$$

$$= 1-\alpha.$$

example unbiased estimator, different MSE

$X_i \sim N(\mu, \sigma^2)$

$\hat{\theta} = X_1$

$E(\hat{\theta}) = \mu$

$Var(\hat{\theta}) = \sigma^2$

$s.e.(\hat{\theta}) = \sigma$

$\hat{\theta} = \bar{X}_n = \frac{\Sigma X_i}{n}$

$E(\hat{\theta}) = E(\Sigma \frac{X_i}{n}) = \Sigma \frac{E(X_i)}{n} = \mu$

$Var(\hat{\theta}) = E((\frac{\Sigma X_i}{n} - \mu)^2)$

$X_i = Z_i + \mu$

$Z_i \sim N(0, \sigma^2)$

$= E(\frac{\Sigma(Z_i + \mu)}{n} - \mu)^2$

$= E\left[\left(\frac{\Sigma(Z_i) + \cancel{n\mu} - \cancel{n\mu}}{n}\right)^2\right]$

$= E\left[\frac{(\Sigma Z_i)^2}{n}\right] = \frac{1}{n^2} E((\Sigma Z_i)^2)$

$= \frac{1}{n^2} E(\Sigma Z_i^2)$ ind

$= \frac{1}{n^2} n \, E(Z_1^2)$ $Var(Z) = E(Z^2) - (E(Z))^2$

$= \frac{1}{n} \cdot Var(Z_1)$

$= \frac{1}{n} \sigma^2$

$s.e(\hat{\theta}) = \frac{1}{\sqrt{n}} \sigma$

example unbiased estimator for variance $\sigma^2$

the sample variance $S_n = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$

the corrected sample variance $S_{n-1} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 = \frac{n}{n-1} S_n$ : unbiased

$E(S_n) = \mathbb{E}\left[\frac{1}{n} \Sigma (X_i - \bar{X})^2\right]$

$= \mathbb{E}\left[\frac{1}{n} \Sigma (X_i - \mu + \mu - \bar{X})^2\right]$

$= \mathbb{E}\left[\frac{1}{n} \Sigma ((X_i - \mu) - (\bar{X} - \mu))^2\right]$

$= \mathbb{E}\left[\frac{1}{n} \Sigma \left((X_i - \mu)^2 + (\bar{X} - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu)\right)\right]$

$= \frac{1}{n}\left[\sum_{i=1}^{n}(\mathbb{E}(X_i - \mu)^2) + \underset{n\times}{E(\bar{X} - \mu)^2} - 2n\,\mathbb{E}(\bar{X} - \mu)^2\right] = \frac{1}{n}\left[\sum_{i=1}^{n} \mathbb{E}(X_i - \mu)^2 - n\, E(\bar{X} - \mu)^2\right]$

$= \frac{1}{n}\left[\sum_{i=1}^{n} Var(X_i)\right] - Var(\bar{X}) = \frac{1}{n} n \cdot \sigma^2 - \frac{1}{n}\sigma^2 = (1 - \frac{1}{n})\sigma^2$

Example

Let $X_1, \ldots, X_n \overset{iid}{\sim} \text{Poisson}(\lambda)$

$$\hat{\lambda}_n = \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\text{bias ? } E(\bar{X}_n) = \frac{E(\Sigma X_i)}{n} = \frac{\Sigma E(X_i)}{n} = \frac{n \cdot \lambda}{n} = \lambda$$

$$E(\bar{X}_n) - \mu = 0$$

$$Var(\bar{X}_n) = Var\left(\frac{\Sigma X_i}{n}\right) = \frac{1}{n^2} var(\Sigma X_i) = \frac{n}{n^2} var(X_1) = \frac{1}{n}\lambda$$

$$MSE = \frac{1}{n}\lambda \longrightarrow 0 \quad \text{as } n \to \infty$$

Thm if bias $\to 0$ and s.e $\to 0$ as $n \to \infty$

then $\hat{\theta}_n$ is consistent, i.e. $\hat{\theta}_n \xrightarrow{P} \theta$

p.f) $P(|\hat{\theta}_n - \theta| > \varepsilon) \to 0$

$$MSE = bias^2 + var \longrightarrow 0 \quad \text{as } n \to \infty \quad (\text{by assumption})$$

$$\mathbb{E}((\hat{\theta} - \theta)^2) \longrightarrow 0$$

we show $\hat{\theta}_n \xrightarrow{qm} \theta$

$\therefore \hat{\theta}_n \xrightarrow{P} \theta$ by Thm 5.4

*Markov's inequality

$$P(|\hat{\theta}_n - \theta| > \varepsilon) = P(|\hat{\theta}_n - \theta| > \varepsilon) \leq \frac{E(\hat{\theta}_n - \theta)^2}{\varepsilon^2}$$

example In the above Poisson example,

$\hat{\lambda}_n = \bar{X}_n$ is consistent

example $X_1, \ldots, X_n \overset{iid}{\sim} \text{Poisson}(\lambda)$

an approximate 95% confidence interval for $\lambda$

$\hat{\theta}_n = \hat{\lambda}$, $s.e(\hat{\theta}_n) = \sqrt{\frac{\hat{\lambda}}{n}}$, where $\hat{\lambda} = \frac{\Sigma X_i}{n}$

therefore $\left(\hat{\lambda} - 1.96\sqrt{\frac{\hat{\lambda}}{n}},\ \hat{\lambda} + 1.96\sqrt{\frac{\hat{\lambda}}{n}}\right)$

example $X_1, \ldots X_n \sim \text{Bernoulli}(p)$, an approximate $1-\alpha$ confidence interval?

$\hat{p}_n = \frac{\sum X_i}{n}$

$E(\hat{p}_n) = \frac{E(\sum X_i)}{n} = p$

$var(\hat{p}_n) = \mathbb{E}\left((\hat{p}_n - p)^2\right) = \mathbb{E}\left(\left(\frac{\sum X_i}{n} - p\right)^2\right) = \frac{1}{n^2} E\left(\sum (X_i - p)^2\right) = \frac{1}{n^2} n \cdot var(X_i)$

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

$s.e(\hat{p}_n) \Leftarrow \sqrt{var(\hat{p}_n)} = \sqrt{\frac{p(1-p)}{n}}$

$\hat{s.e.}(\hat{p}_n) = \sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}$

an approximate $1-\alpha$ confidence interval

$\hat{p}_n \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}$

example $X_1, \ldots, X_n \sim \text{unif}(0, \theta)$

$\hat{\theta}_n = \max\{X_1, \ldots, X_n\}$

CDF $(\hat{\theta}_n)$ ?

$F_{\hat{\theta}_n}(x) = \mathbb{P}(\max\{X_1, \ldots X_n\} \leq x)$

$= \mathbb{P}(X_1 \leq x, \ldots, X_n \leq x)$

$= \mathbb{P}(X_1 \leq x) \times \cdots \times \mathbb{P}(X_n \leq x)$

$= \prod_{i=1}^{n} \mathbb{P}(X_i \leq x) = \left(\frac{x}{\theta}\right)^n$

p.d.f $f(x) = n \cdot \frac{x^{n-1}}{\theta^n}$

$\mathbb{E}(\hat{\theta}_n) = \int_0^\theta x \cdot n \frac{x^{n-1}}{\theta^n} = \int_0^\theta n \cdot \frac{x^n}{\theta^n} = \frac{n}{n+1}\theta$

bias $= \mathbb{E}(\hat{\theta}_n) - \theta = \frac{1}{n+1}\theta$

(* Later, we will show that MLE for $\theta$ is $\max(X_1, \ldots X_n)$. $\therefore$ MLE of $\theta$ is biased )

$Var(\hat{\theta}_n) =$ ~~$\mathbb{E}(\hat{\theta}_n - \mathbb{E}(\hat{\theta}))$~~

$= \mathbb{E}(\hat{\theta}_n^2) - (\mathbb{E}(\hat{\theta}_n))^2$

$= \int_0^\theta x^2 \, n \frac{x^{n-1}}{\theta^n} dx - \left(\frac{n}{n+1}\theta\right)^2 = \int_0^\theta \frac{n}{\theta^n} x^{n+1} dx - \left(\frac{n}{n+1}\theta\right)^2 = \left(\frac{n}{n+2}\right)\theta^2 - \left(\frac{n}{n+1}\right)^2\theta^2$

$= \frac{n(n+1)^2 - n^2(n+2)}{(n+2)(n+1)^2}\theta^2 = \frac{n}{(n+2)(n+1)^2}\theta^2$

## Empirical distribution function

$\hat{F}_n$ is the CDF that put mass $\frac{1}{n}$ at each data point $X_i$

$$= \frac{\sum_{i=1}^{n} I(X_i \le x)}{n} \quad \text{where} \quad I(X_i \le x) = \begin{cases} 1 & \text{if } X_i \le x \\ 0 & X_i > x \end{cases}$$

**Thm**

$$E(\hat{F}_n(x)) = F(x)$$

$$V(\hat{F}_n(x)) = \frac{F(x)(1-F(x))}{n}$$

$$MSE(\hat{F}_n(x)) = \frac{F(x)(1-F(x))}{n} \longrightarrow 0$$

$$\hat{F}_n(x) \xrightarrow{P} F(x)$$

pf) $E(I(X_i \le x)) = 1 \times P(X_i \le x) + 0 = P(X_i \le x) = F(x)$

$$E\left(\frac{\sum_{i=1} I(X_i \le x)}{n}\right) = F(x) \quad \text{(by linearity)}$$

$$Var\left(\frac{\sum_{i=1} I(X_i \le x)}{n}\right) = \frac{1}{n^2} n \cdot Var(I(X_1 \le x)) = \frac{1}{n}\left[E(I(X_1 \le x)^2) - E(I(X_1 \le x))^2\right]$$

$$= \frac{1}{n}\left[F(x) - F(x)^2\right] = \frac{1}{n}F(x)(1-F(x))$$

$$\hat{F}_n(x) \xrightarrow{P} F(x) \quad (\text{by Thm } 6.10)$$

**Thm** (The Glivenko- Cantelli Theorem) $X_1, \dots, X_n \sim F$

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

**Thm** (The Dvoretzky - Kiefer- Wolfowitz (DKW) inequality)

for any $\varepsilon > 0$

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

A nonparametric $1-\alpha$ Confidence band for $F$

for any $\varepsilon > 0$, $L(x) = \max\{\hat{F}_n(x) - \varepsilon_n, 0\}$

$U(x) = \min\{\hat{F}_n(x) + \varepsilon_n, 1\}$

where $\varepsilon_n = \sqrt{\frac{1}{2n}\log\left(\frac{2}{\alpha}\right)}$

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

Recall that

$\hat{\theta}_n(X_1, \dots, X_n)$ : an estimator for $\theta$

$\hat{\theta}_n = T(\hat{F}_n)$ an plug-in estimator of $\theta = T(F)$

a statistical functional

(* Lehmann E. 1998)

$\mu = \int x \, dF(x)$

$\sigma^2 = \int (x-\mu)^2 dF(x)$

median $= F^{-1}\left(\frac{1}{2}\right)$

example of a linear functional

df) if $T(F) = \int r(x)\, dF(x)$ for some function $r(x)$

Thm The plug-in estimator for linear functional $T(F) = \int r(x)\, dF(x)$ ?

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

---

Normal-based interval

in many cases, $T(\hat{F}_n) \approx N(T(F), \hat{se}^2)$

s.t. $T(\hat{F}_n) \pm z_{\alpha/2}\, \hat{se}$

<u>example</u> $X_1, \dots, X_n \overset{iid}{\sim} F$

the plug-in estimator

(the mean) $T(F) = \int x \, dF(x) \to \int x \, d\hat{F}_n(x) = \frac{\sum X_i}{n}$

the expected value of $\exp(X_1)$ $= \int \exp(x_1) \, dF(x) \to \frac{\sum \exp(X_i)}{n}$

(the variance) $T(F) = V(X) = \int x^2 dF(x) - \left( \int x \, dF(x) \right)^2$

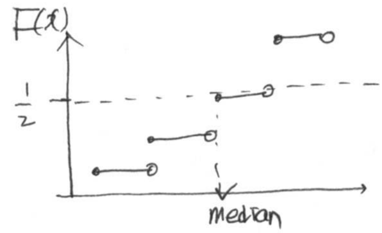
$$\to \frac{\sum X_i^2}{n} - \left( \frac{1}{n} \sum X_i \right)^2$$

$$= \frac{\sum (X_i - \bar{X}_n)^2}{n}$$

the (corrected) sample variance

$$= \frac{1}{n-1} \sum (X_i - \bar{X}_n)^2$$

(the median) $T(F) = F^{-1}(\tfrac{1}{2}) \longrightarrow \hat{F}_n^{-1}(\tfrac{1}{2}) = \inf \{ x : \hat{F}_n(x) \geq \tfrac{1}{2} \}$




<u>example</u> $X_1, \dots, X_n \sim F_1$, $Y_1, \dots, Y_n \sim F_2$

$\mu_1 = \int x \, dF_1(x)$ $\mu_2 = \int y \, dF_2(y)$

the plug-in estimator $\hat{\mu}_1 = \frac{\sum X_i}{n}$ $\hat{\mu}_2 = \frac{\sum Y_i}{n}$

s.e. of the sample mean?

$$s.e.(\hat{\mu}_1) = \sqrt{\mathbb{V}\left(\frac{1}{n} \sum X_i\right)} = \sqrt{\frac{1}{n^2} \sum_{i=1} \mathbb{V}(X_i)} = \sqrt{\frac{1}{n} \sigma^2} = \frac{\sigma}{n}$$

$\widehat{se}(\hat{\mu}) = \frac{\hat{\sigma}}{n}$ where $\hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1} (x_i - \bar{x})^2}$

$\boxed{\theta = T(F_2) - T(F_1)}$ $\hat{\mu}_1 - \hat{\mu}_2 = \frac{\sum X_i}{n} - \frac{\sum Y_i}{n} = \bar{X}_n - \bar{Y}_n$, $se(\hat{\theta}) = \sqrt{Var(\bar{X}_n) + Var(\bar{Y}_n)} = \sqrt{se(\hat{\mu}_1)^2 + se(\hat{\mu}_2)^2}$

$\widehat{se} = \sqrt{\widehat{se}(\hat{\mu}_1) + \widehat{se}(\hat{\mu}_2)}$