Stochastiek 2

Inleiding in de Mathematische Statistiek

Algemene setup

- Data: stochastische grootheid X. Vaak $X = (X_1, \ldots, X_n)$.
- Model: collectie kansverdelingen $\{P_\theta : \theta \in \Theta\}$ die de mogelijke verdelingen van de data beschrijven.

Algemene setup

- Data: stochastische grootheid X. Vaak $X = (X_1, \ldots, X_n)$.
- Model: collectie kansverdelingen $\{P_\theta : \theta \in \Theta\}$ die de mogelijke verdelingen van de data beschrijven.

- H. 2: grafische methoden om op basis van de data X een eerste idee te krijgen over wat een redelijk model $\{P_\theta : \theta \in \Theta\}$ zou kunnen zijn.

Algemene setup

- Data: stochastische grootheid X. Vaak $X = (X_1, \ldots, X_n)$.
- Model: collectie kansverdelingen $\{P_\theta : \theta \in \Theta\}$ die de mogelijke verdelingen van de data beschrijven.

- H. 2: grafische methoden om op basis van de data X een eerste idee te krijgen over wat een redelijk model $\{P_\theta : \theta \in \Theta\}$ zou kunnen zijn.

- H. 3: nu is het model $\{P_\theta : \theta \in \Theta\}$ *gegeven* en willen we op basis van de data X de ware waarde van de parameter θ schatten.

Schatters

Schatters - abstracte definitie

Gegeven:
- data: stochastische grootheid X met waarden in $\mathcal{X}$,
- model: collectie $\{P_\theta : \theta \in \Theta\}$ van kansmaten op $\mathcal{X}$.

(Meestal: $\mathcal{X} \subset \mathbb{R}^n$, $\Theta \subset \mathbb{R}^k$.)

Definitie.
Een schatter is een stochastische grootheid $T(X)$, voor zekere $T : \mathcal{X} \to \Theta$. Als we voor X de realisatie x hebben, dan is $T(x)$ de bijbehorende schatting.

Schatters - vragen

Idee:
Als een zekere $\theta_0 \in \Theta$ de "ware parameter" is, i.e. $X \sim P_{\theta_0}$, dan willen we $T(X) \approx \theta_0$ hebben.

Vragen:

- Hoe maken we dit precies?
- Hoe construëren we "goede" schatters.

Voorbeeld 3.2

We observeren $X_1, \ldots, X_n$, o.o., hom$[0, \theta]$-verdeeld, voor $\theta > 0$.
Doel: schatten van θ.

Poging 1.

Voorbeeld 3.2

We observeren $X_1, \ldots, X_n$, o.o., hom$[0, \theta]$-verdeeld, voor $\theta > 0$.
Doel: schatten van θ.

Poging 1.

We hebben $E_\theta X_1 =$

Voorbeeld 3.2

We observeren $X_1, \ldots, X_n$, o.o., hom$[0, \theta]$-verdeeld, voor $\theta > 0$.
Doel: schatten van θ.

Poging 1.

We hebben $E_\theta X_1 = \theta/2$. Dus (WGA): $\bar{X} \to \theta/2$ onder P_θ.

Voorbeeld 3.2

We observeren $X_1, \ldots, X_n$, o.o., hom$[0, \theta]$-verdeeld, voor $\theta > 0$.
Doel: schatten van θ.

Poging 1.

We hebben $E_\theta X_1 = \theta/2$. Dus (WGA): $\bar{X} \to \theta/2$ onder P_θ.

Redelijke schatter lijkt daarom (althans voor grote n): $\hat{\theta}_1 = 2\bar{X}$.

Voorbeeld 3.2

We observeren $X_1, \ldots, X_n$, o.o., hom$[0, \theta]$-verdeeld, voor $\theta > 0$.
Doel: schatten van θ.

Poging 2.

Voorbeeld 3.2

We observeren $X_1, \ldots, X_n$, o.o., hom$[0, \theta]$-verdeeld, voor $\theta > 0$.
Doel: schatten van θ.

Poging 2.

Een enkele "grote" waarneming in de buurt van θ heeft weinig invloed op de schatter $2\bar{X}$, maar is wel heel informatief over de waarde θ.

Voorbeeld 3.2

We observeren $X_1, \dots, X_n$, o.o., hom$[0, \theta]$-verdeeld, voor $\theta > 0$.
Doel: schatten van θ.

Poging 2.

Een enkele "grote" waarneming in de buurt van θ heeft weinig invloed op de schatter $2\bar{X}$, maar is wel heel informatief over de waarde θ.

Nieuwe poging: $\hat{\theta}_2 = X_{(n)} = \max\{X_1, \dots, X_n\}$.

Voorbeeld 3.2

We observeren $X_1, \ldots, X_n$, o.o., hom$[0, \theta]$-verdeeld, voor $\theta > 0$.
Doel: schatten van θ.

Poging 3.

Voorbeeld 3.2

We observeren $X_1, \ldots, X_n$, o.o., hom$[0, \theta]$-verdeeld, voor $\theta > 0$.
Doel: schatten van θ.

Poging 3.

Onder P_θ hebben we altijd met kans 1 dat $X_{(n)} < \theta$. De schatter $\hat{\theta}_2$ levert dus altijd een onderschatting.

Voorbeeld 3.2

We observeren $X_1, \ldots, X_n$, o.o., hom$[0, \theta]$-verdeeld, voor $\theta > 0$.
Doel: schatten van θ.

Poging 3.

Onder P_θ hebben we altijd met kans 1 dat $X_{(n)} < \theta$. De schatter $\hat{\theta}_2$ levert dus altijd een onderschatting.

Nieuwe poging:

$$\hat{\theta}_3 = \frac{n+2}{n+1} X_{(n)}.$$

$\longrightarrow$ R illustratie

Verwachte kwadratische fout

Idee: we geven de voorkeur aan een schatter die zo goed mogelijk "geconcentreerd" ligt rond de ware parameter.

Definitie.
Stel $\mathcal{X} \subseteq \mathbb{R}^n$, $\Theta \subseteq \mathbb{R}^k$. De verwachte kwadratische fout van een schatter $T = T(X)$ voor θ is

$$MSE(\theta; T) = E_\theta \|T - \theta\|^2, \quad \theta \in \Theta.$$

N.B. We kunnen de MSE van een schatter T opvatten als een functie $\theta \mapsto MSE(\theta; T)$.

We geven de voorkeur aan schatters waarvoor de MSE voor alle $\theta \in \Theta$ klein is.

Onzuiverheid en variantie

Als $\Theta \subseteq \mathbb{R}$, dan hebben we de decompositie

$$\begin{aligned} MSE(\theta; T) &= E_\theta(T - \theta)^2 \\ &= E_\theta((T - E_\theta T) + (E_\theta T - \theta))^2 \\ &= \underbrace{Var_\theta T}_{\text{variantie}} + (\underbrace{E_\theta T - \theta}_{\text{onzuiverheid}})^2. \end{aligned}$$

Definitie.
De schatter $T = T(X)$ heet zuiver als voor alle $\theta \in \Theta$,

$$E_\theta T = \theta.$$

N.B. Zuiverheid zegt niets over de variantie van een schatter, daarom hebben zuivere schatters niet noodzakelijk onze voorkeur.

$\longrightarrow$ vb 3.6, 3.7

Maximum Likelihood-Schatters

Voorbeeld 3.9

Experiment: gooi 10 maal met een onzuivere munt, kans p op kop. Zij X het aantal malen "kop". Stel, we observeren 3 maal kop. Hoe kunnen we p schatten?

Als we p kennen, dan is de kans (likelihood) op de geobserveerde uitkomst

$$P_p(X = 3) = \binom{10}{3} p^3 (1-p)^7.$$

Idee van maximum likelihood: schat p zodanig dat de kans op de geobserveerde data maximaal wordt. M.a.w., definieer de schatter $\hat{p}$ als het punt waar

$$p \mapsto P_p(X = 3)$$

maximaal is. (Hier $\hat{p} = 3/10$).

Algemene definitie - 1

Gegegeven: waarneming X en model $\{P_\theta : \theta \in \Theta\}$.

Als X discreet is onder P_θ, definieer dan p_θ als de *kansmassa functie* van X. Dus als X waarden $x_1, x_2, \ldots$ aanneemt, dan

$$p_\theta(x_i) = P_\theta(X = x_i), \qquad i = 1, 2, \ldots.$$

Als X absoluut continu is onder P_θ, definieer dan p_θ als de *dichtheid* van X.

We noemen p_θ in beide gevallen de kansdichtheid van X.

Algemene definitie - 2

Gegegeven: waarneming X en model $\{P_\theta : \theta \in \Theta\}$. Zij p_θ de kansdichtheid van X onder P_θ.

Definitie.
Voor een vaste realisatie x van X definiëren we

$$L(\theta; x) = p_\theta(x), \quad \theta \in \Theta.$$

De functie $\theta \mapsto L(\theta; x)$ op Θ heet de likelihood-functie.

Algemene definitie - 3

Definitie.

Voor een realisatie x van X definiëren we de maximum likelihood-schatting als het punt $T(x)$ waar de likelihood $\theta \mapsto L(\theta; x)$ maximaal is. De maximum likelihood-schatter is de bijbehorende schatter $T(X)$.

N.B.

- Uiteraard is de MLS alleen goed gedefinieerd als het maximum van de likelihood bestaat.
- De likelihood kan meerdere maxima hebben. In dat geval is het beter om van een MLS te spreken.
- Deze gevallen zullen we in dit college niet tegenkomen.

Berekenen van maximum likelihood-schatters - 1

1. Bepaal de (simultane) dichtheid p_θ van de waarneming X.
2. Beschouw de likelihood $L(\theta; x) = p_\theta(x)$.

Als de likelihood een positieve, gladde functie op Θ is:

3. calculus.

In andere gevallen:

3. ad hoc argument, plaatje.

Berekenen van maximum likelihood-schatters - 2

Geval van een gladde, positive likelihood. Stel $\Theta \subseteq \mathbb{R}$.

- Bereken de likelihood $L(\theta; x)$. (Vaak: product van marginale dichtheden.)
- Berekenen de log-likelihood $\ell(\theta; x) = \log L(\theta; x)$.
- Berekenen de score functie $\dot{\ell}(\theta; x) = \frac{\partial}{\partial \theta} \ell(\theta; x)$.
- Nulstellen en controleren of er een maximum is gevonden.

$\longrightarrow$ vb 3.12, 3.15, 3.13, 3.14