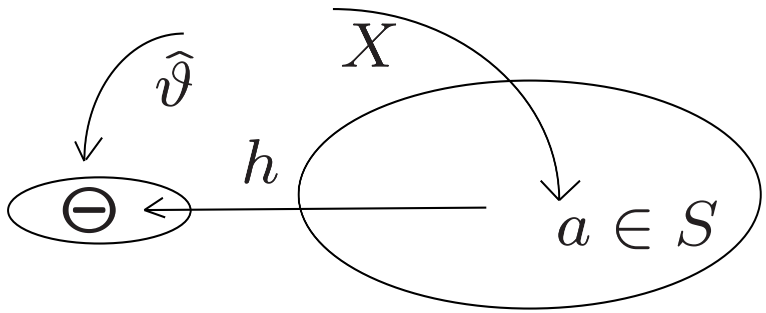
Nachtrag zu Vorlesung 13b

Maximum-Likelikood-Schätzung

$$\mathbf{P}_\vartheta(X \in da) = f_\vartheta(a)\, da, \quad \vartheta \in \Theta$$



Wähle $h(a)$ so, dass

$$f_{h(a)}(a) = \max_{\vartheta \in \Theta} f_\vartheta(a).$$

Die Zufallsvariable $\widehat{\vartheta} := h(X)$ nennt man dann

Maximum-Likelikood-Schätzer für ϑ auf der Basis von X.

7. Beispiel:

Zweiseitige Exponentialverteilung.

Für $\vartheta \in \mathbb{R}$ seien $X_1, \ldots, X_n$

unabhängig und identisch verteilt mit Dichte

$$g_\vartheta(x) := \frac{1}{2}e^{-|x-\vartheta|},\ x \in \mathbb{R}.$$

Was ist der ML-Schätzer für ϑ?

$$f_\vartheta(a_1, \ldots, a_n) := \left(\tfrac{1}{2}\right)^n \prod_{i=1}^{n} e^{-|a_i - \vartheta|},\ a_i \in \mathbb{R}$$

Betrachten wir erst einmal den Fall $n = 2$,

und für festes $a_1, a_2 \in \mathbb{R}$ die Funktion

$$\ell : \vartheta \mapsto |a_1 - \vartheta| + |a_2 - \vartheta|.$$

Angenommen, $a_1 < a_2$.

Für $\vartheta < a_1$ hat ℓ die Steigung -2,

für $\vartheta > a_2$ hat ℓ die Steigung $+2$,

für $\vartheta \in (a_1, a_2)$ hat ℓ die Steigung 0.

Also ist jedes $\vartheta \in [a_1, a_2]$ Minimalstelle von ℓ.

$$f_\vartheta(a_1, \ldots, a_n) := \left(\tfrac{1}{2}\right)^n \prod_{i=1}^{n} e^{-|a_i - \vartheta|},\ a_i \in \mathbb{R}$$

Betrachten wir jetzt den Fall $n = 3$,

und für festes $a_1, a_2, a_3 \in \mathbb{R}$ die Funktion

$\ell : \vartheta \mapsto |a_1 - \vartheta| + |a_2 - \vartheta| + |a_3 - \vartheta|$.

Angenommen, $a_1 < a_2 < a_3$.

Für $\vartheta < a_2$ hat ℓ negative Steigung,

für $\vartheta > a_2$ hat ℓ positive Steigung.

Also ist a_2 die einzige Minimalstelle von ℓ.

Was 2 und 3 recht ist, soll n billig sein.

Definition. Seien $a_1, \ldots, a_n \in \mathbb{R}$.

Eine Zahl heißt **Median** von $a_1, \ldots, a_n$,

wenn

ebenso viele der a_i links wie rechts von ihr liegen.

Für ungerades n führt die Definition auf einen einzigen Wert,

für gerades n führt sie auf ein Intervall.

Fazit:

Im Besipiel der zweisitigen Exponentialverteilung ist der ML-Schätzer für den Lageparameter (das “Zentrum”) ϑ von der Form

$$h(X_1, \ldots, X_n) := \text{ Median von } X_1, \ldots, X_n$$

(und nicht, wie man auf die Schnelle vielleicht vermuten würde, der arithmetische Mittelwert der X_i).

Nachtrag zu Vorlesung 12a

Ein Konfidenzintervall für den Median einer Verteilung

Seien $X_1, X_2, \ldots, X_n$ unabhängig, mit Verteilung ρ.

Es gibt Situationen, in denen die Schätzung der “Lage” von ρ

über den Stichprobenmittelwert problematisch ist –

etwa wenn ρ so viel Masse weit draußen hat,

dass die Varianz von ρ sehr groß ist

(oder sogar der Erwartungswert von ρ gar nicht existiert)

In diesem Fall ist es günstig,

einen “robusteren” Lageschätzer zu verwenden:

Definition.

Eine Zahl ν heißt *Median* der Verteilung ρ auf $\mathbb{R}$,

wenn $\rho((-\infty, \nu]) \geq 1/2$ und $\rho([\nu, \infty)) \geq 1/2$ gilt.

Wenn es nur eine Zahl ν gibt mit $\rho((-\infty, \nu]) = 1/2$

(also z. B. wenn $\rho([\ell, r]) = 1$ gilt

und ρ eine strikt positive Dichtefunktion besitzt),

dann ist ν **der** Median von ρ.

Definition.

Eine Zahl ν heißt *Median* der Verteilung ρ auf $\mathbb{R}$,

wenn $\rho((-\infty,\nu]) \geq 1/2$ und $\rho([\nu,\infty)) \geq 1/2$ gilt.

Wenn es mehrere Zahlen ν gibt mit $\rho((-\infty,\nu]) = 1/2$,

dann ist jede dieser Zahlen **ein** Median von ρ.

Bsp: Für die uniforme Verteilung auf $[0,1] \cup [2,3]$

ist jedes $\nu \in [1,2]$ ein Median.

Wie schätzt man den Median?

Die *Ordnungsstatistiken* $X_{(1)} \leq \cdots \leq X_{(n)}$
sind die aufsteigend geordneten $X_1, \ldots, X_n$.

Ein Kandidat für ein Konfidenzintervall für den Median ist

$$[X_{(1+j)}, X_{(n-j)}]$$

mit $0 \leq j < n/2$.

$$\mathbf{P}_\rho\big(\nu \notin [X_{(1)}, X_{(n)}]\big) = \mathbf{P}_\rho(X_{(1)} > \nu) + \mathbf{P}_\rho(X_{(n)} < \nu) \ .$$

$$\mathbf{P}_\rho(X_{(1)} > \nu) \leq 2^{-n}$$

$$\mathbf{P}_\rho(X_{(n)} < \nu) \leq 2^{-n}.$$

Also:

$$\mathbf{P}_\rho\big(\nu \in [X_{(1)}, X_{(n)}]\big) \geq 1 - \frac{1}{2^{n-1}} \ .$$

$$\mathbf{P}_\rho\big(\nu \in [X_{(1)}, X_{(n)}]\big) \geq 1 - \frac{1}{2^{n-1}}\,.$$

Bsp: $n = 6$

$$\mathbf{P}_\rho\big(\nu \in [X_{(1)}, X_{(6)}]\big) \geq 1 - \frac{1}{32} = 0.97 \;\; :$$

Das Konfidenzintervall $[X_{(1)}, X_{(6)}]$ für den Median hält die Überdeckungswahrscheinlichkeit 0.97 ein.

Mit ähnlichen Überlegungen lässt sich zeigen:

$$\mathbf{P}_\rho\big(\nu \in [X_{(1+j)}, X_{(n-j)}]\big) \geq \mathbf{1} - \mathbf{2P}(Y \leq j)$$

mit Y Bin($n, 1/2$)-verteilt.

Man beachte:

Mit wachsendem j wird das Intervall kürzer

und die Überdeckungsw'keit nimmt ab!

Vorlesung 14a

Eine Klausur aus dem Vorjahr

mit Besprechung der Lösungen an der Tafel

1. Berechnen Sie beim fairen Münzwurf die erwartete Anzahl von Würfen, bis erstmals das Muster `0101` erscheint.*

(Beispielsweise ist bei der Realisierung `00101` *diese Anzahl gleich 5.)*

Finden Sie dafür zuerst einen passenden Graphen.

*Lösung: 20

2. a) Es sei K eine Binomial$(20,\frac{1}{2})$-verteilte Zufallsvariable. Berechnen Sie die Wahrscheinlichkeit des Ereignisses $\{|\,K - \mathbf{E}[K]\,| \geq 8\}$.

b) Wir betrachten eine auf $[0, \infty)$ definierte, strikt positive Dichtefunktion f. Die Zufallsvariablen $X_1, \ldots, X_{20}$ seien unahängig und haben Dichte $f(a)\, da$, $a \geq 0$.

Es wurde folgendes Ergebnis berichtet: “Zwei der X_i sind kleiner ausgefallen als 4, und die restlichen 18 größer als 4”.

i) Wie wahrscheinlich ist unter der Annahme

$$(H_0) \qquad \int_0^4 f(a)\, da = \frac{1}{2}$$

ein so “extremes” Ergebnis wie das berichtete? *(Man beachte: Unter der Annahme (H_0) wäre das Ergebnis “Zwei der X_i sind größer ausgefallen als 4, und die restlichen 18 kleiner als 4” ebenso “extrem” wie das oben berichtete.)*

ii) *Im Jargon der Statistik (und mit unserer Annahme an f) bedeutet* (H_0)*, dass die Zahl* 4 der Median *der Verteilung mit Dichte* $f(a)\,da$ *ist.* Zu welchem p-Wert können Sie aufgrund des in i) berichteten Ergebnisses

“Zwei der X_i sind kleiner ausgefallen als 4,
und die restlichen 18 größer als 4”

die Hypothese ablehnen, dass der Median gleich 4 ist?

3. U sei uniform verteilt auf $[0, 2] \cup [3, 7] \cup [8, 10]$.
Denken Sie sich U als zweite Stufe eines Zufallsexperimentes dargestellt, dessen erste Stufe X Wertebereich $\{1, 2, 3\}$ hat und dessen Übergangsverteilungen $P(1, .)$, $P(2, .)$, $P(3, .)$ jeweils uniforme Verteilungen auf einem Intervall sind.

a) Wie ist die Verteilung von X zu wählen, damit die oben beschriebene Darstellung klappt?

b) Berechnen Sie mittels dieser Darstellung die Varianz von U.

4. Es sei X eine reellwertige Zufallsvariable, und $s > 0$.

a) Begründen Sie unter Verwendung der Ungleichung von Markov, warum für alle $b \in \mathbb{R}$ gilt:

$$\mathbf{P}(X \geq b) \leq \frac{\mathbf{E}[e^{sX}]}{e^{sb}}.$$

Schreiben Sie dazu erst einmal das Ereignis $\{X \geq b\}$ *auf passende Weise um.*

b) Es sei $n \in \mathbb{N}$ und $p \in [0, 1]$. Berechnen Sie $\mathbf{E}[e^{sX}]$ für

(i) Binomial$(1, p)$

(ii) Binomial(n, p)

–verteiltes X.

(*Bei (ii) ist die altbekannte Formel* $e^{u+v} = e^u e^v$ *hilfreich. Erinnern Sie sich auch daran, dass sich eine Binomial*(n, p)*-verteilte Zufallsvariable als Anzahl der Erfolge bei einem* n*-fachen* p*-Münzwurf auffassen lässt.*)

5. S und R seien zwei unabhängige reellwertige Zufallsvariable, $M := S + R$. (S steht für *Signal*, R für *Rauschen* und M für *Messung*.) Die Zufallsvariable S sein N$(10, 4)$-verteilt, und die Zufallsvariable R sei N$(1, 1)$-verteilt.

a) Berechnen Sie die beste affin lineare Prognose $h(M)$ von S auf der Basis von M.

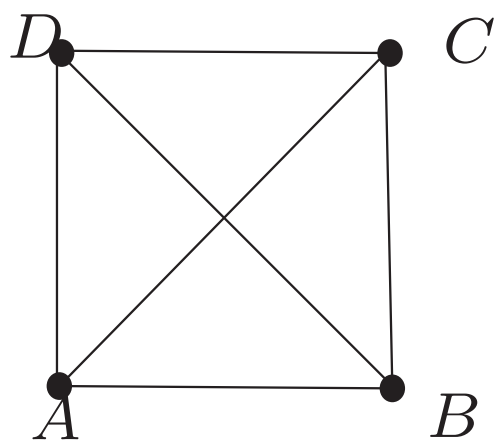
b) Finden Sie den Erwartungswert des Quadrates der Differenz $h(M) - S$.

5. S und R seien zwei unabhängige reellwertige Zufallsvariable, $M := S + R$. Die Zufallsvariable S sein N$(10, 4)$-verteilt, und die Zufallsvariable R sei N$(1, 1)$-verteilt.

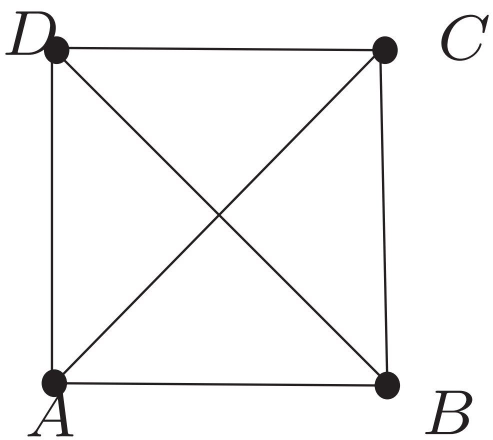
c) *Wie in einer unserer Übungsaufgaben kann man sich überzeugen, dass* S *gegeben* $\{M = b\}$ *die Verteilung* $\mathrm{N}(g(b), v)$ *hat, wobei der bedingte Erwartungswert* $g(b)$ *affin linear von* b *abhängt und die bedingte Varianz* v *nicht von* b *abhängt. Das sollen Sie hier nicht überprüfen, aber zu Folgendem verwenden:*

Geben Sie ein aus M gewonnenes zufälliges Intervall J an, für das gilt: $\mathbf{P}(S \in J | M = b) \approx 0.95$.

6. Wir betrachten die gewöhnliche Irrfahrt $X = (X_0, X_1, \ldots,)$ auf dem skizzierten ungerichteten Graphen (mit 4 Knoten und 6 Kanten), startend in ihrer Gleichgewichtsverteilung π.



a) Bestimmen Sie $\mathbf{P}_\pi((X_0, X_2) = (A, C))$.



b) Berechnen Sie die erwartete Anzahl der Zeitpunkte $i \in \{0, \ldots, 100\}$, für die $(X_i, X_{i+2}) = (D, B)$ gilt.

Nächste Woche in unseren **Tutorien**:

Besprechung einer weiteren Klausur aus dem Vorjahr.

Und:
Das **Lernzentrum Informatik** bietet während der Klausurenphase zu folgenden Zeiten fachliche Beratung zu den Inhalten der Basismodule:

18.02. - 06.03.2019: Mo - Fr 12:00 - 15:00 Uhr,
ausgenommen Freitag, 01.03.2019

27.03. - 09.04.2019: Mo - Fr 12:00 - 15:00 Uhr