3. Vorlesung, 7. Okt. 2013

Letzte Woche: → Schritt für Schritt Anleitung zur Parameterabschätzung
→ Informationszuwachs durch wiederholte Messung derselben Grösse
→ Ergebnis: hinreichend "scharfe" Posteriorverteilung für den/die Parameter

Wie fassen wir dieses Ergebnis in einem wissenschaftlichen Bericht zusammen?

Die Posteriorverteilung gibt alles wieder, was wir über den/die Parameter wissen. Daher wäre es am einfachsten, die Formel für die Verteilung anzugeben, oder sie graphisch darzustellen.

In der Physik üblich sind jedoch Darstellungen der Form

$$\boxed{x = x_0 \pm \sigma} \qquad (15)$$

Doch welche Zahlenwerte wählen wir für x_0 und σ? Der Wert x_0 soll natürlich möglichst nahe am wahren Wert des Parameters liegen. Wir brauchen also ein "Abstandsmass".

Der Mittelwert einer Verteilungsfunktion

Folie

Um ein solches zu finden, stellen wir uns ein Spiel vor: der Geizkragen Dagobert Duck hat eine Wette gegen den Erfinder Daniel Düsentrieb verloren. Deshalb muss er nun durch Messung die Aktivität einer ^{22}Na-Quelle aus Düsentriebs Labor bestimmen, deren Aktivität Düsentrieb natürlich genau kennt. Je weiter der Wert, den Dagobert ermittelt, vom wahren Wert abweicht, desto mehr Geld muss Dagobert an Daniel bezahlen. Sie vereinbaren zum Beispiel, dass der Betrag B (in Goldtalern) durch

$$B = (A_0 - A_W)^2$$

gegeben ist, wobei A_0 Dagoberts Schätzwert und A_W Daniels wahrer Wert ist. Welchen Wert wird Dagobert am besten angeben, wenn er die Posteriorwahrscheinlichkeit prob(A|k) ermittelt hat? Tick, Trick und Track helfen ihm dabei: sie zücken "Das schlaue Buch" und minimieren den Erwartungswert des Verlusts

Folie

$$\langle B(A_0)\rangle = \int (A-A_0)^2\, \mathrm{pdf}(A|k)\, dA$$

also

$$\frac{d\langle B(A_0)\rangle}{dA_0} = -2\int (A-A_0)\,\mathrm{pdf}(A|k)\,dA = -2\big(\langle A\rangle - A_0\big) = 0 \;\Rightarrow\; \boxed{A_0 = \langle A\rangle := \int A\,\mathrm{pdf}(A|k)\,dA} \qquad (16)$$

Dagobert wird für A_0 also den Erwartungswert der Aktivität angeben.

Beispiel radioaktiver Zerfall: Die Posteriorwahrscheinlichkeit ist durch die Gammafunktion (14) gegeben. Der Erwartungswert ist (siehe Formelblatt)

$$\langle A\rangle = \frac{1}{n\Delta t}\sum_{k=1}^{n} k_n + \frac{1}{n\Delta t} = \frac{\langle k_n\rangle + 1/n}{\Delta t}.$$

Für grosse n geht $\langle A\rangle$, wie zu erwarten war, gegen den Wert $\langle k_n\rangle/\Delta t$.

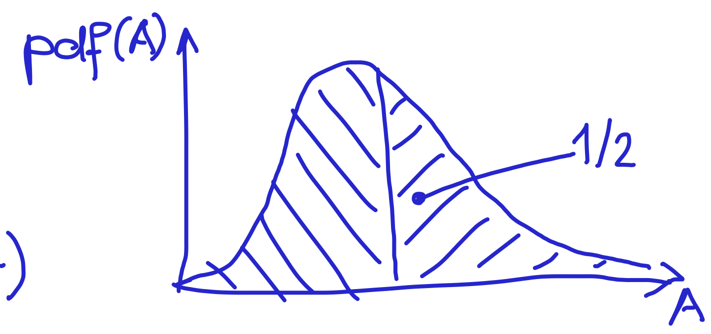
Der Median einer Verteilungsfunktion

Stellen wir uns vor, Dagobert Duck und Daniel Düsentrieb hätten eine andere Regel vereinbart: nehmen wir

$$B = |A_0 - A_W|.$$

Wie sieht dann die optimale Strategie aus? Wieder minimieren wir den Erwartungswert von B

$$\langle B(A_0)\rangle = \int dA\, |A-A_0|\, \mathrm{pdf}(A|k)$$



Diese Minimierung führt zur Bedingung:

$$\boxed{\int_0^{A_0} \mathrm{pdf}(A|k)\,dA = \int_{A_0}^{\infty} \mathrm{pdf}(A|k)\,dA = \frac{1}{2}}. \qquad (17)$$

D.h. Werte $A > A_0$ sollen mit gleicher Wahrscheinlichkeit auftreten, wie Werte $A < A_0$. Das so definierte A_0 heisst Median von pdf(A|k).

Für die Gammaverteilung (14) gibt es keinen analytischen Ausdruck für den Median; er müsste numerisch berechnet werden.

Die Mode einer Verteilungsfunktion

Betrachten wir eine dritte Spielregel, die wir als 'alles oder nichts' bezeichnen:

$$B(A_0) = \begin{cases} 0, & \text{falls } A \in [A_0, A_0 + \Delta A] \\ 10^6 & \text{sonst} \end{cases}$$

In diesem Fall wird Dagobert den wahrscheinlisten Wert für A angeben, also das Maximum der Verteilung bestimmen. Dieser Wert heisst Mode der Verteilung. Man findet ihn als Lösung von

$$\boxed{\frac{d}{dA}\, pdf(A|k) = 0.} \qquad (18)$$

↗ Folie

Mit der Gammafunktion (14) ergibt das die Mode $A_0 = \frac{\langle k_n \rangle}{\Delta t}$

Die Varianz der Verteilungsfunktion

In der Physik ist es üblich, das quadratische Abstandsmass zu wählen und den Mittelwert der Verteilung anzugeben. Dann macht es natürlich auch Sinn, die erwartete Abweichung vom Mittelwert, also

$$\boxed{\langle (x - \langle x \rangle)^2 \rangle = \int dx\, (x - \langle x \rangle)^2\, pdf(x) \equiv Var(x)} \qquad (19)$$

anzugeben. Man nennt diese Grösse die Varianz der Verteilungsfunktion pdf (x). Die Grösse $\sigma = \sqrt{Var(x)}$ hat dieselbe Einheit, wie der Erwartungswert $x_0 = \langle x \rangle$. Wir schreiben daher (per Konvention)

$$\boxed{x = \langle x \rangle \pm \sqrt{Var(x)}} \qquad (20)$$

Dabei werden Mittelwert und Varianz mit Hilfe der Posteriorwahrscheinlichkeit für x berechnet. Die Grösse $\sigma = \sqrt{Var(x)}$ heisst Standardabweichung von x. In unserem Beispiel ist nach n-maliger Messung die Varianz der Posteriorwahrscheinlichkeit (14)

$$Var(A) = \int dA\, (A - \langle A \rangle)^2\, pdf(A|k_1 \dots k_n, \Delta t, M, I) = \left(\sum_{j=1}^{n} k_j + 1\right) \frac{1}{n^2 \Delta t^2} = \frac{\langle A \rangle}{n \Delta t}.$$

Damit haben wir die Standardabweichung von A

$$\sigma = \sqrt{\frac{\langle A \rangle}{n \Delta t}}$$

↗ Folie

Da ⟨A⟩ für eine grosse Zahl von Messungen n gegen den festen Wert $\langle k_j \rangle / \Delta t$ strebt, nimmt die Standardabweichung von A gemäss $\frac{1}{\sqrt{n}}$ ab. Die Verteilung wird also immer schärfer, und der Schätzwert ⟨A⟩ immer verlässlicher.

Das Abfallen der Standardabweichung mit $1/\sqrt{n}$ kann in der Parameterabschätzung als durchaus typisch bezeichnet werden.

1. Bemerkung: Die sog. Ungleichung von Tschebyschiff besagt, dass die Gesamtwahrscheinlichkeit P im Intervall

$\langle x \rangle \pm t\sqrt{Var(x)}$ der Ungleichung

$$P \geq 1 - \frac{1}{t^2}$$

genügt, unabhängig von der genauen Form von pdf (x). So wissen wir zum Beispiel, dass mehr als 8/9 = 88% der Wahrscheinlichkeit im Intervall $\langle x \rangle \pm 3\sqrt{Var(x)}$ konzentriert ist.

2. Bemerkung: Es ist klar, dass die abgekürzte Schreibweise $\langle x \rangle \pm \sqrt{Var(x)}$ nicht die volle Information der Posteriorwahrscheinlichkeit vermittelt. Wenn möglich, ist deren Angabe daher von grossem Wert. Zum Beispiel können aus dem Mittelwert und der Varianz einer Gamma-Verteilung die Parameter α und β bestimmt werden. Man muss aber wissen, dass es sich um eine Gammaverteilung handelt. Wenn die Verteilungsfunktion nicht angegeben wird, nimmt man implizit eine Normalverteilung mit den Parametern $\mu = \langle x \rangle$ und $\sigma^2 = Var(x)$ an.

Mittelwert, Median und Mode bezeichnet man auch als Schätzwerte (Estimators) des gesuchten Parameters x.

Graphische Darstellung mit Fehlerbalken

Zahlenwerte der Form $x = x_0 \pm \sigma$ werden häufig grafisch als Punkte mit Fehlerbalken dargestellt.

↗ Folie

→ darauf achten, dass genau angegeben ist, was dargestellt ist. z. B. $x = \langle x \rangle \pm 3\sigma$
Verbreitet sind Darstellungen von $\langle x \rangle \pm \sqrt{Var(x)}$ oder $\langle x \rangle \pm 3\sqrt{Var(x)}$.

Näherung durch eine Normalverteilung

Es ist sehr häufig auch möglich, die Posteriorverteilung durch eine Normalverteilung anzunehmen. In diesem Fall entwickelt man den Logarithmus der Posteriorverteilung $pdf(x) = f(x)$ um die Mode x_0 herum:

$$\ln f(x) \approx \ln f(x_0) + \frac{1}{2} \frac{d^2}{dx^2} \ln f(x)\Big|_{x=x_0} (x-x_0)^2$$

Definieren wir $\frac{1}{\sigma^2} = -\frac{d^2}{dx^2} \ln f(x)\Big|_{x=x_0}$, dann finden wir die Normalverteilung

$$\boxed{pdf(x) = f(x) = f(x_0) \exp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right)} \qquad (21)$$

mit der Kurzdarstellung

$$\boxed{x = x_0 \pm \sigma}$$

↗ Folie

Die Näherung einer Verteilungsfunktion durch eine Normalverteilung ist besonders für Verteilungen geeignet, die aus einer grossen Zahl von wiederholten Messungen hervorgegangen ist. Die Verteilungen tendieren gegen die Normalverteilung. Darüber hinaus bewährt sich die Näherung zur Berechnung von Integralen, die ansonsten nicht analytisch, oder nur mit grösserem Aufwand berechnet werden könnten.

Entropie einer Verteilung und quantitative Information

Claude Shannon entwarf 1948 eine mathematische Theorie der Information. Dabei ordnete er diskreten Wahrscheinlichkeitsverteilungen $prob(n)$ (n ganzzahlig) die Informationsentropie

$$\boxed{H[prob(n)] = -\sum_n prob(n) \log_2 prob(n)} \qquad (22)$$

zu. Die Einheit dieser Entropie wird "Bit" (binary digit) genannt. Beim Werfen einer Münze mit zwei möglichen Ergebnissen $n \in \{0,1\}$ und $prob(0) = prob(1) = \frac{1}{2}$ haben wir tatsächlich $H = 1$ bit. Das bedeutet, dass wir in der Tat ein Bit Speicherplatz in einem Computer aufwenden müssen, um das Ergebnis eines Münzwurfes zu speichern. Die Shannon Entropie beschreibt demnach die Unsicherheit, die wir über die Grösse n gemäss der Verteilung $prob(n)$ haben.

Die Verallgemeinerung auf kontinuierliche Wahrscheinlichkeitsdichten $pdf(x) = f(x)$ lautet

$$\boxed{H[pdf(x)] = -\int f(x) \log_2 f(x).} \qquad (23)$$

Diese Definitionen geben der Spirale des Informationszuwachses einen quantitativen Sinn: verändert sich nämlich die Posteriorverteilung $f(x)$ durch Wiederholung der Messung zu $g(x)$, dann ist der Informationszuwachs durch

$$\boxed{\Delta I = H[f(x)] - H[g(x)]} \qquad (24)$$

gegeben. Diese Zahl gibt uns ganz konkret an, wieviele zusätzliche Bits wir zum Speichern des verfeinerten Schätzwerts zusätzlich bereitstellen müssen. Dabei entsprechen ca. 3.32 bit einer Dezimalstelle. Gewinne ich durch wiederholtes Messen beispielsweise $\Delta I = 6.64$ bit Information über einen Parameter, dann kann ich den Parameterwert mit der grösseren Genauigkeit von 2 zusätzlichen Dezimalstellen angeben.

Mantra des Messens

Durch Wiederholung von Messungen kann die Unsicherheit über den Wert des gesuchten Parameters verringert, und die Information über den gesuchten Parameter vergrössert werden. Dafür bezahlt man als Preis den Zeitaufwand für das erneute Messen.

↗ Folie