Lernziele Lektion 3

- Sie wissen, dass die Posteriorverteilung das vollständige Ergebnis einer Parameterabschätzung ist, wohingegen Schätzwerte und ihre Unsicherheiten nur eine näherungsweise Beschreibung dieser Verteilung sind.
- Sie kenne die Definitionen der Begriffe Mode, Mittelwert, Median und Varianz einer Warscheinlichkeitsverteilung und wissen, unter welchen Bedingungen sie jeweils der optimale Schätzwert sind.
- Sie wissen, wie man diese Schätzwerte aus einer gegebenen Posteriorverteilung extrahiert.
- Sie kennen die Konventionen beim angeben von Schätzwerten und ihren Unsicherheiten (Fehlern).
- Sie kennen den Unterschied zwischen der Ungenauigkeit (dem Fehler) von Messwerten und der Unsicherheit (dem Fehler) von Schätzwerten (Mittelwerten).
- Sie können Messergebnisse und ihre Fehler graphisch darstellen.
- Sie kennen die Näherung von Laplace.
- Sie kennen die Grundzüge des Zusammenhangs zwischen Wahrscheinlichkeitsverteilungen und dem Begriff der Information.

3 Schätzwerte und ihre Unsicherheiten ('Fehler')

In den Beispielen der zwei vorangegangenen Lektionen haben wir gesehen, dass eine Parameterabschätzung ausgehend von der Likelihood-Funktion mit Hilfe des Bayes'schen Theorems zu einer Posteriorverteilung für die gesuchten Parameter des Modells führt. Die Posteriorverteilung gibt dabei alles wieder, was wir über den/die Parameter wissen. Daher wäre es am einfachsten, die Formel für die Posteriorverteilung als Ergebnis anzugeben, oder sie graphisch darzustellen.

In der Physik üblich sind jedoch Darstellungen der Form

$$x = x_0 \pm \sigma, \tag{27}$$

wobei man x_0 den *Schätzwert* des Parameters x, und σ die Unsicherheit (oder den 'Fehler') des Schätzwertes nennt. Doch welche Zahlenwerte wählen wir für x_0 und σ, und wie extrahieren wir sie aus der Posteriorverteilung? Der Wert x_0 soll natürlich möglichst nahe am wahren Wert des Parameters liegen. Wir brauchen also ein 'Abstandsmass'. Der Wert von σ wird wohl irgendwie mit der Breite der Posteriorverteilung zu tun haben, aber wie extrahieren wir sie aus der Verteilungsfunktion?

3.1 Der Mittelwert einer Verteilungsfunktion

Um ein Abstandsmass zu finden, stellen wir uns ein Spiel vor: der Geizkragen Dagobert Duck hat eine Wette gegen den Erfinder Daniel Düsentrieb verloren. Als Wetteinsatz 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 der Geizkragen Dagobert an Daniel bezahlen. Sie vereinbaren zum Beispiel, dass der Betrag B (in Goldtalern) durch

$$B = (A_0 - A_\mathrm{w})^2$$

bestimmt ist, wobei A_0 Dagoberts Schätzwert und A_w Daniels wahrer Wert ist.

Frage: Welchen Wert wird Dagobert am besten angeben, wenn er die Posteriorwahrscheinlichkeit $\mathrm{pdf}(A|k, M, I)$ ermittelt hat?

Antwort: Tick, Trick, und Track helfen ihm dabei: sie zücken das 'schlaue Buch' und minimieren den Erwartungswert des Verlusts

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

also

$$\frac{d\langle B(A_0)\rangle}{dA_0} = -2 \int (A - A_0) \mathrm{pdf}(A|k) dA = -2(\langle A\rangle - A_0) = 0,$$

woraus folgt:

$$A_0 = \langle A\rangle := \int A \mathrm{pdf}(A|k) dA. \tag{28}$$

Dagobert wird für A_0 also den *Erwartungswert* der Aktivität bezüglich der Posteriorverteilung $\mathrm{pdf}(A|k)$ angeben.

Beispiel radioaktiver Zerfall: Gemäss Gl. (10) haben wir

$$\mathrm{pdf}(A|k, \Delta t, M, I) = \Gamma(A; k+1, 1/\Delta t),$$

wobei $\Gamma(x; \alpha, \beta)$ die Gammaverteilung ist. Der Erwartungswert der Gammaverteilung ist (siehe z.B. Formelsammlung) durch das Produkt $\alpha\beta$ gegeben, so dass sich in unserem Fall ergibt

$$A_0 = \langle A \rangle = \int A\Gamma(A; k+1, 1/\Delta t) = \frac{k+1}{\Delta t}.$$

Bei $k = 20$ gemessenen Events innerhalb von 1 s haben wir hier also das interessante Ergebnis, dass der Schätzwert von Dagobert $A_0 = 21\,\mathrm{Bq}$ *nicht* dem wahrscheinlichsten Wert (dem Maximum der Posteriorverteilung bei $k/\Delta t$) entspricht. Wir sehen aber auch, dass für eine grosse Zahl $k \gg 1$ von gemessenen Events, dieser Schätzwert praktisch identisch mit dem Maximum der Verteilung wird.

3.2 Der Median einer Verteilungsfunktion

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

$$B = |A_0 - A_\mathrm{w}|.$$

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

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

d.h. Werte $A > A_0$ sollten mit der gleichen Wahrscheinlichkeit auftreten, wie Werte $A < A_0$. Diese Bedingung ist in Abb. 10 für das Beispiel des radioaktiven Zerfalls graphisch veranschaulicht. Der so definierte Wert A_0 heisst *Median* der Verteilungsfunktion. Für die Gammaverteilung (10) gibt es keinen analytischen Ausdruck für den Median; er muss numerisch berechnet werden.

3.3 Die Mode einer Verteilungsfunktion

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

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

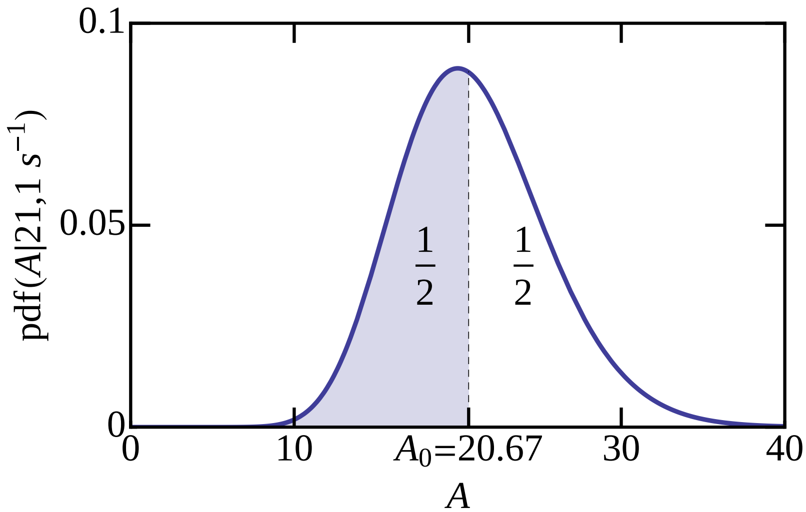

Abbildung 10: Konzept des Medians. Die Fläche unter der Wahrscheinlichkeitsverteilung links und rechts des Medianwertes ist jeweils 1/2.

wobei ΔA eine sehr kleine Zahl ist. In diesem Fall wird Dagobert den wahrscheinlichsten Wert für A angeben, also das Maximum der Verteilung bestimmen. Dieser Wert heisst *Mode* der Verteilung. Man findet ihn als Lösung von

$$\frac{d}{dA}\mathrm{pdf}(A|k) = 0.$$

Für die Gammaverteilung (10) ist die Mode $(\alpha-1)\beta$ (für $\alpha > 1$), wir haben also den Wert der Mode $A_0 = k/\Delta t$, den wir bereits zuvor verwendet haben.

3.4 Zusammenfassung: welchen Schätzwert gibt man an?

Wir haben gesehen, dass es verschiedene Schätzwerte gibt, die jeweils unter bestimmten Bedingungen als die besten angesehen werden können. Diese drei Schätzwerte sind in Abb. 11 zusammengefasst. Würden wir noch mehr Spielregeln in Betracht ziehen, könnten wir noch weitere Schätzwerte mit gutem Grund angeben.

Der übliche Schätzwert: In der Physik ist es üblich, das quadratische Abstandsmass zu wählen und den Mittelwert der Verteilung anzugeben.

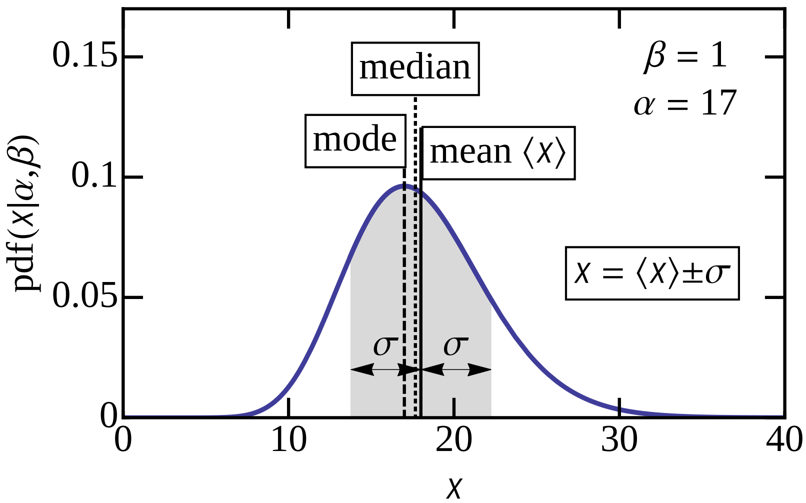

Abbildung 11: A Gamma distribution function with its mean, mode and median indicated.

3.5 Die Varianz der Verteilungsfunktion

Vor diesem Hintergrund macht es nun Sinn für die Unsicherheit des Mittelwertes, die erwartete Abweichung vom Mittelwert, also

$$\left\langle (x - \langle x \rangle)^2 \right\rangle = \int dx\, (x - \langle x \rangle)^2 \mathrm{pdf}(x) \equiv \mathrm{Var}(x) \tag{29}$$

anzugeben. Man nennt diese Grösse die *Varianz* der Verteilungsfunktion $\mathrm{pdf}(x)$. Die Grösse

$$\sigma = \sqrt{\mathrm{Var}(x)} \tag{30}$$

hat dieselbe Einheit wie der Erwartungswert $x_0 = \langle x \rangle$. Wir schreiben daher (per Konvention)

$$x = \langle x \rangle \pm \sqrt{\mathrm{Var}(x)} = \langle x \rangle \pm \sigma.$$

Dabei werden Mittelwert und Varianz mit Hilfe der Posteriorwahrscheinlichkeit für x berechnet. Die Grösse σ in Gl. (30) heisst *Standardabweichung* von x.

Beispiel radioaktiver Zerfall: Die Varianz der Gammaverteilung ist $\alpha\beta^2$, so dass die Varianz der Aktivität gemäss Gl. (10) durch

$$\mathrm{Var}(A) = \frac{k+1}{\Delta t^2} = \frac{\langle A \rangle}{\Delta t}.$$

gegeben ist. Damit erhalten wir die Standardabweichung

$$\sigma = \sqrt{\frac{k+1}{\Delta t^2}} = \sqrt{\frac{\langle A \rangle}{\Delta t}}.$$

Da der Mittelwert $\langle A \rangle$ für sehr lange Messdauer gegen einen festen Wert strebt, nimmt die Standardabweichung von A mit zunehmender Messzeit Δt gemäss $1/\sqrt{\Delta t}$ ab. Die Posteriorverteilung wird also immer schärfer und der Schätzwert $\langle A \rangle$ wird immer verlässlicher.

3.6 Was man darüberhinaus wissen sollte

Fehler der Messwerte vs. Fehler des Mittelwertes: Bei der linearen Regression haben wir eine Normalverteilung für die Messfehler ϵ_i einzelner Messwerte y_i angenommen. Diese Messfehler der einzelnen Datenpunkte hatten wir ebenfalls mit dem Buchstaben σ bezeichnet. Dieser Wert ändert sich allerdings mit zunehmender Zahl der Datenpunkte nicht! Egal wie oft wir messen, die Unsicherheit jedes einzelnen Messpunktes bleibt stets gleich gross. Zur Unterscheidung von der Standardabweichung σ eines Schätzwertes führen wir die Notation σ_y ein. Während σ_y unabhängig von der Zahl der Messwerte ist, nimmt die Standardabweichung eines Schätzwertes (Mittelwertes der Posteriorverteilung) σ in der Regel proportional zu $1/\sqrt{N}$ ab, wobei N die Zahl der Messpunkte ist.

Ungleichung von Tschebyscheff: Die sog. Ungleichung von Tschebyscheff besagt, dass die Gesamtwahrscheinlichkeit P im Intervall $\langle x \rangle \pm t\sqrt{\mathrm{Var}(x)}$ der Ungleichung

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

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

Schätzwert oder volle Posteriorverteilung? Es ist klar, dass die abgekürzte Schreibweise $\langle x \rangle \pm \sqrt{\mathrm{Var}(x)}$ nicht die volle Information über die Posteriorwahrscheinlichkeit vermittelt. Wenn möglich, ist deren Angabe daher von grossem Wert. Zum Beispiel können aus dem Mittelwert und der Varianz der Gammaverteilung 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 $\sigma^2 = \mathrm{Var}(x)$ an.

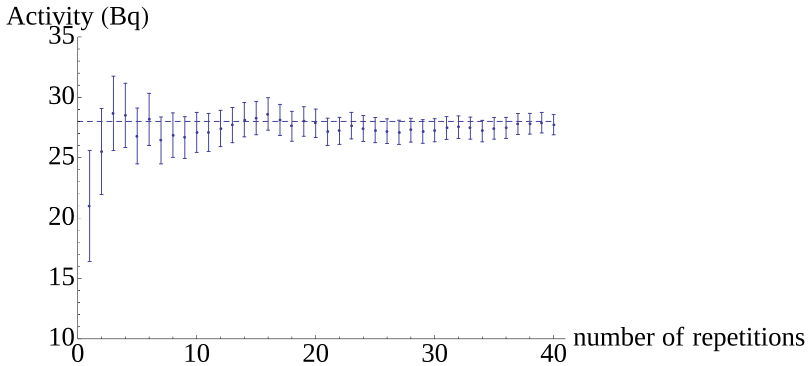

Abbildung 12: Entwicklung der Mittelwerte und ihrer $\pm\sigma$-Fehlerbalken bei wiederholter Messung der Aktivität einer radioaktiven Probe. Die gestrichelte Linie stellt den wahren Wert der Aktivität dar. Der Mittelwert nähert sich mit zunehmender Messzeit dem wahren Wert an, die Fehlerbalken werden zunehmend kleiner.

Graphische Darstellung mit Fehlerbalken: Zahlenwerte der Form $x = \langle x \rangle \pm \sigma$ werden häufig grafisch als Punkte mit Fehlerbalken dargestellt. Dabei ist in der Bildunterschrift genau anzugeben, welche Länge die Fehlerbalken haben. Verbreitet sind Darstellungen von $\langle x \rangle \pm \sqrt{\mathrm{Var}(x)}$ oder $\langle x \rangle \pm 3\sqrt{\mathrm{Var}(x)}$. Abbildung 12 zeigt als Beispiel die Entwicklung des Mittelwerts (Schätzwerts) der Aktivität und seines Fehlers bei zunehmender Zahl wiederholter Messungen von je 1 s. Durch die N-fache Wiederholung der Messung erhöht sich die Messzeit faktisch auf N Sekunden, so dass die Standardabweichung $\sigma = \sqrt{\langle A \rangle / \Delta t}$ proportional zu $1/\sqrt{N}$ kleiner wird. Entsprechend kommt der Mittelwert $\langle A \rangle$ dem wahren Wert immer näher.

4 Näherung von Verteilungsfunktionen durch eine Gausskurve: Methode von Laplace

Es ist sehr häufig möglich, die Posteriorverteilung durch eine Gausskurve anzunähern. In diesem Fall entwickelt man den Logarithmus der Verteilungsfunktion $\mathrm{pdf}(x) = f(x)$ um die Mode x_0 herum in eine Taylorreihe bis zur 2. Ordnung:

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

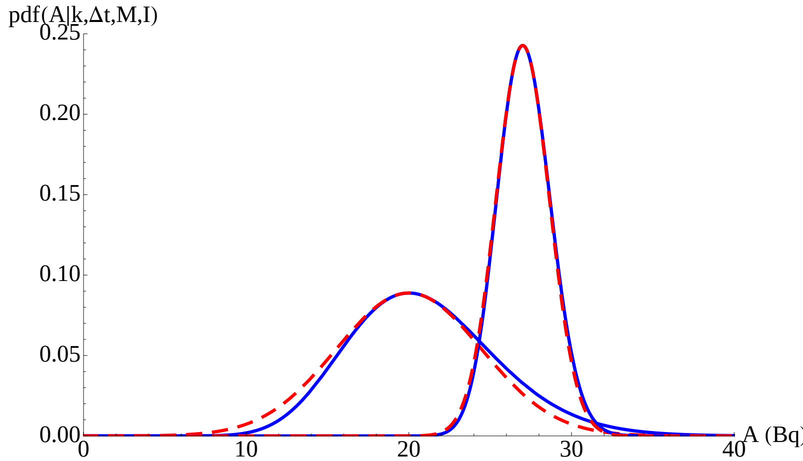

Abbildung 13: Gauss'sche Näherung (rot gestrichelt) von Gammaverteilungen (blau durchgezogen) bei einer Messzeit $\Delta t = 1\,\mathrm{s}$ und $\Delta t = 10\,\mathrm{s}$. Bei der grösseren Messzeit ist auch die Näherung besser.

Definiert man

$$\frac{1}{\sigma^2} = -\left.\frac{d^2}{dx^2}\ln f(x)\right|_{x=x_0},$$

so findet man die Näherung

$$\mathrm{pdf}(x) = f(x) \approx f(x_0)\exp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right).$$

Diese Näherung ist vor allem für Verteilungsfunktionen geeignet, die aus einer grossen Zahl von wiederholten Messungen hervorgegangen ist. Diese Verteilungen tendieren gegen die Normalverteilung. Als Beispiel sehen wir in Abb. 13 die Gauss'sche Näherung von Gammaverteilungen, die zu verschiedenen Messzeiten im Experiment des radioaktiven Zerfalls gehören. Bei der Messzeit $\Delta t = 1\,\mathrm{s}$ ist die Näherung weniger gut, als bei der Messzeit $\Delta t = 10\,\mathrm{s}$.

Insbesondere beim Lösen von Integralen über $\mathrm{pdf}(x)$ ist diese Näherung oft sehr hilfreich, weil Gaussche Integrale recht gut bekannt sind. Zum Beispiel hat man näherungsweise

$$\int dx\, h(x)\mathrm{pdf}(x) \approx \sqrt{\frac{2\pi}{-\left.\frac{d^2}{dx^2}\ln f(x)\right|_{x=x_0}}}\,h(x_0)\mathrm{pdf}(x_0)$$

Die Methode lässt sich auch auf Verteilungsfunktionen anwenden, die nur numerisch bekannt sind.

5 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

$$H[\text{prob}(n)] = -\sum_n \text{prob}(n) \log_2 \text{prob}(n)$$

zu. Die Einheit dieser Entropie wird 'Bit' (binary digit) genannt. Beim Werfen einer Münze mit zwei möglichen Ergebnissen $n \in \{0, 1\}$ und $\text{prob}(0) = \text{prob}(1) = 1/2$ haben wir tatsächlich $H = 1\,\text{bit}$. 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 lautet

$$H[\text{pdf}(x)] = -\int dx \text{pdf}(x) \log_2 \text{pdf}(x).$$

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

$$\Delta I = H[f(x)] - H[g(x)]$$

gegeben. Diese Zahl gibt uns ganz konkret an, wie viele zusätzliche Bits wir zum Speichern des verfeinerten Schätzwertes zusätzlich bereitstellen müssen. Dabei entsprechen ca. 3.23 bit einer Dezimalstelle. Gewinnt man zum Beispiel durch wiederholtes Messen beispielsweise $\Delta I = 6.64\,\text{bit}$ Information über einen Parameter, dann kann ich den Parameterwert mit der grösseren Genauigkeit von 2 zusätzlichen Dezimalstellen angeben.

Abbildung 14 zeigt den Informationszuwachs, der mit der Erhöhung der Messzeit beim radioaktiven Zerfall einhergeht. Wie bereits weiter oben erklärt, erhöht die Zahl N der Wiederholungen von 1 s Messungen die Messzeit faktisch auf N Sekunden. Dieser zusätzliche Zeitaufwand beim Messen schlägt sich in der Zunahme der Information nieder, die man über die Aktivität gewinnt. In der Regel wächst die Information logarithmisch in N.

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.

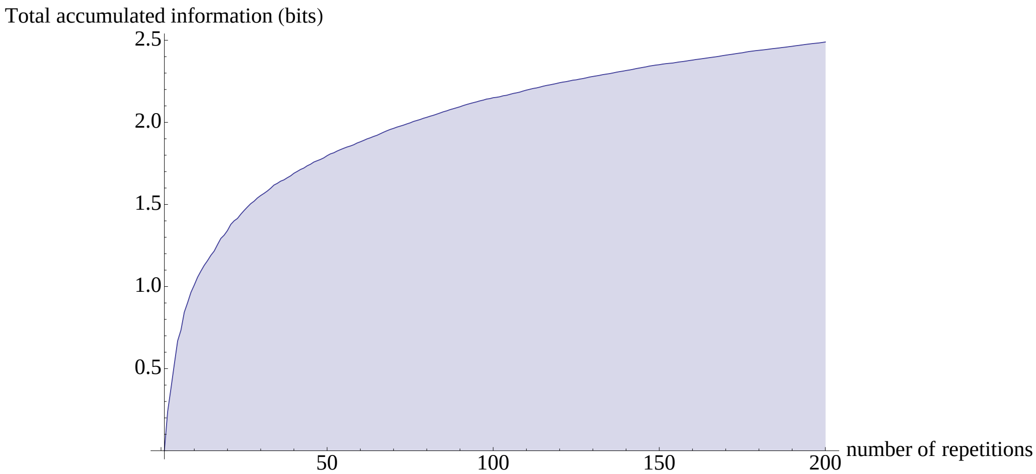

Abbildung 14:

6 Aufgaben und Fragen zum tieferen Verständnis

1. Beschreiben Sie den Unterschied zwischen den folgenden Begriffen möglichst genau:

   a) dem *empirischen Mittelwert* und dem *Mittelwert einer Verteilungsfunktion*

   b) der *empirischen Varianz* und der *Varianz einer Verteilungsfunktion*

   c) dem *Fehler der Messwerte* und dem *Fehler des Mittelwerts*.

2. Eine Verteilungsfunktion $\mathrm{pdf}(x)$ habe ein einziges Maximum bei x_0, eine endliche Breite, aber eine Asymmetrie, die Werten $x - x_0 > 0$ eine grössere Wahrscheinlichkeitsdichte zuordnet, als Werten $x - x_0 < 0$, bei denen $|x - x_0|$ jeweils gleich ist. Welche Auswirkung hat das auf die Lage des Medians im Vergleich zur Mode? Welche Auswirkung hat das auf die Lage des Mittelwerts (wir nehmen an, dass er existiert) im Vergleich zur Mode?

3. Berechnen Sie die Varianzen der Normalverteilung, der Gammaverteilung, der Student'schen t-Verteilung und der inversen Gammaverteilung.

4. Die Cauchyverteilung (Lorentzverteilung, Breit–Wigner-Verteilung) hat die Form

$$\mathrm{pdf}(x|\mu, \sigma) = \frac{1}{\pi\sigma} \left(1 + \frac{(x-\mu)^2}{\sigma^2}\right)^{-1}.$$

Bestimmen Sie die Mode, den Median, den Mittelwert und die Varianz dieser Verteilung.

5. Gilt für Verteilungsfunktionen, die symmetrisch bezüglich der Mode sind, stets, dass der Mittelwert mit der Mode zusammen fällt? Wie verhält es sich mit dem Median?