Vorlesung zur Datenanalyse

Datenanalyse ist die mathematische Form des 'gesunden Menschenverstands'—Pierre Simon de Laplace.

Thomas Ihn

5. Oktober 2014

Lernziele Lektion 1

- Sie kennen die Binomialverteilung, die Poissonverteilung, sowie die Gammaverteilung und ihre Anwendung auf den radioaktiven Zerfall.
- Sie kennen fundamentalen Begriffe und Regeln der Wahrscheinlichkeitsrechnung: Produktregel, Marginalisierungsregel, Bayes'sches Theorem, Normierung von Wahrscheinlichkeiten, bedingte Wahrscheinlichkeit.
- Sie verstehen anhand des Beispiels, wie man mit Hilfe des Bayes'schen Theorems von einer Vorhersage für Messergebnisse bei bekannten Parametern zu einer Abschätzung der Parameter bei bekannten Messergebnissen gelangen kann (Invertieren einer bedingten Wahrscheinlichkeit).
- Sie können die Aktivität einer radioaktiven Quelle aus einer Teilchendetektormessung abschätzen.

1 Radioaktiver Zerfall als Bernoulliexperiment

1.1 Das physikalische Modell des Experiments

Hintergrundinformation *I***.** Wir betrachten eine radioaktive Quelle, die eine grosse (aber unbekannte) Zahl N von ^{22}Na-Atomen enthält. Die Atome zerfallen gemäss

$$^{22}\mathrm{Na} \longrightarrow {}^{22}\mathrm{Ne} + \mathrm{e}^{+} + \nu_{\mathrm{e}}, \quad \text{Halbwertszeit: } 2.6027\,\mathrm{y},$$

unter Emission eines Positrons und eines Elektronneutrinos. Jedes emittierte Positron gibt 1 'Event' im Detektor (siehe Abb. 1). Wir haben innerhalb von 1 Sekunde 20 Events gezählt.

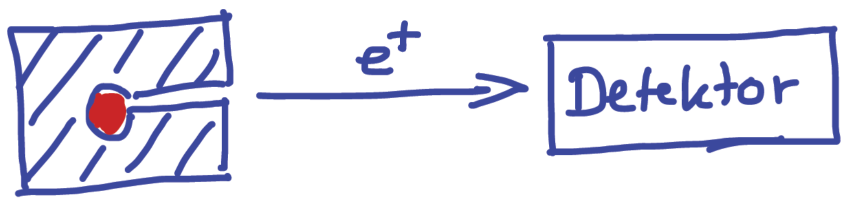

Abbildung 1: Versuchsaufbau, bei dem eine radioaktive Quelle Positronen in Richtung eines Detektors emittiert. Der Detektor zählt die emittierten Positronen über ein gewisses Zeitintervall Δt.

Frage: Wie gross ist die Aktivität der Quelle?

Modell M**:** Jedes ^{22}Na-Atom zerfällt im Zeitintervall Δt mit der Wahrscheinlichkeit $r\Delta t$, wobei r die Zerfallsrate ist. Die Aktivität der Quelle entspricht der mittleren Anzahl von Zerfällen pro Sekunde und ist gegeben durch

$$A = Nr \geq 0 \quad \text{Einheit: } 1\,\text{Becquerel} = 1\,\mathrm{Bq} = 1\,\mathrm{s}^{-1}.$$

Da das Messintervall viel kleiner ist als die Halbwertszeit, ist die Aktivität der Quelle zeitlich praktisch konstant. Der Zerfall eines einzelnen Atoms geschieht zufällig. Wir fassen das als Zufallsexperiment mit zwei möglichen Ergebnissen auf:

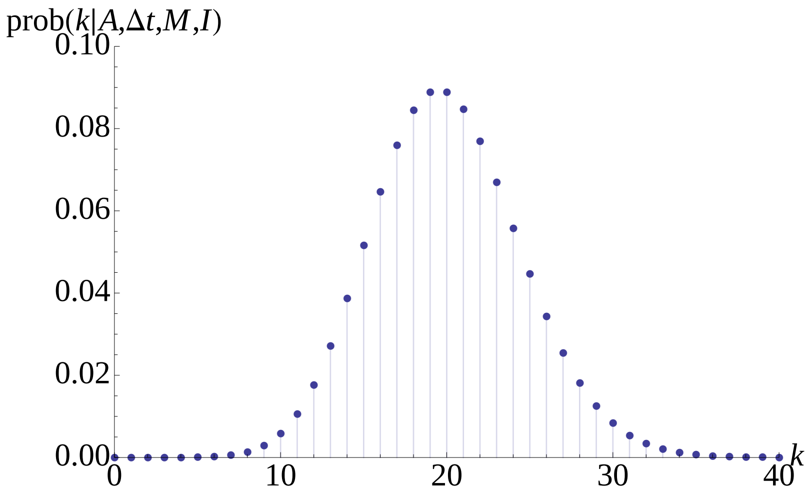

Abbildung 2: Binomialverteilung (1) für $N = 20'000$ und $p = r\Delta t = 0.001$.

Ergebnis	Wahrscheinlichkeit
Atom zerfällt	$p = r\Delta t$
Atom zerfällt nicht	$1 - p = 1 - r\Delta t$

Da wir das Zufallsexperiment mit N Atomen machen, ist das wie beim N-fachen Werfen einer Münze ein Bernoulli-Experiment. Die Wahrscheinlichkeit dabei k Treffer zu erzielen ist durch die Binomial-Verteilung

$$\text{prob}(k|N, r\Delta t, M, I) = B(k; N, r\Delta t) = \begin{pmatrix} N \\ k \end{pmatrix} (r\Delta t)^k (1 - r\Delta t)^{N-k} \qquad (1)$$

gegeben. Wir lesen $\text{prob}(k|N, r\Delta t, M, I)$ als 'die Wahrscheinlichkeit k Events zu messen, wenn (unter der Bedingung dass) wir N Atome haben (d.h. wir kennen die Zahl N), die Zerfallswahrscheinlichkeit eines Atoms $r\Delta t$ ist (d.h. wir kennen $p = r\Delta t$), dass unser physikalisches Modell M stimmt, und dass unsere Hintergrundinformation I korrekt ist'. Das Modell M und die Hintergrundinformation I sind oben gekennzeichnet.

Die Wahrscheinlichkeit (1) macht eine Vorhersage über die Zahl der zu erwartenden Events in einem Experiment. Ein Beispiel der Binomialverteilung ist in Abb. 2 abgebildet.

Bevor wir die obige Frage nach der Aktivität der Quelle weiter verfolgen, entwickeln wir noch eine sehr nützliche und gebräuchliche Näherung der Binomialverteilung für grosse N

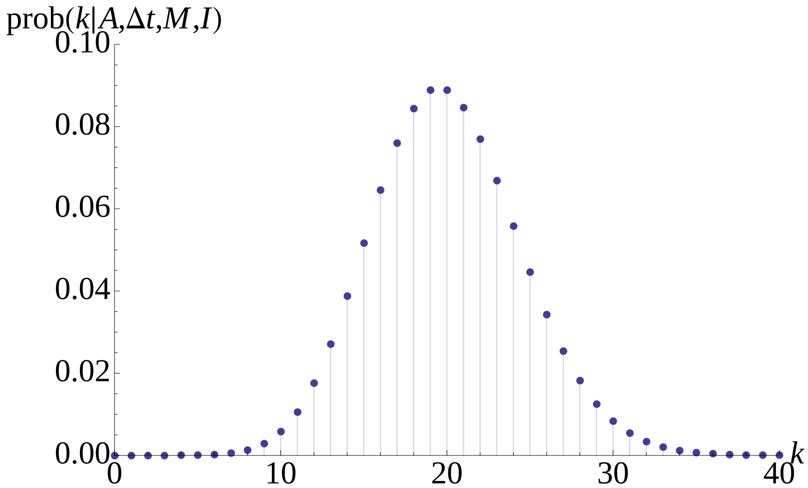

Abbildung 3: Poissonverteilung (2) für $\mu = 20$.

und kleine Δt, unter der Annahme, dass $\mu = N r \Delta t$ endlich bleibt. Wir haben

$$\begin{pmatrix} N \\ k \end{pmatrix} (r\Delta t)^k = \frac{N!}{k!(N-k)!}(r\Delta t)^k = \frac{1}{k!} \underbrace{N(N-1)(N-2)\dots(N-k+1)}_{k \text{ Faktoren}} \left(\frac{\mu}{N}\right)^k$$

$$= \frac{1}{k!}\left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\dots\left(1-\frac{k-1}{N}\right)\mu^k \approx \frac{\mu^k}{k!}$$

und

$$(1 - r\Delta t)^{N-k} = \left(1 - \frac{\mu}{N}\right)^{N-k} \approx e^{-\mu},$$

so dass

$$\mathrm{prob}(k|N, r\Delta t, M, I) \approx P(k;\mu) = \frac{\mu^k}{k!}e^{-\mu}. \qquad (2)$$

Die Verteilung $P(k;\mu)$ heisst Poissonverteilung. Wir erinnern uns, dass $\mu = A\Delta t$, wobei A die Aktivität der Quelle und Δt das Messintervall ist. Ein Beispiel für die Poissonverteilung ist in Abb. 3 abgebildet.

Gleichung (2) sagt [wie auch (1)] voraus, wie plausibel es ist k Events zu messen, wenn die Aktivität der Quelle und auch das Messintervall bekannt sind. Wir stellen aber die umgekehrte Frage: in unserem Fall ist die Zahl der gemessenen Events bekannt, aber die Aktivität ist gesucht. Wie können wir die Aktivität abschätzen?

1.2 Zentrale Regeln und Begriffe der Wahrscheinlichkeitstheorie

Eine Analogie: Wir betrachten zwei Paare Aussagen über die Wähler eines Landes.

G:	ein Wähler wählt grün
$\overline{\text{G}}$:	ein Wähler wählt nicht grün
US:	ein Wähler befürwortet die Umweltsteuer
$\overline{\text{US}}$:	ein Wähler ist gegen die Umweltsteuer

Wir bilden eine Wahrheitstabelle der Form

	US	$\overline{\text{US}}$	
G	prob(G,US)	prob(G,$\overline{\text{US}}$)	prob(G)
$\overline{\text{G}}$	prob($\overline{\text{G}}$,US)	prob($\overline{\text{G}}$,$\overline{\text{US}}$)	prob($\overline{\text{G}}$)
	prob(US)	prob($\overline{\text{US}}$)	

mit Zahlen:

	US	$\overline{\text{US}}$	
G	7.5%	2.5%	10%
$\overline{\text{G}}$	60%	30%	90%
	67.5%	32.5%	100%

Dabei bezeichnet prob(G,US) die gemeinsame Wahrscheinlichkeit (Verbundwahrscheinlichkeit), dass ein Wähler grün wählt und die Umweltsteuer befürwortet (logisches 'und'). Wir sehen unmittelbar, dass

$$\begin{aligned}\text{prob(G)} &= 7.5\% + 2.5\% = 10\% \\ \text{prob(US)} &= 7.5\% + 60\% = 67.5\%\end{aligned}$$

und leiten daraus die allgemeine Marginalisierungsregel der Wahrscheinlichkeitsrechnung

$$\begin{aligned}\text{prob(G)} &= \text{prob(G, US)} + \text{prob(G, }\overline{\text{US}}) \\ \text{prob(US)} &= \text{prob(G, US)} + \text{prob(}\overline{\text{G}}\text{, US)} \qquad (3)\end{aligned}$$

ab. In diesem Zusammenhang können wir prob(G) lesen als 'die Wahrscheinlichkeit, dass ein Wähler grün wählt, egal ob er die Umweltsteuer befürwortet oder nicht'. 'Marginalisieren' bedeutet, dass wir eine Grösse, die uns nicht interessiert (hier: die Haltung zur Umweltsteuer) wegsummieren.[1]

Mit Hilfe der selben Tabelle führen wir den Begriff der bedingten Wahrscheinlichkeit ein: die Wahrscheinlichkeit, dass ein Wähler die Umweltsteuer befürwortet, wenn (unter der Bedingung dass) er grün wählt, ist

$$\text{prob(US|G)} = \frac{7.5\%}{10\%} = 75\%.$$

[1] in der Quantenmechanik sagt man, dass man eine Grösse 'ausintegriert'

Entsprechend ist die Wahrscheinlichkeit, dass ein Wähler, der die Umweltsteuer befürwortet, grün wählt

$$\mathrm{prob}(\mathrm{G}|\mathrm{US}) = \frac{7.5\%}{67.5\%} = 11.1\%.$$

Daraus leiten wir die allgemeine Definition der bedingten Wahrscheinlichkeit ab:

$$\mathrm{prob}(\mathrm{US}|\mathrm{G}) = \frac{\mathrm{prob}(\mathrm{G},\mathrm{US})}{\mathrm{prob}(\mathrm{G})}. \tag{4}$$

Daraus folgt die Produktregel der Wahrscheinlichkeitsrechnung

$$\begin{aligned}\mathrm{prob}(\mathrm{G},\mathrm{US}) &= \mathrm{prob}(\mathrm{G})\mathrm{prob}(\mathrm{US}|\mathrm{G}) \\ &= \mathrm{prob}(\mathrm{US})\mathrm{prob}(\mathrm{G}|\mathrm{US}) \end{aligned} \tag{5}$$

An unserem Beispiel sehen wir auch, dass

$$\mathrm{prob}(\overline{\mathrm{US}}|\mathrm{G}) = \frac{2.5\%}{10\%} = 25\%,$$

so dass

$$\mathrm{prob}(\mathrm{US}|\mathrm{G}) + \mathrm{prob}(\overline{\mathrm{US}}|\mathrm{G}) = 1, \tag{6}$$

was die Tatsache ausdrückt, dass sich Wahrscheinlichkeiten zu 1 (= 100%) addieren (integrieren). Wichtig ist es zu bemerken, dass prob(US|G) eine Wahrscheinlichkeit für US, nicht jedoch für G angibt. Daher ist diese bedingte Wahrscheinlichkeit bezüglich G auch nicht normiert.

Wegen der Gleichungen (3), (4) und (5) gilt

$$\begin{aligned}\mathrm{prob}(\mathrm{US}|\mathrm{G}) = \frac{\mathrm{prob}(\mathrm{US},\mathrm{G})}{\mathrm{prob}(\mathrm{G})} &= \frac{\mathrm{prob}(\mathrm{US})\mathrm{prob}(\mathrm{G}|\mathrm{US})}{\mathrm{prob}(\mathrm{G},\mathrm{US}) + \mathrm{prob}(\mathrm{G},\overline{\mathrm{US}})} \\ &= \frac{\mathrm{prob}(\mathrm{US})\mathrm{prob}(\mathrm{G}|\mathrm{US})}{\mathrm{prob}(\mathrm{US})\mathrm{prob}(\mathrm{G}|\mathrm{US}) + \mathrm{prob}(\overline{\mathrm{US}})\mathrm{prob}(\mathrm{G}|\overline{\mathrm{US}})}.\end{aligned}$$

Dies ist das Bayes'sche Theorem, das zusammengefasst lautet

$$\mathrm{prob}(\mathrm{US}|\mathrm{G}) = \frac{\mathrm{prob}(\mathrm{US})\mathrm{prob}(\mathrm{G}|\mathrm{US})}{\mathrm{prob}(\mathrm{US})\mathrm{prob}(\mathrm{G}|\mathrm{US}) + \mathrm{prob}(\overline{\mathrm{US}})\mathrm{prob}(\mathrm{G}|\overline{\mathrm{US}})}. \tag{7}$$

Das Bayes'sche Theorem erlaubt uns die Invertierung der bedingten Wahrscheinlichkeit prob(G|US) → prob(US|G). Es spielt eine Schlüsselrolle beim Rückschliessen von Messwerten auf unbekannte Parameter.

1.3 Abschätzung der Aktivität der Quelle

Zurück zu unserem Messproblem: Wir kennen $\mathrm{prob}(k|A)$ und können mit Hilfe des Bayes'schen Theorems $\mathrm{prob}(A|k)$ bestimmen. Auf unseren Fall angewendet lautet es[2]

$$\mathrm{prob}(A|k) = \frac{\mathrm{prob}(A)\mathrm{prob}(k|A)}{\sum_A \mathrm{prob}(A)\mathrm{prob}(k|A)}.$$

Der Nenner ist unabhängig von A und sorgt dafür, dass die Wahrscheinlichkeitsverteilung $\mathrm{prob}(A|k)$ normiert ist. Wenn wir uns zunächst einmal auf die funktionale Abhängigkeit der Wahrscheinlichkeit $\mathrm{prob}(A|k)$ von A beschränken, dann ist

$$\mathrm{prob}(A|k) \propto \mathrm{prob}(A)\mathrm{prob}(k|A).$$

Eine Schwierigkeit bei der Anwendung dieser Formel ist, dass wir $\mathrm{prob}(A)$ nicht kennen. Was können wir über die Plausibilität verschiedener Werte von A sagen, <u>bevor</u> wir die Zahl k der Events kennen?

Wir wissen zumindest, dass $A \geq 0$, denn die Aktivität der Quelle kann per Definition nie negativ sein. Ansonsten drücken wir die Tatsache, dass wir keine weiteren Anhaltspunkte für den Wert von A haben dadurch aus, dass

$$\mathrm{prob}(A) = \begin{cases} \text{const.} & \text{für } A \geq 0 \\ 0 & \text{für } A < 0. \end{cases}$$

Dies nennt man das <u>Indifferenzprinzip</u> (Prinzip vom unzureichenden Grund, siehe Laplace 1812).

Nun ist also

$$\mathrm{prob}(A|k) \propto \mathrm{prob}(k|A) = \frac{(A\Delta t)^k}{k!} e^{-A\Delta t}.$$

Wir können diese Verteilungsfunktion für A normieren, indem wir fordern, dass $\sum_A \mathrm{prob}(A|k) = 1$ [siehe Gl. (6)]. Wir lösen also

$$\sum_A \mathcal{N} \frac{(A\Delta t)^k}{k!} e^{-A\Delta t} = \mathcal{N} \int_0^\infty \frac{(A\Delta t)^k}{k!} e^{-A\Delta t} dA = \frac{\mathcal{N}}{k!\Delta t} \int_0^\infty d\mu \mu^k e^{-\mu} = \frac{\mathcal{N}}{k!\Delta t} \times k! = \frac{\mathcal{N}}{\Delta t} = 1$$

und finden

$$\mathcal{N} = \Delta t.$$

Daher ist

$$\mathrm{prob}(A|k) = \Delta t \frac{(A\Delta t)^k}{k!} e^{-A\Delta t} dA \tag{8}$$

[2] Detail zur Notation: $\mathrm{prob}(A|k) = \mathrm{pdf}(A|k)dA$, wobei $\mathrm{pdf}(A|k)$ die Wahrscheinlichkeitsdichte von A ist.

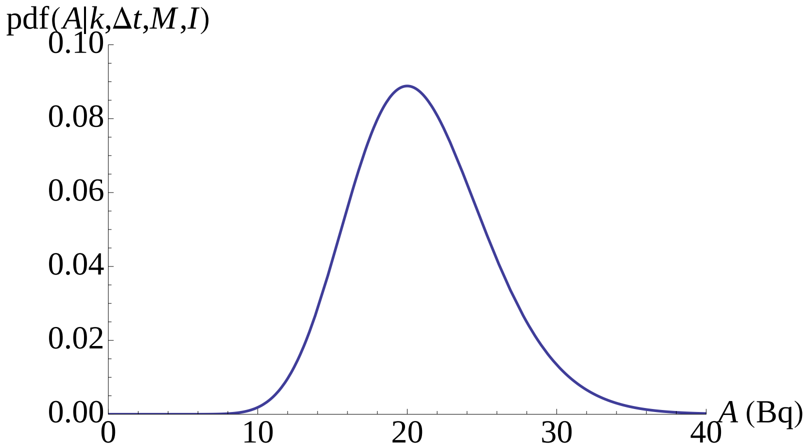

Abbildung 4: Gammaverteilung (9) für $k = 20$ und $\Delta t = 1\,\mathrm{s}$.

Das ist eine Wahrscheinlichkeitsverteilung für die Aktivität A, bei gegebener Zahl von gemessenen Events. Da die Werte von A beliebige positive reelle Zahlen (einschliesslich der Null) sein können, bezeichnen wir

$$\mathrm{pdf}(A|k) = \Delta t \frac{(A\Delta t)^k}{k!} e^{-A\Delta t}$$

als Wahrscheinlichkeitsdichte von A. Diese Wahrscheinlichkeitsdichte ist in Abb. 4 dargestellt.

Diese Wahrscheinlichkeitsdichteverteilung gehört zur Familie der Gamma-Verteilungen

$$\Gamma(\mu;\alpha,\beta) = \begin{cases} \frac{(\mu/\beta)^{\alpha-1} e^{-\mu/\beta}}{\beta\Gamma(\alpha)} & \text{für } \mu > 0 \\ 0 & \text{sonst.} \end{cases} \tag{9}$$

mit den Parametern α (Formparameter) und β (Skalierungsparameter). Wir haben $\alpha = k+1$ und $\beta = 1/\Delta t$, so dass

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

Die Messung erlaubt uns also eine Wahrscheinlichkeitsaussage über die Aktivität A zu machen. Den wahrscheinlichsten Wert für A finden wir als Lösung von

$$\frac{d}{dA}\mathrm{pdf}(A|k) = 0 \Rightarrow A = \frac{k}{\Delta t} = \frac{20}{1\,\mathrm{s}} = 20\,\mathrm{Bq}.$$

Allerdings hat die Verteilung eine beachtliche Breite um das Maximum herum. Der Wert von A liegt z.B. mit einer Wahrscheinlichkeit von 99.85% im Intervall $[9\,\mathrm{Bq}, 38\,\mathrm{Bq}]$, denn

$$\int_9^{38} \mathrm{pdf}(A|k)dA = 0.9985.$$

Aufgaben und Fragen zum tieferen Verständnis:

1. Wie würde sich im obigen Beispiel das Ergebnis für $\mathrm{pdf}(A|k, \Delta t, M, I)$ verändern, wenn wir für $\mathrm{prob}(A)$ anstelle des Indifferenzprinzips die Verteilung $\Gamma(A; 25, 1\,\mathrm{Bq})$ verwenden würden? Bemerkung: wir könnten diese Verteilung zum Beispiel aus einer früheren Messung erhalten haben.

2. Im obigen Beispiel haben wir die Binomialverteilung zunächst durch die Poissonverteilung genähert, und dann mit Hilfe des Bayes'schen Theorems invertiert, um zur Gammaverteilung für die Aktivität A zu gelangen. Führen sie dieselbe Rechnung direkt mit der Binomialverteilung durch, ohne die Näherung durch die Poissonverteilung zu verwenden. Recherchieren Sie, wie die Familie der Verteilungsfunktionen heisst, die Sie als Ergebnis erhalten.

3. **The false positive paradox:** Mediziner haben einen neuen Test erfunden, mit dem sie eine bestimmte seltene Krankheit diagnostizieren wollen. Die Wahrscheinlichkeit, dass eine Person der Bevölkerung diese Krankheit hat ist 0.0001%. Der Test fällt bei Personen, welche die Krankheit nicht haben nur in 1% der Fälle positiv aus, während er bei Personen, welche die Krankheit haben, nur in 0.00001% der Fälle negativ ausfällt. Sie lassen sich testen und der Test fällt positiv aus. Wie gross ist die Wahrscheinlichkeit, dass sie die Krankheit tatsächlich haben? Welche Bedingung muss ein Test erfüllen, damit er die seltene Krankheit mit grosser Sicherheit diagnostizieren kann? Diskutieren sie die Implikationen Ihres Ergebnisses für die Entwicklung von Tests zur Diagnostizierung seltener Krankheiten.