

5 Fehlerfortpflanzung

Diese Lektion folgt D.S. Sivia, *Data Analysis: A Bayesian Tutorial*, Oxford University Press, Chapter 3.6.

Lernziele Lektion 5

- Sie wissen, wie man Wahrscheinlichkeitsverteilungen von einer oder mehreren Variablen x auf eine oder mehrere Variablen y transformiert.
- Sie sind in der Lage solche Transformationen für ein- und zweidimensionale Wahrscheinlichkeitsverteilungen durchzuführen.
- Sie kennen die Näherungen, die zur Gauss'schen Fehlerfortpflanzung führen und können entscheiden, ob ein gegebenes Problem mit dieser Näherung behandelt werden kann.

5.1 Die Fragestellung

Beginnen wir mit einem Beispiel: bei der Messung des Halleffektes haben wir die Steigung R_H der Hallgeraden und ihre Standardabweichung mit Hilfe der linearen Regression bestimmt. Wie gross sind nun die Elektronendichte $n_s = 1/eR_H$ und ihre Standardabweichung?

Im Allgemeinen führen häufig die Parameter des Modells für eine Messung nicht unmittelbar auf die physikalischen Grössen, die durch die Messung ermittelt werden sollen. Vielmehr müssen diese Grössen mit Hilfe von Messergebnissen gemäss einer Formel berechnet werden. Dabei stellt sich die Frage, wie sich die Posteriorverteilungen der Parameter, die wir bereits bestimmt haben, in die Verteilungen der physikalischen Grössen von Interesse transformieren. Insbesondere interessiert uns auch, wie sich die Standardabweichung dabei verändert. Diese Fragestellung lässt sich folgendermassen präzisieren:

Fragestellung: Wir haben die Posteriorverteilung $\text{pdf}(x, y|I)$ der Parameter x und y aus Messdaten bestimmt. Wie finden wir $\text{pdf}(z|I)$, wobei $z = h(x, y)$?

5.2 Erster Fall: nur ein Parameter

Gegeben sei $\text{pdf}(x|I) = f(x)$, gesucht sei $\text{pdf}(z|I) = g(z)$, wobei $z = h(x)$ ist. Jedem Wert x_0 entspricht ein korrespondierender Wert $z_0 = h(x_0)$. Betrachten wir ein kleines Intervall $[x_0 - \delta x/2, x_0 + \delta x/2]$, dann ist die Wahrscheinlichkeit, dass x in diesem Intervall liegt, gegeben durch

$$\text{prob}(x_0 - \delta x/2 \leq x \leq x_0 + \delta x/2|I) = \int_{x_0 - \delta x/2}^{x_0 + \delta x/2} dx \text{pdf}(x|I) \approx \text{pdf}(x_0|I)\delta x.$$

Diese Wahrscheinlichkeit ist im linken Teil der Abb. 25 grau schraffiert dargestellt. Diesem

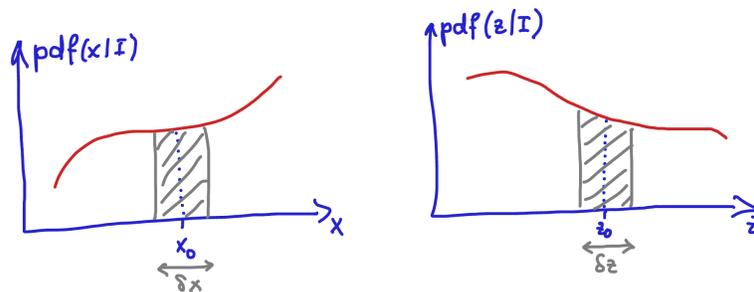


Abbildung 25: Schematische Veranschaulichung der Koordinatentransformation. Grau schraffierte Flächen entsprechen einander und repräsentieren die gleiche Wahrscheinlichkeit.

Intervall δx in x entspricht ein Intervall δz in z (siehe Abb. 25, rechts). Entsprechend ist

$$\text{prob}(z_0 - \delta z/2 \leq z \leq z_0 + \delta z/2|I) \approx \text{pdf}(z_0|I)\delta z.$$

Wir fordern nun, dass diese beiden Wahrscheinlichkeiten im Grenzfall beliebig kleiner Intervalle δx und δz gleich sind, also

$$\text{pdf}(x|I)\delta x = \text{pdf}(z|I)\delta z. \quad (33)$$

Daraus folgt sofort

$$f(x) = g[h(x)] \times \left| \frac{dh(x)}{dx} \right|.$$

Das ist ein einfaches Beispiel für eine Koordinatentransformation.

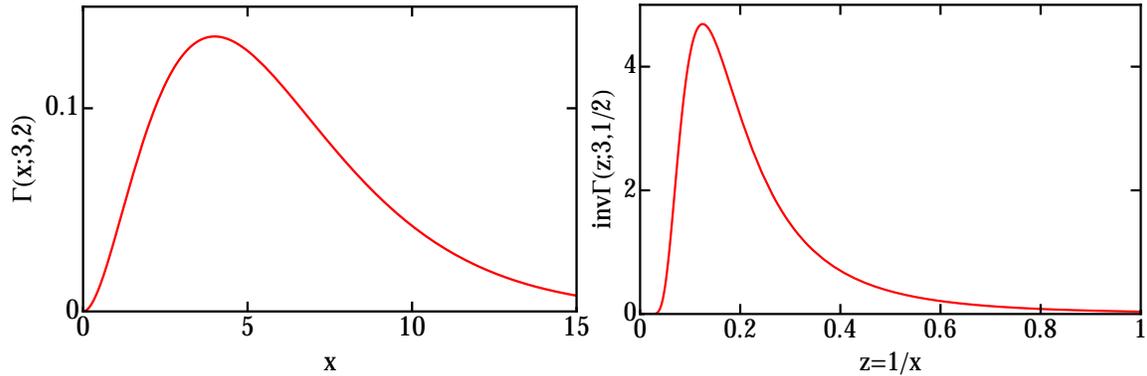


Abbildung 26: Links: die Gammaverteilung mit Parametern $\alpha = 3$ und $\beta = 2$. Rechts: die Inverse Gammaverteilung mit Parametern $\alpha = 3$ und $\tilde{\beta} = 1/\beta = 1/2$.

Beispiel: Gegeben sei die Gammaverteilung (siehe Abb. 26, links)

$$\Gamma(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta} \left(\frac{x}{\beta}\right)^{\alpha-1} e^{-x/\beta} \quad \text{für } x > 0.$$

Wie lautet die Verteilung für $z = 1/x$ ($z > 0$)?

Wir haben

$$f(x)dx = \frac{1}{\Gamma(\alpha)\beta} \left(\frac{x}{\beta}\right)^{\alpha-1} e^{-x/\beta} dx = g(z)dz,$$

so dass

$$g(z) = \frac{1}{\Gamma(\alpha)\beta} \left(\frac{1}{\beta z}\right)^{\alpha-1} e^{-1/\beta z} \times \left|-\frac{1}{z^2}\right| = \underbrace{\frac{1}{\Gamma(\alpha)\tilde{\beta}} \left(\frac{\tilde{\beta}}{z}\right)^{\alpha+1} e^{-\tilde{\beta}/z}}_{\text{inv}\Gamma(z; \alpha, \tilde{\beta})},$$

mit $\tilde{\beta} = 1/\beta$ (siehe Abb. 26, rechts). Beachte, dass das Maximum der abgebildeten Gammaverteilung bei 4 ist, wohingegen das Maximum der inversen Gammaverteilung *nicht* bei $1/4$, sondern bei $1/8$ ist. Ähnlich verhält es sich mit dem Mittelwert: der Mittelwert der abgebildeten Gammaverteilung ist $\langle x \rangle = 6$, wohingegen der Mittelwert der inversen Gammaverteilung bei $\langle z \rangle = 1/4 \neq 1/\langle x \rangle$ ist.

5.3 Zweiter Fall: M Parameter werden in M andere Parameter transformiert

In diesem Fall lautet die Bedingung, die Gl. (33) entspricht

$$\text{pdf}(x_1, x_2, \dots, x_M) \delta x_1 \delta x_2 \dots \delta x_M = \text{pdf}(z_1, z_2, \dots, z_M) \delta^M \text{Vol}(\{z_j\}).$$

Hier ist $\Delta^M \text{Vol}(\{z_j\})$ das infinitesimale Volumenelement, das die Abbildung des Hyperkubus $\delta x_1 \dots \delta x_M$ ist. Man findet

$$\text{pdf}(x_1, \dots, x_M) = \text{pdf}(z_1, \dots, z_M) \times \underbrace{\det \left(\frac{\partial(z_1 \dots z_M)}{\partial(x_1 \dots x_M)} \right)}_{\text{Determinante der Jacobi-Matrix}}.$$

Beispiel: Gegeben sei die zweidimensionale isotrope Normalverteilung (siehe Abb. 27, links)⁴

$$\text{pdf}(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right).$$

Wie lautet die Verteilung in Polarkoordinaten (R, θ)?

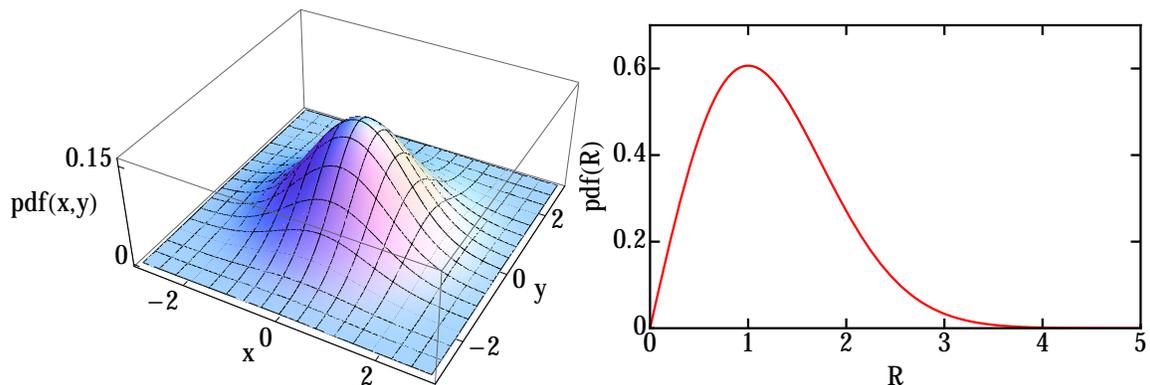


Abbildung 27: Links: zweidimensionale isotrope Normalverteilung. Rechts: Verteilungsfunktion für die Radiuskoordinate R in Polarkoordinaten.

Wir haben

$$x = R \cos \theta \quad \text{und} \quad y = R \sin \theta.$$

⁴Diese Wahrscheinlichkeitsverteilung könnte zum Beispiel die Aufenthaltswahrscheinlichkeit eines Teilchens im Grundzustand eines zweidimensionalen quantenmechanischen harmonischen Oszillators beschreiben.

Die Determinante der Jacobi-Matrix ist dann

$$\det \left(\frac{\partial(x, y)}{\partial(R, \theta)} \right) = \begin{vmatrix} \cos \theta & -R \sin \theta \\ -R \sin \theta & R \cos \theta \end{vmatrix} = R(\cos^2 \theta + \sin^2 \theta) = R.$$

Daher ergibt sich

$$\text{pdf}(R, \theta) = \text{pdf}(x, y) \times R = \frac{R}{2\pi\sigma^2} \exp\left(-\frac{R^2}{2\sigma^2}\right).$$

5.4 Dritter Fall: Wir transformieren von M auf $N < M$ Parameter

Dieser Fall lässt sich folgendermassen lösen:

1. Transformiere zunächst von M auf M Parameter (siehe oben).
2. Marginalisiere die überschüssigen Parameter.

Beispiel: Wie oben sei $\text{pdf}(x, y)$ die isotrope zweidimensionale Normalverteilung für die kartesischen Koordinaten (x, y) , die im linken Teil der Abb. 27 gezeigt ist. Wie lautet die Verteilung für R ?

1. Transformiere zunächst von (x, y) auf (R, θ) (siehe oben). Das ergibt

$$\text{pdf}(R, \theta) = \frac{R}{2\pi\sigma^2} \exp\left(-\frac{R^2}{2\sigma^2}\right).$$

2. Marginalisiere θ :

$$\text{pdf}(R) = \int_0^{2\pi} d\theta \text{pdf}(R, \theta) = \frac{R}{\sigma^2} \exp\left(-\frac{R^2}{2\sigma^2}\right).$$

Diese Verteilung ist in Abb. 27 rechts dargestellt. Interessanterweise geht die Wahrscheinlichkeitsdichte für $R \rightarrow 0$ linear gegen Null, obwohl die ursprüngliche Verteilung bei $(x, y) = (0, 0)$ gerade ihr Maximum hatte. Hier zeigt sich der grosse Einfluss der Determinante der Jacobimatrix, die den Faktor R beisteuert, der die transformierte Verteilung gegen Null gehen lässt.

5.5 Eine mathematisch formale Prozedur

Mit den obigen Beispielen haben wir die zentralen Elemente der Fehlerfortpflanzung kennen gelernt: Koordinatentransformation und Marginalisierung, oder eine Kombination der beiden. Mathematisch formal lässt sich das für die Transformation $z = h(x, y)$ folgendermassen ausdrücken:

$$\begin{aligned} \text{pdf}(z|I) &= \int dx \int dy \text{pdf}(x, y, z|I) = \int dx \int dy \text{pdf}(x, y|I) \text{pdf}(z|x, y, I) \\ &= \int dx \int dy \text{pdf}(x, y|I) \delta(z - h(x, y)). \end{aligned} \quad (34)$$

Diese Formel lässt sich leicht auf Verteilungen höherer Dimensionen verallgemeinern.⁵ Gleichung (34) hat eine einfache anschauliche Deutung. Betrachten wir die Funktion $h(x, y)$ als Oberfläche über der zweidimensionalen Parameterebene (x, y) , dann selektiert die Deltafunktion bei gegebenem z eine bestimmte Höhenlinie dieser Oberfläche. Die Wahrscheinlichkeitsdichte für z ergibt sich als Integral von $\text{pdf}(x, y)$ entlang dieser Höhenlinie.

Beispiel: Es sei $z = x + y$ und $\text{pdf}(x, y|I) = \text{pdf}(x|I)\text{pdf}(y|I)$ mit $\text{pdf}(x|I) = f(x)$ und $\text{pdf}(y|I) = g(y)$. Dann ist

$$\begin{aligned} \text{pdf}(z|I) &= \int dx \int dy \text{pdf}(x, y|I) \delta(z - x - y) = \int dx \int dy f(x)g(y) \delta(z - x - y) \\ &= \int dx f(x)g(z - x) \end{aligned}$$

Die Verteilung für die Summe zweier Parameter ist demnach die *Faltung* der beiden Verteilungen der Parameter. Nehmen wir insbesondere an, dass $f(x)$ und $g(y)$ Normalverteilungen

$$f(x) = \mathcal{N}(x; x_0, \sigma_x) \quad \text{und} \quad g(y) = \mathcal{N}(y; y_0, \sigma_y)$$

sind, dann ist

$$\text{pdf}(z = x + y|I) = \mathcal{N}\left(z; x_0 + y_0, \sqrt{\sigma_x^2 + \sigma_y^2}\right),$$

das heisst die Mittelwerte addieren sich gemäss

$$z_0 = x_0 + y_0$$

und die Fehler addieren sich quadratisch gemäss

$$\sigma_z^2 = \sigma_x^2 + \sigma_y^2.$$

⁵Diese Formulierung ist in gewisser Weise verwandt mit dem Konzept der Zustandsdichte eines quantenmechanischen Systems. Dabei spielen x und y die Rolle von Wellenvektoren, und $\text{pdf}(x, y)$ repräsentiert die konstante Dichte von Zuständen im k -Raum. Die z -Koordinate entspricht der Energie, während $h(x, y)$ die Dispersionsrelation ist.

5.6 Eine nützliche Näherung: das Gauss'sche Fehlerfortpflanzungsgesetz

Betrachten wir Gl. (34) unter dem Aspekt, dass wir Mittelwert und Standardabweichung von z ausrechnen wollen. Sei also $\text{pdf}(x, y|I) = f(x, y)$ gegeben, mit einem scharfen Maximum bei (x_0, y_0) . Zudem sei $z = h(x, y)$.

Das scharfe Maximum erlaubt uns eine Entwicklung von $\ln f(x, y)$ um (x_0, y_0) , so dass in guter Näherung

$$-\ln f(x, y) \approx -\ln f(x_0, y_0) - \frac{1}{2} \begin{pmatrix} x \\ y \end{pmatrix} H \begin{pmatrix} x \\ y \end{pmatrix}.$$

gilt. Die Matrix H ist die Hesse-Matrix der zweiten Ableitungen von $\ln f(x, y)$, ausgewertet bei (x_0, y_0) . In dieser Näherung ist $f(x, y)$ gaussförmig in zwei Dimensionen, so dass

$$\begin{aligned} \langle x \rangle &= x_0 \quad \text{und} \quad \langle y \rangle = y_0, \\ \langle (x - x_0)^2 \rangle &= \sigma_x^2 = \text{Var}(x), \\ \langle (y - y_0)^2 \rangle &= \sigma_y^2 = \text{Var}(y), \\ \langle (x - x_0)(y - y_0) \rangle &= \sigma_x \sigma_y \rho = \text{cov}(x, y). \end{aligned}$$

Die Grösse $\text{cov}(x, y)$ heisst Kovarianz der Parameter x und y . Ebenso entwickeln wir

$$h(x, y) \approx h(x_0, y_0) + \left. \frac{\partial h}{\partial x} \right|_{x_0, y_0} (x - x_0) + \left. \frac{\partial h}{\partial y} \right|_{x_0, y_0} (y - y_0).$$

Damit ist

$$\begin{aligned}
\langle z \rangle &\approx \int dz \int dx \int dy z f(x, y) \delta(z - h(x, y)) \\
&\approx \int dx \int dy h(x, y) f(x, y) \\
&\approx \int dx \int dy f(x, y) \left[h(x_0, y_0) + \frac{\partial h}{\partial x} \Big|_{(x_0, y_0)} (x - x_0) + \frac{\partial h}{\partial y} \Big|_{(x_0, y_0)} (y - y_0) \right] \\
&\approx h(x_0, y_0) \\
\langle z^2 \rangle &\approx \int dx \int dy f(x, y) \left[h(x_0, y_0) + \frac{\partial h}{\partial x} \Big|_{(x_0, y_0)} (x - x_0) + \frac{\partial h}{\partial y} \Big|_{(x_0, y_0)} (y - y_0) \right]^2 \\
&\approx h^2(x_0, y_0) + \left(\frac{\partial h}{\partial x} \Big|_{(x_0, y_0)} \right)^2 \langle (x - x_0)^2 \rangle + \left(\frac{\partial h}{\partial y} \Big|_{(x_0, y_0)} \right)^2 \langle (y - y_0)^2 \rangle \\
&\quad + \frac{\partial h}{\partial x} \Big|_{(x_0, y_0)} \frac{\partial h}{\partial y} \Big|_{(x_0, y_0)} \langle (x - x_0)(y - y_0) \rangle \\
\langle (z - \langle z \rangle)^2 \rangle &\approx \langle z^2 \rangle - \langle z \rangle^2 \\
&\approx \left(\frac{\partial h}{\partial x} \Big|_{(x_0, y_0)} \right)^2 \sigma_x^2 + \left(\frac{\partial h}{\partial y} \Big|_{(x_0, y_0)} \right)^2 \sigma_y^2 + \frac{\partial h}{\partial x} \Big|_{(x_0, y_0)} \frac{\partial h}{\partial y} \Big|_{(x_0, y_0)} \text{cov}(x, y).
\end{aligned}$$

Wir finden auf diese Weise das Gauss'sche Fehlerfortpflanzungsgesetz

$$\sigma_z = \sqrt{\left(\frac{\partial h}{\partial x} \Big|_{(x_0, y_0)} \right)^2 \sigma_x^2 + \frac{\partial h}{\partial x} \Big|_{(x_0, y_0)} \frac{\partial h}{\partial y} \Big|_{(x_0, y_0)} \text{cov}(x, y) + \left(\frac{\partial h}{\partial y} \Big|_{(x_0, y_0)} \right)^2 \sigma_y^2}.$$

Man sollte sich jedoch darüber im Klaren sein, dass wir einige Annahmen benötigt haben, um dieses Fehlerfortpflanzungsgesetz abzuleiten. Bei der praktischen Anwendung ist daher Vorsicht geboten. Wir demonstrieren das mit Hilfe der Gammaverteilung $\Gamma(x; 3, 2)$ und der Transformation $z = h(x) = 1/x$, die uns oben auf die inverse Gammaverteilung geführt hat. Die folgende Tabelle gibt in der ersten Spalte den Mittelwert und die Standardabweichung der zuvor betrachteten Gammaverteilung an. Die zweite Spalte gibt die exakten Ergebnisse für die transformierte Verteilung. Die dritte Spalte gibt das Ergebnis wieder, das durch Anwendung der Gausschen Fehlerfortpflanzung erzielt würde. Letzteres liegt offensichtlich weit neben den exakten Ergebnissen, weil die zugrunde liegende Gammafunktion nicht gut durch eine Normalverteilung genähert werden kann.

	$\Gamma(x; 3, 2)$	$\text{inv}\Gamma(z; 3, 1/2)$	Gauss'sche Näherung
Mittelwert $\langle x \rangle$	6	1/4	1/6
Standardabweichung σ_x	$\sqrt{12}$	1/4	$\sqrt{12}/36$

Dies ändert sich für grosse Werte von α . Zum Beispiel ist für $\alpha = 100$ die Gauss'sche Näherung schon recht gut brauchbar.

Beispiele für die Anwendung der Gauss'schen Fehlerfortpflanzung:

1.

$$z = h(x, y) = x \pm y \Rightarrow \sigma_z = \sqrt{\sigma_x^2 + \sigma_y^2} \quad (\text{siehe oben})$$

2.

$$\begin{aligned} z = h(x, y) = xy &\Rightarrow \sigma_z = \sqrt{y_0^2 \sigma_x^2 + 2x_0 y_0 \text{cov}(x, y) + x_0^2 \sigma_y^2} \\ &\Rightarrow \frac{\sigma_z}{z_0} = \sqrt{\left(\frac{\sigma_x}{x_0}\right)^2 + 2 \frac{\text{cov}(x, y)}{x_0 y_0} + \left(\frac{\sigma_y}{y_0}\right)^2} \end{aligned}$$

3.

$$\begin{aligned} z = h(x, y) = \frac{x}{y} &\Rightarrow \sigma_z = \sqrt{\frac{\sigma_x^2}{y_0^2} - 2 \frac{x_0}{y_0^3} \text{cov}(x, y) + \frac{x_0^2 \sigma_y^2}{y_0^4}} \\ &\Rightarrow \frac{\sigma_z}{z_0} = \sqrt{\left(\frac{\sigma_x}{x_0}\right)^2 - 2 \frac{\text{cov}(x, y)}{x_0 y_0} + \left(\frac{\sigma_y}{y_0}\right)^2} \end{aligned}$$

5.7 Aufgaben und Fragen zum tieferen Verständnis

1. Betrachten Sie die Normalverteilung $\mathcal{N}(x, x_0, \sigma)$. Leiten Sie die Verteilung für $z = e^x$ ab. Vergleichen Sie das Ergebnis mit der log-Normalverteilung.
2. Betrachten Sie die Normalverteilung $\mathcal{N}(x, 0, 1)$. Leiten Sie die Verteilung für $z = x^2$. Achten Sie bei dieser Transformation darauf, dass die Abbildung von z nach x nicht eindeutig ist. Vergleichen Sie das Ergebnis mit der χ^2 -Verteilung.
3. Betrachten Sie die Gammaverteilung $\Gamma(x; \alpha, \beta)$ für grosse Werte von α . Zeigen Sie, dass die Gammaverteilung in diesem Grenzfall gut durch eine Normalverteilung angenähert werden kann. Welche Konsequenzen hat das für die Anwendbarkeit der Gauss'schen Fehlerfortpflanzung?