

Lernziele Lektion 2

- Sie kennen die Normalverteilung, die Studentsche t-Verteilung, sowie die inverse Gamma-Verteilung und können mit deren Maximalwerten, Mittelwerten und Varianzen umgehen.
- Sie kennen und verstehen das additive Fehlermodell.
- Sie wissen was es bedeutet, dass der statistische Fehler von Datenpunkten statistisch unkorreliert ist und haben eine Vorstellung, worauf man bei einer Messung achten muss, um das zu erreichen.
- Sie kennen die Definitionen der statistischen Grössen Mittelwert, Varianz und linearer Korrelationskoeffizient und können diese Grössen berechnen.
- Sie können gegebenen Daten eine lineare Regressionsgerade anpassen und die Regressionsparameter, sowie deren Fehler berechnen.
- Sie kennen die Begriffe Likelihood, Priorwahrscheinlichkeit und Posteriorwahrscheinlichkeit.

2 Lineare Regression am Beispiel des Hall-Effekts

2.1 Das Physikalische Modell des Experiments

Hintergrundinformation *I*. Bei der Messung des Hall-Effekts (siehe Abb. 5) wird ein elektrischer Strom I durch einen dünnen länglichen Metallfilm getrieben. Zwei Spannungskontakte an den gegenüberliegenden Kanten des Metallfilms erlauben das Abgreifen der Hallspannung V_H , die mit einem Voltmeter gemessen wird. Zusätzlich wird ein homogenes Magnetfeld B senkrecht zur Ebene des Metallfilms angelegt. Der Zusammenhang zwischen den beteiligten Grössen ist

$$V_H = \frac{IB}{en_s},$$

wobei n_s die Flächendichte der Elektronen im Metallfilm ist. Misst man die Hallspannung bei festem Magnetfeld B für eine Reihe verschiedener Werte für den Strom I , so hat man eine lineare Beziehung ($V_H \propto I$). Auch wenn man die Hallspannung bei festem Strom I für eine Reihe von Magnetfeldwerten B misst, hat man eine lineare Beziehung ($V_H \propto B$).

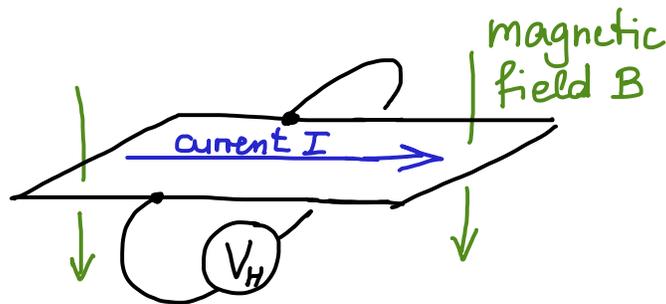


Abbildung 5: Schematischer Versuchsaufbau zur Messung des Halleffektes.

Lineare Regression wendet man im Allgemeinen an, wenn zwischen einer Stellgrösse x (in unserem Beispiel der Strom I oder das Magnetfeld B) und der Messgrösse y (in unserem Beispiel die Hallspannung V_H) ein linearer Zusammenhang der Form

$$y = ax + b$$

besteht und man aus einer Serie von N Messpunkten (x_i, y_i) mit $i = 1, \dots, N$ die Grössen a und b bestimmen will. Die Methode entspricht dem ‘fitten’ der Messpunkte mit einer Geraden.

Die gemessenen Grössen y_i werden im Allgemeinen fehlerbehaftet sein, z.B. weil der Zeiger des Voltmeters beim Ablesen geschwankt hat, oder allgemeiner, weil die Messapparatur ein gewisses statistisches Rauschen aufweist. Man berücksichtigt das Rauschen, indem man die lineare Beziehung in der Form

$$y_i = A(x_i - \bar{x}_i) + B + \epsilon_i \quad (11)$$

schreibt, wobei ϵ_i der Messfehler des Messpunktes i ist. Da der Messfehler hier als additive Grösse auftritt, spricht man von einem *additiven Fehlermodell*. Die hier eingeführte Grösse

$$\bar{x}_i := \frac{1}{N} \sum_{i=1}^N x_i \quad (12)$$

nennt man den *empirischen Mittelwert* der Stellgrössen x_i . Wir haben diese Grösse in Gl. (11) aus vorausschauender Bequemlichkeit eingeführt, um uns spätere Rechnungen zu erleichtern. Das Ziel unserer Analyse ist es, aus einem Satz von Messpunkten (x_i, y_i) die Steigung A und den y -Abschnitt B in Gl. (11) zu bestimmen.

Frage: Wie gross sind die Steigung A und der y -Abschnitt B in der Beziehung (11)?

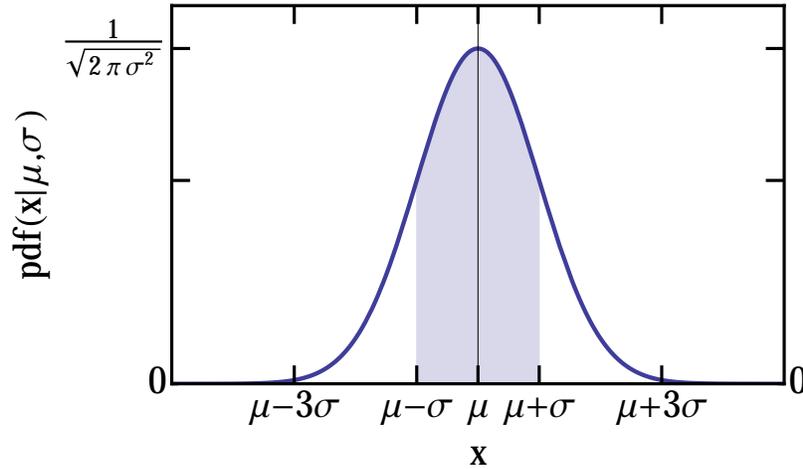


Abbildung 6: Normalverteilung $\mathcal{N}(x; \mu, \sigma)$. Mittelwert und Maximum dieser Verteilung liegen bei $x = \mu$, die Breite der Verteilung ist durch den Wert von σ bestimmt.

Modell M: Für die statistische Schwankungsgrößen ϵ_i benötigen wir an dieser Stelle noch eine statistische Beschreibung. Wie gross ist die Wahrscheinlichkeit, dass die Grösse ϵ_i beim Messen des Messpunktes (x_i, y_i) einen bestimmten Wert annimmt? Im Rahmen der Gauss'schen Fehlerrechnung nimmt man an, dass die ϵ_i der Normalverteilung

$$\text{pdf}(\epsilon_i|\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\epsilon_i^2}{2\sigma^2}\right) =: \mathcal{N}(\epsilon_i; 0, \sigma) \quad (13)$$

genügen, die in Abb. 13 gezeigt ist. Das Symbol pdf steht hier für 'Wahrscheinlichkeitsdichte' (engl.: probability density function). Dabei taucht σ^2 als neuer Parameter auf, der die Ungenauigkeit der Messwerte charakterisiert. Wir nehmen hierbei an, dass diese Ungenauigkeit für alle Messpunkte dieselbe ist, d.h. dass σ^2 unabhängig vom Index i des Messpunktes ist. Damit enthält unser Modell [Gln. (11) und (13)] die drei Parameter A , B und σ^2 , die wir hoffen aus der Messung bestimmen zu können.

Da wir insgesamt N Messpunkte aufnehmen und gleichzeitig auswerten wollen, benötigen wir die Verbundwahrscheinlichkeit, den Satz von Schwankungswerten $\{\epsilon_i\}$ zu messen. Gemäss der Produktregel der Wahrscheinlichkeitsrechnung finden wir

$$\text{pdf}(\epsilon_1, \epsilon_2, \dots, \epsilon_N|\sigma^2) = \text{pdf}(\epsilon_1|\sigma^2)\text{pdf}(\epsilon_2|\epsilon_1, \sigma^2) \dots \text{pdf}(\epsilon_N|\epsilon_1, \dots, \epsilon_{N-1}, \sigma^2).$$

An dieser Stelle machen wir eine weitere wichtige Annahme: wir gehen davon aus, dass der Messfehler der verschiedenen Messpunkte *statistisch unkorreliert* ist. Das bedeutet zum

Beispiel, dass der Wert ϵ_2 nicht von ϵ_1 abhängt, so dass $\text{pdf}(\epsilon_2|\epsilon_1, \sigma^2) = \text{pdf}(\epsilon_2|\sigma^2)$. In einem praktischen Experiment könnte man folgendermassen vorgehen, um diese Bedingung zu erfüllen: die Schwankungen der Messgrösse finden auf einer bestimmten typischen Zeitskala statt. Nehmen wir als Beispiel einen Fall, in dem der Zeiger des Messgeräts auf einer Zeitskala von etwa 1 s statistisch schwankt. Der Messfehler benachbarter Datenpunkte wird dann statistisch unabhängig sein, wenn der zeitliche Abstand Δt , in dem wir diese Datenpunkte aufnehmen, wesentlich grösser ist, als 1 s. Im Allgemeinen spricht man von der Korrelationszeit τ_c des Rauschens, und die Bedingung für statistische Unabhängigkeit lautet $\Delta t \gg \tau_c$. Unter dieser Annahme können wir die Verbundwahrscheinlichkeit vereinfachen und finden

$$\text{pdf}(\epsilon_1, \epsilon_2, \dots, \epsilon_N|\sigma^2) = \prod_{i=1}^N \text{pdf}(\epsilon_i|\sigma^2) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{1}{2} \sum_{i=1}^N \frac{\epsilon_i^2}{\sigma^2}\right). \quad (14)$$

Im letzten Schritt haben wir hier die Normalverteilung (13) eingesetzt.

Mit Hilfe von Gln. (14) und (11) können wir nun eine Verbundwahrscheinlichkeit für die gemessenen Grössen $\{y_i\}$ aufschreiben. Zu diesem Zweck lösen wir Gl. (11) nach ϵ_i auf und setzen dieses Ergebnis in Gl. (14) ein. Das ergibt

$$\text{pdf}(y_1, y_2, \dots, y_N|\{x_i\}, A, B, \sigma^2, M, I) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{1}{2} \sum_{i=1}^N \frac{(y_i - A(x_i - \bar{x}_i) - B)^2}{\sigma^2}\right). \quad (15)$$

Vergleichen wir die Vorgehensweise bei der linearen Regression mit dem Beispiel des radioaktiven Zerfalls aus der ersten Vorlesung, so sind wir jetzt an dem Punkt, an dem wir dort mit Gl. (2) waren. Wir haben eine Wahrscheinlichkeitsverteilung, die vorhersagt wie plausibel ein Messergebnis $\{y_i\}$ von N Messwerten ist, wenn die Stellgrössen $\{x_i\}$, sowie die Parameter A , B und σ^2 bekannt sind und wenn unser Modell M , sowie unsere Hintergrundinformation I korrekt sind. Betrachten wir den Ausdruck auf der rechten Seite von Gl. (15) als Funktion der Parameter A , B und σ^2 , so wird diese Funktion in der Datenanalyse als ‘Likelihood’ oder Likelihood-Funktion bezeichnet. Als Verteilung von den Messwerten $\{y_i\}$ betrachtet, sprechen wir hingegen von der *Stichprobenverteilung*.

2.2 Mittelwert, Varianz und linearer Korrelationskoeffizient

Bevor wir uns der Abschätzung der Parameter A , B und σ^2 zuwenden, wollen wir zunächst den Exponenten in Gl. (15) etwas genauer ansehen und ihn geschickt umformen. Im Rahmen unseres linearen Modells definieren wir die sogenannte χ^2 -Funktion als

$$\chi^2(A, B, \sigma^2) := \sum_{i=1}^N \frac{(y_i - A(x_i - \bar{x}_i) - B)^2}{\sigma^2}. \quad (16)$$

Unser Ziel ist es nun, die χ^2 -Funktion als quadratische Form der beiden Parameter A und B zu schreiben. Zu diesem Zweck betrachten wir

$$\begin{aligned}\sigma^2\chi^2 &= \sum_{i=1}^N (y_i - A(x_i - \bar{x}_i) - B)^2 \\ &= \sum_{i=1}^N (y_i^2 + A^2(x_i - \bar{x}_i)^2 + B^2 - 2A(x_i - \bar{x}_i)y_i - 2By_i + 2AB(x_i - \bar{x}_i)) \\ &= N \times \frac{1}{N} \sum_{i=1}^N (y_i^2 + A^2(x_i - \bar{x}_i)^2 + B^2 - 2A(x_i - \bar{x}_i)y_i - 2By_i + 2AB(x_i - \bar{x}_i))\end{aligned}$$

Die Summe über alle Messpunkte normiert auf deren Anzahl N erinnert uns an die Definition des empirischen Mittelwerts (12). In diesem Sinne schreiben wir

$$\sigma^2\chi^2 = N \times \left(\overline{y_i^2} + A^2 \overline{(x_i - \bar{x}_i)^2} + B^2 - 2A \overline{(x_i - \bar{x}_i)y_i} - 2B \overline{y_i} + 2AB \overline{(x_i - \bar{x}_i)} \right).$$

Wir sehen anhand der Definition des Mittelwerts, dass der letzte Term dieses Ausdrucks Null ist.

Varianz: Die Grösse

$$\text{Var}(x_i) := \overline{(x_i - \bar{x}_i)^2} = \overline{x_i^2} - \bar{x}_i^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x}_i)^2 \quad (17)$$

nennt man *empirische Varianz* oder auch *Schwankungsquadrat* der Werte $\{x_i\}$.

Entsprechend dieser Definition können wir schreiben $\overline{y_i^2} = \text{Var}(y_i) + \bar{y}_i^2$. Damit ergibt sich

$$\begin{aligned}\sigma^2\chi^2 &= N \times \left(\text{Var}(x_i)A^2 - 2A \overline{(x_i - \bar{x}_i)y_i} + B^2 - 2B \bar{y}_i + \bar{y}_i^2 + \text{Var}(y_i) \right) \\ &= N \times \left(\text{Var}(x_i)A^2 - 2A \overline{(x_i - \bar{x}_i)y_i} + (B - \bar{y}_i)^2 + \text{Var}(y_i) \right).\end{aligned}$$

Zudem sehen wir, dass $0 = \overline{(x_i - \bar{x}_i)\bar{y}_i}$, so dass der zweite Term in der Klammer geschrieben werden kann als

$$\overline{(x_i - \bar{x}_i)y_i} = \overline{(x_i - \bar{x}_i)(y_i - \bar{y}_i)}.$$

Linearer Korrelationskoeffizient: Wir definieren nun den *linearen Korrelationskoeffizienten* der Datenpunkte (x_i, y_i) als

$$\rho := \frac{\overline{(x_i - \bar{x}_i)(y_i - \bar{y}_i)}}{\sqrt{\text{Var}(x_i)\text{Var}(y_i)}}.$$

Damit lässt sich $\sigma^2\chi^2$ weiter umschreiben als

$$\begin{aligned} \sigma^2\chi^2 &= N \times \left(\text{Var}(x_i)A^2 - 2A\overline{(x_i - \bar{x}_i)(y_i - \bar{y}_i)} + (B - \bar{y}_i)^2 + \text{Var}(y_i) \right) \\ &= N \times \left[\text{Var}(x_i) \left(A - \sqrt{\frac{\text{Var}(y_i)}{\text{Var}(x_i)}}\rho \right)^2 + (B - \bar{y}_i)^2 + \text{Var}(y_i)(1 - \rho^2) \right]. \end{aligned}$$

Wir sehen, dass die χ^2 -Funktion im Exponenten von Gl. (15) auf eine relativ übersichtliche Weise von den Varianzen der x_i und y_i , dem Mittelwert \bar{y}_i , sowie vom Korrelationskoeffizienten ρ abhängt. Diese statistischen Grössen sind daher ausreichend, um die gesamten statistischen Eigenschaften der Datenpunkte (x_i, y_i) im Rahmen der linearen Regression zu beschreiben. Man nennt Gesamtheit dieser Grössen daher eine *hinreichende Statistik* der Datenpunkte. Wir schreiben Gl. (15) nun als

$$\begin{aligned} \text{pdf}(y_1, y_2, \dots, y_N | \{x_i\}, A, B, \sigma^2, M, I) = \\ \frac{1}{(2\pi\sigma^2)^{N/2}} \exp \left\{ -\frac{N}{2\sigma^2} \left[\text{Var}(x_i) \left(A - \sqrt{\frac{\text{Var}(y_i)}{\text{Var}(x_i)}}\rho \right)^2 + (B - \bar{y}_i)^2 + \text{Var}(y_i)(1 - \rho^2) \right] \right\} \end{aligned} \quad (18)$$

2.3 Abschätzung der Parameter A , B und σ^2

Um die gesuchten Parameter A , B und σ^2 aus gemessenen Datenpunkten (x_i, y_i) abzuschätzen, benutzen wir in kompletter Analogie zur ersten Vorlesung das Bayes'sche Theorem (7). Zunächst führen wir jedoch die neue Variable $\tau = 1/\sigma^2$ ein. Angewendet auf unsere Situation erlaubt uns das Bayes'sche Theorem, den Übergang von

$$\text{pdf}(\{y_i\} | \{x_i\}, A, B, \tau) \rightarrow \text{pdf}(A, B, \tau | \{x_i\}, \{y_i\})$$

zu machen. Wir können schreiben

$$\begin{aligned} \text{pdf}(A, B, \tau | \{x_i\}, \{y_i\}, M, I) = \\ \frac{\text{pdf}(A, B, \tau | M, I) \text{pdf}(\{y_i\} | \{x_i\}, A, B, \tau, M, I)}{\int dA dB d\tau \text{pdf}(A, B, \tau | M, I) \text{pdf}(\{y_i\} | \{x_i\}, A, B, \tau, M, I)}. \end{aligned} \quad (19)$$

Bei derartigen Anwendungen des Bayes'schen Theorems stehen auf der rechten Seite der Gleichung nur Funktionen, die bereits vor der Messung bekannt sind. Für die Funktion $\text{pdf}(\{y_i|\{x_i\}, A, B, \tau, M, I)$ hatten wir bereits den Begriff der *Likelihood* eingeführt. Diese Wahrscheinlichkeitsdichte sagt voraus, wie wahrscheinlich ein Datensatz $\{y_i\}$ ist, wenn die Parameter A, B und τ bekannt sind. Den Faktor $\text{pdf}(A, B, \tau|M, I)$ nennt man *Priorwahrscheinlichkeit*. Er gibt an, was wir vor der Messung über die Werte der Parameter A, B und τ wissen. Der Nenner des Bruches auf der rechten Seite der Gleichung stellt die Normierung der Wahrscheinlichkeitsverteilung sicher. Er ergibt sich aus dem Zähler durch Integration über die Parameter. Die Wahrscheinlichkeitsverteilung $\text{pdf}(A, B, \tau|\{x_i\}, \{y_i\}, M, I)$ auf der linken Seite der Gleichung fasst zusammen, was wir über die Werte der Parameter A, B und τ nach der Messung wissen. Daher nennt man sie *Posteriorwahrscheinlichkeit* der Parameter. In diesem Sinne beschreibt das Bayes'sche Theorem, wie wir aus den Messdaten Information über die unbekannt Parameter gewinnen. In anderen Worten: es beschreibt, wie wir aus den Daten *lernen*.

Die Priorwahrscheinlichkeit. In der letzten Vorlesung hatten wir die Priorwahrscheinlichkeit mit Hilfe des Indifferenzprinzips angegeben. Jedem möglichen Wert der Parameter wird dieselbe Wahrscheinlichkeit zugeordnet, weil wir keinen Grund haben, irgendeinem Wert eine höhere Wahrscheinlichkeit zu geben, als einem anderen (dies muss nicht immer so sein, aber wir wollen es hier annehmen). Daher verwenden wir

$$\text{pdf}(A, B, \tau|M, I) = \begin{cases} \text{const.} & \text{für } \tau > 0 \\ 0 & \text{sonst.} \end{cases}$$

Posteriorwahrscheinlichkeit. Mit dieser Priorwahrscheinlichkeit lässt sich das Bayes'sche Theorem (19) auswerten, wobei wir die Likelihood (18) einsetzen. Das Ergebnis für die Posteriorverteilung der Parameter ist

$$\begin{aligned} \text{pdf}(A, B, \tau|\{x_i\}, \{y_i\}, M, I) &= \frac{N\sqrt{\text{Var}(x_i)}}{2\pi\Gamma(N/2)} \left(\frac{N\text{Var}(y_i)(1-\rho^2)\tau}{2} \right)^{N/2} \times \\ &\exp \left\{ -\frac{N\tau}{2} \left[\text{Var}(x_j) \left(A - \sqrt{\frac{\text{Var}(y_i)}{\text{Var}(x_i)}}\rho \right)^2 + (B - \bar{y}_i)^2 + \text{Var}(y_i)(1-\rho^2) \right] \right\} \quad (20) \end{aligned}$$

Das Maximum dieser Wahrscheinlichkeitsverteilung liegt bei

$$A = \sqrt{\frac{\text{Var}(y_i)}{\text{Var}(x_i)}}\rho, \quad B = \bar{y}_i, \quad \tau = \frac{1}{\text{Var}(y_i)(1-\rho^2)},$$

diese Werte für die Parameter sind also bei gegebenen Daten die wahrscheinlichsten.

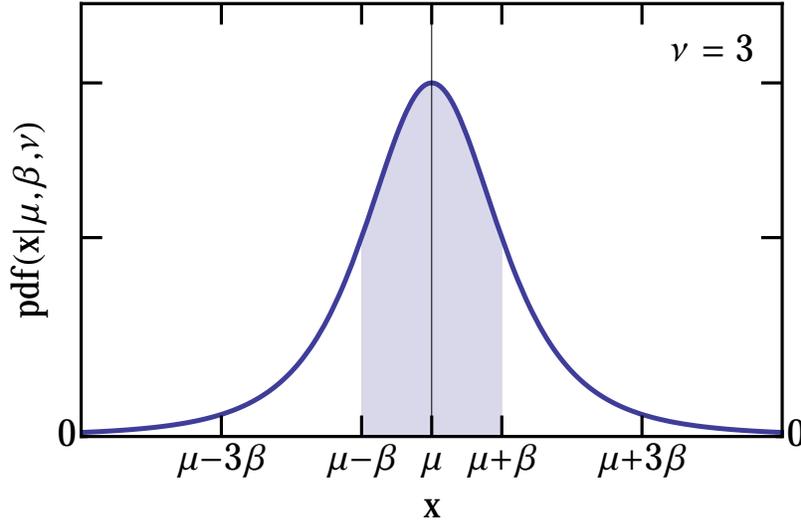


Abbildung 7: Student'sche t-Verteilung mit 3 Freiheitsgraden.

Marginale Verteilungsfunktionen. Aus der Verbundwahrscheinlichkeitsverteilung für die Parameter, $\text{pdf}(A, B, \sigma | \{y_i\}, \{x_i\}, M, I)$, lassen sich die Verteilungsfunktionen der einzelnen Parameter mit Hilfe der Marginalisierungsregel (3) bestimmen. Wir erhalten zum Beispiel

$$\begin{aligned} \text{pdf}(A | \{y_i\}, \{x_i\}, M, I) &= \int_0^\infty d\tau \int_{-\infty}^\infty dB \text{pdf}(A, B, \tau | \{y_i\}, \{x_i\}, M, I) \\ &= \mathcal{T} \left(A; \sqrt{\frac{\text{Var}(y_i)}{\text{Var}(x_i)}} \rho, \sqrt{\frac{\text{Var}(y_j)(1 - \rho^2)}{N \text{Var}(x_j)}}, N \right), \end{aligned} \quad (21)$$

wobei

$$\mathcal{T}(x; \mu, \beta, \nu) = \frac{1}{\sqrt{\nu\pi}} \frac{\Gamma[(\nu + 1)/2]}{\Gamma(\nu/2)} \left(1 + \frac{(x - \mu)^2 / \beta^2}{\nu} \right)^{-(\nu+1)/2} \quad (22)$$

die Student'sche t-Verteilung mit dem Mittelwert μ , dem Skalierungsparameter β und ν Freiheitsgraden ist (siehe Abb. 7). Diese Verteilung hat ihr Maximum bei μ , den Mittelwert μ und die Varianz $\beta^2 \nu / (\nu - 2)$.

Wir geben das Ergebnis der linearen Regressionsanalyse für den Parameter A in der Form 'Mittelwert $\pm \sqrt{\text{Varianz}}$ ' an, was

$$A = \sqrt{\frac{\text{Var}(y_i)}{\text{Var}(x_i)}} \rho \pm \sqrt{\frac{\text{Var}(y_j)(1 - \rho^2)}{(N - 2) \text{Var}(x_j)}} \quad (23)$$

für $N \geq 3$ ergibt.

Auf eine analoge Art und Weise finden wir

$$\begin{aligned} \text{pdf}(B|\{y_i\}, \{x_i\}, M, I) &= \int_0^\infty d\tau \int_{-\infty}^\infty dA \text{pdf}(A, B, \tau|\{y_i\}, \{x_i\}, M, I) \\ &= \mathcal{T} \left(B; \bar{y}_i, \sqrt{\frac{\text{Var}(y_j)(1-\rho^2)}{N}}, N \right). \end{aligned} \quad (24)$$

und geben das Ergebnis in der Form ($N \geq 3$)

$$B = \bar{y}_i \pm \sqrt{\frac{\text{Var}(y_j)(1-\rho^2)}{N-2}} \quad (25)$$

an.

Für die Verteilung von τ findet man

$$\begin{aligned} \text{pdf}(\tau|\{y_i\}, \{x_i\}, M, I) &= \int_{-\infty}^\infty dB \int_{-\infty}^\infty dA \text{pdf}(A, B, \tau|\{y_i\}, \{x_i\}, M, I) \\ &= \Gamma \left(\tau; \frac{N}{2}, \frac{2}{N\text{Var}(y_j)(1-\rho^2)} \right), \end{aligned} \quad (26)$$

wobei

$$\Gamma(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta} \left(\frac{x}{\beta} \right)^{\alpha-1} \exp \left(-\frac{x}{\beta} \right)$$

die bereits bekannte Gammaverteilung mit dem Formparameter α und dem Skalierungsparameter β ist. Die entsprechende Verteilung für $\sigma^2 = 1/\tau$ ist die inverse Gammaverteilung

$$\text{pdf}(\sigma^2|\{y_i\}, \{x_i\}, M, I) = \text{inv}\Gamma \left(\sigma^2; \frac{N}{2}, \frac{N\text{Var}(y_j)(1-\rho^2)}{2} \right), \quad (27)$$

mit

$$\text{inv}\Gamma(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta} \left(\frac{\beta}{x} \right)^{\alpha+1} \exp \left(-\frac{\beta}{x} \right)$$

(siehe Abb. 8). Diese Verteilung hat ihr Maximum bei $\beta/(\alpha+1)$, den Mittelwert $\beta/(\alpha-1)$ und die Varianz $\beta^2/[(\alpha-1)^2(\alpha-2)]$.

Daher geben wir das Ergebnis für σ^2 an als ($N \geq 5$)

$$\sigma^2 = \frac{N\text{Var}(y_j)(1-\rho^2)}{N-2} \left(1 \pm \sqrt{\frac{2}{N-4}} \right).$$

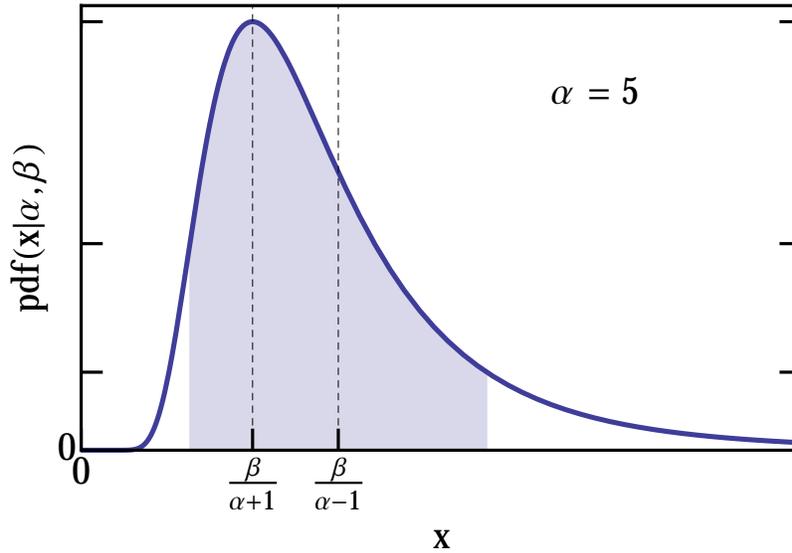


Abbildung 8: Inverse Gammaverteilung mit dem Formparameter $\alpha = 5$. Die grau schattierte Fläche unter der Verteilung kennzeichnet das Gebiet um den Mittelwert $\beta/(\alpha - 1)$, das weniger als die Wurzel aus der Varianz vom Mittelwert abweicht. Die Verteilung ist für diesen Wert von α , der $N = 10$ Datenpunkten entspricht, noch sehr asymmetrisch.

Die Gerade der Form $y = A(x - \bar{x}_i) + B$, die durch die Parameter A in Gl. (23) und B in Gl. (25) bestimmt ist nennt man *Ausgleichsgerade*. Abbildung 9 zeigt ein Beispiel für einen Datensatz, an den eine Ausgleichsgerade angepasst wurde. Die Ausgleichsgerade wird immer durch den Punkt (\bar{x}_i, \bar{y}_i) gehen. Durch Einsetzen der Parameter sehen wir, dass die Gleichung der Ausgleichsgeraden geschrieben werden kann als

$$\frac{y - \bar{y}_i}{\sqrt{\text{Var}(y_i)}} = \rho \frac{x - \bar{x}_i}{\sqrt{\text{Var}(x_i)}}.$$

Die Größen

$$\xi = \frac{x - \bar{x}_i}{\sqrt{\text{Var}(x_i)}}, \quad \text{und} \quad \eta = \frac{y - \bar{y}_i}{\sqrt{\text{Var}(y_i)}}$$

bezeichnet man als *Normalkoordinaten*. Mit ihrer Hilfe lautet die Gleichung der Ausgleichsgeraden $\eta = \rho\xi$. Wir sehen, dass der empirische Korrelationskoeffizient ρ die Steigung der Ausgleichsgeraden in Normalkoordinaten ist.

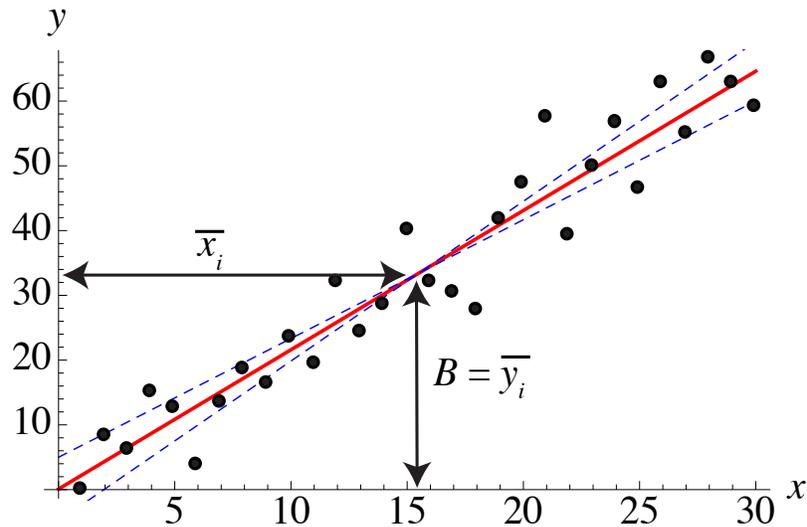


Abbildung 9: Beispiel für die graphische Darstellung einer linearen Regressionsanalyse. Schwarze Punkte repräsentieren die gemessenen Datenwerte. Die rote Gerade hat Steigung und y -Abschnitt gemäss der Abschätzung aus der linearen Regressionsanalyse. Die beiden gestrichelten Linien zeigen Geraden deren Steigungen gerade um plus oder minus die Wurzel aus einer Varianz des Steigungsparameters von der optimalen Steigung abweichen.

Aufgaben und Fragen zum tieferen Verständnis:

1. Wie viele Datenpunkte sind mindestens nötig, um aus einer linearen Regressionsanalyse die Parameter A und B mit ihren Fehlern abzuschätzen? Wie viele Datenpunkte sind hingegen mindestens nötig, um die Streuung σ^2 der Datenpunkte mit ihrem Fehler abzuschätzen?
2. Betrachten Sie eine Situation, bei der durch N -mal wiederholte Messung ein und derselben Messgrösse y (bei konstanten Parametern) die Genauigkeit für den Schätzwert dieser Messgrösse erhöht werden soll. Das Modell für diese Messung lautet $y = a$, wobei a der zu bestimmende Wert der Messgrösse ist. Benutzen Sie ein additives Fehlermodell. Nehmen Sie zudem an, dass die Messfehler normalverteilt und statistisch unabhängig sind. Darüber hinaus gehen Sie in kompletter Analogie zur linearen Regression vor:

a) Geben Sie eine geeignete Likelihood pdf($y_1, y_2, \dots, y_N | a, \sigma, M, I$) an.

- b) Definieren Sie in Analogie zur linearen Regression eine $\chi^2(a, \sigma)$ -Funktion. Wie lässt sich $\sigma^2 \chi^2$ quadratische Form von a ausdrücken? Welche statistischen Größen, die aus den Messwerten y_i bestimmt werden können, kommen dabei vor?
- c) Geben Sie die Priorwahrscheinlichkeit an, die sich aus dem Indifferenzprinzip ergibt.
- d) Wie lautet die Posteriorwahrscheinlichkeit $\text{pdf}(a, \sigma^2 | \{y_i\}, M, I)$?
- e) Geben Sie die marginalen Verteilungsfunktionen $\text{pdf}(a | \{y_i\}, M, I)$ und $\text{pdf}(\sigma^2 | \{y_i\}, M, I)$ an.
- f) Geben Sie die Schätzwerte für a und σ^2 und deren Fehler an. Wie erhöhen sich die Genauigkeiten von a und σ^2 mit zunehmender Zahl N von Messwerten?