Datenanalyse Notizen

Benjamin Dropmann

April 7, 2025

1 TLDR

Name	Bedeutung	Formel
Messfehler	Der Messfehler ist die abweichung vom gemessenen Wert x_1 zum theoretischen oder erwateten Wert$\tilde{x}$	$e = \tilde{x} - x_1$
Probability Density Function	Dies ist eine Funktion f (Probability Mass Funktion falls diskret) die die wahrscheinlichkeit $P_{a,b}(x)$ dass eine variable x zwischen zwei punkte a, b fällt	$P_{a,b}(x) = \int_a^b f(x)dx$
Normalverteilung	Die Normalverteilung ist eine Funktion f mit variablen μ :Erwartungswert und σ :Standardabweichung, die sehr oft die verbreitung zufälliger Datenaufnahmen spiegelt	$f(x,\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$
Binomialverteilung	Die Binomialverteilung beschreibt die wahrscheinlichkeit dass in n Versuchen das ereigniss mit Wahrscheinlichkeit $P(u) = p$ genau k mal vorkommt	$B(k\|p,n) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$
Mittelwertfehler	Der Fehler des Mittelwerts bei einer wiederholung einer Messwiederhulung auf N mal, wobei σ die Standardabweichung ist	$\Delta\mu = \frac{\sigma}{\sqrt{N}}$

Messfehler

Messen wir einen Wert von welchen wir schon eine Theoretische einschätzung $\tilde{x}$ haben, sei der gemessen Wert x_1 dann haben wir einen **Massfehler** $e = \tilde{x} - x_1$. Dieser Fehler kann systematisch (es kommt bei jeder Messung gleich vor) oder Zufällig (es war ein Zufall dass es diesmal den wert e angenommen hat, und es könnte nächstes mal irgeindein Wert annehmen).

Wiederholtes Messen

Mit einer Wiederholung des Messens, kann man dieses Zufällige Fehler minimiseren. Mit viele Messwerte kann man den **Empirischen Mittelwert**:

$$\overline{x} = \frac{1}{N} \sum_{n=1}^{N} x_n$$

und die **Empirische Varianz**:

$$\Delta_x^2 = \frac{1}{N-1} \sum_{n=1}^{N} (x_n - \overline{x})^2$$

Berechnen. Die **Empirische Standardabweichung** ist ähnlich definiert: $\Delta_x = \sqrt{\Delta_x^2}$
Wenn wir Viele Messungen haben, können wir in einem Histogramm darstellen in dem wir Bins (wertebereiche) defnieren und die Anzahl messung in jedes Bin darstellen.

Probability Density Function

Wir können dann dieses Histogramm eine Funktion zuordnen die ungefähr diese anpasst. So eine Funktion heisst Probability Mass Funktion und hat die Eigenschaften dass: $0 \leq P(x_n) \leq 1$ und $\sum P(x_n) = 1$ wobei x_n eine bin ist und $P(x_n)$ Die wahrscheinlichkeit dass eine Messung in diese Bin reinfällt. Wenn diese Funktion nicht mehr Diskret definiert ist sondern Kontinuirlich dann nennt man diese Probability Density Funktion und beugt sich zur Eigencshaften $\int_{-\infty}^{\infty} f(x) = 1$ und

$0 \leq f(x)$ hier ist die Wahrscheilichkeit dass $a \leq x \leq b$ vom integral $P(a \leq x \leq b) = \int_a^b f(x)dx$ Diese PDF hat mehrere Werte die in korrelation damit sind:

- Modus: $x_{\text{mode}} : \frac{df(x)}{dx}\Big|_{x_{\text{mode}}} = 0$ ist der Wert bei dem die function ihr maximum annimt
- Median $x_{\text{median}} : \int_{-\infty}^{x_{\text{median}}} f(x)dx = \frac{1}{2}$ ist der Wert bei welchem die Wahrscheinlichkeiten auf jeder seite dieses Punktes zu legen gleich sind
- Halbwertsbreite: $f(a) = f(b) = \frac{1}{2}f(x_{\text{mode}}) \Longrightarrow H_{\text{breite}} = b - a$ Ist die breite der Verteilung beim Halben wert des Maximalwertes.
- Erwartungswert: $\mu = E(x) = \int_{-\infty}^{\infty} xf(x)dx$ der wert x für welches $f(x) = x_{\text{mode}}$
- Varianz: $\text{var}(x) = E((x-\mu)^2) = \int_{-\infty}^{\infty}(x-\mu)^2 f(x)dx$ Wie zuvor. die Standardabweichung ist einfach $\sqrt{\text{var}(x)}$

Momente können auch definiert werden aber sind nicht sehr nutzlich.

Normalverteilung

Wenn wir Zuruck zum Histogramm gehen dann ist die Funktion die diese Folgt am aller meistens eine Normalverteilung. Wenn wir viele Messungen mit einen Zufälligen Messfehler Betrachten, wird dieser Histogramm eine Normalverteilung folgen. Die Definition dieser Normalverteilung ist:

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2 \cdot \pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Wobei μ der Erwatungswert ist und σ die Standardabweichung, diese Werte müssen Experimentell gefunden werden.

Binomialverteilung

Wenn wir zwei ereignisse u und d mit wahrscheilickeiten $P(u) = p$ und $P(d) = 1 - p = q$ respektiv haben. Dann kann man, erstens die anzahl der unterschiedlichen Reihenfolgen von k objekten in n plätze definieren durch n "*tief*" k

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

z.B. 5 Studente in 7 arbeitsplätze haben $\binom{7}{5} = 21$ unterschiedliche sitzplane. Dann kommt Binomialverteilung die die Wahrscheinlichkeit beschreibt dass in n versuchen k mal den ereigniss u mit wahrscheinlichkeit $P(u) = p$ vorkommt:

$$B(k|p, n) = \binom{n}{k} p^k (1-p)^{n-k}$$

Hier beschreibt der teil $p^k(1-p)^{n-k}$ die wahrscheinlichkeit von genau eine distribution der Ereignisse. Die Poissonverteilung ist die Binomiallverteilung für $\lim_{k \to k}$

Mittelwertfehler

Wenn wir eine Messung vielmals wiederholen, dann wird die standardabweichung ab einen Punkt nicht mehr viel variieren, doch mit mehr messpunkte, sind wir uns sicherer und sicherer dass unser Mittelwert korrekt ist. Deswegen ist der Fehler des Mittelwerts nicht einfach die Standardabweichung sondern durch:

$$\Delta\mu = \frac{\sigma}{\sqrt{N}}$$

Fehlerfortplanzung

Wenn wir experimentell was bestätigen wollen misst man meistens nicht sofort die wichtige Grösse sonderne eine Andere mit deren man die wichtige grössen rechnen kann. Doch die unsicherheiten können sehr schnell schwierige beiträge zur rechnung bringen. Deswegen gibt es eine einfachere Methode die uns hilft diese Fehlerfortpflanzung abzuschätzen im fall von einer Normalverteilten PDF wo die funktion f nur von einen fehlerbehafteten Messgrösse x abhängt. Dann kann man die Taylorentwicklung des ersten Grades benutzen um folgendes Zusammenhang zu finden

$$\Delta f(x) = f(\overline{x} - \Delta x) - f(\overline{x}) = \left.\frac{\partial f}{\partial x}\right|_{\overline{x}} \Delta x$$

Wobei Δ einen Fehler bezeichnet, $\overline{x}$ der punkt um deren wir den Fehler berechnen. Die standardabweichung lässt sich auch derselben weise rechnen:

$$\sigma_f = \left.\frac{\partial f}{\partial x}\right|_{\overline{x}} \sigma_x$$

Mehrdimensional Fehlerfortplanzung

Wir können es erweitern bis zum fall wo die funktion f von mehrere Fehler behafteten Messgrössen. In diesem Fall hängt der Fehler Δf auch von der Kovarianz zwischen x und y die zwei Messgrössen.

$$\sigma_f^2 = (\partial_x, \partial_y) f \begin{pmatrix} \sigma_x^2 & \sigma_{xy}^2 \\ \sigma_{xy}^2 & \sigma_y \end{pmatrix} \begin{pmatrix} \partial_x \\ \partial_y \end{pmatrix} f$$

Wenn die kovarianz $\sigma_{xy} = 0$ dann ist dies einfach eine Diagonale Matrix und es wird zu summe:

$$\sigma_f^2 = \sum_{n=1}^{N} \left(\frac{\partial f}{\partial x_n} \right) \sigma_{x_n}^2$$

Fouriertranform

Wir können jede Funktion auf einem beschränkte Zetiperiode als eine Summe von Trigonometrischen funktionen defnieren. Wenn ich eine funktion x_{t_k} habe dann kann ich eine frequenz hier einsetzen: (in theorie ist dies ein Integral aber hier handelt es sich um diskrete daten)

$$X(f_n) = \frac{1}{N} \sum_{k=0} N - 1 x(t_k) e^{-i2\pi f_n t_k}$$

und ich bekomme ein anteil der trigonometrischen funktionen von dieser fräquenz in der funktion.
Wie man das praktisch macht, man misst für alle frequenzen die Gleichung und dann bekommt man den anteil dieser fräquenz auf der funktion welche wir den Fouriertransform machen

Aliasing

Man muss aufpassen wenn man eine Welle misst, wenn $\Delta t > \frac{1}{\nu}$ dann ist unsere messung okay, man braucht also mindestens zwei messpunkte pro oszillation der Welle, sonst kann ein Fall von Aliasing passieren wo wir eine höhere fräquenz messen als was eigentilch ist.

Gabors limit

Wenn sag ich dass eine fräquenz nicht im Transform auftaucht? Gabors limit besagt dass je langer ein signal ist in der zeit, desto besser kann die präzision der messung vom signal sein: $\sigma_t \cdot \sigma_f \geq \frac{1}{2}$. Wobei σ die auflösung der Messung ist. Daher limitiert die Messzeit t_{tot} σ_t. Wenn man den Fouriertransform von einer funktion mit eine auflösung σ_t, mit eine kleinere auflösung misst, bekommt man nicht mehr information, die zwei werte sind dann auf distanzen kleine als σ_t korreliert und bringen nicht neue information. Zusätslich ist die Maximale frequenz die ich messen kann ist $f_{max} = \frac{1}{2\Delta t}$ die minimale frequenz lässt sich analog finden aber die negative Fourierwerte haben nicht mehr information als die positive