Gaußsche Zufallsvariable

Autor und Herausgeber

Alexander Jung*, Department of Computer Science, Aalto University, Espoo, Finland (ORCID 0000-0001-7538-0990)
Editors: Konstantina Olioumtsevits (ORCID 0000-0003-2697-8659), Ekkehard Schnoor (ORCID 0000-0001-8252-1499), Aalto University, Espoo, Finland
*Corresponding author: alex.jung@aalto.fi

Zusammenfassung

Eine Gaußsche Zufallsvariable (Gaußsche ZV) ist eine reellwertige Zufallsvariable (ZV), deren Wahrscheinlichkeitsdichte proportional zum Exponential einer negativen quadratischen Funktion ist. Die Wahrscheinlichkeitsdichte wird durch zwei Zahlen bestimmt, den Mittelwert μ und die Varianz σ^2 der ZV; die Wahrscheinlichkeitsverteilung mit diesen beiden Parametern wird mit $\mathcal{N}\left(\mu, \sigma^2\right)$ bezeichnet. Die standardmäßige Gaußsche ZV ist der Spezialfall $\mu = 0$ und $\sigma^2 = 1$; eine allgemeine Gaußsche ZV wird daraus durch Skalierung und Verschiebung gewonnen. Ein Gaußscher Zufallsvektor ist eine affine Transformation unabhängiger standardmäßiger Gaußscher Zufallsvariablen. Gaußsche Zufallsvariablen entstehen als Grenzwerte von Mittelwerten vieler unabhängiger Zufallsvariablen durch den central limit theorem (CLT). Unter allen Zufallsvektoren mit einer gegebenen Kovarianzmatrix maximiert der Gaußsche Zufallsvektor die differential entropy.

Definition

Der Messfehler eines Sensors, wie z.B. eines Thermometers, wird üblicherweise als Gaußsche Zufallsvariable (ZV) modelliert. Eine reellwertige ZV z ist eine standardmäßige Gaußsche ZV, wenn ihre Wahrscheinlichkeitsdichte (Bertsekas and Tsitsiklis, 2008, Sect. 3.3; Gray, 2009; Papoulis and Pillai, 2002)

$$p^{(z)}\left(\eta\right) = \frac{1}{\sqrt{2\pi}} \exp\left(-\eta^2/2\right) \text{ ist.}$$

Gegeben eine standardmäßige Gaußsche ZV z, kann eine allgemeine Gaußsche ZV x mit Mittelwert μ und Varianz σ^2 über $x := \sigma z + \mu$ konstruiert werden, sodass die Wahrscheinlichkeitsverteilung von x die Normalverteilung $\mathcal{N}\left(\mu, \sigma^2\right)$ ist, bezeichnet als $x \sim \mathcal{N}\left(\mu, \sigma^2\right)$ (Bertsekas and Tsitsiklis, 2008, Sect. 3.3). Die Wahrscheinlichkeitsdichte von x ist für jedes $\eta \in \mathbb{R}$

$$p^{(x)}\left(\eta\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(\eta-\mu)^2}{2\sigma^2}\Big).$$

Die standardmäßige Gaußsche ZV ist der Spezialfall $\mu = 0$ und $\sigma^2 = 1$.

Die beiden Parameter von $\mathcal{N}\left(\mu, \sigma^2\right)$ wirken getrennt auf die Wahrscheinlichkeitsdichte, wie in Abb. 1 gezeigt: Der Parameter μ ist der Mittelwert von x, und

der Parameter σ ist ihre Standardabweichung, die Quadratwurzel der Varianz σ^2. Das Hinzufügen von μ zur skalierten ZV σz verschiebt die Wahrscheinlichkeitsdichte entlang der horizontalen Achse um μ, ohne ihre Form zu ändern. Für $\sigma > 0$ streckt die Multiplikation von z mit σ die Wahrscheinlichkeitsdichte horizontal um den Faktor σ und senkt ihren Scheitelpunkt um denselben Faktor, sodass die Fläche darunter 1 bleibt.

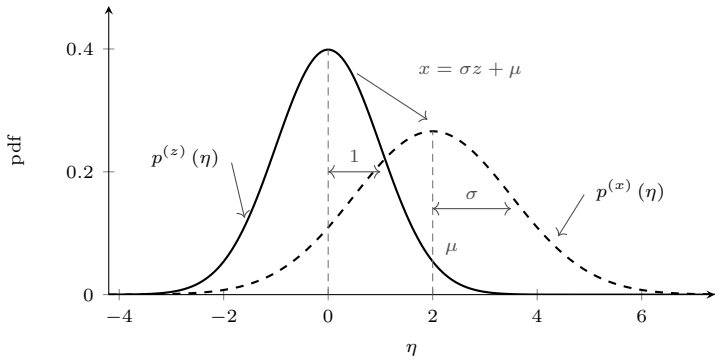


Abbildung 1: Die durchgezogene Kurve ist die Wahrscheinlichkeitsdichte einer standardmäßigen Gaußschen ZV z. Die gestrichelte Kurve ist die Wahrscheinlichkeitsdichte der Gaußschen ZV $x = \sigma z + \mu$ mit $\mu = 2$ und $\sigma = 1.5$. Der Verschiebungsparameter μ verschiebt den Mittelwert von 0 nach μ, während der Skalierungsparameter σ die Standardabweichung von 1 nach σ und die Varianz von 1 nach σ^2 ändert. Ein vertikales Lineal markiert den Mittelwert jeder Kurve, bei $\eta = 0$ und bei $\eta = \mu$, und jeder horizontale Doppelpfeil verläuft vom Mittelwert seiner eigenen Kurve bis zu einer Standardabweichung Entfernung, 1 für die durchgezogene Kurve und σ für die gestrichelte. Der graue Pfeil von Spitze zu Spitze markiert die Verschiebung und Skalierung, die die eine Wahrscheinlichkeitsdichte in die andere überführen.

Zwei gängige Methoden zur Erzeugung einer Gaußschen ZV aus anderen Zufallsquellen sind in Gebrauch. Für die exakte Erzeugung sei U eine ZV, die gleichmäßig auf dem offenen Intervall $(0, 1) \subset \mathbb{R}$ verteilt ist, und Φ sei die Verteilungsfunktion einer standardmäßigen Gaußschen ZV. Dann ist $\mu + \sigma\Phi^{-1}(U) \sim \mathcal{N}\left(\mu, \sigma^2\right)$, dieselbe Wahrscheinlichkeitsverteilung wie x, aber aus einer anderen Zufallsquelle gewonnen.

Für die approximative Erzeugung sei m eine positive ganze Zahl und $v^{(1)}, \ldots, v^{(m)}$ seien unabhängig und identisch verteilt (u.i.v.) Zufallsvariablen, nicht notwendigerweise Gaußsch, mit Mittelwert 0 und endlicher, nicht-null Varianz τ^2. Dann konvergiert die normierte Summe

$$\frac{1}{\tau\sqrt{m}} \sum_{i=1}^{m} v^{(i)}$$

in Verteilung gegen eine standardmäßige Gaußsche ZV, wenn m zunimmt (Bertsekas and Tsitsiklis, 2008, Sect. 5.4; Ross, 2014, Thm. 3.1, p. 391). Somit macht der central limit theorem (CLT) eine Gaußsche Approximation für Mittelwerte vieler unabhängiger Beiträge angemessen, weshalb der Messfehler eines Sensors üblicherweise als Gaußsch modelliert wird, obwohl eine endliche Summe nicht Gaußsch sein muss (Ross, 2014). Abb. 2 zeigt beide Hälften dieser Aussage für Münzwürfe, wobei jeder die Werte $+1$ und -1 mit Wahrscheinlichkeit $1/2$ annimmt. Die probability mass function (pmf) der normierten Summe von $m = 8$ von ihnen folgt bereits der Form der standardmäßigen Gaußschen Wahrscheinlichkeitsdichte, und die Summe nimmt immer noch nur 9 Werte an. Die convergence in distribution ist die convergence der Verteilungsfunktionen, sodass der größte Abstand zwischen der Verteilungsfunktion der normierten Summe und der einer standardmäßigen Gaußschen ZV misst, wie weit die Approximation noch ist. Er beträgt 0.137 für $m = 8$ und 0.050 für $m = 64$.

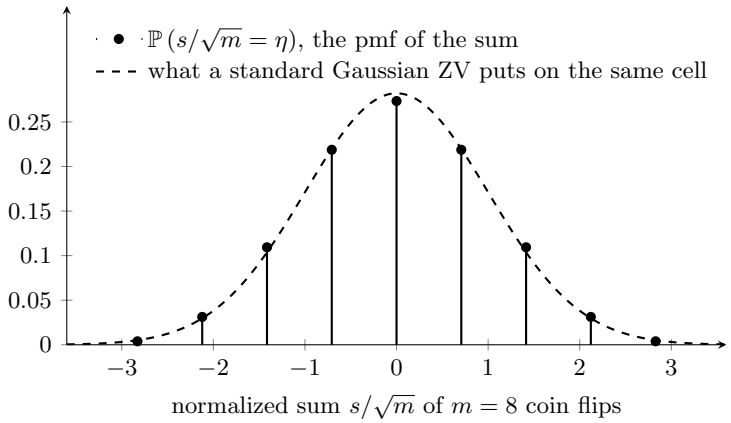


Abbildung 2: Hier ist s die Summe von $m = 8$ u.i.v. Zufallsvariablen, wobei jede die Werte $+1$ und -1 mit Wahrscheinlichkeit $1/2$ annimmt (Münzwürfe). Die Stiele tragen die pmf der normierten Summe $s/\sqrt{m}$, die 9 Werte im Abstand von $2/\sqrt{m}$ annimmt, sodass die normierte Summe nicht Gaußsch ist. Die gestrichelte Kurve ist die Wahrscheinlichkeit, die eine standardmäßige Gaußsche ZV auf die Zelle dieser Breite um jeden Wert legt, ihre Wahrscheinlichkeitsdichte dort multipliziert mit $2/\sqrt{m}$. Daten generiert von `pythondemos/gaussrv.py`

Ein Gaußscher Zufallsvektor $\mathbf{x} \in \mathbb{R}^d$ mit Kovarianzmatrix $\mathbf{C}$ und Mittelwert $\boldsymbol{\mu}$ kann konstruiert werden als (Gray, 2009; Lapidoth, 2017; Papoulis and Pillai, 2002)

$$\mathbf{x} := \mathbf{A}\mathbf{z} + \boldsymbol{\mu}.$$

Hier ist $\mathbf{z} := \big(z_1, \ldots, z_d\big)^\top$ ein Vektor von u.i.v. standardmäßigen Gaußschen Zufallsvariablen. Die Matrix $\mathbf{A} \in \mathbb{R}^{d\times d}$ ist eine Matrix Quadratwurzel von $\mathbf{C}$, d.h. jede Matrix mit $\mathbf{A}\mathbf{A}^\top = \mathbf{C}$ (zum Beispiel ein Cholesky-Faktor). Die Wahr-

scheinlichkeitsverteilung eines Gaußschen Zufallsvektor wird als multivariate Normalverteilung bezeichnet, notiert als $\mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$. Ein Gaußscher Zufallsvektor muss keine unabhängigen Komponenten haben; seine Komponenten sind genau dann unabhängig, wenn seine Kovarianzmatrix diagonal ist.

Ein Gaußscher Zufallsvektor $\mathbf{x} = \big(x_1, \ldots, x_d\big)^\top$ kann als ein stochastic process interpretiert werden, das durch die endliche Menge $\{1, \ldots, d\}$ indiziert ist. Allgemeiner ist ein Gaussian process (GP) ein stochastic process über einer Indexmenge $\mathcal{I}$, dessen Einschränkung auf jede endliche Teilmenge von $\mathcal{I}$ ein Gaußscher Zufallsvektor ist (Rasmussen and Williams, 2006).

Eine standardmäßige Gaußsche ZV z erfüllt die Gaußsche Schwanzschranke $\mathbb{P}(|z| \geq t) \leq 2\exp(-t^2/2)$ für jede positive Zahl t (Wainwright, 2019, Beispiel 2.1, Gl. (2.9)). Diese Schranke kontrolliert auch Funktionen von u.i.v. Zufallsvariablen. Betrachten Sie einen standardmäßigen Gaußschen Zufallsvektor $\mathbf{z} \in \mathbb{R}^d$ und eine Funktion $f : \mathbb{R}^d \to \mathbb{R}$, die Lipschitz mit der Konstanten L ist, d.h. $\big|f(\mathbf{z}) - f(\mathbf{z}')\big| \leq L\|\mathbf{z} - \mathbf{z}'\|$ für jedes Paar von Vektoren $\mathbf{z}, \mathbf{z}' \in \mathbb{R}^d$. Die Abweichung von $f(\mathbf{z})$ von ihrem Erwartungswert ist dann eine sub-Gaussian random variable (sub-Gaussian RV) mit Parameter von höchstens L. Für jede positive Zahl t,

$$\mathbb{P}\left(\big|f(\mathbf{z}) - \mathbb{E}\big\{f(\mathbf{z})\big\}\big| \geq t\right) \leq 2\exp\left(\frac{-t^2}{2L^2}\right).$$

Die Anzahl d der Argumente erscheint nicht in dieser concentration inequality (Wainwright, 2019, Thm. 2.26): Die Schranke ist die gleiche für zwei Argumente und für tausend. Sie beschränkt die Schwänze von $f(\mathbf{z})$ und besagt nicht, dass $f(\mathbf{z})$ eine Normalverteilung hat. Die Euclidean norm ist ein Beispiel mit $L = 1$: Die Länge $\|\mathbf{z}\|$ hat einen Erwartungswert von höchstens $\sqrt{d}$ und weicht von diesem um mehr als t mit einer Wahrscheinlichkeit von höchstens $2\exp(-t^2/2)$ ab, egal wie groß d ist.

Unter allen Zufallsvektoren mit einer gegebenen Kovarianzmatrix $\mathbf{C}$ maximiert der Gaußsche Zufallsvektor $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$ die differential entropy (Cover and Thomas, 2006, Thm. 8.6.5). Diese Eigenschaft ist ein Grund, warum additives Gaußsches Messrauschen ein weit verbreitetes Wahrscheinlichkeitsmodell in der statistischen Analyse von machine learning (ML)-Methoden ist: Wenn nur die Varianz des Rauschens bekannt ist, ist eine Gaußsche Verteilung die am wenigsten restriktive Annahme.

Siehe auch: multivariate Normalverteilung, GP, Wahrscheinlichkeitsmodell, CLT, differential entropy, concentration inequality, sub-Gaussian RV.

Querverweise

Literatur

1. Bertsekas and Tsitsiklis (2008). Introduction to Probability. 2nd ed., Athena Scientific, Belmont, MA, USA.

2. Gray (2009). Probability, Random Processes, and Ergodic Properties. 2nd ed., Springer Science+Business Media, New York, NY, USA.

3. Lapidoth (2017). A Foundation in Digital Communication. 2nd ed., Cambridge Univ. Press, Cambridge, U.K.

4. Rasmussen and Williams (2006). Gaussian Processes for Machine Learning. MIT Press, Cambridge, MA, USA.

5. Wainwright (2019). High-Dimensional Statistics: A Non-Asymptotic Viewpoint. Cambridge Univ. Press, Cambridge, U.K.

6. Cover and Thomas (2006). Elements of Information Theory. 2nd ed., Wiley, Hoboken, NJ, USA.

7. Papoulis and Pillai (2002). Probability, Random Variables, and Stochastic Processes. 4th ed., McGraw-Hill Higher Education, New York, NY, USA.

8. Ross (2014). A First Course in Probability. 9th ed., Pearson Education, Boston, MA, USA.