Wahrscheinlichkeitsrechnung und Statistik

6. Vorlesung - 2017

Verteilungsfunktion (VF)

Definition 11

Die **Verteilungsfunktion** (VF) einer Zufallsgröße X ist $F : \mathbb{R} \to \mathbb{R}$ definiert als

$$F(x) := P(\{\omega \in \Omega \ : \ X(\omega) \leq x\}) = P(\ X \leq x\ ) \text{ für jedes } x \in \mathbb{R}.$$

Verteilungsfunktion (VF)

Definition 11

Die **Verteilungsfunktion** (VF) einer Zufallsgröße X ist $F : \mathbb{R} \to \mathbb{R}$ definiert als

$$F(x) := P(\{\omega \in \Omega \ : \ X(\omega) \leq x\}) = P(\ X \leq x \ ) \text{ für jedes } x \in \mathbb{R}.$$

Satz 10 - Eigenschaften der Verteilungsfunktion F der ZG X

1. F ist monoton wachsend, d.h. aus $x_1 < x_2$ folgt $F(x_1) \leq F(x_2)$.
2. $\lim\limits_{x \to \infty} F(x) = 1 \quad$ und $\quad \lim\limits_{x \to -\infty} F(x) = 0$
3. F ist rechtsseitig stetig, d.h. $\lim\limits_{x \searrow x_0} F(x) = F(x_0)$ für $\forall \ x_0 \in \mathbb{R}$.
4. $P(a < X \leq b) = F(b) - F(a)$

Jede Funktion $F : \mathbb{R} \to \mathbb{R}$, welche die Eigenschaften 1., 2. und 3. des obigen Satzes erfüllt, ist eine Verteilungsfunktion.

Für die Verteilungsfunktion einer diskreten ZG X mit den Werten $x_i, i \in I$ gilt

$$F(x) = \sum_{i \in I : x_i \leq x} P(X = x_i) \; \forall \; x \in \mathbb{R}.$$

Dichtefunktion (DF) einer stetigen ZG

Zur Erinnerung: Eine Zufallsgröße heißt stetig, wenn sie überabzählbar viele Werte in einem Intervall oder $\mathbb{R}$ annehmen kann!

Definition 12

Die Funktion $f : \mathbb{R} \to \mathbb{R}$ heißt **Dichtefunktion** (DF) der **stetigen Zufallsgröße** X, falls für alle $a, b \in \mathbb{R}$ mit $a < b$ gilt

$$P(X \in (a, b]) = \int_a^b f(x)\, dx.$$

Beispiel: Dichtefunktion der Standardnormalverteilung (Gaußverteilung)

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \; x \in \mathbb{R}$$

Satz 11

Sei f Dichtefunktion einer stetigen Zufallsgröße X. Dann gilt

1. $f(x) \geq 0 \quad$ für alle $x \in \mathbb{R}$.
2. $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = 1.$
3. $F(x) = P(X \leq x) = \displaystyle\int_{-\infty}^{x} f(t)\,dt \;\forall x \in \mathbb{R}.$
4. F ist eine stetige Funktion.
5. Wenn F an der Stelle x differenzierbar ist $\Rightarrow \; F'(x) = f(x)$.

Satz 11

Sei f Dichtefunktion einer stetigen Zufallsgröße X. Dann gilt

1. $f(x) \geq 0 \quad$ für alle $x \in \mathbb{R}$.
2. $$\int\limits_{-\infty}^{\infty} f(x)\,dx = 1.$$
3. $$F(x) = P(X \leq x) = \int\limits_{-\infty}^{x} f(t)\,dt \ \forall x \in \mathbb{R}.$$
4. F ist eine stetige Funktion.
5. Wenn F an der Stelle x differenzierbar ist $\Rightarrow \ F'(x) = f(x)$.

Jede Funktion $f : \mathbb{R} \to \mathbb{R}$, welche die Eigenschaften 1. und 2. des obigen Satzes erfüllt, ist eine Dichtefunktion.

Diskrete und stetige ZG

- eine diskrete ZG. X wird vollständig durch ihre Wahrscheinlichkeitsverteilung beschrieben

$$X \sim \begin{pmatrix} x_1 & x_2 & \dots & x_i & \dots \\ p_1 & p_2 & \dots & p_i & \dots \end{pmatrix} = \begin{pmatrix} x_i \\ p_i \end{pmatrix}_{i \in I}$$

$$P(X \in (a,b]) = \sum_{\substack{i \in I \\ a < x_i \leq b}} P(X = x_i)$$

- eine stetige ZG. X wird vollständig durch ihre Dichtefunktion f beschrieben: für alle $a, b \in \mathbb{R}$ mit $a < b$ gilt

$$P(X \in (a,b]) = \int_a^b f(x)\, dx$$

Definition 13

Der **Erwartungswert** oder **Mittelwert** einer **stetigen** Zufallsgröße X ist gegeben durch

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx,$$

für eine **stetige** Zufallsgröße X mit der Dichtefunktion f, vorausgesetzt, dass $\int_{-\infty}^{\infty} |x| f(x) dx < \infty$ gilt.

Satz 12 - Eigenschaften des Erwartungswertes

Es seien X und Y stetige Zufallsgrößen. Dann gilt

1. $E(X+Y) = E(X) + E(Y)$ (Additivität)
2. $E(aX) = aE(X)$ für $a \in \mathbb{R}$ (Homogenität)
3. $|E(X)| \leq E(|X|)$
4. $X \leq Y \Rightarrow E(X) \leq E(Y)$ (Monotonie)
5. Der Erwartungswert der stetigen ZG $H(X)$ (mit $H : \mathbb{R} \to \mathbb{R}$) ist $E(H(X)) = \int\limits_{-\infty}^{\infty} H(x)f(x)dx$, für eine **stetige** ZG X mit der Dichtefunktion f, wenn $\int_{-\infty}^{\infty} |H(x)|f(x)dx < \infty$ gilt.

$$H(z) = (z - EX)^2 \longrightarrow E[H(X)] = E[(X - EX)^2]$$
$$H(z) = z^n \longrightarrow E[H(X)] = E[X^n]$$

Definition 14

Sei X diskrete oder stetige Zufallsgröße.

$Var(X) = E[(X - EX)^2]$	**Varianz** bzw. **Streuung**,
$St(X) = \sqrt{Var(X)}$	**Standardabweichung**,
$E(X^n)$	n-**tes Moment**,
$E[(X - EX)^n]$	n-**tes zentrales Moment**,
	der Zufallsgröße X, $n = 1, 2, \ldots$

Gleichmäßige Verteilung $Unif[a,b]$

Bezeichnung: $X \sim Unif[a,b]$

Parameter: $a, b \in \mathbb{R}$ mit $a < b$

Eine stetige Zufallsgröße X heißt auf dem Intervall $[a,b]$ **gleichverteilt**, falls ihre Dichtefunktion gegeben ist durch

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{für } x \in [a,b] \\ 0 & \text{sonst} \end{cases}$$

- in Octave/Matlab: *unifrnd, unifpdf, unifcdf*
- $X \sim \mathcal{U}[0,1] \Rightarrow \; Y = a + (b-a)X \sim \mathcal{U}[a,b]$.

Beispiel:
Die Verkehrsnachrichten melden einen Unfall auf der Autobahn zwischen KM 60 und 100. Da keine weiteren Informationen verfügbar sind, betrachten wir die genaue Position des Unfalls als gleichverteilt auf dem genannten Streckenabschnitt.
Wie groß ist die Wahrscheinlichkeit dafür, dass sich der Unfall zwischen KM 80 und 85 ereignet hat?

Exponentialverteilung

Bezeichnung	$X \sim Exp(\mu)$
Parameter	$\mu > 0$

Eine stetige Zufallsgröße X heißt **exponentialverteilt** mit dem (Intensitäts-) Parameter $\mu > 0$, falls ihre Dichtefunktion gegeben ist durch

$$f(x) = \begin{cases} \mu e^{-\mu x} & \text{für } x \geq 0 \\ 0 & \text{für } x < 0 \end{cases}$$

$X \sim Exp(\mu)$, man berechne ihre Verteilungfunktion!

$$F(z) = \int_{-\infty}^{z} f(x)dx = \begin{cases} 0 & \text{für } z < 0 \\ 1 - e^{-\mu z} & \text{für } z \geq 0, \end{cases}$$

Bespiel:

In einer Kfz-Werkstatt sei die Reparaturzeit eines Autos exponentialverteilt mit dem Parameter $\mu = 1/2\ [h^{-1}]$.

Wie groß ist die Wahrscheinlichkeit dafür, dass die Reparatur eines zufällig ausgewählten Autos länger als 3 Stunden dauert?

Übung: Nichtalterungseigenschaft / Gedächtnislosigkeit

Sei $X \sim Exp(\mu)$, dann gilt für $s, t > 0$

$$P(X > t + s \, | X > t) = P(X > s).$$

$P(X > t)$: Wahrscheinlichkeit für Lebensdauer größer t

$P(X > t + s | X > t)$: Wahrscheinlichkeit dafür, weitere s Zeiteinheiten zu überleben, wenn bereits *das Alter t überlebt* wurde.

Die letztgenannte Wahrscheinlichkeit ist nach der obigen Übung *unabhängig vom Alter t*.

Gaußsche Normalverteilung

Bezeichnung	$X \sim \mathcal{N}(\mu, \sigma^2)$		
Parameter	$\mu \in \mathbb{R}$	Erwartungswert	(Lageparameter)
	$\sigma > 0$	Standardabweichung	(Formparamter)
	σ^2	Varianz	

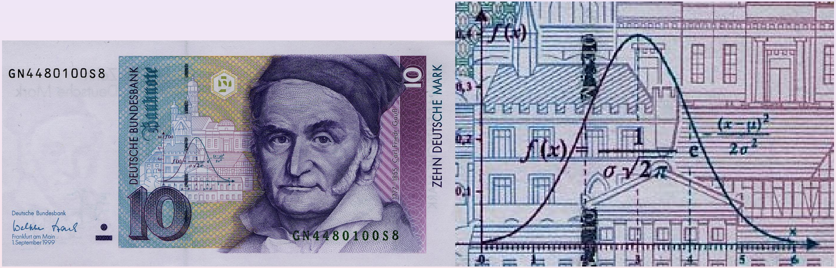
Eine stetige Zufallsgröße X heißt **normalverteilt** bzw. **Gauß-verteilt** mit den Parametern $\mu \in \mathbb{R}$ und $\sigma > 0$, falls ihre Dichtefunktion gegeben ist durch

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right), \quad x \in \mathbb{R}.$$

Für $\mu = 0$ und $\sigma = 1$, d.h. $X \sim \mathcal{N}(0,1)$, spricht man von der **Standardnormalverteilung**.

Gaußsche Normalverteilung

Wegen der Gestalt ihres Graphen wird die Dichtefunktion f auch *Gaußsche Glockenkurve* genannt. Sie findet sich auf dem letzten Zehnmarkschein vor der Euro-Einführung neben dem Bildnis von Carl Friedrich Gauß (1777-1855).



Anwendungen

- zufällige Schwankungen von Messfehlern (Fehlerrechnung)
- Brownsche Molekularbewegung
- Mathematische Statistik

Die **Verteilungsfunktion** $F(z) = \int_{-\infty}^{z} f(x)\,dx$ ist bekanntlich eine Stammfunktion zur Dichtefunktion f. Obwohl diese existiert, kann sie jedoch nicht mit Hilfe anderer elementarer Funktionen ausgedrückt werden. Daher verwendet man Näherungen, welche für die *Standardnormalverteilung* $\mathcal{N}(0,1)$ tabelliert bzw. in Softwares implementiert sind (in Matlab: *normcdf*).

Satz 13

1. Es seien $a, b \in \mathbb{R}$, $X \sim \mathcal{N}(\mu, \sigma^2)$

$$\Rightarrow \quad Y = aX + b \sim \mathcal{N}(a\mu + b,\ a^2\sigma^2).$$

2.

$$X \sim \mathcal{N}(\mu, \sigma^2) \Longleftrightarrow \frac{X - \mu}{\sigma} \sim \mathcal{N}(0,1)$$

Beispiel: Der Kraftstoffverbrauch eines PKW auf 100 km kann durch eine normalverteilte Zufallsgröße X mit den Parametern $\mu = 7$ (Liter) und $\sigma = 0.2$ (Liter) beschrieben werden, d.h. $X \sim \mathcal{N}(7, 0.04)$. Wie groß ist die Wahrscheinlichkeit dafür, dass der Verbrauch

1. zwischen 6.5 und 7.3 Liter liegt?
2. größer als 7.3 Liter ist?

Gamma-Verteilung

Bezeichnung	$X \sim Gamma(\alpha, \beta)$	
Parameter	$\alpha > 0$	(Formparameter)
	$\beta > 0$	(Skalenparameter)

Eine stetige Zufallsgröße X mit Werten in $(0, \infty)$ heißt **Gamma-verteilt** mit den Parametern $\alpha, \beta > 0$, falls ihre Dichtefunktion gegeben ist durch

$$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}} & \text{für } x > 0 \\ 0 & \text{für } x \leq 0 \end{cases}$$

1. Die Gamma-Funktion ist für $a > 0$ definiert durch

$$\Gamma(a) := \int_0^\infty t^{a-1} e^{-t}\, dt$$

und kann als Erweiterung der Fakultät auf nichtganzzahlige Argumente betrachtet werden. Es gilt

$$\Gamma(n+1) = n!, \quad n \in \mathbb{N}.$$

2. Mit der Gamma-Verteilung werden z.B. modelliert: Schadenhöhen von Versicherungsschäden, Lebensdauern.

3. **Spezialfälle**

$Gamma(1, \frac{1}{\beta})$: Exponentialverteilung $Exp(\beta)$

$Gamma(\frac{n}{2}, 2)$: χ^2-Verteilung mit Parameter $n \in \mathbb{N}$

χ^2-Verteilung

Eine stetige Zufallsgröße X mit Werten in $(0, \infty)$ heißt χ^2-**verteilt** mit den Parameter $n \in \mathbb{N}$, falls ihre Dichtefunktion gegeben ist durch

$$f(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & \text{für } x > 0 \\ 0 & \text{für } x \leq 0 \end{cases}$$

Anwendungen: Mathematische Statistik

- Anpassungstest (man prüft, ob die Daten der Stichprobe eine bestimmte Verteilung haben)
- Unabhängigkeitstest (man prüft, ob zwei Merkmale unabhängig sind)
- Homogenitätstest (man prüft, ob zwei oder mehr Stichproben dieselbe Verteilung haben)