Poisson-Prozess

Wir hatten bei der Diskussion der geometrischen und der Poisson-Verteilung festgestellt:

Wenn der zeitliche Abstand der Treffer geometrisch verteilt ist, so ist ihre Anzahl in einer festen Zeitspanne binomialverteilt.

Im Grenzwert $n \to \infty$, wobei wir die Trefferwahrscheinlichkeit mit $p_n = \lambda/n$ ansetzen, konvergiert die Binomialverteilung gegen die Poisson-Verteilung und die geometrische Verteilung gegen die Exponentialverteilung. Im Grenzwert $n \to \infty$ erwarten wir deshalb die folgende Aussage:

Wenn man Ereignisse zählt, deren zeitlicher Abstand exponentialverteilt ist, so ist die Anzahl dieser Ereignisse in einer festen Zeitspanne Poisson-verteilt.

Seien $T_1, T_2 \ldots$ unabhängige exponentialverteilte Zufallsvariablen mit Parameter λ. Die Zufallsvariable T_i modelliert die Zeit, die zwischen Treffer $i-1$ und i vergeht.

Für den Zeitpunkt $t > 0$ definieren wir

$$X(t) := \max\{n \in \mathbb{N} \mid T_1 + \ldots + T_n \leq t\}.$$

$X(t)$ gibt also an, wie viele Treffer sich bis zur Zeit t (von Zeit Null ab) ereignet haben. Es gilt:

Fakt 111

Seien $T_1, T_2, \ldots$ unabhängige Zufallsvariablen und sei $X(t)$ für $t > 0$ wie oben definiert. Dann gilt: $X(t)$ ist genau dann Poisson-verteilt mit Parameter $t\lambda$, wenn es sich bei $T_1, T_2, \ldots$ um exponentialverteilte Zufallsvariablen mit Parameter λ handelt.

Zum Zufallsexperiment, das durch $T_1, T_2, \ldots$ definiert ist, erhalten wir für jeden Wert $t > 0$ eine Zufallsvariable $X(t)$. Hierbei können wir t als Zeit interpretieren und $X(t)$ als Verhalten des Experiments zur Zeit t. Eine solche Familie $(X(t))_{t>0}$ von Zufallsvariablen nennt man allgemein einen stochastischen Prozess. Der hier betrachtete Prozess, bei dem $T_1, T_2, \ldots$ unabhängige, exponentialverteilte Zufallsvariablen sind, heißt Poisson-Prozess und stellt ein fundamentales und zugleich praktisch sehr bedeutsames Beispiel für einen stochastischen Prozess dar.

Beispiel 112

Wir betrachten eine Menge von Jobs, die auf einem Prozessor sequentiell abgearbeitet werden. Die Laufzeiten der Jobs seien unabhängig und exponentialverteilt mit Parameter $\lambda = 1/30[1/s]$. Jeder Job benötigt also im Mittel $30s$.

Gemäß Fakt 111 ist die Anzahl von Jobs, die in einer Minute vollständig ausgeführt werden, Poisson-verteilt mit Parameter $t\lambda = 60 \cdot (1/30) = 2$.

Die Wahrscheinlichkeit, dass in einer Minute höchstens ein Job abgearbeitet wird, beträgt in diesem Fall ($t\lambda = 2$)

$$e^{-t\lambda} + t\lambda e^{-t\lambda} \approx 0{,}406 \,.$$

3.4 Summen von Zufallsvariablen

Satz 113

Seien X und Y unabhängige kontinuierliche Zufallsvariablen. Für die Dichte von $Z := X + Y$ gilt

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(z - x) \,\mathrm{d}\, x\,.$$

Beweis:

Nach Definition der Verteilungsfunktion gilt

$$F_Z(t) = \Pr[Z \leq t] = \Pr[X + Y \leq t] = \int_{A(t)} f_{X,Y}(x, y) \,\mathrm{d}\, x \mathrm{d}\, y$$

wobei $A(t) = \{(x, y) \in \mathbb{R}^2 \mid x + y \leq t\}$.

Beweis (Forts.):

Aus der Unabhängigkeit von X und Y folgt

$$F_Z(t) = \int_{A(t)} f_X(x) \cdot f_Y(y) \,\mathrm{d}\,x \mathrm{d}\,y$$
$$= \int_{-\infty}^{\infty} f_X(x) \cdot \left(\int_{-\infty}^{t-x} f_Y(y) \,\mathrm{d}\,y \right) \mathrm{d}\,x.$$

Mittels der Substitution $z := x + y$, d z = d y ergibt sich

$$\int_{-\infty}^{t-x} f_Y(y) \,\mathrm{d}\,y = \int_{-\infty}^{t} f_Y(z - x) \,\mathrm{d}\,z$$

und somit

$$F_Z(t) = \int_{-\infty}^{t} \left(\int_{-\infty}^{\infty} f_X(x) f_Y(z - x) \,\mathrm{d}\,x \right) \mathrm{d}\,z\,.$$

□

Satz 114 (Additivität der Normalverteilung)

Die Zufallsvariablen $X_1, \ldots, X_n$ *seien unabhängig und normalverteilt mit den Parametern* μ_i, σ_i *(*$1 \leq i \leq n$*). Es gilt: Die Zufallsvariable*

$$Z := a_1 X_1 + \ldots + a_n X_n$$

ist normalverteilt mit Erwartungswert $\mu = a_1\mu_1 + \ldots + a_n\mu_n$ *und Varianz* $\sigma^2 = a_1^2\sigma_1^2 + \ldots + a_n^2\sigma_n^2$.

Beweis:

Wir beweisen zunächst den Fall $n = 2$ und $a_1 = a_2 = 1$. Nach Satz 113 gilt für $Z := X_1 + X_2$, dass

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_{X_1}(z-y) \cdot f_{X_2}(y) \,\mathrm{d}\, y \\ &= \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2} \underbrace{\left(\frac{(z-y-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2} \right)}_{=:v} \right) \mathrm{d}\, y. \end{aligned}$$

Beweis (Forts.):

Wir setzen

$$\begin{aligned} \mu &:= \mu_1 + \mu_2 \\ \sigma^2 &:= \sigma_1^2 + \sigma_2^2 \\ v_1 &:= (z - \mu)/\sigma \\ v_2^2 &:= v - v_1^2 \end{aligned}$$

Damit ergibt sich unmittelbar

$$v_2^2 = \frac{(z - y - \mu_1)^2}{\sigma_1^2} + \frac{(y - \mu_2)^2}{\sigma_2^2} - \frac{(z - \mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2}\,,$$

woraus wir

$$v_2 = \frac{y\sigma_1^2 - \mu_2\sigma_1^2 + y\sigma_2^2 - z\sigma_2^2 + \mu_1\sigma_2^2}{\sigma_1\sigma_2\sigma}$$

ermitteln.

Beweis (Forts.):

Damit folgt für die gesuchte Dichte

$$f_Z(z) = \frac{1}{2\pi \cdot \sigma_1 \cdot \sigma_2} \cdot \exp\left(-\frac{v_1^2}{2}\right) \cdot \int_{-\infty}^{\infty} \exp\left(-\frac{v_2^2}{2}\right) \mathrm{d}\, y.$$

Wir substituieren noch

$$t := v_2 \text{ und } \mathrm{d}\, t = \frac{\sigma}{\sigma_1 \sigma_2} \mathrm{d}\, y$$

und erhalten

$$f_Z(z) = \frac{1}{2\pi \cdot \sigma} \cdot \exp\left(-\frac{(z-\mu)^2}{2\sigma^2}\right) \cdot \int_{-\infty}^{\infty} \exp\left(-\frac{t^2}{2}\right) \mathrm{d}\, t.$$

Mit Lemma 99 folgt, dass $f_Z(z) = \varphi(z; \mu, \sigma)$ ist.

Beweis (Forts.):

Daraus erhalten wir die Behauptung für $n = 2$, denn den Fall $Z := a_1X_1 + a_2X_2$ für beliebige Werte $a_1, a_2 \in \mathbb{R}$ können wir leicht mit Hilfe von Satz 100 auf den soeben bewiesenen Fall reduzieren. Durch Induktion kann die Aussage auf beliebige Werte $n \in \mathbb{N}$ verallgemeinert werden. $\square$

3.5 Momenterzeugende Funktionen für kontinuierliche Zufallsvariablen

Für diskrete Zufallsvariablen X haben wir die momenterzeugende Funktion

$$M_X(s) = \mathbb{E}[e^{Xs}]$$

eingeführt. Diese Definition kann man unmittelbar auf kontinuierliche Zufallsvariablen übertragen. Die für $M_X(s)$ gezeigten Eigenschaften bleiben dabei erhalten.

Beispiel 115

Für eine auf $[a, b]$ gleichverteilte Zufallsvariable U gilt

$$M_U(t) = \mathbb{E}[e^{tX}] = \int_a^b e^{tx} \cdot \frac{1}{b-a} \,\mathrm{d}\, x$$

$$= \left[\frac{e^{tx}}{t(b-a)} \right]_a^b$$

$$= \frac{e^{tb} - e^{ta}}{t(b-a)} \,.$$

Beispiel (Forts.)

Für eine standardnormalverteilte Zufallsvariable $N \sim \mathcal{N}(0,1)$ gilt

$$
\begin{aligned}
M_N(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{t\xi} e^{-\xi^2/2} \,\mathrm{d}\,\xi \\
&= e^{t^2/2} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-(t-\xi)^2/2} \,\mathrm{d}\,\xi \\
&= e^{t^2/2}\,.
\end{aligned}
$$

Beispiel (Forts.)

Daraus ergibt sich für $Y \sim \mathcal{N}(\mu, \sigma^2)$ *wegen* $\frac{Y-\mu}{\sigma} \sim \mathcal{N}(0,1)$

$$\begin{aligned} M_Y(t) &= \mathbb{E}[e^{tY}] \\ &= e^{t\mu} \cdot \mathbb{E}[e^{(t\sigma)\cdot\frac{Y-\mu}{\sigma}}] \\ &= e^{t\mu} \cdot M_N(t\sigma) \\ &= e^{t\mu+(t\sigma)^2/2}\,. \end{aligned}$$

Weiterer Beweis von Satz 114:

Beweis:

Gemäß dem vorhergehenden Beispiel gilt

$$M_{X_i}(t) = e^{t\mu_i + (t\sigma_i)^2/2}\,.$$

Wegen der Unabhängigkeit der X_i folgt

$$\begin{aligned} M_Z(t) = \mathbb{E}[e^{t(a_1X_1+\cdots+a_nX_n)}] &= \prod_{i=1}^n \mathbb{E}[e^{(a_it)X_i}] \\ &= \prod_{i=1}^n M_{X_i}(a_it) \\ &= \prod_{i=1}^n e^{a_it\mu_i+(a_it\sigma_i)^2/2} \\ &= e^{t\mu+(t\sigma)^2/2}\,, \end{aligned}$$

mit $\mu = a_1\mu_1 + \cdots + a_n\mu_n$ und $\sigma^2 = a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2$. $\square$

4. Zentraler Grenzwertsatz

Satz 116 (Zentraler Grenzwertsatz)

Die Zufallsvariablen $X_1, \ldots, X_n$ besitzen jeweils dieselbe Verteilung und seien unabhängig. Erwartungswert und Varianz von X_i existieren für $i = 1, \ldots, n$ und seien mit μ bzw. σ^2 bezeichnet ($\sigma^2 > 0$).
Die Zufallsvariablen Y_n seien definiert durch $Y_n := X_1 + \ldots + X_n$ für $n \geq 1$. Dann folgt, dass die Zufallsvariablen

$$Z_n := \frac{Y_n - n\mu}{\sigma\sqrt{n}}$$

asymptotisch standardnormalverteilt sind, also $Z_n \sim \mathcal{N}(0, 1)$ für $n \to \infty$.

Etwas formaler ausgedrückt gilt: Die Folge der zu Z_n gehörenden Verteilungsfunktionen F_n hat die Eigenschaft

$$\lim_{n\to\infty} F_n(x) = \Phi(x) \text{ für alle } x \in \mathbb{R}.$$

Wir sagen dazu auch: Die Verteilung von Z_n konvergiert gegen die Standardnormalverteilung für $n \to \infty$.

Dieser Satz ist von großer Bedeutung für die Anwendung der Normalverteilung in der Statistik. Der Satz besagt, dass sich die Verteilung einer Summe beliebiger unabhängiger Zufallsvariablen (mit endlichem Erwartungswert und Varianz) der Normalverteilung umso mehr annähert, je mehr Zufallsvariablen an der Summe beteiligt sind.