20251211_Statistik_WIN2

Aufgabe 1: Gleichverteilung – Startverzögerung eines Batch-Jobs

In einem Rechenzentrum startet ein nächtlicher ETL-Batch-Job wegen Scheduler-Jitter nicht exakt um 02:00, sondern mit einer zufälligen Verzögerung X (in Minuten).
Man modelliert:

$$X \sim U(0, 12)$$

also gleichverteilt zwischen 0 und 12 Minuten.

Zusätzlich verursacht jede Verzögerungsminute im Reportingprozess Opportunitätskosten von **40 € pro Minute**.
Definiere die Kosten $C = 40X$.

1. Berechne $P(X \leq 5)$.
2. Berechne $E[X]$, $\mathrm{Var}(X)$ und die Standardabweichung s_X.
3. Bestimme ein **95%-Quantil** $q_{0,95}$, so dass $P(X \leq q_{0,95}) = 0{,}95$. (Interpretation: so viel Puffer, dass der Job in 95% der Nächte gestartet ist.)
4. Berechne den Erwartungswert der Kosten $E[C]$.
5. Berechne $P(C > 300)$.

Aufgabe 2: Normalverteilung – API-Antwortzeit und SLA

Die Antwortzeit X (in ms) eines zentralen API-Endpoints wird während Peak-Last näherungsweise normalverteilt modelliert:

$$X \sim \mathcal{N}(220,\ 40^2)$$

Also: $\mu = 220$ ms, $\sigma = 40$ ms.

1. Berechne $P(X > 300)$ (SLA-Verstoß bei 300 ms).
2. Berechne $P(200 \leq X \leq 260)$.
3. Bestimme den Wert t, so dass 95% der Antwortzeiten **unter** t liegen (95%-Quantil).
4. Ein SLA fordert: „Mindestens 95% der Requests müssen unter 280 ms sein." Prüfe, ob das Modell dies erfüllt.

Hinweis: Standardnormalwerte (typisch aus Tabelle/Excel):

$$\Phi(2{,}00) = 0{,}9772,\ \Phi(1{,}00) = 0{,}8413,\ \Phi(0{,}50) = 0{,}6915,\ \Phi(1{,}50) = 0{,}9332,\ z_{0,95} = 1{,}645$$

und $\Phi(-z) = 1 - \Phi(z)$.

Aufgabe 3: Binomialverteilung – Fehler in Datenimport (ETL/ERP)

Beim Import von Datensätzen ins ERP sind einzelne Datensätze mit Wahrscheinlichkeit $p = 0{,}08$ fehlerhaft (unabhängig voneinander).
Es werden $n = 12$ Datensätze stichprobenartig geprüft.

Sei X = Anzahl fehlerhafter Datensätze in der Stichprobe.

$$X \sim \mathrm{Bin}(n = 12,\ p = 0{,}08)$$

1. Berechne $P(X = 0)$.
2. Berechne $P(X \leq 1)$ (Qualitätsregel: „maximal 1 Fehler erlaubt").
3. Berechne $P(X \geq 3)$ (Alarmregel: „mindestens 3 Fehler").
4. Berechne $E[X]$ und $\mathrm{Var}(X)$ und interpretiere sie.

Aufgabe 4: Poisson-Verteilung – Anzahl kritischer Incidents pro Woche

Ein IT-Service-Team beobachtet kritische Incidents (Severity 1). Im Mittel treten

$$\lambda = 2{,}5$$

Incidents pro Woche auf. Modell:

$$X \sim \mathrm{Pois}(2{,}5)$$

1. Berechne $P(X = 0)$.
2. Berechne $P(X \leq 2)$.
3. Berechne $P(X \geq 5)$.
4. Bestimme $E[X]$ und $\mathrm{Var}(X)$.
5. Ein Incident verursacht im Mittel 800 € Aufwand. Bestimme den erwarteten Wochenaufwand.

Aufgabe 5: Exponentialverteilung – Zeit bis zum nächsten Incident

Ein Monitoring-System registriert kritische Incidents. Im Mittel treten **2 Incidents pro Stunde** auf.
Die Wartezeit T bis zum nächsten Incident wird modelliert als exponentialverteilt mit Rate

$$\lambda = 2 \text{ pro Stunde}$$

also:

$$T \sim \mathrm{Exp}(2)$$

1. Berechne $E[T]$ (in Minuten).
2. Berechne $P(T \leq 15 \text{ Minuten})$.
3. Berechne $P(T > 60 \text{ Minuten})$.
4. Gedächtnislosigkeit: Berechne
$$P(T > 50 \text{ Minuten} \mid T > 20 \text{ Minuten})$$
5. Bestimme die **Median-Wartezeit** $t_{0,5}$, so dass $P(T \leq t_{0,5}) = 0{,}5$.

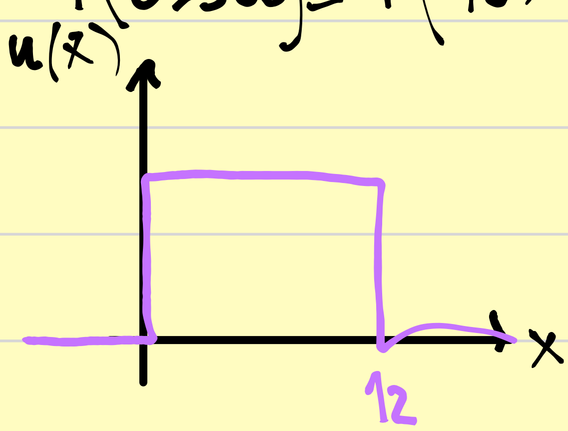
1. $P(X \leq x) = \frac{x-a}{b-a} = \frac{x-0}{12-0} = \frac{x}{12}$ $\quad P(X \leq 5) = \frac{5}{12} = 0'42$

$E[X] = \frac{a+b}{2} = \frac{12}{2} = 6$ $\quad VAR[X] = \frac{(b-a)^2}{12} = \frac{144}{12} = 12$ $\quad \sigma(X) = \sqrt{12} = 3'46$

$P(X \leq q_{0'95}) = 0'95 = \frac{x}{12} \rightarrow x_{0'95} = 12 \cdot 0'95 = 11'4$ Minuten

$E(C) = E[40 \cdot X] = 40 \cdot E[X] = 40 \cdot 6 = 240$ €

$P(C > 300) = P(40x > 300) = P\left(X > \frac{300}{40}\right) = P(X > 7'5) = \frac{12-7'5}{12} = 0'37$



2. $Z = \frac{x-\mu}{\sigma} = \frac{x-220}{40} \sim N(0,1)$

$P(X > 300) = P\left(Z > \frac{300-220}{40}\right) = P(Z > 2) = 1 - P(Z < 2) = 0'0228$

$P(200 < X < 260) = P\left(\frac{200-220}{40} < Z < \frac{260-220}{40}\right) = P(-0'5 < Z < 1) =$

$= P(Z < 1) - P(Z < -0'5) = 0'5328$

$$P(Z < z_{0'95}) = 0'95 \xrightarrow{\text{Tabelle}} z_{0'95} = 1'645 = \frac{x-220}{40} \rightarrow x_{0'95} = 286 \text{ ms}$$

FORDERUNG

$$P(X \leq 280) \geq 0'95 \longrightarrow P\left(Z \leq \frac{280-220}{40}\right) = 0'933 < 0'95$$

Das Modell erfüllt die Anforderung NICHT.

3. $$P(X=k) = \binom{12}{k} (0'08)^k (1-0'08)^{12-k}$$

$$P(X=0) = \underbrace{\binom{12}{0}}_{1} \underbrace{(0'08)^0}_{1} (0'92)^{12} = 0'3677$$

$$P(X \leq 1) = P(X=0) + P(X=1) = 0'3677 + \binom{12}{1}(0'08)^1(0'92)^{11} = 0'75$$

$$P(X \geq 3) = 1 - P(X<3) = 1 - \left[P(X=0) + P(X=1) + P(X=2)\right] =$$
$$= 1 - \left[0'75 + \binom{12}{2}(0'08)^2(0'92)^{10}\right] = 0'065$$

$$E(X) = n \cdot p = 12 \cdot 0'08 = 0'96$$
$$VAR(X) = n \cdot p \cdot (1-p) = 12 \cdot 0'08 \cdot 0'92 = 0'883$$

4. $$P(X=k) = e^{-2'5} \cdot \frac{2'5^k}{k!}$$

$$P(X=0) = e^{-2'5} \cdot \frac{2'5^0}{0!} = 0'082$$

$$P(X \leq 2) = P(X=0) + P(X=1) + P(X=2) = 0'082 + e^{-2'5} \cdot \frac{2'5^1}{1!} + e^{-2'5} \cdot \frac{2'5^2}{2!}$$
$$= 0'544$$

$$P(X>5) = 1 - P(X\leq 4) = 1 - \left[0'544 + P(X=3) + P(X=4)\right] =$$

$$= 1 - \left[0'544 + e^{-2'5}\cdot\frac{2'5^3}{3!} + e^{-2'5}\cdot\frac{2'5^4}{4!}\right] = 0'11$$

$$E(X) = VAR(X) = \lambda = 2'5$$

$$E(\text{KOSTEN}) = E(800X) = 800\cdot E(X) = 800\cdot\lambda = 2000€$$

5. $\lambda = 2$ pro Stunde $\longrightarrow \begin{cases} 15\text{ Min} \longrightarrow \lambda^*_{15} = 0'5 \\ 20\text{ Min} \longrightarrow \lambda^*_{20} = \frac{2}{3} \\ 50\text{ Min} \longrightarrow \lambda^*_{50} = \frac{5}{6}2 = \frac{5}{3} \end{cases}$

$$E[T] = \frac{1}{\lambda} = 0'5\text{ Std} = 30\text{ Minuten}$$

$$P(T\leq t) = 1 - e^{-\lambda t} = 1 - e^{-2t}$$

$$P(T\leq 0'25) = 1 - e^{-2\cdot 0'25} = 0'39$$

$$P(T>t) = 1 - \left[1 - e^{-\lambda t}\right] = e^{-\lambda t}$$

$$P(T>60) = e^{-2\cdot 1} = e^{-2} = 0'135$$

$$\boxed{\text{Gedächtnislosigkeit: } P(T>s+t \mid T>s) = P(T>t)}$$

$$P(T>50 \mid T>20) = P(\underset{\text{Min}}{t = 50-20 = 30}) = P(T>0'5h) =$$

$$= e^{-2\cdot 0'5} = e^{-1} = 0'368$$

Interpretation: auch wenn schon 20 Min nichts passiert ist, bleibt die Restwartezeitverteilung gleich.

$$P(T < t_{0'5}) = 0'5 = 1 - e^{-2 \cdot t_{0'5}} \rightarrow e^{-2 \cdot t_{0'5}} = 0'5 \rightarrow$$

$$\rightarrow -2 \cdot t_{0'5} = \ln 0'5 \rightarrow t_{0'5} = \frac{\ln 0'5}{-2} = 0'3466\,h \simeq 20'8 \text{ Minuten}$$