9 Verschiedenes

9.1 Wiederkehrperiode

Seien $X_1, X_2, \dots$ unabhängige und identisch verteilte Zufallsvariablen und sei $u \in \mathbb{R}$ ein Schwellenwert. Es sei $p := \mathbb{P}[X_i > u] \neq 0$. Wir definieren

$$T_1 := \min\{i \in \mathbb{N} : X_i > u\} \text{ und } T_{n+1} := \min\{i > T_n : X_i > u\}.$$

Man kann sich $T_1, T_2, \dots$ qals Zeitpunkte denken, zu denen ein bestimmtes Ereignis (eine Überflutung) eintritt. Wir definieren außerdem $W_n = T_n - T_{n-1}$, wobei $T_0 = 0$ gesetzt wird.

Theorem 9.1. *Für alle $n \in \mathbb{N}$ gilt $W_n \sim$ Geo(p), d.h. $\mathbb{P}[W_n = k] = p(1-p)^{k-1}$ für alle $k = 1, 2, \dots$.*

Bemerkung 9.2. Es folgt $\mathbb{E}W_n = \frac{1}{p}$. Die Zahl $\frac{1}{p}$ heißt die Wiederkehrperiode. Es ist der erwartete Abstand zwischen zwei aufeinanderfolgenden Überflutungen.

Beweis von Satz 9.1. Wir beweisen den Satz nur für $n = 1$: $\mathbb{P}[W_1 = k] = \mathbb{P}[X_1 \leq u, \dots, X_{k-1} \leq u, X_k > u] = (1-p)^{k-1}p$. □

Beispiel 9.3. Per Definition ist die Wahrscheinlichkeit, dass ein 100-Jahresereignis in einem Jahr eintritt, gleich $p = 1/100$. Die Wiederkehrperiode ist somit 100 Jahre. Wie groß ist die Wahrscheinlichkeit, dass ein 100-Jahresereignis in den nächsten 100 Jahren nicht eintritt? Diese Wahrscheinlichkeit ist

$$\left(1 - \frac{1}{100}\right)^{100} \approx \frac{1}{e} \approx 0,3679.$$

Außerdem können wir direkt angeben, wieviele 100-Jahresereignisse etwa in 100 Jahren statt finden. Diese Zahl ist nämlich $\mathrm{Bin}(100, \frac{1}{100})$, also approximativ $\mathrm{Poi}(1)$-verteilt.

9.2 Gumbel-Überschreitungsmethode

Seien $X_1, X_2, \dots$ u.i.v Zufallsvariablen mit Verteilungsfunktion F. Die absteigend angeordneten Ordnungsstatistiken von $X_1, \dots, X_n$ seien mit $M_n^{(1)} \geq M_n^{(2)} \geq \dots \geq M_n^{(n)}$

bezeichnet. Wir definieren

$$S_{k,r,n} = S = \sum_{i=1}^{r} 1_{X_{n+i} > M_n^{(k)}}.$$

Somit ist $S_{k,r,n}$ die Anzahl der Überschreitungen des Schwellenwerts $M_n^{(k)}$ in Zeitintervall $n+1, \dots, n+r$.

Theorem 9.4. *Die Verteilung von $S_{k,r,n}$ ist gegeben durch*

$$\mathbb{P}[S = j] = \frac{\binom{r+n-k-j}{n-k}\binom{j+k-1}{k-1}}{\binom{r+n}{n}}, \qquad j = 0, \dots, r.$$

Beweis. Unter Anwendung der totalen Wahrscheinlichkeit gilt:

$$\mathbb{P}[S = j] = \int_{\mathbb{R}} \mathbb{P}\left[S = j \middle| M_n^{(k)} = u\right] dF_{n,k}(u).$$

Dabei ist $F_{n,k}$ die Verteilungsfunktion von $M_n^{(k)}$. Es ist zu beachten, dass

$$S \big| M_n^{(k)} = u \sim Bin(r, \bar{F}(u))$$

beziehungsweise

$$dF_{n,k}(u) = \frac{n!}{(k-1)!(n-k)!} F^{n-k}(u) \bar{F}^{k-1}(u) dF(u).$$

Es folgt:

$$\mathbb{P}[S = j] = \int_{\mathbb{R}} \binom{r}{j} \bar{F}^j(u) F^{r-j}(u) \frac{n!}{(k-1)!(n-k)!} F^{n-k}(u) \bar{F}^{k-1}(u) dF(u),$$

was sich unter der Substitution $F(u) = t$ wie folgt schreiben lässt:

$$\mathbb{P}[S = j] = \int_0^1 \binom{r}{j} \frac{n!}{(k-1)!(n-k)!} t^{r-j+n-k} (1-t)^{j+k-1} dt,$$

was sich schließlich unter Beachtung der bekannten Formel für die Beta-Funktion

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1} dt = \frac{(\alpha-1)!(\beta-1)!}{(\alpha+\beta-1)!}$$

zu Folgendem umschreiben lässt:

$$\mathbb{P}[S = j] = \binom{r}{j} \frac{n!}{(k-1)!(n-k)!} \frac{(k+j-1)!(r-j+n-k)!}{(r+n)!}.$$

$\square$

Beispiel 9.5. Die maximalen jährlichen Wasserstände an einem Fluss seien seit $n = 100$ Jahren bekannt und mit $X_1, \ldots, X_n$ bezeichnet. Wie hoch muss ein Deich sein, bei dem die Wahrscheinlichkeit einer Überflutung in den nächsten $r = 12$ Jahren $P = 0,3$ sein soll? Wir setzen die gesuchte Deichhöhe als $M_n^{(k)}$ mit einem unbekannten k an. Im obigen Satz wurde gezeigt, dass

$$\mathbb{P}[S_{k,r,n} = 0] = \frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+1)}{(r+n) \cdot (r+n-1) \cdot \ldots \cdot (r+n-k+1)} \stackrel{!}{=} 0,7.$$

Als Lösung der obigen Gleichung mit $n = 100$, $r = 12$ ergibt sich $k = 3$. Die Deichhöhe kann als die dritte Ordnungsstatistik $M_n^{(3)}$ gewählt werden. Der Vorteil dieser Methode besteht darin, dass die (unbekannte) Verteilungsfunktion von X_i keine Rolle spielt (und nicht geschätzt werden muss). Der Nachteil ist, dass die Methode bei einer kleineren Überflutungswahrscheinlichkeit P nicht funktioniert.