Aufgabe 12 (Randomisierter LQ-Test)

Für eine Erhebung zur Untersuchung einer Vogelkrankheit beschäftigen Sie 20 Mitarbeiter, die jeweils so viele Tiere untersuchen, bis sie zum ersten Mal eine Erkrankung festgestellt haben. Die i.i.d. Zufallsvariablen $X_1, \dots, X_{20}$ geben jeweils die Anzahl der untersuchten Tiere bis einschließlich zum ersten kranken Tier an. Ein Forscher behauptet, dass 13% der Vögel erkrankt sind. Sie vermuten, dass er das Vorkommen der Krankheit überschätzt.

Insgesamt haben Ihre Mitarbeiter 640 Vögel untersucht. Können Sie die Behauptung des Forschers zum 5%-Niveau zugunsten Ihrer Alternative, dass 10% der Vögel erkrankt sind, widerlegen? Formulieren Sie das Testproblem und konstruieren Sie den Test zum exakten Niveau α, der auf der Alternative die größte Güte hat.

Hinweis: Die Summe von n unabhängigen geometrisch verteilten Zufallsvariablen mit Parameter π ist negativ binomialverteilt mit Parametern n und π.

Aufgabe 13 (Multiples Testen)

Sei T eine Teststatistik mit Realisation t und bekannter stetiger Verteilungsfunktion F_{H_0} unter der Nullhypothese H_0. Zeigen Sie zunächst, dass der p-Wert $p = \mathbb{P}_{H_0}(T \geq t)$ auf dem Intervall $[0, 1]$ gleichverteilt ist. Benutzen Sie dazu die Eigenschaft,

(1) dass für eine stetige, monoton wachsende Transformation g die Gleichheit

$$\mathbb{P}(X < x) = \mathbb{P}(g(X) < g(x))$$

gilt;

(2) dass eine stetige Zufallsvariable U genau dann auf dem Bereich $\mathbb{D} = [0, 1]$ gleichverteilt ist, wenn für alle $u \in \mathbb{D}$ des Definitionsbereich $\mathbb{P}(U \leq u) = u$ gilt.

Simulieren Sie nun $k = 1010$ voneinander unabhängige, standardnormalverteilte Zufallsvektoren $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k$ der Länge $n = 50$ in R. Modifizieren Sie die Vektoren $\boldsymbol{x}_1, \dots, \boldsymbol{x}_{10}$ so, dass deren Verteilung jeweils nicht mehr um Null zentriert ist (bspw. durch Addieren einer Konstanten).

(a) Testen Sie, ob der Mittelwert von $\boldsymbol{x}_{11}, \dots, \boldsymbol{x}_k$ jeweils signifikant von Null verschieden ist ($\alpha = 0.05$) und visualisieren Sie die sich ergebende Verteilung der p-Werte. Welche Verteilung sollte sich für $k \to \infty$ ergeben? Testen Sie nun auch, ob sich der Mittelwert von $\boldsymbol{x}_1, \dots, \boldsymbol{x}_{10}$ jeweils signifikant von Null unterscheidet ($\alpha = 0.05$).

(b) Wiederholen Sie den datengenerierenden Prozess sowie die Tests aus Teilaufgabe (a) 100-mal und schätzen Sie anschließend die *False Discovery Rate* (FDR), also den erwarteten Anteil an Tests mit signifikantem Ergebnis, die fälschlicherweise die Nullhypothese abgelehnt haben.

(c) Welche FDR ergibt sich nach Adjustierung der p-Werte mit der Bonferroni-Prozedur?

*Aufgabe 9 (Klassische Testtheorie)

Seien $X_1, \ldots, X_n \overset{i.i.d}{\sim} \mathrm{Po}(\lambda)$ mit Parameter $\lambda > 0$. Betrachten Sie die Hypothesen

$$H_0 : \lambda = \lambda_0 \quad \text{vs.} \quad H_1 : \lambda = \lambda_1$$

mit $\lambda_0 > \lambda_1 > 0$.

(a) Zeigen Sie, dass der entsprechende Likelihood-Quotient die Form

$$\Lambda_{\boldsymbol{x}}(\lambda_1, \lambda_0) = \left(\frac{\lambda_1}{\lambda_0}\right)^z \exp\bigl(-n(\lambda_1 - \lambda_0)\bigr)$$

für $\boldsymbol{x} = (x_1, \ldots, x_n)^\top$ und $x_1, \ldots, x_n \in \mathbb{N}_0$ und $z = \sum_{i=1}^n x_i$ hat.

(b) Begründen Sie, weshalb ein Test zum exakten Niveau α, der auf der Alternative die größte Güte hat, äquivalent ist zu

$$\phi(z) \quad = \quad \begin{cases} 1 & \text{, falls } z < c_\alpha \\ \gamma_\alpha(z) & \text{, falls } z = c_\alpha \\ 0 & \text{, falls } z > c_\alpha. \end{cases}$$

Wieso ist eine Randomisierung notwendig?

Hinweis: Die Summe von n i.i.d. $\mathrm{Po}(\lambda)$-verteilten Zufallsvariablen ist $\mathrm{Po}(n\lambda)$-verteilt.

(c) Bestimmen Sie den kritischen Wert c_α sowie die Randomisierungskonstante $\gamma_\alpha(z)$.