Das Testproblem

- Die grundsätzliche Frage beim Testproblem ist zu beurteilen, welche von zwei sich widersprechenden Hypothesen wahr ist.
- Etwa
  $H : p = p_0$ gegen $K : p = p_1$
  oder
  $H_0 : p \leq p_0$ gegen $H_1 : p > p_0$
- Problem:Egal, wie wir uns entscheiden, können wir zwei Fehler begehen, nämlich
- den Fehler 1. Art: H_0 ist wahr und wird verworfen
- den Fehler 2. Art: H_0 ist falsch und wird angenommen.

Das Testproblem

- Offensichtlich ist es unmöglich, beide Fehler gleichzeitig in den Griff zu bekommen.
- Versucht man beispielsweise den Fehler erster Art kleinzuhalten, indem man das Ablehnen von H_0 erschwert, so vergrößert dies unausweichlich die Chancen auf einen Fehler zweiter Art.
- In der Statistik hat sich nun eingebürgert, nach einem Test zu suchen, bei dem der Fehler 1. Art kleiner ist als ein vorgegebenes *Signifikanzniveau* α.
- Typische Werte für α sind 5%, 2,5%, 1% oder 0,5%.
- Die Wahl des Werts für α hängt natürlich von der jeweiligen Situation ab: je gravierendere Folgen ein Irrtum hätte, desto kleiner sollten wir α wählen.
- In unserem Beispiel hat ein Fehler 1. Art keine desaströsen Auswirkungen und wir wählen $\alpha = 5\%$.

Das Testproblem

- Das folgende Testverfahren liegt nahe:
- Wir schätzen p aus den Daten mithilfe des naiven Schätzers $\hat{p}$.
- Wenn $\hat{p}$ so aussieht, wie wir es unter H_0 vermuten würden (also z.B. nicht zu weit entfernt von $1/2$ liegt, wenn $H : p = 1/2$), so nehmen wir H_0 an, sonst H_1.
- Hierbei tolerieren wir eine Wahrscheinlichkeit von α für einen Fehler 1. Art. Genauer sieht das Testverfahren so aus:
- Schätze p durch $\hat{p}$
- Bestimme eine Schranke Γ für p
- Akzeptiere H_0, falls $\hat{p} < \Gamma$ und verwerfe H_0 andernfalls.
- Der Wert für die Schranke $\Gamma = \Gamma(\alpha)$ berechnet sich dabei so, dass die Wahrscheinlichkeit für einen Fehler 1. Art höchstens gleich α ist.

Das Testproblem

Formal:

Definition

Zu testen sei die Hypothese $H \subset \Theta$ gegen die Alternative $K \neq \emptyset$. Ein Test is eine Abbildung

$$\phi : \{0,1\}^n \to \{0,1\}.$$

$\phi(x) = 0$ soll bedeuten, dass wir uns für H entscheiden, während $\phi(x) = 1$ bedeutet, wir entscheinden uns für K (wir lehnen die Hypothese ab). Ein Test ist vollständig festgelegt durch das Gebiet $R \subseteq \Theta$, auf dem wir die Hypothese verwerfen (R ist das Verwerfungsgebiet von ϕ), d.h. $\phi(x) = 1 \Leftrightarrow x \in R$.

Das Testproblem

Neben gewöhnlichen Tests betrachten wir auch randomisierte Tests:

Definition

Ein randomisierter Test ist eine Abbildung.

$$\phi : \{0,1\}^n \to [0,1].$$

$\phi(x)$ ist die Wahrscheinlichkeit H abzulehnen.

Das Testproblem

- Natürlich will man zwei gegebene Tests der gleichen Hypothese und Alternative vergleichen.
- Dies geht einerseits über das Niveau des Tests

$$\max_{p \in H} P_p(x \in R)$$

- Dies möchte man i.a. durch die gegeben Schranke $\alpha > 0$ kontrollieren.
- Sind zwei Test zu einem Niveau α vorgelegt, so bietet sich der Fehler zweiter Art als Vergleichskriterium an.
- Man definiert daher (äquivalent)

$$\beta(p) = P_p(x \in R).$$

als die Macht eines Tests mit Verwerfungsbereich R in $p \in K$.

Das Testproblem

- Wir werden nun eine Untersuchung der Güte der oben diskutierten Tests im einfachst möglichen Fall präsentieren, dem Fall, in dem sowohl die Hypothese H als auch die Alternative K aus einem einzigen Punkt bestehen.
- Im Fall einer Folge von i.i.d. Bernoulli Variablen mit unbekanntem Erfolgsparameter p testen wir also die einfache Hypothese
$$H : \{p = p_0\}$$
gegen die einfache Alternative
$$K : \{p = p_1\}.$$

Das Testproblem

- Da wir uns in diesem Fall auch mit randomisierten Tests befassen wollen, verallgemeinern wir die Begriffe des Niveaus und der Macht rasch auf diesen Fall: Für einen randomisierten ϕ ist
$$E_H(\phi) = \sum_x \phi(x) P_H(x)$$
das Niveau des Tests.
- $$E_K(\phi) = \sum_x \phi(x) P_K(x)$$
ist seine Macht.
- Bemerke, dass diese Definitionen konsistent sind mit den Definitionen für nicht-randomisierte Tests.

Wir interessieren uns nun dafür unter allen Tests $\{\phi : E_H(\phi) \leq \alpha\}$ denjenigen Test ϕ^* mit maximaler Macht zu finden.

Definition

Ein Test ϕ^* heißt Neyman-Pearson Test, falls es eine Konstante c^*, $0 \leq c^* \leq \infty$ gibt, so dass

- $\phi^*(x) = 1$ falls $P_K(x) > c^* P_H(x)$
- $\phi^*(x) = 0$ falls $P_K(x) < c^* P_H(x)$
- Auf $\{P_K(x) = c^* P_H(x)\}$ darf der Test ϕ^* beliebige Werte $0 \leq \gamma(x) \leq 1$ annehmen.

Das Testproblem

Definition

Wir werden im folgenden einen Test ϕ_1 *schärfer* nennen als einen Test ϕ_2, falls

$$E_K(\phi_1) > E_K(\phi_2).$$

gilt, die Chance H zu verwerfen, wenn K vorliegt bei ϕ_1 somit größer ist als bei ϕ_2.

Satz (Neyman-Pearson Lemma)

In der Situation des n-fachen Münzwurfs mit Parameter p sei die Hypothese

$$H : \{p = p_0\}$$

gegen die Alternative

$$K : \{p = p_1\}$$

zu testen. Dann gilt

- *Falls* ϕ^* *ein Neyman-Pearson Test ist, dann ist* ϕ^* *schärfer als jeder andere Test* ϕ *mit*

$$E_H(\phi) \leq E_H(\phi^*).$$

- *Für jedes* $0 \leq \alpha \leq 1$ *gibt es einen (randomisierten) Neyman-Pearson Test* ϕ^* *zum Niveau* α, *also mit* $E_H(\phi^*) = \alpha$.

Für einseitige Test, d.h. Tests einer Hypothese H die komplett links (oder komplett rechts) von der Alternative K liegt, überträgt sich die Optimalität mit Hilfe des folgenden

Lemma

Sei X Binomial-verteilt zu den Parametern n und p und $x < n$. Dann ist

$$p \mapsto P_p(X \leq x)$$

stetig und strikt fallend in p und

$$P_0(X \leq x) = 1 \quad \text{und} \quad P_1(X \leq x) = 0.$$