

Schließende Statistik

Neil Mañibo

Universität Bielefeld

28.06.2024

Schließende Statistik

- In den vorherigen Kapiteln: Die Verteilungen sind schon bekannt gegeben
(z.B. Binomial für mehrfachen Münzwurf, Hypergeometrisch für das Ziehen aus einer Urne ohne Zurücklegen, u.s.w.)
- In der Praxis kennt man normalerweise die Verteilung nicht, hat aber eine **Stichprobe**!
- Ziel: diese Stichprobe zu nutzen, um Aussagen über die unbekannte Verteilung zu gewinnen
(z.B. durch Punkt- und Intervallschätzung)
- Normalapproximation der Verteilung möglich für große n
- Einführung zu statistischen Tests
(statistische Signifikanz, Fehlertypen)

Notation

- $X \sim f$, X : ZV, f : Verteilung
“Die ZV X ist genau f -verteilt/folgt die Verteilung f ”.
- $X \approx \mathcal{N}(\mu, \sigma^2)$
“Die ZV X ist *näherungsweise* normalverteilt, mit EW μ und Varianz σ^2 ”.
- $x \simeq x'$, für $x, x' \in \mathbb{R}$
“ x und x' sind ungefähr gleich”.

6.1: Mittelwert, empirische Varianz und Punktschätzung

Erwartungstreuer und konsistenter Schätzer

Definition

Eine ZV $\hat{U}$ ist ein **erwartungstreu** Schätzer für eine Größe u , wenn

$$\mathbb{E}(\hat{U}) = u.$$

Definition

Eine ZV $\hat{U}$ ist ein **konsistent** Schätzer für eine Größe u , wenn

$$\lim_{n \rightarrow \infty} V(\hat{U}) = 0.$$

Wiederholung: Zentraler Grenzwertsatz

Satz

Seien $X_1, X_2, \dots, X_n$ unabhängige und identisch verteilte (i.i.d.)^a ZV mit

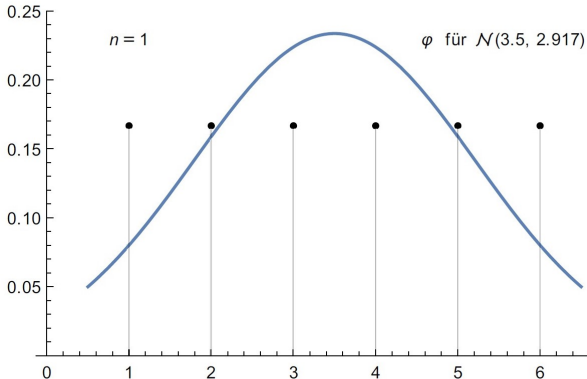
- Erwartungswert: μ , mit $|\mu| < \infty$
- Varianz: $0 < \sigma^2 < \infty$.

Dann ist ihr Mittelwert $\bar{X} := \frac{1}{n} \sum_{i=1}^n X_i$ für großes n näherungsweise **normalverteilt** mit Erwartungswert μ und Varianz $\frac{\sigma^2}{n}$.

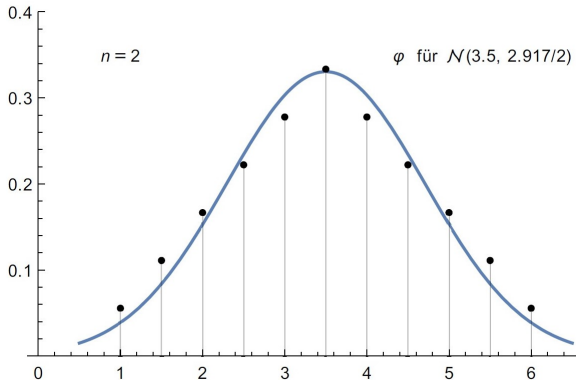
^aindependent and identically distributed

Anders ausgedruckt: $\bar{X} \approx \mathcal{N}(\mu, \frac{\sigma^2}{n})$

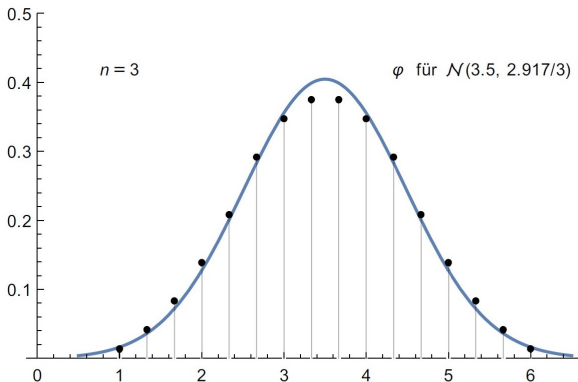
Beispiel: Idealer Würfel ($\mu = 3.5, \sigma^2 = 2.917$)



Beispiel: Idealer Würfel ($\mu = 3.5, \sigma^2 = 2.917$)



Beispiel: Idealer Würfel ($\mu = 3.5, \sigma^2 = 2.917$)



Faustregeln für die Normalapproximation

- X : **binomialverteilt** $\iff X \sim B(n, p)$
Falls $\sigma^2 = np(1-p) \geq 9$ gilt

$$X \approx \mathcal{N}(np, np(1-p))$$

- X : **Poisson-verteilt** $\iff X \sim \text{Poi}(\lambda), \lambda > 0$
Falls $\lambda = \mu = \sigma^2 \geq 10$ gilt

$$X \approx \mathcal{N}(\lambda, \lambda)$$

- X : **hypergeom. verteilt** $\iff X \sim \text{Hyper}(W, S, n)$
Falls $\frac{n}{W+S} \leq 0.05$ und $np(1-p) \geq 9$ (wobei $p = \frac{W}{W+S}$) gilt

$$X \approx \mathcal{N}(np, np(1-p))$$

6.2: Schätzung von Konfidenzintervallen

Punktschätzung vs Konfidenzintervalle

- Letzten Abschnitt: Ermittlung von **einzelnen Schätzwerte** für unbekannte Parameter (Punktschätzung)
- Diesen Abschnitt: Mit diesen Punktschätzungen als Basis, wollen wir **Intervalle** erzeugen, innerhalb derer der Parameter mit einer gewissen Sicherheit liegt.

Wahrscheinlichkeit, dass $Z \in [-c, c]$, $Z \sim \mathcal{N}(0, 1)$

Frage (1)

Was ist die Wahrscheinlichkeit, dass eine standardnormalverteilte ZV Z ein vorgegebenes symmetrische Intervall $[-c, c]$ antrifft?

Wahrscheinlichkeit, dass $Z \in [-c, c]$, $Z \sim \mathcal{N}(0, 1)$

Frage (1)

Was ist die Wahrscheinlichkeit, dass eine standardnormalverteilte ZV Z ein vorgegebenes symmetrische Intervall $[-c, c]$ antrifft?

$$\begin{aligned}\mathbb{P}(Z \in [-c, c]) &= \mathbb{P}(Z \leq c) - \mathbb{P}(Z \leq -c) \\ &= \Phi(c) - \Phi(-c) \\ &= 2\Phi(c) - 1,\end{aligned}$$

$\Phi(x)$: Verteilungsfunktion von $\mathcal{N}(0, 1)$

Bem: $\Phi(-c) = 1 - \Phi(c)$ wegen Symmetrie

Jetzt umgekehrt!

Frage (2)

Mit einer vorgegebenen Wahrscheinlichkeit $1 - \alpha$, wie bestimmen wir ein symmetrisches Intervall $[-c, c]$, so dass die Werte von Z mit Wahrscheinlichkeit $1 - \alpha$ in $[-c, c]$ hineinfallen?

$$2\Phi(c) - 1 = 1 - \alpha$$

$$2\Phi(c) = 2 - \alpha$$

$$\Phi(c) = 1 - \frac{\alpha}{2}$$

Gesucht: der Wert c , sodass $\Phi(c) = 1 - \frac{\alpha}{2}$

Hierfür nutzen wir die Standardnormalverteilungstabelle!

q -Quantil

Definition

Derjenige Wert z_q , für den

$$\Phi(z_q) = q$$

gilt bezeichnet man als das **q -Quantil** der Standardnormalverteilung.

Anders ausgedrückt: $c = z_q = \Phi^{-1}(q)$

(z_q eindeutig, da Φ *streng* monoton steigend ist!)

q -Quantil

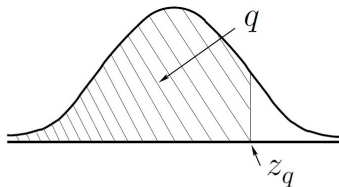


Abbildung 6.2: Der Wert z_q ist das q -Quantil der Verteilung. Die Masse der Verteilung bis z_q ist gerade q .

Beispiele: q -Quantile der Standardnormalverteilung

❶ $z_{0.5} = 0$ (warum?)

❷ $z_{0.975} = ?$

Beispiele: q -Quantile der Standardnormalverteilung

② $z_{0.975} = 1.90 \text{ (Zeile)} + 0.06 \text{ (Spalte)} = 1.96$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890

③ Analog gilt $z_{0.9115} = 1.35$.

Zurück zur Frage (2)

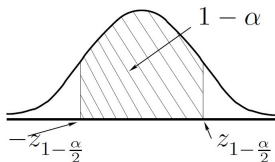


Abbildung 6.3: Die Werte der ZV Z sollen mit Wahrscheinlichkeit $1 - \alpha$ in ein symmetrisches Intervall um 0 fallen. Es sind die entsprechenden Quantile eingetragen.

$$\mathbb{P}\left(Z \in \left[-z_{1-\frac{\alpha}{2}}, z_{1-\frac{\alpha}{2}}\right]\right) = 1 - \alpha \quad \text{für } Z \in \mathcal{N}(0, 1)$$