Vorlesung 7b

Mittelwerte

Die Wahrscheinlichkeit typischer Abweichungen:

Approximative Normalität

Die Wahrscheinlichkeit großer Abweichungen:

Chernoff-Schranken

Teil I

Die Wahrscheinlichkeit typischer Abweichungen beim zufälligen Mittelwert:

Approximative Normalität

1. Populationsmittelwert und Stichprobenmittelwert

Denken wir an eine Liste (eine “Population”)

von reellen Daten

$$w_1, \ldots, w_g$$

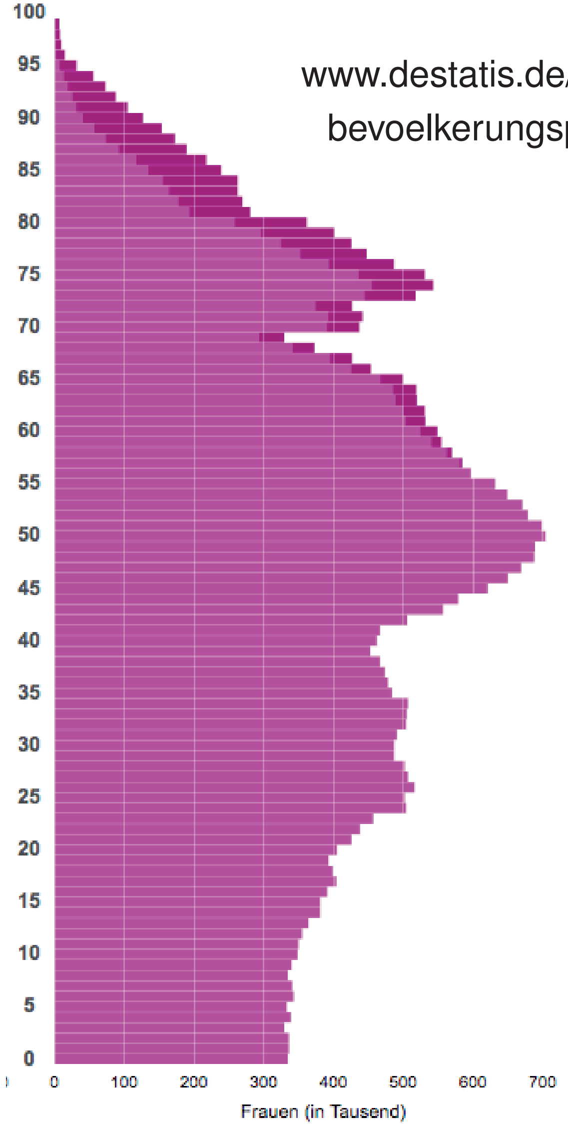
z. B. die Lebensalter aller Frauen

in der deutschen Bevölkerung 2014.

w_i ist der Wert (die Kenngröße) des Individuums i.

Eine komprimierte Darstellung liefern die

Besetzungszahlen (hier der Altersklassen von 1 bis 100+):



www.destatis.de/
bevoelkerungspyramide/

Angenommen man möchte den Populationsmittelwert

$$\mu := \frac{1}{g} \sum_{j=1}^{g} w_j.$$

schätzen,

und zwar aus den Werten einer

aus der Population gezogenen Stichprobe

$$x_1, \ldots, x_n$$

(sagen wir für $n = 100$).

Als *Schätzwert* für μ bietet sich an:

$$m := \frac{1}{n}(x_1 + \ldots + x_n)$$

Wie zuverlässig ist diese Schätzung?

Goldene Idee der Statistik:

Man fasst $x_1, \ldots, x_n$ auf als Ergebnis eines rein zufälligen Ziehens aus der Population:

$$X_1 := w_{J_1},\ X_2 := w_{J_2}, \ldots$$

mit $J_1, J_2, \ldots$ rein zufällige Wahl aus $\{1, \ldots, g\}$ ("Ziehen mit Zurücklegen").

Wir setzen hier g als (sehr) groß gegenüber n voraus,

damit entstehen *auch*

beim n-maligen Ziehen *mit* Zurücklegen

Kollisionen nur mit (verschwindend) kleiner W'keit.

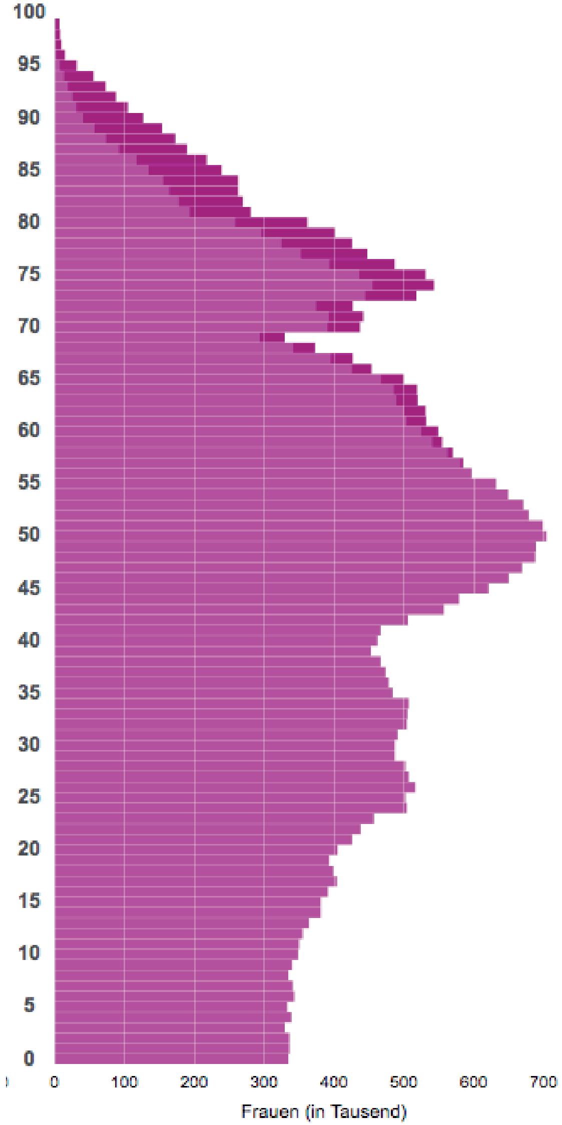
Das führt auf die Vorstellung:

$x_1, \dots, x_n$ sind entstanden

durch n-maliges unabhängiges Ziehen $X_1, \dots, X_n$

aus der Verteilung ρ auf $\mathbb{R}$

mit $\rho([c,d]) := \frac{1}{g}\#\{i : w_i \in [c,d]\}$


100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
0
0
100
200
300
400
500
600
700
Frauen (in Tausend)

$$m = \frac{1}{n}(x_1 + \ldots + x_n)$$

fasst man also auf

als *eine* Realisierung (*einen* beobachteten Ausgang)

der **Zufallsvariable**

$$M := \bar{X} := \frac{1}{n}(X_1 + \ldots + X_n)$$

(des Mittelwertes der zufälligen Stichprobe $(X_1, \ldots, X_n)$).

$$M := \bar{X} := \frac{1}{n}(X_1 + \ldots + X_n)$$

Es gilt:

$$\mathbf{E}[X_j] = \frac{1}{g}\sum_{i}^{g} w_i = \mu$$

und damit auch

$$\mathbf{E}[M] = \mu.$$

Der Erwartungswert des Stichprobenmittelwertes

ist gleich dem Populationsmittelwert.

2. Populationsvarianz
und Varianz des Stichprobenmittelwertes

Zur Erinnerung: X_1 war der Wert eines (“des ersten”) aus der Population rein zufällig gezogenen Individuums.

$\mathbf{E}[X_1]$ ist gleich dem Populationsmittelwert μ.

Und

$$\mathbf{Var}[X_1] = \frac{1}{g}\sum_{j=1}^{g}(w_j - \mu)^2 =: \sigma^2.$$

Diese Zahl σ^2 nennt man auch die **Populationsvarianz**.

Der **Stichprobenmittelwert** war

$$M = \frac{1}{n}(X_1 + \ldots + X_n)$$

$$M = \frac{1}{n}(X_1 + \ldots + X_n)$$

$$\mathbf{E}[M] = \mu$$

$$\mathbf{Var}[M] = ?$$

Wird mit Zurücklegen gezogen, dann sind die X_i unabhängig, und es ergibt sich

$$\mathbf{Var}[M] = \frac{\sigma^2}{n}$$

Wird ohne Zurücklegen gezogen
und ist die Populationsgröße g nicht sehr groß
gegenüber der Stichprobengröße n, dann hat es Sinn,
die *Korrektur für endliche Populationen* zu berücksichtigen
(vgl. Aufgabe 20):

$$\mathbf{Var}[M] = \frac{\sigma^2}{n} \cdot \frac{g-n}{g-1}$$

Diese Korrektur werden wir
für den Rest dieser Vorlesung vernachlässigen (wir denken
an großes g, bzw. – wie schon gesagt – an ein wiederholtes
unabhängiges Ziehen aus einer Verteilung).

Für ein n-maliges unabhängiges Ziehen gilt:

$$\mathbf{Var}[M] = \frac{\sigma^2}{n};$$

die Standardabweichung des Stichprobenmittelwertes M ist also $\frac{\sigma}{\sqrt{n}}$.

3. Approximative Verteilung des Stichprobenmittelwertes

Wie ist (für nicht zu kleines n)

der Stichprobenmittelwert M verteilt?

Der Zentrale Grenzwertsatz gibt eine Antwort:

In der oben beschriebenen Situation gilt

M ist approximativ $N(\mu, \frac{\sigma^2}{n})$-verteilt.

4. Die Stichprobenvarianz als Schätzung für die Populationsvarianz

Ein Problem in der Praxis: Im Allgem. kennt man σ^2 nicht.

Auch σ^2 muss man dann schätzen.

Zwei Vorschläge für die

(aus der Stichprobe) **geschätzte** (Populations-)**Varianz**:

(i) die *Stichprobenvarianz*

$$\hat{\sigma}^2 := \frac{1}{n} \sum_{i=1}^{n} (x_i - m)^2.$$

(ii) die *modifizierte Stichprobenvarianz*

$$s^2 := \frac{1}{n-1} \sum_{i=1}^{n} (x_i - m)^2$$

Es gibt theoretische Begründung für beide Vorschläge

(vgl. Buch S. 124, S. 138).

Wir halten uns hier erst einmal an den Vorschlag (ii):

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - m)^2$$

Die Standardabweichung des Stichprobenmittelwertes M ist

$$\frac{\sigma}{\sqrt{n}}.$$

Die geschätzte Standardabweichung

des Stichprobenmittelwertes M ist

$$\boxed{s/\sqrt{n} =: f}$$

Diese Größe nennen wir auch den *Standardfehler*.

M ist approximativ $N(\mu, \frac{\sigma^2}{n})$-verteilt.

Und (gut für die Praxis):

M ist approximativ $N(\mu, f^2)$-verteilt.

Teil II

Die Wahrscheinlichkeit großer Abweichungen beim zufälligen Mittelwert:

Chernoff-Schranken

5. Die Chernoff-Ungleichung für Binom(n, p)

Zur Wahrscheinlichkeit

großer Abweichungen vom Erwartungswert

bei der Münzwurf-Trefferquote

Sei X_n Binomial(n, p)-verteilt, und $\alpha > p$.

Wir wissen schon aus dem Gesetz der großen Zahlen:

$$\mathbf{P}\left(\frac{X_n}{n} > \alpha\right) \to 0 \quad \text{für} \quad n \to \infty.$$

Wie schnell ist diese Konvergenz?

Die Chebyshev-Ungleichung liefert nur die Ordnung $O(1/n)$.

Gibt es eine asymptotisch "scharfe" Abschätzung

für $\mathbf{P}\left(\frac{X_n}{n} > \alpha\right)$?

Es stellt sich heraus:

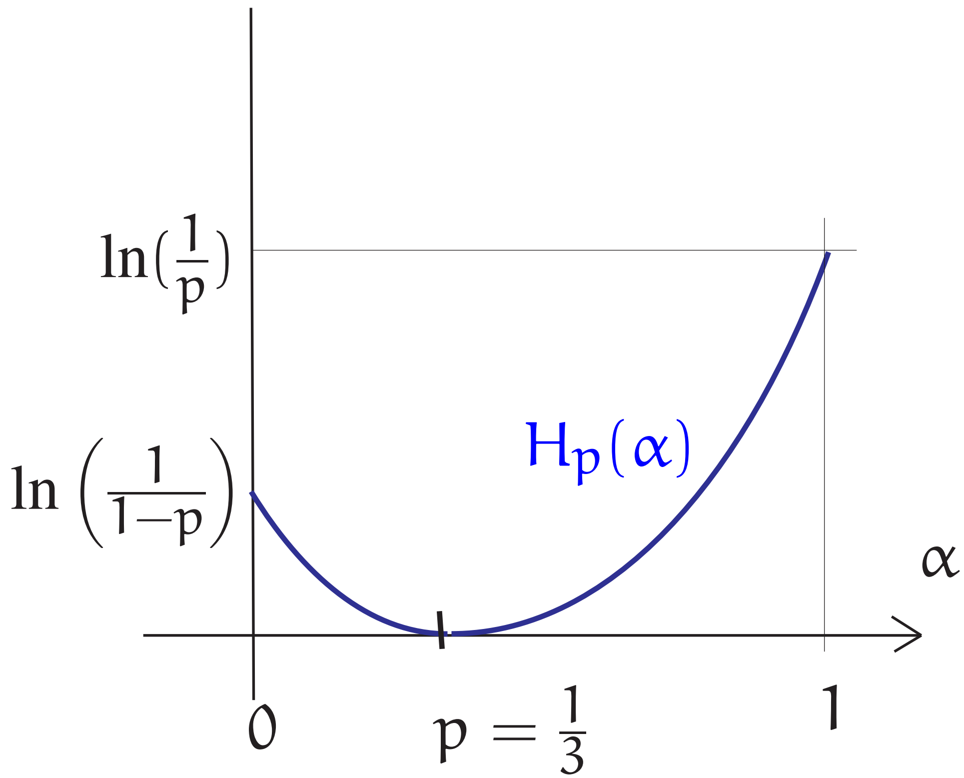
Die Konvergenz (gegen Null) ist exponentiell schnell.

Genauer: Man hat die Chernoff-Ungleichung

$$\mathbf{P}(X_n/n > \alpha) \leq e^{-nH_p(\alpha)}$$

mit $H_p(\alpha) := \alpha \ln\left(\frac{\alpha}{p}\right) + (1-\alpha)\ln\left(\frac{1-\alpha}{1-p}\right) > 0$

... die relative Entropie von Bernoulli(α) bzgl. Bernoulli(p)



$\ln(\frac{1}{p})$
$\ln\left(\frac{1}{1-p}\right)$
$H_p(\alpha)$
α
0
$p = \frac{1}{3}$
1

Ein Zahlenbeispiel:

$$n = 10000,\ p = 0.5, \quad \alpha = 0.6$$

$$H_{0.5}(0.6) = 0.0201$$

Die Wahrscheinlichkeit,
bei einem 10000-maligen fairen Münzwurf
mindestens 6000 Erfolge zu erzielen, ist nicht größer als

$$e^{-nH_p(\alpha)} = e^{-201} \approx 5 \cdot 10^{-88}$$

6. Die exponentielle Markov-Ungleichung

Als Nächstes wenden wir uns dem

Beweis der Chernoff-Ungleichung zu.

Ein wesentlicher Schritt dabei ist die

exponentielle Markov-Ungleichung:

X sei eine rellwertige ZV'e. Dann gilt für alle $b \in \mathbb{R}$ und $t > 0$:

$$\mathbf{P}(X \geq b) = \mathbf{P}(e^{tX} \geq e^{tb}) \leq \frac{1}{e^{tb}} \mathbf{E}[e^{tX}].$$

Also:

$$\boxed{\mathbf{P}(X \geq b) \leq \inf_{t \geq 0} \frac{1}{e^{tb}} \mathbf{E}[e^{tX}]}$$

7. Herleitung der Chernoff-Ungleichung

für die Wahrscheinlichkeit großen Abweichungen

bei der Trefferquote des p-Münzwurfs

aus der exponentiellen Markov-Ungleichung:

$$\mathbf{P}(X \geq b) \leq \inf_{t\geq 0} \frac{1}{e^{tb}}\mathbf{E}[e^{tX}]$$

Wir berechnen die rechte Seite für $X = X_n := Z_1 + \cdots + Z_n$,

mit einem p-Münzwurf (Z_i). Es gilt:

$$\mathbf{E}[e^{tZ_i}] = (1-p) + pe^t.$$

Aus der Produktformel für Erwartungswerte folgt:

$$\mathbf{E}[e^{tX}] = ((1-p) + pe^t)^n.$$

Mit $b := \alpha n$ folgt:

$$\frac{1}{e^{tb}}\mathbf{E}[e^{tX}] = e^{-t\alpha n}((1-p)+pe^t)^n = ((1-p)e^{-t\alpha}+pe^{t(1-\alpha)})^n$$

$$\frac{1}{e^{tb}}\mathbf{E}[e^{tX}] = ((1-p)e^{-t\alpha} + pe^{t(1-\alpha)})^n$$

Für welches t wird $g(t) := (1-p)e^{-t\alpha} + pe^{t(1-\alpha)}$ minimal?

g konvergiert für $t \to \infty$ nach ∞;

$$g'(t) = -\alpha(1-p)e^{-t\alpha} + p(1-\alpha)e^{t(1-\alpha)}$$

$$= e^{-t\alpha}\left(-\alpha(1-p) + p(1-\alpha)e^{t}\right)$$

ist negativ bei $t = 0$ und verschwindet genau bei

$$e^t = \frac{\alpha}{1-\alpha}\frac{1-p}{p}.$$

$e^t = \frac{\alpha}{1-\alpha}\frac{1-p}{p}$ eingesetzt in $\dfrac{(1-p+pe^t)^n}{e^{t\alpha n}}$

ergibt

$$\frac{\left((1-p)+(1-p)\frac{\alpha}{1-\alpha}\right)^n}{\left(\frac{\alpha}{1-\alpha}\frac{1-p}{p}\right)^{\alpha n}} = \left(\frac{p}{\alpha}\right)^{\alpha n}\left(\frac{1-p}{1-\alpha}\right)^{(1-\alpha)n}$$

$$= e^{-n\left(\alpha \ln\left(\frac{\alpha}{p}\right)+(1-\alpha)\ln\left(\frac{1-\alpha}{1-p}\right)\right)}$$

$$= e^{-nH_p(\alpha)}. \quad \square$$

8. Die Chernoff-Ungleichung für Gamma(k)

Übungsaufgabe extra – für die stillste Zeit im Jahr:

$X := Y_1 + \ldots + Y_k$

mit $Y_1, \ldots, Y_k$ unabhängig und Exp(1)-verteilt*

(i) Zeigen Sie für $t < 1$: $\mathbf{E}[e^{tX}] = \dfrac{1}{(1-t)^k}$

(ii) Zeigen Sie für $\alpha > 1$:

$$\mathbf{P}(X > \alpha k) \leq e^{-k(\alpha - 1 - \ln \alpha)}.$$

(iii) Finden Sie für $k = 10000$ die Chernoff-Schranke für

$$\mathbf{P}(X > 11000).$$

*Ein solches X heißt Gamma(k)-verteilt