Kapitel 11

DIE NORMAL-VERTEILUNG

Fassung vom 17. Februar 2006

11.1 Definition der Normal-Verteilung

Bisher haben wir nur diskret verteilte Zufallsvariable betrachtet. Bei physikalischen Messungen, wie z.B. Länge, Gewicht, usw., muß man aber *kontinuierlich* verteilte Zufallsvariable X zulassen, deren Wertebereich $\mathbb{R}$ oder Intervalle in $\mathbb{R}$ sind. Dementsprechend ist der Wahrscheinlichkeitsraum sehr kompliziert. Im Allgemeinen ist man aber nur an der Wahrscheinlichkeit P von Ereignissen der Gestalt

$$\{\omega \in \Omega \mid X(\omega) \in I\} \ ,$$

wobei I ein Intervall in $\mathbb{R}$ ist, interessiert, z.B.

$$P\left(a \leqslant X \leqslant b\right) := P\left(\{\omega \in \Omega \mid a \leqslant X(\omega) \leqslant b\}\right)$$

oder

$$P(X \leqslant b) := P\left(\{\omega \in \Omega \mid X(\omega) \leqslant b\}\right)$$

für $a, b \in \mathbb{R}$, und dazu genügt es, die Verteilung von X zu kennen. Die wichtigste Verteilung bei kontinuierlich verteilten Zufallsvariablen X ist die Normalverteilung.

DEFINITION 1 Die Funktion $\Phi : \mathbb{R} \longrightarrow \mathbb{R} : z \longmapsto \Phi(z)$ definiert durch

$$\Phi(z) := \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{z} e^{-\frac{t^2}{2}}\, dt$$

heißt *Standard-Normal-Verteilung* . Die Funktion $\varphi : \mathbb{R} \longrightarrow \mathbb{R}$ definiert durch

$$\varphi(t) := \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}}$$

nennt man die *Dichtefunktion* von Φ .

BEMERKUNG 1 $\Phi(z)$ ist nicht durch eine einfache Formel, sondern in Tabellen gegeben. Da nach Beispiel 7.6.4

$$\frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}}\, dt \ = \ 1,$$

folgt mit der Zerlegungsregel

$$\Phi(-z) = 1 - \Phi(z) \quad \text{für alle } z \in \mathbb{R} \ .$$

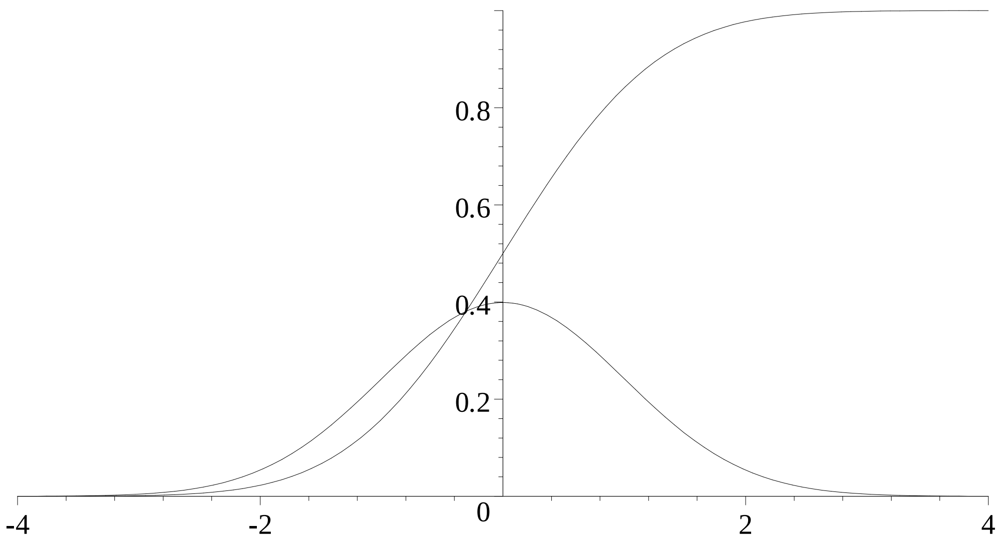
Demnach wird nur $\Phi(z)$ für $z \geqslant 0$ tabelliert :

z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$
0	.5	.2	.57926	.4	.65542	.6	.72575	.8	.78814
.01	.50399	.21	.58317	.41	.6591	.61	.72907	.81	.79103
.02	.50798	.22	.58706	.42	.66276	.62	.73237	.82	.79389
.03	.51197	.23	.59095	.43	.6664	.63	.73565	.83	.79673
.04	.51595	.24	.59483	.44	.67003	.64	.73891	.84	.79955
.05	.51994	.25	.59871	.45	.67364	.65	.74215	.85	.80234
.06	.52392	.26	.60257	.46	.67724	.66	.74537	.86	.80511
.07	.5279	.27	.60642	.47	.68082	.67	.74857	.87	.80785
.08	.53188	.28	.61026	.48	.68439	.68	.75175	.88	.81057
.09	.53586	.29	.61409	.49	.68793	.69	.7549	.89	.81327
.1	.53983	.3	.61791	.5	.69146	.7	.75804	.9	.81594
.11	.5438	.31	.62172	.51	.69497	.71	.76115	.91	.81859
.12	.54776	.32	.62552	.52	.69847	.72	.76424	.92	.82121
.13	.55172	.33	.6293	.53	.70194	.73	.7673	.93	.82381
.14	.55567	.34	.63307	.54	.7054	.74	.77035	.94	.82639
.15	.55962	.35	.63683	.55	.70884	.75	.77337	.95	.82894
.16	.56356	.36	.64058	.56	.71226	.76	.77637	.96	.83147
.17	.56749	.37	.64431	.57	.71566	.77	.77935	.97	.83398
.18	.57142	.38	.64803	.58	.71904	.78	.7823	.98	.83646
.19	.57535	.39	.65173	.59	.7224	.79	.78524	.99	.83891

z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$
1.0	.84134	1.2	.88493	1.4	.91924	1.6	.9452	1.8	.96407
1.01	.84375	1.21	.88686	1.41	.92073	1.61	.9463	1.81	.96485
1.02	.84614	1.22	.88877	1.42	.9222	1.62	.94738	1.82	.96562
1.03	.84849	1.23	.89065	1.43	.92364	1.63	.94845	1.83	.96638
1.04	.85083	1.24	.89251	1.44	.92507	1.64	.9495	1.84	.96712
1.05	.85314	1.25	.89435	1.45	.92647	1.65	.95053	1.85	.96784
1.06	.85543	1.26	.89617	1.46	.92785	1.66	.95154	1.86	.96856
1.07	.85769	1.27	.89796	1.47	.92922	1.67	.95254	1.87	.96926
1.08	.85993	1.28	.89973	1.48	.93056	1.68	.95352	1.88	.96995
1.09	.86214	1.29	.90147	1.49	.93189	1.69	.95449	1.89	.97062
1.1	.86433	1.3	.9032	1.5	.93319	1.7	.95543	1.9	.97128
1.11	.8665	1.31	.9049	1.51	.93448	1.71	.95637	1.91	.97193
1.12	.86864	1.32	.90658	1.52	.93574	1.72	.95728	1.92	.97257
1.13	.87076	1.33	.90824	1.53	.93699	1.73	.95818	1.93	.9732
1.14	.87286	1.34	.90988	1.54	.93822	1.74	.95907	1.94	.97381
1.15	.87493	1.35	.91149	1.55	.93943	1.75	.95994	1.95	.97441
1.16	.87698	1.36	.91309	1.56	.94062	1.76	.9608	1.96	.975
1.17	.879	1.37	.91466	1.57	.94179	1.77	.96164	1.97	.97558
1.18	.881	1.38	.91621	1.58	.94295	1.78	.96246	1.98	.97615
1.19	.88298	1.39	.91774	1.59	.94408	1.79	.96327	1.99	.9767

z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$
2.0	.97725	2.2	.9861	2.4	.9918	.6	.99534	2.8	.99744
2.01	.97778	2.21	.98645	2.41	.99202	.61	.99547	2.81	.99752
2.02	.97831	2.22	.98679	2.42	.99224	.62	.9956	2.82	.9976
2.03	.97882	2.23	.98713	2.43	.99245	.63	.99573	2.83	.99767
2.04	.97932	2.24	.98745	2.44	.99266	.64	.99585	2.84	.99774
2.05	.97982	2.25	.98778	2.45	.99286	.65	.99598	2.85	.99781
2.06	.9803	2.26	.98809	2.46	.99305	.66	.99609	2.86	.99788
2.07	.98077	2.27	.9884	2.47	.99324	.67	.99621	2.87	.99795
2.08	.98124	2.28	.9887	2.48	.99343	.68	.99632	2.88	.99801
2.09	.98169	2.29	.98899	2.49	.99361	.69	.99643	2.89	.99807
2.1	.98214	2.3	.98928	2.5	.99379	.7	.99653	2.9	.99813
2.11	.98257	2.31	.98956	2.51	.99396	.71	.99664	2.91	.99819
2.12	.983	2.32	.98983	2.52	.99413	.72	.99674	2.92	.99825
2.13	.98341	2.33	.9901	2.53	.9943	.73	.99683	2.93	.99831
2.14	.98382	2.34	.99036	2.54	.99446	.74	.99693	2.94	.99836
2.15	.98422	2.35	.99061	2.55	.99461	.75	.99702	2.95	.99841
2.16	.98461	2.36	.99086	2.56	.99477	.76	.99711	2.96	.99846
2.17	.985	2.37	.99111	2.57	.99492	.77	.9972	2.97	.99851
2.18	.98537	2.38	.99134	2.58	.99506	.78	.99728	2.98	.99856
2.19	.98574	2.39	.99158	2.59	.9952	.79	.99736	2.99	.99861

z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$
3.0	.99865	3.2	.99931	3.4	.99966	3.6	.99984	3.8	.99993
3.01	.99869	3.21	.99934	3.41	.99968	3.61	.99985	3.81	.99993
3.02	.99874	3.22	.99936	3.42	.99969	3.62	.99985	3.82	.99993
3.03	.99878	3.23	.99938	3.43	.9997	3.63	.99986	3.83	.99994
3.04	.99882	3.24	.9994	3.44	.99971	3.64	.99986	3.84	.99994
3.05	.99886	3.25	.99942	3.45	.99972	3.65	.99987	3.85	.99994
3.06	.99889	3.26	.99944	3.46	.99973	3.66	.99987	3.86	.99994
3.07	.99893	3.27	.99946	3.47	.99974	3.67	.99988	3.87	.99995
3.08	.99896	3.28	.99948	3.48	.99975	3.68	.99988	3.88	.99995
3.09	.999	3.29	.9995	3.49	.99976	3.69	.99989	3.89	.99995
3.1	.99903	3.3	.99952	3.5	.99977	3.7	.99989	3.9	.99995
3.11	.99906	3.31	.99953	3.51	.99978	3.71	.9999	3.91	.99995
3.12	.9991	3.32	.99955	3.52	.99978	3.72	.9999	3.92	.99996
3.13	.99913	3.33	.99957	3.53	.99979	3.73	.9999	3.93	.99996
3.14	.99916	3.34	.99958	3.54	.9998	3.74	.99991	3.94	.99996
3.15	.99918	3.35	.9996	3.55	.99981	3.75	.99991	3.95	.99996
3.16	.99921	3.36	.99961	3.56	.99981	3.76	.99992	3.96	.99996
3.17	.99924	3.37	.99962	3.57	.99982	3.77	.99992	3.97	.99996
3.18	.99926	3.38	.99964	3.58	.99983	3.78	.99992	3.98	.99997
3.19	.99929	3.39	.99965	3.59	.99983	3.79	.99992	3.99	.99997



$\varphi : t \longmapsto \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}}$ und $\Phi : z \longmapsto \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{z} e^{-\frac{t^2}{2}}\, dt$

DEFINITION 2 Eine Zufallsvariable X heißt *normalverteilt* mit Erwartungswert μ und Varianz σ^2 , man schreibt dann

$$X \text{ ist } N(\mu, \sigma) \text{-verteilt,}$$

wenn

$$P(X \leqslant b) = N(\mu, \sigma)(b) := \Phi\left(\frac{b-\mu}{\sigma}\right) \quad \text{für alle } b \in \mathbb{R}$$

gilt.

BEMERKUNG 2 Ist X $N(\mu, \sigma)$-verteilt, so gilt

$$P(X \leqslant b) = P(X < b) = 1 - P(X \geqslant b) = 1 - P(X > b)$$

und

$$P(a \leqslant X \leqslant b) = P(X \leqslant b) - P(X \leqslant a) \ .$$

Insbesondere gilt

$$P\Big(\mu - r\cdot\mu \leqslant X \leqslant \mu + r\cdot\mu\Big) = P\Big(X \leqslant \mu + r\cdot\mu\Big) - P\Big(X \leqslant \mu - r\cdot\mu\Big) =$$

$$= \Phi\left(\frac{r\cdot\mu}{\sigma}\right) - \Phi\left(\frac{-r\cdot\mu}{\sigma}\right) = 2\cdot\Phi\left(\frac{r\cdot\mu}{\sigma}\right) - 1 \ .$$

BEISPIEL Bei einem Präparat sei das Gewicht X des Wirkstoffes in den Tabletten $N(\mu, \sigma)$-verteilt mit $\mu = 80$ mg und $\sigma = 2$ mg .

Wie groß ist die Wahrscheinlichkeit, daß bei einer Tablette die Menge an Wirkstoff um höchstens 5% von μ abweicht ? Es ist

$$P\left(\mu - 0.05 \cdot \mu \leqslant X \leqslant \mu + 0.05 \cdot \mu\right) = 2 \cdot \Phi\left(\frac{0.05 \cdot \mu}{\sigma}\right) - 1 = 2 \cdot \Phi(2) - 1 =$$

$$= 2 \cdot 0.97725 - 1 = 0.955 = 95.5\%$$

mit Hilfe der Tabelle.

Wie groß ist die Standard-Abweichung σ zu wählen, damit mindestens 99% der Tabletten höchstens 5% Abweichung von μ haben ? Nach obiger Rechnung muß gelten

$$2 \cdot \Phi\left(\frac{0.05 \cdot \mu}{\sigma}\right) - 1 \geqslant 0.99 \ ,$$

d.h.

$$\Phi\left(\frac{4}{\sigma}\right) \geqslant \frac{1.99}{2} = 0.995 \ .$$

Nach der Tabelle ist dies gerade erfüllt, wenn

$$\frac{4}{\sigma} \geqslant 2.58$$

gilt, also folgt

$$\sigma \leqslant \frac{4}{2.58} = 1.55 \ .$$

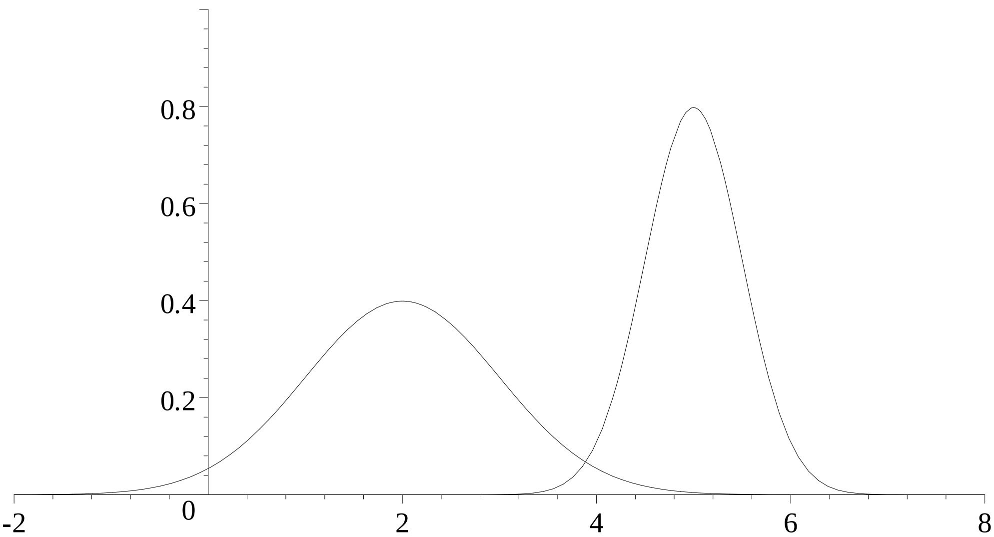
11.2 Berechnung der Normal-Verteilung

Ist X eine $N(\mu,\sigma)$-verteilte Zufallsvariable, so gilt

$$P(X \leqslant b) = \Phi\left(\frac{b-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}} \cdot \int_{-\infty}^{\frac{b-\mu}{\sigma}} e^{-\frac{t^2}{2}}\, dt =$$

$$= \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot \int_{-\infty}^{b} e^{-\frac{(t-\mu)^2}{2\cdot\sigma^2}}\, dt = \frac{1}{\sigma} \cdot \int_{-\infty}^{b} \varphi\left(\frac{t-\mu}{\sigma}\right)\, dt$$

nach der Verschiebungs- und Streckungsregel. Die $N(\mu,\sigma)$-Verteilung $b \longmapsto \Phi\left(\frac{b-\mu}{\sigma}\right)$ hat also die Dichte $t \longmapsto \frac{1}{\sigma} \cdot \varphi\left(\frac{t-\mu}{\sigma}\right)$.



$t \longmapsto \varphi(t-2) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{(t-2)^2}{2}}$, $t \longmapsto 2 \cdot \varphi(2 \cdot (t-5)) = \frac{2}{\sqrt{2\pi}} \cdot e^{-2\cdot(t-5)^2}$

Die Wendepunkte dieser Dichten liegen bei $\mu \pm \sigma$, d.h. 2 ± 1 und 5 ± 0.5 .

ANWENDUNG Zu einer Zufallsvariable X ist eine Meßreihe $x_1, x_2, \cdots, x_n$ gegeben, und man vermutet, daß sie normalverteilt ist.

(a) Zuerst werden μ und σ mit Hilfe dieser Meßdaten geschätzt :

$$\mu := \frac{1}{n} \cdot \sum_{j=1}^{n} x_j \quad \text{und} \quad \sigma^2 := \frac{1}{n} \cdot \sum_{j=1}^{n} (x_j - \mu)^2 \ .$$

(b) Dann untersucht man graphisch, ob eine Normal-Verteilung vorliegt. Man unterteilt das Messintervall $[a,b]$ (mit $x_j \in [a,b]$ für alle j) in m Teilintervalle $I_1, I_2, \cdots, I_m$ der gleichen Länge L , wobei $m < n$, am besten so, daß $m \simeq \sqrt{n}$. Man definiert dann eine Treppenfunktion $\tilde{\varphi}$

durch

$$\tilde{\varphi} = \frac{1}{L} \cdot \frac{k}{n} \text{ auf } I_l \quad \text{falls } k \text{ der Meßwerte } x_1, \ldots, x_n \text{ in } I_l \text{ liegen.}$$

Die Normierung ist gerade so gewählt, dass das Integral über $\tilde{\varphi}$, ebenso wie das über $\frac{1}{\sigma} \cdot \varphi\left(\frac{x-\mu}{\sigma}\right)$, den Wert 1 hat.

Ist $\tilde{\varphi}$ eine gute Approximation von $\frac{1}{\sigma} \cdot \varphi\left(\frac{x-\mu}{\sigma}\right)$, wobei μ und σ wie in (a) bestimmt sind, so geht man davon aus, daß X normalverteilt ist.

BEMERKUNG Eine naturwissenschaftliche Zufallsvariable X (z.B. das Geburtsgewicht) setzt sich aus vielen Faktoren X_j zusammen (z.B. genetischen Faktoren, Ernährung der Mutter, ...). Es gilt der *Zentrale Grenzwertsatz :*

HAUPTSATZ *Ist $X = X_1 + X_2 + \cdots$, und sind alle X_j klein gegen X und paarweise unabhängig, so ist X normal verteilt.*

Dies erklärt das häufige Auftreten von Normal-Verteilungen bei naturwissenschaftlichen Experimenten. Eine typische Situation wird auch durch Satz 11.3 beschrieben.

BEISPIEL Gewicht der Gesamtbevölkerung : Es ist nicht normalverteilt. Z.B. sind Alter und Geschlecht nicht kleine Einflußgrößen.

11.3 Normal- und Binomial-Verteilung

Ein typisches Problem der Statistik ist die Schätzung von Wahrscheinlichkeiten. Wir diskutieren dies am Beispiel der Binomial-Verteilung. Ausgangspunkt ist der folgende

SATZ *Ist X eine binomialverteilte Zufallsvariable mit Erwartungswert $\mu = n \cdot p$ und Varianz $\sigma^2 = n \cdot p \cdot q$, und ist $\sigma^2 \gg 1$, (i.a. genügt $\sigma > 3$), so ist X annähernd $N(\mu, \sigma)$-verteilt, d.h es gilt*

$$P(X \leqslant k) = \sum_{j=0}^{k} \binom{n}{j} \cdot p^j \cdot q^{n-j} \simeq \Phi\left(\frac{k-\mu}{\sigma}\right) .$$

Insbesondere gilt

$$P\Big(n \cdot p - z \cdot \sigma \leqslant X \leqslant n \cdot p + z \cdot \sigma\Big) \simeq 2 \cdot \Phi\left(\frac{z \cdot \sigma}{\sigma}\right) - 1 = 2 \cdot \Phi(z) - 1 =: \gamma(z) .$$

Die letzte Aussage folgt Bemerkung 11.1.2. □

DEFINITION 1 Die Zahl $\gamma(z) := 2 \cdot \Phi(z) - 1$ wird als *Konfidenzzahl* bezeichnet.

BEMERKUNG 1 Aus der Tabelle für Φ liest man die folgenden, häufig benutzten Werte für $\gamma(z)$, und damit Näherungswerte für $P\Big(n \cdot p - z \cdot \sigma \leqslant X \leqslant n \cdot p + z \cdot \sigma\Big)$ ab :

z	1	1.96	2.58	2.97	3.29
$\gamma(z)$	0.683	0.95	0.99	0.997	0.999

Frage Man möchte die Zahl $p \in [0,1]$, die praktisch nicht bekannt ist, mit Hilfe einer Stichprobe ω schätzen. Man wählt

$$\widehat{p} := \frac{X(\omega)}{n}$$

als Schätzwert für p .

Für die Stichprobe ω gilt also mit Wahrscheinlichkeit $\gamma(z)$

$$n \cdot p - z \cdot \sigma \leqslant X(\omega) \leqslant n \cdot p + z \cdot \sigma ,$$

d.h.

$$p - \frac{z \cdot \sigma}{n} \leqslant \widehat{p} \leqslant p + \frac{z \cdot \sigma}{n} \qquad \text{oder} \qquad -\frac{z \cdot \sigma}{n} \leqslant p - \widehat{p} \leqslant \frac{z \cdot \sigma}{n} ,$$

und mit $\sigma^2 = n \cdot p \cdot (1-p)$ auch

$$(p - \widehat{p})^2 \leqslant \left(\frac{z \cdot \sigma}{n}\right)^2 = \frac{z^2 \cdot [n \cdot p \cdot (1-p)]}{n^2} = \frac{z^2}{n} \cdot p(1-p) . \qquad (*)$$

Es folgt

$$p^2 - 2p \cdot \widehat{p} + \widehat{p}^2 \leqslant \frac{z^2}{n} \cdot p(1-p) = \frac{z^2}{n} \cdot p - \frac{z^2}{n} \cdot p^2 ,$$

also

$$f(p) := \left(1+\frac{z^2}{n}\right)\cdot p^2 - \left(2\widehat{p}+\frac{z^2}{n}\right)\cdot p + \widehat{p}^2 \leqslant 0 \ .$$

Somit gilt

$$f(p) \leqslant 0 \quad \text{mit Wahrscheinlichkeit} \quad \gamma(z) \ .$$

Die zwei Nullstellen des Polynoms f werden mit $p_{\pm}$ bezeichnet :

$$p_{\pm} = \frac{1}{2\left(1+\frac{z^2}{n}\right)}\cdot\left(2\widehat{p}+\frac{z^2}{n} \pm \sqrt{\left(2\widehat{p}+\frac{z^2}{n}\right)^2 - 4\left(1+\frac{z^2}{n}\right)\cdot\widehat{p}^2}\right) \ .$$

Weiterhin ist

$$f(0) = \widehat{p}^2 > 0 \qquad \text{und} \qquad f(1) = \left(1+\frac{z^2}{n}\right) - \left(2\widehat{p}+\frac{z^2}{n}\right) + \widehat{p}^2 = (1-\widehat{p})^2 > 0$$

und

$$f(\widehat{p}) = \left(1+\frac{z^2}{n}\right)\cdot\widehat{p}^2 - \left(2\widehat{p}+\frac{z^2}{n}\right)\cdot\widehat{p} + \widehat{p}^2 = \frac{z^2}{n}\cdot\widehat{p}^2 - \frac{z^2}{n}\cdot\widehat{p} = \frac{z^2}{n}\cdot\widehat{p}\cdot(\widehat{p}-1) < 0 \ ,$$

da die Werte 0 und 1 für $\widehat{p}$ unwahrscheinlich sind. Damit haben wir gezeigt, daß

$$0 < p_- < p_+ < 1$$

gilt.

Da $f(p) \leqslant 0$ zu $p \in [p_-, p_+]$ äquivalent ist, folgt die erste Behauptung im folgenden

KOROLLAR *Seien X eine $b_{n,p}$-verteilte Zufallsvariable mit $\sigma^2 \gg 1$, ω eine Stichprobe und*

$$\widehat{p} := \frac{X(\omega)}{n}$$

den Schätzwert von p .

(i) Mit Wahrscheinlichkeit $\gamma(z)$ gilt (näherungsweise):

$$p \in [p_-, p_+] \ ,$$

sowie

$$p \in \left[\widehat{p} - \frac{z}{2\sqrt{n}}, \widehat{p} + \frac{z}{2\sqrt{n}}\right] \ .$$

(ii) Ist $0 < p_0 \leqslant \frac{1}{2}$ und besitzt man die Vorinformation $p \leqslant p_0$ über p , so gilt

$$p \in \left[\widehat{p} - \frac{z}{\sqrt{n}}\cdot\sqrt{p_0(1-p_0)}, \widehat{p} + \frac{z}{\sqrt{n}}\cdot\sqrt{p_0(1-p_0)}\right]$$

mit Wahrscheinlichkeit $\gamma(z)$.

Für die zweite Aussage in (i) folgt aus $(*)$ und nachstehendes Lemma, daß mit Wahrscheinlichkeit $\gamma(z)$ gilt

$$(p-\widehat{p})^2 \leqslant \frac{z^2}{n}\cdot p(1-p) \leqslant \frac{z^2}{n}\cdot\frac{1}{4} \ ,$$

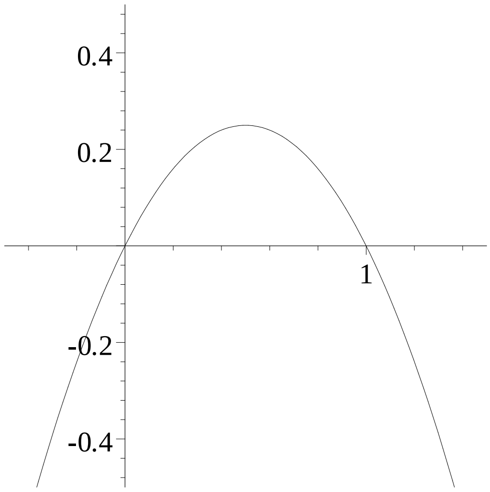
also $|p-\widehat{p}| \leqslant \frac{1}{2}\cdot\frac{z}{\sqrt{n}}$, d.h. $-\frac{1}{2}\cdot\frac{z}{\sqrt{n}} \leqslant p-\widehat{p} \leqslant \frac{1}{2}\cdot\frac{z}{\sqrt{n}}$ oder $p \in \left[\widehat{p} - \frac{z}{2\sqrt{n}}, \widehat{p} + \frac{z}{2\sqrt{n}}\right]$.

Gilt $p \leqslant p_0 \leqslant \frac{1}{2}$, so folgt

$$(p - \widehat{p})^2 \leqslant \frac{z^2}{n} \cdot p\,(1-p) \leqslant \frac{z^2}{n} \cdot p_0\,(1-p_0) \ ,$$

mit Wahrscheinlichkeit $\gamma(z)$, woraus die Behauptung sich wie im ersten Teil ergibt. — □

LEMMA *Für das Polynom* $x \longmapsto x \cdot (1-x) : \mathbb{R} \longrightarrow \mathbb{R}$



$x \longmapsto x\,(1-x)$

gilt

$$0 \leqslant x \cdot (1-x) \leqslant \frac{1}{4} \qquad \textit{für alle } x \in [0,1] \ .$$

In der Tat ist

$$0 \leqslant x \cdot (1-x) = x - x^2 = \frac{1}{4} - \left(x - \frac{1}{2}\right)^2 \leqslant \frac{1}{4} \qquad \text{für alle } x \in [0,1] \ .$$

□

BEMERKUNG 2 Die zweite Aussage ist einfacher und ungenauer, d.h. das Intervall ist größer, dagegen aber sicherer !

DEFINITION 2 Ein Intervall $[a,b] \subset [0,1]$, so daß gilt

$$p \in [a,b] \quad \text{mit Wahrscheinlichkeit} \quad \gamma(z) \ ,$$

heißt *Konfidenzintervall* zum Konfidenzniveau $\gamma(z)$.

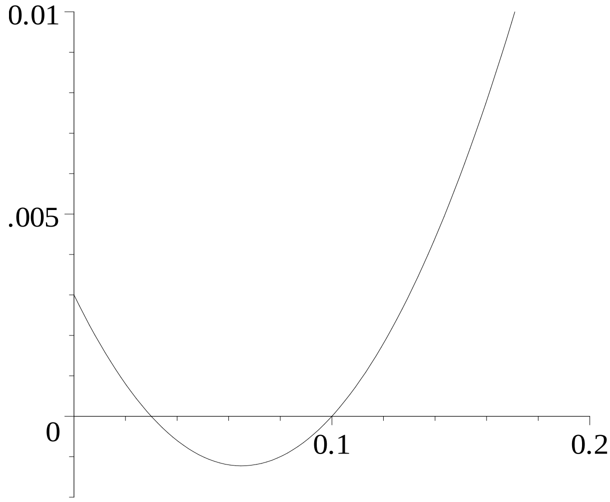
BEISPIEL 1 Innerhalb einer Population tritt eine allergische Reaktion auf eine Chemikalie ein.

Von 200 zufällig ausgewählten Personen reagieren 11 allergisch, es ist also

$$X(\omega) = 11 \quad \text{und} \quad \widehat{p} = \frac{11}{200} = 0.055 \ .$$

Zur Konfidenzzahl $\gamma = 0.95$, also $z = 1.96$, erhält man das Polynom (gerundet)

$$f(p) = p^2 - 0.13 \cdot p + 0.003 \leqslant 0 \ .$$



$$p \longmapsto p^2 - 0.13 \cdot p + 0.003$$

Das gesuchte Konfidenzintervall ergibt sich aus den Nullstellen $p_\pm$ von f :

$$p_\pm = \frac{0.13}{2} \pm \sqrt{\left(\frac{0.13}{2}\right)^2 - 0.003} \ ,$$

und damit $p_- = 0.03$ und $p_+ = 0.1$.

Mit 95% Sicherheit liegt also die Wahrscheinichkeit p für eine allergische Reaktion zwischen 0.03 und 0.1 .

BEISPIEL 2 Man möchte die Schätzung von p aus dem Beispiel 1 durch eine Vergrößerung der Stichprobe verbessern. Um 100 % Sicherheit zu bekommen, müsste man fast alle Personen untersuchen, man wählt deshalb wieder eine Konfidenzzahl $\gamma < 1$. Wir benutzen weiter den Wert aus dem Beispiel 1, also $\gamma = 0.95$ und $z = 1.96$, und suchen ein n so, daß mit Wahrscheinlichkeit γ der Wert von p im Konfidenzintervall $[\widehat{p} - \alpha, \widehat{p} + \alpha]$ liegt, wobei z.B. $\alpha = 0.02$ sei. Der Schätzwert $\widehat{p}$ sei wieder durch $\frac{X(\omega)}{n} = \frac{11}{200} = 0.055$ gegeben.

Mit Hilfe des Korollars genügt es n so zu wählen, daß

$$\frac{z}{2\sqrt{n}} \leqslant \alpha$$

gilt, d.h.

$$n \geqslant \frac{z^2}{4 \cdot \alpha^2} = \frac{1.96^2}{4 \cdot 0.02^2} = 2401 \ .$$

BEMERKUNG 3 Besitzt man Vorinformationen über p , z.B. wird man nach Beispiel 1 $p \leqslant 0.1$ annehmen können, so genügt es n so zu wählen, daß

$$\frac{z}{\sqrt{n}} \cdot \sqrt{0.1 \cdot (1 - 0.1)} \leqslant \alpha = 0.02 \ ,$$

oder

$$n \geqslant \frac{1.96^2}{0.02^2} \cdot 0.09 = 864.36 \ ,$$

d.h. $n = 865$ zu nehmen.