

1.6 Wichtige Verteilungsmodelle

Literatur: Fahrmeir et. al., 2004, 5.3, 6.3.1

Hier nur Binomial- und Normalverteilung. Einige weitere Verteilungsmodelle direkt dort, wo sie benötigt werden. (Für weitere Modelle sei auf die Literatur verwiesen.)

1.6.1 Binomialverteilung

Def 1.41 Seien n und k natürliche Zahlen oder 0.

$$\text{i) } n! := n \cdot (n - 1) \cdot (n - 2) \cdot \dots \cdot 2 \cdot 1 \quad (1.6.1)$$

$$0! := 1$$

$n!$ bezeichnet man als „ n Fakultät.“

$$\text{ii) } \binom{n}{k} := \frac{n!}{k!(n-k)!} \text{ heißt Binomialkoeffizient.}$$

Anmerkungen:

- i) $n!$ gibt die Anzahl aller möglichen Anordnungen von n verschiedenen Elementen an.
 $\binom{n}{k}$ ist die Anzahl aller k -elementigen Auswahlen aus n ohne Berücksichtigung der Ziehungsreihenfolge.
z.B. beim Lotto gibt es $\binom{49}{6}$ verschiedene Ergebnisse (ohne Zusatz- und Superzahl).
- ii) Bei der konkreten Berechnung von Binomialkoeffizienten empfiehlt es sich häufig, „vorher zu kürzen“:

Bem 1.42 *Das Grundmodell der Binomialverteilung*

- n unabhängige Wiederholungen eines Experiments, bei dem ein Ereignis A mit Wahrscheinlichkeit π eintritt.
- Betrachtet wird die Zufallsvariable X , die die Anzahl der Versuche zählt, bei denen A eintritt.
- häufig A : "Treffer", X : Anzahl der "Treffer".
- Es gilt für alle $x \in \{0, 1, \dots, n\}$

$$P(X = x) = \binom{n}{x} \cdot \pi^x \cdot (1 - \pi)^{n-x} \quad (1.6.2)$$

(und $P(X = x) = 0$ für alle anderen x)

- X heißt *binomialverteilt*; das Wahrscheinlichkeitsmaß aus (1.6.2) heißt *Binomialverteilung*. Abkürzung: $X \sim B(n, \pi)$
- Es gilt für binomialverteiltes X

- $\mathbb{E}X = n \cdot \pi$ (1.6.3)

- $VarX = n \cdot \pi \cdot (1 - \pi)$ (1.6.4)

Beispiel 1.43

Risikobereite Slalomfahrer stürzen mit Wahrscheinlichkeit 10%, vorsichtigere mit 2%.

- a) Wie groß sind jeweils die Wahrscheinlichkeiten, dass von je 20 Fahrern mindestens einer stürzt?
- b) Vergleichen Sie die jeweils durchschnittlich zu erwartende Anzahl von Stürzen von je 100 Rennläufern!

kleiner Exkurs: Zur Problematik der Argumentation mittels „natürlicher Häufigkeiten“, wie sie Gigerenzer erfolgreich empfohlen hat. Man würde demgemäß die Wahrscheinlichkeit $\pi_r=0,1$ kommunizieren als „von 100 stürzen 10 Rennläufer“. Diese Interpretation läuft Gefahr, die beträchtliche Variabilität zufälliger Prozesse zu verschleiern. In der Tat ist hier die Wahrscheinlichkeit, dass genau 10 von 100 Läufern stürzen,

$$P(X = 10) = \binom{100}{10} \cdot 0.1^{10} \cdot 0.9^{90} = 0.13,$$

also lediglich etwa 13%.

Bem 1.44 *Zwei weitere Eigenschaften der Binomialverteilung:*

- Symmetrieeigenschaft:
(Vertausche Rolle von Treffer und Nichttreffer)

- Summeneigenschaft:
Seien $X \sim B(n, \pi)$ und $Y \sim B(m, \pi)$;

Bem 1.45 *Tabellierung der Binomialverteilung*

Tabelliert ist oft $P(X \leq x)$
z.B. Fahrmeier et. al.

$\pi = 0.3$	$n = 11$	$n = 12$...
$x \leq 0$	0.0198	0.0138	
1	0.1130	0.0850	...
2	0.3127	0.2528	
\vdots	\vdots	\vdots	

$$\begin{aligned}
 P(X = x) &= P(X \leq x) - P(X \leq x - 1), \quad x \in \mathbb{N}_0 \\
 \text{z.B. } P(X = 2) &= P(X \leq 2) - P(X \leq 1) = \\
 &= 0.1997 (=) \\
 &= \binom{11}{2} \cdot 0.3^2 \cdot 0.7^9.
 \end{aligned}$$

Wegen der Symmetrieeigenschaft gibt es meist nur Tabellen für $\pi \leq 0.5$.
Für großes n verwendet man Approximationen durch die Normalverteilung.

1.6.2 Normalverteilung

Def 1.46 Definition

Eine stetige Zufallsvariable X heißt *normalverteilt* mit den Parametern μ und σ^2 , in Zeichen $X \sim \mathcal{N}(\mu, \sigma^2)$, wenn für ihre Dichte gilt:

$$f(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right), \quad x \in \mathbb{R}, \quad (1.6.5)$$

und *standardnormalverteilt*, in Zeichen $X \sim \mathcal{N}(0; 1)$, falls $\mu = 0$ und $\sigma^2 = 1$.

Bem. 1.47 (grundlegende Eigenschaften)

- a) Die *Dichte der Standardnormalverteilung* wird oft mit $\varphi(x)$ bezeichnet, also

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right); \quad (1.6.6)$$

die zugehörige Verteilungsfunktion $\int_{-\infty}^x \varphi(u) du$ mit $\Phi(x)$.

- b) $\Phi(x)$ lässt sich nicht in geschlossener Form durch bekannte Funktionen beschreiben \implies numerische Berechnung, Tabellierung.
- c) μ und σ^2 sind genau der Erwartungswert und die Varianz, also, wenn $X \sim \mathcal{N}(\mu, \sigma^2)$, dann

$$\mathbb{E}X = \mu \quad \text{und} \quad \text{Var}X = \sigma^2.$$

(Jede Normalverteilung ist durch μ und σ^2 eindeutig bestimmt.)

- d) Die Dichte ist symmetrisch um μ , d.h.

$$f(\mu - x) = f(\mu + x).$$

Bem 1.48 Grundlegendes zum Rechnen mit Normalverteilungen

- $\Phi(-x) = 1 - \Phi(x)$ (1.6.7)
- Ist $X \sim \mathcal{N}(\mu, \sigma^2)$, so ist die zugehörige standardisierte Zufallsvariable

$$Z = \frac{X - \mu}{\sigma}$$

standardnormalverteilt.

- ganz wichtig für Tabellierung! Es reicht die Standardnormalverteilung zu tabellieren. Normalverteilte Zufallsvariablen mit Erwartungswert μ und Varianz σ^2 muss man zuerst standardisieren, dann kann man aber auch die Standardnormalverteilungsfunktionstabelle verwenden.
→ s.a. Übungsaufgaben, hier auch ein Beispiel

Tabelliert sind die Werte der Verteilungsfunktion $\Phi(z) = P(Z \leq z)$ für $z \geq 0$.

Ablesebeispiel: $\Phi(1.75) = 0.9599$

Funktionswerte für negative Argumente: $\Phi(-z) = 1 - \Phi(z)$

Die z -Quantile ergeben sich über die Umkehrfunktion.

Beispielsweise ist $z_{0.9599} = 1.75$ und $z_{0.9750} = 1.96$.

	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	0.7704	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
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993

Wie bestimmt man bei $X \sim \mathcal{N}(\mu, \sigma^2)$ die Wahrscheinlichkeiten $P(X \leq a)$ aus der Tabelle?

Prop 1.49 *Abgeschlossenheit gegenüber Linearkombinationen:*

Seien X_1 und X_2 unabhängig und $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, $i = 1, 2$. Ferner seien b, a_1, a_2 feste reelle Zahlen. Dann gilt

$$Y_1 := a_1 X_1 + b \sim \mathcal{N}(a_1 \mu_1 + b; a_1^2 \sigma_1^2) \quad (1.6.9)$$

und

$$Y_2 := a_1 X_1 + a_2 X_2 \sim \mathcal{N}(a_1 \mu_1 + a_2 \mu_2; a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2) \quad (1.6.10)$$

Beispiel 1.50 (Fahrmeir et al.)

- Schultischhöhe: $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$, $\mu_Y = 113$, $\sigma_Y^2 = 16$
Stuhlhöhe: $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$, $\mu_X = 83$, $\sigma_X^2 = 25$
- optimale Sitzposition: Tisch zwischen 27 und 29 cm höher als Stuhl.
- Wie groß ist die Wahrscheinlichkeit, dass ein zufällig ausgewähltes Paar zueinander gut passt?