Inhalt von Abschnitt 3

Inhalt von Abschnitt 3 (Forts.)

Summen spezieller Zufallsvariablen

Diskrete Zufallsvariablen

Definition 3.1 (Zufallsvariable)

Eine Variable oder ein Merkmal X, dessen Werte oder Ausprägungen die Ergebnisse eines Zufallsvorgangs sind, heißt *Zufallsvariable X*. Die Zahl $x \in \mathbb{R}$, die X bei einer Durchführung des Zufallsvorgangs annimmt, heißt *Realisierung* oder Wert von X.

Beispiele:

- Würfelsumme bei Werfen eines Doppelwürfels: $X \in \{2, 3, 4, \ldots, 12\}$
- Anzahl Erfolge in Bernoullikette der Länge n: $X \in \{0, 1, 2, \ldots, n\}$
- Aktienkurse $X \in \mathbb{R}_+$

- Von Interesse: *Ereignisse* der Art
  - $\{X = x\}$
  - $\{X \neq x\}$
  - $\{X \leq x\}$
  - $\{X > x\}$
  - $\{a \leq X \leq b\}$
  - $\{X \in I\}$

  Weitere durch X definierte zulässige Bereiche: durch Komplementbildung und abzählbare Durchschnitts- und Vereinigungsbildung
- *Wahrscheinlichkeitsverteilung:*
  $P\{X \in B \mid B \text{ zulässiger Bereich}\}$

Definition 3.2 (Diskrete Zufallsvariable)

Eine Zufallsvariable heißt *diskret*, falls sie nur endlich oder *abzählbar unendlich* viele Werte $x_1, x_2, \ldots$ annehmen kann. Die *Wahrscheinlichkeitsverteilung* von X ist gegeben durch die Wahrscheinlichkeiten

$$P(X = x_i) = p_i, \quad i = 1, 2, \ldots .$$

- Die Wertemenge von X heißt *Träger* von X und wird bezeichnet mit $\mathcal{T} = \{x_1, x_2, \ldots\}$.
- aus Kolmogoroff:

$$0 \leq p_i \leq 1$$

$$p_1 + p_2 + \ldots = \sum_{i \geq 1} p_i = 1.$$

- $$P(X \in A) = \sum_{i: x_i \in A} p_i$$
- Insbesondere
$$P(a \leq X \leq b) = \sum_{i: a \leq x_i \leq b} p_i.$$

Definition 3.3 (Bernoulli-Variable (Verteilung))

$$X = \begin{cases} 1, & \text{falls } A \text{ eintritt,} \\ 0, & \text{falls } A \text{ nicht eintritt.} \end{cases}$$

Wenn $P(A) = \pi \Longrightarrow P(X = 1) = \pi,\ P(X = 0) = 1 - \pi$.

Mehrkategorielle Zufallsvariable

Sind $A_1, \ldots, A_k$ sich gegenseitig ausschließende Ereignisse mit $\bigcup A_i = \Omega$. Definiere

$$X = i, \text{ falls } A_i \text{ eintritt}, i = 1, \ldots k.$$

Verteilung von X ist durch $P(X = i)$ gegeben.

Definition 3.4 (Wahrscheinlichkeitsfunktion)

Die *Wahrscheinlichkeitsfunktion* einer diskreten Zufallsvariable X ist für $x \in \mathbb{R}$ definiert durch

$$f(x) = \begin{cases} P(X = x_i) = p_i & \text{für } x = x_i \in \{x_1, x_2, \ldots\}, \\ 0 & \text{sonst.} \end{cases}$$

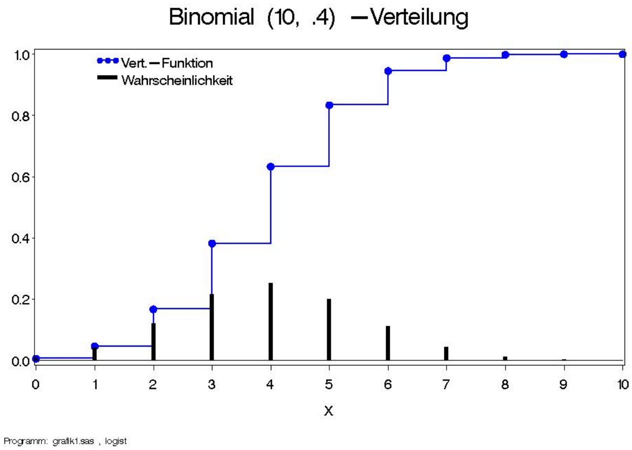
- Darstellung per Stabdiagramm oder Häufigkeitshistogramm

Definition 3.5 (Verteilungsfunktion einer diskreten ZV)

$$F(x) := P(X \leq x) = \sum_{i: x_i \leq x} f(x_i)$$

Eigenschaften:

- rechtsseitig stetig,
- $x < x_1 \Longrightarrow F(x) = 0$,
- $x \longrightarrow \infty \Longrightarrow F(x) \longrightarrow 1$.


Binomial (10, .4) —Verteilung
Vert.—Funktion
Wahrscheinlichkeit
1.0
0.8
0.6
0.4
0.2
0.0
0
1
2
3
4
5
6
7
8
9
10
X
Programm: grafik1.sas , logist

Definition 3.6 (Diskrete Gleichverteilung)

Eine diskrete Zufallsvariable X heißt gleichverteilt auf dem Träger $\mathcal{T} = \{x_1, x_2, \ldots, x_k\}$, wenn für $1 \leq i \leq k$ gilt

$$P(X = x_i) = \frac{1}{k}.$$

Definition 3.7 (Geometrische Verteilung)

Ein Bernoulli-Versuch werde bei gleichbleibender W-keit $P(A) = \pi$ solange wiederholt, bis zum ersten Mal das Ereignis A eintritt. Dann heißt die ZVe

$$X = \text{"Anzahl Versuche, bis zum ersten Mal A eintritt"}$$

geometrisch verteilt mit Parameter π, $X \sim G(\pi)$. Der Träger von X ist $\mathcal{T} = \{1, 2, \ldots\} = \mathbb{N}$, die Wahrscheinlichkeiten sind

$$P(X = i) = \pi \cdot (1 - \pi)^{i-1}.$$

- Einfaches Modell für diskrete Lebensdauer oder Wartezeitverteilung.

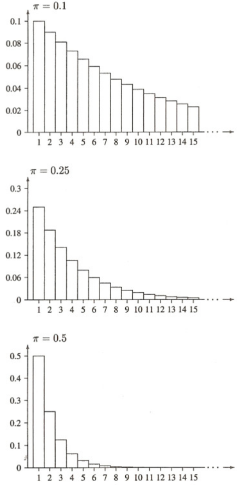


ABBILDUNG 5.9: Wahrscheinlichkeitshistogramme zur geometrischen Verteilung

Unabhängigkeit von diskreten Zufallsvariablen

Bei Zufallsvorgängen interessiert man sich oft für mehrere ZVen. In vielen Fällen beeinflussen sich solche ZVen gegenseitig, sodass sie als voneinander abhängig angesehen werden müssen. Dies geschieht zum Beispiel, wenn im Rahmen einer zufälligen StiPro mehrere Merkmale an den Untersuchungseinheiten erhoben werden.

Falls sich zwei oder mehr ZVen nicht gegenseitig beeinflussen, spricht man von *unabhängigen* ZVen. Die wichtigste derartige Situation liegt vor, wenn Zufallsvorgänge unabhängig voneinander wiederholt werden.

Definition 3.8 (Unabhängigkeit von diskreten Zufallsvariablen)

Zwei diskrete ZVen X und Y mit Trägern $\mathcal{T}_X = \{x_1, x_2, \ldots\}$ und $\mathcal{T}_Y = \{y_1, y_2, \ldots\}$ heißen unabhängig, wenn für beliebige $x \in \mathcal{T}_X$ und $y \in \mathcal{T}_Y$ gilt:

$$P(X = x, Y = y) = P(X = x) \cdot P(Y = y).$$

Allgemein heißen n ZVen $X_1, \ldots, X_n$ unabhängig, wenn für beliebige Werte aus ihren Trägern gilt:

$$P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i).$$

Parameter einer diskreten Verteilung

Definition 3.9 (Erwartungswert einer diskreten Zufallsvariable)

Der *Erwartungswert* einer diskreten ZVe X mit Werten $x_1, x_2, \ldots$ und Wahrscheinlichkeiten $p_1, p_2, \ldots$ ist

$$\mathbb{E}(X) := \sum_{i \geq 1} x_i p_i.$$

Äquivalent: $$\mathbb{E}(X) := \sum_{i \geq 1} x_i f(x_i).$$

Statt $\mathbb{E}(X)$ schreibt man auch μ_X oder μ.

- μ_X charakterisiert *Lage* der Verteilung, ohne dass Daten vorliegen.

Beispiel 3.10 (Geometrische Verteilung)

$$\begin{aligned}\mathbb{E}(X) &= \sum_{i=1}^{\infty} x_i f(x_i) = \sum_{i=1}^{\infty} i \cdot \pi \cdot (1-\pi)^{i-1} \\ &= \frac{1}{\pi}.\end{aligned}$$

- Anwendung der Geometrischen Verteilung: NNT (Number Needed to Treat) beim Vergleich der Wirksamkeit zweier Medikamente.

Transformationsregel für Erwartungswerte

Sei $g(x)$ eine reelle Funktion. Dann gilt für $Y = g(X)$:

$$\mathbb{E}(Y) = \mathbb{E}(g(X)) = \sum_{i \geq 1} g(x_i) p_i = \sum_{i \geq 1} g(x_i) f(x_i).$$

Lineare Transformationsregel für E-Werte

Für $Y = aX + b$ ist

$$\mathbb{E}(Y) = a \cdot \mathbb{E}(X) + b.$$

Erwartungswert einer Summe von ZVen

Für ZVen X und Y gilt:

$$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y).$$

Allgemein, mit beliebigen Konstanten $a_1, \ldots, a_n$:

$$\mathbb{E}(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i \mathbb{E}(X_i).$$

Produktregel für unabhängige ZVen

Für zwei *unabhängige* ZVen X und Y gilt:

$$\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y).$$

Andere Lageparameter: *Median & Quantile*

Median und Quantile setzen ein ordinales Skalenniveau voraus.

Definition 3.11 (Quantil)

Jeder Wert x_p mit $0 < p < 1$, für den

$$P(X \leq x_p) = F(x_p) \geq p \text{ und } P(X \geq x_p) \geq 1 - p$$

heißt p-*Quantil* der diskreten ZVe X mit Verteilungsfunktion $F(x)$.
Für $p = 0.5$ heißt $x_{0.5}$ *Median*.

Andere Lageparameter: *Median & Quantile* (Forts.)

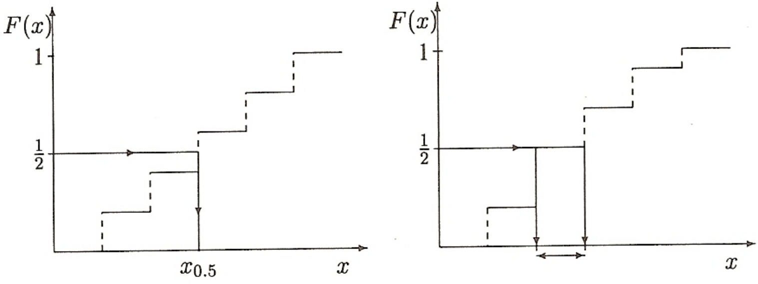


ABBILDUNG 5.11: Eindeutiger Median (links) und nicht eindeutiger Median (rechts)

Figure: aus Fahrmeir et al.

Varianz und Standardabweichung

Definition 3.12 (Varianz und Standardabweichung)

Die *Varianz* einer diskreten ZVe X ist:

$$\begin{aligned}\sigma^2 = \mathbb{V}(X) &= (x_1-\mu)^2 p_1 + \ldots + (x_k-\mu)^2 p_k + \ldots \\ &= \sum_{i\geq 1} (x_i-\mu)^2 f(x_i) \\ &= \mathbb{E}[(X-\mu)^2].\end{aligned}$$

Die *Standardabweichung* ist

$$\sigma = +\sqrt{\mathbb{V}(X)}.$$

Statt $\mathbb{V}(X)$ schreibt man auch $Var(X)$.

Varianz und Standardabweichung (Forts.)

Verschiebungsregel

$$\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \mathbb{E}(X^2) - \mu^2.$$

allgemeiner:

$$\mathbb{V}(X) = \mathbb{E}((X-c)^2) - (\mu - c)^2.$$

Lineare Transformation

Für $Y = aX + b$ ist

$$\mathbb{V}(Y) = \mathbb{V}(aX + b) = a^2\mathbb{V}(X) \text{ und } \sigma_Y = |a|\sigma_X.$$

Varianz und Standardabweichung (Forts.)

Varianz der Summe von unabhängigen ZVen

Für unabhängige ZVen X und Y bzw. $X_1, \ldots, X_n$ gilt

$$\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y)$$

und mit beliebigen Konstanten $a_1, \ldots, a_n$:

$$\mathbb{V}(\sum_i a_i X_i) = \sum_i a_i^2 \mathbb{V}(X_i).$$

Geometrische Verteilung

Genese: Anzahl der unabhängigen Wiederholungen eines Bernoulli-Experiments mit konstanter Wahrscheinlichkeit π, bis zum ersten Mal "Erfolg" eintritt.

Definition 3.13 (Geometrische Verteilung)

X ist *geometrisch* verteilt mit Parameter π $(X \sim G(\pi))$, wenn

$$P(X = i) = \pi \cdot (1 - \pi)^{i-1}.$$

Der Träger von X ist $\mathcal{T} = \{1, 2, 3, \ldots\} = \mathbb{N}$. Erwartungswert und Varianz ergeben sich zu

$$\mathbb{E}(X) = \frac{1}{\pi}, \quad \mathbb{V}(X) = \frac{1 - \pi}{\pi^2}.$$

Binomialverteilung

Genese: Anzahl der Erfolge bei n unabhängigen Wiederholungen eines Bernoulli-Experiments mit konstanter Wahrscheinlichkeit π (Bernoulli-Kette der Länge n).

Definition 3.14 (Binomialverteilung)

X ist *binomial* verteilt mit Parametern n und π $(X \sim B(n, \pi))$, wenn

$$P(X = k) = \begin{cases} \binom{n}{k} \pi^k (1-\pi)^{n-k} & \text{, falls } k = 0, 1, 2, \ldots, n \\ 0 & \text{, sonst.} \end{cases}$$

Der Träger von X ist $\mathcal{T} = \{0, 1, 2, \ldots, n\}$. Erwartungswert und Varianz ergeben sich zu

$$\mathbb{E}(X) = n\pi, \quad \mathbb{V}(X) = n\pi(1-\pi).$$

Negative Binomialverteilung (I)

Genese: Anzahl der unabhängigen Wiederholungen eines Bernoulli-Experiments mit konstanter Wahrscheinlichkeit π, bis r mal "Erfolg" eintritt.

Definition 3.15 (Negative Binomialverteilung I)

X ist *negativ binomial* verteilt mit Parametern r und π $(X \sim NB(r, \pi))$, wenn

$$P(X = n) = \begin{cases} \binom{n-1}{r-1}\pi^r(1-\pi)^{n-r} & \text{, falls } n = r, r+1, r+2, \ldots, \\ 0 & \text{, sonst.} \end{cases}$$

Der Träger von X ist $\mathcal{T} = \{r, r+1, r+2, \ldots\}$. Erwartungswert und Varianz ergeben sich zu

$$\mathbb{E}(X) = \frac{r}{\pi}, \quad \mathbb{V}(X) = \frac{r(1-\pi)}{\pi^2}.$$

Hypergeometrische Verteilung

Genese: In einer Urne mit N Objekten sind M markiert. In einer Stichprobe (ohne Zurücklegen) vom Umfang n ist die Anzahl der markierten Elemente hypergeometrisch verteilt.

Definition 3.16 (Hypergeometrische Verteilung)

X ist *hypergeometrisch* verteilt mit Parametern n, M und N ($X \sim H(n, M, N)$), wenn

$$P(X = i) = \begin{cases} \frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n}} & \text{, falls } i \in \mathcal{T}, \\ 0 & \text{, sonst.} \end{cases}$$

Der Träger von X ist $\mathcal{T} = \{\max(0, n - (N - M)), \ldots, \min(n, M)\}$. Erwartungswert und Varianz ergeben sich zu

$$\mathbb{E}(X) = n\frac{M}{N}, \quad \mathbb{V}(X) = n\frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1}.$$

Poisson-Verteilung

Genese: Die Poisson-Verteilung eignet sich zur Modellierung von *Zählvorgängen*. Dabei werden Ereignisse gezählt, die innerhalb eines festen, vorgegebenen Zeitintervalls eintreten können. Die mögliche Anzahl der Ereignisse ist nicht nach oben begrenzt.

Definition 3.17 (Poisson-Verteilung)

X ist *Poisson* verteilt mit Parameter μ ($X \sim Po(\mu)$), wenn

$$P(X = i) = \begin{cases} e^{-\mu} \frac{\mu^i}{i!} & \text{für } i \in \{0, 1, 2, \ldots\}, \\ 0 & \text{sonst.} \end{cases}$$

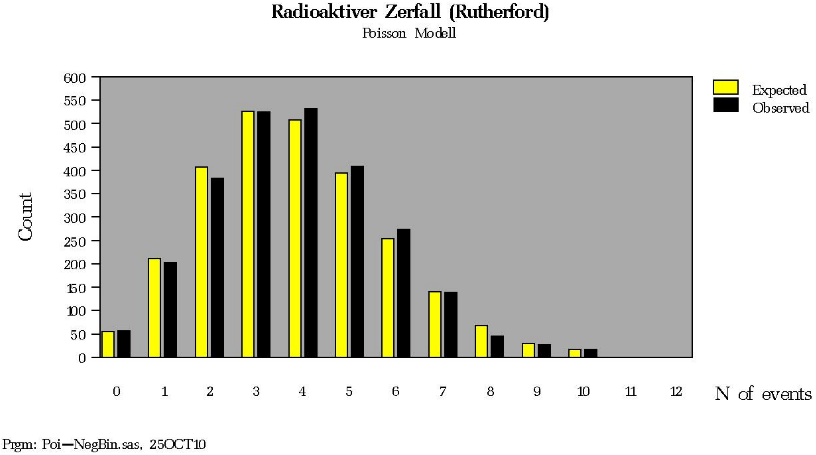
Der Träger von X ist $\mathcal{T} = \mathbb{N}_0$. Erwartungswert und Varianz ergeben sich zu

$$\mathbb{E}(X) = \mathbb{V}(X) = \mu.$$

Beispiel 3.18 (Radioaktiver Zerfall (Rutherford))

2608 Zeitintervalle a 7.5 Sek. mit insgesamt 10086 Ereignissen.

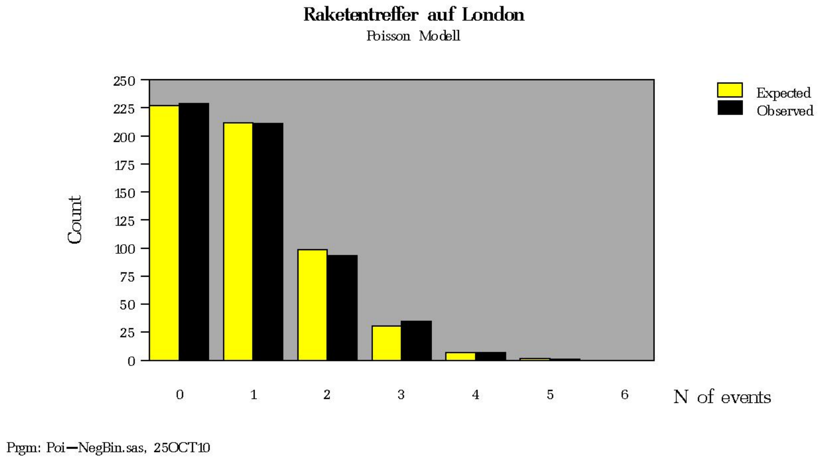
k	N_k	$N \cdot p(k; 3.87)$	k	N_k	$N \cdot p(k; 3.87)$
0	57	54.40	5	408	393.52
1	203	210.52	6	273	253.82
2	383	407.36	7	139	140.33
3	525	525.50	8	45	67.88
4	532	508.42	9	27	29.19
			≥ 10	16	17.08
				2608	2608


Radioaktiver Zerfall (Rutherford)
Poisson Modell
Count
600
550
500
450
400
350
300
250
200
150
100
50
0
0
1
2
3
4
5
6
7
8
9
10
11
12
N of events
Expected
Observed
Prgm: Poi−NegBin.sas, 25OCT10

Beispiel 3.19 (Raketen-Treffer in London)

576 Planquadrate a 1/4 km^2 mit insgesamt 535 Einschlägen.

k	0	1	2	3	4	5+
N_k	229	211	93	35	7	1
$N \cdot p(k; 0.9323)$	226,74	211.39	98.54	30.62	7.14	1.57


Raketentreffer auf London
Poisson Modell
Expected
Observed
Count
250
225
200
175
150
125
100
75
50
25
0
0
1
2
3
4
5
6
N of events
Prgm: Poi—NegBin.sas, 25OCT10

Negative Binomialverteilung (II)

Statt X, die Anzahl der Bernoulli-Experimente (mit Wahrsch.keit π) bis zum $r-$ten "Erfolg" zu zählen (siehe Def. 3.15), bezeichnet man auch die ZVe X', Anzahl der "Misserfolge" bis zum Eintreffen des $r-$ten Erfolgs, als negativ binomialverteilt.

Definition 3.20 (Negative Binomialverteilung II)

X' ist *negativ binomial* verteilt mit Parametern r und μ $(X \sim NB(r, \mu))$, wenn

$$P(X' = k) = \begin{cases} \binom{r+k-1}{k}\left(\frac{r}{r+\mu}\right)^r\left(\frac{\mu}{r+\mu}\right)^k & , \text{ falls } k = 0, 1, 2, \dots, \\ 0 & , \text{ sonst.} \end{cases}$$

Der Träger von X' ist $\mathcal{T} = \{0, 1, 2, \dots\} = \mathbb{N}_0$. Erwartungswert und Varianz ergeben sich zu

$$\mathbb{E}(X') = \mathbb{E}(X) - r = \frac{r(1-\pi)}{\pi} = \mu, \quad \mathbb{V}(X') = \mu + \mu^2/r.$$

- Eine andere Herleitung dieser Verteilung ergibt sich über das *Poisson-Gamma Mischmodell* (s.u.). Das folgende Beispiel kann man als Anwendung sehen.

Beispiel 3.21 (Von Karies befallene Zähne bei Kindern)

Bei 467 österreichischen Kindern im Alter von 5-6 Jahren wurden die Anzahlen gefüllter/kariöser Flächen gezählt (Quelle: http://www.hillel.de/karies/karies.htm).

Die folgenden zwei Grafiken zeigen

(a) Verteilung der Daten und Anpassung unter einem Poisson Modell

(b) Verteilung der Daten und Anpassung unter negativer Binomialverteilung

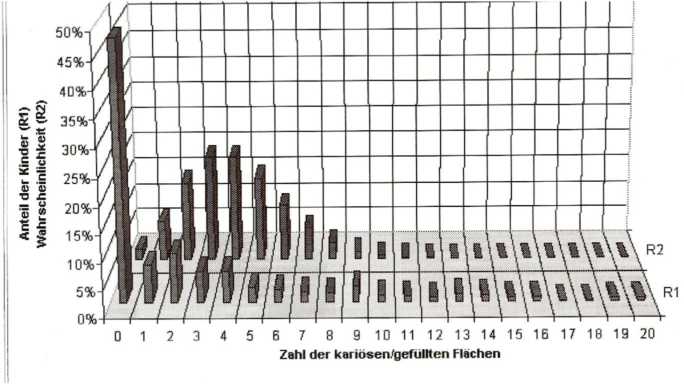


Figure: (a) Poisson Anpassung

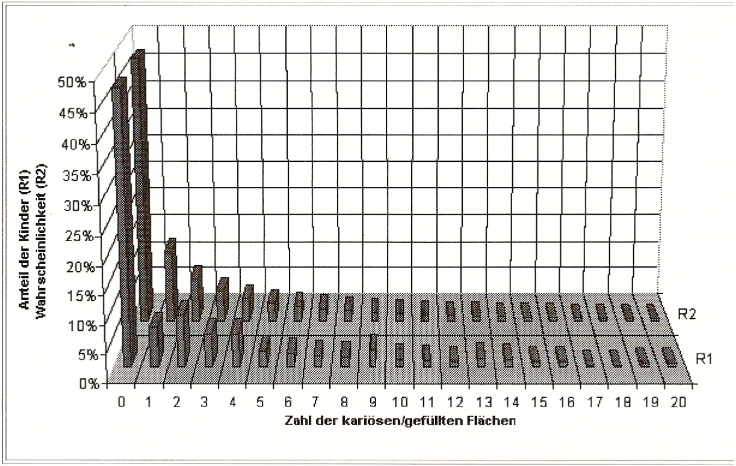


Figure: (b) negative Binomial Anpassung

Multinomialverteilung

Genese: Eine Serie von n unabhängigen Versuchen wird durchgeführt.In jedem Versuch kann eins von k sich jeweils ausschließenden Ereignissen A_i auftreten, wobei $\pi_i = P(A_i)$ über die Serie konstant bleibt, $\sum \pi_i = 1$.

Definition 3.22 (Multinomialverteilung)

Der Vektor $X = (X_1, \ldots, X_k)$ ist *multinomial* verteilt mit Parametern n und $\pi = (\pi_1, \ldots, \pi_k)$ $(X \sim M(n, (\pi_1, \ldots, \pi_k)))$, wenn

$$P(X_1 = x_1, \ldots, X_k = x_k) = n! \prod_{i=1}^{k} \frac{\pi_i^{x_i}}{x_i!}$$

Der Träger von X ist $\mathcal{T} = \{0, 1, \ldots, n\} \times \ldots \times \{0, 1, \ldots, n\}$.

Multinomialverteilung (Forts.)

Erwartungswert und Varianz und Kovarianz der Komponenten von X ergeben sich zu

$$\mathbb{E}(X_i) = n\pi_i,$$

$$Cov(X_i, X_j) = \begin{cases} n\pi_i(1-\pi_i) & \text{für } i = j, \\ -n\pi_i\pi_j & \text{für } i \neq j. \end{cases}$$

Summen spezieller Zufallsvariablen

- Binomialverteilte ZVen:
  - Sind $X \sim B(n, \pi)$, $Y \sim B(m, \pi)$ unabhängig, dann ist $Z = X + Y \sim B(m + n, \pi)$.
  - Unter denselben Voraussetzungen: $P(X = x \mid X + Y = z) \sim H(z, n, m + n)$.
- Poissonverteilte ZVen:
  Sind $X \sim Po(\mu_1)$, $Y \sim Po(\mu_2)$ unabhängig, dann ist $Z = X + Y \sim Po(\mu_1 + \mu_2)$.

(Weitere Relationen siehe Bishop/Holland/Fienberg)