Übungen mit dem Applet „Diskrete Verteilungen“

1 Statistischer Hintergrund

1.1 Diskrete Verteilungen

Ein Messwert kann (zumindest im Prinzip) beliebige Werte auf einer kontinuierlichen Skala annehmen (Applet „Stetige Verteilungen“). Als Ergebnis einer Zählung sind dagegen nur die diskreten Werte 0, 1, 2, 3, … möglich. Zwischenwerte, wie z.B. 1,35 können nicht auftreten. Da bei der Verteilung von Zählergebnissen nur diskrete Werte auftreten, spricht man von „diskreten Verteilungen“.

Entnimmt man aus einer Grundgesamtheit (Population) von N Objekten (z.B. den 49 Zahlen beim Lotto oder einer Lieferung von 10 000 Motoren oder den ca. 60 Mio. Wahlberechtigten in Deutschland), von denen M ein bestimmtes Merkmal tragen (z. B. 6 angekreuzte Zahlen oder 200 defekte Motoren oder 21 Mio. Wähler der Partei xy) **zufällig** eine Stichprobe von n Objekten (z.B. die 6, die bei der nächsten Ziehung gezogen werden, oder die 100 Motoren der Wareneingangsprüfung oder die 2000 bei der Meinungsumfrage Befragten), so kann man zeigen, dass die Wahrscheinlichkeit dafür, dass genau x der n Objekte in der Stichprobe das Merkmal tragen,

$$g(x) = \frac{\binom{N-M}{n-x} \cdot \binom{M}{x}}{\binom{N}{n}}$$ beträgt – die Dichte der **hypergeometrischen Verteilung**.

Ist der Umfang der Stichprobe n<<N (diese Bedingung ist z.B. bei der Wareneingangsprüfung und der Wahlumfrage erfüllt, aber nicht beim Lotto), so erhält man näherungsweise

$$g(x) = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}$$ – die Dichte der **Binomialverteilung**

wobei p = M/N = Anteil der Merkmalsträger in der Grundgesamtheit. Formal erhält man die Binomialverteilung auch, wenn man die Objekte nach der Untersuchung immer wieder zurücklegt – aber das macht natürlich niemand (wer legt bei der Wareneingangsprüfung einen defekten Motor wieder zurück ins Lieferlos?).

Kann das Merkmal an einem Objekt auch **mehrmals** auftreten und ist die Verteilung **zufällig**, so ist die Wahrscheinlichkeit dafür, dass es genau x-mal auftritt

$$g(x) = \frac{\mu^x}{x!} \cdot e^{-\mu}$$ – die Dichte der **Poissonverteilung**

wobei μ = mittlere Anzahl von Merkmalen pro Einheit in der Grundgesamtheit. Beispiele: Staubpartikel/cm^3 Luft, Isolationsfehler/Drahtrolle, Kratzer/Glasplatte. Die Poissonverteilung ergibt sich aus der Binomialverteilung, wenn bei festem Wert $\mu = np$ $n \to \infty$ und $p \to 0$ gehen.

1.2 Verteilungsfunktion und Anwendungen

Die Dichte g(x) der Verteilungsfunktion gibt an, mit welcher Wahrscheinlichkeit genau das Ergebnis x auftritt. Die

$$\text{Verteilungsfunktion } G(x) = g(0) + g(1) + g(2) + ... + g(x) = \sum_{i=0}^{x} g(i)$$

gibt an, mit welcher Wahrscheinlichkeit das Ergebnis 0 oder 1 oder 2 … oder x auftritt, d.h. die Wahrscheinlichkeit, dass ein Ergebnis bis zu x oder höchstens x auftritt. Analog kann man auch andere Fragestellungen beantworten:

- genau x: $g(x)$
- bis zu x, höchstens x: $G(x)$
- weniger als x: $G(x-1)$
- mehr als x: $1 - G(x)$
- mindestens x, x oder mehr: $1 - G(x-1)$
- zwischen x_a und x_b: $G(x_b) - G(x_a)$

Eine andere häufig auftretende Frage ist: In welchem Bereich liegen mindestens 95% oder 99% der Werte? Dies ist der so genannte **zweiseitige 95%-Zufallsstreubereich** (bzw. 99%-). Da die Verteilungen diskret sind, erreicht man die gewünschten 95% nicht genau. Der zweiseitige 95%-Zufallsstreubereich ist der x-Bereich, für den die Wahrscheinlichkeit eines x-Wertes unterhalb des Bereichs höchstens 2,5% beträgt und die Wahrscheinlichkeit eines x-Wertes oberhalb ebenfalls höchstens 2,5% (eine Kompensation wird nicht zugelassen). Analog gibt es einseitig oben oder unten begrenzte Zufallsstreubereiche.

2 Visualisierungen mit dem Applet

Das Applet zeigt die Binomialverteilung ($\boxed{\textbf{B}}$) und die Poissonverteilung ($\boxed{\textbf{P}}$), die Länge der Stäbe ist g(x).

2.1 Form der Verteilungen

Vergleichen Sie die Binomialverteilungen für n=5, p=0,2 mit n=10, p=0,1 und n=100, p=0,01 und mit der Poissonverteilung für μ=1. Wiederholen Sie den Vergleich für n=10, p=0,5 mit n=100, p=0,05 und n=250, p=0,02 und mit der Poissonverteilung für μ=5. Die Form der Binomialverteilung hängt wesentlich von np ab, der Schwerpunkt liegt bei x=np und das Maximum liegt in der Nähe von x=np. Für kleine Werte von p ist die Binomialverteilung sehr ähnlich zur Poissonverteilung mit μ=np.

Vergleichen Sie für den festen Wert n=10 die Binomialverteilung für p=0,1 mit p=0,9 und beobachten Sie, dass die Verteilungen symmetrisch zueinander sind – x=0 wird zu x=10. Dies ist verständlich, da p=0,1 bedeutet, dass 10% der Teile das Merkmal tragen, p=0,9 bedeutet, dass 90% das Merkmal tragen und damit 10% es nicht tragen – Merkmalsträger werden zu Nicht-Merkmalsträgern, x wird zu n−x. Wiederholen Sie den Vergleich für p=0,3 mit p=0,7.

Verändern Sie np bei der Binomialverteilung bzw. μ bei der Poissonverteilung und beobachten Sie:

- Für np<<1 bzw. μ<<1 treten fast nur die Werte x=0 und x=1 auf, g(1)≈np.
- Für np>>1 bzw. μ>>1 wird die Verteilung symmetrisch zu x=np bzw. μ und die Breite der Verteilung nimmt zu (wie $\sqrt{np(1-p)}$ bzw. wie $\sqrt{\mu}$). Die Form ähnelt immer mehr einer Glocke – der Normalverteilung.

2.2 Typische Fragestellungen

Wählen Sie „Wahrscheinlichkeit“ und vergleichen Sie für verschiedene Werte von np bzw. μ die Grafiken für verschiedene Fragestellungen („Art“). Die Summe der grünen Balkenlängen ist jeweils die gesuchte Wahrscheinlichkeit.

g(x) ist die Länge eines Stabes, G(x) ist die Summe der Stablängen von 0 bis einschließlich x.

Durch geeignete Umrechnungen können auch die anderen Fragestellungen beantwortet werden. Überzeugen Sie sich z.B. davon, dass „weniger als 5“ dasselbe bedeutet wie „höchstens 4“, nämlich G(4) und dass „mehr als 4“ und „mindestens 5“ gerade die anderen Werte sind (rot und grün vertauscht, also 1–G(4)).

Vergleichen Sie die Ergebnisse mit dem Applet auch mit Ablesebeispielen mit dem Larson-Nomogramm (Binomialverteilung) bzw. dem Thorndyke-Nomogramm.

2.3 Zufallsstreubereich

Bei der Berechnung des Zufallsstreubereichs wird die Fragestellung umgekehrt – die (Mindest-)Summe der grünen Stablängen wird vorgegeben. Wählen Sie „Zufallsstreubereich“ und geben Sie „zweiseitig“ und 0.95 (=95%) vor. Die Ober- und die Untergrenze werden so gewählt, dass der rote Bereich unten und oben jeweils höchstens 2,5% beträgt (kein Ausgleich wird zugelassen). Da die Hinzunahme eines Stabes den Wert wesentlich ändert, schafft man die angestrebten 2,5% nur selten genau, z.T. ist der Wert deutlich kleiner. Der grüne Bereich ist damit normalerweise deutlich größer als der Zielwert von 0,95 (die tatsächlichen Werte werden angezeigt) – im Gegensatz zu den stetigen Verteilungen.

Schalten Sie zwischen 0.99 und 0.95 hin und her und überzeugen Sie sich, dass der 99%-Zufallsstreubereich breiter ist als der 95%-Zufallsstreubereich.

Bei einseitig oben oder unten abgegrenzten Bereichen liegt der gesamte rote Bereich auf einer Seite der Verteilung.