Berner Fachhochschule

Frühlingssemester 2018
Marx Stampfli, Vidushi Bigler-Maillart
Klassen I3p, I3q

Diskrete Verteilungen mit R-Code

Spezielle diskrete Verteilungen

Wir haben bis jetzt sehr allgemein über diskrete Zufallsvariablen gesprochen. Es gibt jedoch einige spezielle Verteilungen, welche sich in der Anwendung als sehr hilfreich erweisen. Die wichtigsten diskreten Verteilungen lernen wir auf diesem Übungsblatt kennen.

Bevor wir uns den einzelnen Verteilungen widmen, führen wir noch den Begriff der *Verteilungsfunktion* ein. Die Verteilungsfunktion F_X einer beliebigen (diskreten oder stetigen[a]) Zufallsvariable X ist definiert als

$$F_X(x) = P(X \leq x).$$

$F_X(x)$ ist also einfach die Wahrscheinlichkeit, dass X höchstens den Wert x annimmt. Es ist klar, dass eine Verteilungsfunktion immer monoton wachsend mit $\lim_{x\to-\infty} F_X(x) = 0$ und $\lim_{x\to\infty} F_X(x) = 1$ ist.

Für diskrete Zufallsvariablen, wie wir sie bisher betrachtet haben, kann die Verteilungsfunktion offensichtlich als die Summe

$$F_X(x) = \sum_{\text{alle } i \text{ mit } x_i \leq x} P(X = x_i)$$

berechnet werden und ist eine monoton wachsende Treppenfunktion mit Sprüngen der Höhe $P(X = x_i)$ an den Stellen x_i. In Abbildung 1 sehen Sie als Beispiel die Wahrscheinlichkeitsverteilung und die Verteilungsfunktion der Zufallsvariable X, die dem Wurf eines fairen Würfels entspricht.

Gleichverteilung

Nimmt eine diskrete Zufallsvariable X alle möglichen Werte $x_1, x_2, \ldots, x_n$ mit derselben Wahrscheinlichkeit an, so heisst diese gleichverteilt.

Beispiel. Die Augenzahl X beim Wurf eines fairen Würfels ist offensichtlich gleichverteilt. Die Werte $1, \ldots, 6$ werden jeweils mit der Wahrscheinlichkeit $P(X = x_i) = \frac{1}{6}$ (für alle $i = 1, \ldots, 6$) angenommen.

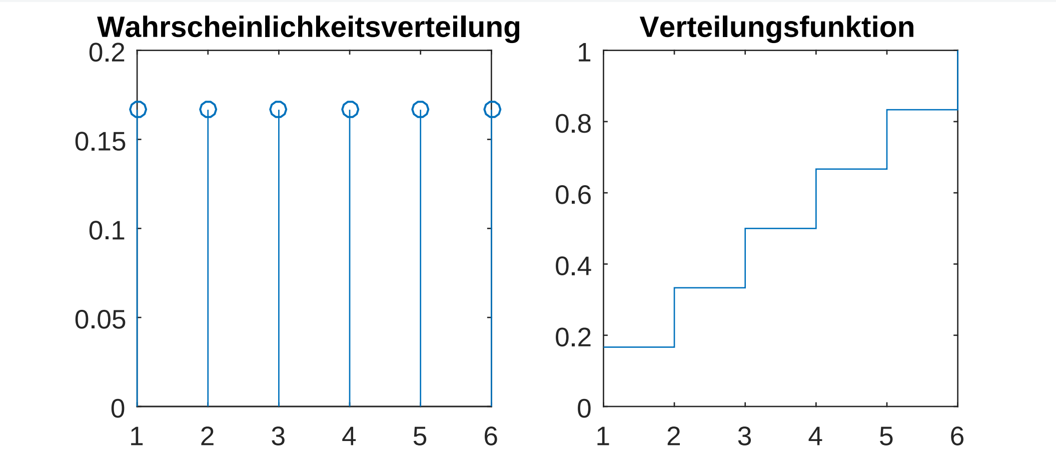

Abbildung 1: Wahrscheinlichkeitsverteilung und Verteilungsfunktion eines fairen Würfels.

Allgemein gilt für die Wahrscheinlichkeitsverteilung einer gleichverteilten Zufallsvariable

$$P(X = x_i) = \frac{1}{n}, \quad i = 1, 2, \dots, n.$$

Für den Erwartungswert und die Varianz bedeutet dies

$$\mathrm{E}(X) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\mathrm{Var}(X) = \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right).$$

In Aufgabe S5.8 haben Sie den Spezialfall $x_1 = 1, x_2 = 2, \dots x_n = n$ bereits betrachtet. Wir erinnern uns: Hier vereinfachen sich Erwartungswert und Varianz zu

$$\mathrm{E}(X) = \frac{n+1}{2}$$

$$\mathrm{Var}(X) = \frac{n^2-1}{12}.$$

Wir haben bereits auf dem letzten Übungsblatt gesehen, wie wir mit R eine gleichverteilte Zufallsvariable erzeugen können, die die Werte $1, 2, \dots, n$ annimmt:

```
>sample(x = 6, size=100, replace=TRUE)
```

z.B. erzeugt einen Vektor mit 100 Zufallszahlen zwischen 1 und 6 (Simulation eines fairen Würfels). Auch gleichverteilte Zufallszahlen, die nicht bei 1 beginnen und nicht Schrittweite 1 haben, lassen sich damit erzeugen:

```
>3*(sample(x = 6, size=100, replace=TRUE)-1)
```

z.B. erzeugt einen Vektor mit 100 gleichverteilten Zufallszahlen aus der Menge {0, 3, 6, 9, 12, 15}.

Leider ist die diskrete Gleichverteilung in R einbisschen stiefmütterlich behandelt worden und es gibt nicht vorprogrammierte Funktionen für die Wahrscheinlichkeitsverteilung und die Verteilungsfunktion im diskreten Fall. Aber diese lassen sich relativ einfach folgendermassen definieren:

```
> dunif_discrete <-function(x, min=0, max=1) {
ifelse(x>=min & x<=max & round(x)==x, 1/(max-min+1), 0)
}
> punif_discrete <-function(q, min=0, max=1) {
ifelse(q<min, 0, ifelse(q>=max, 1, (floor(q)-min+1)/(max-min+1)))
}
```

Beispiel. Damit lassen sich die Wahrscheinlichkeiten, dass ein fairer Würfel *genau* eine 4 bzw. *höchstens* eine 4 würfelt, bestimmen

```
> dunif_discrete(x = 4, min = 1, max = 6)
> punif_discrete(x = 4, min = 1, max = 6)
```

Die Resultate 0.1667 bzw. 0.6667 sind nicht weiter überraschend. Bei komplizierteren Verteilungen werden Sie aber hin und wieder froh sein, Wahrscheinlichkeiten mit Hilfe von R-Funktionen berechnen zu können.

Wahrscheinlichkeitsverteilung und Verteilungsfunktion der Gleichverteilung sind (für den Spezialfall eines fairen Würfels) in Abbildung 1 dargestellt.

Binomialverteilung

Die Binomialverteilung beschäftigt sich mit Ereignissen, bei denen lediglich zwei Ausgänge möglich sind, wie z.B. beim Münzwurf (Kopf oder Zahl). Bereits in Aufgabe S5.5 haben wir „unbewusst“ mit einer binomialverteilten Zufallsvariable gerechnet! Dabei war X die Anzahl „Kopf“ beim dreimaligen Münzwurf ($n = 3$) mit „Erfolgswahrscheinlichkeit“ (d.h. Münze zeigt Kopf) $p = 2/3$.

Wir werden allgemein von *Erfolg* und *Misserfolg* sprechen und verwenden die Abkürzungen 0 und 1:

- Erfolg (1) mit Wahrscheinlichkeit $p \in [0, 1]$
- Misserfolg (0) mit Wahrscheinlichkeit $1 - p \in [0, 1]$

Wir wiederholen unser Zufallsexperiment n-mal und gehen davon aus, dass die einzelnen Wiederholungen unabhängig voneinander sind (!). Eine binomialverteilte Zufallsvariable X beschreibt die Anzahl Erfolge bei n Versuchen. X kann somit die Werte $0, 1, 2, \ldots n$ annehmen.

Beispiel. Zwei Spieler Andreas und Beni spielen Tischtennis. Der bessere Spieler Andreas gewinnt ein Spiel mit Wahrscheinlichkeit 60% (unentschieden sei ausgeschlossen). Sieger des Turniers (3 Spiele) ist der Spieler, der die Mehrzahl der Spiele gewonnen hat. Die Zufallsvariable X sei die Anzahl der von Andreas gewonnenen Spiele. Dann ist X binomialverteilt mit $n = 3$ und $p = 0.6$. Wie gross ist die Wahrscheinlichkeit, dass der schlechtere Spieler (Beni) das Turnier gewinnt? Sie lässt sich berechnen als

$$\begin{aligned} F_X(1) &= P(X \leq 1) = P(X = 0) + P(X = 1) \\ &= \binom{3}{0} p^0 (1-p)^3 + \binom{3}{1} p^1 (1-p)^2 \\ &= 0.4^3 + 3 \cdot 0.6 \cdot 0.4^2 = 0.352 \end{aligned}$$

Es werden also nur etwa 35% der Turniere von Beni gewonnen, obwohl dieser 40% der einzelnen Spiele gewinnt!

Allgemein gilt für die Wahrscheinlichkeitsverteilung einer binomialverteilten Zufallsvariablen X

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k \in \{0, 1, \ldots, n\}.$$

Der Binomialkoeffizient beschreibt hierbei die Anzahl Möglichkeiten, die es gibt, k Erfolge bei gesamthaft n unabhängigen Versuchen zu haben. (Äquivalent hierzu könnte man natürlich die Möglichkeiten der Misserfolge betrachten, da $\binom{n}{k} = \binom{n}{n-k}$.)

Für den Erwartungswert und die Varianz kann man zeigen, dass

$$\begin{aligned} \mathrm{E}(X) &= np \\ \mathrm{Var}(X) &= np(1-p). \end{aligned}$$

Die Aussage für den Erwartungswert ist auch intuitiv klar: Im Schnitt erwarten wir in n Experimenten $n \cdot p$-mal einen Erfolg.

Wahrscheinlichkeitsverteilung und Verteilungsfunktion der Binomialverteilung sehen Sie in der folgenden Abbildung 2.

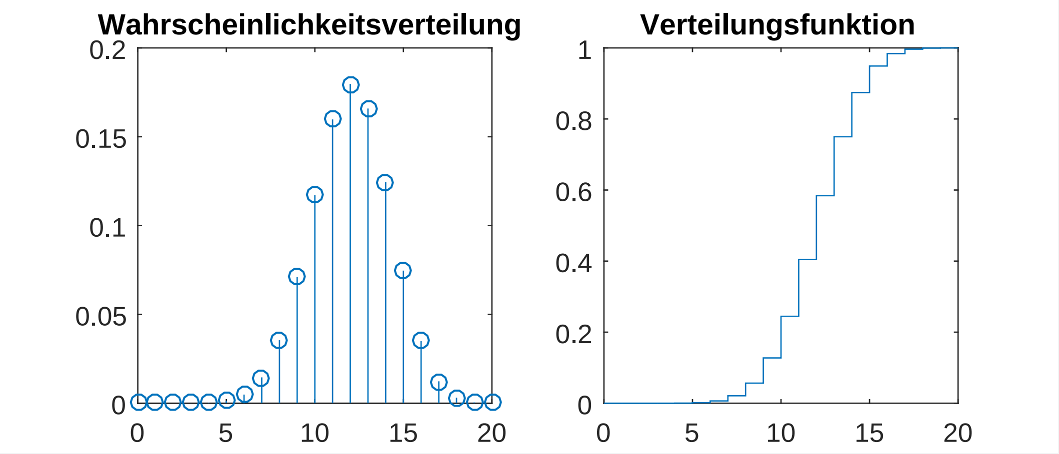

Abbildung 2: Wahrscheinlichkeitsverteilung und Verteilungsfunktion der Binomialverteilung für $n = 20$ und $p = 0.6$.

Beispiel. Die zwei Tischtennisspieler Andreas und Beni aus dem Beispiel oben treffen während eines Jahres an 12 Turnieren aufeinander. Der stärkere Spieler Andreas schreibt sich nach jedem Turnier auf, wie viele Spiele er gegen Beni gewonnen hat. Wie könnte eine solche Sequenz von 12 „Siegzahlen“ aussehen? Wir können so eine Sequenz in R simulieren:

```
> rbinom(n= 1:12, size=3, prob=0.6)
```

Achtung: bitte verwechseln Sie das n in der Funktion `rbinom` nicht mit der n in unseren Gleichungen (das ist nämlich das Input `size`).

Zur Erinnerung: Wenn Sie diese Zeile mehrmals nacheinander in der R-Konsole eingeben, erhalten Sie in jedem Durchlauf eine andere Sequenz von 12 Zahlen. Wenn Sie den Zufallszahlengenerator vor jedem Aufruf von `random` gleich initialisieren (z.B. durch `set.seed(42)`), erhalten Sie jedes Mal dieselbe Sequenz.

Im Beispiel oben haben wir ausgerechnet, wie wahrscheinlich es ist, dass Beni ein Turnier gewinnt. In R können wir diese Wahrscheinlichkeit sehr schnell berechnen:

```
> pbinom(q= 1, size=3, prob=0.6)
```

Und wie wahrscheinlich ist es, dass Andreas *alle* drei Spiele eines Turniers gewinnt?

```
> dbinom(x= 3, size=3, prob=0.6)
```

(Antwort: ca. 21.6%)

Aufgabe S7.1

a) Versuchen Sie, die Bedeutung der Argumente von `rbinom`, `dbinom` und `pbinom` in den Beispielen oben zu „erraten“. Schauen Sie danach in der R-Hilfe nach.

b) Führen Sie den Befehl

```
> rbinom(n= 1:42, size=3, prob=0.6)
```

einige Male nacheinander aus, ohne den Zufallszahlengenerator jedes Mal neu zu initialisieren. Was erwarten Sie im Unterschied zu den Ergebnissen, die Sie mit mehreren Aufrufen von

```
>sample(x = 4, size=42, replace=TRUE)-1
```

erhalten?

Hinweis: Der Befehl `table` stellt Ihnen die Zufallszahlen übersichtlich in tabellarischer Form dar.

Geometrische Verteilung

Wir betrachten wie bei der Binomialverteilung ein Ereignis mit zwei möglichen Ausgängen (Erfolg und Misserfolg). Beschreibt nun die Zufallsvariable X die Anzahl Misserfolge vor dem ersten Erfolg, so ist X geometrisch verteilt.

Beispiel. Insulinpumpen werden im Minutentakt kurzzeitigen Belastungen ausgesetzt. Der Einfachheit halber müssen wir annehmen, dass eine Belastung keine Nachwirkung auf die Funktionstüchtigkeit der Pumpe hat und somit die einzelnen Belastungen unabhängig voneinander sind. Wir interessieren uns nun für die Anzahl Belastungen X, die vergehen, bevor eine Pumpe kaputt ist. Sei $p = 0.001$ die Wahrscheinlichkeit, dass das Gerät eine Belastung nicht übersteht. Dann gilt

$$P(X = k) = p(1-p)^k = 0.001 \cdot 0.999^k$$

für $k \in \{0, 1, 2, ...\}$. Die Wahrscheinlichkeit, dass eine Pumpe frühstens nach 5 Belastungen defekt ist, ist somit $P(X \geq 5) = 1 - P(X < 5) = 1 - F_X(4) = 0.999^5 \approx 0.995$.

In R hätten wir dies wie folgt berechnen können:

```
> 1 - pgeom(q = 0.4, prob = 0.001)
ans =
    0.99501
```

Eine mögliche Sequenz von Anzahl Belastungen bis eine Pumpe defekt ist können wir für eine Charge von 10 Insulinpumpen wie folgt simulieren:

```
>set.seed(42)
>rgeom(n=1:10, prob=0.001)
ans =
         211   42  323  469 1299  111   26  687 1261  602
```

Eine geometrisch verteilte Zufallsvariable X besitzt allgemein folgende Wahrscheinlichkeitsverteilung:

$$P(X = k) = (1-p)^k p, \quad k \in \mathbb{N}_0$$

Mit Hilfe der Partialsumme einer geometrischen Reihe erhält man für die Verteilungsfunktion folgende geschlossene Form:

$$\begin{aligned} F_X(x) = P(X \leq x) &= \sum_{k=0}^{x} P(X = k) = \sum_{k=0}^{x} p(1-p)^k = p \cdot \sum_{k=0}^{x} (1-p)^k = p \cdot \frac{1-(1-p)^{x+1}}{1-(1-p)} \\ &= 1-(1-p)^{x+1} \end{aligned}$$

Dieses Ergebnis macht auch intuitiv Sinn: Die gesuchte Wahrscheinlichkeit entspricht gerade dem Gegenereignis von „die ersten $x + 1$ Versuche sind Misserfolge“.

Für den Erwartungswert und die Varianz kann man folgende Formeln zeigen:

$$\mathrm{E}(X) = \frac{1-p}{p} = \frac{1}{p} - 1$$
$$\mathrm{Var}(X) = \frac{1-p}{p^2} = \frac{1}{p^2} - \frac{1}{p}$$

Wahrscheinlichkeitsverteilung und Verteilungsfunktion der geometrischen Verteilung sehen Sie in Abbildung 3. Beachten Sie, dass die geometrische Verteilung im Unterschied zur Binomialverteilung nicht beschränkt ist – eine geometrisch verteilte Zufallsvariable nimmt beliebig grosse natürliche Zahlen mit positiver (wenn auch möglicherweise sehr kleiner) Wahrscheinlichkeit an!

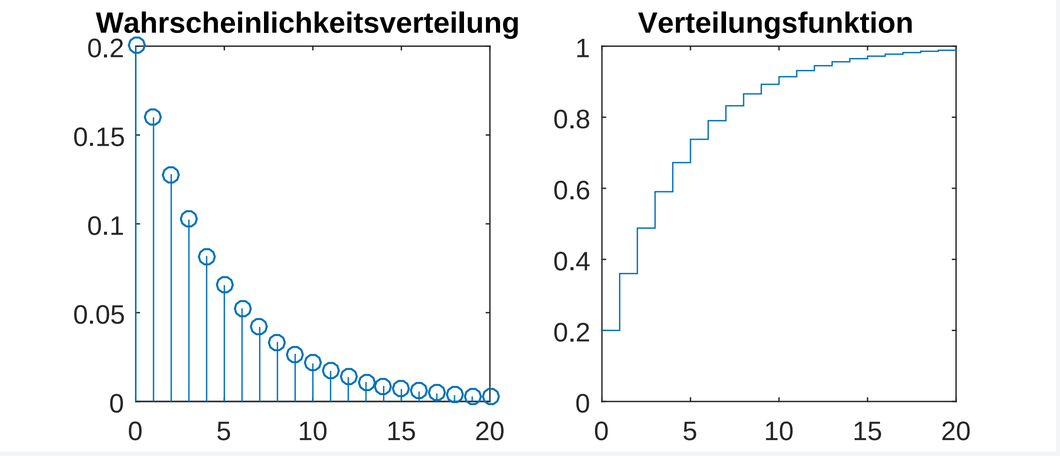

Abbildung 3: Wahrscheinlichkeitsverteilung und Verteilungsfunktion der geometrischen Verteilung für $p = 0.2$.

Bemerkung:
In einigen Lehrbüchern ist die geometrische Verteilung etwas anders definiert, nämlich als die Anzahl benötigter Versuche, bis der erste Erfolg eintritt. Der erfolgreiche Versuch wird hier also mitgezählt! Die so definierte Zufallsvariable X' (= unser X von oben plus 1) hat die Wahrscheinlichkeitsverteilung

$$P(X' = k) = (1-p)^{k-1}p, \quad k \in \mathbb{N}_0.$$

Erwartungswert und Varianz von X' sind

$$\mathrm{E}(X') = \mathrm{E}(X) + 1 = \frac{1}{p}$$
$$\mathrm{Var}(X') = \mathrm{Var}(X) = \frac{1-p}{p^2} = \frac{1}{p^2} - \frac{1}{p}.$$

[a] Eine stetige Zufallsvariable kann *alle* reellen Werte in einem Intervall annehmen. Wir werden uns auf dem nächsten Übungsblatt intensiv mit stetigen Zufallsvariablen beschäftigen.

Aufgabe S7.2 Ein fairer Würfel wird 100-mal geworfen. Geben Sie für die folgenden Wahrscheinlichkeiten jeweils eine Formel an, und berechnen Sie die tatsächlichen Werte anschliessend mit R:

a) Was ist die Wahrscheinlichkeit, dass keine einzige Sechs erscheint?

b) Was ist die Wahrscheinlichkeit, dass genau eine Sechs erscheint?

c) Was ist die Wahrscheinlichkeit, dass höchstens 12 Sechsen erscheinen?

d) Was ist die Wahrscheinlichkeit, dass mindestens 17 Sechsen erscheinen?

Aufgabe S7.3 Eine manipulierte Münze wird 100-mal geworfen. Die Wahrscheinlichkeit für Kopf ist 0.52, und die Wahrscheinlichkeit für Zahl ist demzufolge 0.48. Geben Sie für die folgenden Wahrscheinlichkeiten jeweils eine Formel an, und berechnen Sie die tatsächlichen Werte anschliessend mit R:

a) Was ist die Wahrscheinlichkeit, dass kein einziges Mal Kopf erscheint?

b) Was ist die Wahrscheinlichkeit, dass mindestens einmal Kopf erscheint?

c) Was ist die Wahrscheinlichkeit, dass höchstens 45mal Kopf erscheint?

d) Was ist die Wahrscheinlichkeit, dass bei einer fairen Münze höchstens 45mal Kopf erscheint?

Aufgabe S7.4 Jedes Mitglied eines Komitees mit 9 Mitgliedern kommt mit einer Wahrscheinlichkeit von 0.5 zur Versammlung. Wie gross ist die Wahrscheinlichkeit, dass eine Zweidrittelmehrheit (d.h. sechs oder mehr Personen) anwesend ist?

Sie dürfen davon ausgehen, dass die einzelnen Mitglieder unabhängig voneinander anwesend sind.

Aufgabe S7.5 Jemand wettet, dass er bei 12 Würfen einer Münze genau sechsmal Zahl erzielt. Wie gross ist die Gewinnwahrscheinlichkeit?

Aufgabe S7.6 Ihr Nachbar hat einen Modem-Zugang zum Internet über den Provider U-underline. Wegen Netzüberlastung sind seine Einwahlerfolge, bzw. -misserfolge zufällig und unabhängig voneinander. Die Wahrscheinlichkeit, dass er sich bei einem Anruf erfolgreich einwählt beträgt 0.85.

a) Wie gross ist die Wahrscheinlichkeit, dass er sich erst im dritten Versuch erfolgreich einwählen kann?

b) Wie gross ist die Wahrscheinlichkeit, dass er sich dreimal nacheinander erfolglos versucht einzuwählen?

Aufgabe S7.7 In einer Autowerkstatt ist bekannt, dass 20% der Fahrzeuge, die in die Werkstatt kommen, einen Motorschaden haben. Es bezeichne X die Anzahl der ankommenden Fahrzeuge ohne Motorschaden bis zum ersten Fahrzeug mit Motorschaden.

a) Wie und mit welchen Parametern ist die Zufallsvariable X verteilt?

b) Berechnen und interpretieren Sie den Erwartungswert von X.

c) Berechnen und zeichnen Sie die Wahrscheinlichkeitsverteilung $P(X = x)$ im Intervall $[0, 5]$.

Die Poissonverteilung

Bis jetzt haben wir Verteilungen spezielle Namen gegeben, die bereits in verschiedenen Aufgabenstellungen aufgetaucht sind. Die Poissonverteilung jedoch haben Sie bis anhin noch nicht angetroffen.

Treten zufällige Ereignisse in einem bestimmten Intervall unabhängig voneinander und mit einer *konstanten Rate* λ ein (d.h. erwartete Anzahl Ereignisse pro Zeiteinheit ist konstant), dann ist die Anzahl Ereignisse X in demselben Intervall poissonverteilt mit Parameter λ. Beispiele für solche Ereignisse: Ausfall einer Komponente einer Maschine; Fehler bei der Produktion eines elektronischen Bauteils; Mutationen auf einem Chromosom; Unfälle auf einem bestimmten Strassenabschnitt; etc.

Beispiel. Auf einem Strassenstück ereignen sich im Mittel 0.4 Unfälle pro Monat. Die Poissonverteilung liefert uns Antworten auf Fragen wie „Mit welcher Wahrscheinlichkeit ereignet sich auf diesem Strassenstück mehr als 1 Unfall pro Monat?“.

Die Wahrscheinlichkeitsverteilung einer poissonverteilten Zufallsvariable X ist

$$P(X = k) = \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}, \quad k \in \mathbb{N}_0$$

mit Erwartungswert und Varianz

$$\mathrm{E}(X) = \lambda$$
$$\mathrm{Var}(X) = \lambda.$$

Der Normierungsfaktor $\mathrm{e}^{-\lambda}$ ist plausibel, da wie wir wissen die Reihe $\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = \mathrm{e}^{\lambda}$ ist und wir wissen, dass $\sum_{k=0}^{\infty} P(X = k) \overset{!}{=} 1$ ergeben muss.

Wahrscheinlichkeitsverteilung und Verteilungsfunktion der Poissonverteilung sehen Sie in Abbildung 4. Beachten Sie, dass die Poissonverteilung wie die geometrische Verteilung *nicht* beschränkt ist.

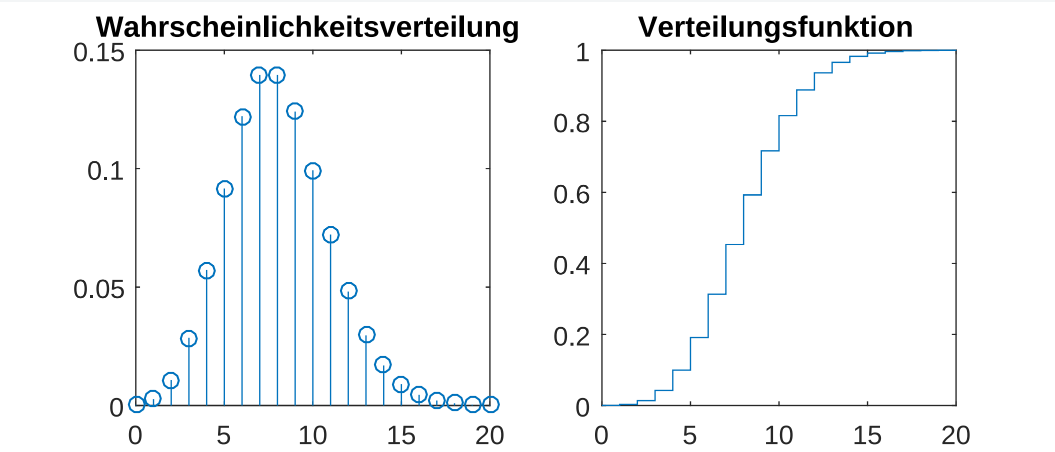

Abbildung 4: Wahrscheinlichkeitsverteilung und Verteilungsfunktion der Poissonverteilung für $\lambda = 5$.

Beispiel. Für unser Beispiel mit den Unfällen bedeutet dies, dass $\lambda = \mathrm{E}(X) = 0.4$, und daraus ergibt sich $P(X > 1) = 1 - P(X \leq 1) = 1 - F_X(1) = 1 - \big(P(X = 0) + P(X = 1)\big) \approx 0.0616$. Mit Hilfe von Matlab erhalten wir diese Zahl viel einfacher:

```
>> 1 - cdf('Poisson', 1, 0.4)
```

Eine mögliche Sequenz monatlicher Unfallzahlen über ein ganzes Jahr können wir wie folgt simulieren:

```
>> random('Poisson', 0.4, [1 12])
```

Aufgabe S7.8 Eine Stadt hat durchschnittlich zwei schwere Unfälle pro Woche. Vereinfacht nehmen wir an, dass die Anzahl schwerer Unfälle poissonverteilt ist.

a) Wie gross ist die Wahrscheinlichkeit für mehr als fünf Unfälle pro Woche?

b) Wie gross ist die Wahrscheinlichkeit für mehr als einen Unfall pro *Tag*?

Überlegen Sie sich zuerst, wie die Zufallsvariable Y=„Anzahl Unfälle pro Tag" verteilt ist.

Aufgabe S7.9 Eine Telefonauskunft wird durchschnittlich 12-mal pro Stunde angewählt. Wir nehmen an, dass die Anzahl Anrufe poissonverteilt ist. Mit welcher Wahrscheinlichkeit erhält die Auskunft

a) in den nächsten 5 Minuten keinen Anruf?

b) innerhalb von 10 Minuten mindestens 5 Anrufe?

c) innerhalb von 20 Minuten höchstens 6 Anrufe?

Aufgabe S7.10 Zeichnen Sie in Matlab die Wahrscheinlichkeitsverteilung und die Verteilungsfunktion der folgenden Verteilungen:

a) Binomialverteilung für $n = 8$ und $p = 0.5$

b) Binomialverteilung für $n = 5$ und $p = 0.05$

c) Poissonverteilung für $\lambda = 0.5$ (wählen Sie für die horizontale Achse den Bereich $[0, 5]$)

d) Poissonverteilung für $\lambda = 5$ (wählen Sie für die horizontale Achse den Bereich $[0, 10]$)

e) Binomialverteilung für $n = 500$ und $p = 0.01$ (wählen Sie für die horizontale Achse den Bereich $[0, 10]$)

Hinweis: Eine Wahrscheinlichkeitsverteilung lässt sich gut mit der Funktion `stem` plotten, eine Verteilungsfunktion mit `stairs`.

Die Poissonverteilung als Approximation der Binomialverteilung

Man kann die Poissonverteilung auch als Approximation der Binomialverteilung betrachten. Hierzu lassen wir für eine binomialverteilten Zufallsvariable X die Anzahl Versuche n gegen unendlich streben und nehmen gleichzeitig an, dass $\mathrm{E}(X) = np = \lambda$ konstant bleibt. Dann ist X im Grenzfall eine poissonverteilte Zufallsvariable mit möglichen Werten $0, 1, 2, \ldots$.

Die Forderung, dass $np = \lambda$ im Grenzprozess konstant bleibt, impliziert, dass die Wahrscheinlichkeit p für einen „Erfolg“ gegen 0 strebt. Als Faustregel gilt, dass man die Poissonverteilung als Approximation der Binomialverteilung für $n \geq 50$ und $p \leq 0.05$ verwenden darf (das ist aber wirklich nur eine Faustregel und keine scharfe Bedingung). Diese Näherung ist praktisch, weil die Verteilungsfunktion einer Binomialverteilung mit steigendem n schnell rechenaufwändig wird.

Beispiel. Die Anzahl Druckfehler X pro Seite in einem Buch kann als poissonverteilte Zufallsvariable betrachtet werden. Wir können annehmen, dass jeder Buchstabe einer Seite eine kleine Wahrscheinlichkeit p besitzt, falsch gesetzt zu werden. Da die Anzahl Buchstaben pro Seite n typischerweise gross ist, gilt die Poissonapproximation mit Parameter $\lambda = np$.

Aufgabe S7.11 Eine Fabrik produziert Schrauben, die mit einer Wahrscheinlichkeit von $p = 0.001$ defekt sind. Wie gross ist die Wahrscheinlichkeit, dass eine Sendung von 500 Schrauben mindestens zwei defekte enthält?

a) Lösen Sie die Aufgabe exakt mit einer Binomialverteilung.

b) Lösen Sie die Aufgabe näherungsweise mit einer Poissonverteilung.

Aufgabe S7.12 Die Wahrscheinlichkeit, dass ein Fluggast, der einen Platz reserviert hat, nicht zum Flug erscheint, beträgt 4%. Die Fluggesellschaft weiss dies und verkauft 75 Tickets für 73 verfügbare Plätze. Wie gross ist die Wahrscheinlichkeit, dass alle Plätze besetzt sind und niemand auf den nächsten Flug warten muss?

a) Lösen Sie die Aufgabe exakt mit einer Binomialverteilung.

b) Lösen Sie die Aufgabe näherungsweise mit einer Poissonverteilung.

Lösungen

Lösung S7.2 Die Zufallsvariable X sei die Anzahl 6-er eines fairen Würfels. Dann ist X binomialverteilt mit $p = \frac{1}{6}$.

a) $P(X = 0) = \left(\frac{5}{6}\right)^{100}$

b) $P(X = 1) = 100 \cdot \left(\frac{1}{6}\right) \cdot \left(\frac{5}{6}\right)^{99}$

c) $P(X \leq 12) = \sum_{k=0}^{12} \binom{100}{k} \cdot \left(\frac{1}{6}\right)^k \cdot \left(\frac{5}{6}\right)^{100-k}$

d) $P(X \geq 17) = 1 - P(X \leq 16) = 1 - \sum_{k=0}^{16} \binom{100}{k} \cdot \left(\frac{1}{6}\right)^k \cdot \left(\frac{5}{6}\right)^{100-k}$

In R berechnen wir die Lösungen wie folgt:

```
> dbinom(x= 0, size=100, prob=1/6)
ans =
   1.207467e-08
> dbinom(x= 1, size=100, prob=1/6)
ans =
   2.414935e-07
> pbinom(q= 12, size=100, prob=1/6)
ans =
   0.1296708
> 1 - pbinom(q= 16, size=100, prob=1/6)
ans =
   0.505841
```

Lösung S7.3 Die Zufallsvariable X sei die Anzahl Kopf beim 100-maligen Münzwurf.

a) $P(X = 0) = (0.48)^{100}$

b) $P(X \geq 1) = 1 - P(X = 0) = 1 - (0.48)^{100}$

c) $P(X \leq 45) = \sum_{k=0}^{45} \binom{100}{k} \cdot (0.52)^k \cdot (0.48)^{100-k}$

d) $\sum_{k=0}^{45} \binom{100}{k} \cdot (0.5)^k \cdot (0.5)^{100-k} = \sum_{k=0}^{45} \binom{100}{k} \cdot (0.5)^{100}$

In R finden wir die folgenden Werte:

```
> dbinom(x= 0, size=100, prob=0.52)
ans =
   1.3308e-32
> 1 - dbinom(x= 0, size=100, prob=0.52)
ans =
   1
> pbinom(q= 45, size=100, prob=0.52)
ans =
    0.0967
> pbinom(q= 45, size=100, prob=0.5)
ans =
    0.1841
```

Lösung S7.4 Die Anzahl der anwesenden Personen ist binomialverteilt mit Parametern 9 und 0.5. Die gesuchte Wahrscheinlichkeit ergibt sich als $\sum_{k=6}^{9}\binom{9}{k}(0.5)^9 = \frac{\sum_{k=6}^{9}\binom{9}{k}}{2^9} \approx 0.2539$.

Lösung S7.5 Die Anzahl Zahl ist binomialverteilt mit Parametern 12 und 1/2. Die gesuchte Wahrscheinlichkeit ist $\binom{12}{6}\left(\frac{1}{2}\right)^{12} \approx 0.2256$.

Lösung S7.6

a) X die Anzahl Misserfolge vor dem ersten Erfolg ist geometrisch verteilt mit $p = 0.85$.

Also $P(X = 2) = 0.85 \cdot (1 - 0.85)^2 \approx 0.0191$.

In R:

```
> dgeom(x=2, prob = 0.85)
ans =
   0.0191
```

b) Hier können wir nicht mehr mit der Formel für die geometrische Verteilung rechnen, da der Nachbar beim 4. Versuch nicht garantiert erfolgreich ist. Die gesuchte Wahrscheinlichkeit berechnet sich jedoch einfach als $(1 - 0.85)^3 \approx 0.0034$.

Lösung S7.7

a) X ist geometrisch verteilt mit $p = 0.2$.

b) $E(X) = \frac{1-0.2}{0.2} = 4$

Im Schnitt werden also 4 Fahrzeuge ohne Motorschaden angeliefert bevor das erste Fahrzeug mit Motorschaden ankommt. Auch intuitiv ist klar, dass für ein $p = 1/5$ im Erwartungswert gerade das fünfte Auto einen Motorschaden hat.

c)

$P(X=0)$	$P(X=1)$	$P(X=2)$	$P(X=3)$	$P(X=4)$	$P(X=5)$
0.20	0.16	0.13	0.10	0.08	0.07

Lösung S7.8

a) Sei X die Anzahl Unfälle pro Woche. X ist poissonverteilt mit $\lambda_X = 2$; die gesuchte Wahrscheinlichkeit ist also $P(X > 5) = 1 - P(X \leq 5) \approx 0.017$.

b) Y ist poissonverteilt mit Parameter $\lambda_Y = \frac{2}{7}$; die gesuchte Wahrscheinlichkeit ist also $P(Y > 1) = 1 - P(Y \leq 1) \approx 0.034$.

Lösung S7.9

a) Die Anzahl Anrufe pro 5 Minuten X ist poissonverteilt mit Parameter $\lambda_X = \frac{12}{60} \cdot 5 = 1$. Also beträgt die gesuchte Wahrscheinlichkeit $P(X = 0) \approx 0.368$.

b) Die Anzahl Anrufe pro 10 Minuten Y ist poissonverteilt mit Parameter $\lambda_Y = 2 \cdot \lambda_X = 2$. Also beträgt die gesuchte Wahrscheinlichkeit $P(Y \geq 5) = 1 - P(X < 5) = 1 - P(X \leq 4) \approx 0.053$. Wir erhalten den Zahlenwert am einfachsten in Matlab:

```
>> 1 - cdf('Poisson', 4, 2)
```

c) Die Anzahl Anrufe pro 20 Minuten Z ist poissonverteilt mit Parameter $\lambda_Z = 4 \cdot \lambda_X = 2 \cdot \lambda_Y = 4$. Also beträgt die gesuchte Wahrscheinlichkeit $P(Z \leq 6) \approx 0.889$. Auch hier bietet sich die Berechnung in Matlab an:

```
>> cdf('Poisson', 6, 4)
```

Lösung S7.11

a) Die Anzahl defekter Schrauben X ist exakt binomialverteilt mit Parametern $n = 500$ und $p = 0.001$. Somit ist die gesuchte Wahrscheinlichkeit

$$P(X \geq 2) = 1 - P(X < 2) = 1 - F_X(1) = 1 - \sum_{k=0}^{1} \binom{500}{k} 0.001^k \cdot 0.999^{500-k} \approx 0.0901$$

b) Die Anzahl defekter Schrauben ist approximativ poissonverteilt mit Parameter $n \cdot p = 500 \cdot 0.001 = 0.5$. Somit beträgt die gesuchte Wahrscheinlichkeit $P(X \geq 2) \approx 1 - \sum_{k=0}^{1} \mathrm{e}^{-0.5} \frac{0.5^k}{k!} \approx 0.0902$.

Lösung S7.12 Sei X die Anzahl Passagiere, die zum Flug erscheinen, $n = 75$ und $p = 0.96$.

a) Ist X binomialverteilt, so ergibt sich $P(X = 73) \approx 0.226$

b) Da die Wahrscheinlichkeit, dass ein Fluggast erscheint sehr gross ist, muss man für eine vernünftige Poissonapproximation das Gegenereignis Y betrachten, welches die Anzahl Passagiere beschreibt, die nicht zum Flug erscheinen. Dann ist Y poissonverteilt mit Parameter $\lambda = 75 \cdot 0.04$ und es folgt $P(Y = 2) \approx 0.224$.