Stochastik

– Aufgaben zur Serie 5 –

Ervin Mazlagić

7. April 2013

Inhaltsverzeichnis

1 Aufgabe 1 **3**

2 Aufgabe 2 **4**

3 Aufgabe 3 **5**

4 Aufgabe 4 **7**

1 Aufgabe 1

Verwende R um folgende Grössen zu berechnen. Es sei $X \sim \text{Poisson}(200)$ die Zufallsvariable, die die Anzahl Unfälle in einem Jahr beschreibt.

(a) Wie gross ist die Wahrscheinlichkeit, dass in einem Jahr genau 200 Unfälle passieren?

(b) Wie gross ist die Wahrscheinlichkeit, dass in einem Jahr höchstens 210 Unfälle passieren?

(c) Wie gross ist die Wahrscheinlichkeit, dass in einem Jahr zwischen 190 und 210 Unfälle passieren (beide Grenzen eingeschlossen)?

a)

Um dies zu berechnen kann in R der Befehl `dpois()` verwendet werden.

```
> dpois(200,200)
```

```
[1] 0.02819773
```

Erläuterung zur Funktion: `dpois(`Erwartungwert, λ`)`

b)

Um dies zu berechnen werden alle Ergebnisse der Wahrscheinlichkeiten addiert für die Werte von 0 bis 210 Unfällen.

```
> sum(dpois(0:210,200))
```

```
[1] 0.772708
```

In R gibt es dafür aber auch einen eigenen Befehl `ppois()`, dieser summiert von 0 bis zur angegebnen Zahl (erster Parameter der Funktion).

```
> ppois(210,200)
```

```
[1] 0.772708
```

c)

Hier wird wie in Aufgabe 1b) die Summe aller Werte im Bereich 190 bis 210 gebildet um die Wahrscheinlichkeit zu berechnen.

```
> sum(dpois(190:210,200))
```

```
[1] 0.5422097
```

Arbeitet man mit `ppois()` muss man die Differenz bilden von der höheren zur tieferen Grenze. Die untere Grenze ist hier nicht 190 sondern eine Zahl Tiefer, sonst wird dieser Wert abgezogen!

```
> ppois(210,200) - ppois(189,200)
```

```
[1] 0.5422097
```

2 Aufgabe 2

Die Zufallsvariable, die die Anzahl eingehender Telefonanrufe in einer Telefonzentrale innerhalb von 10 Minuten beschreibt, nennen wie X. Sie folgt einer Poissonverteilung mit Erwartngswert $\lambda = 2$, d.h. $X \sim$ Poisson(λ).

(a) Wie gross ist die Wahrscheinlichkeit, dass es in einer bestimmten 10-Minuten-Periode keinen einzigen Anruf gibt?

(b) Wie gross ist die Wahrscheinlichkeit, dass es nicht mehr als drei Telefonanrufe in einer bestimmten 10-Minuten-Periode gibt?

(c) Wie gross ist die Wahrscheinlichkeit, dass es mehr als drei Telefonanrufe in einer bestimmten 10-Minuten-Periode gibt?

(d) Angenommen, die Anzahl Anrufe in einer 10-Minuten-Periode ist von der Anzahl Anrufe in einer anderen 10-Minuten-Periode unabhängig. Die Zufallsvariable, die die Anzahl Anrufe in einer Stunde beschreibt bezeichnen wir mit Y . Welcher Verteilung folgt Y ?

a)

Es gilt zunächst mal $\lambda = 2 \rightarrow X \sim$ Pois(λ)

Hier kann wieder wie in Aufgabe 1 die Funktion `dpois()` verwendet werden.

```
> dpois(0,2)
```

```
[1] 0.1353353
```

b)

Hier müssen wieder alle Wert für 0 bis 3 aufsummiert werden

```
> sum(dpois(0:3,2))
```

```
[1] 0.8571235
```

oder man verwendet die Implementierung von R

```
> ppois(3,2)
```

```
[1] 0.8571235
```

c)

Die Wahrscheinlichkeit, dass es in einer bestimmten 10-Minuten-Periode mehr als drei Anrufe gibt, ist doch genau der Wert den man erhält, wenn man die maximale Wahrscheinlichkeit ($\Omega = 1$) nimmt und den Wert für die Wahrscheinlichkeit von maximal 3 Anrufen subtrahiert.

```
> 1 - sum(dpois(0:3,2))
```

```
[1] 0.1428765
```

oder mit der vorbereiteten Funtkion

```
> 1 - ppois(3,2)
```

```
[1] 0.1428765
```

d)

Da die Ereignisse unabhängig sind, folgt Y einer Poissonverteilung mit einem $\lambda = 6 \cdot 2$ da es in einer Stunde 6 10-Minuten-Abschnitte gibt und zu jedem dieser Abschnitte das "Ursprungs-λ" gilt.

```
> Wahrscheinlichkeit <- dpois(0:40,12)
> plot(Wahrscheinlichkeit, type='l')
```

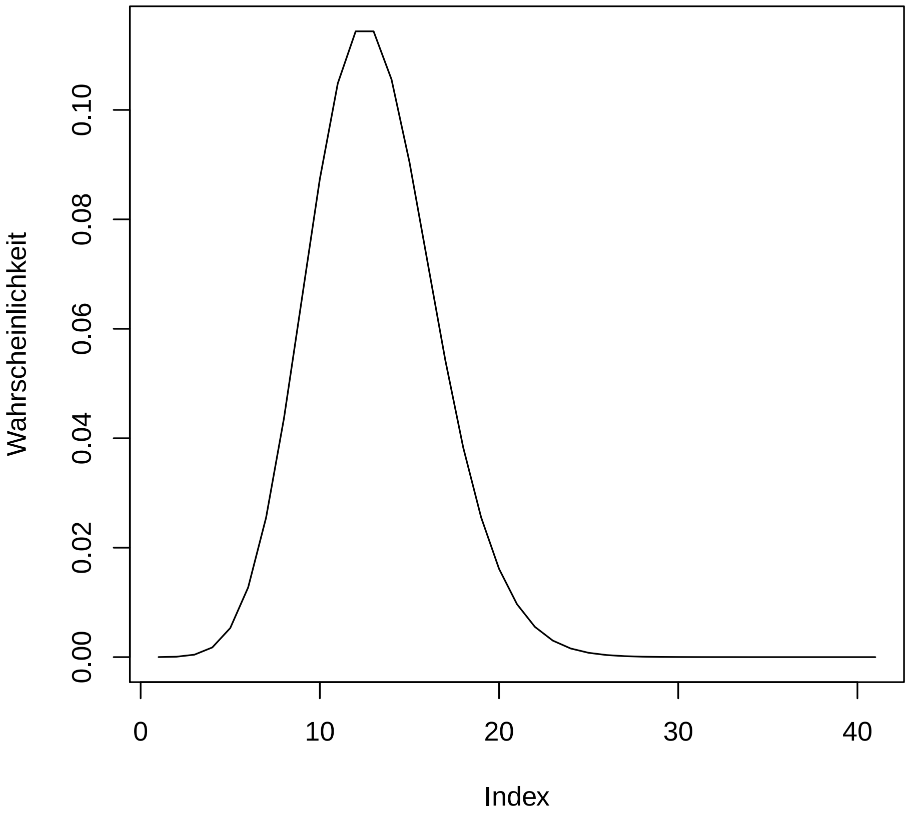

Abbildung 1: Poissonverteilung von Y (Aufgabe 2-d)

3 Aufgabe 3

In der Vorlesung haben wir gesehen, wie man die Erfolgswahrscheinlichkeit π einer Binomialverteilung mit der Maximum-Likelihood-Methode schätzen kann, wenn man die Anzahl Versuche und die Anzahl Gewinne kennt. In dieser Aufgabe kombinieren wir mehrere solcher Beobachtungen zu einer Schätzung. Angenommen Sie gehen über den Jahrmarkt und kaufen bei einer Losbude 30 Lose. Unter den 30 Losen sind 2 Gewinne. Am nächsten Tag erzählt Ihnen Ihr Studienkollege, dass er am Vorabend bei der gleichen Losbude 50 Lose gekauft hat und darunter 4 Gewinne hatte. Wie kombinieren Sie die beiden Ergebnisse, um mit der Maximum-Likelihood-Methode eine möglichst gute Schätzung der Erfolgswahrscheinlichkeit zu erhalten?

(a) Sei X_1 die Zufallsvariable, die die Anzahl Gewinne unter 30 Losen beschreibt ("Ihre Gewinne"). Wenn wir annehmen, dass jedes Los unabhängig von jedem anderen Los ein Gewinn oder eine Niete ist, dann folgt X einer Binomialverteilung mit $n_1 = 30$ und unbekanntem Erfolgsparameter π. Abgekürzt schreiben wir: $X_1 \sim Bin(n_1, \pi)$. Analog sei X_2 die Zufallsvariable, die die Gewinne Ihres Kollegen beschreibt: $X_2 \sim Bin(n_2, \pi)$ mit $n_2 = 50$ und dem gleichen Wert für die Erfolgswahrscheinlichkeit wie bei X_1. Angenommen, die Anzahl Gewinne, die Sie gezogen haben, ist unabhängig von der Anzahl Gewinne, die Ihr Kollege gezogen hat. Wie lässt sich dann $P(X_1 = x_1 \cap X_2 = x_2)$ schreiben?

(b) Wie lässt sich $log(P(X_1 = x_1 \cap X_2 = x_2))$ schreiben? Versuchen Die diesen Term in eine Summe mit mehreren Termen umzuschreiben. Welche Terme hängen von π ab und welche nicht?

(c) Der Maximum-Likelihood-Schätzer für π ist derjenige Zahlenwert, der, wenn man ihn anstelle von π einsetzt, den grösstmöglichen Wert für $log(P(X_1 = x_1 \cap X_2 = x_2))$ (oder $P(X_1 = x_1 \cap X_2 = x_2)$ das ist egal, weil die Funktion log monoton ist) liefert. Finden Sie duch Ableiten und gleich Null setzen den Wert von π in Abhängigkeit von n_1, n_2, x_1 und x_2 der $log(P(X_1 = x_1 \cap X_2 = x_2))$ maximiert.

Allgemeines

Was wissen wir? Nun, wir wissen, dass wir eine Binominalverteilung haben und nehmen mal vorerst die Schätzung mittels der "Momentenmethode" an.

$$X_1 \sim \text{Bin}(n_1 = 30, \widehat{\pi_1}) \text{ wobei } \widehat{\pi_1} = \tfrac{2}{30}$$

Zudem haben wir eine weitere Binominalveteilung durch unseren Kollegen mit den vier Gewinnen.

$$X_2 \sim \text{Bin}(n_2 = 50, \widehat{\pi_2}) \text{ wobei } \widehat{\pi_2} = \tfrac{4}{50}$$

Mit der sog. "Maximum-Likelihood" Methode ist die Abschätzung der Variable $\widehat{\pi}$ nicht ganz so einfach. Bei dieser Methode geht es darum eine (eigentlich beliebige) Anzhal von möglichen Werten von $\widehat{\pi}$ zu verlgeichen und den "passendsten" Wert zu finden. Um dies für alle möglichen Werte zu tun kann der folgende Ausdruck aufgestellt werden.

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

n steht hier für die Anzahl gekaufter Loose und x für die Gewinne. Um nun auf π zu gelangen, muss die Ableitung dieses Ausdrucks zu Null gleichgesetzt werden und natürlich nach π aufgelöst werden. Falls der Ausdruck kompliziert und mühsam zum Ableiten ist, kann versucht werden über eine *Trick* zur Lösung zu gelangen. Der Trick basiert darauf, dass jedes Extremum der Funktion $f(x)$ auch ein Extremum der Fuktion $\log(f(x))$ ist.

$$\log(P[X = x]) = \log\left(\binom{n}{x} \cdot \pi^x \cdot (1 - n)^{n-x}\right)$$

Der eignetliche Trick besteht darin, aus den Faktoren Summen zu bilden. Diese können dann einzeln und unabhängig abgeleitet werden. Das ist natürlich angenehmer zum Ableiten.

$$0 \overset{!}{=} \frac{d}{d\pi} \log\left(\binom{n}{x}\right) + x \cdot \log(\pi) + (n - x) \cdot \log(1 - \pi)$$

Wenn wir dies nun auflösen nach π erhalten wir

$$\pi = \frac{x}{n}$$

Dies stellt einen einfachen Fall dar in welchem das Ergebnis für π mit dem übereinstimmt, welches man für die Momnetenmethode erhalten würde.

a)

Die Anzahl Gewinne ist unabhängig und wird beschrieben durch

$$P(X_1 = x_1 \cap X_2 = x_2)$$

was wie immer durch die Multiplikation der beiden Wahrscheinlichkeiten entsteht.

b)

$$P(X_1 = x_1 \cap X_2 = x_2) = P(X_1 = x_1) \cdot P(X_2 = x_2)$$
$$\log(P(X_1 = x_1 \cap X_2 = x_2)) = \log(P(X_1 = x_1) \cdot P(X_2 = x_2))$$
$$\log(P(X_1 = x_1 \cap X_2 = x_2)) = \log(P(X_1 = x_1)) + \log(P(X_2 = x_2))$$

$$= \log\left(\binom{n_1}{x_1} \cdot \pi^{x_1} \cdot (1-\pi)^{n_1-x_1}\right) + \log\left(\binom{n_2}{x_2} \cdot \pi^{x_2} \cdot (1-\pi)^{n_2-x_2}\right)$$

$$= \log\left(\binom{n_1}{x_1}\right) + x_1 \cdot log(\pi) + (n_1 - x_1) \cdot \log(1-\pi)$$
$$+ \log\left(\binom{n_2}{x_2}\right) + x_2 \cdot log(\pi) + (n_2 - x_2) \cdot \log(1-\pi)$$

Nun kann man sehen, dass π nicht in allen Termen der Summe vorkommnt.

c)

Der ermittelte Ausdruk bei 2-b) kann nun abgeleitet, gleich Null gesetzt und nach π aufgelöst werden.

$$0 \stackrel{!}{=} 0 + \frac{x_1}{\pi} + \frac{n_1 - x_1}{\pi - 1} + 0 + \frac{x_2}{\pi} + \frac{n_2 - x_2}{\pi - 1}$$
$$0 \stackrel{!}{=} \frac{x_1 + x_2}{\pi} + \frac{(n_1 - x_1) + (n_2 - x_2)}{\pi - 1}$$

Habe gerade keine Ahnung wie weiter rechnen ...

4 Aufgabe 4

Das Pharmaunternehmen Life Co. hat ein neues Medikament zur Bekämpfung von ADHS entwickelt. Um die Wirksamkeit festzustellen wurde das Medikament mit n = 10 Patienten getestet. Die derzeitige Standardmethode zeigt bei 30% der behandelten Patienten eine Wirkung.

(a) Angenommen das neue Medikament ist genauso wirksam wie die Standardmethode, wie gross ist die Wahrscheinlichkeit, dass die Behandlung bei genau 2 Patienten eine Wirkung zeigt? Wie gross ist die Wahrscheinlichkeit, dass sie bei höchstens 2 Patienten eine Wirkung zeigt?

(b) Die Behandlung mit dem neuen Medikament war bei 4 Patienten erfolgreich. Führen Sie einen einseitigen Hypothesentest durch um festzustellen ob das neue Medikament wirksamer ist als die Standardmethode (bei einem Signifikanzniveau von 5%). Geben Sie explizit alle Schritte an.

(c) Wie ist die Macht eines Hypothesentests definiert? Geben Sie die Macht an für den Test $H_0 : \pi = 0.3$ vs. $H_A : \pi = 0.6$ (π ist die Wirksamkeit).

a)

Die Wahrscheinlichkeit, dass das Medikament bei 2 behandelten Patienten eine Wirkung zeigt liegt bei

```
> dbinom(2,size=10,0.3)
```

```
[1] 0.2334744
```

Die Wahrscheinlichkeit, dass das Medikament bei höchstens 2 behandelten Patienten eine Wirkung zeigt liegt bei

```
> sum(dbinom(0:2,size=10,0.3))
```

```
[1] 0.3827828
```

Die kann auch mit der Funtkion `pbinom` gerechnet werden

```
> pbinom(2,size=10,0.3)
```

```
[1] 0.3827828
```

b)

Um einen Hypothesentest durchzuführen muss zuerst ein Modell erstellt werden.

$X =$ Anzahl erfolgreich behandelte Patienten

$X \sim Bin(10, \pi)$

Weiter muss die Null-Hypothese und die Alternativ-Hypothese definiert werden.

$H_0 : \pi_0 = 0.3$ Neues Medikament hat die selbe Wirkung

$H_A : \pi_A > \pi_0 = 0.3$ Neues Medikament hat höhere Wirkung

Als nächstes wird die Teststatistik definiert.

$$T : P(T = t|H_0) = (\binom{10}{t}) \cdot 0.3^t \cdot 0.7^1 0 - t$$

Weiter muss das Signifikanzniveau bestimmt werden (aus der Aufgabenstellung geht hervor dass dieses 5% beträgt).

$\alpha = 0.05$

Nun fehlt noch der sog. "Verwerfungsbereich". Dieser Bereich gibt die Werte wieder in welchem die Hypothese, wie der Name schon sagt, verworfen wird. Wir suchen also nach dem Wert, bei welchem die in der Hypothese geltende Wahrscheinlichkeit nicht unterschritten wird. Erstellt man alle Werte im geltenden Bereich d.h. vom 0 bis 10 so erhält man folgende Werte.

```
> 1-pbinom(0:10,size=10,0.3)
```

```
 [1] 0.9717524751 0.8506916541 0.6172172136 0.3503892816 0.1502683326
 [6] 0.0473489874 0.0105920784 0.0015903864 0.0001436859 0.0000059049
[11] 0.0000000000
```

Hier kann man nun erkennen, dass ab dem 6. Element (da von 1 aus Nummeriert wird ist dies das 5. Element) der Wert unter den spezifizierten 5% liegt. Somit ist unser Verwerfungsbereich definiert als

$K = [6, 7, 8, 9, 10]$

Nun muss das Ergebnis interpretiert bzw. ausgewertet werden: Liegt der beobachtete Wert im Verwerfungsbereich? Nein das tut er nicht. Was heisst das nun für unseren Test? Nun ja dies sagt uns, dass die Nullhypothese noch gilt für diese Beobachtung und diese somit nicht verworfen wird.

c)

Die Macht eines statistischen Tests ist die Wahrscheinlichkeit, dass die Nullhypothese verworfen wird, wenn die Alternative stimmt, d.h. $P(T \in K|H_A)$ oder ander formuliert: Macht $= 1 - P$(Fehler 2. Art).

Um die Macht für den den Test aus der Aufgabenstellung zu errechnen kann R benutzt werden.

```
> pbinom(10,size=10,prob=0.6)-pbinom(5,size=10,prob=0.6)

[1] 0.6331033

> plot(dbinom(0:10,size=10,prob=0.3),type='l',ylab="P()")
> lines(dbinom(0:10,size=10,prob=0.6),type='l')
```