Allgemeine Fragen

Ich verstehe noch nicht so ganz, wann ich in meinen Rechnungen multipliziere oder addiere. Ich weiß, dass ich unabhängige Experimente koppeln (also multiplizieren) muss. Aber laut dem 3. Axiom von Kolmogorov müssen manche $P(A)$ auch addiert werden. Könnten wir diesen Unterschied jeweils noch etwas klarer machen?

Kopplung unabhängiger Ereignisse:

Unter der Annahme stochastischer Unabhängigkeit gilt:

$P(A \cap B) = P(A) * P(B)$

Das heißt, die Wahrscheinlichkeit dafür, dass Ereignis A *und* Ereignis B eintritt, lässt sich durch die Multiplikation der beiden Einzelwahrscheinlichkeiten bestimmen (aber eben nur unter Annahme von Unabhängigkeit).

3. Axiom vom Kolmogorov:

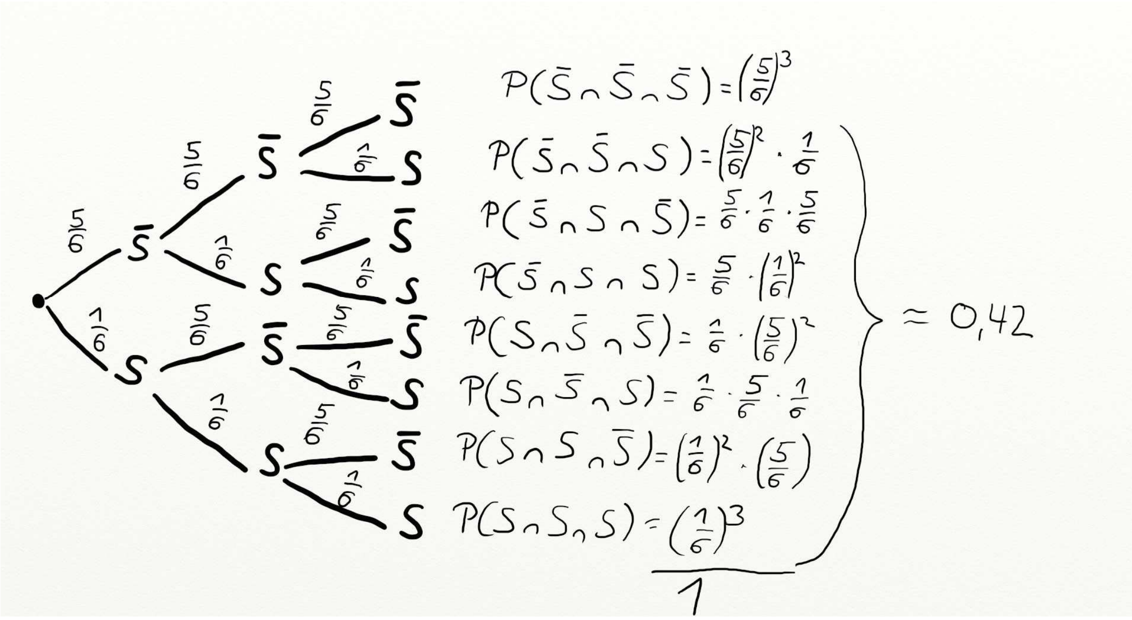
Falls $A \cap B = \emptyset$; dann gilt $P(A \cup B) = P(A) + P(B)$

Sind die Mengen A und B also disjunkt (d. h. sie enthalten keine gemeinsamen Elemente), entspricht die Wahrscheinlichkeit dafür, dass *mindestens eines* der beiden Ereignisse eintritt, der Summe der beiden Einzelwahrscheinlichkeiten.

Beispiel:

Wir spielen Mensch-ärgere-dich-nicht und interessieren uns für die Wahrscheinlichkeit, innerhalb von drei Würfen mindestens eine sechs zu würfeln.

Zum besseren Verständnis erstellen wir zunächst ein Baumdiagramm.



Im Diagramm wird deutlich, dass mehrere Wege unsere Bedingung mindestens einmal eine Sechs zu würfeln, erfüllen (genau genommen alle bis auf einer – dreimal keine Sechs zu würfeln).

Zunächst bestimmen wir die Eintrittswahrscheinlichkeit für jeden Weg. Das ist die Kopplung unabhängiger Ereignisse.

Anschließend addieren wir die Wahrscheinlichkeiten der Wege, die unsere Bedingung erfüllen, um unsere ursprüngliche Frage beantworten zu können.

Wie berechne ich den Bias in der Varianz für einen nicht erwartungstreuen Schätzer. Wie funktioniert das mit MSE?

<u>MSE</u>

Betrachtet wird die Schätzfunktion T für den Parameter ϑ.

Die mittlere quadratische Abweichung der Schätzfunktion ist wie folgt definiert:

$MSE_\vartheta(T) := E_\vartheta(T - \vartheta)^2 = Var_\vartheta(T) + (Bias_\vartheta(T))^2$, mit

$Bias_\vartheta(T) = E_\vartheta(T) - \vartheta$. (Man subtrahiert also den zu schätzenden Parameter vom ermittelten Erwartungswert der Schätzfunktion.)

Ist T erwartungstreu, gilt per Definition $Bias_\vartheta(T) = 0$ und damit $MSE_\vartheta(T) = Var_\vartheta(T)$.

Wird nur dann standardisiert wenn π und μ nicht 1 und 0 sind?

Achtung, zwei unterschiedliche „Arten“ der Standardisierung!

- π nicht 1 ist ja eigentlich fast immer der Fall. Ziel der Standardisierung bei der Binomialverteilung: Approximation der Binomialverteilung durch die Normalverteilung (-> W'keiten lassen sich in Tabelle ablesen)

- Standardisierung von beliebigen Normalverteilungen (also μ nicht 0 und σ^2 nicht 1):

  Die Standardisierung stellt eine unkomplizierte Möglichkeit dar, um die Wahrscheinlichkeiten einer beliebigen Normalverteilung zu ermitteln (da die Werte tabelliert sind).

  Gilt $X \sim \mathcal{N}(\mu, \sigma^2)$, so ist die zugehörige standardisierte Zufallsvariable $Z = \frac{X-\mu}{\sigma}$ standardnormalverteilt, mit $Z \sim \mathcal{N}(0, 1)$. Eine normalverteilte Zufallsvariable, für die gilt, dass $\mu = 0$ und $\sigma^2 = 1$ ist also bereits standardnormalverteilt und muss nicht mehr umgerechnet werden.

Was ist der Unterschied zwischen S^2 und σ^2 ?

σ^2 steht für die **Varianz** eines Merkmals X.

Da diese nicht immer bekannt ist, muss sie unter Umständen durch die **Stichprobenvarianz** S^2 geschätzt werden.

Wie gehe ich beim Ableiten mit dem Summenzeichen um? Ich weiß nicht, wie ich es verändern soll oder ob überhaupt.

Summenregel: $(g \pm h)' = g' \pm h'$

für endliche Summen gilt damit:

$(\sum_{i=1}^{n} f_i(x))' = \sum_{i=1}^{n} f_i'(x)$
Das Summenzeichen bleibt also durch die Ableitung unverändert.

Warum nimmt man die log-likelihood Berechnung her?

Man verwendet zur Bestimmung des Maximum-Likelihood-Schätzers üblicherweise die log-Likelihood, da hierbei das Produkt in eine Summe umgewandelt wird und man damit die Produktregel beim Differenzieren umgeht. Das Summenzeichen kann für die Ableitung "ignoriert" werden (s. o.).
Das Maximum bleibt unverändert, da es sich beim Logarithmus um eine monotone Transformation handelt.

Was sind posteriori und priori odds und Wahrscheinlichkeiten?

Denken Sie an das Diagnoseverfahren für eine Krankheit, von der man weiß, dass ein bestimmter Prozentsatz der Bevölkerung (bspw. 2%) betroffen ist. Weiterhin ist bekannt, dass der Test mit einer bestimmten Wahrscheinlichkeit (bspw. 99%) positiv ausfällt, wenn der Patient erkrankt ist, aber auch bei 3% der nicht-Erkrankten anschlägt.
Die a priori Wahrscheinlichkeit beschreibt die Wahrscheinlichkeit, dass ein Patient an dieser Krankheit leidet, bevor die Ergebnisse des Diagnoseverfahrens bekannt sind. Diese entspricht in diesem Fall 2%
Nach Bekanntwerden der Testergebnisse, ändert sich diese Wahrscheinlichkeit abhängig vom Ergebnis des Tests. Das ist die posteriori Wahrscheinlichkeit.

Erklärung von p-Wert

Der p-Wert hilft zu entscheiden, ob eine Stichprobe gegen die aufgestellte Nullhypothese spricht. Er entspricht der Wahrscheinlichkeit, einen ebenso oder noch extremeren Wert (der Teststatistik, die aus der Stichprobe berechnet wird) zu beobachten, wenn die Nullhypothese korrekt ist. Unterschreitet diese Wahrscheinlichkeit eine bestimmte, zuvor festgelegte Grenze, spricht man davon, dass die Nullhypothese abgelehnt werden kann.

Chi-Quadrat-Unabhängigkeitstest: Ich hatte mir aufgeschrieben, dass dieser Test nur bei nominalskalierten Merkmalen durchgeführt wird. Stimmt das?

Grundsätzlich entscheidet das Merkmal mit dem „schwächsten" Skalenniveau über die möglichen Tests. Der χ^2-Unabhängigkeitstest wird bei diskreten bzw. diskretisierten Merkmalen verwendet. Diese müssen nicht unbedingt nominalskaliert sein, die Information einer „höheren" Skalierung allerdings wird von diesem Test nicht ausgenutzt (siehe Statistik I!).

Wenn ich ceteris paribus argumentiere, dass bei wachsenden n, die Wsk. größer wird, weil mehr Wsk.Masse abgebildet wird und die Varianz und das Intervall kleiner werden, gehe ich dann von dem schwachen Gesetz der großen Zahlen aus?

Warte auf Antwort wegen Rückfrage.

Fragen zu den Übungen

Blatt 2

Aufgabe 2a): Warum ist die Aufgabe mit den Hotelgästen mit Reihenfolge? Ist es, weil die Reisenden wahrscheinlich nicht unabhängig sind (zB Singles) sondern auch meistens Paar und Familien, die eventuell zum selben Zimmer (Betten nebeneinander) zugeteilt werden müssen?

Mit Reihenfolge, weil die Gäste (gedacht) nacheinander ankommen. Spezialsituationen wie Paare oder Familien wurden nicht berücksichtigt.

Blatt 4

Aufgabe 2 c): Warum wird hier die Wahrscheinlichkeit nur über zwei Zustände bzw. Schritte errechnet und nicht zum Beispiel auch über drei?

Weil nach einer zwei-Schritt-Wahrscheinlichkeit („in zwei Schritten“) gefragt wurde.

Aufgabe 2c): Hängen in diesem Fall alle bedingten Wahrscheinlichkeit nur vom unmittelbar vorherigem Zustand ab und wenn ja warum?

Ja. Weil es ein Markovmodell ist.

Aufgabe 3 d): Woher weiß ich, dass ich hier den Satz von Bayes anwenden muss? bzw. woher weiß ich das im Allgemeinen?

Wenn man sich aufschreibt, was gegeben ist und was gesucht ist, dann liegt genau die Situation für die Anwendung des Satzes von Bayes vor. Alternative Lösungen sind über den Wahrscheinlichkeitsbaum möglich, ist aber rechentechnisch meist aufwändiger.

Aufgabe 3 e): Wie komme ich auf den Nenner im Bruch?

Allgemein: Satz von der totalen Wahrscheinlichkeit. Hier: $P(-)$ ist das Gegenereignis von $P(+)$.

Blatt 5

Aufgabe 1a): In der Übung haben wir eine Lösung aufgeschrieben mit dem Binomialkoeffizienten, die mittlerweile nicht mehr für mich nachvollziehbar ist bzw. ich auch mit Hilfe der Formelsammlung nicht auf unseren Rechenweg komme. Wir hatten uns (für die Wahrscheinlichkeit, dass die 4 “Großen” sich nicht begegnen) folgendes aufgeschrieben: (4 über 1) * (4 über 1) geteilt durch (8 über 2). Dies gilt für die Wahrscheinlichkeit beim ersten Zug zwei Mannschaften zu ziehen, wobei eine der “Großen” und eine der “Kleinen” dabei ist.

Dem ganzen liegt eine Laplace-Wahrscheinlichkeit zugrunde: Anzahl günstige/ Alle möglichen. Und diese Anzahlen werden mit Hilfe kombinatorischer Überlegungen getätigt (Vorstellung Urnenmodell: vier rote, vier blaue Kugeln. Wie groß ist die Wahrscheinlichkeit, genau eine rote und eine blaue Kugel zu ziehen)? Hier mit der Formel 'Ziehen ohne Zurücklegen ohne Reihenfolge', die bei $n = 1$ identisch ist zu 'Ziehen ohne Zurücklegen mit Reihenfolge' (In der Formelsammlung unter 1.2.2).
Ein alternativer Rechenweg ist in dem verlinkten Artikel dargestellt.

Aufgabe 1 b): Könnten Sie die Formeln von den a priori und posteriori nochmal zusammenfassen und erklären wo die Zahl '35' in der Lösung dauernd herkommt ?
Die Formeln wurden aus dem verlinkten Artikel übernommen. In den Formel mit 'a priori' steckt nur das Expertenwissen drin, also die Vermutung über die Manipulation (noch bevor überhaupt irgendeine Auslosung stattfand). In den Formeln mit 'a posteriori' wurde dann die Vermutung mit Hilfe der Auslosungsergebnisse (also Daten) aktualsisiert/aufdatiert.

Die Zahl '35' kommt aus der Wahrscheinlichkeit dafür, dass die großen vier Mannschaften nicht aufeinander treffen.

Blatt 6

Aufgabe 1: ich weiß gar nicht wie ich anfangen soll. eine sehr schwere Aufgabe!
So ganz von vorne können wir bei unseren Antworten auch nicht anfangen. Es ist oft hilfreich sich aufzuschreiben, was gegeben und was gesucht ist und dann eine Lösung anzusetzen und diese Schritt für Schritt zu berechnen mit den Rechenregeln, die Sie

kennengelernt haben, und den Informationen aus der Angabe.

Aufgabe 3a): Die Geradengleichung für den Fall $x \in (5;10]$ war laut Lösung in der Übung $f(x) = 1/25x - 1/5$. t ist doch der y-Abschnitt/Höhe, aber wenn man nun in die Geradengleichung von $x \in [0;5]$ $x = 5$ einsetzt ist das Ergebnis 0 und somit müsste t doch für die 2. Geradengleichung 0 und nicht $-1/5$ sein oder?

Der y-Achsenabschnitt (das, was Sie als t bezeichnen) bezieht sich immer auf die Stelle $x = 0$! Das heißt: Wenn Sie sich die Strecke für $x \in [0;5]$ als Gerade vorstellen, dann schneidet diese die y-Achse bei $-1/5$. Um das rechnerisch zu überprüfen, setzen Sie einfach $x = 0$ in die Geradengleichung ein.

Aufgabe 4a) (Träger einer Zufallsvariablen): Was genau versteht man "anschaulich" unter dem Träger einer diskreten Zuvallsvariablen X? Mir ist klar, dass ich das so aus der Formelsammlung abschreiben könnte, aber ich habe noch nicht verstanden, was es in Bezug auf die Zufallsvariable bedeutet. Heißt es, dass die WSK-Verteilung von diskreten Zufallsvariablen immer größer Null ist?

Ja - siehe Definition 1.10. Der Träger sind diejenigen Werte einer Zufallsvariablen X, für die die Wahrscheinlichkeit positiv ist. Und nur dafür kann dann die Wahrscheinlichkeitsverteilung positive Werte annehmen werden

Blatt 7

Aufgaben 2 b) und c): nicht verstanden

Hier wurden Erwartungswerte und Varianzen unter Anwendung der Rechenregeln (Formelsammlung 1.5.3) berechnet.

das chuck-a-luck ist mir ebenfalls unklar. von vorn bis hinten...

Eine (ausführliche) Erklärung würde hier den Rahmen sprengen. Vielleicht wird es

Ihnen mit folgender, ausführlicherer Angabe, klarer:

Zeigt keiner der Würfel die gesetzte Zahl, so ist der Einsatz verloren. Anderenfalls erhält der Spieler (**zusätzlich zu seinem Einsatz**) für jeden Würfel, der die gesetzte Zahl zeigt, einen Betrag in Höhe des Einsatzes (hier eine Einheit).

Blatt 8

Aufgabe 4 c, was ist diese Z-variable und wie berechne ich daraus den erwartungswert und die varianz?

Die Zufallsvariablen Z wurde einfach so „festgelegt“ (definiert) über die angegeben Rechnung. Zum Berechnen von Erwarungswert und Varianz müssen die Rechenregeln für Erwartungswerte und Varianzen (Formelsammlung 1.5.3) und die Information, dass X und Y unabhängig sind, angewendet werden.

Aufgabe 4c): Mir ist nicht ganz klar wie die Lösungsidee hier ist mit 3 Variablen. Mit zwei ist es mir bekannt.

Sie setzen an mit: $E(Z) = E(3X + Y)$ und rechnen diese dann aus (sie Frage oben).

Blatt 10

Aufgabe 1): Wann benutzt man die Stetigkeitskorrektur und welche von beiden? Es gibt nur eine Stetigkeitskorrektur (oder man rechnet eben ohne Stetigkeitskorrektur).

Wann muss man bei der Approx. der Binomialvtlg. (ForSa. S 16) die Stetigkeitskorrektur also x+0,5 anwenden und wann nicht. Bei den Aufgaben 1 und 2 auf Übungsblatt 10 haben wir die Korrektur in Aufg. 2 angewendet, in der 1 nicht. Dabei ist X doch beide Male diskret (einmal Anhänger, einmal Mitglieder die zustimmen).

Grundsätzlich ist mit Stetigkeitskorrektur immer besser (Begründung in den Vorlesungsfolien in Kap. 1.7)! Aus didaktischen Gründen haben wir die Rechnung einmal mit, einmal ohne durchgeführt. In der Klausur geht aus der Angabe hervor, ob mit oder ohne gerechnet werden soll

Aufgaben 3 b) und c): nicht verstanden / sind ebenfalls völlig unklar / Ceteris Paribus

Zu 3b) Um die Werte der Verteilungsfunktion in der Standardnormalverteilungstabelle nachsehen zu können, muss zunächst eine Standardisierung durchgeführt werden (Standardisierung von $\bar{X}_n$ (=Z) sowie von oberer und unterer Grenze). Nach der allgemeinen Berechnung (also in Abhängigkeit von c und n) der Wahrscheinlichkeit $P(\bar{X}_n \in [-c, c])$ kann dann eine certeris paribus Analyse für n (also: Was passiert, wenn n erhöht wird und alle anderen Größen konstant gehalten werden?) durchgeführt werden. Hierbei ist es wichtig zu wissen, dass $\Phi(\cdot)$ streng monoton steigend ist. Da die Aufgabenstellung hier recht allgemein formuliert ist und hier keine Rechnung gefordert wird, wäre auch eine anschauliche Erklärung möglich.

Zu 3c) Analog zu 3b). Auch hier wird aus oben genanntem Grund eine Standardisierung durchgeführt, wofür Erwartungswert und Standardabweichung benötigt wird. Der Erwartungswert einer bernoulliverteilten Größe ist π. Die Varianz (und die Standardabweichung ergibt sich dann aus der Wurzel der Varianz) ist in der Aufgabe angegeben. Auch hier wird an die Berechnung der gesuchten Wahrscheinlichkeit eine ceteris paribus Analyse angeschlossen.

Blatt 11

Aufgabe 1): Wäre es vielleicht möglich die Lösung der Selbststudiumaufgabe online zu stellen? Ist erledigt.

Blatt 13

Wann wird welcher Test angewandt und wie forumliert man H1 und H0? Wir haben eine Art Baum aufgemalt, bei Übung 13, Aufg. 1 waren teilweise auch Tests dabei, die nicht mit im Baum waren, weshalb ich nicht ganz verstanden habe, warum diese angewendet wurden. Und beim Lagevergleich werden bei der verbundenen STP keine Tests aufgeführt anders als bei der unverbundenen.

Der Baum in der Übung sollte nur ein Skizze sein, um grob aufzuzeigen, wie man starten könnte, um sich eigenständig im Rahmen der Vorbereitung auf die Klausur das Testkapitel überblicksmäßig zusammenzustellen. In der Aufgabe 1 von ÜB 13 wurden auch Tests abgefragt die nicht in der Vorlesung behandelt wurden. Für die Klausur sind nur die in der Vorlesung behandelten Tests relevant.

Aufgabe 2: Wieso ist die statistische Hypothese Fall 1 gegeben? In der Angabe steht doch ... Der Anteil größer als 0,18. Folgt daraus nicht Fall 2, also Pi größer/gleich Pi 0 = 0,18?

Das, was ich zeigen möchte, schreibe ich in die Alternativhypothese!

Aufgabe 3: Wie kommt man auf die Umwandlung der Testgrößen-formel, dass am Ende (D-0 / sD) * wurzel n da steht? Und wie kommt man dann auf die Kovarianz und warum?

In der Situation mit verbundenen Stichproben wird aus der Stichprobe als neue zu testende Größe die Differenz für jede einzelne Beobachtung berechnet, in der Aufgabe mit D_i bezeichnet. Deshalb wird statt $\bar{X}$ der Mittelwert der Differenzen $\bar{D}$ eingesetzt. Die 0 stammt aus den Hypothesen. Die Kovarianz wird in der Aufgabe nicht gebraucht, diese dürfte in der Lösung auch nicht auftauchen.