

Lieses Verehrer oder das „Sekretärinnen-Problem“

Eine Aufgabe zur stochastischen Simulation und deren analytische Lösung

Timm Grams, Hochschule Fulda, www.hs-fulda.de/~grams

2. Februar 2008 (überarbeitet: 22.12.08)¹

Kommt ein Reitersmann daher durch die grüne Wiese,
hat ein Wams von Seide an, neigt sich vor der Liese:
Jungfrau so lieblich, Jungfrau so schön, tanzen wir ein wenig?
Mag nicht tanzen, danke schön, wart' auf einen König!

Kommt ein Kaufmannssohn daher, durch die grüne Wiese,
kommt daher mit stolzem Schritt, neigt sich vor der Liese:
Jungfrau so lieblich ...

Kommt ein Schneiderlein daher durch die grüne Wiese,
kommt daher mit leichtem Schritt, neigt sich vor der Liese:
Jungfrau so lieblich ...

Liese wartet Jahr und Tag auf der grünen Wiese,
doch kein König kommen mag, der da spricht zur Liese:
Jungfrau so lieblich, Jungfrau so schön, tanzen wir ein wenig?
Jeden sieht man weiter gehen, niemals kommt ein König.

Kommt der Schweinehirt daher, Johann Christoph Stoffel!
Hat nicht Schuh noch Strümpfe an, trägt nur Holzpantoffel.
"Lieber Stoffel tanz mit mir auf der grünen Wiese!"
Und der Stoffel tanzt mit ihr, mit der dummen Liese.

Aufgabenstellung

Die Liese, von der in dem Volkslied „Kommt ein Reitersmann daher“ die Rede ist, hat keine gute Auswahlstrategie. Gibt es eine Strategie, nach der Liese den besten Bewerber aus einem rein zufälligen Strom verschieden guter Bewerber hätte wählen können? Wenn auch nicht mit Sicherheit, so doch mit der größtmöglichen Wahrscheinlichkeit?

Nehmen wir einmal an, dass Liese die Zahl n der Bewerber kennt. Diese kommen nacheinander und kehren niemals zurück. Die ersten k Bewerber lässt sie abblitzen, merkt sich aber, wie gut der beste unter ihnen war. Dann entscheidet sie sich im weiteren Verlauf für den ersten, der besser ist als jener. Damit kann sie immer noch bei einem Mittelmäßigen oder gar bei Johann Christoph Stoffel landen. Aber es gibt eine gewisse Wahrscheinlichkeit dafür, dass sie tatsächlich den besten unter den Bewerbern („den König“) auswählt.²

Bestimmen Sie diese Trefferwahrscheinlichkeit für jedes denkbar k mittels stochastischer Simulation. Geben Sie die bestmögliche Strategie – das optimale k – an. Führen Sie die Simulation für mehrere n durch und ermitteln Sie jeweils den optimalen Wert für den Quotienten k/n . Was stellen Sie fest?

Wenn Sie die stochastische Simulationsaufgabe zu Ihrer Zufriedenheit gelöst haben, vergleichen Sie Ihre Lösung mit der analytischen Lösung des Problems. Es wird in der Literatur unter dem Namen Mitgift-Problem (Dowry Problem) oder auch *Sekretärinnen-Problem* behandelt. (Der Chef steht vor demselben Problem wie Liese: Aus einem Strom von unterschiedlich guten Bewerberinnen muss er seine zukünftige Sekretärin auswählen.)

¹ [Link zur Erstversion](#)

² Bruss, F. T.: Mathematische Unterhaltungen. Strategien der besten Wahl. Spektr. d. Wiss. (2004) 5, 102-104.

Führen Sie die Simulation auch mit anderen (realistischeren) Zielkriterien durch. Es muss ja nicht unbedingt der König sein. Ein Graf wäre ja auch schon ganz nett³:

- Minimierung der Chance, einen der unteren 10% der Bewerber zu erwischen.
- Maximierung der Chance, einen der oberen 10% der Bewerber zu bekommen.

Eine Formel für die Trefferwahrscheinlichkeit im allgemeinen Fall

Für das verallgemeinerte Sekretärinnen-Problem liefert die stochastische Simulation eine Näherungslösung. Aber man kann die Sache auch mit einer gehörigen Portion Kombinatorik und Wahrscheinlichkeitsrechnung angehen. Ein Student, Matthias Friedrich, regte mich zu der folgenden Berechnung an.

Als Variablen wähle ich

n für die Anzahl der Verehrer insgesamt,

k für die Anzahl der frühen (abgewiesenen) Verehrer, die den Maßstab bestimmen,

m für die Anzahl der akzeptablen Bewerber (die Besten),

α für die (unbekannte) Position des besten Bewerbers.

Für die Herleitung empfiehlt es sich, Mengen zu bezeichnen.

K ist die Testmenge, das ist die Menge der Positionen der frühen (abgewiesenen) Bewerber. Es ist $K = \{1, 2, \dots, k\}$.

M ist die (unbekannte) Menge der Positionen der akzeptablen Bewerber. Sie hat die Mächtigkeit $|M| = m$.

Bei der Berechnung der Trefferwahrscheinlichkeit hat es sich als praktisch erwiesen, zunächst zu unterscheiden, ob der beste Bewerber unter den frühen abgewiesenen Bewerbern ist ($\alpha \in K$) oder nicht ($\alpha \notin K$).

Fall 1 ($\alpha \in K$). Dieser Fall hat die Wahrscheinlichkeit $p_1 = k/n$. Da der Beste den Maßstab bildet, fallen alle anderen durch. Ein Treffer ist unmöglich.

Fall 2 ($\alpha \notin K$). Dieser Fall hat die Wahrscheinlichkeit $p_2 = 1 - p_1 = 1 - k/n$. Der Beste ist nicht in der Testmenge K enthalten. Jetzt sind wiederum zwei Fälle zu unterscheiden: Ist unter den ersten k Bewerbern überhaupt ein akzeptabler ($K \cap M \neq \Phi$) oder nicht ($K \cap M = \Phi$). Nehmen wir zunächst den Fall, dass keiner der Akzeptablen in der Testmenge erscheint.

Fall 2.1 ($K \cap M = \Phi$). Dieser Fall hat die bedingte Wahrscheinlichkeit

$$p_{2.1} = \frac{\binom{n-k-1}{m-1}}{\binom{n-1}{m-1}}.$$

Im Zähler steht die Anzahl aller Kombinationen, bei denen alle Akzeptablen auf den Plätzen $k+1, k+2, \dots, n$ zu finden sind. Dabei ist vorab klar, dass der Beste bereits dort steht, dass es also nur noch um die restlichen $m-1$ Bewerber und auch nur noch um $n-k-1$ zu vergebende Plätze geht. Im Nenner stehen alle möglichen Kombinationen – wieder mit Ausnahme des Besten und dessen Platz.

Wir unterscheiden den Fall 2.1 weiter hinsichtlich der Position j des ersten akzeptablen Bewerbers. Es gibt also für jedes $j = k+1, k+2, \dots, n-m+1$ einen untergeordneten Fall, nämlich

³ Todd, P. M.; Miller, G. F.: From Pride and Prejudice to Persuasion. Satisficing in Mate Search. In "Simple Heuristics that make us smart", S. 287-308, Herausgeber: Gigerenzer, G.; Todd, P. M.; and the ABC Research Group. Oxford University Press, New York, 1999

Fall 2.1.j ($j = \min(M)$). Die bedingte Wahrscheinlichkeit dieses Falles, nämlich dass der erste akzeptable Bewerber auf Position j erscheint, unter der Bedingung, dass keiner der akzeptablen Bewerber zur Testmenge gehört, ist gleich

$$p_{2.1.j} = \frac{m}{n-k} \cdot \frac{\binom{n-j}{m-1}}{\binom{n-k-1}{m-1}} = \frac{\binom{n-j}{m-1}}{\binom{n-k}{m}}.$$

Der Ausdruck kommt so zustande: Die Wahrscheinlichkeit, dass einer der akzeptablen Bewerber auf der Position j steht, ist gleich $\frac{m}{n-k}$. Die Anzahl der Platzkombinationen für die restlichen akzeptablen Bewerber oberhalb der Position j ergibt sich zu $\binom{n-j}{m-1}$. Und $\binom{n-k-1}{m-1}$ ist die Anzahl der insgesamt noch möglichen Platzkombinationen für die restlichen Bewerber. Der Quotient der Binomialkoeffizienten ist die Wahrscheinlichkeit, dass die restlichen akzeptablen Bewerber oberhalb der Position j zu finden sind unter der Bedingung, dass keiner von ihnen zur Testmenge gehört.

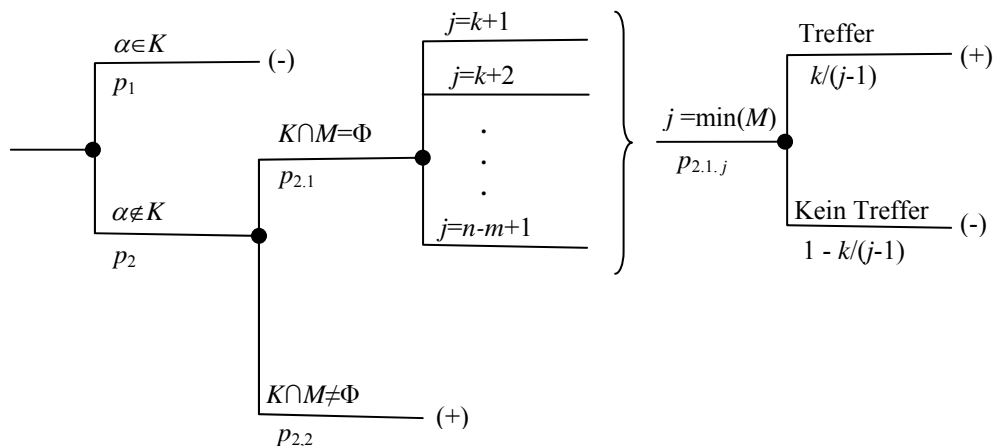
Zu einem Treffer kommt es, wenn der beste derjenigen Bewerber, die vor der Position j liegen, bereits in der Testmenge ist. Dieser Fall hat die (bedingte) Wahrscheinlichkeit $k/(j-1)$.

Damit ist der Fall 2.1 vollständig analysiert und wir kommen zur Alternative, nämlich

Fall 2.2 ($K \cap M \neq \emptyset$). Die bedingte Wahrscheinlichkeit dieses Falles ist gleich $1 - p_{2.1}$. Da in der Testmenge bereits ein akzeptabler Kandidat anzutreffen ist, wird Liese spätestens mit dem Besten – der ja außerhalb der Testmenge anzutreffen ist – einen akzeptablen Bewerber finden. Ein Treffer ist also garantiert.

Ereignisbaum und Trefferwahrscheinlichkeit

Die Fälle des vorigen Abschnitts und deren Wahrscheinlichkeiten lassen sich als Ereignisbaum visualisieren. Die Beschriftung der Zweige geschieht so: Oberhalb eines Zweiges steht die Bedingung des jeweiligen Falles, und unterhalb steht die zugehörige Wahrscheinlichkeit (jeweils unter der Bedingung der bis dahin durchlaufenen Zweige). An den Endzweigen (Blättern) ist angegeben, ob es sich um einen Treffer handelt (+) oder nicht (-).



Nach den Rechenregeln für Baumdiagramme⁴ erhalten wir für die Trefferwahrscheinlichkeit p die Formel

⁴ Sachs, L.: Angewandte Statistik. Anwendung statistischer Methoden. Springer, Berlin, Heidelberg 1992

$$p = p_2 \cdot (p_{2,2} + p_{2,1} \cdot \sum_{j=k+1}^{n-m+1} p_{2,1,j} \cdot \frac{k}{j-1}).$$

Jetzt sind nur noch die oben bereits errechneten bedingten Wahrscheinlichkeiten einzusetzen und nach einigen kleineren Umformungen steht das Resultat fest:

$$p = \frac{n-k}{n} - \frac{\binom{n-k}{m}}{\binom{n}{m}} + \frac{k}{\binom{n}{m}} \sum_{j=k+1}^{n-m+1} \frac{\binom{n-j}{m-1}}{j-1}$$

Für die numerische Berechnung mit dem Computer empfiehlt es sich, die Formel umzuschreiben. Anstelle der direkten Berechnung der Binomialkoeffizienten tritt die Berechnung der Quotienten aus je zwei Binomialkoeffizienten. Dann lassen sich vorab viele Faktoren kürzen und die Werte und Zwischenwerte der Rechnung bleiben im überschaubaren Rahmen. Hilfreich ist eine Funktion $f(a, b, c)$, die als Rückgabewert den Quotienten zweier Binomialkoeffizienten liefert und die Beziehung $\binom{a}{c} / \binom{b}{c} = \prod_{i=0}^{c-1} \frac{a-i}{b-i}$ ausnutzt.

Der Spezialfall $m = 1$

Für $m=1$ vereinfacht sich die Formel zu $p = \frac{1}{n} \sum_{j=k+1}^n \frac{k}{j-1} = \frac{k}{n} \sum_{i=k}^{n-1} \frac{1}{i}$. Die Summe kann man als

Riemannsche Summe eines Integrals auffassen. Das ergibt die Näherungsdarstellung

$$\sum_{i=k}^{n-1} \frac{1}{i} \approx \int_k^n \frac{1}{x} dx = \ln n - \ln k = \ln \frac{n}{k}. \text{ Insgesamt erhält man die Formel } p \approx -\frac{k}{n} \cdot \ln \frac{k}{n} \text{ für die Tref-}$$

ferwahrscheinlichkeit. Dieser Wert nimmt sein Maximum für $k/n = 1/e$ an. Die bestmögliche Chance, den „König“ zu bekommen, hat Liese, wenn sie sich die ersten etwa 37% der Bewerber anschaut und danach den nächsten wählt, der diese Bewerber übertrifft. Ihre Chance auf den Besten liegt dann bei ebenfalls etwa 37%.

Modifikation der Annahmeregeln

Bei der bisher analysierten Strategie kann Liese auch „leer ausgehen“. Liese könnte aber auch entscheiden, bevor sie ganz alleine dasteht, den letzten der Bewerber auf jeden Fall zu nehmen. Schlimmstenfalls landet sie dann eben beim Schweinehirten. Das nenne ich die „Johann-Christoph-Stoffel-Variante“.

Wir ändern die Annahmeregeln also derart, dass im Falle, dass die Regel bis Position $n-1$ noch keinen Treffer ergeben hat, der letzte der Bewerber auf jeden Fall genommen wird.

Es kann also auch im Fall 1 zu Treffern kommen, und zwar mit der bedingten Wahrscheinlichkeit $\frac{m-1}{n-1}$. Beim Übergang auf die Johann-Christoph-Stoffel-Variante vergrößert sich die

Wahrscheinlichkeit p demnach um den Summanden $\frac{k}{n} \cdot \frac{m-1}{n-1}$.

Ergebnisse

Die Grafik zeigt das Simulationsergebnis aus 1000 Versuchen für die Parameter $n=100$ und $m=10$ mit dem 2σ -Intervall (Binomialverteilung je Schätzwert). Außerdem ist der exakte Kurvenverlauf eingezeichnet und zum Vergleich die Kurven für die modifizierte Regel und für die Suche nach dem „König“ ($m=1$).

