Zu Abschnitt 20.6: Eine optimale Auswahlstrategie

Problemstellung

Häufig begegnen einen Auswahlsituationen, bei denen sich einer nach dem Anderen vorstellt bzw. ein Produkt nach dem Anderen angeschaut wird und man sich letztendlich für einen Kandidaten entscheiden muss – und zwar sofort nach der Vorstellung bzw. Inaugenscheinnahme: Beispiele sind Bewerbungen – und zwar von Arbeitgeber- wie auch von Bewerberseite her, Kauf eines neuen Autos, Wahl des Traumprinzen etc. (Im Falle von Bewerbungen entsprechen Ablehnungen jeweils einem schlechteren "Score" als jede angenommene Bewerbung; nur bei den angenommenen Bewerbungen kommt es auf die Arbeitgeberqualität an.) Die Auswahl wird am Beispiel der Wahl des Traumprinzen folgendermaßen spezifiziert:

- n Prinzen stellen sich nacheinander vor. Einer von ihnen ist der Beste ("Traumprinz" T), doch es ist völlig unbekannt, ob er sich am Anfang, am Ende oder irgendwann in der Mitte vorstellt.
- Die Prinzessin kann sofort feststellen, welcher von zwei Prinzen der "bessere" ist.
- Die Prinzession muss sich sofort nach der jeweiligen Vorstellung entscheiden, ehe der nächste kommt: Ein abgewiesener oder hingehaltener Prinz ist beleidigt und bewirbt sich kein zweites Mal.
- Die Prinzessin ist nur mit dem Besten, dem Traumprinzen T zu frieden. Der zweitbeste ist für sie so schlecht wie der schlechteste.

Welche Strategie muss man wählen, damit der Traumprinz mit maximaler Wahrscheinlichkeit gewählt wird, $P(T)$ =max! Wir schränken die Suche auf folgende Klasse von Strategien S_k ein:

$$\textbf{Strategie } \boldsymbol{S_k}\textbf{:} \quad \text{wähle} \begin{cases} \text{keinen in den ersten } k \text{ Runden} \\ \text{den bisher Besten danach} \end{cases}$$

Die ersten k Runden dienen also der "Markterkundung", auf der sich die Prinzessin umschaut, welche Qualität von Prinzen ihr überhaupt angeboten werden bzw. im Falle von Bewerbungen dem Testen des "Marktwertes" und des "Angebots". Mit dieser "Marktkenntnis" kann sie danach gezielt in der folgenden "Selektionsphase" auswählen.
Wir definieren folgende Ereignisse:

- **Ereignis $\boldsymbol{T_t}$:** Traumprinz erscheint als t-ter Kandidat, es gilt

$$P(T_t) \equiv P(\text{Traumprinz an Stelle } t) = \frac{1}{n}$$

- **Ereignis $\boldsymbol{T}$:** Traumprinz wird gewählt

Folgende Beispiele verdeutlichen das Optimierungspotenzial:

- Die Strategie S_0 heißt: "wähle den ersten Kandidat". Hier ist
$$P(T|S_0) = P(T_0) = \frac{1}{n},$$
ansonsten bekommt frau einen minderwertigen Prinzen.

- Bei der Strategie S_{n-1} ist die Orientierungsphase erst beim letzten Prinzen vorbei. Auch hier ist
$$P(T|S_{n-1}) = P(T_n) = \frac{1}{n},$$
da die Prinzessin nur dann glücklich ist, wenn der Traumprinz tatsächlich an n-ter Stelle erscheint.

- Sicher gibt es bessere Strategien. Die Aufgabe ist nun folgende: Wähle $k \in \{0, 1, \ldots, n-1\}$ so, dass gilt
$$P(T|S_k) = \max_k!$$

Um $P(T|S_k)$ zu ermitteln, muss man die bedingten Wahrscheinlichkeiten für alle möglichen Positionen t des Traumprinzen separat betrachten. Da sich die enstprechenden Ereignisse T_t gegenseitig ausschließen, kann man dann den Satz von der totalen Wahrscheinlichkeit anwenden:

$$P(T|S_k) = \sum_{t=1}^{n} P(T_t) P(T|S_k, T_t).$$

Dabei ist $P(T|S_k, t)$ die bedingte Wahrscheinlichkeit dafür, dass der Traumprinz bei der Strategie S_k gefunden wird, falls er als t-ter Prinz erscheint. Diese Wahrscheinlichkeit ist nach der klassischen (Laplace'schen) Wahrscheinlichkeitsdefinition gegeben durch

$$P(T|S_k, t) = \begin{cases} 0 & \text{falls } t \leq k \\ \frac{k}{t-1} & \text{sonst.} \end{cases}$$

Dies macht man sich wie folgt klar:

- Im ersten Fall $t \leq k$ erscheint der Traumprinz in der Orientierungsphase und wird deshalb immer abgewiesen.

- Ansonsten erscheint der Traumprinz in der Selektionsphase, kann also potenziell gewählt werden. Ob er tatsächlich gewählt wird, hängt von der Position b des besten Prinzen B ab, welcher *vor* dem Traumprinz erscheint. Prinzipiell kann B auf irgendeiner der $(t-1)$ Positionen von 1 bis $t-1$ erscheinen, und zwar gleichwahrscheinlich: Nur, wenn B nicht in der "Markterkundungsphase" erscheint, kommt er dem Traumprinzen zuvor. In k von $(t-1)$ Fällen taucht B jedoch in der Erkundungsphase auf, so dass T mit einer (klassischen) Wahrscheinlichkeit von $k/(t-1)$ gewählt wird.

Merke, dass der *zweitbeste* Z der Prinzen, welcher vor dem Traumprinz erscheint, sowie alle weiteren Prinzen *keine* Rolle spielen. Zwar können auch die durchaus zum Zuge kommen (z.B. wenn erst Z, dann B und dann erst der Traumprinz T selbst kommen und alle drei in der Selektionsphase $i > k$ erscheinen) doch hat das auf die Wahrscheinlichkeit, T zu wählen, keinen Einfluss: Wenn es in dieser Konfiguration Z nicht gäbe oder er, entgegen den Regeln, nicht angenommen würde, käme immer noch B dem Traumprinzen zuvor ... Damit können wir $P(T|S_k)$ schließlich ausrechnen:

$$\begin{aligned} P(T|S_k) &= \sum_{t=1}^{n} P(t) P(T|S_k, t) \\ &= \sum_{t=k+1}^{n} \frac{1}{n} \frac{k}{t-1} \\ &\overset{t=j+1}{=} \frac{1}{n} \sum_{j=k}^{n-1} \frac{k}{j} \end{aligned}$$

und damit schließlich

$$\boxed{P(T|S_k) = \frac{k}{n} \sum_{j=k}^{n-1} \frac{1}{j}}$$

Das Optimum k_{opt} wird nun durch Maximieren von $P(T|S_k)$ bezüglich k gefunden. Dazu berechnen wir das Inkrement

$$\begin{aligned} \Delta P(T|S_k) &= P(T|S_k) - P(T|S_{k-1}) \\ &= \frac{k}{n} \sum_{j=k}^{n-1} \frac{1}{j} - \frac{k-1}{n} \sum_{j=k-1}^{n-1} \frac{1}{j} \\ &= \frac{1}{n} \left(\sum_{j=k}^{n-1} \frac{1}{j} - (k-1) \frac{1}{(k-1)} \right) \\ &= \frac{1}{n} \left(\sum_{j=k}^{n-1} \frac{1}{j} - 1 \right). \end{aligned}$$

Wir fangen nun bei $k = 1$ an und erhöhen k solange, wie das Inkrement positiv bleibt. Das maximale k, für welches das der Fall ist, ist offensichtlich das optimale (da bei weiterer Erhöhung wegen der negativen Inkremente $P(T|S_k)$ wieder abfällt). Dies führt direkt zu folgender "Goldenen Regel":

Wähle k so, dass $\sum\limits_{j=k}^{n-1} \frac{1}{j}$ gerade noch ≥ 1 ist!

Für $n = 5$ Prinzen gilt z.B. $\underline{\underline{k_{\text{opt}} = 2}}$, da $\sum_{j=2}^{4}(1/j) = 1/2 + 1/3 + 1/4 = 13/12 > 1$, aber $\sum_{j=3}^{4}(1/j) = 1/3 + 1/4 = 7/12 < 1$ ist. Außerdem ist die Wahrscheinlichkeit, den Traumprinzen zu finden, gegeben durch

$$P(T|k_{\text{opt}}, n = 5) = \frac{2}{5}\frac{13}{12} = \underline{\underline{\frac{13}{30}}}$$

Kontinuierlicher Grenzfall

Im Folgenden zeigen wir, dass für hinreichend viele Bewerber ($n > 30$) die "Drittel-Drittel-Regel" gilt:

> Optimal ist es, etwa 1/3 der Kandidaten zur "Markterkundung" zu nehmen; Den optimalen Kandidaten bekommt man dann mit einer Wahrscheinlichkeit, die etwas oberhalb 1/3 liegt

Im Grenzfall sehr vieler Kandidaten wird einiges leichter: Wir ersetzen alle Positionsvariablen wie j durch relative Positionen $x_j = j/n$ und die Summe durch ein Integral gemäß folgender Regel [1]

$$\sum_j f_j = \sum_j \Delta j f_j \approx \int_j f(j)\ \mathrm{d}j.$$

Also für große n:

$$\begin{aligned} P(T|S_k) &= \frac{k}{n} \sum_{j=k}^{n-1} \frac{1}{j} \\ &= x_k \sum_{j=nx_k}^{n-1} \frac{1}{j} \\ &\approx x_k \int_{j=nx_k}^{n-1} \frac{1}{j}\ \mathrm{d}j \\ &\overset{j=nx_j}{=} x_k \int_{x_k}^{\frac{n-1}{n}} \frac{1}{x_j} dx_j \end{aligned}$$

und im Limes $n \to \infty$ mit der Integrationsvariablen $x_j = x$ und der Strategie $S_k \to S_y$, bei der ein Bruchteil $y = x_k = \frac{k}{n}$ aller Kandidaten der Markterkundung dient:

$$P(T|S_y) = y \int_y^1 \frac{1}{x} dx = \underline{\underline{-y \ln y.}}$$

Das Maximum bezüglich des relativen Anteils y der Markterkundungsphase ist nun gegeben durch

$$\frac{\mathrm{d}}{\mathrm{d}y} P(T|S_y) = -\ln y - 1 = 0$$

also

$$\underline{\underline{y_{\text{opt}} = e^{-1}.}}$$

Der optimale Anteil der Erkundungsphase ist also $1/e$ mit $e = 2.7...$, also ungefähr ein Drittel. Die Wahrscheinlichkeit, den Traumprinz auch bei unendlich vielen Kandidaten dennoch zu bekommen, ist

$$\underline{\underline{P(T|S_{y_{\text{opt}}}) = e^{-1}}},$$

also ebenfalls etwa ein Drittel!

[1] $\Delta j = 1$ sagt nichts weiteres aus, als dass sich j von einem zum nächsten Summanden um 1 erhöht. Bei anderen Inkrementen muss man das berücksichtigen, da man nur die Summe *mit* Inkrement für große n angenähert durch das Integral ersetzen darf.