Vorlesung 11a

Bayes'sche Anteilschätzung und die Pólya-Urne

Teil 4

Die Aktualisierung des Bayes-Schätzers aufgefasst als ein Pólya-Schritt

(Buch S. 113-114)

Wir erinnern hier an den direkten Beweis von

$$\mathbf{P}(Z_{n+1} = 1 | K_n = k) = \frac{k+1}{n+2}$$

in unserer Übungsaufgabe 36:

Das Kernargument war:

Gegeben dass U_0 das $(k+1)$-t kleinste der $U_0, \ldots, U_n$ ist,

gibt es für das Einfügen von U_{n+1}

$k+1$ Slots links von U_0

und

$n+2-(k+1)$ Slots rechts von U_0.

Hier noch einmal langsamer an einem Beispiel:

Wie ist Z_3 verteilt, gegeben $\{K_2 = 2\}$?

Wir übersetzen das in die Darstellung durch U_0, U_1, U_2, U_3.

Wie wahrscheinlich ist es, dass $U_3 < U_0$,

gegeben U_0 ist größer als U_1 und als U_2?

Der (Größen-)Rang von U_3 in U_0, U_1, U_2, U_3 ist uniform verteilt, also fügt sich U_3 je mit Wkeit $1/4$ in einen der 4 von U_0, U_1, U_2 aufgemachten Slots (einer links vom Minimum, zwei in der Mitte, einer rechts vom Maximum) ein.

Gegeben $U_0 = \max(U_0, U_1, U_2)$
(gleichbedeutend damit: gegeben $\{K_2 = 2\}$),
führen 3 dieser Slots auf $U_3 < U_0$ (und damit auf $Z_3 = 1$)
und einer auf $U_3 > U_0$ (und damit auf $Z_3 = 0$).

Fazit:

$$\mathbf{P}(Z_3 = 1 | K_2 = 2) = 3/4, \quad \mathbf{P}(Z_3 = 0 | K_2 = 2) = 1/4.$$

Folgende Frage zu rein zufälligen Permutationen ist

dasselbe in Grün:

Gegeben in einer rein zufälligen Permutation Π von 0,1,2,3

ist $\Pi(0)$ größer als $\Pi(1)$ und als $\Pi(2)$.

Was ist dann die W'keit von $\{\Pi(3) < \Pi(0)\}$?

Antwort: 3 der 4 gleich wahrscheinlichen Slots für $\Pi(3)$

führen auf $\{\Pi(3) < \Pi(0)\}$, einer auf $\{\Pi(3) > \Pi(0)\}$.

Also:

$$\mathbf{P}(\Pi(3) < \Pi(0) | \Pi(1) < \Pi(0), \Pi(2) < \Pi(0)) = 3/4.$$

Fazit

(vgl. Buch S. 113-114):

$(Z_1, Z_2, \ldots)$ ist so verteilt wie eine zufällige Pólya-Folge ,
d.h. wie die Farbfolge (F_i) der Züge aus einer Pólya-Urne
(bzw. die Farbfolge der Zugänge in einer Pólya-Urne)
mit anfänglich
einer Kugel von Farbe 1 und einer Kugel von Farbe 2
und $F_i := I_{\{i-\text{te gezogene Kugel hat Farbe } 1\}}$.

Diese Einsicht können Sie an der folgenden mit unserer A36 verwandten Übungsaufgabe rekapitulieren:

Zusatzaufgabe auf Blatt 10 der "Stochastik für die Informatik", WiSe 19/20: U_0, U_1, U_2, U_3 seien unabhängig und uniform auf [0, 1] verteilt.

(a) Für $i = 0, 1, 2, 3$ setzen wir $R(i) := \sum_{j\in\{0,1,2,3\}\setminus\{i\}} I_{\{U_j<U_i\}}$.
Begründen Sie, warum $(R(0), R(1), R(2), R(3))$ eine rein zufällige Permutation von 0,1,2,3 ist.

b) Bestimmen Sie $\mathbf{P}(U_1 < U_0, U_2 < U_0)$ und
$\mathbf{P}(U_1 < U_0, U_2 < U_0, U_3 \geq U_0)$.

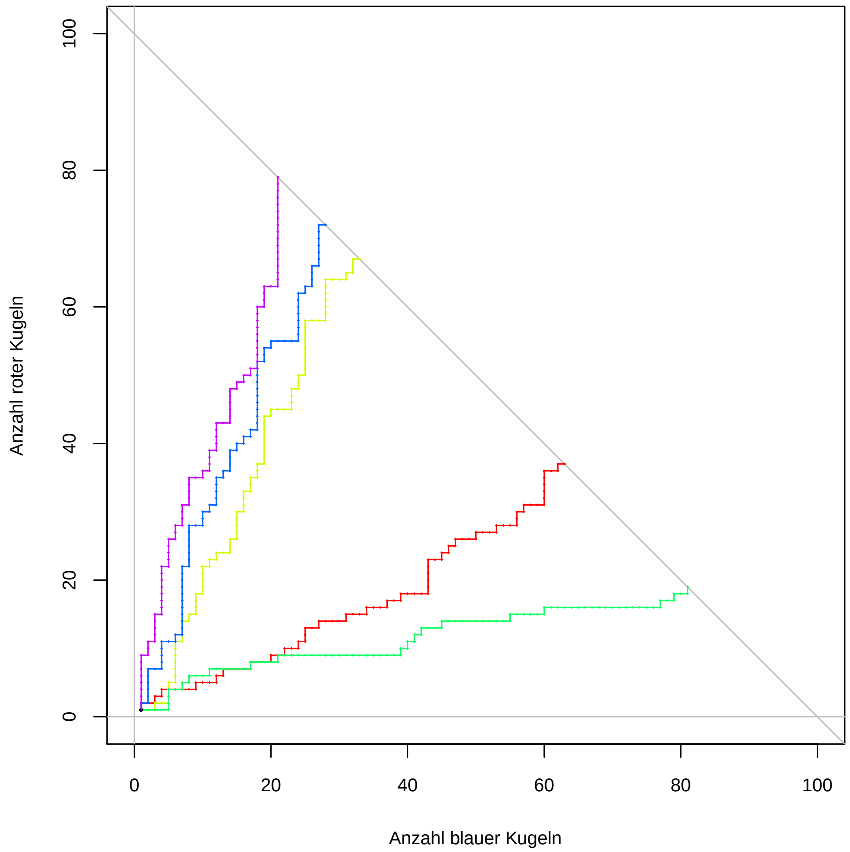
c) Wir definieren $Z_1 := I_{\{U_1<U_0\}}, Z_2 := I_{\{U_2<U_0\}}, Z_3 := I_{\{U_3<U_0\}}$.
(i) Berechnen Sie $\mathbf{P}(Z_2 = 1|Z_1 = 1)$ und $\mathbf{P}(Z_3 = 0|Z_1 = 1, Z_2 = 1)$.
(ii) Tragen Sie die Gewichte der Übergangsverteilungen des durch (Z_1, Z_2, Z_3) beschriebenen 3-stufigen Experiments in einen Baum der Tiefe 3 ein.

Die Asymptotik der Farbanteile in der Pólya-Urne.

Wegen der von uns schon erkannten Verteilungsgleichheit einer Münzwurffolge mit uniform verteiltem Erfolgsparameter und einer Pólya-Folge $(F_1, F_2, \ldots)$ ist das folgende Simulationsergebnis nicht mehr ganz so erstaunlich.

Dort werden 5 Pólya-Pfade $(X_n) = (X_n^{(1)}, X_n^{(2)})$, $n = 1, 2, \ldots$ simuliert,
mit $X_n^{(1)} := 1 + F_1 + \cdots + F_n$
die Anzahl der Kugeln von Farbe 1 nach n Zügen
und $X_n^{(2)} := 1 + (1 - F_1) + \cdots + (1 - F_n)$
die Anzahl der Kugeln von Farbe 2 nach n Zügen.

5 Realisierungen von Pólya-Pfaden $(X_i)_{i=1,\dots 100}$



$\frac{1}{n}(F_1 + \cdots + F_n)$ scheint zu konvergieren -

allerdings gegen einen zufälligen Grenzwert!

Tatsächlich gilt: Bedingt unter U_0 ist

$$I_{\{U_1<U_0\}}, I_{\{U_2<U_0\}}, \cdots$$

Münzwurffoge zum Parameter U_0.

Nach dem Gesetz der großen Zahlen konvergiert

$$\frac{1}{n}\left(I_{\{U_1<U_0\}} + \cdots + I_{\{U_n<U_0\}}\right)$$

gegen U_0.

5 Realisierungen von Pólya-Pfaden $(X_i)_{i=1,\ldots 1000}$