Vorlesung 13a

“Mit welcher Wahrscheinlichkeit

geht morgen die Sonne auf?”

Bayes’sche Schätzung von p und die Pólya-Urne

Buch S. 113/114, S. 127

Wohl mit einem Augenzwinkern stellte

Pierre-Simon Laplace (1729-1847) die folgende Frage:

Stellen wir uns vor, es wird jeden Morgen

mit einer uns unbekannten, konstanten

Erfolgswahrscheinlichkeit p eine Münze geworfen.

Bei Erfolg geht die Sonne auf,

bei Misserfolg ist der jüngste Tag gekommen.

Angenommen es ist bisher die Sonne 1 Million mal

aufgegangen. Wie wahrscheinlich ist es,

dass sie auch morgen wieder aufgeht?

Unser Lieblingsschätzer für p ist $\widehat{p} = H = \frac{K}{n}$

(Anzahl Erfolge geteilt durch Anzahl Versuche).

Der ergäbe hier den Schätzwert 1,

auch schon für $n = 10$,

und auch schon sogar für $n = 1$,

anstelle von $n = 10^6$.

Hmmm...

Das gibt Anlass zum Nachdenken!

Laplace schlug vor,
an ein zweistufiges Zufallsexperiment zu denken.

In der ersten Stufe wird die Erfolgswahrscheinlichkeit P uniform aus $[0, 1]$ gewählt.

In der zweiten Stufe wird, gegeben $P = p$, ein wiederholter p-Münzwurf $(Z_1, Z_2, \ldots)$ durchgeführt.

Sei $K_n := Z_1 + \cdots + Z_n$ die Anzahl der Erfoge in den ersten n Versuchen.

Fragen:

1a. $\mathbf{P}(Z_{n+1} = 1 \mid K_n = k) =?$

Was ist

1b. der bedingte Erwartungswert

und

2. die bedingte Verteilung

der zufälligen Erfolgswahrscheinlichkeit P,

gegeben $\{K_n = k\}$?

Die Antwort auf 1a und 1b ist dieselbe:

$$\mathbf{P}(Z_{n+1} = 1 | K_n = k) = \mathbf{E}[P | K_n = k).$$

Erinnerung an eine Übungsaufgabe:

43. S (Frei nach dem Eingangsbeispiel im 2. Vortrag der diessemestrigen Ringvorlesung *Algorithmen, Maschinelles Lernen, Quantencomputing*):
Jemand führt einen Münzwurf vor. Aus gewissen Gründen kommt nur in Frage, dass er entweder die ganze Zeit eine faire $\mathtt{01}$-Münze verwendet (d.h. $p = 1/2$) oder eine mit $p = 8/10$.
Bevor er zu werfen beginnt, beginnt, schätzen Sie die Wahrscheinlichkeit, dass er eine faire Münze verwendet, mit 0.9 ein.
Wie aktualisieren Sie *Ihre Einschätzung* der Wahrscheinlichkeit, dass es sich um eine faire Münze handelt, nachdem

(i) beim ersten Wurf eine Eins
(ii) bei den beiden ersten Würfen eine Eins

geworfen wurde?

Übergangsgewichte:

	0	1
0.5	0.5	0.5
0.8	0.2	0.8

Gemeinsame Verteilungsgewichte:

		0	1
0.9	0.5	$0.9 \cdot 0.5$	$0.9 \cdot 0.5$
0.1	0.8	$0.1 \cdot 0.2$	$0.1 \cdot 0.8$

Gemeinsame Verteilungsgewichte:

	0	1
0.5	$0.5 \cdot 0.9$	$0.5 \cdot 0.9$
0.8	$0.2 \cdot 0.1$	$0.8 \cdot 0.1$

$$\mathbf{P}(P = 0.5 | K_1 = 1) = \frac{0.5 \cdot 0.9}{0.5 \cdot 0.9 + 0.8 \cdot 0.1} \approx 0.85$$

Übergangsgewichte:

	0	1	2
0.5	0.25	0.5	0.25
0.8	$(0.2)^2$	$2 \cdot 0.8 \cdot 0.2$	$(0.8)^2$

Gemeinsame Verteilungsgewichte:

		0	1	2
0.9	0.5	$0.9 \cdot 0.25$	$0.9 \cdot 0.5$	$0.9 \cdot 0.25$
0.1	0.8	$0.1 \cdot (0.2)^2$	$0.1 \cdot 2 \cdot 0.8 \cdot 0.2$	$0.1 \cdot (0.8)^2$

Gemeinsame Verteilungsgewichte:

	0	1	2
0.5	$0.25 \cdot 0.9$	$0.5 \cdot 0.9$	$0.25 \cdot 0.9$
0.8	$(0.2)^2 \cdot 0.1$	$2 \cdot 0.8 \cdot 0.2 \cdot 0.1$	$(0.8)^2 \cdot 0.1$

$$\mathbf{P}(P = 0.5 | K_2 = 2) = \frac{0.25 \cdot 0.9}{0.25 \cdot 0.9 + 0.64 \cdot 0.1} \approx 0.78$$

In der eben diskutierten Übungsaufgabe

hatte die a-priori Verteilung von P

die Gewichte

0.9 in $p = 1/2$ und 0.1 in $p = 8/10$.

Jetzt nehmen wir - auf den Spuren von Laplace -

die a-priori Verteilung von P

als uniform auf $[0, 1]$ an.

Unsere Fragen waren:

1a. $\mathbf{P}(Z_{n+1} = 1 \mid K_n = k) =?$

1b. $\mathbf{E}[P \mid K_n = k] =?$

2. Was ist die bedingte Verteilung von P, gegeben $\{K_n = k\}$?

Beginnen wir mit Frage 2:

Was ist die bedingte Verteilung von P,

gegeben $\{K_n = k\}$?

Dazu ein Simulationsexperiment für $n = 2$:

Für 3000 uniform aus $[0, 1]$ gewählte p

wird jeweils ein 2-maliger p-Münzwurf durchgeführt.

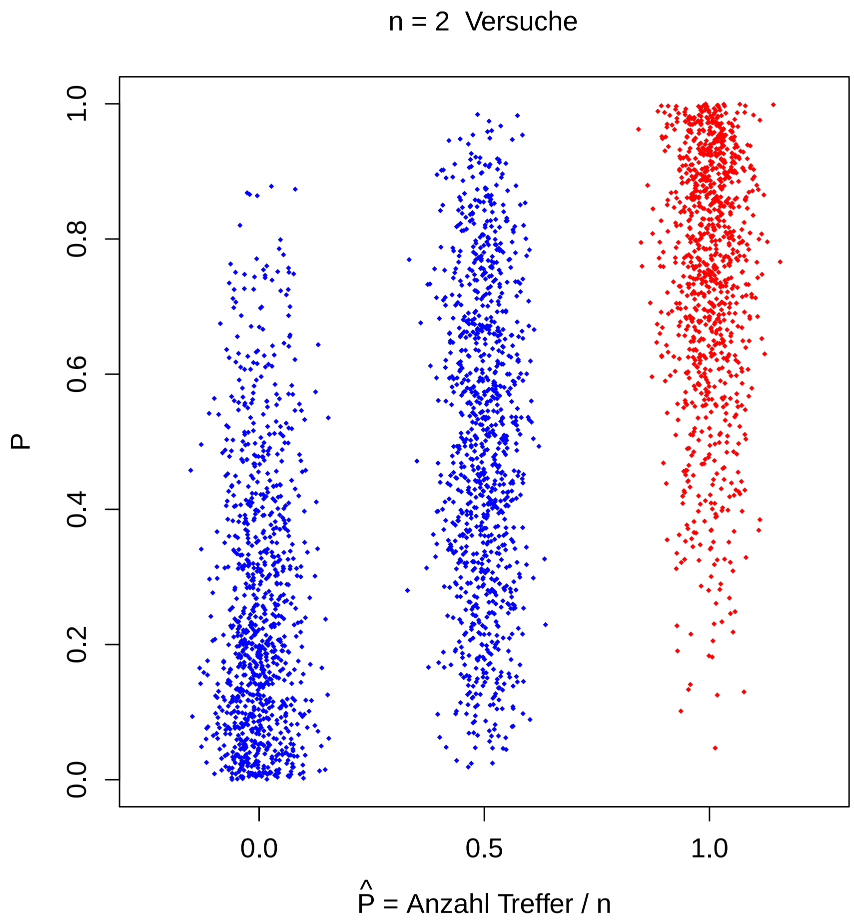
Jeder der 3000 Punkte

gibt eine gemeinsame Realisierung

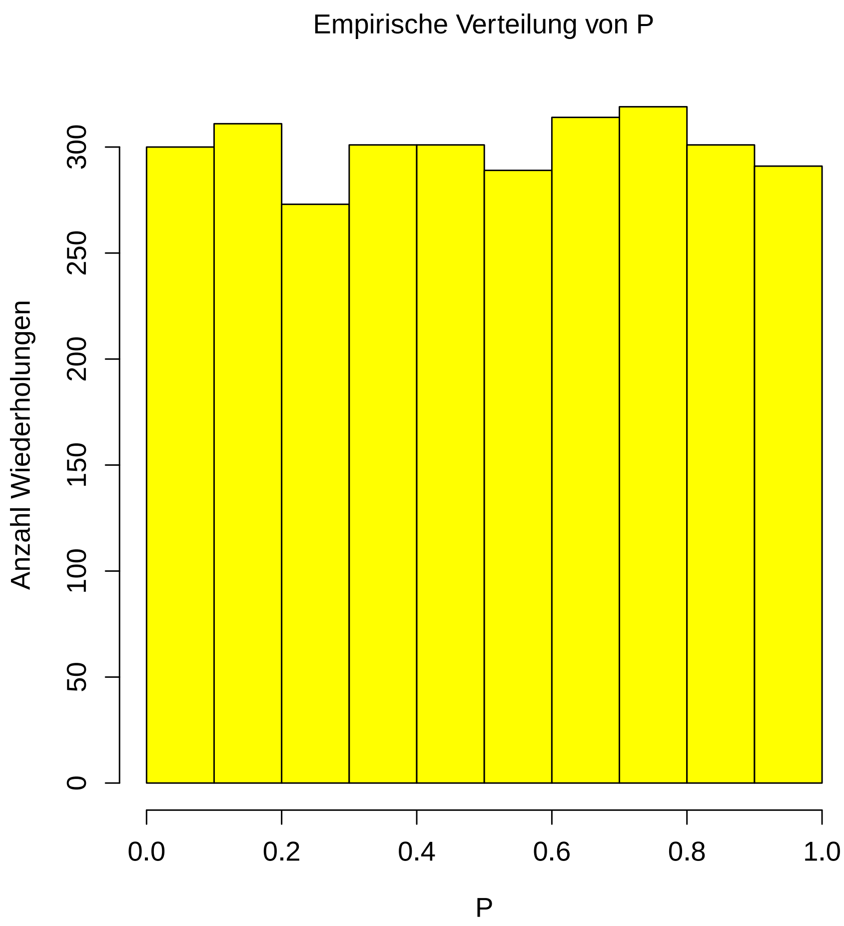
von P (vertikal) und $K_2/2$ (horizontal).

Diese und auch die folgenden Illustrationen stammen aus Ideen und Programmen von Brooks Ferebee.

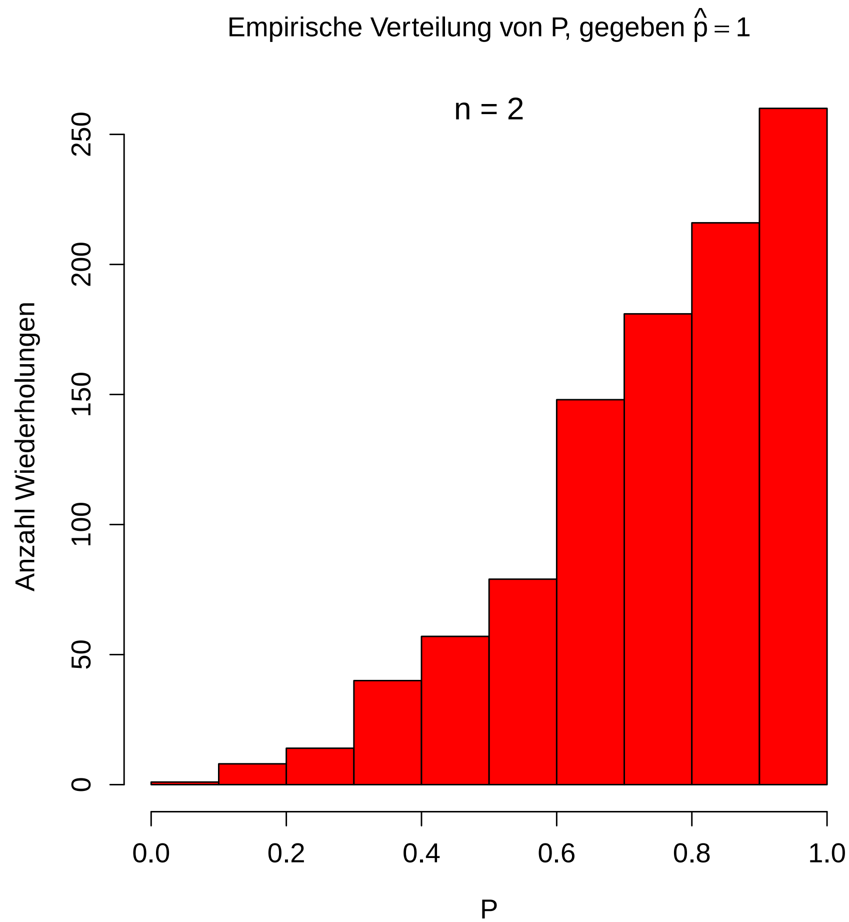
Gemeinsame Verteilung von $(P, K_2/2)$



A-Priori Verteilung von P



Bedingte Verteilung von P gegeben $\{K_2 = 2\}$



Jetzt für $n = 10$:

Für 11000 uniform aus $[0, 1]$ gewählte p

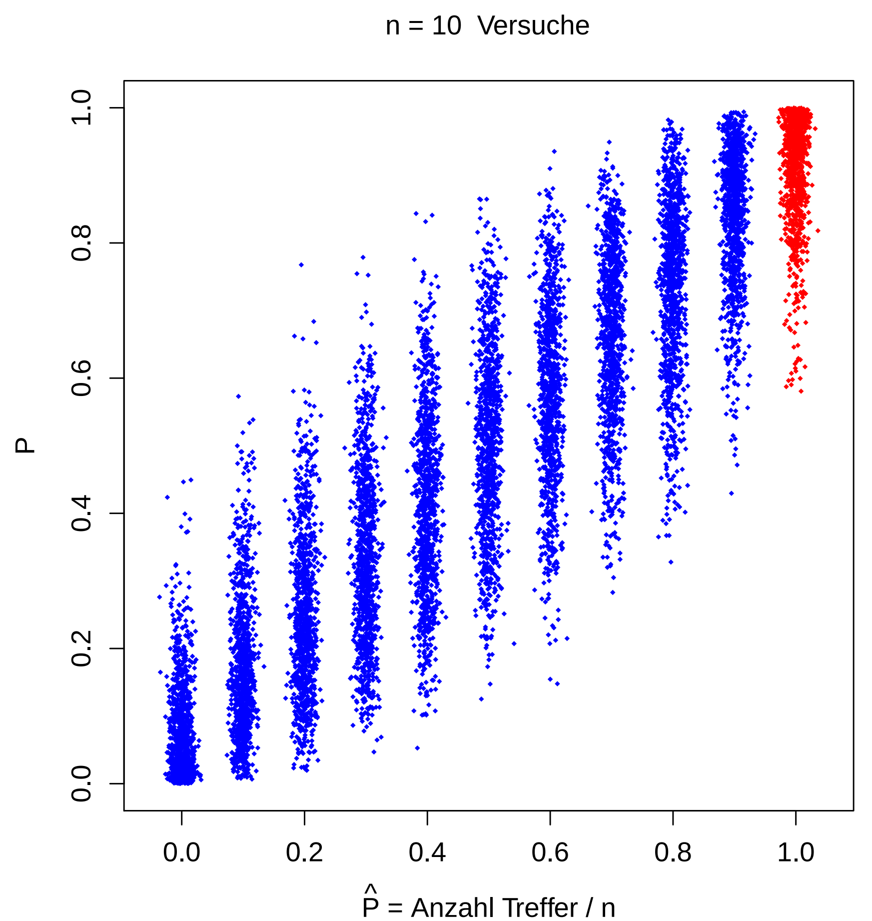
wird jeweils ein 10-maliger p-Münzwurf durchgeführt.

Jeder der 11000 Punkte

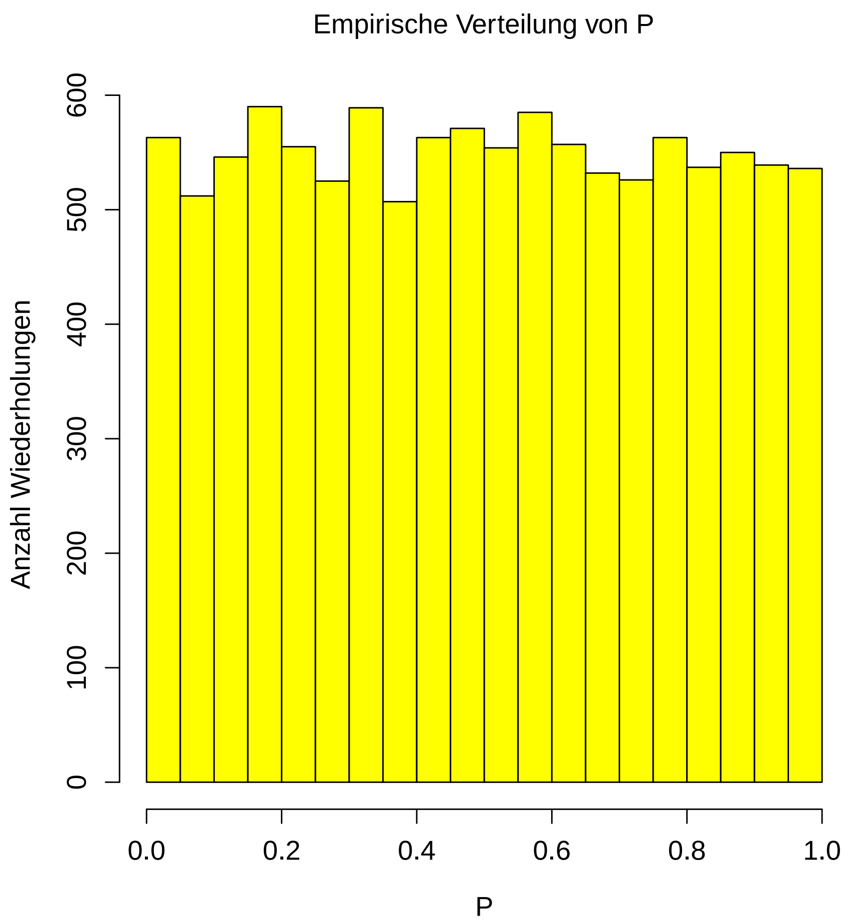
gibt eine gemeinsame Realisierung

von P (vertikal) und $\frac{K_{10}}{10}$ (horizontal).

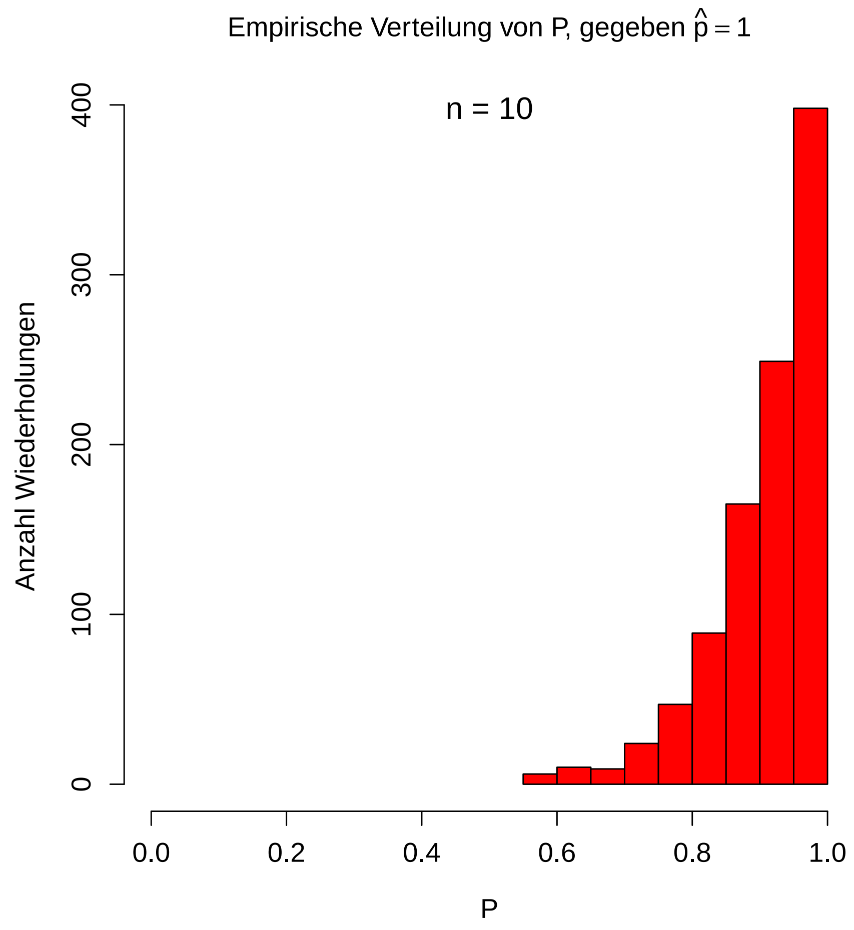
Gemeinsame Verteilung von $(P, \frac{K_{10}}{10})$



A-Priori Verteilung von P



Bedingte Verteilung von P gegeben $\{K_{10} = 10\}$



Wie kann man das elegant verstehen und “nachrechnen”?

Betrachten wir den Fall $n = 2$:

Stellen wir (P, Z_1, Z_2) dar

mittlels dreier

unabhängiger, uniform auf $[0, 1]$ verteilter Zufallsvariabler

U_0, U_1, U_2:

$$P := U_0, \quad Z_1 := I_{\{U_1 < U_0\}}, \quad Z_2 := I_{\{U_2 < U_0\}}$$

Bis auf ein Ereignis von Wahrscheinlichkeit Null gilt:

$$\{K_2 = 2\} = \{U_0 = \max(U_0, U_1, U_2)\}.$$

Wie ist das Minimum von 3 unabhängigen

Unif([0, 1])-verteilten ZV'en verteilt?

Die Verteilungsfunktion ist

$$\mathbf{P}(\max(U_0, U_1, U_2) \le b) = b^3, \quad 0 \le b \le 1.$$

Die Dichte ist $\quad 3b^2\, db, \ \ 0 \le b \le 1.$

Das ist auch die bedingte Dichte von P, gegeben $\{K_2 = 2\}$.

Was $n = 2$ recht ist, soll einem allgemeinen n billig sein:

Die Verteilungsfunktion von $\max(U_0, U_1, \dots, U_n)$ ist

$$\mathbf{P}(\max(U_0, U_1, \dots, U_n) \le b) = b^{n+1}, \quad 0 \le b \le 1.$$

Die Dichte ist $\quad (n+1)b^n \, db, \;\; 0 \le b \le 1.$

Das ist auch die bedingte Dichte von P, gegeben $\{K_n = n\}$.

Der bedingte Erwartungswert von P gegeben $\{K_n = n\}$ ist

$$\int_0^1 b \, (n+1)b^n \, db = \frac{n+1}{n+2}.$$

Was $n = 2$ recht ist, soll einem allgemeinen n billig sein:

Die Verteilungsfunktion von $\max(U_0, U_1, \ldots, U_n)$ ist

$$\mathbf{P}(\max(U_0, U_1, \ldots, U_n) \le b) = b^n, \quad 0 \le b \le 1.$$

Die Dichte ist $\quad (n+1)b^n\, db, \;\; 0 \le b \le 1.$

Das ist auch die bedingte Dichte von P, gegeben $\{K_n = n\}$.

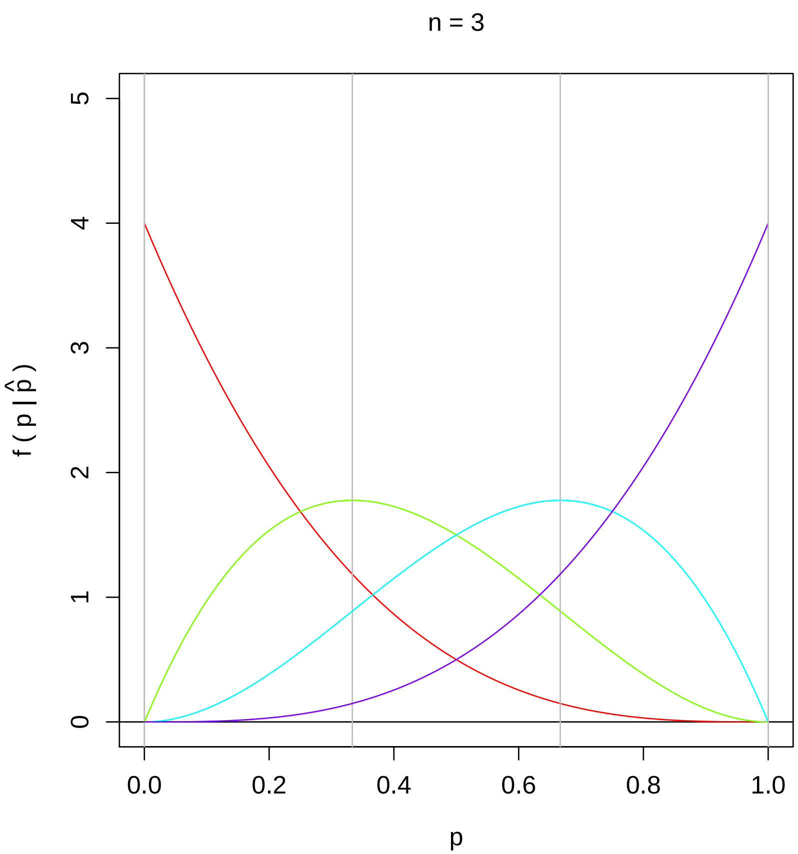
Der bedingte Erwartungswert von P gegeben $\{K_n = n\}$ ist

$$\mathbf{E}[P \,|K_n = n] = \tfrac{n+1}{n+2}.$$

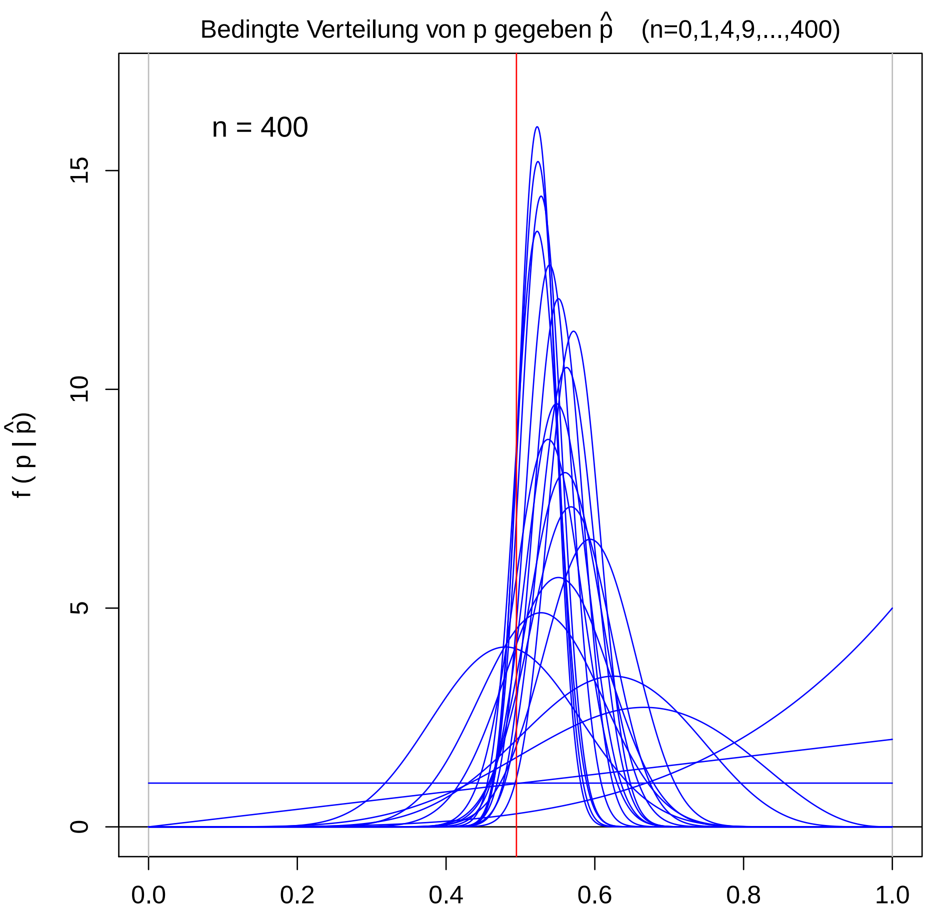
Dies gibt auch die Antwort auf die Frage von Laplace:

$$\mathbf{P}[Z_{n+1} = 1 | K_n = n] = \tfrac{n+1}{n+2}.$$

Bedingte Dichten von P, gegeben $\{K_3 = k\}$, $k = 0, 1, 2, 3$:



Bedingte Dichten von P entlang zufälligem $(Z_1, Z_2, \ldots)$:

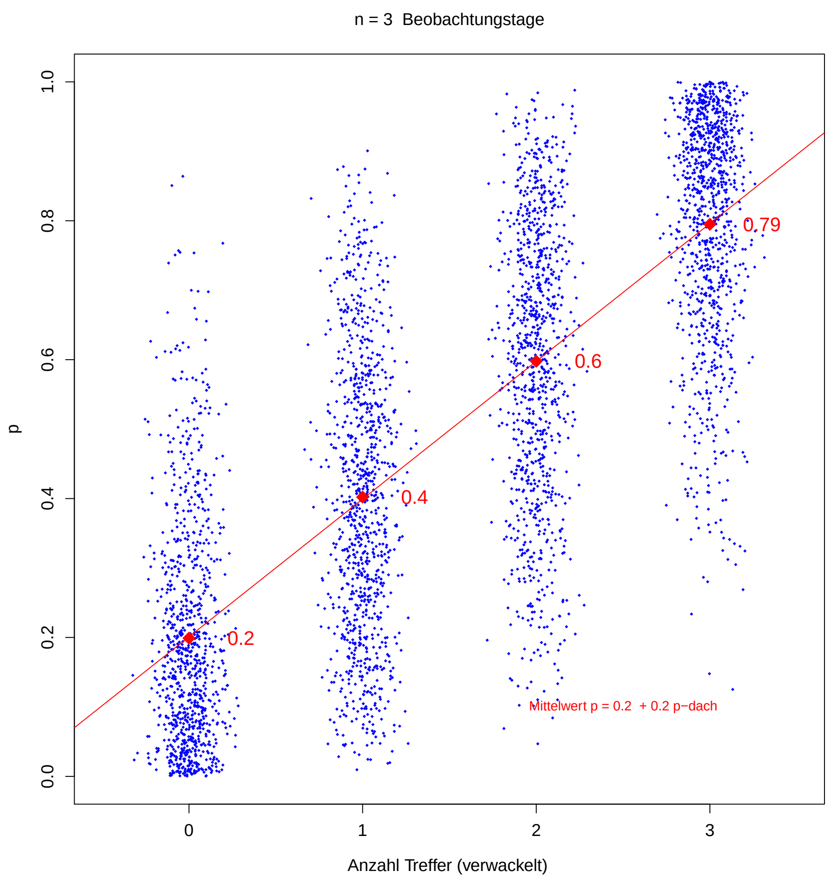


Unsere Frage 1b war:

$$\mathbf{E}[P|K_n = k] = ?$$

Emprischer Befund für $n = 3$:

$$\mathbf{E}[P|K_3 = k], \quad k = 0, 1, 2, 3$$



Vermutung:

$$\mathbf{E}[P|K_n = k] = \tfrac{k+1}{n+2}.$$

Eleganter Beweis über die U_i:

Gegeben sei, dass U_0 das k-t größte der $U_0, U_1, \ldots, U_n$ ist.

Dann gibt es für U_{n+1}

$k+1$ Slots links von U_0

und

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

Weil Frage 1a dieselbe Antwort hat wie Frage 1b, gilt:

$$\mathbf{P}[Z_{n+1} = 1 | K_n = k] = \tfrac{k+1}{n+2}.$$

Das lässt Erinnerungen wach werden ....

Wie ist Z_3 verteilt, gegeben $\{Z_2 = 2\}$? Auch das übersetzen wir 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, zwei in der Mitte, einer rechts) ein.

Gegeben $U_0 = \max(U_0, U_1, U_2)$
(gleichbedeutend damit: gegeben $\{Z_2 = 2\}$),
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:

$$\mathbf{P}(Z_3 = 1 | Z_2 = 2) = 3/4, \quad \mathbf{P}(Z_3 = 1 | Z_2 = 2) = 1/4.$$

Folgende einfache 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:

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

Wir erinnern hier an eine weitere unserer Übungsaufgaben:

44. U_0, U_1, U_2, U_3 seien unabhängig und uniform auf $[0, 1]$ verteilt.

a) Geben Sie eine Abbildung h von $[0, 1]^4$ auf die Menge $\{0, 1, 2, 3\}^4$ an, so dass $h(U_0, U_1, U_2, U_3)$ 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.

Fazit:

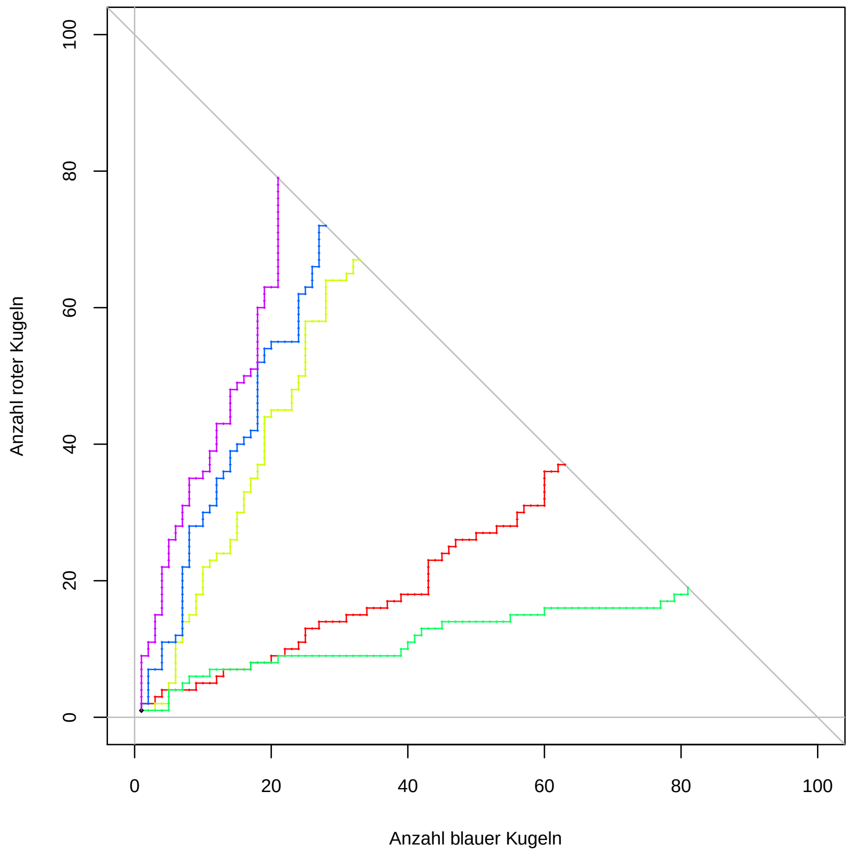
$(Z_1, Z_2, \ldots)$ ist so verteilt wie die Zuwächse eines zufälliger Pólya-Pfades,

d.h. wie die Farbfolge der Züge aus (bzw Zugänge in)

eine(r) Pólya-Urne

mit anfänglich einer roten und einer blauen Kugel.

5 Realisierungen von Pólya-Pfaden bis $n = 100$



$\frac{1}{n}(Z_1 + \cdots + Z_n)$ 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 bis $n = 1000$