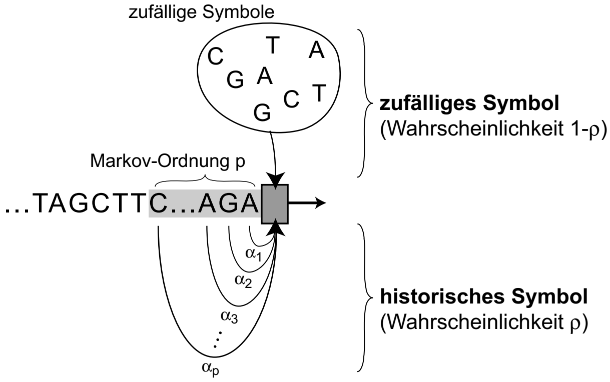

Abb. 4.6. Schematische Darstellung des DAR(p)-Prozesses aus Gleichung (4.16). Ein neues Symbol (Kasten am rechten Sequenzende) wird der Sequenz entweder durch Ziehen eines zufälligen Symbols (oberer Bildteil; Wahrscheinlichkeit $1-\rho$) oder durch Rückgriff auf ein Vorgängersymbol (unterer Bildteil; Wahrscheinlichkeit ρ) bestimmt. In diesem unteren Zweig geschieht mit der Wahrscheinlichkeit α_i ein Rückgriff um i Stellen. Die maximale Rückgriffweite ist durch die festgelegte Markov-Ordnung p gegeben

Wir haben zu Beginn dieses Kapitels Markov-Sequenzen eingeführt, um die Eigenschaften der Entropie und der Transinformation zu verdeutlichen. Wir vertauschen nun die Rollen und nutzen die Werkzeuge Entropie und Transinformation, um die Parameter des DAR(p)-Prozesses besser zu verstehen. Den generierten Sequenzen liegt dabei immer ein Alphabet mit $\lambda = 4$ Buchstaben zugrunde und wir nutzen eine Gleichverteilung als Marginalverteilung π.

Als Erstes untersuchen wir DAR(1)-Prozesse, deren Folge $\{X_n\}$ bestimmt wird durch

$$X_n = V_n X_{n-1} + (1 - V_n) Y_n , \qquad n = 1, 2, \ldots \tag{4.17}$$

und verallgemeinern die Ergebnisse dann auf DAR(p)-Prozesse mit $p > 1$. In Abb. 4.7 sieht man die bedingten Entropien h_n für verschiedene $\rho \in [0, 1]$. Die Sequenzlänge L beträgt mit 1.05×10^6 ungefähr λ^n mit $\lambda = 4$ und $n = 10$. Die Länge L der Sequenz ist damit so gewählt, dass sie gerade die Anzahl der verschiedenen n-Worte darstellt, die man (theoretisch) bei überlappender Zählung in der Sequenz finden kann. Man erkennt sofort die charakteristische Signatur einer Markov-Sequenz erster Ordnung im Maß der bedingten Entropie. Es ist der Knick bei h_1, also bei Erreichen der Ordnung $p = 1$. Die Folge h_n bleibt (theoretisch, also für unendlich lange Sequenzen) für alle $n \geq p = 1$ und festes ρ konstant. Der Parameter ρ erlaubt lediglich, die Unbestimmtheit des Prozesses zu variieren. Die Höhe des Plateaus (also den Wert von h_n bei großem n) bezeichnet man daher auch als die Entropie h des Prozesses. Die unterschiedlichen Plateaus der Folge h_n für $n \geq p = 1$ sind also nur ein Ausdruck der Stochastizität des Prozesses und stehen nicht direkt mit der Markov-Eigenschaft der Sequenz in Verbindung. Dabei bedeutet $\rho = 0$ maximale