Reward Variance in Markov Chains

Tom Verhoeff

Technische Universiteit Eindhoven

Faculty of Math. & Computing Science

Software Construction

`T.Verhoeff@TUE.NL`

`http://www.win.tue.nl/~wstomv/`

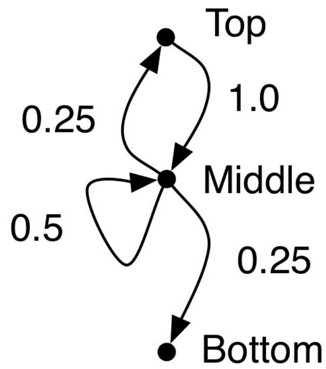
Spider on Cube

A spider walks randomly on the faces of a cube:

What is the expected time for the spider to get off the cube?

What is the corresponding variance?

Markov Chain



Finite state machine with transition probabilities and rewards

Expected Reward

Transition probabilities $p.s.t$ and rewards $r.s.t$ for states $s, t \in \Omega$

Lift to walks $w = s_0\, s_1\, \ldots\, s_n$

Probability of walk w: $P.w = p.s_0.s_1 * \cdots * p.s_{n-1}.s_n$

Reward of walk w: $R.w = r.s_0.s_1 + \cdots + r.s_{n-1}.s_n$

Walks from s to absorption in A: $W.s = \biguplus_{t \in \Omega} \{\, stv \mid tv \in W.t \,\}$ $(s \notin A)$

Expected reward from s to absorption in A:

$$\mathcal{E}_{W.s}[R] \;=\; \sum_{w \in W.s} P.w * R.w$$

Expected Walk Length for Spider

Obtain equations by generalizing and conditioning.

Expected walk lengths $\mu_s = \mathcal{E}_{W.s}[R]$ from face s to bottom:

$$\begin{aligned}
\mu_{\mathrm{T}} &= 1 + \mu_{\mathrm{M}} \\
\mu_{\mathrm{M}} &= 0.25\left(1 + \mu_{\mathrm{T}}\right) + 0.5\left(1 + \mu_{\mathrm{M}}\right) + 0.25\left(1 + \mu_{\mathrm{B}}\right) \\
\mu_{\mathrm{B}} &= 0
\end{aligned}$$

Solution:

$$\begin{aligned}
\mu_{\mathrm{T}} &= 6 \\
\mu_{\mathrm{M}} &= 5 \\
\mu_{\mathrm{B}} &= 0
\end{aligned}$$

Conditioning on First Transition to A

For $s \notin A$:

$$
\begin{array}{cl}
 & \mathcal{E}_{W.s}\left[X\right] \\
= & \quad \{ \text{ definition of } \mathcal{E} \text{ } \} \\
 & \sum_{w \in W.s} P.w * X.w \\
= & \quad \{ \text{ write } w = sv \text{ for } v \in W.t\text{, because } s \notin A \text{ } \} \\
 & \sum_{t \in \Omega} \sum_{v \in W.t} P.sv * X.sv \\
= & \quad \{ \text{ recurrence for walk probability: } P.sv = p.s.t * P.v \text{ for } v \in W.t \text{ } \} \\
 & \sum_{t \in \Omega} \sum_{v \in W.t} p.s.t * P.v * X.sv \\
= & \quad \{ \text{ distribute } p.s.t * \text{ outside } \sum_v \text{ } (p.s.t \text{ does not depend on } v) \text{ } \} \\
 & \sum_{t \in \Omega} p.s.t * \sum_{v \in W.t} P.v * X.sv \\
= & \quad \{ \text{ definition of } \mathcal{E} \text{ } \} \\
 & \mathcal{E}_{t \in \Omega}\left[\mathcal{E}_{v \in W.t}\left[X.sv\right]\right]
\end{array}
$$

Equations for Expected Reward until Absorption

For $s \notin A$:

$$
\begin{array}{cl}
 & \mathcal{E}_{W.s}[R] \\
= & \quad \{ \text{ conditioning on first state } t \text{ after state } s\text{, using } s \notin A \text{ } \} \\
 & \mathcal{E}_{t \in \Omega}\left[\mathcal{E}_{v \in W.t}\left[R.sv\right]\right] \\
= & \quad \{ \text{ recurrence for walk reward: } R.sv = r.s.t + R.v \text{ for } v \in W.t \text{ } \} \\
 & \mathcal{E}_{t \in \Omega}\left[\mathcal{E}_{v \in W.t}\left[r.s.t + R.v\right]\right] \\
= & \quad \{ \text{ linearity of expectation } (r.s.t \text{ is independent of } v) \text{ } \} \\
 & \mathcal{E}_{t \in \Omega}\left[r.s.t + \mathcal{E}_{v \in W.t}[R.v]\right] \\
= & \quad \{ \text{ simplify notation } \} \\
 & \mathcal{E}_{t \in \Omega}\left[r.s.t + \mathcal{E}_{W.t}[R]\right]
\end{array}
$$

Linear equations with unknowns $\mu_s = \mathcal{E}_{W.s}[R]$:

$$\mu_s \;=\; \sum_{t \in \Omega} p.s.t * \left(r.s.t + \mu_t\right)$$

Variance

Definition:

$$\mathcal{V}[X] \;=\; \mathcal{E}\left[(X - \mathcal{E}[X])^2\right]$$

Often computed via second moment:

$$
\begin{aligned}
\mathcal{V}[X] &= \mathcal{E}\left[(X - \mathcal{E}[X])^2\right] \\
&= \mathcal{E}\left[X^2 - 2X\mathcal{E}[X] + \mathcal{E}^2[X]\right] \\
&= \mathcal{E}\left[X^2\right] - 2\mathcal{E}[X]\mathcal{E}[X] + \mathcal{E}^2[X] \\
&= \mathcal{E}\left[X^2\right] - \mathcal{E}^2[X]
\end{aligned}
$$

Numerical disadvantage: loss of accuracy through cancellation

Basic Properties of Variance

A constant offset does not affect the variance:

$$\mathcal{V}[c + X] \;=\; \mathcal{E}\left[(c + X - \mathcal{E}[c + X])^2\right] \;=\; \mathcal{E}\left[(X - \mathcal{E}[X])^2\right] \;=\; \mathcal{V}[X]$$

Via second moment:

$$\mathcal{V}[c + X] = \mathcal{E}\left[(c + X)^2\right] - (c + \mathcal{E}[X])^2$$

Combine:

$$\mathcal{E}[(c + X)^2] \;=\; (c + \mathcal{E}[X])^2 + \mathcal{V}[X] \qquad (1)$$

Reward Variance on Walks until Absorption

For $s \notin A$, we calculate

$$
\begin{array}{cl}
 & \mathcal{V}_{W.s}[R] \\
= & \quad \{ \text{ definition of } \mathcal{V} \ \} \\
 & \mathcal{E}_{W.s}\left[(R - \mathcal{E}_{W.s}[R])^2\right] \\
= & \quad \{ \text{ conditioning on first state } t \text{ after state } s\text{, using } s \notin A \ \} \\
 & \mathcal{E}_{t\in\Omega}\left[\mathcal{E}_{v\in W.t}\left[\left(R.sv - \mathcal{E}_{W.s}[R]\right)^2\right]\right] \\
= & \quad \{ \text{ recurrence for reward: } R.sv = r.s.t + R.v \text{ for } v \in W.t \ \} \\
 & \mathcal{E}_{t\in\Omega}\left[\mathcal{E}_{v\in W.t}\left[\left(r.s.t + R.v - \mathcal{E}_{W.s}[R]\right)^2\right]\right] \\
= & \quad \{ \text{ (1), using that } r.s.t - \mathcal{E}_{W.s}[R] \text{ does not depend on } v \ \} \\
 & \mathcal{E}_{t\in\Omega}\left[\left(r.s.t + \mathcal{E}_{W.t}[R] - \mathcal{E}_{W.s}[R]\right)^2 + \mathcal{V}_{W.t}[R]\right]
\end{array}
$$

Equations for Reward Variance

System of linear equations with unknowns $\sigma_s^2 = \mathcal{V}_{W.s}[R]$ for $s \in \Omega$,

involving $\mu_s = \mathcal{E}_{W.s}[R]$ as parameters:

$$\boxed{\sigma_s^2 = \sum_{t\in\Omega} p.s.t * \left(\left(r.s.t + \mu_t - \mu_s\right)^2 + \sigma_t^2\right)}$$

Variance in Walk Length for Spider

Variance in walk length $\sigma_s^2 = \mathcal{V}_{W.s}[R]$ from face s to bottom:

$$
\begin{array}{rcl}
\sigma_{\mathsf{T}}^2 & = & \left(1 + \mu_{\mathsf{M}} - \mu_{\mathsf{T}}\right)^2 + \sigma_{\mathsf{M}}^2 \\
\sigma_{\mathsf{M}}^2 & = & 0.25\left(\left(1 + \mu_{\mathsf{T}} - \mu_{\mathsf{M}}\right)^2 + \sigma_{\mathsf{T}}^2\right) + \\
 & & 0.5\ \left(\left(1 + \mu_{\mathsf{M}} - \mu_{\mathsf{M}}\right)^2 + \sigma_{\mathsf{M}}^2\right) + \\
 & & 0.25\left(\left(1 + \mu_{\mathsf{B}} - \mu_{\mathsf{M}}\right)^2 + \sigma_{\mathsf{B}}^2\right) \\
\sigma_{\mathsf{B}}^2 & = & 0
\end{array}
$$

Solution:

$$
\begin{array}{rcl}
\sigma_{\mathsf{T}}^2 & = & 22 \\
\sigma_{\mathsf{T}} & \approx & 4.69
\end{array}
$$

Conclusion

- Importance of the variance
- Simple direct formula to calculate variance in reward until absorption in Markov chain
- Numerically attractive
- Applied to score variance of optimal strategy for Yahtzee
  Acyclic Markov chain with $\approx 10^9$ states