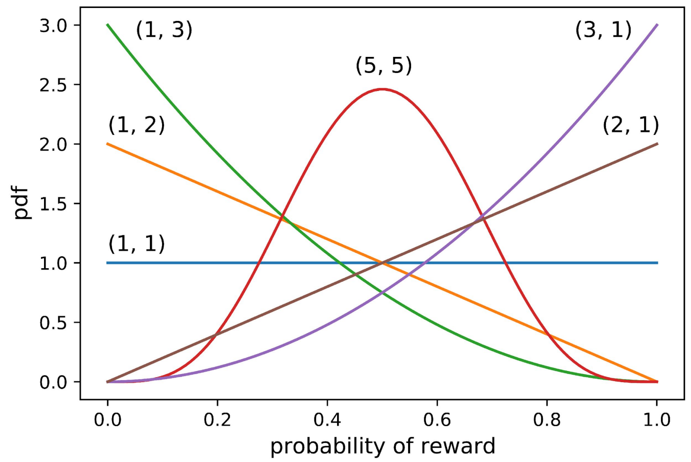
Gittins indices for Bernoulli bandits

Beta priors become Beta posteriors



Fun fact: $1+\gamma+\gamma^2+\ldots=\frac{1}{1-\gamma}$

Bellman's solution - brute force via backward recursion

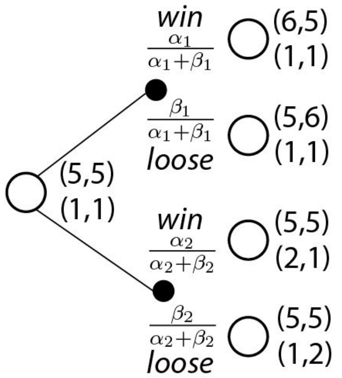
optimal choice value

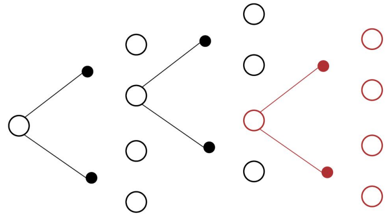
$$R(\alpha_1,\beta_1,\alpha_2,\beta_2)=\max\{\tfrac{\alpha_1}{\alpha_1+\beta_1}[1+\gamma R(\alpha_1+1,\beta_1,\alpha_2,\beta_2)]$$

$$+\tfrac{\beta_1}{\alpha_1+\beta_1}[0+\gamma R(\alpha_1,\beta_1+1,\alpha_2,\beta_2)],$$

$$\tfrac{\alpha_2}{\alpha_2+\beta_2}[1+\gamma R(\alpha_1,\beta_1,\alpha_2+1,\beta_2)]$$

$$+\tfrac{\beta_2}{\alpha_2+\beta_2}[0+\gamma R(\alpha_1,\beta_1,\alpha_2,\beta_2+1)]\}$$





Number of operations (N is horizon with n arms):

$$\frac{(N-1)!}{(2n)!(N-2n-1)!}$$

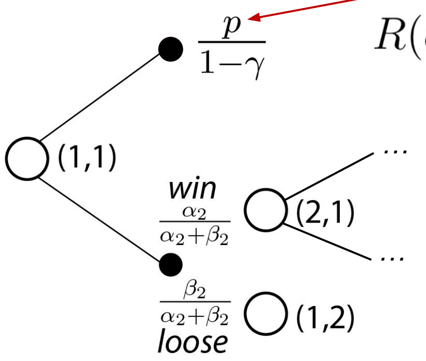
Full backward recursion is expensive for many arms

Gittins indeces from Gittins '11:

Table 8.7 Bernoulli reward process, index values, $a = 0.9$.

β \ α	1	2	3	4	5	6	7	8	9	10
1	.7029	.8001	.8452	.8723	.8905	.9039	.9141	.9221	.9287	.9342
2	.5001	.6346	.7072	.7539	.7869	.8115	.8307	.8461	.8588	.8695
3	.3796	.5163	.6010	.6579	.6996	.7318	.7573	.7782	.7956	.8103
4	.3021	.4342	.5184	.5809	.6276	.6642	.6940	.7187	.7396	.7573
5	.2488	.3720	.4561	.5179	.5676	.6071	.6395	.6666	.6899	.7101
6	.2103	.3245	.4058	.4677	.5168	.5581	.5923	.6212	.6461	.6677
7	.1815	.2871	.3647	.4257	.4748	.5156	.5510	.5811	.6071	.6300
8	.1591	.2569	.3308	.3900	.4387	.4795	.5144	.5454	.5723	.5960
9	.1413	.2323	.3025	.3595	.4073	.4479	.4828	.5134	.5409	.5652
10	.1269	.2116	.2784	.3332	.3799	.4200	.4548	.4853	.5125	.5373

Gittins's solution – find a bribe for each arm that makes you indifferent to play



$$R(\alpha,\beta,p)=\max\{\tfrac{p}{1-\gamma},\tfrac{\alpha}{\alpha+\beta}[1+\gamma R(\alpha+1,\beta,p)]$$

$$+\tfrac{\beta}{\alpha+\beta}[0+\gamma R(\alpha,\beta+1,p)]\}$$

Number of operations for 1 arm: $\frac{1}{2}(N-1)(N-2)$