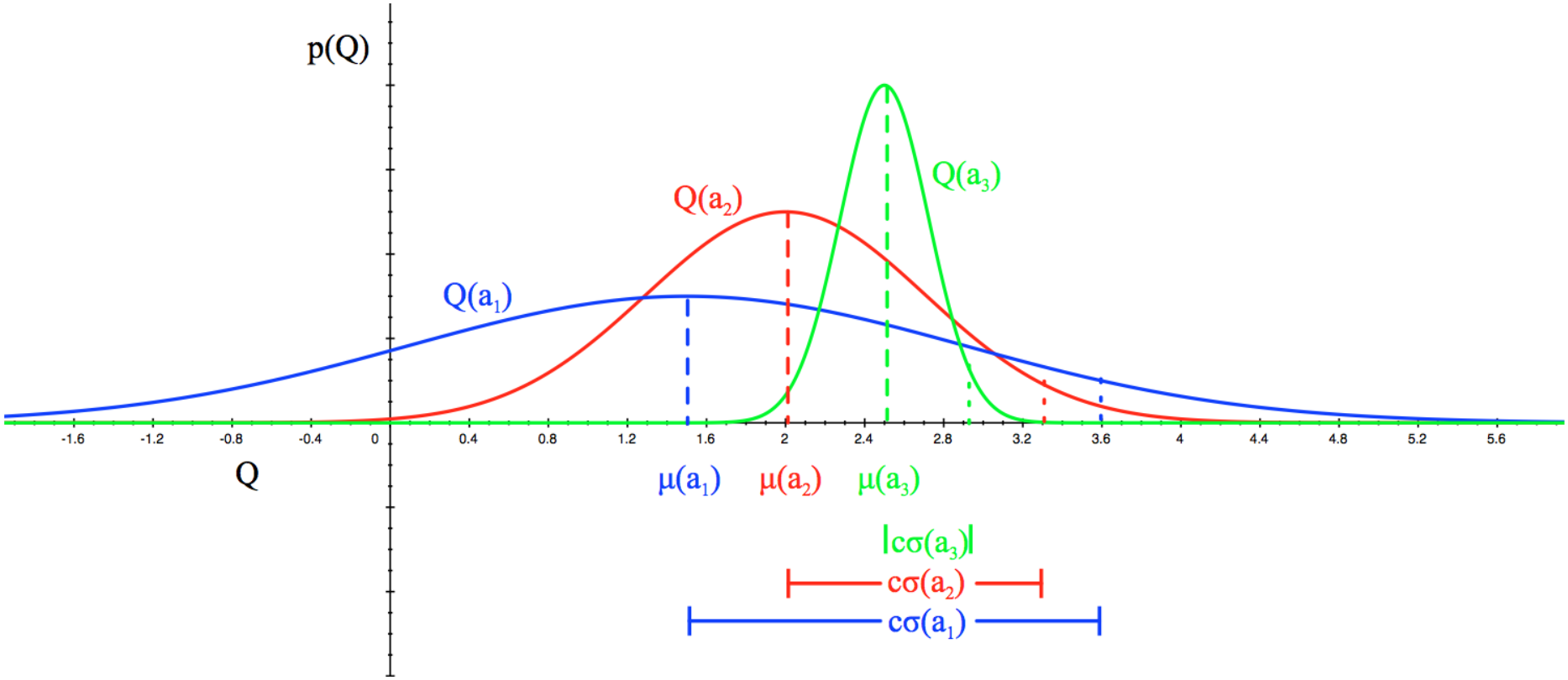

p(Q)
Q(a1)
Q(a2)
Q(a3)
-1.6
-1.2
-0.8
-0.4
0
0.4
0.8
1.2
1.6
2
2.4
2.8
3.2
3.6
4
4.4
4.8
5.2
5.6
Q
μ(a1)
μ(a2)
μ(a3)
|cσ(a3)|
cσ(a2)
cσ(a1)

[Silver, 2020]

Valor esperado

$$E[X] = \sum_{i=1}^{\infty} x_i p(x_i)$$

Variância

$$\operatorname{Var}(X) = \operatorname{E}\left[(X - \mu)^2\right]$$

Desvio padrão

$$\sigma := \sqrt{\operatorname{Var}[X]}$$

Distribuição de Boltzmann

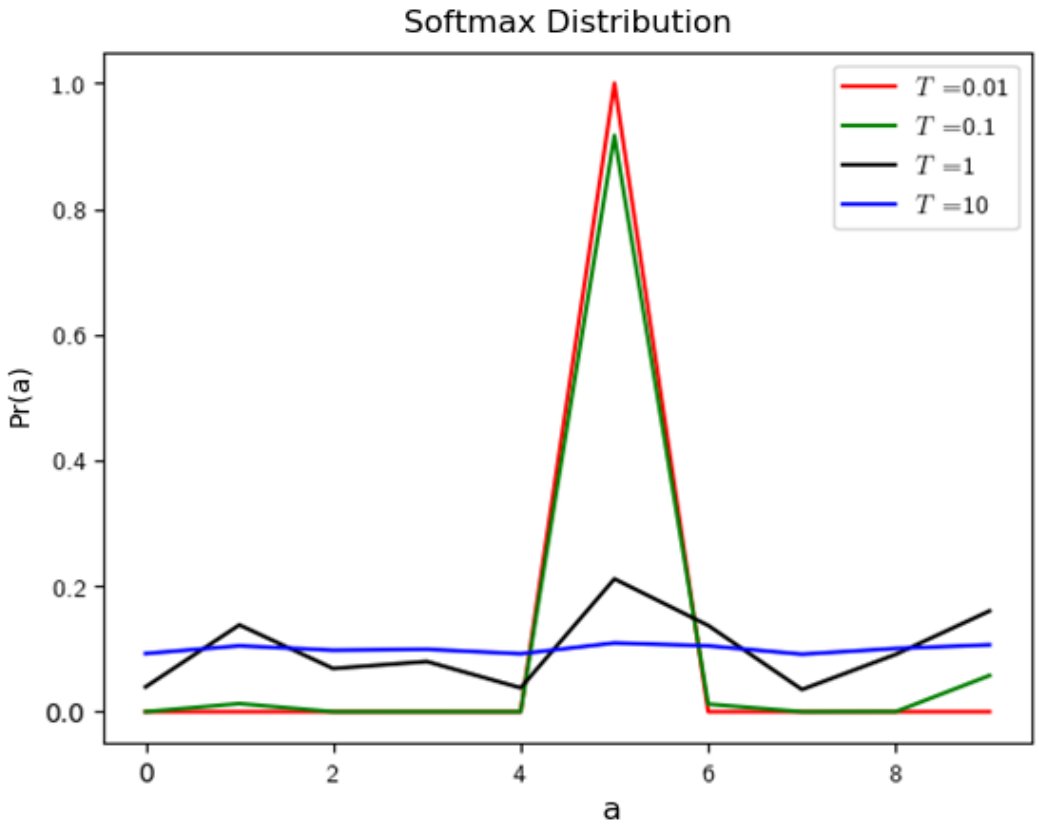
$$p_i \propto e^{-\frac{\varepsilon_i}{kT}}$$

$$p_i = \frac{e^{-\varepsilon_i/kT}}{\sum_{j=1}^{M} e^{-\varepsilon_j/kT}}$$

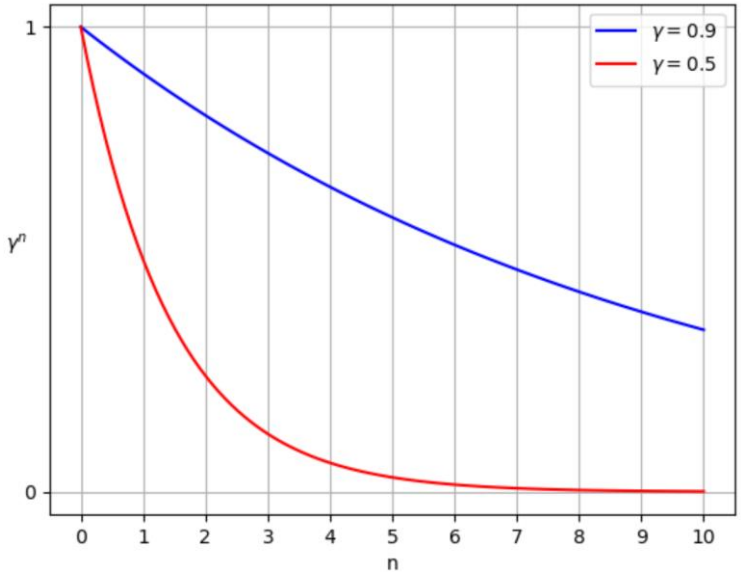
p_i - probabilidade do sistema estar no estado *i*
ε_i - energia do estado
k - constante de Boltzmann
T - temperatura
M - número de estados

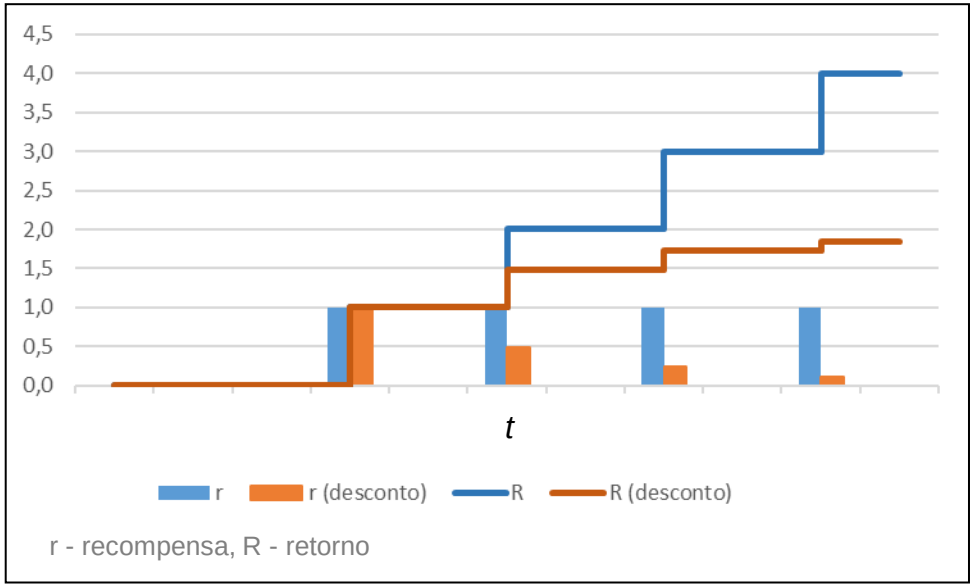
Distribuição *Soft-max*

$$\Pr\{A_t = a\} = \frac{e^{Q_t(a)/\tau}}{\sum_{b=1}^{k} e^{Q_t(b)/\tau}}$$



γ - Factor de desconto





REFERÊNCIAS

[Deisenroth *et al.*, 2020]
M. Deisenroth, A. Faisal, C. Ong, “Mathematics For Machine Learning”, Cambridge University Press, 2020

[Silver, 2020]
D. Silver, UCL Course on RL, *https://www.davidsilver.uk/teaching*, 2020