SHUBHAM THAKRAL, 2017193

Date: / /
Page No.

Q3. Exercice 2.7

$\bar{o}_n = \bar{o}_{n-1} + \alpha(1-\bar{o}_{n-1})$, $n \geq 0$, $\bar{o}_0 = 0$

$\beta_n = \frac{\alpha}{\bar{o}_n}$

$\to \bar{o}_n = \alpha + (1-\alpha)\bar{o}_{n-1}$

$\Rightarrow \bar{o}_n - \alpha = (1-\alpha)\bar{o}_{n-1}$ ——— (1)

Given step size as β_n:

$$
\begin{aligned}
Q_n &= Q_{n-1} + \beta_n(R_n - Q_{n-1}) \\
&= \beta_n R_n + (1-\beta_n)Q_{n-1} \\
&= \beta_n R_n + (1-\beta_n)(Q_{n-2} + \beta_{n-1}(R_{n-1} - Q_{n-2})) \\
&= \beta_n R_n + (1-\beta_n)(\beta_{n-1}R_{n-1} + (1-\beta_{n-1})Q_{n-2}) \\
&= \beta_n R_n + (1-\beta_n)\beta_{n-1}R_{n-1} + (1-\beta_n)(1-\beta_{n-1})Q_{n-2} \\
&\vdots \\
&= Q_0 \cdot \prod_{i=0}^{n-1}(1-\beta_{n-i}) + \sum_{i=0}^{n-1}\left(\beta_{n-i} \cdot R_{n-i} \cdot \prod_{j=0}^{i-1}(1-\beta_{n-j})\right)
\end{aligned}
$$

To prove that is has no bias on initial choice of Q_0, we've to prove that coeff. of $Q_0 = 0$.

$$
\begin{aligned}
\prod_{i=0}^{n-1}(1-\beta_{n-i}) &= \prod_{i=0}^{n-1}\left(1 - \frac{\alpha}{\bar{o}_{n-i}}\right) = \prod_{i=0}^{n-1}\left(\frac{\bar{o}_{n-i} - \alpha}{\bar{o}_{n-i}}\right) \\
&= \prod_{i=0}^{n-1}\frac{(1-\alpha)\bar{o}_{n-i-1}}{\bar{o}_{n-i}} \quad \text{(from (1))} \\
&= \frac{(1-\alpha)\bar{o}_{n-1}}{\bar{o}_n} \times \frac{(1-\alpha)\bar{o}_{n-2}}{\bar{o}_{n-1}} \times \cdots \times \frac{(1-\alpha)\bar{o}_0}{\bar{o}_1} \\
&= 0 \quad (\text{as } \bar{o}_0 = 0)
\end{aligned}
$$

So this method has no initial bias.

Let's now simplify reward term:

$$\sum_{i=0}^{n-1} R_{n-i}\left(\beta_{n-i}\prod_{j=0}^{i-1}(1-\beta_{n-j})\right);$$

Simplifying this term:

$$= \frac{\alpha}{\bar{o}_{n-i}} \times \left(1-\frac{\alpha}{\bar{o}_n}\right) \times \left(1-\frac{\alpha}{\bar{o}_{n-1}}\right) \times \cdots \times \left(1-\frac{\alpha}{\bar{o}_{n-i+1}}\right)$$

Using ① :

$$= \cancel{\frac{\alpha}{\bar{o}_{n-i}}} \times \frac{(1-\alpha)\cancel{\bar{o}_{n-i}}}{\cancel{\bar{o}_{n-i+1}}} \times \cdots \times \frac{(1-\alpha)\cancel{\bar{o}_{n-1}}}{\bar{o}_n}$$

$$= \frac{\alpha(1-\alpha)^{i}}{\bar{o}_n}$$

$$\sum_{i=0}^{n-1} \frac{\alpha(1-\alpha)^{i}}{\bar{o}_n} R_{n-i}$$

$$\bar{o}_n = \alpha + (1-\alpha)\bar{o}_{n-1} = \alpha + (1-\alpha)\left(\bar{o}_{n-2} + \alpha(1-\bar{o}_{n-2})\right)$$

$$= \alpha + (1-\alpha)\left(\alpha + (1-\alpha)\bar{o}_{n-2}\right)$$

$$= \alpha + \alpha(1-\alpha) + (1-\alpha)^2\bar{o}_{n-2}$$

$$\vdots$$

$$= \alpha + \alpha(1-\alpha) + \alpha(1-\alpha)^3 + \cdots + \alpha(1-\alpha)^{n-1}$$

$$= \cancel{\alpha} \cdot \frac{\left(1-(1-\alpha)^n\right)}{\cancel{\left(1-(1-\alpha)\right)}} = 1-(1-\alpha)^n$$

$$\sum_{i=0}^{n-1} \frac{\alpha\cdot(1-\alpha)^{i}}{1-(1-\alpha)^n} \cdot R_{n-i}$$

~~Hence rewards are exponentially recency weighted average without initial bias.~~

Date : / /
Page No.

Now we just have to prove that ~~it~~ it is weighted average

i.e. $\displaystyle\sum_{i=0}^{n-1} \frac{\alpha \cdot (1-\alpha)^i}{1-(1-\alpha)^n} = 1$

To prove $\Rightarrow \displaystyle\sum_{i=0}^{n-1} \alpha \cdot (1-\alpha)^i = 1-(1-\alpha)^n$

Let's expand this :

$\alpha + \alpha(1-\alpha) + \alpha(1-\alpha)^2 + \dots + \alpha(1-\alpha)^{n-1}$

$= \dfrac{\cancel{\alpha}\left(1-(1-\alpha)^n\right)}{\cancel{(1-(1-\alpha))}}$ (using GP formula)

$= 1-(1-\alpha)^n$

Hence Proved.

Hence the method makes Q_n as exponential recency weighted average without initial bias.