**p. 107** – Lo de que *“gradually shifted closer and closer to a deterministic optimal policy”* solo ocurriría (que no lo dice) si cuanto

En (5.2) la notación no es la más apropiada. Ese toma la forma de una ya que puede ser cualquier acción (pues se trata de un *e-soft policy*). Si bien es cierto que te lo compara con el *policy improvement theorem* de la p.83, no llega a ser lo mismo ya que en la p.83, la policy era determinista (y por tanto sí tenía sentido hablar de un ).

**p.108** – Este desarrollo no lleva a nada. La explicación que hace desde el final de la p. 107 (“*Consider a new environment that is just like the original environment, except with the requirement that policies be e-soft “moved inside” the environment.*”) hasta el final de la p.108 parece recursiva e incompleta. No aporta nada. ¿qué significa moved inside the environment? Creo que la idea que intenta transmitir no está muy bien aclarada.

**p.112** – “*Thus the value of s under the target policy is 1*” ¿debería ser 0.1?