Problema Grupal

# 3. La fàbrica de píndoles I

**La companyia farmacèutica Smart Pills ha construït una cinta transportadora que porta dues classes de píndoles (adequades per dos tipus de malalties diferents), que anomenem C1 i C2. Aquestes píndoles surten en un ombrejat de colors que va del *yellow* al *white* (que és detectat per una càmera, donant un valor continu en [0, 2]). La companyia fabrica píndoles en proporcions p(C1) = 1/3 , p(C2) = 2/3. Se’ns facilita també informació sobre la distribució (contínua) del color per cada classe:**

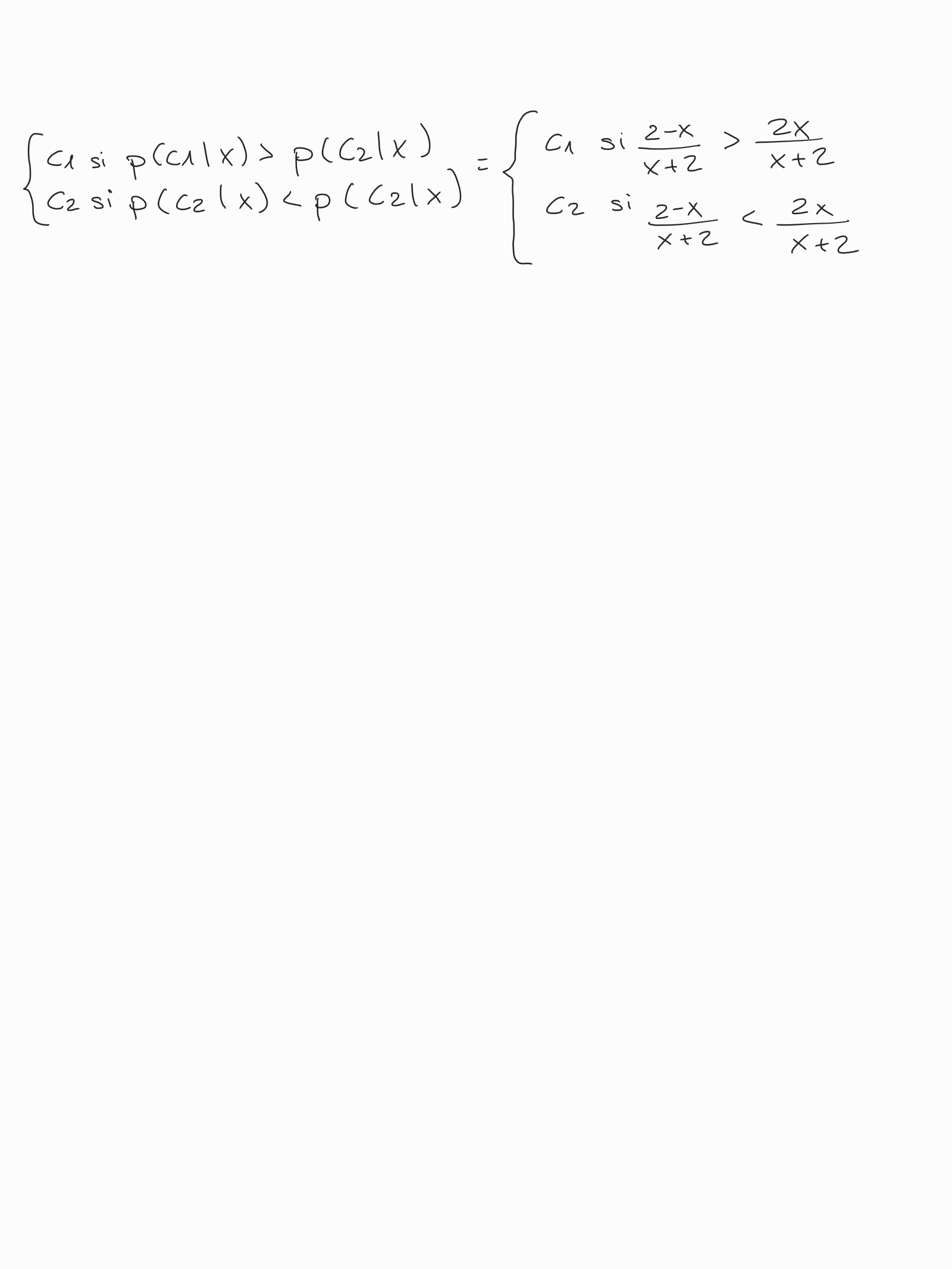


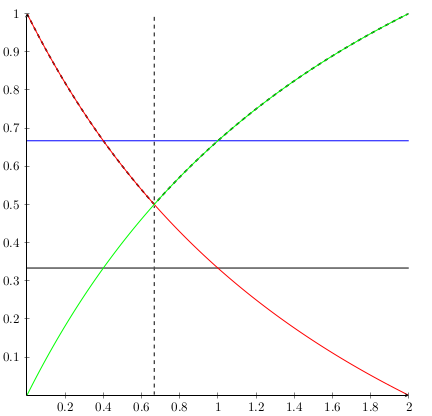
* **(a) Quina és la probabilitat d’error si no s’utilitza el color per classificar?**

Com que la probabilitat de C2 és més alta, si no classifiquem per colors podem considerar que totes les píndoles són de C2. I per tant la probabilitat d’error és que surti de C1:

* **(b) Calcular la distribució incondicional del color p(x) = p(C1) p(x|C1) + p(C2) p(x|C2).**
* **(c) Calcular les distribucions de probabilitat p(C1|x) i p(C2|x).**
* **(d) Quina és la classificació òptima en funció del color?**

La regla per a decidir la classificació òptima en funció del color seria:





**Llegenda:**

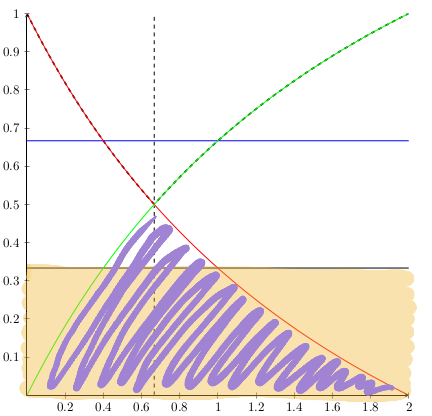
* corba verda: p(C2|x)
* corba vermella: p(C1|x)
* recta blava: p(C2)
* recta negra: p(C1)
* **(e) Quina és la probabilitat d’error si s’utilitza el color per classificar? Perquè és millor que la del primer apartat?**

Si classifiquem per colors, es donaria una probabilitat d’error de:

min

Per a poder calcular l’error total, primer hem de buscar el punt d'intersecció entre les probabilitats de C1 i C2:

Llavors podem integrar la corbes per a trobar l’error quan s’utilitza la classificació per colors:



**Llegenda:**

* franja taronja: p error (Rcolor)
* franja rallada lila: p error (Rclasse)

Si el comparem amb l’error total quan no s’utilitza el color:

Veiem que l’error total és inferior quan utilitzem colors per classificar. 0.3333333 (nomes classes) > 0.2222222 (amb colors)