Probabilités conditionnelles
Problème de recherche : le filtre anti-spam

Une entreprise développe un filtre anti-spam pour sa messagerie. Elle a analysé un grand nombre de courriels reçus :

— 60 % des courriels reçus sont des spams ;
— le mot « gratuit » apparaît dans 30 % des spams ;
— le mot « gratuit » apparaît dans 2 % des courriels qui ne sont pas des spams.

La mission

1. Un courriel contient le mot « gratuit ». Quelle est la probabilité que ce soit un spam ?
2. Un courriel ne contient pas le mot « gratuit ». Quelle est la probabilité que ce soit un spam ?
3. Le filtre doit-il envoyer automatiquement dans les spams tous les courriels qui contiennent le mot « gratuit » ? Discuter.

Pour aller plus loin

Une autre entreprise reçoit seulement 10 % de spams (les autres pourcentages sont les mêmes). La présence du mot « gratuit » est-elle encore un bon indice ?

Coups de pouce

Bloqué ? Des indices progressifs sont disponibles en ligne. N'en ouvrir qu'un à la fois, et seulement après avoir cherché !