M1-ISL : option Initiation à la Recherche.

Protection d'une réponse individuelle. Analyse d'une proposition et extensions.

Jean-François Couchot

Université de Franche-Comté, UFR-ST

Plan

Idée de réponse randomisée

Evaluation de la proposition

Plan

Idée de réponse randomisée

Evaluation de la proposition

Sondage avec une question embarrassante [1]

Obj. : publier le pourcentage des étudiants ayant triché au moins une fois.

- Question Q_1 : « Avez-vous triché au moins une fois durant vos études ? »
- Embarras : tentation pour un étudiant de ne pas répondre honnêtement.

1. https ://fr.coursera.org/lecture/stanford-statistics/warners-randomized-response-model-ck65q

Méthode de Warner [2]

Rappel

Question Q_1 : « Avez-vous triché au moins une fois durant vos études ? »

Extension de la question

- Chaque étudiant lance 2 fois une pièce de monnaie $\{\text{Pile}, \text{Face}\}$ sans montrer les 2 résultat successif t_1 et t_2.
- Ajout de la question Q_2 : « Est-que t_2 est égal à Pile ? ».
  - Si t_1 vaut Pile, l'étudiant répond honnêtement à la question Q_1.
  - Sinon ($t_1 = \text{Face}$), l'étudiant répond honnêtement à la question Q_2

Analyse de l'extension

- Réponse partiellement aléatoire : on ne sait pas si une réponse OUI d'un étudiant provient d'une tricherie estudiantine ou d'un Pile au second tirage.
- Cela devrait aider l'étudiant à répondre honnêtement.

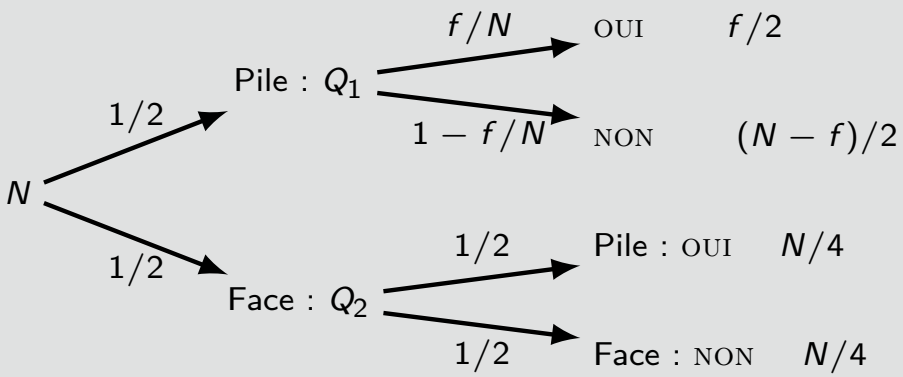
2. Warner, S. L. (1965). Randomized response : A survey technique for eliminating evasive answer bias. Journal of the American Statistical Association, 60(309), 63-69.

Estimation du pourcentage de tricheurs

Point clef

- Un OUI *individuel* : on ne connaît pas exactement son origine.
- Après calcul du pourcentage *global* des OUI : on doit pouvoir estimer le pourcentage des étudiants ayant triché au moins une fois.

Estimateur pour N personnes et $f < N$ tricheurs



- Nombre de OUI : $r \approx N/4 + f/2$
- Estimation $\hat{f}$ du nbre. original de OUI : $\hat{f} = 2r - N/2$
- Xpl. : #OUI=27, #NON=30.
  - $\hat{f} = 2 \times 27 - \frac{57}{2} \approx 25.5$.

Plan

Plan

Estimateur non biaisé ?

En moyenne, l'estimateur $\hat{f}$ doit retourner la bonne réponse.

- Evaluation pratique du calcul de la moyenne de $\hat{f}$ sur des données connues.
  - $\rightsquigarrow$ Nécessite de coder la proposition.
  - $\rightsquigarrow$ Nécessite des données.
  - $\rightsquigarrow$ Nécessite de répéter l'expérience un grand nombre de fois.
- Evaluation théorique de l'espérance de l'estimateur $\hat{f}$.
  - $\rightsquigarrow$ Nécessite quelques compétences en stats/proba

Estimateur avec dispersion faible ?

La variance de p devrait être réduite.

- Evaluation pratique de la variance de $\hat{f}$.
  - $\rightsquigarrow$ Mêmes prérequis que plus haut.
- Evaluation théorique de la variance de $\hat{f}$.
  - $\rightsquigarrow$ Mêmes prérequis que plus haut.

Plan

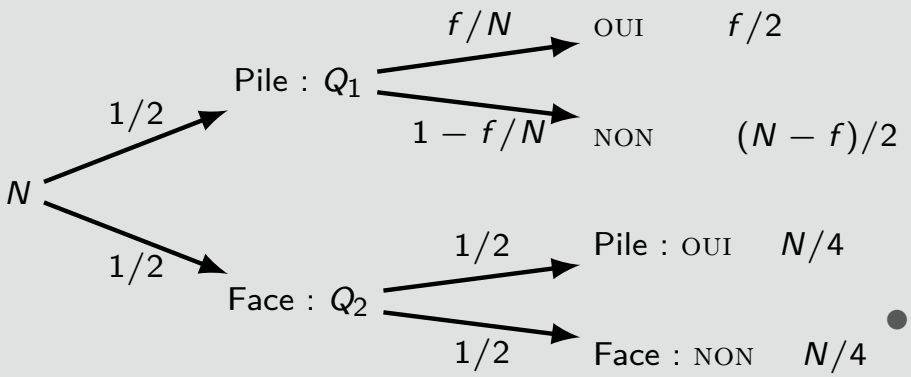
Mesure probabiliste de la fuite d'information

Formalisation probabiliste : théorie

- $\mathcal{M} : \{\text{OUI}, \text{NON}\} \to \{\text{OUI}, \text{NON}\}$: mécanisme de réponse randomisée
- Étant donnée une sortie y de ce mécanisme, quelle est la probabilité qu'elle soit l'image de x_1, de x_2 ?
- Quelle est la variation maximale entre ces 2 probabilité ?

$$\frac{\Pr[\mathcal{M}(x_1) = y]}{\Pr[\mathcal{M}(x_2) = y]} \text{ pour } x_1, x_2, r \in \{\text{OUI}, \text{NON}\}$$

Formalisation probabiliste : résultats



x \ y	OUI	NON
OUI	3/4	1/4
NON	1/4	3/4

- $\frac{\Pr[\mathcal{M}(x_1)=y]}{\Pr[\mathcal{M}(x_2)=y]} \leq \frac{\Pr[\mathcal{M}(\text{OUI})=\text{OUI}]}{\Pr[\mathcal{M}(\text{OUI})=\text{OUI}]} \leq 3$

Plan

Modification : choix et conséquences

Pourquoi lancer une pièce ?

Modification : choix et conséquences

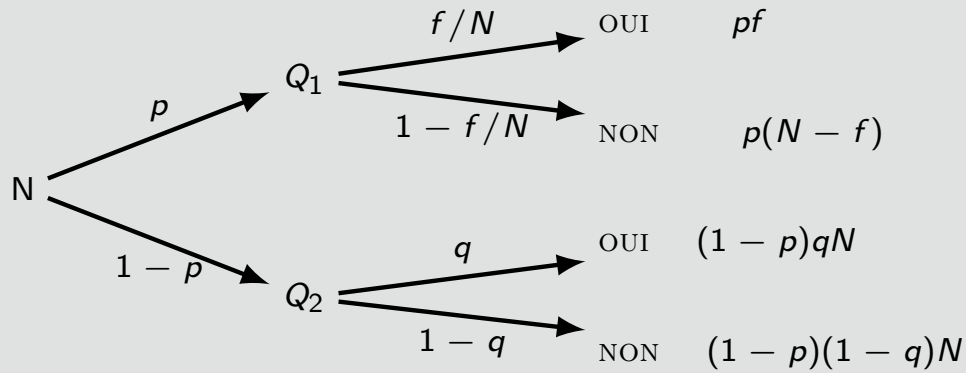
Pourquoi lancer une pièce ?

- Modification des probabilités dans l'arbre

Modification : choix et conséquences

Pourquoi lancer une pièce ?

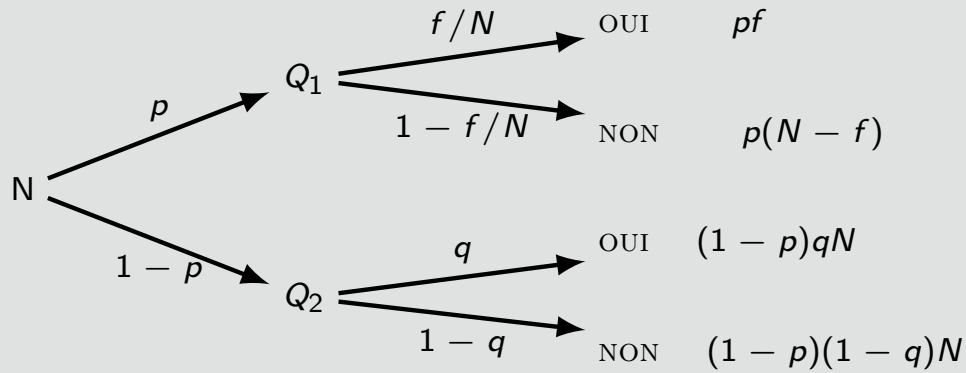
- Modification des probabilités dans l'arbre



Modification : choix et conséquences

Pourquoi lancer une pièce ?

- Modification des probabilités dans l'arbre

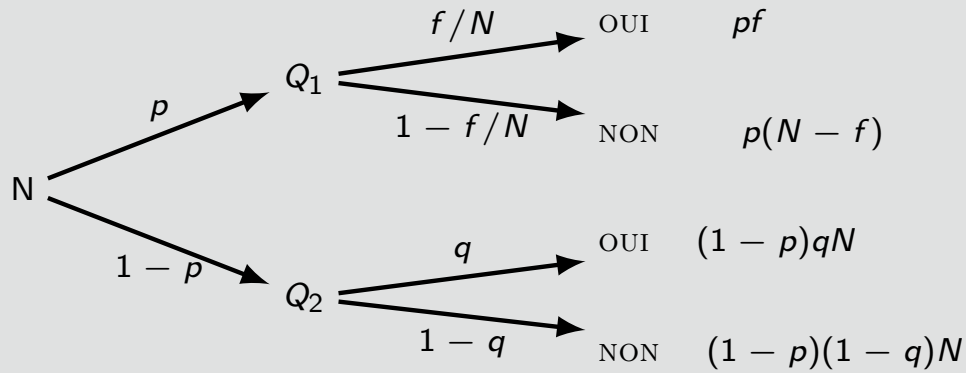


- Conséquences sur l'estimateur $\hat{f}$

Modification : choix et conséquences

Pourquoi lancer une pièce ?

- Modification des probabilités dans l'arbre

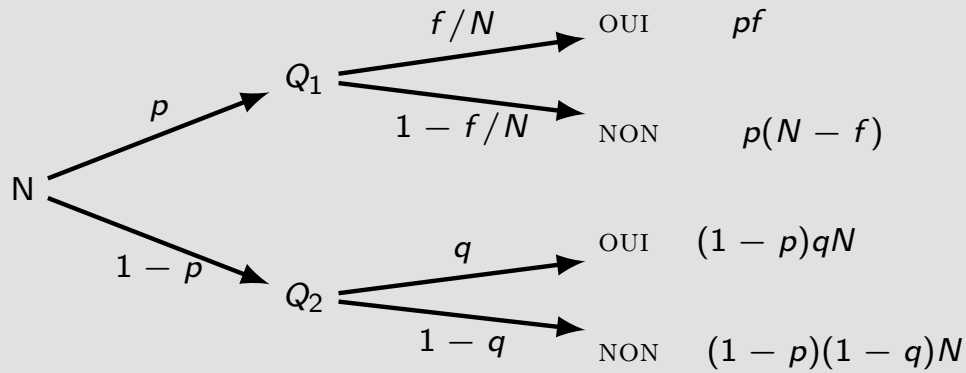


- Conséquences sur l'estimateur $\hat{f}$
  - Nécessaire de le redéfinir.

Modification : choix et conséquences

Pourquoi lancer une pièce ?

- Modification des probabilités dans l'arbre

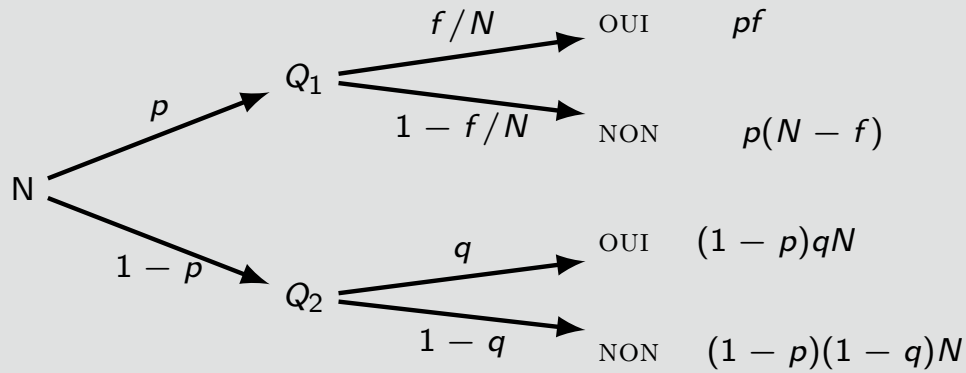


- Conséquences sur l'estimateur $\hat{f}$
  - Nécessaire de le redéfinir.
  - Nécessaire de vérifier qu'il est non biaisé.

Modification : choix et conséquences

Pourquoi lancer une pièce ?

- Modification des probabilités dans l'arbre

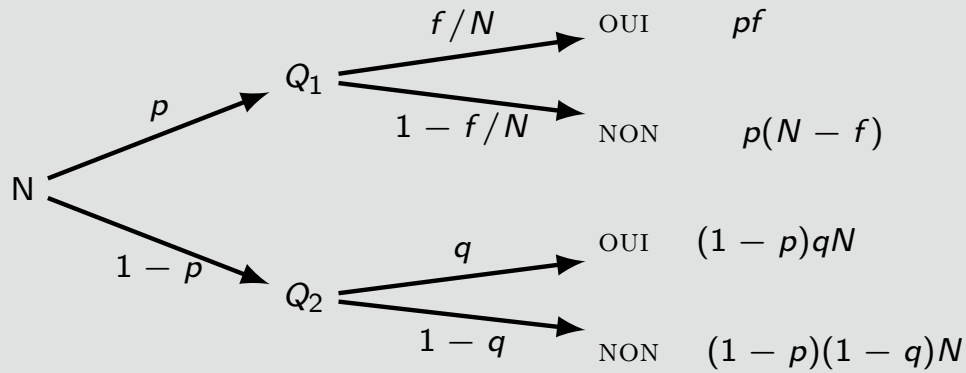


- Conséquences sur l'estimateur $\hat{f}$
  - Nécessaire de le redéfinir.
  - Nécessaire de vérifier qu'il est non biaisé.
  - Nécessaire de recalculer sa variance.

Modification : choix et conséquences

Pourquoi lancer une pièce ?

- Modification des probabilités dans l'arbre



- Conséquences sur l'estimateur $\hat{f}$
  - Nécessaire de le redéfinir.
  - Nécessaire de vérifier qu'il est non biaisé.
  - Nécessaire de recalculer sa variance.
  - Quelles sont les valeurs de p et q qui minimisent la variance ?