Que déduire d'un générateur d'aléa qui passe des tests statistiques avec succès ?

Guenaëlle De Julis

API Hour #13

UNIVERSITÉ DE GRENOBLE

coffreo
créateur de domiciles numériques

Aléa en cryptographie

- **Utilisation massive en cryptographie** :
  $\rightarrow$ besoin d'une graine, d'un token, d'un masque aléatoire, de padding, d'une clef, ...
- **Deux types** de sources brutes sont distinguées :
  - déterministes (DRBG) :
    $\rightarrow$ graine + algorithme
  - non déterministes (NDRBG) :
    $\rightarrow$ phénomène physique + mécanisme d'extraction

Aléa et signature ECDSA

La signature a besoin, entre autres, d'un aléa r et d'une clef privée k.

- **Aléa et sécurité** de la signature :
  $\rightarrow$ si r est connu, alors k est compromise
  $\rightarrow$ si le même r est utilisé pour signer 2 messages différents, alors r peut être facilement calculé
- **Sony PS3** (2010)
  $\rightarrow$ r était toujours le même ...
  $\rightarrow$ forger des signatures de jeux valides
- **Bitcoin, Java PRNG** (2013)
  $\rightarrow$ trop de collisions, donc des r prédictibles
  $\rightarrow$ vol de compte

Comment évaluer une source d'aléa ?

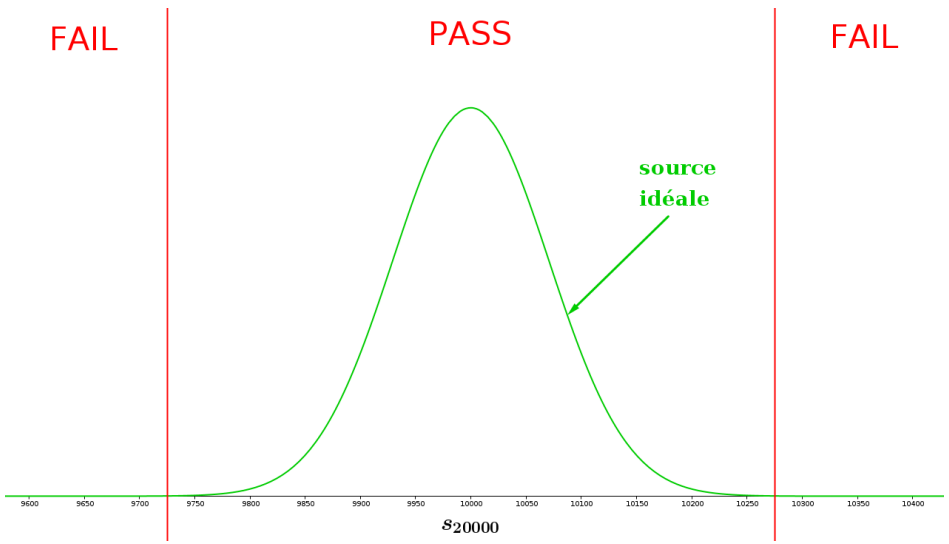
- Bonne ou mauvaise source ?
  $\rightarrow$ si j'observe une suite de 50 zéros ?
  $\rightarrow$ si je n'observe jamais 50 zéros à la suite ?
- source idéale : « jeu de pile ou face infini,
  avec pièce équilibrée et lancers indépendants »
- par séries de tests statistiques :
  - standards NIST : FIPS 140-2 (2002), SP800-22(2010),
  - standard BSI (2011) : AIS31,
- évaluations «en aveugle»
  $\rightarrow$ modèle de la source en entrée inconnu
  $\rightarrow$ obj : donner un degré de vraisemblance de l'entrée par rapport à la source idéale.

Tests d'hypothèse : théorie

	Entrée		**Sortie**
Données :	$X_1, \ldots, X_n$ n variables aléatoires	$\rightarrow$ fonction $\rightarrow$	S_n (statistique de test)
Théorie :	source idéale	$\Rightarrow$	S_n de loi $\mathcal{D}_S$
Utilisation :	$b_1, \ldots, b_n$ n observations		s_n **(s-valeur)**

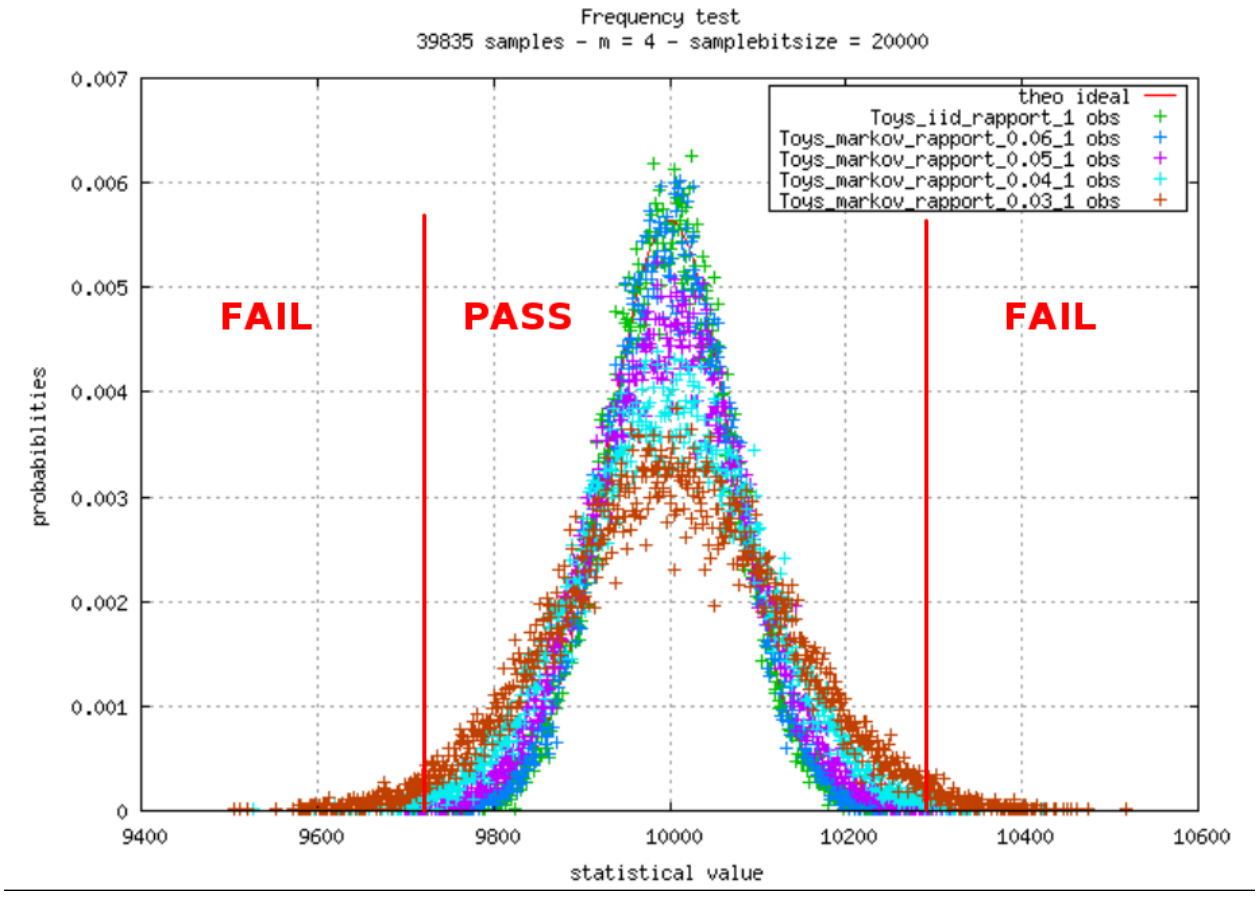
Tests d'hypothèse : pratique

1. Comment interpréter s_n
   pour décider si la source en entrée est idéale ?
   $\rightarrow$ PASS/FAIL par zone de rejet $+$ taux de réussite
2. Exemple : test de fréquence sur $n = 20\,000$ bits
   - entrée : $b_1 \ldots b_{20\,000}$
   - sortie : $s_{20\,000}$ = nombre de $'1'$ $\in [0, 20\,000]$
   - décision pratiquée :
     PASS si $s_{20\,000} \in [9725\ ,\ 10\,275]$,
     FAIL sinon
   - ce que dit la théorie pour une source idéale en entrée :
     la sortie suit la loi binomiale $B\left(20\,000, \frac{1}{2}\right)$.



Escroquerie au test de fréquence #1

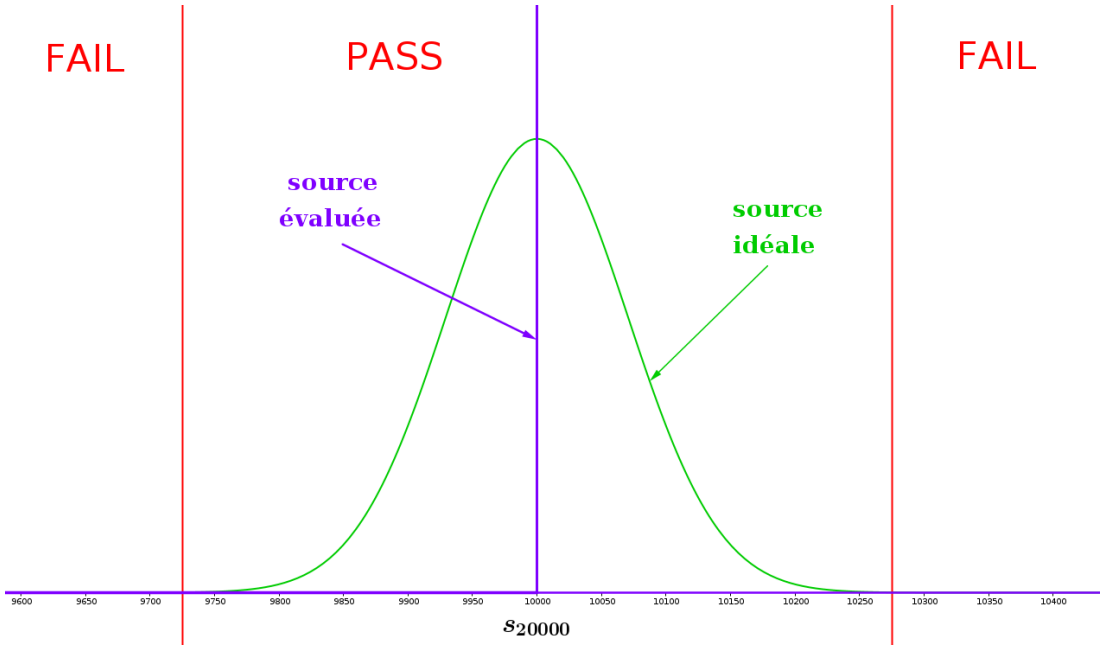
Source évaluée, non idéale : la pièce est équilibrée mais les lancers ne sont pas indépendants.



→ **taux de réussite élevé : la zone de rejet n'est pas un critère suffisant**

Escroquerie au test de fréquence #2

Source évaluée, non idéale : $\underbrace{0\ldots0}_{10000\ \textit{fois}}\ \overbrace{1\ldots1}^{10000\ \textit{fois}}\ \underbrace{0\ldots0}_{10000\ \textit{fois}}\ \ldots$

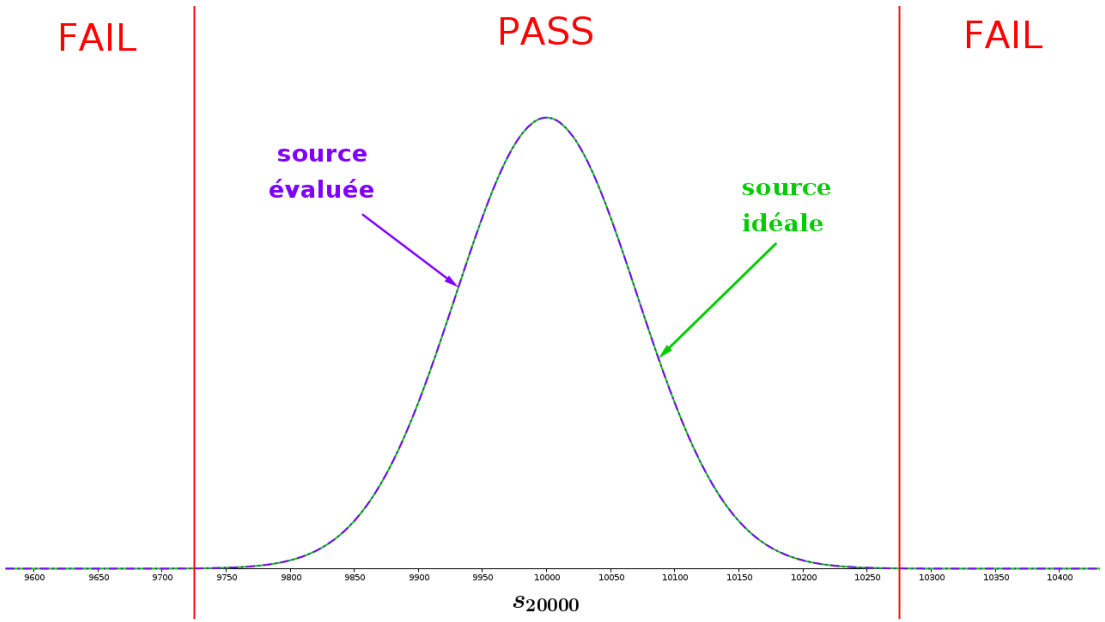


→ **taux de réussite** 100% **: la zone de rejet n'est pas un critère pertinent**

Escroquerie au test de fréquence #3

Source non idéale : lancers indépendants par groupe de m bits avec $m+1$ motifs au lieu des 2^m possibles selon la règle

- $0 \ldots 0 \rightarrow$ probabilité $\frac{1}{2^m}$
- $0 \ldots 01 \rightarrow$ probabilité $k \times \frac{1}{2^m}$
- $0 \ldots 011 \rightarrow$ probabilité $\binom{m}{2} \times \frac{1}{2^m}$
- $0 \ldots 0111 \rightarrow$ probabilité $\binom{m}{3} \times \frac{1}{2^m}$
- $\ldots$



$\rightarrow$ loi en sortie identique à la loi attendue

Que déduire d'un RNG qui «passe FIPS 140-2» ?

- Pas grand chose . . .
  $\rightarrow$ il existe des sources non idéales en entrée qui se comporte comme la source idéale en sortie
  $\rightarrow$ les zones de rejets ne sont pas suffisantes
  $\rightarrow$ les taux de réussite ne garantissent rien

- Les risques viennent de la façon d'interpréter la sortie :
  - quelques valeurs ne représentent pas la loi en sortie
  - la théorie dit « $A \Rightarrow B$ » et non « $B \Rightarrow A$ » !

Merci pour votre attention