Case Study: Robust Randomness Extraction from a 40-Card Deck (Uniform 128-bit Key)

Author: Jesús Garví | Github | Linkedin | Contact: jesus.garvigu@gmail.com

Date: 08/14

---

1. Problem Statement

This note measures information content from a single physical source: draws from a 40-card Spanish deck with replacement. We compute bits per draw, the total entropy across repeated draws, and the mathematically correct method for mapping these outcomes to a uniform 128-bit value. We also show why the case without replacement requires more draws to reach the same entropy threshold.

---

2. Method & Model

For a system with Ω equally likely outcomes, the Shannon entropy per event is:

$$H = \log_2(\Omega)$$

Across *n* independent draws with replacement:

$$M_{\text{with}} = \Omega^n, \quad H_{\text{total, with}} = n\log_2(\Omega)$$

Across *n* draws without replacement (where order matters), the total state space is the number of n-permutations from Ω symbols:

$$M_{\text{wo}} = P(\Omega, n) = \frac{\Omega!}{(\Omega - n)!}, \quad H_{\text{total, wo}} = \log_2\left(\frac{\Omega!}{(\Omega - n)!}\right)$$

To generate a uniform 128-bit value from a physical outcome X ∈ [0, M−1], we use two extractors:

1. Rejection sampling (information-theoretic):
   Define

   $$T = \left\lfloor \frac{M}{2^{128}} \right\rfloor \cdot 2^{128}$$

   If X < T, accept and output Y = X mod 2^{128}; otherwise, resample. This yields an exactly uniform key on [0, 2^{128}−1].

2. Cryptographic hash (pragmatic):
   Serialize the draw sequence, hash with SHA-256, and truncate to 128 bits (standard PRF/KDF practice).

3. Implementation uses arbitrary-precision integers and a vetted OS-level CSPRNG; both are recommended for correctness and security.

---

3. Results & Diagnostics

The analysis is focused on a 40-card deck.

- Entropy per draw: $\log_2(40) \approx 5.3219$ bits.
- 25 draws, with replacement: $H_{\text{total}} \approx 25 \cdot \log_2(40) \approx 133.048$ bits. Rejection sampling acceptance probability ≈ 0.99736 (99.736%).
- 25 draws, without replacement: $H_{\text{total}} \approx \log_2\big(P(40{,}25)\big) \approx 118.909$ bits — insufficient for uniform 128-bit extraction.
- Minimum draws, without replacement: A minimum of n=28 is required to exceed 128 bits; the acceptance probability is ≈ 0.99885 (99.885%).

Context (entropy per event):

- Coin: 1.000 bits
- Die (6-sided): 2.585 bits
- 40-card Deck: 5.322 bits
- 52-card Deck: 5.700 bits

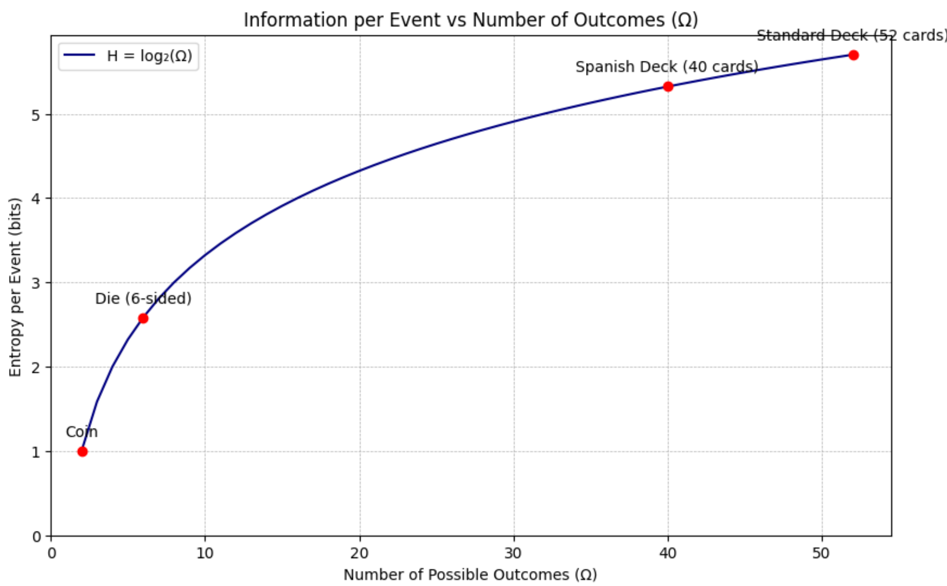

Figure 1 — Entropy per event vs number of outcomes. Entropy scales logarithmically in Ω; larger outcome spaces deliver more bits per draw.

---

4. Quant Takeaway

- Signal Quality: Entropy checks assumptions (uniformity, independence), but predictive quality hinges on relevant information (e.g., correlation or mutual information with the target) and out-of-sample information coefficient — not on raw source entropy.
- Model Assumptions: Equiprobability and independence for draws "with replacement" mirror core quantitative assumptions like stationarity or no data leakage. Violations invalidate the entropy calculation and any downstream guarantees.
- Extraction Correctness: The extractor is as critical as the source. A biased mapping can destroy an otherwise sound signal — the data pipeline and mapping logic must be provably correct.

---

5. Repository & Full Code

The full, reproducible Jupyter notebook (analysis, visualization, and key-derivation implementation) is available for verification.

---

6. References

1. Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3), 379–423.
2. National Institute of Standards and Technology (NIST). Special Publication 800-90A Rev. 1: *Recommendation for Random Number Generation Using Deterministic Random Bit Generators*.
3. National Institute of Standards and Technology (NIST). Special Publication 800-90B: *Recommendation for the Entropy Sources Used for Random Bit Generation*.

Appendix A: Reproducibility Log (Simulated Draws)

Simulations used an OS-level CSPRNG to model physical draws and confirmed theoretical calculations.

- Source: 40-symbol deck, with replacement, n=25 draws.
    - Total Entropy: ≈ 133.048 bits (≥ 128 bits).
    - Theoretical Acceptance Probability: ≈ 0.99736.
- Source: 40-symbol deck, without replacement, minimum n=28 draws.
    - Total Entropy: ≈ 135.535 bits (≥ 128 bits).
    - Theoretical Acceptance Probability: ≈ 0.99885.