COPING WITH THE STOCHASTIC BIOMETRICS

Tomáš Rosa

Raiffeisenbank, a.s.

SIGNALS PRIMER

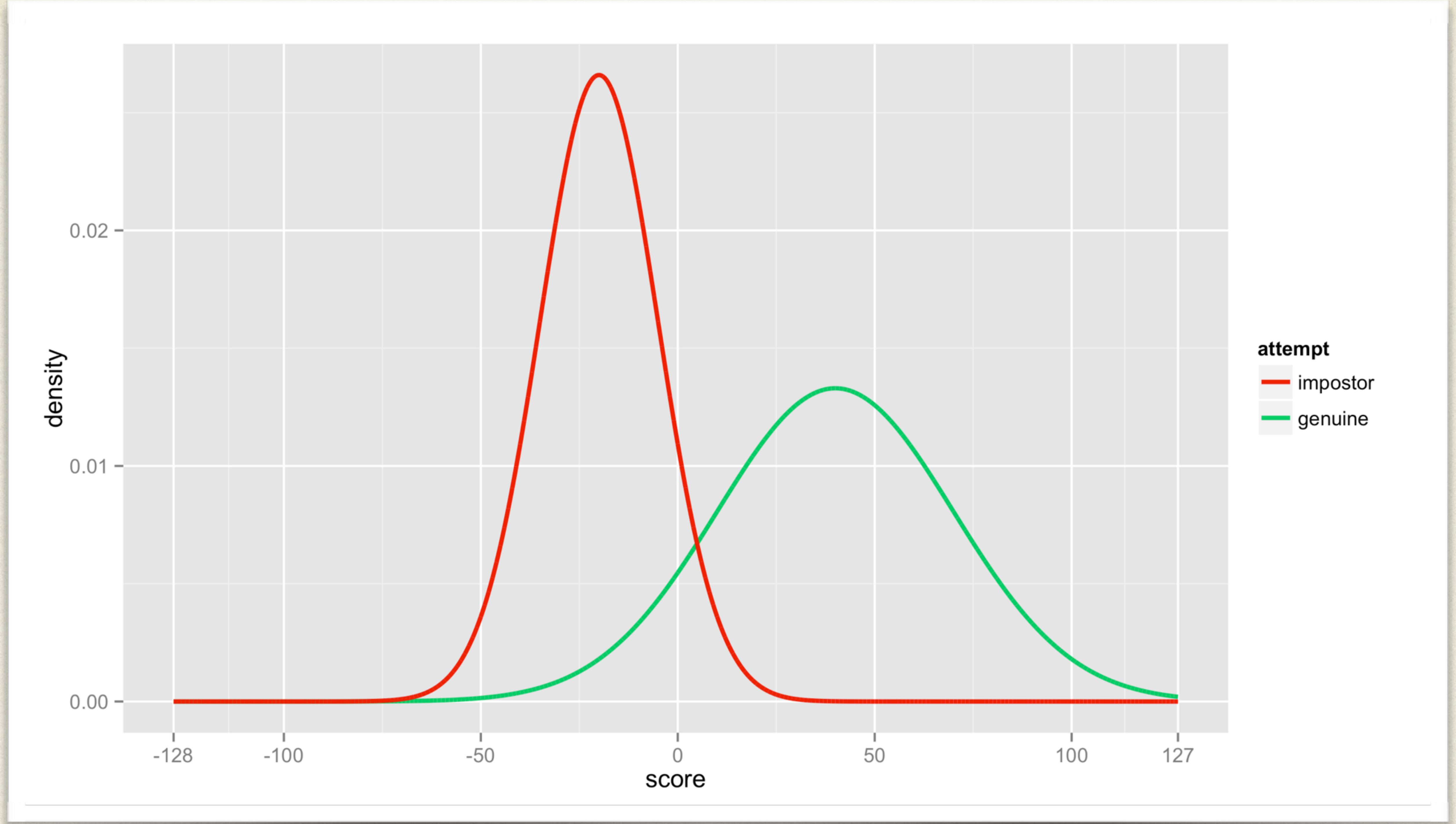
- Let a signal be any measurable space-time varying quantity conveying information about a physical phenomena.
- Signal detection is then an ability to discern between information-bearing patterns (signals) and random patterns (noise) that distract from the information.

MATCH SCORE

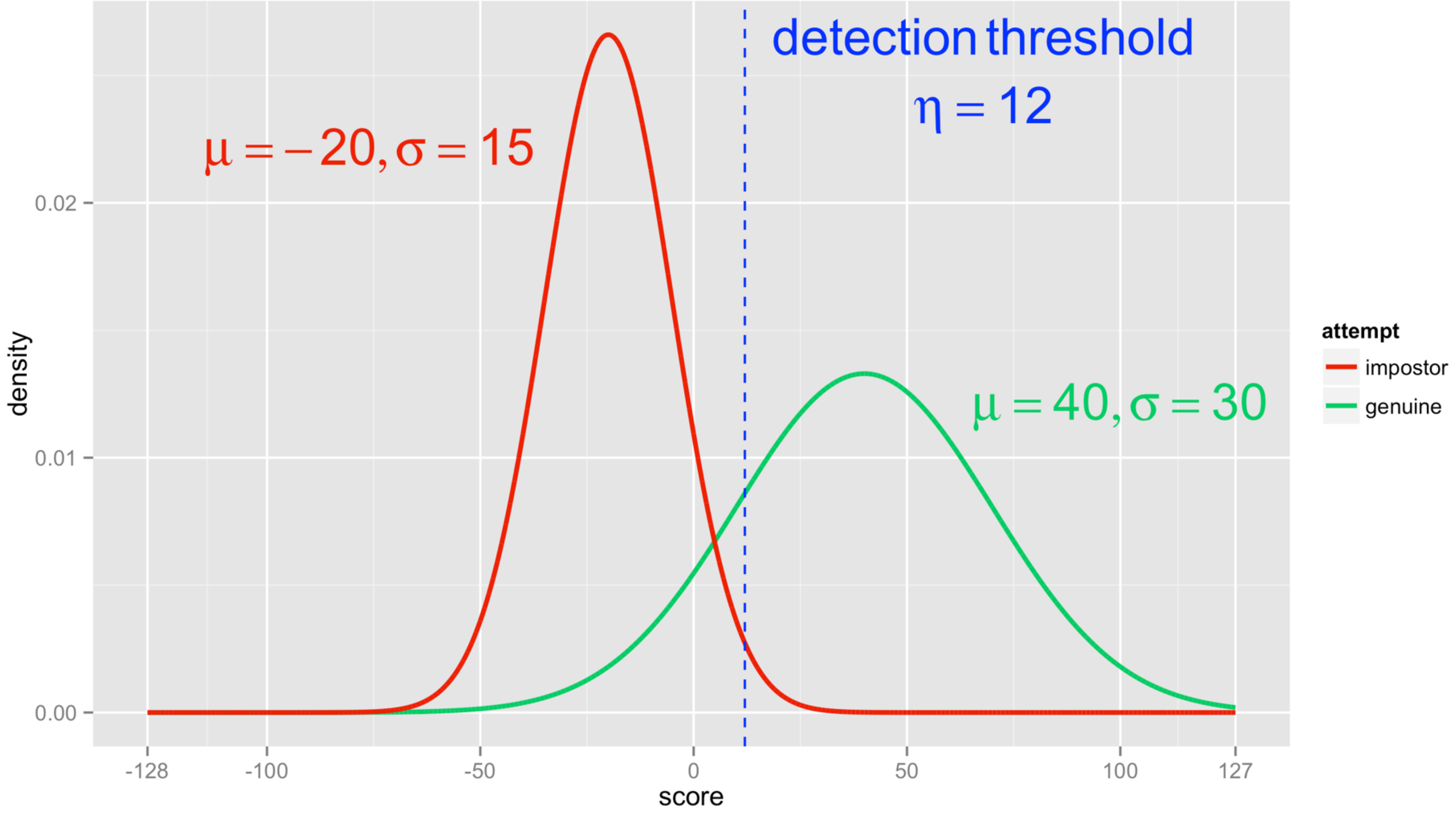
- It would be nice if we had a simple true-false result.
  - As in conventional crypto.
  - But we cannot...

- All we have is a value of random variable X that follows two conditional distributions.
  - $f(x \mid \text{impostor})$
  - $f(x \mid \text{genuine})$

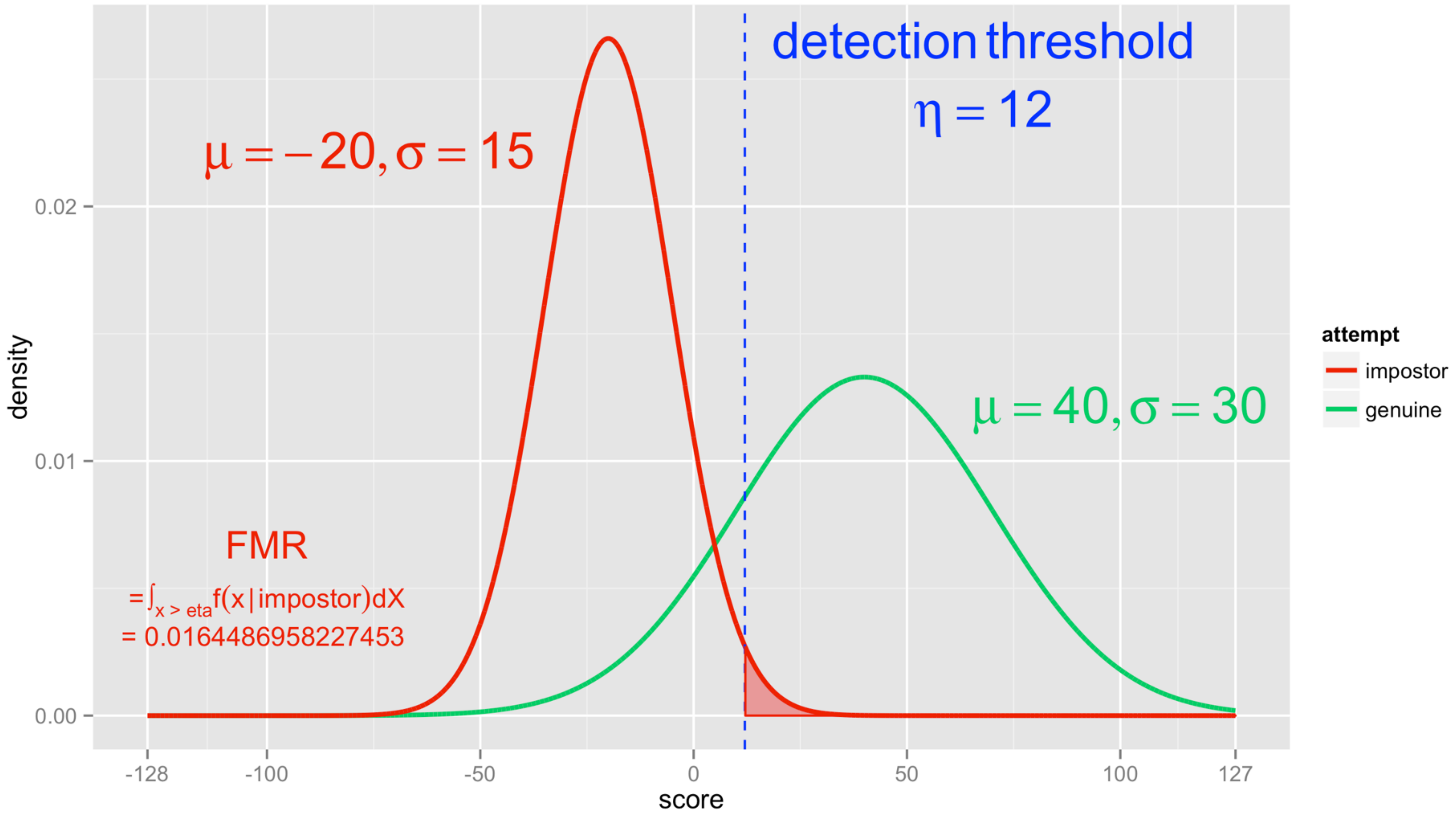
BASE "CAMEL" GRAPH



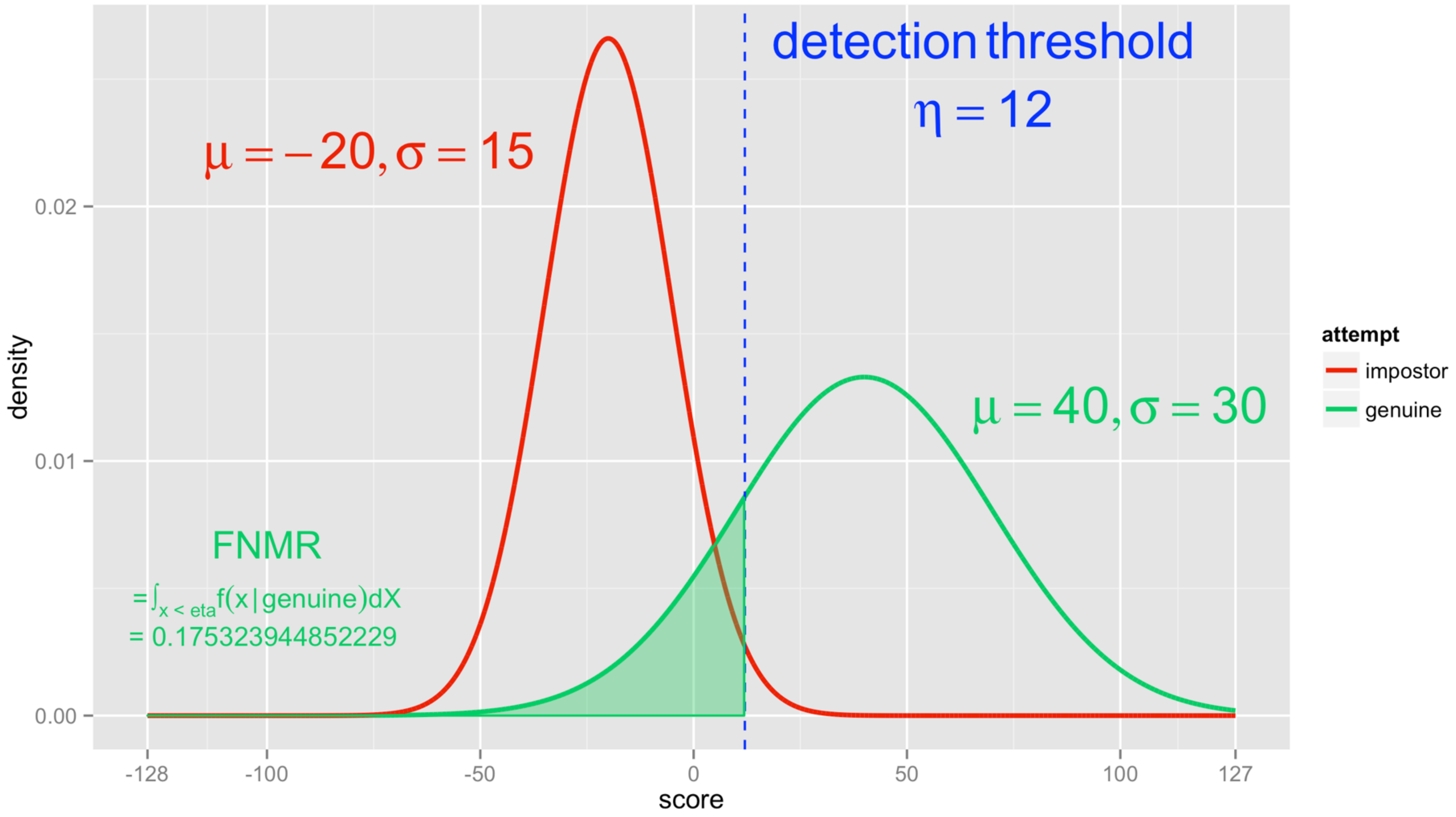
SIGNAL DETECTION APPROACH



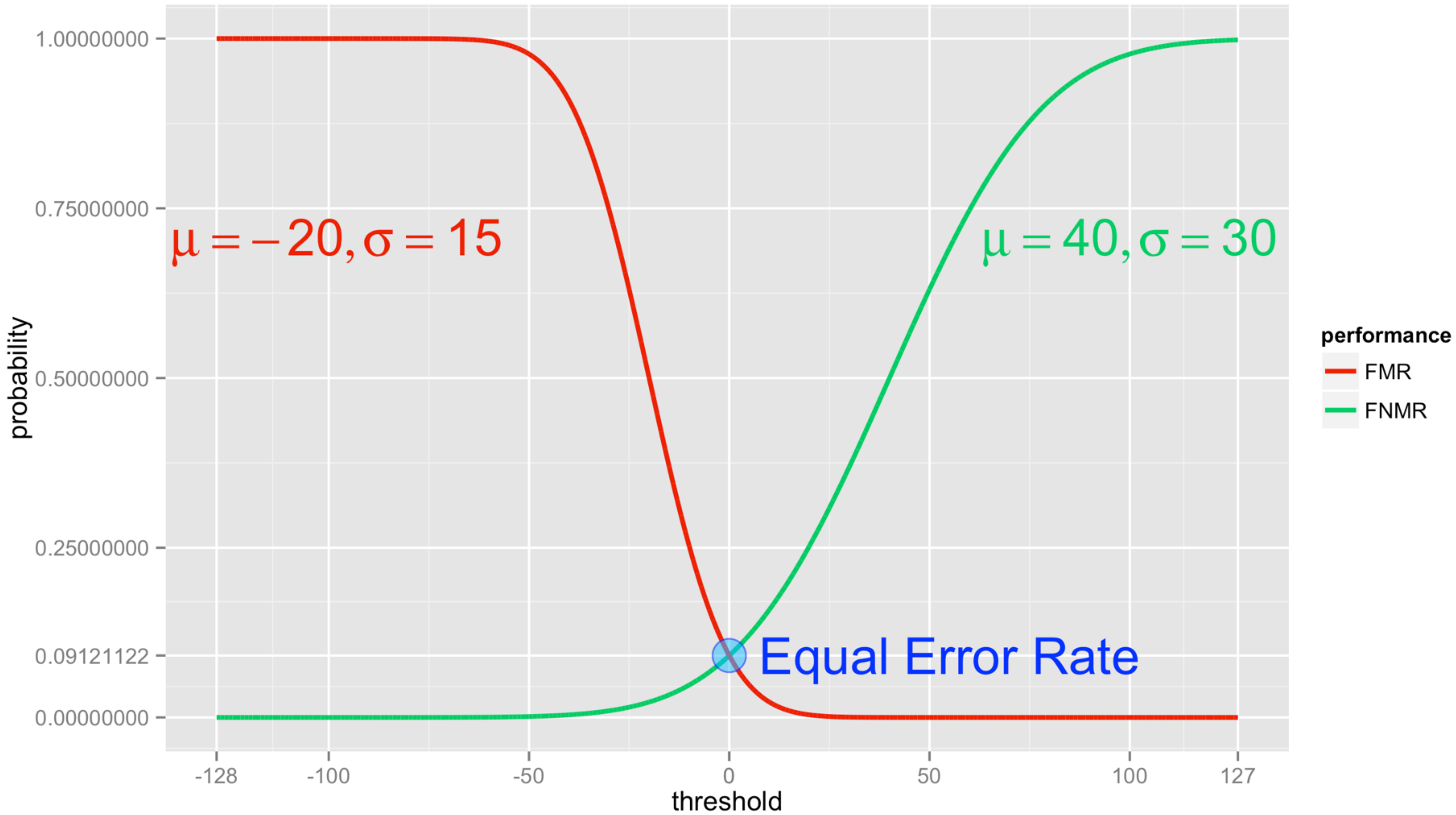
FALSE MATCH RATE



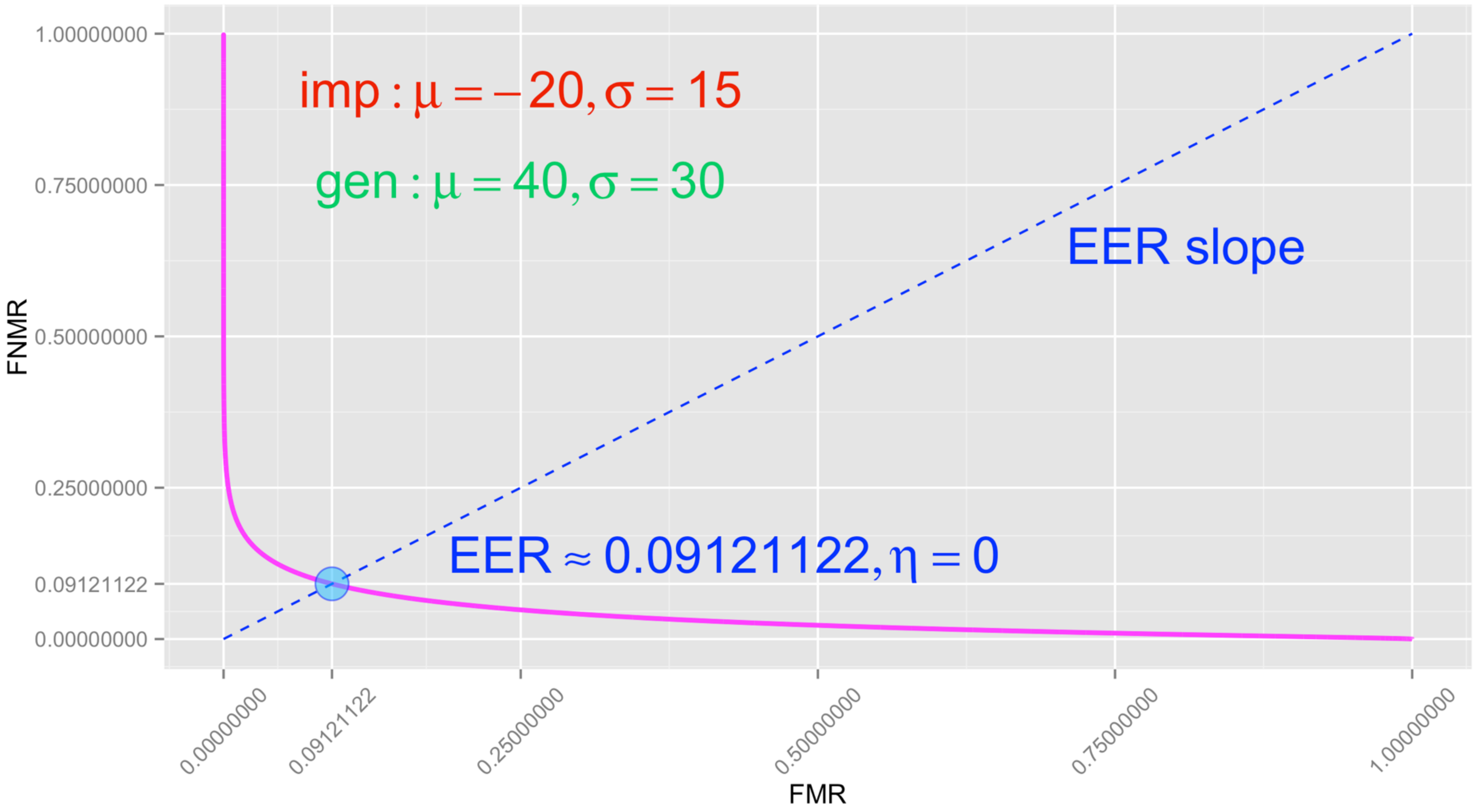
FALSE NON-MATCH RATE



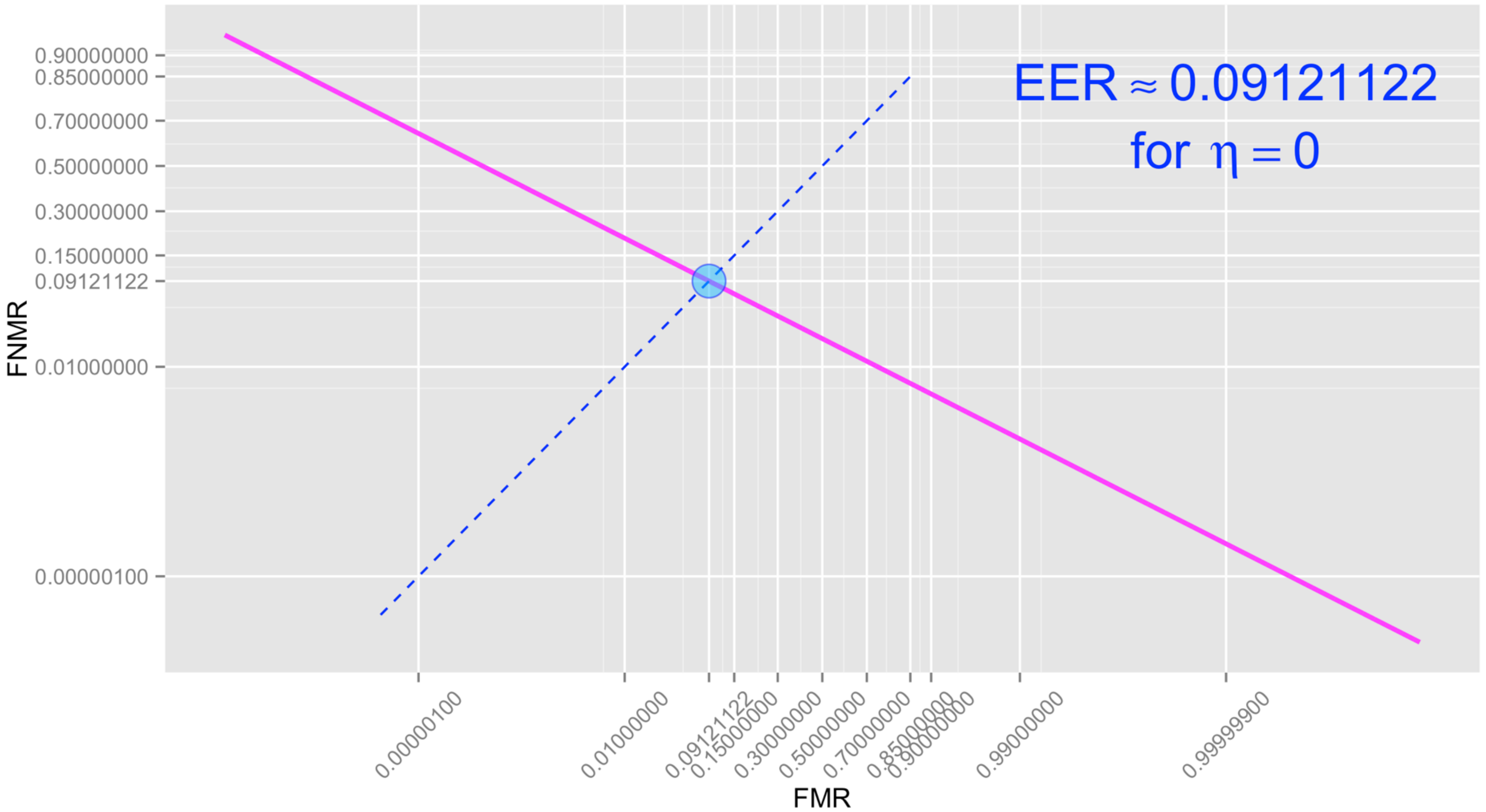
ERROR DISTRIBUTION FUNCTIONS



RECEIVER OPERATING CHARACTERISTICS



DETECTION ERROR TRADE-OFF



ISO/IEC 19795

- Performance test methodologies for different life-cycle phases:
  - technology evaluation
  - scenario evaluation
  - operational evaluation

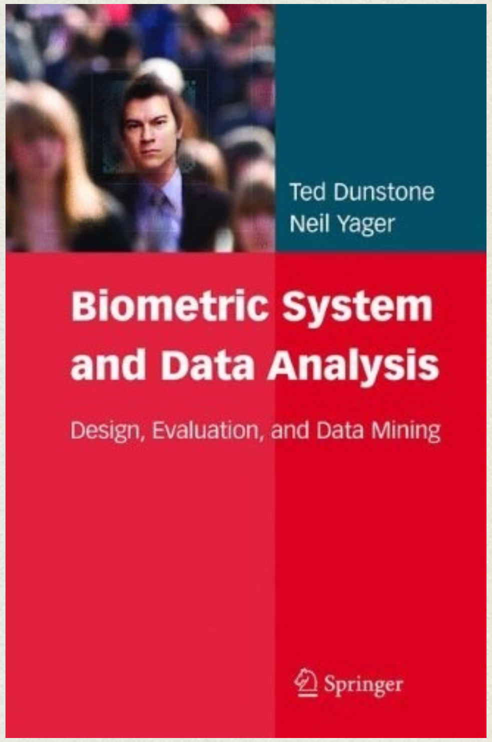
- We get comparable results with plausible confidence intervals.

BUNCH OF PARAMETERS

- False Match Rate / False Non-Match Rate
  - attempt oriented
- False Acceptance Rate / False Rejection Rate
  - transactional version of FMR/FNMR
- Failure To Acquire
- Failure To Enroll
  - both attempt and txn-oriented versions

BIOMETRIC DATA MINING

- In any life-cycle phase, we shall gather as much data as we can to estimate the performance or check we are still operating in expected margins.
- Anomalies may indicate a component malfunction or even a fraud.
- Again, be careful about confidence.
- Misleading statistics can be worse than none!



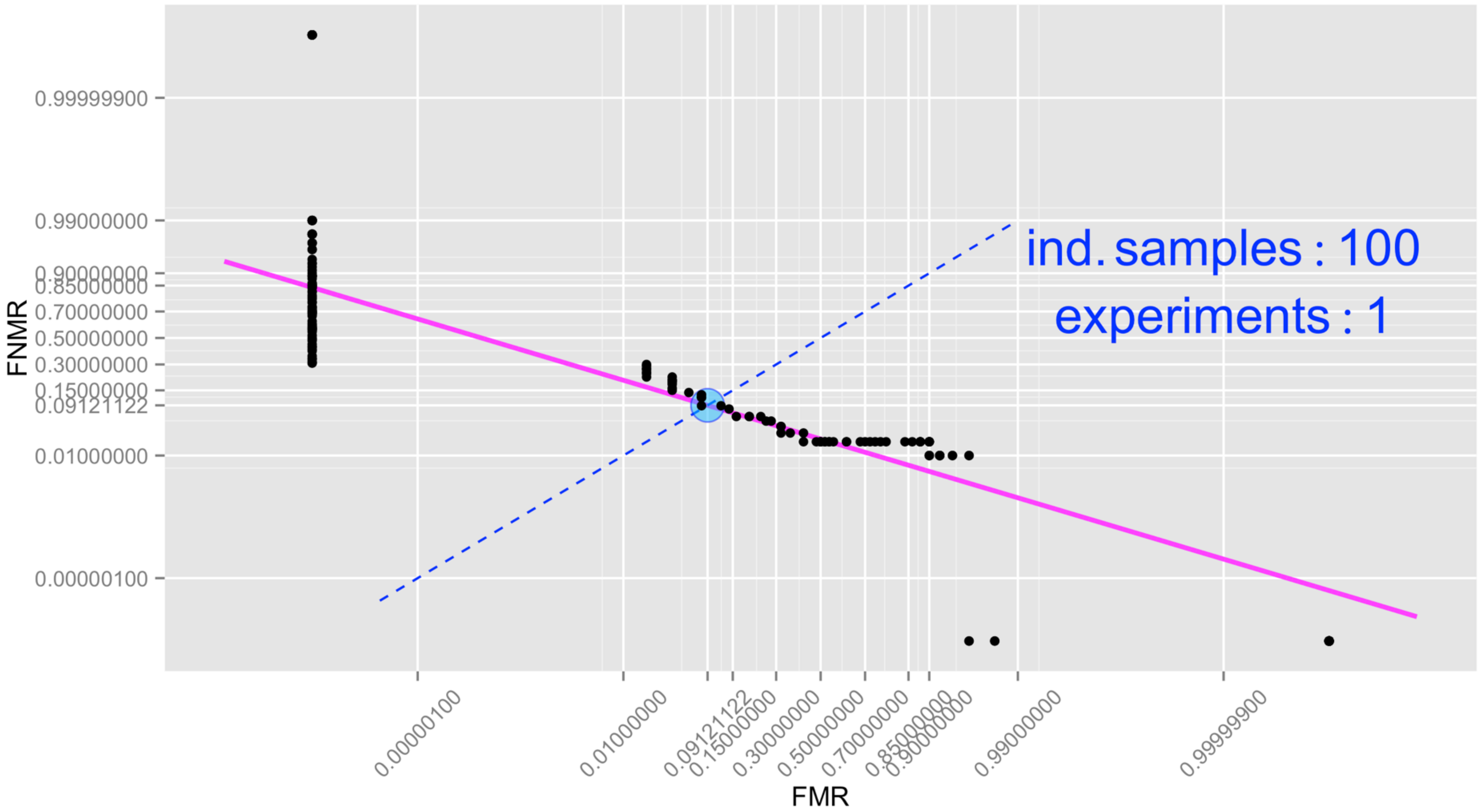
Ted Dunstone
Neil Yager

Biometric System and Data Analysis

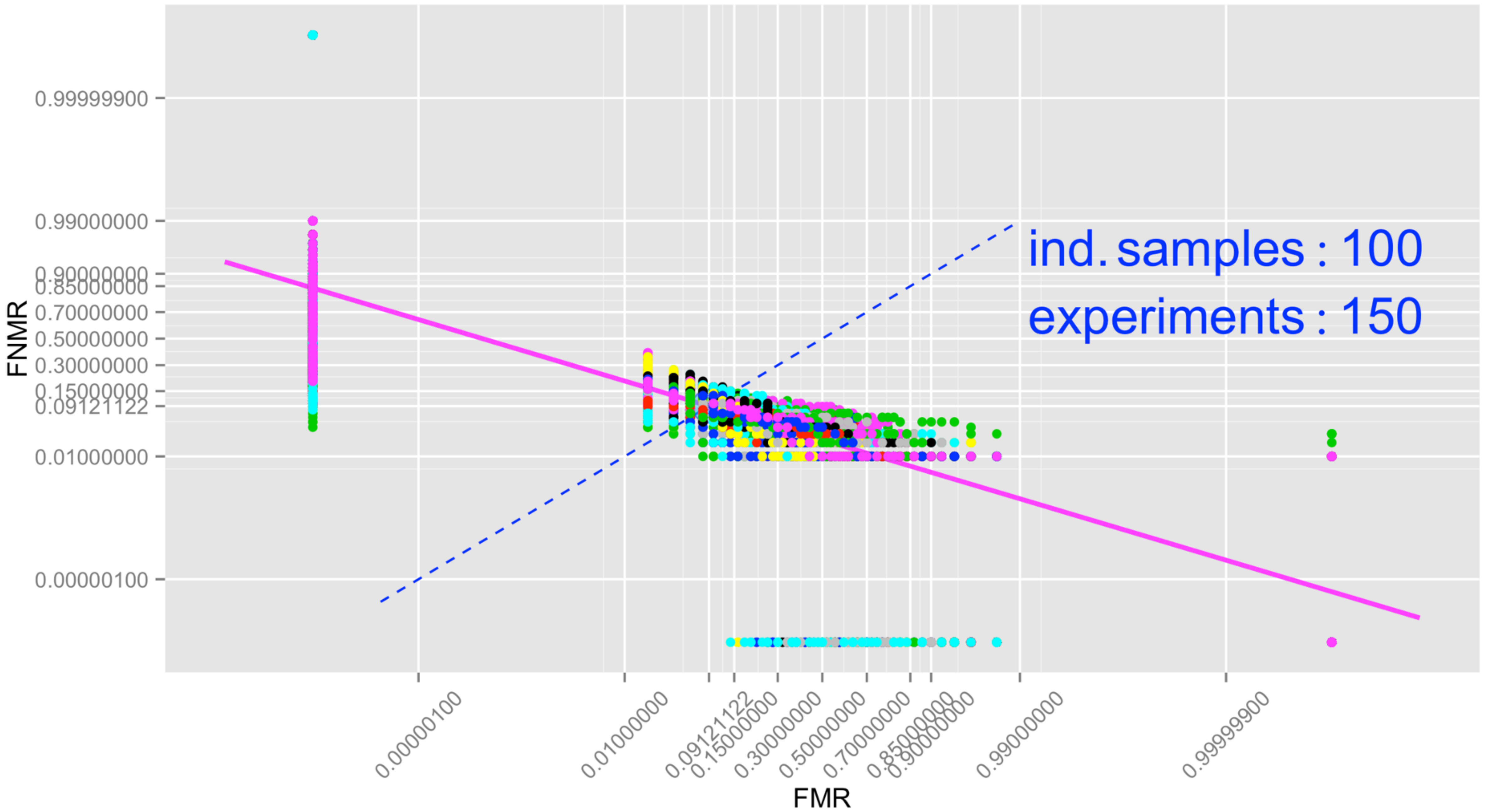
Design, Evaluation, and Data Mining

Springer

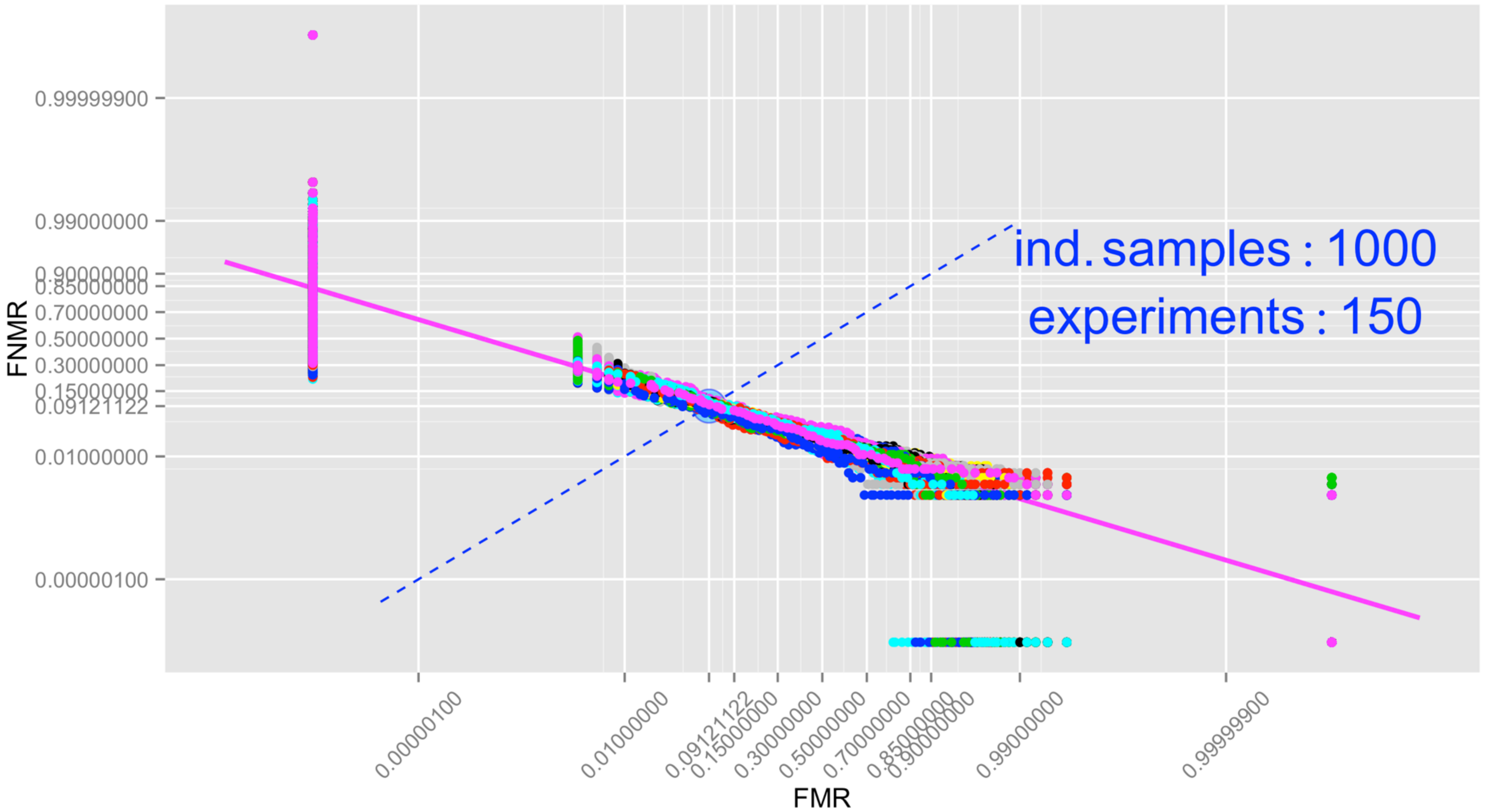
DET ESTIMATION SIMULATION



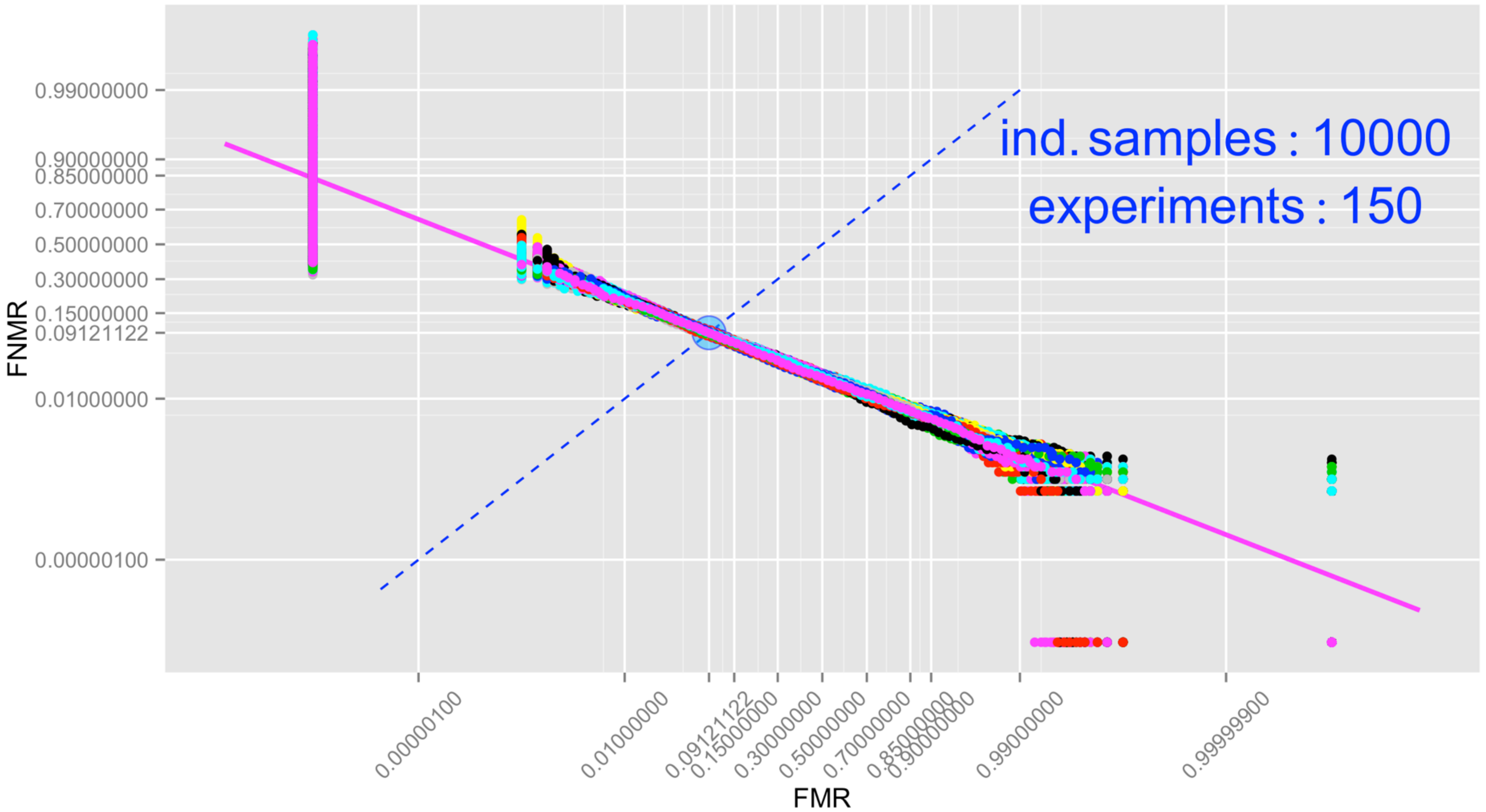
CONFIDENCE INTERVALS?!



ANY CONFIDENCE, YET?



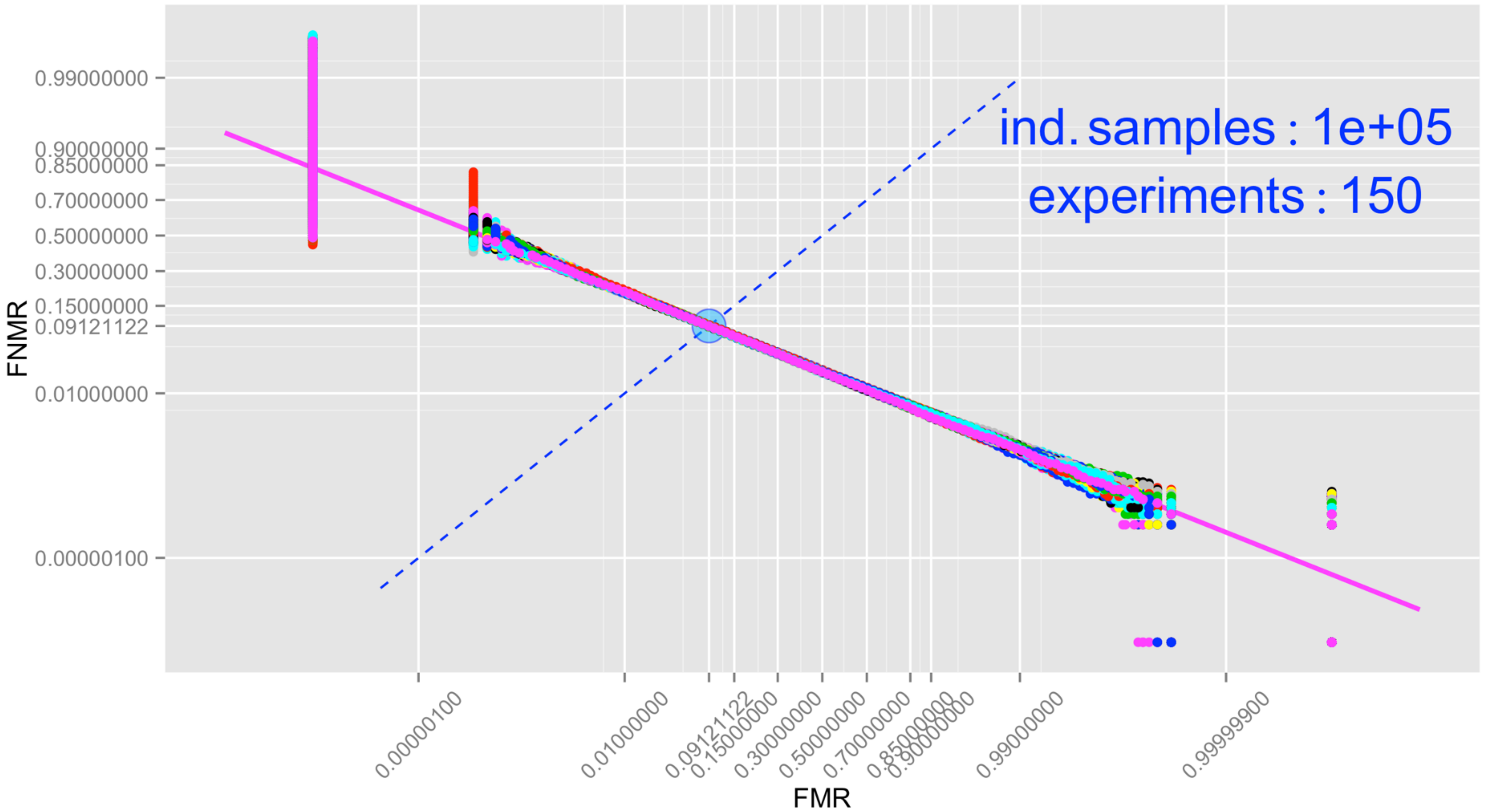
FAIR CONFIDENCE



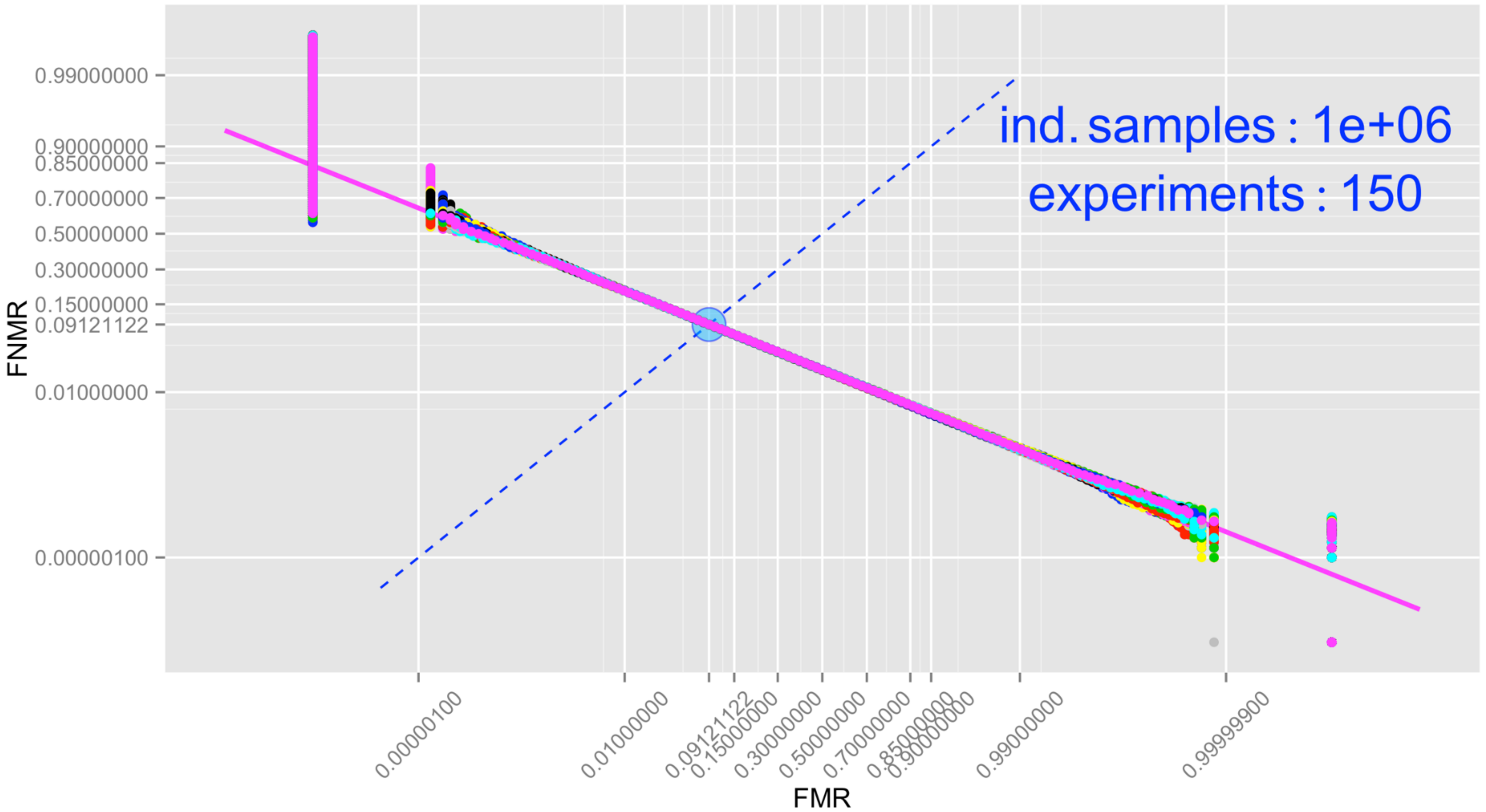
ind. samples : 10000

experiments : 150

WE CAN BE PROUD



JUST A DREAM...

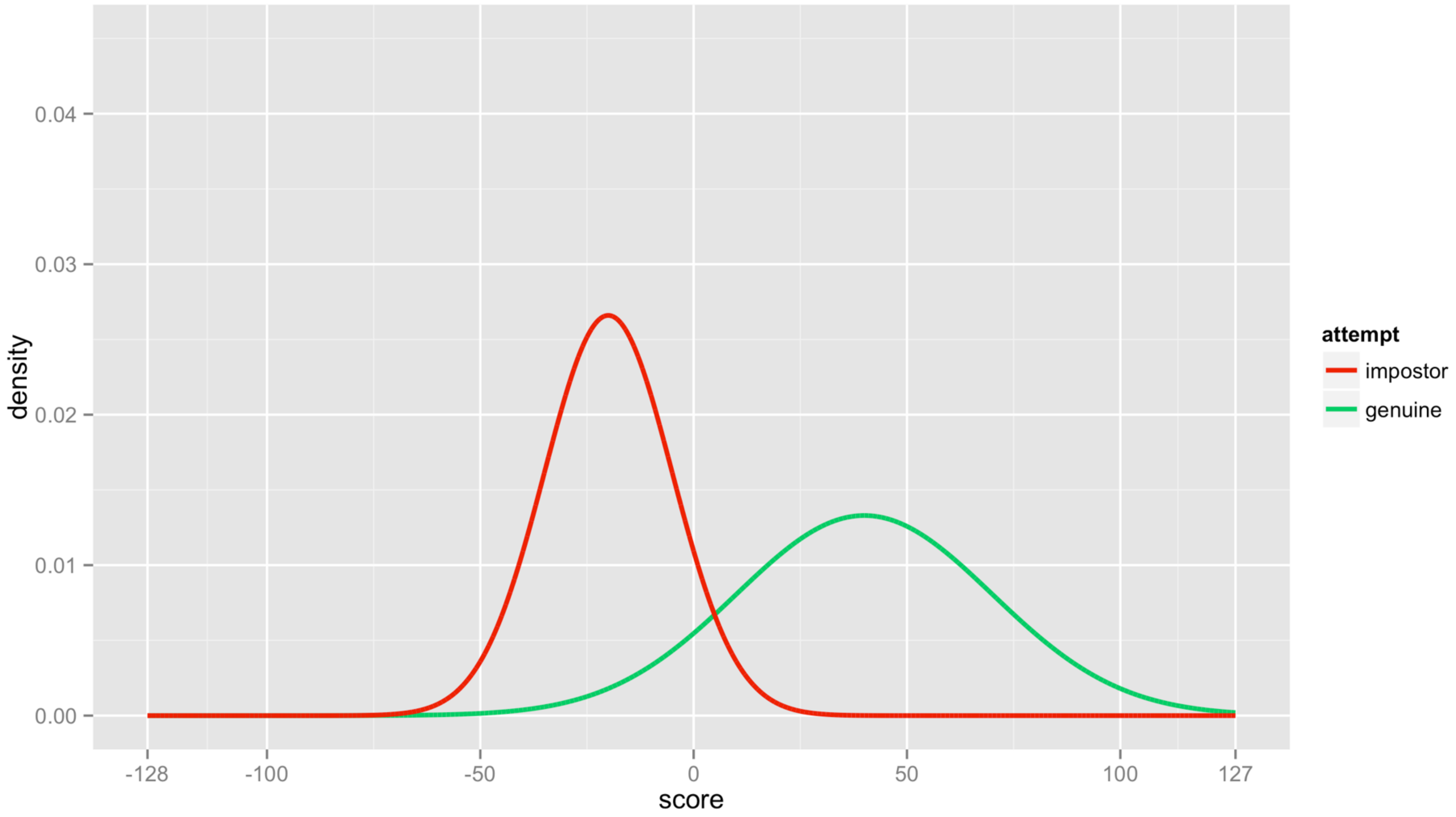


ind. samples : 1e+06
experiments : 150
FNMR
0.99000000
0.90000000
0.85000000
0.70000000
0.50000000
0.30000000
0.15000000
0.09121122
0.01000000
0.00000100
FMR
0.00000100
0.01000000
0.09121122
0.15000000
0.30000000
0.50000000
0.70000000
0.85000000
0.90000000
0.99000000
0.99999900

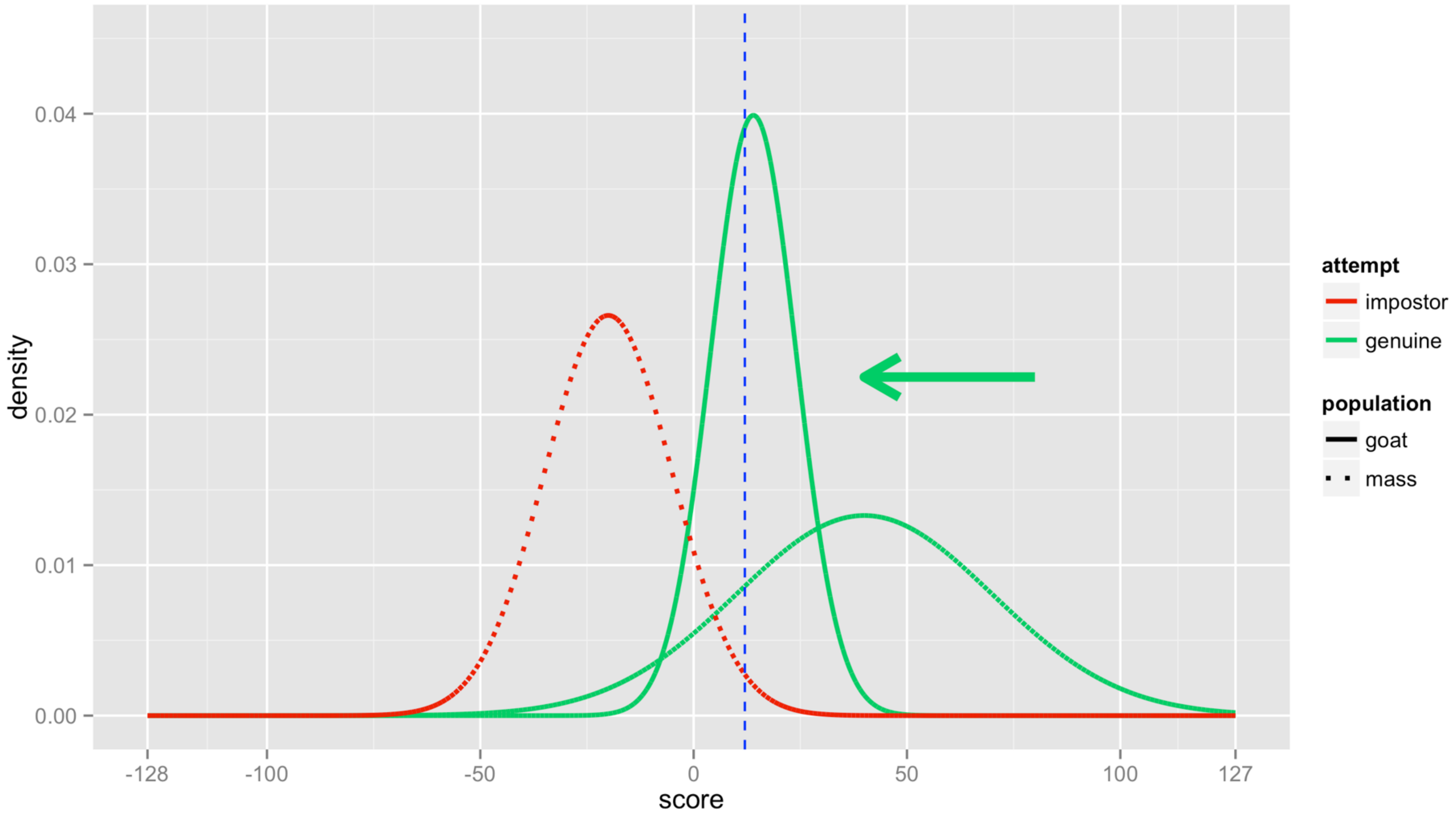
BIOMETRIC MENAGERIE

- To further complicate biometrics testing, those score distributions are usually not person-independent.
  - That means the performance is not the same for all people.
- There are plenty of anomalies out there we shall be aware of to interpret the system behaviour correctly.

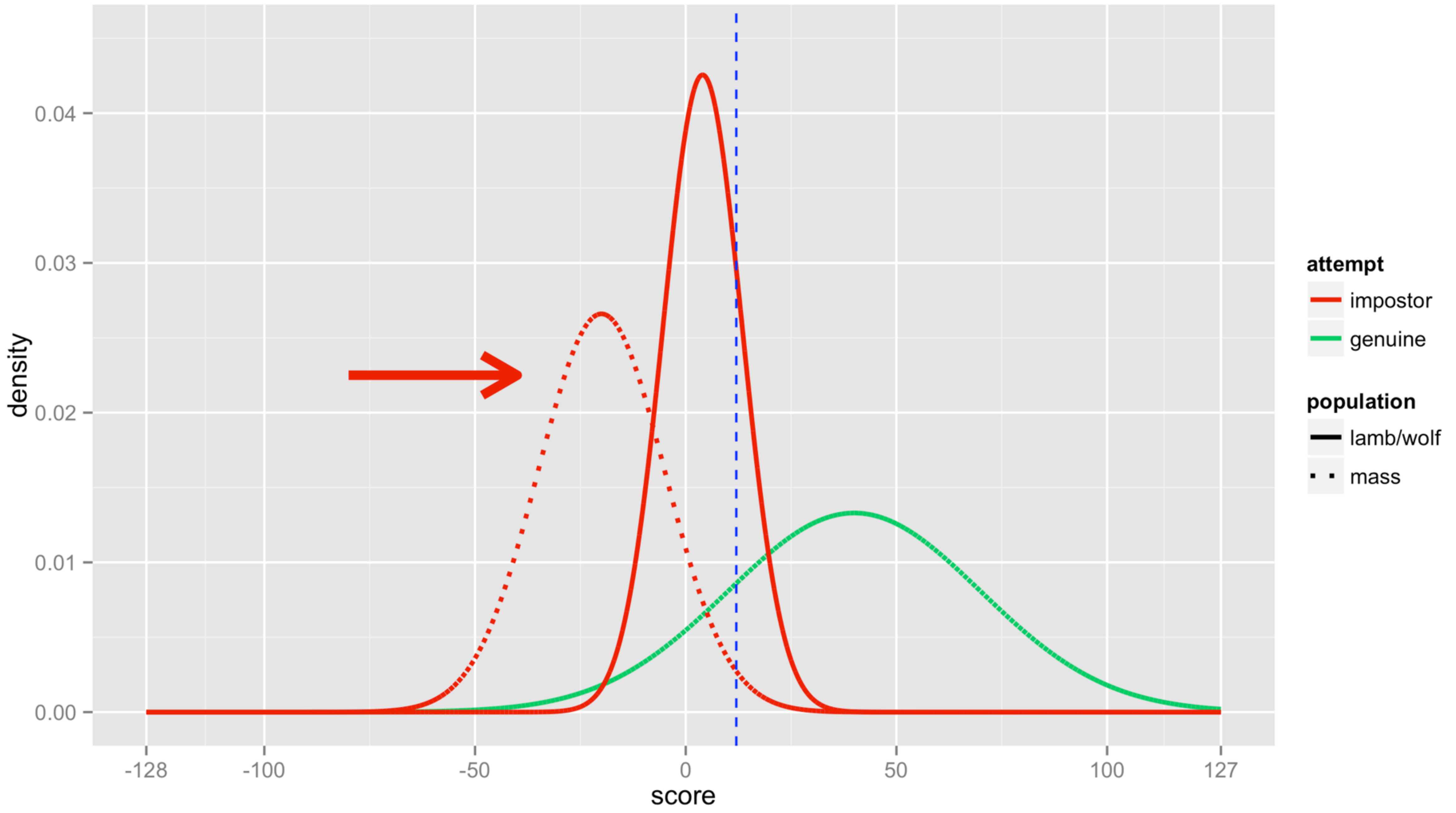
SHEEP: AN ORDINARY USER



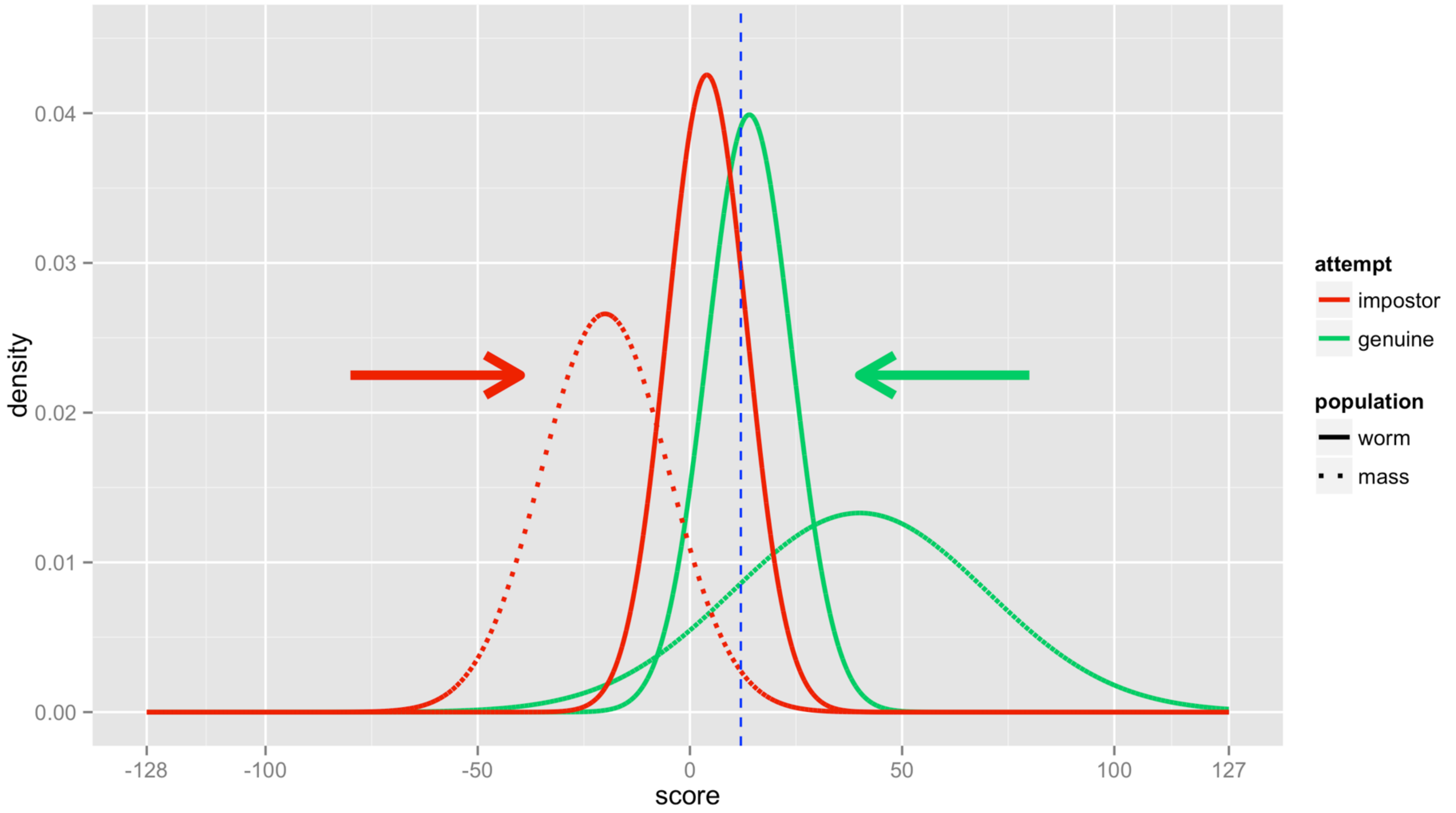
GOAT: PROBLEMATIC FNMR



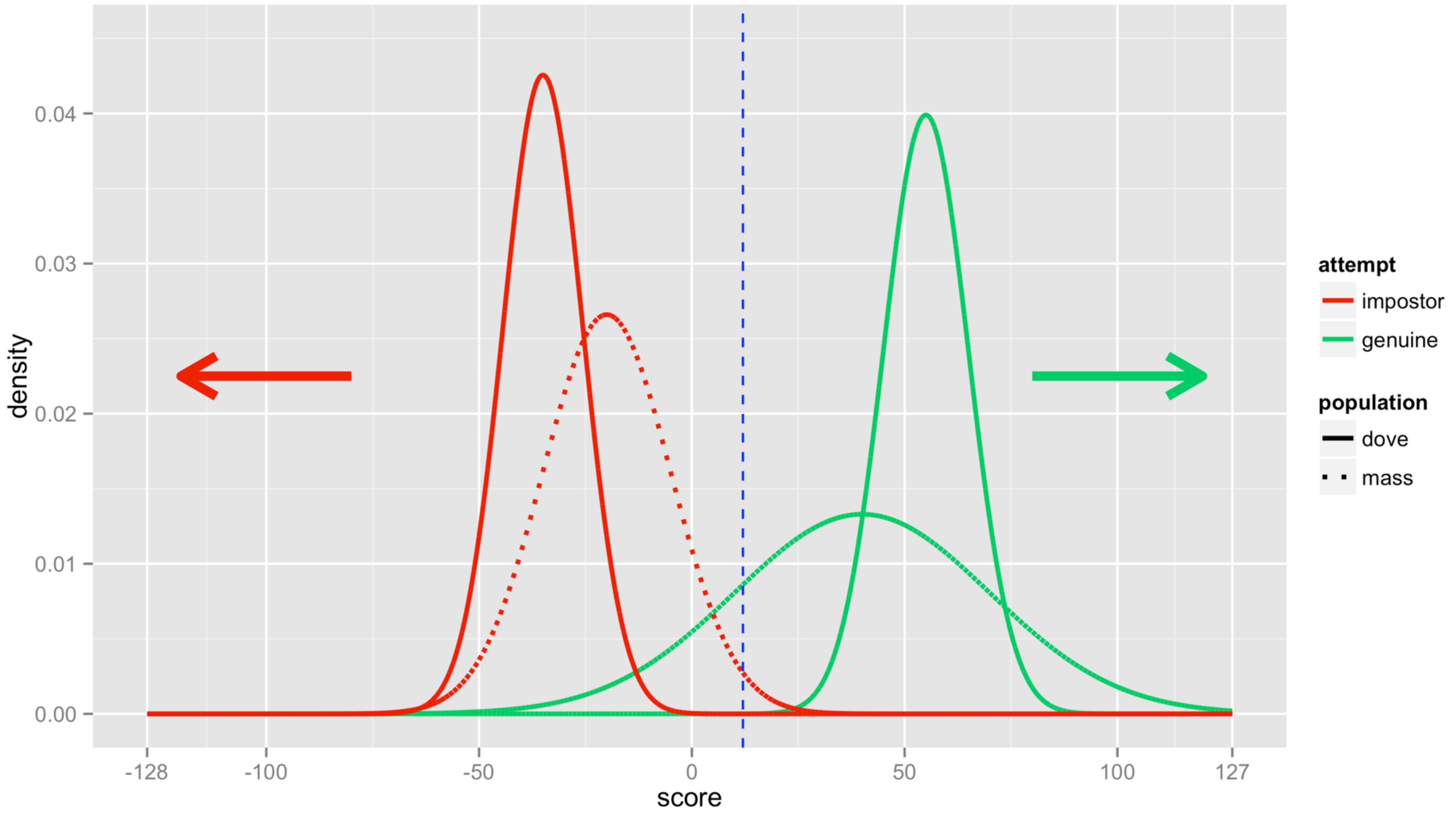
LAMB/WOLF: EASY TARGET AND-OR EFFECTIVE PREDATOR



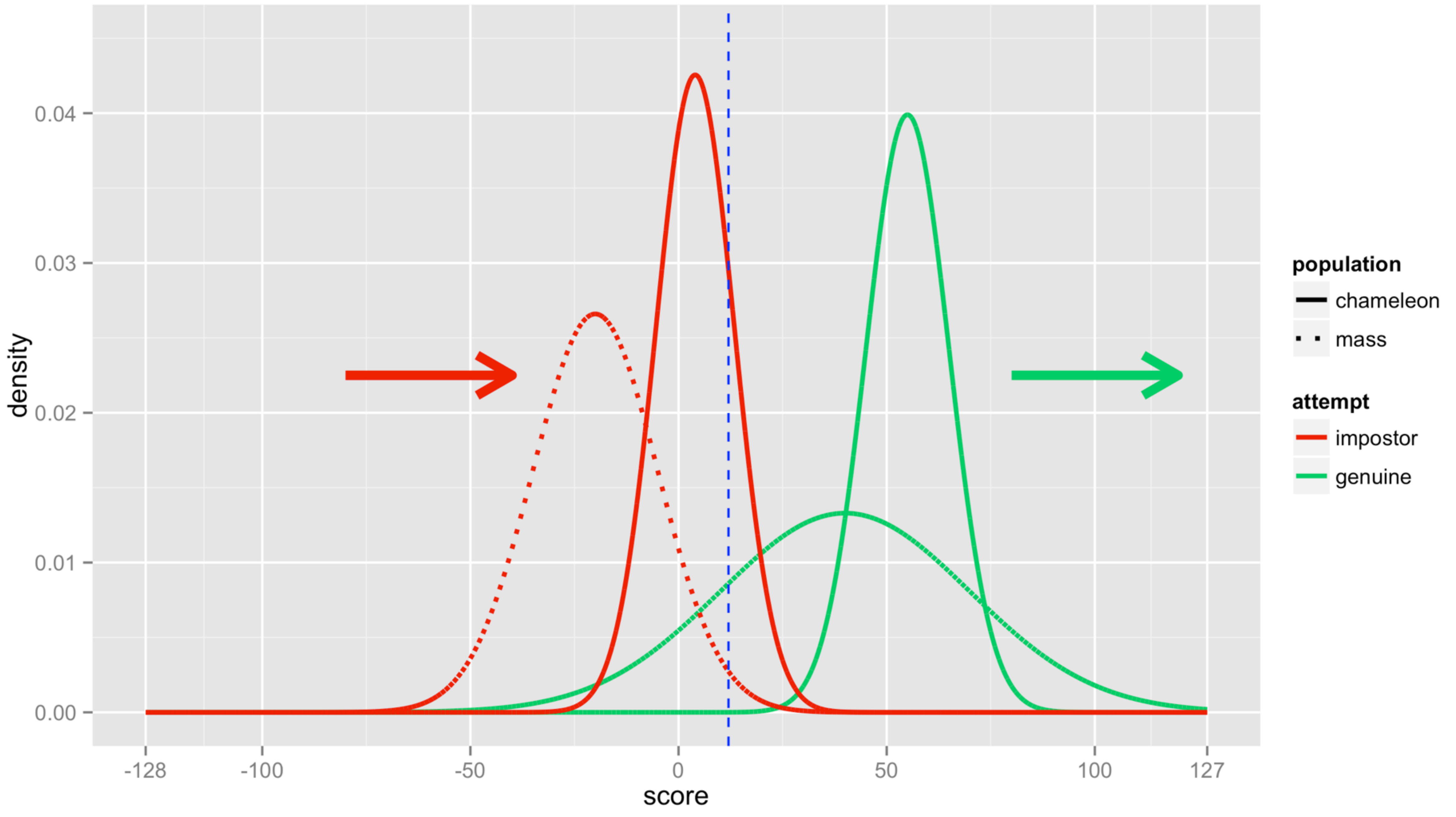
WORMS: BOTH FNMR AND FMR INCREASED



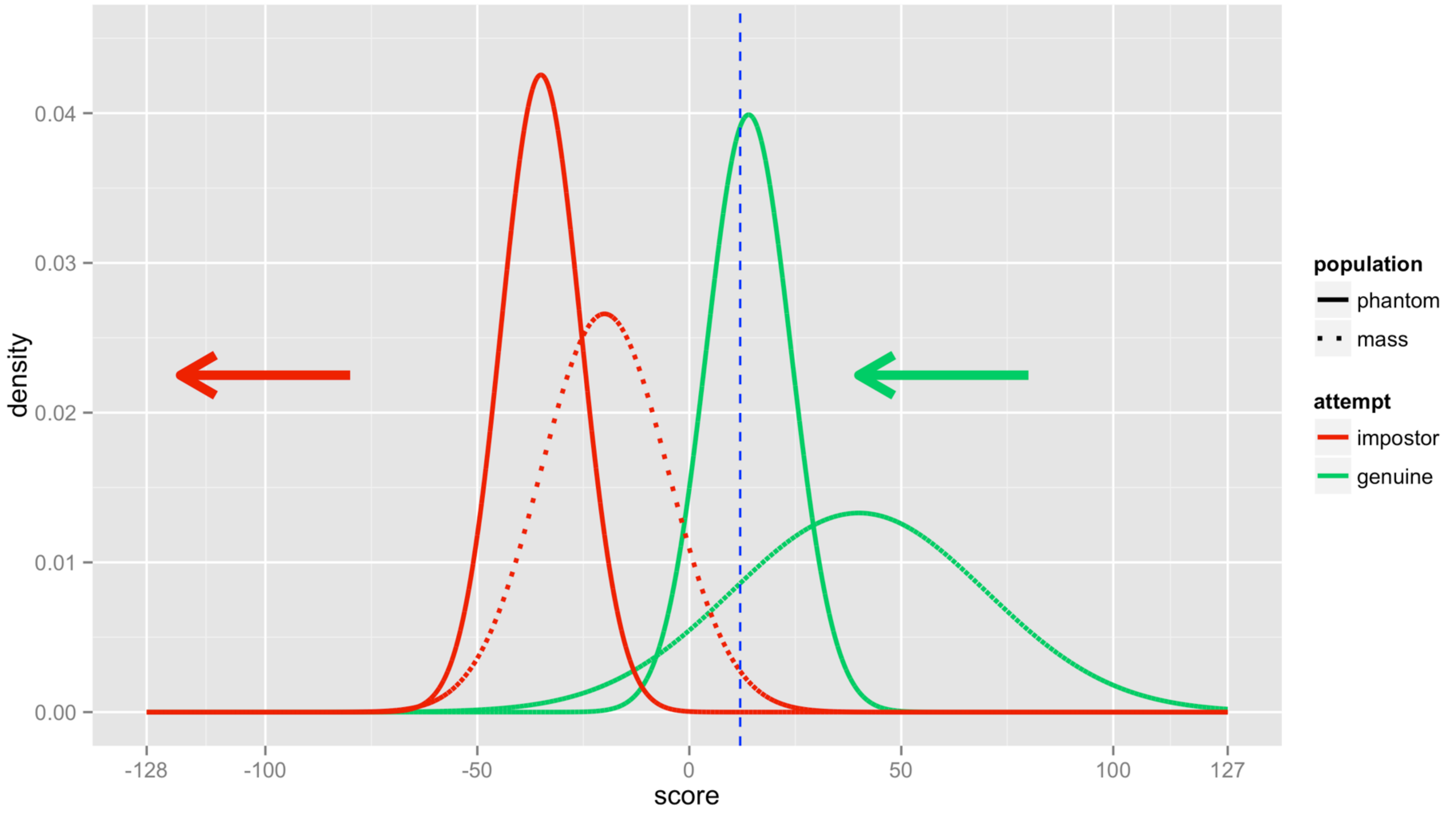
DOVE: EXCELLENT USER



CHAMELEON: EXCELLENT SCORES, ANYWAY(!)



PHANTOM: PROBLEMATIC MATCHING, ANYWAY



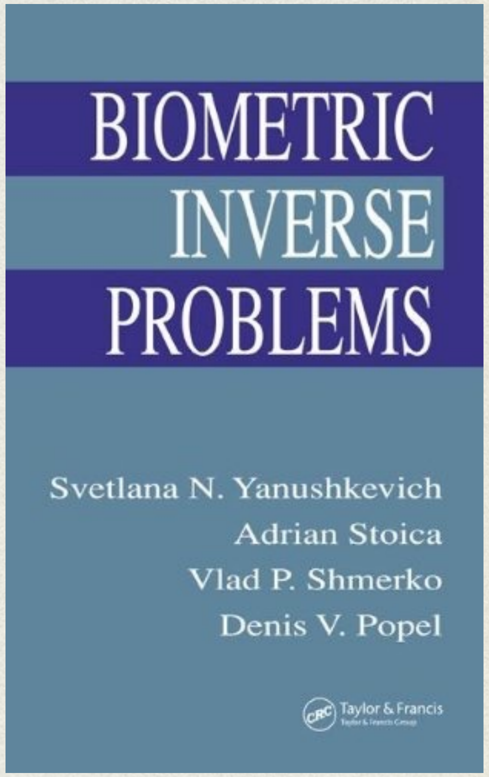
SECRET FILES ON BIOMETRICS

BIO BRUTE FORCE ATTACK

- Randomly generate plausible circca 1/FMR samples and put them to the test.

  - Also termed “Zero-Effort”, denoting that the attacker makes no special effort to imitate the original person characteristic.

- Synthetic samples generation is quite feasible today.



BIOMETRIC INVERSE PROBLEMS

Svetlana N. Yanushkevich
Adrian Stoica
Vlad P. Shmerko
Denis V. Popel

Taylor & Francis

CRYPTANALYSIS-LIKE ATTACKS

- Masquerade attacks, can be a variant of “Hill-Climbing” denoting the attacker iteratively improves the BIO sample data based on:
  - scoring feedback (side channels)
  - stolen template (pre-image attacks)
  - independent template trained from intercepted BIO samples (correlation attacks)
  - known scoring anomaly (differential analysis)
  - implementation faults (general hacking)

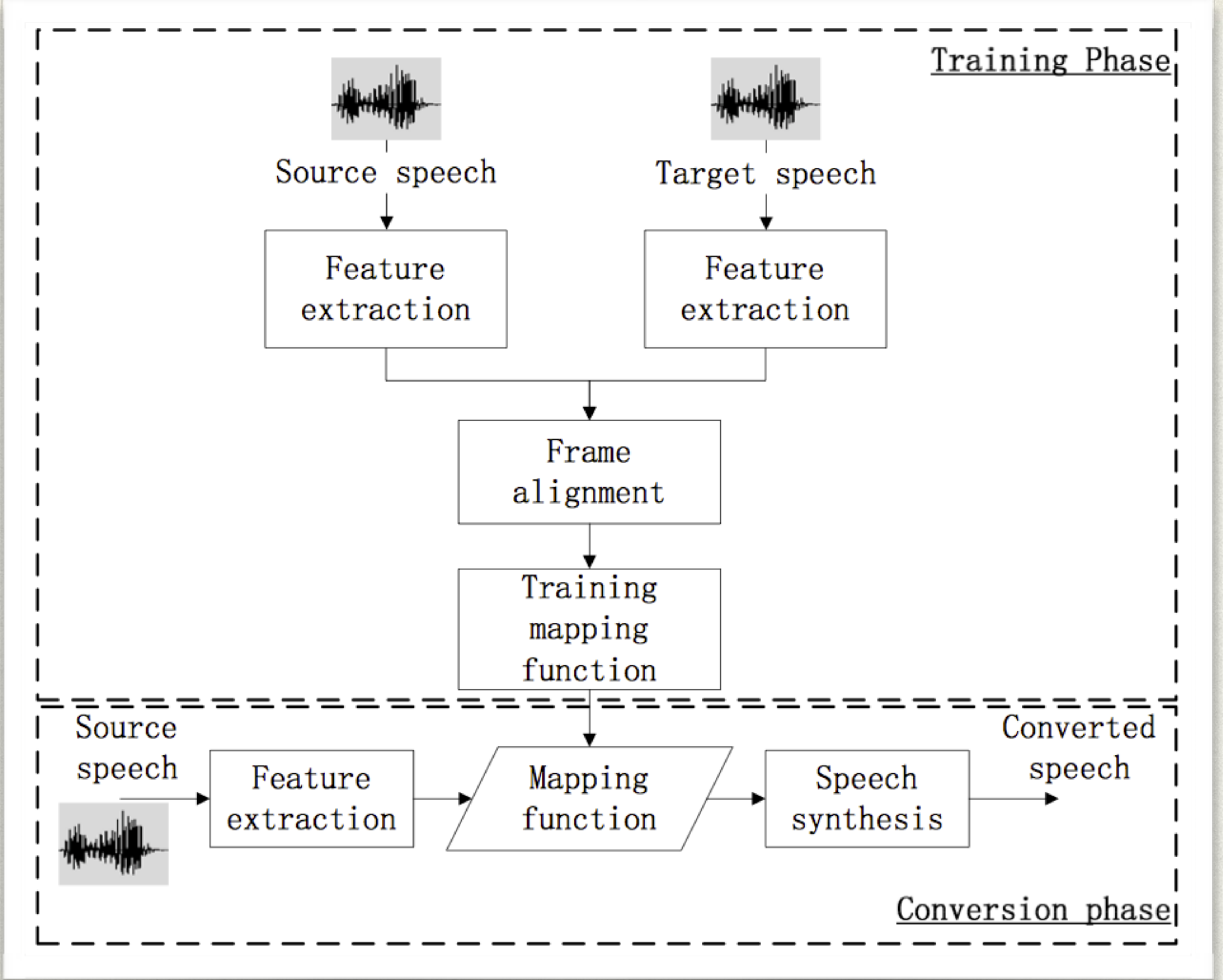
SPOOFING

- *The process of defeating a biometric system through the introduction of fake biometric samples.*
    - *(Schuckers, Adler et al., 2010)*
- Particular modus operandi on how to deploy the attacking data vectors.
    - Can be seen as being orthogonal to the aforementioned ways of gaining fake samples.

SENSOR-BYPASS ATTACKS

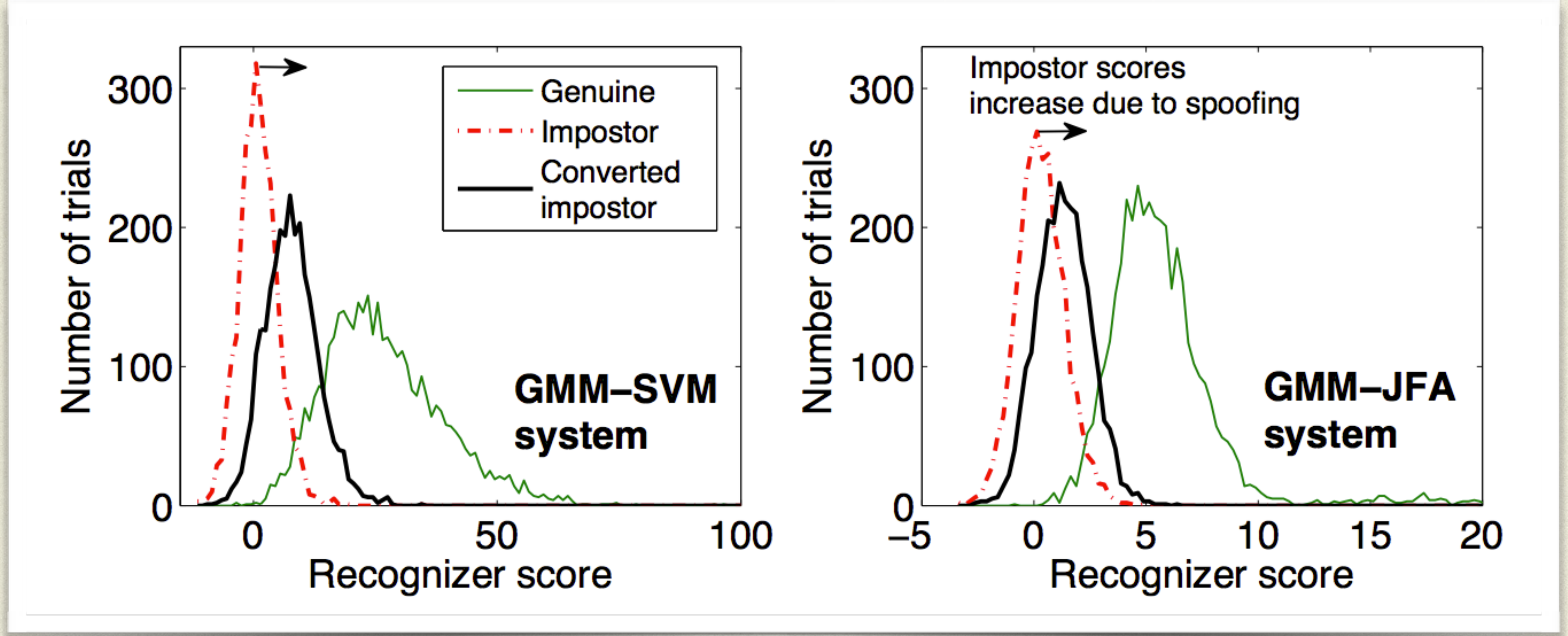
- Do not expose API service for unrestricted automated sample verification!
  - Recall the zero-effort attack complexity is often trivial.
  - Furthermore, masquerade attacks can shift FMR significantly.

CONVERSION ATTACK EXAMPLE



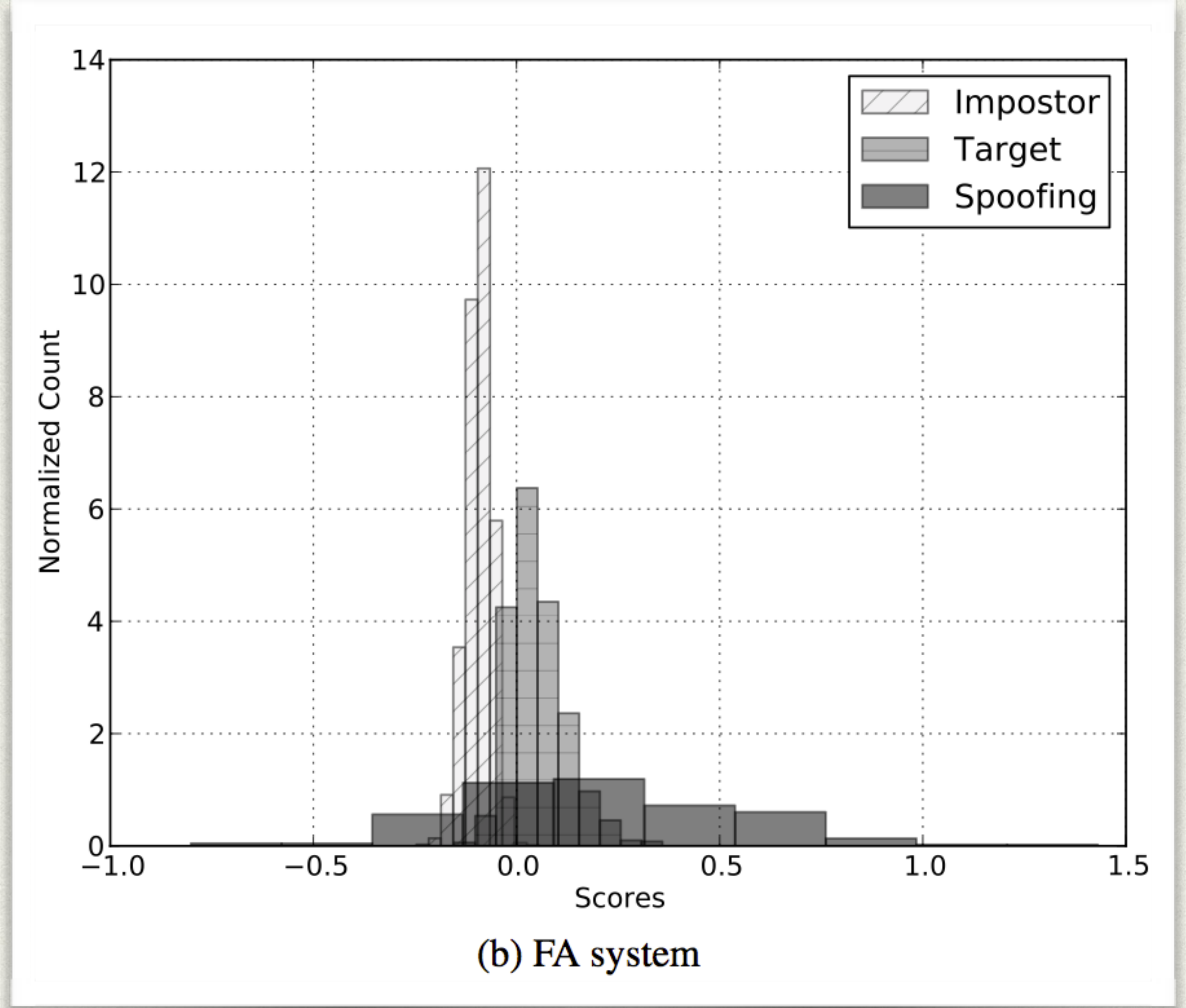
Kinnunen et al., ICASSP 2012

REPORTING ATTACK IMPACT



Kinnunen et al., ICASSP 2012

ARTIFICIAL SIGNALS IMPACT



(b) FA system

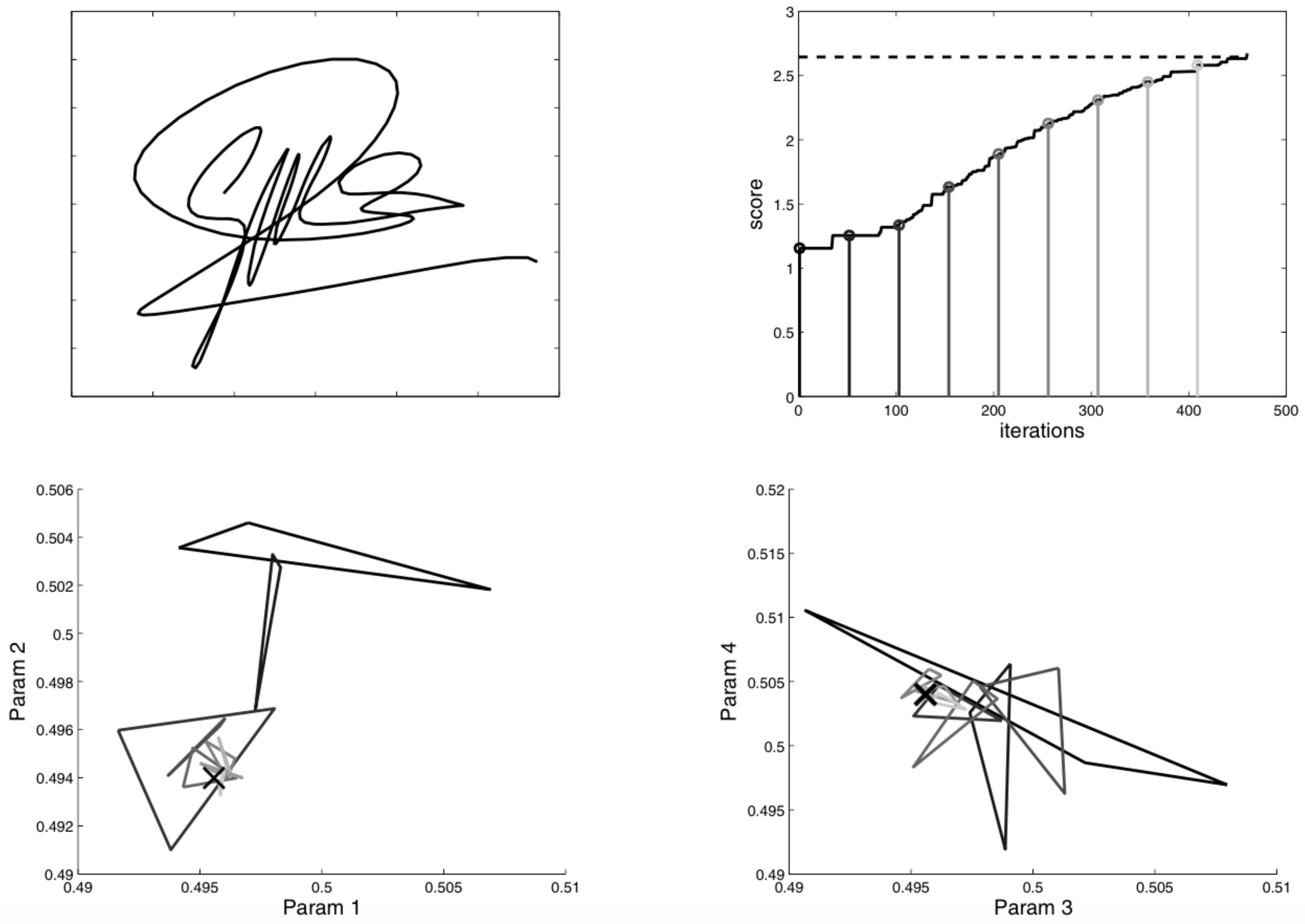
Alegre et al., EUSIPCO 2012-13

BIOMETRIC SIGNATURE MASQUERADE

- Hill-Climbing attack based on the Uphill Simplex algorithm and its application to signature verification
    - Gomez-Barrero, M., Galbally, J., Fierrez, J., and Garcia, J.-O. at BioID 2011

FMR 0-effort	ϕ(#trials) 0-effort	FMR’ masq.	ϕ(#iters) masq.
0.05%	2 000	91.76%	1 556
0.01%	10 000	89.58%	1 678
0.0025 %	40 000	87.82%	1 805

SUBSPACE CONVERGENCE ILLUSTRATED



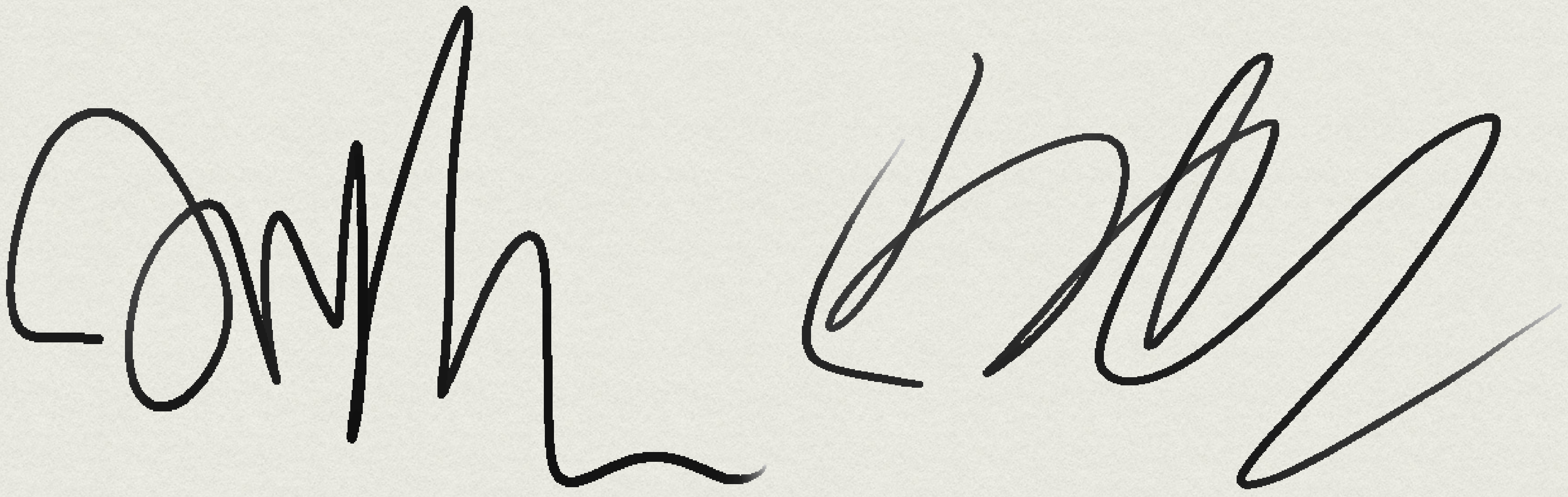
X-TALK SIGNAL LEAKAGE

- Furthermore, there is a certain link in between online (sign-pad made) and offline (pen-and-paper made) signatures.
  - Btw., we also hope to exploit this link should it come to a trial.
  - On the other hand, the amount of information being cross-transferred in between these two signal forms is yet to be discovered!

PDF SIGNATURE LEAKAGE

- When signing a PDF using online signature data, we often put a human readable picture into the PDF annotation.
    - This is just to make the technology more user-friendly.
- This is, however, usually an offline plaintext projection of the (encrypted) online signature data.
    - How much information is leaking this way?

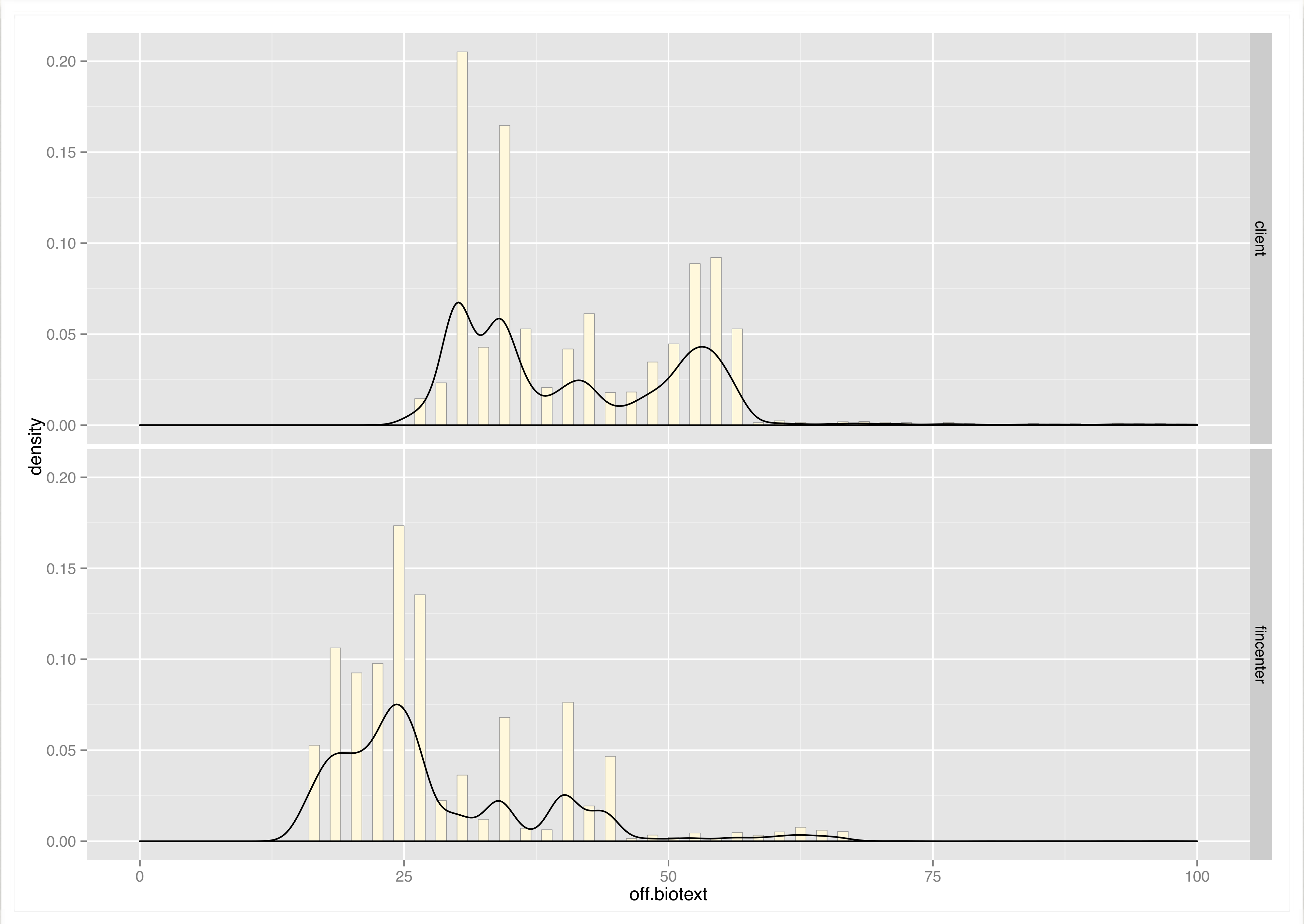
OFFLINE PROJECTION EXAMPLE



fincenter

client

OFFLINE SIGNAL BRIEF - THERE IS SOMETHING!



ISO/IEC 24745 REQUIREMENTS

- Renewability
  - allows multiple independent biometric references created ad hoc
  - a particular leaked template does not compromise the other ones (provably!)
- Revocability
  - user can revoke the ability of being successfully verified by a particular template from now on
- Biocryptography is an effective way on how to achieve these goals.

BIOMETRIC CRYPTOGRAPHY?

CRYPTOGRAPHY EXACTNESS

Let $y = AES_K(x)$ for a random K.

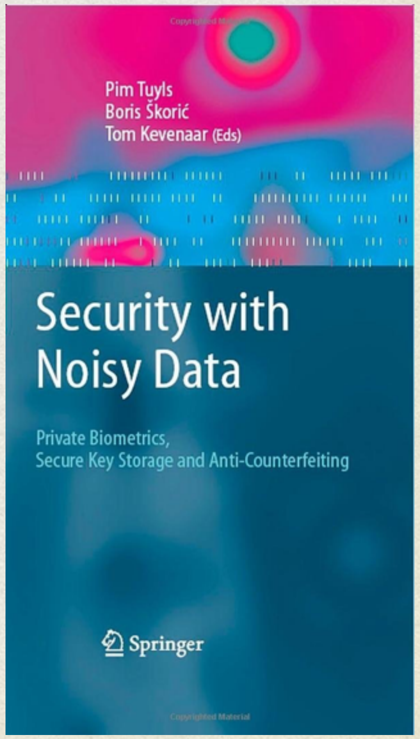
Then $AES_K^{-1}(y) = x$, while

$AES_{K\oplus 1}^{-1}(y) \neq x$ (probability ≈ 1).

- *The better the algorithm is the more randomized response we get for even one-bit error.*

BIOMETRICS FUZZINESS

- We seldom get the same data in the subsequent scans of the very same person.
    - Actually, this is usually a clear sign of a spoofed sample.
- To overcome this (intra-class) variability, we can employ the *biometric cryptography*.



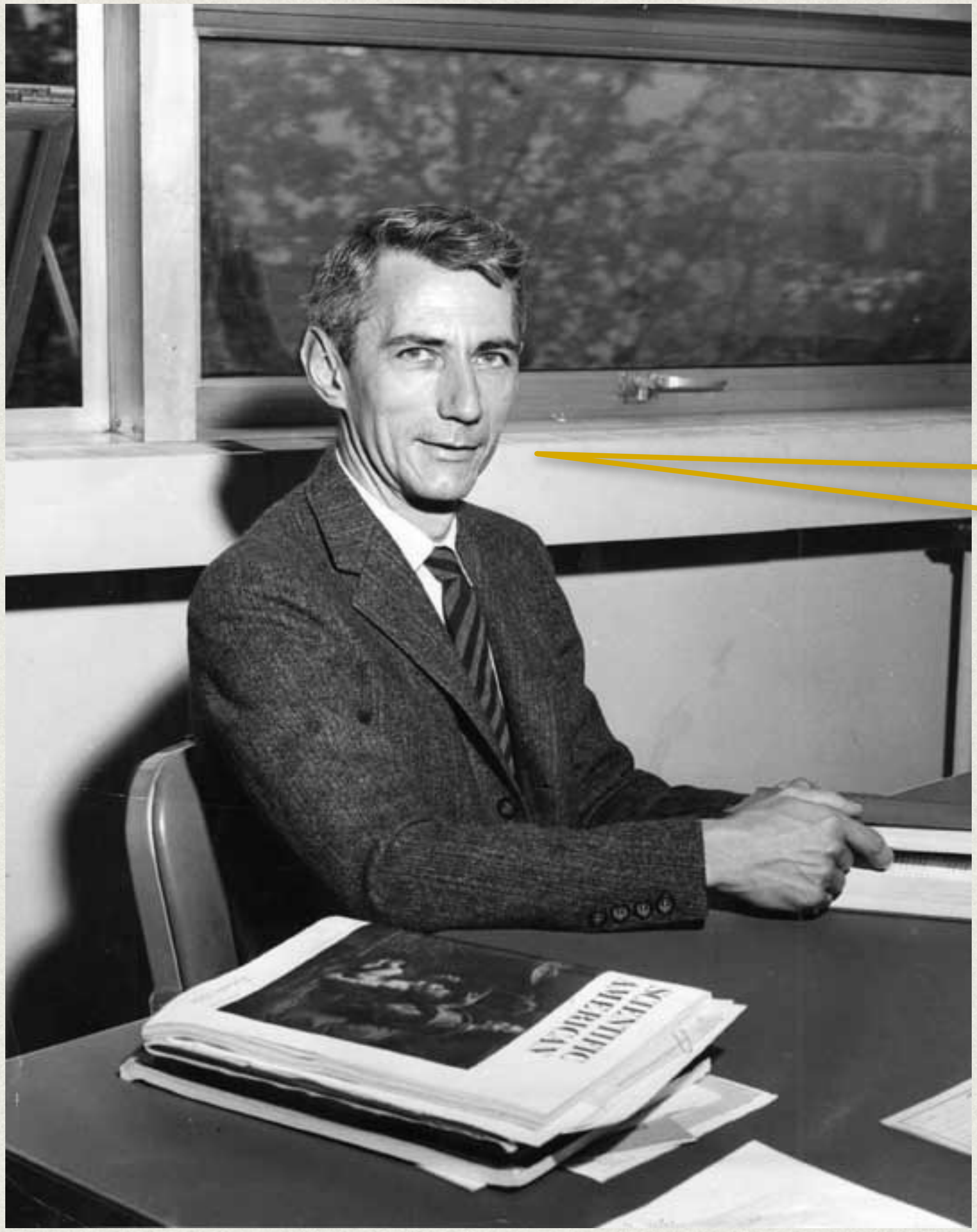
Pim Tuyls
Boris Škorić
Tom Kevenaar (Eds)

Security with
Noisy Data

Private Biometrics,
Secure Key Storage and Anti-Counterfeiting

Springer

BACK TO THE ORIGIN



1. analyse the entropy gain from inter-class variation
2. use an error-correction code to cope with intra-class noise

Claude Elwood Shannon, 1948-49

ERROR-CORRECTING CODE C

Let (F,ρ) be a metric space, $\rho : F \times F \to [0,\infty)$.

translation invariant metric: $\rho(x,y) = \rho(0,x-y)$

Error correcting code is $C \subset F, C = \{c_1, c_2, \ldots\}$.

$decode : F \to C$

t-error correction capability:

Let $\rho(c_i, y) \leq t$, then $decode(c_i) = decode(y) = c_i$.

We assume *decode*() always returns
a (possibly wrong) codeword.

ENROLMENT

i) randomly choose $c_{key} \in \boldsymbol{C} \subset \boldsymbol{F}$

ii) get BIO features vector $w \in \boldsymbol{F}$

iii) let $\xi = w - c_{key}$

iv) let $BIO\_key = hash(c_{key})$

v) template = (ξ)

ENROLMENT

i) randomly choose $c_{key} \in \boldsymbol{C} \subset \boldsymbol{F}$

ii) get BIO features vector $w \in \boldsymbol{F}$

iii) let $\xi = w - c_{key}$

iv) let $BIO\_key = hash(c_{key})$

v) template = (ξ)

More involved entropy extractors can be used here…

VERIFICATION

i) get BIO features vector $w' \in \boldsymbol{F}$

ii) let $y = w' - \xi$

iii) let $c_{key}' = decode(y)$

iv) let $BIO\_key' = hash(c_{key}')$

v) use BIO_key' in the upper-layer protocol

VERIFICATION

i) get BIO features vector $w' \in \boldsymbol{F}$

ii) let $y = w' - \xi$

iii) let $c_{key}' = decode(y)$

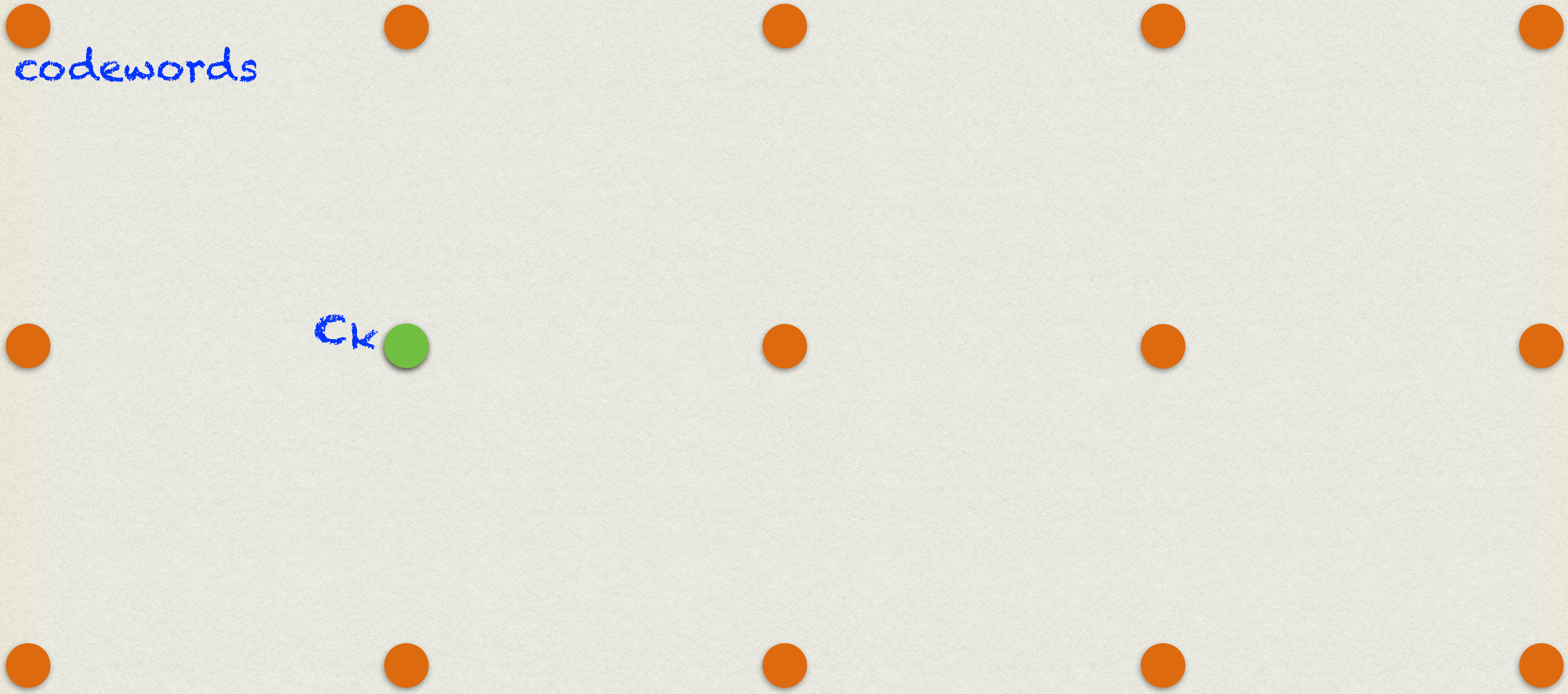
iv) let $BIO\_key' = hash(c_{key}')$

v) use BIO_key' in the upper-layer protocol

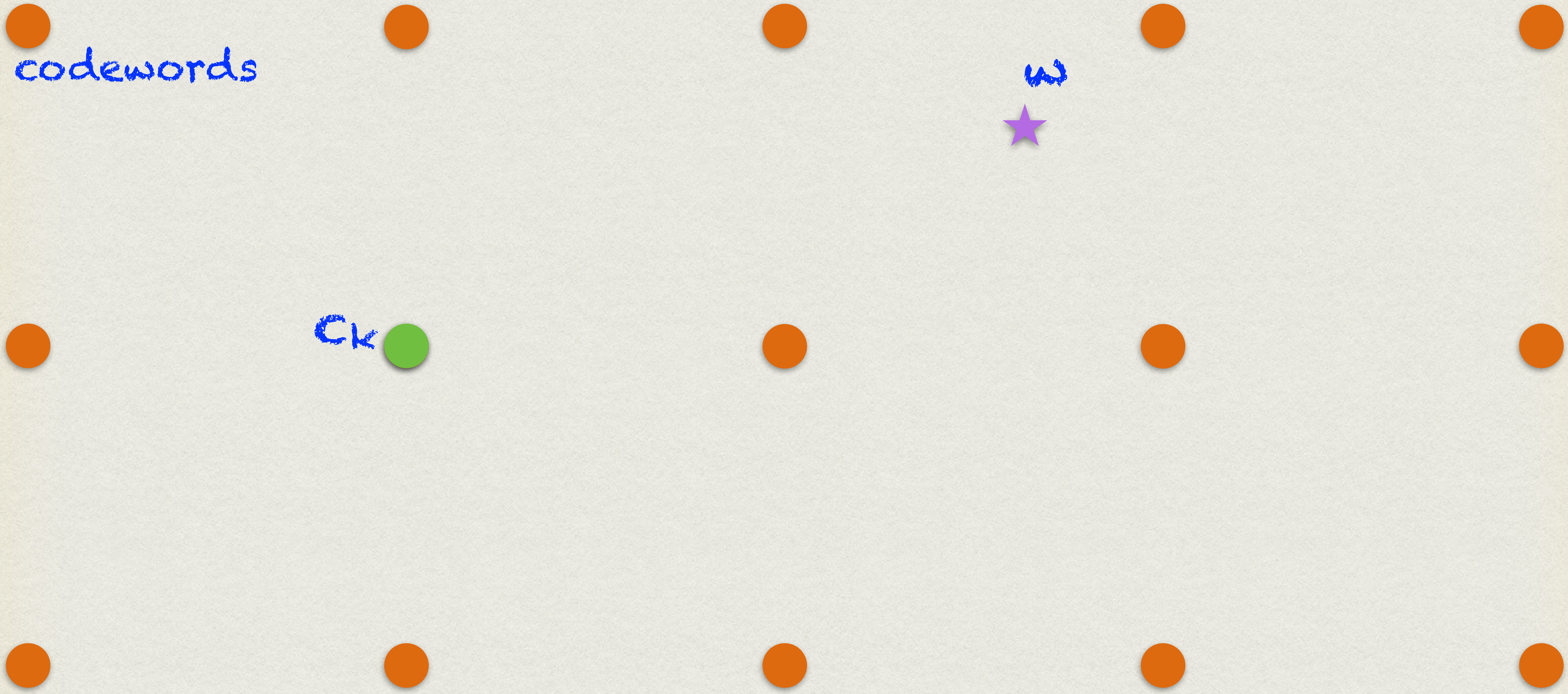
We have an ordinary crypto key, now…

CORE PRINCIPLE ILLUSTRATED

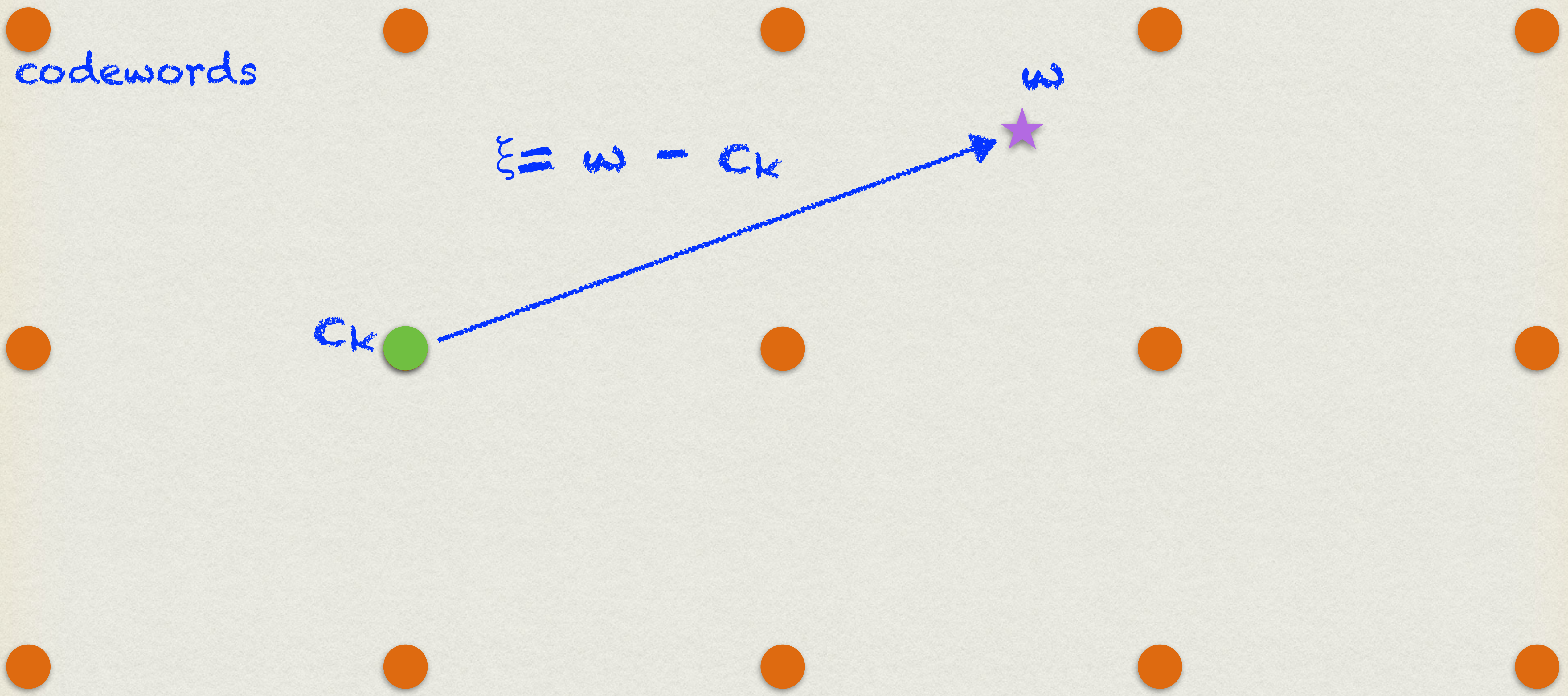
CORE PRINCIPLE ILLUSTRATED



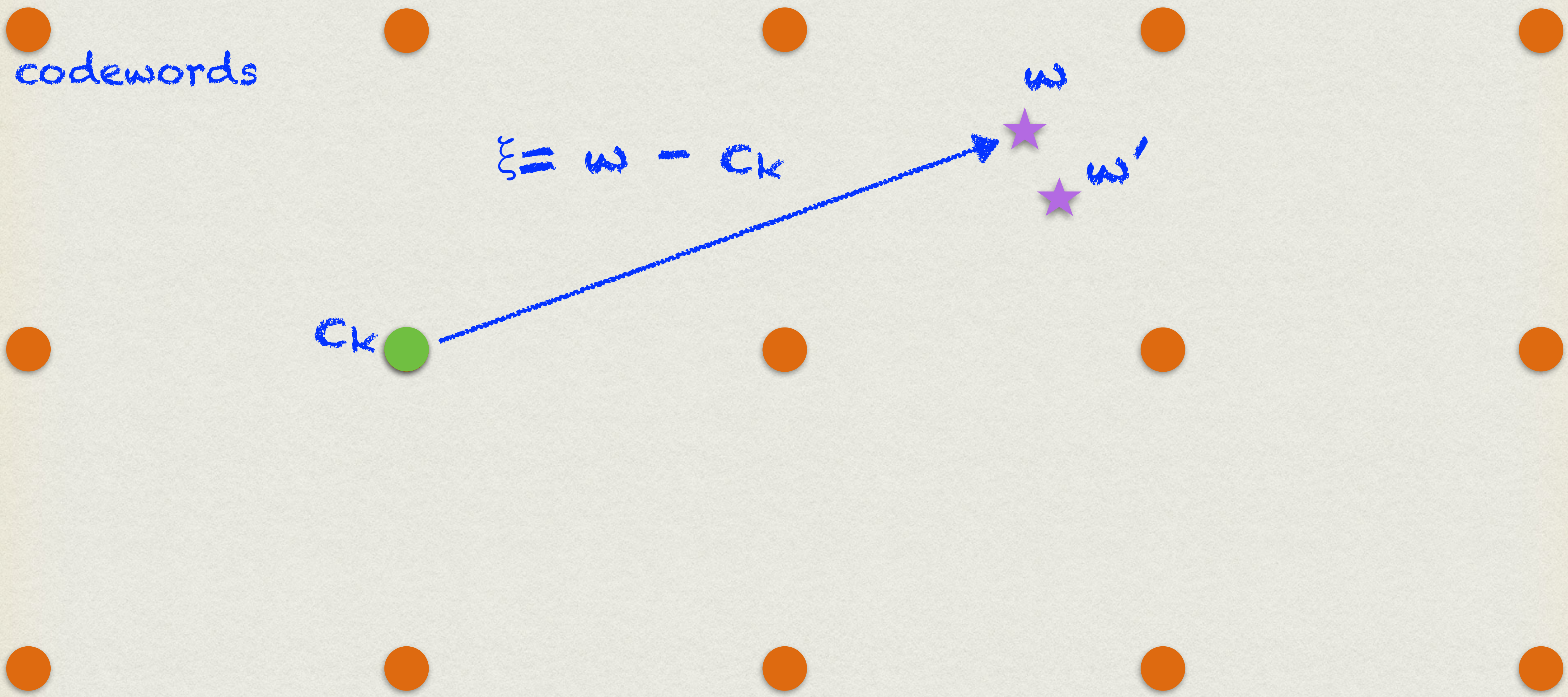
CORE PRINCIPLE ILLUSTRATED



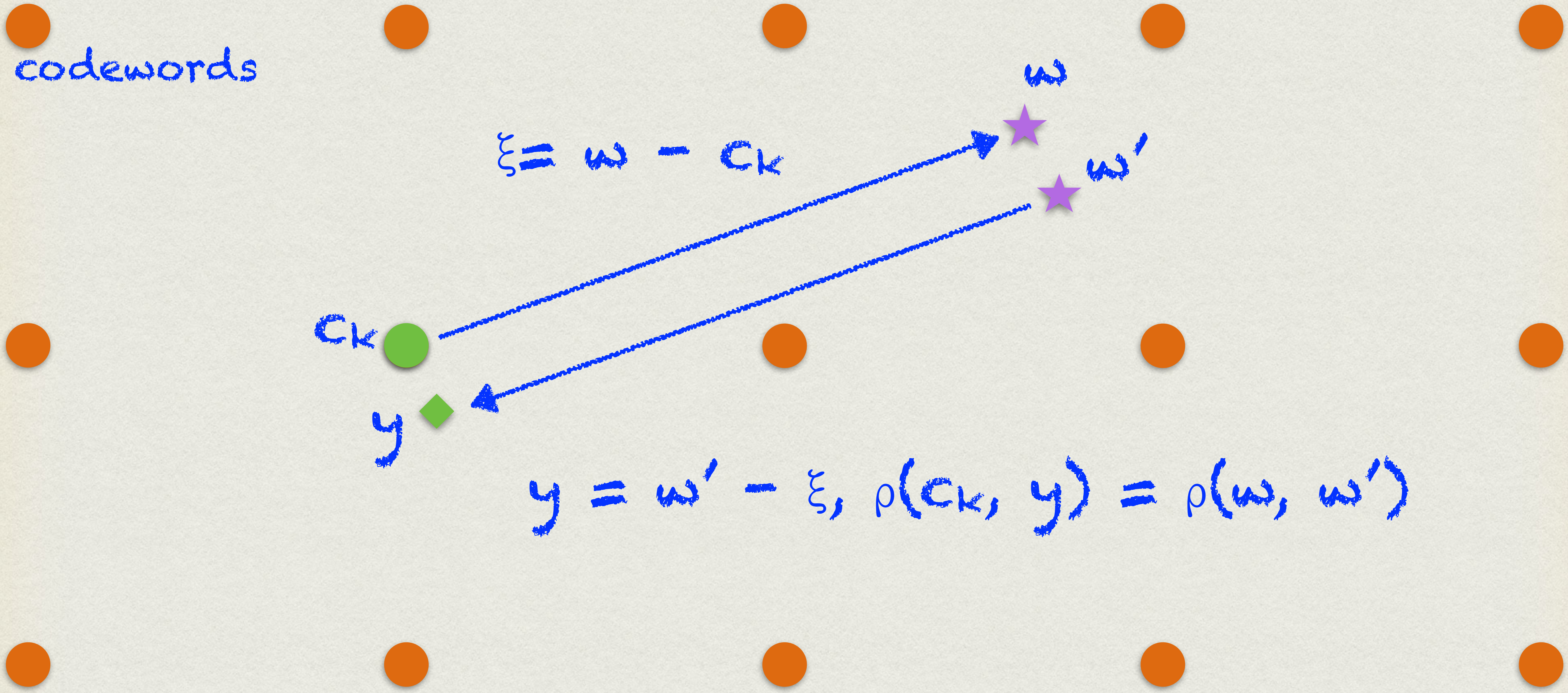
CORE PRINCIPLE ILLUSTRATED



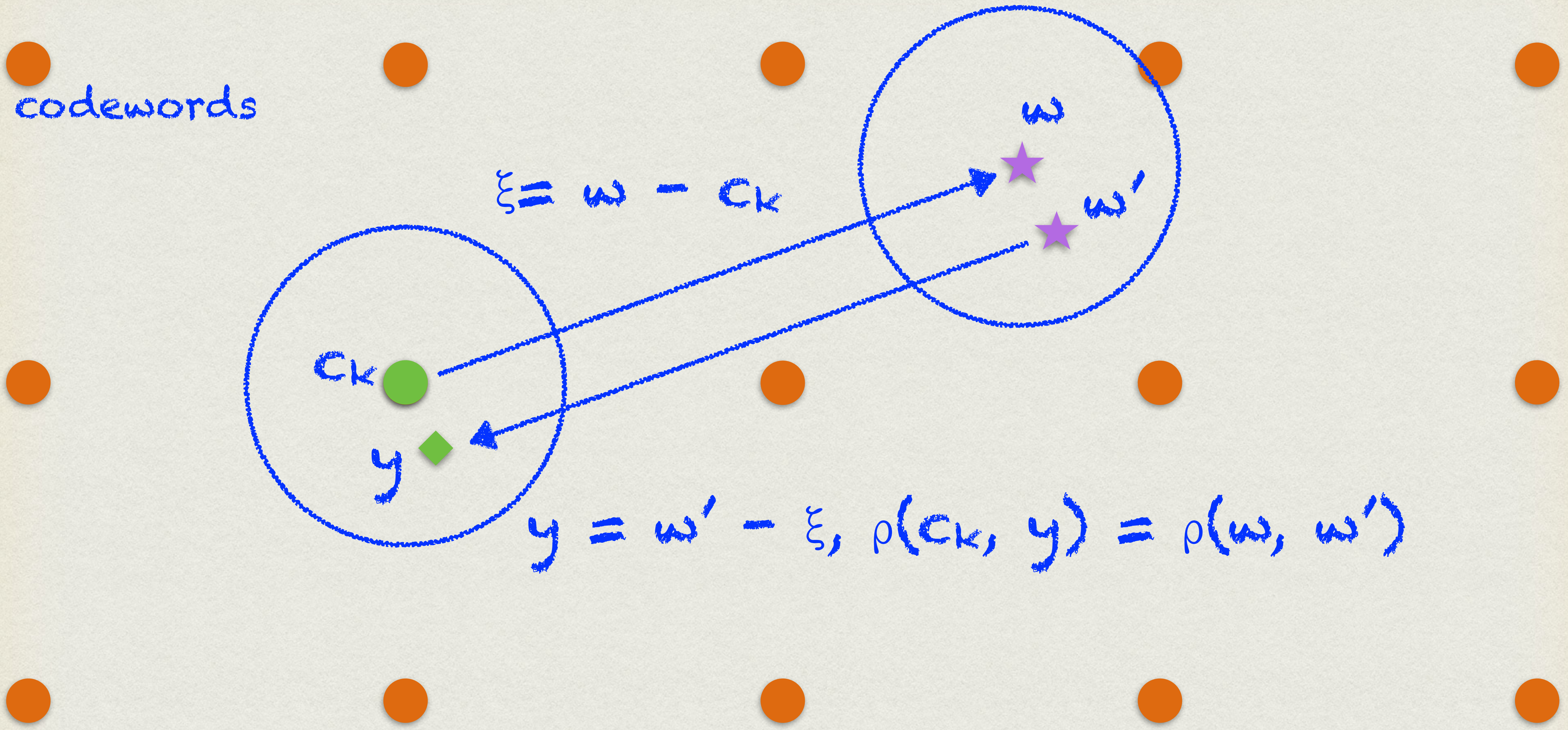
CORE PRINCIPLE ILLUSTRATED



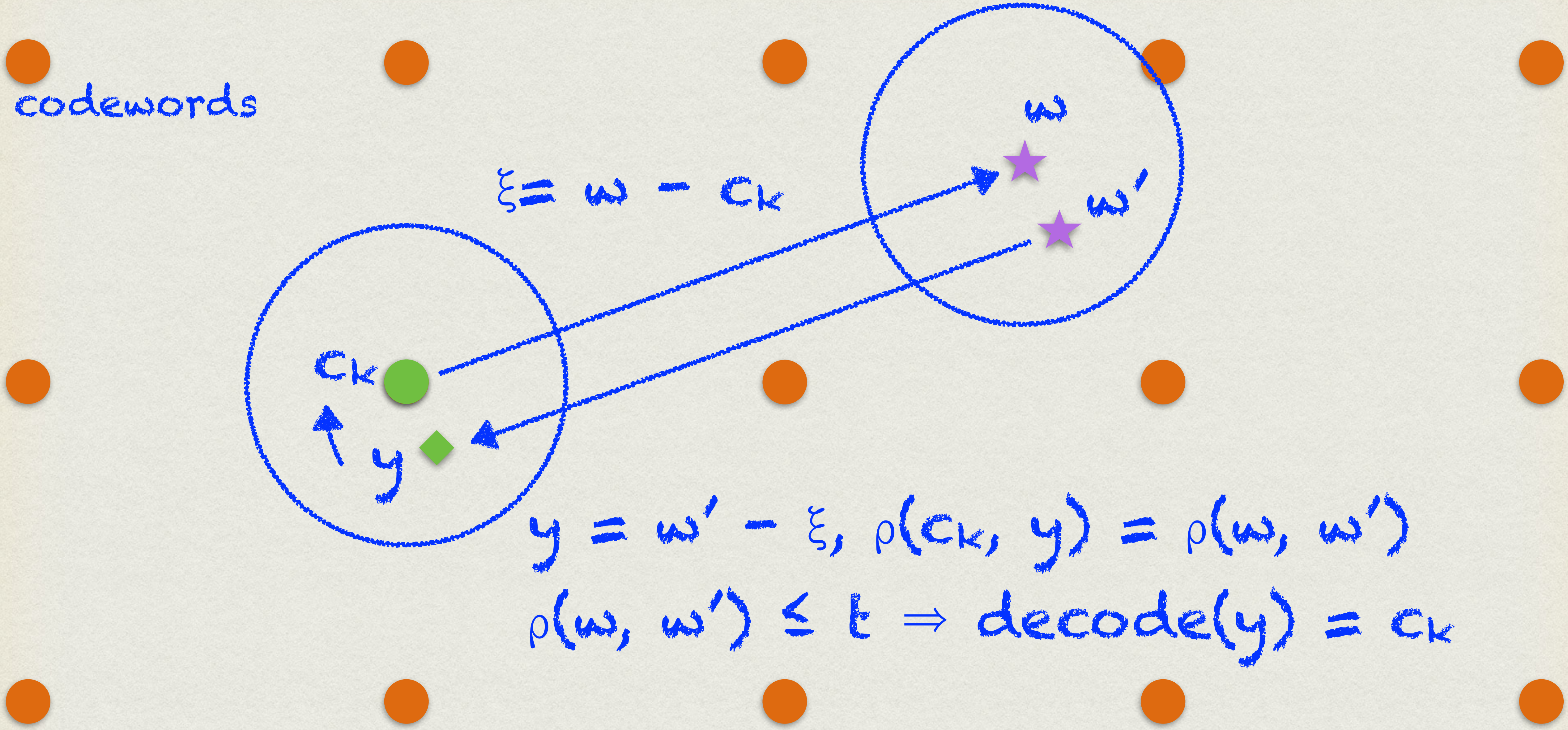
CORE PRINCIPLE ILLUSTRATED



CORE PRINCIPLE ILLUSTRATED



CORE PRINCIPLE ILLUSTRATED



$$y = w' - \xi,\ \rho(c_k, y) = \rho(w, w')$$

$$\rho(w, w') \le t \Rightarrow \text{decode}(y) = c_k$$

IS IT ENOUGH?

- Template protection in contemporary systems is often quite questionable (to be polite).
- On the other hand, is it the only one problem?
  - No. We shall not push the concept of bio-keys too hard anyway.

BIO-SKIMMING

- Once biometric systems become ubiquitous, this will be a fruitful attack vector.
  - Attackers use a fake sensor (or hack into an original one) to skim the "bio-master-key".
  - At the end of the day, how many eyes, fingers, faces, vocal tracts (etc.) do we have?
  - It is like having few master-keys for a whole life.
  - Furthermore, we prove the master-key possession by simply handing it over to almost any device that asks so (again, again, ...and again).

SPOOFING STILL MATTERS!

- That said, liveness detection will be always important!
  - Remember, biometrics is a signal detection.
  - It all works as long as we can assume the signal is coming from a particular human being!
    - *Apparently, the biometric signal detector output shall be just one out of many inputs into an authentication process (itself being another multidimensional signal detection problem).*

TAMPER-RESISTANT SENSOR

- It signs the biometric signal samples with its private key to indicate it already has sampled that signal from a living individual.

    - Furthermore, the sample shall be then processed as soon as possible.

    - Otherwise, we have to mitigate the risk of a sensor compromise in the intermediate time by a further time-stamping: Long Term Validation of bio-samples.

    - This concept is all too often neglected in the emerging handwritten signature biometrics!

ANYWAY, DO THE PENTEST!

CONCLUSION

- We shall require ISO 19795 methodology during biometric application selection, comparison, and operational testing.
- Use an independent penetration test to verify:
  - zero-effort attack complexity

    –*beware of automated APIs!*
  - masquerade attacks
  - spoofing possibilities
  - template security
  - system security in general

    –*threshold settings, template tampering*