BIOMETRIJSKI SUSTAVI - GREŠKE I RANJIVOSTI

Zlatko Sirotić, dipl.ing.
Istra informatički inženjering d.o.o.
Pula

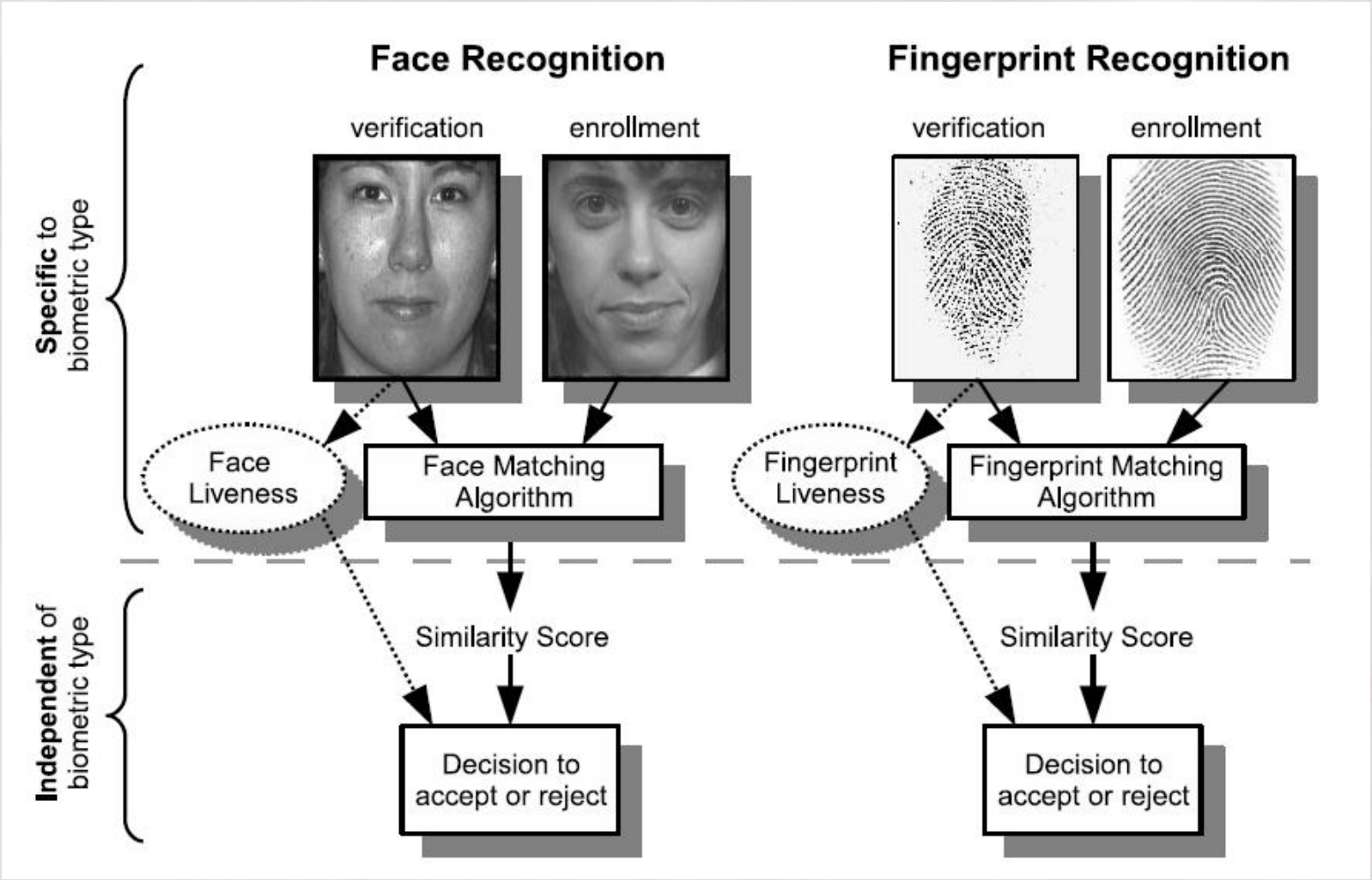
Biometrija

- Ozbiljna biometrijska istraživanja počela su šezdesetih godina 20.stoljeća. Od sredine devedesetih godina biometrija se značajno primjenjuje u praksi.
- Biometrija koristi **fiziološke ili ponašajne** (engl.behavioral) **karakteristike** određene ljudske individue da bi ga automatski identificirala.
- Postoje različite **biometrijske tehnologije**. Najstarija je tehnologije ona otiska prsta, nastala oko 1960., komercijalizirana oko 1980.
- Slijedile su je mnoge druge biometrijske tehnologije, npr. mrežnice i šarenice oka, geometrije lica, geometrije ruke, otiska dlana, geometrije krvnih žila ruke ili prsta, glasa i dr.

Opći biometrijski sustav

- Svaka od ovih tehnologija koristi različite **biometrijske senzore** i različite **algoritme za uparivanje** (engl.matching) biometrijskih podataka pročitanih senzorom i (prije) snimljenih biometrijskih podataka.
- Bez obzira na različitost, sve ove biometrijske tehnologije imaju nešto zajedničko, a to je proces: ulaz podataka – obrada podataka – izlaz podataka. Pritom se biometrija obilato koristi oruđem **matematičke statistike**.
- U prezentaciji se daje prikaz biometrije neovisan od tehnologije, a naglasak se daje na **greške i ranjivosti općeg biometrijskog sustava**.

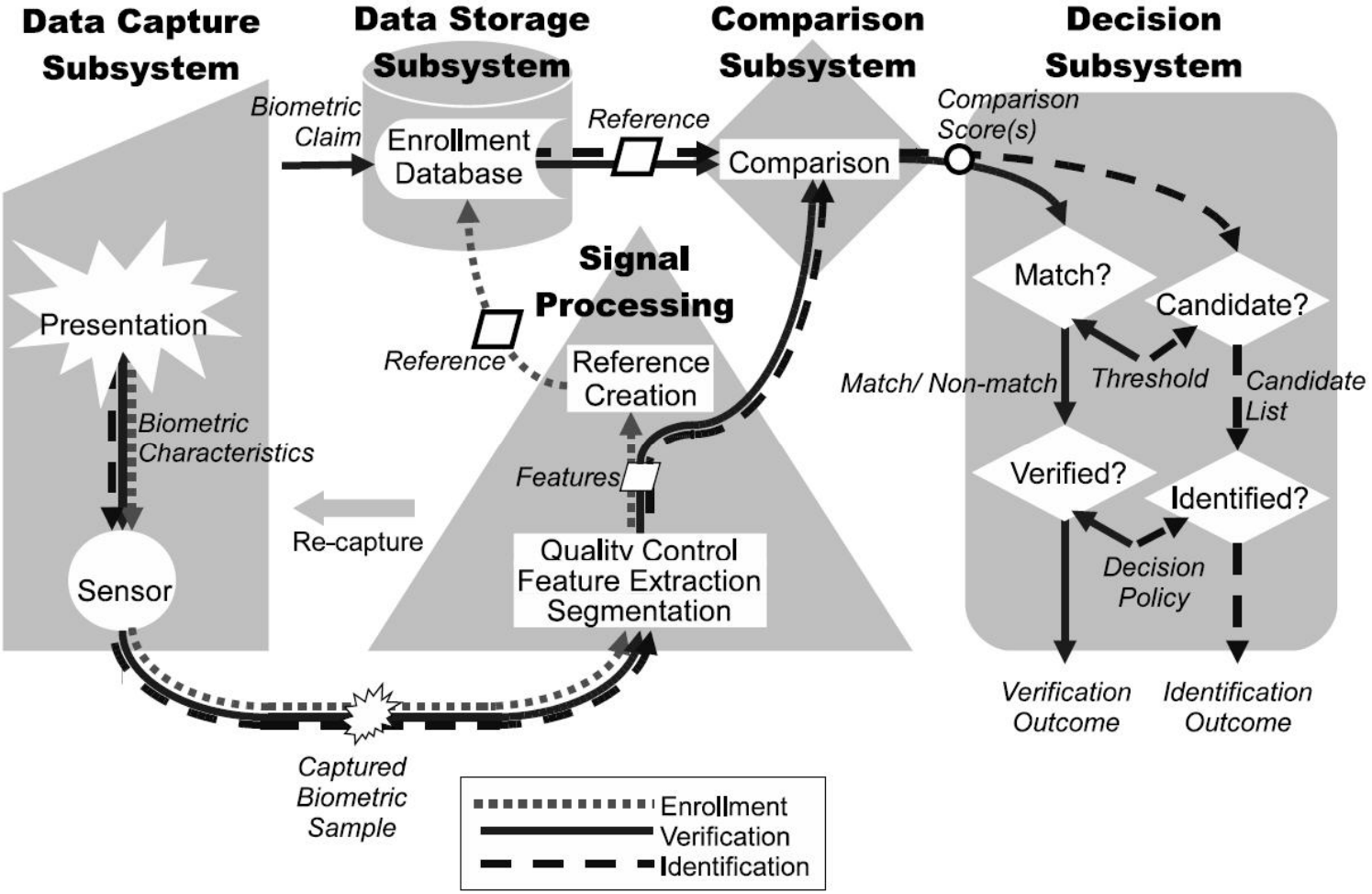
Što je specifično, a što nezavisno kod rada sa određenom biometrijskom karakteristikom



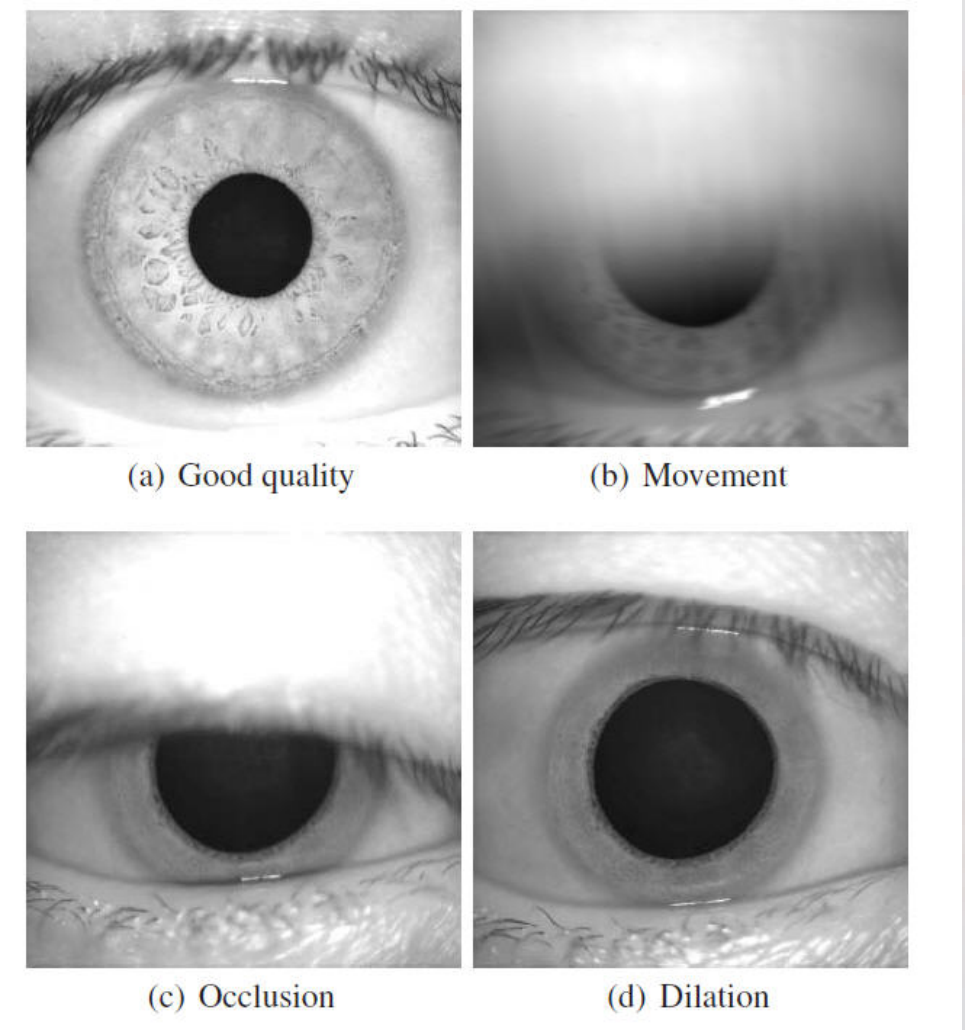
Neki pojmovi

- **Biometrija** (engl. biometrics) je automatska identifikacija osobe na temelju njenih fizioloških (bioloških) ili ponašajnih karakteristika.
- **Biometrijska karakteristika** (engl. biometric) je mjerljiva fiziološka (biološka) ili ponašajna karakteristika koja se koristi za prepoznavanje osobe. Biometrijska karakteristika ima ova četiri svojstva: svaka je osoba mora imati, treba biti dovoljno različita kod različitih osoba, treba ostati konstantna kroz vrijeme, mora biti mjerljiva kvantitativno (a ne samo opisno).
- **Biometrijski modalitet** (engl.modality) je vrsta biometrijske karakteristike. Npr., tri uobičajena biometrijska modaliteta su lice, otisak prsta, glas.
- **Multibiometrija** (engl.multibiometrics) je automatska identifikacija osobe na temelju dvije ili više biometrijskih karakteristika.

Biometrijski sustav kao sustav za obradu podataka



Biometrijski uzorak dobre i loše kvalitete - na primjeru šarenice (irisa) oka



Matematička statistika je temelj biometrije

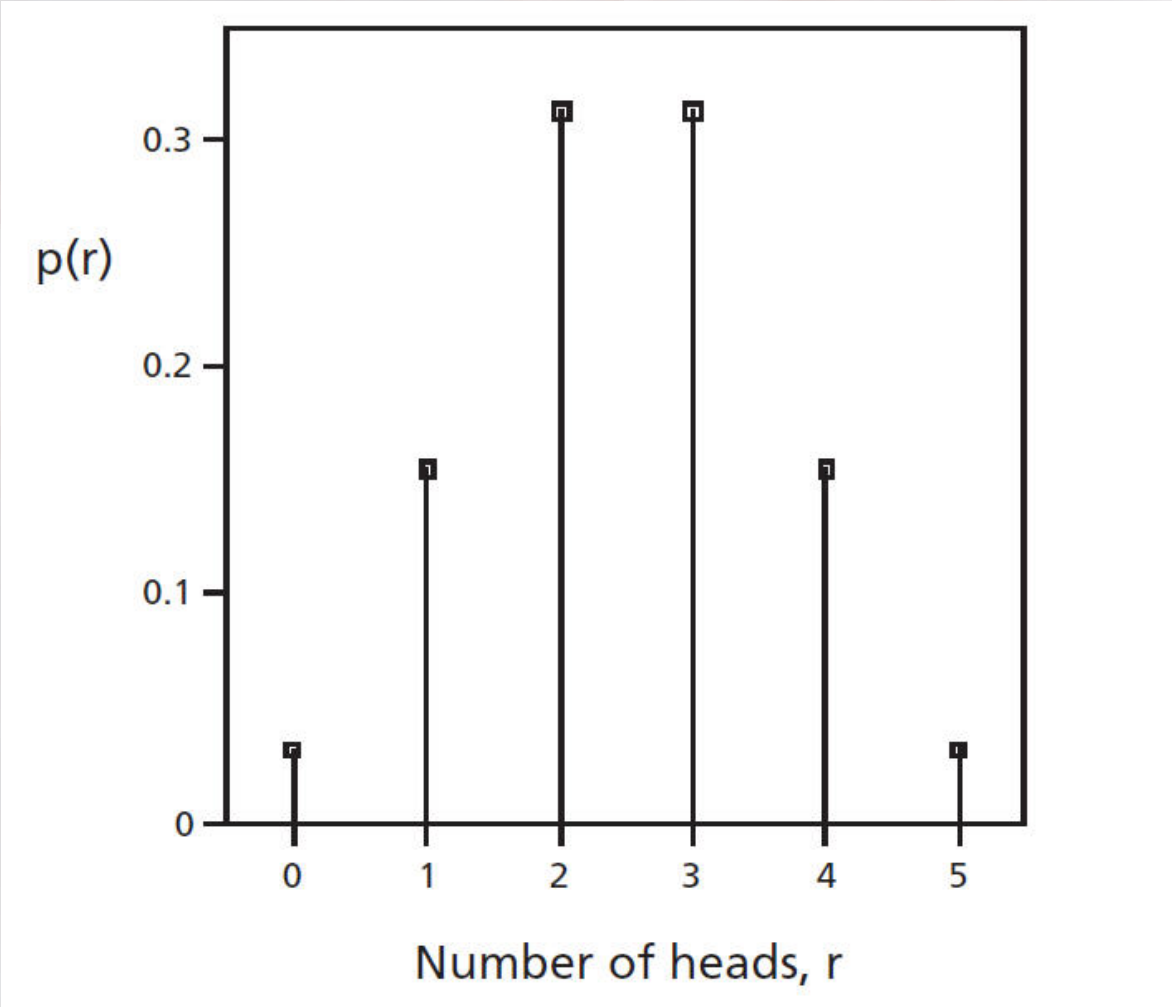
- Postoje diskontinuirane (diskretne) i kontinuirane slučajne varijable.

- **Diskontinuirana slučajna varijabla** takva je varijabla *x* koja
  - prima niz vrijednosti x_1, x_2, …
  - ali svaku od njih s određenom vjerojatnošću $p(x_1)$, $p(x_2)$, …
  - pri čemu vjerojatnosti $p(x_i)$ zadovoljavaju jednakost

    $$\sum p(x_i) = 1$$

- Zakon p(x) po kojem svakoj vrijednosti x_i pripada vjerojatnost $p(x_i)$ zovemo **funkcijom vjerojatnosti slučajne varijable** x.

Primjer:
Funkcija vjerojatnosti varijable r
(broj glava kod bacanja pet novčića)

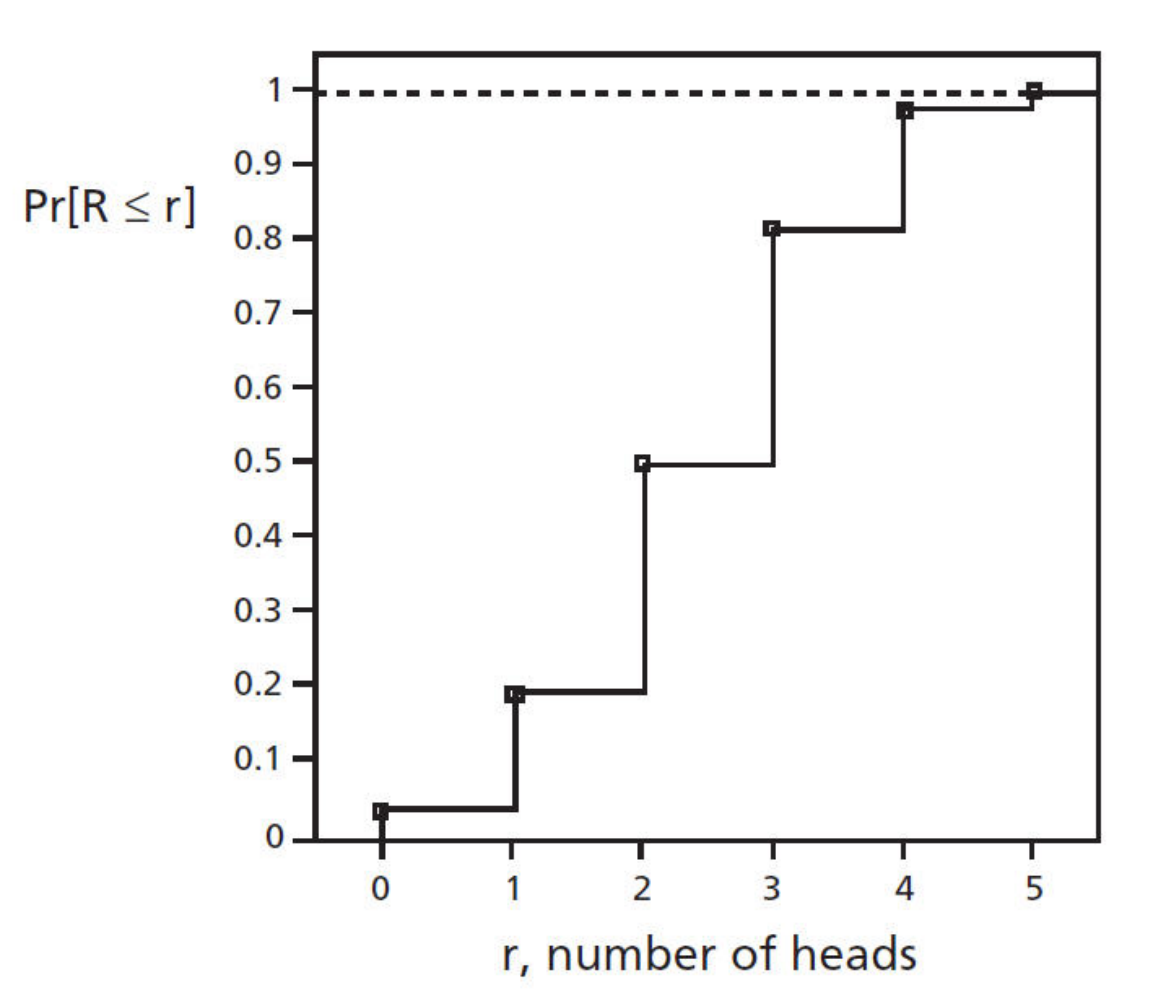


Funkcija distribucije slučajne varijable

- Osim funkcije vjerojatnosti, kod diskretnih slučajnih varijabli važna je **funkcija distribucije slučajne varijable** (ili funkcija kumulativne distribucije, engl. Cumulative Distribution Functions).
- Ona pokazuje kolika je vjerojatnost da slučajna varijabla x poprimi bilo koju vrijednost $\leq x_0$:

$$F(x_0) = \sum_{x_i \leq x_0} p(x_i) \qquad \text{tj. } F(x_0) = P\{x \leq x_0\}$$

Primjer:
Funkcija distribucije varijable r
(broj glava kod bacanja pet novčića)

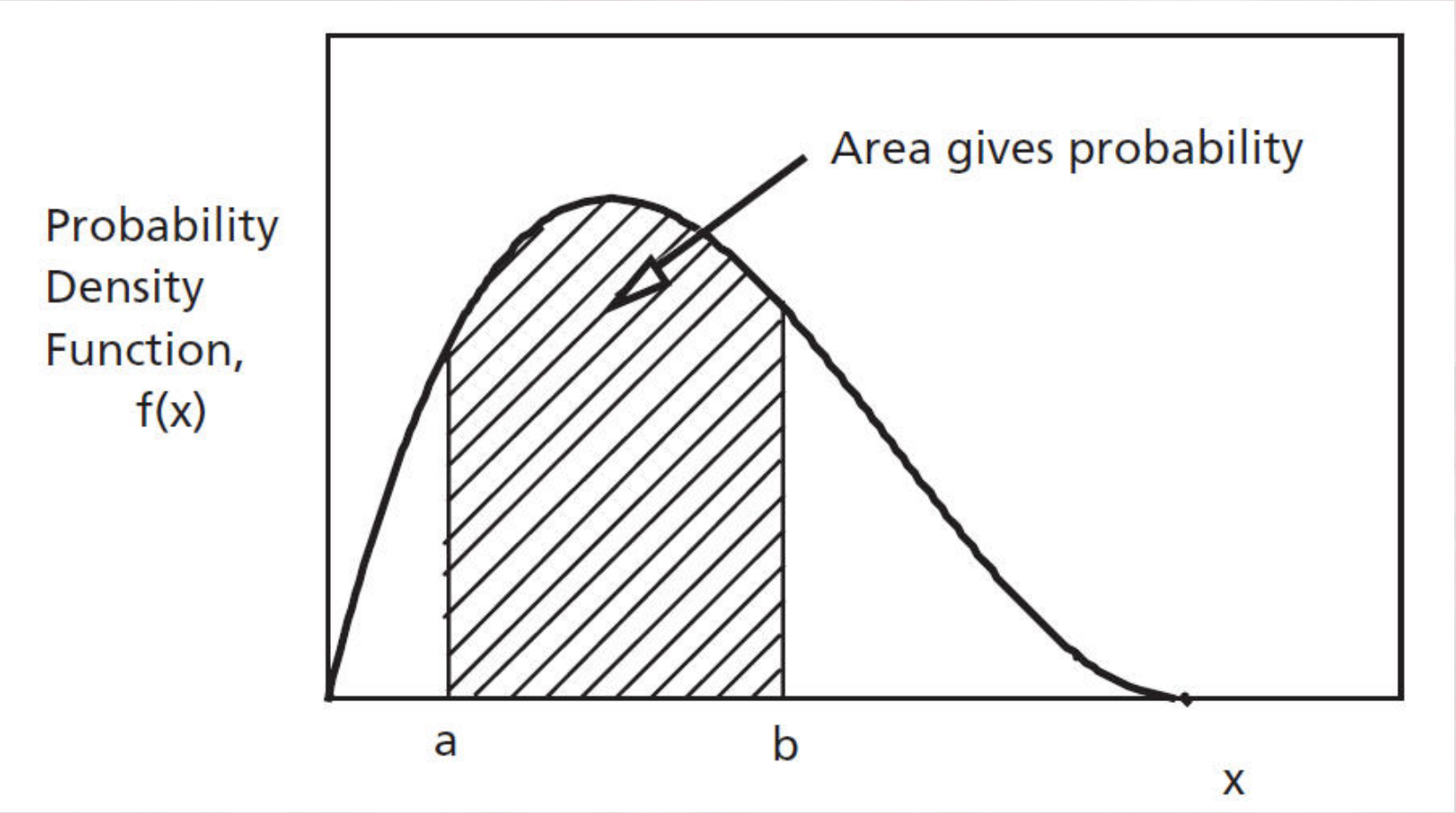


Kontinuirane slučajne varijable

- **Funkcija vjerojatnosti kontinuirane slučajne varijable** x (ili funkcija gustoće vjerojatnosti, engl.Probability Density Function) je takva funkcija f(x) koja ima svojstva:

1. $f(x) \geq 0$ za svaki x iz domene funkcije [a, b]
   (a može biti $-\infty$, b može biti $+\infty$)
2. $\int_a^b f(x)\,dx = 1$
   (površina ispod funkcije unutar domene [a, b] je 1)
3. $\int_{x1}^{x2} f(x)\,dx = P\{x_1 \leq x \leq x_2\}$
   (površina ispod funkcije unutar domene $[x_1, x_2]$ jednaka je vjerojatnosti da varijabla poprimi vrijednost iz te domene).

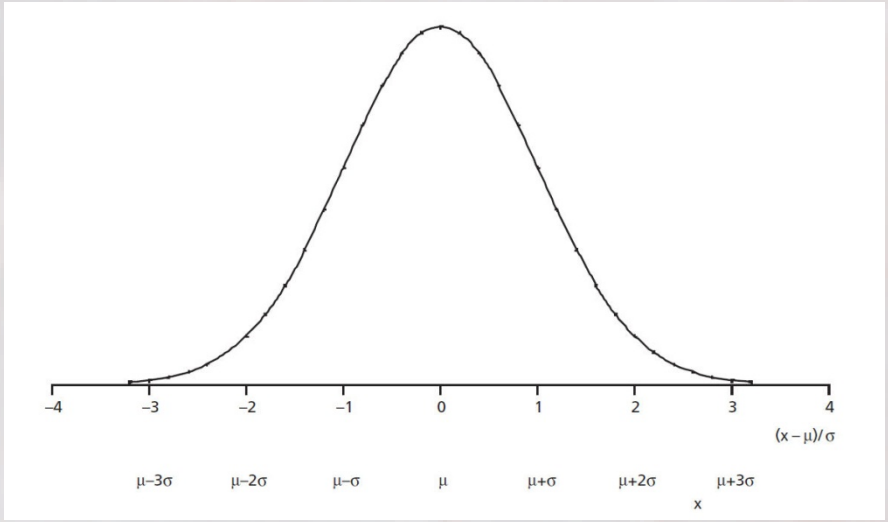
Vjerojatnost kod kontinuirane slučajne varijable

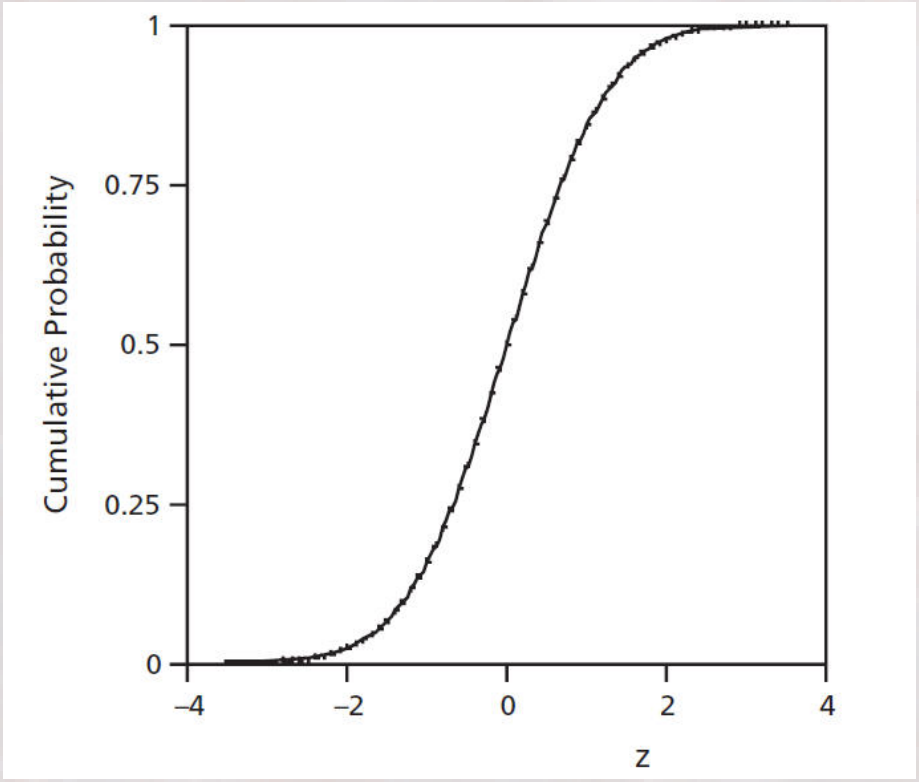


Normalna razdioba (Gaussova krivulja)

- Jedna od najpoznatijih funkcija vjerojatnosti (kontinuirane varijable) je tzv. **normalna razdioba** (poznata i kao Gaussova krivulja, po matematičaru Gaussu)
- Važna je po tome što mnoge druge razdiobe (diskontinuirane i kontinuirane) graniče prema njoj ako neki parametri rastu u beskonačnost.
- Posebno postoji tzv. **jedinična (ili standardna) normalna razdioba**, kod koje je matematičko očekivanje = 0, a standardna devijacija = 1.

Funkcija (gustoće) vjerojatnosti i funkcija distribucije vjerojatnosti kod (jedinične) normalne razdiobe





Testiranje statističkih hipoteza

- Funkcije vjerojatnosti slučajne varijable x ovise o parametrima, npr. parametri ne-jedinične normalne razdiobe su matematičko očekivanje i standardna devijacija.
- Ako jedan **nepoznati parametar promatramo kao varijablu**, a ostale kao konstantu, onda možemo postaviti **hipotezu** H_0 da je vrijednost tog parametra npr. P_0, te alternativnu hipotezu H_1, da je vrijednost parametra P_1 (moguće su i drugačije varijante postavljanja hipoteza).
- Odluku o tome da li prihvaćamo hipotezu H_0 ili H_1 donosimo na temelju **testiranja uzorka**, koji je uvijek konačan. Kod testiranja, **moguće su četiri situacije, dvije u kojima smo donijeli ispravnu odluku i dvije u kojima smo donijeli pogrešnu odluku.**

Testiranje statističkih hipoteza – mogući ishodi

Hipoteza H_0	Istinita	Neistinita
Odbacuje se	**Greška 1.vrste (vjerojatnost je α)**	Pravilan zaključak
Prihvaća se	Pravilan zaključak	**Greška 2.vrste (vjerojatnost je β)**

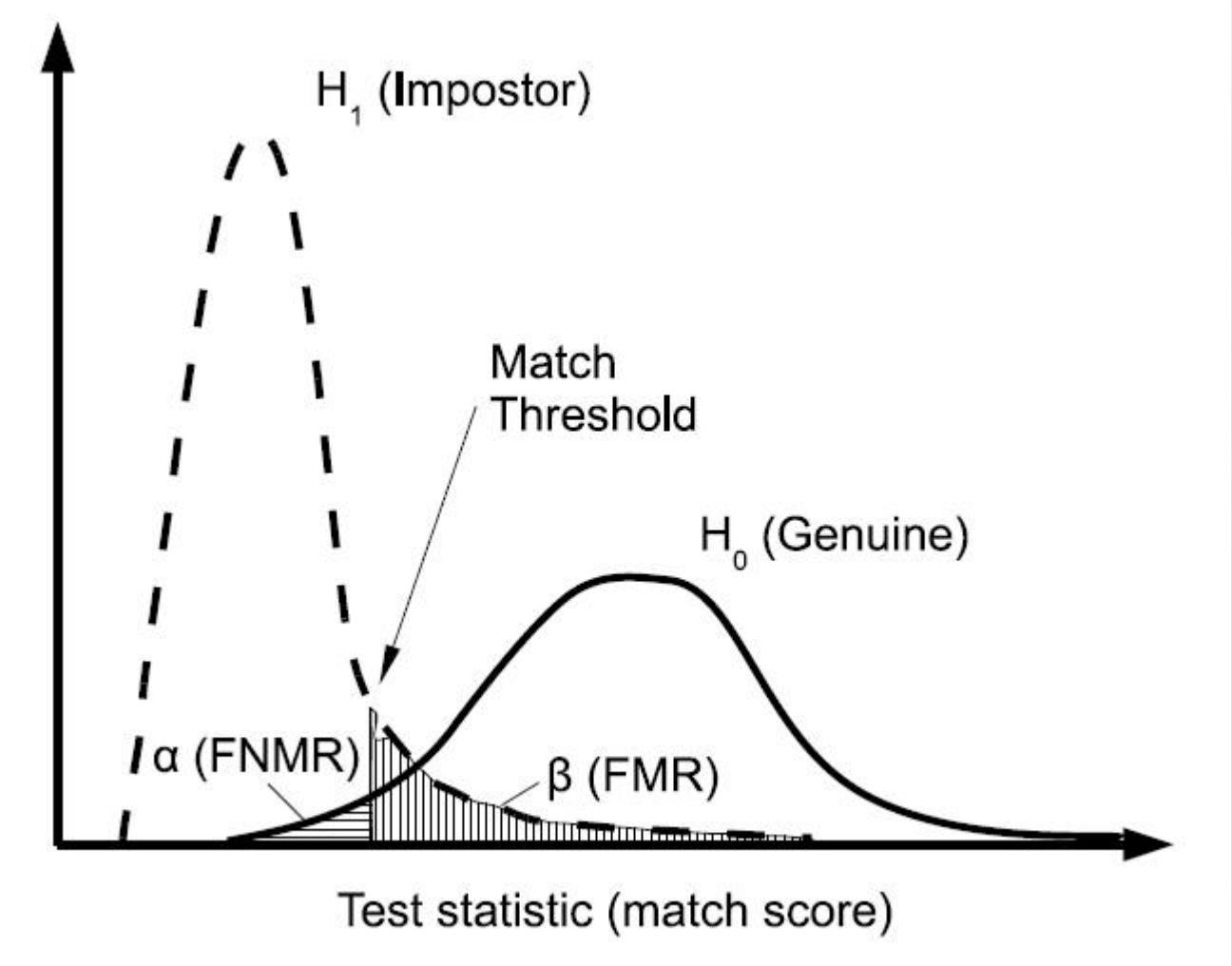
Nepoznate definicije nekih poznatih pojmova ☺

- **Optimist** – onaj koji radije prihvaća vrlo veliku grešku β (pogrešno prihvaćanje krivog), nego grešku α (pogrešno odbacivanje ispravnog)
- **Pesimist** - onaj koji radije prihvaća vrlo veliku grešku α (pogrešno odbacivanje ispravnog) nego grešku β (pogrešno prihvaćanje krivog)
- **Realist** – nalazi pravu mjeru između α i β
- **Idealist** – vjeruje da se istovremeno mogu imati mali α i β

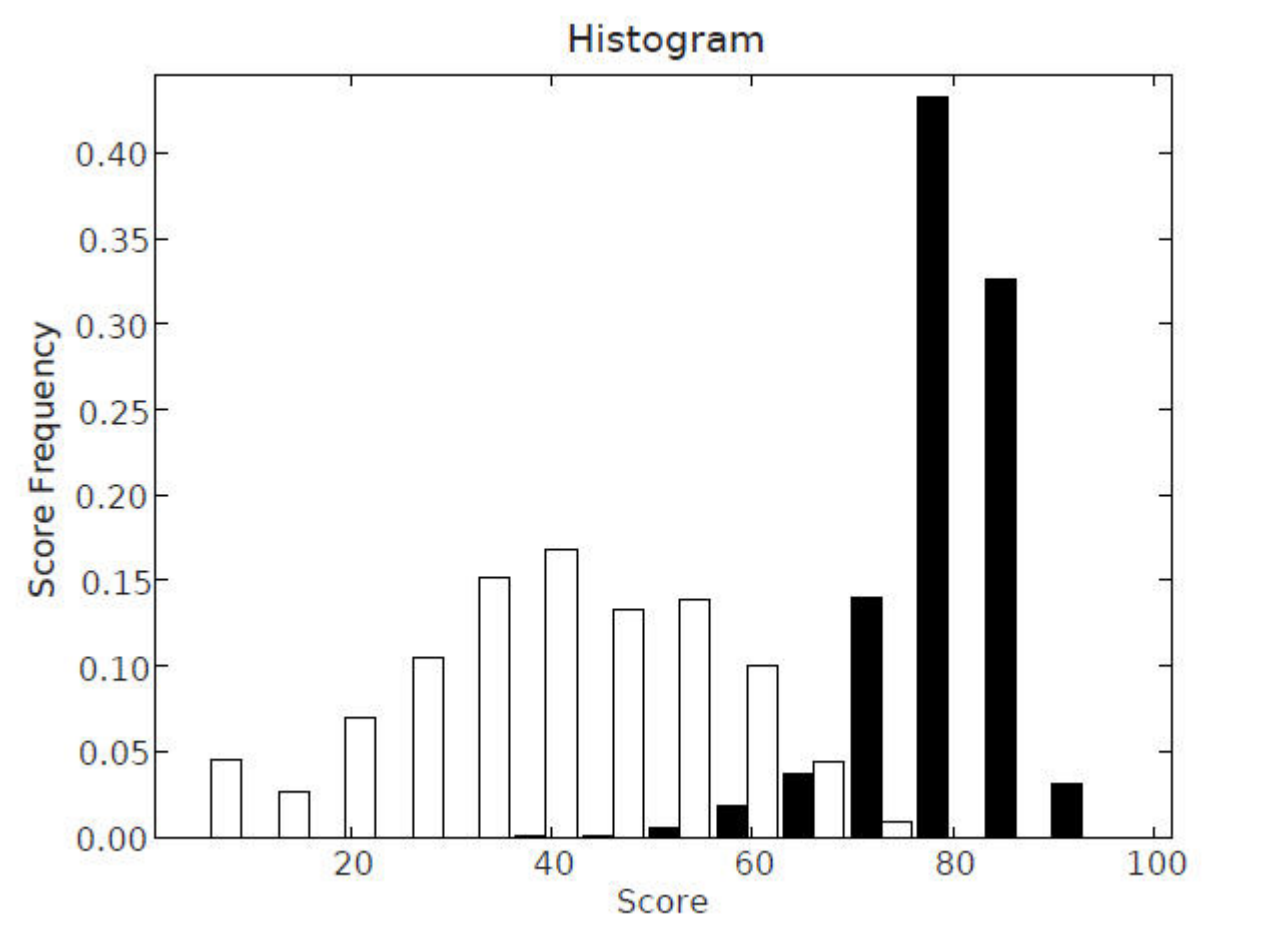
Biometrijska verifikacija (autentifikacija)

- Želimo utvrditi da li je osoba koja je dala biometrijski uzorak **zaista ta osoba kojom se predstavlja**.
- Postavljamo (biometrijsku) hipotezu H_0 da je riječ o pravoj osobi. Ako je rezultat usporedbe veći ili jednak od zadanog **praga** (eng.threshold), prihvaćamo hipotezu H_0, inače prihvaćamo hipotezu H_1 (smatramo da je osoba lažna).
- U biometriji se vjerojatnost greške 1.vrste (α) naziva **FNMR (engl.False Non-Match Rate),** slobodno prevedeno - stopa pogrešnog odbacivanja, a vjerojatnost greške 2.vrste (β) **FMR (eng.False Match Rate)**, slobodno prevedeno - stopa pogrešnog prihvaćanja.
- Funkcija vjerojatnosti (kontinuirane) slučajne varijable (koja predstavlja rezultat usporedbe) **jedna je u slučaju da je osoba prava, a druga u slučaju da je osoba lažna!**

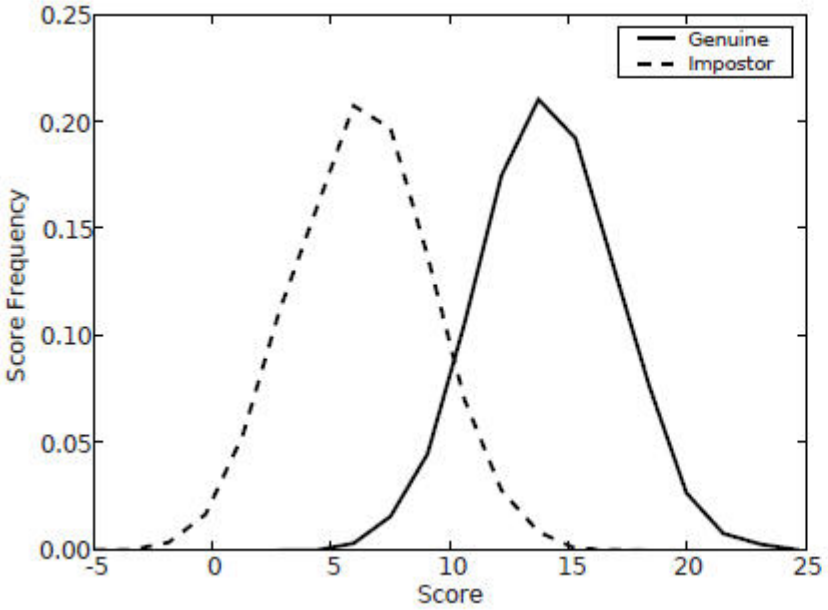
Imamo dvije funkcije vjerojatnosti – za pravu i lažnu osobu

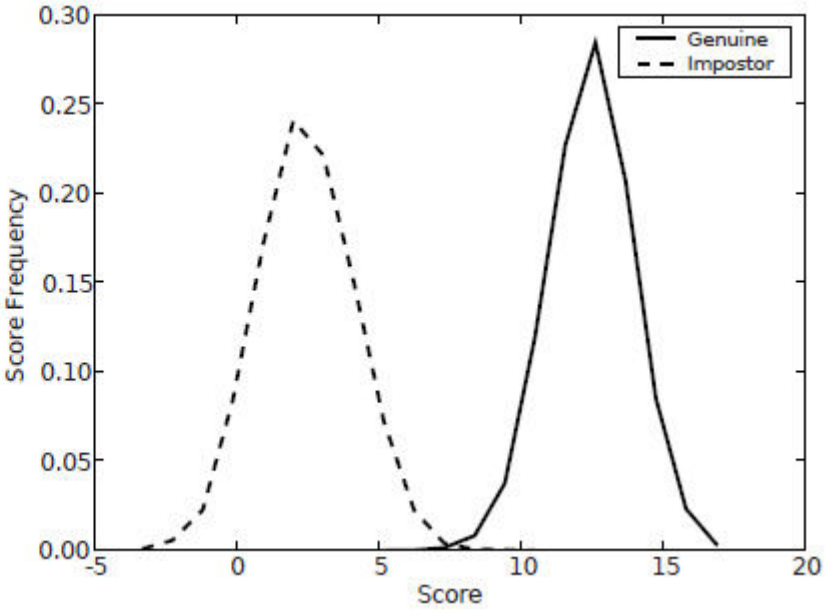


U praksi (uglavnom) ne raspolažemo s gotovim funkcijama vjerojatnosti - kreiramo ih pomoću uzoraka

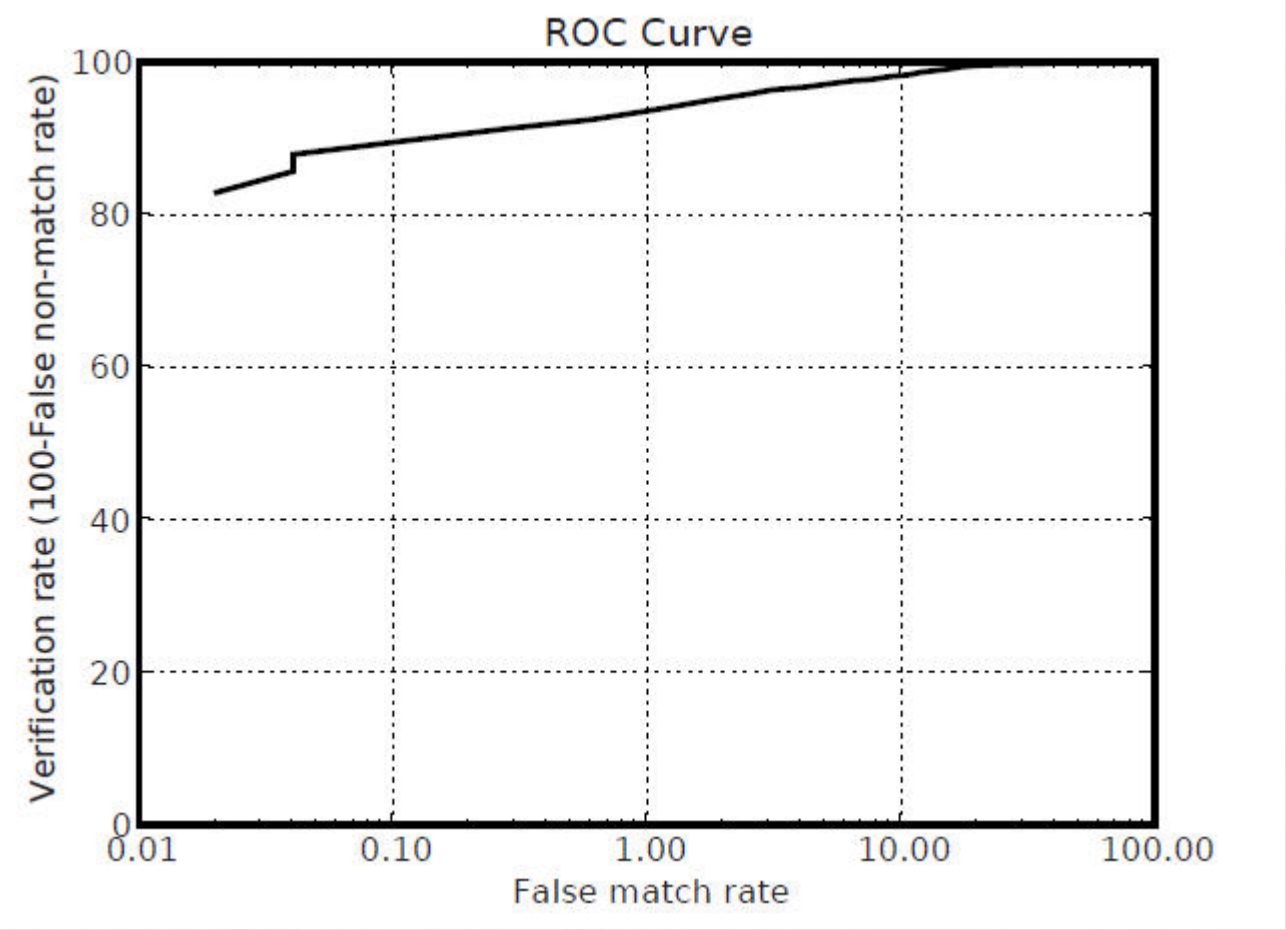


Dva primjera preklapanja funkcija vjerojatnosti – veliko preklapanje (loše) i malo preklapanje (dobro)





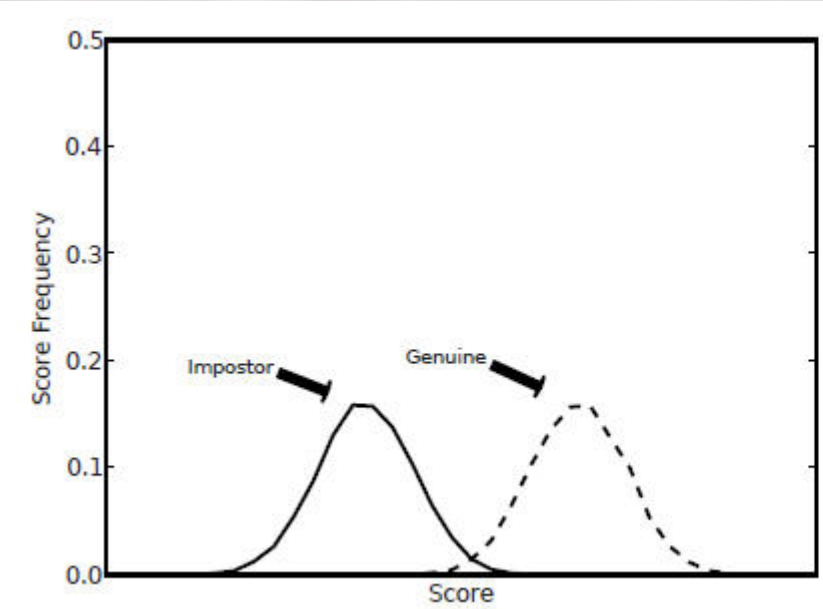
Evaluacija sustava za verifikaciju pomoću ROC krivulje (Receiver Operating Characteristic)

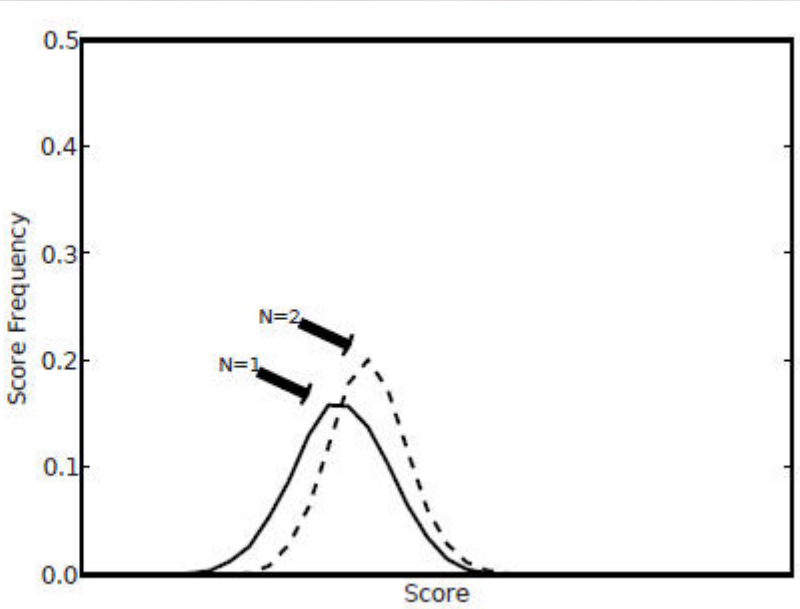


Biometrijska identifikacija – složenija je od verifikacije

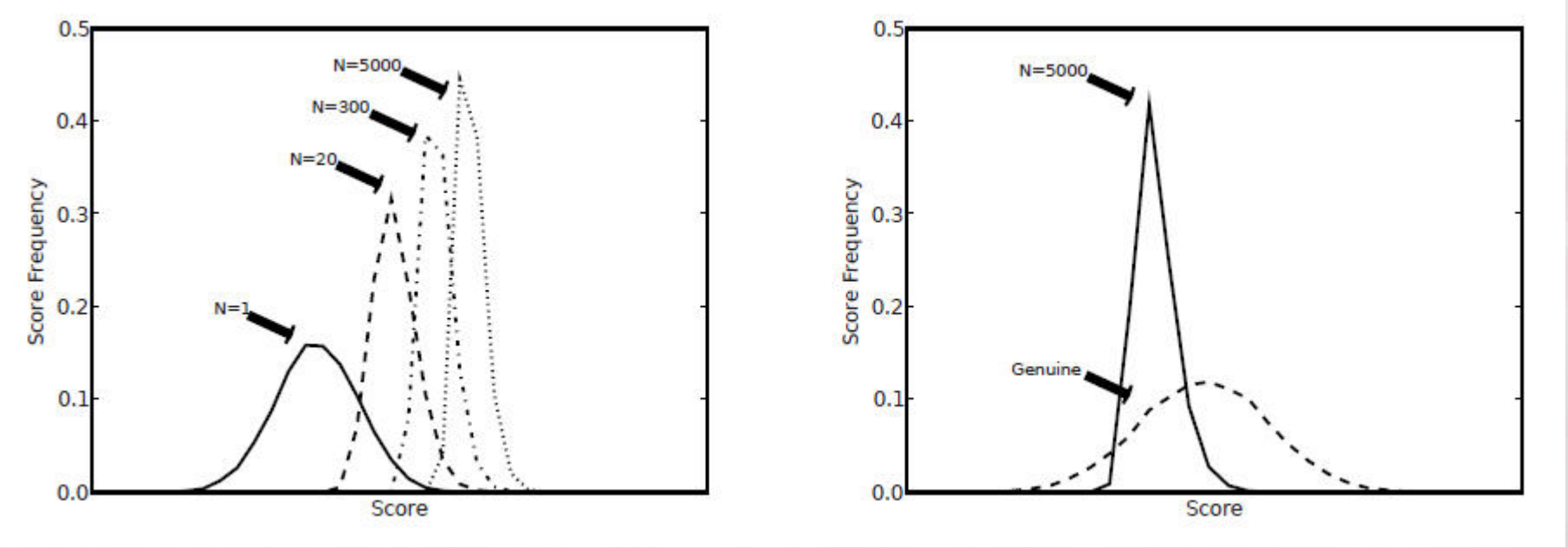
- Kod identifikacije želimo utvrditi **"Tko je ova osoba?"**
- Moguće je da osoba koja se prijavljuje u biometrijski sustav **uopće nije upisana u biometrijsku bazu**.
- Verifikacija je 1:1 usporedba između biometrijskog uzorka i predloška iz biometrijske baze. **Identifikacija je 1:N usporedba**, jer se biometrijski uzorak u općem slučaju mora usporediti sa svakim predloškom iz baze.
- Identifikacijski sustav je bitno različit od verifikacijskog sa stanovišta preciznosti. I kod verifikacije, preciznost je ovisila o preklapanju razdioba za pravu i lažnu osobu. No, **kod identifikacije će preciznost padati s porastom veličine biometrijske baze**.

Biometrijska identifikacija - preklapanje između funkcija vjerojatnosti se povećava s povećavanjem baze





Biometrijska identifikacija - preklapanje između funkcija vjerojatnosti se povećava s povećavanjem baze

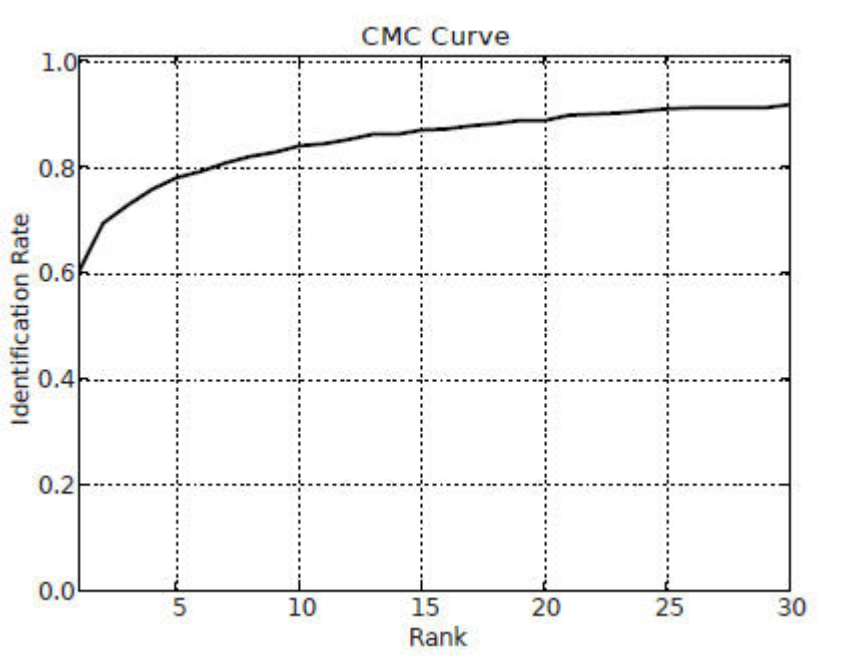


Sustavi za identifikaciju temeljeni na rangu

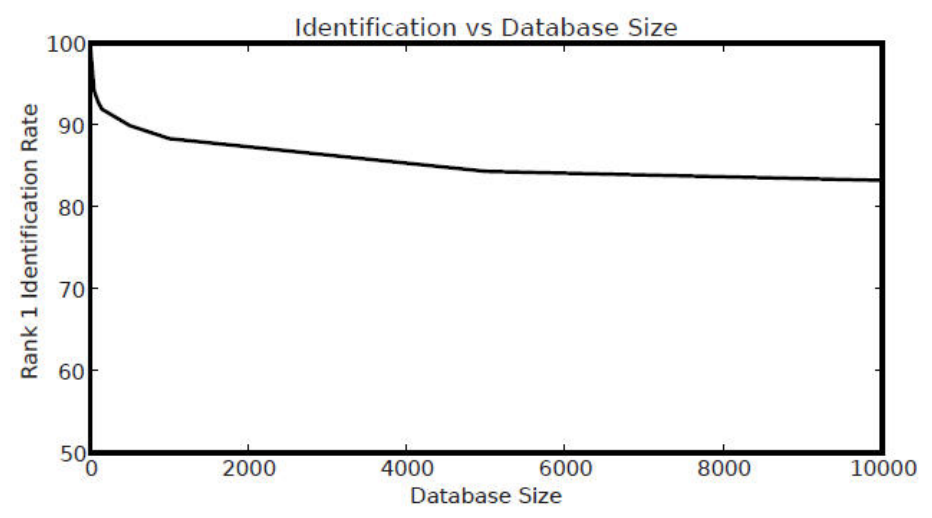
- Identifikacijski sustavi najčešće rade tako da **operator pogleda rezultate** koje mu je predočio biometrijski sustav.
- Sustav daje rezultate na **dva uobičajena načina**. Jedan je da se operateru prikaže prvih nekoliko osoba (tzv.**rang**) koje imaju najveći rezultat usporedbe, a drugi je da se prikažu sve osobe čiji je **prag usporedbe** veći od zadanog praga
- Na prvi način rade **sustavi za identifikaciju temeljeni na rangu** (rank-based) i najčešće se koriste kada se identifikacija radi na zatvorenom skupu (osoba sigurno postoji u bazi, što je u praksi rjeđa situacija).
- Najvažnija krivulja koja se tada koristi je **CMC krivulja (eng.Cumulative Match Characteristic)**, koja prikazuje ovisnost identifikacijske stope od veličine ranga (veličine kandidacijske liste, broja kandidata za identifikaciju).

a) CMC krivulja (Cumulative Match Characteristic) b) Ovisnost identifikacijske stope od veličine baze

a)



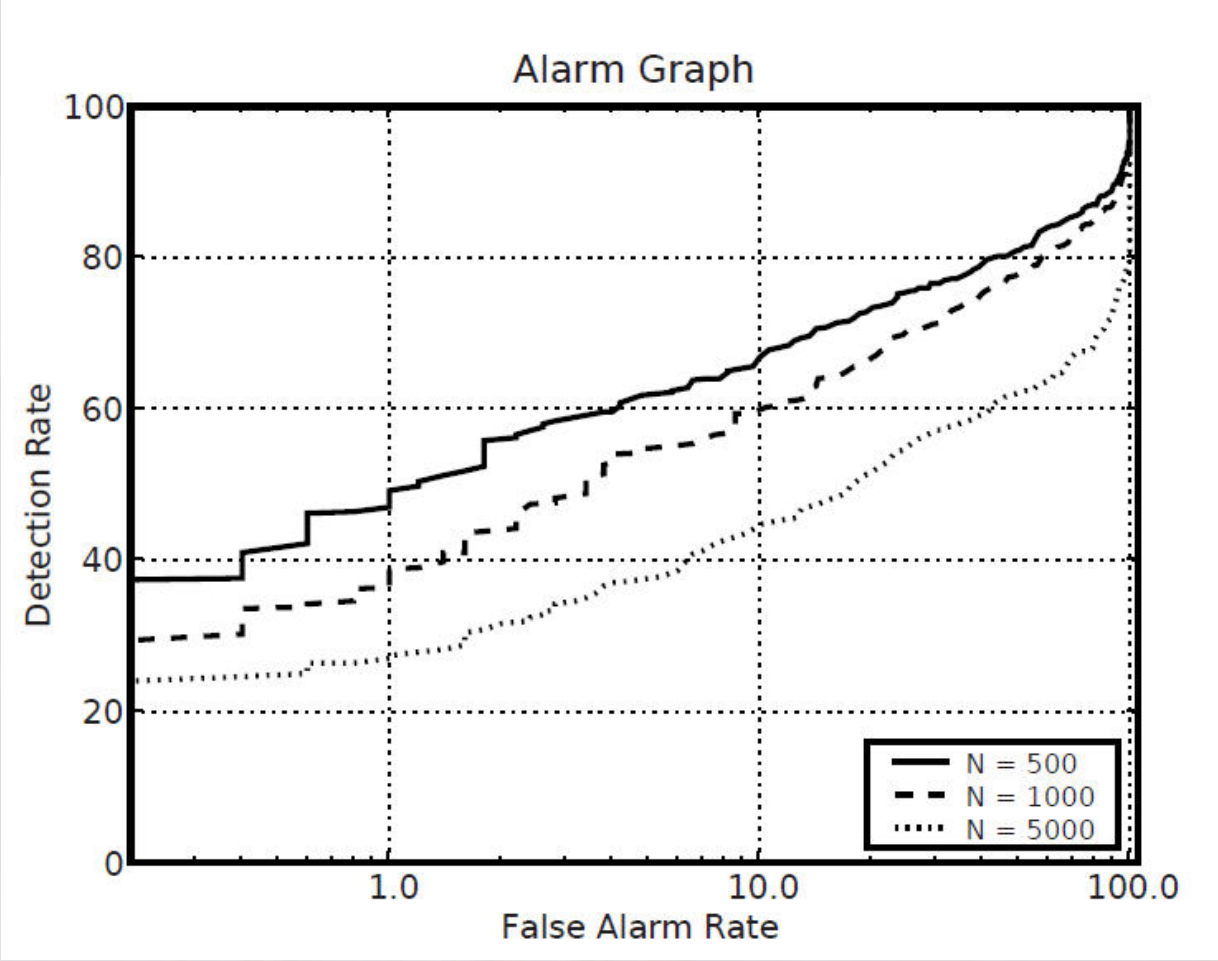
b)



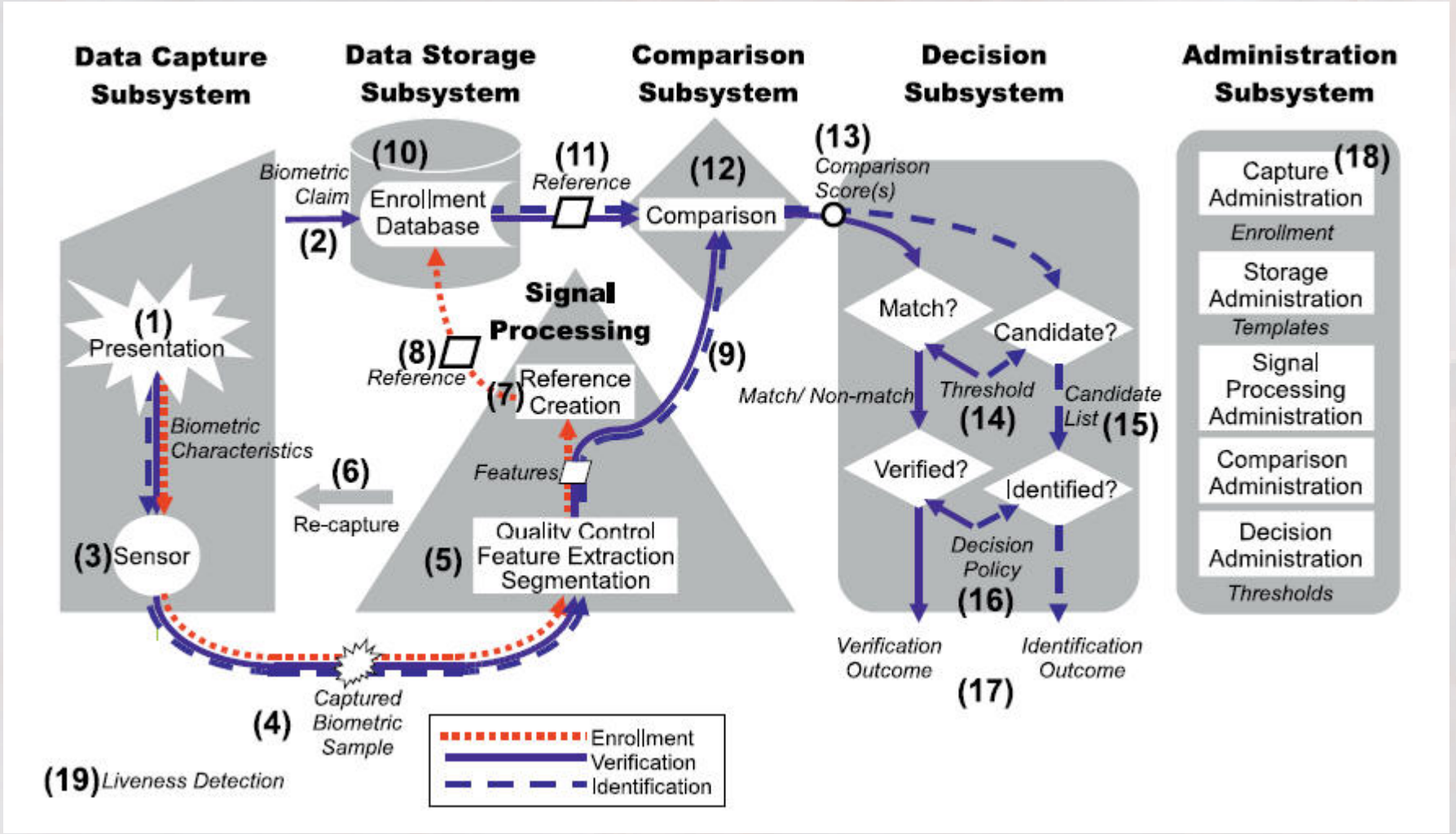
Sustavi za identifikaciju temeljeni na pragu (threshold-based)

- Kod identifikacije na **otvorenom skupu** ne znamo da li je osoba koju želimo identificirati uopće zapisana u bazu.
- Kod identifikacije na zatvorenom skupu, imali smo vrlo veliku vjerojatnost da prava osoba bude uključena u listu kandidata, samo smo trebali odabrati dovoljno veliki rang.
- Kod otvorenog skupa to ne vrijedi (jer osoba možda nije zapisana u bazu podataka), pa nam je pogodnija identifikacija na temelju praga.
- Kod evaluacije sustava za identifikacije temeljenu na pragu koristi se **krivulja alarmiranja (eng.Alarm Curve)**. Ta je krivulja posebna vrsta ROC krivulje.
- Krivulja alarmiranja **prikazuje ovisnost stope detekcije od stope lažnog alarma** (ali za određenu veličinu baze).

Različite krivulje alarmiranja (iz istog biometrijskog sustava), za različite veličine baze podataka



Ranjivost biometrijskog sustava (nezavisna je od biometrijske preciznosti)



Točke ranjivosti biometrijskog sustava

1. **Prezentacija** (Presentation)
2. **Zahtjev za identifikacijski dokument** (Identity Claim)
3. **Senzor** (Sensor)
4. **Prijenos uzorka** (Transmission – Sample)
5. **Kontrola kvalitete i ekstrakcija biometrijskih značajki** (Quality control and feature extraction)
6. **Ponovno preuzimanje uzorka** (Re-capture)
7. **Kreiranje reference** (Reference creation)
8. **Prijenos od reference do predloška** (Transmission - Reference to enrollment)
9. **Prijenos biometrijskih značajki u bazu** (Transmission - Features to database)
10. **Baza predložaka** (Enrollment database)

Točke ranjivosti biometrijskog sustava - nastavak

11. **Prijenos reference iz baze** (Transmission - Reference from database)
12. **Proces usporedbe** (Comparison process)
13. **Prijenos rezultata** (Transmission – Score)
14. **Proces definiranja praga** (Threshold process)
15. **Lista kandidata** (Candidate list)
16. **Politika odlučivanja** (Decision policy)
17. **Prijenos ishoda** (Transmission – Outcome)
18. **Administracija** (Administration)
19. **Detekcija životnosti** (Liveness detection)

Zaključak

- **Biometrijska verifikacija i identifikacija** su najvažniji biometrijski procesi. To nisu deterministički procesi, nego stohastički, proučavaju se pomoću matematičke statistike.
- Kod verifikacije se javljaju **dvije vrste grešaka**: pogrešno odbacivanje prave osobe, čija je stopa FNMR (engl.False Non-Match Rate), te pogrešno prihvaćanje lažne osobe, čija je stopa FMR (eng.False Match Rate). Jedno od glavnih pitanja je kako naći mjeru između dvije greške - manji FNMR obično znači veći FMR, i obrnuto.
- **Identifikacija je puno složenija od verifikacije**, jer je kod identifikacije je stopa pogrešnog prihvaćanja FMR_N ovisna o veličini baze - veća baza, veća greška.
- Sve veće korištenje biometrije u praksi utječe na to da se i **ranjivosti biometrijskog sustava** posvećuje veća pažnja.