Setup: a dataset w/ 2 populations

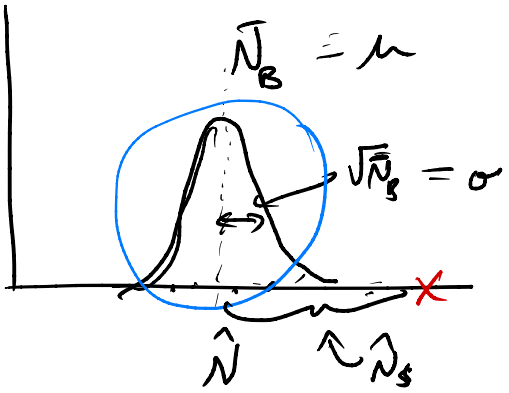
- Example 1: Web store front
  - customers ≡ Background
  - Bots ≡ Signal
- Example 2: SUSY
  - SUSY ≡ signal
  - Bkg to SUSY ≡ Bkg

⇒ Estimated the expected Background
- Ex 1: previous data
- Ex 2: Physics calc + simulation

} ⇒ $\overline{N}_B$

Expect $\bar{N}_B$

Observe $\hat{N}$



Counting $\Rightarrow$ Poisson Stat

Standard Deviation of N is the most likely value

$\Rightarrow$ standard deviation is $\sqrt{N} \equiv \sigma$

$\bar{N}_B = 100 \Rightarrow \sqrt{\bar{N}_B} = 10$

90, 100, 110 } w/in 1 STD

150, 180 $\Rightarrow$ very unlikely

$\hat{N}_S = \hat{N} - \bar{N}_B$ ← very good estimate $\Rightarrow$ almost no error

$\sigma_{\hat{N}_S} = \text{STD}\ \hat{N}_S = \sqrt{\hat{N}}$

$\approx \sqrt{\hat{N}_S + \bar{N}_B}$

$\hat{N}_S = 50 \quad \sigma_{\hat{N}_S} = \sqrt{150}$

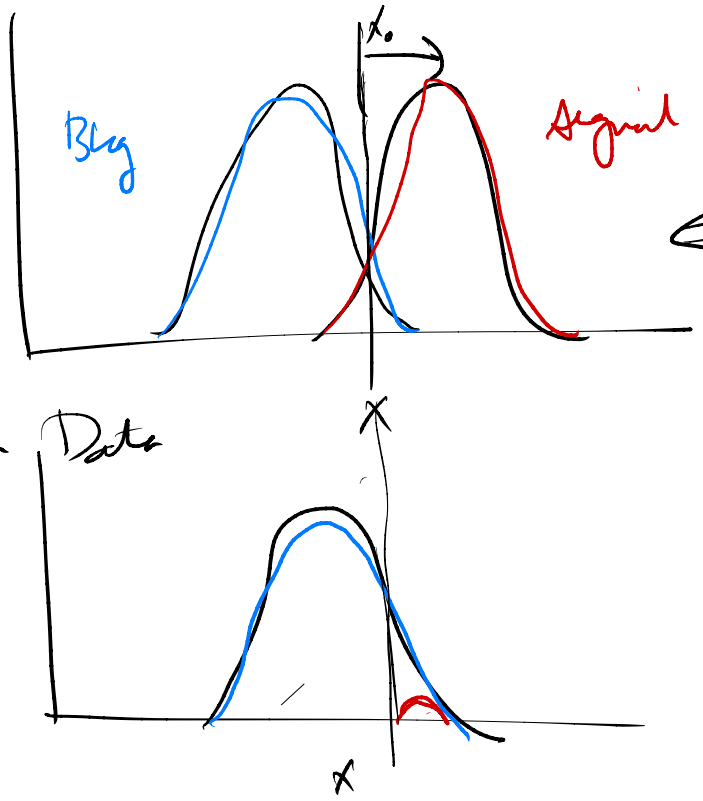
$$\text{significance} = \frac{\hat{N}_S}{\sigma_{\hat{N}_S}} \approx \frac{\hat{N}_S}{\sqrt{\bar{N}_S + \bar{N}_B}}$$

≥ 4 observation

≥ 5 discovery

Imagine that there is a observable (feature, raw feature)

x



⇐ normalized to unit area (prob dist)

True positive rate (TPR)

ϵ_S = fraction of signal events passing selection of $x > x_0$

False Positive Rate (FPR) in Data

ϵ_B = " " Bkg

$\epsilon_S \gg \epsilon_B$

N_B ← after selection

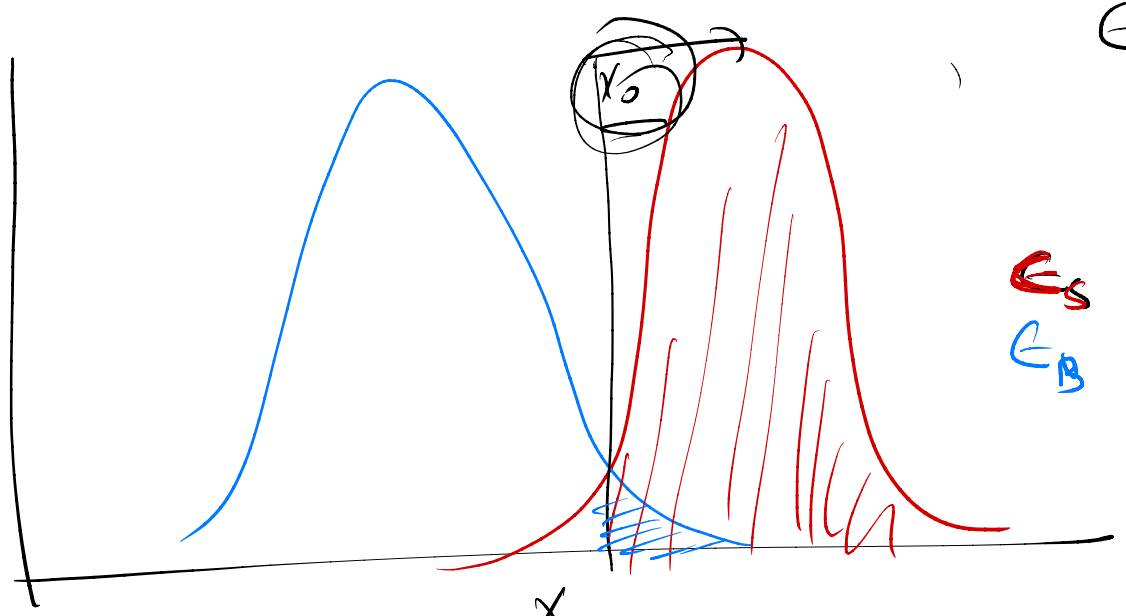
$$N_B = \epsilon_B N_B \Rightarrow N' = N_S' + N_B' \Leftarrow \text{expected}$$

$$N_S' = \epsilon_S N_S \Rightarrow \hat{N}' \Leftarrow \text{observed}$$

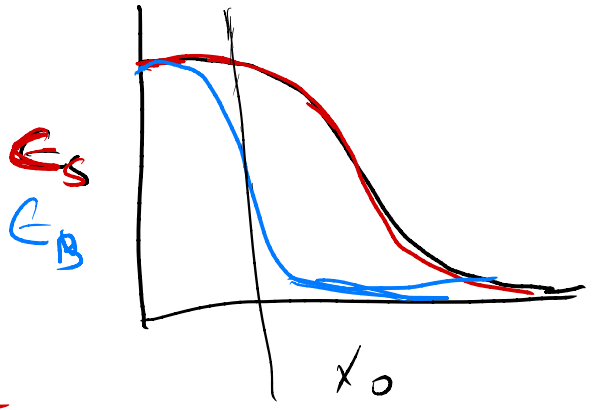
$$\hat{N}_S' = \hat{N}' - N_B'$$

$$\sigma_{\hat{N}_S'} = \sqrt{\hat{N}'} \simeq \sqrt{N'} = \sqrt{\epsilon_B N_B + \epsilon_S N_S}$$

$$\text{Significance} = \frac{\hat{N}_S}{\sqrt{\epsilon_S \bar{N}_S + \epsilon_S N_B}} \approx \frac{\epsilon_S \bar{N}_S}{\sqrt{\epsilon_S N_S + \epsilon_B N_B}}$$

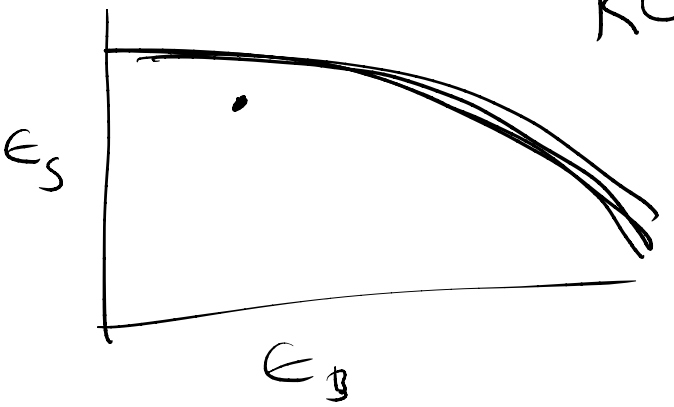


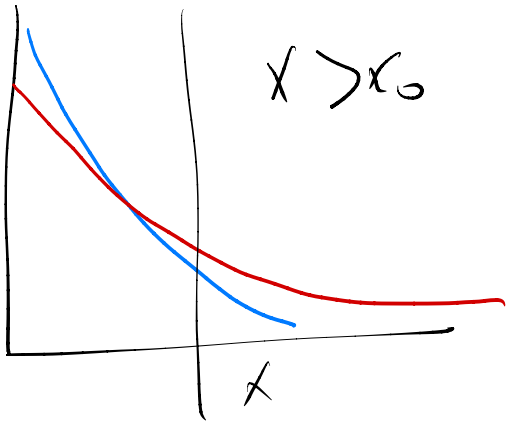
$\epsilon_S \gg \epsilon_B$

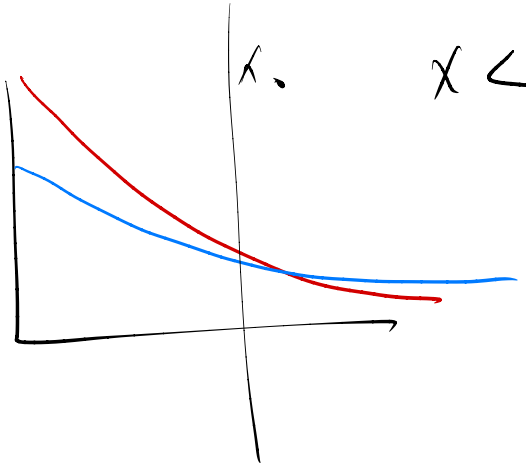


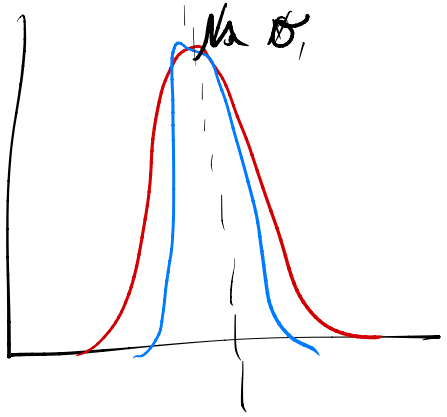
ROC

Optimization $\Rightarrow$ select x_0 such that s_{ig} is maximum.

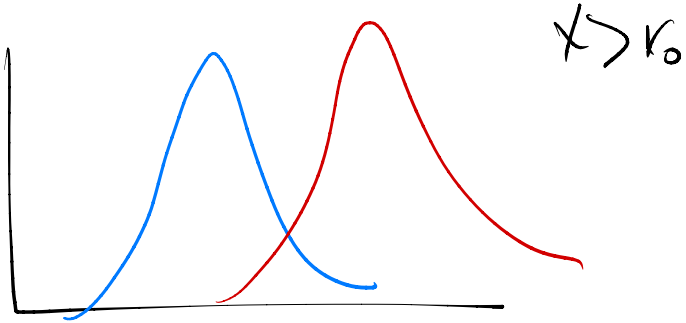


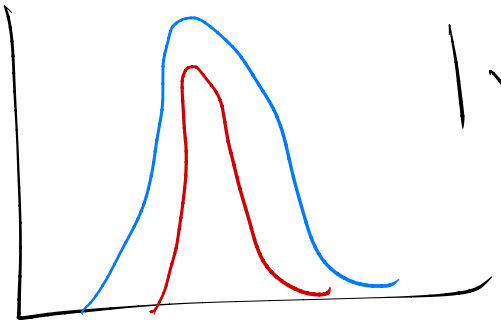






$|x - \mu| > \sigma$





$|x - \mu| < \sigma$