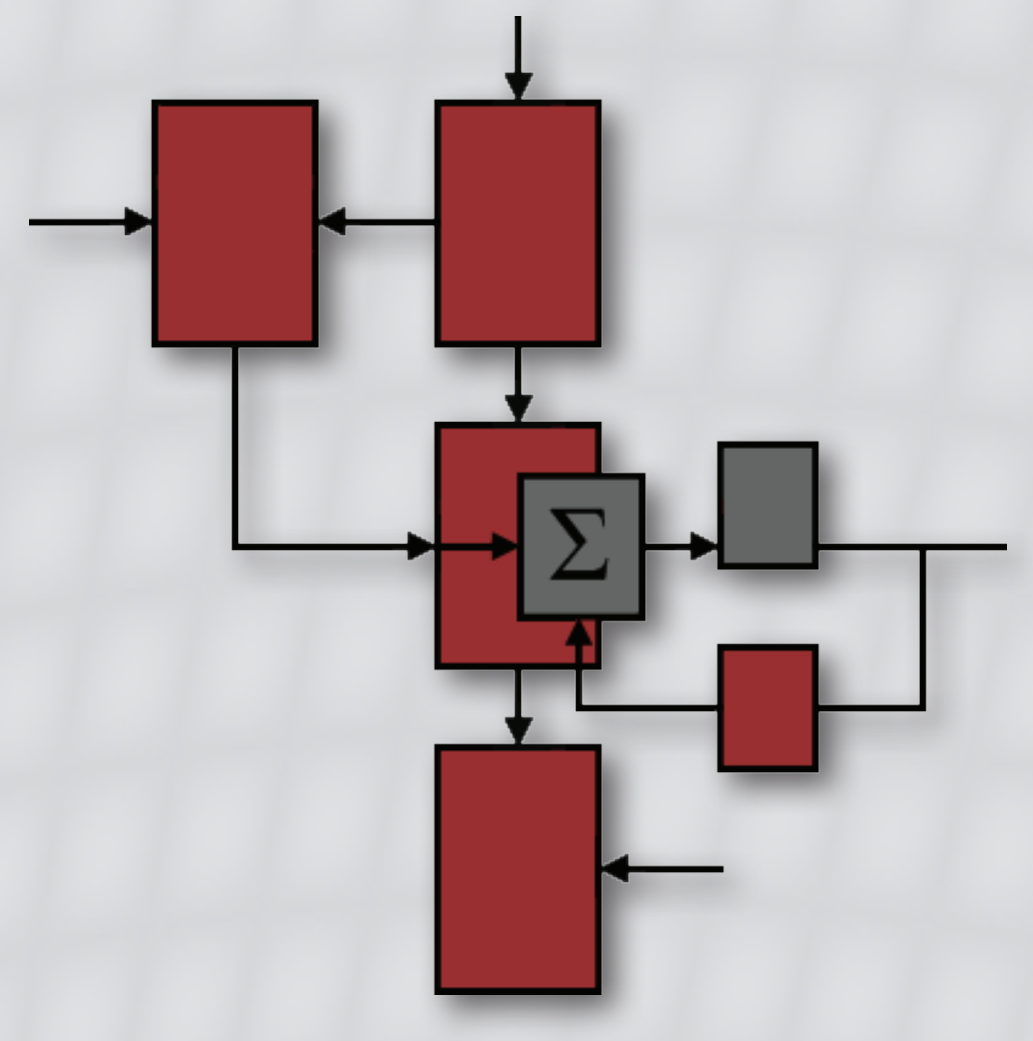


Minimum Mean Bayes Risk Error Quantization of Prior Probabilities

Kush R. Varshney and Lav R. Varshney

Laboratory for Information and Decision Systems, Massachusetts Institute of Technology

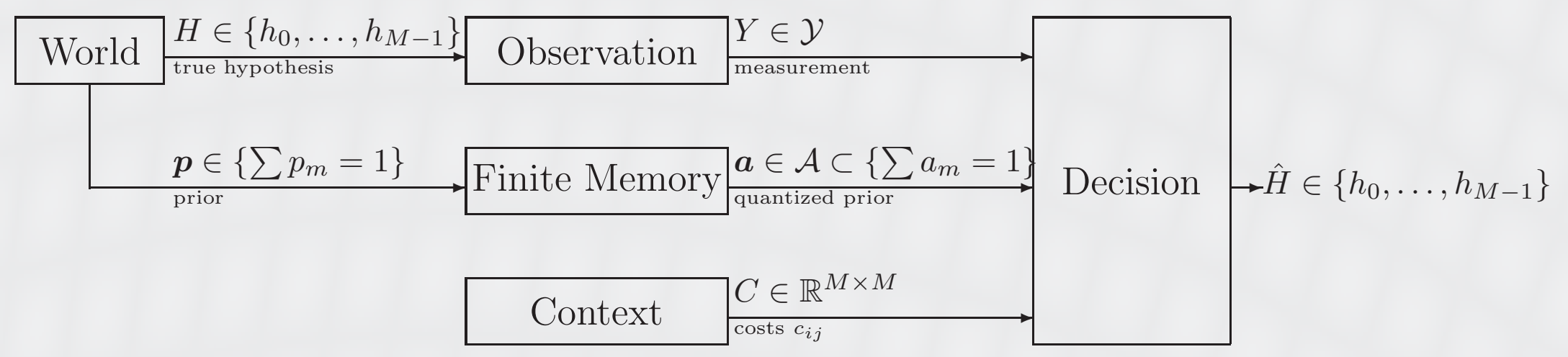
Problem Overview

Hypothesis Testing Scenario

- Object in one of M states: $\{h_0,\dots,h_{M-1}\}$ with $\Pr[H=h_m]=p_m$
- Prior probability vector $\boldsymbol{p}=\begin{bmatrix}p_0 & \cdots & p_{M-1}\end{bmatrix}^T$
- Population of objects — each object's prior probability drawn from $f_{\boldsymbol{P}}(\boldsymbol{p})$
- Noisy measurement Y
- Task: for a given object with prior probability $\boldsymbol{p}$, estimate h from y under Bayes risk objective with costs c_{ij}
- Constraint: $\boldsymbol{P}$ is quantized
- Goal: design best K-point quantizer to minimize average Bayes risk error over the population $f_{\boldsymbol{P}}(\boldsymbol{p})$

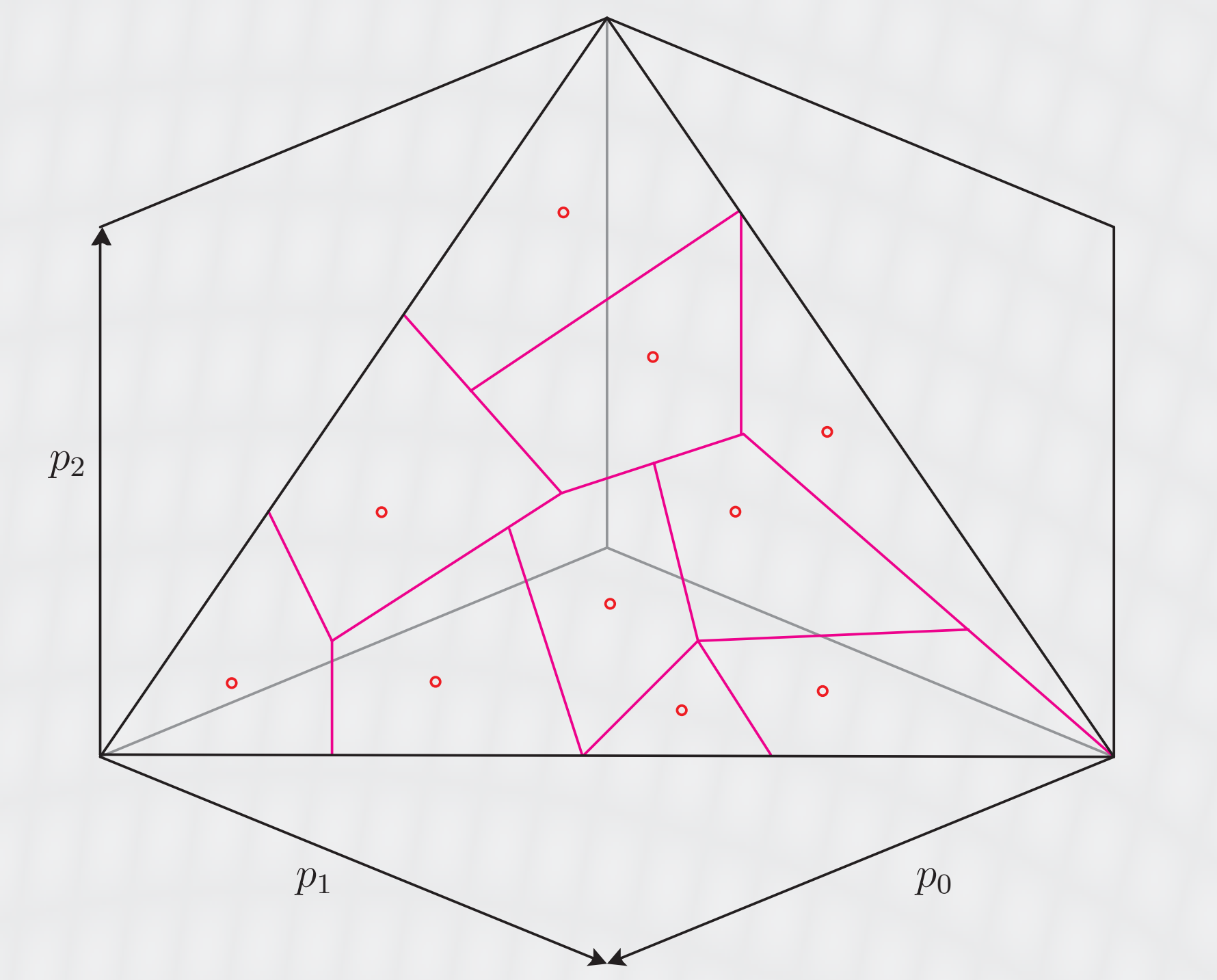
• •

Block Diagram



• •

A Quantizer over the Population $f_{\boldsymbol{P}}(\boldsymbol{p})$, $M=3$



Background

Binary Bayesian Hypothesis Testing

- Hypotheses h_0, h_1, and prior probabilities p_0, p_1, such that $p_0=\Pr[H=h_0]$ and $p_1=\Pr[H=h_1]=1-p_0$
- Noisy measurement Y with likelihoods $f_{Y|H}(y|h_0)$ and $f_{Y|H}(y|h_1)$
- Function $\hat{h}(y)$ maps every possible y to either h_0 or h_1
- $\hat{h}(\cdot)=\arg\min_{f(\cdot)} E\left[c\left(H,f(Y)\right)\right]$ with solution $\frac{f_{Y|H}(y|h_1)}{f_{Y|H}(y|h_0)} \underset{\hat{h}(y)=h_0}{\overset{\hat{h}(y)=h_1}{\gtrless}} \frac{p_0(c_{10}-c_{00})}{p_1(c_{01}-c_{11})}$, where $c_{ij}=c(h_i,h_j)$
- Errors: $p_E^{\mathrm{I}}=\Pr[\hat{h}(Y)=h_1|H=h_0]$ and $p_E^{\mathrm{II}}=\Pr[\hat{h}(Y)=h_0|H=h_1]$
- Bayes risk: $J=(c_{10}-c_{00})p_0p_E^{\mathrm{I}}+(c_{01}-c_{11})p_1p_E^{\mathrm{II}}+c_{00}p_0+c_{11}p_1$

• •

Scalar Quantization

- K-point quantizer $v_K(\cdot)$ for $f_{P_0}(p_0)$, $p_0\in[0,1]$ partitions the domain into intervals $\mathcal{R}_1=[0,b_1]$, $\mathcal{R}_2=(b_1,b_2]$, $\mathcal{R}_3=(b_2,b_3]$, $\dots$, $\mathcal{R}_K=(b_{K-1},1]$
- $v_K(p_0)=a_k$ for $p_0\in\mathcal{R}_k$
- Design $v_K(\cdot)$ to minimize $D=E[d(P_0,v_K(P_0))]=\int d(p_0,v_K(p_0))f_{P_0}(p_0)dp_0$
- Conditions for quantizer optimality: nearest neighbor condition, centroid condition, zero probability of boundary condition [Gersho & Gray, 1992]

Bayes Risk Error Distortion

Bayes Risk Function Properties

- Assuming $M=2$ and $c_{00}=c_{11}=0$, Bayes risk as a function of p_0 is $J(p_0)=c_{10}p_0p_E^{\mathrm{I}}(p_0)+c_{01}(1-p_0)p_E^{\mathrm{II}}(p_0)$
- $J(p_0)$ is zero at $p_0=0$ and $p_0=1$
- $J(p_0)$ is positive, concave, and continuous in the interval $(0,1)$

• •

Mismatched Bayes Risk

- If the true prior probability is p_0, but $\hat{h}(y)$ is designed using some other value a, there is mismatch
- Mismatched Bayes risk function $\tilde{J}(p_0,a)=c_{10}p_0p_E^{\mathrm{I}}(a)+c_{01}(1-p_0)p_E^{\mathrm{II}}(a)$
- $\tilde{J}(p_0,a)$ is a linear function of p_0 with slope $(c_{10}p_E^{\mathrm{I}}(a)-c_{01}p_E^{\mathrm{II}}(a))$ and intercept $c_{01}p_E^{\mathrm{II}}(a)$
- $\tilde{J}(p_0,a)$ is tangent to $J(p_0)$ at $p_0=a$
- $\tilde{J}(p_0,a)\geq J(p_0)\geq 0$

• •

Bayes Risk Error

- Define distortion function for quantization
- Bayes risk error: $d(p_0,a)=\tilde{J}(p_0,a)-J(p_0)$
- Nonnegative and only equal to zero when $p_0=a$
- Convex and continuous in p_0 for all a

• •

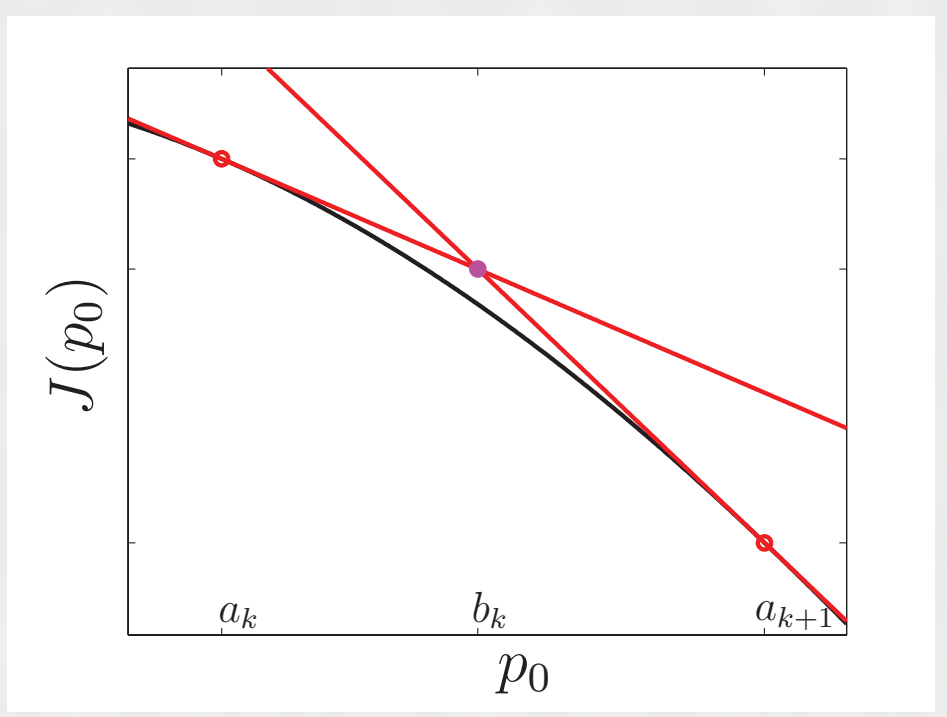
Comparison to Previous Work

- Previous work combining detection and quantization quantizes y, not p_0 [Kassam, 1977], [Poor & Thomas, 1977], [Gupta & Hero, 2003]
- Also only approximates Bayes risk function instead of using it directly

Quantizer Optimality Conditions

Nearest Neighbor Condition

- For fixed representation points $\{a_k\}$, derive optimal boundary points $\{b_k\}$
- Given $p_0\in[a_k,a_{k+1}]$, if $\tilde{J}(p_0,a_k)<\tilde{J}(p_0,a_{k+1})$, then the Bayes risk error is minimized when p_0 represented by a_k and vice versa
- $b_k\in[a_k,a_{k+1}]$ is abscissa of intersection between $\tilde{J}(p_0,a_k)$ and $\tilde{J}(p_0,a_{k+1})$



- $b_k=\frac{c_{01}\left(p_E^{\mathrm{II}}(a_{k+1})-p_E^{\mathrm{II}}(a_k)\right)}{c_{01}\left(p_E^{\mathrm{II}}(a_{k+1})-p_E^{\mathrm{II}}(a_k)\right)-c_{10}\left(p_E^{\mathrm{I}}(a_{k+1})-p_E^{\mathrm{I}}(a_k)\right)}$

• •

Centroid Condition

- For fixed boundary points $\{b_k\}$, derive optimal representation points $\{a_k\}$
- Minimize $D=\sum_{k=1}^{K}\int_{\mathcal{R}_k}\left(\tilde{J}(p_0,a_k)-J(p_0)\right)f_{P_0}(p_0)dp_0$ for each interval separately
- Let $I_k^{\mathrm{I}}=\int_{\mathcal{R}_k}p_0f_{P_0}(p_0)dp_0$ and $I_k^{\mathrm{II}}=\int_{\mathcal{R}_k}(1-p_0)f_{P_0}(p_0)dp_0$
- $a_k=\arg\min_a\left\{c_{10}I_k^{\mathrm{I}}p_E^{\mathrm{I}}(a)+c_{01}I_k^{\mathrm{II}}p_E^{\mathrm{II}}(a)\right\}$
- Uniquely minimize by setting derivative equal to zero since $d(p_0,a)$ has exactly one stationary point
- a_k is solution to $c_{10}I_k^{\mathrm{I}}\frac{dp_E^{\mathrm{I}}(a_k)}{da_k}+c_{01}I_k^{\mathrm{II}}\frac{dp_E^{\mathrm{II}}(a_k)}{da_k}=0$

Gaussian Examples

Signal and Noise Model

- Model: $Y=s_m+W$, where $s_0=0$, $s_1=\mu$, and W is a zero-mean Gaussian random variable with variance σ^2
- The likelihood functions are $f_{Y|H}(y|h_0)=\mathcal{N}(y;0,\sigma^2)=\frac{1}{\sigma\sqrt{2\pi}}e^{-y^2/2\sigma^2}$ and $f_{Y|H}(y|h_1)=\mathcal{N}(y;\mu,\sigma^2)=\frac{1}{\sigma\sqrt{2\pi}}e^{-(y-\mu)^2/2\sigma^2}$

• •

Optimality Conditions

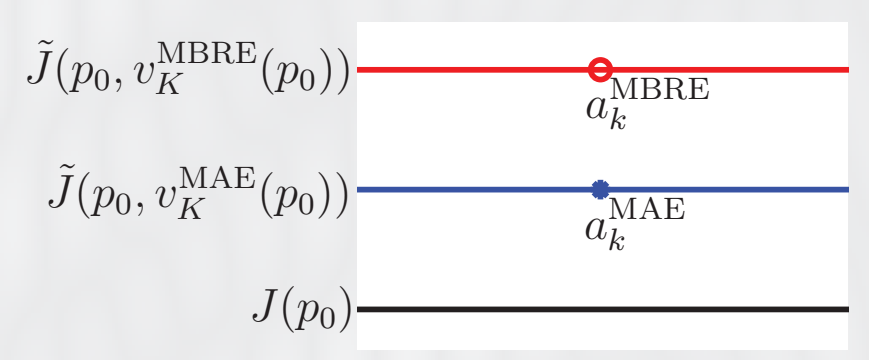
- The two error probabilities are $p_E^{\mathrm{I}}(p_0)=Q\left(\frac{\mu}{2\sigma}+\frac{\sigma}{\mu}\ln\left(\frac{c_{10}p_0}{c_{01}(1-p_0)}\right)\right)$ and $p_E^{\mathrm{II}}(p_0)=Q\left(\frac{\mu}{2\sigma}-\frac{\sigma}{\mu}\ln\left(\frac{c_{10}p_0}{c_{01}(1-p_0)}\right)\right)$, where $Q(\alpha)=\frac{1}{\sqrt{2\pi}}\int_\alpha^\infty e^{-x^2/2}dx$
- Substitute error probabilities into nearest neighbor condition expression
- Derivatives are $\left.\frac{dp_E^{\mathrm{I}}(p_0)}{dp_0}\right|_{p_0=a_k}=-\frac{1}{\sqrt{2\pi}}\frac{\sigma}{\mu}\frac{1}{a_k(1-a_k)}e^{-\frac{1}{2}\left(\frac{\mu}{2\sigma}+\frac{\sigma}{\mu}\ln\left(\frac{c_{10}a_k}{c_{01}(1-a_k)}\right)\right)^2}$ and $\left.\frac{dp_E^{\mathrm{II}}(p_0)}{dp_0}\right|_{p_0=a_k}=+\frac{1}{\sqrt{2\pi}}\frac{\sigma}{\mu}\frac{1}{a_k(1-a_k)}e^{-\frac{1}{2}\left(\frac{\mu}{2\sigma}-\frac{\sigma}{\mu}\ln\left(\frac{c_{10}a_k}{c_{01}(1-a_k)}\right)\right)^2}$
- Substitute derivatives into centroid condition expression and solve to obtain $a_k=\frac{I_k^{\mathrm{I}}}{I_k^{\mathrm{I}}+I_k^{\mathrm{II}}}$

• •

Setup of Examples

- Minimum mean Bayes risk error (MBRE) quantizer designed for $\mu=1$ and $\sigma=1$ using Lloyd-Max algorithm
- Compared to minimum mean absolute error (MAE) quantizer ($d(p_0,a)=|p_0-a|$)
- Plots show $f_{P_0}(p_0)$; MBRE as a function of K for both quantizers; and $\tilde{J}(p_0,v_K(p_0))$ for four values of K for both quantizers

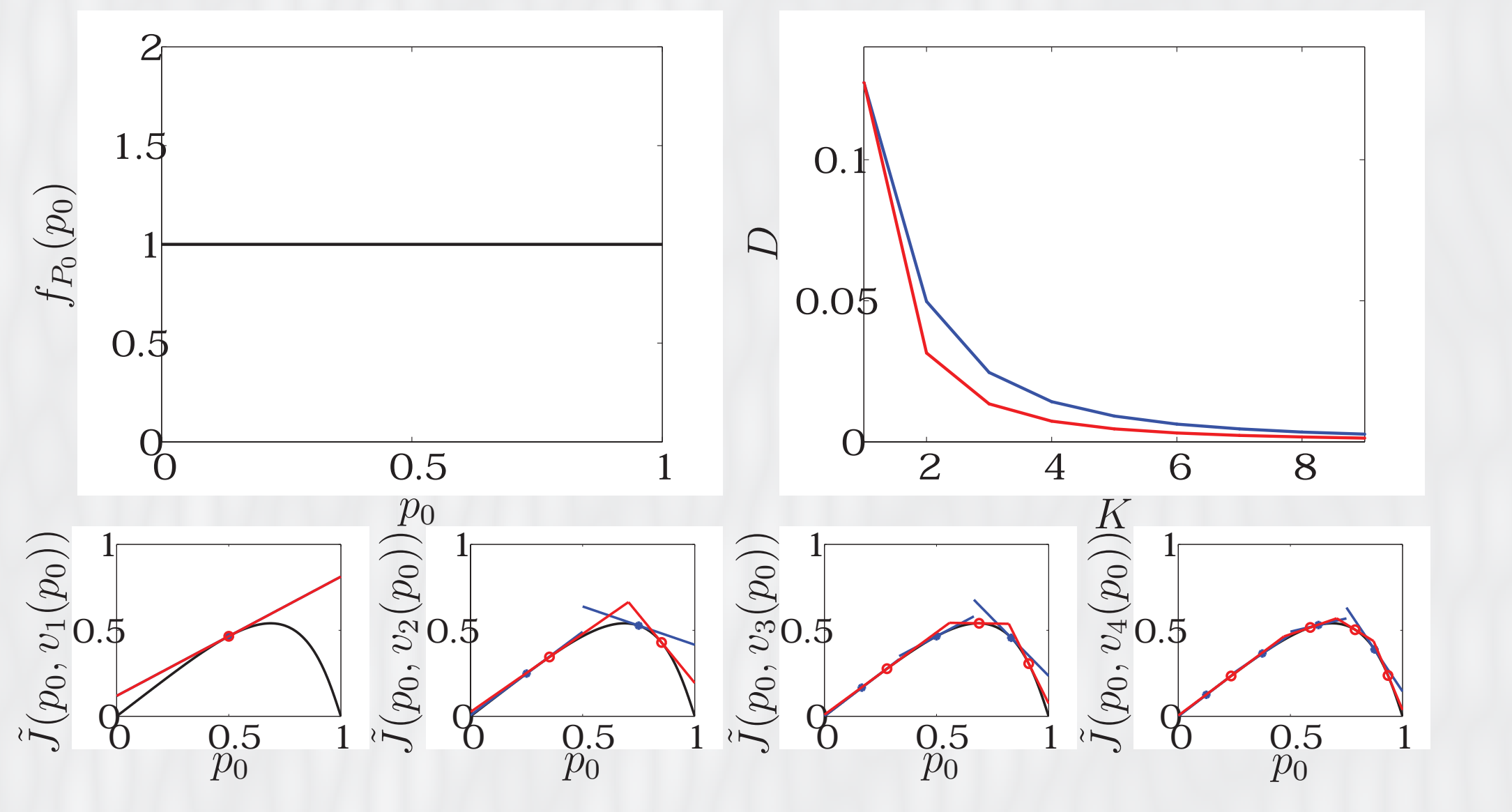
Legend



In mean Bayes risk error plot, red line is for MBRE-optimal quantizer and blue line is for MAE-optimal quantizer.

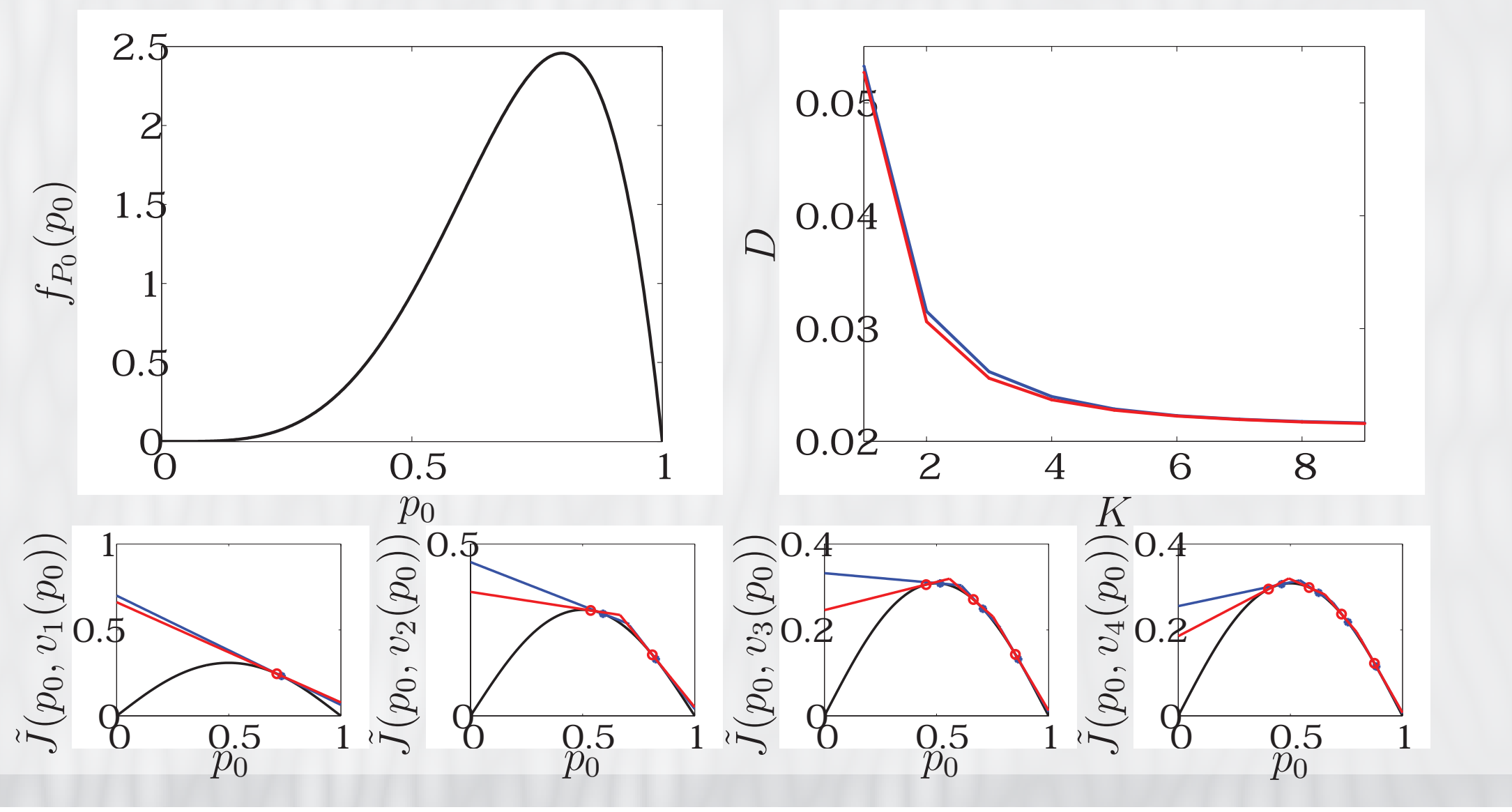
• •

$f_{P_0}(p_0)$ ~ Uniform; $c_{10}=1$, $c_{01}=4$



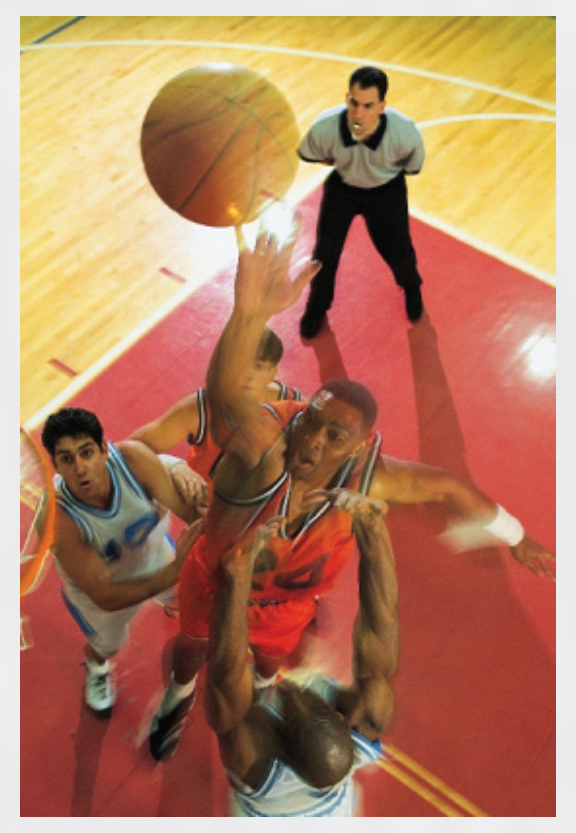
• •

$f_{P_0}(p_0)$ ~ Beta(5,2); $c_{10}=1$, $c_{01}=1$



Human Decision Making

Human Decision Makers

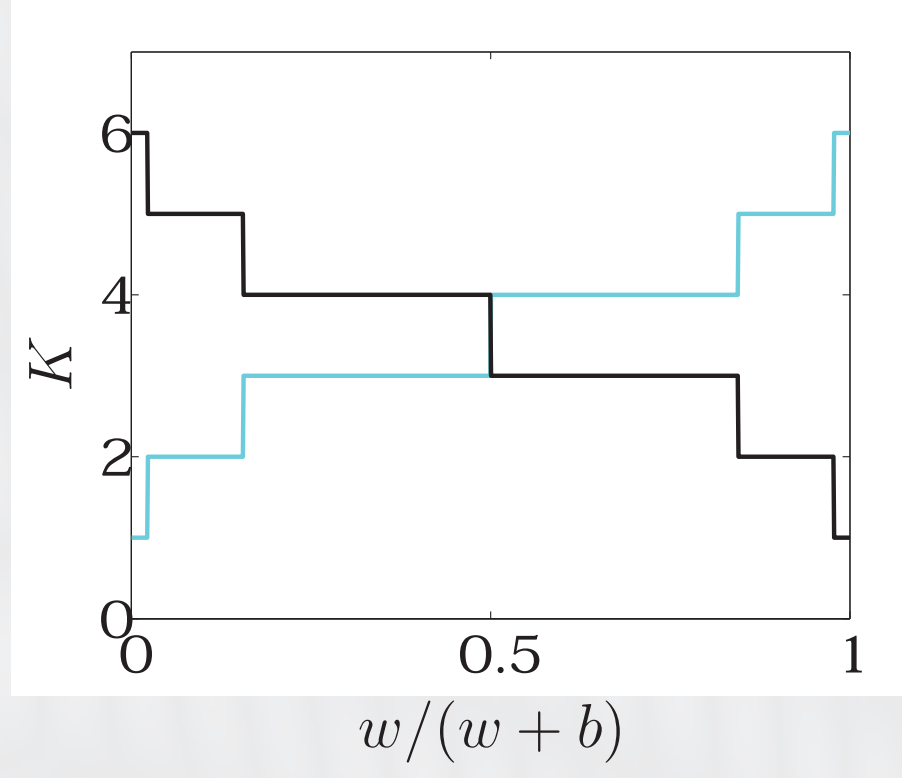


- Limited information processing capacity
- Handle 7 ± 2 categories without getting confused [Miller, 1956]
- Assumption: perform optimal Bayesian hypothesis testing with quantized priors
- Essentially automatic race and gender categorization, particularly when lacking the time, motivation, or cognitive capacity to think deeply [Macrae & Bodenhausen, 2000]
- Assumption: separate quantizers for different races and genders with total point constraint K_t

• •

Two Population Quantization

- Two identical populations B and W with common $f_{P_0}(p_0)$
- Separate quantizers $v_{K_b}(\cdot)$ and $v_{K_w}(\cdot)$ with total constraint $K_t=K_w+K_b$
- Exposure of decision maker to populations not identical: b and w
- Extend MBRE to two population case: $D^{(2)}=\frac{w}{w+b}E[\tilde{J}(P_0,v_{K_w}(P_0))]+\frac{b}{w+b}E[\tilde{J}(P_0,v_{K_b}(P_0))]-E[J(P_0)]$
- Optimization involves point allocation and optimal separate quantizer

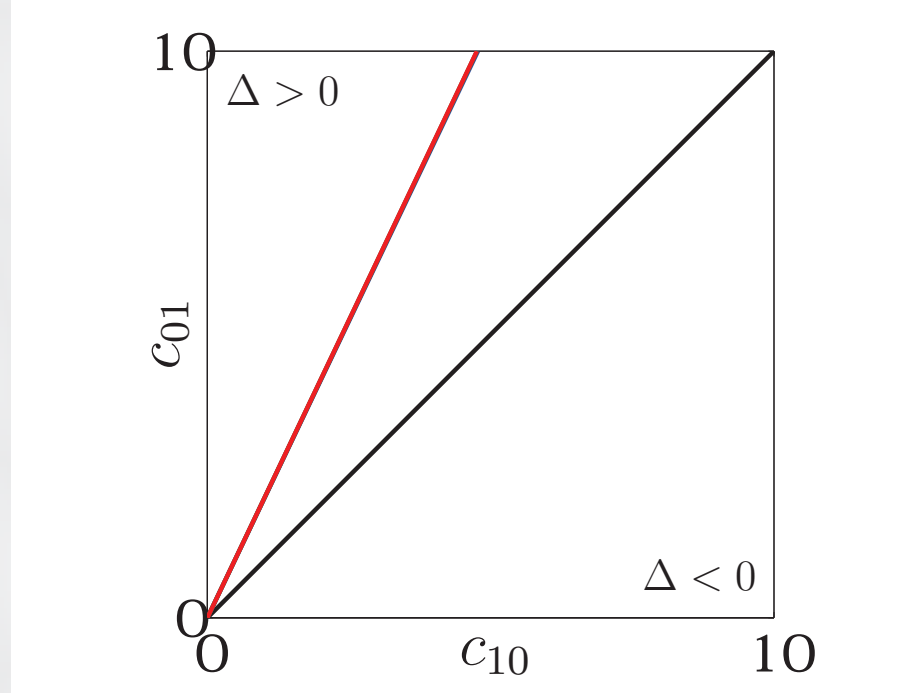


- Optimal point allocation as a function of $w/(w+b)$ for $f_{P_0}(p_0)$ ~ Beta(5,2), $c_{10}=c_{01}=1$, and $K_t=7$
- K_b = black line, K_w = cyan line
- Better Bayes risk performance for population W when $K_b<K_w$ and vice versa
- Predictions from model consistent with face recognition experiments [Meissner & Brigham, 2001]

• •

Effect of Bayes Costs

- Bayes risk performance includes both Type I and Type II errors
- Often, only detection rate is observable experimentally: $\Pr[\tilde{h}_K(Y)=h_1]=1-p_0+p_0p_E^{\mathrm{I}}(v_K(p_0))-(1-p_0)p_E^{\mathrm{II}}(v_K(p_0))$
- If decision maker is member of W with $w>b$ and $\Delta>0$, where $\Delta=E\left[\Pr[\tilde{h}_{K_b}(Y)=h_1]-\Pr[\tilde{h}_{K_w}(Y)=h_1]\right]$, then out-of-population bias; if $\Delta<0$, then in-population bias



- Δ is a function of c_{10} and c_{01}
- Dividing line between in-population bias and out-of-population bias is plotted for Uniform (black) and Beta(5,2) (colored) distributions
- Precautionary principle, $c_{10}\gg c_{01}$, yields out-of-population bias
- Pushful principle, $c_{01}\gg c_{10}$, yields in-population bias
- Empirical evidence of human decision making shows $\Delta>0$ [Donohue & Levitt, 2001], [Price & Wolfers, 2007] — consistent with precautionary principle

Conclusion

Memory Constraints Limit Decision Making

- Formulated hypothesis testing under memory constraints rather than communication constraints
- Derived optimal quantizer of prior probability distribution under Bayes risk objective and determined performance
- Quantized prior hypothesis testing combined with theories of social cognition and empirical facts about segregation leads to generative model of discrimination
- Decision making biased despite identical population distributions and no malicious intent by decision maker

Presented at IEEE International Conference on Acoustics, Speech, and Signal Processing, March 30–April 4, 2008