Maximin Decision Rule

Robert M. Haralick

Computer Science, Graduate Center
City University of New York

Proposition

Images of linear functions of convex sets are convex.

Proof.

Let C be a convex set and $f : C \to \mathbb{R}^N$ be a linear function. Define $D = \{y \in \mathbb{R}^N \mid y = f(x), x \in C\}$ Let $y_1, y_2 \in D$ and let $0 \leq \lambda \leq 1$. Then there exists $x_1, x_2 \in C$ such that $y_1 = f(x_1)$ and $y_2 = f(x_2)$.

$$\begin{aligned} \lambda y_1 + (1 - \lambda y_2) &= \lambda f(x_1) + (1 - \lambda) f(x_2) \\ &= f(\lambda x_1 + (1 - \lambda) x_2) \end{aligned}$$

But $x_1, x_2 \in C$ and C is convex. Therefore $\lambda x_1 + (1 - \lambda) x_2 \in C$. Hence, $f(\lambda x_1 + (1 - \lambda) x_2) \in D$. And this makes $\lambda y_1 + (1 - \lambda y_2) \in D$ □

Dependence on Prior Class Probabilities

Proposition

Expected economic gain for a decision rule is an affine function of the expected economic conditional gains with coefficients $P(c^1), \ldots, P(c^{K-1})$.

Proof.

$$
\begin{aligned}
E[e; f] &= \sum_{j=1}^{K} E[e \mid c^j; f] P(c^j) \\
&= \sum_{j=1}^{K-1} E[e \mid c^j; f] P(c^j) + E[e \mid c^K; f](1 - \sum_{j=1}^{K-1} P(c^j)) \\
&= \sum_{j=1}^{K-1} \{E[e \mid c^j; f] - E[e \mid c^K; f]\} P(c^j) + E[e \mid c^K; f]
\end{aligned}
$$

$\square$

Example

e	Assigned	
True	c^1	c^2
c^1	2	-1
c^2	-1	2

$P(d \mid c)$	Measurement		
True Class	d^1	d^2	d^3
c^1	.2	.3	.5
c^2	.5	.4	.1

$$E[e \mid c^j; f] = \sum_{d \in D} \sum_{k=1}^{K} e(c^j, c^k) P(d \mid c^j) f_d(c^k)$$

	Measurements			Conditional Gain	
f	d^1	d^2	d^3	$E[e\|c^1; f]$	$E[e\|c^2; f]$
f^1	c^1	c^1	c^1	2.0	-1.0
f^2	c^1	c^1	c^2	.5	-.7
f^3	c^1	c^2	c^1	1.1	.2
f^4	c^1	c^2	c^2	-.4	.5
f^5	c^2	c^1	c^1	1.4	.5
f^6	c^2	c^1	c^2	-.1	.8
f^7	c^2	c^2	c^1	.5	1.7
f^8	c^2	c^2	c^2	-1.	2.0

Expected Conditional Gain and Expected Gain

$$E[e \mid c^j; f] = \sum_{d \in D} \sum_{k=1}^{K} e(c^j, c^k) P(d \mid c^j) f_d(c^k)$$

$$\begin{aligned}
E[e; f] &= \sum_{j=1}^{K} E[e \mid c^j; f] P(c^j) \\
&= \left[\sum_{j=1}^{K-1} E[e \mid c^j; f] P(c^j) \right] + \left[E[e \mid c^K; f](1 - \sum_{j=1}^{K-1} P(c^j)) \right] \\
&= \left[\sum_{j=1}^{K-1} \{ E[e \mid c^j; f] - E[e \mid c^K; f] \} P(c^j) \right] + E[e \mid c^K; f]
\end{aligned}$$

$$\begin{array}{rcccc}
E[e; f^1] & = & [2 - (-1)]P(c^1) + (-1) & = & 3.0P(c^1) - 1 \\
E[e; f^2] & = & [.5 - (-.7)]P(c^1) + (-.7) & = & 1.2P(c^1) - .7 \\
E[e; f^3] & = & [1.1 - .2]P(c^1) + .2 & = & 0.9P(c^1) + .2 \\
E[e; f^4] & = & [-.4 - .5]P(c^1) + .5 & = & -0.9P(c^1) + .5 \\
E[e; f^5] & = & [1.4 - .5]P(c^1) + .5 & = & 0.9P(c^1) + .5 \\
E[e; f^6] & = & [-.1 - .8]P(c^1) - .8 & = & -0.9P(c^1) + .8 \\
E[e; f^7] & = & [.5 - 1.7]P(c^1]1.7 & = & -1.2P(c^1) + 1.7 \\
E[e; f^8] & = & [-1.0 - 2.0]P(c^1) + 2.0 & = & -3.0P(c^1) + 2.0
\end{array}$$

Expected Conditional Gain and Expected Gain

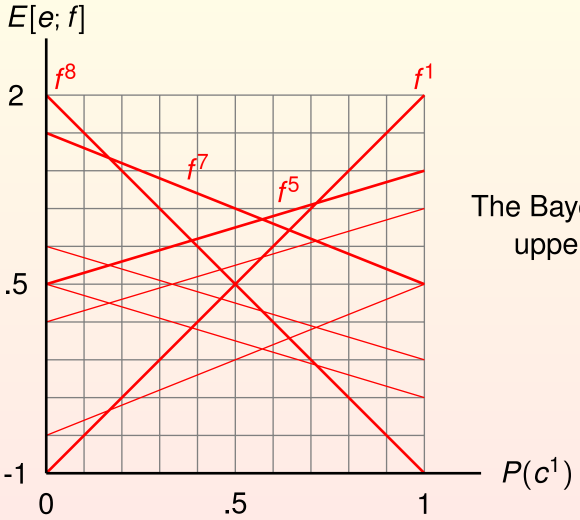
	Measurements			Expected Conditional Gain		Expected Gain
f	d^1	d^2	d^3	$E[e\|c^1; f]$	$E[e\|c^2; f]$	$E[e, f]$
f^1	c^1	c^1	c^1	2.0	-1.0	$3P(c^1) - .7$
f^2	c^1	c^1	c^2	.5	$-.7$	$1.2P(c^1) - .7$
f^3	c^1	c^2	c^1	1.1	.2	$.9P(c_1) + .2$
f^4	c^1	c^2	c^2	$-.4$	.5	$-.9P(c_1) + .5$
f^5	c^2	c^1	c^1	1.4	.5	$.9P(c^1) + .5$
f^6	c^2	c^1	c^2	$-.1$	.8	$-.9P(c^1) + .8$
f^7	c^2	c^2	c^1	.5	1.7	$-1.2P(c^1) + 1.7$
f^8	c^2	c^2	c^2	$-1.$	2.0	$-3P(c^1) + 2.0$

Dependence on Prior Class Probabilities

$$E[e; f] = \sum_{j=1}^{K-1} \{E[e \mid c^j; f] - E[e \mid c^K; f]\} P(c^j) + E[e \mid c^K; f]$$

	Measurements			Conditional Gain		Expected Gain
f	d^1	d^2	d^3	$E[e\|c^1; f]$	$E[e\|c^2; f]$	$E[e; f, P(c^1)]$
f^1	c^1	c^1	c^1	2.0	-1.0	$3P(c^1) - 1$
f^2	c^1	c^1	c^2	.5	-.7	$1.2P(c^1) - .7$
f^3	c^1	c^2	c^1	1.1	.2	$.9P(c^1) + .2$
f^4	c^1	c^2	c^2	-.4	.5	$-.9P(c^1) + .5$
f^5	c^2	c^1	c^1	1.4	.5	$.9P(c^1) + .5$
f^6	c^2	c^1	c^2	-.1	.8	$-.9P(c^1) + .8$
f^7	c^2	c^2	c^1	.5	1.7	$-1.2P(c^1) + 1.7$
f^8	c^2	c^2	c^2	-1.	2.0	$-3P(c^1) + 2.0$

Dependence on Class Prior Probabilities



The Bayes Gain is the upper envelope

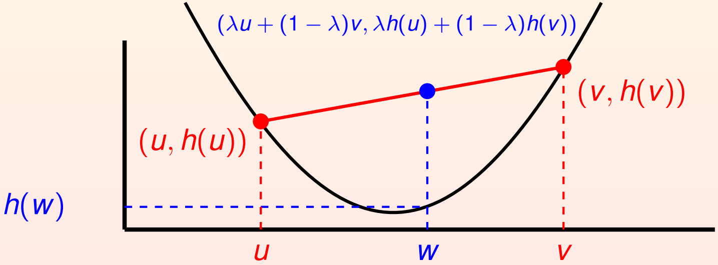
Convex Functions

Definition

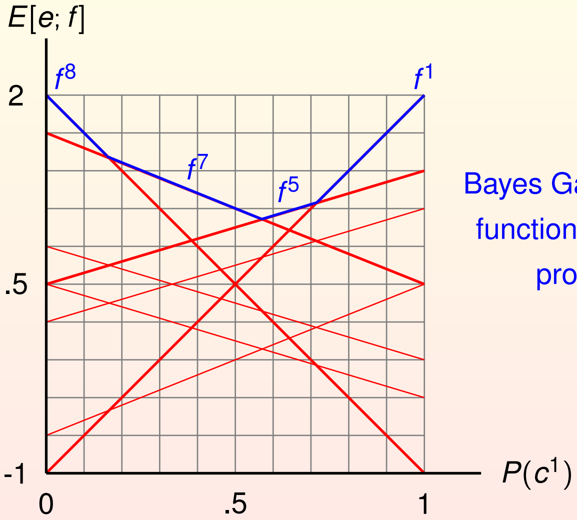
A function h, $h : \mathbb{R}^N \to \mathbb{R}$, is a convex function if and only if for every λ, $0 \le \lambda \le 1$,

$$h(\lambda(x_1, \ldots, x_N) + (1-\lambda)(y_1, \ldots, y_N)) \le \lambda h(x_1, \ldots, x_N) + (1-\lambda) h(y_1, \ldots, y_N)$$

$$w = \lambda u + (1-\lambda) v$$



Bayes Gain is Convex



Bayes Gain is a convex function of class prior probabilities

Bayes Gain Is Convex

$$
\begin{aligned}
E[e; f] &= \sum_{j=1}^{K} E[e \mid c^j; f]P(c^j)] \\
G_B &= \max_f E[e; f] \text{ Bayes Gain}
\end{aligned}
$$

Let $f^n, \ n = 1, \ldots N$ be the $N = |C|^{|D|}$ deterministic decision rules.

Define for $j = 1, \ldots, K$

$$
\begin{aligned}
a_{jn} &= E[e \mid c^j; f^n] \\
p_j &= P(c^j) \\
G_B(P(c^1), \ldots, P(c^K)) &= \max_n \sum_{j=1}^{K} E[e \mid c^j; f^n]P(c^j) \\
G_B(p_1, \ldots, p_K) &= \max_n \sum_{j=1}^{K} a_{jn} p_j
\end{aligned}
$$

Bayes Gain Is Convex

Theorem

Let $p = (p_1, \ldots, p_K)$ *and* $q = (q_1, \ldots q_K)$*. Let* $0 \le \lambda \le 1$*.*

$$G_B(\lambda p + (1-\lambda)q) \quad \le \quad \lambda G_B(p) + (1-\lambda)G_B(q)$$

Proof.

$$\begin{aligned}
G_B(\lambda p + (1-\lambda)q) &= \max_n \sum_{j=1}^{K} a_{jn}(\lambda p_j + (1-\lambda)q_j) \\
&= \max_n \left\{ \lambda \sum_{j=1}^{K} a_{jn}p_j + (1-\lambda) \sum_{j==1}^{K} a_{jn}q_j \right\} \\
&\le \left[\max_n \lambda \sum_{j=1}^{K} a_{jn}p_j \right] + \left[\max_n (1-\lambda) \sum_{j=1}^{K} a_{jn}q_j \right] \\
&\le \lambda G_B(p) + (1-\lambda)G_B(q)
\end{aligned}$$

□

Definition

Let $f : \mathbb{R}^N \to \mathbb{R}$. The epigraph of f, denoted Epi(f) is the set of points lying on or above the graph of f.

$$Epi(f) = \{(x, u) \in \mathbb{R}^N \times \mathbb{R} \mid u \geq f(x)\}$$

Epigraph and Convex Sets

Proposition

If a function is convex then its epigraph is a convex set.

Proof.

Suppose f *is convex. Let* $(x, u), (y, v) \in Epi(f)$ *and* $0 \leq \lambda \leq 1$. *Then by definition of* $Epi(f)$, $f(x) \leq u$, $f(y) \leq v$ *and, therefore,* $\lambda f(x) + (1-\lambda)f(v) \leq \lambda u + (1-\lambda)v$. *Since* f *is convex,* $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$. *But* $\lambda f(x) + (1-\lambda)f(v) \leq \lambda u + (1-\lambda)v$. *Now by definition of* $Epi(f)$, $(\lambda x + (1-\lambda)y, \lambda u + (1-\lambda)v) \in Epi(f)$ *making* $Epi(f)$ *convex.* $\square$

Epigraph and Convex Sets

Proposition

If the epigraph of a function is a convex set, then the function is convex.

Proof.

Suppose $Epi(f)$ *is a convex set. Then by definition of* $Epi(f)$, $(x, f(x)) \in Epi(f)$ *and* $(y, f(y)) \in Epi(f)$. *Since* $Epi(f)$ *is convex,* $\lambda(x, f(x)) + (1 - \lambda)(y, f(y)) \in Epi(f)$. *Hence* $(\lambda x + (1 - \lambda y), \lambda f(x) + (1 - \lambda)f(y)) \in Epi(f)$. *By definition of* $Epi(f)$, $f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y)$. *And by definition of a convex function, this implies that* f *is convex.* $\square$

Theorem

A function is convex if and only if its epigraph is a convex set.

Basin sets of Convex Functions

Definition

Let $f : \mathbb{R}^N \to \mathbb{R}$ and $c \in \mathbb{R}$. A basin set of f is any set of the form

$$L = \{x \in \mathbb{R}^N \mid f(x) \leq c\}$$

Theorem

Let C be a convex set, h be a convex function on C and $L = \{c \in C \mid h(c) \leq b\}$. Then L is a convex set.

Proof.

Let $x, y \in L$ so that $h(x) \leq b$ and $h(y) \leq b$ and let $0 \leq \lambda \leq 1$. Since $x, y \in L \subseteq C$ and since C is a convex set, $\lambda x + (1-\lambda)y \in C$. Then since h is a convex function,

$$h(\lambda x + (1-\lambda)y) \leq \lambda h(x) + (1-\lambda)h(y) \leq \lambda b + (1-\lambda)b = b$$

This implies by definition of L that $\lambda x + (1-\lambda)y \in L$. $\square$

Minima Set of A Convex Function is Convex

Corollary

Let $C \subset \mathbb{R}^N$ be a closed and bounded convex set. Let $h : C \to \mathbb{R}$ be a convex function. Suppose $b = \min_{c \in C} h(c)$. Then $M = \{x \in C \mid h(x) = b\}$ is a convex set.

Proof.

Note that since $b = \min_{c \in C} h(c)$, $M = \{x \in C \mid h(x) \leq b\}$. C being closed and bounded is needed because the minima of h may be on the boundary. $\square$

For Convex Functions Local Minima are Global Minima

Theorem

Let C be a convex set and h be a convex function on C. Suppose h has a local minima at $x_0 \in C$. Then for any $x \in C,\ h(x_0) \leq h(x)$.

Proof.

Let $x \in C$ and $1 \geq \alpha > 0$ be sufficiently small so that $(1-\alpha)x_0 + \alpha x \in C$. Then,

$$
\begin{aligned}
h(x_0) &\leq h((1-\alpha)x_0 + \alpha x) \leq (1-\alpha)h(x_0) + \alpha h(x) \\
0 &\leq \alpha(h(x) - h(x_0)) \\
h(x_0) &\leq h(x)
\end{aligned}
$$

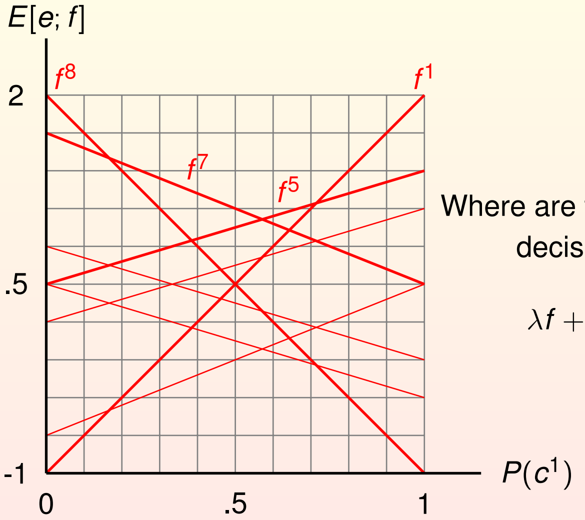
$\square$

Dependence on Prior Class Probabilities

$$E[e;f] = \sum_{j=1}^{K-1} \{E[e \mid c^j;f] - E[e \mid c^K;f]\} P(c^j) + E[e \mid c^K;f]$$

	Measurements			Conditional Gain		Expected Gain
f	d^1	d^2	d^3	$E[e\|c^1;f]$	$E[e\|c^2;f]$	$E[e;f,P(c^1)]$
f^1	c^1	c^1	c^1	2.0	-1.0	$3P(c^1) - 1$
f^2	c^1	c^1	c^2	.5	-.7	$1.2P(c^1) - .7$
f^3	c^1	c^2	c^1	1.1	.2	$.9P(c^1) + .2$
f^4	c^1	c^2	c^2	-.4	.5	$-.9P(c^1) + .5$
f^5	c^2	c^1	c^1	1.4	.5	$.9P(c^1) + .5$
f^6	c^2	c^1	c^2	-.1	.8	$-.9P(c^1) + .8$
f^7	c^2	c^2	c^1	.5	1.7	$-1.2P(c^1) + 1.7$
f^8	c^2	c^2	c^2	-1.	2.0	$-3P(c^1) + 2.0$

Dependence on Class Prior Probabilities



Where are the probabilistic decision rules?

$$\lambda f + (1 - \lambda) g$$

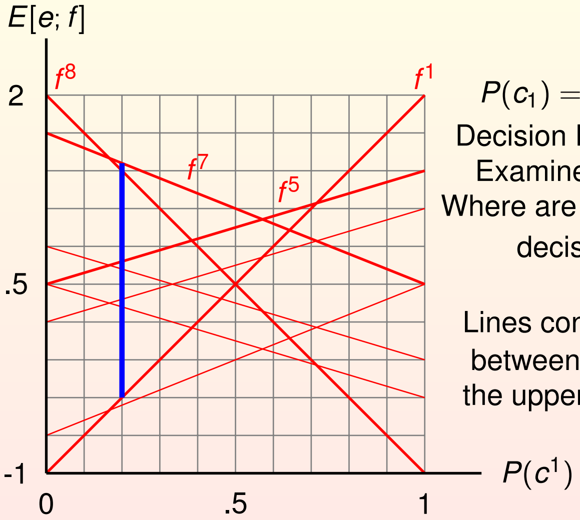
Probabilistic Decision Rules

- Pick a prior probability $P(c^1)$
- For decision rule f there is an Expected Gain $E[e; f]$
- For decision rule g there is a Expected Gain $E[e; g]$
- For decision rule $\lambda f + (1 - \lambda)g$, the Expected Gain is

$$\lambda E[e; f] + (1 - \lambda)E[e; g]$$

  - In between the Expected Gain for f and the Expected Gain for g

Dependence on Class Prior Probabilities



$P(c_1) = .2$

Decision Rules f^7 and f^1

Examine the blue line

Where are the probabilistic decision rules?

Lines contained in the area between the lower and the upper envelopes

Probabilistic Decision Rules Are In Between

Expected gain of a mixed decision rule is the mixture of the expected gains of the component decision rules.

$$E[e; \lambda f + (1-\lambda) g, P(c^1)] \quad = \quad \lambda E[e; f, P(c^1)] + (1-\lambda) E[e; g; P(c^1)]$$

A Mixed Decision Rule is an affine function of $P(c^1)$

Two Class Case

Proposition

Let $0 \leq \lambda \leq 1$. Let f_1 and f_2 be two decision rules and Suppose there are two classes, then $E[e; \lambda f_1 + (1-\lambda) f_2, P(c^1)]$ is an affine function of $P(c^1)$.

Proof.

$$
\begin{aligned}
E[e; f_1, P(c^1)] &= \alpha_1 P(c^1) + \beta_1 \\
E[e; f_2, P(c^1)] &= \alpha_2 P(c^1) + \beta_2 \\
E[e; \lambda f_1 + (1-\lambda) f_2, P(c^1)] &= \lambda(\alpha_1 P(c^1) + \beta_1) + (1-\lambda)(\alpha_2 P(c^1) + \beta_2) \\
&= (\lambda \alpha_1 + (1-\lambda)\alpha_2) P(c^1) + \lambda \beta_1 + (1-\lambda)\beta_2
\end{aligned}
$$

$\square$

Probabilistic Decision Rules Are In Between

Proposition

Fix $P(c^1)$*. Let* $0 \leq \lambda \leq 1$*.*
If $E[e; f, P(c^1)] \leq E[e; g, P(c^1)] =$ *then*

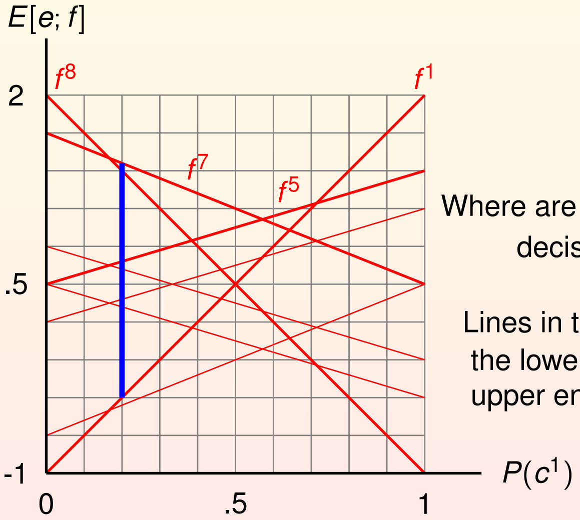
$$E[e; f, P(c^1)] \leq E[e; \lambda f + (1-\lambda) g] \leq E[e; g, P(c^1)]$$

Proof.

$$\begin{aligned}
E[e; f, P(c^1)] &= \lambda E[e; f, P(c^1)] + (1-\lambda) E[e; f, P(c^1)] \\
E[e; f, P(c^1)] &\leq \lambda E[e; f, P(c^1)] + (1-\lambda) E[e; g, P(c^1)] \leq E[e; g, P(c^1)) \\
E[e; f, P(c^1)) &\leq E[e; \lambda f + (1-\lambda) g] \leq E[e; g, P(c^1)]
\end{aligned}$$

$\square$

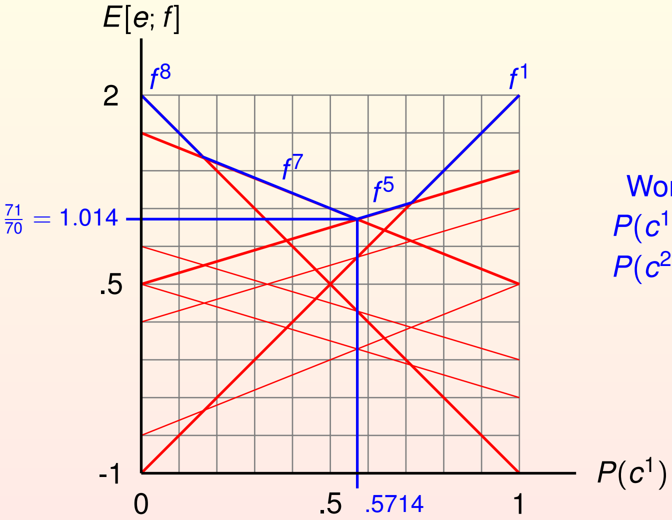
Dependence on Class Prior Probabilities



Where are the probabilistic decision rules?

Lines in the area between the lower and the upper envelopes

Dependence on Class Prior Probabilities



Worst Class Priors

$P(c^1) = 4/7 \approx .5714$

$P(c^2) = 3/7 \approx .4286$

Two Class Case
Decision Rules of Mixture are Known

$$
\begin{aligned}
E[e; f_5; P(c^1)] &= .9P(c^1) + .5 \\
E[e; f_7; P(c^1)] &= -1.2P(c^1) + 1.7 \\
\text{Set } E[e; f_5; P(c^1)] &= E[e; f_7; P(c^1)] \\
.9P(c^1) + .5 &= -1.2P(c^1) + 1.7; \\
2.1P(c^1) &= 1.2 \\
P(c^1) &= \frac{1.2}{2.1} = \frac{4}{7} \\
P(c^2) &= 1 - P(c^1) = \frac{3}{7}
\end{aligned}
$$

Dependence of a Probabilistic Decision Rule on Priors

Suppose we know the deterministic decision rules to make up the mixture: f_5 and f_7
Since $E[e;f] = E[e;c^1,f]P(c^1) + E[e;c^2,f]P(c^2)$

$$E[e;\lambda f^5 + (1-\lambda)f^7] \quad = \quad E[e|c^1;\lambda f^5 + (1-\lambda)f^7]P(c^1) + E[e|c^2;\lambda f^5 + (1-\lambda)f^7]P(c^2)$$

Since Expectation is a linear operator $E[e|c;\alpha f + \beta g] = \alpha E[e|c;f] + \beta E[e|c;g]$

$$\begin{aligned} E[e;\lambda f^5 + (1-\lambda)f^7] &= \left(\lambda E[e|c^1;f^5] + (1-\lambda)E[e|c^1;f^7]\right)P(c^1) + \\ &\quad \left(\lambda E[e|c^2;f^5] + (1-\lambda)E[e|c^2;f^7]\right)(1-P(c^1)) \\ &= \left\{\left(\lambda E[e|c^1;f^5] + (1-\lambda)E[e|c^1;f^7]\right) - \right. \\ &\quad \left.\left(\lambda E[e|c^2;f^5] + (1-\lambda)E[e|c^2;f^7]\right)\right\}P(c^1) + \\ &\quad E[e|c^2;\lambda f^5 + (1-\lambda)f^7] \end{aligned}$$

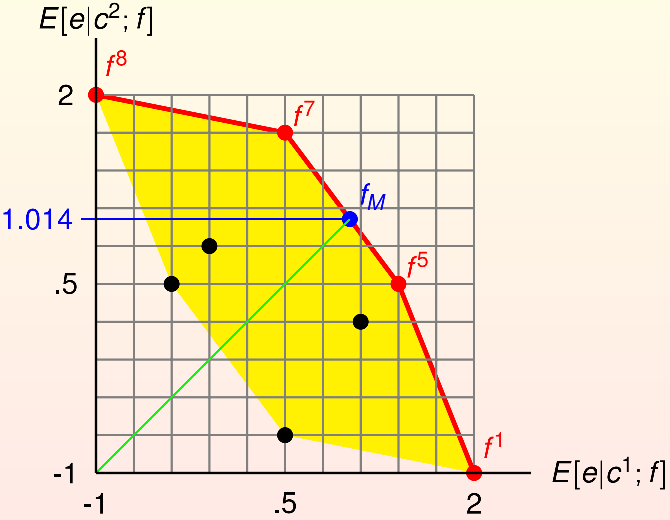
When there is no dependence on priors, the coefficient of $P(c^1)$ must be zero

$$\left(\lambda E[e|c^1;f^5] + (1-\lambda)E[e|c^1;f^7]\right) - \left(\lambda E[e|c^2;f^5] + (1-\lambda)E[e|c^2;f^7]\right) \quad = \quad 0$$

The class conditional expected gains must be equal

$$\begin{aligned} \lambda E[e|c^1;f^5] + (1-\lambda)E[e|c^1;f^7] &= \lambda E[e|c^2;f^5] + (1-\lambda)E[e|c^2;f^7] \\ E[e|c^1;\lambda f_5 + (1-\lambda)f_7] &= E[e|c^3;\lambda f_5 + (1-\lambda)f_7] \end{aligned}$$

Maximin Decision Rule



Dependence of a Probabilistic Decision Rule on Priors

$$\lambda E[e|c^1;f^5] + (1-\lambda)E[e|c^1;f^7] - \left(\lambda E[e|c^2;f^5] + (1-\lambda)E[e|c^2;f^7]\right) = 0$$

$$\lambda \left(E[e|c^1;f^5] - E[e|c^1;f^7] - E[e|c^2;f^5] + E[e|c^2;f^7]\right) = E[e|c^2;f^7] - E[e|c^1;f^7]$$

$$\begin{aligned}\lambda &= \frac{E[e|c^2;f^7] - E[e|c^1;f^7]}{E[c^1;f^5] - E[c^1;f^7] - E[e|c^2;f^5] + E[e|c^2;f^7]} \\ &= \frac{1.7 - .5}{1.4 - .5 - .5 + 1.7} = \frac{1.2}{2.1} = \frac{4}{7}\end{aligned}$$

Require $0 \le \lambda \le 1$

$$\lambda = \frac{E[e|c^2; f^7] - E[e|c^1; f^7]}{E[c^1; f^5] - E[c^1; f^7] - E[e|c^2; f^5] + E[e|c^2; f^7]}$$

$\lambda \geq 0$ implies

$$Sign\left(E[e|c^2; f^7] - E[e|c^1; f^7]\right) = Sign\left(E[c^1; f^5] - E[c^1; f^7] - E[e|c^2; f^5] + E[e|c^2; f^7]\right)$$

$\lambda \leq 1$ implies

$$|E[e|c^2; f^7] - E[e|c^1; f^7]| \leq |E[c^1; f^5] - E[c^1; f^7] - E[e|c^2; f^5] + E[e|c^2; f^7]|$$

If either of these inequality cannot be satisfied, it implies that the mixture of f_5 and f_7 is wrong

Expected Gain As A Function of Priors

The Expected economic gain can be related to the class conditional expected economic gain and prior probabilities.

$$
\begin{aligned}
E[e; f] &= \sum_{j=1}^{K} E[e \mid c^j; f] P(c^j) \\
&= \left\{ \sum_{j=1}^{K-1} E[e \mid c^j; f] P(c^j) \right\} + E[e \mid c^K; f] P(c^K) \\
&= \left\{ \sum_{j=1}^{K-1} E[e \mid c^j; f] P(c^j) \right\} + E[e \mid c^K; f] \left(1 - \sum_{j=1}^{K-1} P(c_j) \right) \\
&= \left\{ \sum_{j=1}^{K-1} E[e \mid c^j; f] P(c^j) \right\} + E[e \mid c^K; f] - \sum_{j=1}^{K-1} E[e \mid c^K; f] P(c^j)
\end{aligned}
$$

$$
E[e; f, P(c^1), \ldots, P(c^{K-1})] = \left\{ \sum_{j=1}^{K-1} \left(E[e \mid c^j; f] - E[e \mid c^K; f] \right) P(c^j) \right\} + E[e \mid c^K; f]
$$

Expected Gain As A Function Of Priors

$$E[e; f, P(c^1), \ldots, P(c^{K-1})] \quad = \quad \left\{ \sum_{j=1}^{K-1} \Big(E[e \mid c^j; f] - E[e \mid c^K; f] \Big) P(c^j) \right\} + E[e \mid c^K; f]$$

Two Class Case

$$\begin{aligned}
E[e; f, P(c^1)] &= \Big(E[e \mid c^1; f] - E[e \mid c^2; f] \Big) P(c^1) + E[e \mid c^2; f] \\
&= \alpha P(c^1) + \gamma \\
E[e; f_1, P(c^1)] &= \alpha_{11} P(c^1) + \gamma_1 \\
E[e; f_2, P(c^1)] &= \alpha_{21} P(c^1) + \gamma_2
\end{aligned}$$

When the expected gains of f_1 and f_2 are the same

$$\begin{aligned}
E[e; f_1, P(c^1)] &= E[e; f_2, P(c^1)] \\
\alpha_{11} P(c^1) + \gamma_1 &= \alpha_{21} P(c^1) + \gamma_2 \\
(\alpha_{11} - \alpha_{21}) P(c^1) &= \gamma_2 - \gamma_1 \\
P(c^1) &= \frac{\alpha_{11} - \alpha_{21}}{\gamma_2 - \gamma_1}
\end{aligned}$$

Three Class Case

$$\begin{aligned} E[e; f_i; P(c^1), P(c^2)] &= \alpha_{i1} P(c^1) + \alpha_{i2} P(c^2) + \gamma_i, \;\; i = 1, 2, 3 \\ E[e; f_i; P(c^1), P(c^2)] &= E[e; f_3; P(c^1), P(c^2)], \;\; i = 1, 2 \end{aligned}$$

$$\begin{aligned} \alpha_{11} P(c^1) + \alpha_{12} P(c^2) + \gamma_1 &= \alpha_{31} P(c^1) + \alpha 32 P(c^2) + \gamma_3 \\ \alpha_{21} P(c^1) + \alpha_{22} P(c^2) + \gamma_2 &= \alpha_{31} P(c^1) + \alpha 32 P(c^2) + \gamma_3 \end{aligned}$$

$$\begin{pmatrix} \alpha_{11} - \alpha_{31} & \alpha_{12} - \alpha_{32} \\ \alpha_{21} - \alpha_{31} & \alpha_{22} - \alpha_{32} \end{pmatrix} \begin{pmatrix} P(c^1) \\ P(c^2) \end{pmatrix} = \begin{pmatrix} \gamma_1 - \gamma_3 \\ \gamma_2 - \gamma_3 \end{pmatrix}$$

K Class Case

$$
\begin{aligned}
E[e; f_k; P(c^1), \ldots, P(c^{K-1})] &= \sum_{i=1}^{K-1} \alpha_{ki} P(c^i) + \gamma_k,\ k = 1, \ldots, K \\
E[e; f_k; P(c^1), \ldots, P(c^{K-1})] &= E[e; f_K; P(c^1), \ldots, P(c^{K-1})],\ k = 1, \ldots, K-1
\end{aligned}
$$

$$
\begin{pmatrix}
\alpha_{11} - \alpha_{K1} & \alpha_{12} - \alpha_{K2} & \cdots & \alpha_{1,K-1} - \alpha_{K,K-1} \\
 & & \vdots & \\
\alpha_{K-1,1} - \alpha_{K1} & \alpha_{K-1,2} - \alpha_{K2} & \cdots & \alpha_{K-1,K-1} - \alpha_{K,K-1}
\end{pmatrix}
\begin{pmatrix}
P(c^1) \\ \vdots \\ P(c^{K-1})
\end{pmatrix}
=
\begin{pmatrix}
\gamma_1 - \gamma_K \\ \vdots \\ \gamma_{K-1} - \gamma_K
\end{pmatrix}
$$

$$
\begin{aligned}
0 &\leq P(x^k) \leq 1,\ k = 1, \ldots, K-1 \\
\sum_{k=1}^{K} P(c^k) &= 1
\end{aligned}
$$

Two Class Case

$$\lambda E[e|c^1;f^5] + \Big(1-\lambda)E[e|c^1;f^7] - (\lambda E[e|c^2;f^5] + (1-\lambda)E[e|c^2;f^7]\Big) \quad = \quad 0$$

$$E[e;\lambda f^5 + (1-\lambda)f^7] \quad = \quad E[e|c^1;\lambda f^5 + (1-\lambda)f^7]P(c^1) + E[e|c^2;\lambda f^5 + (1-\lambda)f^7]P(c^2)$$

Find $P(c^1)$ that solves $\alpha_{11}P(c^1)+\gamma_1 = \alpha_{21}P(c^1)+\gamma_2$. Call the solution $P_0(c^1)$. Consider the expected gain of a mixed decision rule that has expected gain $\alpha_{21}P_0(c^1)+\gamma_2$ for any prior $P(c^1)$.

$$\lambda(\alpha_{11}P(c^1)+\gamma_1) + (1-\lambda)(\alpha_{21}P(c^1)+\gamma_2) \quad = \quad \alpha_{21}P_0(c^1)+\gamma_2$$

$$\begin{aligned}(\lambda\alpha_{11} + (1-\lambda)\alpha_{21})P(c^1) \quad &= \quad \alpha_{21}P_0(c^1)+\gamma_2 - \lambda\gamma_1 - (1-\lambda)\gamma_2 \\ &= \quad \alpha_{21}P_0(c^1) - \lambda(\gamma_1+\gamma_2)\end{aligned}$$

Therefore, $\lambda\alpha_{11} + (1-\lambda)\alpha_{21} = 0$ and $\lambda = \frac{-\alpha_{21}}{\alpha_{11}-\alpha_{21}} = \frac{\alpha_{21}P_0(c^1)}{\gamma_1+\gamma_2}$

Two Class Case
Identity in $P(c^1)$ meaning For all $P(c^1)$

$$\begin{aligned} 0 \leq \lambda_1, \lambda_2 \leq 1 \\ \lambda_1 + \lambda_2 = 1 \\ \lambda_1(\alpha_{11}P(c^1) + \gamma_1) + \lambda_2(\alpha_{21}P(c^1) + \gamma_2) &= \alpha_{21}P_0(c^1) + \gamma_2 \end{aligned}$$

$$(\lambda_1\alpha_{11} + \lambda_2\alpha_{21})P(c^1) = \alpha_{21}P_0(c^1) + \gamma_2 - \lambda_1\gamma_1 - \lambda_2\gamma_2$$

This implies

$$\begin{aligned} \lambda_1\alpha_{11} + \lambda_2\alpha_{21} &= 0 \\ \lambda_1\gamma_1 + \lambda_2\gamma_2 &= \alpha_{21}P_0(c^1) + \gamma_2 \\ \lambda_1 + \lambda_2 &= 1 \end{aligned}$$

Finding the Convex Combination

$$
\begin{aligned}
\lambda_1 \alpha_{11} + \lambda_2 \alpha_{21} &= 0 \\
\lambda_1 \gamma_1 + \lambda_2 \gamma_2 &= \alpha_{21} P_0(c^1) + \gamma_2 \\
\lambda_1 + \lambda_2 &= 1
\end{aligned}
$$

$$
\begin{aligned}
\lambda_2 &= -\lambda_1 \frac{\alpha_{11}}{\alpha_{21}} \\
\lambda_1 + \lambda_2 &= \lambda_1 \left(1 - \frac{\alpha_{11}}{\alpha_{21}}\right) = 1 \\
\lambda_1 &= \frac{\alpha_{21}}{\alpha_{21} - \alpha_{11}}
\end{aligned}
$$

Finding the Convex Combination: Consistency Check

$$\begin{aligned} 0 \leq \quad \lambda_1, \lambda_2 \quad &\leq 1 \\ \lambda_1 \quad &= \quad \frac{\alpha_{21}}{\alpha_{21} - \alpha_{11}} \end{aligned}$$

Either $\alpha_{21} - \alpha_{11} > 0$ *or* < 0.
If $\alpha_{21} - \alpha_{11} > 0$ then

$$\begin{aligned} \alpha_{21} &> \alpha_{11} \\ \alpha_{21} &> 0 \end{aligned}$$

If $\alpha_{21} - \alpha_{11} < 0$ then,

$$\begin{aligned} \alpha_{21} &< \alpha_{11} \\ \alpha_{21} &< 0 \end{aligned}$$

Finding the Convex Combination

Once λ_1 and λ_2 are known, the exact value for $P_0(c^1)$ can be determined.

$$
\begin{aligned}
\lambda_1\gamma_1 + (1-\lambda_1)\gamma_2 &= \alpha_{21}P_0(c^1) + \gamma_2 \\
\lambda_1(\gamma_1 - \gamma_2) &= \alpha_{21}P_0(c^1) \\
P_0(c^1) &= \frac{\lambda_1(\gamma_1 - \gamma_2)}{\alpha_{21}} \\
&= \frac{\alpha_{21}}{\alpha_{21}-\alpha_{11}} \frac{\gamma_1 - \gamma_2}{\alpha_{21}} \\
&= \frac{\gamma_1 - \gamma_2}{\alpha_{21}-\alpha_{11}}
\end{aligned}
$$

Finding The Convex Combination

K Class Case
Identity in $P(c^1)\ldots, P(c^{K-1})$

$$\begin{aligned}
\sum_{k=1}^{K} \lambda_k \left(\sum_{i=1}^{K-1} \alpha_{ik} P(c^i) + \gamma_k \right) &= \sum_{i=1}^{K} \alpha_{Ki} P_0(c^i) + \gamma_K \\
\sum_{i=1}^{K-1} \left(\sum_{k=1}^{K} \lambda_k \alpha_{ik} \right) P(c^i) &= \sum_{i=1}^{K} \alpha_{Ki} P_0(c^i) + \gamma_K - \sum_{k=1}^{K} \lambda_k \gamma_k
\end{aligned}$$

Implies

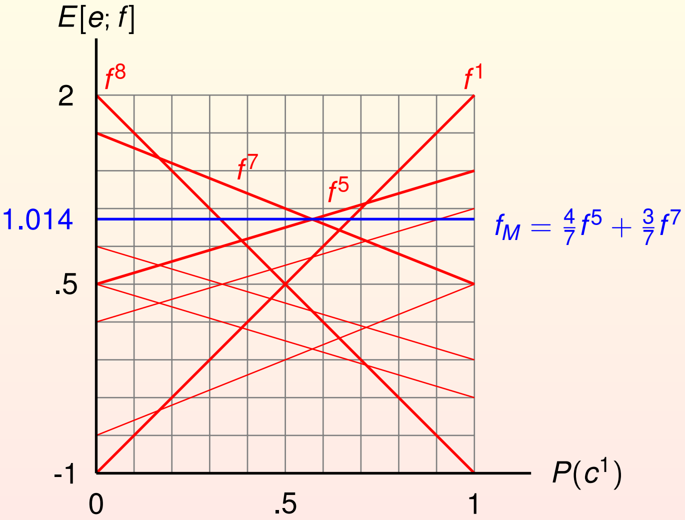
$$\begin{aligned}
\sum_{k=1}^{K} \lambda_k \alpha_{ik} &= 0,\ i = 1, \ldots, K-1 \\
\sum_{k=1}^{K} \lambda_k &= 1
\end{aligned}$$

Finding The Convex Combination

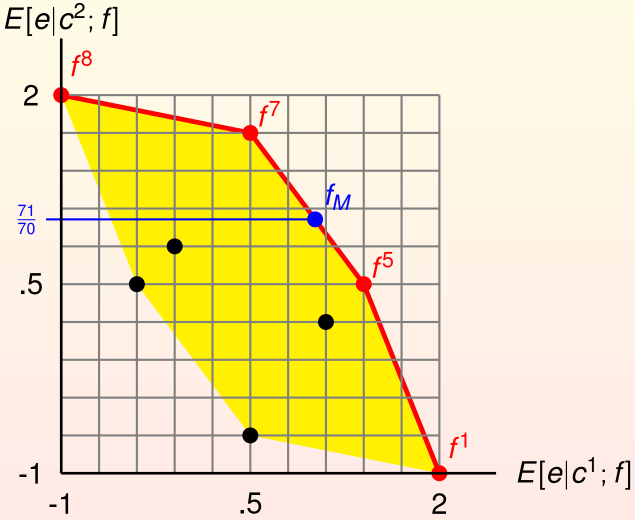
- Each component decision rule of the mixture has an expected gain that is a hyperplane in the axes $P(c^1) \ldots, P(c^{K-1})$
- The first $K-1$ rows of the i^th column consists of the coefficients of $P(c^1) \ldots, P(c^{K-1})$ for the i^{th} hyperplane

$$\begin{pmatrix} \alpha_{11} & \alpha_{21} & \ldots & \alpha_{K1} \\ \alpha_{21} & \alpha_{22} & \ldots & \alpha_{K2} \\ & & \vdots & \\ \alpha_{K-1,1} & \alpha_{K-1,2} & \ldots & \alpha_{K-1,K} \\ 1 & 1 & \ldots & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_{K-1} \\ \lambda_K \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

Dependence on Class Prior Probabilities



Conditional Expected Gains: All Decision Rules



Two Entity Game

The game is played for a large number of trials.

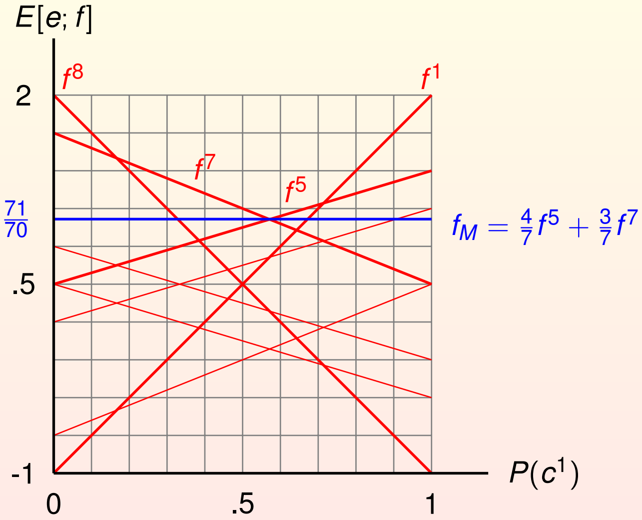
- Nature chooses class c in accordance with class priors $P(c^1), \ldots, P(c^K)$
- A measurement d is sampled in accordance with $P(d \mid c)$
- Bayes chooses decision rule to maximize expected gain under given class priors

Suppose nature chooses class priors so that the Bayes gain is minimized. Bayes chooses to maximize expected gain under worst priors. But suppose nature does not choose c in accordance with worst priors.

Maximin Decision Rule

There is a mixed decision rule that guarantees that regardless of what class priors nature chooses, the expected gain is equal to the Bayes gain under the worst class priors. This is the maximin decision rule.

Dependence on Class Prior Probabilities



Maximin Decision Rule

Definition

A decision rule f is a Maximin Decision Rule if and only if

$$\min_{P(c^1),\ldots,P(c^K)} \sum_{j=1}^{K} E[e \mid c^j; f]P(c^j) \geq \min_{P(c^1),\ldots,P(c^K)} \sum_{j=1}^{K} E[e \mid c^j; g]P(c^j)$$

for any decision rule g where

$$E[e \mid c^j; f] \;=\; \sum_{d \in D} \sum_{k=1}^{K} e(c^j, c^k) P(d \mid c^j) f_d(c^k)$$

Determining the Maximin Decision Rule

$$
\begin{aligned}
E[e;f] &= E[e|c^1;f]P(c^1) + E[e|c^2;f]P(c^2) \\
&= E[e|c^1;f]P(c^1) + E[e|c^2;f](1 - P(c^1)) \\
&= (E[e|c^1;f] - E[e|c^2;f])P(c^1) + E[e|c^2;f]
\end{aligned}
$$

Since a maximin decision rule has no dependence on the prior probability, we must have

$$
\begin{aligned}
E[e|c^1;f] - E[e|c^2;f] &= 0 \\
E[e|c^1;f] &= E[e|c^2;f]
\end{aligned}
$$

In this case,

$$
\begin{aligned}
E[e;f] &= E[e|c^1;f] \\
&= E[e|c^2;f]
\end{aligned}
$$

Maximin Decision Rule

Theorem

A decision rule f is a maximin decision rule if and only if

$$\min_{j=1,\ldots,K} E[e \mid c^j; f] \geq \min_{j=1,\ldots,K} E[e \mid c^j, g]$$

for any decision rule g.

Maximin Decision Rule

Theorem

A decision rule f is a maximin decision rule if and only if

$$\min_{P(c^1),\ldots,P(c^K)} E[e; f, P(c^1), \ldots, P(c^K)] \geq \min_{P(c^1),\ldots,P(c^K)} E[e; g, P(c^1), \ldots, P(c^K)]$$

for any decision rule g.

Proof.

Recall

$$E[e; f, P(c^1), \ldots, P(c^K)] = E[e; f] = \sum_{j=1}^{K} E[e \mid c^j; f] P(c^j)$$

$\square$

Maximin Decision Rule

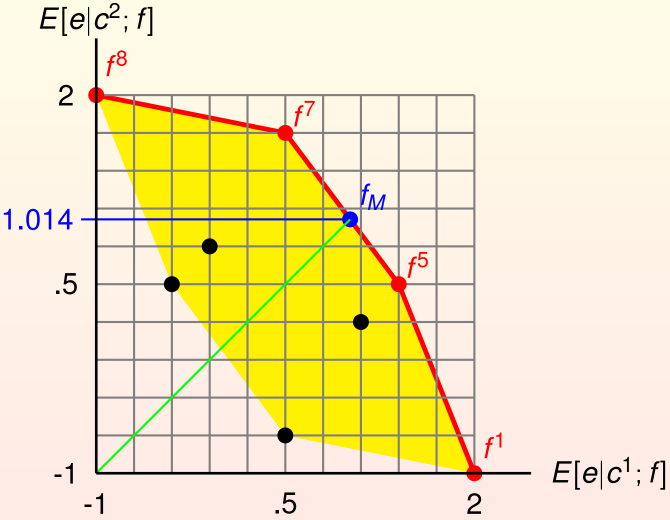
A decision rule f is a maximin decision rule if and only if the expected gain of f is the same as the expected gain of the Bayes rule under the worst possible prior class probabilities.

Theorem

Let G be the Bayes Economic Gain under the worst prior class probabilities. Then f is a maximin decision rule if and only if

$$E[e \mid c^j; f] = G, \; j = 1, \ldots, K$$

Maximin Decision Rule



Maximin Decision Rule

Let $P(c^1), \ldots, P(c^K)$ be given class prior probabilities. Let $f^m,\ m = 1, \ldots, M$ be M deterministic decision rules satisfying

$$G = \sum_{j=1}^{K} E[e \mid c^j; f^m] P(c^j),\ m = 1, \ldots, M$$

Then there exists $\lambda_m,\ \lambda_m \geq 0,\ m = 1, \ldots, M$, and $\sum_{m=1}^{M} \lambda_m = 1$ satisfying

$$G = E[e \mid c^j; \sum_{m=1}^{M} \lambda_m f^m],\ j = 1, \ldots, K$$

Note:

$$E[e \mid c^j; \sum_{m=1}^{M} \lambda_m f^m] = \sum_{m=1}^{M} \lambda_m E[e \mid c^j; f^m]$$

Maximin Decision Rule

- Let $P(c^1), \ldots, P(c^K)$ be the worst priors
- Let G_w be the worst Bayes gain
- Let f^m be deterministic decision rules, $m = 1, \ldots, M$
  - $G_w = \sum_{j=1}^K E[e \mid c^j; f^m] P(c^j)$
- Find convex combination $\lambda_1, \ldots, \lambda_M$
  - $G_w = E[e \mid c^k; \sum_{m=1}^M \lambda_m f^m] = \sum_{m=1}^M \lambda_m E[e \mid c^j; f^m], \; j = 1, \ldots, K$
- Let $a_{jm} = E[e \mid c^j; f^m]$
- Find convex combination $\lambda_1, \ldots, \lambda_M$ satisfying
  - $G_w = \sum_{m=1}^M \lambda_m a_{jm}, \; j = 1, \ldots, K$

Existence of Mixed Decision Rule Strategy

Theorem

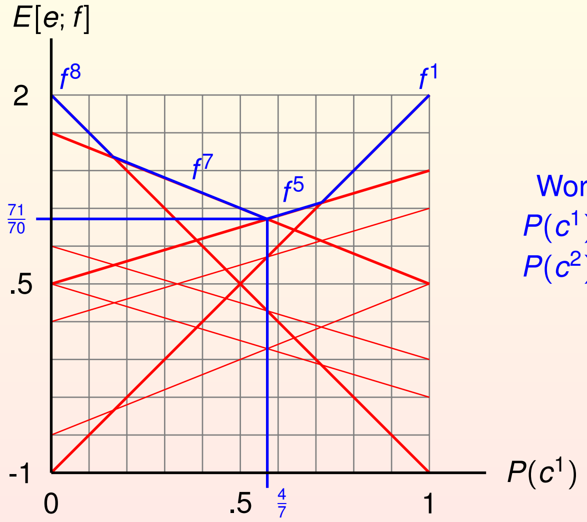
Let a_{jm} be a real numbers, $j = 1, \ldots, K$; $m = 1, \ldots, M$. Let $p_j \geq 0$ and $\sum_{j=1}^{K} p_j = 1$. Suppose

$$G = \sum_{j=1}^{K} p_j a_{jm}, \;\; m = 1, \ldots, M$$

Then there exists $\lambda_m, \;\; m = 1, \ldots, M$, $\lambda_m \geq 0$ and $\sum_{m=1}^{M} \lambda_m = 1$ satisfying

$$G = \sum_{m=1}^{M} a_{jm}\lambda_m, \;\; j = 1, \ldots, K$$

Dependence on Class Prior Probabilities



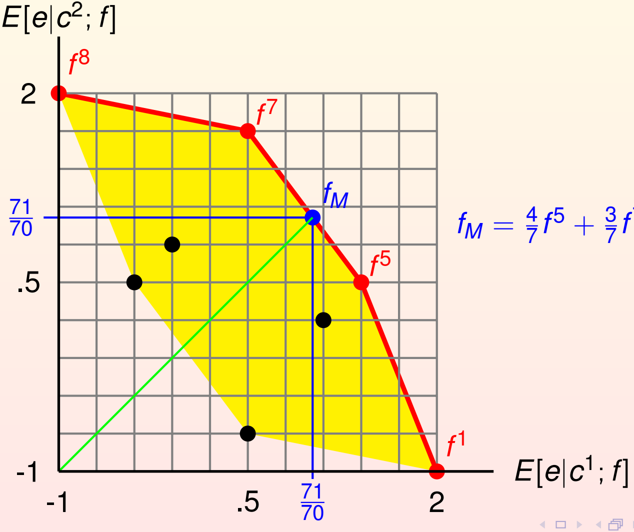
Worst Class Priors

$P(c^1) = 4/7 \approx .5714$

$P(c^2) = 3/7 \approx .4286$

Maximin Decision Rule

$$E[e|c^1; f_M] = E[e|c^2; f_M]$$



$$f_M = \frac{4}{7} f^5 + \frac{3}{7} f^7$$