Statistical Estimation

Parametric distribution estimation

- distribution estimation problem: estimate probability density $p(y)$ of a random variable from observed values

- parametric distribution estimation: choose from a family of densities $p_x(y)$, indexed by a parameter x

Parametric distribution estimation

- distribution estimation problem: estimate probability density $p(y)$ of a random variable from observed values

- parametric distribution estimation: choose from a family of densities $p_x(y)$, indexed by a parameter x

Maximum likelihood estimation

observed value

maximize (over x) $\quad \log p_x(y)$

log-likelihood function

Parametric distribution estimation

If we have prior information about x, such as $x \in C \subseteq \mathbf{R}^n$

add constraints $x \in C$ explicitly, or define $p_x(y) = 0$ for $x \notin C$

$$\begin{array}{ll} \text{maximize} & l(x) = \log p_x(y) \\ \text{subject to} & x \in C, \end{array}$$

Parametric distribution estimation

If we have prior information about x, such as $x \in C \subseteq \mathbf{R}^n$

add constraints $x \in C$ explicitly, or define $p_x(y) = 0$ for $x \notin C$

$$
\begin{array}{ll}
\text{maximize} & l(x) = \log p_x(y) \\
\text{subject to} & x \in C,
\end{array}
$$

a convex optimization problem if $\log p_x(y)$ is concave in x for fixed y

C can be described by a set of linear equality and convex inequality constraints

Linear measurements with IID noise

linear measurement model

$$y_i = a_i^T x + v_i, \quad i = 1, \ldots, m$$

- $x \in \mathbf{R}^n$ is vector of unknown parameters
- v_i is IID measurement noise, with density $p(z)$

Linear measurements with IID noise

linear measurement model

$$y_i = a_i^T x + v_i, \quad i = 1, \ldots, m$$

- $x \in \mathbf{R}^n$ is vector of unknown parameters
- v_i is IID measurement noise, with density $p(z)$
- y_i is measurement: $y \in \mathbf{R}^m$ has density $p_x(y) = \prod_{i=1}^m p(y_i - a_i^T x)$

Linear measurements with IID noise

linear measurement model

$$y_i = a_i^T x + v_i, \quad i = 1, \ldots, m$$

- $x \in \mathbf{R}^n$ is vector of unknown parameters
- v_i is IID measurement noise, with density $p(z)$
- y_i is measurement: $y \in \mathbf{R}^m$ has density $p_x(y) = \prod_{i=1}^m p(y_i - a_i^T x)$

maximum likelihood estimate:

$$\text{maximize} \quad l(x) = \sum_{i=1}^m \log p(y_i - a_i^T x)$$

Linear measurements with IID noise

Examples:

- Gaussian noise $\mathcal{N}(0, \sigma^2)$: $p(z) = (2\pi\sigma^2)^{-1/2} e^{-z^2/(2\sigma^2)}$,

Linear measurements with IID noise

Examples:

- Gaussian noise $\mathcal{N}(0, \sigma^2)$: $p(z) = (2\pi\sigma^2)^{-1/2} e^{-z^2/(2\sigma^2)}$,

$$l(x) = -\frac{m}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{m}(a_i^T x - y_i)^2$$

  ML estimate is LS solution

- Laplacian noise: $p(z) = (1/(2a))e^{-|z|/a}$,

Linear measurements with IID noise

Examples:

- Gaussian noise $\mathcal{N}(0, \sigma^2)$: $p(z) = (2\pi\sigma^2)^{-1/2} e^{-z^2/(2\sigma^2)}$,

$$l(x) = -\frac{m}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{m}(a_i^T x - y_i)^2$$

  ML estimate is LS solution

- Laplacian noise: $p(z) = (1/(2a))e^{-|z|/a}$,

$$l(x) = -m\log(2a) - \frac{1}{a}\sum_{i=1}^{m}|a_i^T x - y_i|$$

  ML estimate is ℓ_1-norm solution

- uniform noise on $[-a, a]$:

Linear measurements with IID noise

Examples:

- Gaussian noise $\mathcal{N}(0,\sigma^2)$: $p(z) = (2\pi\sigma^2)^{-1/2}e^{-z^2/(2\sigma^2)}$,

$$l(x) = -\frac{m}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{m}(a_i^Tx - y_i)^2$$

  ML estimate is LS solution

- Laplacian noise: $p(z) = (1/(2a))e^{-|z|/a}$,

$$l(x) = -m\log(2a) - \frac{1}{a}\sum_{i=1}^{m}|a_i^Tx - y_i|$$

  ML estimate is ℓ_1-norm solution

- uniform noise on $[-a,a]$:

$$l(x) = \begin{cases} -m\log(2a) & |a_i^Tx - y_i| \leq a, \quad i = 1,\ldots,m \\ -\infty & \text{otherwise} \end{cases}$$

  ML estimate is any x with $|a_i^Tx - y_i| \leq a$

Logistic regression

random variable $y \in \{0, 1\}$ with distribution

$$p = \mathbf{prob}(y = 1) = \frac{\exp(a^T u + b)}{1 + \exp(a^T u + b)}$$

- a, b are parameters; $u \in \mathbf{R}^n$ are (observable) explanatory variables

Logistic regression

random variable $y \in \{0, 1\}$ with distribution

$$p = \mathbf{prob}(y = 1) = \frac{\exp(a^T u + b)}{1 + \exp(a^T u + b)}$$

- a, b are parameters; $u \in \mathbf{R}^n$ are (observable) explanatory variables
- estimation problem: estimate a, b from m observations (u_i, y_i)

Logistic regression

random variable $y \in \{0, 1\}$ with distribution

$$p = \mathbf{prob}(y = 1) = \frac{\exp(a^T u + b)}{1 + \exp(a^T u + b)}$$

- a, b are parameters; $u \in \mathbf{R}^n$ are (observable) explanatory variables
- estimation problem: estimate a, b from m observations (u_i, y_i)

log-likelihood function for $u_1, \ldots, u_q$, the outcome is $y = 1$, and for $u_{q+1}, \ldots, u_m$ the outcome is $y = 0$

$$\prod_{i=1}^{q} p_i \prod_{i=q+1}^{m} (1 - p_i),$$

$$l(a, b) = \sum_{i=1}^{q} \log p_i + \sum_{i=q+1}^{m} \log(1 - p_i)$$

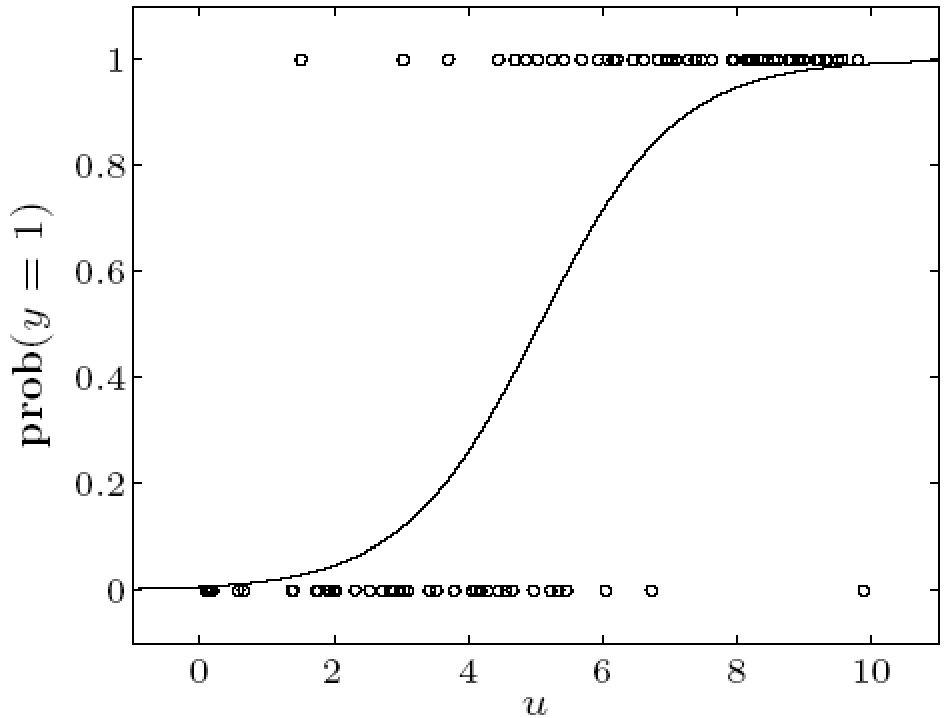
Logistic regression

$$\begin{aligned} l(a,b) &= \sum_{i=1}^{q} \log \frac{\exp(a^T u_i + b)}{1+\exp(a^T u_i + b)} + \sum_{i=q+1}^{m} \log \frac{1}{1+\exp(a^T u_i + b)} \\ &= \sum_{i=1}^{q} (a^T u_i + b) - \sum_{i=1}^{m} \log(1+\exp(a^T u_i + b)). \end{aligned}$$

$\boxed{\text{concave in } a,\, b}$

Logistic regression

example ($n = 1$, $m = 50$ measurements)



- circles show 50 points (u_i, y_i)
- solid curve is ML estimate of $p = \exp(au + b)/(1 + \exp(au + b))$

Maximum a posteriori probability estimation

- a Bayesian version of maximum likelihood estimation
- a prior probability density on the underlying parameter x.
- the vector to be estimated and the observation are random variables

Prior density

$$p(x, y)$$

$$p_x(x)$$

$$p_{y|x}(x, y) = \frac{p(x, y)}{p_x(x)}$$

Maximum a posteriori probability estimation

- a Bayesian version of maximum likelihood estimation
- a prior probability density on the underlying parameter x.
- the vector to be estimated and the observation are random variables

Prior density

$$p(x, y)$$

$$p_x(x)$$

$$p_{y|x}(x, y) = \frac{p(x, y)}{p_x(x)}$$

posterior density

$$p_{x|y}(x, y) = \frac{p(x, y)}{p_y(y)} = p_{y|x}(x, y)\frac{p_x(x)}{p_y(y)}$$

Maximum a posteriori probability estimation

$$\hat{x}_{\mathrm{map}} \quad = \quad \mathrm{argmax}_x p_{x|y}(x, y) \quad = \quad \mathrm{argmax}_x p_{y|x}(x, y) p_x(x)$$

$$\hat{x}_{\mathrm{map}} = \mathrm{argmax}_x (\log p_{y|x}(x, y) + \log p_x(x)).$$

Maximum a posteriori probability estimation

$$\hat{x}_{\mathrm{map}} = \mathrm{argmax}_x p_{x|y}(x,y) = \mathrm{argmax}_x p_{y|x}(x,y) p_x(x)$$

Different from ML

$$\hat{x}_{\mathrm{map}} = \mathrm{argmax}_x (\log p_{y|x}(x,y) + \log p_x(x)).$$

penalizes choices of x that are unlikely

Linear measurements with I ID noise

$$y_i = a_i^T x + v_i, \quad i = 1, \ldots, m,$$

$x \in \mathbf{R}^n$ and $y \in \mathbf{R}^m$

v_i are IID with density p_v on $\mathbf{R}$

x has prior density p_x on $\mathbf{R}^n$

Linear measurements with I ID noise

$$y_i = a_i^T x + v_i, \quad i = 1, \ldots, m,$$

$x \in \mathbf{R}^n$ and $y \in \mathbf{R}^m$

v_i are IID with density p_v on $\mathbf{R}$

x has prior density p_x on $\mathbf{R}^n$

$$p(x, y) = p_x(x) \prod_{i=1}^{m} p_v(y_i - a_i^T x)$$

$$\boxed{\text{maximize} \quad \log p_x(x) + \textstyle\sum_{i=1}^m \log p_v(y_i - a_i^T x).}$$

Linear measurements with I ID noise

$$\boxed{\text{maximize} \quad \log p_x(x) + \sum_{i=1}^{m} \log p_v(y_i - a_i^T x).}$$

v_i are uniform on $[-a, a]$

x is Gaussian with mean $\bar{x}$ and covariance Σ

Linear measurements with I ID noise

$$\boxed{\text{maximize} \quad \log p_x(x) + \sum_{i=1}^m \log p_v(y_i - a_i^T x).}$$

v_i are uniform on $[-a, a]$

x is Gaussian with mean $\bar{x}$ and covariance Σ

$$\begin{array}{ll} \text{minimize} & (x-\bar{x})^T \Sigma^{-1} (x-\bar{x}) \\ \text{subject to} & \|Ax - y\|_\infty \leq a, \end{array}$$

MAP with perfect linear measurements

m perfect (noise free, deterministic) linear measurements

$$y = Ax$$

MAP estimate

MAP with perfect linear measurements

m perfect (noise free, deterministic) linear measurements

$$y = Ax$$

MAP estimate

$$\begin{array}{ll} \text{maximize} & \log p_x(x) \\ \text{subject to} & Ax = y. \end{array}$$

MAP with perfect linear measurements

m perfect (noise free, deterministic) linear measurements

$$y = Ax$$

MAP estimate

$$\begin{array}{ll} \text{maximize} & \log p_x(x) \\ \text{subject to} & Ax = y. \end{array}$$

xi are IID with density p

$$\begin{array}{ll} \text{maximize} & \sum_{i=1}^n \log p(x_i) \\ \text{subject to} & Ax = y, \end{array}$$

$$\phi(u) = -\log p(u)$$

extension: least-penalty problem

$$\begin{array}{ll} \text{minimize} & \phi(x_1) + \cdots + \phi(x_n) \\ \text{subject to} & Ax = b \end{array}$$

$\phi : \mathbf{R} \to \mathbf{R}$ is convex penalty function

Nonparametric distribution estimation

a random variable X with values in the finite set $\{\alpha_1, \dots, \alpha_n\} \subseteq \mathbf{R}$.

$p \in \mathbf{R}^n$, with $\mathbf{prob}(X = \alpha_k) = p_k$.

all possible probability distributions

$$\boxed{\{p \in \mathbf{R}^n \mid p \succeq 0,\ \mathbf{1}^T p = 1\}}$$

Nonparametric distribution estimation

a random variable X with values in the finite set $\{\alpha_1, \ldots, \alpha_n\} \subseteq \mathbf{R}$.

$p \in \mathbf{R}^n$, with $\mathbf{prob}(X = \alpha_k) = p_k$.

all possible probability distributions

$$\boxed{\{p \in \mathbf{R}^n \mid p \succeq 0, \ \mathbf{1}^T p = 1\}}$$

Prior information

$$\mathbf{E}\, X = \alpha, \qquad \mathbf{E}\, X^2 = \beta$$

Nonparametric distribution estimation

a random variable X with values in the finite set $\{\alpha_1, \ldots, \alpha_n\} \subseteq \mathbf{R}$.

$p \in \mathbf{R}^n$, with $\mathbf{prob}(X = \alpha_k) = p_k$.

all possible probability distributions

$$\boxed{\{p \in \mathbf{R}^n \mid p \succeq 0,\ \mathbf{1}^T p = 1\}}$$

Prior information

linear equalities

$$\mathbf{E}\, X = \alpha, \qquad \mathbf{E}\, X^2 = \beta$$

$$\mathbf{E}\, X = \sum_{i=1}^n \alpha_i p_i = \alpha, \qquad \mathbf{E}\, X^2 = \sum_{i=1}^n \alpha_i^2 p_i = \beta, \qquad \mathbf{E}\, f(X) = \sum_{i=1}^n p_i f(\alpha_i)$$

Nonparametric distribution estimation

a random variable X with values in the finite set $\{\alpha_1, \ldots, \alpha_n\} \subseteq \mathbf{R}$.

$p \in \mathbf{R}^n$, with $\mathbf{prob}(X = \alpha_k) = p_k$.

all possible probability distributions

$$\boxed{\{p \in \mathbf{R}^n \mid p \succeq 0,\ \mathbf{1}^T p = 1\}}$$

Prior information

linear equalities

$$\mathbf{E}\,X = \alpha, \qquad \mathbf{E}\,X^2 = \beta$$

$$\mathbf{E}\,X = \sum_{i=1}^n \alpha_i p_i = \alpha, \qquad \mathbf{E}\,X^2 = \sum_{i=1}^n \alpha_i^2 p_i = \beta, \qquad \mathbf{E}\,f(X) = \sum_{i=1}^n p_i f(\alpha_i)$$

Nonparametric distribution estimation

Prior information

$$\mathbf{prob}(X \geq 0) \leq 0.3. \qquad \sum_{\alpha_i \geq 0} p_i \leq 0.3,$$

$$\mathbf{prob}(X \in C) = c^T p, \qquad c_i = \begin{cases} 1 & \alpha_i \in C \\ 0 & \alpha_i \notin C. \end{cases}$$

Nonparametric distribution estimation

Prior information

$$\mathbf{prob}(X \geq 0) \leq 0.3. \qquad \sum_{\alpha_i \geq 0} p_i \leq 0.3,$$

$$\mathbf{prob}(X \in C) = c^T p, \qquad c_i = \begin{cases} 1 & \alpha_i \in C \\ 0 & \alpha_i \notin C. \end{cases}$$

Some types of prior information can be expressed in terms of nonlinear convex inequalities. Such as a min for entropy

$$-\sum_{i=1}^{n} p_i \log p_i,$$

Concave function

a set of linear equalities and convex inequalities.

express the prior information about the distribution p as $p \in \mathcal{P}$

Nonparametric distribution estimation

Goal : min $\mathbf{E}\, f(X)$

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{n} f(\alpha_i) p_i \\ \text{subject to} & p \in \mathcal{P}. \end{array}$$

Nonparametric distribution estimation

Goal : min $\mathbf{E}\, f(X)$

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{n} f(\alpha_i) p_i \\ \text{subject to} & p \in \mathcal{P}. \end{array}$$

Maximum entropy

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{n} p_i \log p_i \\ \text{subject to} & p \in \mathcal{P}. \end{array}$$

Nonparametric distribution estimation

Goal : min $\mathbf{E} f(X)$

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{n} f(\alpha_i) p_i \\ \text{subject to} & p \in \mathcal{P}. \end{array}$$

Maximum entropy

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{n} p_i \log p_i \\ \text{subject to} & p \in \mathcal{P}. \end{array}$$

Maximum likelihood estimation

observe N independent samples $x_1, \ldots, x_N$

k_i denote the number of these samples with value α_i

$$k_1 + \cdots + k_n = N$$

$$\begin{array}{ll} \text{maximize} & l(p) = \sum_{i=1}^{n} k_i \log p_i \\ \text{subject to} & p \in \mathcal{P}, \end{array}$$

Optimal detector design and hypothesis testing

Suppose X is a random variable with values in $\{1, \ldots, n\}$, with a distribution that depends on a parameter $\theta \in \{1, \ldots, m\}$.

Hypotheses

a matrix $P \in \mathbf{R}^{n \times m}$

$$p_{kj} = \mathbf{prob}(X = k \mid \theta = j)$$

hypothesis testing

$\boxed{\text{estimating } \theta, \text{ based on an observed sample of } X}$

- ❖one of the hypotheses corresponds to some normal situation
- ❖others correspond to some abnormal event.

Detection

Deterministic and randomized detectors

A (deterministic) estimator or detector is a function ϕ from {1,... ,n} (the set of possible observed values) into {1,..., m}

$$\hat{\theta} = \psi(k)$$

1. maximum likelihood detector

$$\hat{\theta} = \psi_{\mathrm{ml}}(k) = \underset{j}{\operatorname{argmax}}\, p_{kj}$$

estimate of θ, given an observed value of X, is random $\longrightarrow$ *randomized detector*

a random variable $\hat{\theta} \in \{1, \ldots, m\}$, with a distribution that depends on the observed value of X

$T \in \mathbf{R}^{m \times n}$ $\qquad$ $t_{ik} = \mathbf{prob}(\hat{\theta} = i \mid X = k)$

designing the matrix T

the columns t_k of T must satisfy $\qquad t_k \succeq 0, \qquad \mathbf{1}^T t_k = 1$

Detection probability matrix

Detection probability matrix

$$D = TP \qquad D_{ij} = (TP)_{ij} = \mathbf{prob}(\hat{\theta} = i \mid \theta = j)$$

✓characterizes the performance of the randomized detector
✓diagonal and off-diagonal entry
✓Perfect detector

Detection probabilities → $$P_i^{\mathrm{d}} = D_{ii} = \mathbf{prob}(\hat{\theta} = i \mid \theta = i).$$

Error probabilities → $$P_i^{\mathrm{e}} = 1 - D_{ii} = \mathbf{prob}(\hat{\theta} \neq i \mid \theta = i)$$

$$P_i^{\mathrm{e}} = \sum_{j \neq i} D_{ji}.$$

Optimal detector design

Limits on errors and detection probabilities

$$P_j^{\mathrm{d}} = D_{jj} \geq L_j, \qquad D_{ij} \leq U_{ij},$$

1. Minimax detector design

minimax detector minimizes the worst-case (largest) probability of error over all m hypotheses.

$$\begin{array}{ll} \text{minimize} & \max_j P_j^{\mathrm{e}} \\ \text{subject to} & t_k \succeq 0, \quad \mathbf{1}^T t_k = 1, \quad k = 1, \ldots, n, \end{array}$$

can be reformulated as an LP.

Optimal detector design

2. Bayes detector design

In Bayes detector design, we have a prior distribution for the hypotheses

$$q_i = \mathbf{prob}(\theta = i)$$

The probability of error for the detector is then given by $q^T P^{\mathrm{e}}$

$$\begin{array}{ll} \text{minimize} & q^T P^{\mathrm{e}} \\ \text{subject to} & t_k \succeq 0, \quad \mathbf{1}^T t_k = 1, \quad k = 1, \ldots, n. \end{array}$$

an LP

Binary hypothesis testing

Detection problem

given observation of a random variable $X \in \{1, \ldots, n\}$, choose between:

- hypothesis 1: X was generated by distribution $p = (p_1, \ldots, p_n)$
- hypothesis 2: X was generated by distribution $q = (q_1, \ldots, q_n)$

randomized detector

- a nonnegative matrix $T \in \mathbf{R}^{2 \times n}$
- if we observe $X = k$, we choose hypothesis 1 with probability t_{1k}, hypothesis 2 with probability t_{2k}
- if all elements of T are 0 or 1, it is called a deterministic detector

Binary hypothesis testing

detection probability matrix:

$$D = TP \qquad D_{ij} = (TP)_{ij} = \mathbf{prob}(\hat{\theta} = i \mid \theta = j)$$

$$D = \left[\begin{array}{cc} Tp & Tq \end{array}\right] = \left[\begin{array}{cc} 1 - P_{\mathrm{fp}} & P_{\mathrm{fn}} \\ P_{\mathrm{fp}} & 1 - P_{\mathrm{fn}} \end{array}\right]$$

- P_{fp} is probability of selecting hypothesis 2 if X is generated by distribution 1 (false positive)
- P_{fn} is probability of selecting hypothesis 1 if X is generated by distribution 2 (false negative)

multicriterion formulation of detector design

variable $T \in \mathbf{R}^{2\times n}$

$$\begin{array}{ll} \text{minimize (w.r.t. } \mathbf{R}_+^2) & (P_{\mathrm{fp}}, P_{\mathrm{fn}}) = ((Tp)_2, (Tq)_1) \\ \text{subject to} & t_{1k} + t_{2k} = 1, \quad k = 1, \ldots, n \\ & t_{ik} \geq 0, \quad i = 1, 2, \quad k = 1, \ldots, n \end{array}$$

Binary hypothesis testing

scalarization (with weight $\lambda > 0$)

$$\begin{array}{ll} \text{minimize} & (Tp)_2 + \lambda(Tq)_1 \\ \text{subject to} & t_{1k} + t_{2k} = 1, \quad t_{ik} \geq 0, \quad i = 1, 2, \quad k = 1, \ldots, n \end{array}$$

an LP with a simple analytical solution

$$(t_{1k}, t_{2k}) = \begin{cases} (1, 0) & p_k \geq \lambda q_k \\ (0, 1) & p_k < \lambda q_k \end{cases}$$

$\boxed{\textit{likelihood ratio threshold test}}$

Example

$$P = \begin{bmatrix} 0.70 & 0.10 \\ 0.20 & 0.10 \\ 0.05 & 0.70 \\ 0.05 & 0.10 \end{bmatrix}$$

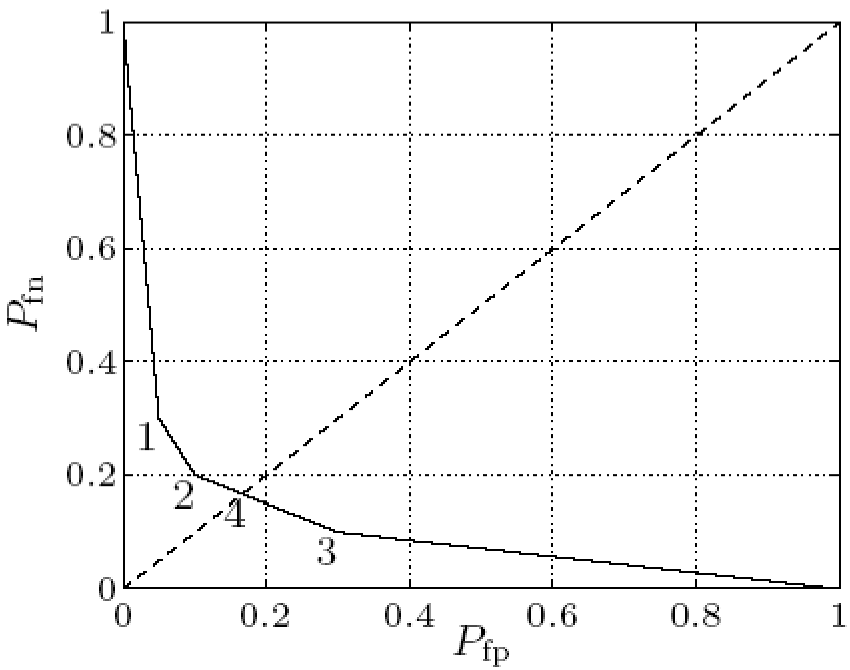
$$T^{(1)} = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

$$T^{(2)} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix},$$

$$T^{(3)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix},$$

$$T^{(4)} = \begin{bmatrix} 1 & 2/3 & 0 & 0 \\ 0 & 1/3 & 1 & 1 \end{bmatrix},$$

either a false positive or false negative probability that exceeds 1/6,



Experiment design

m linear measurements $y_i = a_i^T x + w_i$, $i = 1, \ldots, m$ of unknown $x \in \mathbf{R}^n$

- measurement errors w_i are IID $\mathcal{N}(0, 1)$
- ML (least-squares) estimate is $x^\star = (A^T A)^{-1} A^T b$

$$\hat{x} = \left(\sum_{i=1}^{m} a_i a_i^T \right)^{-1} \sum_{i=1}^{m} y_i a_i$$

- error $e = \hat{x} - x$ has zero mean and covariance

$$E = \mathbf{E}\, ee^T = \left(\sum_{i=1}^{m} a_i a_i^T \right)^{-1}$$

experiment design: choose $a_i \in \{v_1, \ldots, v_p\}$ (a set of possible test vectors) to make E 'small'

Experiment design

Let m_j denote the number of experiments for which a_i is chosen to have the value v_j, so we have

$$m_1 + \cdots + m_p = m.$$

$$E = \left(\sum_{i=1}^{m} a_i a_i^T\right)^{-1} = \left(\sum_{j=1}^{p} m_j v_j v_j^T\right)^{-1}$$

error covariance depends only on the numbers of each type of experiment chosen

$$\begin{array}{ll} \text{minimize} & E = \left(\sum_{j=1}^{p} m_j v_j v_j^T\right)^{-1} \\ \text{subject to} & m_i \geq 0, \quad m_1 + \cdots + m_p = m \\ & m_i \in \mathbf{Z}, \end{array}$$

- variables are m_k

a vector optimization problem

Experiment design

relaxing, the constraint that the m_i are integers.

assume $m \gg p$, use $\lambda_k = m_k/m$

$$E = \frac{1}{m}\left(\sum_{i=1}^{p} \lambda_i v_i v_i^T\right)^{-1}$$

$\lambda \in \mathbf{R}^p$, $\lambda \succeq 0$, $\mathbf{1}^T\lambda = 1$

each λ_i is an integer multiple of $1/m$. By ignoring this last constraint, we arrive at the problem

minimize	$E = (1/m)\left(\sum_{i=1}^{p} \lambda_i v_i v_i^T\right)^{-1}$
subject to	$\lambda \succeq 0, \quad \mathbf{1}^T\lambda = 1,$

relaxed experiment design problem

$$m_i = \mathbf{round}(m\lambda_i), \quad i = 1, \ldots, p.$$

$$\tilde{\lambda}_i = (1/m)\mathbf{round}(m\lambda_i), \quad i = 1, \ldots, p.$$

$$|\lambda_i - \tilde{\lambda}_i| \le 1/(2m)$$

for m large, we have $\lambda \approx \tilde{\lambda}$

- common scalarizations: minimize $\log\det E$, $\mathbf{tr}\, E$, $\lambda_{\max}(E)$, ...