<u>Ref.</u> Statistical methods in experimental physics, F. James.

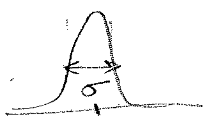
<u>03/08/208</u>. Statistics (for inverse problems)

model parameters

$$y = A_\eta(x) + e \qquad y \in \mathbb{R}^n$$

data, nuisance, noise

<u>ex.</u>(typical) $e_i \sim N(0, \sigma^2)$ iid



means $p(t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-t^2/2\sigma^2}$ for each e_i

so $\mathbb{P}[e_i \le t] = \int_{-\infty}^{t} p(t)\,dt$

$\mathbb{E} e_i = 0$, $\mathrm{Var}\, e_i = \sigma^2$

reasonable choice if one does not know better,
- because of CLT
- because of ease of computation

Seek an <u>estimator</u> $\hat{x}$ of x, function of y only
$\Rightarrow$ random, like y.

Two points of view

a) <u>frequentist</u> / classical:
- probabilities are <u>frequencies of repeatable events</u>, in the limit $n \to \infty$.
- the desired parameter x is unknown, but fixed (not random).
- $\hat{x}$ is consistent when $\hat{x} \xrightarrow{n\to\infty} x$ (i.e. $\forall \varepsilon > 0, \forall \delta > 0, \exists N : \mathbb{P}(|\hat{x}_n - x| > \varepsilon) < \delta$)
- data distribution function $p(y|x)$ given x
  likelihood

b) Bayesian

- probabilities are <u>degrees of belief</u>
  → subjective
- the desired parameter is itself random.
- prior belief about x: $p(x)$
  posterior distribution of x: Bayes's theorem

$$\underbrace{p(x|y)}_{\text{posterior}} = \frac{\underbrace{p(y|x)}_{\text{data likelihood}}\,\underbrace{p(x)}_{\text{prior}}}{\underbrace{p(y)}_{\text{normalization, indep. of } x}} \sim p(y|x)\,p(x)$$

<u>Def.</u> $\hat{x} = \arg\max_x p(y|x)$ is called

maximum likelihood estimator (MLE)

$\hat{x} = \arg\max_x p(y|x)\,p(x)$ is called

maximum a posteriori (MAP)

⚠ uniform prior does not always mean absence of prior. (x and $f(x)$ do not have the same distribution)

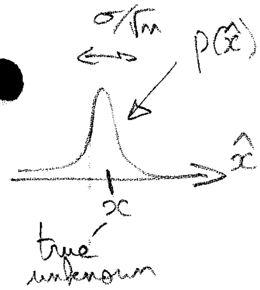
<u>ex.</u> $y_i = x + e_i$, $e_i \sim N(0,\sigma^2)$ iid

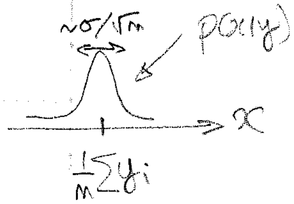
$y|x \sim N(x, \sigma^2 I)$

$$p(y|x) = \frac{1}{(2\pi\sigma^2)^{M/2}} \exp\left(-\|y-x\|^2/2\sigma^2\right) = \prod_i p(y_i|x)$$

<u>MLE:</u> $\max_x p(y|x) \Leftrightarrow \min_x \|y-x\|^2$

$$\Rightarrow \hat{x} = \frac{1}{M}\sum_{i=1}^{M} y_i \quad \text{(sample mean)}$$





$$\hat{x} \sim N\left(x, \frac{\sigma^2}{m}\right) \Rightarrow p(\hat{x}) = \frac{1}{\sqrt{2\pi\frac{\sigma^2}{m}}} \exp\left(-(\hat{x}-x)^2 / 2\left(\sigma^2/m\right)\right)$$

<u>MAP</u>: $p(x|y) \sim p(y|x)\,p(x)$

$$= \frac{1}{(2\pi\sigma^2)^{m/2}} \exp\left(-\|y-x\|^2/2\sigma^2\right) p(x)$$

$$\hat{x} = \underset{x}{\operatorname{argmax}}\; p(x|y)$$

$$= \underset{x}{\operatorname{argmin}}\; \|x-y\|^2 \underbrace{- 2\sigma^2 \log p(x)}_{\text{regularization term}}$$

(if $p(x) = Ce^{-\|x\|^2/2\alpha^2}$, then Tikhonov).

<u>Def.</u> $\hat{x}$ is unbiased if $\mathbb{E}\hat{x} = x \quad \forall m$ (sample size)
(biased otherwise, and $\mathrm{Bias}(\hat{x}, x) = \mathbb{E}\hat{x} - x$).

<u>ex.</u>
- prior regularization creates a bias
- $\mathbb{E}\,\frac{1}{m}\sum_{i=1}^{m} y_i = \frac{1}{m}\sum_{i=1}^{m} \mathbb{E}y_i = \frac{1}{m}\sum_{i=1}^{m} x = x$ → unbiased

<u>Def.</u> Mean squared error / Risk

$$MSE = \mathbb{E}_{\hat{x}} \|\hat{x} - x\|^2,$$

where $\hat{x} = f(y)$ with $y \sim p(y|x)$

<u>Prop.</u> $MSE = \mathrm{Var}\,\hat{x} + (\mathrm{Bias}(\hat{x}, x))^2$ when $x, \hat{x} \in \mathbb{R}$

<u>Pf.</u> $\mathbb{E}_{\hat{x}}(\hat{x}-x)^2 = \mathbb{E}_{\hat{x}}(\hat{x} - \mathbb{E}\hat{x} + \mathbb{E}\hat{x} - x)^2$

$$= \mathbb{E}_{\hat{x}}(\hat{x} - \mathbb{E}\hat{x})^2 + 2\,\mathbb{E}_{\hat{x}}\left((\hat{x} - \mathbb{E}\hat{x})(\mathbb{E}\hat{x} - x)\right) + (\mathbb{E}\hat{x} - x)^2$$

$\square$

ex. $y = \overset{\text{small}}{\varepsilon} x + e$, $e \sim N(0, \overset{\text{small}}{\sigma^2})$

$y|x \sim N(\varepsilon x, \sigma^2)$

MAP: $\min_x \frac{(y-\varepsilon x)^2}{\sigma^2} + \underbrace{\frac{x^2}{\alpha^2}} \longrightarrow \hat{x}_\alpha$

prior on size of x: $p(x) \sim \exp(-x^2/2\alpha^2)$

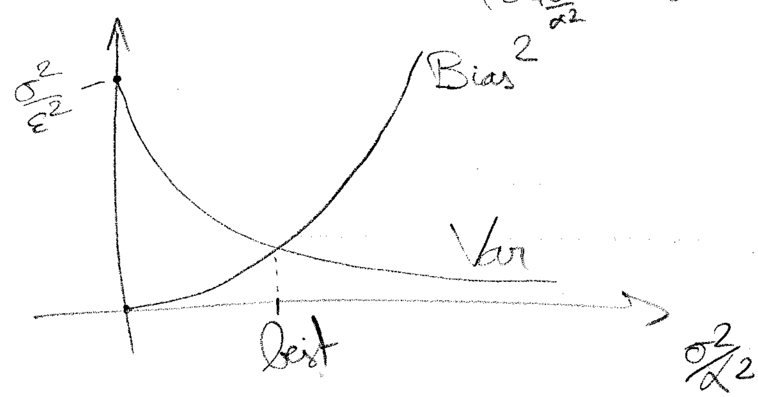
$$y^2 - 2\varepsilon x y + \left(\varepsilon^2 + \frac{\sigma^2}{\alpha^2}\right) x^2$$

$$-2\varepsilon y + 2\left(\varepsilon^2 + \frac{\sigma^2}{\alpha^2}\right)\hat{x} = 0$$

$$\Rightarrow \quad \hat{x} = \frac{\varepsilon}{\varepsilon^2 + \frac{\sigma^2}{\alpha^2}}\, y$$

$$\hat{x} \sim N\Bigg(\underbrace{\frac{\varepsilon^2}{\varepsilon^2 + \sigma^2/\alpha^2}\, x}_{\mathbb{E}\hat{x}} \; , \; \underbrace{\frac{\varepsilon^2}{\left(\varepsilon^2 + \frac{\sigma^2}{\alpha^2}\right)^2}\, \sigma^2}_{\mathrm{Var}\,\hat{x}} \Bigg)$$

$$\mathrm{Bias}(\hat{x}, x)^2 = \left(\frac{\varepsilon^2}{\varepsilon^2 + \frac{\sigma^2}{\alpha^2}}\, x - x \right)^2 = \frac{\alpha^2}{\left(\varepsilon^2 + \frac{\sigma^2}{\alpha^2}\right)^2}\, x^2$$



~~Thm. Cramér-Rao lower bound. Assume $\hat{x}$ is unbiased, then~~

$$\cancel{\mathrm{Var}\,\hat{x} \geq 1/I_y(x)}$$

$$\cancel{I_y(x) = \mathbb{E}\left[\left(\frac{\partial l(y,x)}{\partial x} \right)^2 \right]}$$ ~~(Fisher information "of x contained in y")~~

$$\cancel{l(y,x) = \log p(y|x)}$$ ~~(log likelihood)~~