Bayesian Approach to MLE:

$$p(x|\mathcal{X}) = p(x|x_1,\dots,x_n) = \int p(x,\theta|x_1,\dots,x_n)\,d\theta$$

$$= \int p(x|\theta,x_1,\dots,x_n)\,p(\theta|x_1,\dots,x_n)\,d\theta$$

$$= \int p(x|\theta)\,p(\theta|x_1,\dots,x_n)\,d\theta$$

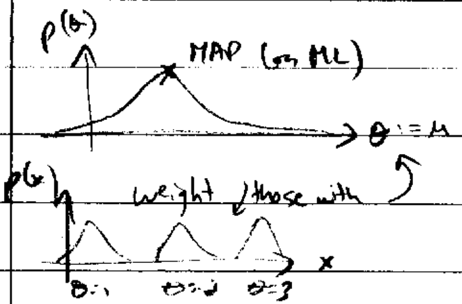
Bayes

$$= \int p(x|\theta)\,\frac{p(x_1,\dots,x_n|\theta)\,p(\theta)}{p(x_1,\dots,x_n)}\,d\theta$$

i.i.d.

$$\propto \int p(x|\theta)\,\prod p(x_i|\theta)\,p(\theta)\,d\theta$$

use all models.



rarely solvable

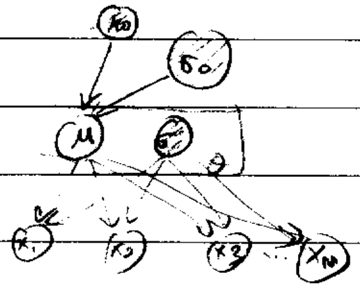
So approx. MAP, ML, or Expected θ

logarithmic monotonic

Gaussian Mean in 1 dim: $N(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\,e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$

canonical for scalar vars

$\theta = \mu$ only for now



$$p(\mu) = \frac{1}{\sqrt{2\pi\tau^2}}\,e^{-\frac{1}{2\tau^2}(\mu-\mu_0)^2}$$

$$p(x|\mu) = \frac{1}{\sqrt{2\pi\sigma^2}}\,e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$$

$$p(\mathcal{X},\mu) = \frac{1}{(2\pi\sigma^2)^{N/2}}\,e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\mu)^2}\;\frac{1}{\sqrt{(2\pi\tau^2)}}\,e^{-\frac{1}{2\tau^2}(\mu-\mu_0)^2}$$

do log - MAP

do log - ML

pos. semi. def.

$$\text{Gaussian Multi-D:} \quad N(\vec{x} \mid \mu, \Sigma) = \frac{1}{(2\pi)^{D/2}\sqrt{|\Sigma|}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}$$

$$\log P(X \mid \theta) = -\sum_i \log (2\pi)^{D/2} - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(x_i-\mu)^T \Sigma^{-1}(x_i-\mu)$$

$$\frac{\partial}{\partial \mu} \quad \sum_i \Sigma^{-1}(x_i-\mu) = 0$$

$$\underline{\underline{\mu = \frac{1}{N}\sum_i \vec{x}_i}}$$

$$\frac{\partial}{\partial \Sigma^{-1}} \quad \frac{\partial}{\partial \Sigma^{-1}}\left(\frac{N}{2}\log|\Sigma^{-1}| - \frac{1}{2}\operatorname{tr}\left(\sum_i (x_i-\mu)(x_i-\mu)^T \Sigma^{-1}\right)\right)$$

sloppy
Symm. +ve def.
Magnus & Neudecker
Rustagi

$$\frac{N}{2}(\Sigma^{-1})^{-T} \quad -\frac{1}{2}\sum_i (x_i-\mu)(x_i-\mu)^T = 0$$

$$N/2\, \Sigma - \frac{1}{2}\sum_i (x_i-\mu)(x_i-\mu)^T = 0$$

$$\underline{\underline{\Sigma = \frac{1}{N}\sum_i (x_i-\mu)(x_i-\mu)^T}}$$

$$\frac{\partial}{\partial A}\log|A| = (A^{-1})^T$$

$$\frac{\partial}{\partial A}\operatorname{tr}(BA) = B^T$$

if square matrix

$$\operatorname{tr}(AB) = \operatorname{tr}(BA)$$

$$\operatorname{tr}(BAB^{-1}) = \operatorname{Tr}(A)$$

$$\operatorname{tr}(A) = \operatorname{tr}(A^T)$$

Bernoulli: coin flips:

$$p(x \mid \theta) = \theta^x (1-\theta)^{1-x}$$

$$\sum_{i=1}^{n} x_i \log\theta + (1-x_i)\log(1-\theta)$$

$$\frac{1}{\theta}\sum x_i - \sum_i \frac{1-x_i}{1-\theta} = 0$$

Multinomial: discrete case, equiv to tables

binary: $\begin{array}{c} 0 \\ 1 \end{array}\begin{bmatrix} 0.4 \\ 0.1 \end{bmatrix}$ multiple discrete $x = \begin{array}{c} A \\ B \\ C \\ D \end{array}\begin{bmatrix} .4 \\ .2 \\ .2 \\ .1 \end{bmatrix}$ θ_1 ... θ_3

$$A = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad B = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \ldots$$

$$p(x \mid \theta) = \prod_{j=1}^{d} \theta_j^{x^j} \qquad \sum_i \theta_i = 1$$

$$\log p(\mathcal{X} \mid \theta) = \sum_{i=1}^{n} \log \prod_{j=1}^{d} \theta_j^{x_i^j}$$

$$\sum_{i=1}^{N} \log \prod_{j=1}^{d} \theta_j^{x_i^j} + \lambda\left(1-\sum_j \theta_j\right)$$

θ_k

$$\frac{\partial}{\partial\theta_k}\left(\sum_{i=1}^{N}\sum_{j=1}^{d} x_i^j \log\theta_j + \lambda\left(1-\sum_j\theta_j\right)\right) = 0$$

$$\sum_{i=1}^{N} x_i^k \frac{1}{\theta_k} - \lambda = 0$$

$$\frac{\sum_{i=1}^{N} x_i^k}{\lambda} = \theta_k$$

$$\sum_j \theta_j = 1 \implies \sum_j \frac{\sum_{i=1}^{N} x_i^j}{\lambda} = 1$$

$$\lambda = \frac{1}{\sum_j\sum_i x_i^j}$$

$$\theta_k = \frac{\sum_{i=1}^{N} x_i^k}{\underbrace{\sum_{i=1}^{N}\sum_{j=1}^{d} x_i^j}_{N}} = \frac{\sum_{i=1}^{N} x_i^k}{N} = \text{ratio}$$

Bayesian 1d Gauss: $p(x|\mathcal{X}) = \dfrac{\int p(x|\theta)\prod_i p(x_i|\theta)\,p(\theta)\,d\theta}{\prod_i p(x_i)}$

usually prior on μ has diff. covar. than σ^2

$$p(x|\mathcal{X}) \propto \int p(x|\theta)\prod_i p(x_i|\theta)\,p(\theta)\,d\theta$$

$$\propto \int \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} \times \left(\prod_{i=1}^{N}\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}}\right) \times \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(\mu-\mu_0)^2}{\sigma^2}}\right) d\mu$$

$$\propto \int e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2} - \frac{1}{2}\sum_i\frac{(x_i-\mu)^2}{\sigma^2} - \frac{1}{2}\frac{(\mu-\mu_0)^2}{\sigma^2}}\,d\mu$$

$$\propto \int e^{-\frac{1}{2\sigma^2}\left(\mu^2 - 2x\mu + x^2 + N\mu^2 - 2\sum_i x_i\mu + \sum_i x_i^2 + \mu^2 - 2\mu_0\mu + \mu_0^2\right)}\,d\mu$$

$$\propto \int e^{-\frac{1}{2\sigma^2}\left[(N+2)\mu^2 - 2\mu\left[x + \sum_i x_i + \mu_0\right] + x^2\right]}\,d\mu$$

$$p(x|X) \propto \int e^{-\frac{N+2}{2\sigma^2}\left[\mu^2 - 2\mu\left[\frac{x+\sum_i x_i+\mu_0}{N+2}\right] + \frac{x^2}{N+2}\right]} d\mu$$

$$\propto \int e^{-\frac{N+2}{2\sigma^2}\left[\mu^2 - 2\mu\left[\frac{x+\sum_i x_i+\mu_0}{N+2}\right] + \left[\frac{x+\sum_i x_i+\mu_0}{N+2}\right]^2 - [\ ]^2 + \frac{x^2}{N+2}\right]} d\mu$$

$$\propto e^{-\frac{N+2}{2\sigma^2}\left(-\left[\frac{x+\sum_i x_i+\mu_0}{N+2}\right]^2 + \frac{x^2}{N+2}\right)}$$

$$\propto e^{-\frac{1}{2\sigma^2}\left(-\frac{[x+\sum_i x_i+\mu_0]^2}{N+2} + x^2\right)}$$

$$\propto e^{-1/2\sigma^2\left(\frac{-x^2 - (\sum_i x_i)^2 - \mu_0^2}{N+2} + \frac{2x\sum_i x_i + 2x\mu_0 + 2\sum_i x_i \mu_0}{N+2} + x^2\right)}$$

$$\propto e^{-\frac{1}{2\sigma^2}\left(\frac{(N+2)x^2 - x^2}{N+2} + \frac{(\sum_i x_i)^2 - \mu_0^2}{N+2} - \frac{2x(\sum_i x_i + \mu_0)}{N+2}\right)}$$

$$\propto e^{-1/2\sigma^2\left(\frac{N+1}{N+2}x^2 - 2x\frac{\sum_i x_i + \mu_0}{N+2}\right)}$$

$$\propto e^{-1/2\sigma^2 \frac{N+1}{N+2}\left(x^2 - 2x\left(\frac{\sum_i x_i + \mu_0}{N+1}\right)\right)}$$

$$\propto e^{-1/2\sigma^2 \frac{N+1}{N+2}\left(x^2 - 2x\left(\frac{\sum_i x_i + \mu_0}{N+1}\right) + \left(\frac{\sum_i x_i + \mu_0}{N+1}\right)^2\right)}$$

$$\propto e^{-1/2\tilde{\sigma}^2 (x-\tilde{\mu})^2}$$

$$= \frac{1}{\sqrt{2\pi\tilde{\sigma}^2}} e^{-1/2\tilde{\sigma}^2 (x-\hat{\mu})^2}$$

$$\frac{1}{2\tilde{\sigma}^2} = \frac{1}{2\sigma^2}\frac{N+1}{N+2}$$

$$\hat{\sigma}^2 = \frac{N+2}{N+1}\sigma^2$$

$$\hat{\mu} = \frac{\sum_i x_i + \mu_0}{N+1}$$