$X = (x_1, x_2, \dots x_N)$

$P(X) = P(x_1, x_2 \dots x_N)$

$= q(x_1)\, q(x_2) \cdots q(x_N)$

$= \prod_{i=1}^{N} q(x_i)$

$\log P(X) = \sum_{i=1}^{N} \log q(x_i)$

ML Estimate:

$x_1 \dots x_N$

ML estimate of μ, σ $\Rightarrow$ $P(X) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$

Take log, take $\frac{d}{dx} = 0$

$\mu_{ML} = \frac{1}{N} \sum_{i=1}^{N} x_i$

$q_1(x_i) \equiv$ gaussian (μ, σ) , $q_2(x_i) \equiv unf(a,b)$

$u(x|a,b)$ $\Rightarrow P(X) = \prod_{i=1}^{N} \begin{cases} \frac{1}{b-a} & \text{, if } a \le x < b \\ 0 & \text{otherwise} \end{cases}$

$a \le \min(x_i)$

$b > \max(x_i)$

Example 1: Exponential distribution

* $P_\lambda(x) = \begin{cases} \lambda e^{-\lambda x} & , x \ge 0 \\ 0, & x < 0 \end{cases}$

$\rightarrow$ log likelihood $L_\lambda(X) = \sum_i \log P_\lambda(x_i)$

$\rightarrow$ ~~Differentiating~~

Statistics is a function of the sample

$T(X) = T(x_1, \dots, x_N)$

For some distributions, computing a few statistics is sufficient for MLE estimate.

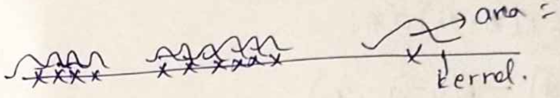
$\rightarrow$ Ex- of sufficient statistics:

* sample mean and variance for gaussian distribution.
* sample mean for exponential distribution.
* max and min for uniform distribution.

Non-Parametric density fun:

if there is N points.

$\rightarrow$ area $= 1/N$ each kernel area $= \frac{1}{N}$



$P_X(x) = \frac{1}{N} \sum_{i=1}^{N} K(x - x_i)$

$K(x) \ge 0, \quad \forall x$

$\int_{-\infty}^{\infty} K(x)\,dx = 1$

Assume gaussian kernel:

$K(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$

$\rightarrow$ Select σ based on a rule of thumb that takes onto account the given data.

$\rightarrow$ MLE for parametric distribution.

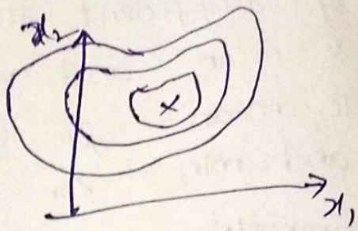
$\rightarrow$ Bayesian est.

Take into account a prior be liet over parameters

Assume: Prior belief $P_0(\theta)$ Instead of makingly $\int \theta L_\theta(X)$ o

Maximize $\frac{\int \theta P_0(\theta) L_\theta(x)\, d\theta}{\int P_\theta(\theta) L_\theta(x) d\theta}$

Multivariate distribution:



$P(x_1, x_2) \geq 0$

$\iint P(x_1, x_2)\, dx_1\, dx_2 = 1$

$P(x_2) = \int_{-\infty}^{\infty} P(x_1, x_2)\, dx_1$

Marginal: $P(x_2) = \int_{-\infty}^{\infty} P(x_1, x_2)\, dx_1$

conditional: $P(x_2 | x_1 = a) = \dfrac{P(x_1 = a, x_2)}{P(x_1 = a)}$

Type and coding of variables can be different:

→ Integers can be used to code:

- Nominal / categorical (species, postal codes)
- ordinal
- Tune numerical.