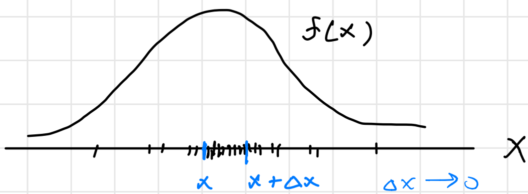
Lecture 10

- Gaussian Process:

- Review of Probability

  - Density $X \sim f(x)$
    $\|$
    population of equally likely possibilities
    $\{x_1, x_2, \ldots, x_i, \ldots, x_N \approx \infty\}$
    e.g. heights of 300 million people.

  - Scatterplot



$N(x) = \#$ of people in $[x, x+\Delta x)$

$$f(x) = \frac{N(x)/N}{\Delta x} \xrightarrow{\text{random sampling}} = \frac{\Pr(X \in [x, x+\Delta x))}{\Delta x}$$

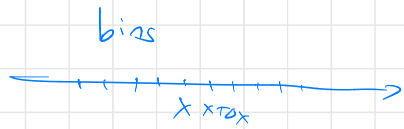
e.g.

$$\text{Density}(LA) = \frac{\#\text{ of people in LA} / \text{total }\#\text{ in U.S.}}{\text{size of LA}}$$

$$\frac{N(x)}{N} = f(x)\,\Delta x = \Pr(X \in [x, x+\Delta x))$$

- If $X \sim f(x)$, then

$$\mathbb{E}(h(x)) = \frac{1}{N}\sum_{i=1}^{N} h(x_i)$$



$$= \frac{1}{N}\sum_{\text{bins}(x,\, x+\Delta x)} h(x)\, N(x)$$

$$\frac{N(x)}{N} = f(x)\,\Delta x$$

$$= \sum_{\text{bins}(x,\, x+\Delta x)} h(x)\, f(x)\, \Delta x \xrightarrow{\Delta x \to 0} \int h(x)\, f(x)\, dx$$

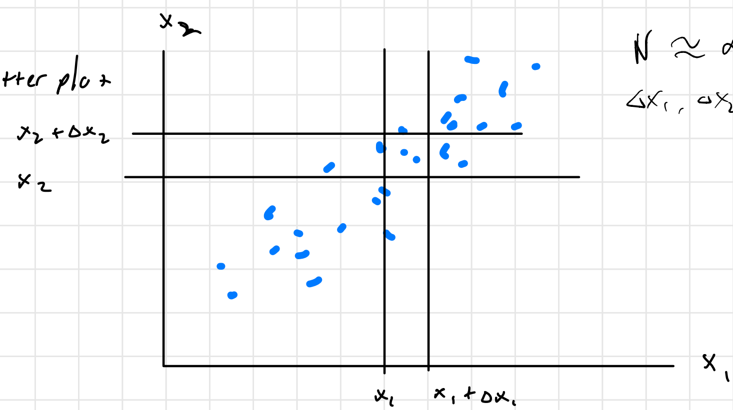
- $\text{Var}(X) = \mathbb{E}\left(\left(x - \mathbb{E}(x)\right)^2\right)$

$\text{Var}(X) \to \sigma^2$, $\quad \mathbb{E}(x) \to \mu$

- 2D $\quad X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

$$f(x) = \frac{P_r\left(x_1 \in (x_1,\, x_1+\Delta x_1),\; x_2 \in (x_2,\, x_2+\Delta x_2)\right)}{\Delta x_1\, \Delta x_2}$$

Scatter plot



$N \approx \infty$

$\Delta x_1, \Delta x_2 \to 0$

- $N(x) = N(x_1, x_2) = \#$ of points in $(x_1, x_1 + \Delta x_1) \times (x_2, x_2 + \Delta x_2)$

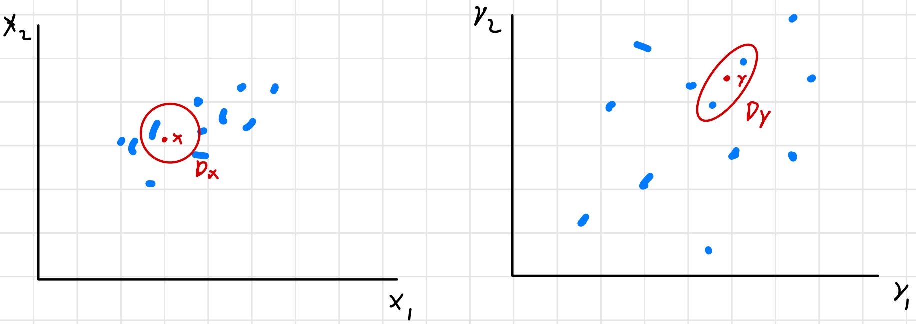
$$f(x) = f(x_1, x_2) = \frac{N(x_1, x_2)/N}{\Delta x_1 \, \Delta x_2} \qquad \frac{N(x_1, x_2)}{N} = f(x_1, x_2)\,\Delta x_1\,\Delta x_2$$

- $$\mathbb{E}(h(x)) = \int h(x) f(x)\, dx = \iint h(x_1, x_2) f(x_1, x_2)\, dx_1\, dx_2$$

- Transformation : $Y = AX$

change of variable

$$X \sim f_X(x) \qquad , \qquad Y \sim f_Y(y)$$



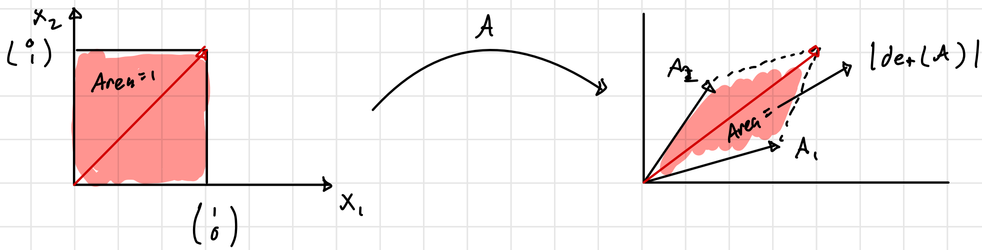
$$N(X) = \# \text{ points in } D_x \quad = \quad N(Y) = \# \text{ points in } D_Y$$

$$N(X)/N = f_X(x)\,|D_x| \qquad N(Y)/N = f_Y(y)\,|D_Y|$$

So $\quad f_X(x)\,|D_x| = f_Y(y)\,|D_Y|$

A^{-1} $\quad A$ $\quad A^{-1}$ $\quad A$

So $$\frac{|D_y|}{|D_x|} = |\det(A)|$$

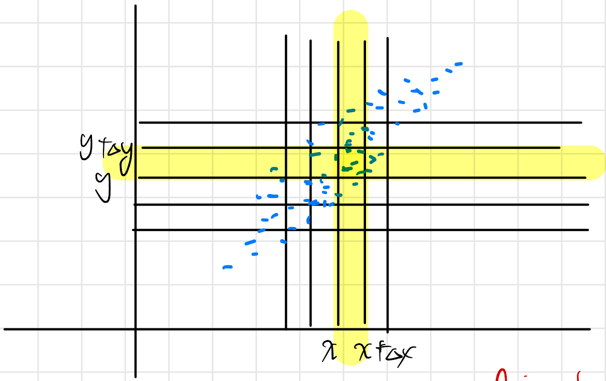


$$y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = Ax = (A_1, A_2)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Make LA the size of Texas $\longrightarrow$ density is smaller

Make LA the size of UCLA $\longrightarrow$ density is larger

- change notation $(X, Y) \sim f(x, y)$ joint



$N \approx \infty$ points

$\Delta x, \Delta y \to 0$

Operation 1 : Marginalization

- $f(x) = \dfrac{N(x)/N}{\Delta x} = \dfrac{\sum_y N(x,y)/N}{\Delta x} = \dfrac{\sum_y f(x,y)\,\Delta x \Delta y}{\Delta x}$

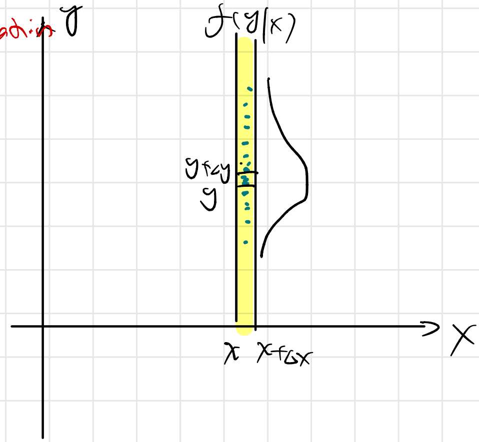
$\longrightarrow \int f(x,y)\,dy$

$N(x) = \sum_y N(x,y)$

- Similarly $f(y) = \int f(x,y)\,dx$

Operation 2 : Conditioning / normalization

- Conditional density $f(y \mid x)$



$$f(y) = \frac{N(y)/N}{\Delta y}$$

- $$f(y \mid x) = \frac{N(x,y)/N(x)}{\Delta y}$$

$$= \frac{\frac{N(x,y)}{N} \Big/ \frac{N(x)}{N}}{\Delta y} = \frac{f(x,y)\,\Delta x\,\Delta y \,/\, f(x)\,\Delta x}{\Delta y}$$

$$= \frac{f(x,y)}{f(x)}$$

Operation 3 : factorization

- Factorization :

$$\begin{aligned} f(x,y) &= f(x)\, f(y \mid x) \\ &= f(y)\, f(x \mid y) \end{aligned}$$

- Independence : $X \perp Y$

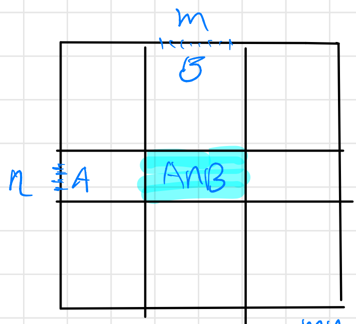
$$\begin{aligned} f(y \mid x) &= f(y) \\ f(x \mid y) &= f(x) \\ f(x,y) &= f(x)\, f(y) \end{aligned}$$

- $X \sim f(x) \sim \{x_1, \dots, x_n\}$
  $Y \sim f(y) \sim \{y_1, \dots, y_m\}$

  $X \perp Y$

  $(X, Y) \sim f(x) f(y) \sim N \times M$ equally likely pairs

$$\begin{array}{c|c}
 & y_1, \dots, y_j, \dots y_M \\
\hline
x_1 & \\
\vdots & \\
x_i & (x_i, y_j) \\
\vdots & \text{population} \\
x_N &
\end{array}$$



$X \perp Y$

$P(X \in A \ \& \ Y \in B) = \dfrac{mn}{MN} = \dfrac{m}{M} \dfrac{n}{N}$

$= P(X \in A) P(Y \in B)$

- Multivariate Statistics:

  $X_{m \times n} \sim f(x)$

- $\mathbb{E}(X) = \int x f(x) dx = \begin{pmatrix} & & \\ & E(x_{ij}) & \\ & & \end{pmatrix}_{m \times n}$

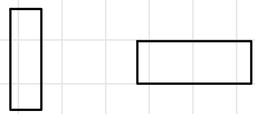
  A, B constant

- $\mathbb{E}(AX) = \int A x f(x) dx = A \int x f(x) dx = A \, \mathbb{E}(x)$
- $\mathbb{E}(XB) = \mathbb{E}(X) B$

Vector: $X_{d\times 1} \sim f(x)$

$$\mu_{d\times 1} = \mathbb{E}(x)$$

$$Var(x) = \mathbb{E}\left((x-\mu)(x-\mu)^T\right)_{d\times d}$$



$d\times 1$ $\quad$ $1\times d$

$$= \begin{pmatrix} \ddots & & \\ & Var(x_i) & \\ Cov(x_i, x_j) & & \ddots \end{pmatrix}_{d\times d}$$

$Var(x_i) \rightarrow E((x_i-\mu_i)^2)$

$Cov(x_i,x_j) \downarrow E((x_i-\mu_i)(x_j-\mu_j))$

$$Var(Ax) = E((Ax - A\mu)(A - A\mu)^T)$$

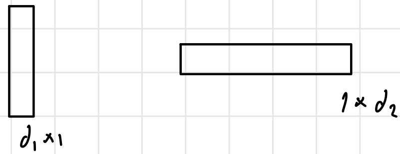
$$= E(A(x-\mu)(x-\mu)^T A^T)$$

$$= A\, var(x) A^T$$

- Note $Var(a^T x) = a^T Var(x)\, a \geq 0$

Thus $Var(x) \geq 0$

- $\mathrm{Cov}(X, Y) = E\left((X - E(X))(Y - E(Y))^T \right)$



- Gaussian Density:

$$f(x) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2} (x-\mu)^T \Sigma^{-1} (x-\mu) \right)$$

$$X \sim N(\mu, \Sigma)$$

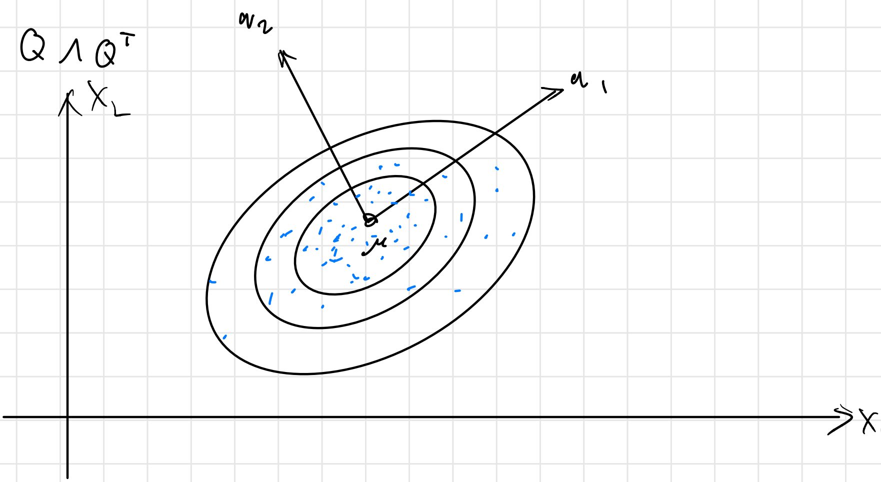
$$E(X) = \mu \quad , \quad Var(X) = \Sigma$$

- Consider The Contours:

$$(x-\mu)^T \Sigma^{-1} (x-\mu) \longrightarrow y^T \Lambda^{-1} y = \sum_{j=1}^{d} \frac{y_j^2}{\lambda_j}$$

$$\begin{array}{ll} x-\mu & \xrightarrow{\Sigma} \\ Q \uparrow & \\ y & \xrightarrow{\Lambda} \end{array}$$

$$\Sigma = Q \Lambda Q^T$$



- Note: $X = \mu + QY$

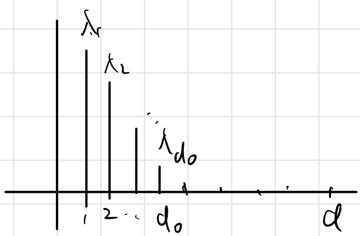
$$= \mu + (q_1, \dots, q_i, \dots, q_d)\begin{pmatrix} Y_1 \\ \vdots \\ Y_i \\ \vdots \\ Y_d \end{pmatrix}$$

- $$f(Y) = \frac{1}{(2\pi)^{d/2}\prod_{j=1}^{d} \lambda_j^{1/2}} e^{-\frac{1}{2}\sum_{j=1}^{d}\frac{Y_j^2}{\lambda_j}}$$

$$= \prod_{j=1}^{d} \frac{1}{\sqrt{2\pi\lambda_j}} e^{-\frac{Y_j^2}{2\lambda_j}}$$

so Y_j are independent with $Y_j \overset{ind}{\sim} N(0, \lambda_j)$

- plot the λ_j:



if λ_j close to 0 then Y_j close to 0

dim reduction $\quad \mu + \sum_{j=1}^{d_0} q_j Y_j \quad$ principal component analysis

- Note : $Y \sim N(0, \Lambda)$

$$\left.\begin{aligned} \mathbb{E}(Y) &= 0 \\ Var(Y) &= \Lambda \end{aligned}\right\} \text{ uncorrelated elements}$$

$$X = \mu + QY \sim N(\mu, \Sigma)$$

$$E(X) = \mu + QE(Y) = \mu$$

$$Var(X) = Var(QY) = Q\, Var(Y)\, Q^T$$
$$= Q \Lambda Q^T = \Sigma$$

Change of variable

$$X \sim N(\mu, \Sigma)$$

$$Y = \underset{d\times d}{A} X \sim N(A\mu, A\Sigma A^T)$$

$$E(Y) = A\mu$$
$$Var(Y) = A \Sigma A^T$$

$$f_X(x)\, |D_x| = f_Y(y)\, |D_y|$$

$$f_X(x) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

$$f_Y(y) = f_X(x) \frac{1}{|D_y|/|D_x|} = f_X(A^{-1}y) \frac{1}{|A|}$$

$$= \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}} |A|} \exp\left(-\frac{1}{2}(A^{-1}y-\mu)^T \Sigma^{-1} (A^{-1}y-\mu)\right)$$

$$= \frac{1}{(2\pi)^{\frac{d}{2}} |\underbrace{A\Sigma A^T}_{Var}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(Y-\underbrace{A\mu}_{E})^T (\underbrace{A\Sigma A^T}_{Var})^{-1} (Y-\underbrace{A\mu}_{E})\right)$$

- Gaussian Process:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim \mathcal{N}\left(0, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\right)$$

$$[x_2 \mid x_1] \sim \mathcal{N}(\Sigma_{21}\Sigma_{11}^{-1}x_1, \ \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$$

Proof: Let $x_2 = A x_1 + \varepsilon$, $\varepsilon \perp x_1$

$$\varepsilon = x_2 - A x_1$$

$$\begin{pmatrix} x_1 \\ \varepsilon \end{pmatrix} = \begin{pmatrix} I & 0 \\ -A & I \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

So $\mathrm{Var}\begin{pmatrix} x_1 \\ \varepsilon \end{pmatrix} = \begin{pmatrix} I & 0 \\ -A & I \end{pmatrix}\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\begin{pmatrix} I & -A^T \\ 0 & I \end{pmatrix}$

$$= \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ -A\Sigma_{11} + \Sigma_{21} & -A\Sigma_{12} + \Sigma_{22} \end{pmatrix}\begin{pmatrix} I & -A^T \\ 0 & I \end{pmatrix} \to 0$$

$$= \begin{pmatrix} \Sigma_{11} & \cancel{-\Sigma_{11}A^T + \Sigma_{12}} \\ \cancel{-A\Sigma_{11} + \Sigma_{21}} & \Sigma_{22} - A\Sigma_{12} \end{pmatrix}$$

$$= \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$$

We want $-A\Sigma_{11} + \Sigma_{21} = 0$

$$\Sigma_{21} = -A\Sigma_{11}$$

$$A = \Sigma_{21}\Sigma_{11}^{-1}$$

Thus

$$\begin{pmatrix} X_1 \\ \varepsilon \end{pmatrix} \sim N\left(0, \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \end{pmatrix}\right)$$

So that

- $X_1 \sim N(0, \Sigma_{11})$

  $X_2 = AX_1 + \varepsilon, \quad \varepsilon \perp X_1, \quad \varepsilon \sim N(0, \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$

  Given x_1 (i.e. fix x_1), ε has the same distn

- $[X_2 | x_1] \sim N(\Sigma_{21}\Sigma_{11}^{-1}x_1, \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$