Lecture 26 - Probability Overview

- So far in class
  - $\Rightarrow$ assumed we could sense $q, \dot{q} \in \mathbb{R}^n$
  - $\Rightarrow$ assumed that sensing was noise free

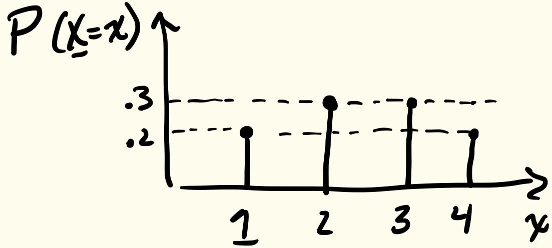
How do we estimate the state $q, \dot{q}$ given partial/noisy measurements

$$y = h(q, \dot{q}) + \text{"noise"} \qquad y \in \mathbb{R}^p \qquad p < 2n$$

- Today: Probability Basics
  - Probability mass/density
  - Gaussian Distributions (Univariate/Multivariate)

- Next lecture: Using these fundamentals to determine maximum likelihood estimates

Simple Introduction: 1D discrete probability



- Let $\underline{X}$ a random variable (1D)

- We denote $P(\cdot)$ as the probability of a random event. e.g.
    - $P(\underline{X}=2)=.2$
    - $P(\underline{X}\leq 2)=.3$

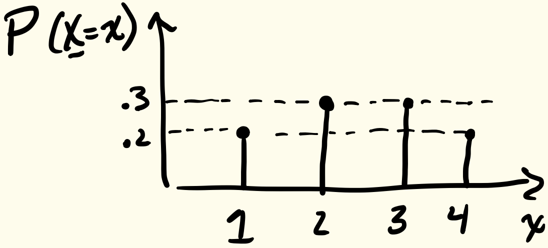
- The probability of $\underline{X}$ taking a specific value is denoted by the probability mass function (PMF)

$$P_{\underline{X}}(x)=P(\underline{X}=x)$$

- $\sum_{x} P_{\underline{X}}(x)=1$, $P_{\underline{X}}(x)\geq 0 \ \forall x$

Simple Introduction: 1D discrete probability

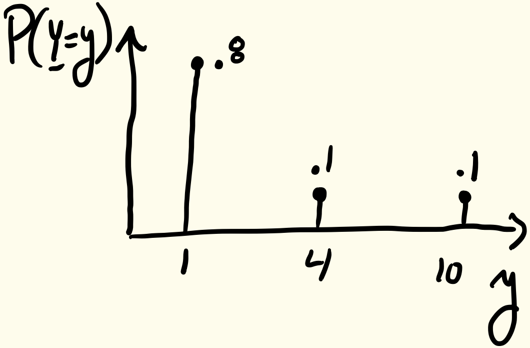
- Denote the Expected Value of a random variable as



$$E[\underline{X}] = \text{"average value of } \underline{X}\text{"}$$
$$= \sum_x x \, p_{\underline{X}}(x)$$

Example:

$$E[\underline{X}] = 2.5 \qquad E[\underline{Y}] = .8 + .4 + 1 = 2.2$$



Denote the Joint PMF

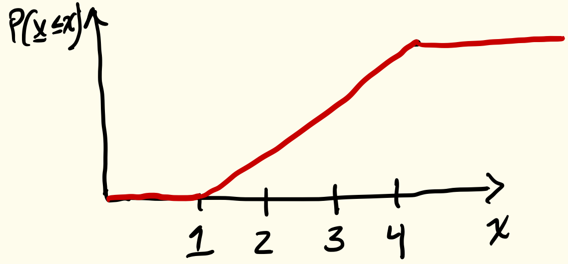
$$p_{\underline{X}\underline{Y}}(x,y) = P(\underline{X}=x \text{ and } \underline{Y}=y)$$

If $\underline{X}$ and $\underline{Y}$ independent

$$p_{\underline{X}\underline{Y}}(x,y) = p_{\underline{X}}(x)\, p_{\underline{Y}}(y)$$

Continuous Random Variables:

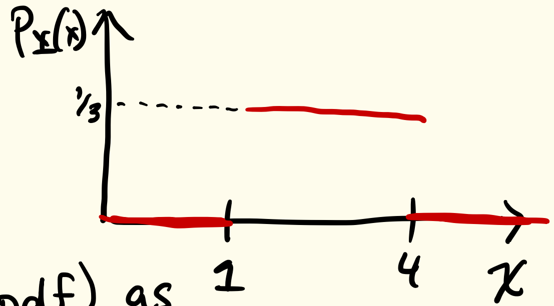
Consider $\underline{X}$ uniformly distributed on $[1, 4]$



- We call the function

$$F_{\underline{X}}(x) = P(\underline{X} \leq x)$$

The cumulative distribution function.



- Define the probability density function (pdf) as

$$p_{\underline{X}}(x) = \frac{\partial F_{\underline{X}}}{\partial x} \Rightarrow P(x \leq \underline{X} \leq x + \Delta x) = p_{\underline{X}}(x) \cdot \Delta x$$

Let $\underline{X}$ a random variable (1D):

The PDF satisfies:

(1) $$P(x_1 \le \underline{X} \le x_2) = \int_{x_1}^{x_2} p_{\underline{X}}(x)\,dx$$ (Probability of $X \in [x_1, x_2]$)

(2) $$\int_{-\infty}^{\infty} p_{\underline{X}}(x) = 1$$

Expected Value of a funtion of $\underline{X}$:

$$E[f(\underline{X})] = \int_{-\infty}^{\infty} f(x)\, p_{\underline{X}}(x)\, dx$$

Examples

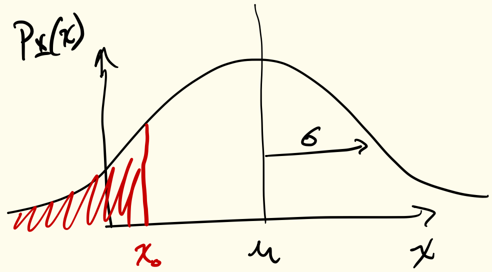
mean: $\mu = E[\underline{X}]$

Standard deviation

Variance: $\sigma^2 = E[(\underline{X} - \mu)^2]$

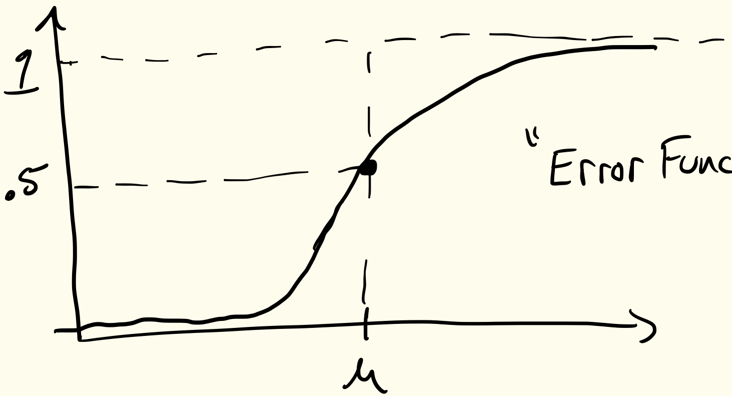
Gaussians in 1 D:

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



$F_{\underline{x}}(x_0) =$ "Area in red"

Draw CDF shape for 1 D Gaussian



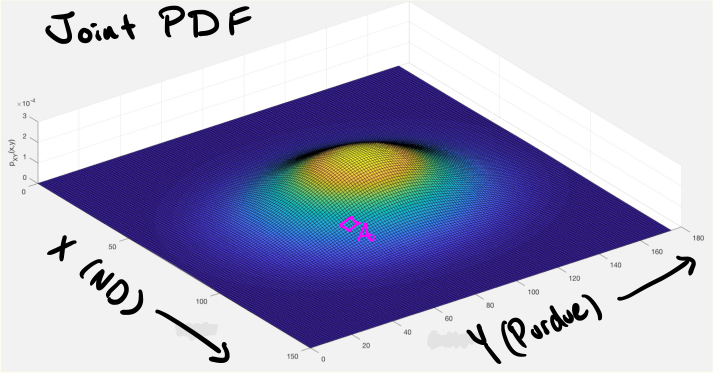
"Error Function"

Multivariate Gaussians:

Example: Pick 2 people at random
1 @ ND and 1 @ Purdue

$\underline{X}$ = weight of Person @ ND $\sim \mathcal{N}(\mu_x, \sigma_x^2)$

$\underline{Y}$ = weight of Person @ Purdue $\sim \mathcal{N}(\mu_y, \sigma_y^2)$



Joint Probability density:

$$\boxed{p_{\underline{X}\underline{Y}}(x,y)} \qquad P((x,y) \in A) = \iint_A p_{\underline{X}\underline{Y}}(x,y)\, dA$$

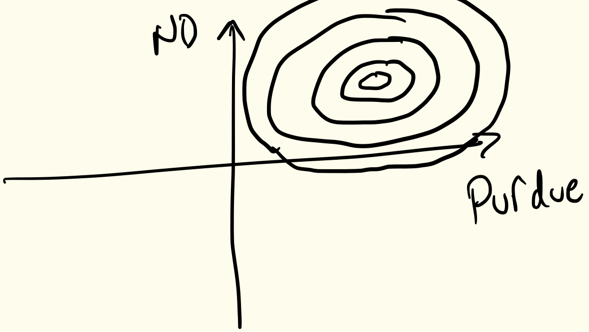
In this case:

Marginal Densities

$$p_{\underline{X}\underline{Y}}(x,y) = \boxed{p_{\underline{X}}(x)\; p_{\underline{Y}}(y)}$$

- The probability density factors since these two variables are independent

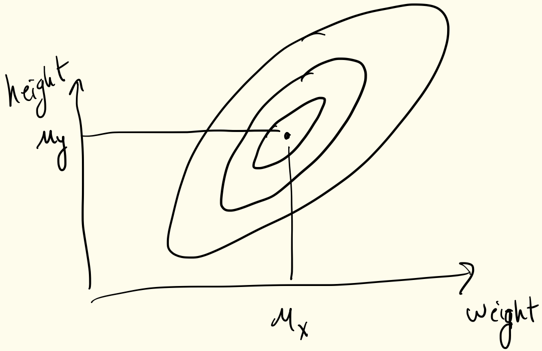
Contours of $p_{\underline{X}\underline{Y}}$

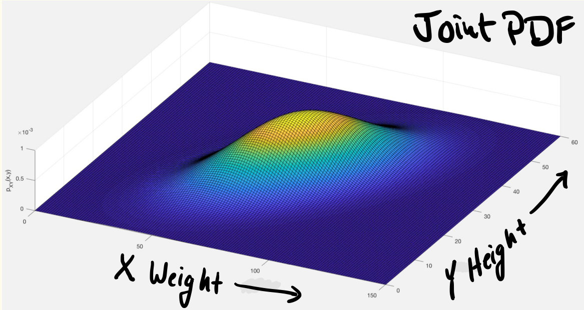


Multi-Dim gaussians - Non independent

Example: Pick a random person @ ND

X = their weight Y = their height





- Let $\underline{Z} = \begin{bmatrix} X \\ Y \end{bmatrix}$

- We denote the co-variance of $\underline{Z}$ by the 2x2 matrix

$$\Sigma = E\left[(\underline{Z} - E[\underline{Z}])(\underline{Z} - E[\underline{Z}])^T\right] = \begin{bmatrix} \sigma_x^2 & \sigma_x \sigma_y \rho \\ \sigma_x \sigma_y \rho & \sigma_y^2 \end{bmatrix}$$

$-1 \le \rho \le 1$ Correlation coefficient

If $\underline{z} \sim \mathcal{N}(\mu, \Sigma)$:

$z \in \mathbb{R}^n$ $\quad p_{\underline{z}}(z) = \dfrac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{1/2}} e^{-\frac{1}{2}(z-\mu)^T \Sigma^{-1} (z-\mu)}$

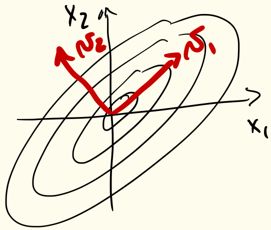
- $(z-\mu)^T \Sigma^{-1} (z-\mu)$ "measures" how unlikely z is
- This measure is a quadratic form!
- $\Rightarrow$ Contours of $p_{\underline{z}}(z)$ have the same shape as contours of

  $(z-\mu)^T \Sigma^{-1} (z-\mu)$

Suppose $X \sim \mathcal{N}(\mu, \Sigma)$ (multi-D)

① $p_{\underline{X}}(x) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$

Contours of $p_x(x)$



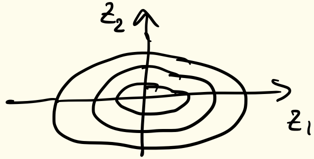
② Let Σ's eig vec $v_1, \dots, v_n$ $\|v_i\| = 1$

$\lambda_1, \dots, \lambda_n$

$x = Vz + \mu$ $V = [v_1 \dots v_n]$

$\Sigma = \sum v_i v_i^T \lambda_i$

$\Sigma^{-1} = \sum v_i v_i^T \frac{1}{\lambda_i}$



$\sqrt{\lambda_i}$ gives the standard deviation along v_i

③ $p_x(x) = \prod \frac{1}{\sqrt{2\pi\lambda_i}} e^{-\frac{1}{2}z_i^2/\lambda_i}$

Properties of Multi-D Gaussians

Assume:

$$\underline{X} \sim \mathcal{N}(\mu, \Sigma)$$

$$\underline{Y} \sim \mathcal{N}(r, \Omega)$$

① If X, Y independent and same size

$$\underline{X} + \underline{Y} \sim \mathcal{N}(\mu + r, \Sigma + \Omega)$$

② $$A\underline{X} + b \sim \mathcal{N}(A\mu + b, A\Sigma A^T)$$

$$E[A\underline{X} + b] = E[A\underline{X}] + b = A\,E[\underline{X}] + b$$

$$E\left[\left(A\underline{x} + b - E[A\underline{x} + b]\right)\left(A\underline{x} + b - E[A\underline{x} + b]\right)^T\right]$$

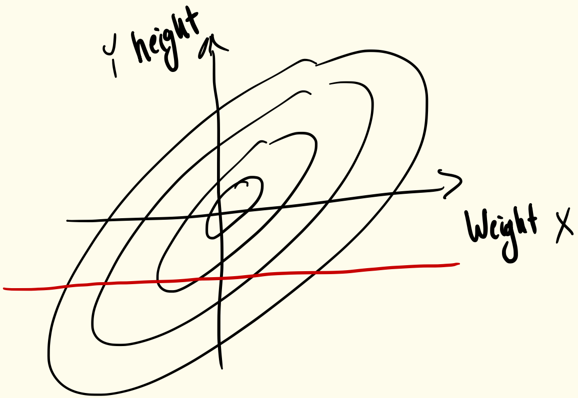
$$E\left[\left(A\underline{x} \cancel{+b} - (A\mu \cancel{+b})\right)\left(A\underline{x} \cancel{+b} - (A\mu \cancel{+b})\right)^T\right]$$

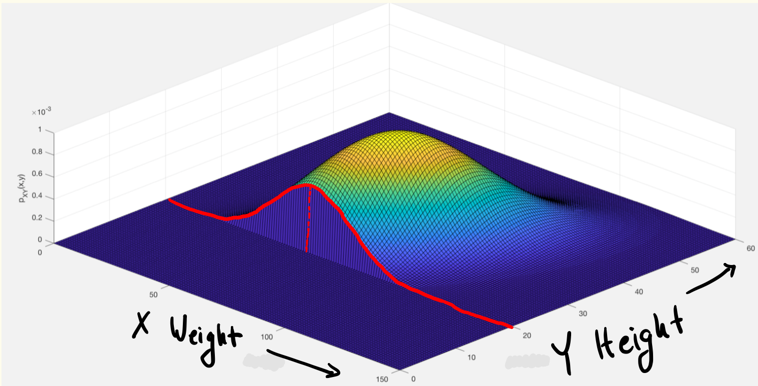
$$E\left[A(\underline{x} - \mu)(\underline{x} - \mu)^T A^T\right] = A\,E\left[(\underline{x} - \mu)(\underline{x} - \mu)^T\right]A^T$$

$$= A\Sigma A^T$$

Conditional Distributions: If you gain information, how does it affect uncertainty for other variables

Example: Suppose height and weight are jointly Normal. If you find out someone's weight, then ~~that affects~~ narrows down your uncertainty in their height...

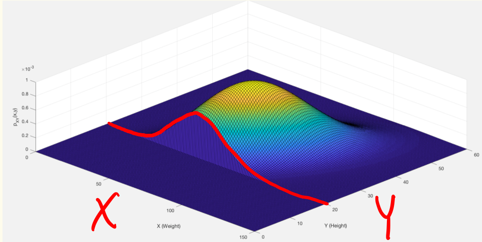




Conditional Probability:

- Let $p_{\underline{X}|\underline{Y}}(x|y)$ denote the PDF of $\underline{X}$ under the restriction that $\underline{Y}=y$.

- Conditional Gaussians



$$\begin{bmatrix} \underline{X} \\ \underline{Y} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix} \right)$$

$$\Sigma_{xy} = \Sigma_{yx}^T$$

reduction in uncertainty

$$p_{\underline{X}|\underline{Y}}(x|y) \sim \mathcal{N}\left(\mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y-\mu_y),\ \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} \right)$$

$$\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} \succeq 0$$

Preview:

$$x_{k+1} = A x_k + B u_k + w_k$$ gaussian noise

$$y_{k+1} = C x_{k+1} + v_k$$

- Given measurements $y_0 \cdots y_N$ what is the best estimate for x_N?

- Next time:
  - Kalman filtering via conditional gaussians
  - Sensor Fusion for legged locomotion