Will only provide enough background for discussion of

1) Covariance matrix
2) Principal Component Analysis (PCA)
3) Monte Carlo methods

Familiarity with basic statistics of random events, samples + random variables is assumed

Mean: Average value of n-samples

Probability: Likelihood of a particular sample to occur

Differentiate between the mean of a set of samples and the true mean or expected value of a random variable

Let there be N total random samples x_i for $i \in [1, N]$

Sample mean: $\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$

If these random samples can take n discrete values y_i for $i \in [1, n]$, each with probability p_i,

then Expected value: $\mu = E[X] = \sum_{i=1}^{n} p_i y_i$

with $\sum_{i=1}^{n} p_i = 1$

random variable X

Example: A six-sided die has equal probability of returning 1, 2, 3, 4, 5, 6 → each w/ $p_i = 1/6$

$$\mu = E[X] = \frac{1}{6}(1) + \frac{1}{6}(2) + \frac{1}{6}(3) + \frac{1}{6}(4) + \frac{1}{6}(5) + \frac{1}{6}(6) = 3.5$$

Expected Value of random variable X

You roll the die five times to obtain 2, 5, 1, 1, 3

For $N = 5$

$$\bar{x} = \frac{2+5+1+1+3}{5} = 2.4 \neq E[X]$$

Sample mean

However, the Law of Large Numbers states that as the number of samples increases ($N \to \infty$), then

$\bar{x} \to E[X]$, that is

$$\lim_{N\to\infty} \bar{x}_N = E[X]$$

Variance: Expected distance squared of a random sample from the expected value $E[X]$

Again, differentiate between sample variance S^2 and true variance σ^2

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

sample variance

Note: Divide by $N-1$, rather than N, to create an unbiased estimator of σ^2

$$\text{var}[X] = \sigma^2 = E[(X-\mu)^2] = E[(X - E[X])^2]$$

true variance

$$= \sum_{i=1}^{n} p_i (x_i - \mu)^2$$

for discrete values x_i

As before, from the Law of Large Numbers

As before, from the Law of Large Numbers

$$\lim_{N \to \infty} S_N^2 = \sigma^2 \qquad \text{Note, also } \sigma^2 \geq 0$$

The square root of the variance is the Standard Deviation σ, which is the expected spread about the mean

Example: A six-sided die has equal probability of returning 1, 2, 3, 4, 5, 6 → each w/ $p_i = 1/6$

$$\sigma^2 = \sum_{i=1}^{6} p_i (x_i - \mu)^2 = \frac{1}{6}\Big[(1-3.5)^2 + (2-3.5)^2 + (3-3.5)^2 + (4-3.5)^2 + (5-3.5)^2 + (6-3.5)^2\Big] = 2.91\bar{6}$$

You roll the die five times to obtain $\underbrace{2, 5, 1, 1, 3}_{\text{same experiment as above}}$ $(N = 5)$

$$S^2 = \frac{1}{5-1} \sum_{i=1}^{5} (x_i - \underbrace{\bar{x}}_{2.4})^2 = \frac{1}{4}\Big[(2-2.4)^2 + (5-2.4)^2 + (1-2.4)^2 + (1-2.4)^2 + (3-2.4)^2\Big] = 2.8$$

Note: Unless you have a large data set or you know μ, the best you have is $\bar{x}$

Probability Distribution Function (pdf)

In some cases, a sample only has a finite number of possible values (e.g., die). In these cases, there are a finite number of probabilities, which are called Discrete Probability Functions

Examples:

1. Bernoulli: Takes a value of 1 with probability p and 0 with probability $1-p$
2. Binomial: Yes/No with a probability ½ for each (e.g., coin toss)
3. Boltzmann: Probability a system will be in a certain state at a given temperature & pressure

However, most situations do not admit a finite number

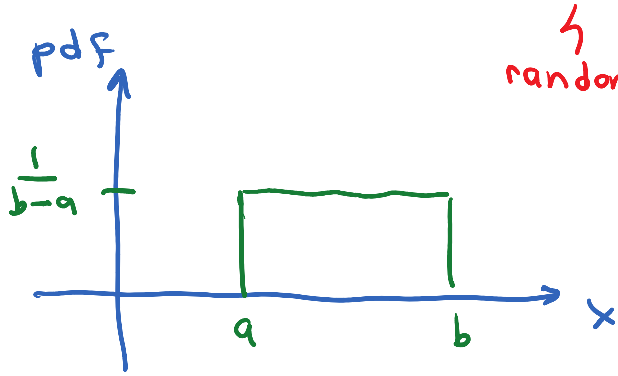
However, most situations do not admit a finite number of possible cases

Example: How long does it take to drive from North Campus to Niagara Falls? Too variable for discrete possibilities, although time will usually be clustered around 25 minutes.

This requires Continuous Distribution Functions

① Uniform: All values $x \in [a, b]$ are equally likely

random variable



$$p(x) = \begin{cases} \frac{1}{b-a} & \text{for } x \in [a,b] \\ 0 & \text{otherwise} \end{cases}$$

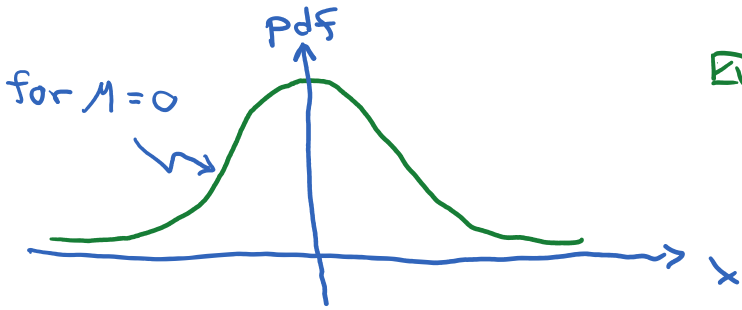
$$\int_{-\infty}^{\infty} p(x)\,dx = \int_a^b p(x)\,dx = 1$$

② Normal:

$$p(x) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right) e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad x \in (-\infty, \infty)$$

with mean (expected value) μ + std. deviation σ

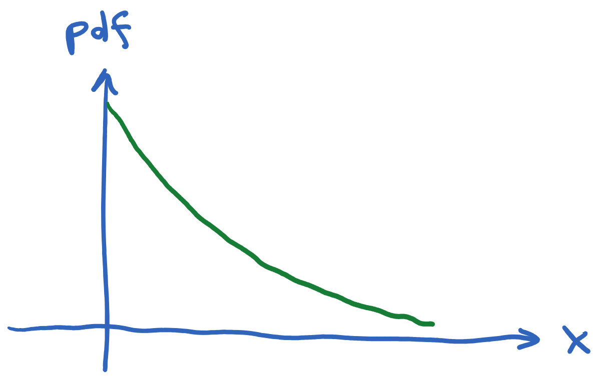
Informally called Bell Curve



Even function of x

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

③ Exponential: $P(x) = \lambda e^{-\lambda x}$ $\quad x \in [0, \infty)$



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

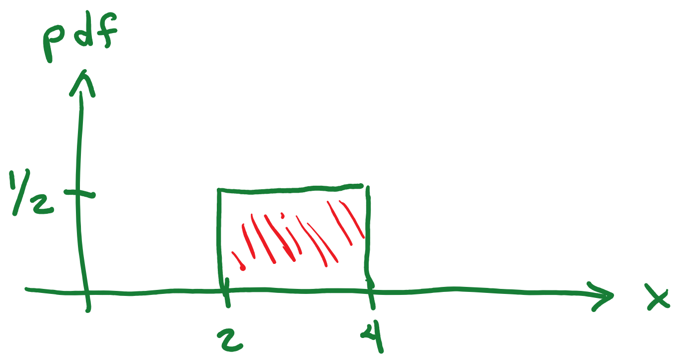
(Informal) Central Limit Theorem: The average of N samples of any pdf approaches a normal distribution as $N \to \infty$

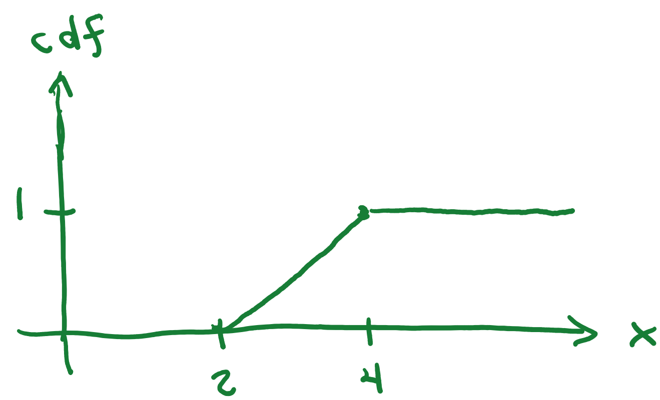
$\Rightarrow$ Analysis of Normal Distributions is very useful

Cumulative Distribution Function (cdf): Probability that a random sample y has a value $y \lesssim x$

$$F(x) = \int_{-\infty}^{x} p(y)\,dy$$

Example: Uniform distribution w/ $x \in [2,4]$





If $p(x)$ is continuous, then the probability that $x \in [a,b]$ is

$$P(a \leq x \leq b) = \int_a^b p(x)\,dx = F(b) - F(a)$$

Mean and Variances of pdfs

A sample has probability $p(x)$ to have a value of x

$$\Rightarrow \mu = E[X] = \int_{-\infty}^{\infty} p(x)\, x \, dx$$

Discrete case: $\mu = E[X] = \sum_{i=1}^{n} p_i y_i$

$$\Rightarrow \sigma^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} p(x)\,(x-\mu)^2\, dx$$

Discrete case: $var[X] = \sigma^2 = \sum_{i=1}^{n} p_i (x_i - \mu)^2$

Note: Some portion of $x \in (-\infty, \infty)$ may be zero

Notice analogy with μ & σ^2 for discrete case. Integrals replace summations!

Now look at systems where you may have many different samples, which might influence each other (i.e., samples are correlated)

Example: Age & height of children → typically older children are taller → age provides information on what height to expect

Covariance: A measure of joint variability of two random variables X & Y

If when X increases, then so does Y → Covariance is $+$

If when X increases, then Y decreases → Covariance is $-$

If X changes, but Y does not → Covariance is 0

Covariance of X & Y: $\mathrm{cov}[X,Y] = E[(X-\mu_x)(Y-\mu_y)]$

$= \sigma_{xy}$

where $\mu_x = E[X]$, $\mu_y = E[Y]$

What is $\mathrm{cov}[X,X]$?

$\mathrm{cov}[X,X] = E[(X-\mu_x)(X-\mu_x)]$

$$= E\left[(X-\mu_x)^2\right] = \sigma_{xx} = \sigma_x^2 > 0$$

How does

X influence X?

variance

Additionally,

$$\operatorname{cov}[X,Y] = E\left[(X-E[X])(Y-E[Y])\right]$$

$$= E\left[XY - XE[Y] - YE[X] + E[X]E[Y]\right]$$

$$= E[XY] - E[X]E[Y] - E[Y]E[X] + E[X]E[Y]$$

$$= E[XY] - E[X]E[Y]$$

Joint Probability: P_{ij} equals the probability that experiment 1 produces x_i and experiment 2 produces y_j

Then, covariance can be defined as

$$\operatorname{cov}[X,Y] = \sum_i \sum_j P_{ij}(x_i-\mu_x)(y_i-\mu_y)$$

for $i \in$ all possibilities of X

$j \in$ all possibilities of Y

Example:

Let X be either 1 or 2 with equal probability

Let Y be either 1 or 2 with equal probability

$\Rightarrow x_1 = 1,\ x_2 = 2,\ y_1 = 1,\ y_2 = 2 \Rightarrow \mu_x = \mu_y = 3/2$

Let X & Y be independent $\Rightarrow X$ & Y do not influence each other

$\Rightarrow p_{ij} = p_i\, p_j = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$ for all combinations of $i, j = \{(1,1), (1,2), (2,1), (2,2)\}$

Note: $\sum_i \sum_j p_{ij} = 1$

$$\mathrm{cov}[X, Y] = \sum_i \sum_j p_{ij} (x_i - \mu_x)(y_i - \mu_y)$$

$$= p_{11}(x_1 - \mu_x)(y_1 - \mu_y) + p_{12}(x_1 - \mu_x)(y_2 - \mu_y)$$

$$+ p_{21}(x_2 - \mu_x)(y_1 - \mu_y) + p_{22}(x_2 - \mu_x)(y_2 - \mu_y)$$

$$= \frac{1}{4}\left(1 - \frac{3}{2}\right)\left(1 - \frac{3}{2}\right) + \frac{1}{4}\left(1 - \frac{3}{2}\right)\left(2 - \frac{3}{2}\right)$$

$$+ \frac{1}{4}\left(2 - \frac{3}{2}\right)\left(1 - \frac{3}{2}\right) + \frac{1}{4}\left(2 - \frac{3}{2}\right)\left(2 - \frac{3}{2}\right)$$

$$+\frac{1}{4}\left(2-\frac{3}{2}\right)\left(1-\frac{3}{2}\right)+\frac{1}{4}\left(2-\frac{3}{2}\right)\left(2-\frac{3}{2}\right)$$

$$\therefore \ \mathrm{cov}[X,Y]=\sigma_{xy}=\frac{1}{4}\left(\frac{1}{4}-\frac{1}{4}-\frac{1}{4}+\frac{1}{4}\right)=0 \quad \underline{\text{uncorrelated}}$$

As expected, due to independence of X + Y

Example: Now look at situation with same possible states for X + Y, where if $x=1$, then $y=1$ + if $x=2$, then $y=2$

Again, $\mu_x=\mu_y=3/2$, but $p_{11}=\frac{1}{2}$, $p_{12}=p_{21}=0$

$p_{22}=\frac{1}{2}$

Note: $\sum_i \sum_j p_{ij}=1$

$$\mathrm{cov}[X,Y]=p_{11}(x_1-\mu_x)(y_1-\mu_y)+p_{22}(x_2-\mu_x)(y_2-\mu_y)$$

$$=\frac{1}{2}\left(1-\frac{3}{2}\right)\left(1-\frac{3}{2}\right)+\frac{1}{2}\left(2-\frac{3}{2}\right)\left(2-\frac{3}{2}\right)$$

$$\sigma_{xy}=\frac{1}{2}\left(\frac{1}{4}+\frac{1}{4}\right)=\frac{1}{4} \Leftarrow \text{correlated}$$

$\mu_x, \mu_y \qquad \underline{\mu} = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}$

$\sigma_{xx}, \sigma_{xy}, \sigma_{yx}, \sigma_{yy} \qquad \underline{V} = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{bmatrix}$

Covariance

Covariance Matrix

Covariance of X & Y: $\text{cov}[X,Y] = E[(X-\mu_x)(Y-\mu_y)] = \sigma_{xy}$

$$\mu_x = E[X], \quad \mu_y = E[Y]$$

Consider problem with two random variables X & Y

$$\text{cov}[x,x] = \sigma_x^2 = \sigma_{xx}$$

$$\text{cov}[y,y] = \sigma_y^2 = \sigma_{yy}$$

$$\text{cov}[x,y] = E[(x-\mu_x)(y-\mu_y)] = \sigma_{xy}$$

$$\text{cov}[y,x] = E[(y-\mu_y)(x-\mu_x)]$$

$$= E[(x-\mu_x)(y-\mu_y)] = \sigma_{xy}$$

Collect terms into a Co-variance Matrix

$$\underline{V} = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}$$

(Also represented as $\underline{K}$ or $\underline{\Sigma}$)

Example:

Let X be either 1 or 2 with equal probability $\Rightarrow \mu_x = 3/2$

Let Y be either 1 or 2 with equal probability $\Rightarrow \mu_y = 3/2$

Let X & Y be independent $\Rightarrow \sigma_{xy} = 0$ (from previous work)

$$\sigma_x^2 = \sum_i P_i (x_i - \mu_x)^2 = \frac{1}{2}\left(1-\frac{3}{2}\right)^2 + \frac{1}{2}\left(2-\frac{3}{2}\right)^2 = \frac{1}{4}$$

$$\sigma_y^2 = \sum_i P_i (y_i - \mu_y)^2 = \frac{1}{2}\left(1-\frac{3}{2}\right)^2 + \frac{1}{2}\left(2-\frac{3}{2}\right)^2 = \frac{1}{4}$$

$$\Rightarrow \underline{V} = \begin{bmatrix} 1/4 & 0 \\ 0 & 1/4 \end{bmatrix}$$

Example: Now look at situation with same possible states for X & Y, where if $x=1$, then $y=1$

& if $x=2$, then $y=2$

Again, $\mu_x = \mu_y = 3/2$, but $P_{11} = \frac{1}{2}$, $P_{12} = P_{21} = 0$

$P_{22} = \frac{1}{2}$

From previous work: $\sigma_{xy} = 1/4 = \sigma_{yx}$

Also, from above: $\sigma_x^2 = \sigma_y^2 = 1/4$

$$\Rightarrow \underline{V} = \begin{bmatrix} 1/4 & 1/4 \\ 1/4 & 1/4 \end{bmatrix}$$

Next, extend ideas to m random variables (e.g., age, height, weight)

Let $\underline{X}$ be the vector of m random variables $X_1, X_2, \ldots, X_m$

such that

$$\underline{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{bmatrix}$$

with mean $\underline{\bar{x}}$

Then

$$\underline{V} = E\left[(\underline{X} - \underline{\bar{x}})(\underline{X} - \underline{\bar{x}})^T\right] = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1m} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2m} \\ \vdots & \vdots & & \vdots \\ \sigma_{1m} & \sigma_{2m} & \cdots & \sigma_m^2 \end{bmatrix}$$

If $\underline{X} = X_1$ (one variable only), then

$$V = E\left[(X_1 - \bar{x}_1)(X_1 - \bar{x}_1)^T\right] = E\left[(X_1 - \bar{x}_1)(X_1 - \bar{x}_1)\right]$$

$$= E\left[(X_1 - \bar{x}_1)^2\right] = \sigma_1^2$$

Consider a linear combination of the random vector $\underline{X}$

$$\underline{c}^T \underline{X} = c_1 X_1 + c_2 X_2 + \dots + c_m X_m \,, \quad \underline{c} \in \mathbb{R}^m$$

Expected value of $\underline{c}^T \underline{X}$:

$$E[\underbrace{\underline{c}^T \underline{X}}_{\text{scalar}}] = \underline{c}^T E[\underline{X}] = c_1 E[X_1] + c_2 E[X_2] + \dots + c_m E[X_m]$$

constants!

Variance of $\underline{c}^T \underline{X}$:

$$\begin{aligned}
\operatorname{var}[\underline{c}^T \underline{X}] &= \operatorname{cov}[\underline{c}^T \underline{X}, \underline{c}^T \underline{X}] \\
&= E[(\underline{c}^T \underline{X} - E[\underline{c}^T \underline{X}])(\underline{c}^T \underline{X} - E[\underline{c}^T \underline{X}])^T] \\
&= E[(\underline{c}^T \underline{X} - \underline{c}^T E[\underline{X}])(\underline{c}^T \underline{X} - \underline{c}^T E[\underline{X}])^T] \\
&= E[(\underline{c}^T \underline{X} - \underline{c}^T \bar{\underline{x}})(\underline{c}^T \underline{X} - \underline{c}^T \bar{\underline{x}})^T] \\
&= E[\underline{c}^T (\underline{X} - \bar{\underline{x}})(\underline{c}^T (\underline{X} - \bar{\underline{x}}))^T] \\
&= E[\underline{c}^T (\underline{X} - \bar{\underline{x}})(\underline{X} - \bar{\underline{x}})^T \underline{c}] \\
&= \underline{c}^T E[(\underline{X} - \bar{\underline{x}})(\underline{X} - \bar{\underline{x}})^T] \underline{c} \\
&= \underline{c}^T \underline{V} \underline{c} \geq 0 \quad \Rightarrow \quad \underline{V} \text{ is positive semi-definite}
\end{aligned}$$

$= \underline{c}^T \underline{V}\, \underline{c} \geq 0 \Rightarrow \underline{V}$ is positive semi-definite

$\underline{V}$ is a real, symmetric matrix

$\Rightarrow \underline{V}$ is normal

$\underline{V}$ has an eigendecomposition, in which

all eigenvalues $\underline{\lambda}$ are real

Due to positive semi-definiteness

$\lambda_i \geq 0$ for $i = 1, 2, \dots, m$

Eigendecomposition of $\underline{V}$

$\underline{V} = \underline{\Phi}\, \underline{\Lambda}\, \underline{\Phi}^T$ where $\underline{\Phi}$ unitary ($\underline{\Phi}^T \underline{\Phi} = \underline{I}$)

$\underline{\Lambda}$ diagonal with λ_i's on diagonal