AINT351 - Distributions

Date: 03-10-16

More on Distributions

Expected value

- Discrete case:

$$E(X) = \sum_{\text{all x}} x_i p(x_i)$$

- The sample mean:

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n}$$

- Continuous case:

$$E(X) = \int_{\text{all x}} x_i p(x_i) dx$$

Variance

- Calculate mean before calculating the variance

- Discrete case:

$$Var(X) = \sigma^2 = \sum_{\text{all } x} (x_i - \mu)^2 p(x_i)$$

- The sample variance:

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

- Continuous case:

$$Var(X) = \sigma^2 = \int_{-\infty}^{\infty} (x_i - \mu)^2 p(x_i) dx$$

Covariance: joint prob

- The covariance measures the strength of the linear relationship between two variables
- This comes into practise when having two deminsional data that isn't independant

- The covariance us given by:

$$E[(x-\mu_x)(y-\mu_y)]$$

$$\sigma_{xy} = \sum_{i=1}^{N} (x_i - \mu_x)(y_i - \mu_y)\, P(x_i, y_i)$$

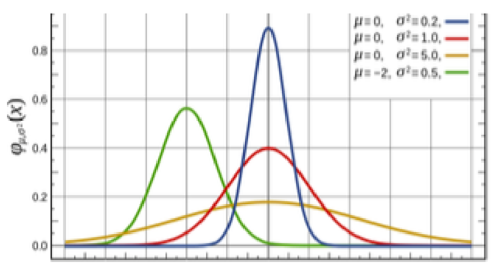
- The sample covariance:

$$\sigma_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{X})(y_i - \bar{Y})}{n-1}$$

- E = expectation

Normal distributions

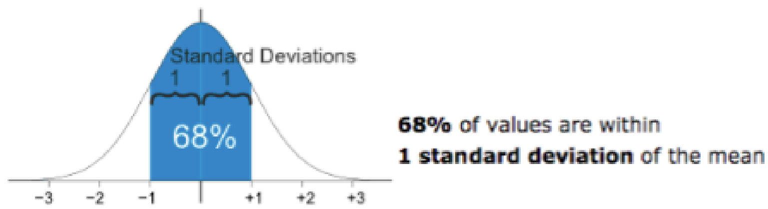
- Quite often continuous values will be characterised by a Normal (or Gaussian) distribution
- Sufficient statistics:
  - mean
  - standard deviation
- if you change the standard deviation you change the spread of the curve

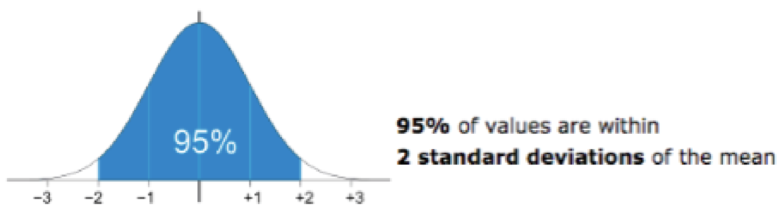


Sufficient statistics:
- mean
- standard deviation

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(a-\mu)^2/2\sigma^2} \qquad \mu, \sigma \text{ are given constants}$$

Effect of standard deviation

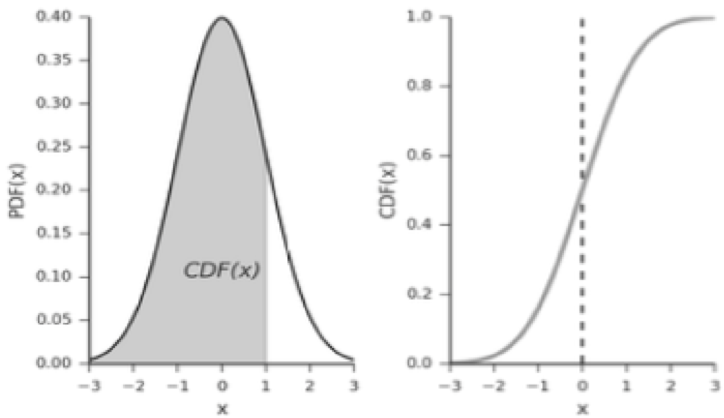




Cumulative distribution function

- For a continuous random variable X
  - The cumulative distribution function CDF(x) represents the ara under the probability densitiy function P(x) to the left of X
- not looking at the point x, what is the likelihood of all that is happening at that point and at all points previous
- area from that point - previous

$$CFD(x) = P(X \leq x)$$



Exponential distribution

- when the probability is above 0
  - not guassian (can or is negative)

Interpreting covariance

- If we calculate the covariance between two random variables
  - Will have two sets of results for everything, mean etc
- If cov(X,Y) is positive
  - Positiviely Correlated
- If cov(X,Y) is negative
  - Negatively correlated
- if cov(X,Y) is 0
  - Independant

To remember

- Mostly caring about the area under a curve

to do

- look at all functions, try to find videos explaning them

– go to the math place in library if can't understand