Probability Review II

CSCI-P556 Applied Machine Learning
Lecture 9

D.S. Williamson

Agenda and Learning Outcomes

- **Announcements**:
  - Everyone needs to create their own Github Repo. See the Piazza post instructions
  - Read and follow the homework instructions ***before*** starting the assignment
  - Notetakers
  - Information about project will come out in a week or so

Agenda and Learning Outcomes

- **Today's Topics**
  - Define continuous random variables, and name common families of CRVs
  - Compute moments of random variables (Expected value, variance, etc.)
  - Understand joint probability distributions (discrete/continuous, bivariate/multivariate)
  - Describe conditional probability and see how its used for random variables
  - Understand Bayes Rule
  - Determine independence between random variables
  - Compute covariance and correlation and understand what they mean for random variables

Recall: Types of Random Variables

- **Discrete Random Variables (DRV)**
  - If the RV can take integer (discrete) values $\{x_1, x_2, \ldots, x_m\}$ only, no values in between
  - **Example**:
    - X = the number of car accidents in Bloomington next weekend
    - *Possible values:* 0, 1, 2, ... -> X is a discrete RV

- **Continuous Random Variables (CRV)**
  - If the RV can take any value between two limits, where the number of possible values is uncountable
  - **Example**
    - X = a random person's height (in inches), measured to an infinite degree of accuracy.
    - *Possible values*: Any number in the interval [20, 100]

Continuous Random Variables

Continuous RVs take any value between two limits.

- Probability model, $f_X(X=x)$, is called a ***probability density function (***PDF)
- PDF satisfies

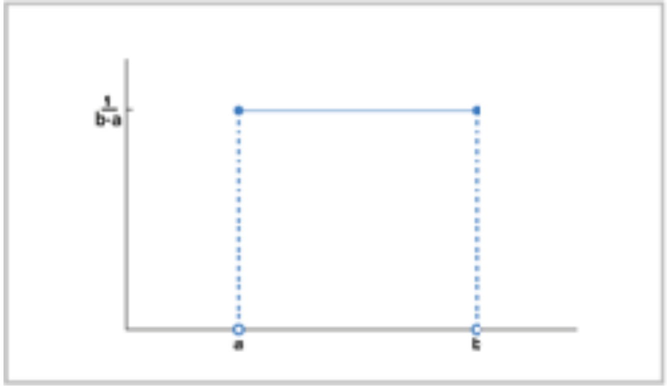
$$P(a \leq x \leq b) = \int_a^b f(x)dx \quad \text{whenever a} \leq \text{b}$$

- **Other properties of f(x):**
    - $f(x) \geq 0$
    - $\int_{-\infty}^{\infty} f(x)dx = 1$
    - $P(X = x) = 0$, for all x

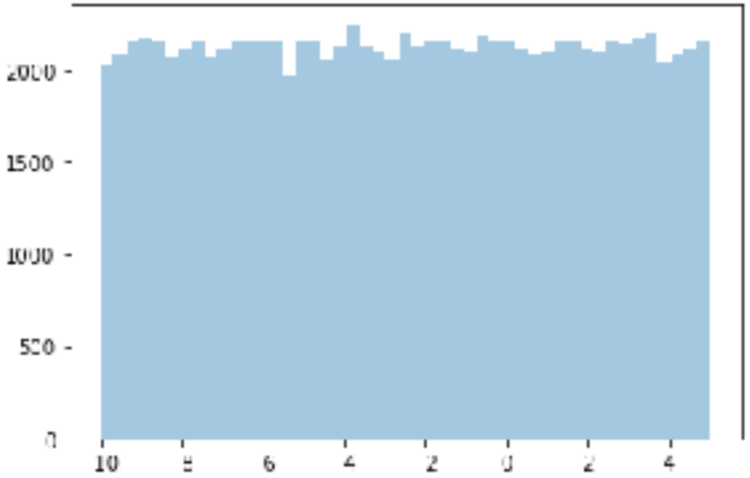
Important PDFs

- **<u>Continuous Uniform Distribution</u>**
  - Pretty much the same as the PMF for DRVs
  - Denoted as X ~ Uniform(a,b)

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a < x < b \\ 0, & \text{otherwise} \end{cases}$$




```
import numpy as np
s = np.random.uniform(-10,5,100000)
sns.distplot(s, hist=True, kde=False)
plt.show()
```

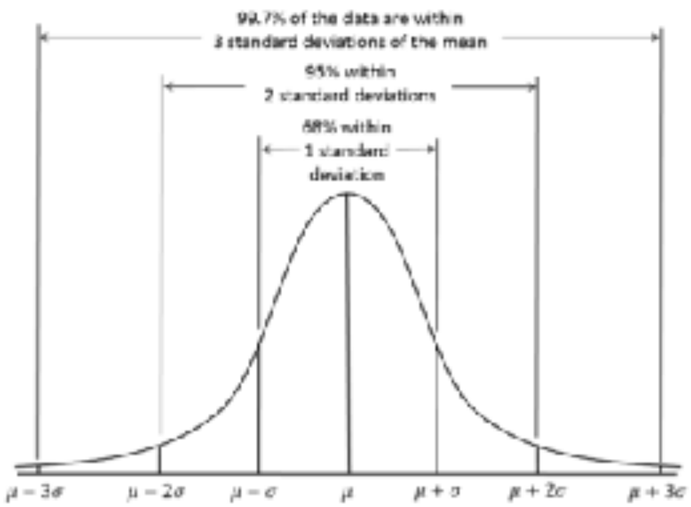


Important PDFs

Gaussian Distribution

- **<u>Normal/Gaussian Distribution</u>**
  - X ~ $N(\mu, \sigma)$
  - X has normal distribution with parameters μ and σ
  - X has PDF f(x) with

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



  - μ is the mean of the distribution
  - σ is the standard deviation of the distribution (more on these later)

Important PDFs

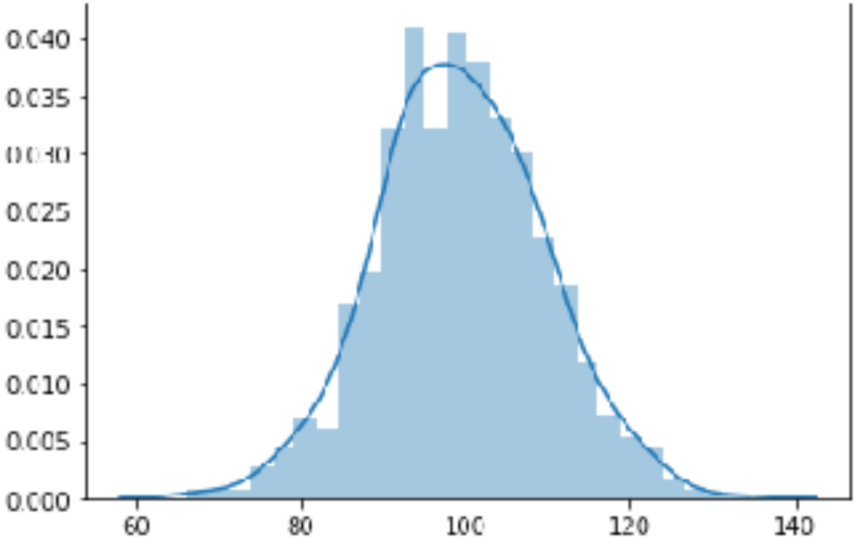
Gaussian Distribution

- **<u>Normal/Gaussian Distribution</u>**
  - X ~ $N(\mu, \sigma)$
  - X has normal distribution with parameters μ and σ
  - X has PDF f(x) with

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

  - μ is the mean of the distribution
  - σ is the standard deviation of the distribution (more on these later)

```
mu, sigma = 100, 10 # mean and standard deviation
s = np.random.normal(mu, sigma, 1000)
sns.distplot(s, hist=True, kde=True)
plt.show()
```



Moments of Random Variables

Expected Value

First Moment of a Random Variable

- Average (mean) value of X or any function of X, g(X)
    - Also known as the first moment
    - Denoted as E[X], μ, $\bar{X}$
- <u>If X is CRV, with PDF f(x), then</u>

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$$

- <u>If X is DRV, with PMF P(X), then</u>

$$E[X] = \sum_{x \in X} x P_X(x)$$

Higher Moments of a Random Variable

General computation of nth Moment

- Definition: n^{th} moment of RV X

$$E(X^n) = \begin{cases} \sum_{x \in X} x^n P_X(x), & \text{if discrete} \\ \int_{-\infty}^{\infty} x^n f_X(x)dx, & \text{if cont.} \end{cases}$$

- Definition: n^{th} central moment of RV X

$$\mu_n = E[(X-\mu)^n]$$

  - μ = E[X]

- Variance: 2^{nd} central moment

$$Var(X) = \sigma^2 = E[(X-\mu)^2]$$
$$= \begin{cases} \sum_{x \in X} (x-\mu)^2 P_X(x), & \text{if discrete} \\ \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)dx, & \text{if cont.} \end{cases}$$

Mean and Variance Examples

- Bernoulli distribution:

$$P_X(x) = \begin{cases} 1-p, & \text{if } x = 0 \\ p, & \text{if } x = 1 \\ 0, & \text{otherwise} \end{cases}$$

  - E[X] = 1*p + 0*(1-p) = p
  - Var(X) = (1-p)*(0 – p)^2 + p*(1-p)^2 = p(1-p)

- Binomial(n,p) Distribution

$$\binom{n}{x} p^x (1-p)^{n-x}$$

  - E[X] = np
  - Var(X) = np(1-p)

Verify these on your own in Python

Mean and Variance Examples

- <u>Continuous Uniform(a,b) distribution:</u>

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a < x < b \\ 0, & \text{otherwise} \end{cases}$$

  - E[X] = (b + a)/2
  - Var(X) = $(b-a)^2/12$

- <u>Normal/Gaussian Distribution</u>

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

  - E[X] = μ
  - Var(X) = σ^2

Verify these on your own in Python

Joint Distributions

Joint Distributions

Distributions involving two or more Random Variables

- Suppose have two RVs X and Y (i.e. bivariate case). How to describe their joint distribution?

- If Discrete RVs – Joint PMF

$$P_{XY}(x, y) = P(X = x, Y = y)$$

- Example: Toss 2 fair nickels and a fair dime
  - X = total money (in cents) when coins land on heads
  - Y = # tosses that are heads
  - Solution
    - Sample space, S, has 8 equally likely outcomes
    - S = {TTT, HTT, THT, HHT, TTH, HTH, THH, HHH}
    - So Joint PMF is?

		X			
P(x,y)	0	5	10	15	20
Y = 0	1/8	0	0	0	0
1	0	2/8	1/8	0	0
2	0	0	1/8	2/8	0
3	0	0	0	0	1/8

Discrete Joint Distributions

- Joint PMFs can be extended for more than two RVs (Multivariate Distributions)

$$P(X_1, X_2, X_3, ..., X_N) = P(X_1 = x_1, X_2 = x_2, ..., X_N = x_N)$$

- <u>General Rules</u>

$$P(X_1, X_2, X_3, ..., X_N) \geq 0$$

$$\sum_{X_1} \sum_{X_2} ... \sum_{X_N} P(X_1, X_2, X_3, ..., X_N) = 1$$

<u>Vector Notation</u>

$$\boldsymbol{X} = \{X_1, X_2, \ldots, X_N\}$$

$$P(\boldsymbol{X}) \triangleq P(X_1, X_2, X_3, \ldots, X_N)$$

Marginal PMFs

- Given a Joint PMF, *P(X,Y),* how to determine ***marginal*** PMFs, *P(X)* and *P(Y)*?
- If Discrete Bivariate RVs – Joint PMF

$$P(X) = \sum_{Y} P_{XY}(X,Y)$$
$$P(Y) = \sum_{X} P_{XY}(X,Y)$$

- If Discrete Multivariate RVs – Joint PMF

$$P_{X_1}(x_1) = \sum_{X_2}\sum_{X_3}\cdots\sum_{X_N} P(X_1, X_2, ..., X_N)$$
$$P_{X_2}(x_2) = \sum_{X_1}\sum_{X_3}\cdots\sum_{X_N} P(X_1, X_2, ..., X_N)$$
$$\vdots$$
$$P_{X_i}(x_i) = \sum_{X_1}\cdots\sum_{X_{i-1}}\sum_{X_{i+1}}\cdots\sum_{X_N} P(X_1, X_2, ..., X_N)$$

Continuous Joint Distributions

- <u>If Continuous RVs – Joint PDF</u>

$$f_{XY}(x, y) = f(X = x, Y = y) \quad \text{Bivariate case}$$

$$f(x_1, x_2, ..., x_N) = f(X_1 = x_1, ..., X_N = x_N) \quad \text{Multivariate case}$$

<u>Vector Notation</u>

$$\boldsymbol{X} = \{X_1, X_2, \ldots, X_N\}$$

$$f(\boldsymbol{X}) \triangleq f(X_1, X_2, X_3, \ldots, X_N)$$

- <u>General Rules</u>

$$f(x_1, x_2, \ldots, x_N) \geq 0$$

$$\int_{X_1} \cdots \int_{X_N} f(x_1, x_2, \ldots, x_N) = 1$$

<u>Marginal PDFs</u>

$$f_{X_i}(x_i) = \int_{X_1} \cdots \int_{X_{i-1}} \int_{X_{i+1}} \cdots \int_{X_N} f(X_1, ..., X_N) dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_N$$

Conditional Probability

Conditional Probability

- Have 2 events A and B. Given that B has occurred, what is the probability of A occurring?
  - Proportion of time that B occurs is P(B) (prior probability of B)
  - Proportion of time that A and B both occur is $P(A \cap B)$ or P(A,B), i.e., the joint probability of A and B
- ***Conditional Probability of A given B:***

$$P(A|B) = \frac{P(A,B)}{P(B)}$$

Notation:
$P(A,B) \equiv P(AB) \equiv P(A \cap B)$
$P(A \cup B) \equiv P(A + B)$

Conditional Probability: Example

- **Example**: Have 3 cards (where each card has 2 faces or sides).
  - One card has 2 red faces
  - One card has 1 red and 1 green face
  - One card has 2 green faces
  - Select a card and a face on it at random. The face is red.
  - What is the conditional probability that the flip side is also red?

$$P(A|B) = \frac{P(A,B)}{P(B)}$$

Conditional Probability: Example

Solution

Example: Have 3 cards (where each card has 2 faces or sides).

- One card has 2 red faces
- One card has 1 red and 1 green face
- One card has 2 green faces
- Select a card and a face on it at random. The face is red.
- What is the conditional probability that the flip side is also red?

- Let R_1 be the event, the first face is red
- R_2 be the event, the flip side is red
- $P(R_2|R_1) = ?$ $= \frac{P(R_2, R_1)}{P(R_1)}$ => $P(R_1) = \frac{3}{6}$ 3 faces are red, 6 total faces
- $P(R_2, R_1) = \frac{1}{3}$ Probability of picking the first card (3 cards total)

=>

$$P(R_2|R_1) = \frac{1/3}{1/2} = \frac{2}{3}$$

Conditional (Bivariate) Distributions

- X and Y are RVs with joint PMF P(X,Y)
- X has a marginal PMF P(X), where P(x) > 0 for all x
- The conditional PMF of Y given that X = x is

$$P(Y = y|X = x) = \frac{P(x, y)}{P(x)}$$

- Similarly for continuous RVs

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

Conditional (Bivariate) Distributions

- <u>Example</u>
  - X = # of touchdowns in a quarter
  - Y = # of field goals in a quarter
  - X and Y are discrete RVs, with Joint PMF
  - Find P(y|x)?
    - <u>Solution</u>
      - We have P(x,y)
      - Need to find P(x)

X \ Y	0	1	2	3
0	.30	.12	.08	.02
1	.15	.10	.05	.01
2	.08	.03	.01	0
3	.04	.01	0	0

Conditional (Bivariate) Distributions

$$P(Y = y|X = x) = \frac{P(x, y)}{P(x)} \qquad P(X) = \sum_{Y} P_{XY}(X, Y)$$

- Example
  - X = # of touchdowns in a quarter
  - Y = # of field goals in a quarter
  - X and Y are discrete RVs, with Joint PMF
  - Find P(y|x)?
    - Solution
      - We have P(x,y)
      - Need to find P(x)

X \ Y	0	1	2	3	P(X)
0	.30	.12	.08	.02	.52
1	.15	.10	.05	.01	.31
2	.08	.03	.01	0	.12
3	.04	.01	0	0	.05

$$P(y = 0|x = 1) = \frac{P(x = 1, y = 0)}{P(x = 1)} = \frac{.15}{.31} = .48$$

$$P(y = 1|x = 1) = \frac{P(x = 1, y = 1)}{P(x = 1)} = \frac{.10}{.31} = .32$$

$$P(y = 2|x = 1) = .16$$

$$P(y = 3|x = 1) = .03 \qquad P(y|x = 1) = 0 \text{ if } y \neq 0, 1, 2, 3$$

Law of Multiplication

For Conditional Probabilities

- The same axioms that apply to probability, also apply to conditional probability (e.g. $1 \geq P(B|A) \geq 0$,…)

- Law of Multiplication $P(B|A) = \dfrac{P(A,B)}{P(A)}$ => $P(A,B) = P(A)P(B|A)$

- General Law of Multiplication (prove own your own)

$$P(A_1, A_2, ..., A_N) = P(A_1)P(A_2|A_1)P(A_3|A_1, A_2)...P(A_N|A_1, ..., A_{N-1})$$

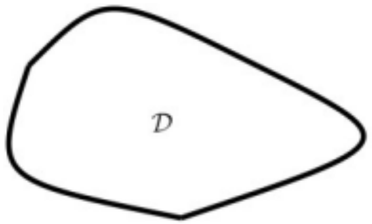
Conditional Probability: Example

- Example:
  - A Box has 3 red and 2 blue balls
  - Pick one, record its color, replace it in the box along with 2 additional balls of the same color
  - If do this 4 times, find the probability: first 3 are red and the 4th is blue
- Solution:
  - Let R_i = {i^{th} pick is red}
  - B_i = {i^{th} pick is blue}
  - $P(R_1,R_2,R_3,B_4) = P(R_1)P(R_2|R_1)P(R_3|R_1,R_2)P(B_4|R_1,R_2,R_3)$

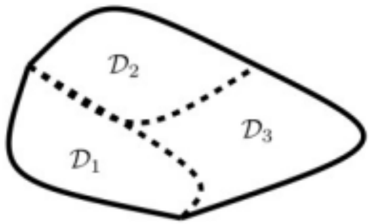
$$= \frac{3}{5} \times \frac{5}{7} \times \frac{7}{9} \times \frac{2}{11}$$
$$= \frac{6}{99} = \frac{2}{33}$$

Law of Total Probability

Combines Marginalization and Law of Multiplication



(a) Sample Space



(b) Partition of Sample Space

- <u>Law of Total Probability</u>
  - Let A_1, A_2, … be a partition of S and let B be any event, meaning they do not overlap but they combine to complete the sample space.

$$P(B) = \sum_{j=1}^{\infty} P(B, A_j)$$

Law of partition using marginalization

$$= \sum_{j=1}^{\infty} P(A_j)P(B|A_j)$$

Law of Multiplication

- <u>Note</u>: A is any event
  - Then A and A^c are a partition!
  - Law of Total Probability becomes

$$P(B) = P(A)P(B|A) + P(A^c)P(B|A^c)$$

Bayes Rule (Theorem)

Conditional Probability can be re-written

(a) Sample Space (b) Partition of Sample Space

- $A_1, A_2, \ldots$ form a partition
- $P(A_j) > 0$ for all j
- Let B be any event s.t. $P(B) > 0$
- then

$$P(A_i \mid B) = \frac{P(A_i, B)}{P(B)}$$

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{j=1}^{\infty} P(A_j)P(B|A_j)}$$

$P(A_i,B)$ -> Law of Multiplication

for any i=1,2,…

P(B) -> Law of Total Probability

Bayes Rule: Example

Suppose

- 50% of voters are Democrats
- 30% of voters are Republican
- 20% of voters are Independent

- 40% of Democrats voted for candidate X
- 80% of Republicans voted for candidate X
- 50% of Independents voted for candidate X

What fraction of candidate X's votes came from Republications?

- Solution
  - A_1 = {Democrats}, A_2 = {Republican}, A_3 = {Independent}
  - B = { voted for X}
  - $P(A_2 | B)$ = ?
  - $P(A_1) = 0.5$, $P(A_2) = 0.3$, $P(A_3) = 0.2$
  - $P(B | A_1) = 0.4$, $P(B | A_2) = 0.8$, $P(B | A_3) = 0.5$
  - A_1, A_2, and A_3 form a partition

=>

$$\begin{aligned} P(A_2|B) &= \frac{P(A_2, B)}{P(B)} \\ &= \frac{P(A_2)P(B|A_2)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3)} \\ &= \frac{0.3 * 0.8}{0.5 * 0.4 + 0.3 * 0.8 + 0.2 * 0.5} = \frac{4}{9} \end{aligned}$$

Independence

Independence

- X and Y are independent if and only if: $$f(x,y) = f_X(x) f_Y(y) \quad \text{for CRVs}$$ $$P(x,y) = P_X(x) P_Y(y) \quad \text{for DRVs}$$
- Equivalent condition
    - if X and Y are independent for all x and y, then

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

- Similarly
    - X and Y are independent ⇔ $f(x,y) = g(x)h(y)$ for some functions g and h
    - In other words,
        - if the joint distribution (PMF or PDF) can be written as a product of a function of RV X and a function of RV Y, then X and Y are independent
        - If X and Y are independent, then the joint distribution is a product of a function of X and a function of Y

Independence: Example

- Are X and Y independent?

$$f_{XY}(x,y) = \begin{cases} 2e^{-(x+2y)}, & \text{if } \quad 0 \leq x, 0 \leq y \\ 0, & \text{otherwise} \end{cases}$$

  - Solution
    - Can show that

$$\begin{aligned} 2e^{-(x+2y)} &= e^{-x}(2e^{-2y}) \\ &= f_X(x) f_Y(y) \end{aligned}$$

  - X and Y are independent

Conditional Independence

- Have random variables *X, Y, Z*

- Consider $P(x,y \mid z)$

- *X* and *Y* are <u>conditionally independent</u> given *Z* if:

$$P(x, y|z) = P(x|z)P(y|z)$$

Covariance

- <u>Definition</u>
  - The covariance between two RVs (*X* and *Y*) is a measure of their association

$$Cov(X,Y) = E[(X-\mu_x)(Y-\mu_y)] \quad \text{where } \mu_x = E[X]$$
$$\mu_y = E[Y]$$

  - So

$$\begin{aligned} Cov(X,Y) &= E[(X-\mu_x)(Y-\mu_y)] \quad \text{where} \\ &= E[XY - \mu_x Y - \mu_y X + \mu_x\mu_y] \\ &= E[XY] - \mu_x E[Y] - \mu_y E[X] + \mu_x\mu_y \\ &= E[XY] - 2\mu_x\mu_y + \mu_x\mu_y \\ &= E[XY] - \mu_x\mu_y \end{aligned}$$

Sign of Cov(X,Y)

- If Cov(X,Y) > 0
  - High values of X tend to occur with high values of Y
  - Low values of X tend to occur with low values of Y

$$Cov(X,Y) = E[XY] - \mu_x \mu_y$$

- If Cov(X,Y) < 0
  - High values of X tend to occur with low values of Y
  - Low values of X tend to occur with high values of Y
- If Cov(X,Y) = 0
  - X and Y are uncorrelated

Covariance and Correlation

- Correlation of X and Y

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} \qquad -1 \leq \rho_{XY} \leq 1$$

- Useful facts
  - $\rho_{XY} = 1$ <=> perfect positive linear association.
    with probability of 1, Y = aX + b for some $a > 0$
  - $\rho_{XY} = -1$ <=> perfect negative linear association.
    with probability of 1, Y = aX + b for some $a < 0$

Covariance and Correlation

- Correlation of X and Y

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

- Useful facts
    - If X and Y are independent, then Cov(X,Y) = ρ_{XY} = 0
        - Proof:
        - Cov(X,Y) = E[XY] – E[X]E[Y]
        - = E[X]E[Y] – E[X]E[Y] = 0
        - The reverse of this is not true: Cov(X,Y) = 0 then X and Y are independent (wrong!)

- **Next Class**

Naive Bayes Classification