Graph Mining Class Notes
Amin Assareh

Probability

Definition 1: Probability Space (Ω, F, P)

- Ω=Sample space: all possible results
  Example: roll a six sided die, results is Ω={1 2 3 4 5 6}

- F - power set of Ω: set of all possible subsets.

$$|F| = 2^{|\Omega|}$$

- P - probability function

Definition 2: Probability Function

Probability Function is any function that maps a subset of the sample space to a probability:

$$P : F \longrightarrow R$$

if E is an event(E $\subseteq$ F), then:
A) $0 \leq P(E) \leq 1$
B) $P(\Omega) = 1$
C) $P(E1 \cup E2) = P(E1) + P(E2)$ if $E1 \cap E2 = \oslash$

$\Rightarrow P(E1 \cup E2 \cup E3 \cdots En) = P(E1) + P(E2) + P(E3) + \cdots + P(En)$ if $Ei \cap Ej = \oslash$ $\forall i \neq j$
Lemma 1 :P(E1$\cup$E2) $= P(E1) + P(E2) - P(E1 \cap E2)$

Definition 3: Conditional Probability
Events $A, B \in F$
$P(A|B)$ - Probability of event A given event B is true

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

$$\Rightarrow P(A_1 \cap A_2 \cap \cdots \cap A_k) = \underbrace{\underbrace{P(A_1).P(A_2|A_1)}_{P(A_1 \cap A_2)}.P(A_3|A_1 \cap A_2)}_{P(A_1 \cap A_2 \cap A_3)}.\cdots.P(A_k|A_1 \cap A_1 \cap ... \cap A_{k-1})$$

Definition 4: Random Variables
Random variable X is a function which maps an event from sample space to a real number:

$$X : \Omega \to R$$

$$\Longrightarrow X = a \text{ means } \{y \in \Omega | X(y) = a\}$$

Example: roll a six-sided die twice, sum the results
Ω = {1-1, 1-2, 1-3, 1-4, 1-5, 1-6, 2-1, 2-2, 2-3, 2-4, 2-5, 2-6, 3-1, 3-2, 3-3, 3-4, 3-5, 3-6, 4-1, 4-2, 4-3, 4-4, 4-5, 4-6, 5-1, 5-2, 5-3, 5-4, 5-5, 5-6, 6-1, 6-2, 6-3, 6-4, 6-5, 6-6}
$|\Omega| = 36$
Each event has 1/36 probability.
X=4 is {1-3, 2-2, 3-1} $\Rightarrow$ P(X=4) = 3/36 = 1/12
X=2 is {1-1} $\Rightarrow$ P(X=2) = 1/36
X=7 is {1-6, 2-5, 3-4, 4-3, 5-2, 6-1} $\Rightarrow$ P(X=7) = 6/36 = 1/6

Definition 5: Independent events
X and Y are independent if

$$\forall x, y : P((X = x) \cap (Y = y)) = P(X = x).P(Y = y)$$

Definition 6: Expectation
The expectation of a random variable X is

$$E(X) = \sum_i i.P(X = i)$$

Example: Pick 10 cards. What is the expected number of red cards? (10×0.5 = 5)

Lemma 2: Linearity of Expectation Given events X1, X2,$\cdots$, Xk (mutually exclusive):

$$E(\sum X_i) = \sum E(X_i)$$

Proof: We need to show that E(X+Y)=E(X)+E(Y)
E(X+Y)=$\sum_{X,Y}(X + Y)P(X \wedge Y)$
$= \sum_{X,Y} X.P(X \wedge Y) + \sum_{X,Y} Y.P(X \wedge Y)$

$= \sum_X \sum_Y X.P(X \wedge Y) + \sum_Y \sum_X X.P(X \wedge Y)$
$= E(X) + E(Y)$

Definition 7: Variance of Random Variable X
Var(X) = E(X - E(X))^2
$= E(X^2 - 2XE(X) + (E(X))^2)$
$= E(X^2) - E(2XE(X)) + E(E(X)^2)$
$= E(X^2) - 2E(X)E(X) + (E(X)^2)$
$Thus Var(X) = E(X - E(X))^2 = E(X^2) - (E(X))^2$

Standard deviation : $\sigma(X) = \sqrt{Var(X)}$

Definition 8: Covariance of Two Random Variables X,Y

Cov(X,Y) = E((X-E(X)).(Y-E(Y))
= E(X.Y - X.E(Y) - Y.E(X) + E(X).E(Y))
= E(X.Y) - E(X.E(Y)) - E(Y.E(X) + E(X).E(Y))
= E(X.Y) - E(X).E(Y)

Lemma 3: Var(X+Y) = Var(X) + Var(Y) + 2.Cov(X,Y)
Proof:
Var(X+Y) = E(X+Y - E(X+Y))^2
$= E((X + Y)^2 - 2(X + Y)E(X + Y) + (E((X + Y))^2)$
$= E((X + Y)^2) - (E(X + Y))^2$
$= E(X^2) + 2E(X.Y) + E(Y^2) - (E(X + Y))^2$
$= E(X^2) + 2E(X.Y) + E(Y^2) - (E(X) + E(Y))^2$
$= E(X^2) + 2E(X.Y) + E(Y^2) - (E(X)^2 + 2E(X)E(Y) + E(Y)^2$
$= (E(X^2) - (E(X)))^2 + (E(Y^2) - (E(Y))^2 + (2E(X.Y) - 2E(X)E(Y))$
$= Var(X) + Var(Y) + 2Cov(X, Y)$

Lemma 4: if events X,Y are independent then
E(X.Y) = E(X).E(Y)
Cov(X,Y) = 0
Var(X+Y) = Var(X) + Var(Y)

Example: Bernoulli and Binomial distributions
Bernoulli distribution is a discrete probability distribution, which takes value 1 with success probability p and value 0 with failure probability q=1-p :
P(Y=1) = p
P(Y=0) = 1 - p
E(Y) = p

Var(Y) = p.(1-p)

Binomial distribution (Binomial(n,p)) is a sequence of n independent bernoulli trials(Y_k) with success probability p:

X=Binomial(n,p)=$\sum_{k=1}^{n} Y(k)$
$P(X = m) = \binom{n}{m} p^m (1 - p)^{n-m}$
$E(X) = E(Y1 + Y2 + Yn) = np$
$Var(X) = Var(Y1 + Y2 + \cdots + Yn)$
$= Var(Y1) + Var(Y2) + \cdots + Var(Yn)$
$= np(1 - p)$

Example: Coupon collector's problem
Suppose that there n coupons, from which coupons are being collected with replacement. What is the expectation of number of trials needed to collect all n coupons?

n=number of coupon types
k=order of successful experiment
i=order of desired coupon
EX_i=expected time for collecting i-th type of coupons
EX=expected time for collecting all n types of coupons

EX=EX$_0 + \cdots + EX_{n-1}$
$EX_i = \sum_1^\infty k.P(X_i = k)$
$P(X_i = k) = (\frac{i}{n})^{k-1}(1 - \frac{i}{n})$
suppose $p_i = 1 - \frac{i}{n} \rightarrow P(X_i = k) = (1 - p_i)^{k-1} p_i$
$\Rightarrow EX_i = \sum_{k=1}^\infty k(1 - p_i)^{k-1} p_i = p_i \sum_{k=1}^\infty k(1 - p_i)^{k-1}$
$= p_i \left(\frac{d}{dp} (- \sum_{k=1}^\infty (1 - p_i)^k) \right) = p_i \left(\frac{d}{dp} (- \frac{1}{p_i}) \right) = \frac{1}{p_i}$
$\Rightarrow EX = \frac{1}{p_0} + \cdots + \frac{1}{p_i} + \cdots + \frac{1}{p_{n-1}} = 1 + \cdots + \frac{n}{n-i} + \cdots + n =$
$n(\frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{n}) \cong n.logn$

Example: The Movie "21" Door Replacement Problem:

Problem: There are 3 doors. Behind one of them is gold. The other two have no prize.You chose Door No.1, Someone else opens Door No.3, and it turns out that nothing is behind it; Should you change your choice now?

The original version of this problem is called Monty Hall problem based on the American television game show "Let's Make a Deal".A well-known

statement of the problem was published in Parade magazine:
Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, Do you want to pick door No. 2? Is it to your advantage to switch your choice?

The answer is Yes, you should switch your choice. Players initially have a 1/3 chance of choosing the car and a 2/3 chance of choosing the goat. Players who stick to their original choice therefore have only a 1/3 chance of winning the car (and a 2/3 chance of getting a goat). Players who switch always get the opposite of their original choice so they have a 2/3 chance of getting a car (and 1/3 chance of getting a goat).