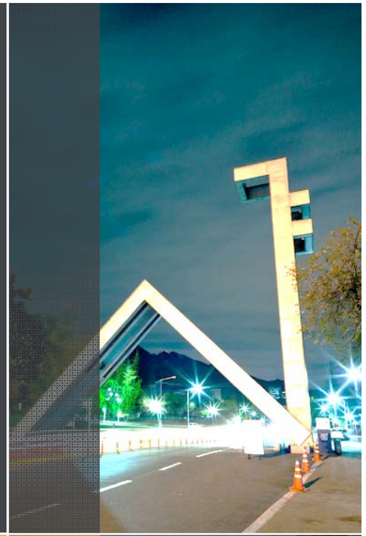
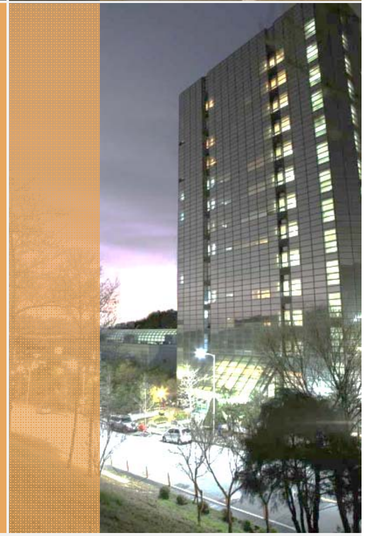
Seoul National University



Review for Midterm



Name: Chong-kwon Kim

SCONE Lab.

Probability

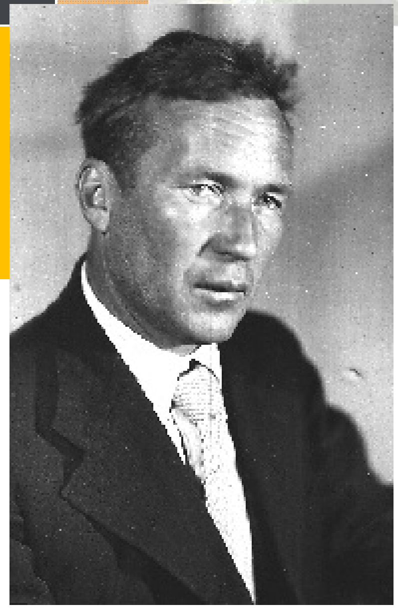
Cardano (1501~1576) was an Italian polymath, gambler
He invented idea of probability (odds), independence, binomial coefficients, Pascal refined later

- Cardano, a gambler & mathematician, first introduced the notion of probability
  - Roll two dice. How many times should you try until (6, 6) occurs?
- Probability
  - Trial: Roll two dice
  - Sample space: {(1,1), … (6, 6)}
  - Event: E1={6, 6}
  - Probability of E1 = 1/36

Axioms of Probability

SCONE Lab.

Kolmororov(1903~1987) was a Russian Mathematician
One of most important researchers in probability
At age 5, he discovered that sum of non-negative odd numbers is equal to square of a number (1+3+5+7 = 4^2)



- **Axioms of Probability**

  A1: $0 \le \Pr(E) \le 1$

  A2: $Pr(\Omega) = 1$

  A3: If E_1 and E_2 are mutually exclusive ($E_1 \cap E_2 = \varnothing$),

  then $Pr(E_1 \cup E_2) = Pr(E_1) + Pr(E_2)$

  → For any sequence of pairwise mutually disjoint events $E_1, E_2, \cdots, E_n$

  $Pr(\bigcup_{i=1}^{n} E_i) = \sum_{i=1}^{n} Pr(E_i)$

Lemmas

- Lemma1.1
  - For any events E_1 & E_2
    $Pr(E_1 \cup E_2) = Pr(E_1) + Pr(E_2) - Pr(E_1 \cap E_2)$

- Lemma 1.2: Union Bound
  - For any sequence of events E_i
  - $Pr(\bigcup_{\geq 1} E_i) \leq \sum_{i \geq 1} Pr(E_i)$

- Lemma 1.3: *Inclusion-exclusion principle*
  - $Pr(\bigcup_{i=1}^{n} E_i) = \sum_{i=1}^{n} Pr(E_i) - \sum_{i<j} Pr(E_i \cap E_j) + \cdots$
    $(-1)^{l+1} \sum_{i_1<i_2<\cdots<i_l} Pr(\bigcap_{r=1}^{l} E_{i_r}) + \cdots$

Independence, Conditional Prob.

Note E[X•Y] = E[X]•E[Y]

- Independence
  - Two events E and F are **independent** iff

$$\Pr(E \cap F) = \Pr(E) \cdot \Pr(F)$$

- Conditional probability, Pr(E|F)
  - Probability that E occurs *given* that F has already occurred

- Chain rule

What if E and F are independent?

$$\Pr(E|F) = \frac{\Pr(E \cap F)}{\Pr(F)} \text{ where } \Pr(F) > 0$$

$$\rightarrow \Pr(E \cap F) = \Pr(E \mid F) \bullet \Pr(F)$$

- More generally, $\Pr(E_1 \cap E_2 \cap \cdots \cap E_n)$

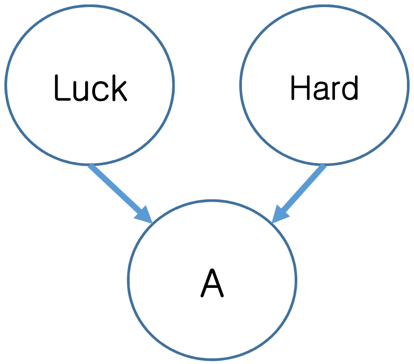
$$= \Pr(E_1)\Pr(E_2 \mid E_1)\Pr(E_3 \mid E_1 \cap E_2) \cdots \Pr(E_n \mid E_1 \cap E_2 \cap \cdots \cap E_{n-1})$$

Bayes' Theorem

- $\Pr(F \mid E) = \dfrac{Pr(E \cap F)}{\Pr(E)} = \dfrac{Pr(E \mid F)\Pr(F)}{\Pr(E)}$

$$= \frac{Pr(E \mid F)\Pr(F)}{\Pr(E \mid F)\Pr(F) + \Pr(E \mid \bar{F})\Pr(\bar{F})}$$

- Example
  - A student receives A grade either he is lucky(L) or work hard(H)
  - Given a student receives A, what is the probability the cause is luck?



Bayesian Network
Graphical model

Problem Solving

- Given a problem. Among many methods, which one to use?
  - Reiterate the problem as probability expressions
  - Properly define events, conditions and etc
  - Use proper notations

- Example – Car insurance company problem
  - There are two types of drivers: Careful (0.6) and Careless(0.4)
  - Probability that careful and careless customer have accidents in a year is 0.2 and 0.4, respectively
  - Events to have accidents in each year are independent (Depends only on the driving types)
  - Given a new customer have accidents in the first year, What is the probability that the customer have accidents in the second year?

Random Variables & Expectation

- A Random Variable X is a real-valued function defined on sample space

  X: Ω → R

- Independent
  - Two random variables X and Y are independent iff

  Pr((X=a) ∩ (Y=b)) = Pr(X=a) Pr(Y=b) for all a and b

- E[X]: Expectation of a rv X

  $$E[X] = \sum_i x_i \Pr(X = x_i)$$

Properties of Expectations

SCONE Lab.

- E[g(X)]=E[Y] = $\sum_j y_j$Pr($Y = y_j$)

  = $\sum_j \sum$ g(x_i)Pr(x_i)

- Example
  - n-th moment of X:

  E[X^n] = $\sum_i {x_i}^n \, Pr(X = x_i)$

- Linearity of Expectation
  - For any finite collection of discrete rv $X_1, X_2, \cdots, X_n$

  $E[\sum_{i=1}^{n} X_i] = \sum_{i=i}^{n} E[X_i]$

Bernoulli & Binomial RV

- Run an experiment
  - Success probability = p and Failure probability = (1−p)
- Bernoulli (Indicator) random variable Y is
  - $Y = \begin{cases} 1, & if\ success \\ 0, & if\ failure \end{cases}$
- E[Y] = p = Pr(Y=1)
- Now, perform the experiment n times. Random Variable X= the number of successes in n experiments
- Definition: Binomial random variable X with parameter n and p, B(n,p), is

  $\Pr(X=j) = \binom{n}{j} \cdot p^j (1-p)^{n-j}$

Geometric Distribution

- X: # coin flips until the first heads
- Definition: A Geometric random variable X with parameter p is given by the following probability distribution for n=1, 2,⋯

  Pr(X=n) = $(1-p)^{n-1} \cdot p$

- Properties
  - $\sum_{n \geq 1}$ Pr(X=n) = 1
  - Memoryless property: Given you tried k times w/o heads, how many more trials until the first success?
- Lemma: Pr(X=n+k | X>k) = Pr(X=n)

Conditional Expectation

SCONE Lab.

- Definition: E[Y | Z=z] = $\sum_y y \cdot \Pr(Y = y \mid Z = z)$

- Lemma: For any random variables X and Y,

  $$E[X] = \sum_y \Pr(Y = y) \cdot E[X|Y = y]$$

- Examples
  - Expectation and Variance of Geometric random variable
  - Y: result of the first flip = {0, 1}

  - E[X]= E[X | Y=0] Pr(Y=0) +
    E[X | Y=1] Pr(Y=1)
    = E[X+1]$\cdot$ $(1 - p)$+ 1$\cdot$ p

Conditional Expectation as a R.V.

- Definition: Expression $E[Y \mid Z]$ is a r.v. $f(Z)$ that takes on the value $E[Y \mid Z=z]$ when $Z=z$

- Theorem: $E[Y] = E[E[Y \mid Z]\}$

Moments, Variance

SCONE Lab.

- Definition: k-th moment of X $\equiv$ E$[X^k]$

- Definition: Variance

  $$\text{Var}[X] = \text{E}[(X - E[X])^2]$$
  $$= \text{E}[X^2] - E[X]^2$$

- Definition: Standard deviation

  $$\sigma[X] = \sqrt{\text{Var}[X]}$$

Properties of Moments & Variances

- Note that $E[X+Y] = E[X] + E[Y]$ holds even if X and Y are dependent

- $E[X \cdot Y] \equiv E[X] \cdot E[Y]$?
  - True only if X and Y are independent

- $Var[X+Y] \equiv Var[X]+Var[Y]$?

- Covariance of two r v X and Y

  $Cov(X, Y) = E[(X-E[X]) \cdot (Y-E[Y])]$

- Theorem: $Var[X+Y] = Var[X] + Var[Y] + 2 \cdot Cov(X,Y)$

MGF

SCONE Lab.

- Function that generate moments

  $M_X(\mathrm{t}) = \mathrm{E}[\ e^{tX}\ ] = \sum_i e^{tx_i} \cdot \Pr(x_i)$

- $\mathrm{E}[\ X^n] = {M_X}^{(n)}(0)$
  - where ${M_X}^{(n)}(\mathrm{t})$ is n th derivative of $M_X(\mathrm{t})$

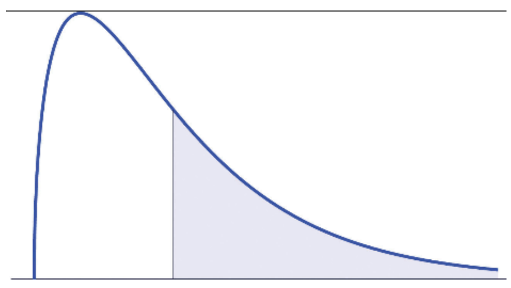
- If two random variables X and Y have the same MGF, then X ≡ Y

- If X and Y are independent r.v., then

  $M_{X+Y}(\mathrm{t}) = M_X(\mathrm{t})\ M_Y(\mathrm{t})$

Bounds

- We are interested in "Tail Bound", like $\Pr(X \geq a)$



- Markov
  - Only E[X] is given
- Chebyshev
  - E[X] and Var[X] are known
- Chernoff
  - MGF based

Bounds

SCONE Lab.

- Markov Ineqaulity
- Let X assumes only non-negative values.

  For any a > 0, $\Pr(X \geq a) \leq \frac{E[X]}{a}$

- Chebyshev's Inequality
- Also known as Weak Law of Large Number
- For any a > 0,

  $\Pr(|X - E[X]| \geq a) \leq \frac{Var(X)}{a^2}$

Let X_i: r v with mean μ and variance σ^2
Let $\overline{X_n} = (X_1+X_2+\ldots+X_n)/n$
$\lim_{n\to\infty} \Pr(|\overline{X_n} - \mu| > \varepsilon) = 0$

Chernoff Bounds

- Apply Markov inequality to e^{tX}
- From Markov inequality, for any t >0
  - Pr (X ≥ a) = Pr($e^{tX} \geq e^{ta}$) $\leq \frac{E[\,e^{tX}]}{e^{ta}}$
  - In particular, Pr (X ≥ a) $\leq \min_{t>0} \frac{E[\,e^{tX}]}{e^{ta}}$

MGF

Find appropriate *t* that minimizes the bound

- Similarly, for t < 0
  - Pr (X ≤ a) = Pr($e^{tX} \geq e^{ta}$) $\leq \frac{E[\,e^{tX}]}{e^{ta}}$
  - Hence, Pr (X ≤ a) $\leq \min_{t<0} \frac{E[\,e^{tX}]}{e^{ta}}$

Chernoff Bound for Poisson Trials

SCONE Lab.

Bernoulli trial: Each experiment has the same distribution

- Poisson trial
  - A sequence of experiments(trials) each of which has different distribution
  - Let $X_1, X_2, \ldots, X_n$ be a sequence of independent Poisson trials with $\Pr(X_i=1) = p_i$
  - $X = X_1 + X_2 + .. + X_n$
  - Let $\mu = E[X] = E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} p_i$

- Bounds of $\Pr(X \geq (1+\delta)\mu)$ and $\Pr(X \geq (1-\delta)\mu)$

  1. For any $\delta > 0$, $\Pr(X \geq (1+\delta)\mu) < \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$
  2. For $0 < \delta \leq 1$, $\Pr(X \geq (1+\delta)\mu) \leq e^{-\frac{\delta^2}{3}}$
  3. For $R \geq 6\mu$, $\Pr(X > R) \leq 2^{-R}$

Balls into Bins

- Throw m balls into n bins
- Hash m objects into a Hash table with n slots
- Collect m coupons each of which is one of n types
- Birthdays of m people where n is # possible birthdays

Poisson R.V.

- A discrete Poisson random variable X with parameter μ, Poi(μ), is

$$\Pr(X=j) = \frac{\mu^j e^{-\mu}}{j!}$$

- Poisson as Limit of Binomial

- Let Xn be Binomial with parameters *n* and *p*, where *p* is function of *n* and $\lambda = \lim_{n\to\infty} np$ is a constant and is independent of *n*. Then for any k

$$\lim_{n\to\infty} \Pr(X_n = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

- $\mathrm{Poi}(\mu_1) + \mathrm{Poi}(\mu_2) = \mathrm{Poi}(\mu_1 + \mu_2)$

Poisson Approximation

- One difficulty is the *dependency* between bins
- If we know that bin i is empty, then the probability that other bins are empty decreases
- Again throw m balls into n bins
- AND, Consider two sets of random variables

$\{X_1^{(m)}, X_2^{(m)}, \cdots, X_n^{(m)}\}$ and $\{Y_1^{(m)}, Y_2^{(m)}, \cdots, Y_n^{(m)}\}$

  - $X_i^{(m)}$ be the # balls in i-th bin
  - $Y_1^{(m)}, Y_2^{(m)}, \cdots, Y_n^{(m)}$ are independent Poisson with mean m/n.

Exact Case: # balls in a bin when m balls are thrown to n bins
Poisson Case: # balls in a bin is Poisson with mean m/n

Poisson Approximation

- Theorem: The distribution of $\{Y_1^{(m)}, Y_2^{(m)}, \cdots, Y_n^{(m)}\}$ conditioned on $\sum_i Y_i^{(m)} = k$ is the same as $\{X_1^{(k)}, X_2^{(k)}, \cdots, X_n^{(k)}\}$, regardless of m.

$Y_i^{(m)}$ is Poisson with mean m/n

- Let f(x_1,.., x_n) be a nonnegative function. Then

$$\mathrm{E}[f(X_1^{(m)}, X_2^{(m)}, \cdots, X_n^{(m)})] \leq e\sqrt{m}\,\mathrm{E}[f(Y_1^{(m)}, Y_2^{(m)}, \cdots, Y_n^{(m)})]$$

Exchange of Differentiation & Expectation

- Theorem 4.1: E[X^n] = ${M_X}^{(n)}$(0)
  - where ${M_X}^{(n)}$(t) is the *n* th derivative of M_X(t)
- While proving the theorem, we assume that exchange of expectation and differentiation is valid (Proof is beyond our scope)
- Proof of exchangeability of differentiation & differentiation is given in MathOverflow.
- Refer to https://math.stackexchange.com/questions/217702/when-can-we-interchange-the-derivative-with-an-expectation

Proof of $\Pr(X_1 \geq \mathrm{M}) \leq \binom{n}{M} \cdot (\frac{1}{n})^M$

SCONE Lab.

- While proving Lemma 5.1, we use the inequality $\Pr(X_1 \geq \mathrm{M}) \leq \binom{n}{M} \cdot (\frac{1}{n})^M$
- Proof
  - Let E1: Event that bin1 receives at least M balls
    $\Pr(\mathrm{E1}) = \Pr(X_1 \geq \mathrm{M}) \leq \binom{n}{M} \cdot (\frac{1}{n})^M$
- Proof
  - There are $\binom{n}{M}$ distinct methods to form size M balls subsets from a set of n balls
  - Let k be one of such subsets
  - Let E_k: Event that balls in subset k all land in bin 1
  - $\Pr(E_k) = (\frac{1}{n})^M$
  - $\Pr(X_1 \geq \mathrm{M}) = \Pr(\bigcup_{k=1}^{\binom{n}{M}} E_k) \leq \binom{n}{M} \cdot (\frac{1}{n})^M$

By Union bound