Aplikasi/Implementasi Probabilitas dalam “Birthday Problem”

Aurelia H B Matondang-13510023
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia
13510023@std.stei.itb.ac.id

Makalah ini berisi pembahasan implementasi dari teori-teori Probabilitas yang telah diajarkan selama senester ganjil tahun ajaran 2011/2012 dalam mata perkuliahan Struktur Diskrit.Pembahasan yang diambil dalma makalah ini adalah Pembahasan dalam masalah ulang tahun. Bagaimana dalam jumlah n orang (yang dipilih secara acak) ada beberapa pasang orang yang memiliki tanggal lahir yang sama.Hal ini dapat dijelaskan dengan beberapa teorema dan ketentuan yang pada akhirnya dapat menjawab (kurang lebih) dari masalh yang dihadapkan.

Kata Kunci- Fungsi Hash,”Birthday Attack”,”Pigeonhole Principle”.

I. PENDAHULUAN

Dari sekian jumlah penduduk di suatu daerah yang dilahirkan setiap hari dan setiap detiknya, pasti terdapat beberapa kasus dimana beberapa dari manusia tersebut memiliki waktu (tanggal lahir) yang sama. Kelahiran tersebut dapat didata dan dapat di prediksikan berapa persen kemungkinan kejadian tersebut terjadi dalam suatu daerah dengan jumlah penduduk yang dikaji dalam jumlah tertentu.
Perhitungan mengenai bagaimana mengetahui seberapa besar kemungkinan tersebut dapat dicari berdasarkan beberapa teori da hokum tertentu.
Pada bagian berikutnya akan dibahas mengenai beberapa teori dan hokum yang dipakai sebagai dasar penentuan kemungkinan kejadian ini terjadi.

II. BIRTHDAY PROBLEM

Dalam birthday problem ini yang akan dibahas adalah kemungkinan beberapa pasang orang bisa memiliki tanggal lahir yang sama. Contohnya dalam 24 orang yang dipilih secara acak, dengan membandingkan tanggal ulang tahun orang pertama dari 24 orang yang dipilih tersebut dengan 23 orang lainnya, orang pertama tersebut memiliki 23 kesempatan untuk menemukan kesamaan dalam tanggal lahir. Sedangkan orang ke dua memiliki kesempatan 22 untuk menemukan kesamaan tanggal ulang tahun dengan lainnya. Secara umum untuk menemukan beberapa kemungkinan terbentuknya pasangan (tanpa memperdulikan tanggal ulang tahun dan beberapa hal lain yang dapat mempengaruhi perhitungan probabilitas) dapat dilakukan dengan mencari kombinasi dari jumlah orang yang dipilih secara acak dengan jumlah orang yang akan dipasangkan (2 orang).

$$_{24}C_2 = (24.23)/(2) = 276$$

Untuk memilih sebuah tanggal untuk dibandingkan dengan tanggal lahir orang lainnya dalam 24 orang tersebut, terdapat 1/365 kemungkinan yang dimiliki oleh tanggal lahir dari setiap orang tersebut (dengan tidak memperhitungkan tanggal 29 Febuari). Perhitungan ini membuat perhitungan kemungkinan adanya tanggal lahir yang sama dari 24 orang tersebut secara statistic tidak ekuivalen. Perhitungan tersebut pada akhirnya akan mengarahkan perhitungan banyaknya pasangan yang mungkin, bukan sebagai jumlah dari orang yang memliki tanggal lahir yang sama.

III. PERHITUNGAN DAN TEORI

A. Perhitungan kemungkinan

Untuk kasus dimana dari 24 orang tersebut terdapat paling sedikit 2 orang dengan tanggal lahir yang sama. Untuk mempermudah distribusi darii variasi yang akan terbentuk, abaikan adanya kemungkinan dari 24 orang tersebut ada yang kembar dan kemungkinan ada yang memiliki tanggal ulang tahun 29 Febuari. Dengan menganggap semua tanggal memiliki kemungkina yang sama, maka perhitungan mengenai kemungkinan terdapat 2 orang memliki tanggal uang tahun yang sama dan yang tidak memiliki tanggal ulang tahun yang sama dapat dilakukan dengan cara

$$P(A) = 1 - P(A')$$

Dengan;
P(A) = kemungkinan terdapat 2 orang yang memiliki tanggal ulang tahun yang sama
P(A’) = kemungkinan tidak terdapat 2 orang yang memiliki tanggal ulang tahun yang sama.

Apabila kemungkinan P(A) dari 24 orang tersebut sebesar 50% maka jumlah 24 oarng tersebut dapat diambil sebagi contoh .

Ketika kejadian merupakan kejadian yang independen, maka kemungkinan seluruh kejadian terjadi sama dengan kemungkinan salah satu dari kejadian-kejadian tersebut terjadi. Sehingga, dalam kasus 24 orang setiap kejadian adalah kejadian dimana masing masing orang dibandingkan tanggal lahirnya (sesuai dengan kasus tertentu) dengan tanggal lahir 23 orang lainnya keudian dihitung kemnugkinan nya. Dan pada akhirnya kemungkinan terjadinya seluruh kejadian adalah dengan mengalikan setiap kejadian yang terjadi menyangkut kejadian utama tersebut.

Secara penulisan matematik, perhitungan menjadi :

$$P(A') = P(1).P(2).P(3)\ldots.P(23)$$

Untuk kasus pertama , dimana sebelumnya tidak ada orang yang dianalisis.Maka untuk orang pertama yang dianalisis mengalami ketidaksamaan tanggal ulang tahun dengan orang sebelumnya memliki kemungkinan sebesar 100%(365/365) (dengan tidak memperdulikan tahun kabisat).

Untuk kasus kedua, dimana orang kedua dibandingkan dengan orang yang dianalisis sebelumnya, yaitu orang pertama, maka kemungkinan orang tersebut untuk tidak memiliki kesamaan tanggal ulang tahun dengan orang sebelumnya adalah (364/365).Kemungkinan ini didasarkan apabila orang kedua memiliki tanggal lahir yang berbeda dari 364 hari lainnya selain tanggal kelahiran orang pertama.Untuk kasus berikutnya perhitungan kemungkinan dilakukan dengan cara yang sama.

Sehingga kemungkinan terjadinya ketidaksamaan tanggal lahir pada 24 orang yang dipilih secara acak adalah:

P(A')= 365/365. 364/365. 363/365… .342/365
P(A')= 0.4616557421
Maka,
P(A)=1-0.4616557421=
0.5383442579(53.834426%)

Proses ini dilakukan dengan memandang kemungkinan terbeut kemungkinan dalam sebuah grup dengan n orang 2 diantaranya memiliki tanggal lahir yang sama

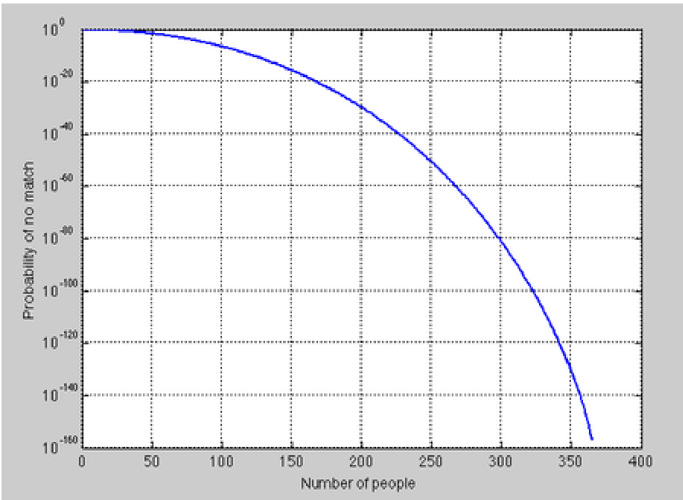
Untuk memempermudah perhitungan, berdasarkan teori pigeonhole:

$$\begin{aligned}\bar{p}(n) &= 1 \times \left(1-\frac{1}{365}\right) \times \left(1-\frac{2}{365}\right) \times \cdots \times \left(1-\frac{n-1}{365}\right) \\ &= \frac{365 \times 364 \times \cdots \times (365-n+1)}{365^n} \\ &= \frac{365!}{365^n(365-n)!} = \frac{n! \cdot \binom{365}{n}}{365^n}\end{aligned}$$

Dengan *p(n)* adalah kemungkinan tidak terdapat pasangan dengan tanggal ulangtahun yang sama. Dan ketika n>365 *p(n)* sama dengan 0. Sama seperti cara mengitung kemungkinan sedikitnya 2 orang yang memiliki tanggal ulang tahun yang sama maka kemungkinan tersebut didapatkan dengan cara:

$$p(n) = 1 - \bar{p}(n).$$

Apabila dibentuk tabel yang nantinya akan menggambarkan grafik kemungkinan, maka grafiknya akan berbentuk :



Dengan Tabel penunjuk grafik:

n	***p*(*n*)**
10	11.7%
20	41.1%
24	53.834%
30	70.6%
50	97.0%
57	99.0%
100	99.99997%
200	99.9999999999999999999999999998 %
300	$(100 - (6\times10^{-80}))\%$
350	$(100 - (3\times10^{-129}))\%$
365	$(100 - (1.45\times10^{-155}))\%$
366	100%

B. Aproksimasi

Probabilitas dari kejadian tersebut dapat dicari lagi dengan cara lain dengan hasil yang lebih sederhana dan akurat. Salah satu cara untuk mendapatkan hasil tersebut adalah dengan menggunakan cara eksponensasi sederhana:

$$p(n) \approx 1 - \left(\frac{364}{365}\right)^{C(n,2)} .$$

Dengan angka (364/365) merupakan probabilitas dari dua orang (acak) yang tidak memiliki kesamaan tanggal ulang tahun. Sedangkan kombinasi yang menjadi pangkat dari angka tersebut menandakan jumlah pasangan/ jumlah kejadian yang ada untuk menentukan kemungkinan total kejadian tersebut.

C. Generalisasi (Berdasarkan Ahli)

Dengan menganggap sebagai collision problem:

Diberikan n bilangan integer acak dari discrete uniform distribution dengan range [1,d].Berapa kemungkinan p(n;d) dengan paling sedikit 2 angka sama?(d=365)
Maka,

$$n(p;d) \approx \sqrt{2d \cdot \ln\left(\frac{1}{1-p}\right)}.$$

Dimana dengan konteks ini, birthday problem lebih umum diaplikasikan menurut fungsi hash.

Teori ini digunakan oleh Zoe Schnabel dibawah nama “capture-recapture” statistiv untuk menghitung populasi ikan di danau.

D. Teori yang digunakan untuk pemecahan masalah.

Dalam mencari kemungkinan terjadinya kejadian, terdapat beberapa teori yang mendasari perhitungan dan cara mencari kemungkinan.

a. Pigeonhole Principle
“Apabila n item diletakkan ke dalam m pigeonholes dengan n>m, maka sedikitnya satu pigeonhole harus terdiri dari lebih dari1 item”
Pembentukan ide ini dilakukan oleh Johann Dirichlet ada tahun 1834 dengan judul “Schubfachprinzip”.
Meskipun aplikasi yang paling mudah adalah dengan set terbatas, prinsip in juga digunakan untuk set tidak terbatas yang tidak bisa diletakkan dengan korespondensi satu-satu.Untuk melakukan hal demikian digunakan pernyataan formal bahwa”Fungsi injeksi tidak ada dalam set terbatas dengan kodomainnya lebih kecil daripada domainnya”.
Siegel’s lemma dibentuk dengan prinsip (secara umum) ini.

b. Birtday Attack
Sebuah birthday attack merupakan salah satu tipe dari kriptoanalisi yang mengatur/ merancang perhitungan matematikuntuk pemecahan masalah “Birthday Problem”.

c. Fungsi Hash
Fungsi Hash merupakan algoritma yang memetakan set data besar, yang juga disebu kunci, ke set data yang lebih kecil.
Contohnya untuk bilangan integer tunggal dapat berfungsi sebagai indeks daripada sebuah array. Nilai yang dikembalikan oleh fungsi hash disebut nilai hash, kode-kode hash, jumlah hash, atau hash-hash sederhana.
Fungsi hash terhubung (dan sering dibingungkan dengan) checksums, sek digit-digit, fingerprints(sidik jari), pengacakan fungsi-fungsi,kode-kode pengoreksi eror dan fungsi

hash kriptografi. Meskipun begitu, konsep ini "bertumpukan" sampai batas tertentu,dimana masing-masing kondep memiliki kegunaan tertentu dan syarat-syarat serta dirancang dan dioptimasi secara berbeda.

IV. BEBERAPA MASALAH LAINNYA DALAM "BIRTHDAY PROBLEM"

Masalah terbalik.

Untuk probabilitas tertentu p:

- Cari *n* terbesar dengan probablitas p(n) lebih kecil dari pada probabilitas p ,atau
- Cari n terkecil dengan probabilitas p lebih besar dari probabilitas p.

Dengan rumus yang juga digunakan dalam "birthday problem" (untuk *d* = 365):

$$n(p;365) \approx \sqrt{2 \times 365 \ln\left(\frac{1}{1-p}\right)}.$$

Contoh Hasil Perhitungan :

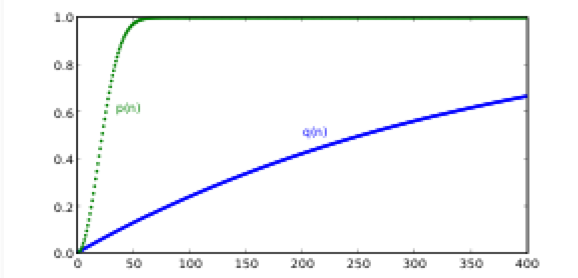
p	*n*	*n*↓	*p*(*n*↓)	*n*↑	*p*(*n*↑)
0.01	0.14178√365 = 2.70864	2	0.00274	3	0.00820
0.05	0.32029√365 = 6.11916	6	0.04046	7	0.05624
0.1	0.45904√365 = 8.77002	8	0.07434	9	0.09462
0.2	0.66805√365 = 12.76302	12	0.16702	13	0.19441
0.3	0.84460√365 = 16.13607	16	0.28360	17	0.31501
0.5	1.17741√365 = 22.49439	22	0.47570	23	0.50730
0.7	1.55176√365 = 29.64625	29	0.68097	30	0.70632
0.8	1.79412√365 = 34.27666	34	0.79532	35	0.81438
0.9	2.14597√365 = 40.99862	40	0.89123	41	0.90315
0.95	2.44775√365 = 46.76414	46	0.94825	47	0.95477
0.99	3.03485√365 = 57.98081	57	0.99012	58	0.99166

Catatan: beberapa nilai yang berada diluar batas telah diberi warna. Ini menunjukkan aproksimasi ttidak selalu tepat.

"First match"

Untuk kasus dimana sekelompok orang masuk ke dalam suatu ruangan secara bersamaan, timbul pertanyaan siapa yang lebih cocok untuk ditunjuk atau disebut sebagai orag pertama yang memiliki tanggal lahir yang sama dengan orang yang sudah lebih dahulu berada di ruangan tersebut?
Berkaitan dengan pertanyaan sebelumnnya, maka jawabannya adalah 20. Dengan n yang membuat p(n)-p(n-1) maksimum.

Tanggal Lahir yang sama dengan Anda



Membandingkan *p*(*n*) = probabilitas kecocokan tanggal lahir dengan *q*(*n*) = probabilitas mencocokkan tanggal lahir Anda.

Ingat bahwa dalam birthday problem, yang lain dari kedua orang dipilih secara khusus. Maka, probabilitas *q*(*n*) bahwa seseorang dalam sebuah ruangan dengan n orang lainnya memiliki tanggal lahir yang sama dengan orang tertentu adalah:

$$q(n) = 1 - \left(\frac{365-1}{365}\right)^n$$

Dan untuk d umum dengan,

$$q(n;d) = 1 - \left(\frac{d-1}{d}\right)^n.$$

Dalam kasus standar untuk *d* = 365 penggantian *n* = 24 menghasilkan 6.4%, dimana lebih kecil dari pada 1 kesempatan 16 kejadian. Untuk kesemaptan yang lebih dari 50% bahwa seseorang dalam ruangan yang penuh dengan orang memiliki tanggal lahir yang sama dengan Anda, n harus sebesar (minimal) 276. Ingat bahwa angka ini secar signifikan lebih besar daripada 365/2 = 182.5: dimana menjadi alasan bahwa ada kemungkinan terjadinya kecocokan tanggal lahir denga seseorang yang ada di dalam ruangan tersebut.

Kecocokan dekat

Persoalan umum lainnya mempertanyakan berapa banyak orang yang dibutuhkan untuk memiliki kesempatan lebih dari 50% dimana 2 orang memiliki tanggal lahir yang sama dalam satu hari satu dengan yang lainnya , atau dua hari , atau tiga hari , dan seterusnya satu sama lain. Persoalan yang lebih rumit ini membutuhkan penggunaan prinsip inklusi dan eksklusi Jumlah orang yang dibutuhkan sehingga probabilitas dari beberapa pasangan akan memiliki tanggal ulang tahun yang terpisah lebih sedikit dari k hari akan lebih tinggi dari 50% adalah:

k	**# people required**
1	24
2	14
3	11
4	9
5	8
6	8
7	7
8	7

Dengan demikian dalam kelompok dengan 7 orang (acak) , lebih terlihat seperti tidak 2 orang dari pasangan tersebu memiliki tanggal lahir dalam satu minggu dalam masing-masing tanggal lahir

REFERENSI

[1] Wikipedia 2011
http://en.wikipedia.org/wiki/Birthday_problem
Waktu akses :10 Desember 2011; pukul 8.38 WIB

[2] Wikipedia 2011
http://en.wikipedia.org/wiki/Hash_function
Waktu akses : 10 Desember 2011;pukul 9.00 WIB

[3] Wikipedia 2011
http://en.wikipedia.org/wiki/Pigeonhole_principle
Waktu akses : 10 Desember 2011; pukul 9.03WIB

[4] Wikipedia 2011
http://en.wikipedia.org/wiki/Birthday_attack
Waktu akses: 10 Desember 2011; pukul 9.05 WIB

PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Bandung, 10 Desember 2011

Aurelia H B Matondang-13510023