Contents

[Conditional Probability Example 2](#_Toc28113321)

[Expected Value 4](#_Toc28113322)

[Normal distribution 7](#_Toc28113323)

[Binomial Distribution 14](#_Toc28113324)

[Exponential 19](#_Toc28113325)

[Poisson 20](#_Toc28113326)

[Uniform distribution 26](#_Toc28113327)

[Regression 28](#_Toc28113328)

[Probability Questions 28](#_Toc28113329)

[What is the probability of getting 50 heads when 100 coins are tossed? 29](#_Toc28113330)

[Throw a fair die 100 times, what is the probability that   you will get more than 75 tails? 29](#_Toc28113331)

[32. In any 15-minute interval, there is a 20% probability that you will see at least one shooting star. What is the probability that you see at least one shooting star in the period of an hour? 30](#_Toc28113332)

[2. Does the frequentist approach always give the same result as the Bayesian approach? 30](#_Toc28113333)

[5. A jar contains 4 marbles. 3 Red & 1 white. Two marbles are removed with replacement after each draw. Find the probability that the same color marble is drawn twice? 30](#_Toc28113334)

[6. In a website offering dating services, users can select 5 out of 24 adjectives to describe themselves. A match is said to be between two users if they match on at least 4 adjectives. 31](#_Toc28113335)

[10. In a class of 30 students, what is the probability that two of the students have their birthday on the same (assuming that it is not a leap year)? 32](#_Toc28113336)

[An example of a favourable event would be students with birthday 3rd Jan 1998 and 3rd Jan 32](#_Toc28113337)

[1997. 32](#_Toc28113338)

[12. A fly has a lifetime of between 4-6 days. What is the probability that the fly will die in exactly 5 days? 32](#_Toc28113339)

[Case Study 33](#_Toc28113340)

[Alice and Bob take turns in rolling a fair dice. Whoever gets "6" first wins the game. Alice starts the game. What are the chances that Alice wins. 33](#_Toc28113341)

[If you draw 2 cards from a shuffled 52 card deck, what is the probability that you'll have a pair? 34](#_Toc28113342)

[There's a game where you are given two fair six-sided dice and asked to roll. If the sum of the values on the dice equals seven, then you win $21. However, you must pay $5 to play each time you roll both dice. Do you play this game? And in follow-up: What is the probability of making money from this game? 34](#_Toc28113343)

[We have two options for serving ads within Newsfeed 34](#_Toc28113344)

[: 1 - out of every 25 stories, one will be an ad 34](#_Toc28113345)

[2 - every story has a 4% chance of being an ad 34](#_Toc28113346)

[For each option, what is the expected number of ads shown in 100 news stories? If we go with option 2, what is the chance a user will be shown only a single ad in 100 stories? What about no ads at all? 34](#_Toc28113347)

[If a PM says that they want to double the number of ads in 35](#_Toc28113348)

[Newsfeed, how would you figure out if this is a good idea or not? 35](#_Toc28113349)

[What is the expectation of the variance? 35](#_Toc28113350)

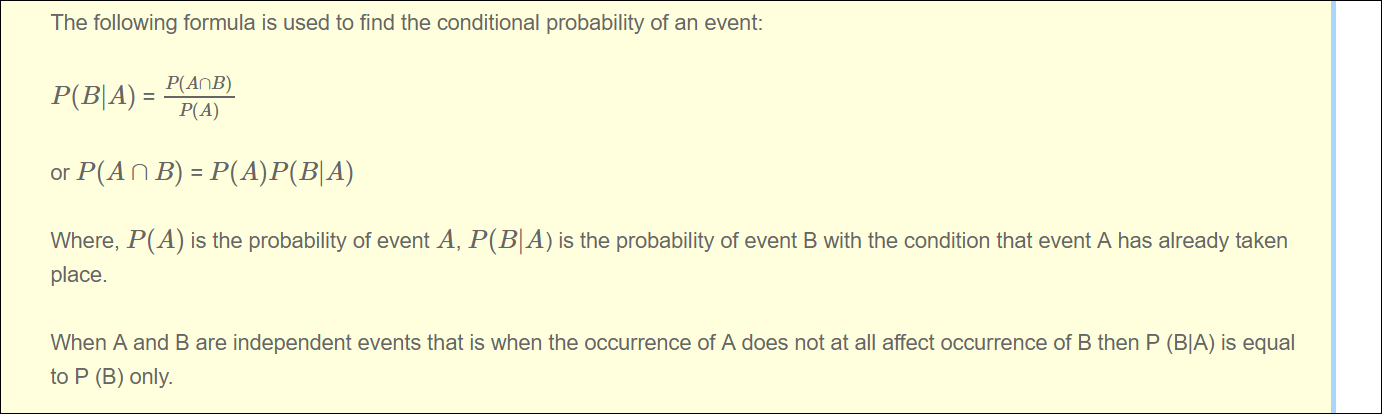
[Bayes Theorem 35](#_Toc28113351)

[You're about to get on a plane to Seattle. You want to know if you should bring an umbrella. You call 3 random friends of yours who live there and ask each independently if it's raining. Each of your friends has a 2/3 chance of telling you the truth and a 1/3 chance of messing with you by lying. All 3 friends tell you that "Yes" it is raining. What is the probability that it's actually raining in Seattle? 37](#_Toc28113352)

[Suppose N students participate in a coin flip experiment, when they get heads they stop, when they get tails they keep going. All students will stop after the second trial no matter the results. Y is the binary indicator of whether they claim they cheated in the experiment. Estimate how many students cheat in this experiment. 37](#_Toc28113353)

[Link 38](#_Toc28113354)

# Conditional Probability Example

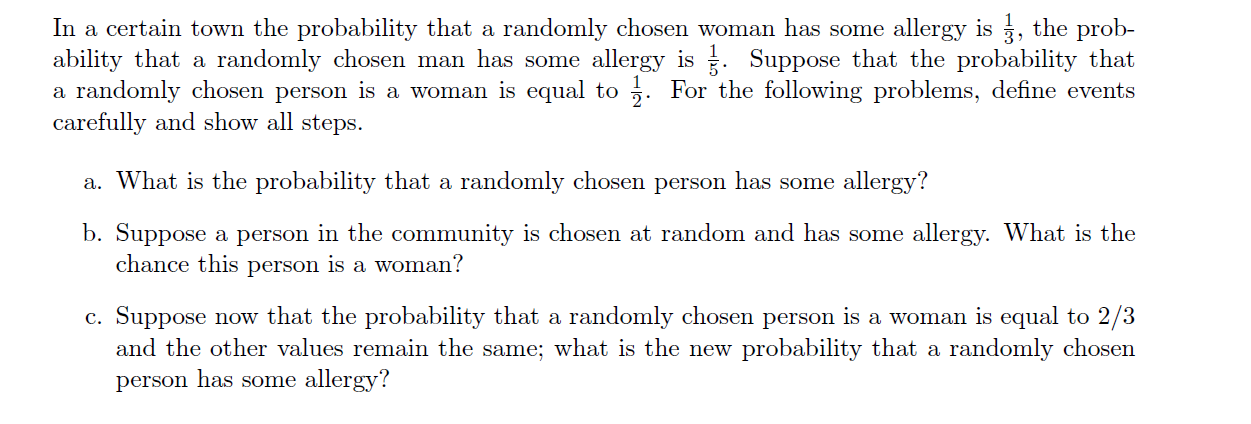


Suppose an individual applying to a college determines that he has an 80% chance of being accepted, and he knows that dormitory housing will only be provided for 60% of all of the accepted students. The chance of the student being accepted *and* receiving dormitory housing is defined by  
*P(Accepted and Dormitory Housing)*

*= P(Dormitory Housing|Accepted)\*P(Accepted)*

= (0.60)\*(0.80) = 0.48.

***P(B) = P(B|A)P(A) + P(B|Ac)P(Ac)*,**



|  |
| --- |
| Let W be the event that a randomly chosen individual in the town is a woman (this means that W’ is the event that a randomly chosen individual in the town is a man).  Let A be the event that a randomly chosen individual in the town has some allergy.  This problem asks for:  P(A)  From the problem statements we know:  P(A|W) = 1/3  P(A|W’) = 1/5  P(W) = 1/2  and so P(W’) = 1/2  Therefore, by conditioning on W and W’, we have  **P(A) = P(A|W) \* P(W) + P(A|W’) \* P(W’)**  P(A) = 1/3 \* 1/2 + 1/5 \* 1/2 |

# Expected Value

The player pays $2 and spins the spinner. If the spinner lands on purple, the player wins $0.50. If the spinner lands on yellow, the player wins $5. If the spinner lands on blue, the player wins $1. If the spinner lands on red, the player wins nothing*.*



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| X | Blue | Yellow | Purple | Red | wager |  |
| Prize amount | 1 | 5 | 0.5 | 0 | -2 |  |
| probability | 3/8 | 2/8 | 2/8 | 1/8 | 1 |  |
| **Expected value=** x1 p1 + x2 p2 + … + xn p | 3/8 | 10/8 | 1/8 | 0\*1/8 | -2 |  |

If the dice match (doubles are rolled), the player pays $5. How much should the house pay if the dice do not match to make the game fair? [Remember, a game is fair if the expected value of the game is $0.]

|  |  |  |  |
| --- | --- | --- | --- |
| X | Dice match | Dice do not match | Total |
| Prize amount | 5 | x |  |
| Probability | 1/6 | 5/6 |  |
| E(X) | 5(1/6) | X(1/6) | 5/6+x/6=0 |

What is the expected win for a $1 bet on black in a single roll of American roulette? The American roulette wheel has 38 numbered fields, two of which are green (0 and 00), 18 red and 18 black

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *X* | *Black* | *Red* | *green* |  |  |
| *Prize* | ***1*** | ***-1*** | ***-1*** |  |  |
| *Prob* | *18/38* | *18/38* | *2/38* |  |  |
| *E(x)* | *18/38* | *-18/38* | *-2/38* |  |  |

***18/38 - 20*** */38 =-2/38*

**Problem 3**: (20 pts) The main-prize for the weekly state lottery is one million dollars. There are five second-prizes of $100,000 each. Suppose that a ticket costs $1 and that ten million people had bought a single ticket each. Furthermore, all six prizes are awarded (i.e., there are no rollovers to the future draws).

Define a random variable representing the amount of money a player can win (or lose), find its probability distribution, and compute the expected win per ticket (in dollars)?

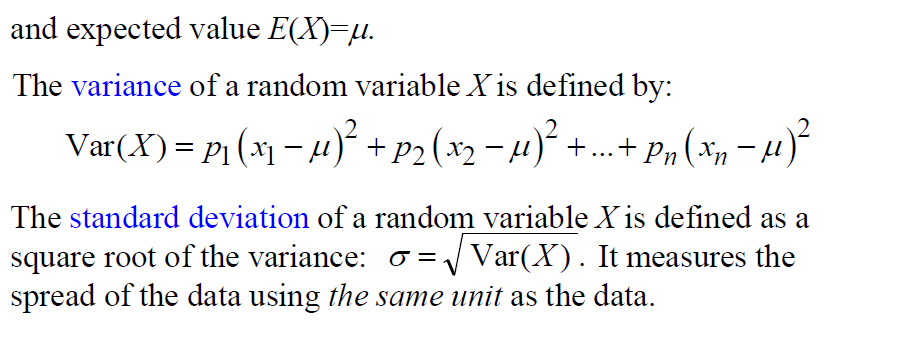
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***X*** | ***First*** | ***Second*** | ***wager*** |  |
| *Price* | *1M* | *100k* | *-1* |  |
| *Probability* | *1/10M* | *5/10M* | *1* |  |
| *E(x)* | *1/10* | *5/100* | *-1* | *-0.85* |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Problem 4**: A black and a white die are rolled. Let *M* be the random variable recording **the larger of the two numbers** showing.   1. (10 pts) Find the probability distribution of the random variable *M*. 2. (5 pts) Find the expected value of *M*.  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | **M** | 1 | 2 | 3 | 4 | 5 | 6 |  | | **Frequency of 2 dice** | 1 | 3 | 5 | 7 | 9 | 11 |  | | **Probability**  **Distribution** | 1/36 | 3/36 | 5/36 | 7/36 | 9/36 | 11/36 |  | | **E(x)** =x1 p1 + x2 p2 + … + xn p | 1/36 | 2\*3/36 | 3\*5(36) | 4\*7/36) | 5\*9/36 | 6\*11/36 | 161/36 |   b) Expected Value = x1 p1 + x2 p2 + … + xn p  = 161/36 =4.472222 |

: Box contains 6 half-dollars, 4 quarters, 10 dimes and 5 nickels. A single coin is drawn from the box.

1. (10 pts) Create the frequency table and probability distribution of the random variable *X* representing the value drawn (in dollars).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***x*** | ***Half dollar*** | ***quarter*** | ***dimes*** | ***Nickels*** |
| Price | 0.5 | 0.25 | 0.1 | 0.05 |
| Probability | 6/25 | 4/25 | 10/25 |  |
|  |  |  |  |  |
| E(x) |  |  |  |  |



|  |
| --- |
|  |

**Problem 4**: A black and a white die are rolled. Let *M* be the random variable recording **the larger of the two numbers** showing.

1. (10 pts) Find the probability distribution of the random variable *M*.
2. (5 pts) Find the expected value of *M*.

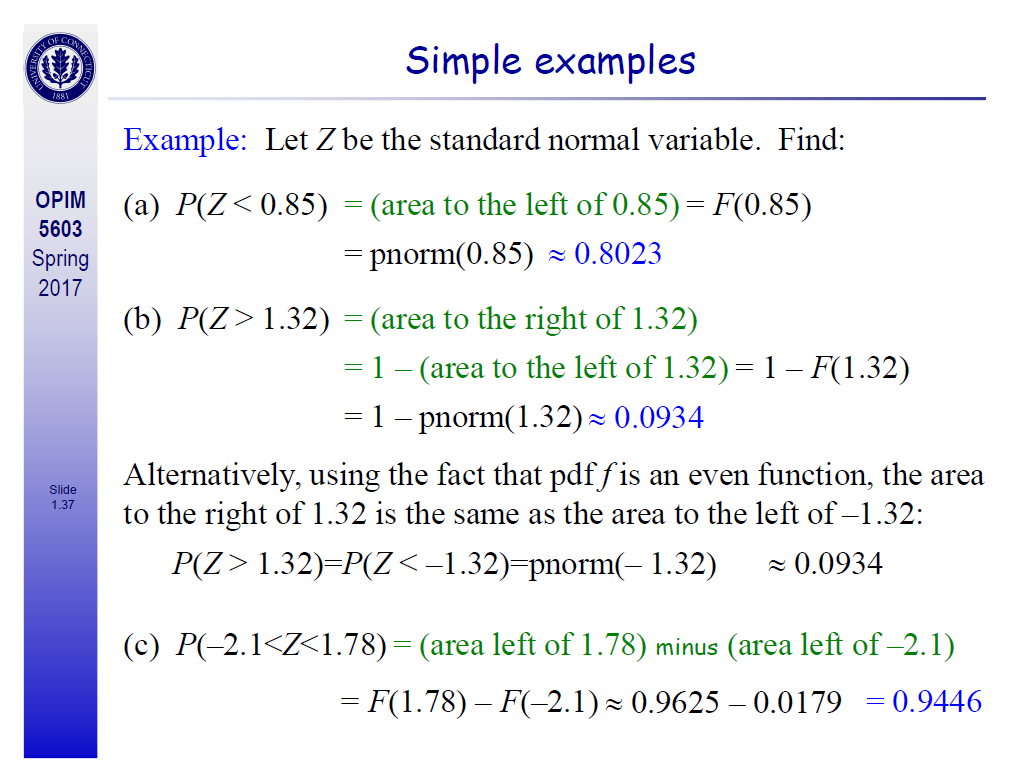
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **M (x)** | 1 | 2 | 3 | 4 | 5 | 6 |  |
| **Frequency of 2 dice** | 1 | 3 | 5 | 7 | 9 | 11 |  |
| **Probability**  **Distribution** | 1/36 | 3/36 | 5/36 | 7/36 | 9/36 | 11/36 |  |
| **E(x)** =x1 p1 + x2 p2 + … + xn p | 1/36 | 2\*3/36 | 3\*5(36) | 4\*7/36) | 5\*9/36 | 6\*11/36 | 161/36 |

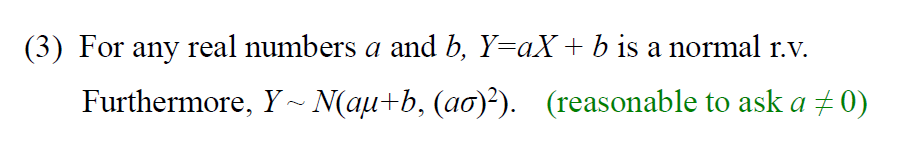
b) Expected Value = x1 p1 + x2 p2 + … + xn p

= 161/36 =4.472222

# Normal distribution

**To check if data is normal – mean should be equal to median , skewness should be equal to 0 or less than 1 ,kurtosis is should be almost equal to 3 , and sd should be 1**





|  |
| --- |
| ***Questio 7***  **Problem 3**: (10 pts) Let *X*~ *N*(3,4). What is the distribution of the random variable 2*X* – 1?  That function will be normal distribute  X ~ N(μ, σ2)  Y(2X-1) ~ N(aμ+b, (aσ)2)  Y(2X-1) ~ N(2(3) +(-1) , (2\*2)2  N(μ, σ2) = N(5,16)  E(2X – 1) = 2EX – 1 = 2・ 3 – 1 = 5  Var(X) = 16 |

|  |
| --- |
| ***Questio 8***  A particular rash has shown up at an elementary school. It has  been determined that the length of time that the rash will last is  normally distributed with μ = 6 days and σ = 1.5 days. Find the  probability that for a student selected at random, the rash will  last for less than 3 days.  P(X < 3)=P(N(6,1.52) < 3)  Pnorm(2,6,1.5) |

|  |
| --- |
| 54. The expected (mean) life of a particular type of light bulb is 1,000 hours with a standard deviation of 50 hours. The life of this bulb is normally distributed. What is the probability that a randomly selected bulb would last longer than 1150 hours?  a) 0.4987  b) 0.9987  c) 0.0013  d) 0.5013  e) 0.5513     1. Pnorm(3)   1-pnorm(1150,1000,50) |

|  |
| --- |
| 43. Let z be a normal random variable with mean 0 and standard deviation 1. What is P(1.3 < z < 2.3)?  a) 0.4032  b) 0.9032  c) 0.4893  d) 0.0861  e) 0.0086  pnorm(2.3) - pnorm(1.3) =0.086 |

|  |
| --- |
| . Let z be a normal random variable with mean 0 and standard deviation 1. What is P(z > -1.1)?  z > -1.1 => z<=F(-1.1) => 1- pnorm(-1.1) |

|  |
| --- |
| 47. Let z be a normal random variable with mean 0 and standard deviation 1. What is  P(-2.25 < z < -1.1)?  pnorm(-1.1) - pnorm(-2.25) |

|  |
| --- |
| A standard normal distribution has the following characteristics:  a) the mean and the variance are both equal to 1  b) the mean and the variance are both equal to 0  c) the mean is equal to the variance  d) the mean is equal to 0 and the variance is equal to 1  e) the mean is equal to the standard deviation  Ans: d  Response: See section 6.2, Normal Distribution  Difficulty: Medium |

|  |
| --- |
| 58. Sure Stone Tire Company has established that the useful life of a particular brand of its automobile tires is normally distributed with a mean of 40,000 miles and a standard deviation of 5000 miles. What is the probability that a randomly selected tire of this brand has a life of at most 30,000 miles?  Atomost 30,000 => F(x<=30000)  pnorm(30000,40000,5000)  If it is discrete F(x<=30000-1) |

|  |
| --- |
| A manufacturer of a special type of one-size glove wants to design the glove to fit at least 99 percent of the population. Hand span is known to be normally distributed with a mean of 195 millimeters and a standard  deviation of 17 millimeters. What range of hand spans must the glove accommodate?  Here probability is given to us which is 99%.  So what is left is 1% together on both side  So 0.5% on each side  To find the value below which 0.5 percent of the population falls, use the command:  > qnorm(0.005, 195, 17)  [1] 151.2109  Similarly, to find the value above which 0.5 percent of the population falls, use the command:  > qnorm(0.005, 195, 17, lower.tail=F)  Or  qnorm(.995, 195, 17)  [1] 238.7891  The remaining 99 percent of the population falls between these two values. So, to accommodate 99 percent of  the population, the gloves must be designed to fit hands with a span between 151 and 239 millimeters. |

|  |
| --- |
| 1. P(X<70) => pnorm(70,79,sqrt(144)) 2. P(64<X<96) => Pnorm(96) -Pnorm(64) |

|  |
| --- |
| Since 20% of the members stay for more than 125 minutes, 80% stay for less than or equal to 125 minutes. For a standard normal distribution,  qnorm(.80) will give the z score so,  80% of the area under the curve lies to the left of z = 0.841621234. We know that  z = (x - µ)/σ   0.841621234 = (125 - 90)/σ   σ = (125 - 90)/0.841621234 = 35/0.841621234 = 41.58640325  b) Now that we know sigma, we can find the z score for 25 minutes,  z = (25 - 90)/41.58640325 = -65/41.58640325 = -1.563010862  The area to the left of that z score is 0.05902502, so that's the probability that a visit will last less then 25 minutes.  **pnorm(25,90,41.5)**  c) Since Tara arrives at 8 PM, she has only two hours (120 minutes) to use the facility. The z score for 120 minutes is  z = (120 - 90)/41.58640325 = 30/41.58640325 = 0.721389629  The area under the curve to the left of that z score is 0.764665087, so we're losing 0.235334913, the area to the right. In other words, our distribution has about 23.5% of its area chopped off, so it's not a suitable model. |

|  |
| --- |
| The random variable X is normally distributed with mean 79 and variance 144: X ~ N (79,144)    Total area is 1  So we 1-0.6463/3 =0.1179  Sd=12 given  qnorm(0.2358,79,12,lower.tail=F) =79+b  b =8.63 |

|  |
| --- |
|  |

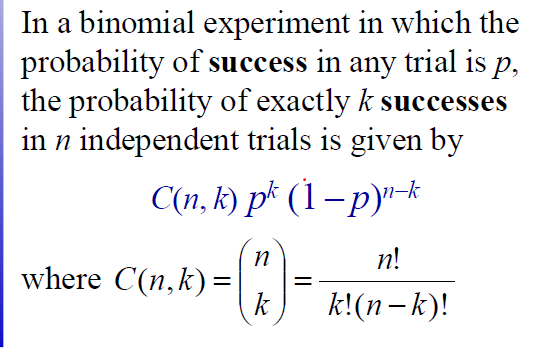
# Binomial Distribution

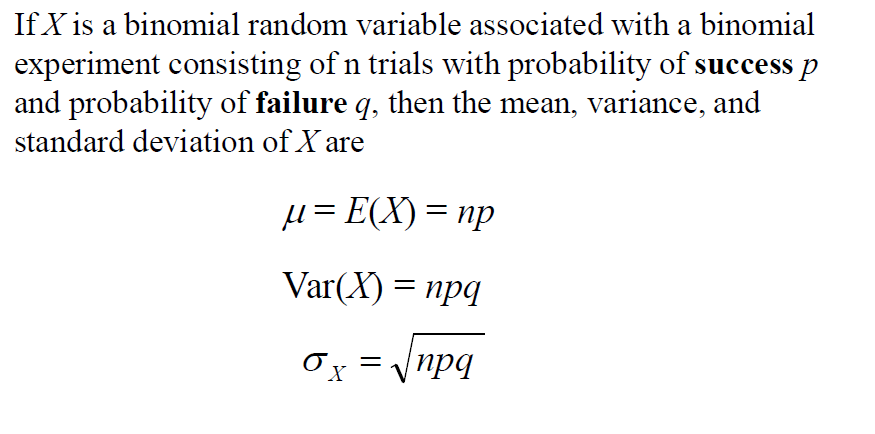
1. Number of trials in the experiment is fixed,

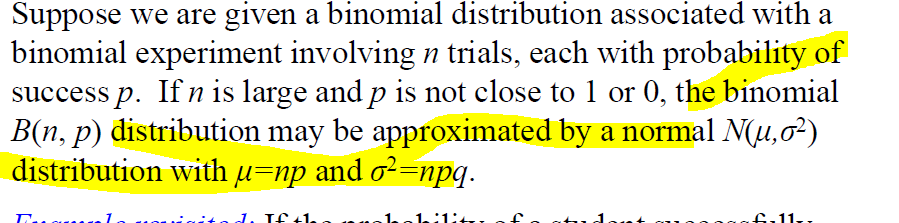
2. The only outcomes are success and failure,

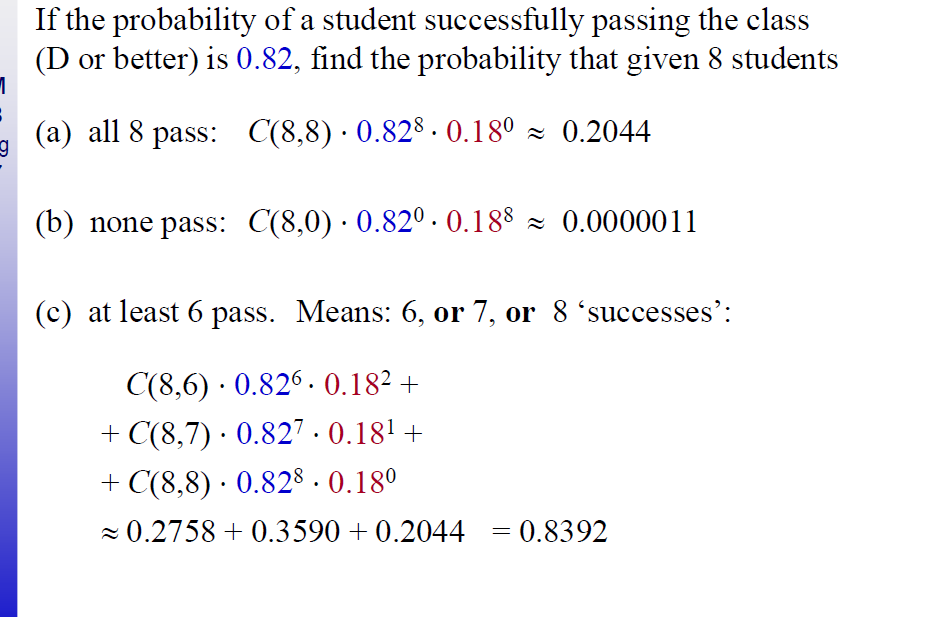
3. In each trial the success probability is the same, and

4. The trials are independent of each other.









**Suppose you toss a fair coin 12 times. What is the probability of getting exactly 7 Heads.**

|  |
| --- |
| ***Questio 5***  Five cards are drawn, with replacement, from a standard 52-card  deck. If drawing a club is considered a success, find the mean,  variance, and standard deviation of X (where X is the number of  successes).  Number of trials are fixed  It is a binomial distribution as each card is either a club or not.  And since it is done with replacement every experiment is independent of other experiment  And probability in each experiment is same  So p =1/4 , q=1-1/4 , mean = np, var = npq |

|  |
| --- |
| **Ten percent of computer parts produced by a certain supplier are defective. What is the probability that a sample of 10 parts contains more than 3 defective ones?**  n= 10,x=3 ,P = probability of success - .10  P(X > 3) = 1 − F(3)  1- pbinom(3,10,1/10)  If the question is atleast 3 parts  P(x>=3) 1-F(2) |

|  |
| --- |
| 1. dbinom(8,8,0.82) 2. dbinom(0,8,0.82) 3. atleast 6 pass =P(x>=6) = 1- Prob(x<=5)   1-pbinom(5,8,0.82) |

|  |
| --- |
| The Stanley Cup winner is determined in the final series between two teams. The  first team to win 4 games wins the Cup. Suppose that Dallas Stars advance to the final series, and they have a probability of 0.55 to win each game, and the game results are independent of each  other. Find the probability that  a) Dallas Stars wins the Stanley Cup  b) seven games are required to determine the Cup winner  (Hint: Without loss of generality, you can assume that the series continues until 7 games are played, even if the Cup winner is determined earlier. This ”change of Stanley Cup rules” will not change the answer to the problem!)  Suppose that the series continues until Dallas Stars win 4 games, even if the other rival wins the Cup earlier.   1. n = 7 , p = 0.55,x=3   P( Dallas wins ) = P(X >=4) = 1 - F(3) = 0.6083     1. pbinom(x,n,p) 2. First team to win 4 games will win the tournament so if 7 games are required to determine te game winner that means no team must have won 4 games   n = 6 ; p = 0.55 ;x=3  pbinom(3,n,p) - pbinom(2,n,p) |

|  |
| --- |
| **Suppose you toss a fair coin 12 times. What is the probability of getting exactly 7 Heads.**  X=7  N=12  P=1/2  dbinom(7,12,1/2)  atmost 7 heads  pbinom(7,12,1/2) |

|  |
| --- |
| ) An internet search engine looks for a certain keywordin a sequence of independent web sites. It is believed that 20% of the sites contain this keyword.  c) Out of the first 10 websites, let Y be the number of sites that contain the keyword. Find the distribution of Y .  d) Compute the expected value and the standard deviation of Y .  e) Compute the probability that at least 5 of the first 10 websites contain the keyword.  f) Compute the probability that the search engine had to visit at least 5 sites inorder to find the first occurrence of a keyword.  c) Y is Binomial(n = 10; p = 0:2).  d) E(Y ) = np = 2 and Std(Y ) =sqrt( np(1-p))  e) From the Binomial Table, P(X >=5) = 1- F(4) = 1 - 0:9672 = 0.0328 .  f) P(Y >= 5) = (1 – p(4) = 0.4096 : |

|  |
| --- |
|  |

# Exponential

|  |  |  |
| --- | --- | --- |
| Exponential | Expected value =1/ λ | Variance=1/ λ2 |

* Interarrival time forms the exponential distribution

|  |
| --- |
| Problem 3: (5 pts) Let X~ Exp(0.3). Write a single line in R that calculates P(1 < X < 4) and execute it. Paste (or rewrite) the textual content of your RStudio console as the solution.    P(1<x<4) = P(X<=4) – P(X<=1)  pexp(4,rate=0.3)-pexp(1,rate=0.3)  [1] 0.439624 |

|  |
| --- |
| Malfunctions in a particular type of electronic device are known to follow an exponential distribution with a mean time of 24 months until the device malfunctions. What is the probability that a randomly selected device will malfunction within the first 6 months?  Expected value =1/lambda  Lambda =1/expected value  You can answer the question using the pexp function  > pexp(6, 1/24) |

|  |
| --- |
| Malfunctions in a particular type of electronic device are known to follow an exponential distribution with a mean  time of 24 months until the device malfunctions.  The probability of malfunction within six months is 0.22 (22%).  What is the probability that a randomly selected device will last more than 5 years (60 months) without malfunction?  > pexp(60, 1/24, lower.tail=F)  Or 1-pexp(60,1/24)  [1] 0.082085  The probability that it will last more than 5 years is approximately 0.08, or 8 percent. |

|  |
| --- |
| Malfunctions in a particular type of electronic device are known to follow an exponential distribution with a mean  time of 24 months until the device malfunctions. After how many months will 40 percent of the devices already  have malfunctioned?  E(x) = 1/lambda  To find the length of time within which 40 percent of devices will have malfunctioned, use the command:  > qexp(0.4, 1/24)  [1] 12.25981  So 40 percent of devices will malfunction within 12.3 months. |

# Poisson

* When we have to find how many have arrived between a particular time then it becomes poisson
* E(x) =λ(end time -start time ) ( expected number of arrivals is lam\*(et-st))

The Poisson random variable satisfies the following conditions:

1. The number of successes in two disjoint time intervals is independent.
2. The probability of a success during a small time interval is proportional to the entire length of the time interval.

Apart from disjoint time intervals, the Poisson random variable also applies to disjoint regions of space.

* the number of deaths by horse kicking in the Prussian army (first application)
* birth defects and genetic mutations
* rare diseases (like Leukemia, but not AIDS because it is infectious and so not independent) - especially in legal cases
* car accidents
* traffic flow and ideal gap distance
* number of typing errors on a page
* hairs found in McDonald's hamburgers
* spread of an endangered animal in Africa
* failure of a machine in one month

|  |
| --- |
| **Problem 4**: Arrival of cars at New Jersey Turnpike toll booth on Saturdays between 3AM and 5AM is modeled as a *Poisson Arrival* process with a rate 2.1 per minute. Let *X* be a random variable that counts the arrivals between 4:00AM and 4:10AM.  (a)(5 pts) Is this variable finite discrete, infinite discrete, or continuous?  Infinite discrete  (b)(5pts) What is the distribution of X?  It is a **Poisson** distribution since all the cars are independent events with Pois(λ ∆t) = Pois(21)  Probability distribution = P(N=k) = e -μ μk/k! = e -2121 k/k !  (c)(5 pts) What is the expected value of X?  E(x) =λ(end time -start time ) = 2.1\*10 =21 |

|  |
| --- |
| 2= mu = λ ∆t  ∆t =5  Lambda =0.4  Atmost 5 errors = Pois(5, mu)  So deltaT =15  Mu = 15 \*0.4 =6  Ppois(5,6) |

|  |
| --- |
| The switchboard in a Denver law office gets an average of 2.5 incoming phonecalls during the noon hour on Thursdays. Experience shows that the existing lunch hour staff can handle up to 5 calls in an hour. They seem to be well covered. But just to test the edges, what is the actual chance that 6 calls will be received during the lunch period, on some particular Thursday?  dpois(6,2.5\*1) |

|  |
| --- |
| Suppose that the number of inquiries arriving at a certain interactive system follows  a Poisson distribution with arrival rate of 12 inquiries per minute.  Find the probability of 10 inquiries arriving  a) in a 1-minute interval;  b) in a 3-minute interval.  c. What is the expectation and the variance of the number of arrivals during each of these  intervals?  λ = 12/min  mu = λ ∆t = 12\*1 =12  x = 10  dpois(10,12)      λ = 12/min  mu = λ ∆t = 12\*3 =36  x = 10  dpois(10,36)   1. . For Poisson distribution with parameter mu, we have E(X) = V ar(X) = mu. |

|  |
| --- |
| To find 3 more flaws  Probability=*P*(*X*≥3) = 1-ppois(2,2.3) |

|  |
| --- |
|  |

|  |
| --- |
| To calculate the rate mu = lambda delta t  3 = lambda (20)  Lambda =3/20  Ppois(1,3/20) |
|  |

|  |
| --- |
| **Problem 4**: Arrival of cars at New Jersey Turnpike toll booth on Saturdays between 3AM and 5AM is modeled as a *Poisson Arrival* process with a rate 2.1 per minute. Let *X* be a random variable that counts the arrivals between 4:00AM and 4:10AM.  (a)(5 pts) Is this variable finite discrete, infinite discrete, or continuous?  Infinite discrete  (b)(5pts) What is the distribution of X?  It is a **Poisson** distribution since all the cars are independent events with Pois(λ ∆t) = Pois(21)  Probability distribution = P(N=k) = e -μ μk/k! = e -2121 k/k !  (c)(5 pts) What is the expected value of X?  E(x) =λ(end time -start time ) = 2.1\*10 =21 |

**Problem 4**: Arrival of cars at New Jersey Turnpike toll booth on Saturdays between 3AM and 5AM is modeled as a *Poisson Arrival* process with a rate 2.1 per minute. Let *X* be a random variable that counts the arrivals between 4:00AM and 4:10AM.

(a)(5 pts) Is this variable finite discrete, infinite discrete, or continuous?

Infinite discrete

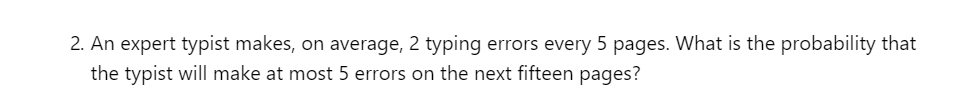
(b)(5pts) What is the distribution of X?

It is a **Poisson** distribution since all the cars are independent events with Pois(λ ∆t) = Pois(21)

Probability distribution = P(N=k) = e -μ μk/k! = e -2121 k/k !

(c)(5 pts) What is the expected value of X?

E(x) =λ(end time -start time ) = 2.1\*10 =21



2= mu = λ ∆t

∆t =5

Lambda =0.4

Atmost 5 errors = Pois(5, mu)

So deltaT =15

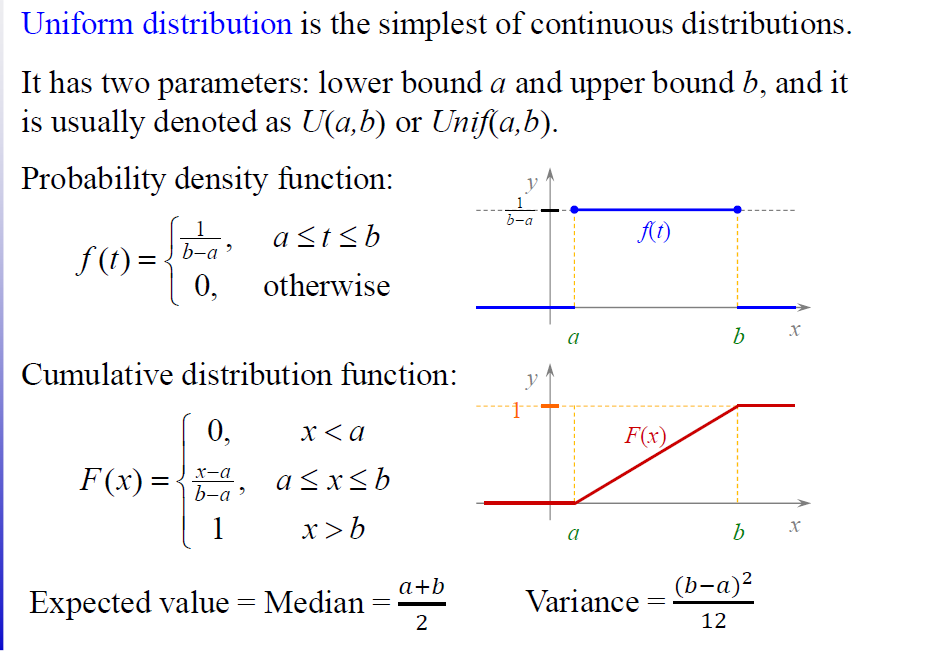
Mu = 15 \*0.4 =6

Ppois(5,6)

2. In any 15-minute interval, there is a 20% probability that you will see at least one shooting star. What is the proba- bility that you see at least one shooting star in the period of an hour?

* 1-(0.8)^4. Or, we can use Poisson processes

# Uniform distribution



Consider the toss of a single die. The outcome of this toss is a random variable that can take on any of six possible values: 1, 2, 3, 4, 5, or 6. Each of these outcomes is equally likely to occur. The probability that any particular outcome will occur is equal to 1/6. Therefore, the outcome from the toss of a single die has a uniform distribution

. Graphically, a uniform distribution has no clear peaks.

#### 6. Given draws from a normal distribution with known parameters, how can you simulate draws from a uniform distribution?

* plug in the value to the CDF of the same random variable

# Regression

#### 15. Let’s say you have a very tall father. On average, what would you expect the height of his son to be? Taller, equal, or shorter? What if you had a very short father?

* Shorter. Regression to the mean

# Probability Questions

#### **I write a program should print out all the numbers from 1 to 300, but prints out Fizz instead if the number is divisible by 3, Buzz instead if the number is divisible by 5, and FizzBuzz if the number is divisible by 3 and 5. What is the total number of numbers that is either Fizzed, Buzzed, or FizzBuzzed?**

100+60-20=140

**If you know for certain that a friend of yours has 2 children and that atleast one of them is a boy, what is the probability that the other is also a boy?**

I think it's 1/3, because the possibilities are: Boy Boy, Boy Girl, Girl Boy, Girl Girl. There are three possibilities with at least one girl. Out of these three, only one has two girls.  
  
**There are 6 marbles in a bag - 1 is white. You reach in the bag 100 times. After drawing a marble, it is placed back in the bag. What is the probability of drawing the white marble at least once?**

P(White at leas **once**)=1-P(None over 100times)=1-[P(Not white)^100]=1-[(5/6)^100]

**Bobo the amoeba has a 25%, 25%, and 50% chance of producing 0, 1, or 2 offspring, respectively. Each of Bobo’s descendants also have the same probabilities. What is the probability that Bobo’s lineage dies out?**

p=1/4+1/4p+1/2p^2 => p=1/2

#### **In any 15-minute interval, there is a 20% probability that you will see at least one shooting star. What is the proba- bility that you see at least one shooting star in the period of an hour?**

1-(0.8)^4.

Or, we can use Poisson processes

**Tossing a coin ten times resulted in 8 heads and 2 tails. How would you analyze whether a coin is fair? What is the p-value?   
In addition, more coins are added to this experiment. Now you have 10 coins. You toss each coin 10 times (100 tosses in total) and observe results. Would you modify your approach to the the way you test the fairness of coins?**

H0 (null hypothesis): the coin is fair

H1 (alternative): the coin is not fair

P(k=8)=10C8\*0.5^(8) \*0.5^(2) = 10\*9/2\*(0.5)^10 = 45/1024 = 0.045( approx.)

Now we can choose the alpha = 0.05 or 0.01

If we choose 0.05 then coin i

# What is the probability of getting 50 heads when 100 coins are tossed?

1. Let’s assume the coins are fair. Now toss the 100 coins and line them up.

We want to know the probability of getting exactly 50 heads.

First, realize that from the 100 positions, you can count how many ways there are to get 50 heads. Since they can be in 50 positions, there are (10050)(10050) ways of arranging 100 coins with exactly 50 heads.

Now we need the total number of possible coin flips. Each flip has 2 possibilities and there are 100 flips, so there are 21002100 possible flips.

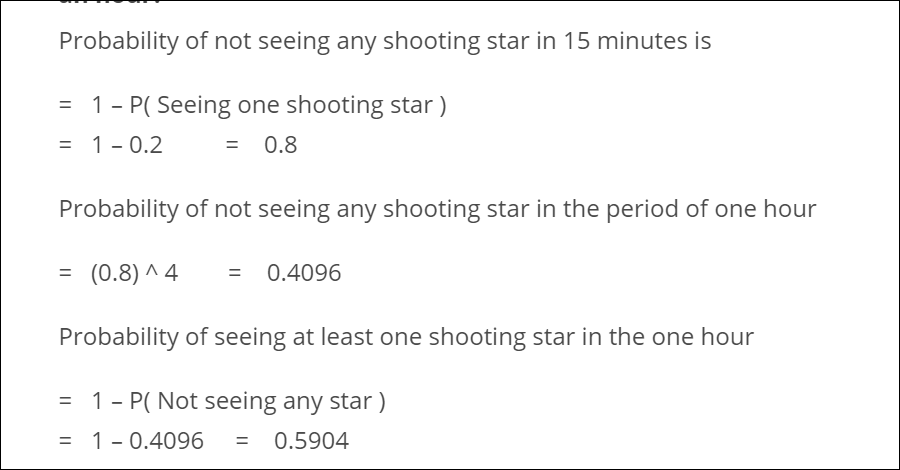
Thus, the probably of flipping exactly 50 heads out of 100 is the quotient of these is (10050)2100=7.96%(10050)2100=7.96%.

nC50 (p)50 q(50)

[What is the probability that if you flip a coin 100 times you will get exactly 50 heads and 50 tails?](https://www.quora.com/What-is-the-probability-that-if-you-flip-a-coin-100-times-you-will-get-exactly-50-heads-and-50-tails)

100C50(0.5^50)(0.50^50) = [100!/(50!\*50!)]\*(0.5^50)(0.50^50)

## Throw a fair die 100 times, what is the probability that   you will get more than 75 tails?

32. In any 15-minute interval, there is a 20% probability that you will see at least one shooting star. What is the proba­bility that you see at least one shooting star in the period of an hour? 

## 2. Does the frequentist approach always give the same result as the Bayesian approach?

No. The frequentist approach depends on how the hypothesis is defined while our prior faiths are updated by Bayesian approach. Therefore the frequentist approach might result in an opposite conclusion if the hypotheses are stated in a different manner. Thus the two approaches may not yield the same results.

## 5. A jar contains 4 marbles. 3 Red & 1 white. Two marbles are removed with replacement after each draw. Find the probability that the same color marble is drawn twice?

Suppose that the marbles are of the same color.Then the calculation will be 3/4 \* 3/4 + 1/4 \* 1/4 = 5/8.

## 6. In a website offering dating services, users can select 5 out of 24 adjectives to describe themselves. A match is said to be between two users if they match on at least 4 adjectives.

If Alice and Bob randomly pick adjectives, what is the probability that they are able to find a match?

The probability here is calculated as

24C5\*(1+5(24-5))/24C5\*24C5 = 4/1771

7. There is 0.1% chance of picking up a coin with both heads, and a 99.9% chance that you pick up a fair coin. A coin is flipped and it comes up heads 10 times. What’s the chance that the fair coin was picked, given the information that you observed?

The possible events are : F = "picked a fair coin", T = "10 heads in a row"

· P(F|T) = P(T|F)P(F)/P(T) (Bayes theorem) -(1)

· P(T) = P(T|F)P(F) + P(T|¬F)P(¬F) (total probabilities) -(2)

Injecting (2) in (1):

P(F|T) = P(T|F)P(F)/(P(T|F)P(F) + P(T|¬F)P(¬F)) = 1 / (1 + P(T|¬F)P(¬F)/(P(T|F)P(F)))

= 1/(1 + 0.001 \* 2^10 /0.999)

With 210 ≈ 1000 and 0.999 ≈ 1 this is approximately equal to ½

8. If a life insurance company sells a $240,000 life insurance policy with a one year term to a 25-year old lady for $210, the probability that she survives the year is .999592. Find the expected value of this policy for the insurance company?

Probability that company loses the money, P(company does not loses the money ) = 0.99592

Probability that company doesn’t lose the moneyP(company lose the money ) = 0.000408

The amount of money company loses in case of loss = 240,000 – 210 = 239790

The money gained is $210

Expected money the company should give = 239790\*0.000408 = 97.8

Expect money company receives = 210.

Therefore the required value = 210 – 98 = $112

9. Alice has 2 children, one of which is a girl. In what probability will the other child be also a girl?

Assuming there are an equal number of males and females in the world, the outcomes for two kids can be {BB, BG, GB, GG}

Since it is given that one of them is a girl, BB option can be removed. Therefore the sample space has 3 options. In those, only one fits the second condition. Therefore the probability that the second child will be a girl too is 1/3.

## 10. In a class of 30 students, what is the probability that two of the students have their birthday on the same (assuming that it is not a leap year)?

## An example of a favourable event would be students with birthday 3rd Jan 1998 and 3rd Jan

## 1997.

The total number of possible combinationsfor no two persons to have the same birthday in a class of 30 is 30 \* (30-1)/2 = 435.

Now, a year has 365 days (if not a leap year). Thus, the probability of two personsto have a different birthday would be 364/365. Out of 870 possible combinations, no two people having the same birthday is (364/365)435 = 0.303.

Thus, the probability of two people having their birthdays on the same date would be 1 – 0.303 = 0.696

## 12. A fly has a lifetime of between 4-6 days. What is the probability that the fly will die in exactly 5 days?

The continuous probabilities here form a mass function. The probability of the event is calculated by finding the area under the curve. Here since we should calculate the probability of the fly expiring at exactly 5 days – the area under the curve will be 0.

*11. According to hospital records, 75% of patients suffering from a disease die from that disease. Find out the probability that 4 out of the 6 randomly selected patients survive.*

This has to be a binomialas there are only 2 outcomes – death or life. Here n =6, and x=4. p=0.25(probability if life) q = 0.75(probability if death)

P(X) = nCx\*p\*q\*(n-x) = 6C4\* (0.25)\*4\*(0.75)\*2 = 0.03295

# Case Study

## Alice and Bob take turns in rolling a fair dice. Whoever gets "6" first wins the game. Alice starts the game. What are the chances that Alice wins.

Let R1 = the value of the die on roll 1, R2 = value on roll 2, etc. Since Alice rolls first, she will win if R1=6.  
  
But what if R1 is not 6? Then Alice will still win if (R1 !=6) and (R2 !=6) (so that Bob doesn't win) and (R3 = 6) (so that Alice wins).  
  
Similarly, what if R3 is not 6? Then Alice will still win if (R1 !=6) and (R2 != 6) and (R3 != 6) (by assumption) and (R4 != 6) (so that Bob doesn't win) and (R5 = 6).  
  
It should be obvious that we can repeat this pattern forever. Also, each of these events are mutually exclusive, since (for example) it's impossible for (R1 = 6) (the first possibility) and (R1 != 6 and R2 != 6 and R3=6) (our second possibility), due to the first one having R1 = 6, and the second one having R1 != 6. Therefore we can just add up the probabilities (using the rule P(A or B) = P(A) + P(B) - P(A and B)).  
  
Each time is one roll which must equal 6, which obviously has a probability of 1/6. Similarly, at the n'th step there are 2n rolls which cannot equal 6, each occurring with a probability of 5/6. Thus adding up all the probabilities gives the infinite sum:  
  
(1/6) \* sum for n from 0 to infinity of (5/6)^(2\*n) = (1/6) \* sum for n from 0 to infinity of (25/36)^n  
  
This is an infinite geometric series. An infinite geometric series going from 0 to infinity adding up r^n has a sum equal to 1 / (1 - r). Thus in this case we get:  
  
(1/6) \* (1 / (1 - 25/36)) = 6 / 11.

## If you draw 2 cards from a shuffled 52 card deck, what is the probability that you'll have a pair?

(4/52)\*(3/51)\*13

## There's a game where you are given two fair six-sided dice and asked to roll. If the sum of the values on the dice equals seven, then you win $21. However, you must pay $5 to play each time you roll both dice. Do you play this game? And in follow-up: What is the probability of making money from this game?

In order to win $21, you need to come up with (1,6), (2,5), (3,4), (4,3), (5,2), (6,1) on two dice. There are 6\*6=36 possible outcomes, so the chance of wining in 6/36=1/6, in which case you actually earn 21-5=$16. The chance of loosing is of course 5/6 and you earn -$5. So, your expected earning is:  
  
1/6\*16 + 5/6\*(-5) = -1.5

## We have two options for serving ads within Newsfeed

## : 1 - out of every 25 stories, one will be an ad

## 2 - every story has a 4% chance of being an ad

## For each option, what is the expected number of ads shown in 100 news stories? If we go with option 2, what is the chance a user will be shown only a single ad in 100 stories? What about no ads at all?

 The expected number of ads for each option:  
  
     Let the r.v. X be the number of ads in 100 stories.  
     - Option 1: X = 4, constant, so E(X) = E(4) = 4.  
     - Option2: X is binomial, with n =100, p = 0.04, so E(X) = n\*p = 100\*(0.04) = 4.

b. The probability that a user sees only 1 ad out of 100 stories (using Option 2):  
  
     For Option 2, X is binomial, with n=100, p=0.04. Then, we use the prob. mass function to get  
  
            P(X=1) = (100 choose 1) \* (0.04)^1 \* (0.96)^99 = 4 \* (0.96)^99, which is approx. 0.074 (~7%).  
  
c. The probability that a user sees zero ads in 100 stories (using Option 2):  
  
          X is binomial, so P(X=0) = (100 choose 0) \* (0.04)^0 \* (0.96)^100 = (0.96)^100, which is  
         approximately 0.0169 (~1.7%).

## If a PM says that they want to double the number of ads in

## Newsfeed, how would you figure out if this is a good idea or not?

0. Ask what Biz Value the PM is looking for  
1. What users will do  
DAU/MAU goes down, App Uninstalls, DAU,MAU goes down. Session Time Goes Down. Start Flagging ads as unhelpful/usefulness goes down. Cognitive load will go up. AD CTRs will go down. Ad Pricing can go coz of more supply. Hide Ad/Report Ad/Why I am seeing this metric  
2. What people will say they will do  
People can say if you start showing me more ads. I may leave.  
Some can say, if you show me Ads that are relevant, I will stick around for longer  
App Reviews  
Feedback on Ad  
  
3. Run an A/B Test – Decide how much % is comfortable looking at it.  
4. People will be blogging about it. Be Careful about it. Can be a massive PR failure  
5. Effort Required to Change the Code  
Recommendations. Use Data to show relevant ads. Maybe, there is a balance between few getting more ads, few getting less ads

## What is the expectation of the variance?

# Bayes Theorem

You randomly draw a coin from 100 coins — 1 unfair coin (head-head), 99 fair coins (head-tail) and roll it 10 times. If the result is 10 heads, what is the probability that the coin is unfair?

Basically we need to find P(unfair|10 heads) ??

By Bayes theorem -:

P(unfair|10 heads) = P(10 heads|unfair) \* P(unfair) / **p(10 head)**

**And by conditional probabililty,**

***P(B) = P(B|A)P(A) + P(B|Ac)P(Ac)*,**

P(unfair|10 heads) =

P(10 heads|unfair) \* P(unfair) /(P(10heads|unfair)\*P(unfair)+P(10heads|fair)\*P(fair))

Suppose a voter poll is taken in three states. In state A, 50% of voters support the liberal candidate, in state B, 60% of the voters support the liberal candidate, and in state C, 35% of the voters support the liberal candidate. Of the total population of the three states, 40% live in state A, 25% live in state B, and 35% live in state C. Given that a voter supports the liberal candidate, what is the probability that she lives in state B?

By Bayes's formula,

*P(Voter lives in state B|Voter supports liberal candidate) =*

*P(Voter supports liberal candidate|Voter lives in state B)P(Voter lives in state B)/*

*(P(Voter supports lib. cand.|Voter lives in state A)P(Voter lives in state A) +*

*P(Voter supports lib. cand.|Voter lives in state B)P(Voter lives in state B) +*

*P(Voter supports lib. cand.|Voter lives in state C)P(Voter lives in state C))*

= (0.60)\*(0.25)/((0.50)\*(0.40) + (0.60)\*(0.25) + (0.35)\*(0.35))

= (0.15)/(0.20 + 0.15 + 0.1225) = 0.15/0.4725 = 0.3175.

## You're about to get on a plane to Seattle. You want to know if you should bring an umbrella. You call 3 random friends of yours who live there and ask each independently if it's raining. Each of your friends has a 2/3 chance of telling you the truth and a 1/3 chance of messing with you by lying. All 3 friends tell you that "Yes" it is raining. What is the probability that it's actually raining in Seattle?

**Bayesian stats: you should estimate the prior probability that it's raining on any given day in Seattle. If you mention this or ask the interviewer will tell you to use 25%. Then it's straight-forward:**

P(raining | Yes,Yes,Yes) = Prior(raining) \* P(Yes,Yes,Yes | raining) / P(Yes, Yes, Yes)

P(Yes,Yes,Yes) = P(raining) \* P(Yes,Yes,Yes | raining) + P(not-raining) \* P(Yes,Yes,Yes | not-raining) = 0.25\*(2/3)^3 + 0.75\*(1/3)^3 = 0.25\*(8/27) + 0.75\*(1/27)  
  
P(raining | Yes,Yes,Yes) = 0.25\*(8/27) / ( 0.25\*8/27 + 0.75\*1/27 )  
  
\*\*Bonus points if you notice that you don't need a calculator since all the 27's cancel out and you can multiply top and bottom by 4.  
  
P(training | Yes,Yes,Yes) = 8 / ( 8 + 3 ) = 8/11

**Second approach**

I thought about this a little differently from a non-bayes perspective.  
  
It's raining if any ONE of the friends is telling the truth, because if they are telling the truth then it is raining. If all of them are lieing, then it isn't raining because they told you that it was raining.  
  
So what you want is the probability that any one person is telling the truth.  
  
Which is simply 1-Pr(all lie) = 26/27

## Suppose N students participate in a coin flip experiment, when they get heads they stop, when they get tails they keep going. All students will stop after the second trial no matter the results. Y is the binary indicator of whether they claim they cheated in the experiment. Estimate how many students cheat in this experiment.

I think the logic should be: The probability of saying yes given a student cheat should be 3/4 (1/2+1/2\*1/2). The probability of saying yes given a student not cheat should be 1/4 (1/2\*1/2). Then we have p(yes) = 0.3. If we set the probability of cheating to be x. Then using Bayes theorem, we have: 3/4\*x+1/4\*(1-x) = 0.3. x should equal to 0.1. The probability of cheating is only 10%.

# Link

<https://www.coursera.org/lecture/inferential-statistics-intro/inference-for-comparing-two-independent-means-wkwlZ>