Chapter 6 & 7
Probability Distribution II

경영대학 재무금융학과
윤선중

Objectives

- 확률변수
  - 이산확률분포(Discrete Random Variables): 셀 수 있는 확률변수
  - 연속확률분포(Continuous Random Variables):셀 수 없는 경우의 수
- 이산확률변수
  - 분포의 대표값
    - 기대치 (Expected Value), 분산 (Variance), 표준편차 (Standard Deviation)
  - 이변량 확률분포 (Bivariate Distribution)
  - 이항분포(Binomial Distribution)와 포아송 분포 (Poisson Distribution)

Objectives (2)

- 연속확률변수
  - 무한개의 값을 가질 수 있으므로 어느 하나의 값과 같을 확률은 0
  - 임의의 구간의 확률
    - 확률밀도함수(Probability Density Function)의 면적
  - 정규확률변수 (Standard Normal Distribution)
    - 정규확률분포표의 사용
  - 지수분포 (Exponential Distribution)
  - Student-t 분포
  - 카이제곱분포 (Chi-squared Distribution)

I. Binomial Probability Distribution

Binomial Experiment

- 베르누이 시행 (Bernoulli trial)
  - 확률실험에서 결과가 오직 두 가지로 구분
  - 각 시행에서 성공할 확률은 p, 실패할 확률은 1-p 로 일정
  - 각각의 시행은 서로 독립
  - 예시
    - 동전 던지기
    - 주사위를 던졌을 때, 짝수/홀수 관찰
    - 내일의 주가의 상승/ 하락
- 이항실험
  - 베르누이 시행을 고정된 횟수(n)만큼 반복하여 시행
  - 예시
    - 동전 던지기 10회
    - 주사위를 10회 던져서 짝수/홀수 관찰
    - 공화당/민주당 두 후보에 대한 선거
    - 100일간 일별 주가의 상승/하락 관찰

Binomial Random Variable

- 이항 확률 변수
  - 이항 실험에서 나타나는 성공 또는 실패 횟수
  - 이산 확률 변수의 일종
- 예시
  - 동전을 3회 던질 때, 앞면이 나타나는 횟수
  - 100명이 민주당/공화당 후보에게 투표한다고 할 때, 민주당 후보의 득표수
  - 100일 간 주가의 일별 수익률을 관찰할 때, 주가의 상승일 수
  - 계량 재무분석 수업을 들은 학생 중 F학점을 받은 학생의 수

Calculation of Probability

- 이항확률분포
  - 이항확률변수의 값과 그 확률을 나타내는 표, 그래프, 또는 공식
  - 예제
    - 총 두 문제의 5지 선다 객관식 시험을 치르는 A
    - A씨가 맞추는 문제의 개수를 이항확률 변수 X라고 정의하면 X의 확률 분포는?
- 이항확률변수의 확률
  - 성공확률 p, 실패확률 1-p, 총 시행횟수 n번인 이항 실험에서, 총 성공횟수를 X라고 하면
  - X: 이항 확률변수;   X~ B(n, p)

$$P(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

  - 교제 Appendix B참조

Expected Value & Variance

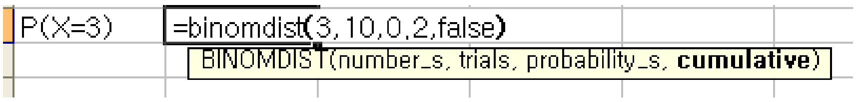
- 이항확률변수의 기대값과 분산
  - X ~ B(n, p)일 때,
  - E(X)=np,   V(X)=np(1-p)
  - 앞의 예제에서 총 5문제의 문제를 푼다고 하면, 정답을 맞출 횟수의 기대값과 분산은?
- 엑셀의 활용
  - X~ B(n, p)일 때,
  - BINOMDIST(x,n,p,false): 이항확률변수의 확률 값
  - BINOMDIST(x,n,p,true): 이항확률변수의 누적 확률 값

Expected Value & Variance

- 예제 6.9: 10문제의 5지선다형 문제에서 정답의 개수를 X라고 하면,
  - n=10 & P(success) = 1/5 = .20

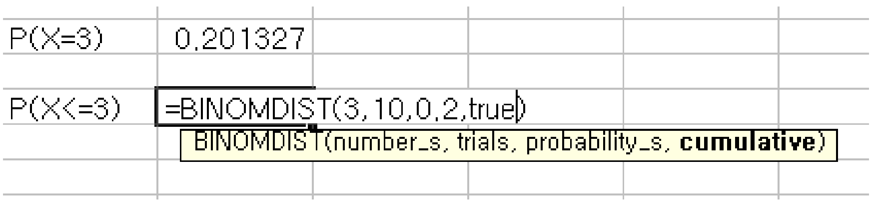
$$P(0) = \frac{10!}{0!(10-0)!}(.2)^0(1-.2)^{10-0}$$
$$= 1(1)(.8)^{10} = .1074$$

$$P(2) = \frac{10!}{2!(10-2)!}(.2)^2(1-.2)^{10-2}$$
$$= 45(.04)(.1678) = .3020$$



Expected Value & Variance

- 누적확률 분포
  - P(X ≤ 4) = P(0) + P(1) + P(2) + P(3) + P(4)



Exercise

- **(1) 보험 계약**
  - **보험회사 직원이 과거 경험으로 볼 때, 고객을 만나 계약을 맺을 확률 =0.1**
  - **하루에 50명의 고객을 만난다고 하면,**
    - 10건의 계약이 성사될 확률은?
    - 하루에 성사될 계약의 기대값과 분산은?
- **(2) 품질 관리**
  - **A 제품의 품질을 검사하기 위해서는 10개의 표본을 추출하여 검사**
  - **이중 불량품이 2개를 초과하지 않아야 제품 전체가 품질 검사를 통과**
  - **제품 전체의 불량률이 10%라고 하면, 제품 전체를 합격으로 판정할 확률은?**

Exercise

- **(3) 항공예약**
  - **A항공사는 예약 후 나타나지 않는 고객들로 인한 손실을 막기 위하여 일반적으로 정원을 초과하여 예약을 받음**
  - **평균적으로 좌석을 예약한 고객의 5%가 나타나지 않는다고 함**
  - **정원 10명의 좌석에 12명의 예약을 받았다고 하면,**
    - 나타난 손님의 수가 정원을 초과할 확률은?
    - 비행기를 타기 위하여 나타난 손님의 수의 기대값과 분산은?
- **(4) 시험**
  - **시험에 패스하기 위해서는 총 10문제 중 7문제 이상을 맞추어야 함**
  - **개별 문제를 맞출 확률은 모두 0.8로 같다고 함**
  - **총 100명의 학생들이 시험을 본다면,**
    - 60명 이상의 학생들이 패스할 확률은?
    - 패스하는 학생 수의 기대값과 분산은?

Binomial Table

n = 10

k	0.01	0.05	0.1	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.75	0.8	0.9	0.95	0.99
0	0.9044	0.5987	0.3487	0.1074	0.0563	0.0282	0.0060	0.0010	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1	0.9957	0.9139	0.7361	0.3758	0.2440	0.1493	0.0464	0.0107	0.0017	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000
2	0.9999	0.9885	0.9298	0.6778	0.5256	0.3828	0.1673	0.0547	0.0123	0.0016	0.0004	0.0001	0.0000	0.0000	0.0000
3	1.0000	0.9990	0.9872	0.8791	0.7759	0.6496	0.3823	0.1719	0.0548	0.0106	0.0035	0.0009	0.0000	0.0000	0.0000
4	1.0000	0.9999	0.9984	0.9672	0.9219	0.8497	0.6331	0.3770	0.1662	0.0473	0.0197	0.0064	0.0001	0.0000	0.0000
5	1.0000	1.0000	0.9999	0.9936	0.9803	0.9527	0.8338	0.6230	0.3669	0.1503	0.0781	0.0328	0.0016	0.0001	0.0000
6	1.0000	1.0000	1.0000	0.9991	0.9965	0.9894	0.9452	0.8281	0.6177	0.3504	0.2241	0.1209	0.0128	0.0010	0.0000
7	1.0000	1.0000	1.0000	0.9999	0.9996	0.9984	0.9877	0.9453	0.8327	0.6172	0.4744	0.3222	0.0702	0.0115	0.0001
8	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9983	0.9893	0.9536	0.8507	0.7560	0.6242	0.2639	0.0861	0.0043
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9990	0.9940	0.9718	0.9437	0.8926	0.6513	0.4013	0.0956

$$P(X = 2) = P(X \le 2) - P(X \le 1) = .6778 - .3758 = .3020$$

$$P(X = k) = P(X \le k) - P(X \le [k-1])$$

$$P(X \ge k) = 1 - P(X \le [k-1])$$

Poisson Distribution

- 정의
  - 이항 확률변수와 같이 유사하게 성공이라 부르는 사건의 발생 횟수
  - 일정한 시간동안 또는 일정한 공간에서 발생하는 성공 횟수
    - 1시간 동안에 도착하는 자동차의 수
    - 한 필지의 옷감에 있는 흠집의 수
    - 고속도로에서 발생하는 사고의 수
  - 포아송 실험의 특성
    - 임의의 일정한 시간구간에서 발생하는 성공횟수는 다른 시간구간과 독립
    - 한번의 성공이 발생할 확률은 모든 시간에서 동일
    - 한번의 성공이 발생할 확률은 시간 구간의 크기와 비례
    - 성공 확률은 시간이 0에 가까워 오며 0으로 근접

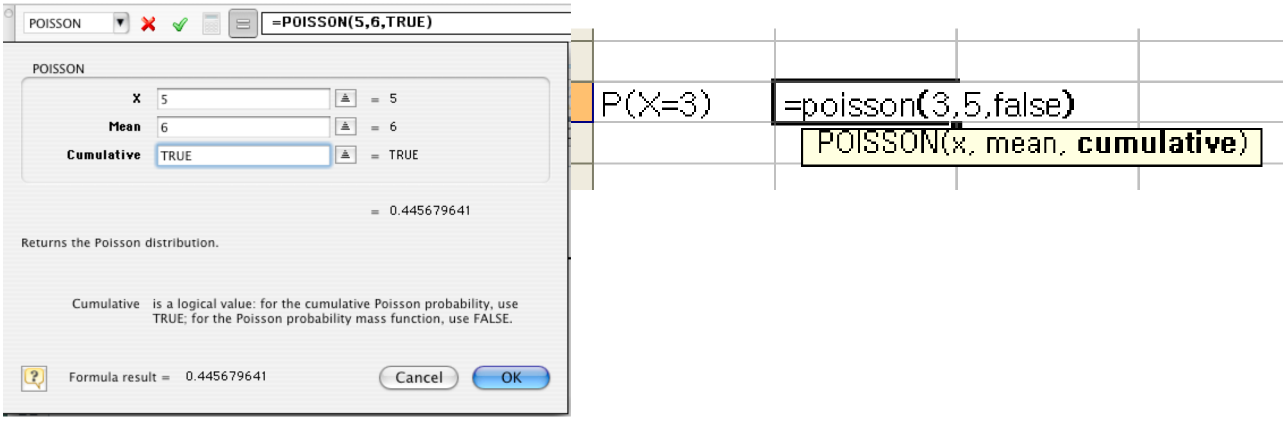
Poisson Distribution

■ 포아송 분포

$$P(x) = \frac{e^{-\mu}\mu^{x}}{x!} \quad for \quad x = 0,\ 1,\ 2, \ldots$$

where μ is the mean number of successes in the interval

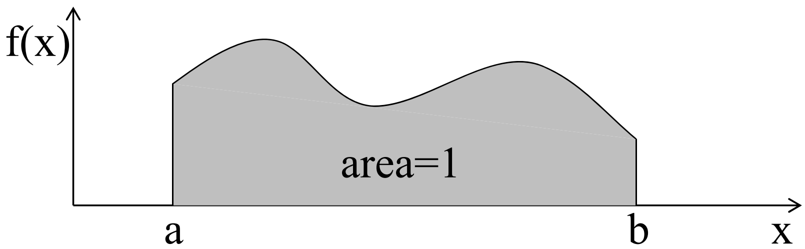
$$E(X) = V(X) = \mu$$



II. Continuous Probability Distributions

■ 확률 밀도 함수

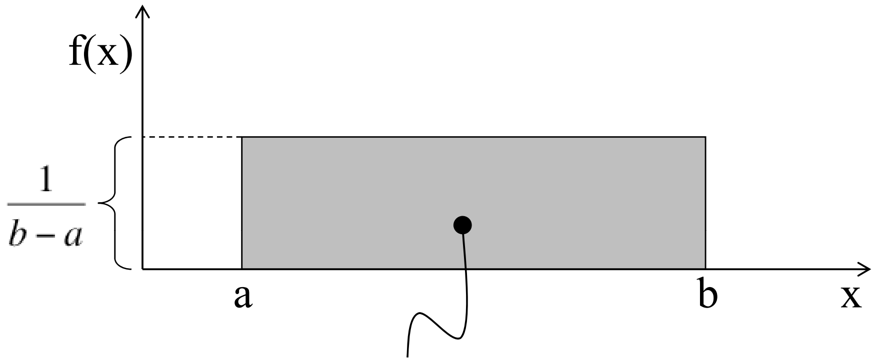
- 이산확률변수와는 다르게 연속확률분포에는 무한히 많은 수의 수가 존재
  - 무한한 수의 값이 있기 때문에 임의의 한점의 확률은 0이다.
- 연속확률과정의 확률을 계산하기 위해서는 "값의 구간 (range of values)" 가 필요
  - 주사위를 던질 때 X=5가 의미가 있지만, 연속확률 하에서 정확히 5는 의미가 없다. : P(X=5)=0
- f(x): 확률밀도함수, f(x)>=0, a와 b사이의 구간의 합(확률)은 항상 1임.



Uniform Distribution

- 일양분포 (uniform distribution)

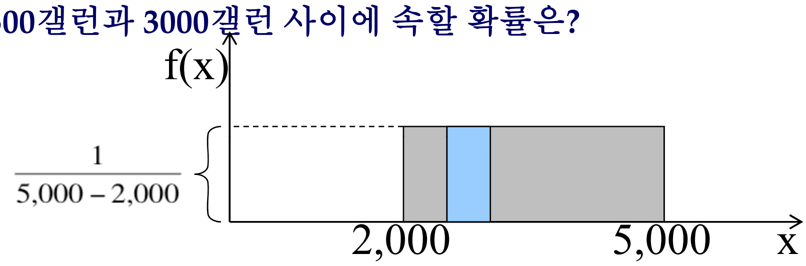
$$f(x) = \frac{1}{b-a}, \quad where \quad a \le x \le b$$



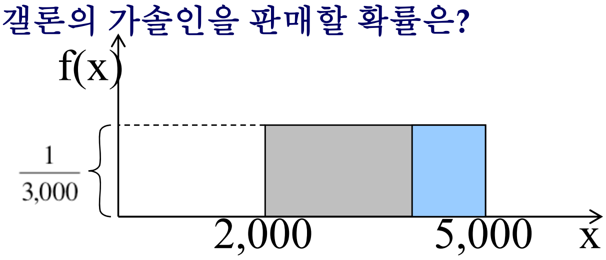
area = width x height = (b – a) x $\frac{1}{b-a}$ = 1

Uniform Distribution

- 예제 7.1: 가솔린 판매량
  - 최소값: 2000, 최대값: 5000의 일양분포
  - 2500갤런과 3000갤런 사이에 속할 확률은?



  - 적어도 4000 갤론의 가솔인을 판매할 확률은?



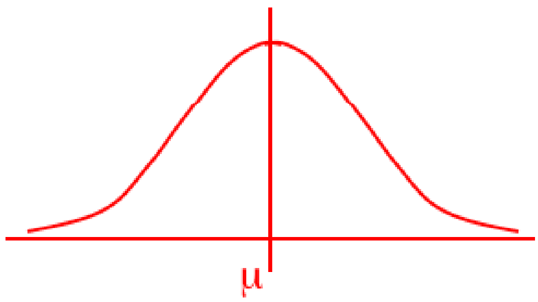
  - 정확히 2500 갤런의 가솔린을 판매할 확률은?

III. Normal Probability Distribution

Introduction

- 정규확률변수
  - X: 평균이 mu, 표분편차가 sigma인 정규확률변수이면,
  - X는 (-inf, inf)인 모든 실수
  - X의 확률밀도함수는 다음과 같이 주어짐
  - 정규분포의 구간: -inf ~ inf

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad -\infty <$$



Two Parameters

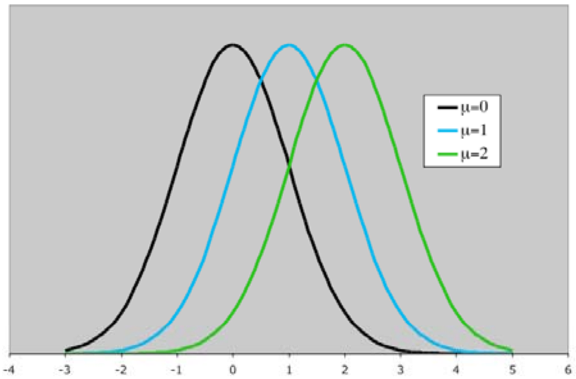
- 정규분포의 표현

  $X \sim N(\mu, \sigma^2)$

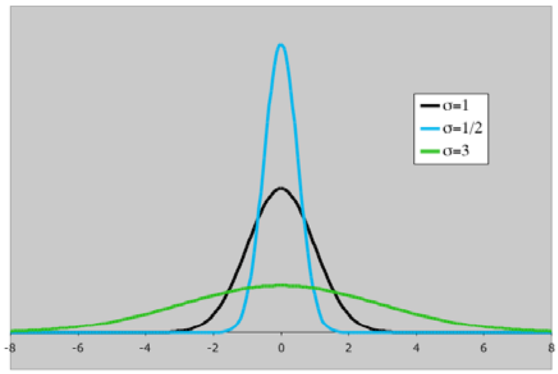
standard deviation   mean

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty < x < \infty$$

Same variance, different means

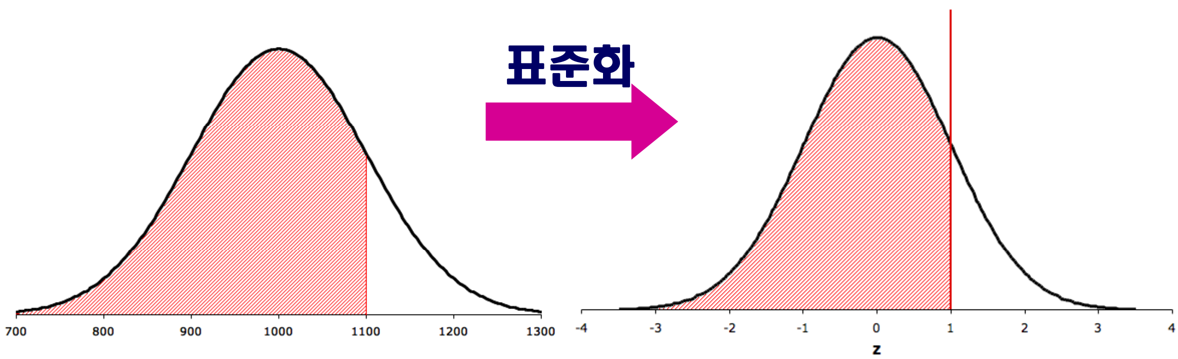


Same mean, different standard deviations



Calculation of Probability

- 예제 7.2: 정규분포를 따르는 가솔린 판매량
  - 평균 1000 갤런, 표준편차 100갤런의 정규분포
  - 현재 가솔린 보유량: 1,100갤런
  - P(X < 1,100)



Standardization

- 표준화
  - $X \sim N(\mu, \sigma^2)$

    $Z = \dfrac{X - \mu}{\sigma} \Rightarrow Z \sim (0,1)$
  - **Z: 표준정규분포(standard normal distribution)**
    - 평균=0; 표준편차 =1

- 예제
  - **X ~ N(1000,100)**
  - $P(X < 1{,}100) \; = P\left(\dfrac{X \text{ - } \mu}{\sigma} < \dfrac{1100 \text{ - } 1000}{100}\right) = \; P(Z < 1.00)$

Standardization

- 표준화과정을 통해

단지 1개의 Table만 필요

- 표준정규분포 표
- Page 286
- Table 7.1

	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0	0.004	0.008	0.012	0.016	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.091	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.148	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.17	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.195	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.219	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.258	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.291	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.334	0.3365	0.3389
1	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.377	0.379	0.381	0.383
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.398	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.437	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.475	0.4756	0.4761	0.4767
2	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.483	0.4834	0.4838	0.4842	0.4846	0.485	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.489
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.492	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.494	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952

Exercise

- 엑셀의 활용
  - **X ~ N(mu, sigma^2)**
  - **NORMDIST (x, mu, sigma, true): 정규확률 변수의 누적확률**
  - **앞서 예에서, P(45< X < 60)?**
    - =normdist(60, 50, 10, true)-normdist(45,50,10,true)
- **(1)** 투자수익률
  - **투자의 평균 수익률은 10%, 표준편차가 5%인 정규분포를 따른다고 함**
  - **손실이 발생할 확률은?**
- **(2)** 성적
  - **계량재무분석의 중간시험 성적은 평균 60점, 표준편차 15점인 정규분포**
  - **상위 10% 안에 포함되면 A+라고 할 때, A+를 받기 위해서는 몇 점 이상?**
  - **Table vs. Norminv(prob, mean, stdev)**

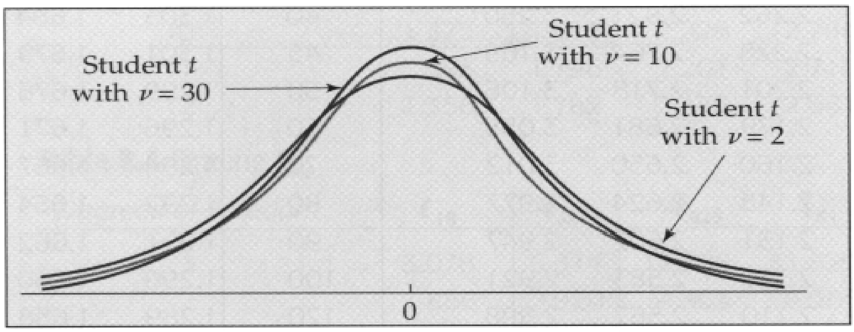
Binomial vs. Normal

- 이항분포 **vs.** 정규분포
  - **이항분포를 따르는 확률변수는 시행 횟수를 무수히 많이 하게 되면, 정규분포로 수렴함**
  - **X ~ B(n, p): n이 커지면 X~N(np, np(1-p))**
- 예제
  - **비행기를 예약한 고객의 10%는 실제 탑승을 하지않는다고 함**
  - **100명이 예약하였을 때, 탑승하지 않는 사람이 5명 이상 10명 이하일 확률**
    - (1) 이항분포확률로 계산하면?
    - (2) 정규분포확률로 계산하면?

Student t Distribution

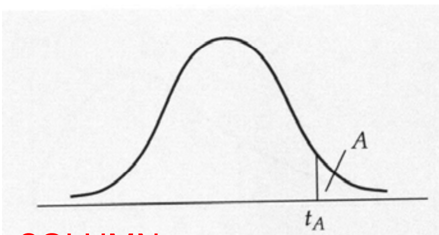
- 정의
  - 표준정규분포와 아주 흡사한 형태의 확률분포
  - t_v(x)
    - v: 자유도 (degree of freedom)이며 2보다 큼
    - t-분포의 평균 =0, 분산 =v/(v-2)
    - v (nu): 자유도 (degree of freedom)
    - Γ (Gamma function): Γ(k)=(k-1)(k-2)…(2)(1)

$$f(t)=\frac{\Gamma[(\nu+1)/2]}{\sqrt{\nu\pi}\,\Gamma(\nu/2)}\left[1+\frac{t^2}{\nu}\right]^{-(\nu+1)/2}$$



Student t Distribution

- **Determining Student *t* Values**
  - $P(t > t_{A,\nu}) = A$
  - **10%, 5%, 2.5%, 1% , 1/2%**에 대응하는 임계치 구하기



Area under the curve value (t_A) : COLUMN

$t_{.05,10}$

$t_{.05,10}=1.812$

Degrees of Freedom : ROW

DEGREES OF FREEDOM	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106

Exercise

■ 예제: X~t(5)

- P(X>1.476)=?
  - (1)엑셀의 활용: tdist (1.476, 5, 1)
  - (2)확률표의 활용

df	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898

Exponential Distribution

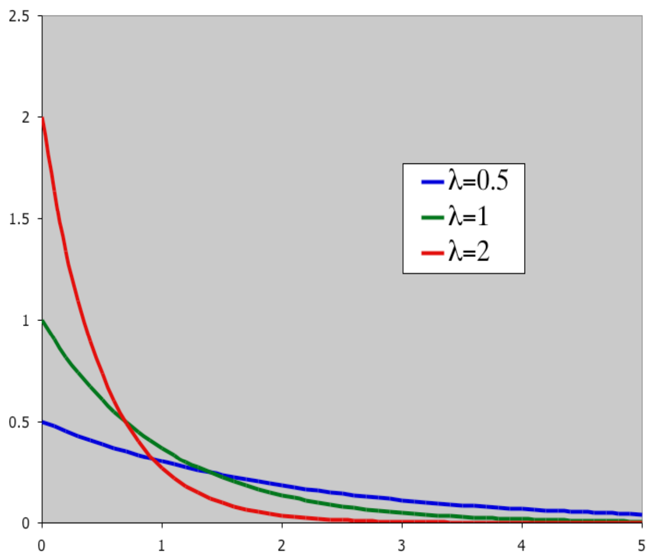
■ 정의

- 확률변수 X의 확률밀도 함수가 다음과 같이 주어지면 X는 지수분포

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

- 지수분포의 평균과 표준편차

$$\mu = \sigma = \frac{1}{\lambda}$$



Exponential Distribution

- 지수분포를 따르는 확률 변수와 관련된 확률공식

$$P(X > x) = e^{-\lambda x}$$
$$P(X < x) = 1 - e^{-\lambda x}$$
$$P(x_1 < X < x_2) = P(X < x_2) - P(X < x_1) = e^{-\lambda x_1} - e^{-\lambda x_2}$$

- 예제) 알카라인 배터리의 수명은 λ = .05인 지수분포를 따름.
  - 평균과 표준편차?
  - 배터리의 수명이 10시간과 15시간 사이일 확률?
  - 수명이 30시간 이상일 확률?
  - Excel 함수: expondist(x, lambda, true)