분포이론과 중심극한정리

CE730 통계와 금융

조남운

목차

- 모집단 분포
- 이산형 확률변수의 확률분포함수
- 연속형 확률분포
- 중심극한정리
- 표본분포

분포에 대한 지식의 의미

- 알고자 하는 것: 모집단의 분포
- 모수적 추론 (parametric inference)
  - 분포함수의 형태는 알고 있지만, 구체적인 계수(모수)를 알고 있지 못하는 경우
  - 표본을 통해 모수를 추정
- 비모수적 추론 (nonparametric inference)
  - 모수는 물론, 분포함수의 형태도 모르는 경우

모집단 분포

Namun Cho/ mailto:namun@snu.ac.kr

모집단의 확률분포를 통해 알 수 있는 것

- 확률분포를 안다 = 변수값(의 범위)의 상대 빈도를 안다
- 따라서 해당 변수값의 확률분포를 알 경우 관측된 값이 출현할 확률을 알거나 추측할 수 있음
- 예: 주사위 눈금의 확률분포 (uniform distribution) 를 안다 ⇒ 6이 10번 나올 가능성이 거의 없다는 것을 안다

모집단의 확률분포를 항상 알 수 있는가?

- 모든 값을 관측했다면 그 관측치의 분포표 그 자체가 모집단의 확률변수
- 이러한 상황은 현실적으로 불가능
- 경우에 따라 영원히 불가능할 수도 있음
  - 예: 주사위 던지기의 모집단은 무한한 시행

확률분포의 표현

- 이산형 확률변수
  - 모든 변수값과 그 변수값의 상대빈도
  - 확률질량함수 (pmf: Probability Mass Function): 변수값과 상대빈도의 관계에 대한 함수
- 연속형 확률변수
  - 모든 변수값의 구간에 대한 확률
  - 확률밀도함수 (pdf: Probability Density Function)

정규분포

$$\mathcal{N}(\mu, \sigma^2)$$

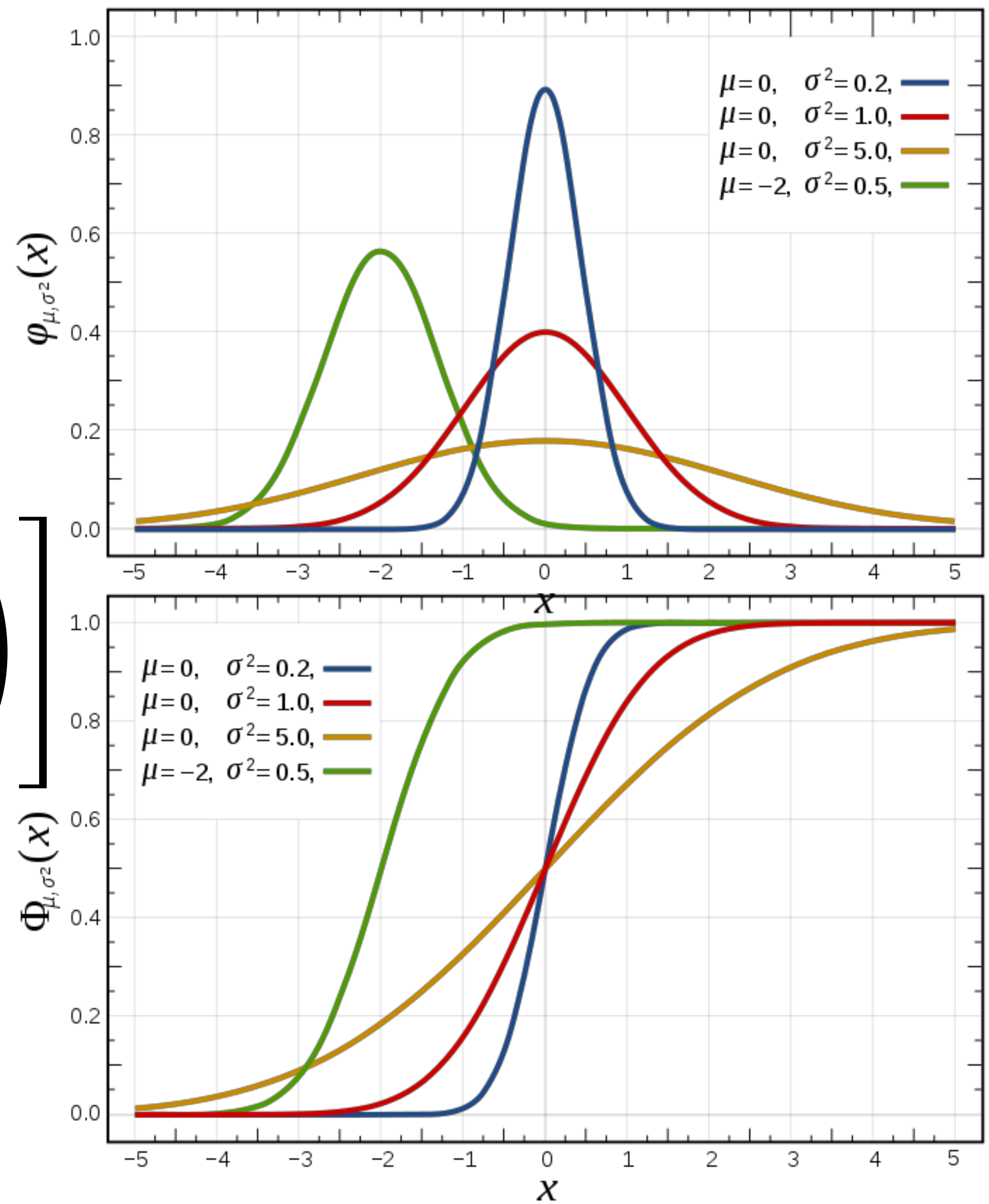
PDF

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

CDF

$$\Pr(X \le x) = \frac{1}{2}\left[1 + erf\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

$$erf(x) := \frac{1}{\sqrt{\pi}}\int_{-x}^{x} e^{-t^2} dt$$



Namun Cho/ mailto:namun@snu.ac.kr

이산형 확률변수의 확률분포함수

Namun Cho/ mailto:namun@snu.ac.kr

분포의 표현

$$X \sim A(param_1, param_2, \cdots)$$

분포의 표현

$$X \sim A(param_1, param_2, \cdots)$$

우측에 기술된
분포를 따르는
확률변수

분포의 표현

$$X \sim A(param_1, param_2, \cdots)$$

우측에 기술된 분포를 따르는 확률변수

확률분포의 이름

분포의 표현

$$X \sim A(param_1, param_2, \cdots)$$

우측에 기술된 분포를 따르는 확률변수

확률분포의 이름

확률분포의 매개변수들

이산형 균일분포
Discrete Uniform Distribution

- 이산형 변수: 변수값의 집합이 셀 수 있는 (countable) 분포임
- 이산형 균일분포: 모든 변수값의 상대빈도가 동일한 분포
  - 예: coin toss, fair dice

이산형 균일분포의 예

$$\Pr(X = 2) = 0.25$$

$$\Pr(0 \leq X < 3) = 0.4$$

X	확률	누적확률
1	0.2	0.2
3	0.2	0.4
2	0.2	0.6
6	0.2	0.8
5	0.2	1

베르누이 시행
Bernoulli Trial

$$\Pr(A = 1) = 1 - \Pr(A = 0) = p$$

- 어떤 시행의 결과가 두 가지인 경우
  - [성공] 이거나 [실패]
  - 확률변수로 표기: A=1 ⇒ 성공! , A=0 ⇒ 실패
- 성공일 확률이 p 라면 실패일 확률은 1-p
  - 성공과 실패는 상호배반이기 때문
- 이러한 시행을 독립적으로 n번 시행했을 때 [성공]의 횟수 ⇒ 베르누이 시행

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0,1,2,\cdots,n$$

베르누이 시행 Xi: 기대값과 분포

$$E(X_i) = 1 \times p + 0 \times (1-p) = p$$

$$E(X_i^2) = 1^2 \times p + 0^2 \times (1-p) = p$$

$$Var(X_i) = E(X_i^2) - E(X_i)^2 = p - p^2 = p(1-p)$$

Namun Cho/ mailto:namun@snu.ac.kr

이항분포: 기본성질
Binomial Distribution

$$\Pr(X=k)=\binom{n}{k}p^k(1-p)^{n-k},\quad k=0,1,2,\cdots,n$$

베르누이 분포: n번의 베르누이 시행에서 1이 나온 횟수의 분포

$$Y := \sum_i^n X_i$$

$$\because Var(X+Y)=Var(X)+Var(Y)+\cancel{2Cov(X,Y)} \xrightarrow{} 0$$

$$Y \sim B(n,p)$$

$$E(Y)=E\left(\sum_i^n X_i\right)=\sum_i^n E(X_i)=\sum_i^n p=np$$

$$Var(Y)=Var\left(\sum_i^n X_i\right)=\sum_i^n Var(X_i)=\sum_i^n p(1-p)=np(1-p)$$

동전 10000번 던지기

$$X \sim B(10000, 0.5)$$

- 베르누이 시행: 동전 1번 던지기
- 이 시행을 10000번 실시하고 앞면이 나온 횟수를 기록한다면 그 횟수는 확률변수
- 이 확률변수는 n=10000, p=0.5 인 이항분포를 따름

동전 10000번 던지기

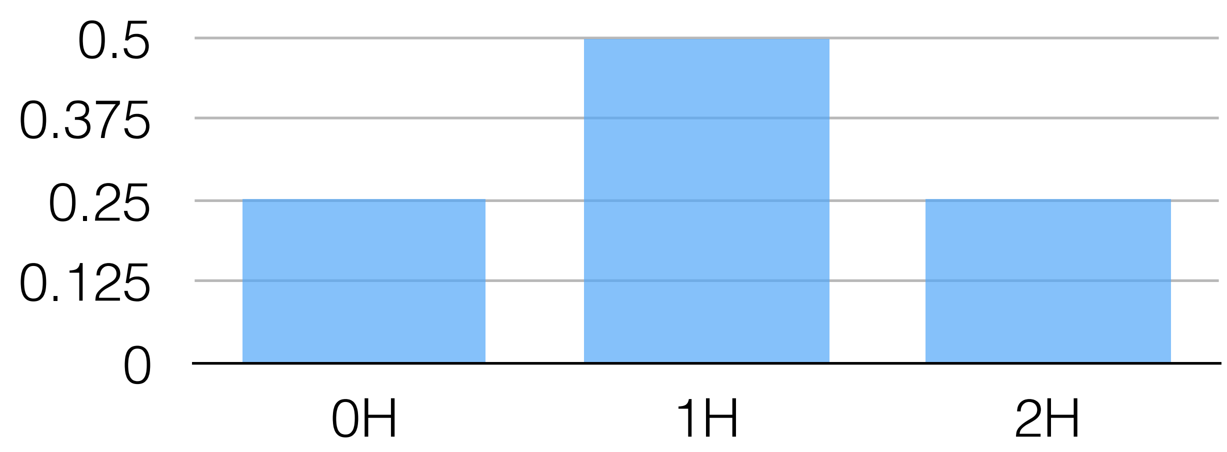
$$X \sim B(10000, 0.5)$$

- 베르누이 시행: 동전 1번 던지기
- 이 시행을 10000번 실시하고 앞면이 나온 횟수를 기록한다면 그 횟수는 확률변수
- 이 확률변수는 n=10000, p=0.5 인 이항분포를 따름

두 번 던진다면?
n=2

- 순서에는 관심이 없이 오로지 H가 몇 번 나오는지에만 관심이 있다면 경우의 수는 총 3가지임
  - H가 2번 나올 경우
  - H가 1번 나올 경우
  - H가 0번 나올 경우
- 각각의 경우의 수는
  - 1/4, 1/2, 1/4

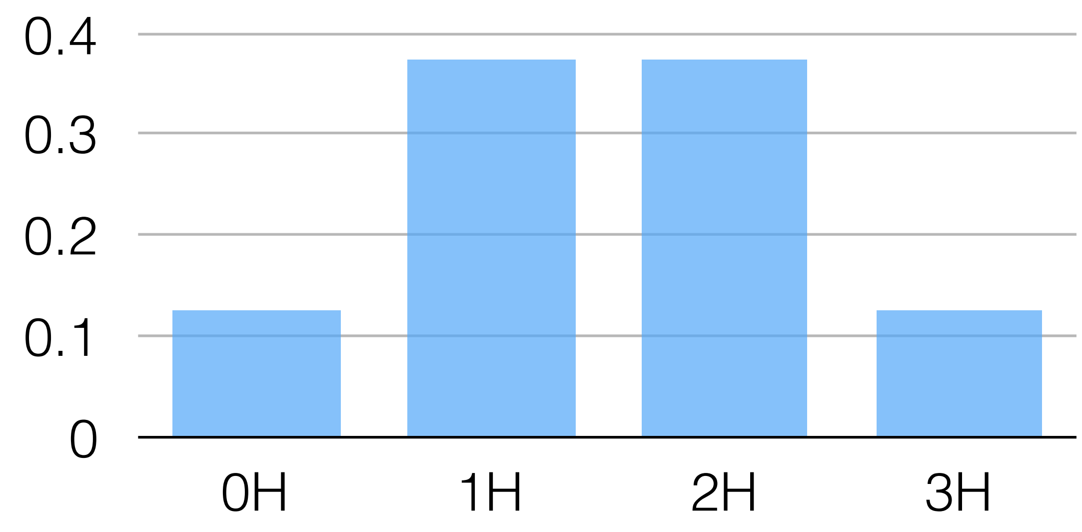
Events	확률
HH	1/4
HT	1/4
TH	1/4
TT	1/4



세 번 던진다면?
n=3

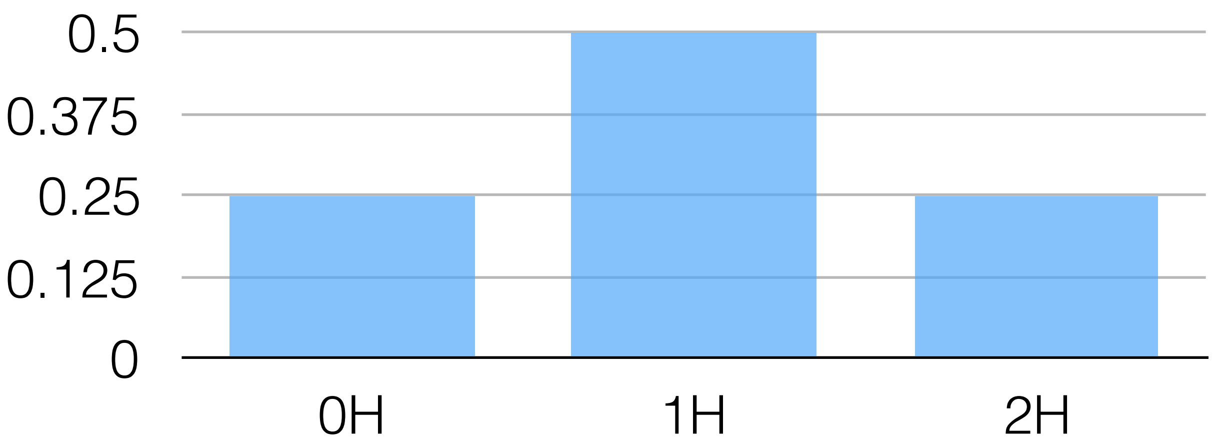
- n번 던진다면 어떻게 될까?

Events	확률
0H	0.125
1H	0.375
2H	0.375
3H	0.125

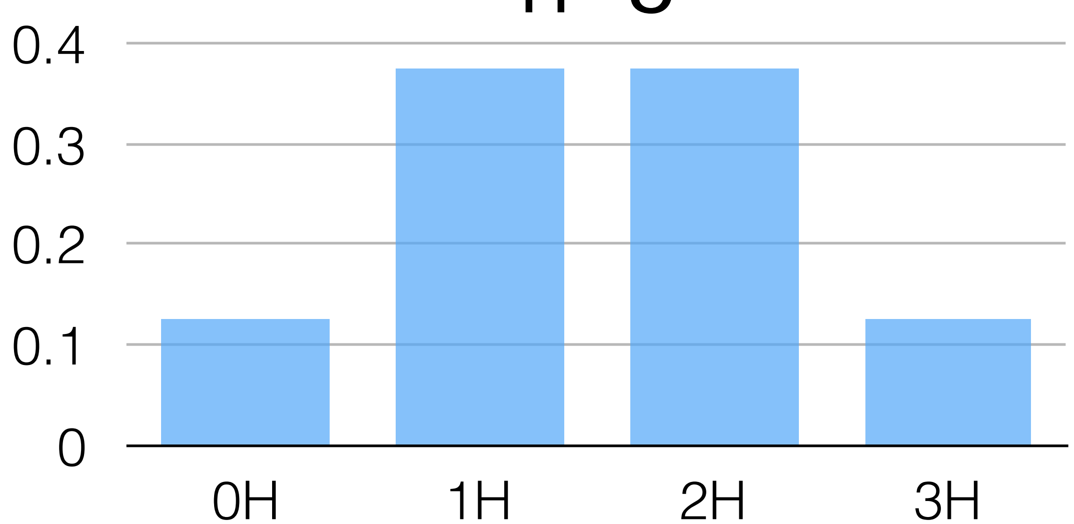


10,000번 동전 던지기

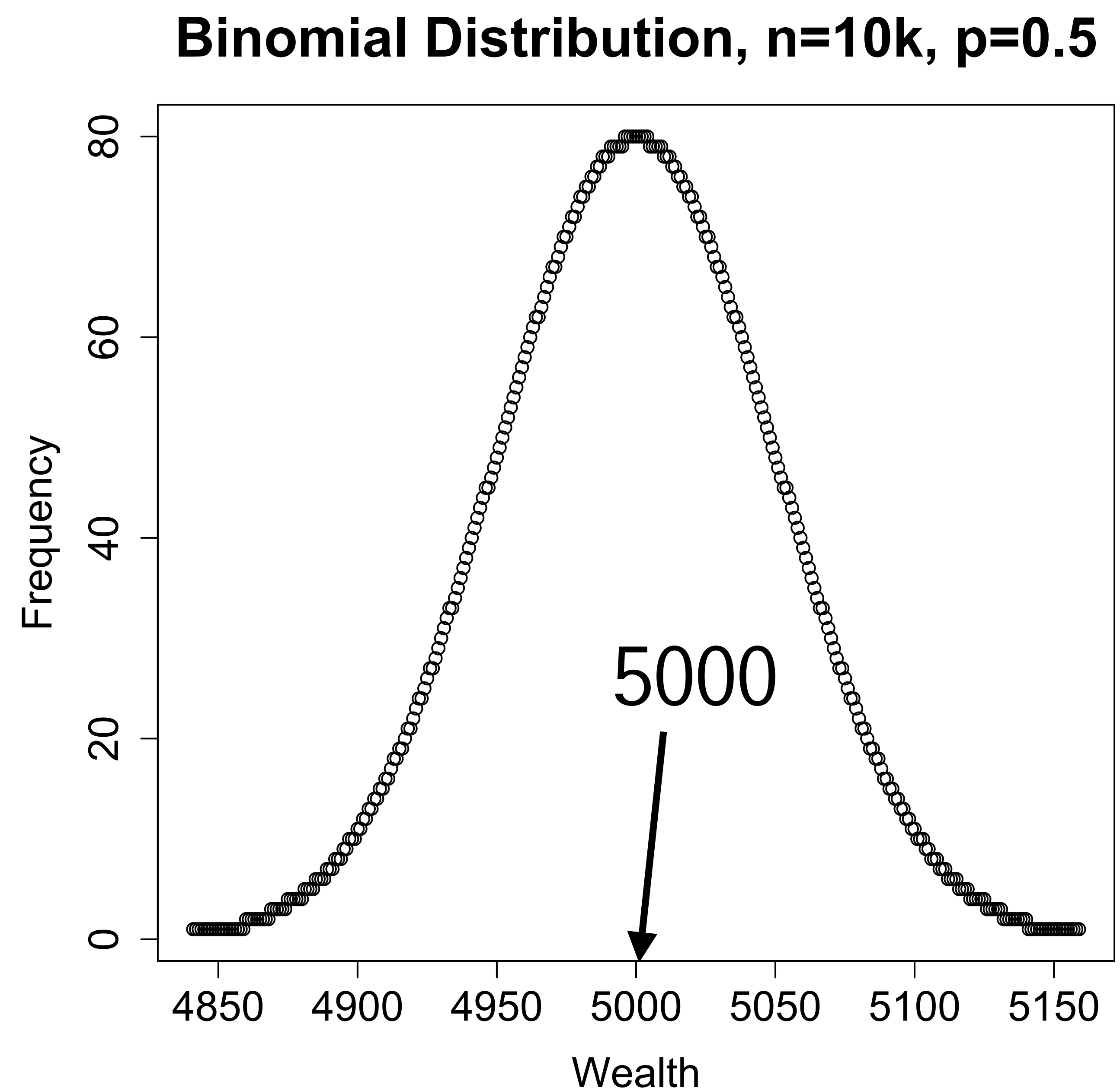
n=2



n=3



# of Head



n=10,000

Namun Cho/ mailto:namun@snu.ac.kr

연속형 확률분포

Namun Cho/ mailto:namun@snu.ac.kr

이산형 분포와의 차이

- 확률변수 X가 연속형 분포일 경우:
- pmf 정의 불가능
  - $P(X = x) = 0$
- 따라서 언제나 일정 범위에 대한 확률의 형태로 기술해야 함: $P(a < X < b) \geq 0$
  - pdf: 범위의 크기가 dx (0에 수렴)

연속형 균일분포

- 모든 값이 출현할 확률이 동일함
  - pdf 가 상수함수
- 값은 반드시 유한해야 함 (bounded)
  - 그렇지 않을 경우 pdf = 0 ⇒ 무의미해짐
- 파라미터: 정의역의 양 끝 a,b
- 가장 간단한 분포 중 하나
  - pdf 가 상수함수이기 때문임
- Notation: U(a,b)

연속형 균일분포 Continuous Uniform Distribution

$$X \sim U(a, b)$$

$$f(x) = \frac{1}{b-a}$$

$$E(X) = \int_a^b xf(x)dx = \int_a^b x\frac{1}{b-a}dx = \frac{1}{b-a}\frac{(b^2-a^2)}{2} = \frac{b+a}{2}$$

$$Var(X) = E(X^2) - E(X)^2 = \int_a^b x^2 f(x)dx - \left(\frac{b-a}{2}\right)^2$$

$$= \frac{b^3-a^3}{3}\frac{1}{b-a} - \left(\frac{b+a}{2}\right)^2 = \frac{1}{12}(b-a)^2$$

정규분포

$$\mathcal{N}(\mu, \sigma^2)$$

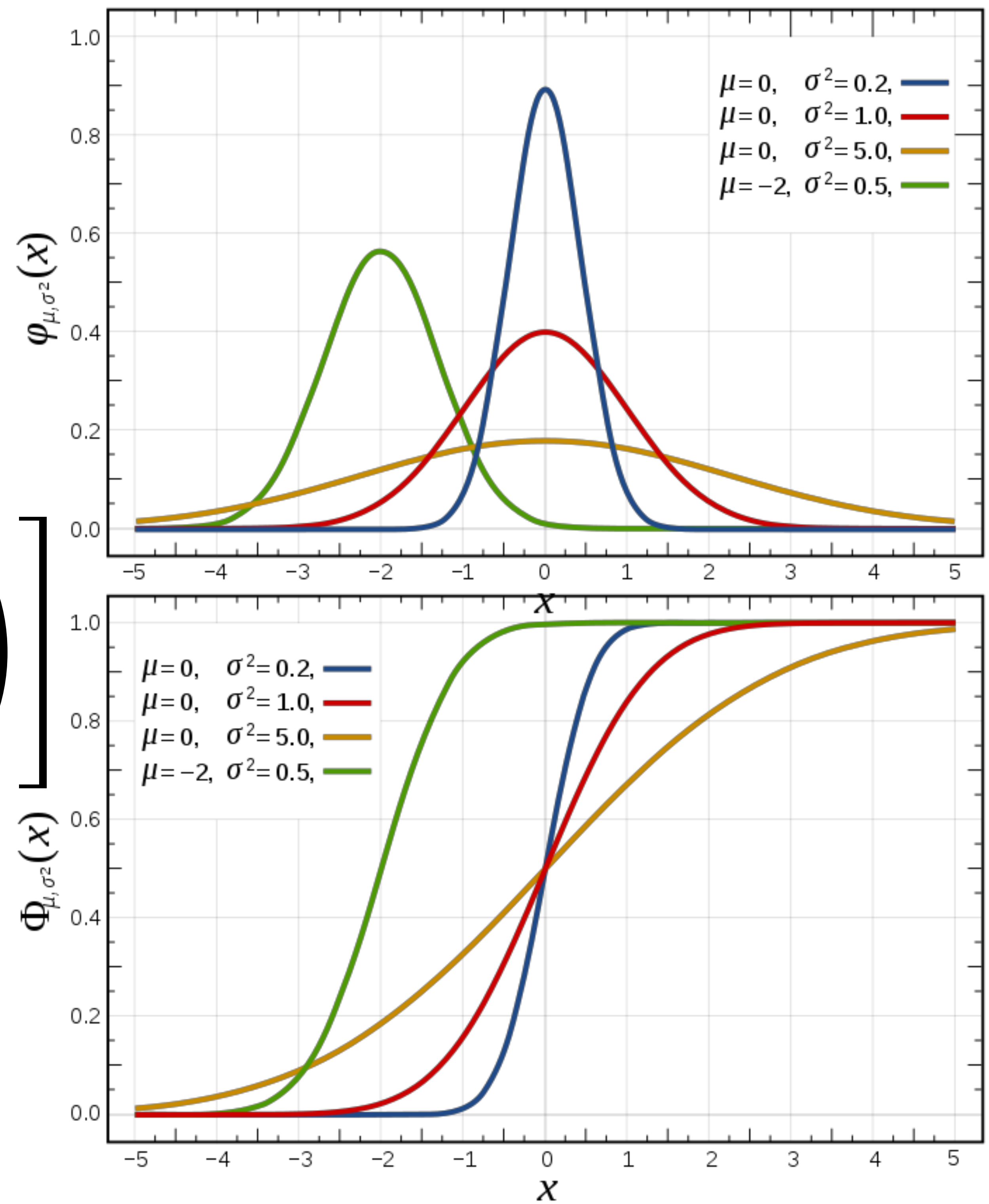
PDF

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

CDF

$$\Pr(X \le x) = \frac{1}{2}\left[1 + erf\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

$$erf(x) := \frac{1}{\sqrt{\pi}} \int_{-x}^{x} e^{-t^2} dt$$



Namun Cho/ mailto:namun@snu.ac.kr

정규분포

- 연속형 분포중 가장 중요한 분포
- 이항분포로부터 근사적으로 도출 가능
- Central Limit Theorem (CLT)

정규분포: 주요 특징

- 중심 (μ)쪽에 데이터의 대부분이 몰려 있음
  - thin tail
  - 극단값을 가질 확률이 매우 적음을 의미
- μ 를 중심으로 좌우대칭 (왜도 = 0)
- 종 모양 형태, 첨도는 3
- 정의역 = 실수 전체
- 모수는 μ 와 $\boldsymbol{\sigma}^2$ 두 가지
- n이 커질수록 이항분포는 정규분포에 수렴

정규분포의 성질

$$E(X) = \mu$$

$$Var(X) = \sigma^2$$

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

$$Z = \frac{X - \mu}{\sigma} \sim N(0,1)$$

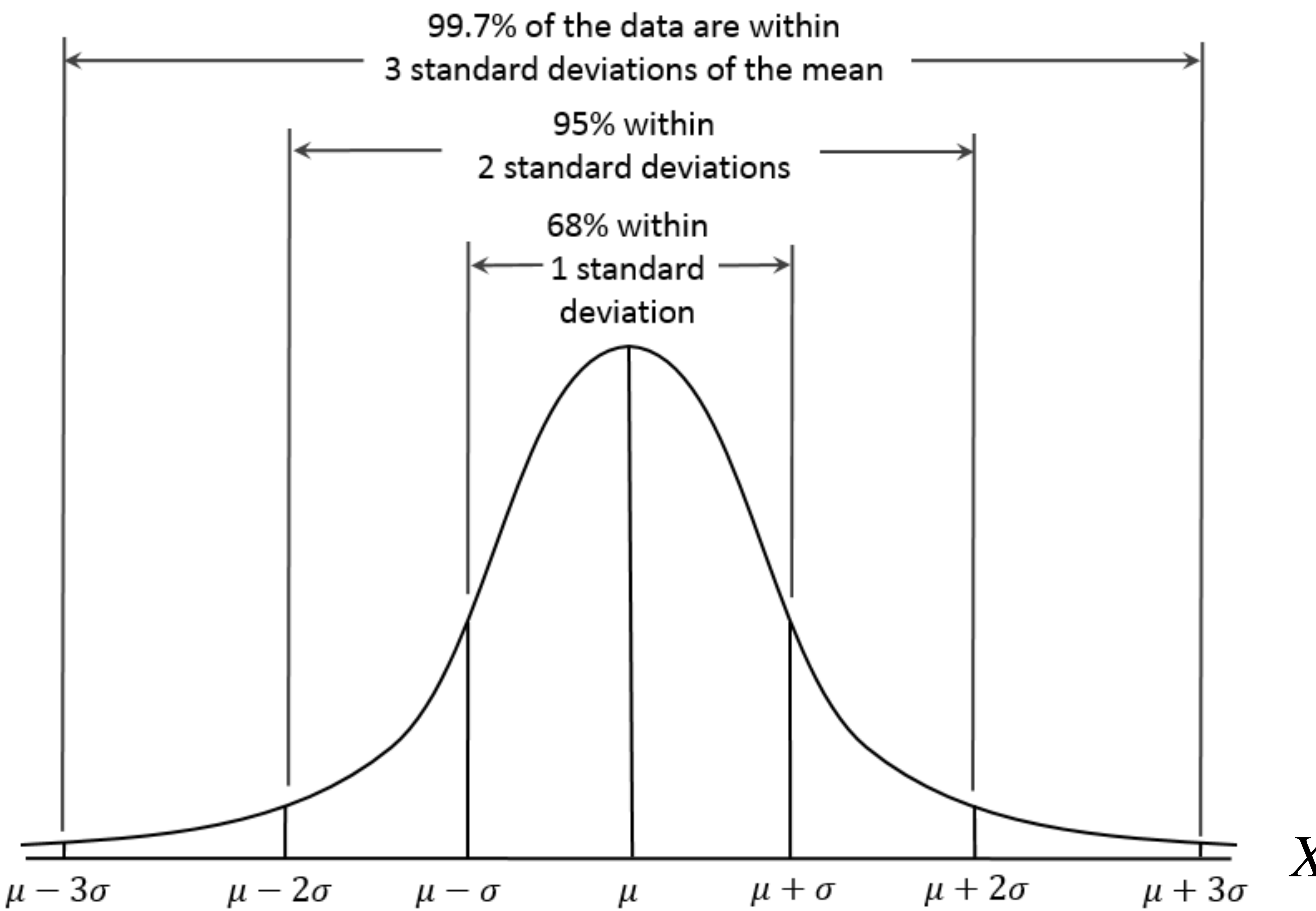
- N(0,1) : 표준정규분포
- Z는 Z-score 라고도 함
- 정규분포를 따르는 확률변수는 덧셈과 스칼라곱을 해도 정규분포를 따름

$$aX \sim N(a\mu, a^2\sigma^2)$$

$$X \sim N(\mu_X, \sigma_X^2), \quad Y \sim N(\mu_Y, \sigma_Y^2)$$

$$\Rightarrow \quad X \pm Y \sim N(\mu_X \pm \mu_Y, \sigma_X^2 + \sigma_Y^2 \pm 2Cov(X, Y))$$

정규분포 구간 확률


99.7% of the data are within
3 standard deviations of the mean
95% within
2 standard deviations
68% within
1 standard
deviation
X
μ − 3σ
μ − 2σ
μ − σ
μ
μ + σ
μ + 2σ
μ + 3σ

Z 변환
Z Transformation

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

$$Z := \frac{X - \mu}{\sigma}$$

$$E(Z) = E\left(\frac{X - \mu}{\sigma}\right)$$

- Z 변환:
- 정규분포를 따르는 확률변수에 모평균을 빼고 표준편차를 나누는 변환
- 평균이 0, 표준편차가 1인 정규분포

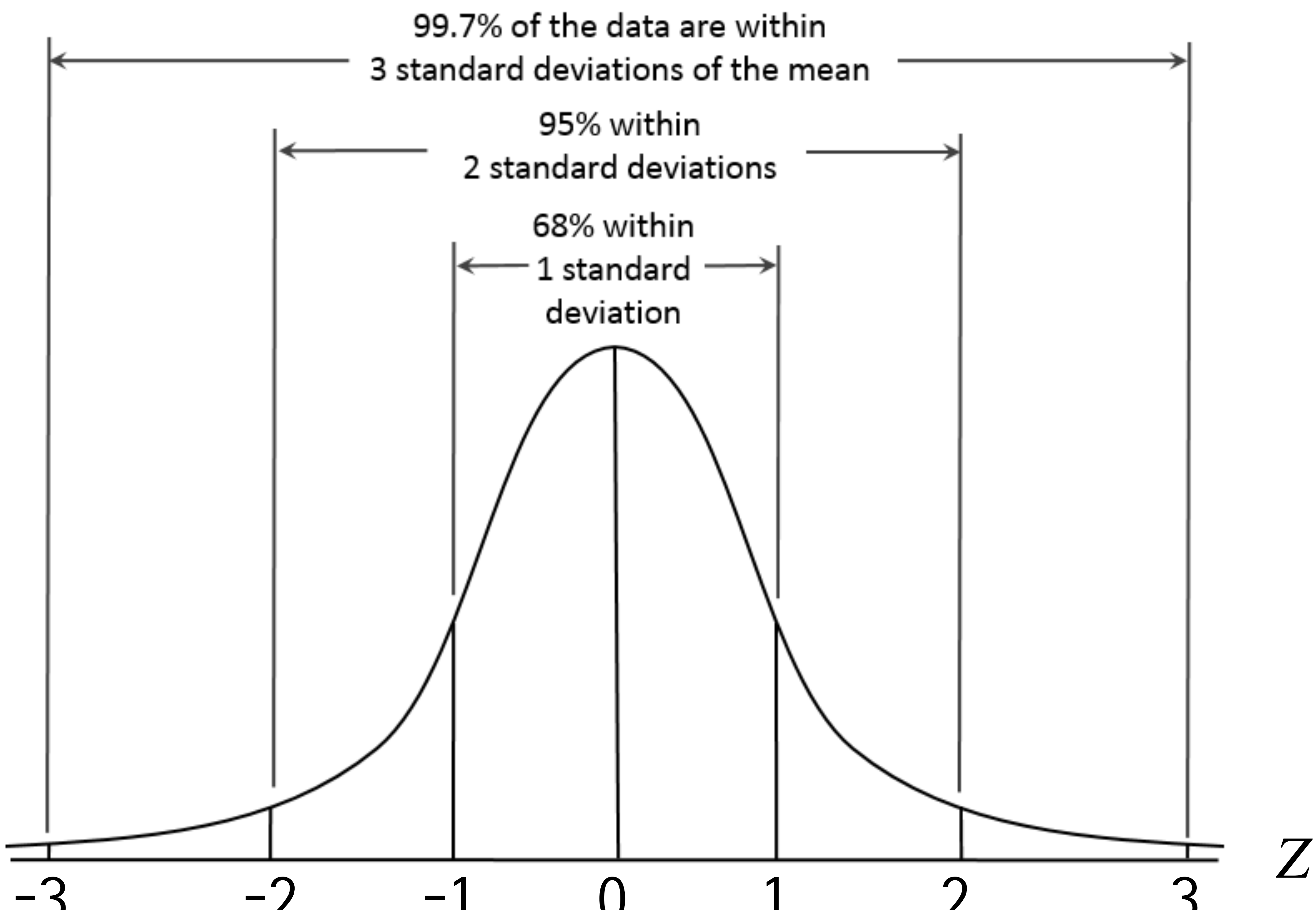
$$E\left(\frac{X}{\sigma}\right) - \frac{\mu}{\sigma} = \frac{\mu}{\sigma} - \frac{\mu}{\sigma} = 0$$

$$Var(Z) = Var\left(\frac{X - \mu}{\sigma}\right) = Var\left(\frac{X}{\sigma}\right) + Var\left(\frac{\mu}{\sigma}\right)$$

$$= \frac{1}{\sigma^2} Var(X) + 0 = 1$$

$$\therefore Z \sim N(0,1)$$

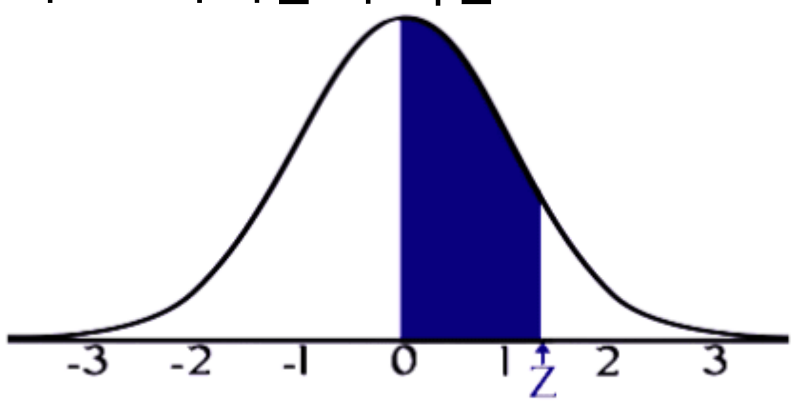
표준정규분포의 구간확률



Namun Cho/ mailto:namun@snu.ac.kr

표준정규분포표

주의: 교과서는 누적분포표



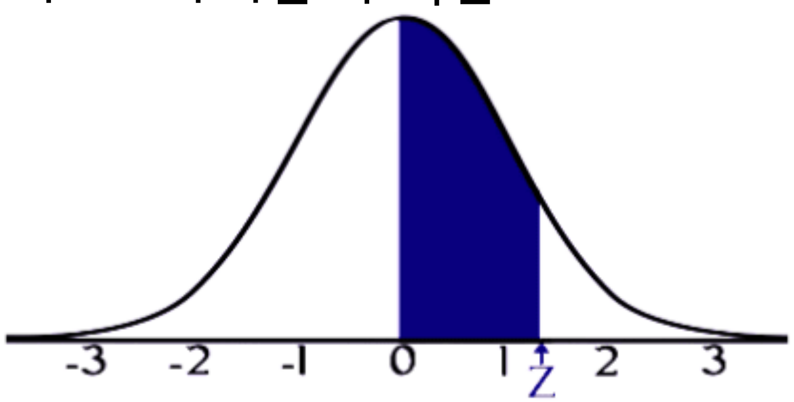
STANDARD NORMAL TABLE (Z)

Entries in the table give the area under the curve between the mean and *z* standard deviations above the mean. For example, for *z* = 1.25 the area under the curve between the mean (0) and *z* is 0.3944.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0190	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	0.0910	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	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	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	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2969	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	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3513	0.3554	0.3577	0.3529	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	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	0.4370	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

표준정규분포표

주의: 교과서는 누적분포표



STANDARD NORMAL TABLE (Z)

Entries in the table give the area under the curve between the mean and z standard deviations above the mean. For example, for $z = 1.25$ the area under the curve between the mean (0) and z is 0.3944.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0190	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	0.0910	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	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	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	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2969	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	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3513	0.3554	0.3577	0.3529	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	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	0.4370	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

예제 5.4

- 어떤 기업의 주당 수익 (EPS)의 기대치: 4000원
- EPS의 표준편차: 400
- EPS ~ $N(4000, 400^2)$
  - Pr(EPS<3200) = ?
    - EPS가 3200원보다 작을 확률
  - Pr(3600 < EPS < 4400) = ?
- 표준정규분포표로 구해보자

Solution: Ex 5.4

- Pr(EPS<3200)
  - EPS를 표준화
  - Z = (EPS-4000)/400
  - (3200-4000)/400=-2
  - Pr(EPS<3200)=Pr(Z<-2) = 0.5-0.4772 = 0.0228
- Pr(3600<EPS<4400)
  - (3600-4000)/400=-1
  - (4400-4000)/400=1
  - Pr(3600<EPS<4400) = Pr(-1<Z<1)=0.3413*2 = 0.6826

z	0.00	0.01	0.02	0.03	0.04	0.05
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0190
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734
0.8	0.2881	0.2910	0.2939	0.2969	0.2995	0.3023
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3513
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842

Solution: Ex 5.4

- Pr(EPS<3200)
  - EPS를 표준화
  - Z = (EPS-4000)/400
  - (3200-4000)/400=-2
  - Pr(EPS<3200)=Pr(Z<-2) = 0.5-0.4772 = 0.0228
- Pr(3600<EPS<4400)
  - (3600-4000)/400=-1
  - (4400-4000)/400=1
  - Pr(3600<EPS<4400) = Pr(-1<Z<1)=0.3413*2 = 0.6826

z	0.00	0.01	0.02	0.03	0.04	0.05
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0190
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734
0.8	0.2881	0.2910	0.2939	0.2969	0.2995	0.3023
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3513
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842

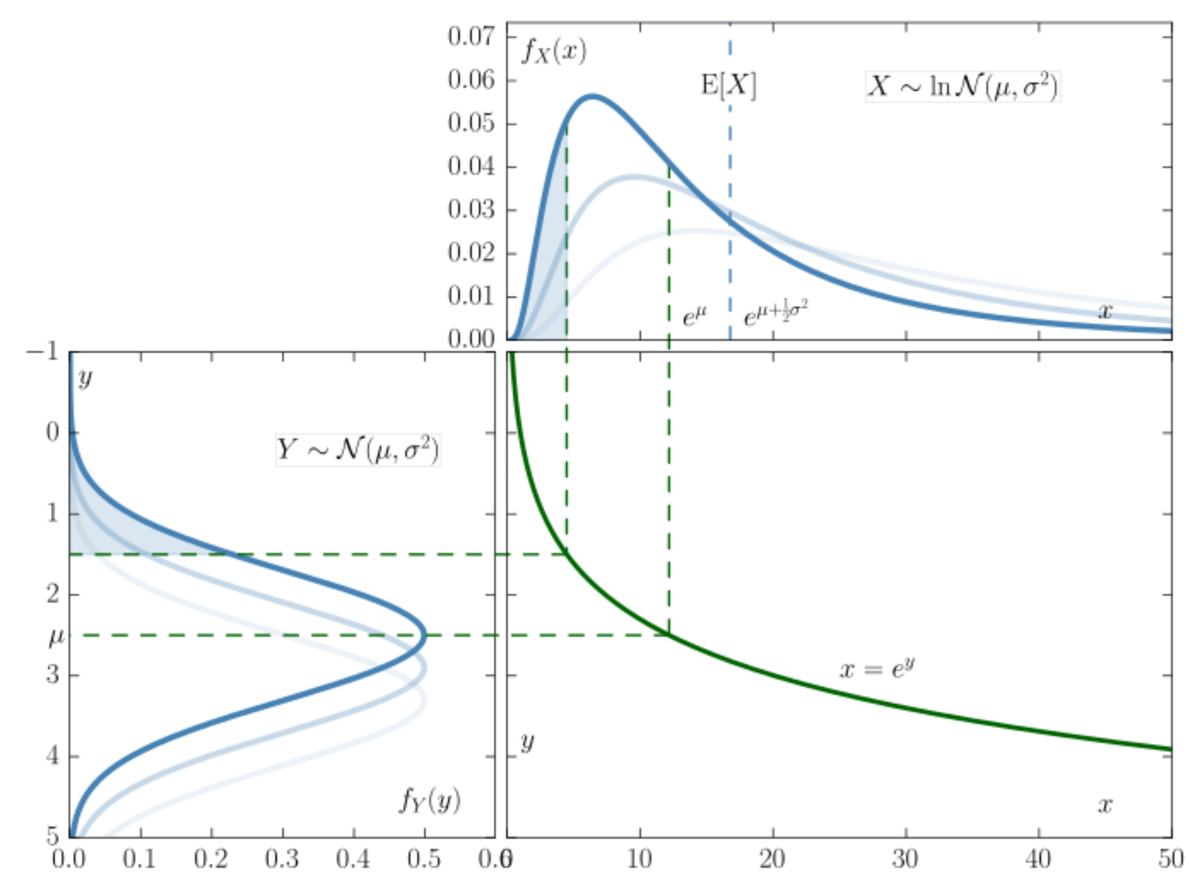
표준누적정규분포표

- 주텍스트 p.124는 표준누적정규분포표임
- 해당 표로도 동일한 방식으로 예제를 풀 수 있음

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

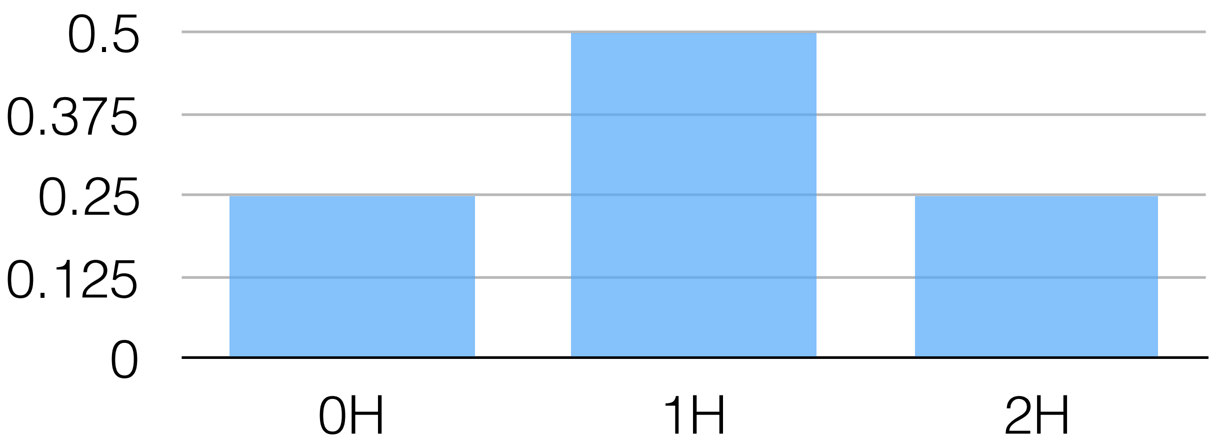
로그 정규 분포
Log Normal Distribution

- 이항분포는 베르누이시행의 결과의 합
- 만일 결과의 합이 아니라 곱이 된다면?
  - 같은 이야기로, 베르누이시행의 로그합이라면?
- 로그 이항분포
- 로그 이항분포 역시 n이 커질수록 로그 정규분포에 수렴

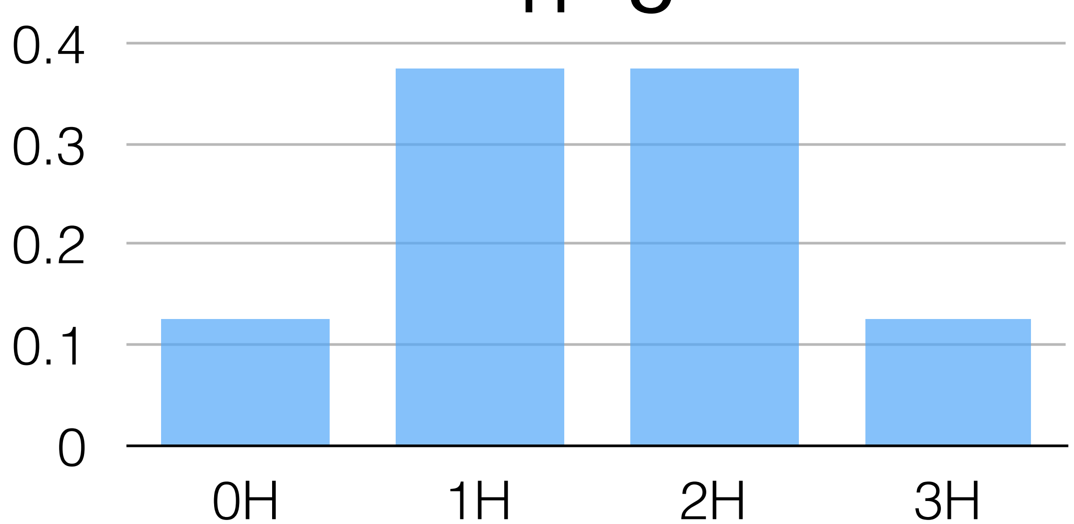


10,000번 동전 던지기

n=2

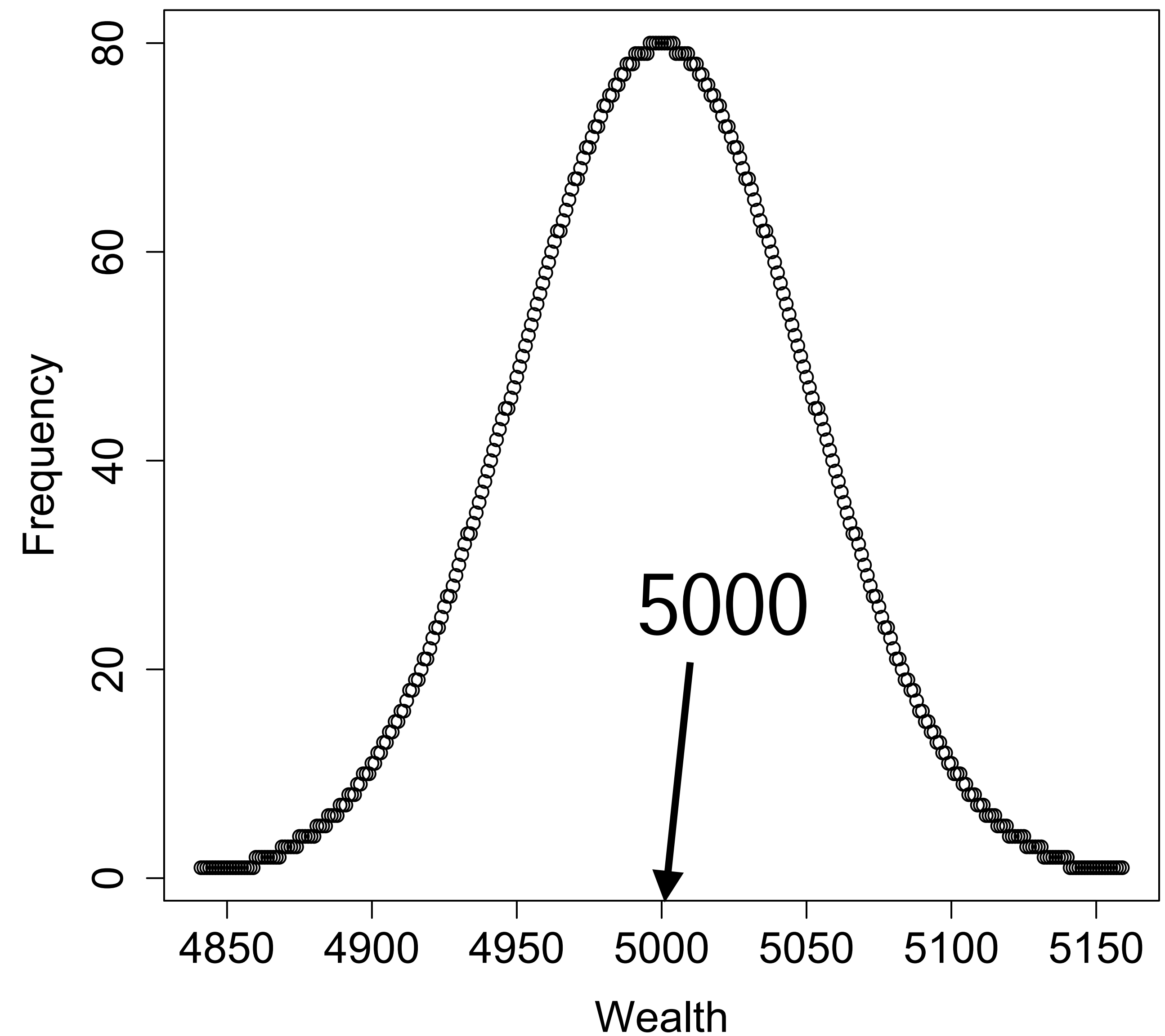


n=3



# of Head

Binomial Distribution, n=10k, p=0.5



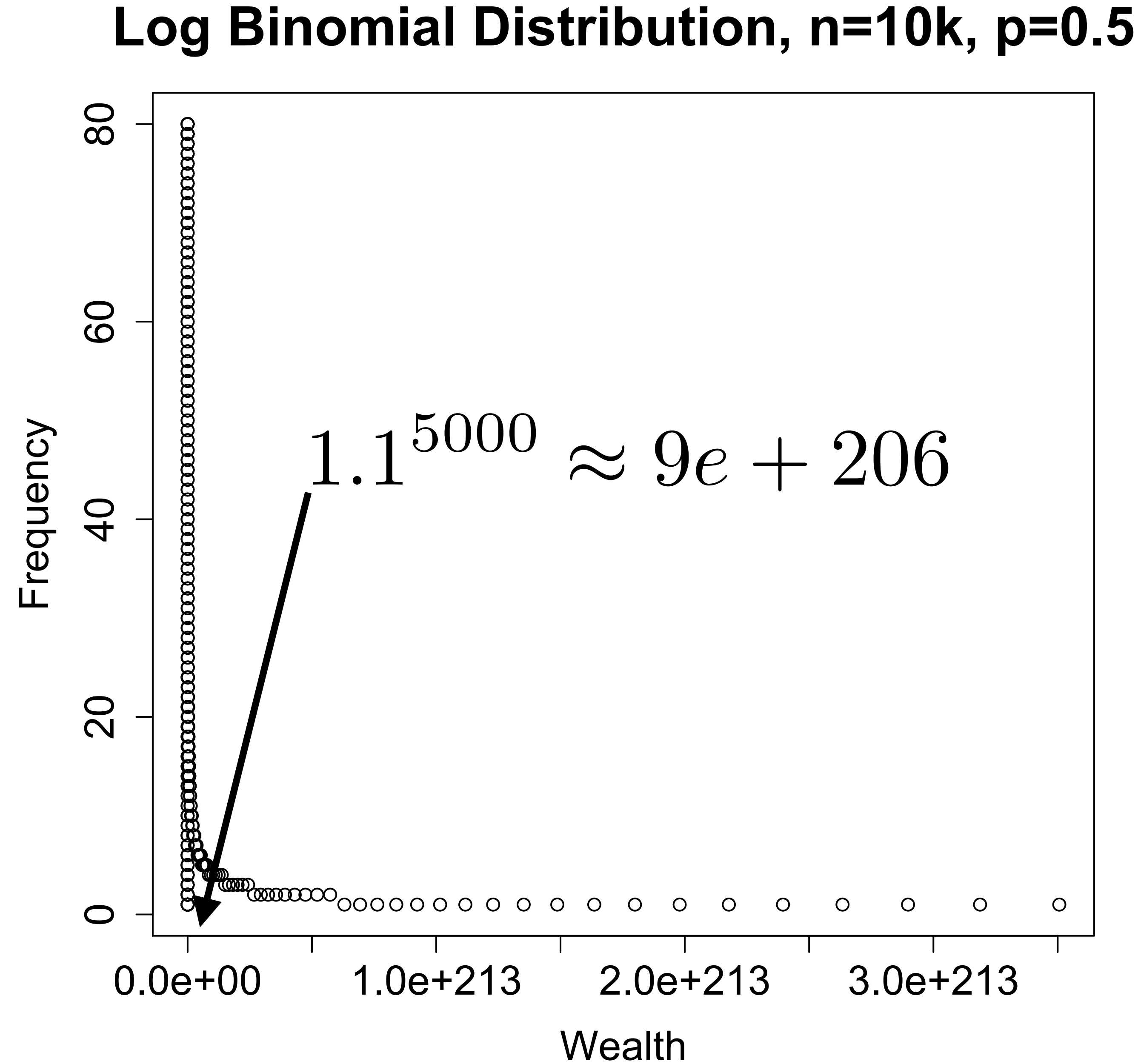
n=10,000

Namun Cho/ mailto:namun@snu.ac.kr

동전던지기: 지수증가의 경우

- 10,000명에게 초기 자금을 동일하게 부여
- 만일 앞면이 나올 때마다 1점씩 부과한다면 분포는 앞에서와 동일
- 하지만 앞면이 나올 때마다 그 사람의 자금에 10%씩을 부과한다면
- 매우 불평등한 부의 분배가 나타남
- 현대의 부의 분포와 유사한 패턴
- 사전적으로는 공정
- 사후적으로는 불평등

Payoff



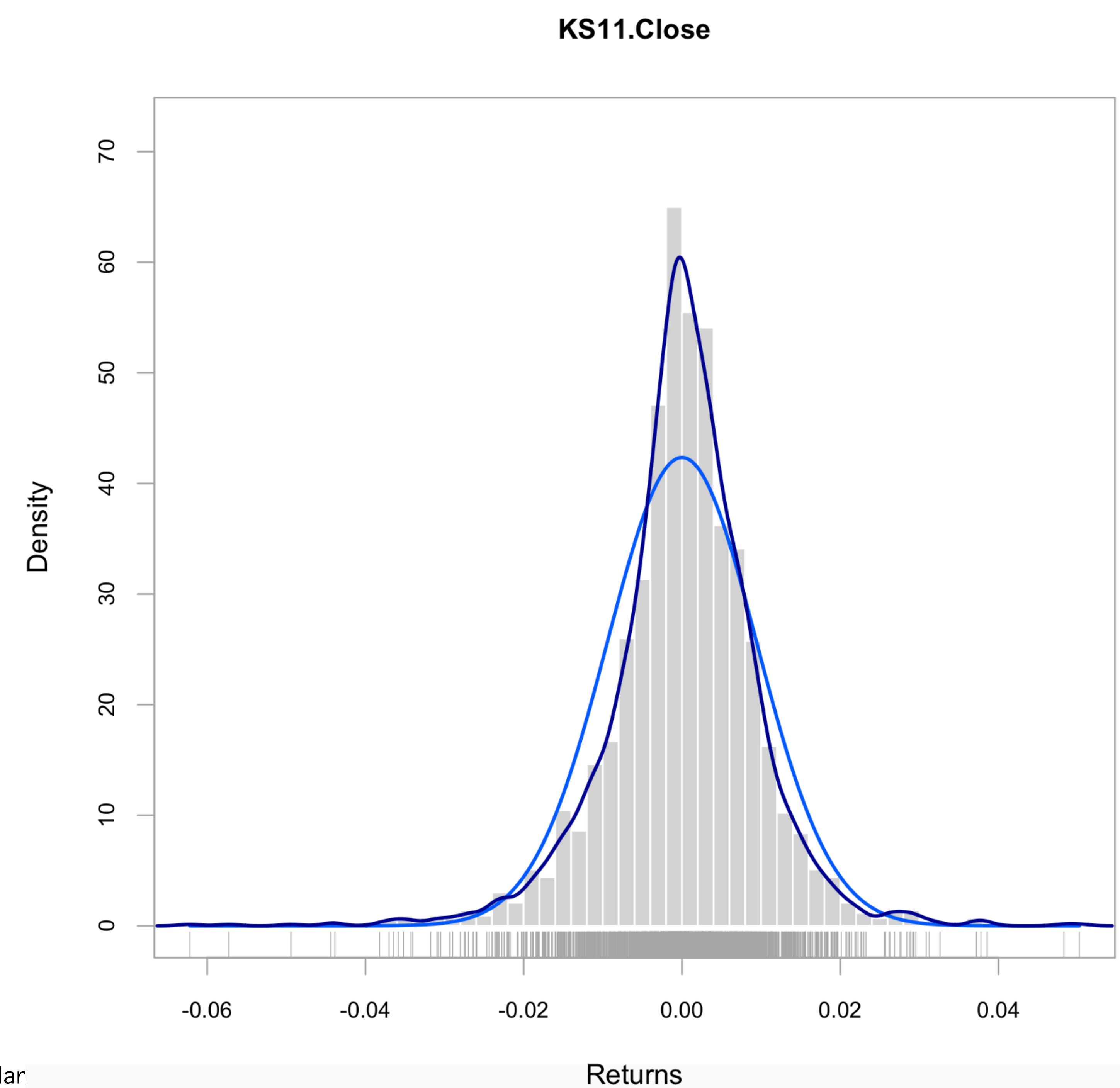
금융과 로그정규분포

- 금융에서는 많은 확률사건의 결과가 지수적 증가로 나타남
- 예: 50% High return (30%), 50% Low return (5%)
- 이러한 사건의 결과는 로그정규분포의 속성을 가지게 됨
- 이제 수익률 r 이 확률변수인 상황을 생각해보자
- P0: 초기자산가치
- Pt: t기후 가치 (원금포함)
- t는 상수이므로 Pt도 확률변수

$$P_t / P_0 := (1 + r)^t$$

주식 평균 수익률 분포

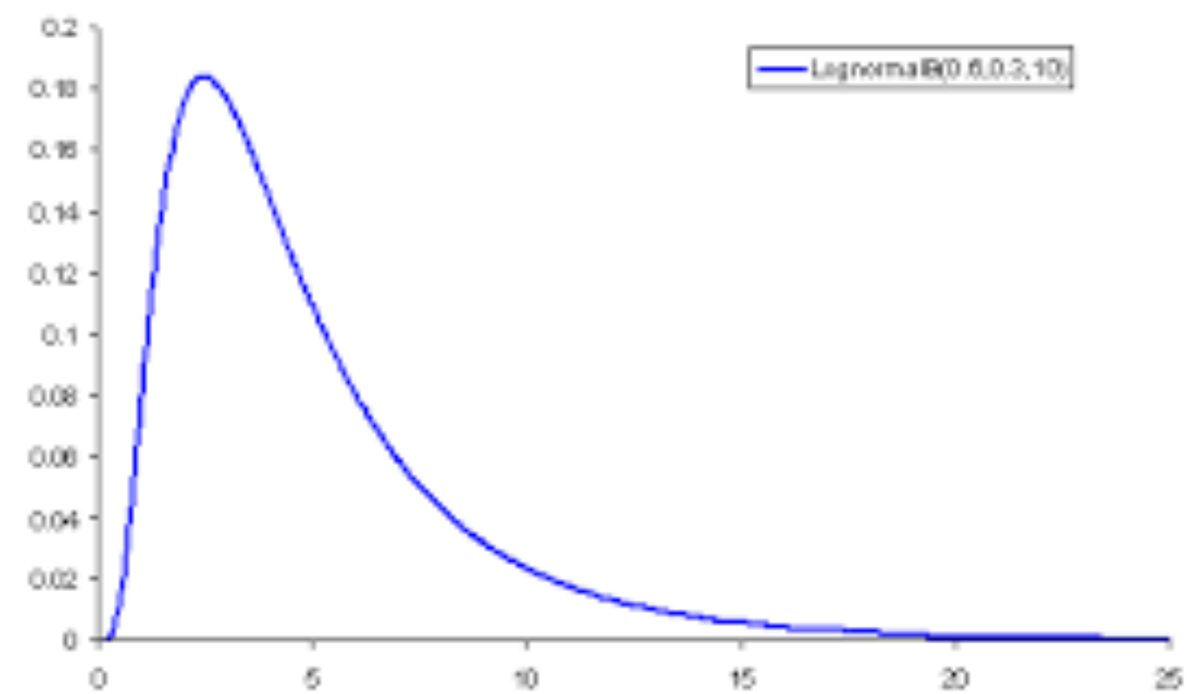
정규분포와 흡사함



Nan

Pt의 max, min

- Maximum
  - 이론적으로는 얼마든지 높은 수익률이 가능함
  - 따라서 max(Pt) = ∞
  - 이때의 r = ∞
- Minimum
  - 아무리 많이 잃어도 원금 이상을 잃을 수는 없음
  - 따라서 min(Pt) = 0
  - 이때의 r = 0
- 로그 변환을 통해 이 비대칭 분포를 대칭화할 수 있음

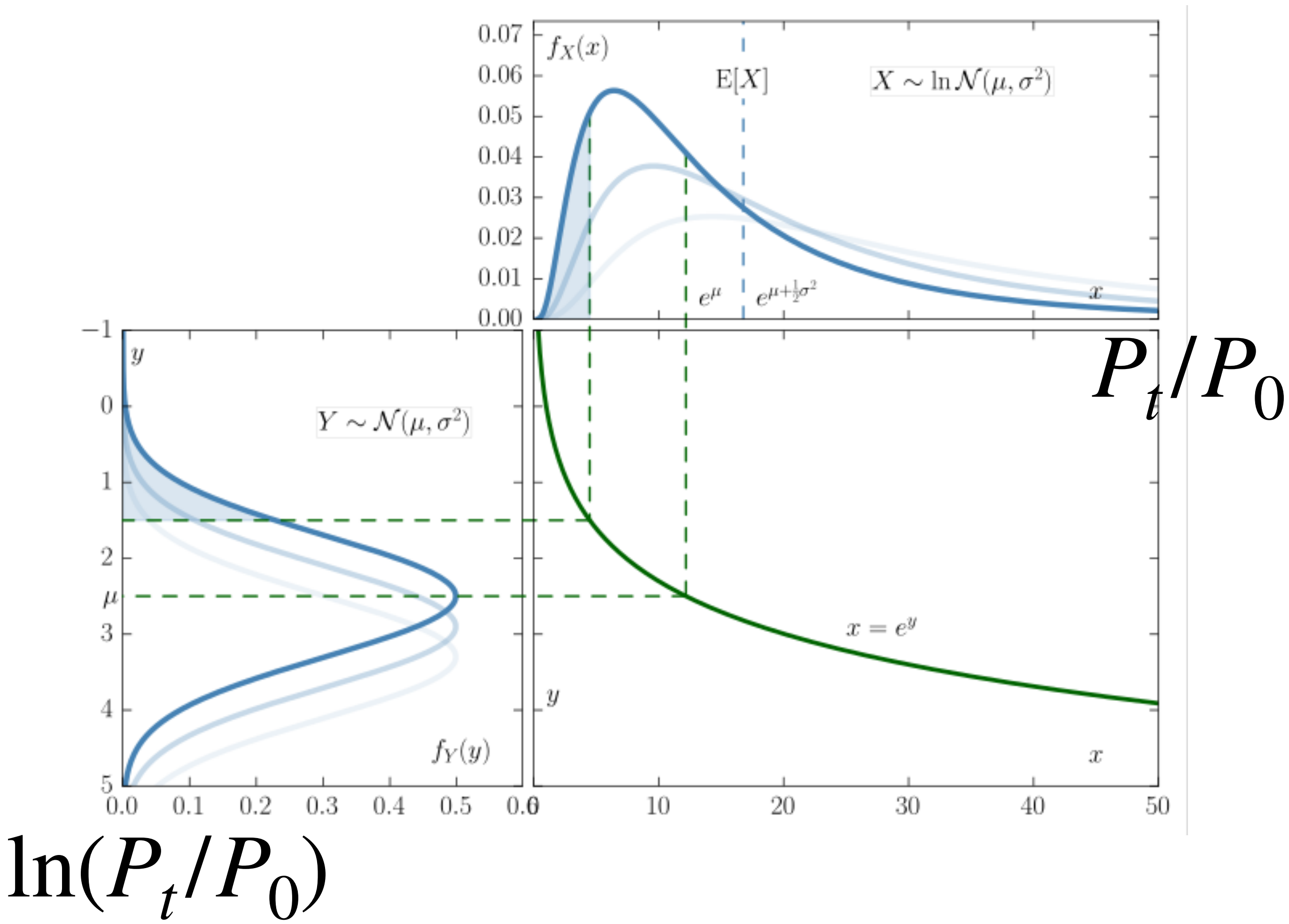


P_t/P_0

$$X \sim N(\bullet,\bullet) \Rightarrow e^X \sim LN(\bullet,\bullet)$$
$$Y \sim LN(\bullet,\bullet) \Rightarrow \ln(Y) \sim N(\bullet,\bullet)$$

로그변환

완전한 분석을 위해서는
시간을 연속화해야 함



보론: 지수와 로그
Exponents and Logarithms

Namun Cho/ mailto:namun@snu.ac.kr

Exponential Functions

Definition (Exponential Function)

$f : \mathbb{R} \to \mathbb{R}$ *is exponential function if* $f(x) = \bar{a}\bar{b}^x, \quad \bar{b} > 0$

- $x \in \mathbb{N} \Rightarrow f(x) := \bar{a} \prod_{i=1}^{x} \bar{b}$
- $f(0) := \bar{a}$
- $f(1/n) := \bar{a} \sqrt[n]{\bar{b}}$
- $f(m/n) := \bar{a} \sqrt[n]{\bar{b}^m}$
- $x < 0 \Rightarrow f(x) = \bar{a}(1/\bar{b})^{|x|}$
- Graph($a > 0$): convex, monotonic increasing ($b > 1$) or decreasing ($b \in (0, 1)$) function (horizontal line when $b = 1$)

Growth of an Account with Interest rate r

Saving Account at $t = \bar{T}$ with Interest rate $\bar{r}$, Initial Endowment $\bar{A}$

$$A_t = \bar{A}\left(1+\bar{r}\right)^{\bar{T}}$$

Compound Interest

If interest is compounded n times per time unit,

$$A_t = \bar{A}\left(1+\frac{\bar{r}}{n}\right)^{n\bar{T}}$$

Continuous Compounding

Compound Interest with $n \to \infty$

$$A_t = \lim_{n\to\infty} \bar{A}\left(1+\frac{\bar{r}}{n}\right)^{n\bar{T}} = \bar{A}e^{\bar{r}\bar{T}}$$

Number e

Definition (The Number e)

$$e := \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = \sum_{n=0}^{\infty}\frac{1}{n!} \approx 2.718281693\cdots$$

e is irrational number.

Theorem (5.1)

$$\lim_{n\to\infty} A\left(1+\frac{r}{n}\right)^{nt} = Ae^{rt}$$

In general, an initial quantity a_0 with growth rate r (per time unit) become a_0e^{rt} at time t (time unit)

Logarithm

Definition (Base b Logarithm)

Base b logarithm is an inverse of exponential function with base b

$$f = b^x \quad \Leftrightarrow \quad x = \log_b f$$

- $a^{\log_a z} = z$
- $\log_a a^y = y$
- Graph: concave, monotonic increasing ($b > 1$) or convex, monotonic decreasing ($b \in (0, 1)$)

Natural Logarithm

Definition (Natural Logarithm)

Base e logarithm is natural logarithm

$$\ln x := \log_e x$$

$$\ln x = y \quad \Leftrightarrow \quad e^y = x$$

$$e^{\ln x} = x$$

$$\ln e^x = x$$

Basic Properties of Exponential functions

$\forall r, s \in \mathbb{R}$,

1. $a^r a^s = a^{r+s}$
2. $a^{-r} := 1/a^r$
3. $a^r/a^s = a^{r-s}$
4. $(a^r)^s = a^{rs}$
5. $a^0 := 1$

Basic Properties of Logarithmic functions

$\forall r, s, a, b, c > 0 \wedge a, c \neq 1,$

1. $\log(rs) = \log r + \log s$
2. $\log(1/s) = -\log s$
3. $\log(r/s) = \log r - \log s$
4. $\log r^s = s \log r$
5. $\log 1 = 0$
6. $\log_a b = \frac{\log_c b}{\log_c a} = \frac{\ln b}{\ln a}$

(Ex5.4) Rule of 70 (or 69)

Derivatives of Exp and Log functions

Theorem (5.2)

$$(e^x)' = e^x$$

$$(\ln x)' = \frac{1}{x}$$

if $u \in \mathbf{C}^1$, *from chain rule,*

$$(e^u)' = (e^u)\, u'$$

$$(\ln u)' = \frac{u'}{u} \qquad (u > 0)$$

Present Value

Present Value (PV)

After time T, A (at $t = 0$) grow to B (at $t = T$)

$$B = Ae^{rT}$$

A is the present value (PV) of B at $t = T$

$$A = Be^{-rT}$$

- PV of annuity

이산형 복리수익률과 연속형 복리수익률

- 이산형 복리수익률: 정해진 기간 단위로 이자를 지급
  - n=1,2,4,12, 365
- 연속형 복리수익률: 기간 단위가 연속
  - n → ∞
- 이산형 복리수익률과 연속형 복리수익률은 다름
- 예: 연 수익률 10%의 경우
- 연수익률(r) 10%에 해당하는 연속형 복리수익률 (r_c)

$$P_t/P_0 = (1+r)^t = e^{r_c t}$$

$$\Rightarrow (1+r) = e^{r_c}$$

$$r_c = \ln(1+r)$$

$$r_c = \ln(1+0.1) = 0.0953101798$$

$$\approx 9.531\ \%$$

수익률이 확률변수일 때 결과값의 확률분포

- 이제 순간 수익률 r이 정규분포를 따른다고 생각해보자
- 그 결과물인 수익률은 로그 정규분포를 따르게 된다

$$P_t/P_0 := e^{rt}$$

$$r \sim N(\mu, \sigma^2)$$
$$\Rightarrow rt \sim N(t\mu, t^2\sigma^2)$$

$$r \sim N(\bullet, \bullet) \Rightarrow e^{rt} \sim LN(\bullet, \bullet)$$

$$\therefore P_t/P_0 \sim LN(\bullet, \bullet)$$

중심극한정리
CLT: Central Limit Theorem

Namun Cho/ mailto:namun@snu.ac.kr

임의 표본 Random Sample

- X1, X2, … , Xn 이 n개의 임의표본이 되기 위해서는
  - 모든 표본들은 동일 모집단에서 독립적으로 추출되어야 함
  - independent: 독립
  - identical: 동일 모표본에서 추출
- 동일 모집단이기만 하다면 어떤 확률분포를 가지더라도 상관없음
- 검토하고자 하는 것은 이 n개의 확률표본의 평균

$$X_i \sim i.i.d.?(\mu, \sigma^2)$$

$$\overline{X} := \frac{\sum_i^n X_i}{n}$$

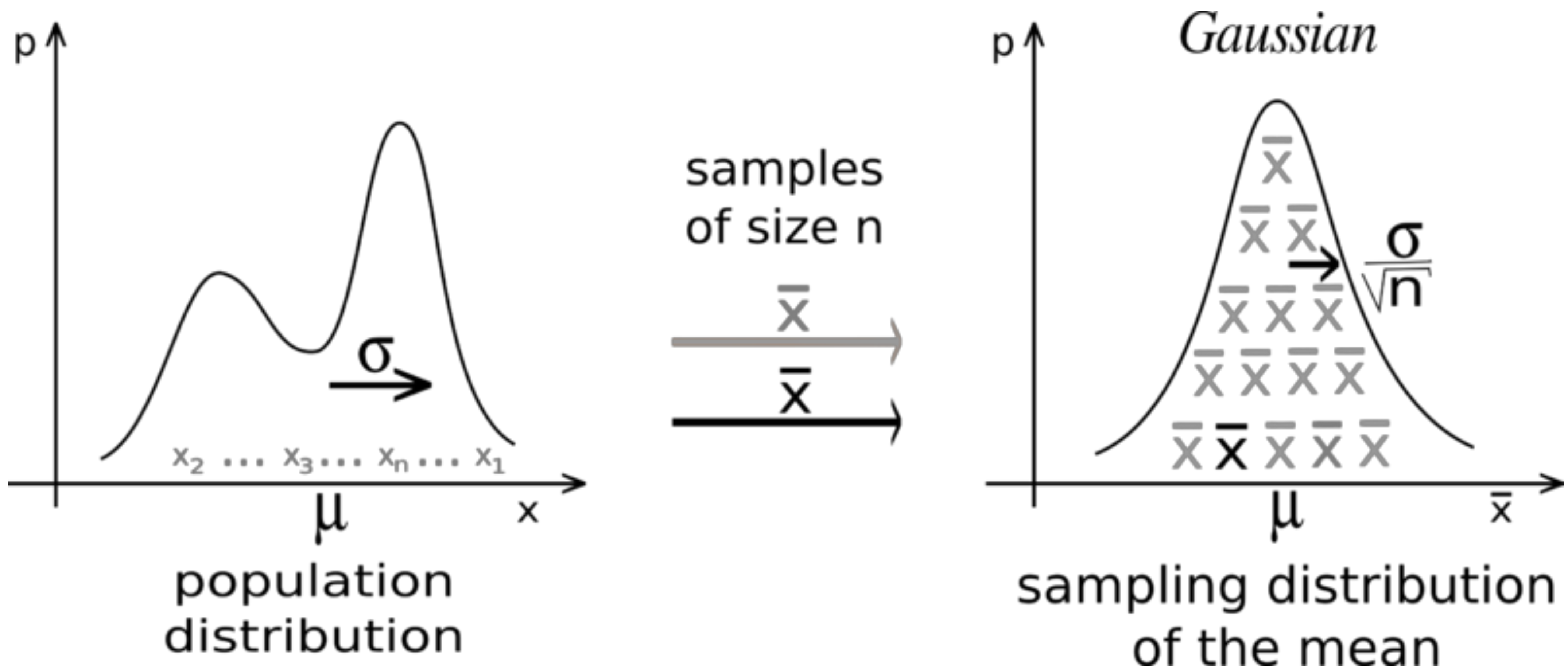
임의표본평균의 분포

- 임의표본의 평균은 확률변수의 합으로 이루어져 있음
  - 따라서 이 또한 확률분포
- 샘플 Xi 가 어떤 분포로부터 나왔던 그런 샘플들의 평균 (표본평균) 의 분포는 정규분포를 따름

$$\overline{X} := \frac{\sum_i^n X_i}{n} \xrightarrow{d} N(\mu, \sigma^2)$$

$$\frac{\overline{X} - \mu}{\sigma} \xrightarrow{d} N(0,1)$$

CLT



p
x_2 ... x_3 ... x_n ... x_1
σ
μ
x
population distribution
samples of size n
$\bar{x}$
$\bar{x}$
Gaussian
p
$\frac{\sigma}{\sqrt{n}}$
μ
$\bar{x}$
sampling distribution of the mean

n의 크기에 대해서

- 대칭분포 + thin tail
  - 경험적으로 20여개 이상이면 정규분포에 근사
- 비대칭분포
  - 경험적으로 50여개 이상이면 정규분포에 근사

예제

- p=0.6인 어떤 시행을 100번 했을 때 50회 이상 success할 확률은?

$$Y \sim B(100,0.6)$$

$$\Rightarrow \Pr(Y \geq 50) = ?$$

풀이

$$\Pr(Y \geq 50) = P\left(\frac{Y}{100} \geq \frac{50}{100}\right)$$

$$\frac{\frac{Y}{100} - E\left(\frac{Y}{100}\right)}{\sqrt{Var\left(\frac{Y}{100}\right)}} \sim N(0,1)$$

(CLT)

$$E(Y/100) = 1/100E(Y) = 1/100 * np = 0.6$$

$$Var(Y/100) = 1/100^2 Var(Y) = 1/100^2 np(1-p) = 24/100$$

$$P\left(\frac{Y}{100} \geq \frac{50}{100}\right) = P\left(\frac{Y/100 - 0.6}{\sqrt{24/100}} \geq \frac{50/100 - 0.6}{\sqrt{24/100}}\right) \approx P\left(Z \geq \frac{50/100 - 0.6}{\sqrt{24/100}}\right)$$

$$= P(Z \geq -0.2041) = 0.5793$$

Next Topics

- 표본분포
- 추정

수고하셨습니다!

Namun Cho/ mailto:namun@snu.ac.kr

수고하셨습니다!

Namun Cho/ mailto:namun@snu.ac.kr