4

Continuous Random Variables and Probability Distributions

LO

Describe the probability distribution and calculate the mean and variance of a continuous random variable

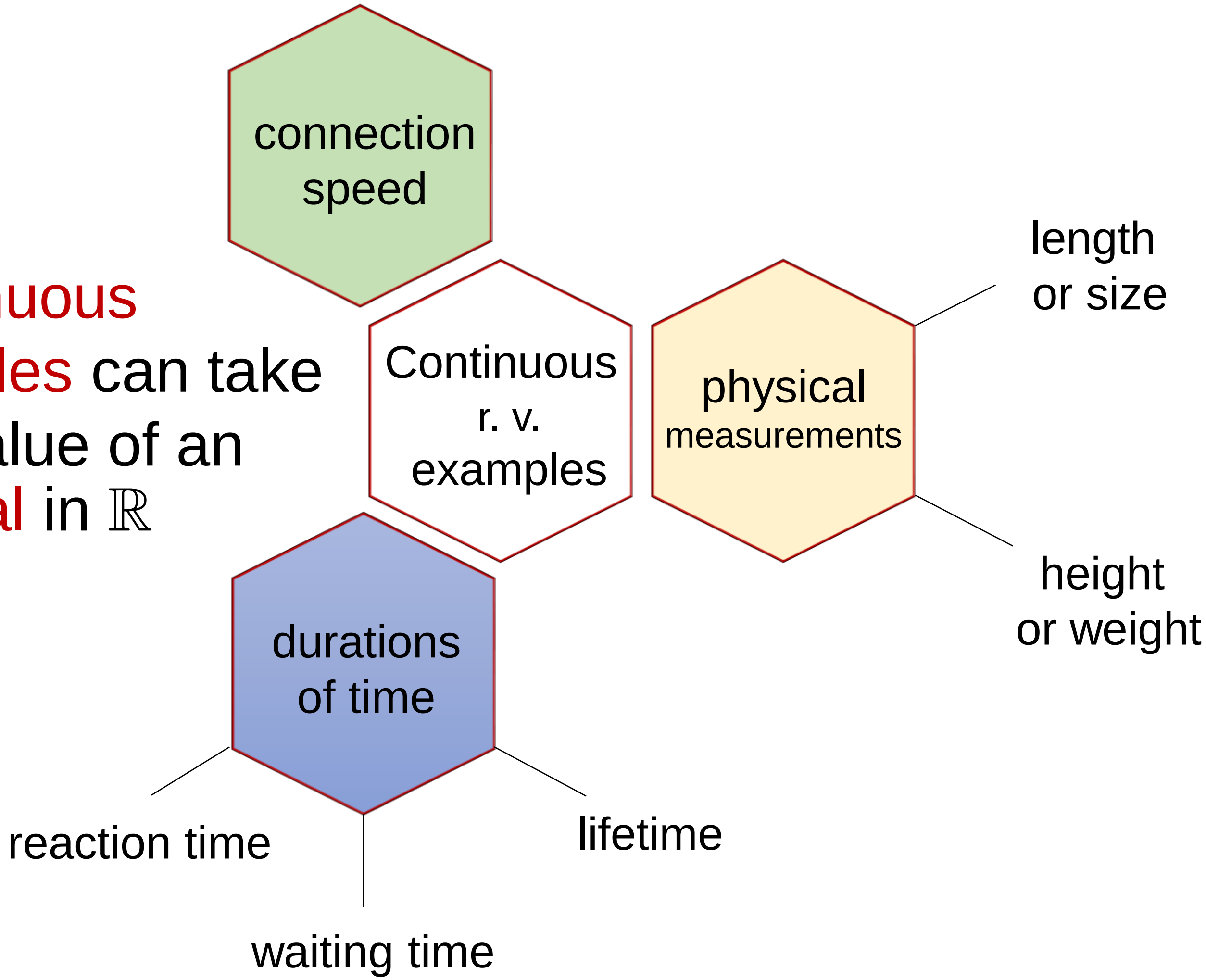
Continuous Uniform Distribution

Normal Random Variable

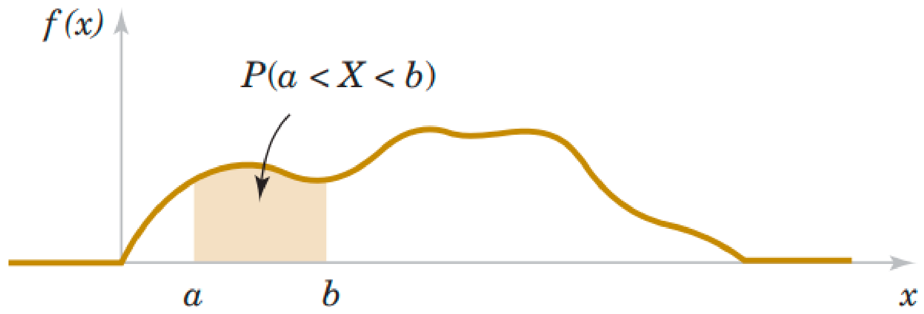
Exponential Distribution

Continuous Random Variables

Continuous variables can take any value of an interval in $\mathbb{R}$



Probability Density Function (pdf)



Probability determined from the area under f(x).

Ex. Let X be a continuous random variable and suppose that $f(x) = cx^2$, for $-1 < x < 2$, and 0 otherwise is the pdf of X.

a/ Find c

b/ Find $P(X \geqslant 0)$

------

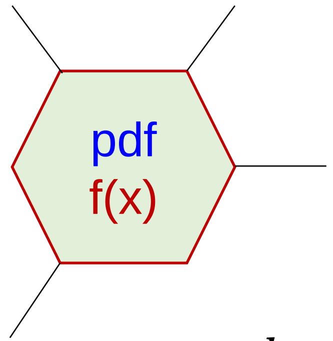
a/ $\int_{-\infty}^{\infty} f(x)dx = \int_{-1}^{2} cx^2 dx$ = c(8/3 +1/3) = 3c

f(x) is the pdf of X $\Leftrightarrow \int_{-\infty}^{\infty} f(x)dx = 1 \Leftrightarrow c = 1/3$

b/ $P(X \geqslant 0) = \int_0^2 (1/3)x^2 dx = 8/9$

"density" is a measure of "probability mass" per unit length

$f(x) \geqslant 0$



$$\int_{-\infty}^{\infty} f(x)dx = 1$$

$P(a \leqslant X \leqslant b) = \int_a^b f(x)dx$

$= P(a \leqslant X < b)$

$= P(a < X \leqslant b)$

$= P(a < X < b)$

$P(X = x_0) = 0$

(different from pmf)

Probability Density Function Pdf – Ex

The lifetime, in years, of some electronic component is a continuous random variable with the probability density function

$$f(x) = \begin{cases} \dfrac{k}{x^3} & for\ x > 1 \\ 0 & otherwise \end{cases}$$

Find k and compute the probability for the lifetime to exceed 5 years.

- $\int_{-\infty}^{\infty} f(x)dx = \int_{1}^{\infty}$ k/x^3dx = -k(0 – 1)/2 = 1 ⇔ k = 2

Cumulative Distribution Function (cdf)

cdf
$F(x)$

$F(x) = P(X \leqslant x)$
$= P(X < x)$

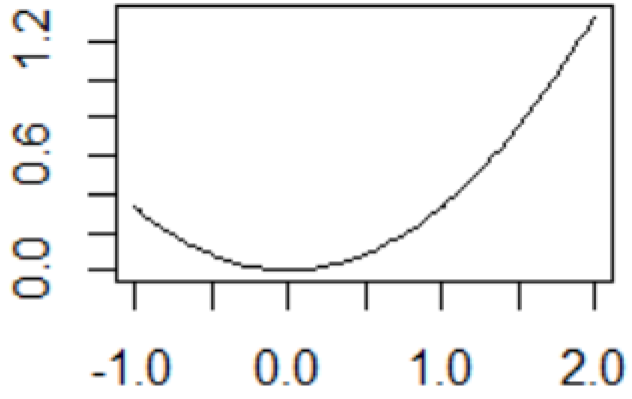
$F'(x) = f(x)$

$F(x) = \int_{-\infty}^{x} f(u)\,du$

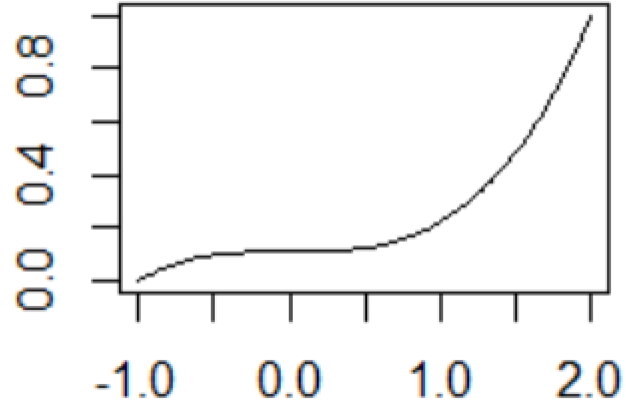
$\lim_{x \to -\infty} F(x) = 0$
$\lim_{x \to \infty} F(x) = 1$

$F(x) \leqslant F(y)$
if $x \leqslant y$
(increasing)

$P(a \leqslant X \leqslant b)$
$= F(b) - F(a)$



Pdf: $f(x) = x^2/3$, for $-1 < x < 2$



Cdf: $F(x) = \int_{-1}^{x} t^2/3\,dt = x^3/9 + 1/9$
for $-1 < x < 2$, and 0 otherwise

Pdf vs Cdf – Ex

cdf
F(x)

$F(x) = P(X \leqslant x)$
$= P(X < x)$

$F'(x) = f(x)$

$P(a \leqslant X \leqslant b)$
$= F(b) - F(a)$

$$F(x) = \int_{-\infty}^{x} f(u)du$$

$F(x) \leqslant F(y)$
if $x \leqslant y$
(increasing)

$\lim_{x \to -\infty} F(x) = 0$
$\lim_{x \to \infty} F(x) = 1$

Ex. The lifetime, in years, of some electronic component is a continuous random variable with the probability density function

$$f(x) = \begin{cases} \frac{2}{x^3} & for\ x > 1 \\ 0 & otherwise \end{cases}$$

Find the cdf F(x) and P(2 < X < 3).

$F(x) = \int_{-\infty}^{x} f(u)du = \int_{1}^{x} f(u)du = \int_{1}^{x} \frac{2}{u^3} du$
$= 1 - 1/x^2$ for $x > 1$

$P(2 < X < 3) = F(3) - F(2)$
$= ¼ - 1/9 = 0.139$

Mean and Variance

Expected value of X

Measure central value

Expected value of g(x)

E[g(x)]

$= \int_{-\infty}^{\infty} g(x) f(x) dx$

Mean

μ

$\mu = E(X)$

$= \int_{-\infty}^{\infty} x f(x) dx$

X

Standard deviation

σ

Variance

σ^2

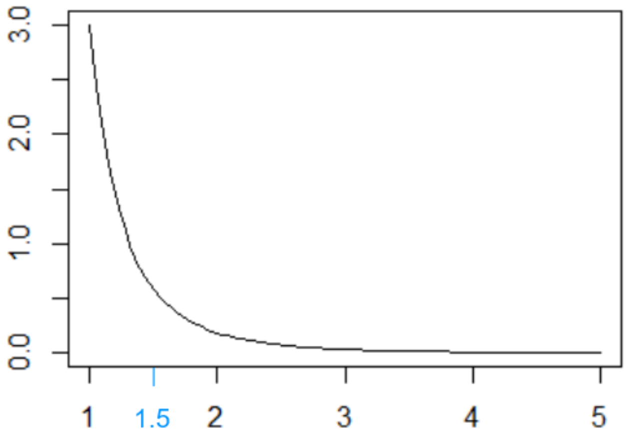
$\sigma^2 = V(X)$

$= \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$

$\sigma = \sqrt{\sigma^2}$

Measure spread of values

$\sigma^2 = E(X^2) - E(X)^2$



Ex. Given the pdf of X,

$$f(x) = \begin{cases} \dfrac{3}{x^4} & for\ x > 1 \\ 0 & otherwise \end{cases}$$

Find the mean and variance of X.

$E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_{1}^{\infty} 3/x^3 dx = 1.5$

$V(X) = E(X^2) - E(X)^2 = 0.75$

4-30. Suppose that $f(x) = x/8$ for $3 < x < 5$. Determine the mean and variance of x.

4-36. The probability density function of the weight of packages delivered by a post office is $f(x) = 70/(69x^2)$ for $1 < x < 70$ pounds.

(a) Determine the mean and variance of weight.

(b) If the shipping cost is \$2.50 per pound, what is the average shipping cost of a package?

(c) Determine the probability that the weight of a package exceeds 50 pounds.

Continuous Uniform Distribution

used for computer simulation of various events and processes

used in any situation when a value is picked "at random" from a given interval

i.e., locations of errors in a program, birthdays throughout a year

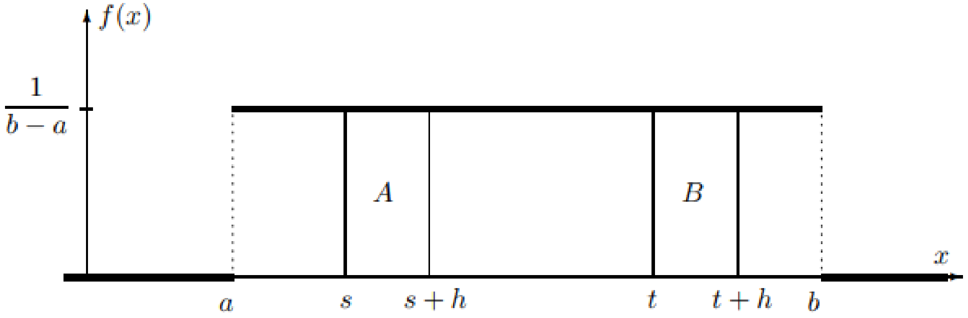
Uniform Distribution

Variance: $\sigma^2 = (b - a)^2/12$

pdf: $f(x) = 1/(b-a)$

Mean: $\mu = (a + b)/2$

For $h > 0$, $P(t < X < t + h) = h/(b-a)$ is independent of t



Ex. Tri is taking a bus from Dist.1 to Thu Duc city, a distance of 15km. His position is uniformly distributed between the two areas. What is the probability that he is past Saigon Bridge, which is 5km from Dist.1?

Let X be Tri's location ➔ $X \sim Unif(0, 15)$ ➔ $P(X > 5) = (5 < X < 15) = (15 - 5)/(15 - 0) = 2/3$.

Continuous Uniform Distribution – Ex

(Uniform Current) Let the continuous random variable X denote the current measured in a thin copper wire in milliamperes. Assume that the range of X is [0, 20 mA], and assume that the probability density function of X is $f(x) = 0.05,\ 0 \leqslant x \leqslant 20$. Find $P(5 < X < 10)$, $E(X)$, $V(X)$.

$P(5 < X < 10) = (10 – 5)(0.05) = 0.25$

$E(X) = (a + b)/2 = (0 + 20)/2 = 10$

$V(X) = (b – a)^2/12 = (20 – 0)^2/12 = 33.33$

Normal Distribution (Gaussian distribution)

Plays a vital role in Probability and Statistics

Good model for physical variables like weight, height, temperature, voltage, pollution level (household incomes or student grades, etc.)

Normal distribution

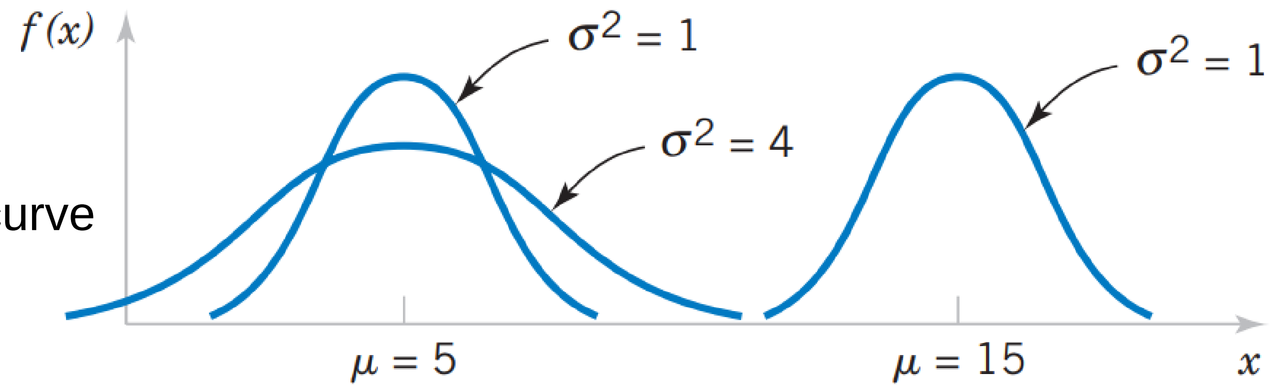
$N(\mu, \sigma^2)$

Related to the Central Limit Theorem (chapter 7)

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

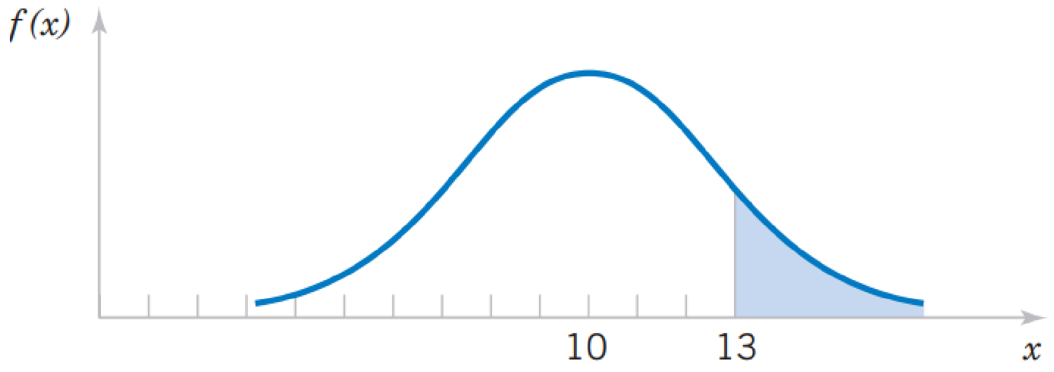
Bell-shaped curve

Not easy to compute with paper and pencil



Normal probability density functions for selected values of μ and σ^2.

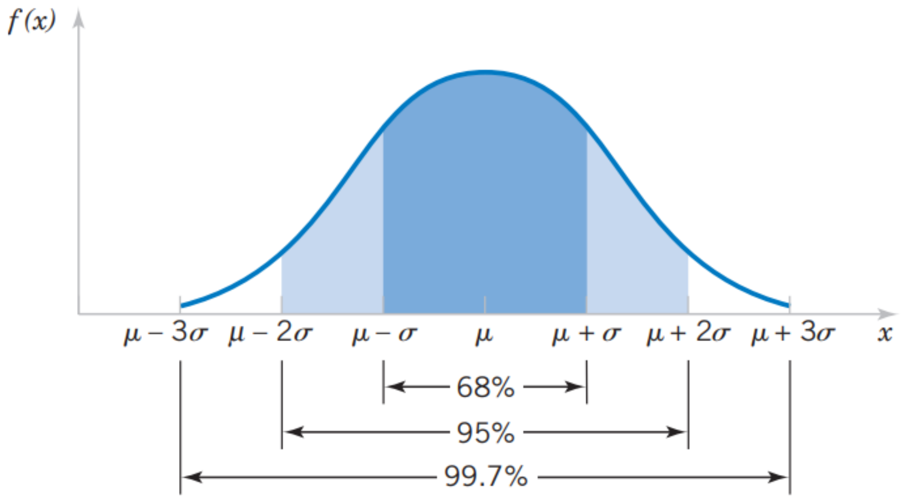
Normal Distribution



Probability that X > 13 for a normal random variable with and $\mu = 10$ and $\sigma^2 = 4$.

For any normal random variable

$$P(\mu - \sigma < X < \mu + \sigma) = 0.6827$$
$$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.9545$$
$$P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.9973$$



6σ is often referred to as the *width* of a normal distribution.

Standard Normal Random Variable

Standardize: If $X \sim N(\mu, \sigma^2)$, then $Z = (X - \mu)/\sigma \sim N(0, 1)$

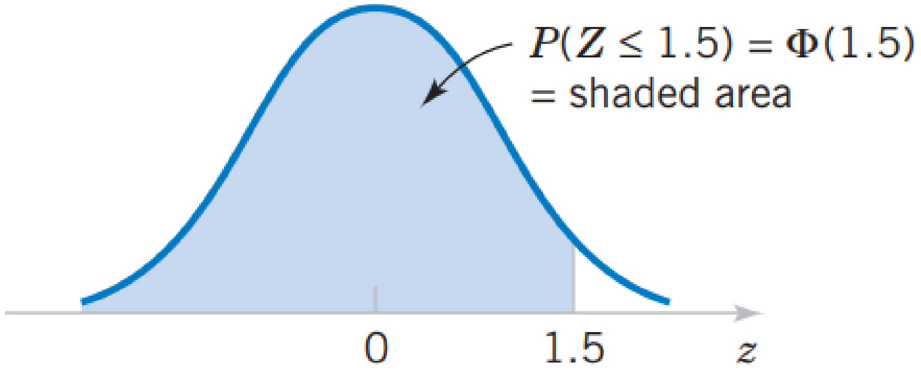
It is the key step to calculating a probability for an arbitrary normal random variable.

$Z \sim N(0, 1)$

Standard normal distribution

Assume Z is a standard normal random variable. Appendix Table III provides probabilities of the form $\Phi(z) = P(Z \leqslant z)$. Read down the z column to the row that equals 1.5. The probability $P(Z \leqslant 1.5)$ is read from the adjacent column, labeled 0.00, to be 0.93319.

$\Phi(z) = P(Z \leqslant z)$



z	0.00	0.01	0.02	0.03
0	0.50000	0.50399	0.50398	0.51197
⋮		⋮		
1.5	0.93319	0.93448	0.93574	0.93699

Using Table III in the text book to find probabilities of Normal distribution

$$\Phi(z) = \boxed{P(Z \le z)} = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}\, du$$

$\Phi(1.34) = P(Z \leqslant 1.34) = 0.909877$

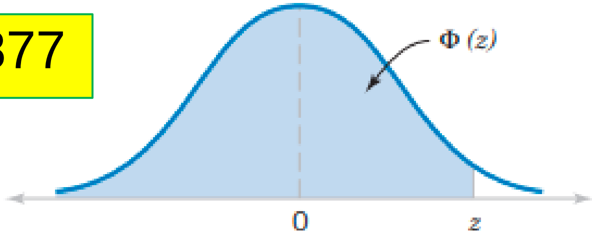
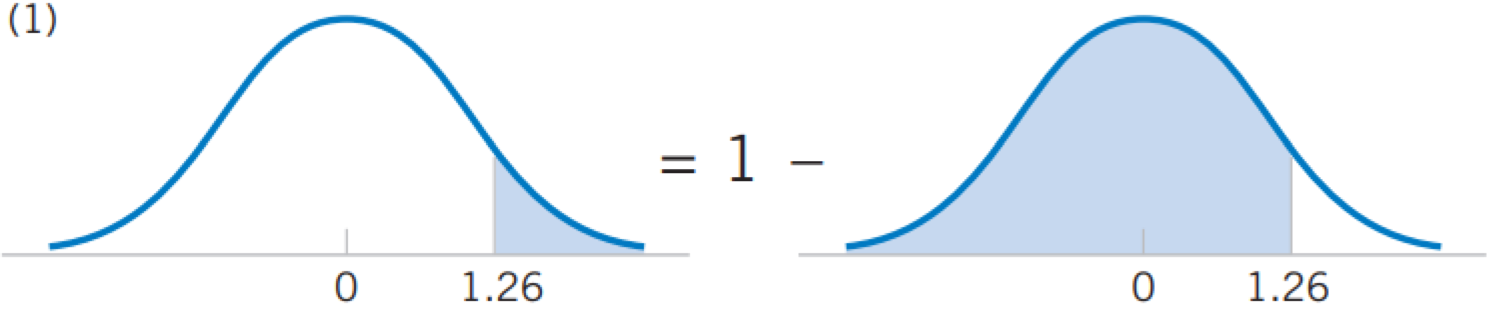


Table III Cumulative Standard Normal Distribution (*continued*)

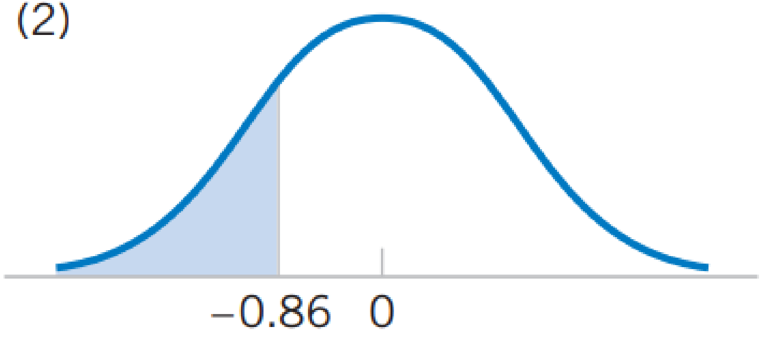
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500000	0.503989	0.507978	0.511967	0.515953	0.519939	0.532922	0.527903	0.531881	0.535856
0.1	0.539828	0.543795	0.547758	0.551717	0.555760	0.559618	0.563559	0.567495	0.571424	0.575345
0.2	0.579260	0.583166	0.587064	0.590954	0.594835	0.598706	0.602568	0.606420	0.610261	0.614092
0.3	0.617911	0.621719	0.625516	0.629300	0.633072	0.636831	0.640576	0.644309	0.648027	0.651732
0.4	0.655422	0.659097	0.662757	0.666402	0.670031	0.673645	0.677242	0.680822	0.684386	0.687933
0.5	0.691462	0.694974	0.698468	0.701944	0.705401	0.708840	0.712260	0.715661	0.719043	0.722405
0.6	0.725747	0.729069	0.732371	0.735653	0.738914	0.742154	0.745373	0.748571	0.751748	0.754903
0.7	0.758036	0.761148	0.764238	0.767305	0.770350	0.773373	0.776373	0.779350	0.782305	0.785236
0.8	0.788145	0.791030	0.793892	0.796731	0.799546	0.802338	0.805106	0.807850	0.810570	0.813267
0.9	0.815940	0.818589	0.821214	0.823815	0.826391	0.828944	0.831472	0.833977	0.836457	0.838913
1.0	0.841345	0.843752	0.846136	0.848495	0.850830	0.853141	0.855428	0.857690	0.859929	0.862143
1.1	0.864334	0.866500	0.868643	0.870762	0.872857	0.874928	0.876976	0.878999	0.881000	0.882977
1.2	0.884930	0.886860	0.888767	0.890651	0.892512	0.894350	0.896165	0.897958	0.899727	0.901475
1.3	0.903199	0.904902	0.906582	0.908241	0.909877	0.911492	0.913085	0.914657	0.916207	0.917736
1.4	0.919243	0.920730	0.922196	0.923641	0.925066	0.926471	0.927855	0.929219	0.930563	0.931888
1.5	0.933193	0.934478	0.935744	0.936992	0.938220	0.939429	0.940620	0.941792	0.942947	0.944083
1.6	0.945201	0.946301	0.947384	0.948449	0.949497	0.950529	0.951543	0.952540	0.953521	0.954486

(1)



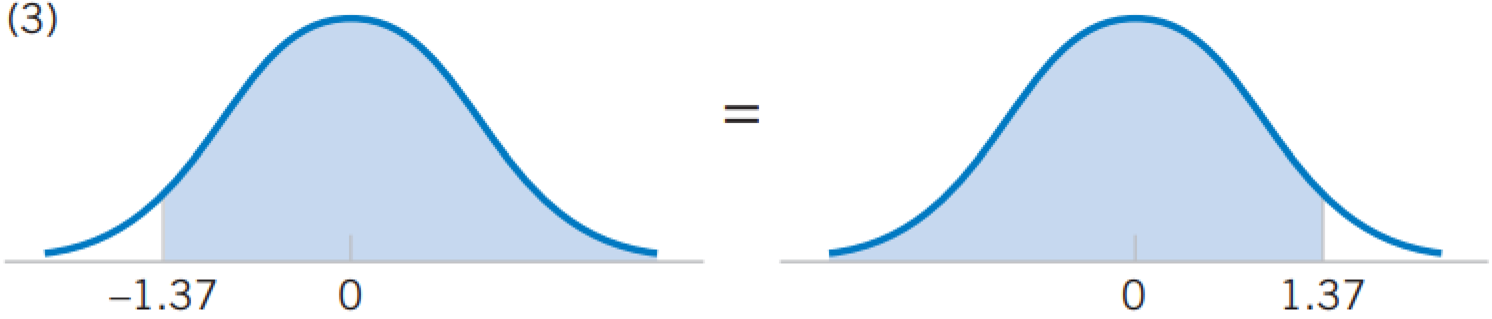
$$P(Z > 1.26) = 1 - P(Z \leq 1.26) = 1 - 0.89616$$
$$= 0.10384$$

(2)



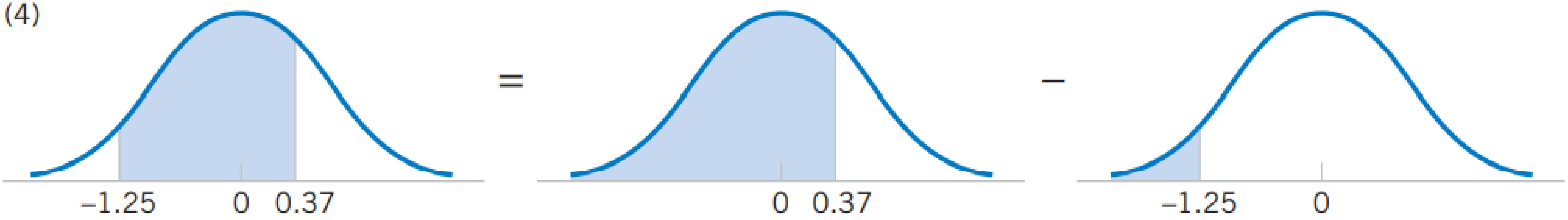
$$P(Z < -0.86) = 0.19490.$$

(3)



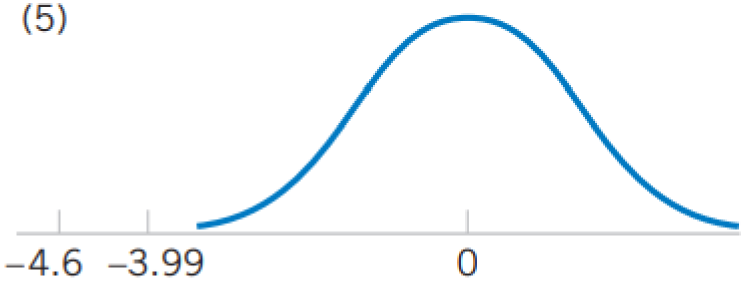
$$P(Z > -1.37) = P(Z < 1.37) = 0.91465$$

(4)



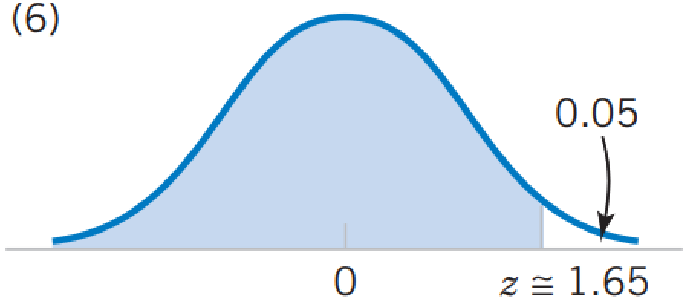
$$P(-1.25 < Z < 0.37) \;=\; P(Z < 0.37) \;-\; P(Z < -1.25)$$

(5)



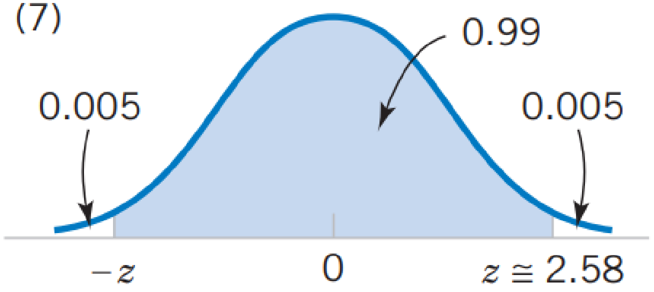
(5) $P(Z < -4.6)$ cannot be found exactly from Appendix Table III. The last entry in the table can be used to find that $P(Z < -3.99) = 0.00003$. So, $0 \approx P(Z < -4.6) < P(Z < -3.99)$.

(6)



(6) Find z such that $P(Z > z) = 0.05$. Equivalently, find z such that $P(Z \leqslant z) = 0.95$

(7)



(7) Find z such that $P(-z < Z < z) = 0.99$.
➔ Find z such that $P(Z \leqslant z) = 0.995$

EXERCISES FOR SECTION 4-6

4-49. Use Appendix Table III to determine the following probabilities for the standard normal random variable Z:

(a) $P(Z < 1.32)$ (b) $P(Z < 3.0)$

(c) $P(Z > 1.45)$ (d) $P(Z > -2.15)$

(e) $P(-2.34 < Z < 1.76)$

4-50. Use Appendix Table III to determine the following probabilities for the standard normal random variable Z:

(a) $P(-1 < Z < 1)$ (b) $P(-2 < Z < 2)$

(c) $P(-3 < Z < 3)$ (d) $P(Z > 3)$

(e) $P(0 < Z < 1)$

4-51. Assume Z has a standard normal distribution. Use Appendix Table III to determine the value for z that solves each of the following:

(a) $P(Z < z) = 0.9$ (b) $P(Z < z) = 0.5$

(c) $P(Z > z) = 0.1$ (d) $P(Z > z) = 0.9$

(e) $P(-1.24 < Z < z) = 0.8$

Standardizing a Normal Random Variable - Ex

Suppose the current measurements in a strip of wire are assumed to follow a normal distribution with a mean of 10 milliamperes and a variance of 4 (milliamperes)^2. What is the probability that a measurement will exceed 13 milliamperes?

Let X denote the current in milliamperes.

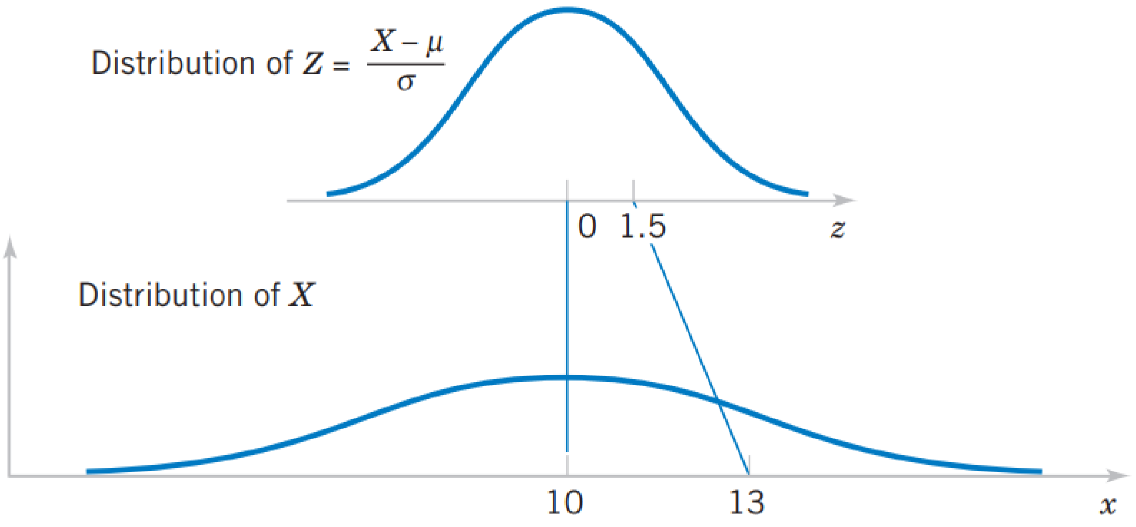
Let Z = (X - 10)/2.

Note that X >13 corresponds to Z >1.5.

Therefore, from Appendix Table III

P(X > 13) = P(Z > 1.5)

= 1 – P(Z $\leqslant$1.5) = 1- 0.93319 = 0.06681



Practical Interpretation: Probabilities for any normal random variable can be computed with a simple transform to a standard normal random variable.

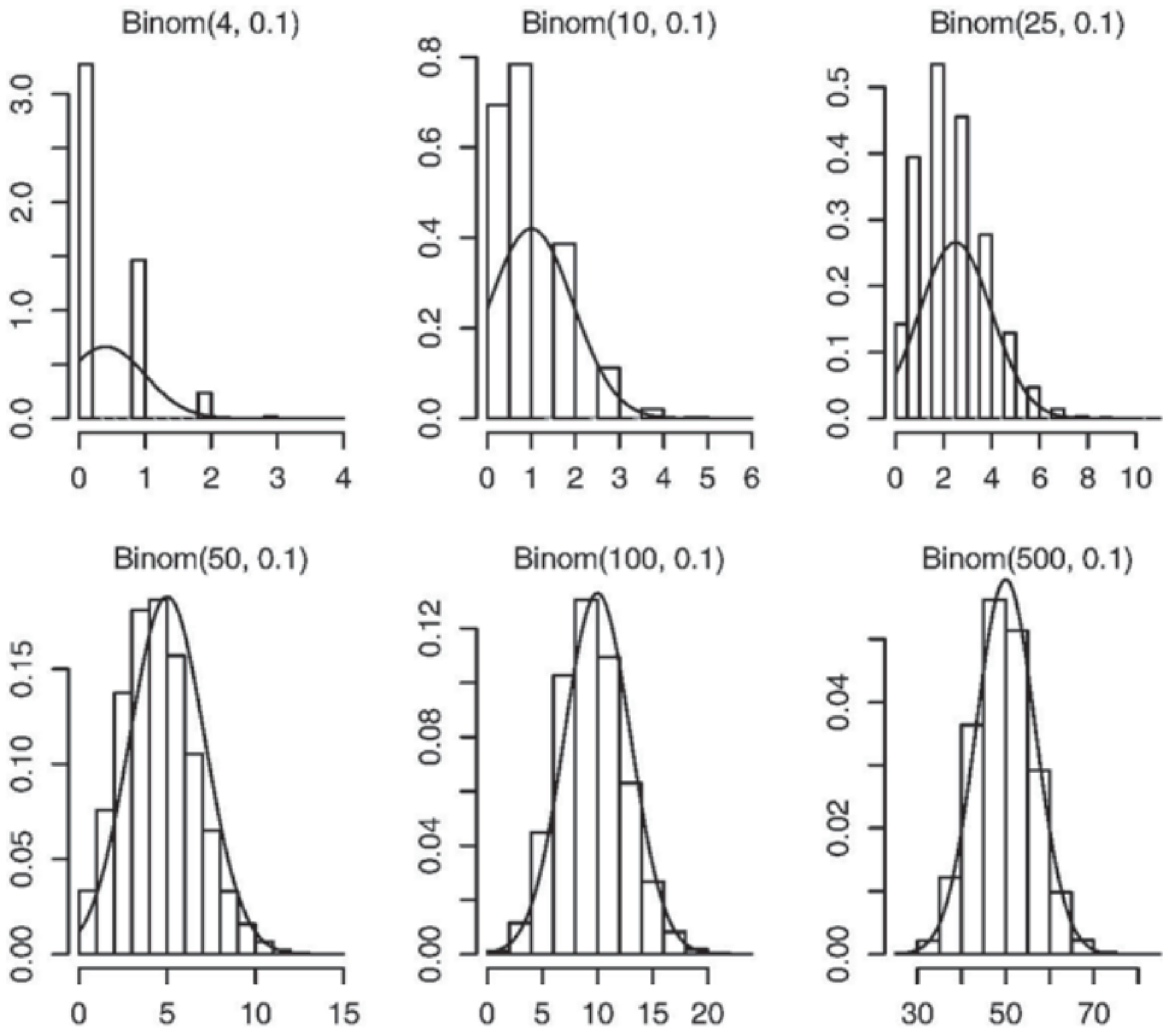
Standardizing to Calculate a Probability

(Normally Distributed Current) Continuing the previous example.

a/ What is the probability that a current measurement is between 9 and 11 milliamperes?

b/ Determine the value for which the probability that a current measurement is below this value is 0.98.

Normal Approximation to the Binomial Distribution



Binom(n, p) ~ N(μ = np, σ^2 = np(1-p))

Good *approximation* with n and p such that

np > 5, n(1 – p) > 5

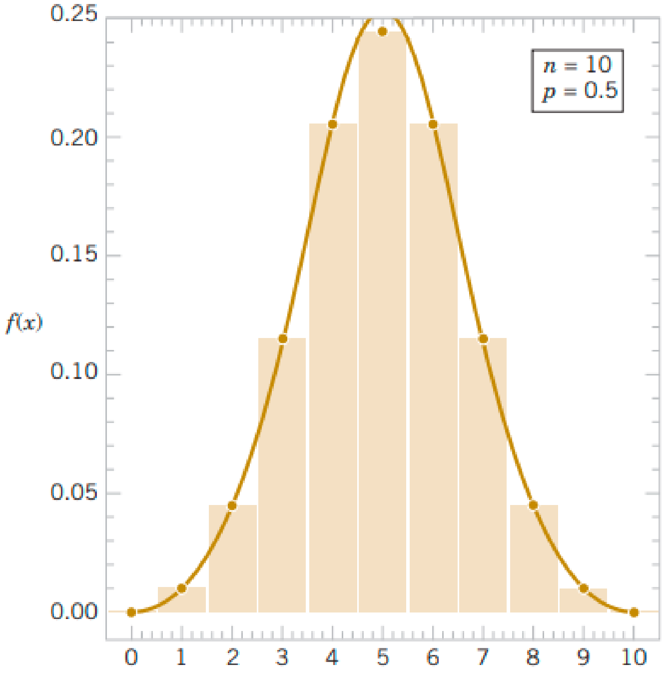
Normal Approximation to the Binomial Distribution

- $X \sim \text{Binom}(n, p)$
- $Z = \dfrac{X - np}{\sqrt{np(1-p)}} \sim N(0, 1)$
- Normal Approximation of Binomial as follow

$$P(X \leqslant x) = P(X \leqslant x + 0.5) = P\left(Z \leqslant \frac{x + 0.5 - np}{\sqrt{np(1-p)}} \right)$$

continuity correction

$$P(x \leqslant X) = P(x - 0.5 \leqslant X) = P\left(\frac{x - 0.5 - np}{\sqrt{np(1-p)}} \leqslant Z \right)$$



Normal Approximation to the Binomial Distribution – Ex

Assume that in a digital communication channel, the number of bits received in error can be modeled by a binomial random variable, and assume that the probability that a bit is received in error is 10^{-5}. If 16 million bits are transmitted, what is the probability that 150 or fewer errors occur?

Let the random variable X denote the number of errors. Then X is a binomial random variable with n = 16×10^6 and p = 10^{-5}, and

$$P(X \le 150) = \sum_{x=0}^{150} \binom{16{,}000{,}000}{x} \left(10^{-5}\right)^x \left(1 - 10^{-5}\right)^{16{,}000{,}000-x}$$

DIFFICULT!

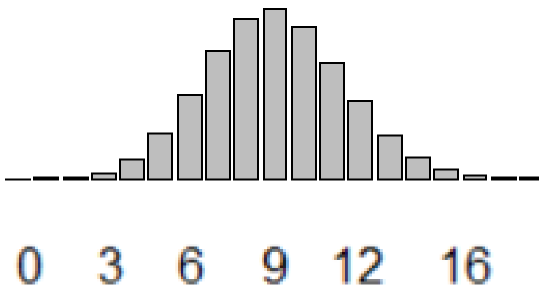
Use normal distribution to approximate

$$P(X \le 150) = P(X \le 150.5) = P\left(\frac{X - 160}{\sqrt{160\left(1 - 10^{-5}\right)}} \le \frac{150.5 - 160}{\sqrt{160\left(1 - 10^{-5}\right)}} \right) \approx P(Z \le -0.75) = 0.227$$

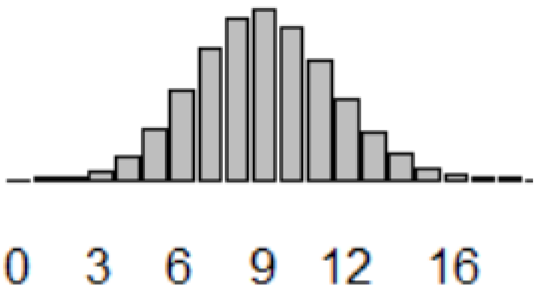
Ex. A new computer virus attacks a folder consisting of 200 files. Each file gets damaged with probability 0.2 independently of other files. What is the probability that fewer than 50 files get damaged?

Conditions for approximating hypergeometric and binomial probabilities

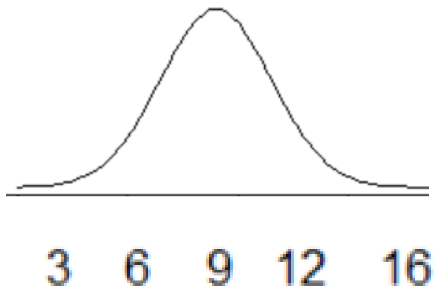
hypergometric distribution	$\approx$	binomial distribution	$\approx$	normal distribution
	$\frac{n}{N} < 0.1$		$np > 5$ $n(1-p) > 5$	



N = 400, K = 120, n = 30



n = 30, p = 0.3



$\mu = np = 9$, $\sigma^2 = np(1\text{-}p) = 6.3$

Normal Approximation to the Poisson Distribution

- X is a *Poisson* random variable with $E(X) = \lambda$ and $V(X) = \lambda$
- $Z = \dfrac{X - \lambda}{\sqrt{\lambda}} \sim N(0, 1)$
- Normal Approximation to Poisson as follow

$$P(X \leqslant x) = P(X \leqslant x + 0.5) = P\left(Z \leqslant \frac{x + 0.5 - \lambda}{\sqrt{\lambda}} \right)$$

$$P(x \leqslant X) = P(x - 0.5 \leqslant X) = P\left(\frac{x - 0.5 - \lambda}{\sqrt{\lambda}} \leqslant Z \right)$$

Normal Approximation to the Poisson Distribution – Ex

Ex. Assume that the number of asbestos particles in a squared meter of dust on a surface follows a Poisson distribution with a mean of 1000. If a squared meter of dust is analyzed, what is the probability that 950 or fewer particles are found?

Let X be the number of particles found. Then X ~ Pois(1000)

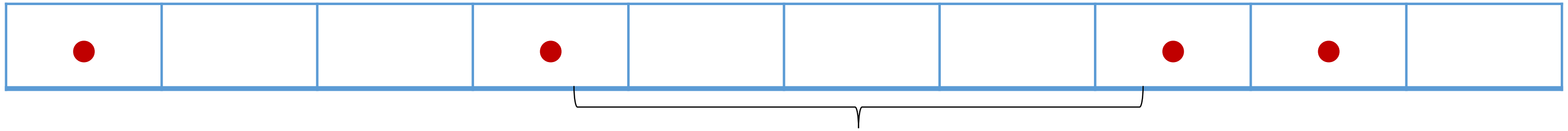
$$P(X \le 950) = \sum_{x=0}^{950} \frac{e^{-1000} 1000^x}{x!}$$

DIFFICULT!

AN APPROXIMATION

$$P(X \le 950) = P(X \le 950.5) \approx P\left(Z \le \frac{950.5 - 1000}{\sqrt{1000}}\right) = P(Z \le -1.57) = 0.058$$

Exponential Distribution



X = *waiting time* for the next event

Consider the sequence of “rare” events, where the number of occurrences has Poisson distribution with parameter λ.

Event “the time X until the next event is greater than x” can be rephrased as “zero events occur by the time x.”

Let N be the number of events during the time interval $[0, x]$ ➔ $N \sim \text{Poisson}(\lambda x)$.

$P(X > x) = P(N = 0) = e^{-\lambda x}(\lambda x)^0/0! = e^{-\lambda x}$

➔ $F(x) = P(X \leqslant x) = 1\ P(X > x) = 1 - e^{-\lambda x}$

➔ $f(x) = F'(x) = \lambda e^{-\lambda x}$

Exponential Distribution

$$f(x) = \lambda e^{-\lambda x} \quad \text{for} \quad x > 0$$

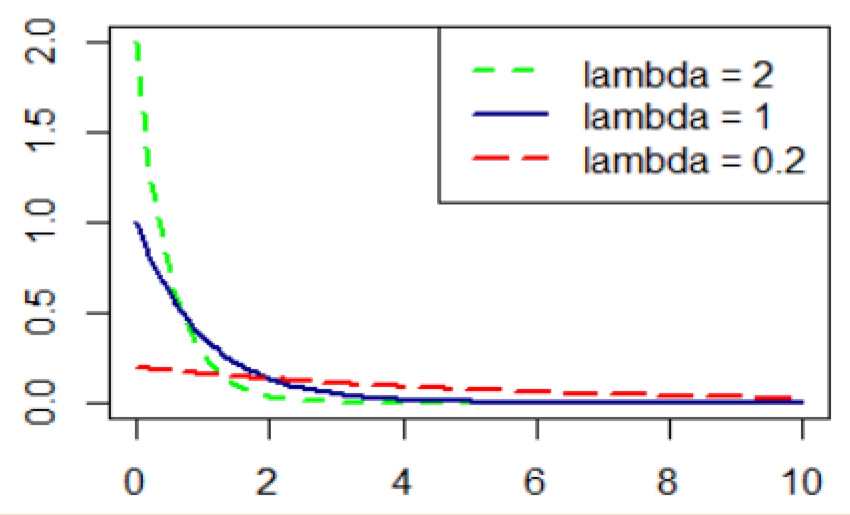
Exp. distribution

Often used to model *waiting time* between rare events: time between telephone calls, accidents, etc. When the number of events is Poisson, the time between events is Exponential.

$$\sigma^2 = V(X) = \frac{1}{\lambda^2}$$

$$F(x) = \int_0^x \lambda e^{-\lambda t}\, dt = \lambda \left(\frac{-1}{\lambda} e^{-\lambda t} \right) \Bigg|_0^x = 1 - e^{-\lambda x}$$

$$\mu = E(X) = \frac{1}{\lambda}$$



For small lambda, it's usual that we need a long waiting time for the next event

Exponential Distribution – Ex

Ex. Questions are sent to the chatbot of a Student Service at an average rate of 12 questions per day.

a/ What is the probability that the next question is sent within 1 hour minutes?

b/ What is the expected time between questions?

Questions arrivals represent rare events, thus the time X between them has an exponential distribution with $\lambda = 12$ (questions per day).

a/ 1hr = 1/24 day, $P(X \leqslant 1/24) = F(1/24) = 1 - e^{-\lambda x} = 1 - e^{-1/2} = 0.3935$

b/ $E(X) = 1/\lambda = 1/12$ (day) = 2 hours

Exponential Distribution – Ex

Suppose that X has an exponential distribution with a mean of 10. Determine the following:

a/ $P(X < 5)$

b/ $P(X < 15 \mid X > 10)$

$E(X) = 1/\lambda = 10 \Rightarrow \lambda = 1/10$

a/ $P(X < 5) = F(5) = 1 - e^{-5\lambda} = 1 - e^{-1/2} = 0.3935$

b/ $P(X < 15 \mid X > 10) = P(10 < X < 15)/P(X > 10)$

$= (F(15) – F(10))/(1 – F(10)) = [(1 - e^{-15\lambda}) – (1 - e^{-10\lambda})]/e^{-10\lambda}$

$= e^{-10\lambda}(1 - e^{-5\lambda})/e^{-10\lambda} = 1 - e^{-5\lambda} = 0.3935 = P(X < 5)$

Lack of Memory Property

For an exponential random variable X,

$$P(X < t_1 + t_2 \mid X > t_1) = P(X < t_2)$$

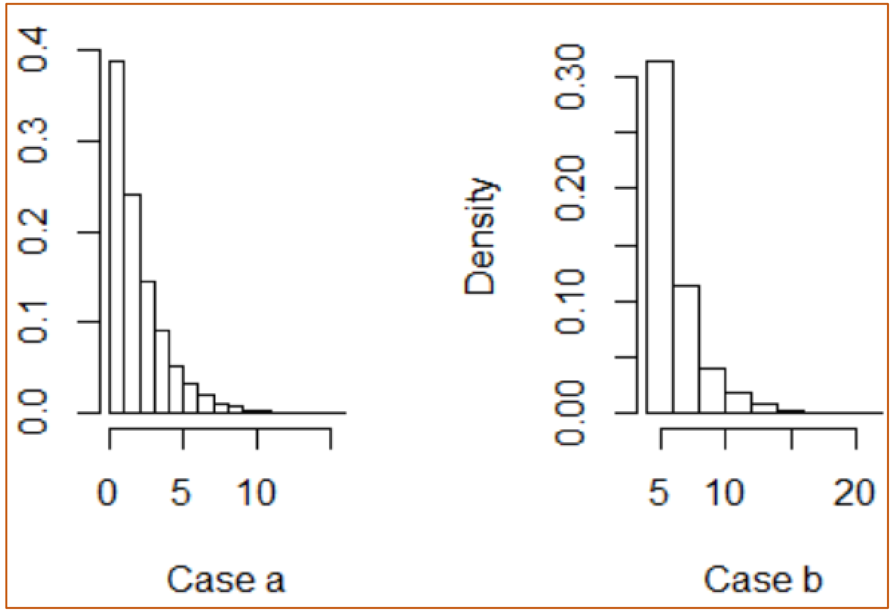
<u>**Ex.**</u> The *time between* the arrival of electronic messages at your computer is *exponentially distributed* with a mean of two hours ($\lambda = 1/2$).

a/ What is the probability that you do not receive a message during a two-hour period?

$P(X > 2) = 1 - P(X \leqslant 2) = e^{-2\lambda} = e^{-1} = 0.368$

b/ If you have not had a message in the last four hours, what is the probability that you do not receive a message in the next two hours?

$P(X > 6 \mid X > 4) = P(X > 2)$



Exercises

$(5x + 19) \bmod 109$

$(7x + 29) \bmod 109$

$(9x + 39) \bmod 109$

SUMMARY

Continuous variables can take any value of an interval in $\mathbb{R}$