21DS636-SM

Probability Theory

Special Continuous Distribution

Dr. Deepa Gupta, Professor
Human Language Technology Lab
Department of Computer Science & Engineering
ASE-Bangalore

Summary From Last Time

Discrete Random Variables

Binomial Distribution

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Probability of number of k success when you do n Bernoulli trials

Mean and variance $\mu = np$ $\sigma^2 = np(1-p)$

Poisson distribution

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$$,x=0,1.2.3……..

Probablily of k randomly occurring events, given average number is λ

Mean and variance $\mu = \lambda$ $\sigma^2 = \mathrm{var}(X) = \lambda$

- It is approximation to Binomial when *n* is large and *p* is small

Multinomial Distribution

$$f(x_1,\ldots,x_k;n,p_1,\ldots,p_k) = \Pr(X_1 = x_1 \text{ and } \ldots \text{ and } X_k = x_k)$$
$$= \begin{cases} \dfrac{n!}{x_1!\cdots x_k!} p_1^{x_1} \times \cdots \times p_k^{x_k}, & \text{when } \sum_{i=1}^k x_i = n \\ 0 & \text{otherwise,} \end{cases}$$

for non-negative integers $x_1, \ldots, x_k$.

Special Probability distribution of Random Variable

- Discrete Random Variable:
    - Bernoulli distribution
    - Binomial Distribution
    - Multinominal Distribution
    - Poisson Distribution
    - …..etc.

- Continuous Random Variable:
    - Uniform Distribution
    - Exponential Distribution
    - Normal Distribution/Gaussian Distribution
    - ……..etc.

Uniform Distribution

A **Uniform Distribution** has equally likely values over the range of possible outcomes.

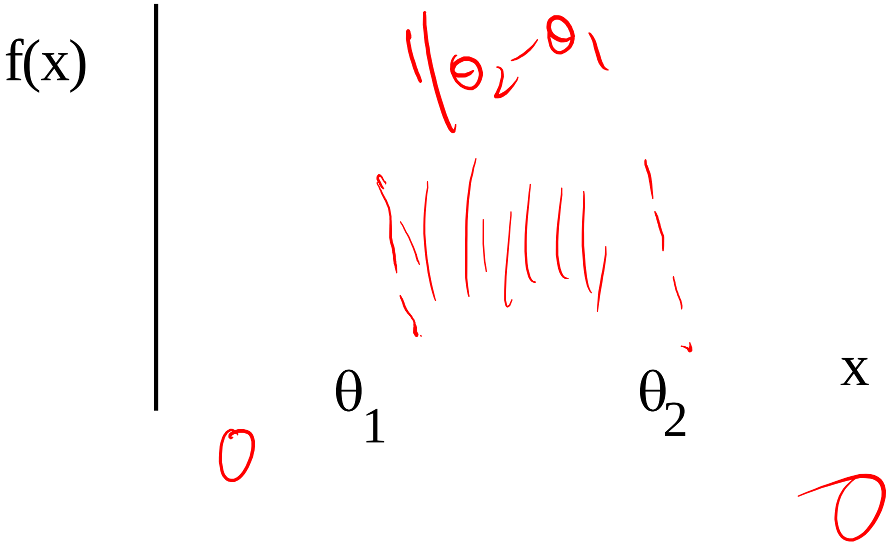
A graph of the uniform probability distribution is a rectangle with area equal to 1.

Uniform distribution

The continuous random variable X has the Uniform distribution between θ_1 and θ_2, with $\theta_1 < \theta_2$ if

$X \sim U(\theta_1, \theta_2)$, for short.

$$f(x) = \begin{cases} \dfrac{1}{\theta_2 - \theta_1} & \theta_1 \le x \le \theta_2 \\ 0 & \text{otherwise} \end{cases}$$



Mean and variance: for $U(\theta_1, \theta_2)$

$$\mu = \frac{(\theta_1 + \theta_2)}{2} \qquad \sigma^2 = \frac{(\theta_2 - \theta_1)^2}{12}$$

Uniform Distribution

Area to the Left of x:

$$P(X<x) = (x-a)\left(\frac{1}{b-a}\right) = \frac{x-a}{b-a}$$

$a \to x$

Area to the Right of x:

$$P(X>x) = (b-x)\left(\frac{1}{b-a}\right) = \frac{b-x}{b-a}$$

$x \to b$

$P(X \le x) = 1$

Area Between c and d:

$$P(c<x<d) = (\text{base})(\text{height}) = (d-c)\left(\frac{1}{b-a}\right)$$

Uniform: $X \sim U(a, b)$ where $a < x < b$

Example1 : The amount of time, in minutes, that a person must wait for a bus is uniformly distributed between zero and 15 minutes, inclusive.

1. What is the probability that a person waits fewer than 12.5 minutes?
2. On the average, how long must a person wait? Find the standard deviation σ.
3. Ninety percent of the time, the time a person must wait falls below what value?

Sol

X : person waiting time

$X \sim U(0, 15) \Rightarrow f(x) = \begin{cases} \frac{1}{15-0} = \frac{1}{15} & 0 \le x \le 15 \\ 0 & \text{otherwise} \end{cases}$

1) $P(X < 12.5) = \int_0^{12.5} \frac{1}{15} dx = \frac{1}{15}(12.5 - 0) = \frac{12.5}{15}$

2) $E[X] = \frac{a+b}{2} = \frac{15}{2} = 7.5$ mins

$\sigma = \sqrt{\frac{(b-a)^2}{12}} = \sqrt{\frac{15^2}{12}} = \boxed{}$

$P(X \le K) = 0.90$

$\Rightarrow \frac{K-0}{15} = 0.90$

$\Rightarrow K = 0.90 \times 15 = \boxed{}$

Example 2: The total duration of baseball games in the major league in the 2011 season is uniformly distributed between 447 hours and 521 hours inclusive.

1. Find *a* and *b* and describe what they represent.
2. Write the distribution.
3. Find the mean and the standard deviation.
4. What is the probability that the duration of games for a team for the 2011 season is between 480 and 500

The Exponential Distribution

- The exponential distribution is often concerned with the **amount of time until some specific random event occurs.**
  - For example, the amount of time (beginning now) until an earthquake occurs has an exponential distribution.
  - Other examples include the length, in minutes, of long-distance business telephone calls, and the amount of time, in months, a car battery lasts.
- **Values for an exponential random variable occur in the following way**.
  - There are fewer large values and more small values.
  - For example, the amount of money customers spend in one trip to the supermarket follows an exponential distribution. There are more people who spend small amounts of money and fewer people who spend large amounts of money.
  - The value of the change that you have in your pocket or purse approximately follows an exponential distribution
- The exponential distribution is widely used in the field of reliability.
  - Reliability deals with the amount of time a product lasts.

The Exponential Distribution

- What is the probability distribution for the time to the first random event/ time between random events? $\Rightarrow$ **Exponential distribution**

- The common element associated with these random variables is that they can only assume positive values.
  - patient survival time after the diagnosis of a particular cancer,
  - Time until the failure of a part
  - the time between births at a hospital,

The Exponential Distribution

- *X* is a **continuous random variable** since time is measured.

Probability Density function

$$f(x) = \frac{1}{\mu} e^{-x/\mu} \text{ for } x \geq 0, \mu > 0$$

μ = mean

the decay parameter, m=1/ μ

e = 2.71828

Mean E(X) = μ

Variance V(X) = μ^2

- The distribution notation is *X* ~ *Exp*(*m*) =Exp(**1/ μ**)

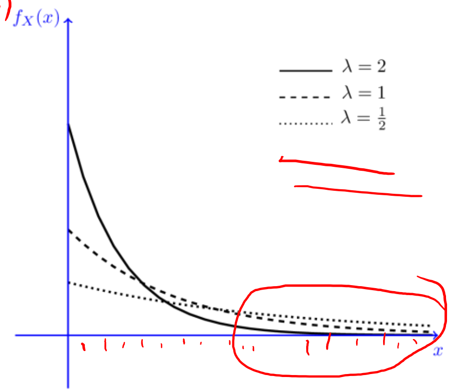
Cumulative probabilities

$$\boxed{P(X \leq x)} = 1 - e^{-x/\mu}$$

$F(X \leq x) = \int_0^x f(t)\,dt$

Where x is some specific value of *X*

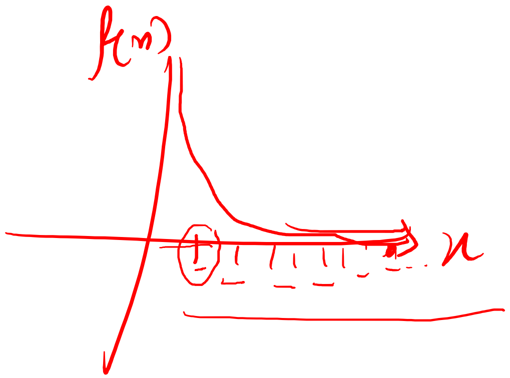
f(x)



Example 6: **Is it exponential?**

Which of the following random variables is best modelled by an exponential distribution?

1. The distance between defects in an optical fibre
2. The number of days between someone winning the National Lottery
3. The number of fuses that blow in the UK today
4. The hours of sunshine in Brighton this week assuming an average of 7.2hrs/day



Solution: Is it exponential?

Which of the following random variables is best modelled by an exponential distribution?

1. The distance between defects in an optical fibre

   - YES: continuous distribution that is the separation between independent random events (the location of the defects)

2. The number of days between someone winning the National Lottery

   - NO: continuous (if you allow fractional days), but draws happen *regularly* on a schedule

$\frac{1}{\mu} e^{-x/\mu}$

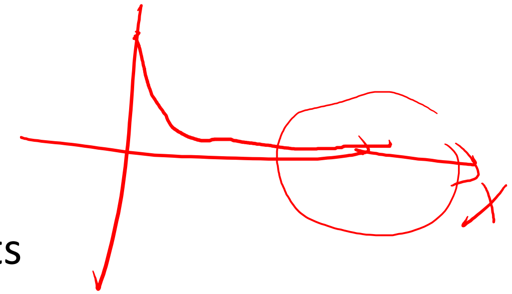
$n \to \infty$

3. The number of fuses that blow in the UK today

   - NO: this is a *discrete* distribution – the number of events is a Poisson distribution (exponential is the distribution of times *between* events)

4. The hours of sunshine in Brighton this week assuming an average of 7.2hrs/day

   - NO: This is a continuous variable, but not the time between independent random events



$\lim_{x\to\infty} e^{-x} = 0$ $\quad e^0 = 1$

Example2: Let X = amount of time (in minutes) a postal clerk spends with his or her customer. The time is known to have an exponential distribution with the average amount of time equal to four minutes.

1. Find the probability that a clerk spends four to five minutes with a randomly selected customer.
2. Half of all customers are finished within how long? (Find the 50th percentile).
3. Which is larger, the mean or the median?

Sol

X = amount of time (in min)

$X \sim Exp(1/4) \Rightarrow m = 1/\mu = \frac{1}{4}$

$E[X] = 4$

$f(x) = \begin{cases} \frac{1}{\mu} e^{-x/\mu} & x \geq 0 \\ 0 & \text{otherwise} \end{cases} = \begin{cases} \frac{1}{4} e^{-x/4} & x \geq 0 \\ 0 & \end{cases}$

$\int e^{ax} dx = \frac{e^{ax}}{a}$

$P(4 \leq X \leq 5) = \int_4^5 \frac{1}{4} e^{-x/4} dx = \frac{1}{4} \left. \frac{e^{-x/4}}{-1/4} \right|_4^5 = -\left[e^{-5/4} - e^{-4/4} \right] = e^{-1} - e^{-5/4}$

2) $P(X \leq x) = 0.50$

$\Rightarrow \int_0^x \frac{1}{4} e^{-t/4} dt = 0.50$

$\Rightarrow e^{-x/4} = 0.50$

$\Rightarrow 0.50 = e^{-x/4}$

$\Rightarrow \ln 0.50 = -x/4$

$\Rightarrow -4\ln(0.50) = x$

$\Rightarrow 2.77$

Memoryless of the Exponential Distribution

- **Memoryless property**

> If X is exponential with parameter $\lambda > 0$, then X is a *memoryless* random variable, that is
>
> $$P(X > x + a \mid X > a) = P(X > x), \quad \text{for } a, x \geq 0.$$

- In fact, exponential distribution is the only continuous distribution with this property

Example7 : Let X = amount of time (in minutes) a postal clerk spends with his or her customer. The time is known to have an exponential distribution with the average amount of time equal to four minutes. Suppose a customer has spent four minutes with a postal clerk. What is the probability that he or she will spend at least an additional three minutes with the postal clerk?

Sol //

$P(X > 3+4 \mid X > 4)$

$=$

$P(X > 3+4 \mid X > 4)$

$= P(X > 3)$

$= 1 - P(X \le 3)$

$= 1 - e^{-x/\mu}$

$= \boxed{1 - e^{-3/4}}$

$x = 3$

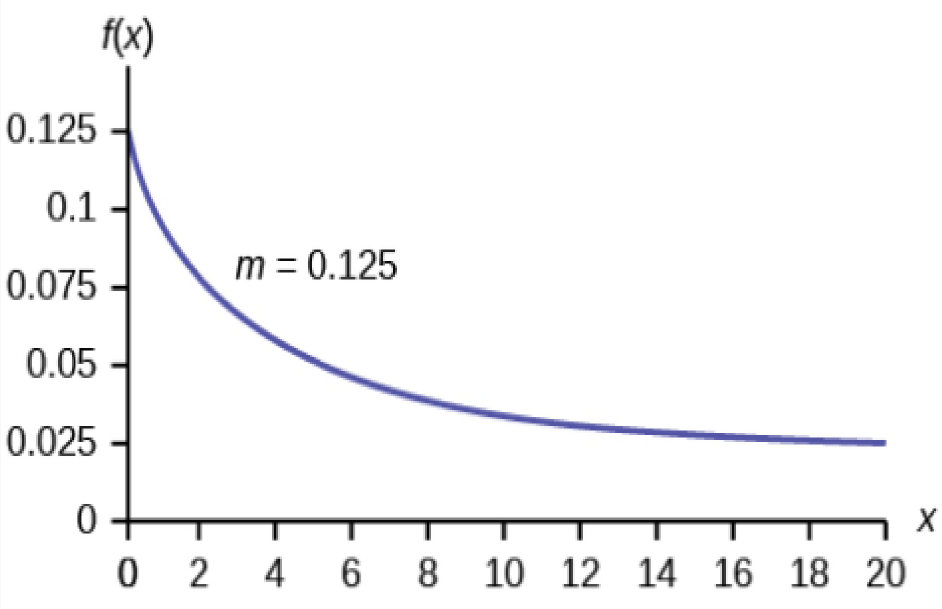
$\mu = 4$

Example 1:

The amount of time spouses shop for anniversary cards can be modeled by an exponential distribution with the average amount of time equal to eight minutes. Write the distribution, state the probability density function, and graph the distribution.

Solution:

$X \sim Exp(0.125)$; $f(x) = 0.125e^{-0.125x}$;



Example: **Reliability:** The time till failure of an electronic component has an Exponential distribution and it is known that 10% of components have failed by 1000 hours.

(a) What is the probability that a component is still working after 5000 hours?

(b) Find the mean and standard deviation of the time till failure.

Answer

Mean E(Y) = $\boldsymbol{\mu = 1/m}$

Let Y = time till failure in hours; $f(y) = me^{-my}$.

(a) **First we need to find m** $P(Y \leq 1000) = \int_0^{1000} me^{-my}$

$$= [-e^{-my}]_0^{1000} = 1 - e^{-1000m}$$

$$P(Y \leq 1000) = 0.1 \Rightarrow \quad 1 - e^{-1000m} = 0.1$$

$$\Rightarrow e^{-1000m} = 0.9$$

$$\Rightarrow -1000m = \ln 0.9 = -0.10536 \quad \Rightarrow m \approx 1.05 \times 10^{-4}$$

If Y is the time till failure, the question asks for $P(Y > 5000)$:

$$P(Y > 5000) = \int_{5000}^{\infty} me^{-my} dy$$

$$= [-e^{-my}]_{5000}^{\infty} \quad = e^{-5000m} \approx 0.59$$

(b) Find the mean and standard deviation of the time till failure.

Answer: Mean = $1/m$ = 9491 hours.

Standard deviation =$\sqrt{\text{Variance}}$

= $\sqrt{\frac{1}{m^2}} = 1/m$ = 9491 hours

Relationship Between Poisson and Exponential Distribution

- There is an interesting relationship between the exponential distribution and the Poisson distribution.
  - Suppose that the time that elapses between two successive events follows the exponential distribution with a mean of μ units of time.
  - Also assume that these times are independent, meaning that the time between events is not affected by the times between previous events.
    - If these assumptions hold, then the number of events per unit time follows a Poisson distribution with mean $\lambda = 1/\mu$.
    - **Conversely, if the number of events per unit time follows a Poisson distribution, then the amount of time between events follows the exponential distribution.**

Relationship between the Poisson and Exponential Distributions

- The Poisson distribution provides an appropriate description of the number of occurrences per interval

- The exponential distribution provides an appropriate description of the length of the interval between occurrences

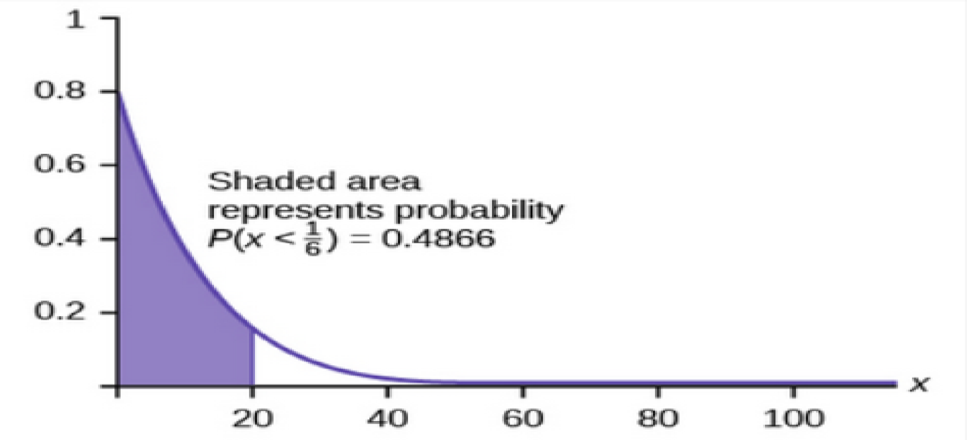
Example:

At a police station in a large city, calls come in at an average rate of four calls per minute. Assume that the time that elapses from one call to the next has the exponential distribution. Take note that we are concerned only with the rate at which calls come in, and we are ignoring the time spent on the phone. We must also assume that the times spent between calls are independent. This means that a particularly long delay between two calls does not mean that there will be a shorter waiting period for the next call. We may then deduce that the total number of calls received during a time period has the Poisson distribution.

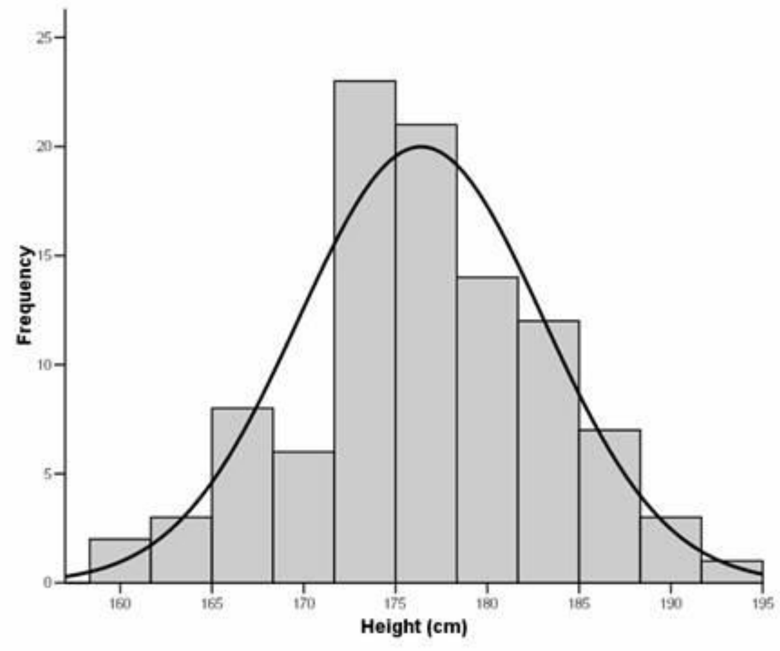
1. Find the average time between two successive calls.
2. Find the probability that after a call is received, the next call occurs in less than ten seconds.
3. Find the probability that exactly five calls occur within a minute.
4. Find the probability that less than five calls occur within a minute.
5. Find the probability that more than 40 calls occur in an eight-minute period.

Solution:

1. On average there are four calls occur per minute, so 15 seconds, or $\frac{15}{60}$ = 0.25 minutes occur between successive calls on average.
2. Let T = time elapsed between calls. From part a, $\mu = 0.25$, so m = $\frac{1}{0.25}$ = 4. Thus, T ~ Exp(4). The cumulative distribution function is $P(T < t) = 1 - e^{-4t}$. The probability that the next call occurs in less than ten seconds (ten seconds = 1/6 minute) is $P(T < \frac{1}{6}) = 1 - e^{-4\frac{1}{6}} \approx 0.4866$



3. Let X = the number of calls per minute. As previously stated, the number of calls per minute has a Poisson distribution, with a mean of four calls per minute. Therefore, $X \sim Poisson(4)$, and so $P(X = 5) = \frac{4^5 e^{-4}}{5!} \approx 0.1563$. (5! = (5)(4)(3)(2)(1))
4. Keep in mind that X must be a whole number, so $P(X < 5) = P(X \leq 4)$.
   To compute this, we could take $P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)$.
   Using technology, we see that $P(X \leq 4) = 0.6288$.
5. Let Y = the number of calls that occur during an eight minute period.
   Since there is an average of four calls per minute, there is an average of (8)(4) = 32 calls during each eight minute period.
   Hence, $Y \sim Poisson(32)$. Therefore, $P(Y > 40) = 1 - P(Y \leq 40) = 1 - 0.9294 = 0.0707$.



Normal distribution

- The normal, a continuous distribution, is the most important of all the distributions.
- **It is widely used and even more widely abused**.
- Its graph is bell-shaped.

Normal distribution

The continuous random variable X has the Normal distribution if the pdf is:

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

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

$X \sim N(\mu, 0.5)$

$P(X \leq 5.5)$

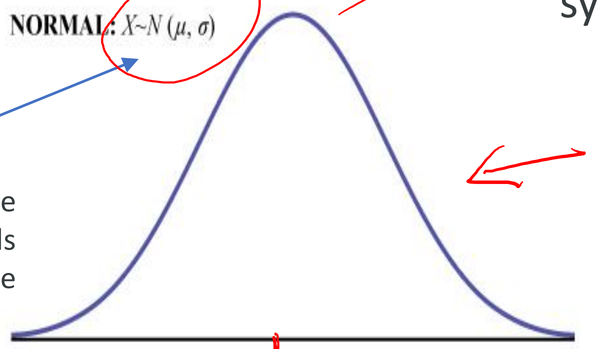
$= \int_{-\infty}^{5.5} \ldots dx$

μ: mean σ: standard deviation

- The probability density function is a rather complicated function. **Do not memorize it**. It is not necessary.
- The curve is symmetrical about a vertical line drawn through the mean, μ.
- In theory, the **mean is the same as the median**, because the graph is symmetric about μ.

Note: *The distribution is also sometimes called a Gaussian distribution*

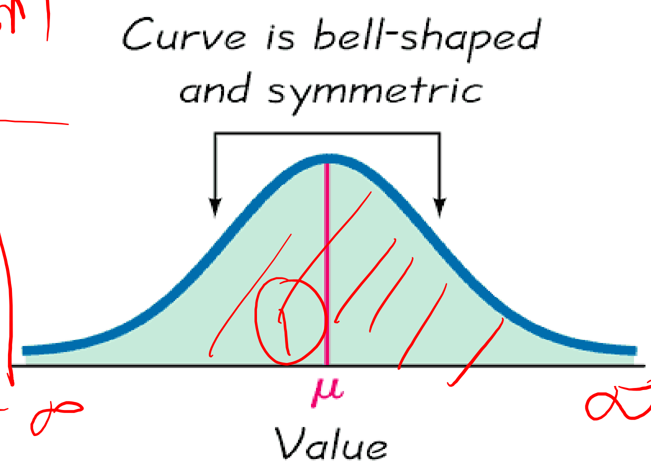
$N(\mu, \sigma)$



As the notation indicates, the normal distribution depends only on the mean and the standard deviation

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

$f(x)$



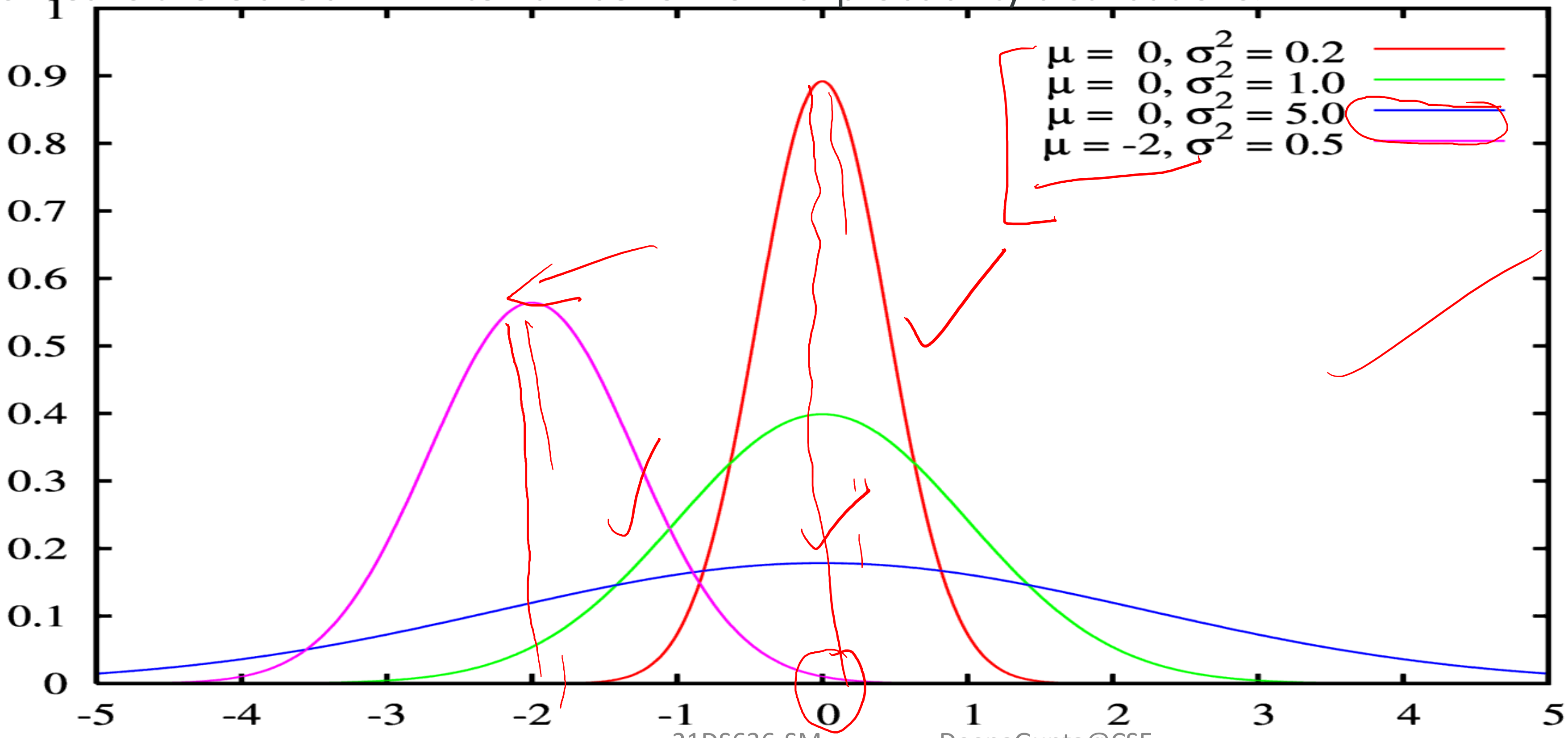
$-\infty$ ∞ x

$\int e^{-x^2} dx$

$\int e^{-x/4} dx = \frac{e^{-x/4}}{-1/4}$

Normal Distributions

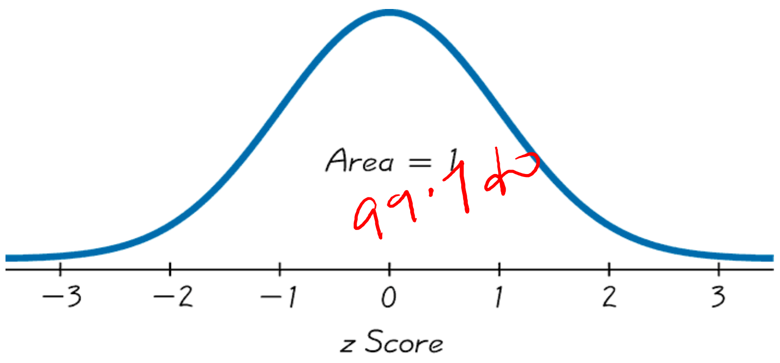
- Since the area under the curve must equal one, a change in the standard deviation, **σ, causes a change in the shape of the curve**; the curve becomes fatter or skinnier depending on σ.
- A change in **μ causes the graph to shift to the left or right**.
- This means there are an infinite number of normal probability distributions.



Standard Normal Distribution

- The **Standard Normal Distribution** is a normal probability distribution that has a mean of 0 and a standard deviation of 1.

$$\mu = 0, \quad \sigma = 1$$



Z-Scores

If X is a normally distributed random variable and $X \sim N(\mu, \sigma)$, then the z-score is:

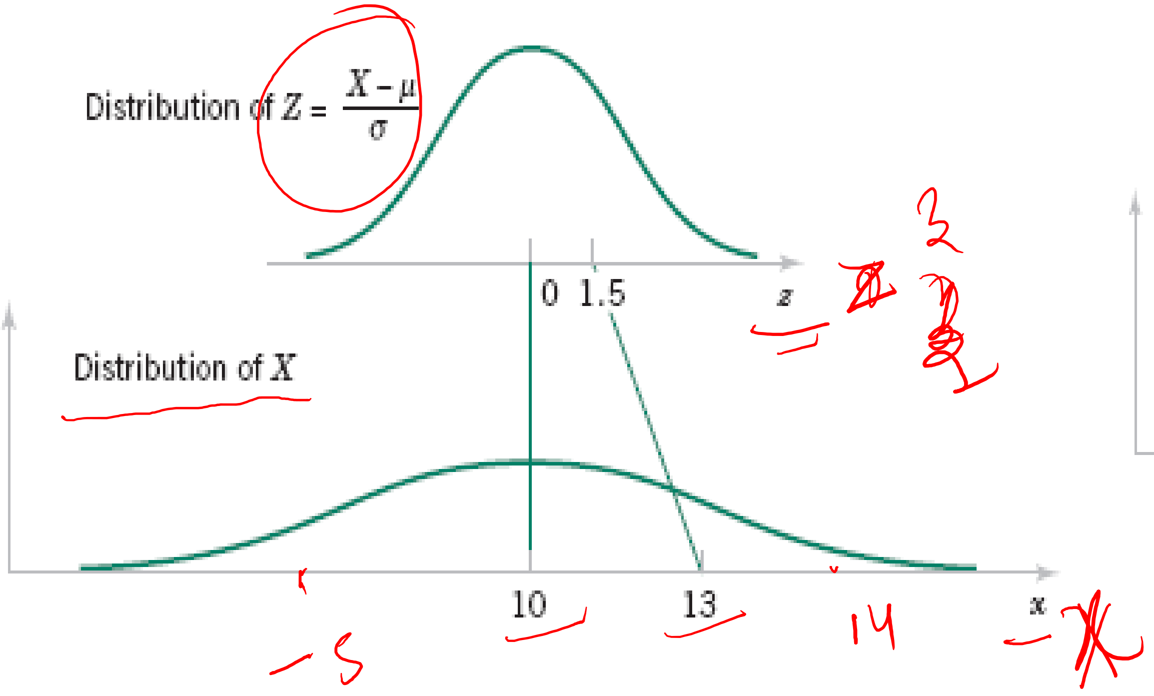
$$\boxed{z = \frac{x - \mu}{\sigma}}$$

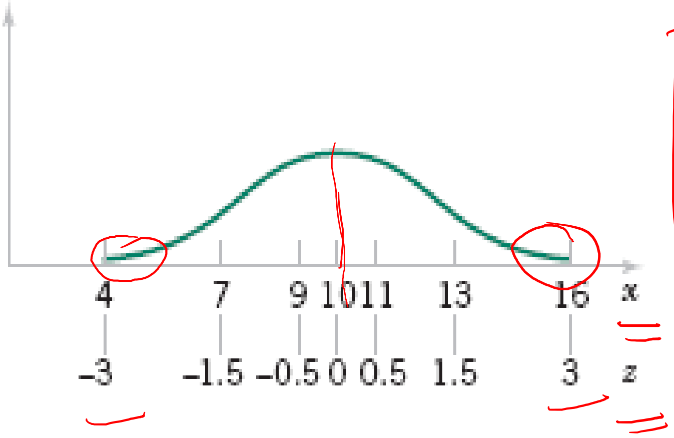
$$Z = \frac{X - \mu}{\sigma} \qquad X \sim N(\mu, \sigma^2)$$

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

- **The *z*-score tells you how many standard deviations the value *x* is above (to the right of) or below (to the left of) the mean, μ.**
- Values of *x* that are larger than the mean have positive *z*-scores, and values of *x* that are smaller than the mean have negative *z*-scores.
- If *x* equals the mean, then *x* has a *z*-score of zero.

Standard Normal Distribution





$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2} \, dz$$

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

The Empirical Rule

If X is a random variable and has a normal distribution with mean μ and standard deviation σ, then the **Empirical Rule** says the following:

- About 68% of the x values lie between -1σ and $+1\sigma$ of the mean μ (within one standard deviation of the mean).
- About 95% of the x values lie between -2σ and $+2\sigma$ of the mean μ (within two standard deviations of the mean).
- About 99.7% of the x values lie between -3σ and $+3\sigma$ of the mean μ(within three standard deviations of the mean). Notice that almost all the$x$ values lie within three standard deviations of the mean.
- The z-scores for $+1\sigma$ and -1σ are $+1$ and -1, respectively.
- The z-scores for $+2\sigma$ and -2σ are $+2$ and -2, respectively.
- The z-scores for $+3\sigma$ and -3σ are $+3$ and -3 respectively.

The empirical rule is also known as the $\boxed{\text{68-95-99.7 rule.}}$

$P(|X-\mu| \le \sigma) = 0.68$

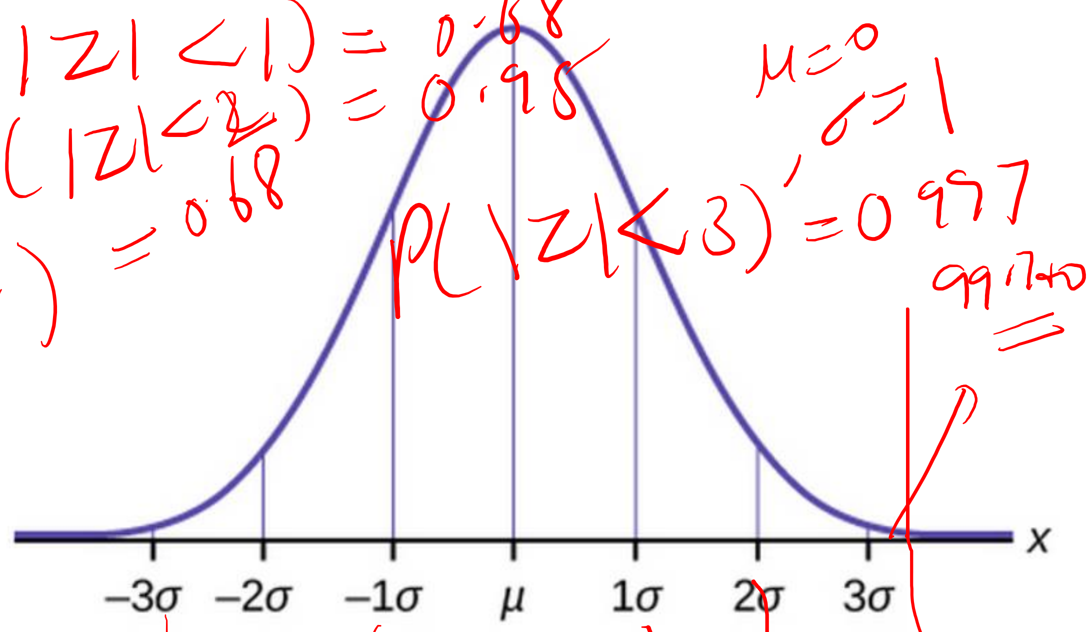
$P(|Z|<1) = 0.68$

$P(|Z|<2) = 0.98$

$\mu = 0, \sigma = 1$

$P(|Z|<3) = 0.997$

99.7%



68%

95%

Note: Almost all the values lie within three standard deviations of the mean.

Key Concept

- The standard normal distribution which has three properties:
  1. It is bell-shaped.
  2. It has a mean equal to 0.
  3. It has a standard deviation equal to 1.
- It is extremely important to develop the skill to find areas (or probabilities or relative frequencies) corresponding to various regions under the graph of the standard normal distribution.
- $P(a < z < b)$ → denotes the probability that the *z* score is between *a* and *b.*
- $P(z > a)$ → denotes the probability that the *z* score is greater than *a.*
- $P(z < a)$ → denotes the probability that the *z* score is less than *a.*

Table A-2

$P(Z \le 3) = \Phi(z)$

TABLE A-2 Standard Normal (z) Distribution: Cumulative Area from the LEFT

3

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.50 and lower	.0001									
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	* .0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	* .0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559

Table A-2

TABLE A-2 (*continued*) Cumulative Area from the LEFT

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	* .9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890

Using Table, A-2

***z* Score :** Distance along horizontal scale of the standard normal distribution; refer to the leftmost column and top row of Table A-2.

Area : Region under the curve; refer to the values in the body of Table A-2.

- Don't confuse *z* scores and areas. *z* scores are distances along the horizontal scale, but areas are regions under the normal curve.
- Table A-2 lists *z* scores in the left column and across the top row, but areas are found in the body of the table.

Finding a *z* Score When Given a Probability Using Table A-2

1. Draw a bell-shaped curve, draw the centerline, and identify the region under the curve that corresponds to the given probability. If that region is not a cumulative region from the left, work instead with a known region that is a cumulative region from the left.

2. Using the cumulative area from the left, locate the closest probability in the body of Table A-2 and identify the corresponding *z* score.

Standard Normal Distribution

$$\Phi(z) = P(Z < z)$$
$$= \int_{-\infty}^{z} f(z)\, dz \, .$$

Here Φ has been used to denote the cumulative probability.

For positive z, the function gives you the probability of being less than z SDs above the mean.

For example, $\Phi(1.0) = 0.84313$, therefore 84.13% of the distribution is less than one SD above the mean.

Example

What is the probability of being less than 1.5 SDs below the mean i.e. $\Phi(-1.5)$?

Example

If $Z \sim N(0,1)$, find

(a) $P(Z > 1.2)$

(b) $P(-2.0 < Z < 2.0)$

(c) $P(-1.2 < Z < 1.0)$

Solution

(a) $P(Z > 1.2) = 1 - \Phi(1.2)$

$= 1 - 0.88493$ (from tables)

$= 0.11507$

(b) $P(-2.0 < Z < 2.0)$

$= P(Z < 2.0) - P(Z < -2.0)$

$= \Phi(2.0) - P(Z > 2.0)$

$= \Phi(2.0) - (1 - P(Z < 2.0))$

$= 2\Phi(2.0) - 1$

$= 2 \times 0.97725 - 1$

$= 0.9545$

(c) $P(-1.2 < Z < 1.0)$

$= P(Z < 1.0) - P(Z < -1.2)$

$= P(Z < 1.0) - P(Z > 1.2)$

$= \Phi(1.0) - (1 - \Phi(1.2))$

(a) $P(Z > 1.2)$

$= 1 - P(Z \le 1.2)$

$= 1 - P(Z \le 1.20)$

$= 1 - 0.88493$

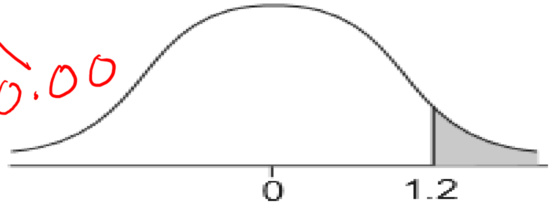
z = integer place . two digits

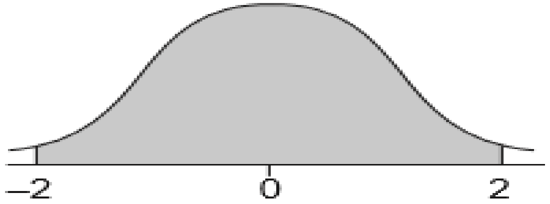
$P(Z \le 3) = \Phi(3)$

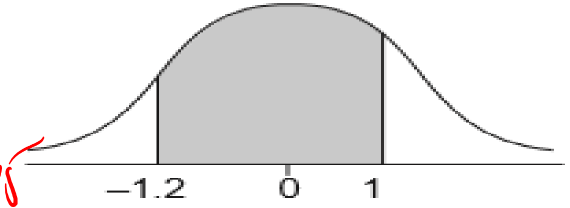
$= P(Z < 2.0) - P(Z < -2.0)$

$0.84134 - 0.11507$

$0.97725 - 0.02275$







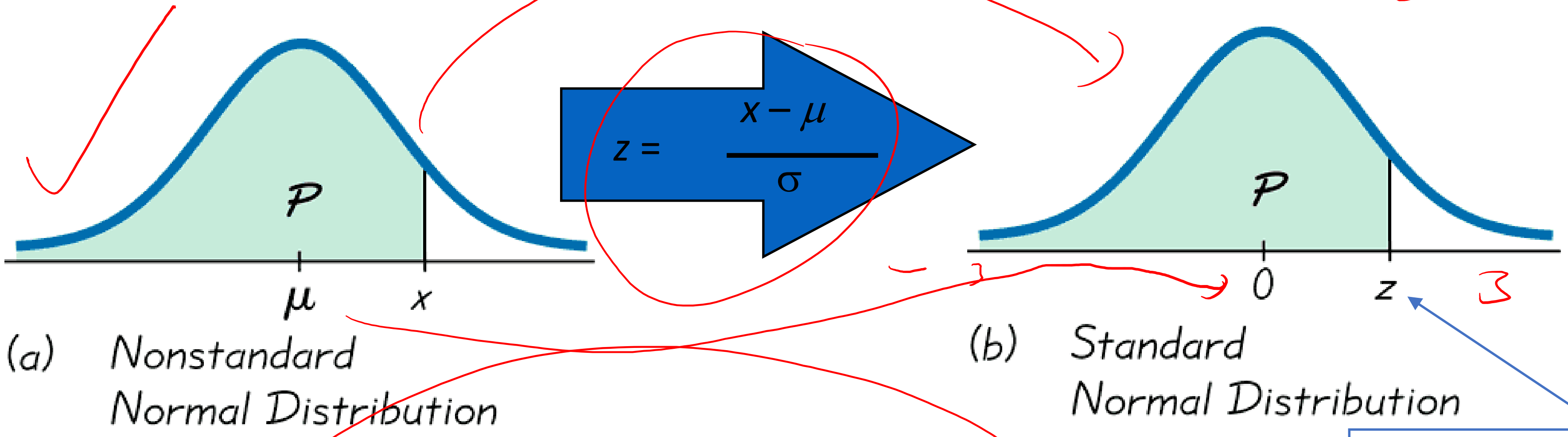
$z = 3.891 \approx 3.89$

$z = 1.758 \approx 1.76$

$z = 50.3 \approx 50.30$

Converting to a Standard Normal Distribution

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



P(X ≤ x) = ?

$P(X \le x) = ?$

P(Z ≤ z) = ?

Round *z* scores to 2 decimal places

$$P(X \le x) = P\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right)$$
$$= P(Z \le 3)$$

Example

A personal computer is used for office work at home, research, communication, personal finances, education, entertainment, social networking, and a ~~myriad of other~~ things. Suppose that the average number of hours a household personal computer is used for entertainment is two hours per day. Assume the times for entertainment are normally distributed and the standard deviation for the times is half an hour.

25 th

1. Find the probability that a household personal computer is used for entertainment between 1.8 and 2.75 hours per day.
2. Find the maximum number of hours per day that the bottom quartile of households uses a personal computer for entertainment.

$$X \sim N\left(2, \left(1/2\right)^2\right)$$

1Q

$$P(1.8 \le X \le 2.75)$$

$$= P\left(\frac{1.8-\mu}{\sigma} < \frac{X-\mu}{\sigma} \le \frac{2.75-\mu}{\sigma}\right)$$

$$= P\left(\frac{1.8-2}{1/2} < Z \le \frac{2.75-2}{0.5}\right)$$

$$= P(-0.4 \le Z \le 1.5)$$

$-0.40,\ 0.00 \qquad 1.50,\ 0.00$

$$= P(Z \le 1.5) - P(Z \le -0.4)$$

$$= 0.93319 - 0.34458$$

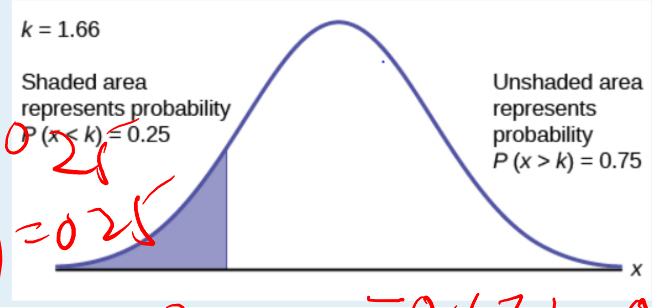
$$= \boxed{}$$

2Q $P(X \le x) = 0.25$

$$\Rightarrow P\left(Z \le \frac{x-\mu}{\sigma}\right) = 0.25$$

$$\Rightarrow P\left(Z \le \frac{x-2}{0.5}\right) = 0.25$$

when $\frac{x-2}{0.5} = \frac{-0.67 + -0.68}{2}$

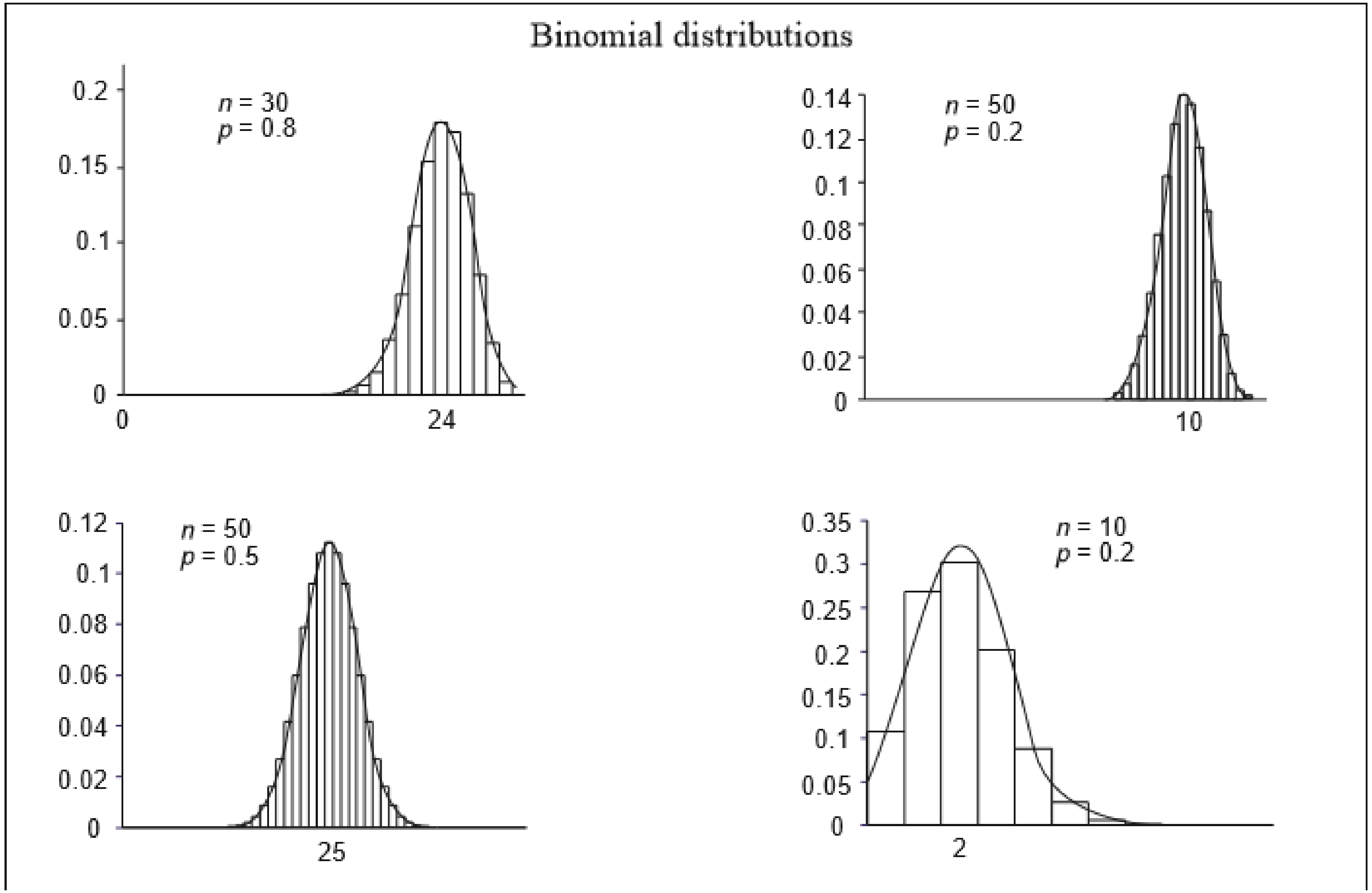


Example

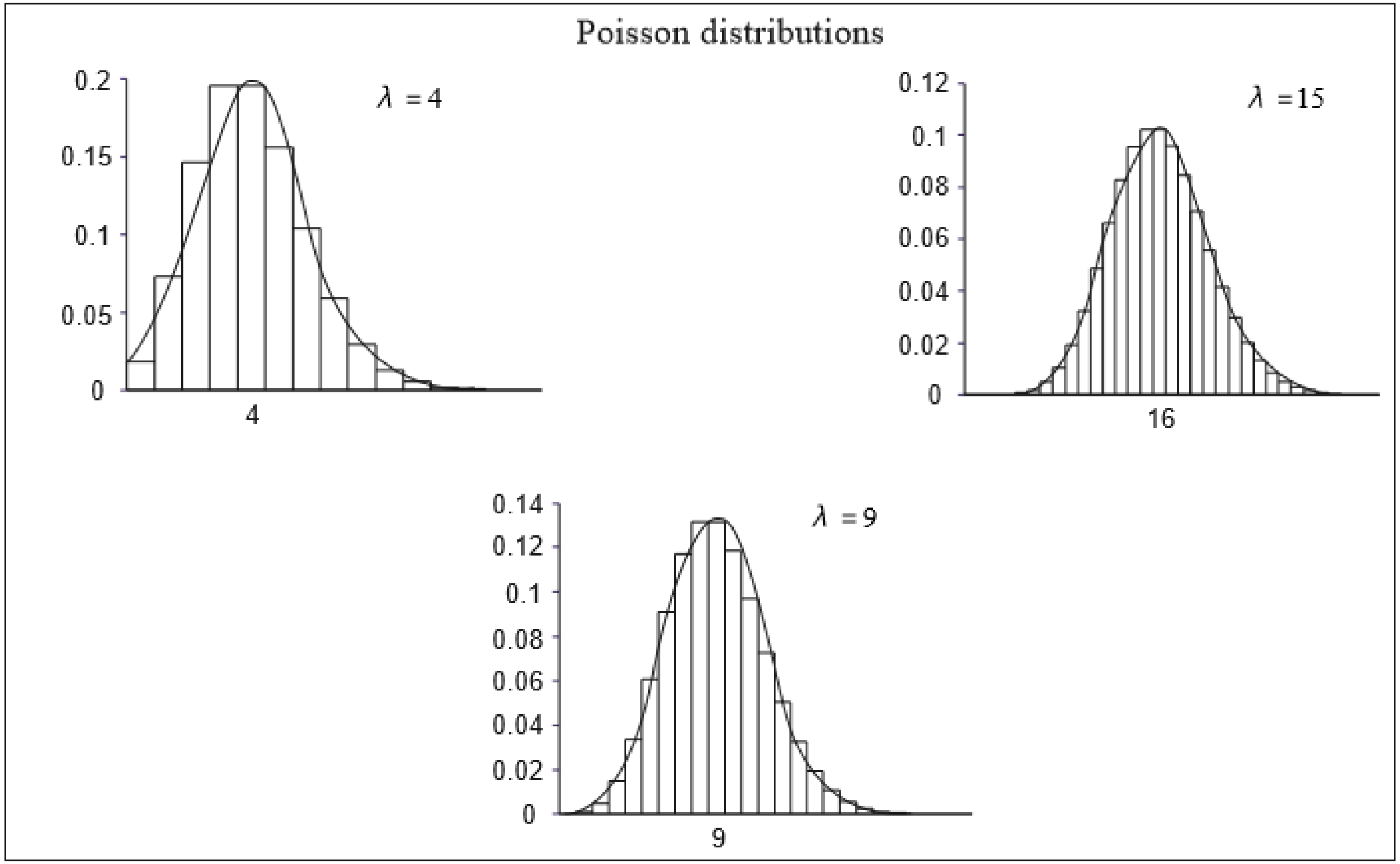
There are approximately one billion smartphone users in the world today. In the United States the ages 13 to 55+ of smartphone users approximately follow a normal distribution with approximate mean and standard deviation of 36.9 years and 13.9 years, respectively.

1. Determine the probability that a random smartphone user in the age range 13 to 55+ is between 23 and 64.7 years old.
2. Determine the probability that a randomly selected smartphone user in the age range 13 to 55+ is at most 50.8 years old.
3. Find the 80th percentile of this distribution, and interpret it in a complete sentence.

Using normal as an approximation to other distribution



Using normal as an approximation to other distribution



To summarise, including the use of the Poisson to approximate to the binomial,

Distribution	Conditions for using	Approximating distribution approximation
$X \sim B(n, p)$	n large (say >50) and p small (say <0.1)	$X \sim Po(np)$
$X \sim B(n, p)$	p close to $\frac{1}{2}$ and $n > 10$ or p moving away from $\frac{1}{2}$ and $n > 30$	$X \sim N(np, npq)$ $(q = 1 - p)$
$X \sim Po(\lambda)$	$\lambda > 20$ (say)	$X \sim N(\lambda, \lambda)$