VI. Probability Distribution

Instructor: Yanlin Qi

Institute of Transportation Studies

Department of Statistics

University of California, Davis

Special Distribution Recap

Special Discrete Distributions

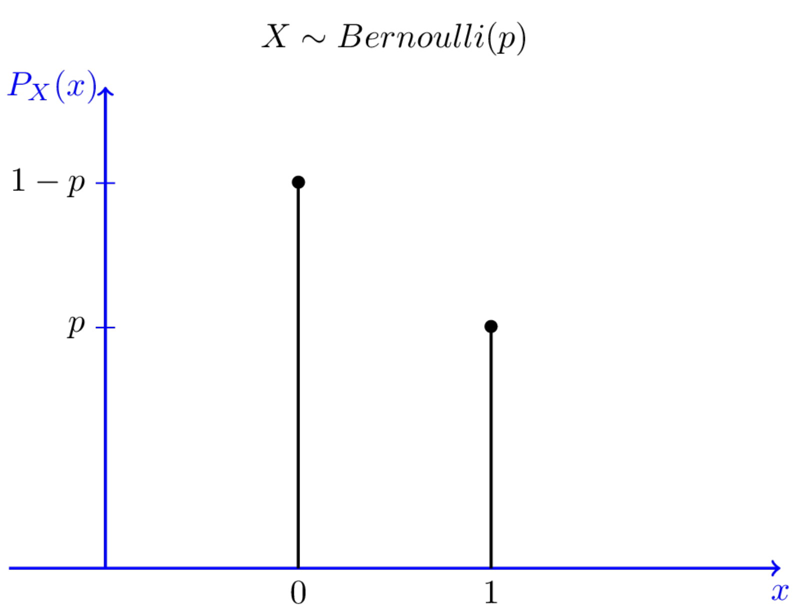
Bernoulli Distribution

Definition 3.4

A random variable X is said to be a *Bernoulli* random variable with *parameter* p, shown as $X \sim Bernoulli(p)$, if its PMF is given by

$$P_X(x) = \begin{cases} p & \text{for } x = 1 \\ 1 - p & \text{for } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

where $0 < p < 1$.



Special Discrete Distributions

PMFs

Discrete Random Variables

Distribution	PMF / Formula	Mean (μ)	Variance (σ^2)
Discrete Uniform	$P(X=x)=\frac{1}{n}$ for $x=1,2,\ldots,n$	$\frac{n+1}{2}$	$\frac{(n^2-1)}{12}$
Binomial	$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$	np	$np(1-p)$
Geometric	$P(X=k)=(1-p)^{k-1}p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Negative Binomial	$P(X=k)=\binom{k-1}{r-1}p^r(1-p)^{k-r}$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$
Hypergeometric	$P(X=k)=\frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$	$\frac{nK}{N}$	$\frac{nK(N-K)(N-n)}{N^2(N-1)}$
Poisson	$P(X=k)=\frac{\lambda^k e^{-\lambda}}{k!}$	λ	λ

Special Continuous Distributions

Special Continuous Distributions

PDFs

Continuous Random Variables

Distribution	PDF / Formula	Mean (μ)	Variance (σ^2)
Continuous Uniform	$f(x) = \frac{1}{b-a}$ for $a \leq x \leq b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential	$f(x) = \lambda e^{-\lambda x}$ for $x \geq 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Normal	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2

CDFs:

$X \sim U(a, b)$

The CDF of X, $F(x)$, is defined as:

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{for } x < a \\ \frac{x-a}{b-a} & \text{for } a \leq x \leq b \\ 1 & \text{for } x > b \end{cases}$$

$X \sim \text{Exponential}(\lambda)$

The CDF of X, $F(x)$, is defined as:

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 - e^{-\lambda x} & \text{for } x \geq 0 \end{cases}$$

Special Distributions – Supplementary Contents

Poisson Distribution

Poisson Distribution Recap

- Standard Poisson Distribution for a fixed interval
- **Formula**:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

- **Explanation**:
- λ is the average number of events per unit time or space.
- k is the number of events.
- This formula gives the probability of exactly k events occurring in a fixed interval when the average rate of events is λ.

Poisson Process

- **A Poisson process** models a sequence of events occurring randomly over time or space.
- Key Properties:
  - **Independent Events**: Number of events in disjoint intervals are independent.
  - **Stationary Increments**: Probability of events depends only on the length of the interval.
  - **No Simultaneous Events**: Probability of multiple events occurring at the same time is zero.
  - **Homogeneity**: In a homogeneous Poisson process, the rate (λ) is constant.

Poisson Process Distribution (1 unit time -> t)

- Poisson distribution within the context of a Poisson process over time t
- **Formula**:

$$P(N(t) = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

- **Explanation**:
  - λ is the average number of events per unit time.
  - t is the length of the time interval.
  - k is the number of events.
  - This formula gives the probability of exactly k events occurring in a time interval of length t when the average rate of events is λ.

Exponential Distribution

- **Purpose:** Time between events in a Poisson process with rate parameter λ.

- PDF: $f(x) = \lambda e^{-\lambda x}, x \geq 0$
- CDF: $F(x) = 1 - e^{-\lambda x}, x \geq 0$

Poisson Dis. and Exponential Dis.

- **Poisson Distribution:** Describes the number of events in a fixed interval of time or space.
- **Exponential Distribution:** Describes the time between successive events in a Poisson process.

Summary:

- **Link:** The parameter λ is common to both distributions, representing the rate of events in a Poisson process.
- **Difference:**
  - In the Poisson distribution, λ describes the average number of events in a fixed interval.
  - In the exponential distribution, λ describes the rate of events per unit time, which is the reciprocal of the average time between events.

Normal Distribution

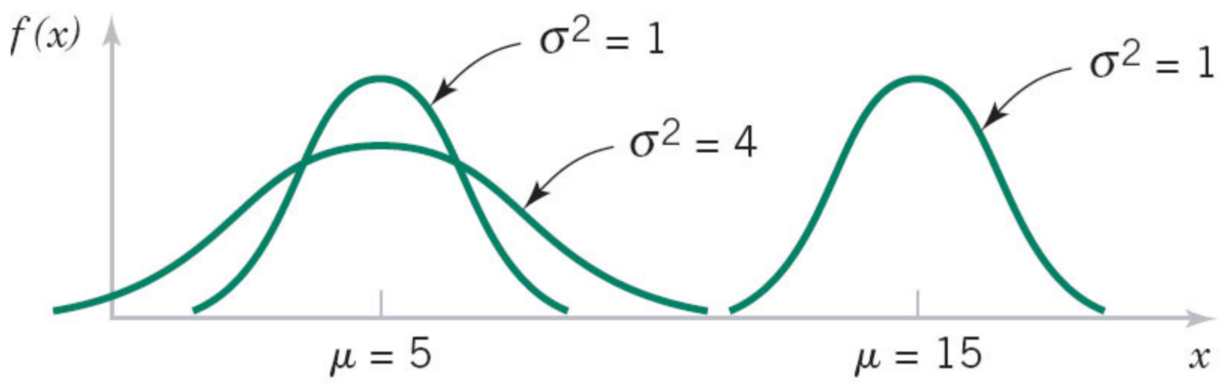
Normal (Gaussian) PDF

- The PDF of a normal RV is

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

- with

$$E(X) = \mu \qquad V(X) = \sigma^2$$



Standard Normal Distribution

- Special case with mean 0 and standard deviation 1
- PDF:

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

- A standard normal RV is often denoted as Z
- The CDF of a normal RV is denoted as

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

- Values for $\Phi(z)$ can be found in many Normal tables

Standard Normal Distribution

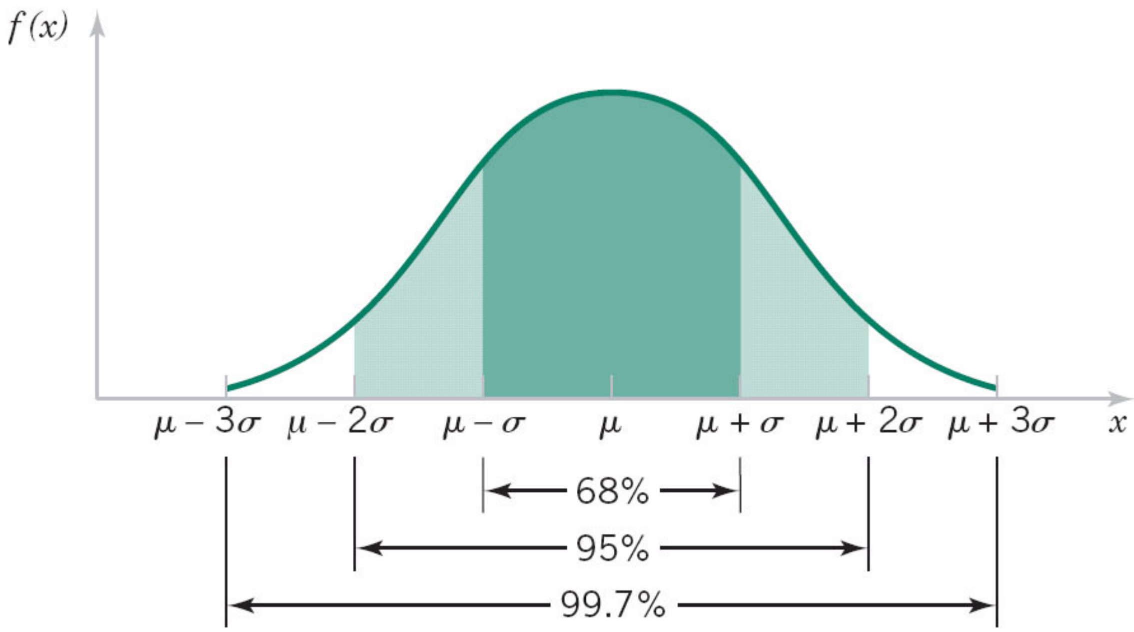
- Special case with mean 0 and standard deviation 1
- PDF:

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

- **Properties:**
  - **68-95-99.7 Rule:**
    - 68% within one standard deviation
    - 95% within two standard deviations
    - 99.7% within three standard deviations

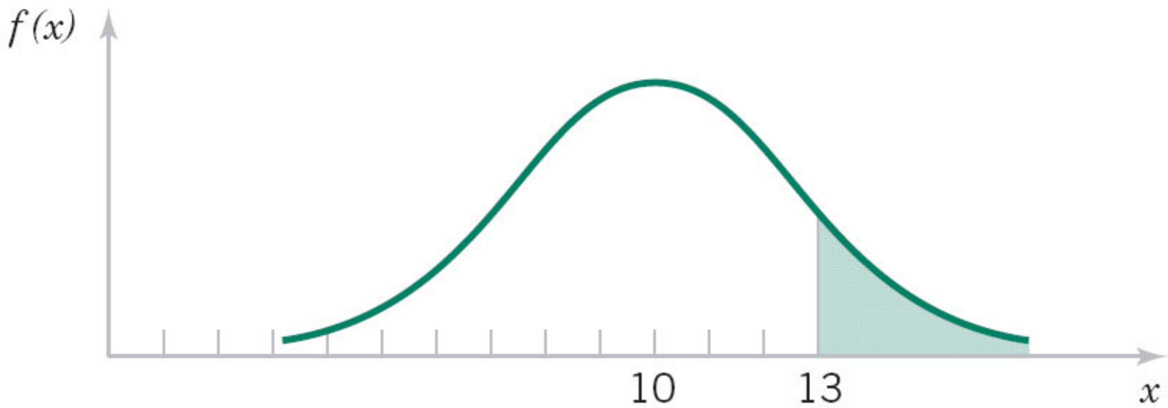
Probabilities Associated to a Normal RV

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



Example [$X = N(\mu,\sigma^2) \rightarrow Z = N(0,1)$]

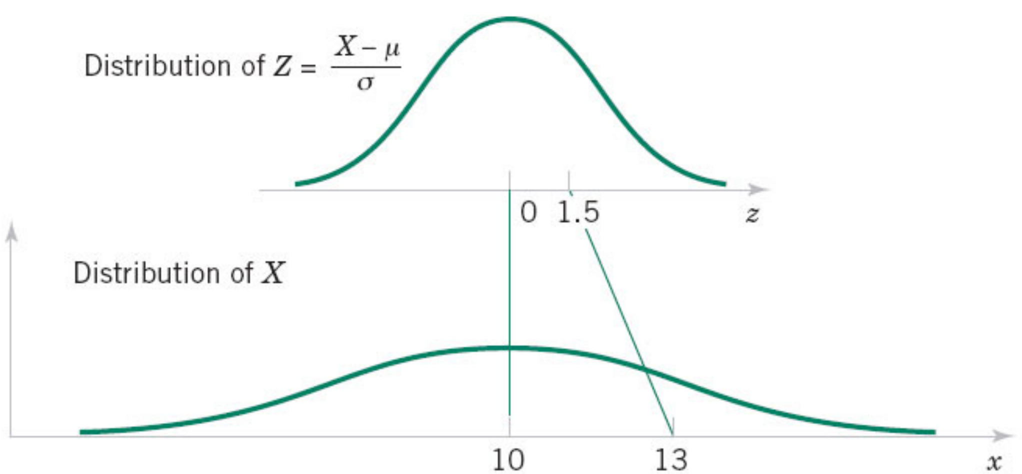
- A concrete strength has mean 10MPa/variance 4MPa^2
- What is the probability that a test result exceeds 13 MPa?

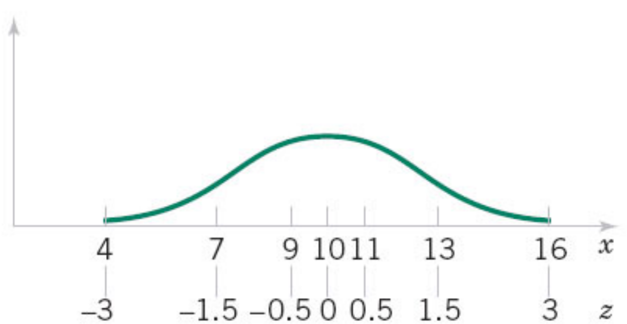


- Unfortunately, there is no closed form for the normal CDF
- To this purpose, tables are used

Example [$X = N(\mu,\sigma^2) \rightarrow Z = N(0,1)$]

- A concrete strength has mean 10MPa/variance 4MPa^2
- What is the probability that a test result exceeds 13 MPa?
  - Standardize: $Z = (X - 10)/2$
  - Then: $X > 13 \rightarrow Z > 1.5$
  - From Table III: $P(Z > 1.5) = 1 - P(Z \leq 1.5) = 1 - 0.93319 = 0.06681$

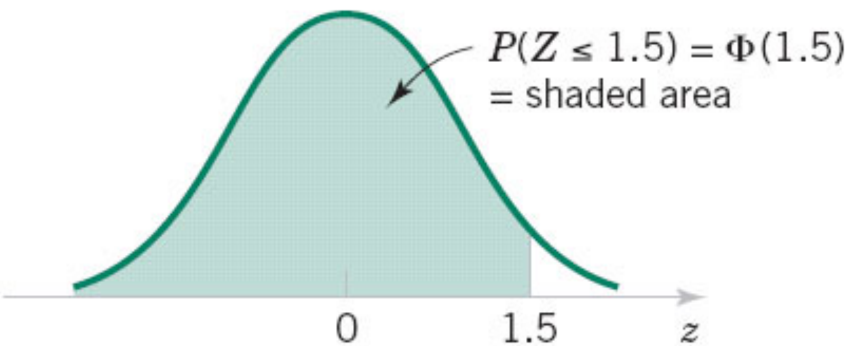




Example [$X = N(\mu,\sigma^2) \rightarrow Z = N(0,1)$]

- Find the probability that a standard NRV is lower than 1.5

$$P(Z \le 1.5) = \Phi(1.5)$$



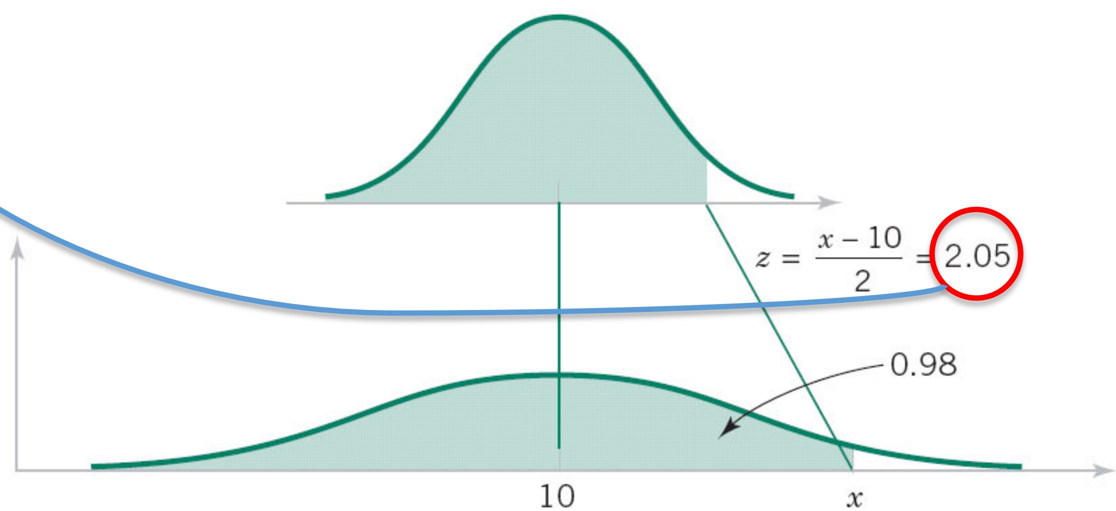
z	0.00	0.01	0.02	0.03
0	0.50000	0.50399	0.50398	0.51197
⋮		⋮		
1.5	0.93319	0.93448	0.93574	0.93699

If you have to find $P(Z \le 1.53)$, go here

Example [$X = N(\mu,\sigma^2) \rightarrow Z = N(0,1)$]

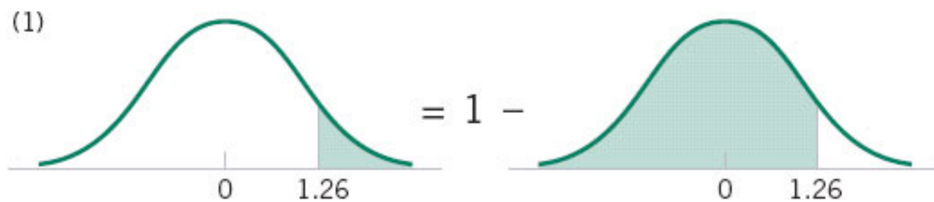
- What is $P(9 < X < 11)$?
  - $P(9 < X < 11) = P(-0.5 < Z < 0.5) = P(Z < 0.5) - P(Z < -0.5)$
    $= 0.69146 - 0.30854 = 0.38292$

- Inverse problem: What is x so that $P(X < x) = 0.98$?
  - $x = 2(2.05) + 10 = 14.1$ MPa

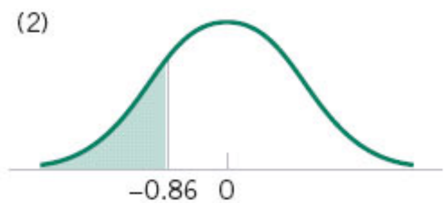


Examples (using basic probability and symmetry)

$P(Z > 1.26)$



$P(Z < -0.86)$

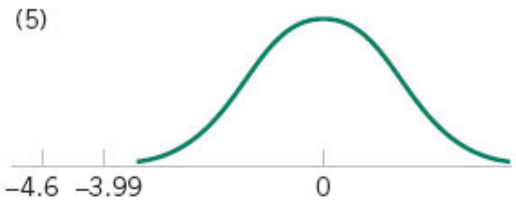


$P(Z > -1.37)$

$P(-1.25 < Z < 0.37)$

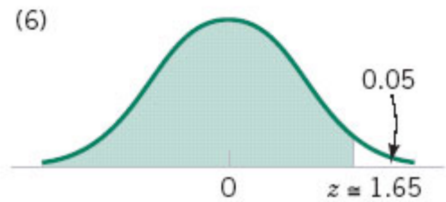
Examples (using basic probability and symmetry)

$P(Z \le -4.6)$

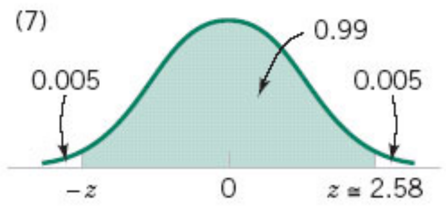


$P(Z > z) = 0.05$

$P(Z \le z) = 0.95$



$P(-z < Z < z) = 0.99$



Example [using basic probability and symmetry]

Assume X is normally distributed with a mean of 10 and a standard deviation of 2. Determine:

a) P(2<X<4)

$$P\left(\frac{2-10}{2} < Z < \frac{4-10}{2}\right)$$
$$= P(-4 < Z < -3)$$
$$= P(Z < -3) - P(Z < -4)$$
$$= 0.00132$$

b) P(X>9)

$$= 1 - P(X < 9)$$
$$= 1 - P(Z < (9-10)/2)$$
$$= 1 - P(Z < -0.5)$$
$$= 0.69146$$

c) P(-2<X<8)

$$= P(X < 8) - P(X < -2)$$
$$= P\left(Z < \frac{8-10}{2}\right) - P\left(Z < \frac{-2-10}{2}\right)$$
$$= P(Z < -1) - P(Z < -6)$$
$$= 0.15866$$

d) P(6<X<14)

$$= P\left(\frac{6-10}{2} < Z < \frac{14-10}{2}\right)$$
$$= P(-2 < Z < 2)$$
$$= P(Z < 2) - P(Z < -2)]$$
$$= 0.9545$$

Review Examples – Special Discrete Distributions

PMF – General Exampe

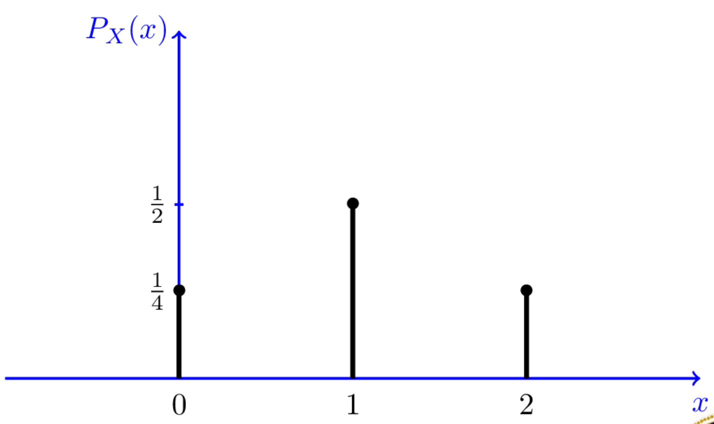
- I toss a fair coin twice, and let X be defined as the number of heads I observe. Find the range of X, R_X, as well as its probability mass function P_X.
- Sol:
  - S={HH,HT,TH,TT}
  - $R_X = \{0,1,2\}, P_X(k) = P(X = k)$ for $\mathrm{k} = \{0,1,2\}$

$$P_X(0) = P(X = 0) = P(TT) = \frac{1}{4},$$

$$P_X(1) = P(X = 1) = P(\{HT, TH\}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2},$$

$$P_X(2) = P(X = 2) = P(HH) = \frac{1}{4}.$$

$$P_X(x) = \begin{cases} P(X = x) & \text{if } x \text{ is in } R_X \\ 0 & \text{otherwise} \end{cases}$$



Binomial vs Hypergeometric Dist.

Binomial vs Hypergeometric Dist. – Example 1

- We draw 10 balls at random from an urn containing 30 red and 20 blue balls, and let X=#red balls obtained.
- Derive the PMFs for the following two scenarios:
  - Drawing balls with replacement
  - Drawing balls without replacement

Binomial vs Hypergeometric Dist. – Example 1

- **Scenario 1: Drawing with Replacement**
- When drawing with replacement, each draw is independent of the previous ones. Therefore, the number of red balls obtained follows a **Binomial distribution**.
- Let's define:
  - Binomial RV: X = #red balls obtained.
  - n=10 as the number of draws.
  - $p = 30/50 = 0.6$ as the probability of drawing a red ball on each draw.
  - PMF: $X \sim B(10{,}0.6)$ $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$
  - Substituting the values:

$$P(X=k) = \binom{10}{k} (0.6)^k (0.4)^{10-k}$$

Binomial vs Hypergeometric Dist. – Example 1

- **Scenario 2: Drawing without Replacement**
- When drawing without replacement, the draws are dependent on each other. Therefore, the number of red balls obtained follows a **Hypergeometric distribution**.
- Let's define:
  - Hypergeometric RV: X = #red balls obtained.
  - N=50 as the total number of balls.
  - K=30 as the number of red balls.
  - n=10 as the number of draws.
- PMF: $X \sim HyperGem(N, k, n) = \; P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$
- Substituting the values:

$$P(X = k) = \frac{\binom{30}{k}\binom{50-30}{10-k}}{\binom{50}{10}}$$

Geometric Distribution

Geometric Distribution – Example 1

- Suppose Max owns a lightbulb manufacturing company and determines that 3 out of every 75 bulbs are defective.
- Q:
  - What is the probability that Max will find the first faulty lightbulb on the 6th one that he tested?
  - What is the number lightbulbs we would expect Max to inspect until he finds his first defective, as well as the standard deviation?

Geometric Distribution – Example 1

- Q1 Sol. P({the first faulty lightbulb on the 6th one})
- Define Geometric RV: X = #tests to find the first faulty lightbulb
- $X \sim Geom(p)$

$$p = \frac{3}{75} = 0.04$$

$$k = 6$$

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

$$P(X = 6) = 0.04(1 - 0.04)^{6-1}$$

$$P(X = 6) = 0.04(0.96)^{5} = 0.0326$$

Geometric Distribution – Example 1

- Q2 Sol. P({Expected #tests until the first defective})
- Define Geometric RV: X = #tests to find the first faulty lightbulb
- $X \sim Geom(p)$

Mean:	$\mu = E(X) = \frac{1}{p}$ $E(X) = \frac{1}{0.04} = 25$	
Variance:	$\sigma^2 = V(X) = \frac{(1-p)}{p^2}$ $V(X) = \frac{(1-0.04)}{(0.04)^2} = 600$	$p = probability\ of\ success$ $p = 0.04$
Standard Deviation	$\sigma = \sqrt{\sigma^2} = \sqrt{600} = 24.49$	

Poisson Distribution

Poisson Distribution – Example 1

- The number of emails that one gets in a weekday can be modeled by a Poisson distribution with an average of 0.2 emails per minute.
  - What is the probability that one gets no emails in an interval of length 5 minutes?
  - What is the probability that one gets more than 3 emails in an interval of length 10 minutes?

Poisson Distribution – Example 1

- Q1
  - Define Poisson RV: X = #emails obtained in the 5-minute interval
  - Then, by the assumption X is a Poisson RV with parameter $\lambda=5\cdot 0.2=1$
  - $X \sim Pois(\lambda = 1)$

$$P(X=0) = P_X(0) = e^{-\lambda}\frac{\lambda^0}{0!} = e^{-1}\cdot\frac{1^0}{1} = e^{-1} \approx 0.3679.$$

- Q2
  - Define Poisson RV: Y = #emails obtained in the 10-minute interval.
    - Then, by the assumption Y is a Poisson RV with $\lambda=10\cdot 0.2=2$
    - $Y \sim Pois(\lambda = 2)$

$$\begin{aligned} P(Y>3) &= 1 - P(Y\le 3) = 1-(P_Y(0)+P_Y(1)+P_Y(2)+P_Y(3)) \\ &= 1-(e^{-2}+2e^{-2}+2e^{-2}+\frac{8}{6}e^{-2}) \\ &= 1-e^{-2}(1+2+2+\frac{8}{6}) \\ &= 1-\frac{19}{3}e^{-2} \approx 0.1429. \end{aligned}$$

Review Examples – Special Continuous Distributions

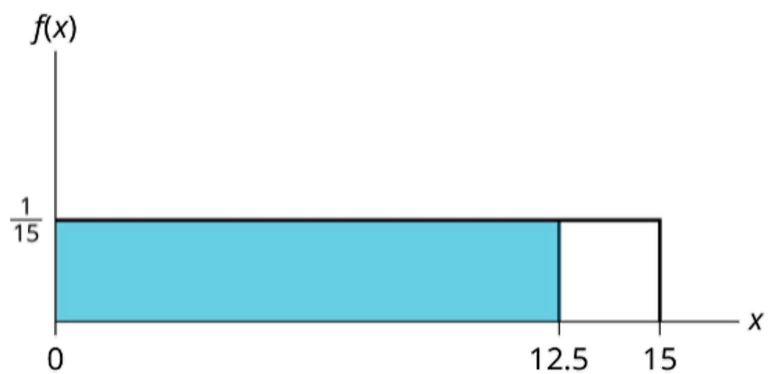
Continuous Uniform Distribution

Continuous Uniform Distribution – Example 1

- The amount of time, in minutes, that a person must wait for a bus is **uniformly distributed** between zero and 15 minutes, inclusive.
- Problem
  - a. What is the probability that a person waits **fewer than 12.5 minutes**?
  - b. On the average, how long must a person wait? Find the mean, μ, and the standard deviation, σ.
  - c. To what value does the wait time fall below 90% of the time?

Continuous Uniform Distribution – Example 1

- Sol. P1a. What is the probability that a person waits **fewer than 12.5 minutes**?
  - Let X = #minutes a person must wait for a bus.
  - a = 0 and b = 15. X ~ U(0, 15).
  - PDF: $f\ (x) = \frac{1}{15-0} = \frac{1}{15}$, for 0 ≤ x ≤ 15.



$$P(x < k) = (\text{base})(\text{height}) = (12.5 - 0)\left(\frac{1}{15}\right) = 0.8333$$

Continuous Uniform Distribution – Example 1

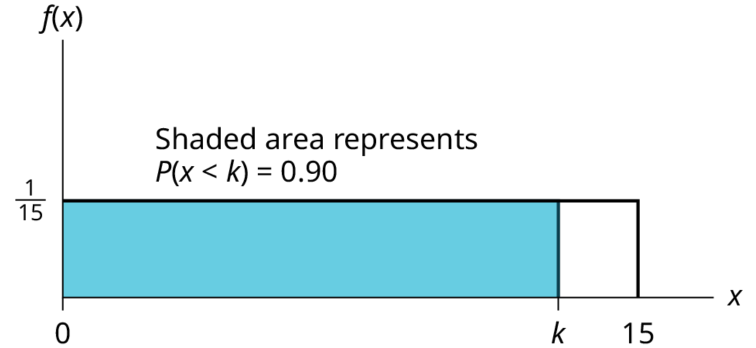
- Sol. P2
  - PDF: $f\ (x) = \frac{1}{15}$, for $0 \le x \le 15$.

b. $\mu = \frac{a+b}{2} = \frac{15+0}{2} = 7.5$. On the average, a person must wait 7.5 minutes.

$\sigma = \sqrt{\frac{(b-a)^2}{12}} = \sqrt{\frac{(15-0)^2}{12}} = 4.3$. The Standard deviation is 4.3 minutes.

- Sol. P3 To what value does the wait time fall below 90% of the time?

- $P(X < k) = F(x = k) = \frac{k-a}{b-a} = (k-0)(\frac{1}{15})$
- $0.90 = (k)\left(\frac{1}{15}\right) 0.90 = (k)\left(\frac{1}{15}\right)$
- $k = (0.90)(15) = 13.5$



- The 90th percentile is 13.5 minutes. Ninety percent of the time, a person must wait at most 13.5 minutes.

Continuous Uniform Distribution – Exercise

- The thickness x of a protective coating applied to a conductor designed to work in corrosive conditions follows a uniform distribution over the interval $[20, 40]$ microns.
- Problems
  - Find the mean, standard deviation and cumulative distribution function of the thickness of the protective coating.
  - Find also the probability that the coating is less than 35 microns thick.

Exponential Distribution

Exponential Dist. vs Poisson Dist. – Example 1

- Suppose, in a call center, calls arrive at an average rate of 3 calls per hour.
- Problems:
  - A. What is the probability that the next call will arrive **within** the next 10 minutes?
  - B. What is the probability that **exactly** 5 calls will arrive in the next hour?

Poisson Dist. vs Exponential Dist. – Example 1

- Sol. Q(A) P({the next call will arrive **within** the next 10 minutes})
- Define:
  - **Exponential RV**: X = The time until the next call arrives at the call center
  - **Parameter:** Average rate λ=3 calls per hour.
  - Time $x = \frac{10}{60} = \frac{1}{6}$ hours (since 10 minutes is $\frac{1}{6}$ of an hour).
- PDF: $X \sim Exp(\lambda = 3), f(x) = \lambda e^{-\lambda x}, x \geq 0$
- CDF: $\mathrm{F}(x) = 1 - e^{-\lambda x}$
- Substitute the values:

$$P(X \leq \frac{1}{6}; 3) = 1 - e^{-3 \times \frac{1}{6}} = 1 - e^{-0.5} \approx 0.3935$$

Poisson Dist. vs Exponential Dist. – Example 1

- Sol. Q(B) P({**exactly** 5 calls will arrive in the next hour})
- Given:
  - **Poisson RV**: X = #calls arriving at the call center in an hour
  - **Parameter:** Average rate λ=3 calls per hour.
  - Number of calls k=5.
- $X \sim Pois(\lambda = 3)$
- Using the Poisson distribution formula:

$$P(X = k; \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$$

- Substitute the values:

$$P(X = 5; 3) = \frac{3^5 e^{-3}}{5!} = \frac{243e^{-3}}{120} \approx 0.1008$$

Poisson Dist. vs Exponential Dist. - Exercise

- Example Scenario: Hospital ER
- In a hospital ER, patients arrive at an average rate of 6 patients per hour.
- Questions:
  - What is the probability that exactly 8 patients will arrive at the ER in the next hour?
  - What is the probability that the next patient will arrive within the next 5 minutes?

(You may leave the answer expressed as a fraction involving e and factorials.)

Normal Distribution

Normal Distribution – Example 1

- **Example Scenario: Heights of Adults**
  - Suppose the heights of adult men in a certain country are normally distributed with a mean $\mu=175$ cm and a standard deviation $\sigma=10$ cm.
- **Question:**
  - What is the probability that a randomly selected adult man is taller than 185 cm?

Normal Distribution – Example 1

- **Sol:**
  - 1.Define the Random Variable:
    - NRV: X be the height of an adult man
  - 2. Given the parameters
    - X~N(μ,σ2) $f(x) = \frac{1}{10\sqrt{2\pi}} e^{-\frac{(x-175)^2}{2\cdot 100}}$
  - 3. Calculate the Z-Score: $Z = \frac{X-\mu}{\sigma}$
    - For x=185 cm $Z = \frac{185-175}{10} = \frac{10}{10} = 1$
    - P(X>185) → P(Z>1) in the standard N(0, 1)

$P(Z > 1) = 1 - P(Z \leq 1)$ $P(Z \leq 1) \approx 0.8413.$

$P(Z > 1) = 1 - 0.8413 = 0.1587$

Normal Distribution – Exercise

- Question:
  - Find the probability that a randomly selected adult man is between 165 cm and 185 cm
  - (Hint: Draw the PDF and use the symmetry of Normal dist.)

$$P(Z \leq 1) \approx 0.8413.$$

$$P(Z > 1) = 1 - 0.8413 = 0.1587$$