HCMC University of Technology

Dung Nguyen

May 16, 2020

Probability and Statistics

Some Special Distributions

Outline I

1. Discrete Distributions

2. Continuous Distributions

1. Discrete Distributions
   - The Binomial Distributions $\mathsf{B}(n, p)$
   - The Hypergeometric Distributions $\mathsf{H}(N, m, n)$
   - The Poisson Distributions $\mathsf{Poisson}(\lambda)$

The Binomial Distributions $\mathrm{B}(n,p)$

Definition

- *Bernoulli trial* $\mathrm{B}(p)$:

u	1	0
$P(Z=u)$	p	$q=1-p$

$\mathrm{E}(Z)=p$ *and* $\mathrm{V}(Z)=pq$.

- *Binomial random variable = the # of successes in n Bernoulli trials (the probability of success in each trial is* $0\le p\le 1$*).*

Examples:

- The # of defective items among 20 independent items with the defective rate 5%.
- The # of winning tickets among 11 independent lottery tickets with the winning rate 1%.
- The # of patients reporting symptomatic relief with a specific medication with the effective rate 80%.

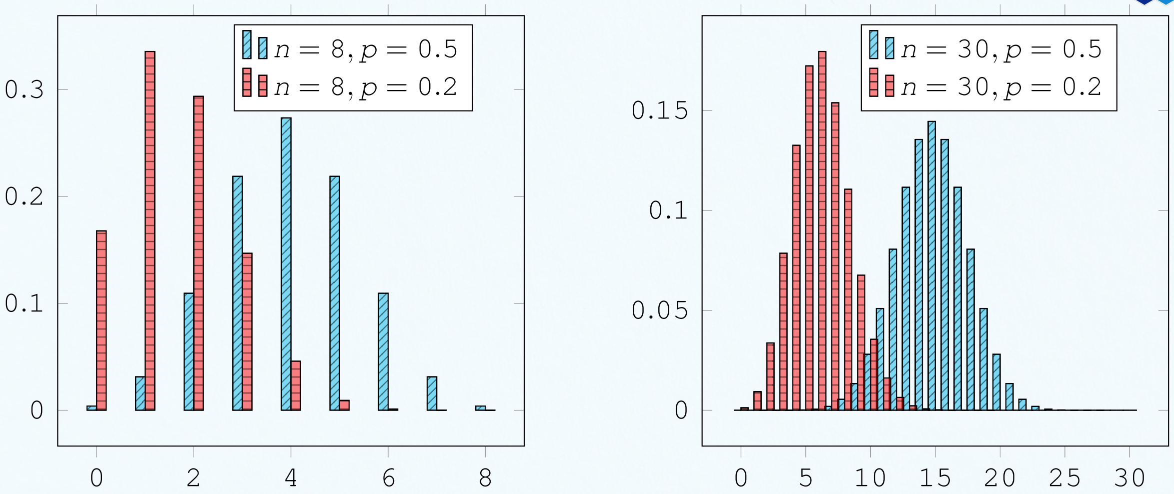


Figure: Pmf of $B(n, p)$

Proposition

Let $X \sim \mathrm{B}(n, p)$. Then

1. X takes values in $\Omega = \{0, 1, \ldots, n\}$ such that

$$f(k) = P(X = k) = C_n^k p^k q^{n-k}.$$

- $k = 0$: $\boxed{q}\ \boxed{q}\ \boxed{q}\ \boxed{\cdots}\ \boxed{q} \implies P(X = 0) = q^n$.
- $k = n$: $\boxed{p}\ \boxed{p}\ \boxed{p}\ \boxed{\cdots}\ \boxed{p} \implies P(X = n) = p^n$.
- $k = 1$: $\boxed{p}\ \boxed{q}\ \boxed{q}\ \boxed{\cdots}\ \boxed{q} \implies P(X = 1) = C_n^1 p q^{n-1}$.
- $k = 2$: $\boxed{p}\ \boxed{p}\ \boxed{q}\ \boxed{\cdots}\ \boxed{q} \implies P(X = 2) = C_n^2 p^2 q^{n-2}$.

2. X is a sum of n independent Bernoulli random variables.

$$X = Z_1 + Z_2 + \cdots + Z_n,$$

$V(x) = E(x^2) - (EX)^2$

where $Z_i = \begin{cases} 1, & \text{if the i-th trial is successful} \\ 0, & \text{otherwise} \end{cases}$.

$E(x) = \sum_{k=0}^{n} k C_n^k p^k \cdot (1-p)^{n-k}$

3. $\mathrm{E}(X) = np$ and $\mathrm{V}(X) = npq$.

$E(x^2) = \dots k^2 \dots$

Example

1. Each sample of water has a 10% chance of containing a particular organic pollutant. Assume that the samples are independent with regard to the presence of the pollutant.

Example

1. Each sample of water has a 10% chance of containing a particular organic pollutant. Assume that the samples are independent with regard to the presence of the pollutant.
   (a) Find the probability that, in the next 18 samples, exactly 2 contain the pollutant.

$$X \sim B(18, 0.1) \Rightarrow P(3 \leq X < 7) = (P=3) + (P=4) + (P=4) + (P=5)$$

$$=$$

Example

1. Each sample of water has a 10% chance of containing a particular organic pollutant. Assume that the samples are independent with regard to the presence of the pollutant.
   a. Find the probability that, in the next 18 samples, exactly 2 contain the pollutant.
   b. Determine the probability that at least 4 samples contain the pollutant.

Example

1. Each sample of water has a 10% chance of containing a particular organic pollutant. Assume that the samples are independent with regard to the presence of the pollutant.
   (a) Find the probability that, in the next 18 samples, exactly 2 contain the pollutant.
   (b) Determine the probability that at least 4 samples contain the pollutant.
   (c) Now determine the probability that $3 \leq X < 7$.

$X \sim B(18, 0{,}1) \Rightarrow P(3 \leq X < 7) =$ 3,4,5,6

$$= \sum_{3}^{6} C_k^{18} \cdot 0{,}1^k \cdot 0{,}9^{(18-k)} = 0{,}265$$

Example

1. Each sample of water has a 10% chance of containing a particular organic pollutant. Assume that the samples are independent with regard to the presence of the pollutant.
   - (a) Find the probability that, in the next 18 samples, exactly 2 contain the pollutant.
   - (b) Determine the probability that at least 4 samples contain the pollutant.
   - (c) Now determine the probability that $3 \leq X < 7$.

$X \sim B(10, 0{,}2) \Rightarrow P(0) = C_{10}^{0}\, 0{,}2^{0} \cdot 0{,}8^{10} \mid P(1) = C_{10}^{1}\, 0{,}2^{1} \cdot 0{,}8^{9}$

2. A certain electronic system contains 10 components. Suppose that the probability that each individual component will fail is 0.2 and that the components fail independently of each other. Given that at least one of the components has failed, what is the probability that at least two of the components have failed?

$$P(X \geq 2 \mid X \geq 1) = \frac{P(X \geq 2, P(X \geq 1))}{P(X \geq 1)} = \frac{1 - P(0) - P(1)}{1 - P(0)}$$

Example

3. A certain binary communication system has a bit-error rate of 0.1; i.e., in transmitting a single bit, the probability of receiving the bit in error is 0.1. If 6 bits are transmitted, then how many bits, on average, will be received in error? Determine the corresponding variance.

Example

- Three men A, B, and C shoot at a target. Suppose that A shoots three times and the probability that he will hit the target on any given shot is 1/8, B shoots five times and the probability that he will hit the target on any given shot is 1/4, and C shoots twice and the probability that he will hit the target on any given shot is 1/3. What is the expected number of times that the target will be hit?

$E(X) = np \; ; \; V(x) = npq$

- $E(X) = 3 \cdot \frac{1}{8} + 5 \cdot \frac{1}{4} + 2 \cdot \frac{1}{3}$

Example 5 –

30 balls	
20 red balls	10 blue balls

X = the # of red balls in the SMP

withdraw 8 balls

8 balls	
5 red balls	3 blue balls

with replacement

$X \sim \mathbf{B}(8, 20/30)$
$P(R_i) = 20/30$
$P(R_2|R_1) = 20/30$
Independent

$$P(X=5) = C_8^5\,(2/3)^5(1/3)^3$$

without replacement

$X \sim \mathbf{H}(30, 20, 8)$
$P(R_i) = 20/30$
$P(R_2|R_1) = 19/29$
Dependent

$$P(X=5) = \frac{C_{20}^5\, C_{10}^3}{C_{30}^8}$$

Generalization

N balls	
m red balls	(N-m) blue balls

X = the # of red balls in the SMP

withdraw n balls

n balls	
k red balls	(n-k) blue balls

with replacement

$X \sim \mathbf{B}(n, p = m/\mathrm{N})$
$P(R_i) = \mathrm{p}$
$P(R_2|R_1) = \mathrm{p}$
$\Rightarrow$ Independent

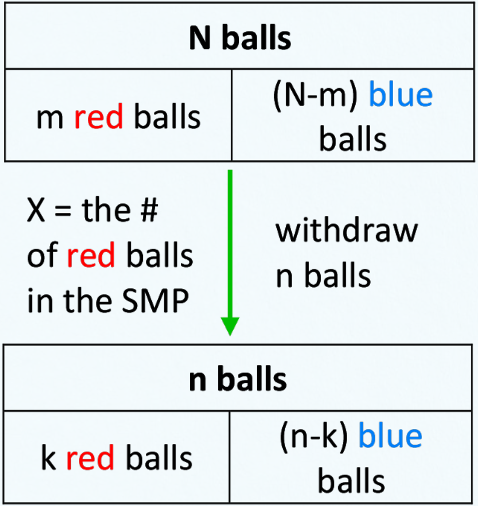
$P(X = \mathrm{k}) = C_n^{\mathrm{k}}\, p^{\mathrm{k}} q^{n\text{-}\mathrm{k}}$

without replacement

$X \sim \mathbf{H}(N, m, n)$
$P(R_i) = p = m/N$
$P(R_2|R_1) = \dfrac{m\text{-}1}{N\text{-}1}$
$\Rightarrow$ Dependent

$$P(X = \mathrm{k}) = \frac{C_m^{\mathrm{k}}\, C_{N\text{-}\mathrm{m}}^{n\text{-}\mathrm{k}}}{C_N^{n}}$$

$P(R_i) = p$



- With replacement

$$P(R_1) = \frac{m}{N}$$

$$P(R_2|R_1) = P(R_2|B_1) = \frac{m}{N}$$

$$P(R_2) = \frac{m}{N}.$$

- Without replacement

$$P(R_1) = \frac{m}{N}$$

$$P(R_2|R_1) = \frac{m-1}{N-1}$$

$$P(R_2|B_1) = \frac{m}{N-1}$$

$$P(R_2) = P(R_2|R_1)P(R_1) + P(R_2|B_1)P(B_1)$$

$$= \frac{m-1}{N-1}\frac{m}{N} + \frac{m}{N-1}\frac{N-m}{N} = \frac{m}{N}.$$

The Hypergeometric Distributions $\mathrm{H}(N, m, n)$

Definition

Suppose that there are n draws from a finite population of size N containing m successes without replacement. Let X be the number of successes. Then X is called a hypergeometric random variable or X has a hypergeometric distribution.

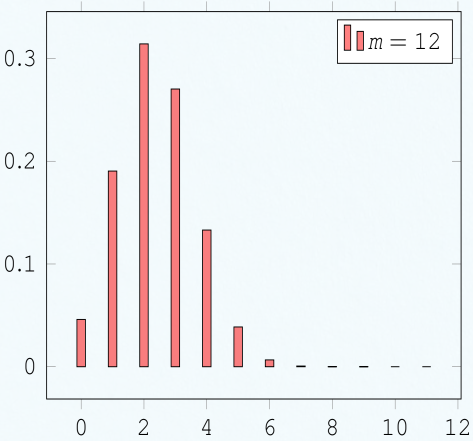
Proposition

Let $X \sim \mathrm{B}(n, p)$. Then

1. *X has the pmf:* $f(k) = \dfrac{\mathrm{C}_m^k \mathrm{C}_{N-m}^{n-k}}{\mathrm{C}_N^n}$.
2. *X is a sum of n dependent Bernoulli random variables.*

$$X = Z_1 + Z_2 + \cdots + Z_n, \quad Z_i = \begin{cases} 1, & \text{if the } i\text{-th trial is successful} \\ 0, & \text{otherwise} \end{cases}.$$

3. $\mathrm{E}(X) = np$ *and* $\mathrm{V}(X) = npq \cdot \dfrac{N-n}{N-1}$ *with* $p = m/N$.



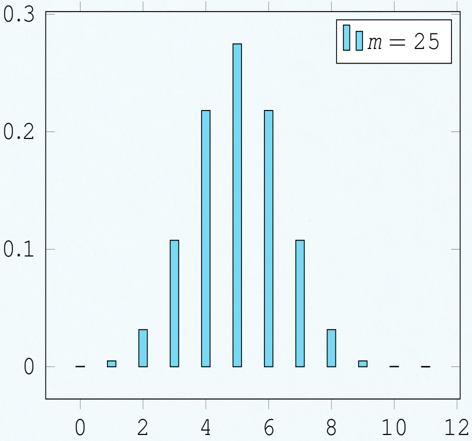


Figure: Pmf of $\mathrm{H}(40, m, 10)$

Example

6. A batch of parts contains 100 from a local supplier of tubing and 200 from a supplier of tubing in the next state.

Example

6 A batch of parts contains 100 from a local supplier of tubing and 200 from a supplier of tubing in the next state.

(a) If four parts are selected randomly and without replacement, what is the probability they are all from the local supplier?

Example

6 A batch of parts contains 100 from a local supplier of tubing and 200 from a supplier of tubing in the next state.

(a) If four parts are selected randomly and without replacement, what is the probability they are all from the local supplier?

(b) What is the probability that two or more parts are from the local supplier?

Example

6 A batch of parts contains 100 from a local supplier of tubing and 200 from a supplier of tubing in the next state.

a. If four parts are selected randomly and without replacement, what is the probability they are all from the local supplier?
b. What is the probability that two or more parts are from the local supplier?
c. What is the probability that at least one part in the sample is from the local supplier?

Example

6. A batch of parts contains 100 from a local supplier of tubing and 200 from a supplier of tubing in the next state.
   a. If four parts are selected randomly and without replacement, what is the probability they are all from the local supplier?
   b. What is the probability that two or more parts are from the local supplier?
   c. What is the probability that at least one part in the sample is from the local supplier?

7. Suppose that seven balls are selected at random without replacement from a box containing five red balls and ten blue balls. If X denotes the proportion of red balls in the sample, what are the mean and the variance of X?

Approximation property

If $m, N-m, n$ are large enough then $\mathsf{H}(N,m,n) \approx \mathsf{B}(n,p)$.

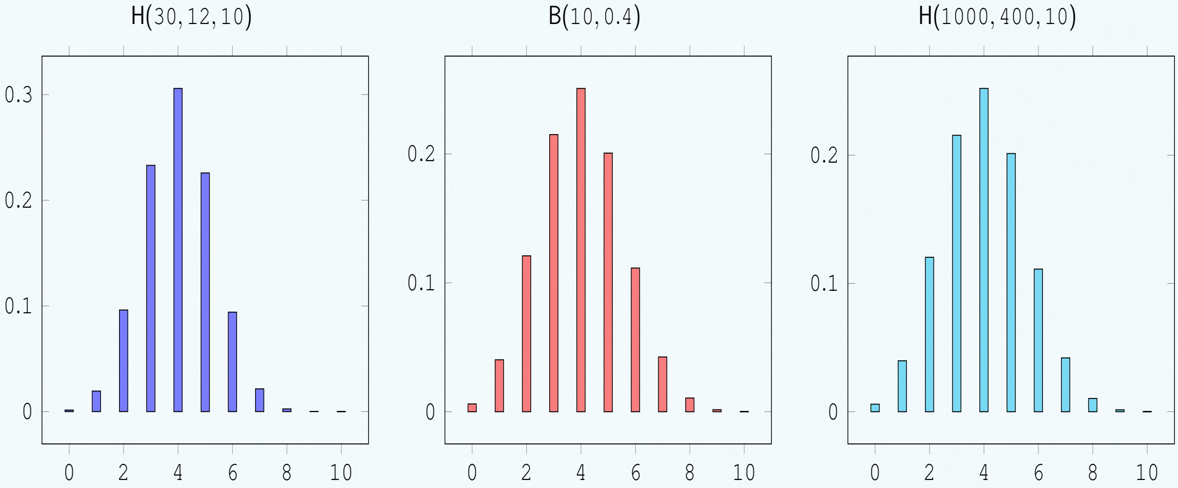
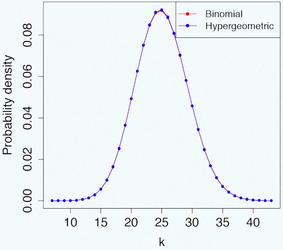


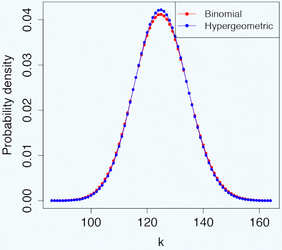
Figure: $\mathsf{B}(n,p)$ vs. $\mathsf{H}(N,m,n)$ with $p = m/N$

Hypergeometric distribution vs. Binomial distribution ($p = 0.25$, Population size $N = 10000$)

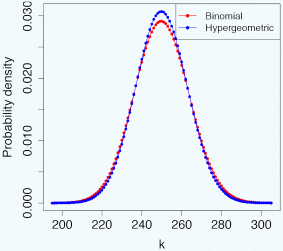
trials $n = 100$, $n/N = 1\%$



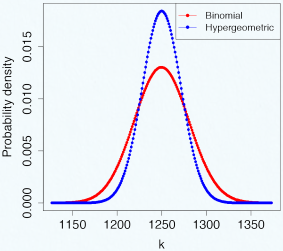
trials $n = 500$, $n/N = 5\%$



trials $n = 1000$, $n/N = 10\%$



trials $n = 5000$, $n/N = 50\%$



Example

8 A list of customer accounts at a large company contains 1,000 customers. Of these, 700 have purchased at least one of the company's products in the last 3 months. To evaluate a new product, 50 customers are sampled at random from the list. What is the probability that more than 45 of the sampled customers have purchased from the company in the last 3 months?

Applications of Poisson Distributions

- Electrical system example: the number of telephone calls arriving in a system, the number of wrong connections to your phone number.
- Astronomy example: the number of photons arriving at a telescope.
- Biology example: the number of mutations on a strand of DNA per unit length, the number of bacteria on some surface or weed in the field,
- Management example: the number of customers arriving at a counter or call centre.
- Civil engineering example: the number of cars arriving at a traffic light.
- Finance and insurance example: the number of Losses/Claims occurring in a given period of time.

The Poisson Distributions

Poisson r.v. = the count of events that occur within an interval.

- Unknown: the # of trials n or the probability of success p
- Known: the average # of successes per time period $\lambda = np$.

$$\lim_{n\to\infty} P(X_n = k) = \lim_{n\to\infty} \frac{n!}{k!(n-k)!}\left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = e^{-\lambda}\cdot\frac{\lambda^k}{k!}.$$

Poisson distribution:

$$\Omega = \{0, 1, 2, \dots\} \quad \text{and} \quad f(k) = P(X = k) = e^{-\lambda}\cdot\frac{\lambda^k}{k!}, x \in \Omega.$$

The mean and variance of the Poisson model are the same.

$$\boxed{E(X) = \lambda \quad \text{and} \quad V(X) = \lambda.}$$

Otherwise, the Poisson distribution would not be a good model.

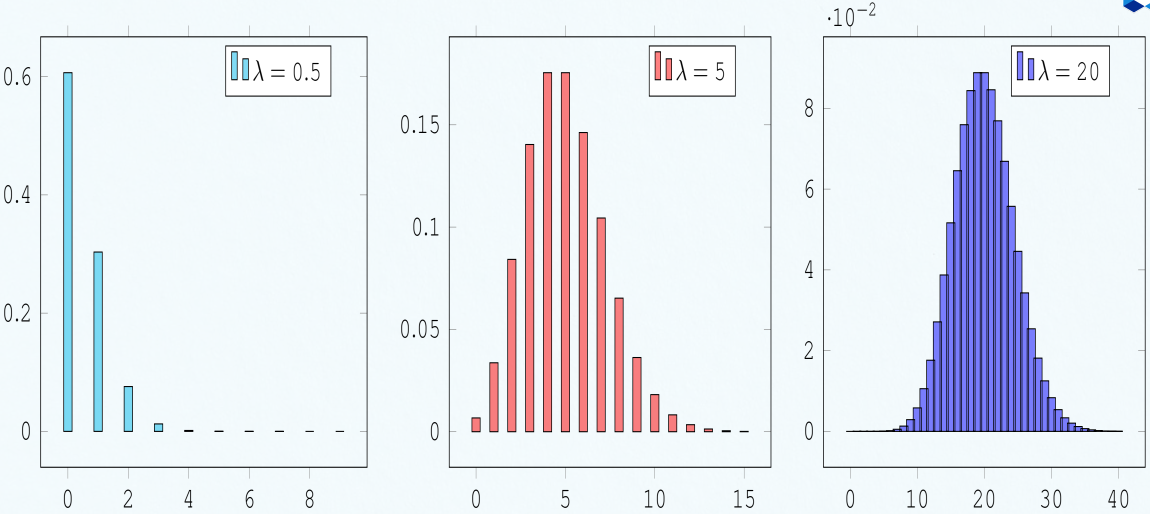


Figure: Pmf of Poisson(λ)

Example

9 Consider an experiment that consists of counting the number of α particles given off in a 1-second interval by 1 gram of radioactive material. If we know from past experience that, on average, 3.2 such α particles are given off, what is a good approximation to the probability that no more than 2 α particles will appear?

Example

9 Consider an experiment that consists of counting the number of α particles given off in a 1-second interval by 1 gram of radioactive material. If we know from past experience that, on average, 3.2 such α particles are given off, what is a good approximation to the probability that no more than 2 α particles will appear?

10 Flaws occur at random along the length of a thin copper wire. Suppose that the number of flaws follows a Poisson distribution with a mean of 2.3 flaws per mm. Find the probability of

Example

9 Consider an experiment that consists of counting the number of α particles given off in a 1-second interval by 1 gram of radioactive material. If we know from past experience that, on average, 3.2 such α particles are given off, what is a good approximation to the probability that no more than 2 α particles will appear?

10 Flaws occur at random along the length of a thin copper wire. Suppose that the number of flaws follows a Poisson distribution with a mean of 2.3 flaws per mm. Find the probability of

a. exactly 2 flaws in 1 mm of wire.

Example

9 Consider an experiment that consists of counting the number of α particles given off in a 1-second interval by 1 gram of radioactive material. If we know from past experience that, on average, 3.2 such α particles are given off, what is a good approximation to the probability that no more than 2 α particles will appear?

10 Flaws occur at random along the length of a thin copper wire. Suppose that the number of flaws follows a Poisson distribution with a mean of 2.3 flaws per mm. Find the probability of

(a) exactly 2 flaws in 1 mm of wire.

(b) exactly 10 flaws in 5 mm of wire.

Example

9 Consider an experiment that consists of counting the number of α particles given off in a 1-second interval by 1 gram of radioactive material. If we know from past experience that, on average, 3.2 such α particles are given off, what is a good approximation to the probability that no more than 2 α particles will appear?

10 Flaws occur at random along the length of a thin copper wire. Suppose that the number of flaws follows a Poisson distribution with a mean of 2.3 flaws per mm. Find the probability of

(a) exactly 2 flaws in 1 mm of wire.
(b) exactly 10 flaws in 5 mm of wire.
(c) at least 1 flaw in 2 mm of wire.

Approximation property

Let $Y \sim \text{Poisson}(\lambda)$ and $X_n \sim \text{B}(n, p_n)$ with $p_n = \lambda/n$. Then

$$\lim_{n\to\infty} \frac{P(Y=k)}{P(X_n=k)} = 1, \quad \forall k = 0, 1, \ldots$$

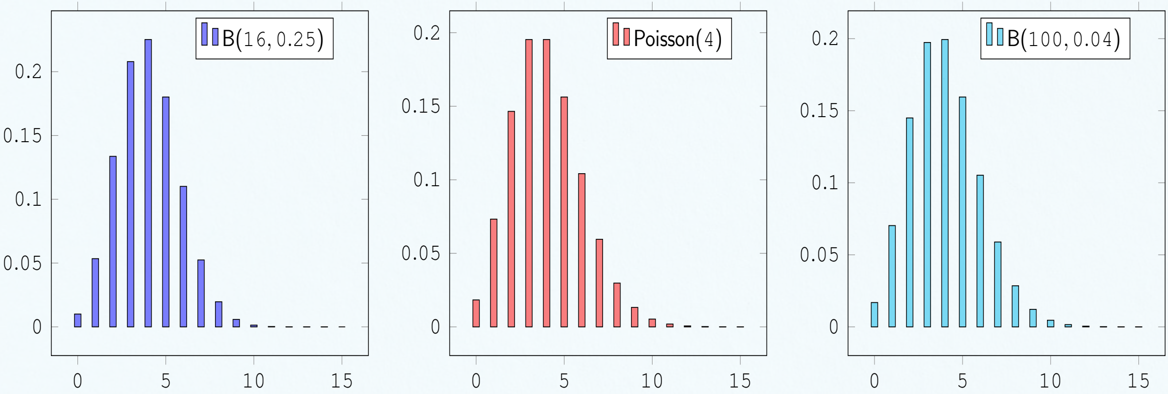


Figure: Poisson(λ) vs. B(n, p) with $np = \lambda$

Example

11 Suppose that 1 in 5000 light bulbs are defective. What is the probability that there are at least 3 defective light bulbs in a group of size 10000?

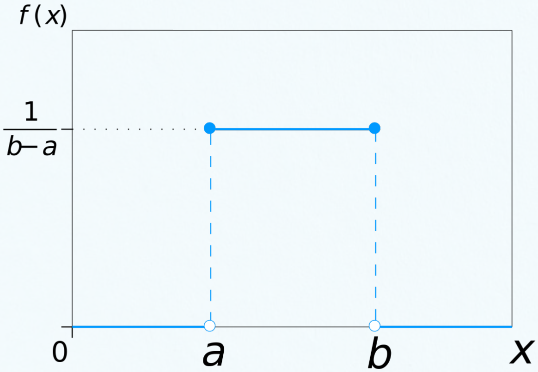
2 Continuous Distributions

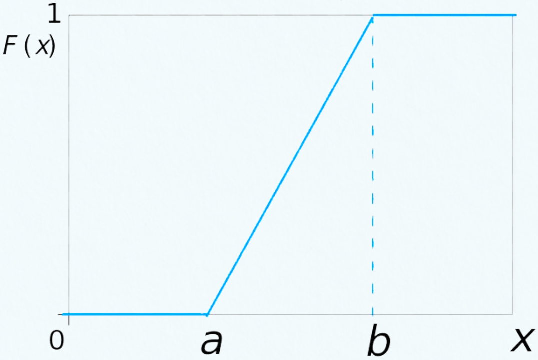
- The Continuous Uniform Distributions $\mathsf{U}(a, b)$
- The Normal Distributions $\mathsf{N}(m, \sigma^2)$

The Continuous Uniform Distributions $\mathrm{U}(a,b)$

This distribution has pdf and cdf

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & x \in [a,b] \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad F(x) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & x \in [a,b) \\ 1, & x \geq b \end{cases}.$$





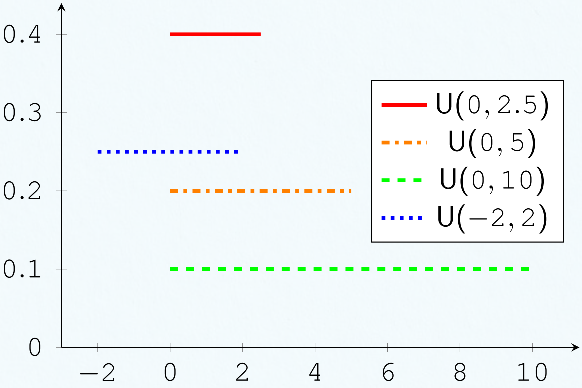


Figure: Pdf of $\mathrm{U}(a, b)$

Proposition (Properties)

1. $\mathrm{E}(X) = \dfrac{a+b}{2}$
2. $\mathrm{Var}(X) = \dfrac{(b-a)^2}{12}$

Example

12 If X is uniformly distributed over (0, 10), calculate the probability that

Example

12 If X is uniformly distributed over (0, 10), calculate the probability that

(a) $X < 3$

Example

12 If X is uniformly distributed over (0, 10), calculate the probability that

ⓐ $X < 3$

ⓑ $X > 6$

Example

12 If X is uniformly distributed over (0, 10), calculate the probability that

(a) $X < 3$ (b) $X > 6$ (c) $3 < X < 8$

Example

12 If X is uniformly distributed over (0, 10), calculate the probability that

(a) $X < 3$ (b) $X > 6$ (c) $3 < X < 8$

13 Let X be a measurement of current, which is a variable following a continuous uniform distribution on $[4.9, 5.1]$. The probability density function of X is $f(x) = 5$, $4.9 \le x \le 5.1$.

Example

12 If X is uniformly distributed over (0, 10), calculate the probability that

(a) $X < 3$ (b) $X > 6$ (c) $3 < X < 8$

13 Let X be a measurement of current, which is a variable following a continuous uniform distribution on $[4.9, 5.1]$. The probability density function of X is $f(x) = 5,\ 4.9 \le x \le 5.1$.

(a) What is the probability that the current is between 4.95 and 5.0 mA?

Example

12 If X is uniformly distributed over (0, 10), calculate the probability that

(a) $X < 3$ (b) $X > 6$ (c) $3 < X < 8$

13 Let X be a measurement of current, which is a variable following a continuous uniform distribution on $[4.9, 5.1]$. The probability density function of X is $f(x) = 5$, $4.9 \leq x \leq 5.1$.

(a) What is the probability that the current is between 4.95 and 5.0 mA?

(b) Calculate the mean and variance.

The Normal Distributions $\mathrm{N}(m, \sigma^2)$

X is called to be of a normal distribution $N(m, \sigma^2)$ if its pdf satisfies

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}, x \in \mathbb{R}$$

We often standardize a normal distribution $X \sim N(m, \sigma^2)$ by

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

In this case, Y is called a random variable of standard normal distribution, or simply a standard score. Its pdf is

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, x \in \mathbb{R}$$

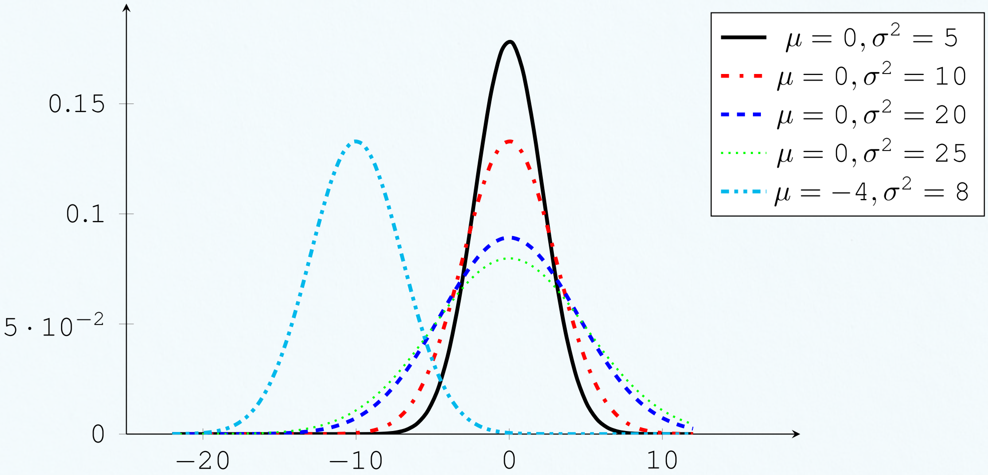


Figure: Pdf of $N(\mu, \sigma^2)$

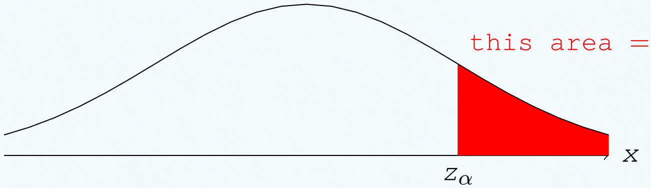
The cdf of $X \sim N(0,1)$

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

satisfies

$$\Phi(-x) = 1 - \Phi(x) \quad \text{and} \quad \Phi^{-1}(p) = -\Phi^{-1}(1-p), \quad \text{for } 0 < p < 1.$$

Denote z_α as the solution to $1 - \Phi(z) = \alpha$



z_α is called the upper α critical point or the $100(1-\alpha)$th percentile.

Example 14 – Normally Distributed Current

Suppose that the current measurements in a strip of wire are assumed to follow a normal distribution with $\mu = 10$ and $\sigma = 2$ mA.

(a) what is the probability that the current measurement is between 9 and 11 mA?

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

Properties

Proposition (Basic properties)

1. $\mathrm{E}(X) = m$ and $\mathrm{Var}(X) = \sigma^2$
2. If $X \sim N(m, \sigma^2)$ and $Y = aX + b, a \neq 0$ then
$$Y \sim \mathrm{N}(am + b, a^2\sigma^2).$$
3. If $X_i \sim N(m_i, \sigma_i^2)$ then
$$\sum_{i=1}^{n} X_i \sim \mathrm{N}\left(\sum_{i=1}^{n} m_i, \sum_{i=1}^{n} \sigma_i^2\right).$$

Example

15 Let X be a $\mathsf{N}(10, 16)$ random variable. Find the numerical values of the following probabilities

Example

15 Let X be a $\mathsf{N}(10, 16)$ random variable. Find the numerical values of the following probabilities

ⓐ $P(X \geq 15)$.

Example

15 Let X be a $N(10, 16)$ random variable. Find the numerical values of the following probabilities

a) $P(X \geq 15)$.

b) $P(X \leq 5)$.

Example

15 Let X be a $\mathsf{N}(10, 16)$ random variable. Find the numerical values of the following probabilities

(a) $P(X \geq 15)$.

(b) $P(X \leq 5)$.

(c) $P(X = 2)$.

Example

15 Let X be a $\mathsf{N}(10, 16)$ random variable. Find the numerical values of the following probabilities

(a) $P(X \geq 15)$.
(b) $P(X \leq 5)$.
(c) $P(X = 2)$.

16 Let $X \sim \mathsf{N}(5, 9)$. What is the distribution of $Y = 2X - 6$?

Example

15 Let X be a $N(10, 16)$ random variable. Find the numerical values of the following probabilities

(a) $P(X \geq 15)$. (b) $P(X \leq 5)$. (c) $P(X = 2)$.

16 Let $X \sim N(5, 9)$. What is the distribution of $Y = 2X - 6$?

17 An expert witness in a paternity suit testifies that the length (in days) of human gestation is approximately normally distributed with $\mu = 270$ and $\sigma^2 = 100$. The defendant in the suit is able to prove that he was out of the country during a period that began 290 days before the birth of the child and ended 240 days before the birth. If the defendant was the father of the child, what is the probability that the mother could have had the very long or very short gestation indicated by the testimony?

Example

18 Data from the National Oceanic and Atmospheric Administration indicate that the yearly precipitation in Los Angeles is a normal random variable with a mean of 12.08 inches and a standard deviation of 3.1 inches. Assume that the precipitation totals for the next 2 years are independent.

Example

18 Data from the National Oceanic and Atmospheric Administration indicate that the yearly precipitation in Los Angeles is a normal random variable with a mean of 12.08 inches and a standard deviation of 3.1 inches. Assume that the precipitation totals for the next 2 years are independent.

a. Find the probability that the total precipitation during the next 2 years will exceed 25 inches.

Example

18 Data from the National Oceanic and Atmospheric Administration indicate that the yearly precipitation in Los Angeles is a normal random variable with a mean of 12.08 inches and a standard deviation of 3.1 inches. Assume that the precipitation totals for the next 2 years are independent.

a. Find the probability that the total precipitation during the next 2 years will exceed 25 inches.
b. Find the probability that next year's precipitation will exceed that of the following year by more than 3 inches.

Example

18 Data from the National Oceanic and Atmospheric Administration indicate that the yearly precipitation in Los Angeles is a normal random variable with a mean of 12.08 inches and a standard deviation of 3.1 inches. Assume that the precipitation totals for the next 2 years are independent.

a. Find the probability that the total precipitation during the next 2 years will exceed 25 inches.
b. Find the probability that next year's precipitation will exceed that of the following year by more than 3 inches.

19 Consider three independent memory chips. Suppose that the lifetime of each memory chip has normal distribution with mean 300 hours and standard deviation 10 hours. Compute the probability that at least one of three chips lasts at least 290 hours.