Continuous Random Variables

CSGE602013 - Statistics and Probability

Fakultas Ilmu Komputer

Universitas Indonesia

2017

Credits

The content was based on previous semester’s course slides created by **all previous lecturers**.

References

- Introduction to Probability and Statistics for Engineers & Scientists, 4th ed.,
  - Sheldon M. Ross, Elsevier, 2009.
- A First Course in Probability, 8th Edition.
  - Sheldon M. Ross
- Applied Statistics for the Behavioral Sciences, 5th Edition,
  - Hinkle., Wiersma., Jurs., Houghton Mifflin Company, New York, 2003.
- Probability and Statistics for Engineers & Scientists, 4^{th} Edition
  - Anthony J. Hayter, Thomson Higher Education

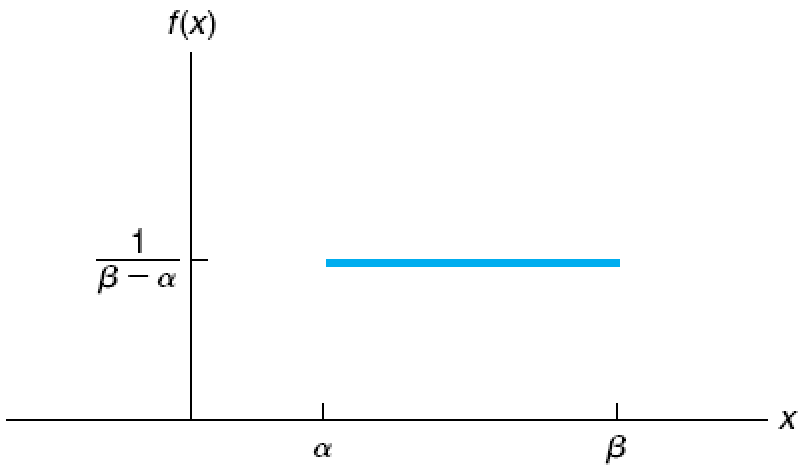
Outline

- Uniform Random Variables
- Exponential Random Variables
- Normal (Gaussian) Random Variables
- Distributions Arising from The Normal

Uniform Random Variables

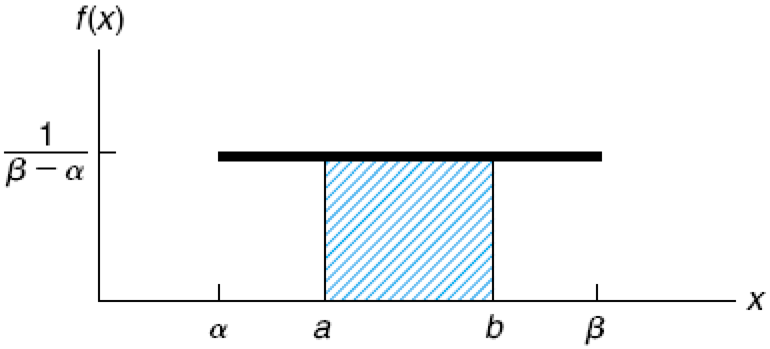
A random variable **X** is said to be **uniformly** distributed over the interval **[α, β]** if its probability density function is given by:

$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha} & \alpha \le x \le \beta \\ 0 & otherwise \end{cases}$$



Graph of $f(x)$ for a uniform [a, β].

$$P(a < x < b) = \frac{1}{\beta-\alpha}\int_a^b dx = \frac{b-a}{\beta-\alpha}$$



$P\{a < X < b\}$,
(a, b) *is a subinterval of* $[\alpha, \beta]$

$$E[X] = \int_\alpha^\beta \frac{x}{\beta-\alpha}\,dx = \frac{\alpha+\beta}{2}$$

$$Var(X) = E[X^2] - (E[X])^2 = \frac{(\beta-\alpha)^2}{12}$$

$$F(x) = P(X \le x) = \begin{cases} 0 & x < \alpha \\ \dfrac{1}{\beta-\alpha}\displaystyle\int_\alpha^x dx = \dfrac{x-\alpha}{\beta-\alpha} & \alpha \le x \le \beta \\ 1 & \alpha > \beta \end{cases}$$

If **X** is **uniformly** distributed over the interval **[0, 10]**, compute:

1) ***F*(x)**
2) ***P*(1 < X < 4)**
3) ***P*(X < 5)**
4) ***P*(X > 6)**
5) ***P*(X > 6 | X > 5)**

1)

$$F(x) = P(X \le x) = \begin{cases} 0 & x < 0 \\ \dfrac{1}{10-0}\displaystyle\int_0^x dx = \dfrac{x}{10} & 0 \le x \le 10 \\ 1 & x > 10 \end{cases}$$

2)

$$P(1 < X < 4) = F(4) - F(1) = \frac{4-1}{10} = 0.3$$

3)

$$P(X < 5) = F(5) = \frac{5}{10} = 0.5$$

4)

$$P(X > 6) = 1 - P(X \le 6) = 1 - F(6) = 1 - \frac{6}{10} = 0.4$$

5)

$$P(X > 6 \mid X > 5) = \frac{P(X > 6, X > 5)}{P(X > 5)}$$

$$= \frac{P(X > 6)}{P(X > 5)} = \frac{1 - F(6)}{1 - F(5)} = \frac{4}{5}$$

Practice

Waktu yang dibutuhkan untuk menyelesaikan sebuah operasi terdistribusi secara uniform pada selang **$30 \leq x \leq 40$** detik.

- Tentukan peluang bahwa operasi membutuhkan waktu lebih dari 35 detik agar selesai !
- Tentukan mean dan variance dari waktu yang dibutuhkan untuk menyelesaikan sebuah operasi !
- Jika diketahui bahwa operasi belum selesai hingga 35 detik, berapa peluang operasi selesai sebelum 3 detik berikutnya ?

Buses arrive at a specified stop at 15-minute intervals starting at **7 A.M.** That is, they arrive at **7**, **7:15**, **7:30**, **7:45**, and so on.

If a passenger arrives at the stop at a time that is uniformly distributed between **7** and **7:30**, find the probability that he waits:

- less than **5 minutes** for a bus
- at least **12 minutes** for a bus

In this example $\alpha = 0, \beta = 30$;

$$f(x) = \begin{cases} \frac{1}{30}, if\ 0 \leq x \leq 30 \\ 0, \qquad\qquad otherwise \end{cases}$$

Note that between 7 and 7:30 there are two schedules for the bus, i.e. 7.15 and 7:30

In (a), for this passenger to wait less than 5 minutes he must arrive between 7:10 and 7:15 or 7:25 and 7:30.

- Thus the probability is
- $P(10 < X < 15) + P(25 < X < 30) = \frac{5}{30} + \frac{5}{30} = \frac{1}{3}.$

In (b), he must arrive between 7:0 and 7:03 or 7:15 and 7:18;

- So the probability is
- $P(0 < X < 3) + P(15 < X < 18) = \frac{3}{30} + \frac{3}{30} = \frac{1}{5}.$

Practice

Anda tiba di bus stop pada pukul 10:00. Anda tahu bahwa waktu kedatangan bus terdistribusi secara uniform antara pukul 10:00 hingga 10:30.

- Berapa probabilitas bahwa Anda harus menunggu lebih dari 10 menit ?
- Jika pada pukul 10:10 bus belum juga datang, berapa probabilitas bahwa Anda harus menunggu paling sedikit 10 menit lagi ?

Exponential Random Variables

The exponential distribution is often used to describe **the amount of time until some specific event occurs**.

For example:

- The amount of time until an earthquake occurs
- The amount of time until a new war breaks out
- The amount of time until a telephone call you receive turns out to be a wrong number
- Time between **two** successive job arrivals to a file server (**interarrival time**)
- Time to failure of a component (**lifetime** of a component)

The random variable **X** is said to be an **exponential** random variable (exponentially distributed) with parameter **λ** if it's **PDF** is given by:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$X \sim Exp(\lambda)$$

$\lambda \rightarrow$

- Arrival rate per **unit** interval
- **Average number** of **events** occuring **per unit** of interval

CDF:

$$F(x) = P(X \le x) = \int_0^x \lambda e^{-\lambda y}\, dy = 1 - e^{-\lambda x} \qquad x \ge 0$$

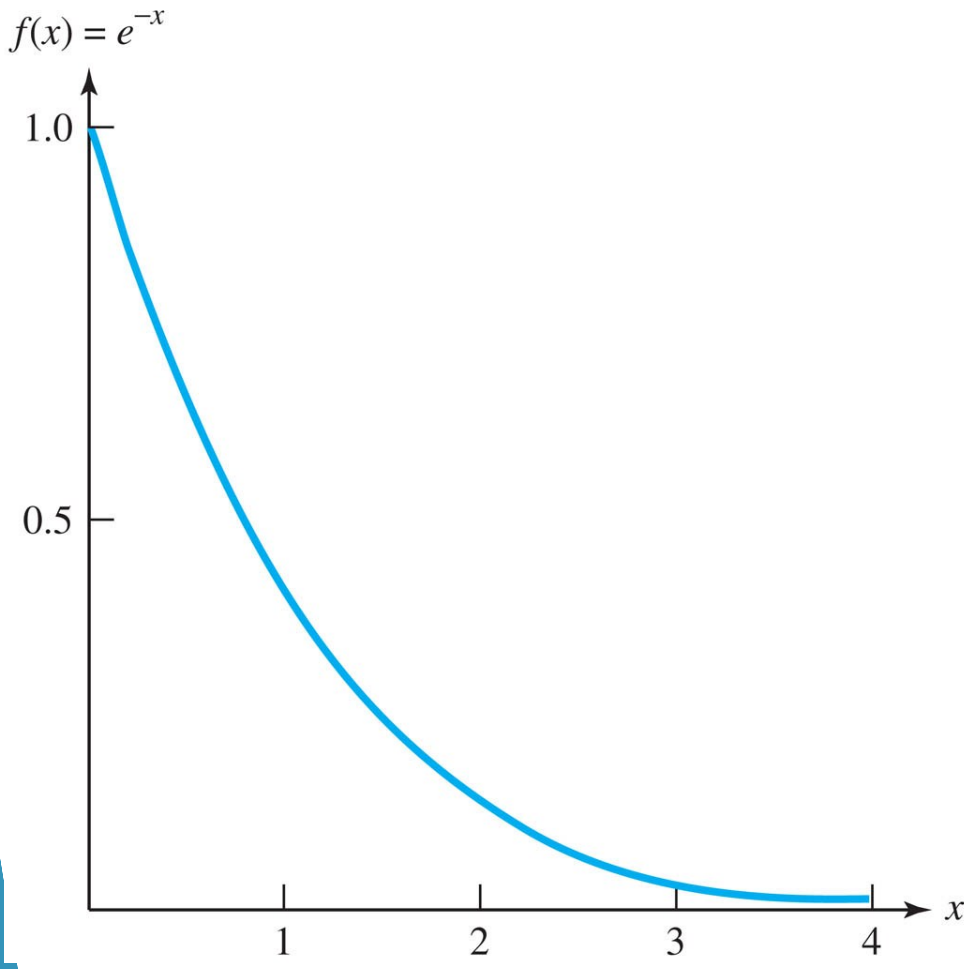
Or, in complete

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \ge 0 \end{cases}$$

MGF (Moment Generating Function) (**optional**):

$$\phi(\theta) = \frac{\lambda}{\lambda - \theta}$$

PDF of an exponential distribution with λ = 1



Mean & Variance:

$$E[X] = \frac{1}{\lambda}$$

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

Prove it using Moment generating function of exponential distribution.

Poisson distribution & **Exponential** distribution

The number of arrivals **within an interval** follows Poisson distribution.

The amount of time between **two successive arrivals** is exponentially distributed.

Jarak antar retakan pada sebuah pipa air mengikuti distribusi eksponensial dengan **rata-rata** terdapat **3 retakan per meter**.

Misal, seorang petugas melakukan inspeksi dimulai dari ujung pipa. Berapakah peluang ia **tidak** menemukan retakan setelah memeriksa sejauh **2 meter** ?

Jawab:

Misal, **X** : R.V. yang menyatakan jarak hingga menemukan sebuah retakan (atau jarak antar retakan).

$$\lambda = 3 \, {}^{retakan}\!/_{meter} \qquad X \sim Exp(3)$$

$$P(X > 2) = 1 - P(X \le 2)$$

$$= 1 - F(2) = 1 - (1 - e^{-3(2)}) = e^{-6}$$

The arrival numbers of the customer in a store is a **Poisson RV** with rate **λ = 5 visitors per hour**.

- What is the probability that the next visitor coming in less than **10 minutes** ?
- What is the probability that the next visitor coming between **minute 10 to 12 from now** ?

Answer:

Let **X** be a R.V. that denotes the amount of time until a customer arrive (or inter-arrival time)

(a)

$$\lambda = 5 \; {}^{visitors}\!/\!_{hour} \qquad X \sim Exp(5)$$

$$10 \min = 1/6 \; hour$$

$$P(X < 1/6) = F(1/6) = 1 - e^{-5/6} = 0.5654$$

(b)

$$10 \min = 1/6 \; hour \qquad 12 \min = 1/5 \; hour$$

$$\begin{aligned} P(1/6 < X < 1/5) &= F(1/5) - F(1/6) \\ &= (1 - e^{-5/5}) - (1 - e^{-5/6}) \\ &= 0.0667 \end{aligned}$$

Misalkan umur sebuah bakteri sebelum mati terdistribusi secara eksponensial dengan rata-rata 10 menit. Berapa peluang bahwa umur bakteri akan lebih dari 20 menit ?

Jawab:

Misal, **X** : lamanya waktu hingga bakteri mati (umur bakteri)

$$E[X] = 10 \; menit$$

$$\lambda = \frac{1}{E[X]} = {}^{1}\!/_{10} \; {}^{kematian}\!/_{menit}$$

$$P(X > 20) = 1 - F(20) = 1 - (1 - e^{-20/10}) = e^{-2}$$

Markov (Memoryless) Property

That is,

$$P(X > s + t \mid X > t) = P(X > s) \qquad s, t \geq 0$$

The **exponential distribution** is the only continuous distribution that has the memoryless property.

The distribution of additional functional life of an item of age t is the same as that of a new item.

There is **no need to remember** the age of a functional item since as long as it is still functional it is "**as good as new.**"

Markov (Memoryless) Property

Proof:

$$
\begin{aligned}
P(X > s+t \mid X > t) \\
&= \frac{P(X > s+t, X > t)}{P(X > t)} \\
&= \frac{P(X > s+t)}{P(X > t)} \\
&= \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} \\
&= e^{-\lambda s} = P(X > s)
\end{aligned}
$$

$$P(X > s+t, X > t) = P(X > s)P(X > t)$$

Suppose that a number of miles that a car run before its battery wears out is exponentially distributed with an average value of **10,000 miles.** If a person desires to take a **5,000** mile trip, what is the probability that he will be able to complete his trip without to replace the battery?

Let **X** be a random variable including the remaining lifetime (in thousand miles) of the battery. Then,

$$E[X] = 1/\lambda = 10 \qquad \Rightarrow \lambda = 1/10$$

$$P(X > 5) = 1 - F(5) = e^{-5\lambda} = e^{-1/2} = 0.604$$

What if **X** is **NOT exponential R. V. ?**

$$P(X > t+5 \mid X > t) = \frac{1 - F(t+5)}{1 - F(t)}$$

Additional information **t** should be known !

Practice

Lamanya radio berfungsi (dalam tahun) terdistribusi secara eksponensial dengan parameter $\boldsymbol{\lambda} = 1/8$.

Jika Cecep membeli sebuah radio **bekas**, berapa probabilitas bahwa radio tersebut akan masih bekerja hingga 10 tahun kedepan ?

Proposition:

If $X_1, X_2, \ldots, X_n$ are independent exponential random variables having respective parameters $\lambda_1, \lambda_2, \ldots, \lambda_n$,

then **min** $(X_1, X_2, \ldots, X_n)$ is the **exponential R.V.** with parameter

$$\sum_{i=1}^{n} \lambda_i$$

Proof:

$$P(\min(X_1, X_2, \ldots, X_n) > x) = P(X_1 > x, X_2 > x, \ldots, X_n > x)$$

$$= \prod_{i=1}^{n} P(X_i > x)$$ **By independence**

$$= \prod_{i=1}^{n} \exp(-\lambda_i x)$$

$$= \exp(-\sum_{i=1}^{n} \lambda_i x)$$

A series system is one that all of its components to function in order for the system itself to be functional. For an **n component series** system in which the component lifetimes are independent exponential random variables with respective parameters $\boldsymbol{\lambda_1, \lambda_2, \ldots, \lambda_n}$.

What is the probability the system survives for a time ***t*** ?

Let **X** : system's lifetime

System's lifetime is equal to the **minimal component life**. Based on previous proposition:

$$X \sim Exp\left(\sum_{i=1}^{n} \lambda_i\right)$$

$$P(X > t) = \exp\left(-\sum_{i=1}^{n} \lambda_i t\right)$$

Practice

Waktu (dalam jam) yang dibutuhkan untuk membenahi sebuah mesin adalah sebuah random variable eksponensial dengan parameter $\lambda = 1$ (rata-rata 1 buah mesin dibenahi **per jam**).

- Berapa probabilitas bahwa waktu yang dibutuhkan untuk membenahi mesin melebihi 2 jam ?
- Berapa probabilitas bahwa waktu yang dibutuhkan untuk membenahi mesin paling sedikit 3 jam (dihitung dari awal mulai), jika diketahui pekerjaan ini sudah dilakukan selama 2 jam ?

Practice

Diketahui bahwa dalam waktu 30 menit, rata-rata terjadi 5 kecelakaan pada jalan raya. Tentukan:

- Peluang bahwa dalam 1.5 jam tidak ada kecelakaan yang terjadi ?
- Misalkan, sebuah kecelakaan telah terjadi. Berapa peluang kecelakaan berikutnya terjadi setelah 1 jam ?
- Diketahui bahwa 1 jam sejak awal pemantauan tidak terjadi kecelakaan, berapakah peluang bahwa dalam 2 jam berikutnya tetap tidak ada kecelakaan ?

Practice

Misalkan suatu sistem mengandung komponen yang umurnya (dalam tahun) terdistribusi secara eksponensial dengan rata-rata waktu hingga mati adalah 5 tahun.

Bila sebanyak 5 komponen tersebut dipasang dalam sistem yang berlainan, berapakah peluang bahwa paling sedikit 2 komponen masih akan berfungsi pada akhir tahun ke-8 ?

0.2

0.26

Normal Random Variables

Definition of Normal R. V.

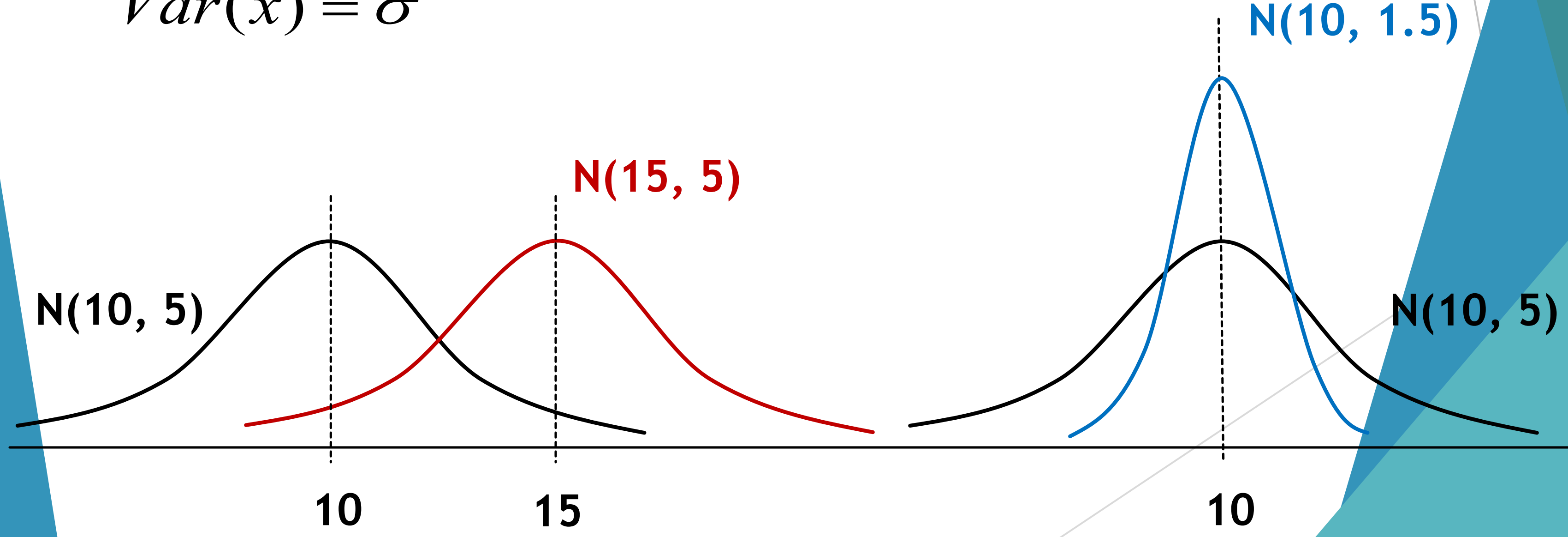
X is a Normal (Gaussian) R.V. with parameter **μ** dan **σ²**:

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

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

$$E[X] = \mu$$

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



Standard Normal Distribution

Standard Normal R.V. **Z** is obtained using:

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

So,

$$E[Z] = \frac{E[X] - \mu}{\sigma} = \frac{\mu - \mu}{\sigma} = 0$$

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

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

PDF:

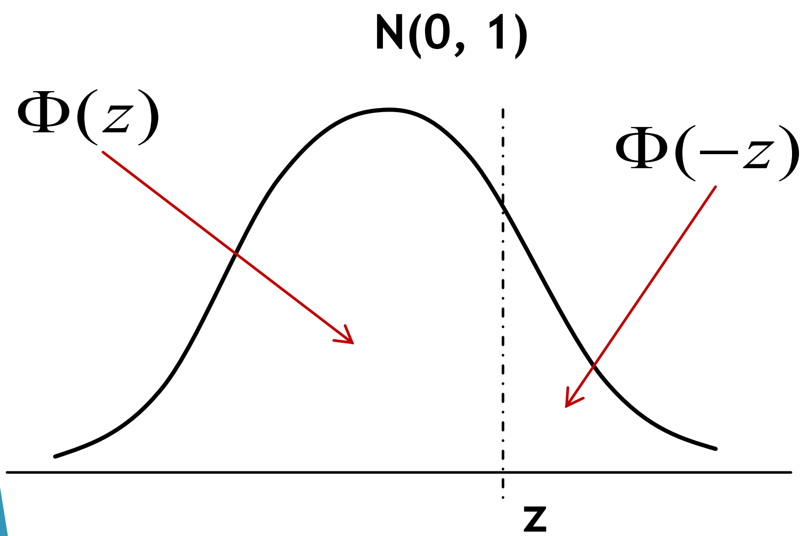
$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \qquad -\infty < z < \infty$$

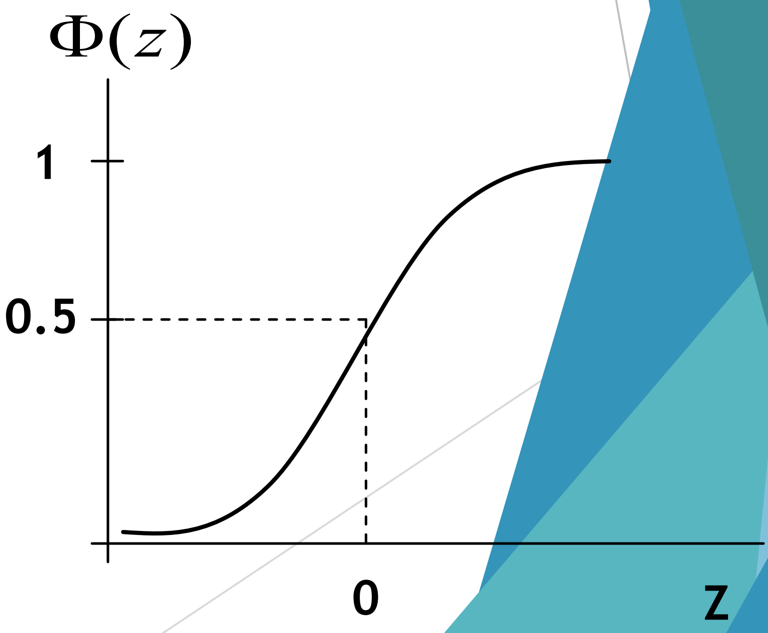
Standard Normal Distribution

CDF:

$$\Phi(z) = \int_{-\infty}^{z} f(y)\, dy$$

$$1 - \Phi(z) = P(Z \geq z) = P(Z \leq -z) = \Phi(-z)$$

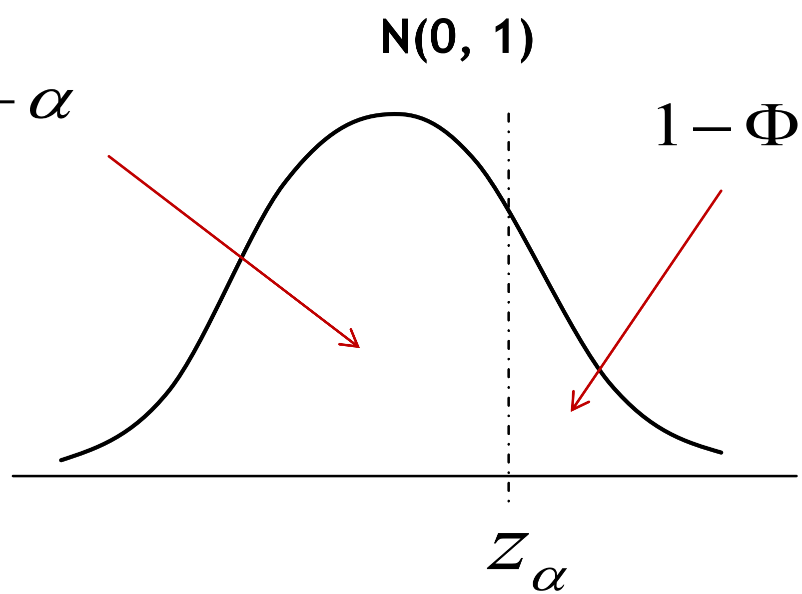




Standard Normal Distribution

Introducing the z_α notation

$$\Phi(z_\alpha) = 1 - \alpha$$



$$1 - \Phi(z_\alpha) = \Phi(-z_\alpha)$$
$$= \alpha$$

Probability that a standard normal R.V. is greater than $\mathbf{z_\alpha}$ is equal to $\boldsymbol{\alpha}$.

How to calculate probability for General Normal R. V ?

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

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

$$P(a \le X \le b) = P\left(\frac{a-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right)$$

$$= P\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right)$$

$$= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$P(\mu - c\sigma \le X \le \mu + c\sigma) = P(-c \le Z \le c)$$

$$P(X \le \mu + \sigma z_\alpha) = P(Z \le z_\alpha) = 1 - \alpha$$

$$P(\mu - \sigma z_{\alpha/2} \le X \le \mu + c z_{\alpha/2}) = P(-z_{\alpha/2} \le Z \le z_{\alpha/2}) = 1 - \alpha$$

Standard Normal Table $\Phi(z)$

To facilitate the calculation (if the calculator is not available), some text books often provide CDF table (e.g., Table C3 in the Kishor book). This table only provide some **$\Phi(z)$**, that are for z from 0.00 to 3.90.

- How to read:
  - to find $\Phi(x.yz)$, go to row $(x.y)$, then in the same row go to the right to find column z.
  - Expecially for $z > 3$ $\Phi(x.y)$, the right column is for y.
  - Other values can be calculated based on the symmetry properties of the function of pdf,
  - for example $\Phi(-0.5) = 1 - \Phi(0.5)$

Standard Normal Table Φ(z)

$$\alpha = 0.33$$
$$z_{\alpha} = 0.44$$
$$\Phi(z_{\alpha}) = 0.67$$

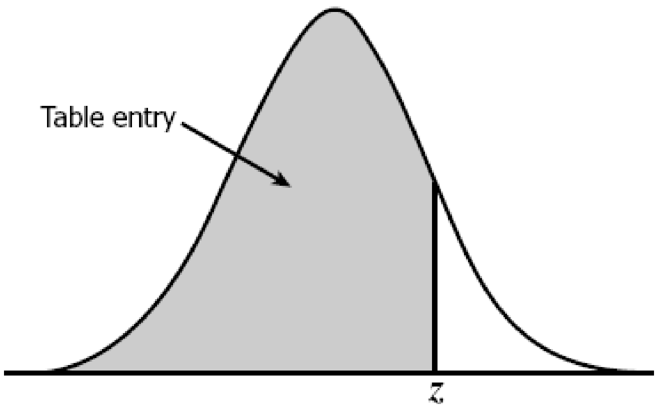


Table entry for z is the area under the standard normal curve to the left of z.

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	$\boxed{.6700}$	.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

Analogue signals received by a detector is modeled as Gaussian RV ***N*(200, 256)** using microvolt unit (μV).

What is the probability the signal received **exceeds** 240 μV ?

Let **X** : analogue signals received by a detector, **X ~ N(200, 256)**

See the Z-table !

$$P(X > 240) = 1 - P(X \leq 240)$$

$$= 1 - \Phi\left(\frac{240 - 200}{16}\right) = 1 - \Phi(2.5) = 0.0062$$

What is the probability the signal received exceeds 240 μV **if** the signal is known larger than 210 μV ?

$$P(X > 240 \mid X > 210) = \frac{P(X > 240)}{P(X > 210)}$$

$$= \frac{1 - P(X \leq 240)}{1 - P(X \leq 210)} = \frac{1 - \Phi(2.5)}{1 - \Phi\left(\frac{210 - 200}{16}\right)} = \frac{1 - \Phi(2.5)}{1 - \Phi(0.625)}$$

$$= 0.02335$$

Suatu jenis baterai mobil rata-rata berumur 3 tahun dengan simpangan baku 0,5 tahun. Bila umur baterai dianggap berdistribusi normal, carilah

- Peluang baterai berumur kurang dari 4 tahun ?
- Peluang baterai berumur kurang dari 2,3 tahun ?
- Peluang baterai berumur lebih dari 3,5 tahun ?
- Peluang baterai berumur antara 2,5 hingga 3,5 tahun ?

Suatu jenis radio mempunyai rata-rata umur 800 jam dan variansi 1600. Umur radio mengikuti distribusi Normal.

Jika Cecep membeli radio bekas yang diketahui sudah berumur 400 jam, berapa peluang radio tersebut masih berfungsi hingga 300 jam setelah pembelian ?

Suatu pengukur dipakai untuk menolak semua suku cadang yang ukurannya tidak memenuhi ketentuan **1,5 ± d**. Diketahui bahwa pengukuran tersebut berdistribusi Normal dengan rataan 1,5 dan simpangan baku 0,2.

Tentukanlah **d** sehingga ketentuan tersebut "mencakup" 95% dari seluruh pengukuran !

The power **W** dissipated in a resistor is proportional to the square of the voltage **V**. that is,

$$W = rV^2$$

where r is a constant. If **r = 3**, and **V** can be assumed (to a very good approximation) to be a normal random variable with **mean 6** and **standard deviation 1**.

$$Var(V) = E[V^2] - (E[V])^2$$

$$V \sim N(6,1)$$

a) **E[W]** ?

$$E[W] = E[rV^2] = rE[V^2]$$
$$= r\left(Var(V) + (E[V])^2\right)$$
$$= 3(1+36) = 111$$

b) ***P*(W > 120)** ?

$$P(W > 120) = P(rV^2 > 120) = P(V^2 > 40)$$
$$= P(V > \sqrt{40})$$
$$= 1 - P(V \le \sqrt{40}) = 1 - \Phi\left(\frac{\sqrt{40} - 6}{1}\right) = 0.3727$$

Linear Combination of Normal Random Variables

Sum of **independent** normal random variables is also a normal random variables !

$$X \sim N(\mu, \sigma^2)$$
$$\Rightarrow Y = aX + b$$
$$\Rightarrow Y \sim N(a\mu + b, a^2\sigma^2)$$

a and **b** are constants

$$X_1 \sim N(\mu_1, \sigma_1^{\;2})$$
$$X_2 \sim N(\mu_2, \sigma_2^{\;2})$$
$$\Rightarrow Y = X_1 + X_2$$
$$\Rightarrow Y \sim N(\mu_1 + \mu_2, \sigma_1^{\;2} + \sigma_2^{\;2})$$

X_1 and **X_2** are **independent !**

Linear Combination of Normal Random Variables

$$X_i \sim N(\mu_i, \sigma_i^{\ 2})$$

for **1 ≤ i ≤ n**, **X_i** are **independent**

$$\Rightarrow Y = a_1 X_1 + a_2 X_2 + \ldots + a_n X_n + b$$

a_i and **b** are constants

$$\Rightarrow Y \sim N(b + \sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^{\ 2} \sigma_i^{\ 2})$$

Proof can be made using moment generating function

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.**

Find the probability that the total precipitation during the **next 2 years** will exceed 25 inches !

We know that $\mathbf{X_1 + X_2}$ is also normal random variable.

$$X_i \sim N(12.08, (3.1)^2)$$

$$X_1 + X_2 \sim N(12.08 + 12.08,\ 2(3.1)^2)$$
$$\sim N(24.16, 19.22)$$

$$P(X_1 + X_2 > 25) = P\left(\frac{X_1 + X_2 - 24.16}{\sqrt{19.22}} > \frac{25 - 24.16}{\sqrt{19.22}}\right)$$
$$= P(Z > 0.1916) = 0.424$$

Find the probability that next year's precipitation will exceed that of the following year by more than 3 inches ?

$$X_i \sim N(12.08, (3.1)^2)$$

We know **X_1 - X_2** is also a normal random variable with mean & variance:

$$E[X_1 - X_2] = E[X_1] - E[X_2] = 0$$

$$Var(X_1 - X_2) = Var(X_1) + (-1)^2 Var(X_2)$$

$$= 2.(3.1)^2 = 19.22$$

$$P(X_1 > X_2 + 3) = P(X_1 - X_2 > 3)$$

$$= P\left(\frac{X_1 - X_2 - 0}{\sqrt{19.22}} > \frac{3-0}{\sqrt{19.22}}\right)$$

$$= P(Z > 0.6843)$$

$$= 0.2469$$

Normal Approximation to The Binomial Distribution

Besides it can be approximated by a **Poisson RV**, Binomial RV can be approximated by the **Normal RV**.

$$X \sim B(n,p) \underset{approx}{\rightarrow} X \underset{approx}{\sim} N(np, np(1-p))$$

This approximation works well as long as **np ≥ 5** and **n(1-p) ≥ 5**.

Normal Approximation to The Binomial Distribution

We need **correction** because Normal RV is continuous RV.

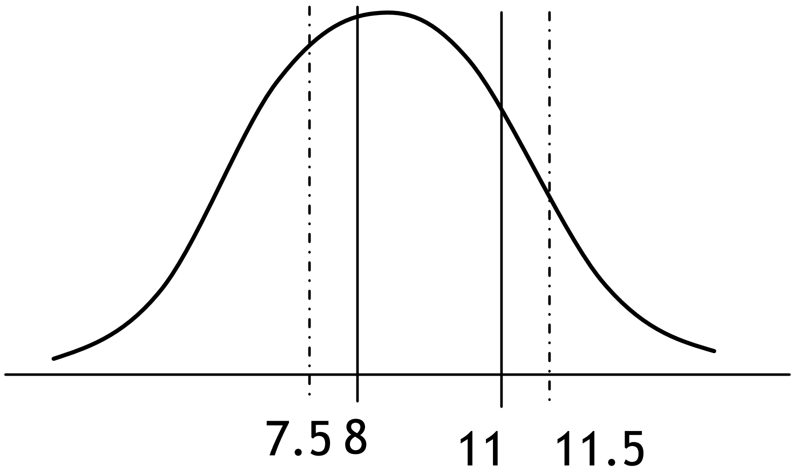
$$X \sim B(n, p) \underset{approx}{\longrightarrow} Y \sim N(np, np(1-p))$$

$$P(X \le x) \approx P(Y < x + 0.5) = \Phi\left(\frac{x + 0.5 - np}{\sqrt{np(1-p)}}\right)$$

$$P(X < x) \approx P(Y < x - 0.5) = \Phi\left(\frac{x - 0.5 - np}{\sqrt{np(1-p)}}\right)$$

$$P(X \ge x) \approx 1 - \Phi\left(\frac{x - 0.5 - np}{\sqrt{np(1-p)}}\right)$$

Normal Approximation to The Binomial Distribution



$$X \sim B(16, 0.5)$$

$$Y \sim N(8, 4)$$

$$P(8 \le X \le 11) = \sum_{x=8}^{11} \binom{16}{x} (0.5)^x (0.5)^{16-x} = 0.5598$$

$$\begin{aligned} P(7.5 \le Y \le 11.5) &= \Phi\left(\frac{11.5-\mu}{\sigma}\right) - \Phi\left(\frac{7.5-\mu}{\sigma}\right) \\ &= \Phi\left(\frac{11.5-8}{2}\right) - \Phi\left(\frac{7.5-8}{2}\right) \\ &= \Phi(1.75) - \Phi(-0.25) = 0.5586 \end{aligned}$$

Peluang seorang penderita sembuh dari suatu penyakit adalah 0,4. Bila diketahui ada 100 orang yang telah terserang penyakit ini, hitunglah:

- Peluang kurang dari 30 yang sembuh ? (exclusive)
- Peluang tepat 20 sembuh ?
- Peluang antara 20 dan 30 yang sembuh ? (inclusive)

- The polls are often conducted with a random selection from a number of people from several places. For instance, it is selected 400 people from a city inhabited by men and the women assumed in the same amount. What is the probability that within the selected people there are 190 women at most?

- This is a Binomial problem with $n = 400$ and $p = 0.5$.

  Find $P(Y \leq 190)$. Manual calculations with a 400! is impossible. While using Normal RV approximation, we obtain:

$$P(Y \leq 190) \approx \Phi\left(\frac{190 + 0.5 - 200}{\sqrt{100}}\right) = \Phi(-0.95) = 0.1711$$

The probability that an oyster produces a pearl is **0.6**.

How many oyster does an oyster farmer need to farm in order to be 99% confident of having at least 1000 pearls ?

X: the number of pearls

X ~ B(n, 0.6) => we approximate using **Y ~ N(0.6n, 0.24n)**

$$P(X \geq 1000) \approx P(Y \geq 1000) = 1 - \Phi\left(\frac{999.5 - 0.6n}{\sqrt{0.24n}}\right) \geq 0.99$$

$$\Phi\left(\frac{999.5 - 0.6n}{\sqrt{0.24n}}\right) \leq 1 - 0.99 = 0.01$$

$$\Rightarrow \frac{999.5 - 0.6n}{\sqrt{0.24n}} \leq -2.33$$

$$\Rightarrow n \geq 1746$$

Distribution Arising from Normal Dist.

1. The Chi-Square Distribution

Definition

If Z_1, Z_2, ..., Z_n are independent standard normal random variables, then **X**, defined by

$$X = Z_1^{\;2} + Z_2^{\;2} + \ldots + Z_n^{\;2}$$

is said to have a ***chi-square distribution*** with **n degrees of freedom**, or we use notation:

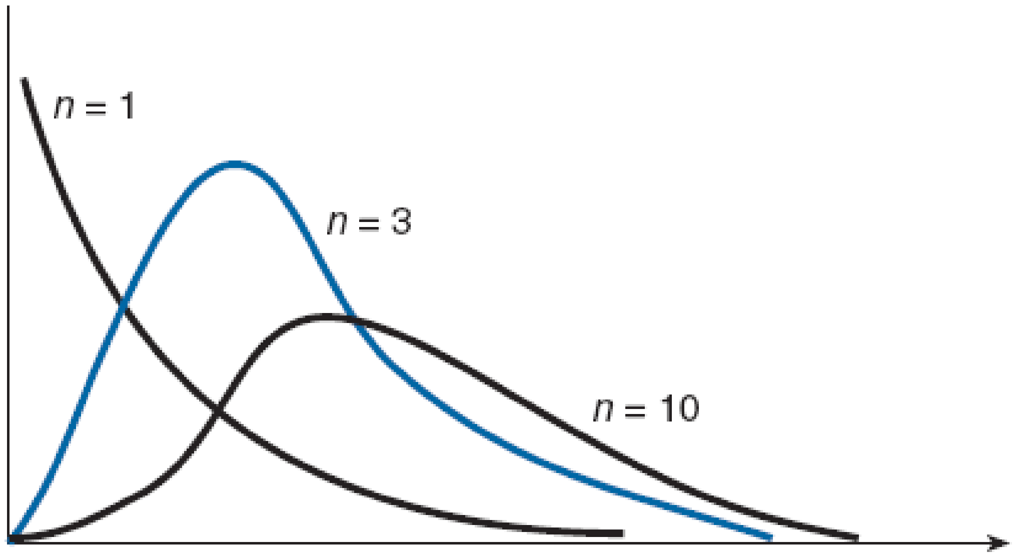
$$X \sim \chi_n^{\;2}$$

$$E[X] = n$$

$$Var(X) = 2n$$

1. The Chi-Square Distribution

Chi-square density function



The chi-square density function with n degrees of freedom.

1. The Chi-Square Distribution

The chi-square distribution has the **additive property**.

If X_1 & X_2 are independent chi-square random variables with n_1 and n_2 degrees of freedom, then $X_1 + X_2$ is chi-square with $n_1 + n_2$ degrees of freedom.

$$X_1 \sim \chi^2_{n1} \qquad\qquad X_2 \sim \chi^2_{n2}$$

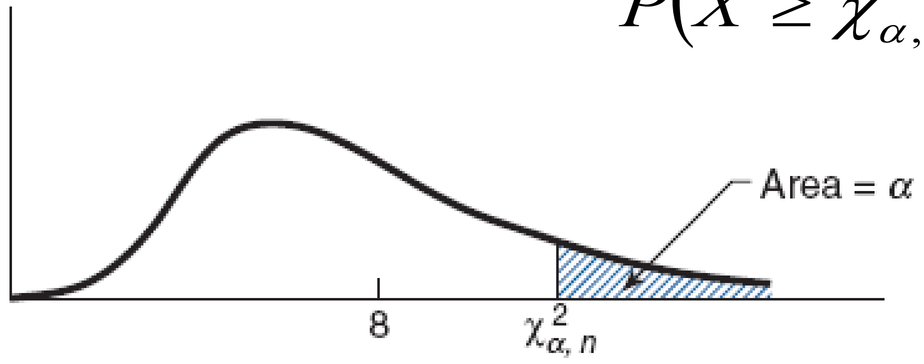
$$X_1 + X_2 \sim \chi^2_{n1+n2}$$

This can be easily proven using the definition of Chi-square.

1. The Chi-Square Distribution

If X is a chi-square random variable with n degrees of freedom, then for any $\alpha \in (0,1)$: the quantity $\chi^2_{\alpha,n}$ is defined to be such that:

$$P\left(X \geq \chi^2_{\alpha,n}\right) = \alpha$$



The chi-square density function with 8 degrees of freedom.

1. The Chi-Square Distribution

Suppose that we are attempting to locate a target in three-dimensional space, and that the three coordinate errors (in meters) of the point chosen are independent normal random variables with mean 0 and standard deviation 2.

Find the probability that the distance between the point chosen and the target exceeds 3 meters.

$$D^2 = X_1^2 + X_2^2 + X_3^2 \qquad Z_i = \frac{X_i - 0}{2} = \frac{X_i}{2}$$

$$\begin{aligned} P\left(D^2 > 9\right) &= P\left(Z_1^2 + Z_2^2 + Z_3^2 > 9/4\right) \\ &= P\left(\chi_3^2 > 9/4\right) \\ &= 0.5222 \end{aligned}$$

2. The t-Distribution

If **Z** and $\mathbf{x_n}^2$ are indepedent random variables, with **Z** having a standard normal distribution and $\mathbf{x_n}^2$ having a chi-square distribution with **n** degrees of freedom, then the random variable $\mathbf{T_n}$ defined by

$$T_n = \frac{Z}{\sqrt{\chi_n^2 / n}}$$

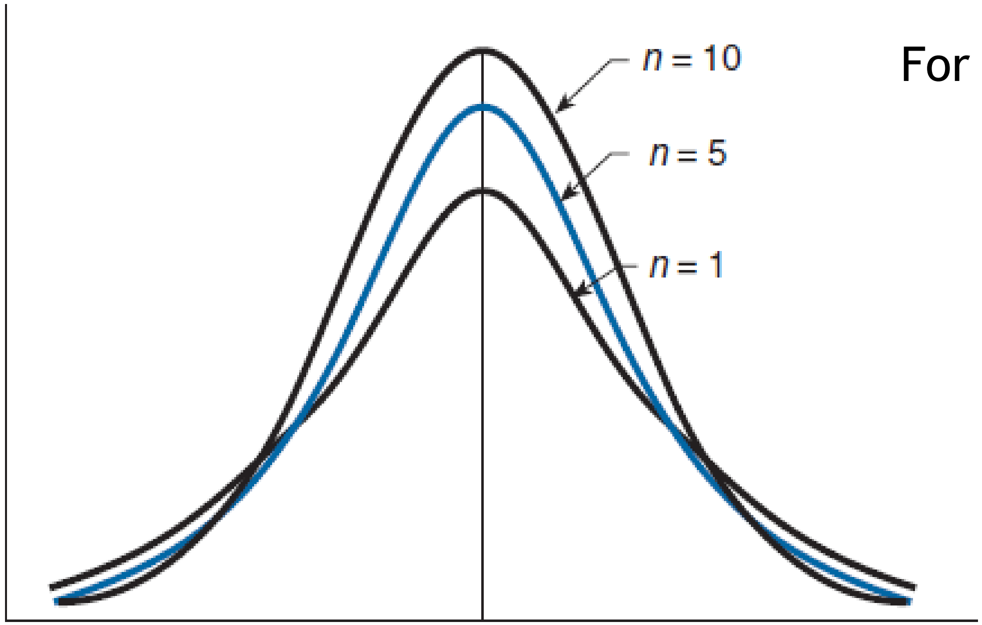
is said to have a ***t-distribution with n degrees of freedom***.

$$E[T_n] = 0 \qquad n > 1$$

$$Var(T_n) = \frac{n}{n-2} \qquad n > 2$$

2. The t-Distribution

$$\frac{\chi_n^2}{n} = \frac{Z_1^2 + \ldots + Z_n^2}{n}$$



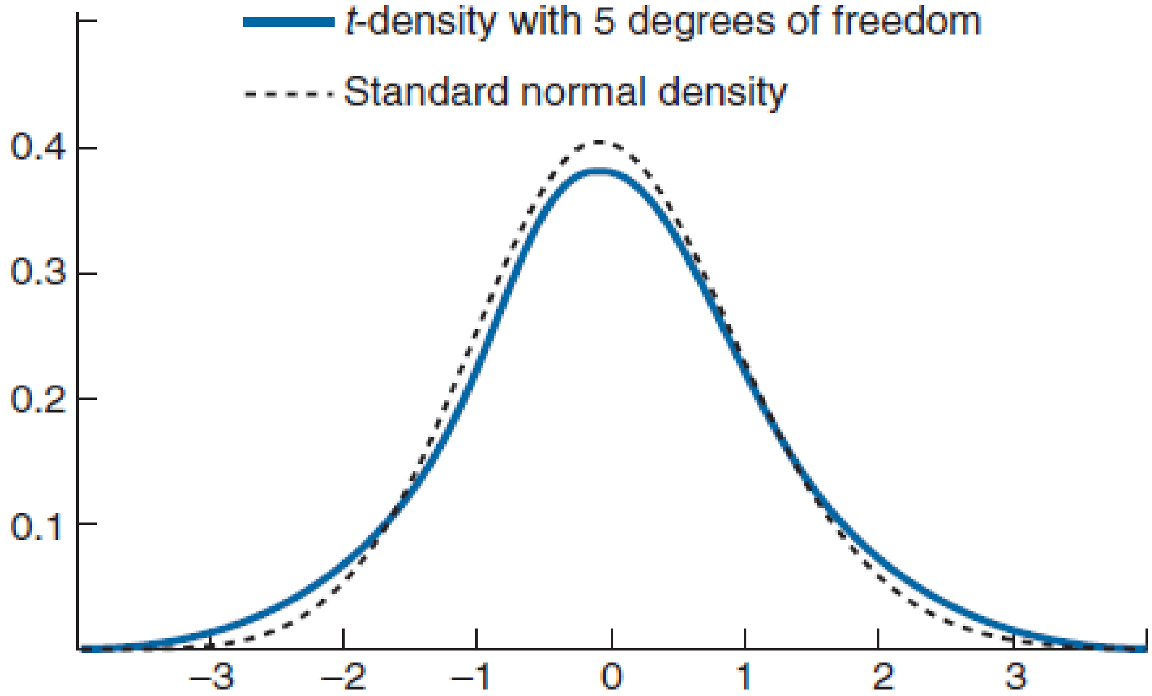
Density function of T_n.

For **large n**,

$$\frac{\chi_n^2}{n} \approx 1$$

Like the standard normal density, the t-density is symmetric about zero. In addition, as n becomes larger, it becomes more and more like a standard normal density.

2. The t-Distribution

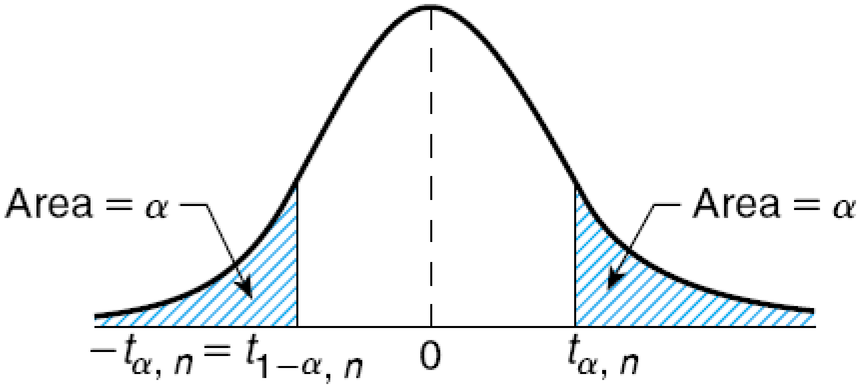


Comparing standard normal density with the density of T_5.

Notice that the t -density has thicker "tails," indicating greater variability, than does the normal density.

2. The t-Distribution

- $P\{T_n \geq t_{\alpha,n}\} = \alpha$
- $\alpha = 1 - P\{T_n > -t_{\alpha,n}\} = P\{T_n < -t_{\alpha,n}\}$



$$t_{1-\alpha,n} = -t_{\alpha,n}.$$

3. The F-Distribution

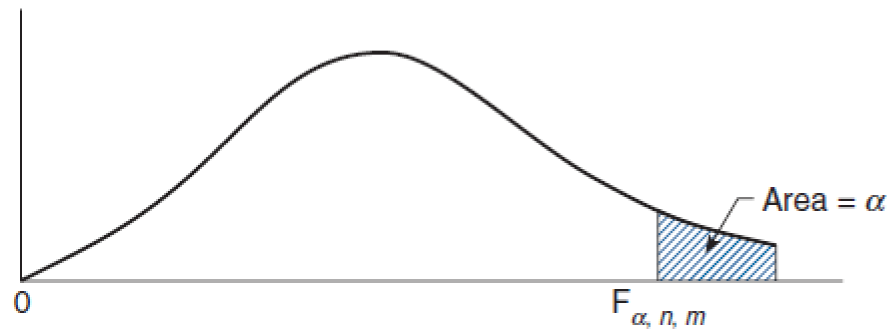
If x_n^2 and x_m^2 are indepedent chi-square random variables with **n** and **m** degrees of freedom, respectively, then the random variable $F_{n,m}$ defined by:

$$F_{n,m} = \frac{\chi_n^2 / n}{\chi_m^2 / m}$$

is said to have an ***F-distribution with n and m degrees of freedom***.

3. The F-Distribution

$$P\left(F_{n,m} > F_{\alpha,n,m}\right) = \alpha$$



Density function of $F_{n,m}$.

$$\frac{1}{F_{\alpha,n,m}} = F_{1-\alpha,m,n}$$

For Example: $F_{0.9,5,7} = \frac{1}{F_{0.1,7,5}}$