ECS315 2012/1 Part IV.1 Dr.Prapun

11 Continuous Random Variables

11.1 From Discrete to Continuous Random Variables

In many practical applications of probability, physical situations are better described by random variables that can take on a *continuum* of possible values rather than a *discrete* number of values. For this type of random variable, the interesting fact is that

- any individual value has probability zero:

$$P[X = x] = 0 \quad \text{for all } x \tag{23}$$

  and that

- the support is always uncountable.

These random variables are called **continuous random variables**.

11.1. We can see from (23) that the pmf is going to be useless for this type of random variable. It turns out that the cdf F_X is still useful and we shall introduce another useful function called *probability density function* (pdf) to replace the role of pmf. However, integral calculus[35] is required to formulate this continuous analog of a pmf.

11.2. In some cases, the random variable X is actually discrete but, because the range of possible values is so large, it might be more convenient to analyze X as a continuous random variable.

[35] This is always a difficult concept for the beginning student.

Example 11.3. Suppose that current measurements are read from a digital instrument that displays the current to the nearest one-hundredth of a mA. Because the possible measurements are limited, the random variable is discrete. However, it might be a more convenient, simple approximation to assume that the current measurements are values of a continuous random variable.

Example 11.4. If you can measure the heights of people with infinite precision, the height of a randomly chosen person is a continuous random variable. In reality, heights cannot be measured with infinite precision, but the mathematical analysis of the distribution of heights of people is greatly simplified when using a mathematical model in which the height of a randomly chosen person is modeled as a continuous random variable. [23, p 284]

Example 11.5. Continuous random variables are important models for

(a) voltages in communication receivers

(b) file download times on the Internet

(c) velocity and position of an airliner on radar

(d) lifetime of a battery

(e) decay time of a radioactive particle

(f) time until the occurrence of the next earthquake in a certain region

Example 11.6. The simplest example of a continuous random variable is the "random choice" of a number from the interval $(0, 1)$.

- In `MATLAB`, this can be generated by the command `rand`. In `Excel`, use `rand()`.
- The generation is "unbiased" in the sense that "any number in the range is as likely to occur as another number."
- Histogram is flat over $(0, 1)$.
- Formally, this is called a uniform RV on the interval $(0, 1)$.

Definition 11.7. We say that X is a **continuous random variable**[36] if we can find a (real-valued) function[37] f such that, for any set B, $P[X \in B]$ has the form

$$P[X \in B] = \int_B f(x)dx. \tag{24}$$

- In particular,

$$P[a \leq X \leq b] = \int_a^b f_(x)dx. \tag{25}$$

  In other words, the **area under the graph** of $f(x)$ between the points a and b gives the probability $P[a \leq X \leq b]$.

- The function f is called the ***probability density function*** (pdf) or simply ***density***.
- When we want to emphasize that the function f is a density of a particular random variable X, we write f_X instead of f.

[36] To be more rigorous, this is the definition for *absolutely* continuous random variable. At this level, we will not distinguish between the continuous random variable and absolutely continuous random variable. When the distinction between them is considered, a random variable X is said to be continuous (not necessarily absolutely continuous) when condition (23) is satisfied. Alternatively, condition (23) is equivalent to requiring the cdf F_X to be continuous. Another fact worth mentioning is that if a random variable is absolutely continuous, then it is continuous. So, absolute continuity is a stronger condition.

[37] Strictly speaking, δ-"function" is not a function; so, can't use δ-function here.

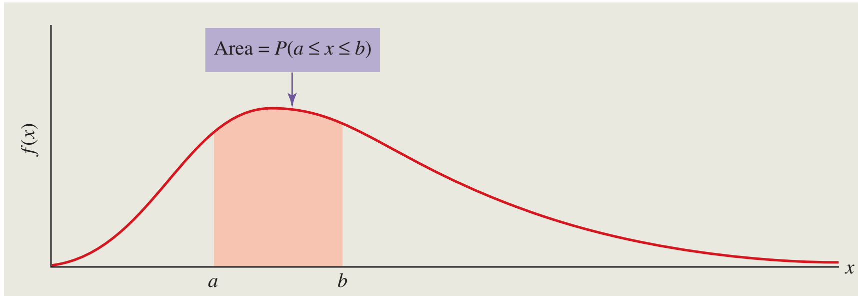

Figure 8: For a continuous random variable, the probability distribution is described by a curve called the probability density function, $f(x)$. The total area beneath the curve is 1.0, and the probability that X will take on some value between a and b is the area beneath the curve between points a and b.

Example 11.8. For the random variable generated by the `rand` command in `MATLAB`[38] or the `rand()` command in `Excel`,

Definition 11.9. Recall that the support S_X of a random variable X is any set S such that $P[X \in S] = 1$. For continuous random variable, S_X is usually set to be $\{x : f_X(x) > 0\}$.

[38]The `rand` command in `MATLAB` is an approximation for two reasons:

(a) It produces pseudorandom numbers; the numbers seem random but are actually the output of a deterministic algorithm.

(b) It produces a double precision floating point number, represented in the computer by 64 bits. Thus MATLAB distinguishes no more than 2^{64} unique double precision floating point numbers. By comparison, there are uncountably infinite real numbers in the interval from 0 to 1.

11.2 Properties of PDF and CDF for Continuous Random Variables

11.10. f_X is determined only almost everywhere[39]. That is, given a pdf f for a random variable X, if we construct a function g by changing the function f at a countable number of points[40], then g can also serve as a pdf for X.

11.11. The cdf of any kind of random variable X is defined as

$$F_X(x) = P\left[X \le x\right].$$

Note that even through there are more than one valid pdfs for any given random variable, the cdf is unique. There is only one cdf for each random variable.

11.12. For continuous random variable, given the pdf $f_X(x)$, we can find the cdf of X by

$$F_X(x) = P\left[X \le x\right] = \int_{-\infty}^{x} f_X(t)dt.$$

11.13. Given the cdf $F_X(x)$, we can find the pdf $f_X(x)$ by

- If F_X is differentiable at x, we will set

$$\frac{d}{dx}F_X(x) = f_X(x).$$

- If F_X is not differentiable at x, we can set the values of $f_X(x)$ to be any value. Usually, the values are selected to give simple expression. (In many cases, they are simply set to 0.)

[39] Lebesgue-a.e, to be exact

[40] More specifically, if $g = f$ Lebesgue-a.e., then g is also a pdf for X.

Example 11.14. For the random variable generated by the `rand` command in `MATLAB` or the `rand()` command in `Excel`,

Example 11.15. Suppose that the lifetime X of a device has the cdf

$$F_X(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{4}x^2, & 0 \le x \le 2 \\ 1, & x > 2 \end{cases}$$

Observe that it is differentiable at each point x except at $x = 2$. The probability density function is obtained by differentiation of the cdf which gives

$$f_X(x) = \begin{cases} \frac{1}{2}x, & 0 < x < 2 \\ 0, & \text{otherwise.} \end{cases}$$

At $x = 2$ where F_X has no derivative, it does not matter what values we give to f_X. Here, we set it to be 0.

11.16. In many situations when you are asked to find pdf, it may be easier to find cdf first and then differentiate it to get pdf.

Exercise 11.17. A point is "picked at random" in the inside of a circular disk with radius r. Let the random variable X denote the distance from the center of the disk to this point. Find $f_X(x)$.

11.18. Unlike the cdf of a discrete random variable, the cdf of a continuous random variable has no jumps and is continuous everywhere.

11.19. $p_X(x) = P[X = x] = P[x \le X \le x] = \int_x^x f_X(t)dt = 0$.

Again, it makes no sense to speak of the probability that X will take on a pre-specified value. This probability is always zero.

11.20. $P[X = a] = P[X = b] = 0$. Hence,

$$P[a < X < b] = P[a \le X < b] = P[a < X \le b] = P[a \le X \le b]$$

- The corresponding integrals over an interval are not affected by whether or not the endpoints are included or excluded.
- When we work with continuous random variables, it is usually not necessary to be precise about specifying whether or not a range of numbers includes the endpoints. This is quite different from the situation we encounter with discrete random variables where it is critical to carefully examine the type of inequality.

11.21. f_X is nonnegative and $\int_{\mathbb{R}} f_X(x)dx = 1$.

Example 11.22. Random variable X has pdf

$$f_X(x) = \begin{cases} ce^{-2x}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

Find the constant c and sketch the pdf.

Definition 11.23. A continuous random variable is called ***exponential*** if its pdf is given by

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & x \le 0 \end{cases}$$

for some $\lambda > 0$

Theorem 11.24. Any nonnegative[41] function that integrates to one is a ***probability density function*** (pdf) of some random variable [9, p.139].

[41]or nonnegative a.e.

11.25. Intuition/Interpretation:

The use of the word "density" originated with the analogy to the distribution of matter in space. In physics, any finite volume, no matter how small, has a positive mass, but there is no mass at a single point. A similar description applies to continuous random variables.

Approximately, for a small Δx,

$$P[X \in [x, x + \Delta x]] = \int_x^{x+\Delta x} f_X(t)dt \approx f_X(x)\Delta x.$$

This is why we call f_X the density function.

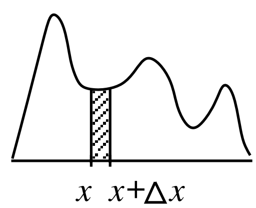

Figure 9: $P[x \leq X \leq x + \Delta x]$ is the area of the shaded vertical strip.

In other words, the probability of random variable X taking on a value in a *small* interval around point c is approximately equal to $f(c)\Delta c$ when Δc is the length of the interval.

- In fact, $f_X(x) = \lim_{\Delta x \to 0} \frac{P[x < X \leq x + \Delta x]}{\Delta x}$
- The number $f_X(x)$ itself is ***not a probability***. In particular, it does not have to be between 0 and 1.
- $f_X(c)$ is a relative measure for the likelihood that random variable X will take on a value in the immediate neighborhood of point c.

  Stated differently, the pdf $f_X(x)$ expresses how densely the probability mass of random variable X is smeared out in the neighborhood of point x. Hence, the name of density function.

11.26. Histogram and pdf [23, p 143 and 145]:

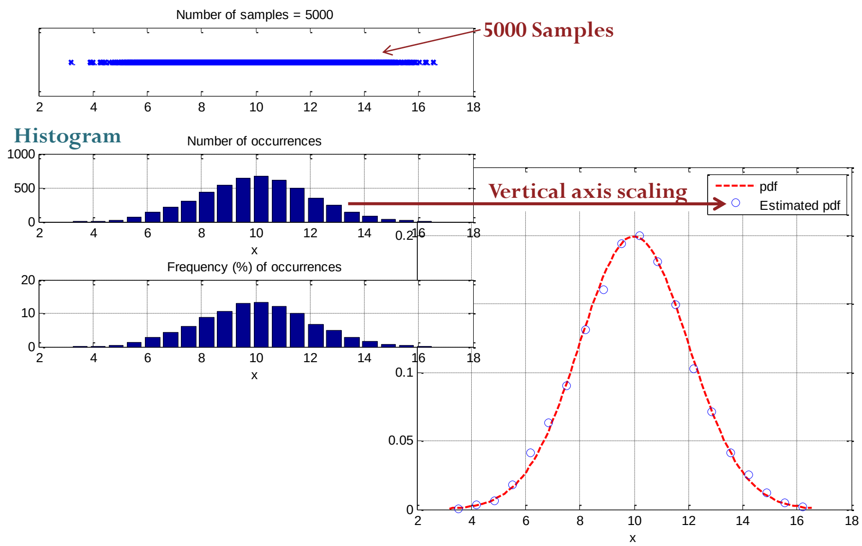

Figure 10: From histogram to pdf.

(a) A (probability) **histogram** is a bar chart that divides the range of values covered by the samples/measurements into intervals of the same width, and shows the proportion (relative frequency) of the samples in each interval.

- To make a histogram, you break up the range of values covered by the samples into a number of disjoint adjacent intervals each having the same width, say width Δ. The height of the bar on each interval $[j\Delta, (j+1)\Delta)$ is taken such that the area of the bar is equal to the proportion of the measurements falling in that interval (the proportion of measurements within the interval is divided by the width of the interval to obtain the height of the bar).
- The total area under the histogram is thus standardized/normalized to one.

(b) If you take sufficiently many independent samples from a continuous random variable and make the width Δ of the base intervals of the probability histogram smaller and smaller, the graph of the histogram will begin to look more and more like the pdf.

(c) Conclusion: A probability density function can be seen as a "smoothed out" version of a probability histogram

11.3 Expectation and Variance

11.27. ***Expectation***: Suppose X is a continuous random variable with probability density function $f_X(x)$.

$$\mathbb{E}X = \int_{-\infty}^{\infty} x f_X(x) dx \tag{26}$$

$$\mathbb{E}\left[g(X)\right] = \int_{-\infty}^{\infty} g(x) f_X(x) dx \tag{27}$$

In particular,

$$\mathbb{E}\left[X^2\right] = \int_{-\infty}^{\infty} x^2 f_X(x) dx$$

$$\operatorname{Var} X = \int_{-\infty}^{\infty} (x - \mathbb{E}X)^2 f_X(x) dx = \mathbb{E}\left[X^2\right] - (\mathbb{E}X)^2.$$

Example 11.28. For the random variable generated by the `rand` command in `MATLAB` or the `rand()` command in `Excel`,

Example 11.29. For the exponential random variable introduced in Definition 11.23,

11.30. If we compare other characteristics of discrete and continuous random variables, we find that with discrete random variables, many facts are expressed as sums. With continuous random variables, the corresponding facts are expressed as integrals.

11.31. All of the properties for the expectation and variance of discrete random variables also work for continuous random variables as well:

(a) Intuition/interpretation of the expected value: As $n \to \infty$, the average of n independent samples of X will approach $\mathbb{E}X$. This observation is known as the "Law of Large Numbers".

(b) For $c \in \mathbb{R}$, $\mathbb{E}[c] = c$

(c) For constants a, b, we have $\mathbb{E}[aX + b] = a\mathbb{E}X + b$.

(d) $\mathbb{E}\left[\sum_{i=1}^{n} c_i g_i(X)\right] = \sum_{i=1}^{n} c_i \mathbb{E}[g_i(X)]$.

(e) $\operatorname{Var} X = \mathbb{E}\left[X^2\right] - (\mathbb{E}X)^2$

(f) $\operatorname{Var} X \geq 0$.

(g) $\operatorname{Var} X \leq \mathbb{E}\left[X^2\right]$.

(h) $\operatorname{Var}[aX + b] = a^2 \operatorname{Var} X$.

(i) $\sigma_{aX+b} = |a|\, \sigma_X$.

11.32. ***Chebyshev's Inequality***:

$$P\left[|X - \mathbb{E}X| \geq \alpha\right] \leq \frac{\sigma_X^2}{\alpha^2}$$

or equivalently

$$P\left[|X - \mathbb{E}X| \geq n\sigma_X\right] \leq \frac{1}{n^2}$$

- This inequality use variance to bound the "tail probability" of a random variable.
- Useful only when $\alpha > \sigma_X$

Example 11.33. A circuit is designed to handle a current of 20 mA plus or minus a deviation of less than 5 mA. If the applied current has mean 20 mA and variance 4 mA^2, use the Chebyshev inequality to bound the probability that the applied current violates the design parameters.

Let X denote the applied current. Then X is within the design parameters if and only if $|X - 20| < 5$. To bound the probability that this does not happen, write

$$P\left[|X - 20| < 5\right] \leq \frac{\operatorname{Var} X}{5^2} = \frac{4}{25} = 0.16.$$

Hence, the probability of violating the design parameters is at most 16%.

11.34. Interesting applications of expectation:

(a) $f_X(x) = \mathbb{E}\left[\delta\left(X - x\right)\right]$

(b) $P\left[X \in B\right] = \mathbb{E}\left[1_B(X)\right]$

11.4 Families of Continuous Random Variables

Theorem 11.24 states that any nonnegative function $f(x)$ whose integral over the interval $(-\infty, +\infty)$ equals 1 can be regarded as a probability density function of a random variable. In real-world applications, however, special mathematical forms naturally show up. In this section, we introduce a couple families of continuous random variables that frequently appear in practical applications. The probability densities of the members of each family all have the same mathematical form but differ only in one or more parameters.

11.4.1 Uniform Distribution

Definition 11.35. For a uniform random variable on an interval $[a, b]$, we denote its family by uniform$([a, b])$ or $\mathcal{U}([a, b])$. Expressions that are synonymous with "X is a uniform random variable" are "X is uniformly distributed" and "X has a uniform distribution". This family is characterized by

$$f_X(x) = \begin{cases} 0, & x < a, x > b \\ \frac{1}{b-a}, & a \le x \le b \end{cases}$$

- The random variable X is just as likely to be near any value in $[a, b]$ as any other value.
- In MATLAB, use `X = a+(b-a)*rand`.

Exercise 11.36. Show that $F_X(x) = \begin{cases} 0, & x < a, x > b \\ \frac{x-a}{b-a}, & a \le x \le b \end{cases}$

Example 11.37 (F2011)**.** Suppose X is uniformly distributed on the interval $(1, 2)$. ($X \sim \mathcal{U}(1, 2)$.)

(a) Plot the pdf $f_X(x)$ of X.

(b) Plot the cdf $F_X(x)$ of X.

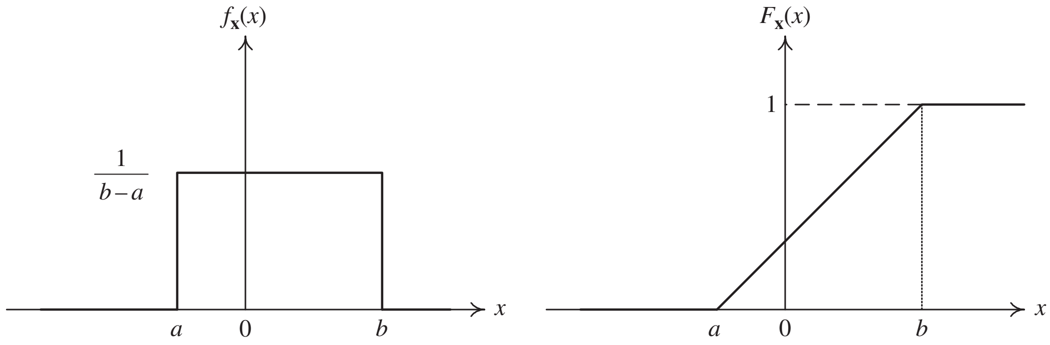

Figure 11: The pdf and cdf for the uniform random variable. [18, Fig. 3.5]

11.38. The uniform distribution provides a probability model for selecting a point at random from the interval $[a, b]$.

- Use with caution to model a quantity that is known to vary randomly between a and b but about which little else is known.

Example 11.39. [9, Ex. 4.1 p. 140-141] In coherent radio communications, the phase difference between the transmitter and the receiver, denoted by Θ, is modeled as having a uniform density on $[-\pi, \pi]$.

(a) $P[\Theta \leq 0] = \frac{1}{2}$

(b) $P\left[\Theta \leq \frac{\pi}{2}\right] = \frac{3}{4}$

Exercise 11.40. Show that $\mathbb{E}X = \frac{a+b}{2}$, $\operatorname{Var} X = \frac{(b-a)^2}{12}$, and $\mathbb{E}\left[X^2\right] = \frac{1}{3}\left(b^2 + ab + a^2\right)$.

11.4.2 Gaussian Distribution

11.41. This is the most widely used model for the distribution of a random variable. When you have many independent random variables, a fundamental result called the central limit theorem (CLT) (informally) says that the sum (technically, the average) of them can often be approximated by normal distribution.

Definition 11.42. ***Gaussian*** random variables:

(a) Often called ***normal*** random variables because they occur so frequently in practice

(b) In `MATLAB`, use `X = σ*randn + m`.

(c) $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$.

- In `Excel`, use `NORMDIST(x,m,σ,FALSE)`.
  In `MATLAB`, use `normpdf(x,m,σ)`.
- Figure 12 displays the famous ***bell-shaped*** graph of the Gaussian pdf. This curve is also called the *normal* curve.

(d) $F_X(x) =$ `normcdf(x,m,σ)` in `MATLAB`

- In `Excel`, use `NORMDIST(x,m,σ,TRUE)`.

(e) We write $X \sim \mathcal{N}\left(m, \sigma^2\right)$.

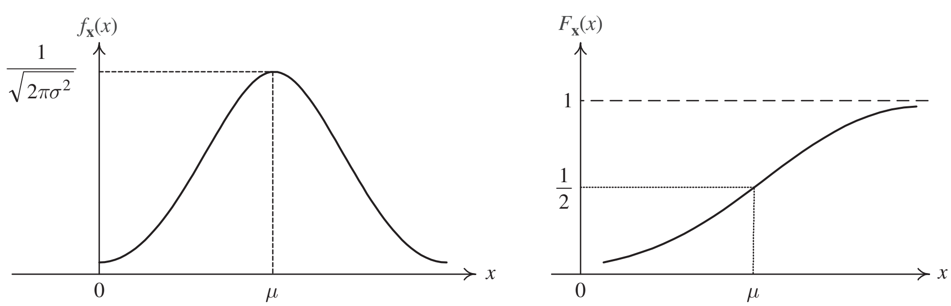

Figure 12: The pdf and cdf of $\mathcal{N}(\mu, \sigma^2)$. [18, Fig. 3.6]

11.43. $\mathbb{E}X = m$ and $\operatorname{Var} X = \sigma^2$.

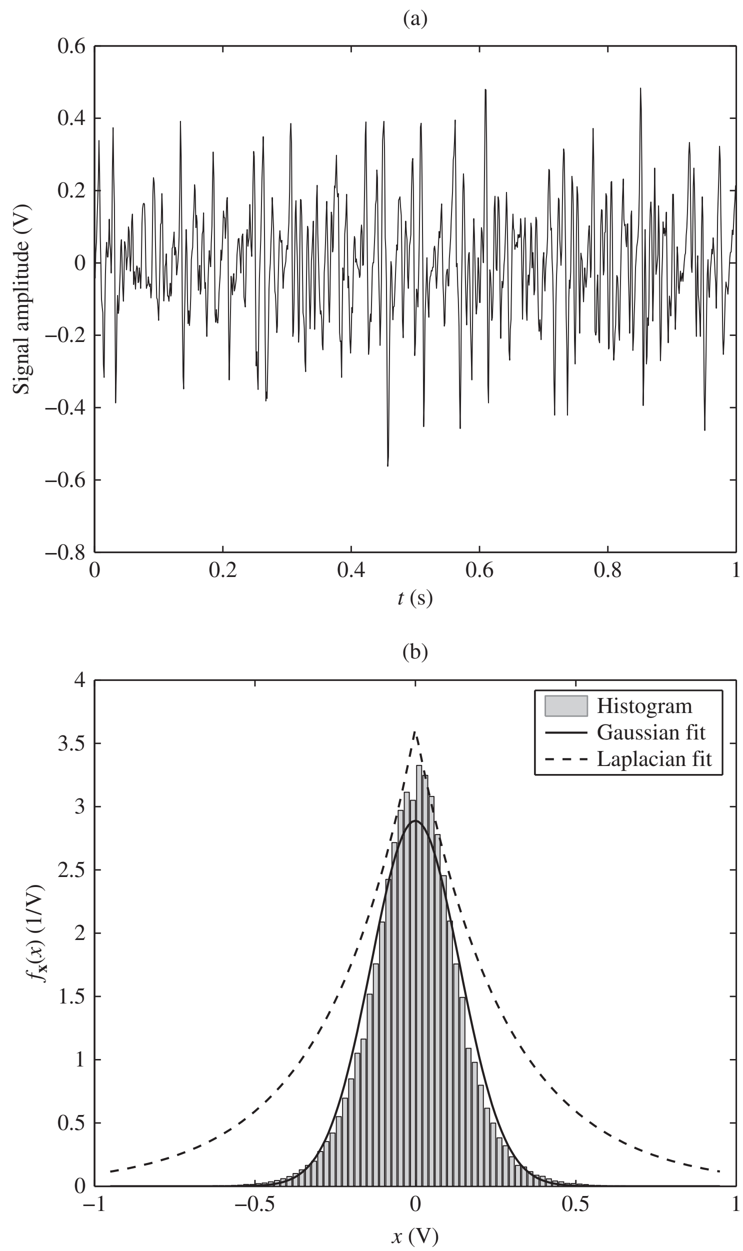

Figure 13: Electrical activity of a skeletal muscle: (a) A sample skeletal muscle (emg) signal, and (b) its histogram and pdf fits. [18, Fig. 3.14]

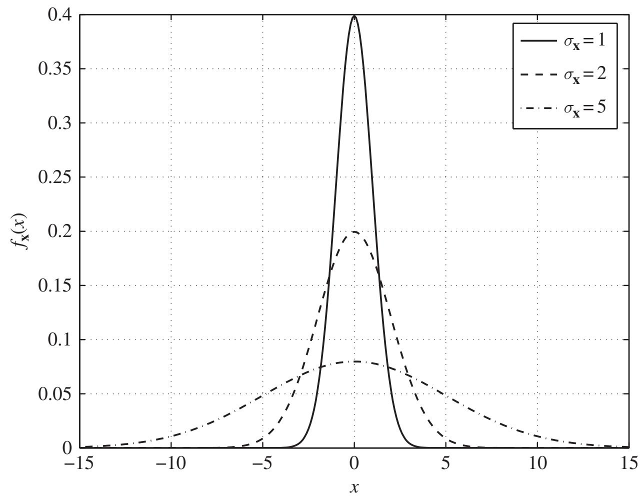

Figure 14: Plots of the zero-mean Gaussian pdf for different values of standard deviation, σ_X. [18, Fig. 3.15]

11.44. Important probabilities:
$P[|X-\mu|<\sigma]=0.6827$;
$P[|X-\mu|>\sigma]=0.3173$;
$P[|X-\mu|>2\sigma]=0.0455$;
$P[|X-\mu|<2\sigma]=0.9545$

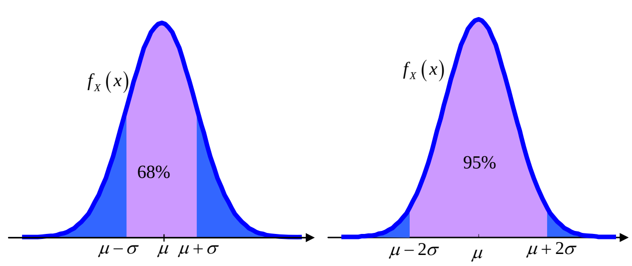

Figure 15: Probability density function of $X \sim \mathcal{N}(\mu, \sigma^2)$.

11.45. $\mathcal{N}(0,1)$ is the ***standard*** Gaussian (normal) distribution.

- In `Excel`, use `NORMSINV(RAND())`.
  In `MATLAB`, use `randn`.
- The standard normal cdf is denoted by $\Phi(z)$. It inherits all properties of cdf. Moreover, note that $\Phi(-z)=1-\Phi(z)$.

11.46. Relationship between $\mathcal{N}(0,1)$ and $\mathcal{N}(m,\sigma^2)$.

(a) An arbitrary Gaussian random variable with mean m and variance σ^2 can be represented as $\sigma Z+m$, where $Z \sim \mathcal{N}(0,1)$.

This relationship can be used to generate general Gaussian RV from standard Gaussian RV.

(b) If $X \sim \mathcal{N}\left(m,\sigma^2\right)$, the random variable

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

is a standard normal random variable. That is, $Z \sim \mathcal{N}(0,1)$.

- Creating a new random variable by this transformation is referred to as ***standardizing***.
- The standardized variable is called "***standard score***" or "***z-score***".
- It is the key step to calculating a probability for an arbitrary normal random variable. Using this relationship, we can show that
    - The CDF of X is

$$F_X(x) = \Phi\left(\frac{x-m}{\sigma}\right).$$

11.47. It is impossible to express the integral of a Gaussian PDF between non-infinite limits (e.g., (25)) as a function that appears on most scientific calculators.

- An old but still popular technique to find integrals of the Gaussian PDF is to refer to tables that have been obtained by numerical integration.
    - One such table is the table that lists $\Phi(z)$ for many values of positive z.

      Note that the table is expressed in "standardized" form.

Example 11.48. Figure 16 compares several deviation scores and the normal distribution:

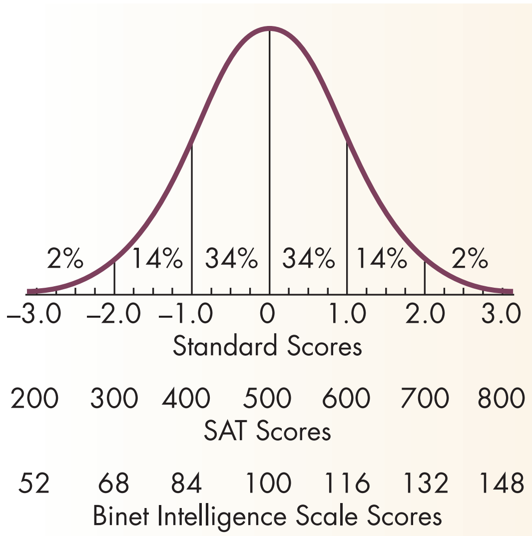

Figure 16: Comparison of Several Deviation Scores and the Normal Distribution

(a) Standard scores have a mean of zero and a standard deviation of 1.0.

(b) Scholastic Aptitude Test scores have a mean of 500 and a standard deviation of 100.

(c) Binet Intelligence Scale[42] scores have a mean of 100 and a standard deviation of 16.

[42]Alfred Binet, who devised the first general aptitude test at the beginning of the 20th century, defined intelligence as the ability to make adaptations. The general purpose of the test was to determine which children in Paris could benefit from school. Binets test, like its subsequent revisions, consists of a series of progressively more difficult tasks that children of different ages can successfully complete. A child who can solve problems typically solved by children at a particular age level is said to have that mental age. For example, if a child can successfully do the same tasks that an average 8-year-old can do, he or she is said to have a mental age of 8. The intelligence quotient, or IQ, is defined by the formula:

$$\text{IQ} = 100 \times (\text{Mental Age}/\text{Chronological Age})$$

There has been a great deal of controversy in recent years over what intelligence tests measure. Many of the test items depend on either language or other specific cultural experiences for correct answers. Nevertheless, such tests can rather effectively predict school success. If school requires language and the tests measure language ability at a particular point of time in a childs life, then the test is a better-than-chance predictor of school performance.

In each case there are 34 percent of the scores between the mean and one standard deviation, 14 percent between one and two standard deviations, and 2 percent beyond two standard deviations. [Source: Beck, Applying Psychology: Critical and Creative Thinking.]

11.49. ***Q-function***: $Q(z) = \int\limits_z^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$ corresponds to $P[X > z]$ where $X \sim \mathcal{N}(0,1)$; that is $Q(z)$ is the probability of the "tail" of $\mathcal{N}(0,1)$. The Q function is then a complementary cdf (ccdf).

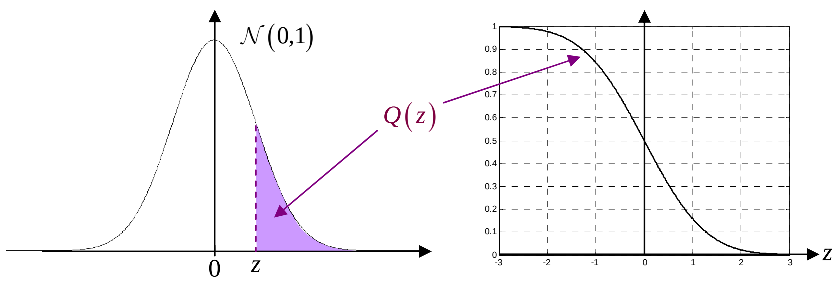

Figure 17: Q-function

(a) Q is a decreasing function with $Q(0) = \frac{1}{2}$.

(b) $Q(-z) = 1 - Q(z) = \Phi(z)$

(c) $Q^{-1}(1 - Q(z)) = -z$

(d) Approximation:

(i) $Q(z) \approx \left[\frac{1}{(1-a)z + a\sqrt{z^2+b}}\right] \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}; \quad a = \frac{1}{\pi}, b = 2\pi$

(ii) $\left(1 - \frac{1}{x^2}\right) \frac{e^{-\frac{x^2}{2}}}{x\sqrt{2\pi}} \le Q(x) \le \frac{1}{2} e^{-\frac{x^2}{2}}$

(iii) $Q(z) \approx \frac{1}{z\sqrt{2\pi}} \left(1 - \frac{0.7}{z^2}\right) e^{-\frac{z^2}{2}}; z > 2$

11.4.3 Exponential Distribution

Definition 11.50. The exponential distribution is denoted by $\mathcal{E}(\lambda)$.

(a) $\lambda > 0$ is a parameter of the distribution, often called the **rate parameter**.

(b) Characterized by

- $f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & x \le 0 \end{cases}$
- $F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0, \\ 0, & x \le 0 \end{cases}$

Exercise 11.51. Exponential random variable as a continuous version of geometric random variable: Show that $\lfloor X \rfloor \sim \mathcal{G}_0(e^{-\lambda})$ and $\lceil X \rceil \sim \mathcal{G}_1(e^{-\lambda})$

Example 11.52. The exponential distribution is intimately related to the Poisson process. It is often used as a probability model for the (waiting) time until a "rare" event occurs.

- time elapsed until the next earthquake in a certain region
- decay time of a radioactive particle
- time between independent events such as arrivals at a service facility or arrivals of customers in a shop.
- duration of a cell-phone call
- time it takes a computer network to transmit a message from one node to another.

Example 11.53. Phone Company A charges \$0.15 per minute for telephone calls. For any fraction of a minute at the end of a call, they charge for a full minute. Phone Company B also charges \$0.15 per minute. However, Phone Company B calculates its charge based on the exact duration of a call. If T, the duration of a call in minutes, is exponential with parameter $\lambda = 1/3$, what are the expected revenues per call $\mathbb{E}[R_A]$ and $\mathbb{E}[R_B]$ for companies A and B?

Solution: First, note that $\mathbb{E}T = \frac{1}{\lambda} = 3$. Hence,

$$\mathbb{E}[R_B] = \mathbb{E}[0.15 \times T] = 0.15\mathbb{E}T = \$0.45.$$

and

$$\mathbb{E}[R_A] = \mathbb{E}[0.15 \times \lceil T \rceil] = 0.15\mathbb{E}\lceil T \rceil.$$

Now, recall that $\lceil T \rceil \sim \mathcal{G}_1\left(e^{-\lambda}\right)$. Hence, $\mathbb{E}\lceil T \rceil = \frac{1}{1-e^{-\lambda}} \approx 3.53$. Therefore,

$$\mathbb{E}[R_A] = 0.15\mathbb{E}\lceil T \rceil \approx 0.5292.$$

11.54. ***Memoryless property***: The exponential r.v. is the only continuous[43] r.v. on $[0,\infty)$ that satisfies the memoryless property:

$$P[X > s + x \,|X > s] = P[X > x]$$

for all $x > 0$ and all $s > 0$ [20, p. 157–159]. In words, the future is independent of the past. The fact that it hasn't happened yet, tells us nothing about how much longer it will take before it does happen.

- Imagining that the exponentially distributed random variable X represents the lifetime of an item, the residual life of an item has the same exponential distribution as the original lifetime, regardless of how long the item has been already in use. In other words, there is no deterioration/degradation over time. If it is still currently working after 20 years of use, then today, its condition is "just like new".
- In particular, suppose we define the set $B+x$ to be $\{x + b : b \in B\}$. For any $x > 0$ and set $B \subset [0,\infty)$, we have

  $$P[X \in B + x|X > x] = P[X \in B]$$

  because

  $$\frac{P[X \in B + x]}{P[X > x]} = \frac{\int_{B+x} \lambda e^{-\lambda t} dt}{e^{-\lambda x}} \overset{\tau = t - x}{=} \frac{\int_B \lambda e^{-\lambda(\tau + x)} d\tau}{e^{-\lambda x}}.$$

[43] For discrete random variable, geometric random variables satisfy the memoryless property.