Continuous Random Variables

Reuy-Lin Sheu

Department of Mathematics, National Cheng Kung University,
Tainan, Taiwan

October, 2023

Recall the Definition of a discrete random variable

- Let $(\Omega, \mathcal{F}, P)$ be a probability space that corresponds to a random experiment and suppose X is a real-valued function from $(\Omega, \mathcal{F}, P)$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.
- We say that X is a discrete random variable if X ONLY takes "*finitely or countably infinite*" many values x_i on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ so that it has a discrete distribution (mass) function

$$p_X(x_i) = P(X = x_i).$$

- A discrete random variable is often a count. For example, count the number of heads in n tosses (Bernoulli(n, p)); count the number of occurrences over a time interval (Poisson(λ)); or count the number of tosses before the first head comes up (Geometric).

Definition of a continuous random variable

- If the image of X is an uncountable set on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (usually an interval $[a, b]$ on $\mathbb{R}$ or the entire $\mathbb{R}$), then X is called a continuous random variable.

- A continuous random variable X is often a measurement. For example, X denotes the measurement of the length of a bar; or X is the length of time before the first occurrence if it occurs according to a Poisson distribution.

- The most important special case of a continuous random variable is the so-called "absolute continuous" random variable which assigns the probability of a Borel set by a probability density function and which must assign the probability of a singleton set to the value 0.

(Review) Induced measure on Borel sets by a random variable

- Let $(\Omega, \mathcal{F}, P)$ be probability space and $X: \ (\Omega, \mathcal{F}, P) \ \mapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a r.v. on Ω, discrete or continuous.
- We can utilize P on $\mathcal{F}$ to define an induced measure $\mathcal{L}_X$ by X on any Borel set $B \in \mathcal{B}(\mathbb{R})$ by

$$\mathcal{L}_X(B) = P(X^{-1}(B)) = P(X \in B).$$

- If X is a discrete r.v. taking values $x_i, \ i = 1, 2, \ldots$, since each singleton set $\{x_i\}$ is Borel, the induced measure $\mathcal{L}_X$ on each x_i : (the "*probability mass function*" of X,)

$$\begin{aligned} p_X(x_i) &\triangleq \mathcal{L}_X(\{x_i\}) = P(X^{-1}(\{x_i\})) \\ &= P(X = x_i) = P(\omega \in \Omega : X(\omega) = x_i). \end{aligned}$$

- For a continuous r.v. $X \in [a, b]$, however, the more important thing on each $x \in [a, b]$ is the "probability density" at x.

Definition of an absolutely continuous random variable taking values on the entire $\mathbb{R}$ (page 58 in the textbook)

- Let X be an absolutely continuous r.v. defined on a probability space $(\Omega, \mathcal{F}, P)$ which assigns Ω to the entire $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.
- There must exist a probability density function

$$f_X : \mathbb{R} \longrightarrow \mathbb{R}_+$$

so that, for any Borel set $I \in \mathcal{B}(\mathbb{R})$, the induced measure $\mathcal{L}_X$ on I is computed through the integration of f_X over I. That is,

$$\mathcal{L}_X(I) = P(X^{-1}(I)) = P(X \in I) = P(\omega \in \Omega : \ X(\omega) \in I) = \int_I f_X(s)ds.$$

- In case that $I = [a, b]$, we have

$$P(X \in [a, b]) = P(a \leq X \leq b) = \int_{[a,b]} f_X(s)ds = \int_a^b f_X(s)ds.$$

- In case that $I = \{a\} = [a, a]$,

$$P(X \in \{a\}) = P(X = a) = \int_{[a,a]} f_X(s)ds = \int_a^a f_X(s)ds = 0.$$

Absolutely Continuous Random Variable

- In case that $I = (-\infty, x]$, we have the cumulative distribution function of X as

$$F_X(x) = P(X \in (-\infty, x]) = P(X \leq x) = \int_{-\infty}^{x} f_X(s) ds. \qquad (1)$$

- For $I = (-\infty, \infty)$,

$$P(X \in (-\infty, \infty)) = P(\omega \in \Omega) = \int_{-\infty}^{\infty} f_X(s) ds = 1.$$

- Since $P(X = a) = 0$ for an absolute continuous random variable X, we have

$$P(X < a) = P(X \leq a) = \int_{-\infty}^{a} f_X(s) ds.$$

Absolutely Continuous Random Variable

- Let X be an absolutely Continuous Random Variable with density $f_X(x)$ and the cumulative distribution function $F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(s)ds.$
- By Fundamental Theorem of Calculus, on an open interval where f is continuous, $F_X^{'}(x) = f_X(x)$ for x in that open interval.
- The reason we say that $f(x)$ is the probability "density" (rate of change of probability at a particular $x \in \mathbb{R}$ with respect to the unit Borel length) is because

$$f_X(x) = F_X^{'}(x) = \lim_{h \to 0} \frac{F_X(x+h) - F_X(x)}{h} = \lim_{h \to 0} \frac{P(X \in [x, x+h])}{h}$$

- By the differential form, we also have the Leibniz notation connecting the cumulative distribution function $F_X(x)$ and the density function $f_X(x)$ of X by

$$dF_X = dF_X(x, dx) = F_X'(x) \cdot dx = f_X(x) \cdot dx.$$

Absolutely Continuous Random Variable

- Not every continuous r.v. is absolutely continuous.
- A continuous random variable X could be "singular." That is, $X'(x) = 0, a.e.$. For example, the Cantor-Lebesgue function. We are not going to discuss singular r.v.'s in this course.
- For an absolutely continuous random variable X taking values on $[a, b]$ with the density $f_X(x)$, we can define its expectation as (where a could be $-\infty$, and b could be ∞ in which case the improper integral is used.):

$$E(X) = \int_a^b x \cdot f_X(x)dx = \int_a^b x \cdot dF_X$$

- By partition $[a, b]$ into $a = x_0 < x_1 < x_2 < \cdots < x_n = b$, the expectation of X can be approximated by Riemann Sum:

$$\begin{aligned}
& x_1 \cdot P(X \in [x_0, x_1]) + x_2 \cdot P(X \in [x_1, x_2]) + \cdots + x_n \cdot P(X \in [x_{n-1}, x_n]) \\
= \quad & x_1 \cdot \frac{P(X \in [x_0, x_1])}{\Delta x_1} \Delta x_1 + x_2 \cdot \frac{P(X \in [x_1, x_2])}{\Delta x_2} \Delta x_2 + \cdots + x_n \cdot \frac{P(X \in [x_{n-1}, x_n])}{\Delta x_n} \Delta x_n \\
\xrightarrow{\max \|\Delta x_i\| \to 0} \quad & \int_a^b x \cdot f_X(x)dx.
\end{aligned}$$

Expectation and Variance of an absolutely continuous random variable

- For a discrete r.v., we have

$$E(X) = \underbrace{\sum_{\omega \in \Omega} X(\omega) P(\omega)}_{\text{sum over sample space}} = \underbrace{\sum_{i=1}^{\infty} x_i p_X(x_i)}_{\text{sum over foreground space}}$$

- and X is a discrete r.v.,

$$E(g(X)) = \underbrace{\sum_{\omega \in \Omega} g(X)(\omega) P(\omega)}_{\text{sum over sample space}} = \underbrace{\sum_{i=1}^{\infty} g(x_i) p_X(x_i)}_{\text{sum over foreground space}}$$

- For a continuous r,v, X, the formula for $E(g(X))$ can be proved to be the integration of $Y = g(X)$ w.r.t. the distribution function of X as

$$E(g(X)) = \int_{g(a)}^{g(b)} y \cdot dF_Y(y) = \int_a^b g(x) \cdot dF_X(x) = \int_a^b g(x) \cdot f_X(x) dx$$

- Variance of X is computed by the same formula:

$$\mathrm{Var}(X) = E(X - \mu_X)^2 = E(X^2) - \mu_X^2 = E(X^2) - (EX)^2.$$

Example 6.1 (page 59 in the textbook)

- Let X be a (absolutely) continuous random variable with the following density function:

$$f_X(x) = \begin{cases} cx, & \text{if } x \in (0,4); \\ 0, & \text{otherwise.} \end{cases}$$

- By the fact that

$$P(-\infty < X < \infty) = P(\Omega) = 1 = \int_0^4 f_X(x) \cdot dx = \int_0^4 cx \cdot dx = c\frac{4^2}{2} = 8c,$$

  it implies that $c = \frac{1}{8}$ and the density of X is $f_X(x) = \frac{x}{8},\ x \in (0,4)$.

- The probability $P(X \in [1,2]) = \int_1^2 \frac{x}{8} dx = \frac{3}{16}$.
- The expectation of X is $E(X) = \int_0^4 x \cdot f_X(x) dx = \int_0^4 \frac{x^2}{8} dx = \frac{8}{3}$
- To compute the variance of X, we first do

$$E(X^2) = \int_0^4 x^2 \cdot f_X(x) dx = \int_0^4 \frac{x^3}{8} dx = 8.$$

  Then, $\mathrm{Var}(X) = E(X^2) - (EX)^2 = 8 - \frac{8^2}{3^2} = \frac{8}{9}$.

Example 6.2 (page 59 in the textbook)

- Let X be an (absolutely) continuous random variable with the following density function:

$$f_X(x) = \begin{cases} 3x^2, & \text{if } x \in [0,1]; \\ 0, & \text{otherwise.} \end{cases}$$

  Compute the density f_Y for $Y = 1 - X^4$.

- We first determine the range of Y to be also between 0 and 1. Then, we compute the accumulative distribution function $F_Y(y)$ for $y \in [0,1]$. That is,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = \int_{-\infty}^{y} f_Y(x)dx \\ &= P(1 - X^4 \leq y) = P(\sqrt[4]{1-y} \leq X) \\ &= \int_{\sqrt[4]{1-y}}^{1} 3x^2 dx. \end{aligned}$$

- Then,

$$f_Y(y) = \frac{d}{dy}F_Y(y) = -3\left(\sqrt[4]{1-y}\right)^2 \frac{1}{4}(1-y)^{-\frac{3}{4}}(-1) = \frac{3}{4\sqrt[4]{1-y}}.$$

Law of the Unconscious Statistician

- Example 6.2 above can be generalized to prove a special case of Law of the Unconscious Statistician.
- Let X be an (absolutely) continuous r.v. defined on defined on a probability space $(\Omega, \mathcal{F}, P)$ with the range set $\mathcal{X}$ having the density function $f_X(x)$ defined on $\mathcal{X}$; and $g : \mathbb{R} \to \mathbb{R}$ is a Borel function so that $Y = g(X)$ is a r.v. also with the range set $\mathcal{Y}$.
- Then, Law of the Unconscious Statistician says that

$$\begin{aligned} E(Y) &= \int_{\mathcal{Y}} y \cdot f_Y(y) \cdot dy \\ &= E(g(X)) = < g(x), f_X(x) > = \int_{\mathcal{X}} g(x) \cdot f_X(x) \cdot dx. \end{aligned}$$

- Here, we only prove for a special case that g is a differentiable monotonic function so that g^{-1} exists and is also monotone. Then, a general Borel function can be approximated a.e. by a sequence of monotonic increase functions.

Law of the Unconscious Statistician

- Since $y = g(x)$ is assumed to be monotonic and differentiable, its inverse function $x = g^{-1}(y)$ exists and also differentiable. In fact, the differential form gives $dx = \frac{d}{dy} g^{-1}(y) \cdot dy$.
- On the other hand, we have

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(g(X) \leq y) \\ &= P(X \leq g^{-1}(y)) = F_X(g^{-1}(y)) \end{aligned}$$

- Then,

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X(g^{-1}(y)) = f_X(g^{-1}(y)) \cdot \frac{d}{dy} g^{-1}(y)$$

- Therefore, by change of variable $y = g(x)$ in integration,

$$\begin{aligned} E(Y) = \int_{\mathcal{Y}} y \cdot f_Y(y) \cdot dy &= \int_{\mathcal{Y}} y \cdot f_X(g^{-1}(y)) \cdot \frac{d}{dy} g^{-1}(y) \cdot dy \\ &= \int_{\mathcal{X}} g(x) \cdot f_X(x) \cdot dx. \end{aligned}$$

Uniform Random Variable (page 60 in the textbook)

- A continuous random variable X on a probability space $(\Omega, \mathcal{F}, P)$ is called a uniform random variable, denoted by $X \sim \mathrm{Unif}[\alpha, \beta]$, if X defined on $(\Omega, \mathcal{F}, P)$ takes values on $[\alpha, \beta], \alpha < \beta \in \mathbb{R}$ with the following density

$$f_X(x) = \begin{cases} \frac{1}{\beta-\alpha}, & \text{if } x \in [\alpha, \beta]; \\ 0, & \text{otherwise.} \end{cases}$$

- That is, the probability density is a constant for all points on $[\alpha, \beta]$, thus the name "uniform."
- Suppose $X \sim \mathrm{Unif}[\alpha, \beta]$. For $\alpha < a < b < \beta$, the probability for the event that X takes some value on $[a, b]$ to happen, is the portion of length of $[a, b]$ in terms of the entire $[\alpha, \beta]$.

$$P(a \le X \le b) = \int_a^b f_X(x)dx = \frac{1}{\beta-\alpha}\int_a^b dx = \frac{b-a}{\beta-\alpha}.$$

- For example, if X is the time at which an event occurred and $X \sim \mathrm{Unif}[\alpha, \beta]$. Then, each interval in $[\alpha, \beta]$ of equal length should have the same probability of containing the event.
- The expectation $EX = \frac{1}{\beta-\alpha}\int_\alpha^\beta x \cdot dx = \frac{\alpha+\beta}{2}$. The variance $\mathrm{Var}(X) = \frac{(\beta-\alpha)^2}{12}$.

Uniform Random Variable (page 60 in the textbook)

- For $[\alpha, \beta] = [0,1]$, $f_X(x) = 1, \forall x \in [0,1]$. In this case, X models an ideal random number generator on a computer[1].
- Assume that $X \sim \mathrm{Unif}[0,1)$. The probability for X to take a value in $\mathbb{Q}$ (let $\{q_1, q_2, \ldots, q_n, \ldots\} \subset [0,1)$ be an enumeration of $\mathbb{Q}$) is $P(X \in \mathbb{Q}) = P(\cup_i \{X = q_i\}) = \sum_{i=1}^{n} P(X = q_i) = 0.$[2]
- Each point $x \in [0,1)$ has the binary expression

$$x = (0.x_1x_2x_3\cdots)_2 = \sum_{i=1}^{\infty} \frac{x_i}{2^i}, \;\; x_i = 0 \vee 1.$$

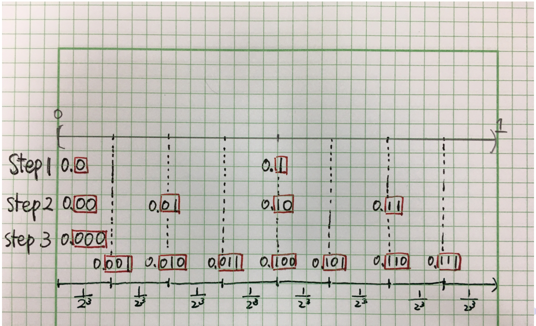
- The set $\{x = (0.0x_2x_3\cdots)_2\}$ has the smallest number 0, the largest one $(0.1)_2 = 0.5$.
- The set $\{x = (0.1x_2x_3\cdots)_2\}$ has the smallest number $(0.1)_2 = 0.5$, but no largest one because $(0.11111\ldots)_2 = 1 \notin [0,1)$.

[1] The existing random number generator on a computer is "far from random" though!

[2] This can be interpreted as the "length" of $\mathbb{Q}$ accounts for 0% of the total length 1 of $[0,1]$, indicating that in $\mathbb{R}$ they are essentially all irrational numbers.

Uniform Random Variable (page 60 in the textbook)

- As the same pattern repeats, the set $\{x = (0.00x_3x_4\cdots)_2\}$ has the smallest number 0, while the largest one $(0.01)_2 = 0.25$.
- The set $\{x = (0.01x_2x_3\cdots)_2\}$ has the smallest number $(0.01)_2 = 0.25$, while the largest one $(0.1)_2 = 0.5$.
- The set $\{x = (0.10x_2x_3\cdots)_2\}$ has the smallest number $(0.1)_2 = 0.5$, while the largest one $(0.11)_2 = 0.75$..
- The set $\{x = (0.11x_2x_3\cdots)_2\}$ has the smallest number $(0.11)_2 = 0.75$, while there is no largest one in the set.



Uniform Random Variable (page 60 in the textbook)

- In general, at the n^{th} step, the interval $[0,1)$ is divided into 2^n subintervals, each of the length $\frac{1}{2^n}$. The first n binary digits of x determine which of the 2^n subintervals x belongs to.
- If X is a uniform random variable on $[0,1)$, any of the 2^n subintervals are equally likely, each with the probability of $\frac{1}{2^n}$ to happen.
- In other words, the binary digits of a uniformly distributed $X \sim \mathrm{Unif}[0,1)$

$$X(\omega) = (0.x_1x_2x_3\cdots)_2 = \sum_{i=1}^{\infty} \frac{x_i}{2^i},\ \ x_i = 0 \vee 1$$

are the result of an infinite sequence of independent fair coin tosses.

Infinite coin tossing space

- Let $\Omega_\infty = \{\omega = (\omega_1, \omega_2, \omega_3, \ldots) : \ \omega_i = H \vee T, \ \forall i = 1, 2, \ldots\}$ be the set of all nonterminating sequences of H and T, modeling the situation that a coin can be tossed repeatedly without stopping.
- Ω_∞ is an uncountably infinite space.
- For each integer n, we define $\mathcal{F}_n$ to be the σ-algebra containing information up to the first n tosses.
- For example,

$$\begin{aligned}\mathcal{F}_2 = \quad \{ \quad & \emptyset, \Omega, A_{HH}, A_{HT}, A_{TH}, A_{TT}, \\ & A_H, A_T, A_{HH} \cup A_{TH}, A_{HT} \cup A_{TT}, A_{HH} \cup A_{TT}, A_{HT} \cup A_{TH}, \\ & A_{HH}^c, A_{HT}^c, A_{TH}^c, A_{TT}^c\}.\end{aligned}$$

where

$$A_{HH} = \{\omega = (H, H, \omega_3, \omega_4, \ldots) : \ \omega_i = H \vee T, \ \forall i = 3, 4, \ldots\},$$

$$A_{HT} = \{\omega = (H, T, \omega_3, \omega_4, \ldots) : \ \omega_i = H \vee T, \ \forall i = 3, 4, \ldots\},$$

and so forth.

- Each $A_{HH}, A_{HT}, A_{TH}, A_{TT}$ consists of an uncountable number of sample points, so do their unions.

Infinite coin tossing space

- We define the σ-algebra $\mathcal{F}_\infty$ on Ω_∞ to be the smallest σ-algebra generated by the union of all $\mathcal{F}_n$'s, denoted by $\mathcal{F}_\infty = \sigma(\bigcup_{n=1}^\infty \mathcal{F}_n)$.
- Notice that $\mathcal{F}_\infty$ contains sets not belonging to $\bigcup_{n=1}^\infty \mathcal{F}_n$.
- For example, the set containing the single sequence

$$\{(H, H, H, \cdots)\} = \{H \text{ on every toss}\} = \bigcap_{n=1}^{\infty} \{H \text{ on the first } n \text{ tosses}\}$$

  is in $\mathcal{F}_\infty$ because the singleton set $\{(H, H, H, \cdots)\}$ is formed by countable intersections of $A_H \in \mathcal{F}_1, A_{HH} \in \mathcal{F}_2, A_{HHH} \in \mathcal{F}_3, \cdots$
- Another example is

$$\{(H, T, H, T, H \cdots)\} = A_H \cap A_{HT} \cap A_{HTH} \cap \cdots$$

- However, either the "singleton" events $\{(H, H, H, \cdots)\}$ or $\{(H, T, H, T, H \cdots)\}$ are not in any of the $\mathcal{F}_n$'s because each element in $\mathcal{F}_n$, $\forall n \in \mathbb{N}$ consisting of an uncountable number of sample points (except for $\emptyset$).

Infinite coin tossing space

- We next construct a probability measure P on $(\Omega_\infty, \mathcal{F}_\infty)$ which corresponds to probability $p \in [0,1]$ for a single toss H and $q = 1 - p$ for T.
- First, for $A \in \mathcal{F}_n$, since it depends on only the first n tosses, $P(A)$ can be defined to be the product of the p's and q's corresponding to the n tosses. For example, we define $P(A_{HH}) = p^2, P(A_{TH}) = qp$ so that $P(A_{HH} \cup A_{TH}) = p^2 + qp = p$.
- In other words, the probability of the event for a H on the second toss (in tossing a coin infinitely many times) is p, the same as the probability to get a H in a single toss.
- For sets $A \in \mathcal{F}_\infty \setminus \bigcup_{n=1}^\infty \mathcal{F}_n$, we define $P(A)$ by the limit.
- For example, we can define $P(\{(H, H, H, \cdots)\}) = \lim_{n\to\infty} p^n$ since $\{(H, H, H, \cdots)\}$ can be represented as the intersection of a sequence of decreasing sets: $A_H, A_{HH}, A_{HHH}, \cdots$
- When $p = 1$, $P(\{(H, H, H, \cdots)\}) = 1$. Otherwise, $P(\{(H, H, H, \cdots)\}) = 0$ for $0 \le p < 1$.

Infinite coin tossing space

- On Ω_∞, let us define a sequence of random variables $Y_1, Y_2, \ldots$ by
$$Y_k(\omega) = \begin{cases} 1, & \omega_k = H, \\ 0, & \omega_k = T. \end{cases}$$
- With $\{Y_k\}_{k=1}^{\infty}$, let us define $X(\omega) = \sum_{k=1}^{\infty} \frac{Y_k(\omega)}{2^k}$.
- By this way, the random variable X sends a sample point $\omega \in \Omega_\infty$ into a value in $[0,1]$ which has the binary expression $X(\omega) = (0.Y_1(\omega)Y_2(\omega)Y_3(\omega)\cdots)_2$
- At the first step, we toss a fair coin to determine which of the two subintervals $[0, 0.5]$, $[0.5, 1)$ the number $X(\omega)$ belongs to.
- Suppose $Y_1(\omega) = T$, $X(\omega)$ belongs to $[0, 0.5]$.
- The event that the infinite coin tossing with the first trial to be tail is $A_T = \{\omega = (T, \omega_2, \omega_3, \omega_4, \ldots) : \ \omega_i = H \vee T, \ \forall i = 2, 4, \ldots\}$, which is sent by the r.v. X to $[0, 0.5]$.

Uniform Random Variable (page 60 in the textbook)

- At the second step, we toss a fair coin again to determine which of the two subintervals $[0, 0.25]$, $[0.25, 0.5]$ the number $X(\omega)$ belongs to.
- Otherwise, if $Y_1(\omega) = H$, $X(\omega)$ belongs to $[0.5, 1)$, at the second step, we toss a fair coin again to determine which of the two subintervals $[0.5, 0.75]$, $[0.75, 1)$ the number $X(\omega)$ belongs to.
- Continue the experiment for infinitely many times. We can then obtain a real number in $[0, 1)$ in an equally likely manner.
- However, since a computer cannot execute a random experiment for infinitely many times, the random number generator is difficulty to achieve.

Homework Exercise

A "dyadic rational number" is a real number of the form $\frac{m}{2^k}$ where k and m are integers. Suppose we set $p = q = \frac{1}{2}$ in the construction for a probability measure on Ω_∞ and $X(\omega) = \sum_{k=1}^{\infty} \frac{Y_k(\omega)}{2^k}$ is a random variable on Ω_∞.

- Show that, the induced measure $\mathcal{L}_X$ by the random variable X on Ω satisfies that, for any positive integers k and m such that $0 \leq \frac{m-1}{2^k} < \frac{m}{2^k} \leq 1$, we have

  $$\mathcal{L}_X[\frac{m-1}{2^k}, \frac{m}{2^k}] = \frac{1}{2^k}.$$

  In other words, the induced measure $\mathcal{L}_X$ on all intervals in $[0,1]$ whose endpoints are dyadic rational numbers is the same as the Lebesgue measure of these intervals. The only possible way is that $\mathcal{L}_X$ is indeed the Lebesque measure.
- Show that, in this case ($p = \frac{1}{2}$), the random variable $X(\omega) = \sum_{k=1}^{\infty} \frac{Y_k(\omega)}{2^k}$ is uniformly distributed on $[0,1]$.

Exponential Random Variable (page 61 in the textbook)

- An exponential random variable, denoted by $X \sim \mathrm{Exp}(\lambda)$, is a continuous random variable taking non-negative values on $x \in [0, \infty)$ while having the following density function with parameter $\lambda > 0$:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \in [0, \infty); \\ 0, & \text{if } x < 0. \end{cases}$$

- The expectation

$$EX = \int_0^\infty \lambda x e^{-\lambda x} dx = \frac{1}{\lambda} \int_0^\infty t e^{-t} dt = \frac{1}{\lambda} \int_0^\infty -t de^{-t} = \frac{1}{\lambda}.$$

- $\mathrm{Var}(X) = \frac{1}{\lambda^2}$. (This is left as an exercise)
- $P(X \geq x) = \int_x^\infty \lambda e^{-\lambda t} dt = e^{-\lambda x}$; $P(X \leq x) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}$.
- (Example 6.5) Let $X \sim \mathrm{Exp}(\lambda)$ be the lifespan of a lightbulb which is random. Assuming that the lightbulbs last on average 100 hours. What is the probability that it lasts less than 50 hours?
- We first note that $\lambda = \frac{1}{\mu_X} = 0.01$. Then, $P(X < 50) = 1 - e^{-0.01 \cdot 50} \approx 0.3935$.

Normal Random Variable (page 61-62 in the textbook)

- A normal random variable, denoted by $X \sim N(\mu, \sigma^2)$, is a continuous random variable taking all real values on $\mathbb{R}$ while having the following density function with parameter μ, σ^2:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \ x \in (-\infty, \infty).$$

- Certainly, for any μ and σ^2, there is

$$\int_{-\infty}^{\infty} f_X(x) dx = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1.$$

- The expectation

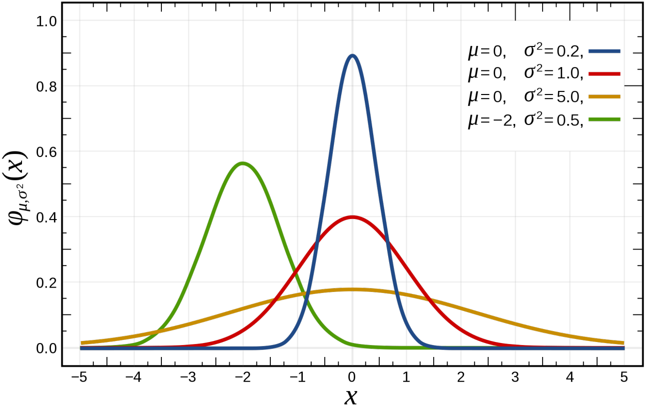
$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} (x-\mu) e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx + \mu \\ &= \int_{-\infty}^{\infty} \underbrace{\frac{1}{\sigma\sqrt{2\pi}} z e^{-\frac{z^2}{2\sigma^2}}}_{\text{odd function}} dz + \mu = \mu \end{aligned}$$

- Variance (calculation omitted): $\mathrm{Var}(X) = \sigma^2$.

Normal Random Variable (page 62 in the textbook)

- Density functions of $X \sim N(\mu, \sigma^2)$ with different parameters.

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \; x \in (-\infty, \infty).$$



Normal Random Variable (page 62 in the textbook)

- Let $X \sim N(\mu, \sigma^2)$ be a normal r.v. and let $Y = \alpha X + \beta$, with $\alpha > 0$, which is a linear transformation on the value of a normal r.v.
- We start by computing the cumulative distribution of Y:

$$
\begin{aligned}
F_Y(y) &= P(Y \leq y) = P(\alpha X + \beta \leq y) \\
&= P(X \leq \frac{y-\beta}{\alpha}) \\
&= \int_{-\infty}^{\frac{y-\beta}{\alpha}} f_X(x)dx
\end{aligned}
$$

- The density of Y

$$
f_Y(y) = F'_Y(y) = f_X(\frac{y-\beta}{\alpha})\frac{1}{\alpha} = \frac{1}{\alpha\sigma\sqrt{2\pi}} e^{-\frac{(y-\beta-\alpha\mu)^2}{2\alpha^2\sigma^2}}
$$

- Then, $Y \sim N(\alpha\mu + \beta, (\alpha\sigma)^2)$ is normal with $EY = \alpha\mu + \beta$ and variance $\mathrm{Var}(Y) = (\alpha\sigma)^2$.

Normal Random Variable (page 63 in the textbook)

- In particular, if $X \sim N(\mu, \sigma^2)$ and let $Z = \frac{X-\mu}{\sigma}$, then Z is also normal with

$$EZ = \frac{EX - \mu}{\sigma} = 0 \text{ and } \mathrm{Var}(Z) = (\frac{1}{\sigma} \cdot \sigma)^2 = 1.$$

- Such a $N(0, 1^2)$ random variable is called *standard* Normal. It has density:

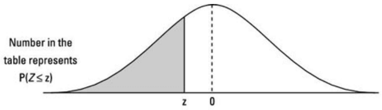
$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \ z \in (-\infty, \infty).$$

- The cumulative distribution of Z is denoted by $\Phi(z)$ with

$$\Phi(z) = F_Z(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{x^2}{2}} dx$$

- The integral for $\Phi(z)$ cannot be computed as an elementary function, so approximate values are given in tables.
- By the fact that $f_Z(z)$ is even, we have $\Phi(-z) = 1 - \Phi(z)$.

Normal Random Variable ($P(Z \leq -2.67) = 0.0038$)



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
–3.6	.0002	.0002	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
–3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
–3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
–3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
–3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
–3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
–3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
–2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
–2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
–2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
–2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
–2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
–2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
–2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
–2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
–2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
–2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
–1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
–1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
–1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
–1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
–1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
–1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
–1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
–1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
–1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
–1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
–0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
–0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867

Normal Random Variable (Example 6.7 page 63 in the textbook)

- What is the probability that a Normal random variable differs from its mean μ by more than σ? more than 2σ? more than 3σ?
- In mathematical symbols, if $X \sim N(\mu, \sigma^2)$, we need to compute $P(|X - \mu| \geq \sigma)$, $P(|X - \mu| \geq 2\sigma)$, and $P(|X - \mu| \geq 3\sigma)$.
- The computation is easier through transforming to a standard normal random variable $Z = \frac{X-\mu}{\sigma} \sim N(0, 1^2)$. That is,

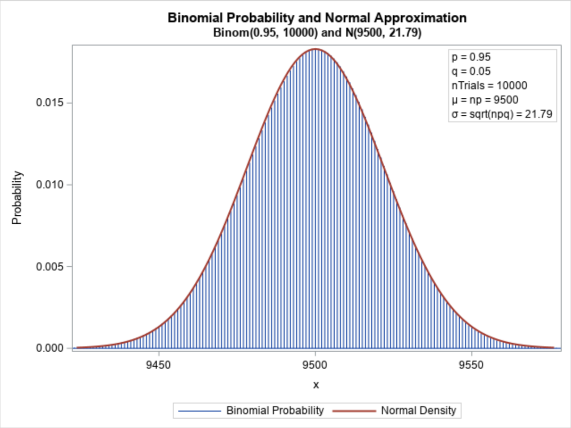
$$
\begin{aligned}
P(|X - \mu| \geq \sigma) &= P(\frac{X - \mu}{\sigma} \geq 1) \\
&= 2P(Z \leq -1) \\
&\approx 2 \cdot 0.1587 = 0.3174.
\end{aligned}
$$

- Similarly, $P(|X - \mu| \geq 2\sigma) = P(|Z| \geq 2) = 2P(Z \leq -2) = 2 \cdot (0.0228) = 0.0456$.

de Moivre-Laplace Central Limit Theorem (1783) (page 64 in the textbook)

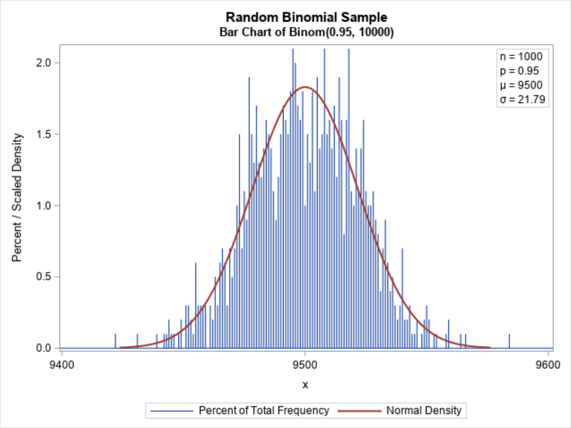
- Let $S_n \sim \mathrm{Binomial}(n, p)$. Recall that its mean it np and its variance is $np(1-p) = npq$.
- If we pretend that S_n is Normal with mean np and variance npq, then
$$\frac{S_n - np}{\sqrt{npq}} \sim N(0, 1).$$
- de Moivre-Laplace Central Limit Theorem assures that such pretending is indeed "quite real" when p is fixed and n is large. That is, the normal distribution can be used to approximate the binomial distribution "under certain conditions."
- For example, if k is very close to np, we can directly comptue
$$\binom{n}{k} p^k q^{n-k} \simeq \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}}$$
by Stirling's formula $n! \simeq n^n e^{-n} \sqrt{2\pi n}$ and $\ln(1+x) \simeq x - \frac{x^2}{2} + \frac{x^3}{3} + \cdots$

de Moivre-Laplace Central Limit Theorem (1783)



- Notice that, the binomial density is discrete, which is defined only for positive integers, whereas the normal density is defined for all real numbers.

de Moivre-Laplace Central Limit Theorem (1783)



- If we take a sample size of 1000 from the binomial distribution $\mathrm{Binomial}(10000, 0.95)$, the distribution of the sample (percent) looks, at first glance, a bit alike to the density curve of normal, but quite different at a closer look.

de Moivre-Laplace Central Limit Theorem (1783) (page 64 in the textbook)

- Let $S_n \sim \mathrm{Binomial}(n, p)$ and m_1, m_2 be two positive integers. The probability that the number of successes between $m_1 < m_2$ has a precise formula:

$$P(m_1 \le S_n \le m_2) = \sum_{i=m_1}^{m_2} \binom{n}{i} p^i q^{n-i}.$$

- For large number of m_1 and m_2, the computation of the precise formula could be tedious.
- However, according to de Moivre-Laplace Central Limit Theorem, $\frac{S_n - np}{\sqrt{npq}} \sim N(0,1)$ so that

$$\begin{aligned} P(m_1 \le S_n \le m_2) &= P(\underbrace{\frac{m_1 - np}{\sqrt{npq}}}_{=\alpha} \le \frac{S_n - np}{\sqrt{npq}} \le \underbrace{\frac{m_2 - np}{\sqrt{npq}}}_{=\beta}) \\ &= \int_\alpha^\beta \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \end{aligned}$$

de Moivre-Laplace Central Limit Theorem

- A die is threw 12000 times. What is the probability that there will be exactly 1800 rolls of 6?
- This is a Binomial trial for $n = 12000$, $p = \frac{1}{6}$. The exact probability is $\binom{12000}{1800}\left(\frac{1}{6}\right)^{1800}\left(\frac{5}{6}\right)^{10200}$, whose exact value is difficult to compute.
- The probability can be approximated by Poisson(np)=Poisson(2000) $= \frac{e^{-2000}2000^{1800}}{1800!}$, which is still very difficult to compute.
- However, if we approximate by de Moivre-Laplace Central Limit Theorem,

$$\binom{12000}{1800}\left(\frac{1}{6}\right)^{1800}\left(\frac{5}{6}\right)^{10200} \approx \underbrace{\frac{1}{\sqrt{2\pi \cdot 1666.67}}}_{=9.772\times 10^{-3}} \cdot \underbrace{e^{-\frac{(1800-2000)^2}{2\cdot 1666.67}}}_{=6.144\times 10^{-6}} \approx 6.004\times 10^{-8}.$$

- (EXERCISE) What is the approximate probability for the number of $6's$ lies in the interval $[1950, 2100]$?