1. Uncountable probability spaces - the resolution

Recall the example of drawing a number at random from $\Omega = [0,1]$. Any interval (a,b) with $0 \le a < b \le 1$ is called a basic set and its probability is defined as $\mathbf{P}(a,b) = b - a$. Is this truly a solution to the question of uncountable spaces? Are we assured of never running into inconsistencies? NOT always!

Example 1. Let $\Omega = [0,1]$ and let intervals (a,b) be open sets with their probabilities defined as $\mathbf{P}(a,b) = \sqrt{b-a}$. This quickly leads to problems. For example, $\mathbf{P}(0,1) = 1$ by definition. But $(0,1) = (0,0.5) \cup (0.5,1) \cup \{1/2\}$ from which the rules of probability would imply that $\mathbf{P}(0,1)$ must be at least $\mathbf{P}(0,1/2) + \mathbf{P}(1/2,1) = \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} = \sqrt{2}$ which is greater than 1. Inconsistency!

Thus, one cannot arbitrarily assign probabilities to basic events. However, if we use the notion of distribution function to assign probabilities to intervals, then no inconsistencies arise.

Theorem 2. *Let $\Omega = \mathbb{R}$ and let intervals of the form $(a,b]$ with $a < b$ be called basic sets. Let F be any distribution function.* Define *the probabilities of basic sets as $\mathbf{P}\{(a,b]\} = F(b) - F(a)$. Then, applying the rules of probability to compute probabilities of more complex sets (obtained by taking countable intersections, unions and complements) will never lead to inconsistency.*

Let F be a CDF. Then, the above consistency theorem really asserts that there exists (a possibly uncountable) probability space and a random variable such that $F(t) = \mathbf{P}\{X \le t\}$ for all t. We say that X has distribution F. However, it takes a lot of technicalities to define what uncountable probability spaces look like and what random variables mean in this more general setting, we shall never define them.

The job of a probabilist consists in taking a CDF F (then the probabilities of intervals are already given to us as $F(b) - F(a)$ etc.) and find probabilities of more general subsets of $\mathbb{R}$. Instead we can use the following simple working rules to answer questions about the distribution of a random variable. Here are the working rules:

(1) For an $a < b$, we set $\mathbf{P}\{a < X \le b\} := F(b) - F(a)$.

(2) If $I_j = (a_j, b_j]$ are countably many pairwise disjoint intervals, and $I = \bigcup_j I_j$, then we define $\mathbf{P}\{X \in I\} := \sum_j F(b_j) - F(a_j)$.

(3) For a general set $A \subseteq \mathbb{R}$, here is a general scheme: Find countably many pairwise disjoint intervals $I_j = (a_j, b_j]$ such that $A \subseteq \cup_j I_j$. Then we define $\mathbf{P}\{X \in A\}$ as the infimum (over all such coverings by intervals) of the quantity $\sum_j F(b_j) - F(a_j)$.

All of probability in another line: Take an (interesting) random variable X with a given CDF F and an (interesting) set $A \subseteq \mathbb{R}$. Find $\mathbf{P}\{X \in A\}$.

There are loose threads here, but they can be safely ignored for this course. We just remark about them for those who are curious to know.

Remark 3. The above method starts from a CDF F and defines $\mathbf{P}\{X \in A\}$ for all subsets $A \subseteq \mathbb{R}$. However, for most choices of F, the countable additivity property turns out to be violated! The sets which violate them rarely arise in practice, and hence we ignore them for the present exposition.

Exercise 4. Let X be a random variable with distribution F. Use the working rules to find the following probabilities:

(1) Show that $\mathbf{P}\{X < a\} = F(a-)$, where $\lim_{h \to 0^+} F(a-h) = F(a-)$.

Hint: $(-\infty, a-1/n]$ increases to $(-\infty, a)$ as $n \to \infty$.

(2) Now, show that $\mathbf{P}\{X = a\} = F(a) - F(a-)$. In particular, this probability is zero unless F has a jump at a.

(3) Write $\mathbf{P}\{a < X < b\}$, $\mathbf{P}\{a \le X < b\}$, $\mathbf{P}\{a \le X \le b\}$ in terms of F.

We now illustrate how to calculate the probabilities of rather non-trivial sets in a special case. It is not always possible to get an explicit answer as here.

Example 5. Let F be the CDF defined below:

$$F(t) = \begin{cases} 0 & \text{if } t \le 0, \\ t & \text{if } 0 < t < 1, \\ 1 & \text{if } t \ge 1. \end{cases}$$

We calculate $\mathbf{P}\{X \in A\}$ for two sets A.

1. $A = \mathbb{Q} \cap [0,1]$. Since A is countable, we may write $A = \cup_n \{r_n\}$ and hence $A \subseteq \cup_n I_n$, where $I_n = (r_n, r_n + \delta 2^{-n}]$ for any fixed $\delta > 0$. Hence, $0 \le \mathbf{P}\{X \in A\} \le \sum_n F(r_n + \delta 2^{-n}) - F(r_n) \le 2\delta$. Since this is true for every $\delta > 0$, we must have $\mathbf{P}\{X \in A\} = 0$. (We stuck to the letter of the recipe described earlier. It would have been simpler to say that any countable set is a countable union of singletons, and by the countable additivity of probability, must have probability zero. Here, we have used the fact that singletons have zero probability since F is continuous).

2. $A =$ Cantor's set. [1] How to find $\mathbf{P}\{X \in A\}$? Let A_n be the set of all $x \in [0,1]$ which do not have 1 in the first n digits of their ternary expansion. Then $A \subseteq A_n$. Further, it is not hard to see that $A_n = I_1 \cup I_2 \cup \cdots \cup I_{2^n}$, where each of the intervals I_j has length equal to 3^{-n}. Therefore, $0 \leq \mathbf{P}\{X \in A\} \leq \mathbf{P}\{X \in A_n\} = 2^n 3^{-n}$ which goes to 0 as $n \to \infty$. Hence, $\mathbf{P}\{X \in A\} = 0$.

[1] To define the Cantor set, recall that any $x \in [0,1]$ may be written in ternary expansion as $x = 0.u_1 u_2 \ldots := \sum_{n=1}^{\infty} u_n 3^{-n}$ where $u_n \in \{0,1,2\}$. This expansion is unique except if x is a rational number of the form $p/3^m$ for some integers p, m (these are called triadic rationals). For triadic rationals, there are two possible ternary expansions, a terminating one and a non-terminating one (for example, $x = 1/3$ can be written as $0.100\ldots$ or as $0.0222\ldots$). For definiteness, for triadic rationals we shall always take the non-terminating ternary expansion. With this preparation, the Cantor set is defined as the set of all x which do not have the digit 1 in their ternary expansion.

2. EXAMPLES OF CONTINUOUS DISTRIBUTIONS

Cumulative distributions (CDF) will also be referred to as simply distribution functions (DF). We start by giving two large classes of CDFs. There are CDFs that do not belong to either of these classes, but for practical purposes they will be ignored (for now).

(1) (CDFs with pmf). Let f be a pmf, i.e., let $t_1, t_2, \ldots$ be a countable subset of reals and let $f(t_i)$ be non-negative numbers such that $\sum_i f(t_i) = 1$. Define $F : \mathbb{R} \to [0,1]$ by

$$F(t) := \sum_{i: t_i \le t} f(t_i).$$

Then, F is a CDF. Indeed, we have seen that it is the CDF of a discrete random variable. A special feature of this CDF is that it increases only in jumps (in more precise language, if F is continuous on an interval $[s,t]$, then $F(s) = F(t)$).

(2) (CDFs with pdf). Let $f : \mathbb{R} \to \mathbb{R}_+$ be a function (convenient to assume that it is a piecewise continuous function) such that $\int_{-\infty}^{+\infty} f(u)du = 1$. Such a function is called a *probability density function* (or, pdf for short). Then, define $F : \mathbb{R} \to [0,1]$ by

$$F(t) := \int_{-\infty}^{t} f(u)du.$$

Again, F is a CDF. Indeed, it is clear that F has the increasing property (if $t > s$, then $F(t) - F(s) = \int_s^t f(u)du$ which is non-negative because $f(u)$ is non-negative for all u), and its limits at $\pm\infty$ are as they should be (Why?). As for right-continuity, F is in-fact continuous. Actually, F is differentiable except at points where f is discontinuous and $F'(t) = f(t)$.

Remark 6. We understand the pmf. For example if X has pmf f, then $f(t_i)$ is just the probability that X takes the value t_i. How to interpret the pdf? If X has pdf f, then as we already remarked, the CDF is continuous and hence $\mathbf{P}\{X = t\} = 0$. Therefore, $f(t)$ cannot be interpreted as $\mathbf{P}\{X = t\}$ (in fact, pdf can take values greater than 1, so it cannot be a probability!).

To interpret $f(a)$, take a small positive number δ and look at

$$F(a+\delta) - F(a) = \int_a^{a+\delta} f(u)du \approx \delta f(a).$$

In other words, $f(a)$ measures the chance of the random variable taking values near a. Higher the pdf, greater the chance of taking values near that point.

Among distributions with pmf, recall that we have seen the Binomial, Poisson, Geometric and Hypergeometric families of distributions. Now, we give many important examples of distributions (CDFs) with densities.

Example 7. Uniform distribution on the interval $[a, b]$: Denoted by Unif($[a, b]$), where $a < b$ is the distribution with density and distribution given by

$$\text{PDF: } f(t) = \begin{cases} \frac{1}{b-a} & \text{if } t \in (a, b) \\ 0 & \text{otherwise} \end{cases} \qquad \text{CDF: } F(t) = \begin{cases} 0 & \text{if } t \leq a \\ \frac{t-a}{b-a} & \text{if } t \in (a, b) \\ 1 & \text{if } t \geq b. \end{cases}$$

Example 8. Exponential distribution with parameter λ: Denoted by Exp(λ), where $\lambda > 0$ is the distribution with density and distribution given by

$$\text{PDF: } f(t) = \begin{cases} \lambda e^{-\lambda t} & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases} \qquad \text{CDF: } F(t) = \begin{cases} 0 & \text{if } t \leq 0 \\ 1 - e^{-\lambda t} & \text{if } t > 0. \end{cases}$$

Example 9. Normal distribution with parameters μ, σ^2: Denoted by N(μ, σ^2), where $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ is the distribution with density and distribution given by

$$\text{PDF: } \varphi_{\mu,\sigma^2}(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(t-\mu)^2} \qquad \text{CDF: } \Phi_{\mu,\sigma^2}(t) = \int\limits_{-\infty}^{t} \varphi_{\mu,\sigma^2}(u)du.$$

There is no closed form expression for the CDF. It is standard notation to write φ and Φ to denote the normal density and CDF when $\mu = 0$ and $\sigma^2 = 1$. N$(0, 1)$ is called the standard normal distribution. By a change of variable, one can check that $\Phi_{\mu,\sigma^2}(t) = \Phi(\frac{t-\mu}{\sigma})$.

We said that the normal CDF has no simple expression, but is it even clear that it is a CDF?! In other words, is the proposed density a true pdf? Clearly $\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$ is non-negative. We need to check that its integral is 1.

Lemma 10. *Fix $\mu \in \mathbb{R}$ and $\sigma > 0$ and let $\varphi(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(t-\mu)^2}$. Then, $\int\limits_{-\infty}^{\infty} \varphi(t)dt = 1$.*

Proof. It suffices to check the case $\mu = 0$ and $\sigma^2 = 1$ (Why?). To find its integral is quite non-trivial. Let $I = \int_{-\infty}^{\infty} \varphi(t)dt$. We introduce the two-variable function $h(t, s) := \varphi(t)\varphi(s) = (2\pi)^{-1} e^{-(t^2+s^2)/2}$. On one hand,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(t, s)dtds = \left(\int_{-\infty}^{+\infty} \varphi(t)dt\right) \left(\int_{-\infty}^{+\infty} \varphi(s)ds\right) = I^2.$$

On the other hand, using polar co-ordinates $t = r\cos\theta$, $s = r\sin\theta$, we see that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(t, s)dtds = \int_{0}^{\infty} \int_{0}^{2\pi} (2\pi)^{-1} e^{-r^2/2} r d\theta dr = \int_{0}^{\infty} r e^{-r^2/2} dr = 1$$

since $\frac{d}{dr}e^{-r^2/2} = -re^{-r^2/2}$. Thus $I^2 = 1$, and hence $I = 1$. ■

Example 11. Gamma distribution with shape parameter ν and scale parameter λ: Denoted by Gamma(ν, λ) with $\nu > 0$ and $\lambda > 0$, is the distribution with density and distribution given by:

$$\text{PDF: } f(t) = \begin{cases} \frac{1}{\Gamma(\nu)}\lambda^\nu t^{\nu-1}e^{-\lambda t} & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases} \qquad \text{CDF: } F(t) = \begin{cases} 0 & \text{if } t \leq 0 \\ \int_0^t f(u)du & \text{if } t > 0. \end{cases}$$

Here, $\Gamma(\nu) := \int_0^\infty t^{\nu-1}e^{-t}dt$. Firstly, f is a density, i.e., it integrates to 1. To see this, make the change of variable $\lambda t = u$ to see that

$$\int_0^\infty \lambda^\nu e^{-\lambda t}t^{\nu-1}dt = \int_0^\infty e^{-u}u^{\nu-1}du = \Gamma(\nu).$$

Thus, $\int_0^\infty f(t)dt = 1$.

When $\nu = 1$, we get back the exponential distribution. Thus, the Gamma family subsumes the exponential distributions.

Exercise 12. For positive integer values of ν, one can actually write an expression for the CDF of Gamma(ν, λ) as

$$F_{\nu,\lambda}(t) = 1 - e^{-\lambda t}\sum_{k=0}^{\nu-1}\frac{(\lambda t)^k}{k!}.$$

Once the expression is given, it is easy to check it by induction (and integration by parts). A curious observation is that the right hand side is exactly $\mathbf{P}(N \geq \nu)$, where $N \sim$ Pois(λt). This is in fact indicating a deep connection between Poisson distribution and the Gamma distributions. The function $\Gamma(\nu)$, also known as Euler's Gamma function, is an interesting and important one and occurs all over mathematics. [2]

[2]**The Gamma function:** The function $\Gamma : (0, \infty) \to \mathbb{R}$ defined by $\Gamma(\nu) = \int_0^\infty e^{-t}t^{\nu-1}dt$ is a very important function that often occurs in mathematics and physics. There is no simpler expression for it, although one can find it explicitly for special values of ν. One of its most important properties is that $\Gamma(\nu + 1) = \nu\Gamma(\nu)$. To see this, consider

$$\Gamma(\nu + 1) = \int_0^\infty e^{-t}t^\nu dt = -e^{-t}t^\nu \Big|_0^\infty + \nu\int_0^\infty e^{-t}t^{\nu-1}dt = \nu\Gamma(\nu).$$

Starting with $\Gamma(1) = 1$ (direct computation) and using the above relationship repeatedly one sees that $\Gamma(\nu) = (\nu - 1)!$ for positive integer values of ν. Thus, the Gamma function interpolates the factorial function (which is defined only for positive integers). Can we compute it for any other ν? The answer is yes, but only for special values of ν. For example,

$$\Gamma(1/2) = \int_0^\infty x^{-1/2}e^{-x}dx = \sqrt{2}\int_0^\infty e^{-y^2/2}dy$$

by substituting $x = y^2/2$. The last integral was computed above in the context of the normal distribution and equal to $\sqrt{\pi/2}$. Hence, we get $\Gamma(1/2) = \sqrt{\pi}$. From this, using again the relation $\Gamma(\nu + 1) = \nu\Gamma(\nu)$, we can compute $\Gamma(3/2) = \frac{1}{2}\sqrt{\pi}$, $\Gamma(5/2) = \frac{3}{4}\sqrt{\pi}$, etc. Yet another useful fact about the Gamma function is its asymptotics as $\nu \to \infty$.

Example 13. Beta distributions: Let $\alpha, \beta > 0$. The Beta distribution with parameters α, β, denoted Beta(α, β), is the distribution with density and distribution given by

$$\text{PDF: } f(t) = \begin{cases} \frac{1}{B(\alpha,\beta)} t^{\alpha-1}(1-t)^{\beta-1} & \text{if } t \in (0,1) \\ 0 & \text{otherwise} \end{cases} \qquad \text{CDF: } F(t) = \begin{cases} 0 & \text{if } t \leq 0 \\ \int_0^t f(u)du & \text{if } t \in (0,1) \\ 0 & \text{if } t \geq 1. \end{cases}$$

Here, $B(\alpha, \beta) := \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt$. Again, for special values of α, β (e.g., positive integers), one can find the value of $B(\alpha, \beta)$, but in general there is no simple expression. However, it can be expressed in terms of the Gamma function!

Proposition 14. *For any $\alpha, \beta > 0$, we have $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.*

Proof. For $\beta = 1$, we see that $B(\alpha, 1) = \int_0^1 t^{\alpha-1} = \frac{1}{\alpha}$ which is also equal to $\frac{\Gamma(\alpha)\Gamma(1)}{\Gamma(\alpha+1)}$ as required. Similarly (or, by the symmetry relation $B(\alpha, \beta) = B(\beta, \alpha)$), we see that $B(1, \beta)$ also has the desired expression.

Now, for any other *positive integer* value of α and real $\beta > 0$ we can integrate by parts and get

$$\begin{aligned} B(\alpha, \beta) &= \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt \\ &= -\frac{1}{\beta} t^{\alpha-1}(1-t)^{\beta} \Big|_0^1 + \frac{\alpha-1}{\beta} \int_0^1 t^{\alpha-2}(1-t)^{\beta}dt \\ &= \frac{\alpha-1}{\beta} B(\alpha-1, \beta+1). \end{aligned}$$

Stirling's approximation: $\frac{\Gamma(\nu+1)}{\nu^{\nu+\frac{1}{2}} e^{-\nu}\sqrt{2\pi}} \to 1$ as $\nu \to \infty$.

A small digression: It was Euler's idea to observe that $n! = \int_0^\infty x^n e^{-x}dx$ and that on the right side n could be replaced by any real number greater than -1. But this was his second approach to defining the Gamma function. His first approach was as follows. Fix a positive integer n. Then for any $\ell \geq 1$ (also a positive integer), we may write

$$n! = \frac{(n+\ell)!}{(n+1)(n+2)\cdots(n+\ell)} = \frac{\ell!(\ell+1)\cdots(\ell+n)}{(n+1)\cdots(n+\ell)} = \frac{\ell!\,\ell^n}{(n+1)\cdots(n+\ell)} \cdot \frac{(\ell+1)\cdots(\ell+n)}{\ell^n}$$

The second factor approaches 1 as $\ell \to \infty$. Hence,

$$n! = \lim_{\ell\to\infty} \frac{\ell!\,\ell^n}{(n+1)\cdots(n+\ell)}.$$

Euler then showed (by a rather simple argument that we skip) that the limit on the right exists if we replace n by any complex number other than $\{-1, -2, -3, \ldots\}$ (negative integers are a problem as they make the denominator zero). Thus, he extended the factorial function to all complex numbers except negative integers! It is a fun exercise to check that this agrees with the definition by the integral given earlier. In other words, for $\nu > -1$, we have

$$\lim_{\ell\to\infty} \frac{\ell!\,\ell^\nu}{(\nu+1)\cdots(\nu+\ell)} = \int_0^\infty x^\nu e^{-x}dx.$$

Note that the first term vanishes because $\alpha > 1$ and $\beta > 0$. When α is an integer, we repeat this for α times and get

$$B(\alpha, \beta) = \frac{(\alpha-1)(\alpha-2)\cdots 1}{\beta(\beta+1)\cdots(\beta+\alpha-2)} B(1, \beta+\alpha-1).$$

But, we already checked that $B(1, \beta+\alpha-1) = \frac{\Gamma(1)\Gamma(\alpha+\beta-1)}{\Gamma(\alpha+\beta)}$ from which we get

$$B(\alpha, \beta) = \frac{(\alpha-1)(\alpha-2)\cdots 1}{\beta(\beta+1)\cdots(\beta+\alpha-2)} \frac{\Gamma(1)\Gamma(\alpha+\beta-1)}{\Gamma(\alpha+\beta)} = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

by the recursion property of the Gamma function. Thus, we have proved the proposition when α is a positive integer. By symmetry, the same is true when β is a positive integer (and α can take any value). We do not prove the proposition for general $\alpha, \beta > 0$ here. ■

Example 15. The standard Cauchy distribution: A distribution with density and distribution given by

$$\text{PDF: } f(t) = \frac{1}{\pi(1+t^2)} \qquad \text{CDF: } F(t) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1} t.$$

One can also make a parametric family of Cauchy distributions with parameters $\lambda > 0$ and $a \in \mathbb{R}$ denoted Cauchy(a, λ) as follows:

$$\text{PDF: } f(t) = \frac{\lambda}{\pi(\lambda^2 + (t-a)^2)} \qquad \text{CDF: } F(t) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{t-a}{\lambda}\right).$$

Remark 16. Does every CDF come from a pdf? Not necessarily. For example any CDF that is not continuous (for example, CDFs of discrete distributions such as Binomial, Poisson, Geometric etc.). In fact, even continuous CDFs may not have densities (there is a good example manufactured out of the $1/3$-Cantor set, but that would take us out of the topic now). However, suppose F is a *continuous* CDF and suppose F is differentiable except at finitely many points and that the derivative is a continuous function. Then, $f(t) := F'(t)$ defines a pdf which by the fundamental theorm of Calculus satisfies $F(t) = \int_{-\infty}^{t} f(u)du$.

3. IS A CDF NECESSARILY DISCRETE, OR CONTINUOUS?

Let X be a random variable defined on a probability space, and F_X be the distribution function (DF) of X.

Exercise 17. F_X will either be continuous everywhere, or it will have countable number of discontinuities. Moreover, the sum of sizes of jumps at the point of discontinuities of F_X will be either 1, or less than 1.

This property of F_X can be used to classify the random variable X into three broad categories:

- The random variable X is said to be of *discrete* type if there exists a non-empty and countable set S_X such that

$$f_X(x) = \mathbf{P}(\{X = x\}) = F_X(x) - F_X(x-) > 0 \, \forall \, x \in S_X,$$

  and $\mathbf{P}\left(S_X\right) = \sum_{x \in S_X} \mathbf{P}(\{X = x\}) = \sum_{x \in S_X} \left[F_X(x) - F_X(x-)\right] = 1$. The function f_X is called the probability mass function (pmf) of the random variable X, and the set S_X is called the support of random variable X (or, of the pmf f_X).
- A random variable X is said to be of *continuous* type if its distribution function F_X is continuous everywhere.
- A random variable X with distribution function F_X is said to be of *absolutely continuous* type if there exists an integrable function $f_X : \mathbb{R} \to \mathbb{R}_+$ such that $f_X(x) \geq 0 \, \forall \, x \in \mathbb{R}$, and

$$F_X(x) = \int_{-\infty}^{x} f_X(t) dt \, \forall \, x \in \mathbb{R}.$$

  The function f_X is called the probability density function (pdf) of the random variable X, and the set $S_X = \{x \in \mathbb{R} : f_X(x) > 0\}$ is called the support of random variable X (or, of the pdf f_X).

Example 18. Now, consider the following DF:

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0 \\ \frac{x}{4}, & \text{if } 0 \leq x < 1 \\ \frac{x}{3}, & \text{if } 1 \leq x < 2 \\ \frac{3x}{8}, & \text{if } 2 \leq x < \frac{5}{2} \\ 1, & \text{if } x \geq \frac{5}{2}. \end{cases}$$

Note that the set of discontinuity points of F_X is $\{1, 2, 5/2\}$ with

$$\begin{aligned} p_1 &= \mathbf{P}(\{X = 1\}) = F_X(1) - F_X(1-) = \tfrac{1}{12}, \\ p_2 &= \mathbf{P}(\{X = 2\}) = F_X(2) - F_X(2-) = \tfrac{1}{12} \text{ and} \\ p_3 &= \mathbf{P}(\{X = 5/2\}) = F_X(5/2) - F_X(5/2-) = \tfrac{1}{16}. \end{aligned}$$

Thus, $p_1 + p_2 + p_3 = \frac{11}{48} < 1$.

We can decompose F_X as $F_X(x) = \alpha H_d(x) + (1-\alpha)H_c(x)$ for $x \in \mathbb{R}$, where $\alpha \in [0,1]$. Here, H_d is a distribution function of some random variable X_d of discrete type, while H_c is a distribution function of some random variable X_c of continuous type.

Let us take $\alpha = p_1 + p_2 + p_3 = \frac{11}{48}$. Thus, $\mathbf{P}(\{X_d = 1\}) = \frac{p_1}{\alpha} = \frac{4}{11}$, $\mathbf{P}(\{X_d = 2\}) = \frac{p_2}{\alpha} = \frac{4}{11}$ and $\mathbf{P}(\{X_d = 5/2\}) = \frac{p_3}{\alpha} = \frac{3}{11}$. This gives us

$$H_d(x) = \begin{cases} 0, & \text{if } x < 1 \\ \frac{4}{11}, & \text{if } 1 \leq x < 2 \\ \frac{8}{11}, & \text{if } 2 \leq x < \frac{5}{2} \\ 1, & \text{if } x \geq \frac{5}{2}, \end{cases}$$

and

$$\begin{aligned} H_c(x) &= \frac{F_X(x) - \alpha H_d(x)}{1-\alpha} \\ &= \begin{cases} 0, & \text{if } x < 0 \\ \frac{12x}{37}, & \text{if } 0 \leq x < 1 \\ \frac{4(4x-1)}{37}, & \text{if } 1 \leq x < 2 \\ \frac{2(9x-4)}{37}, & \text{if } 2 \leq x < \frac{5}{2} \\ 1, & \text{if } x \geq \frac{5}{2}. \end{cases} \end{aligned}$$

Here, the distribution function F_X (equivalently, the random variable X) is neither discrete, nor continuous.

Remark 19. Convex combination of two DFs is also a DF.

4. Change of Variable

Let $h : \mathbb{R} \to \mathbb{R}$ be a function. Given the distribution of X, how will you find the distribution of $h(X)$?

1. CDF technique: The distribution $Z = h(X)$ can be determined by computing the distribution function. Fix $z \in \mathbb{R}$,

$$F_Z(z) = \mathbf{P}(Z \le z) = \mathbf{P}(h(X) \le z).$$

Depending on the properties of the function h, we may, or may not be able to derive this probability in a closed form expression.

Example 20. Let X be a random variable with pmf

$$f_X(x) = \begin{cases} \begin{pmatrix} n \\ x \end{pmatrix} p^x(1-p)^{n-x}, & \text{if } x \in \{0, 1, \dots, n\} \\ 0, & \text{otherwise ,} \end{cases}$$

where n is a positive integer and $p \in (0, 1)$. Find the distribution function of $Y = n - X$.

Note that $S_X = S_Y = \{0, 1, \dots, n\}$. For $y \in S_Y$, we get

$$\mathbf{P}(\{Y \le y\}) = \mathbf{P}(\{X \ge n - y\}) = \sum_{x=n-y}^{n} \begin{pmatrix} n \\ x \end{pmatrix} p^x(1-p)^{n-x} = \sum_{x=0}^{y} \begin{pmatrix} n \\ n-x \end{pmatrix} (1-p)^x p^{n-x}.$$

Thus, the distribution function of Y is

$$F_Y(y) = \begin{cases} 0, & \text{if } y < 0 \\ p^n, & \text{if } 0 \le y < 1 \\ \sum_{j=0}^{i} \begin{pmatrix} n \\ j \end{pmatrix} (1-p)^j p^{n-j}, & \text{if } i \le y < i+1 \text{ for } i = 1, 2, \dots, n-1 \\ 1, & \text{if } y \ge n. \end{cases}$$

Example 21. Let X be random variable with pdf

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

and $T = X^2$. Find the pdf of T.

We have $F_T(t) = \mathbf{P}\left(\left\{X^2 \le t\right\}\right)$ for $t \in \mathbb{R}$. Clearly, $F_T(t) = 0$ for $t < 0$. For $t \ge 0$,

$$\begin{aligned} F_T(t) &= \mathbf{P}(\{-\sqrt{t} \le X \le \sqrt{t}\}) \\ &= \int_{-\sqrt{t}}^{\sqrt{t}} f_X(x)\mathrm{d}x \\ &= \int_0^{\sqrt{t}} e^{-x}\,\mathrm{d}x \\ &= 1 - e^{-\sqrt{t}}. \end{aligned}$$

Therefore, the distribution function of T is

$$F_T(t) = \begin{cases} 0, & \text{if } t < 0 \\ 1 - e^{-\sqrt{t}}, & \text{if } t \ge 0. \end{cases}$$

Clearly, F_T is differentiable everywhere except at $t = 0$. Therefore, T is of absolutely continuous type with pdf $f_T(t) = F_T'(t)$ for $t \ne 0$. At $t = 0$, we may assign any arbitrary nonnegative value to $f_T(0)$. Thus, a pdf of T is

$$f_T(t) = \begin{cases} \frac{e^{-\sqrt{t}}}{2\sqrt{t}}, & \text{if } t > 0 \\ 0, & \text{otherwise.} \end{cases}$$

Example 22. Let X be a random variable with pdf

$$f_X(x) = \begin{cases} \frac{|x|}{2}, & \text{if } -1 < x < 1 \\ \frac{x}{3}, & \text{if } 1 \le x < 2 \\ 0, & \text{otherwise} \end{cases}$$

and $T = X^2$. Find the distribution function of T.

We have $F_T(t) = \mathbf{P}\left(\left\{X^2 \le t\right\}\right)$ for $t \in \mathbb{R}$. Since $\mathbf{P}(\{X \in (-1, 2)\}) = 1$, we have $\mathbf{P}(\{T \in (0, 4)\}) = 1$. Therefore, $F_T(t) = \mathbf{P}(\{T \le t\}) = 0$ for $t < 0$, and $F_T(t) = \mathbf{P}(\{T \le t\}) = 1$ for $t \ge 4$. For $t \in [0, 4)$, we have

$$\begin{aligned} F_T(t) &= \mathbf{P}(\{-\sqrt{t} \le X \le \sqrt{t}\}) \\ &= \int_{-\sqrt{t}}^{\sqrt{t}} f_X(x)\mathrm{d}x \\ &= \begin{cases} \displaystyle\int_{-\sqrt{t}}^{\sqrt{t}} \frac{|x|}{2}\,\mathrm{d}x, & \text{if } 0 \le t < 1 \\ \displaystyle\int_{-1}^{1} \frac{|x|}{2}\,\mathrm{d}x + \int_1^{\sqrt{t}} \frac{x}{3}\,\mathrm{d}x, & \text{if } 1 \le t < 4. \end{cases} \end{aligned}$$

Therefore, the distribution function of T is

$$F_T(t) = \begin{cases} 0, & \text{if } t < 0 \\ \frac{t}{2}, & \text{if } 0 \leq t < 1 \\ \frac{t+2}{6}, & \text{if } 1 \leq t < 4 \\ 1, & \text{if } t \geq 4. \end{cases}$$

Clearly, F_T is differentiable everywhere except at points 0, 1 and 4. It follows that the random variable T is of absolutely continuous type with pdf

$$f_T(t) = \begin{cases} \frac{1}{2}, & \text{if } 0 < t < 1 \\ \frac{1}{6}, & \text{if } 1 < t < 4 \\ 0, & \text{otherwise.} \end{cases}$$

2.a. Change of variable (for discrete probability distributions): Let X be a random variable of discrete type with support S_X and pmf f_X. Define $Z = h(X)$. Then, Z is a random variable of discrete type with support $S_Z = \{h(x) : x \in S_X\}$ with pmf

$$f_Z(z) = \begin{cases} \sum_{x \in A_z} f_X(x), & \text{if } z \in S_Z \\ 0, & \text{otherwise,} \end{cases}$$

where $A_z = \{x \in S_X : h(x) = z\}$.
Corollary: Suppose that $h : \mathbb{R} \to \mathbb{R}$ is one-one with inverse function $h^{-1} : D \to \mathbb{R}$, where $D = \{h(x) : x \in \mathbb{R}\}$. Then, Z is a discrete type random variable with support $S_Z = \{h(x) : x \in S_X\}$ and pmf

$$f_Z(z) = \begin{cases} f_X(h^{-1}(z)), & \text{if } z \in S_Z \\ 0, & \text{otherwise.} \end{cases}$$

Example 23. Let X be a random variable with pmf

$$f_X(x) = \begin{cases} \frac{1}{7}, & \text{if } x \in \{-2, -1, 0, 1\} \\ \frac{3}{14}, & \text{if } x \in \{2, 3\} \\ 0, & \text{otherwise.} \end{cases}$$

Find the pmf and distribution function of $Z = X^2$.

Clearly, $S_X = \{-2, -1, 0, 1, 2, 3\}$ and $S_Z = \{0, 1, 4, 9\}$. Moreover,

$$\begin{aligned} &\mathbf{P}(\{Z = 0\}) = \mathbf{P}\left(\{X^2 = 0\}\right) = \mathbf{P}(\{X = 0\}) = \tfrac{1}{7}, \\ &\mathbf{P}(\{Z = 1\}) = \mathbf{P}\left(\{X^2 = 1\}\right) = \mathbf{P}(X \in \{-1, 1\}) = \tfrac{1}{7} + \tfrac{1}{7} = \tfrac{2}{7}, \\ &\mathbf{P}(\{Z = 4\}) = \mathbf{P}\left(\{X^2 = 4\}\right) = \mathbf{P}(X \in \{-2, 2\}) = \tfrac{1}{7} + \tfrac{3}{14} = \tfrac{5}{14} \text{ and} \\ &\mathbf{P}(\{Z = 9\}) = \mathbf{P}\left(\{X^2 = 9\}\right) = \mathbf{P}(X \in \{-3, 3\}) = 0 + \tfrac{3}{14} = \tfrac{3}{14}. \end{aligned}$$

Therefore, the pmf of Z is

$$f_Z(z) = \begin{cases} \frac{1}{7}, & \text{if } z = 0 \\ \frac{2}{7}, & \text{if } z = 1 \\ \frac{5}{14}, & \text{if } z = 4 \\ \frac{3}{14}, & \text{if } z = 9 \\ 0, & \text{otherwise,} \end{cases}$$

and the distribution function of Z is

$$F_Z(z) = \begin{cases} 0, & \text{if } z < 0 \\ \frac{1}{7}, & \text{if } 0 \le z < 1 \\ \frac{3}{7}, & \text{if } 1 \le z < 4 \\ \frac{11}{14}, & \text{if } 4 \le z < 9 \\ 1, & \text{if } z \ge 9. \end{cases}$$

Example 24. Let X be a random variable with pmf

$$f_X(x) = \begin{cases} \frac{|x|}{2550} & \text{if } x \in \{\pm 1, \pm 2, \ldots, \pm 50\} \\ 0, & \text{otherwise.} \end{cases}$$

Find the pmf and distribution function of $Z = |X|$.

We have $S_X = \{\pm 1, \pm 2, \ldots, \pm 50\}$ and $S_Z = \{1, 2, \ldots, 50\}$. Moreover, for $z \in S_Z$

$$\mathbf{P}(\{Z = z\}) = \mathbf{P}(\{|X| = z\}) = \mathbf{P}(\{X \in \{-z, z\}\}) = \frac{|-z|}{2550} + \frac{|z|}{2550} = \frac{z}{1275}.$$

Therefore, the pmf of Z is

$$f_Z(z) = \begin{cases} \frac{z}{1275}, & \text{if } z \in \{1, 2, \ldots, 50\} \\ 0, & \text{otherwise,} \end{cases}$$

and the distribution function of Z is

$$F_Z(z) = \begin{cases} 0, & \text{if } z < 1 \\ \frac{1}{1275}, & \text{if } 1 \le z < 2 \\ \frac{i(i+1)}{2550}, & \text{if } i \le z < i+1 \text{ for } i = 2, 3, \ldots, 49 \\ 1, & \text{if } z \ge 50. \end{cases}$$

Example 25. Let X be a random variable with pmf

$$f_X(x) = \begin{cases} \begin{pmatrix} n \\ x \end{pmatrix} p^x (1-p)^{n-x}, & \text{if } x \in \{0, 1, \ldots, n\} \\ 0, & \text{otherwise,} \end{cases}$$

where n is a positive integer and $p \in (0, 1)$. Find the pmf and distribution function of $Y = n - X$.

Note that $S_X = S_Y = \{0, 1, \ldots, n\}$. For $y \in S_Y$, we get

$$\mathbf{P}(\{Y = y\}) = \mathbf{P}(\{X = n - y\}) = \begin{pmatrix} n \\ n-y \end{pmatrix} p^{n-y}(1-p)^y = \begin{pmatrix} n \\ y \end{pmatrix} (1-p)^y p^{n-y}.$$

Thus, the pmf of Y is

$$f_Y(y) = \begin{cases} \begin{pmatrix} n \\ y \end{pmatrix} (1-p)^y p^{n-y}, & \text{if } y \in \{0, 1, \ldots, n\} \\ 0, & \text{otherwise,} \end{cases}$$

and the distribution function of Y is

$$F_Y(y) = \begin{cases} 0, & \text{if } y < 0 \\ p^n, & \text{if } 0 \leq y < 1 \\ \sum_{j=0}^{i} \begin{pmatrix} n \\ j \end{pmatrix} (1-p)^j p^{n-j}, & \text{if } i \leq y < i+1 \text{ for } i = 1, 2, \ldots, n-1 \\ 1, & \text{if } y \geq n. \end{cases}$$

2.b. Change of variable (for continuous probability distributions): Let X be a random variable of absolutely continuous type with pdf f_X and support S_X. Let $S_1, S_2, \ldots, S_k$ be open intervals in $\mathbb{R}$ such that $S_i \cap S_j = \emptyset$ if $i \neq j$ and $\cup_{i=1}^{k} S_i = S_X$.

Let $h : \mathbb{R} \to \mathbb{R}$ be a function such that on each S_i, the function $h : S_j \to \mathbb{R}$ is *strictly monotone* and *continuously differentiable* with inverse function (say, h_j^{-1}) for $j = 1, \ldots, k$. Let $h(S_j) = \{h(x) : x \in S_j\}$ so that $h(S_j)$ is an open interval in $\mathbb{R}$ for $j = 1, \ldots, k$.

Then, the random variable $T = h(X)$ is continuous with pdf

$$f_T(t) = \sum_{j=1}^{k} f_X\left(h_j^{-1}(t)\right) \left| \frac{\mathrm{d}}{\mathrm{d}t} h_j^{-1}(t) \right| I_{h(S_j)}(t).$$

Corollary: Let X be a continuous random variable with pdf f_X and support S_X. Suppose that S_X is a finite union of disjoint open intervals in $\mathbb{R}$, and let $h : \mathbb{R} \to \mathbb{R}$ be *differentiable and strictly monotone* on S_X (i.e., either $h'(x) < 0 \;\forall\, x \in S_X$ or $h'(x) > 0 \;\forall\, x \in S_X$). Let $S_T = \{h(x) : x \in S_X\}$. Then, $T = h(X)$ is a continuous random variable with pdf

$$f_T(t) = \begin{cases} f_X(h^{-1}(t)) \left| \frac{\mathrm{d}}{\mathrm{d}t} h^{-1}(t) \right|, & \text{if } t \in S_T \\ 0, & \text{otherwise.} \end{cases}$$

Example 26. Let X be random variable with pdf

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

and $T = X^2$. Find the distribution function of T, and hence find its pdf.

We have $S_X = S_T = (0, \infty)$. Also, $h(x) = x^2$ for $x \in S_X$ is strictly increasing on S_X with inverse function $h^{-1}(x) = \sqrt{x}$ for $x \in S_T$. It follows that $T = X^2$ has pdf

$$f_T(t) = \begin{cases} f_X(\sqrt{t}) \left| \frac{d}{dt}(\sqrt{t}) \right|, & \text{if } t > 0 \\ 0, & \text{otherwise} \end{cases}$$

$$= \begin{cases} \frac{e^{-\sqrt{t}}}{2\sqrt{t}}, & \text{if } t > 0 \\ 0, & \text{otherwise.} \end{cases}$$

Example 27. Let X be a random variable with pdf

$$f_X(x) = \begin{cases} \frac{|x|}{2}, & \text{if } -1 < x < 1 \\ \frac{x}{3}, & \text{if } 1 \leq x < 2 \\ 0, & \text{otherwise} \end{cases}$$

and $T = X^2$. Find the distribution function of T, and hence find its pdf.

We have $S_X = (-1, 0) \cup (0, 2) = S_1 \cup S_2$, say. Now, $h(x) = x^2$ is strictly decreasing in $S_1 = (-1, 0)$ with inverse function $h_1^{-1}(t) = -\sqrt{t}$; and $h(x) = x^2$ is strictly increasing in $S_2 = (0, 2)$ with inverse function $h_2^{-1}(t) = \sqrt{t}$. Note that $h(S_1) = (0, 1)$ and $h(S_2) = (0, 4)$. It now follows that $T = X^2$ has pdf

$$f_T(t) = f_X(-\sqrt{t}) \left| \frac{\mathrm{d}}{\mathrm{d}t}(-\sqrt{t}) \right| I_{(0,1)}(t) + f_X(\sqrt{t}) \left| \frac{\mathrm{d}}{\mathrm{d}t}(\sqrt{t}) \right| I_{(0,4)}(t)$$

$$= \begin{cases} \frac{1}{2}, & \text{if } 0 < t < 1 \\ \frac{1}{6}, & \text{if } 1 < t < 4 \\ 0, & \text{otherwise.} \end{cases}$$

Example 28. Let X be a random variable of absolutely continuous type with pdf

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

and $T = [X]$, where $[x]$ denotes the largest integer not exceeding x for $x \in \mathbb{R}$. Find its pmf.

Note that $S_X = (0, \infty)$. Since $\mathbf{P}(\{X \in S_X\}) = 1$, we have $\mathbf{P}(T \in \{0, 1, 2, \ldots\}) = 1$. Also, for $i \in \{0, 1, 2, \ldots\}$

$$\begin{aligned}
\mathbf{P}(\{T = i\}) &= \mathbf{P}(\{i \leq X < i + 1\}) \\
&= \int_i^{i+1} f_X(x)\mathrm{d}x \\
&= \int_i^{i+1} e^{-x}\,\mathrm{d}x \\
&= \left(1 - e^{-1}\right) e^{-i}.
\end{aligned}$$

Consequently, the random variable T is of discrete type with support $S_T = \{0, 1, 2, \ldots\}$ with pmf

$$f_T(t) = \mathbf{P}(\{T = t\}) = \begin{cases} \left(1 - e^{-1}\right) e^{-t}, & \text{if } t \in \{0, 1, 2, \ldots\} \\ 0, & \text{otherwise.} \end{cases}$$

Remark 29. This example illustrates that in general, a function of an absolutely continuous type random variable may not be of absolutely continuous type.