On Continuous Distributions

Tristan Chaang

January 7, 2022

Contents

1 PDFs and CDFs

Suppose we want to randomly choose a number in the interval $[0,100]$ so that 'every number is equally likely to be chosen'. The natural question to ask is, what do we mean exactly by equally likely? The set $[0,100]$ has infinitely many elements, so the probability of choosing an *exact given value* is zero. How do we resolve this? The answer is by using arbitrary intervals: The probability of choosing a number in the interval $[a,b]$ is $(b-a)/100$ for any $0 \le a \le b \le 100$.

The distribution above is called the *uniform distribution on* $[0,100]$, denoted as $\mathcal{U}_{[0,100]}$. However, that is not the only distribution out there. We can have distribution on other intervals, and even if we are just taking a distribution on $[0,100]$, there can be distributions where it is more likely to choose one number than the other, e.g. choosing a number nearer to 0 being more likely than choosing one near 100.

The way we describe a distribution is by using *probability density functions (pdf)*. For a continuous random variable X with distribution $\mathcal{D}$ (written as $X \sim \mathcal{D}$), the pdf $f_X(x)$ of X is the function such that

$$\mathcal{P}(a \le X \le b) = \int_a^b f_X(x)\ dx$$

for any $a \le b$. For example, the pdf of $X \sim \mathcal{U}_{[0,100]}$ is

$$f_X(x) = \begin{cases} 1/100 & \text{if } 0 \le x \le 100; \\ 0 & \text{otherwise.} \end{cases}$$

There are a few properties to notice. Since $\mathcal{P}(X \in \mathbb{R}) = 1$ and $\mathcal{P}(a \le X \le b) \ge 0$,

$$\int_{\forall x} f_X(x)\ dx = 1 \qquad \text{and} \qquad \forall x \in \mathbb{R} : f_X(x) \ge 0$$

for any pdf[1] $f_X(x)$.

[1] We assume these pdfs are Riemann integrable. You can ignore this if you haven't heard of this. It roughly means it cannot be weird functions such as $f = 0$ at rational numbers and $f = 1$ otherwise. We cannot integrate it the typical way.

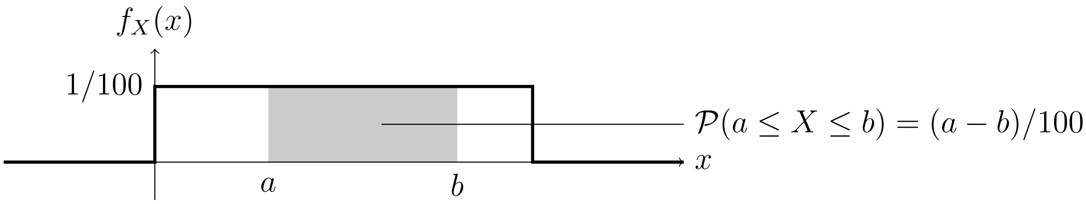

Figure 1: Uniform distribution $\mathcal{U}_{[0,100]}$

The *cumulative distribution function (cdf)*, or distribution function in short, of X is

$$F_X(x) = \mathcal{P}(X \le x) = \int_{-\infty}^{x} f_X(t)\ dt$$

Notice we used t in the integrand instead of x because we have already used the symbol x in the bounds. The cdf for $X \sim \mathcal{U}_{[0,100]}$ is thus

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0; \\ x/100 & \text{if } 0 \le x \le 100; \\ 1 & \text{if } x > 100. \end{cases}$$

A simple corollary is that $F'_X(x) = f_X(x)$. Using this corollary, we can find the pdf of $g(X)$ where g is a function taking X as input:

$$f_{g(X)}(x) = \frac{d}{dx}\mathcal{P}(g(X) \le x).$$

Writing the pdf of $3X+1$ in terms of the pdf of X

$$f_{3X+1}(x) = \frac{d}{dx}P\left(3X+1 \le x\right) = \frac{d}{dx}P\left(X \le \frac{x-1}{3}\right) = \frac{1}{3}f_X\left(\frac{x-1}{3}\right).$$

Writing the pdf of X^2 in terms of the pdf of X

$f_{X^2}(x) = \dfrac{d}{dx}\mathcal{P}(X^2 \le x)$. We have two cases:

$$\begin{aligned} x < 0: \qquad & \frac{d}{dx}\mathcal{P}(X^2 \le x) = \frac{d}{dx}(0) = 0 \\ x \ge 0: \qquad & \frac{d}{dx}\mathcal{P}(X^2 \le x) = \frac{d}{dx}\mathcal{P}(-\sqrt{x} \le X \le \sqrt{x}) \\ & = \frac{d}{dx}(F_X(\sqrt{x}) - F_X(-\sqrt{x})) \\ & = \frac{f_X(\sqrt{x})}{2\sqrt{x}} + \frac{f_X(-\sqrt{x})}{2\sqrt{x}}. \end{aligned}$$

In the last line we used the chain rule. Therefore,

$$f_{X^2}(x) = \begin{cases} 0 & \text{if } x < 0; \\ \dfrac{f_X(x) + f_X(-x)}{2\sqrt{x}} & \text{if } x \ge 0. \end{cases}$$

The *expectation*, or mean, of a continuous random variable X is defined as

$$E(X) = \int_{\forall x} x f_X(x)\ dx$$

The Law of the Unconscious Statistician (LOTUS)

$$E(g(X)) = \int_{-\infty}^{\infty} x f_{g(X)}(x)\ dx = \int_{-\infty}^{\infty} g(x) f_X(x)\ dx$$

The proof of this is complicated. I will write a new article about this.

and its *variance* and *standard deviation* are defined as

$$Var(X) = E(X^2) - E(X)^2 \qquad \text{and} \qquad Sd(X) = \sqrt{Var(X)}$$

2 Multiple Continuous Random Variables

If we have two variables X and Y, we can construct a pdf with two inputs $f_{X,Y}(x,y)$ so that

$$\mathcal{P}((X,Y) \in A) = \iint_A f_{X,Y}(x,y)\ dx\ dy \qquad \text{for any closed region } A.$$

This means, instead of calculating the probability by finding the area under a curve of a 2D pdf graph, we calculate the probability by finding the *volume* under a surface of a 3D pdf graph. Such a pdf must also satisfy similar properties such as the total volume under the surface is 1 and the pdf is always nonnegative.

If $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$ for all x, y, we say that the random variables X and Y are *independent*. We will normally deal with independent variables.

If $X \sim \mathcal{D}_1$ and $Y \sim \mathcal{D}_2$ are independent variables, we would like to find the pdf of $Z = X + Y$:

$$\begin{aligned}
f_Z(z) &= \frac{d}{dz}\mathcal{P}(Y \leq z - X) \\
&= \frac{d}{dz}\int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f_{X,Y}(x,y)\ dy\ dx \\
&= \int_{-\infty}^{\infty} f_X(x) \cdot \frac{d}{dz}\left(\int_{-\infty}^{z-x} f_Y(y)\ dy\right)\ dx \\
&= \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(z-x)\ dx
\end{aligned}$$

This will be important later on. Also, for some positive constant c, the pdf of $Z = cX$ is

$$f_Z(z) = \frac{d}{dz}\mathcal{P}\left(Z \leq z\right) = \frac{d}{dz}\mathcal{P}\left(X \leq \frac{z}{c}\right) = \frac{1}{c}\ f_X\left(\frac{z}{c}\right)$$

Finally, let's find the pdf of $Z = X/Y$.

$$\begin{aligned}
f_Z(z) &= \frac{d}{dz}\mathcal{P}(X \leq Yz) \\
&= \frac{d}{dz}\int_{-\infty}^{\infty}\int_{-\infty}^{yz} f_{X,Y}(x,y)\ dy\ dx \\
&= \int_{-\infty}^{\infty} f_Y(y) \cdot \frac{d}{dz}\left(\int_{-\infty}^{yz} f_X(x)\ dx\right)\ dy \\
&= \int_{-\infty}^{\infty} y \cdot f_Y(y) f_X(yz)\ dy
\end{aligned}$$

3 Normal Distribution

We come to our first famous distribution: the normal distribution. Any continuous random variable X having the pdf in the form[2]

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \qquad \text{for some } \mu \text{ and positive } \sigma$$

is said to belong to a normal distribution, denoted as $X \sim \mathcal{N}(\mu, \sigma^2)$. Figure 2 below shows the pdf of $\mathcal{N}(\mu, \sigma^2)$. Even though this distribution is determined by the mean μ and variance σ^2, we can always apply a transformation to the pdf of $\mathcal{N}(\mu, \sigma^2)$ by denoting $Z = (X - \mu)/\sigma$, transforming $X \sim \mathcal{N}(\mu, \sigma^2)$ into $Z \sim \mathcal{N}(0, 1)$, as shown in Figure 2.

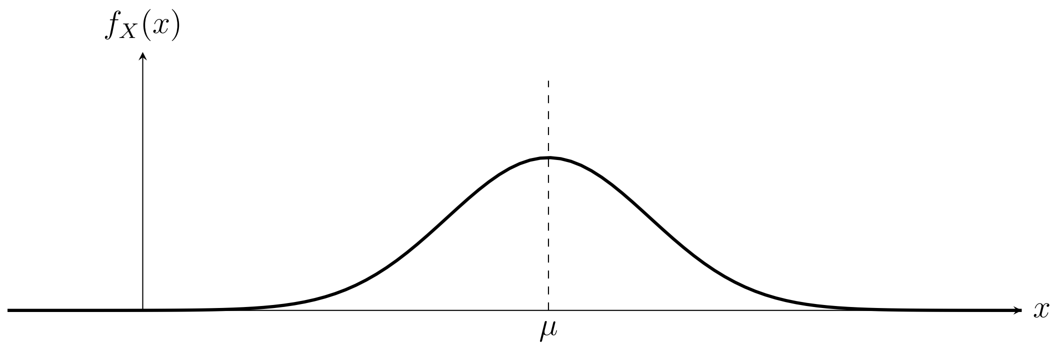

Figure 2: Normal distribution $\mathcal{N}(\mu, \sigma^2)$

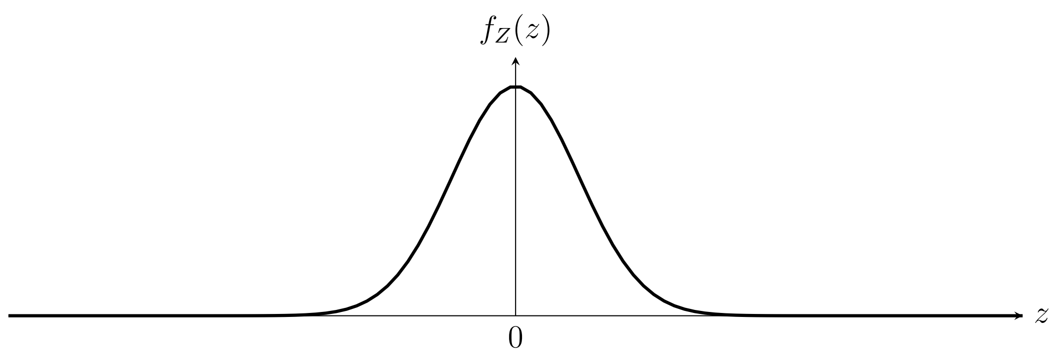

Figure 3: Standard normal distribution $\mathcal{N}(1, 0)$

This transformation is called *standardisation*. The resultant pdf is much simpler, given by

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$$

3.1 Properties of the Normal Distribution

These properties are fun to prove using what we have learnt so far:

- The distribution is symmetric about μ.
- The mean of $X \sim \mathcal{N}(\mu, \sigma^2)$ is μ.
- The variance of $X \sim \mathcal{N}(\mu, \sigma^2)$ is σ^2.

[2]Here $\exp(x)$ means e^x.

3.2 Deriving the pdf for $\mathcal{N}(0,1)$ from $\mathcal{B}(n,p)$

One of the reasons we study normal distributions is due to the binomial distribution $\mathcal{B}(n,p)$. Recall the binomial distribution formula

$$B_{n,p}(x) = \binom{n}{x} p^x q^{n-x} \qquad \text{where } x = 0, \cdots, n \text{ and } q = 1-p.$$

If we plot the graphs of $B_{n,p}(x)$ with fixed p and varying n, we get the following:

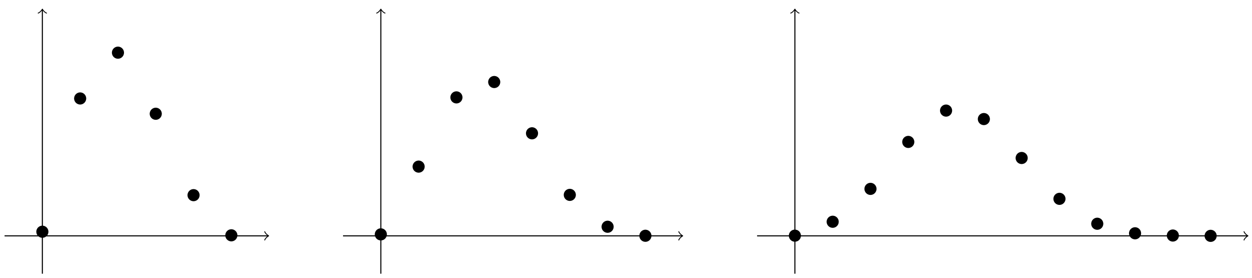

Figure 4: Binomial distributions $\mathcal{B}(5, \frac{2}{5}), \mathcal{B}(7, \frac{2}{5}), \mathcal{B}(11, \frac{2}{5})$.

As we see above, the graph gets wider as n increases because the domain enlarges. The height of the graph is also falling because as the domain enlarges, the sum of the probabilities must still be equal to 1. We know that the mean and standard deviation of $\mathcal{B}(n,p)$ are $\mu = np$ and $\sigma = \sqrt{npq}$ respectively. So, in order to get a converging sequence of graphs, let's shift the distribution of $\mathcal{B}(n,p)$ to the left by μ, shrink it horizontally by σ, and then stretch it vertically by σ to maintain the sum-equals-one property:

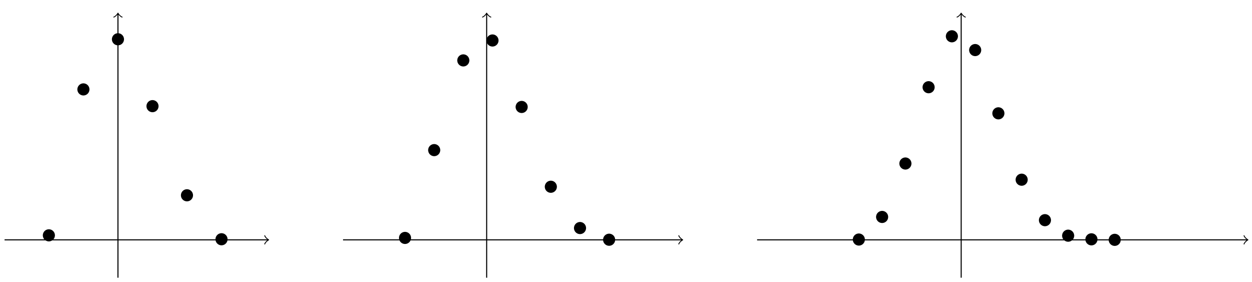

Figure 5: The graphs of $y = \sigma B_{n,p}(\sigma x + \mu)$ for $n = 5, 7, 11$.

Now we're converging (or at least we seem to)! Suppose they converge to $N(x)$. For fixed n, the point $(x, \sigma B_{n,p}(\sigma x + \mu))$ on the adjusted Binomial graph corresponds to $(u, B_{n,p}(u)) = (\sigma x + \mu, B_{n,p}(\sigma x + \mu))$ on the original Binomial graph. Therefore, the gradient of the line connecting $(u, B_{n,p}(u))$ and $(u+1, B_{n,p}(u+1))$ is

$$\frac{B_{n,p}(u+1) - B_{n,p}(u)}{(u+1) - u} = \binom{n}{u+1} p^{u+1} q^{n-u-1} - \binom{n}{u} p^u q^{n-u}$$

This line, after transforming it onto the adjust Binomial graph, is shrinked horizontally by σ but stretched vertically by σ, so its gradient is σ^2 times larger than the expression above, i.e.

$$N'(x) = \lim_{n \to \infty} \sigma^2 \left[\binom{n}{u+1} p^{u+1} q^{n-u-1} - \binom{n}{u} p^u q^{n-u} \right]$$

This is hard to handle, so let's divide this by $N(x) = \lim_{n\to\infty} \sigma B_{n,p}(\sigma x + \mu) = \sigma B_{n,p}(u)$.

$$
\begin{aligned}
\therefore \frac{N'(x)}{N(x)} &= \lim_{n\to\infty} \sigma^2 \cdot \frac{\binom{n}{u+1}p^{u+1}q^{n-u-1} - \binom{n}{u}p^u q^{n-u}}{\sigma\binom{n}{u}p^u q^{n-u}} \\
&= \lim_{n\to\infty} \sqrt{npq} \cdot \frac{np - u - q}{q(u+1)} \\
&= \lim_{n\to\infty} \sqrt{npq} \cdot \frac{np - x\sqrt{npq} - np - q}{qx\sqrt{npq} + npq + q} \\
&= \lim_{n\to\infty} \frac{-(pqx)n - (q\sqrt{pq})n^{1/2}}{(pq)n + (qx\sqrt{pq})n^{1/2} + q} \\
&= -x
\end{aligned}
$$

but $\dfrac{N'(x)}{N(x)} = \dfrac{d}{dx}\ln(N(x))$, so

$$
\begin{aligned}
\ln(N(x)) &= -x^2/2 + c \\
N(x) &= C\exp(-x^2/2)
\end{aligned}
$$

where C is a constant. Now we just have to solve for C in

$$\int_{-\infty}^{\infty} C\exp\left(-\frac{x^2}{2}\right)\, dx = 1.$$

However, $\int_{-\infty}^{\infty}\exp(-x^2)\, dx = \sqrt{\pi}$ is a classic calculus problem (perhaps I will write about it). Therefore, in this case C must be $1/\sqrt{2\pi}$. In conclusion,

$$N(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)$$

is the desired curve.