On Continuous Distributions

Tristan Chaang

February 8, 2022

In this article I will write about continuous distributions, including their properties and their derivations, without using advanced concepts such as measure theory.

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 $0 \leq a \leq b \leq 100$.

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

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

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

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

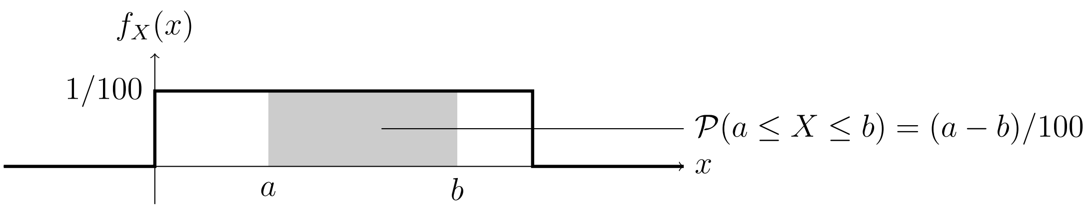

Figure 1: Uniform distribution $\mathcal{U}_{[0,100]}$

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

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

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

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

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

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

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

[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.

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:

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

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

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

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

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

$$x \leq 0: \qquad \frac{d}{dx}\mathcal{P}(X^2 \leq x) = \frac{d}{dx}(0) = 0$$

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

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

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

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 $Z = X + Y$:

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

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 $Z = X/Y$.

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

If X and Y are two random variables (might not be independent!), we have:

Linearity of Expectation

$$\begin{aligned} E(X+Y) &= \iint_{\forall x,y} (x+y) f_{X,Y}(x,y)\ dy\ dx \\ &= \iint_{\forall x,y} x f_{X,Y}(x,y)\ dy\ dx + \iint_{\forall x,y} y f_{X,Y}(x,y)\ dy\ dx \\ &= \int_{\forall x} x\left(\int_{\forall y} f_{X,Y}(x,y)\ dy\right)\ dx + \int_{\forall y} y\left(\int_{\forall x} f_{X,Y}(x,y)\ dx\right)\ dy \\ &= \int_{\forall x} x f_X(x)\ dx + \int_{\forall y} y f_Y(y)\ dy \\ &= E(X) + E(Y). \\ \text{and } E(cX) &= \int_{\forall x} cx f_X(x)\ dx = c\int_{\forall x} x f_X(x)\ dx = cE(X). \end{aligned}$$

If X and Y are two **independent** variables,

Multiplicity of Expectation for Independent Variables

$$E(X)E(Y) = \int_{\forall x} x f_X(x)\ dx \int_{\forall y} y f_Y(y)\ dy = \iint_{\forall x,y} (xy) f_{X,Y}(x,y)\ dx\ dy = E(XY).$$

(Semi)linearity of Variance for Independent Variables

$$\begin{aligned} Var(X+Y) &= E((X+Y)^2) - E(X+Y)^2 \\ &= E(X^2) + 2E(XY) + E(Y^2) - E(X)^2 - 2E(X)E(Y) - E(Y)^2 \\ &= Var(X) + Var(Y). \\ \text{and } Var(cX) &= E(c^2X^2) - E(cX)^2 = c^2 Var(X). \end{aligned}$$

Proving that X and $X+Y$ are Not Independent if X and Y are Independent

$$\begin{aligned} f_{X,X+Y}(x,y) &= \lim_{\delta x \to 0} \lim_{\delta y \to 0} \frac{\mathcal{P}(x \le X \le x + \delta x \text{ and } y - X \le Y \le y - X + \delta y)}{\delta x \delta y} \\ &= \lim_{\delta x \to 0} \lim_{\delta y \to 0} \frac{\int_x^{x+\delta x} f_X(u) \left[\int_{y-u}^{y-u+\delta y} f_Y(v)\ dv\right]\ du}{\delta x \delta y} \\ &= \lim_{\delta x \to 0} \frac{\int_x^{x+\delta x} f_X(u) f_Y(y-u)\ du}{\delta x} \\ &= f_X(x) f_Y(y-x) \neq f_X(x) f_Y(y) \end{aligned}$$

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.

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

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

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

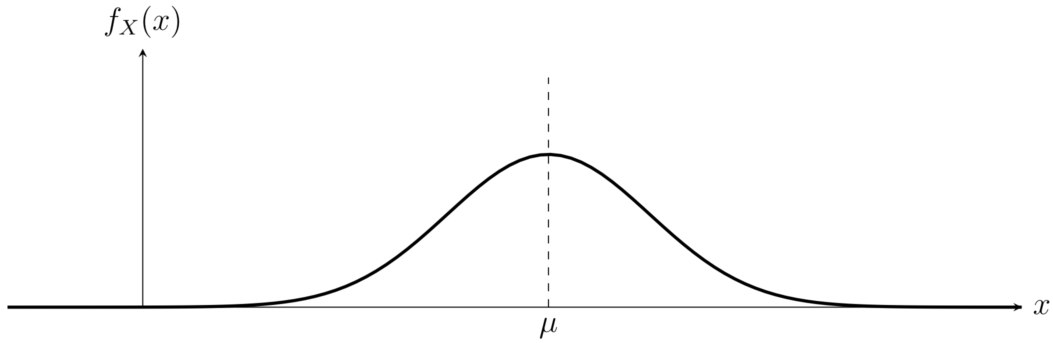

Figure 2: Normal distribution $\mathcal{N}(\mu, \sigma^2)$

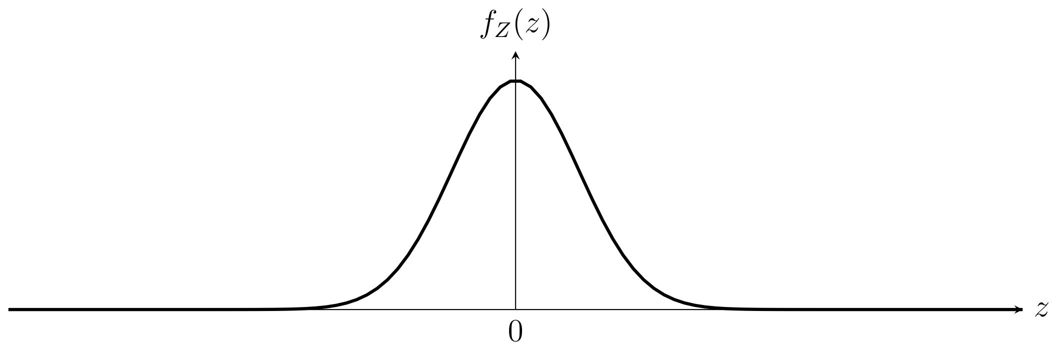

Figure 3: Standard normal distribution $\mathcal{N}(1, 0)$

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.

The Sum of Two Independent Normal Variables is Normal

Say $X \sim \mathcal{N}(\mu_1, \sigma_1^2), Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$, then the pdf of $Z = X + Y$ is

$$\begin{aligned}
f_Z(z) &= \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\ dx \\
&= \int_{-\infty}^{\infty} \frac{\exp(-\frac{1}{2}(\frac{x-\mu_1}{\sigma_1})^2)}{\sigma_1\sqrt{2\pi}} \cdot \frac{\exp(-\frac{1}{2}(\frac{z-x-\mu_2}{\sigma_2})^2)}{\sigma_2\sqrt{2\pi}}\ dx \\
&= \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - \frac{1}{2}\left(\frac{z-x-\mu_2}{\sigma_2}\right)^2\right]\ dx \\
&= (\text{constant}) \int_{-\infty}^{\infty} \exp\left[-A(x + \text{some linear in } z)^2 + \text{some quadratic in } z\right]\ dx \\
&= (\text{constant}) e^{\text{quadratic in } z} \int_{-\infty}^{\infty} e^{-A(x + \text{ some linear in } z)^2}\ dx \\
&= (\text{constant}) e^{\text{quadratic in } z}
\end{aligned}$$

Therefore Z is normal. We do not have to specifically find out what the constants are because we know $E(Z) = E(X) + E(Y) = \mu_1 + \mu_2$ and $Var(Z) = Var(X) + Var(Y) = \sigma_1^2 + \sigma_2^2$. Thus $X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

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:

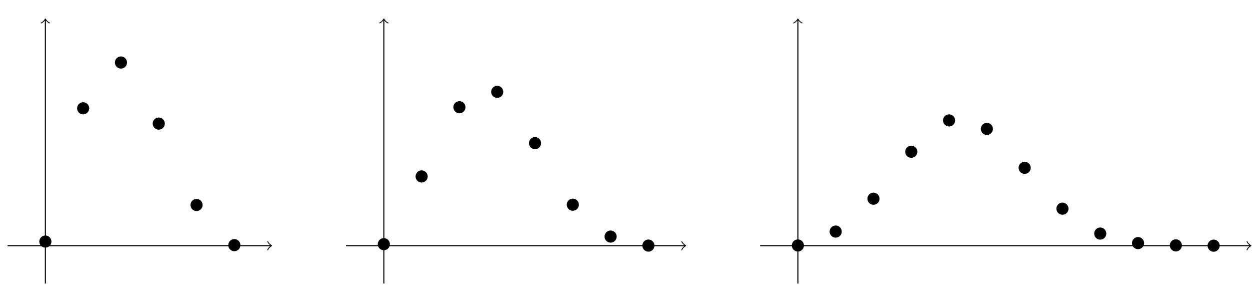

Figure 4: Binomial distributions $\mathcal{B}(5, \frac{2}{5}), \mathcal{B}(7, \frac{2}{5}), \mathcal{B}(11, \frac{2}{5})$.

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:

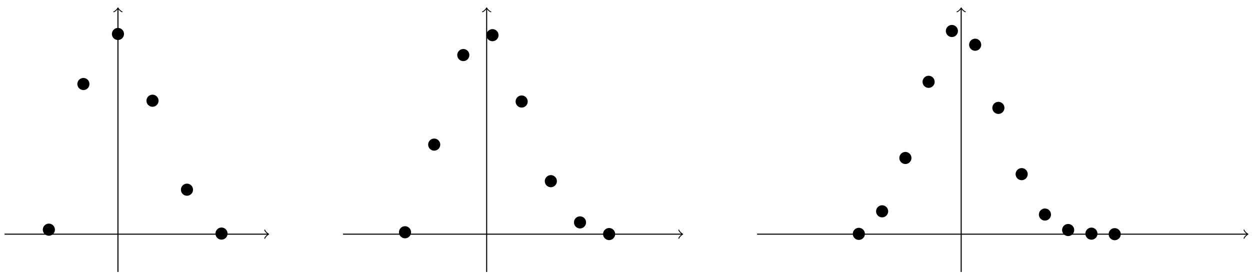

Figure 5: The graphs of $y = \sigma B_{n,p}(\sigma x + \mu)$ for $n = 5, 7, 11$.

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 $N(x) = \lim\limits_{n\to\infty} \sigma B_{n,p}(\sigma x + \mu) = \sigma B_{n,p}(u)$.

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

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.

4 Sample Statistics

Suppose the variables $X_1, \cdots, X_n$ are independent but all have the same distribution $\mathcal{D}$. We say that $X_1, \cdots, X_n$ are *independent observations* of $X \sim \mathcal{D}$. A set of independent observations $\{X_1, \cdots, X_n\}$ is a *sample*. Given a sample S, we can construct any function taking inputs from S. These functions are called *statistics*. For example,

Sample mean: $$\overline{X} = \frac{X_1 + \cdots + X_n}{n}$$

Sample variance: $$s^2 = \frac{(X_1 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2}{n-1}$$

are commonly-used statistics. From section 2, we know that we can find the distribution of $\overline{X}$ by finding its pdf.

Suppose $\mathcal{D} = \mathcal{N}(\mu, \sigma^2)$. Since the sum of two normal variables is normal, by induction, $X_1 + \cdots + X_n$ is normal. By the linearity of expectation and variance,

$$E(\overline{X}) = \frac{1}{n} \sum E(X) = \frac{1}{n} \cdot n\mu = \mu.$$
$$Var(\overline{X}) = \frac{1}{n^2} \sum Var(X) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}.$$

Therefore $\overline{X} \sim \mathcal{N}(\mu, \sigma^2/n)$.

Central Limit Theorem (CLT)

Even when X is not normal, if $X_1, \cdots, X_n$ are independent observations of X then the distribution of $\overline{X}$ converges to $\mathcal{N}(\mu, \sigma^2/n)$ when $n \to \infty$.

This is also a hard theorem to prove, I will write about it in the future.

Suppose we want to estimate the mean μ and variance σ^2 of a population. If we were given access to every possible data value of the population, then we can get μ and σ^2 easily. However, when we only take a sample, we can only evaluate $\overline{X}$ and s^2. First of all, why is the denominator of s^2, the sample variance, $n-1$ instead of n, the number of values taken? Second of all, how are we sure that $\overline{X}$ and s^2 give good estimates of μ and σ^2? Third of all, how confident are we in these estimates?

We answer the first and second question together. We say that a statistic T is an *unbiased estimate* of a value θ if $E(T) = \theta$. For example, $\overline{X}$ is an unbiased estimate of μ because $E(\overline{X}) = \mu$ as shown above. Since the formula of $\overline{X}$ looks exactly the same as that for μ, one would expect that $S^2 = \sum (X_i - \overline{X})^2/n$

is also an unbiased estimate for σ^2. However, since $\sigma^2 = E(X^2) - E(X)^2 = E(X^2) - \mu^2$,

$$\begin{aligned}
E(S^2) &= E\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) - E\left(\left(\frac{\sum_{i=1}^{n} X_i}{n}\right)^2\right) \\
&= E(X^2) - \frac{1}{n^2}E\left(\sum_{i=1}^{n} X_i^2 + 2\sum_{i<j} X_i X_j\right) \\
&= E(X^2) - \frac{1}{n}E(X^2) - \frac{2}{n^2}\cdot\binom{n}{2}\cdot E(X)E(X) \\
&= \left(1 - \frac{1}{n}\right)(\sigma^2 + \mu^2) - \frac{n-1}{n}\mu^2 \\
&= \frac{n-1}{n}\sigma^2 \neq \sigma^2
\end{aligned}$$

suggests otherwise. However, in light of the $(n-1)/n$ term, we see that by replacing the denominator of S to $n-1$, we have $E(s^2) = \sigma^2$ instead. Therefore s^2 is an unbiased estimator for σ^2, while S^2 is not.

To answer the third question, we previously proved that $\overline{X} \sim \mathcal{N}(\mu, \sigma^2/n)$, so for large values of n, the standard deviation is very small and hence we are very confident that $\overline{X}$ is close to μ. The problem is, how confident is confident, especially when n is small? If we want to find an interval so that $\overline{X}$ has a 0.9 probability to lie in it, we just standardise and solve

$$\mathcal{P}\left(-z_{0.95} < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < z_{0.95}\right) = 0.9$$

$$\mathcal{P}\left(\overline{X} - z_{0.95}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{0.95}\frac{\sigma}{\sqrt{n}}\right) = 0.9$$

where $z_{0.95} = 1.645$ is the value of z such that $\mathcal{P}(Z < z) = 0.95$ whereby Z is normal. This is mathematically correct, but we have no idea what the value of σ is in the first place. We will resolve this problem in the t-distribution section.

5 Chi-Squared Distribution

In section 1, we found that the pdf of $Z = X^2$ is

$$f_Z(z) = \begin{cases} 0 & \text{if } z \le 0; \\ \dfrac{f_X(\sqrt{z}) + f_X(-\sqrt{z})}{2\sqrt{z}} & \text{if } z > 0. \end{cases}$$

In this section, we will study the case when $X \sim \mathcal{N}(0,1)$, so $f_X(x) = \exp(-x^2/2)/\sqrt{2\pi}$, giving

$$f_Z(z) = \begin{cases} 0 & \text{if } z \le 0; \\ \dfrac{\exp(-z/2)}{\sqrt{2\pi z}} & \text{if } z > 0. \end{cases}$$

Therefore, the distribution of X^2 looks like Figure 6 (asymptote at $x = 0$):

This is called the chi-squared distribution with 1 degree of freedom. In general, the *chi-squared distribution with n degrees of freedom*, denoted as χ^2_n, is the distribution of $Q_n = X_1^2 + \cdots + X_n^2$ where $X_1, \cdots, X_n$ are independent observations of $X \sim \mathcal{N}(0,1)$.

Firstly, the sum of two independent chi-squared distributed variables with a and b degrees of freedom respectively is chi-squared distributed with $a+b$ degrees of freedom. This is simply because $(X_1^2 + \cdots + X_a^2) + (Y_1^2 + \cdots + Y_b^2)$ is just the sum of squares of $a+b$ standard normal variables.

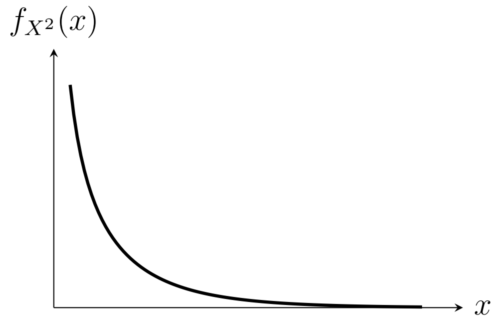

Figure 6: Distribution of X^2 if $X \sim \mathcal{N}(0,1)$

$\dfrac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$ **if** $X_1, \cdots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ **are** n **independent observations**.

Standardise $Z_i = (X_i - \mu)/\sigma$. $(n-1)s^2/\sigma^2$ can be written as a sum of $n-1$ squares:

$$\begin{aligned}
\frac{(n-1)s^2}{\sigma^2} &= \sum_{1 \le i \le n} Z_i^2 - \frac{(Z_1 + \cdots + Z_n)^2}{n} \\
&= \frac{n-1}{n} \sum_{1 \le i \le n} Z_i^2 - \frac{2}{n} \sum_{1 \le i < j \le n} Z_i Z_j \\
&= \left(\sqrt{\frac{n-1}{n}} Z_n - \frac{1}{\sqrt{n(n-1)}} Z_{n-1} - \cdots - \frac{1}{\sqrt{n(n-1)}} Z_1 \right)^2 \\
&\quad + \frac{n-2}{n-1} \sum_{1 \le i \le n-1} Z_i^2 - \frac{2}{n-1} \sum_{1 \le i < j \le n-1} Z_i Z_j \\
&= \sum_{k=2}^{n} \left(\sqrt{\frac{k-1}{k}} Z_k - \frac{1}{\sqrt{k(k-1)}} Z_{k-1} - \cdots - \frac{1}{\sqrt{k(k-1)}} Z_1 \right)^2 \text{ by induction}
\end{aligned}$$

It is a fun exercise to prove that each squared term at the end is standard normal. However, it is extremely difficult to prove that they are mutually independent using the tools we've learnt so far (we can proceed something similar to how we proved X and $X+Y$ are dependent in section 4). We will accept the fact that the above expression is a sum of $n-1$ independent standard normal variables, and hence has the chi-squared distribution with $n-1$ degrees of freedom.

Since its pdf for $x \le 0$ is obviously 0, we have

$$f_{Q_n}(x) = \begin{cases} 0 & \text{if } x \le 0; \\ g_n(x) & \text{if } x > 0 \end{cases}$$

for some function $g_n(x)$. Henceforth we just find $g_n(x)$.

5.1 Deriving the pdf for the Chi-Squared Distribution

Finding the pdf for χ_2^2

From above, $g_1(x) = \dfrac{\exp(-x/2)}{\sqrt{2\pi x}}$. Let's find the pdf for $Q_2 \sim \chi_2^2$ for $x > 0$, which is

$$\begin{aligned} g_2(x) &= \int_{-\infty}^{\infty} f_{X_1^2}(t) f_{X_2^2}(x-t)\ dt \\ &= \int_0^x g_1(t) g_1(x-t)\ dt \\ &= \int_0^x \frac{\exp(-t/2)}{\sqrt{2\pi t}} \cdot \frac{\exp(-(x-t)/2)}{\sqrt{2\pi(x-t)}}\ dt \\ &= \frac{\exp(-x/2)}{2\pi} \int_0^x \frac{1}{\sqrt{t(x-t)}}\ dt \\ &= \frac{\exp(-x/2)}{2\pi} \left[\arcsin\left(\frac{2t}{x} - 1\right)\right]_{t=0}^{x} \\ &= \frac{1}{2}\exp(-x/2). \end{aligned}$$

The distribution for χ_2^2 looks like Figure 8:

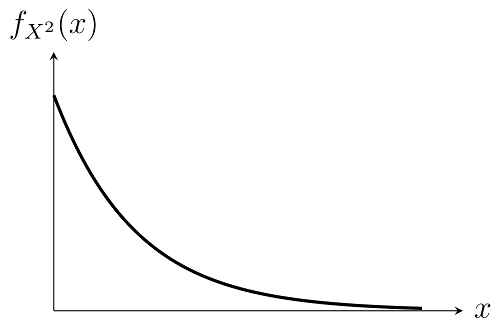

Figure 7: Chi-Squared Distribution χ_2^2

Now that we have g_1 and g_2, we can construct a recursive relation of g_n. Since $X_1^2 + \cdots + X_n^2 = (X_1^2 + \cdots + X_{n-2}^2) + (X_{n-1}^2 + X_n^2)$ and $X_1^2 + \cdots + X_{n-2}^2$ and $X_{n-1}^2 + X_n^2$ are independent, the random variable Q_n is just the sum of two independent variables Q_2 and Q_{n-2}. Hence

$$\begin{aligned} g_n(x) &= \int_{-\infty}^{\infty} f_{Q_{n-2}}(t) \cdot f_{Q_2}(x-t)\ dt \\ &= \frac{1}{2}\int_0^x g_{n-2}(t) \exp\left(-\frac{x-t}{2}\right)\ dt \end{aligned}$$

If we write $g_n(x) = 2^{-n/2} h_n(x) \exp(-x/2)$, the above expression simplifies to

$$h_n(x) = \int_0^x h_{n-2}(t)\ dt$$

Finding the pdf for Chi-Squared Distributions with Even Degree of Freedom

Since $h_2(x) = 1$, we can prove by induction that $h_{2k} = \dfrac{x^{k-1}}{(k-1)!}$, because

$$h_{2k} = \frac{x^{k-1}}{(k-1)!} \qquad \Rightarrow \qquad h_{2(k+1)} = \int_0^x \frac{t^{k-1}}{(k-1)!}\,dt = \frac{x^k}{k!}.$$

In conclusion, we have that the pdf for χ^2_{2k} is

$$\begin{cases} 0 & \text{if } x \le 0; \\ \dfrac{1}{2^k(k-1)!}x^{k-1}e^{-x/2} & \text{if } x > 0. \end{cases}$$

Here, we find the pattern by integrating $h_i(x)$ many times. A similar method can be used to find the pdf of those with odd degrees of freedom.

Finding the pdf for Chi-Squared Distributions with Odd Degree of Freedom

Since $h_1(x) = \dfrac{x^{-0.5}}{\sqrt{\pi}}$, we can prove by induction that $h_{2k+1} = \dfrac{2^{2k}k!x^{k-0.5}}{(2k)!\sqrt{\pi}}$, because

$$h_{2k+1} = \frac{2^{2k}k!x^{k-0.5}}{(2k)!\sqrt{\pi}} \quad \Rightarrow \quad h_{2k+3} = \int_0^x \frac{2^{2k}k!t^{k-0.5}}{(2k)!\sqrt{\pi}}\,dt = \frac{2^{2k}k!x^{k+0.5}}{(k+0.5)(2k)!\sqrt{\pi}} = \frac{2^{2k+2}(k+1)!x^{k-0.5}}{(2k+2)!\sqrt{\pi}}.$$

In conclusion, we have that the pdf for χ^2_{2k+1} is

$$\begin{cases} 0 & \text{if } x \le 0; \\ \dfrac{2^{k-0.5}k!}{(2k)!\sqrt{\pi}}x^{k-0.5}e^{-x/2} & \text{if } x > 0. \end{cases}$$

Combining both cases, we have that the pdf of χ^2_n in general is

$$f_{Q_n}(x) = \begin{cases} 0 & \text{if } x \le 0; \\ \dfrac{1}{2^{n/2}(n/2-1)!} \cdot x^{n/2-1}e^{-x/2} & \text{if } n \text{ is even and } x > 0; \\ \dfrac{2^{n/2-1}(\frac{n-1}{2})!}{(n-1)!\sqrt{\pi}} \cdot x^{n/2-1}e^{-x/2} & \text{if } n \text{ is odd and } x > 0. \end{cases}$$

The Gamma Function

The gamma function $\Gamma : \mathbb{R} \to \mathbb{R}_{\geq 0}$ is defined as

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\ dt.$$

Some properties include

- $\Gamma(1) = 1,\ \Gamma(0.5) = \sqrt{\pi}$
- $\Gamma(x) = (x-1)\Gamma(x-1)$
- $\Gamma(x)\Gamma(x+0.5) = 2^{1-2x}\Gamma(2x)$

Using these facts, we can prove easily by induction that for $n = 0, 1, \cdots,$

- $\Gamma(n+1) = n!$
- $\Gamma(n+0.5) = \dfrac{\sqrt{\pi}(2n)!}{2^{2n}(n-1)!}$

Using the *Gamma function*, you can simplify the pdf as

$$f_{Q_n}(x) = \begin{cases} 0 & \text{if } x \leq 0; \\ \dfrac{1}{2^{n/2}\Gamma(n/2)} \cdot x^{n/2-1} e^{-x/2} & \text{if } x > 0. \end{cases}$$

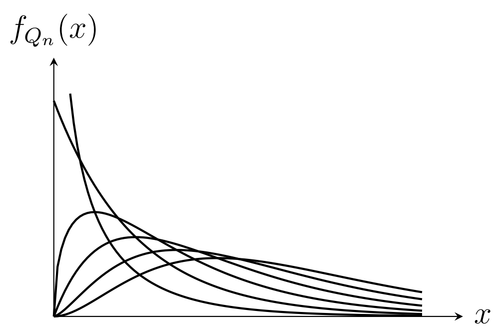

Figure 8: Chi-Squared Distributions $\chi^2_1, \chi^2_2, \chi^2_3, \chi^2_4, \chi^2_5, \chi^2_6$

6 Student's t-Distribution

We answer the question raised at the end of section 4: Let $X_1, \cdots, X_n$ be n independent observations of $\mathcal{N}(0,1)$. How do we find an interval, in terms of $\overline{X}$, such that μ has 0.9 probability to lie in it? Previously, we considered

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$$
$$\Rightarrow \mathcal{P}\left(\overline{X} - z_{0.95}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{0.95}\frac{\sigma}{\sqrt{n}}\right) = 0.9$$

but we noticed that σ is unknown. Therefore, we will consider the distribution of

$$\frac{\overline{X}-\mu}{s/\sqrt{n}} \sim t_{n-1}(0,1)$$

instead, where s is the sample variance. This distribution will differ according to n and is known as the *Student's t-distribution*, or t-distribution in short, with $\nu = n-1$ degrees of freedom.

6.1 Deriving the pdf for the t-Distribution

Note that

$$\frac{\overline{X}-\mu}{s/\sqrt{n}} = \left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right) / \left(\frac{s}{\sigma}\right)$$

Recall from section 4 and 5 that

$$R = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$$
$$T = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{n-1}$$

We will accept the fact that $\overline{X}$ and s^2 are independent without proof (Informally, they measure completely different things: $\overline{X}$ measures size and s^2 measures spreadness, so one does not affect the other). Therefore U and V are independent. It remains to find the pdf of $W = \sqrt{n-1}\cdot R/\sqrt{T}$ where $R \sim \mathcal{N}(0,1)$ and $T \sim \chi^2_{n-1}$. Since the t-distribution is symmetric, let's just study $w > 0$ below. Denote $\nu = n-1$.

$$\begin{aligned}
f_W(w) &= \frac{d}{dw}\mathcal{P}(R \le w\sqrt{T/\nu}) \\
&= \frac{d}{dw}\int_0^\infty f_T(t)\int_{-\infty}^{w\sqrt{t/\nu}} f_R(r)\ dr\ dt \\
&= \int_0^\infty f_T(t)\sqrt{\frac{t}{\nu}}\cdot f_U\left(w\sqrt{\frac{t}{\nu}}\right)\ dt \\
&= \int_0^\infty \frac{t^{\nu/2-1}\exp(-t/2)}{2^{\nu/2}\Gamma\left(\frac{\nu}{2}\right)}\sqrt{\frac{t}{\nu}}\cdot\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{w^2}{2}\cdot\frac{t}{\nu}\right)\ dt \\
&= \frac{1}{2^{\nu/2}\Gamma\left(\frac{\nu}{2}\right)\sqrt{2\pi\nu}}\int_0^\infty t^{(\nu-1)/2}\exp\left[-\underbrace{\frac{1}{2}\left(1+\frac{w^2}{\nu}\right)}_{K}t\right]\ dt \\
&= \frac{1}{2^{\nu/2}\Gamma\left(\frac{\nu}{2}\right)\sqrt{2\pi\nu}\cdot K^{(\nu+1)/2}}\int_0^\infty (Kt)^{(\nu-1)/2}\exp\left(-Kt\right)\ d(Kt) \\
&= \frac{1}{2^{\nu/2}\Gamma\left(\frac{\nu}{2}\right)\sqrt{2\pi\nu}}\cdot K^{-\frac{\nu+1}{2}}\cdot\Gamma\left(\frac{\nu+1}{2}\right) \\
&= \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}}\left(1+\frac{w^2}{\nu}\right)^{-\frac{\nu+1}{2}}
\end{aligned}$$

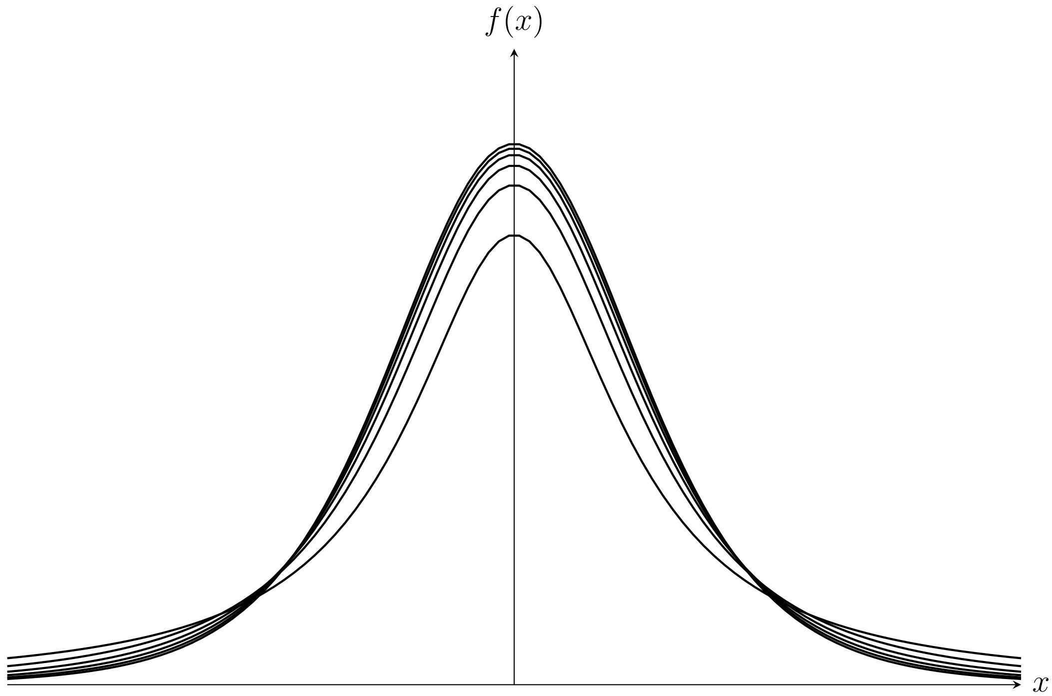

Figure 9: t-Distributions $t_1, t_2, t_3, t_4, t_5, t_6$

In Figure 9, the one with the lower y-intercept has the lower degree of freedom.

References

[1] Law of the Unconscious Statistician
https://en.wikipedia.org/wiki/Law_of_the_unconscious_statistician

[2] Normal Distribution
https://en.wikipedia.org/wiki/Normal_distribution

[3] Chi-Squared Distribution
https://en.wikipedia.org/wiki/Chi-squared_distribution

[4] Gamma Function
https://en.wikipedia.org/wiki/Gamma_function

[5] T-Distribution
https://en.wikipedia.org/wiki/Student%27s_t-distribution