Concepts

listenzcc

January 13, 2021

Abstract

Useful concepts of probability and statistics.

Contents

1 Laws

1.1 Law of total probability

It is common practice to compute the sum of total probability of all available options.

Thinking **forwardly**. Which means starting from the **reason** to the **result**.

Theorem 1.1. *Law of total probability*

For random variables A and B, we have

$$P(A) = \sum P(A|B_i) \cdot P(B_i), \forall B_i \in B$$

It is automatically accepted that all the B_is are all separable, and mutually exclusive with each other, which means

$$P(B_i, B_j) = P(B_i) \cdot P(B_j), i \neq j$$
$$P(B_i, B_i) = P(B_i)$$

It is a prior rule to be accepted, and we will accept it if not specified.

Thinking **backwardly**. If we have already known that A only has one option (noted as a), which is also inevitable ($P(A = a) = 1$).

Proposition 1.1. *Sum of probability of every options is ONE*

Proof. We have $P(a) = 1$ and $P(a|B_i) = 1, \forall B_i \in B$. Thus,

$$1 = \sum P(B_i), \forall B_i \in B$$

Since a can be something that naturally happens regardless of the choice of B, the proposition may not affected by the choice of a. Thus, the total probability of all options of B is 1. □

2 Definitions

2.1 Definition of e

The constant e is defined as the infinity integral of

$$e = \lim_{x \to \infty} (1 + \frac{1}{x})^x \tag{1}$$

the very existence of e is based on the fact that integrand function is monotone bounded.

Proposition 2.1. *The function $f(x)$ is **monotone bounded***

$$f(x) = (1 + \frac{1}{x})^x, x \in \mathcal{R}$$

Proof. Use its discrete sequence version

$$g_n = (1 + \frac{1}{n})^n, n \in \mathcal{N}$$

by expanding it, we have

$$g_n = \sum_{i=0}^{n} (n, i) \cdot \frac{1}{n^i}$$
$$g_{n+1} = \sum_{i=0}^{n+1} (n+1, i) \cdot \frac{1}{(n+1)^i}$$
$$g_{n+1} = \sum_{i=0}^{n} \frac{(n+1, i)}{n+1} \cdot \frac{1}{n^i} + c$$

where $c > 0$. For every i, we have $\frac{(n+1,i)}{n+1} > (n, i)$. Thus, the g_n is **monotone increasing**.

To find a valid upper bound, we enlarge the g_n.

$$g_n < \sum_{i=0}^{n} \frac{1}{i!}$$

the enlarge is based on the idea of $(n, i) < \frac{n^i}{i!}$. Further,

$$g_n < c + \sum_{i=4}^{n} \frac{1}{2^i}$$

where constant $c > 0$. It uses the idea of $2^i < i!, i \geq 4$. Use the sum of geometric series, we can say the g_n is **upper bounded**.

Back to $f(x)$, use **Squeeze Theorem**

$$(1+\frac{1}{n+1})^n < f(x) < (1+\frac{1}{n})^{n+1}$$

where $n = \lfloor x \rfloor$. On the left hand, it equals to $g_{n+1}/(1+\frac{1}{n+1})$; On the right hand, it equals to $g_n * (1+\frac{1}{n})$. Based on the analysis above, they are of the same value which is defined as e.

Hence proved.

□

2.2 Gaussian integral

The gaussian integral is the improper integral defined as

$$I = \int_{-\infty}^{\infty} e^{-x^2} dx \tag{2}$$

and the value equals to $\sqrt{\pi}$.

Lemma 2.1. *The gaussian integral equals to* $\sqrt{\pi}$

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$$

Proof. Use the square-shaped integral

$$I^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-x^2} e^{-y^2} dxdy$$

Transform into polar coordinates

$$I^2 = \int_0^{2\pi} \int_0^{\infty} e^{-r^2} rdrd\theta$$

Evaluate the integral of r

$$\int_0^{\infty} e^{-r^2} rdr = \frac{1}{2} \int_0^{\infty} e^{-u} du = \frac{1}{2}$$

Thus

$$I^2 = \int_0^{2\pi} \frac{1}{2} d\theta = \pi$$

Hence proved.

□

2.3 Gamma function

The Gamma function is defined as

$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt, Re(z) > 0 \tag{3}$$

The Gamma function has useful properties

$$\begin{aligned} \Gamma(1) &= 1 \\ \Gamma(z+1) &= z \cdot \Gamma(z) \\ \Gamma(z) &= 2 \int_0^{\infty} t^{2z-1} e^{-t^2} dt \end{aligned}$$

Based on the gaussian integral in (2), we have

$$\Gamma(\frac{1}{2}) = \sqrt{\pi}$$

There are countless useful properties of Gamma function. The followings are from the website [1].

2.3.1 Euler's reflection formula

The Gamma function satisfies the reflection formula

$$\Gamma(x)\Gamma(1-x) = \frac{\pi}{sin(x\pi)} \tag{4}$$

The proof can be found in [2]. The formula can verify the value of $\Gamma(\frac{1}{2})$.

2.3.2 Euler's product representation

The Gamma function can be expressed as an infinite product

$$\Gamma(x) = \frac{1}{x}\prod_{n=1}^{\infty}(1+\frac{1}{n})^x(1+\frac{x}{n})^{-1} \tag{5}$$

2.3.3 Weierstrass's product representation

The Gamma function can be expressed as another infinite product

$$\Gamma(x) = \frac{exp(-\gamma x)}{x}\prod_{n=1}^{\infty}(1+\frac{x}{n})^{-1}exp(\frac{x}{n}) \tag{6}$$

where γ denotes the Euler-Mascheroni constant

$$\gamma = \lim_{n\to\infty}\sum_{k=1}^{n}\frac{1}{k} - log(n) \tag{7}$$

This definition implies the reflection formula along with the Weierstrass product of sine and is equivalent with Euler's representation when the definition of the Euler-Mascheroni constant is substituted and rearranged.

2.3.4 Riemann's reflection formula

The Gamma function plays a key role in the analytic continuation of the Riemann zeta function to the complex plane

$$\zeta(s) = 2^s\pi^{s-1}sin(\frac{\pi s}{2})\Gamma(1-s)\zeta(1-s) \tag{8}$$

This example shows how closely related the Gamma function is to other functions. Another similar relation with the zeta function is

$$\zeta(z)\Gamma(z) = \int_0^{\infty}\frac{t^{z-1}}{e^t-1}dt$$

[1] `https://math.wikia.org/wiki/Gamma_function`
[2] `https://brilliant.org/discussions/thread/proof-of-euler-reflection-formula/`

2.3.5 Other properties

The relationship between Beta function

$$\int_0^1 t^{n-1}(1-t)^{m-1}dt = \frac{\Gamma(n)\Gamma(m)}{\Gamma(n+m)}$$

Use $t = sin^2\theta$ for convenience, we have

$$\int_0^{\frac{\pi}{2}} cos^{2n-1}(\theta)sin^{2m-1}(\theta)d\theta = \frac{\Gamma(n)\Gamma(m)}{2\Gamma(n+m)}$$

When $n \in \mathcal{N}$, we have

$$\Gamma(n) = (n-1)!$$
$$\Gamma(\frac{n}{2}) = \sqrt{\pi}\frac{(n-2)!!}{2^{(n-1)/2}}$$

where !! denotes the operator of double factorial [3].

2.4 Sterling's formula

In short version, the sterling's formula gives an expression of $n!$

$$n! = n^n e^{-n}\sqrt{2\pi n}[1+\mathcal{O}(\frac{1}{n})] \tag{9}$$

Proof. Proof the Sterling's formula

$$n! = n^n e^{-n}\sqrt{2\pi n}[1+\mathcal{O}(\frac{1}{n})]$$

Formulate new formula based on Gamma function

$$f(x) = \frac{\Gamma(x)e^x\sqrt{x}}{x^x} = 2\int_0^\infty e^{x-u^2}(\frac{u}{\sqrt{x}})^{2x-1}du$$

Changing variables $u = \sqrt{x} + v$

$$f(x) = 2\int_{-\sqrt{x}}^\infty \phi_x(v)e^{-v^2}dv$$

where $\phi_x(v) = e^{-2vx^{1/2}}(1+\frac{v}{\sqrt{x}})^{2x-1}, \forall v \geq -\sqrt{x}$. Additionally, we define $\phi_x(v) = 0, \forall v < -\sqrt{x}$, one gets

$$f(x) = 2\int_{-\infty}^\infty \phi_x(v)e^{-v^2}dv$$

Firstly, using the series expansion

$$log(1+\frac{v}{\sqrt{x}}) = \frac{v}{\sqrt{x}} - \frac{v^2}{2x} + \cdots$$

thus, for a fixed v

$$log\phi_x(v) = -v^2 + \mathcal{O}(\frac{1}{\sqrt{x}}), x \to \infty$$

[3] `https://math.wikia.org/wiki/Double_factorial`

hence

$$\phi_x(v) = e^{-v^2+\mathcal{O}(\frac{1}{\sqrt{x}})}$$
$$\lim_{x\to\infty} \phi_x(v) = e^{-v^2}$$

Use Lebesgue dominated convergence theorem we conclude

$$lim_{x\to\infty}\frac{\Gamma(x)e^x\sqrt{x}}{x^x} = 2\int_{-\infty}^{\infty} e^{-2v^2}dv = \sqrt{2\pi}$$

additionally, the approximation can be expressed as

$$\frac{\Gamma(x)e^x\sqrt{x}}{x^x} = \sqrt{2\pi}\cdot e^{\mathcal{O}(\frac{1}{\sqrt{x}})}$$

Use the condition of $n = x, n \in \mathcal{N}$, and $\Gamma(n) = (n-1)!$, we get

$$n! = n^n e^{-n}\sqrt{2\pi n}[1+\mathcal{O}(\frac{1}{n})]$$

Hence proved.

□

3 Basic concepts

3.1 Expectation and Variation

For a random variable, it follows certain distribution

$$x \approx p(x), \forall 0 < p(x) < 1, \int_x p(x)dx = 1$$

its expectation and variation are computed as

$$\mathcal{E}(x) = \int_x x\cdot p(x)dx$$
$$\mathcal{V}(x) = \int_x (x-\mathcal{E}(x))^2\cdot p(x)dx$$

additionally, the variation can be computed as

$$\mathcal{V}(x) = \mathcal{E}(x^2) - \mathcal{E}^2(x)$$