Chapter 3

Usual Probability Lows

A probability distribution is a mathematical function that theoretically describes a random experiment. Probability distributions are essential in biology for quantifying and predicting variability in various biological processes. They allow biologists to analyze data, formulate hypotheses, and make decisions. These mathematical laws thus contribute to a better understanding of random phenomena in the living world.

Contents

3.1 Random Variable

Definition 3.1.1. *A random variable X is an application from the sample space Ω to $\mathbb{R}$:*

$$\begin{aligned} X: \ \Omega &\to \mathbb{R} \\ w &\mapsto X(w) \end{aligned}$$

There are two types of random variables:

3.1.1 Discrete Random Variable

Definition 3.1.2. *A random variable X is discrete if it can take on a finite number of distinct values.*

Probability low

Definition 3.1.3. *Let X be a random variable on Ω, and $X(\Omega) = \{x_1, x_2, ..., x_n\}$. The probability low of X is given by:*

x_i	x_1	x_2	$\dots$	x_n
$p(X = x_i)$	$p_1 = p(X = x_1)$	$p_2 = p(X = x_2)$	...	$p_n = p(X = x_n)$

Distribution Function

:

Definition 3.1.4. *The distribution function of D.R X is:*

$$F_X(x) = P(X \le x) = \sum_{x_i \le x} p(X = x_i)$$

$$F_X(x) = \begin{cases} 0 & \text{if } x < x_1 \\ p_1 & \text{if } x_1 \le x < x_2 \\ p_1 + p_2 & \text{if } x_2 \le x < x_3 \\ \vdots & \\ 1 & \text{if } x \ge x_n \end{cases}$$

Expected Value

Definition 3.1.5. *The expected value of the random variable X is:*

$$E(X) = \sum_{i=1}^{n} x_i p(X = x_i)$$

Variance and Standard Deviation:

The variance of the random variable X is:

Definition 3.1.6.

$$V(X) = E\left((X - E(X))^2\right) = \sum_{i=1}^{n} (x_i - E(X))^2 p(X = x_i)$$

Alternatively,

$$V(X) = E(X^2) - (E(X))^2 = \sum_{i=1}^{n} x_i^2 p(X = x_i) - \left(\sum_{i=1}^{n} x_i p(X = x_i)\right)^2$$

The standard deviation δ_X is:

Definition 3.1.7.

$$\delta_X = \sqrt{V(X)}$$

Example 3.1.1. *Consider the random experiment "rolling a six-sided die and observing the result." The game is as follows:*

- *If the result is even, you win 2 DA.*
- *If the result is 1, you win 3 DA.*
- *If the result is 3 or 5, you lose 4 DA.*

Define a random variable X which gives the gain in this game.

1. *The sample space is* $\Omega = \{1, 2, 3, 4, 5, 6\}$
2. $X(\Omega) = \{-4, 2, 3\}$

3. *The probability distribution of X is:*

x_i	-4	2	3
$p(X = x_i)$	$\frac{2}{6}$	$\frac{3}{6}$	$\frac{1}{6}$

4. *The DF of X is:*

$$F_X(x) = \begin{cases} 0 & \text{if } x < -4 \\ \frac{2}{6} & \text{if } -4 \leq x < 2 \\ \frac{5}{6} & \text{if } 2 \leq x < 3 \\ 1 & \text{if } x \geq 3 \end{cases}$$

5. *The expected value $E(X)$ is:*

$$E(X) = -4 \times \frac{2}{6} + 2 \times \frac{3}{6} + 3 \times \frac{1}{6} = \frac{1}{6}$$

6. *The variance $V(X)$ is:*

$$V(X) = 8.8 \quad \text{and} \quad \delta_X = \sqrt{V(X)} = 2.97$$

3.1.2 Continuous Random Variable

Definition 3: A random variable X is said to be continuous if it can take any value within an interval.

Probability Density Function (PDF)

Let X be a continuous random variable. We say that $f : \mathbb{R} \to \mathbb{R}^+$ is the probability density function (PDF) of X if:

- $f(x) \geq 0$ for all $x \in \mathbb{R}$,
- f is continuous on $\mathbb{R}$,
- $\int_{-\infty}^{+\infty} f(x)\, dx = 1$.

Distribution Function

The distribution function of a continuous random variable X is defined by:

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

Proposition 3.1.1. • *$F_X(x)$ is positive,*

• $\lim_{x\to-\infty} F_X(x) = 0$ *and* $\lim_{x\to+\infty} F_X(x) = 1$.

Remark 3.1.1. • *$P(X = x) = 0$,*

• *$P(X \leq x) = P(X < x)$,*

• *The probability that $X \in [a, b]$ is given by:*

$$\begin{aligned} P(a \leq X \leq b) &= P(a < X \leq b) \\ &= P(a \leq X < b) \\ &= P(a < X < b) \\ &= \int_a^b f(t)\,dt \\ &= F_X(b) - F_X(a) \end{aligned}$$

Expected Value (Mathematical Expectation)

The expected value of the continuous random variable X is given by:

$$E(X) = \int_{-\infty}^{+\infty} x f(x)\,dx$$

Variance and Standard Deviation

The variance of the continuous random variable X is given by:

$$V(X) = \int_{-\infty}^{+\infty} (X - E(X))^2 f(x)\,dx$$

or,

$$V(X) = E(X^2) - (E(X))^2 = \int_{-\infty}^{+\infty} x^2 f(x)\,dx - \left(\int_{-\infty}^{+\infty} x f(x)\,dx\right)^2$$

3.2 Usual Probability lows

3.2.1 Discrete Lows

Bernoulli Low

- Any random experiment with two possible outcomes: success and failure, is called a **Bernoulli experiment**.
- Probability low: The random variable $X = 1$ in case of success with probability p, and $X = 0$ in case of failure with probability $q = 1 - p$.

The probability distribution is given by:

$$p(X = x) = \begin{cases} p & \text{if } x = 1 \\ q & \text{if } x = 0 \end{cases}$$

Notation: $X \sim B(p)$

$$E(X) = p$$

$$V(X) = pq$$

$$\delta_X = \sqrt{V(X)} = \sqrt{pq}$$

Examples 3.2.1. *We flip a fair coin. If we get heads, it's a success.* $\{\Omega\}$ *={heads, tails},**{A}**= {heads}*

$\bar{A}$ *= {tails}*

Probability distribution:

The probability of success $p = p(A) = \frac{1}{2}$

The probability of failure $q = p(\bar{A}) = \frac{1}{2}$

$$p(X = x) = \begin{cases} \frac{1}{2} & \text{if } x = 1 \\ \frac{1}{2} & \text{if } x = 0 \end{cases}$$

Calculating the values:

$$E(X) = p = \frac{1}{2}$$

$$V(X) = pq = \frac{1}{4}$$

$$\delta_X = \sqrt{V(X)} = \frac{1}{2}$$

Binomial low

- The Binomial low with parameters n and p models the number of successes obtained from n repetitions of identical and independent Bernoulli experiments.
- Probability low:

$$p(X = k) = C_n^k p^k q^{n-k}, \quad k = 1, 2, \ldots, n$$

where $C_n^k = \frac{n!}{k!(n-k)!}$.

Notation: $X \sim B(n, p)$

$$E(X) = np, \quad V(X) = npq, \quad \delta_X = \sqrt{V(X)}$$

Examples 3.2.2. *We roll a fair die 5 times and are interested in the outcome "getting the number 2."*

- *1. What is the probability of getting exactly two 2's?*
- *2. What is the probability of getting at least three 2's?*
- *3. Determine $E(X), V(X), \delta_X$.*

Solution: The random variable X: "getting the number 2," $X \sim B(5, p)$

$$\Omega = \{1, 2, 3, 4, 5, 6\}, \quad A = \{2\}, \quad \bar{A} = \{1, 3, 4, 5, 6\}$$

Probability of success: $p = p(A) = \frac{1}{6}$
Probability of failure: $q = p(\bar{A}) = \frac{5}{6}$

$$p(X = 2) = C_5^2 \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^3 = 0.16$$

$$p(X \geq 3) = 1 - p(X < 3) = 1 - (p(X = 0) + p(X = 1) + p(X = 2)) = 0.036$$

$$E(X) = np = \frac{5}{6}, \quad V(X) = npq = \frac{25}{36}, \quad \delta_X = \frac{5}{6}$$

Poisson low

- The Poisson low models rare events, i.e., events with a low probability of occurrence.
- Probability distribution:

$$p(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad \lambda > 0$$

Notation: $X \sim P(\lambda)$

$$E(X) = \lambda, \quad V(X) = \lambda, \quad \delta_X = \sqrt{\lambda}$$

Examples 3.2.3. *A call center receives an average of 5 calls in minute. What is the probability that the center receives exactly two calls? The random variable X: the number of calls,* $X \sim P(5)$

$$p(X = 2) = e^{-5} \frac{5^2}{2!} = 0.084$$

3.2.2 Continuous low

Normal low

- A random variable X follows a normal low or Gauss-Laplace low with parameters m and δ if:

$$f(x) = \frac{1}{\sqrt{2\pi}\delta} e^{-\frac{1}{2}\left(\frac{x-m}{\delta}\right)^2}$$

Notation: $X \sim N(m, \delta)$

Log-Normal low

- A random variable X follows a log-normal low if $\ln(X)$ follows a normal distribution $N(m, \delta)$:

$$f(x) = \frac{1}{x\sqrt{2\pi}\delta} e^{-\frac{1}{2}\left(\frac{\ln(x)-m}{\delta}\right)^2}, \quad x > 0$$

Notation: $X \sim LN(m, \delta)$

Standard Normal low

- The standard normal low is $N(0,1)$.

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

- If $X \sim N(m, \delta)$, then $Y = \frac{X-m}{\delta} \sim N(0,1)$.

Examples 3.2.4. *Table and calculate the probability*

- $X \sim N(0,1)$, *calculate* $p(X \leq 1.25)$ *and* $p(X \leq 0.67)$
  For $x = 1.25 \rightarrow$ *row*=*1.2 and column*= $0.05 \rightarrow p(X \leq 1.25) = F(1.25)) = 0.8944$ *For* $x = 0.67 \rightarrow$ *row*=0.6 *and column*= $0.07 \rightarrow p(X \leq 0.67) = F(0.67) = 0.7486$

- $X \sim N(0,1)$, *calculate* $p(X \geq 0.87)$ *and* $p(X \leq 0.74)$

$$p(X \geq 0.87) = 1 - p(X \leq 0.87) = 1 - F(0.87) = 1 - 0.8078 = 0.1922$$

$$p(X \geq 0.74) = 1 - p(X \leq 0.74) = 1 - F(0.74) = 1 - 0.7704 = 0.2296$$

- $p(X \leq -x) = 1 - p(X \leq x)$, *calculate* $p(X \leq -1.87)$

$$p(X \leq -1.87) = 1 - p(X \leq 1.87) = 1 - F(1.87) = 1 - 0.9693 = 0.0407$$

- $p(X \geq -x) = p(X \leq x)$, *calculate* $p(X \geq -0.74)$

$$p(X \geq -0.74) = p(X \leq 0.74) = 0.7704$$

- $p(a \leq X \leq b) = F(b) - F(a)$, *calculate* $p(1.15 \leq X \leq 2.25)$ *and* $p(-0.58 \leq X \leq -0.14)$

$$p(1.15 \leq X \leq 2.25) = F(2.25) - F(1.15) = 0.9878 - 0.8749 = 0.1129$$

$$p(-0.58 \leq X \leq -0.14) = F(0.58) - F(0.14) = 0.7190 - 0.5557 = 0.1633$$

- $p(-a \leq X \leq b) = F(b) + F(a) - 1$, *calculate* $p(-1.14 \leq X \leq 2.58)$

$$p(-1.14 \leq X \leq 2.58) = F(2.58) + F(1.14) - 1 = 0.9951 + 0.8729 - 1 = 0.8679$$

- *$p(-a \leq X \leq a) = 2F(a) - 1$, calculate $p(-1 \leq X \leq 1)$ and $p(-1.96 \leq X \leq 1.96)$*

$$p(-1 \leq X \leq 1) = 2F(1) - 1 = 2(0.8413) - 1 = 0.6827$$

$$p(-1.96 \leq X \leq 1.96) = 2F(1.96) - 1 = 2(0.976) - 1 = 0.95$$

Chi-squared low

- If $X_1, X_2, \ldots, X_n$ are n independent random variables, each following a normal low $N(0,1)$, then $Y = X_1^2 + X_2^2 + \cdots + X_n^2$ follows a Chi-squared low with n degrees of freedom.

- Probability low:

$$f(y) = \frac{y^{\frac{n}{2}-1} e^{-y/2}}{2^{n/2}\Gamma\left(\frac{n}{2}\right)}, \text{ with } \Gamma(n) = \int_0^{+\infty} e^x x^{n-1}.$$

Notation: $Y \sim \chi_n^2$

$$E(Y) = n, \quad V(Y) = 2n$$

Example 3.2.1. • *$Y \sim \chi_{18}^2$, calculate $p(Y \leq 28.87)$:*

$p(Y \leq 28.87) = 0.95$

• *$Y \sim \chi_{10}^2$, calculate $p(Y \geq 23.209)$*

$p(Y \geq 23.209) = 1 - p(Y \leq 23.209) = 1 - 0.99 = 0.01$

• *Find y such that $p(Y \leq y) = 0.975$ and $Y \sim \chi_{22}^2$*

$y = 36.781$

Student's t low

If X and Y Two independent random variables, such that: $X \sim N(0,1)$ and $Y \sim \chi_n^2$, then $T = \frac{X}{\sqrt{Y/n}}$ follows a student low with n degrees of freedom. Probability low:

$$f(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi\Gamma\left(\frac{n}{2}\right)\left(1 + \frac{t^2}{n}\right)^{\frac{n+1}{2}}}}$$

Notation:

$T \sim t_n$

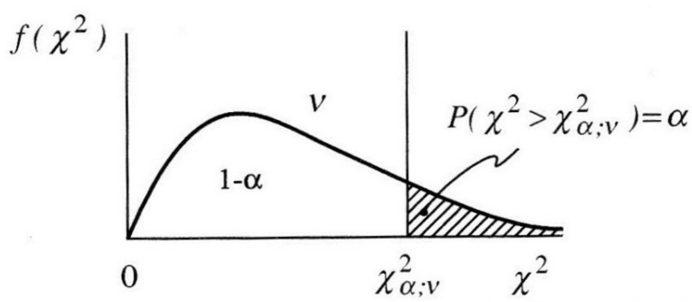
$$E(T) = 0, n > 1.$$

$$V(T) = \frac{n}{n-2} \quad \text{for } n > 2.$$

Standard Normal Distribution Table (Z-Table)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5	0.504	0.508	0.512	0.516	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.591	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.648	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.67	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.695	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.719	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.758	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.791	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.834	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.877	0.879	0.881	0.883
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.898	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.937	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.975	0.9756	0.9761	0.9767

TABLE DE LA LOI KHI-DEUX (χ^2)



ν \ α	**0,995**	**0,975**	**0,95**	**0,9**	**0,5**	**0,1**	**0,05**	**0,025**	**0,01**	**0,005**	**0,001**
1	0,000	0,001	0,004	0,016	0,455	2,706	3,841	5,024	6,635	7,879	10,827
2	0,010	0,050	0,103	0,211	1,386	4,605	5,991	7,378	9,210	10,597	13,815
3	0,072	0,215	0,352	0,584	2,366	6,251	7,815	9,348	11,345	12,838	16,268
4	0,207	0,484	0,711	1,064	3,357	7,779	9,488	11,143	13,277	14,860	18,465
5	0,412	0,831	1,145	1,610	4,351	9,236	11,070	12,832	15,086	16,750	20,517
6	0,676	1,237	1,635	2,204	5,348	10,645	12,592	14,449	16,812	18,548	22,457
7	0,989	1,689	2,167	2,833	6,346	12,017	14,067	16,013	18,475	20,278	24,322
8	1,344	2,179	2,733	3,490	7,344	13,362	15,507	17,535	20,090	21,955	26,125
9	1,735	2,700	3,325	4,168	8,343	14,684	16,919	19,023	21,666	23,589	27,877
10	2,156	3,247	3,940	4,865	9,342	15,987	18,307	20,483	23,209	25,188	29,588
11	2,603	3,815	4,575	5,578	10,341	17,275	19,675	21,920	24,725	26,757	31,264
12	3,074	4,403	5,226	6,304	11,340	18,549	21,026	23,337	26,217	28,300	32,909
13	3,565	5,008	5,892	7,041	12,340	19,812	22,362	24,736	27,688	29,819	34,528
14	4,075	5,628	6,571	7,790	13,339	21,064	23,685	26,119	29,141	31,319	36,123
15	4,601	6,262	7,261	8,547	14,339	22,307	24,996	27,488	30,578	32,801	37,697
16	5,142	6,907	7,962	9,312	15,338	23,542	26,296	28,845	32,000	34,267	39,252
17	5,697	7,564	8,672	10,085	16,338	24,769	27,587	30,191	33,409	35,718	40,790
18	6,265	8,230	9,390	10,865	17,338	25,989	28,869	31,526	34,805	37,156	42,312
19	6,844	8,906	10,117	11,651	18,338	27,204	30,144	32,852	36,191	38,582	43,820
20	7,434	9,590	10,851	12,443	19,337	28,412	31,410	34,170	37,566	39,997	45,315
21	8,034	10,282	11,591	13,240	20,337	29,615	32,671	35,479	38,932	41,401	46,797
22	8,643	10,982	12,338	14,041	21,337	30,813	30,924	36,781	40,289	42,796	48,268
23	9,260	11,688	13,091	14,848	22,337	32,007	35,172	38,076	41,638	44,181	49,728
24	9,886	12,401	13,848	15,659	23,337	33,196	36,415	39,364	42,980	45,558	51,179
25	10,520	13,119	14,611	16,473	24,337	34,382	37,652	40,646	44,314	46,928	52,620
26	11,160	13,843	15,379	17,292	25,336	35,563	38,885	41,923	45,642	48,290	54,052
27	11,808	14,573	16,151	18,114	26,336	36,741	40,113	43,195	46,963	49,645	55,476
28	12,461	15,307	16,928	18,939	27,336	37,916	41,337	44,461	48,278	50,994	56,893
29	13,121	16,047	17,708	19,768	28,336	39,087	42,557	45,722	49,588	52,335	58,302
30	13,787	16,790	18,493	20,599	29,336	40,256	43,773	46,979	50,892	53,672	59,703
40	20,706	24,433	26,051	29,051	39,335	51,805	55,758	59,342	63,691	66,766	73,403
60	35,534	40,481	43,188	46,459	59,335	74,397	79,082	83,298	88,379	91,952	99,608
80	51,171	57,153	60,391	64,278	79,334	96,578	101,879	106,629	112,329	116,321	124,839
100	67,327	74,221	77,929	82,358	99,334	118,498	124,342	129,561	135,807	140,170	149,449

Critical Value Tables: Student's t-distribution and Chi-squared Distribution

Student's t-distribution Critical Values

df	90%	95%	97.5%	99%	99.5%
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.92	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
10	1.372	1.812	2.228	2.764	3.169
20	1.325	1.725	2.086	2.528	2.845
30	1.31	1.697	2.042	2.457	2.75
50	1.299	1.676	2.009	2.403	2.678
100	1.29	1.66	1.984	2.364	2.626